Dynamics of Nuclear Matrix Attachment Regions During 5th Instar Posterior Silk Gland Development in *Bombyx Mori*

Alekhya Rani Chunduri
GITAM Institute of Science: Gandhi Institute of Technology and Management Institute of Science
Resma Rajan
GITAM Institute of Science: Gandhi Institute of Technology and Management Institute of Science
Anugata Lima
GITAM Institute of Science: Gandhi Institute of Technology and Management Institute of Science
Senthilkumar Ramamoorthy
University of Freiburg Faculty of Medicine
Anitha Mamillapalli (amamilla@gitam.edu)
GITAM Institute of Science: Gandhi Institute of Technology and Management Institute of Science
https://orcid.org/0000-0001-8338-6970

Research Article

Keywords:

Posted Date: February 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-763068/v1

Version of Record: A version of this preprint was published at BMC Genomics on March 31st, 2022. See the published version at https://doi.org/10.1186/s12864-022-08446-3.

Abstract

Chromatin architecture is critical for gene expression during development. Matrix attachment regions (MARs) control and regulate chromatin dynamics. The position of MARs in the genome determines the expression of genes in the organism. In this study, we set out to elucidate how MARs temporally regulate the expression of the fibroin heavy chain (FIBH) gene during development. We addressed this by identifying MARs and studying their distribution and differentiation, in the posterior silk glands of Bombyx mori during 5th instar development. Of the MARs identified on three different days, 7.15% MARs were common to all three, whereas, 1.41%, 19.27% and 52.47% MARs were unique to day 1, day 5, and day 7, respectively highlighting the dynamic nature of the matrix associated DNA. Further, significant changes in the MARs in the vicinity of the FIBH gene were found during different days of 5th instar development which implied their role in the regulation and expression of the FIBH gene.

Introduction

The role of the nuclear matrix in the spatial and temporal organization of the genome makes it especially significant in DNA replication, gene expression, and regulation (1–3).The nuclear matrix facilitates the arrangement of chromatin into loop structures by anchoring it to the nucleus (4). These DNA sequences anchor the chromatin loops and are known as scaffold or matrix attached regions (MARs) (5). MARs are often correlated with chromatin features like, AT richness, the origin of replication, etc., and are also hotspots for transposable element insertions (6). Many studies have been carried out determining their role in the maintenance of chromatin architecture, stabilization and regulation of gene expression (7–9) and their influence on gene silencing mechanisms (10–12).

The number of whole genome sequencing of MARs has been carried out by very few studies. A genome-wide analysis of MARs belonging to the euchromatic part of the genome was performed in *Drosophila melanogaster* and identified 7353 MARs (13). The MAR DNA isolated from embryos of *D. melanogaster* identified the LTR transposable element '*roo*' which showed high similarity with the transposable element *gypsy* (14). In *Arabidopsis*, the genome-wide in-silico MAR identification showed a significant correlation between the presence of intragenic S/MARs and transcriptional down-regulation (15). Several studies were also carried out to investigate MARs in vertebrates such as humans, mice, chickens and Chinese hamsters (16–19). The influence of MARs in gene expression and their successful integration and regulation of transgenic genes were also determined (19–22).

The silkworm, *Bombyx mori (B. mori)* is a widely commercialized insect that is agriculturally and economically relevant. The silk produced by this insect has extensive industrial and medical applications. (23). The silk gland is a complex organ that is known for its tissue-specific gene expression and regulation during development (24). The silk glands were found to show an increase in the genomic DNA content by 200-300 times during larval development due to endomitosis (25). The silk glands exhibit a significant and sudden growth during larval development. It was determined that the silkworm, *B. mori* has significantly more cells in its PSGs that result in higher fibroin production when compared to the wild type silkworm due to the up-regulation of genes responsible for division and growth of cells which is attributed

to its history of domestication (26).The compartmentalisation of the silk gland into different regions i.e., ASGs, MGSs and PSGs is attributed to the Hox genes and their role in the regulation of expression in silk related genes (27). The posterior silk gland (PSG) synthesizes the most important protein involved in silk synthesis, Fib-H (28). The spatial organization of chromatin in the multi-nucleated cells of silk glands is facilitated by the existence of the nuclear matrix (29).

The FIBH gene which produces fibroin heavy chain protein is regulated by multiple factors. DNA sequences upstream of the gene were shown to possess transcription modulation signals (13, 30). The 62 kb upstream region of FIBH, Bmmar1, a mariner like element previously identified in *Drosophila* and other insects, as well as Bm1 (a SINE element belonging to tRNA superfamily) and L1Bm (a LINE element within the intronic region of FIBH) were found in the ORF regions (24, 31–34). BmMar1 was found to specifically attach to the nuclear matrix in silk glands and is predicted to be involved in the regulation of its expression, The core region of this MAR consisted of two degenerate sequences derived from LIBM and Bm1 (30, 35). The genome-wide MARs of silk glands and their distribution and differentiation in occurrence and positions during development has not been elucidated.

The current study investigates the genome-wide temporal regulation of MARs in PSGs of *B. mori* on different days of 5^{th} instar larval development. The presence of MARs is explored in the context of their possible role in the expression of FIBH. Chromosome-wide distribution and density of MARs were dynamic during the 5^{th} instar development. MARs varied in number and position flanking FIBH. Many MAR regions identified in the flanking regions of FIBH were found to be developmentally regulated.

Materials And Methods

Isolation and Quantification of nuclear and MAR DNA

Bivoltine, double hybrid, CSR2 X CSR4, *B. mori* 4^{th} moult larvae were collected from the Department of Sericulture, Srikakulam, Government of Andhra Pradesh. 100 Larvae were fed fresh V1 mulberry leaves from day 1 to day 7 (wandering stage) of the 5^{th} instar stage. PSGs were dissected from 5^{th} instar larvae on day 1, day 5, and day 7. PSGs were pooled (day1 (n=30), day 5 (n=15), and day 7(n=10)) from a single rearing, and 1 g tissue from each day was homogenized in nuclear isolation buffer (5 times the volume of the weight of the tissue) and processed for nuclei and nuclear matrix isolation by following the standard protocol for isolation through nuclease digestion and salt extraction (13) **(Supplementary Figure 1)**

The nuclear and MAR pellets were further processed for DNA isolation with HiPurA™ Genomic DNA Purification Kit. Quantification of nuclear and nuclear matrix PSG DNA on day 1, day 5, and day 7 was carried out by UV spectrophotometry **(Figure 1A).** Agarose gel electrophoresis **(Figure 1B)** and Tapestation profiling using Agilent 2200 TapeStation, were performed to check the enrichment of MARs and their fragment size distribution in the libraries respectively **(Supplementary Figure 2).** TheMAR DNA, consists of small but varying size fragments and therefore appears as a smear instead of a clear band on the agarose gel **(Figure 1B).**

Transmission Electron microscopy (TEM) analysis

Transmission electron microscopy was performed to confirm the isolation of the nuclear matrix by nuclease digestion and salt extraction. Sections of day 5 PSGs were made by fixing samples in 2.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.2) for 24 h at 4^{o} C, and washing with PBS for 4 times (each wash for one hour). Post fixation was carried out in aqueous osmium tetroxide for 3 h. Washes were performed 6 times (each for one minute) with deionized distilled water. The samples were dehydrated in a series of graded alcohols, infiltrated, and embedded in Araldite resin. Incubation was done at 80^{o} C for 72 h for complete polymerization. Ultra-thin (60 nm) sections were made with a glass knife on ultra-microtome (Leica Ultra cut UCT-GA-D/E-1/00), mounted on copper grids, and stained with saturated aqueous uranyl acetate (UA) and counterstained with Reynolds lead citrate (LC) (36). JEOL JEM-2100 Electron Microscope was used for TEM Imaging. TEM imaging was carried out at Ruska Labs, Acharya N. G. Ranga Agricultural University, Hyderabad, Telangana.

Library Preparation and Sequencing

MAR DNA sequencing libraries were prepared with Illumina-compatible NEXTflex ChIPSeq Library Prep Kit (Bio Scientific, Austin, TX, U.S.A.) at the Genotypic Technology Pvt. Ltd., Bangalore, India. DNA was sheared using a Covaris S220 system (Covaris, Woburn, Massachusetts, USA) and the fragmented DNA was purified using magnetic beads and subjected to end repair. Adapter ligation was carried out to the end-repaired DNA after 3' adenylation. The adapter-ligated DNA was purified with JetSeq magnetic beads and then amplified for 8 cycles (denaturation at 98^{o} C for 2 min, cycling [98^{o} C for 30 seconds, 65^{o} C for 30 seconds, and 72^{o} C for 1 min] and a final extension at 72^{o} C for 4 min). The final PCR product was purified with JetSeq beads (Bio, # 68031), followed by a library-quality control check. The sequencing library was initially quantified by Qubit (Thermo Fisher Scientific, MA, U.S.A.) **(Supplementary Table 1).**

MAR-Seq data analysis

The raw sequenced data obtained from the Illumina sequencing platform as single-end sequencing were labelled SG 1, SG 5, and SG 7 (day1, day 5, and day 7), respectively, and the quality control for each of these three datasets was carried out using FastQC tool v1.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The adapters were trimmed using cutadapt (Version 1.16) (37)

The MAR library reads of SG 1, SG 5, and SG 7 were aligned to the *B. mori* genome (Silk Base Version 2017.4.21) (38). Mapping was performed using a BOWTIE2 wrapper available on the galaxy platform (39). The wrapper internally uses bowtie2 (version 2.3.4.1) and samtools (version 1.9) (40, 41).

The mapped reads were further analysed for sequencing depth using the plotCoverage tool (Galaxy Version 3.3.200.0) which utilises deeptools (Version 3.3.2) and samtools (Version 1.9).

The MAR peaks were identified using MACS2 callpeak from the galaxy platform which internally uses MACS2 (Version 2.1.1.20160309) and r-base (Version 3.4). The identified MAR peaks were merged for the

overlapping intervals and bookended intervals into a single interval using MergeBED from bedtools (Version 2.29.2). The merged BED files were used for determining the distribution of MARs in different chromosomes and the size of the MARs across the three datasets (SG1, SG 5 and SG 7). The average loop length was determined by calculating the distance between MARs. Analysis of unique and common MAR regions in SG 1, SG 5, and SG 7 was performed with bedtools intersect command (Version 2.29.0).

MAR annotation and biological pathway analysis

The 'bedops/closest features' tool was used for annotating the MARs. The merged peaks of SG 1, SG 5, and SG 7 datasets in BED format and the *B. mori* reference genome (annotated genome version 2017.4.21 from Silkbase) were used for annotating the MAR peaks. The intronic, exonic, and non-genic regions along with the distances of MARs from the transcription start site (TSS) were determined. The MAR-associated genes were also identified in the analysis and were used to identify associated pathways utilizing the 'pathways' tool from the silkworm genome informatics database (42).

Analysis of MAR associated features in the datasets

The merged peaks were used as input and the various motifs found to be associated with MARs such as ORI Signals (ATTA, ATTTA, ATTTTA), topoisomerase II signals (RnYnnGYnGKTnYnY, GTnWAYATTnATnnR), TG Rich Signal (TGTTTTG, TGTTTTTTG, TTTTGGGG), AT-rich signal (WWWWWW), Curved DNA ($AAAAn_7AAAAn_7AAAA$, $TTTTn_7TTTTn_7TTTT$, TTTAAA) and Kinked DNA (TAn_3TGn_3CA, TAn_3CAn_3TG, TGn_3TAn_3CA, TGn_3CAn_3TA, CAn_3TAn_3TG, CAn_3TGn_3TA) were identified in the three developmental datasets with the help of Regulatory Sequence analysis tools (RSAT Metazoa) (43, 44). MAR Recognition Signatures (MRS) (bipartite sequence element that consists of two individual sequences of 8 (AATAAYAA) and 16 bp (AWWRTAANNWWGNNNC) within a 200 bp distance from each other) were identified in the datasets using EMBOSS marscan from galaxy server which internally uses emboss (Version 5.0.0) and Perl (Version 5.26).

Identification of Repeats

The STR repeats were identified for the enriched MAR peaks in the three datasets. The MAR peaks were converted to FASTA format and used for extracting Short Tandem Repeats (STRs) of di-, tri-, tetra-, penta- and hexanucleotides. MISA-web (Version 2.1, updated: 2020-08-25) tool was used for finding tandem repeats (45). The default parameters i.e., minimum number of repetitions set as 5 for their detection by the algorithm and maximum length of sequence between two SSRs to register as compound SSR set as 100 were used for this analysis. The analysis was carried out to identify the signals of the repeat sequences picked throughout the datasets and not in accordance with the number of times the repeat occurs in a specific location in the genome. The compound sequences for tetra, penta and hexa-nucleotides were also included as individual signals.

The FASTA sequences of transposable elements known to be associated with the nuclear matrix were extracted from the NCBI database and aligned against the merged BED files of the datasets SG 1, SG 5, and SG 7 using the BLAST2 tool to find their distribution in the developmental datasets. The hits with

>=90% percentage identity with >=100 bp sequence length were considered. Other transposable elements associated with the genome were identified using rmblast of RepeatMasker (http://www.repeatmasker.org) with the closest available repbase library which was of *D. melanogaster.*

Developmental analysis of FIBH and Br-c flanking MARs.

The MAR peaks were loaded into the integrated genome viewer (Version 2.6.3) and MARs distribution was visualized, among the developmental datasets: SG 1, SG 5, and SG 7, in the flanking regions of the genes FIBH and Br-c.

Statistical Analysis

Student's t-test: 2-sample assuming unequal variances, was used to determine the significance of variation in the nuclear and MAR DNA isolated from PSGs of 5^{th} instar larval stage on day 1, day 5 and day 7. $p \leq 0.5$ was considered as significant; $p > 0.5$ was considered as not significant. The experiment was conducted five times each for isolation and estimation of nuclear and MAR DNA respectively. Pooled samples of PSG tissues of day 1, day 5 and day 7 (1g tissue of PSGs for each day) were used in the experiments.

Results

Enrichment of nuclear matrix-associated DNA of silk glands during 5^{th} instar larval development

The association of DNA to the nuclear matrix is transient and is highly regulated during development to control gene expression. PSGs, where FIBH is specifically expressed, have been shown to have large variations in the expression of genes during different days of 5^{th} instar development (46).

Ultrastructural analysis of silk gland nuclear matrix reveals a fibro-granular network

The ultrastructure of the nuclear matrix reveals the internal architecture and confirms the depletion of chromatin during isolation. To check if nuclease digestion and high salt extraction revealed the inner nuclear skeleton, TEM imaging was performed with the day 5 PSGs of the 5^{th} instar larvae. The PSG nucleus showed three well-defined compartments: the lamina (NL), inter-chromatin network, and a densely stained nucleolus (Nu). Condensed chromatin (dCh) surrounding the nucleolus and lining the inner periphery of the nuclear lamina was observed. In addition, condensed (dCh) and open chromatin (lCh) was observed as dark and light areas throughout the nucleus **(Figure 1C).** The chromatin-depleted nuclear matrix of PSG showed a fibro-granular network spanning across the nucleus. A distinct decrease in the condensed chromatin regions was seen. Many domains retained their positions after the removal of the soluble proteins, and chromatin; and correspond to structures that can be observed by electron microscopy in unfractionated nuclei and nuclear matrix preparations; underscoring the relevance of electron microscopic studies (47, 48). The nuclear matrix preparation still retained a small portion of chromatin along the inner periphery of the nuclear envelope. The residual nucleolus was not observed in the nuclear matrix preparation **(Figure 1D).** This result validated the procedure employed for nuclear matrix isolation. A

resinless section electron microscopy study showed the nuclear matrix to consist of the nuclear lamina (NL) and an internal matrix connected to the lamina. This internal matrix is a network of irregular fibers with intricate fine structures evident in the images obtained.

Evaluation of MAR datasets showed good coverage and depth

The MAR DNA extracted from the three datasets was sequenced to analyse their genome-wide association. Coverage and depth analysis are usually performed for the mapped datasets obtained by next-generation sequencing techniques before downstream analysis. This confirms the uniform coverage of the mapped reads across the genome and gives the number of reads that are aligned to a reference base position. The quality of the mapped reads in PSGs from day 1, day 5, and day 7 were evaluated. Results showed that 12% of the regions sampled were covered at least twice and 10% of the regions were covered at least 5 times. Mean values of SG 1, SG 5, and SG 7 were 4.0, 4.6, and 7.2, respectively **(Figure 2A, left).** The data showed that 20% of the sampled base pairs had up to 10 overlapping reads in terms of sequencing depth, and 5% of the sampled base pairs exhibited more than 40 overlapping reads **(Figure 2A, right)**. Lower number of fraction of bases were observed in SG7. This could be due to the high read number of the SG7 dataset. It could also be attributed to the nature of MAR DNA which is observed to be highly repetitive and higher in their count at this terminal phase of 5^{th} instar development wherein the silk gland development is complete. This could be due to the, strict approach applied by the MAR finding algorithm to avoid reads from non-enriched MAR regions. Thus, the regions with high read coverage were identified using the MACS2. Analysis of the raw sequenced data showed an average of 74.85% unique sequences among the SG 1, SG 5, and SG 7 datasets **(Supplementary Table 3).** Thus the obtained datasets from SG1, SG5, and SG7 were deemed good to proceed for further analysis.

The number of MARs and the unique MARs in the three datasets show an increase in occurrence as development progresses

After the quantitative MAR DNA analysis of the three days (day1, day 5, and day 7), we looked into the number of MARs present in the sequenced nuclear matrix extractions in each of the datasets (SG 1, SG 5, and SG 7). There is a paucity of MAR-associated developmental studies. The number of MARs in each dataset was identified by peak calling of mapped reads. The results showed a total of 1861, 6158, and 11075 MAR regions in SG 1, SG 5, and SG 7, respectively. To understand the dynamics of MARs during different days of 5^{th} instar development,the common and unique regions were identified in three datasets. 1019 MAR regions were found to be common among SG1, SG5, and SG7 (**Figure 2F**). There is an increase in the number of unique MARs from day 1 to day 7 suggesting the role of these MARs in the developmental expression and regulation of genes in silk glands. 7.15% MARs were found to be common to all three days and 1.41%, 19.27%, and 52.47% MARs were unique to day 1, day 5, and day 7, respectively. The unique sequences may hold clues in further understanding the development-based differential gene expression of PSGs.

MAR DNAs showed dynamics in chromosome-wise distribution and with developmental progression

The *B. mori* genome is 460.3 Mb in size and is organized in 28 chromosomes with each chromosome harbouring many genes (38). The chromosome-wise distribution of MARs and MAR density was carried out from the merged call peak datasets derived from the mapped datasets as it is important in understanding the gene expression of chromosomes (18, 49). This information also gives clues on the topological location of the chromosome in the nucleus. The highest MAR density in PSGs was found to vary on different days of 5^{th} instar development, on chromosomes 19, 2, and 5 on day 1, day 5, and day 7, respectively (**Table 1)**. MAR density on chromosomes predicts chromatin loop size (18). The increase in MAR density from SG 1 to SG 5 and SG 5 to SG 7 showed decreased chromatin loop size and an increase in the number of chromatin loops as the development of 5^{th} instar *B. mori* progressed. The distribution of MARs determines the chromatin loop formation and is thus imperative in understanding the gene regulation processes (50). The MAR density has been shown to positively correlate with gene density **(Figure 2D).** The average chromatin loop length decreased from day 1 (253.91 kb) to day 5 (73.54 kb) to day 7 (39.19 kb). A weak positive correlation was found between MAR count and gene density in SG 1, SG 5, and SG 7 datasets **(Supplementary Figure 3).** The MAR DNA sequences associated with the nuclear matrix in the 200-1000 bp were found to be even more enriched in the size range of 200-400 bp **(Figure 2C).** These findings showed that chromosome-wise differences and developmental differences (day-wise) exist in MAR association with the nuclear matrix.

Short Tandem repeats- and transposable elements are associated with PSG nuclear matrix

Several repeat sequences have been associated with the nuclear matrix (51, 52), and STRs in particular play an important role in gene identification (53). Hence, STR analysis was performed to understand their association with the nuclear matrix during PSG development. The results of STR detection showed enrichment of STRs in SG 1 compared to SG 5 and SG 7. TATT, TGAA, and AGTC tetranucleotides; CCTAA, TTAGG, and GTTAG pentanucleotides; and TACCAA, and TTATTG hexanucleotides were the most abundant **(Figure 3A).** SG 1, SG 5, and SG 7 datasets showed AT- richness suggesting a high association of functional genes with the matrix in PSGs.

Several studies support the association between transposable elements (TEs) and gene expression (54, 55). The transposable elements associated with the nuclear matrix are found to ensure the necessary spatial arrangement of promoters and enhancers and aid the binding of MARs with MAR binding proteins. Many regulatory regions are associated with TEs as evident from whole-genome analyses of *Caenorhabditis elegans* (56), *Schizosaccharomyces pombe* (57), and humans (58, 59). Further, retrotransposon sequences of LTRs and LINE elements were enriched in MARs in the *Drosophila* euchromatic genome (13). RepeatMasker analysis showed LINES, LTR elements, and other repeats associated with the nuclear matrix in the PSG datasets. The maximum number of LINE elements and LTR elements were found on day 7. While the LINE elements occupied 4526 bp, 427 bp, and 22760 bp in SG 1, SG 5, and SG 7; the LTR elements were found to occupy 111 bp and 430 bp in SG 5 and SG 7 **(Table 3).** LINEs have internal promoters that initiate transcription upstream of the 5'end of the element or are co-transcribed along with the target site from an external host promoter. The enrichment of MARs in LINE elements on day 7 validates their role in the expression of genes related to the final stage of larval

development and silk production. The transposable element L1Bm (a LINE element), was shown to be matrix-associated. It is located in the upstream region of the FIBH promoter and was previously predicted to be involved in the regulation of the FIBH gene (34, 35). L1Bm was attached to the nuclear matrix of PSGs on all three days of 5th instar larval development. R1Bmks, another LINE element, was found to have regions of association with the nuclear matrix on day 1 and day 7 but not on day 5. Thus the differences in the enrichment and distribution of STRs and TEs of PSG developmental datasets hint at the complexity and multi-factorial regulation of the PSG gene.

Motifs found in the datasets correlated with known MAR Associated features

Motifs also correlate with gene expression and are associated with the nuclear matrix (60–62). The role of transcription factors and their binding to DNA sequences such as the MAR regions and the regulatory elements: promoters, enhancers, ORIs, and silencers, to aid the regulation of gene expression in eukaryotes has also been explored by previous studies (63) The use of MARs in transfected cells regardless of species showed their role in epigenetic maintenance and consistent insulating activity allowing relatively stable gene expression (20, 21). MARs were found to enhance the transgene expression and the MAR assisted transcriptional activation of these transgenes involved AT rich sequences and specific transcription factor binding motifs (22). The 5^{th} instar developmental datasets of SG 1, SG 5, and SG 7 were scanned for the presence of the most abundantly overrepresented DNA motifs, which are most often transcription factor binding sites. The motif patterns were found to be different in SG1, SG5, and SG7 **(Figure 3B).** The MAR-associated motifs such as ORI signals, topoisomerase II signals, TG rich signal, AT-rich signal, curved DNA, and kinked DNA were identified in the three developmental datasets **(Figure 3C).** Of all the MAR motifs, AT richness was found to be the highest in all the datasets with a percentage abundance of 92.18%, 92.28%, and 87.44% and the percentage abundance of ORI signals was 77.46%, 77.08%, and 68.53%, in SG 1, SG 5, and SG 7, respectively. These findings suggest that PSGs are transcriptionally more active in the 5^{th} instar larval stage. The curved DNA signals in the datasets were 19%, 5.68%, and 11.63% in SG 1, SG 5, and SG 7, respectively. The signals for kinked DNA were 12.35%, 6.43%, and 13.02%, in SG 1, SG 5, and SG 7, respectively. TG-rich signals were picked up at 2.11%, 1.98%, and 1.49% in SG 1, SG 5, and SG 7, respectively. The topoisomerase II signals were found to be 0.77%, 0.46%, and 0.57%, in SG 1, SG 5, and SG 7, respectively. The MAR recognition signatures occurred at an abundance of 9.39%, 9.08%, and 5.77% in SG 1, SG 5, and SG 7, respectively. These data show for the first time that MAR association in PSGs during development is dependent on multiple factors. Both sequence and structure of the DNA play a role in the association of chromatin to the nuclear matrix.

Identification of genomic elements and gene associated functions of MARs

MARs have been identified in intronic, exonic, and non-genic regions (64). The number of introns, exons, and non-genic regions associated with the nuclear matrix were identified in SG 1, SG 5, and SG 7 datasets by annotating the datasets against the gene models and gene lists and comparing each of the datasets. 1.11%, 37.54%, and 47.95% of the identified exons of the three datasets were found to be unique to SG 1, SG 5, and SG 7, respectively while 6.31% of the exons were common to the three datasets. 1.21%, 16.62%, and 58.21% of the identified introns of the three datasets were found to be unique to SG 1, SG 5, and SG 7,

respectively while 5.99 % of the introns were common to the three datasets. Of the total 10503 non-genic regions identified, 7.56% were common among the three datasets while 1.50%, 19.81%, and 50.75% unique non-genic regions were found in SG 1, SG 5, and SG 7 **(Figure 4A).** MARs were enriched in non-genic regions followed by intronic and exonic regions. Significant differences were observed in the number of non-genic, intronic, and exonic MARs among SG 1, SG 5, and SG 7. The analysis again confirmed the increased association of MARs from day 1 to day 7 of 5th instar development. The enrichment of MARs in the close proximity of the transcription start site validated their importance in the transcriptional regulation of genes (18). The region spanning the -100 kb to +100 kb region from the transcription start site was checked for the presence of MARs. Results showed that the highest peaks were observed at -10 to +10 kb region flanking TSS in all three datasets (SG 1, SG 5, and SG 7) **(Figure 4B).**

A comparison of the three datasets based on the association between the MAR and the genes involved in various biological pathways revealed that the most significant pathways in the three datasets were the basic biological pathways-the metabolic pathways, the genetic pathways, and the signalling pathways. An increased association was observed between the MAR and the genes involved in apoptosis, ribosome biogenesis, transcription, and DNA repair associated pathways from day 1 to day 7 **(Figure 5).** The transcription and DNA repair-related pathways increased significantly with developmental progression. These results correlate with the increased secretion of silk protein from PSGs.

Dynamic attachment of FIBH and Br- c regions with the nuclear matrix

The association of genes with the nuclear matrix is important in gene regulation (65). To understand the dynamics of the fibroin gene regulation, the 1600 kb (800 kb upstream and 800 kb downstream) region flanking FIBH gene was tested in SG 1, SG 5, and SG 7 for MARs. The MARs were observed to be different in number and position among the three datasets across the genome **(Figure 6A).** The distribution of MARs in the flanking region of FIBH on the three days of development varied. Forty-five MARs were found in this region, of which forty-two MARs were found to be developmentally different. Results show that when development progressed from SG 1 to SG 5, and SG 5 to SG 7, new MARs were produced indicating the formation of new chromatin domains. There were 1, 8, and 24 unique MARs in SG 1, SG 5, and SG 7 datasets, respectively while all three datasets shared 3 common MARs in the 1600 kb FIBH flanking region.

Previous work from our lab showed that the broad-complex negatively regulates fibroin gene expression (66).This was further supported by a study by a recent study that showed that the overexpression of BmBr-c Z2 in the transgenic lines of *B. mori* led to decline in the FIBH, Fib-l and P25 silk proteins. This mechanism was shown to be connected with the juvenile hormone pathway which is linked to the expression of its transcription factors, Bmdimm and BmKr-h1 (67). The expression of Bmdimm was shown to positively regulate FIBH which correlates with the effect of Br-c expression on the regulation of FIBH expression (68). Therefore, the 1000 bp flanking regions of Br-c gene were assessed for MAR association. The number of MARs increased from day 1 to day 7 **(Figure 6B)**. MARs were observed within the gene region of Br-c on day 5 and day 7. The number of MARs associated with the matrix of Br-c increased significantly from SG1 to SG7. Gene silencing has been shown to occur when MARs are located within a gene associated with the nuclear matrix (nuclear matrix attachment-induced silencing) (69). Our

data suggest that the MARs found in the coding regions of Br-c may be involved in the down-regulation of the broad-complex which in turn increase the FIBH expression. This is consistent with studies that showed ectopic expression of Br-c Z2 inhibited FIBH expression (67).

Discussion

The nuclear matrix helps in the maintenance of the higher-order chromatin structure and stability (70, 71). MARs bind and regulate chromatin regions and allow temporal and spatial regulation of genes. Not many studies have analysed the MAR sequences at the genome level. Our study characterized for the first time, tissue-specific genome-wide MARs during 5^{th} instar larval development in *B. mori*, to understand the dynamics associated with chromatin architecture in PSGs. Sequences were found to be enriched at the 200-400 bp range and imaging showed the inner architecture of the PSG nuclear matrix. Both the size of MARs associated with the nuclear matrix and the architecture matched previous studies. The median MAR length enrichment in humans is 596 bp (18), and in *Drosophila* is 400 bp (13).

MARs of the 5^{th} instar PSGs showed common and variable regions across the genome. The common regions are important for the constitutive function of the genome, while variable regions are essential for the dynamic expression of genes during development. Inter-MAR associations aid in the formation of transcriptionally active chromatin loops and regulate gene expression (72). They enable chromatin loop formation which dictates the initiation and regulation of transcriptional mechanisms (73–75). The nuclear matrix-associated transcription factors that bind at the MARs and the actively transcribed genes often coexist with the same transcriptional machinery (76, 77). Hence the chromatin loops are important for chromatin function. The average loop size of chromatin is in the range of 20-200 kb in HCT116 cell line and HeLa cells (78, 79). In silkworms, Li *et al.* demonstrated that chromatin loop size decreased when development progressed from day 1 to day 7 potentially resulting in significant differences in expressed transcripts from the 4^{th} moult until the wandering stage of PSGs. The average chromatin loop size has been shown to vary from 40 kb to 254 kb in different days of 5^{th} instar development. The size of the MAR anchored loops is appears to influence the expression efficiency of the genes. Therefore, a decrease in loop size towards day 7 in PSGs likely promotes gene expression.

To determine the MAR enrichment pattern in the genome, a chromosome-wise MAR distribution analysis was performed and the number of MARs per chromosome in SG 1, SG 5, and SG 7 were determined. The enrichment of MARs was checked for correlation against gene density in each chromosome and a weak positive correlation was seen between the MAR count in each chromosome and its gene density. This corroborates that MARs are relatively enriched in gene-rich regions of the genome.

In a follow up analysis to the current study, we selected specific genes from the protein expression data of previous work from our lab (80), which were differentially expressed, and found that the MARs found within these genes were dynamic in their spatio-temporal association in the NuMat. The variation in the location of the MARs and the occurrence of new MARs with developmental progression was observed among day 1 (SG 1), day 5 (SG 5) and day 7 (SG 7) PSG MAR datasets (unpublished data). These MARs may be

involved in the structural or functional roles during development in the PSGs of the 5th instar larval stage. MARs were also found to be present in the intronic regions. Two examples of such genes and its associated MAR regions are available in Supplementary Table 4 Similar findings were reported in a study on the MAR regions and their correlation with gene expression levels in *Arabidopsis thaliana,* where the expression data from previous studies was used to compare the expression levels of MAR containing genes and genes lacking MARs. It was found that the genes containing MARs had lower expression levels as they were down-regulated due to higher association with regulatory elements. Intragenic/ intronic MARs were said to cause tissue and organ specific regulation of gene expression. The differential expression of MAR containing genes and genes without MARs was explored during different stages in 2 different tissues (root and flower) and it was found that the MAR containing genes were involved in gene regulatory functions.TF families were enriched in S/MARs and a synergistic effect was observed for tissue and organ specific expression of TF genes and the presence of intragenic S/MARs (81).

Previous studies found an association of repeats to be correlated with gene expression (82). Transposable elements are known to be associated with the nuclear matrix. They are known to influence genome plasticity, genome architecture, and chromatin loop formation (44, 45). They often carry embedded transcription factor binding sites and therefore contribute to transcription and gene expression (83). L1Bm repeats were found in the intronic region along with an AT enriched modulator of fibroin (46, 47). L1Bm was also found to be an integral part of Bmmar1, a MAR identified in the 62 kb upstream region of fibroin (30). Its association with the nuclear matrix was further validated by its presence throughout the 5th instar development. The STRs, TTAGG, and CCTAA, enriched in the datasets are known telomeric repeats of *B. mori* and have many LTR repeats integrated into them which include L1Bm (32, 84). Moreover, telomeric repeats are associated with the nuclear matrix (51). Other transposable elements like *gypsy, jockey,* and *roo* involved in establishing chromatin boundaries were also found to be associated with the nuclear matrix (13, 14, 85). The LTR transposon, Ty1 associated with certain regions of chromatin and an important player in genome organisation, along with R1Bmks, a non-LTR transposon of *B. mori*, was also found to be associated with the nuclear matrix. STRs also play a role in gene expression (86) and enrichment of repeat sequences was found in all three datasets.

MAR sequence data sets in the current study on all three days showed AT base enrichment as observed in previous studies (13, 87). In addition, we identified a wide range of motifs that were associated with the nuclear matrix (88). While AT richness is often associated with flexibility, certain recognition sites for protein-DNA binding and narrowing of the minor groove (89); curved and kinked DNA motif signals are identified by their overall shape and are bound by proteins (90, 91); topoisomerase II motif signals are important for supercoiling as well as preventing rotation of the double helix (92); and ORI signals are strongly associated with the nuclear matrix and execute or prevent the replication process and are hence important in gene regulation (93). The MAR recognition signatures which are two proximal degenerate sequences are common to most MARs identified and were also found in all three datasets (94). The three most abundantly overrepresented motifs that we identified were putative transcription factor binding sites. These signature motifs of day 1, day 5, and day 7 were found to be different. The variations in the three most abundantly overrepresented motif sequences among the three datasets indirectly convey the

change in the precedence of the genes transcribed during development and the dynamic role of MARs and their influence in gene expression and regulation. Future work focusing on these motifs will help understand their relevance across other genomes.

The association of the matrix with intergenic regions is involved in protecting genes from being silenced (95). MARs present in the exonic regions may be involved in transcription (13). Although not much is known about MARs present in the intronic regions (96), they mainly occur at the flanks of transcribed regions, in 5'-introns and telomeres (97). To address the genomic context of MARs, their presence in different genomic regions was determined. The results portrayed enrichment of MARs in the order: non-genic>introns>exons. MAR distance from the transcription start site affects the transcription of genes (98). The enrichment of MARs throughout the -100kb to +100 kb region flanking the transcription start site was established with many MARs enriched proximal to the transcription start site. The enrichment of intergenic MARs proximal to the transcription start site may influence the transcription of genes downstream to them, resulting in the enhanced expression of genes in PSGs from day 1 to day 7. The MAR association with genomic elements was also determined similarly in other studies (13, 18). The association of MAR with genomic elements was also found in plant species such as *Arabidopsis* (15, 81).

To further understand the functions of genes regulated by these MARs, the biological pathways of the MAR-associated genes were explored. We found that the metabolic pathway genes were most abundantly associated with the nuclear matrix at all the three-time points of 5^{th} instar larval development. These results confirm that PSGs express more genes involved in metabolic control as they progress towards the wandering stage (80, 99).

The current study identified common and variable MARs flanking the FIBH gene in PSG of 5^{th} instar larvae which showed day-specific variations. Many MAR regions were identified for the first time flanking the FIBH gene. This is consistent with previous literature which shows that actively transcribed genes are associated with the nuclear matrix (65). Since the broad complex was shown to negatively regulate FIBH expression (66), the Br-c gene was analysed for MARs. Though matrix attachment in intergenic regions can protect genes from being silenced, the presence of MARs within genes correlates with gene silencing (69, 95). Br-c showed presence of MARs in the gene region which could contribute towards the downregulation of the genes on day 5 and day 7 of the 5^{th} instar larval stage. The identified MARs likely downregulate the expression of broad complex which in turn upregulates FIBH gene expression on day 5 and day 7 of the 5^{th} instar stage.

Our study is the first to report the developmental analysis of genome-wide MAR sequencing data in the PSGs of *B. mori*. The results of the present analysis show that the organization of the silk gland genome is under the dynamic control of the nuclear matrix. These elements and their association to PSG nuclear matrix can be linked directly or indirectly in the regulation of FIBH expression.

Declarations

DATA AVAILABILITY

The raw sequence data from this study has been submitted to the NCBI Sequence Read Archive under the bio project numbers PRJNA554391 and PRJNA555810 with biosample numbers: SUB6046074 (SG1), SUB6046196 (SG5), and SUB6046343 (SG7).

FUNDING

This work was supported by the Department of Biotechnology, Government of India (BT/PR15319/TDS/121/12/2015) to MA.

AUTHORS' CONTRIBUTIONS

The computational data analysis was performed by ARC. RR and AL were involved in feeding and rearing of worms and maintenance of sericulture facility in the department. SR supervised the computational analysis and participated in the discussion and writing. The conceptualization and supervision of experiments was carried out by AM. All authors read and approved the final manuscript.

ACKNOWLEDGEMENT

The authors thank Mr. K Satya Rao and Mr. G.V. Rajesh, Department of sericulture, Government of Andhra Pradesh for providing silkworms, and Dr. Pranali P. Pathare, 3P Scientific Communications for editorial support.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

References

1. Jackson,D.A. and Cook,P.R. (1985) Transcription occurs at a nucleoskeleton. *EMBO J.*, **4**, 919–925.

2. Liu,J., Bramblett,D., Zhu,Q., Lozano,M., Kobayashi,R., Ross,S.R. and Dudley,J.P. (1997) The matrix attachment region-binding protein SATB1 participates in negative regulation of tissue-specific gene expression. *Mol Cell Biol*, **17**, 5275–5287.

3. Wilson,R.H.C. and Coverley,D. (2013) Relationship between DNA replication and the nuclear matrix. *Genes Cells*, **18**, 17–31.

4. Berezney,R., Mortillaro,M.J., Ma,H., Wei,X. and Samarabandu,J. (1995) The nuclear matrix: a structural milieu for genomic function. *Int Rev Cytol*, **162A**, 1–65.

5. Ottaviani,D., Lever,E., Takousis,P. and Sheer,D. (2008) Anchoring the genome. *Genome Biol*, **9**, 201.

6. Gimenes,F., Takeda,K.I., Fiorini,A., Gouveia,F.S. and Fernandez,M.A. (2008) Intrinsically bent DNA in replication origins and gene promoters. *Genet Mol Res*, **7**, 549–558.

7. Dolgova,A.S. and Dolgov,S.V. (2019) Matrix attachment regions as a tool to influence plant transgene expression. *3 Biotech*, **9**, 176.

8. Zhang,K.-W., Wang,J.-M. and Zheng,C.-C. (2004) [The potential role of nuclear matrix attachment regions (MARs) in regulation of gene expression]. *Sheng Wu Gong Cheng Xue Bao*, **20**, 6–9.

9. Ji,L., Xu,R., Lu,L., Zhang,J., Yang,G., Huang,J., Wu,C. and Zheng,C. (2013) TM6, a novel nuclear matrix attachment region, enhances its flanking gene expression through influencing their chromatin structure. *Mol Cells*, **36**, 127–137.

10. Allen,G.C., Spiker,S. and Thompson,W.F. (2000) Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Mol. Biol.*, **43**, 361–376.

11. Brouwer,C., Bruce,W., Maddock,S., Avramova,Z. and Bowen,B. (2002) Suppression of transgene silencing by matrix attachment regions in maize: a dual role for the maize 5' ADH1 matrix attachment region. *Plant Cell*, **14**, 2251–2264.

12. Linnemann,A.K. and Krawetz,S.A. (2009) Silencing by nuclear matrix attachment distinguishes cell-type specificity: association with increased proliferation capacity. *Nucleic Acids Res*, **37**, 2779–2788.

13. Pathak,R.U., Srinivasan,A. and Mishra,R.K. (2014) Genome-wide mapping of matrix attachment regions in Drosophila melanogaster. *BMC Genomics*, **15**, 1022.

14. Mamillapalli,A., Pathak,R.U., Garapati,H.S. and Mishra,R.K. (2013) Transposable element 'roo' attaches to nuclear matrix of the Drosophila melanogaster. *J. Insect Sci.*, **13**, 111.

15. Rudd,S., Frisch,M., Grote,K., Meyers,B.C., Mayer,K. and Werner,T. (2004) Genome-wide in silico mapping of scaffold/matrix attachment regions in Arabidopsis suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. *Plant Physiol.*, **135**, 715–722.

16. Boulikas,T. (1994) Transcription factor binding sites in the matrix attachment region (MAR) of the chicken alpha-globin gene. *J Cell Biochem*, **55**, 513–529.

17. Glazko,G.V., Koonin,E.V., Rogozin,I.B. and Shabalina,S.A. (2003) A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends in Genetics*, **19**, 119–124.

18. Narwade,N., Patel,S., Alam,A., Chattopadhyay,S., Mittal,S. and Kulkarni,A. (2019) Mapping of scaffold/matrix attachment regions in human genome: a data mining exercise. *Nucleic Acids Res*, **47**, 7247–7261.

19. Chang,M., Liu,R., Jin,Q., Liu,Y. and Wang,X. (2014) Scaffold/matrix attachment regions from CHO cell chromosome enhanced the stable transfection efficiency and the expression of transgene in CHO cells. *Biotechnol Appl Biochem*, **61**, 510–516.

20. Namciu,S.J., Blochlinger,K.B. and Fournier,R.E. (1998) Human matrix attachment regions insulate transgene expression from chromosomal position effects in Drosophila melanogaster. *Mol Cell Biol*, **18**, 2382–2391.

21. Lin,Y., Li,Z., Wang,T., Wang,X., Wang,L., Dong,W., Jing,C. and Yang,X. (2015) MAR characteristic motifs mediate episomal vector in CHO cells. *Gene*, **559**, 137–143.

22. Sun,Q.-L., Zhao,C.-P., Chen,S.-N., Wang,L. and Wang,T.-Y. (2016) Molecular characterization of a human matrix attachment region that improves transgene expression in CHO cells. *Gene*, **582**, 168–172.

23. Bradner,S.A., Partlow,B.P., Cebe,P., Omenetto,F.G. and Kaplan,D.L. (2017) Fabrication of elastomeric silk fibers. *Biopolymers*, **107**.

24. Zhou,C.Z., Confalonieri,F., Medina,N., Zivanovic,Y., Esnault,C., Yang,T., Jacquet,M., Janin,J., Duguet,M., Perasso,R., *et al.* (2000) Fine organization of Bombyx mori fibroin heavy chain gene. *Nucleic Acids Res.*, **28**, 2413–2419.

25. Dong,Z.-C., Li,F.-F., Xiang,Y.-C., Mao,H.-H., Zhang,J., Cui,H.-Z., Min,H.-P. and Lu,C. (2012) DNA replication events during larval silk gland development in the silkworm, Bombyx mori. *J Insect Physiol*, **58**, 974–978.

26. Zhou,Q.-Z., Fu,P., Li,S.-S., Zhang,C.-J., Yu,Q.-Y., Qiu,C.-Z., Zhang,H.-B. and Zhang,Z. (2020) A Comparison of Co-expression Networks in Silk Gland Reveals the Causes of Silk Yield Increase During Silkworm Domestication. *Front Genet*, **11**, 225.

27. Dhawan,S. and Gopinathan,K.P. (2003) Expression profiling of homeobox genes in silk gland development in the mulberry silkworm Bombyx mori. *Dev Genes Evol*, **213**, 523–533.

28. Grzelak,K. (1995) Control of expression of silk protein genes. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.*, **110**, 671–681.

29. Wasąg,P. and Lenartowski,R. (2016) Nuclear matrix - structure, function and pathogenesis. *Postepy Hig Med Dosw (Online)*, **70**, 1206–1219.

30. Zhou,C.-Z., Confalonieri,F., Esnault,C., Zivanovic,Y., Jacquet,M., Janin,J., Perasso,R., Li,Z.-G. and Duguet,M. (2003) The 62-kb upstream region of Bombyx mori fibroin heavy chain gene is clustered of repetitive elements and candidate matrix association regions. *Gene*, **312**, 189–195.

31. Adams,D.S., Eickbush,T.H., Herrera,R.J. and Lizardi,P.M. (1986) A highly reiterated family of transcribed oligo(A)-terminated, interspersed DNA elements in the genome of Bombyx mori. *J Mol Biol*, **187**, 465–478.

32. Ichimura,S., Mita,K. and Sugaya,K. (1997) A major non-LTR retrotransposon of Bombyx mori, L1Bm. *J. Mol. Evol.*, **45**, 253–264.

33. Robertson,H.M. (1993) The mariner transposable element is widespread in insects. *Nature*, **362**, 241–245.

34. Robertson,H.M. and Asplund,M.L. (1996) Bmmar1: a basal lineage of the mariner family of transposable elements in the silkworm moth, Bombyx mori. *Insect Biochem Mol Biol*, **26**, 945–954.

35. Zhou,C.Z. and Liu,B. (2001) Identification and characterization of a silkgland-related matrix association region in Bombyx mori. *Gene*, **277**, 139–144.

36. Bozzola,John.J. and Russell,L.D. (1998) Electron Microscopy Principles and Techniques for Biologists 2nd ed. Jones and Bartlett Publishers, Sudbury, Massachusetts.

37. Martin,M. (2011) Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet J*, **17**, 10–12.

38. Kawamoto,M., Jouraku,A., Toyoda,A., Yokoi,K., Minakuchi,Y., Katsuma,S., Fujiyama,A., Kiuchi,T., Yamamoto,K. and Shimada,T. (2019) High-quality genome assembly of the silkworm, Bombyx mori. *Insect Biochem. Mol. Biol.*, **107**, 53–62.

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

40. Giardine,B., Riemer,C., Hardison,R.C., Burhans,R., Elnitski,L., Shah,P., Zhang,Y., Blankenberg,D., Albert,I., Taylor,J., *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res*, **15**, 1451–1455.

41. Jalili,V., Afgan,E., Gu,Q., Clements,D., Blankenberg,D., Goecks,J., Taylor,J. and Nekrutenko,A. (2020) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res*, **48**, W395–W402.

42. Zhu,Z., Guan,Z., Liu,G., Wang,Y. and Zhang,Z. (2019) SGID: a comprehensive and interactive database of the silkworm. *Database (Oxford)*, **2019**.

43. Medina-Rivera,A., Defrance,M., Sand,O., Herrmann,C., Castro-Mondragon,J.A., Delerce,J., Jaeger,S., Blanchet,C., Vincens,P., Caron,C., *et al.* (2015) RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic Acids Res*, **43**, W50-56.

44. Nguyen,N.T.T., Contreras-Moreira,B., Castro-Mondragon,J.A., Santana-Garcia,W., Ossio,R., Robles-Espinoza,C.D., Bahin,M., Collombet,S., Vincens,P., Thieffry,D., *et al.* (2018) RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res*, **46**, W209–W214.

45. Beier,S., Thiel,T., Münch,T., Scholz,U. and Mascher,M. (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics*, **33**, 2583–2585.

46. Zhong,B., Yu,Y., Xu,Y., Yu,H., Lu,X., Miao,Y., Yang,J., Xu,H., Hu,S. and Lou,C. (2005) Analysis of ESTs and gene expression patterns of the posterior silkgland in the fifth instar larvae of silkworm, Bombyx mori L. *Sci China C Life Sci*, **48**, 25–33.

47. Nickerson,J. (2001) Experimental observations of a nuclear matrix. *Journal of Cell Science*, **114**, 463–474.

48. Nickerson,J.A., Krockmalnic,G., Wan,K.M. and Penman,S. (1997) The nuclear matrix revealed by eluting chromatin from a cross-linked nucleus. *Proc Natl Acad Sci U S A*, **94**, 4446–4450.

49. Fritz,A.J., Sehgal,N., Pliss,A., Xu,J. and Berezney,R. (2019) Chromosome territories and the global regulation of the genome. *Genes Chromosomes Cancer*, **58**, 407–426.

50. Razin,S.V. (2001) The nuclear matrix and chromosomal DNA loops: is their any correlation between partitioning of the genome into loops and functional domains? *Cell Mol Biol Lett*, **6**, 59–69.

51. de Lange,T. (1992) Human telomeres are attached to the nuclear matrix. *EMBO J.*, **11**, 717–724.

52. Pathak,R.U., Mamillapalli,A., Rangaraj,N., Kumar,R.P., Vasanthi,D., Mishra,K. and Mishra,R.K. (2013) AAGAG repeat RNA is an essential component of nuclear matrix in Drosophila. *RNA Biol*, **10**, 564–571.

53. Trochez-Solarte,J.D., Ruiz-Erazo,X., Almanza-Pinzon,M. and Zambrano-Gonzalez,G. (2019) Role of microsatellites in genetic analysis of Bombyx mori silkworm: a review. *F1000Res*, **8**, 1424.

54. Britten,R.J. (1996) Cases of ancient mobile element DNA insertions that now affect gene regulation. *Mol Phylogenet Evol*, **5**, 13–17.

55. Elbarbary,R.A., Lucas,B.A. and Maquat,L.E. (2016) Retrotransposons as regulators of gene expression. *Science*, **351**, aac7247.

56. Ganko,E.W., Bhattacharjee,V., Schliekelman,P. and McDonald,J.F. (2003) Evidence for the contribution of LTR retrotransposons to C. elegans gene evolution. *Mol Biol Evol*, **20**, 1925–1931.

57. Bowen,N.J., Jordan,I.K., Epstein,J.A., Wood,V. and Levin,H.L. (2003) Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of Schizosaccharomyces pombe. *Genome Res*, **13**, 1984–1997.

58. Jordan,I.K., Rogozin,I.B., Glazko,G.V. and Koonin,E.V. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet*, **19**, 68–72.

59. Thornburg,B.G., Gotea,V. and Makałowski,W. (2006) Transposable elements as a significant source of transcription regulating signals. *Gene*, **365**, 104–110.

60. Liebich,I., Bode,J., Reuter,I. and Wingender,E. (2002) Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Res*, **30**, 3433–3442.

61. Hagedorn,C., Gogol-Döring,A., Schreiber,S., Epplen,J.T. and Lipps,H.J. (2017) Genome-wide profiling of S/MAR-based replicon contact sites. *Nucleic Acids Res*, **45**, 7841–7854.

62. Singh,G.B., Kramer,J.A. and Krawetz,S.A. (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res*, **25**, 1419–1425.

63. Boulikas,T. (1994) A compilation and classification of DNA binding sites for protein transcription factors from vertebrates. *Crit Rev Eukaryot Gene Expr*, **4**, 117–321.

64. Anthony,A. and Blaxter,M. (2007) Association of the matrix attachment region recognition signature with coding regions in Caenorhabditis elegans. *BMC Genomics*, **8**, 418.

65. Ciejek,E.M., Tsai,M.J. and O'Malley,B.W. (1983) Actively transcribed genes are associated with the nuclear matrix. *Nature*, **306**, 607–609.

66. Ali,A., Bovilla,V.R., Mysarla,D.K., Siripurapu,P., Pathak,R.U., Basu,B., Mamillapalli,A. and Bhattacharya,S. (2017) Knockdown of Broad-Complex Gene Expression of Bombyx mori by Oligopyrrole Carboxamides Enhances Silk Production. *Sci Rep*, **7**, 805.

67. Cong,J., Tao,C., Zhang,X., Zhang,H., Cheng,T. and Liu,C. (2020) Transgenic Ectopic Overexpression of Broad Complex (BrC-Z2) in the Silk Gland Inhibits the Expression of Silk Fibroin Genes of Bombyx mori. *Insects*, **11**, 374.

68. Zhao,X.-M., Liu,C., Jiang,L.-J., Li,Q.-Y., Zhou,M.-T., Cheng,T.-C., Mita,K. and Xia,Q.-Y. (2015) A juvenile hormone transcription factor Bmdimm-fibroin H chain pathway is involved in the synthesis of silk protein in silkworm, Bombyx mori. *J. Biol. Chem.*, **290**, 972–986.

69. Linnemann,A.K., Platts,A.E. and Krawetz,S.A. (2009) Differential nuclear scaffold/matrix attachment marks expressed genes. *Hum Mol Genet*, **18**, 645–654.

70. Silva-Santiago,E., Pardo,J.P., Hernández-Muñoz,R. and Aranda-Anzaldo,A. (2017) The nuclear higher-order structure defined by the set of topological relationships between DNA and the nuclear matrix is species-specific in hepatocytes. *Gene*, **597**, 40–48.

71. Bera,M. and Sengupta,K. (2020) Nuclear filaments: role in chromosomal positioning and gene expression. *Nucleus*, **11**, 99–110.

72. Wang,L., Di,L.-J., Lv,X., Zheng,W., Xue,Z., Guo,Z.-C., Liu,D.-P. and Liang,C.-C. (2009) Inter-MAR association contributes to transcriptionally active looping events in human beta-globin gene cluster. *PLoS ONE*, **4**, e4629.

73. Iarovaia,O.V., Akopov,S.B., Nikolaev,L.G., Sverdlov,E.D. and Razin,S.V. (2005) Induction of transcription within chromosomal DNA loops flanked by MAR elements causes an association of loop DNA with the nuclear matrix. *Nucleic Acids Res*, **33**, 4157–4163.

74. Getzenberg,R.H. (1994) Nuclear matrix and the regulation of gene expression: tissue specificity. *J Cell Biochem*, **55**, 22–31.

75. Laemmli,U.K., Käs,E., Poljak,L. and Adachi,Y. (1992) Scaffold-associated regions: cis-acting determinants of chromatin structural loops and functional domains. *Curr Opin Genet Dev*, **2**, 275–285.

76. Davie,J.R. (1997) Nuclear matrix, dynamic histone acetylation and transcriptionally active chromatin. *Mol Biol Rep*, **24**, 197–207.

77. Osborne,C.S., Chakalova,L., Brown,K.E., Carter,D., Horton,A., Debrand,E., Goyenechea,B., Mitchell,J.A., Lopes,S., Reik,W., *et al.* (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, **36**, 1065–1071.

78. You,Q., Cheng,A.Y., Gu,X., Harada,B.T., Yu,M., Wu,T., Ren,B., Ouyang,Z. and He,C. (2021) Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution. *Nat Biotechnol*, **39**, 225–235.

79. Jackson,D.A., Dickinson,P. and Cook,P.R. (1990) The size of chromatin loops in HeLa cells. *EMBO J*, **9**, 567–571.

80. Bovilla,V.R., Padwal,M.K., Siripurapu,P., Basu,B. and Mamillapalli,A. (2016) Developmental proteome dynamics of silk glands in the 5th instar larval stage of Bombyx mori L (CSR2×CSR4). *Biochim. Biophys. Acta*, **1864**, 860–868.

81. Tetko,I.V., Haberer,G., Rudd,S., Meyers,B., Mewes,H.-W. and Mayer,K.F.X. (2006) Spatiotemporal expression control correlates with intragenic scaffold matrix attachment regions (S/MARs) in Arabidopsis thaliana. *PLoS Comput Biol*, **2**, e21.

82. Lu,J.Y., Shao,W., Chang,L., Yin,Y., Li,T., Zhang,H., Hong,Y., Percharde,M., Guo,L., Wu,Z., *et al.* (2020) Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep*, **30**, 3296-3311.e5.

83. Chuong,E.B., Elde,N.C. and Feschotte,C. (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet*, **18**, 71–86.

84. Kubo,Y., Okazaki,S., Anzai,T. and Fujiwara,H. (2001) Structural and phylogenetic analysis of TRAS, telomeric repeat-specific non-LTR retrotransposon families in Lepidopteran insects. *Mol. Biol. Evol.*, **18**, 848–857.

85. Nabirochkin,S., Ossokina,M. and Heidmann,T. (1998) A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J. Biol. Chem.*, **273**, 2473–2479.

86. Jakubosky,D., D'Antonio,M., Bonder,M.J., Smail,C., Donovan,M.K.R., Young Greenwald,W.W., Matsui,H., i2QTL Consortium, D'Antonio-Chronowska,A., Stegle,O., *et al.* (2020) Properties of structural variants and short tandem repeats associated with gene expression and complex traits. *Nat Commun*, **11**, 2927.

87. Flickinger,R.A. (1986) Localization of AT-rich sequences at the nuclear matrix. *Cell Biol. Int. Rep.*, **10**, 415–420.

88. Bode,J., Stengert-Iber,M., Kay,V., Schlake,T. and Dietz-Pfeilstetter,A. (1996) Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. *Crit. Rev. Eukaryot. Gene Expr.*, **6**, 115–138.

89. Szemes,M., Gyorgy,A., Paweletz,C., Dobi,A. and Agoston,D.V. (2006) Isolation and characterization of SATB2, a novel AT-rich DNA binding protein expressed in development- and cell-specific manner in the rat brain. *Neurochem Res*, **31**, 237–246.

90. Fukuda,Y. (2000) Interaction of nuclear proteins with intrinsically curved DNA in a matrix attachment region of a tobacco gene. *Plant Mol Biol*, **44**, 91–98.

91. Collis,C.M., Molloy,P.L., Both,G.W. and Drew,H.R. (1989) Influence of the sequence-dependent flexure of DNA on transcription in E. coli. *Nucleic Acids Res*, **17**, 9447–9468.

92. Razin,S.V., Vassetzky,Y.S. and Hancock,R. (1991) Nuclear matrix attachment regions and topoisomerase II binding and reaction sites in the vicinity of a chicken DNA replication origin. *Biochem Biophys Res Commun*, **177**, 265–270.

93. Boulikas,T. (1996) Common structural features of replication origins in all life forms. *J Cell Biochem*, **60**, 297–316.

94. van Drunen,C.M., Sewalt,R.G., Oosterling,R.W., Weisbeek,P.J., Smeekens,S.C. and van Driel,R. (1999) A bipartite sequence element associated with matrix/scaffold attachment regions. *Nucleic Acids Res*, **27**, 2924–2930.

95. Linnemann,A.K. and Krawetz,S.A. (2009) Silencing by nuclear matrix attachment distinguishes cell-type specificity: association with increased proliferation capacity. *Nucleic Acids Res*, **37**, 2779–2788.

96. Romig,H., Ruff,J., Fackelmayer,F.O., Patil,M.S. and Richter,A. (1994) Characterisation of two intronic nuclear-matrix-attachment regions in the human DNA topoisomerase I gene. *Eur J Biochem*, **221**, 411–419.

97. Bode,J., Schlake,T., Ríos-Ramírez,M., Mielke,C., Stengert,M., Kay,V. and Klehr-Wirth,D. (1995) Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. *Int Rev Cytol*, **162A**, 389–454.

98. Schübeler,D., Mielke,C., Maass,K. and Bode,J. (1996) Scaffold/matrix-attached regions act upon transcription in a context-dependent manner. *Biochemistry*, **35**, 11160–11169.

99. Li,J., Cai,Y., Ye,L., Wang,S., Che,J., You,Z., Yu,J. and Zhong,B. (2014) MicroRNA expression profiling of the fifth-instar posterior silk gland of Bombyx mori. *BMC Genomics*, **15**, 410.

Tables

Table 1. Chromosome-wise distribution of MARs. The number of MARs per chromosome were identified and their density was calculated (number of MARs/ Size of Chromosome).

Chromosome	Size (Mb)	No. of genes	Gene density	MAR count			MAR/Mb		
				SG1	SG5	SG7	SG1	SG5	SG7
1	20.66	710	34.37	73	255	454	3.53	12.34	21.97
2	8.39	452	53.87	35	168	206	4.17	20.02	24.55
3	15.21	569	37.41	63	216	375	4.14	14.20	24.65
4	18.73	769	41.06	78	263	490	4.16	14.04	26.16
5	19.06	798	41.87	76	275	577	3.99	14.43	30.27
6	16.65	568	34.11	69	209	374	4.14	12.55	22.46
7	13.94	489	35.08	46	176	320	3.30	12.63	22.96
8	16.26	586	36.04	71	239	430	4.37	14.70	26.45
9	16.79	542	32.28	60	215	432	3.57	12.81	25.73
10	17.61	670	38.05	77	212	461	4.37	12.04	26.18
11	20.44	928	45.40	110	304	515	5.38	14.87	25.20
12	17.58	651	37.03	68	228	476	3.87	12.97	27.08
13	17.73	614	34.63	56	209	394	3.16	11.79	22.22
14	13.34	427	32.01	48	168	314	3.60	12.59	23.54
15	18.44	807	43.76	56	228	477	3.04	12.36	25.87
16	14.33	600	41.87	36	175	320	2.51	12.21	22.33
17	16.84	662	39.31	70	225	427	4.16	13.36	25.36
18	15.69	538	34.29	66	201	378	4.21	12.81	24.09
19	14.80	592	40.00	73	229	361	4.93	15.47	24.39
20	12.37	474	38.32	54	195	339	4.37	15.76	27.41
21	15.31	499	32.59	71	219	384	4.64	14.30	25.08
22	18.48	659	35.66	63	238	430	3.41	12.88	23.27
23	21.46	739	34.44	61	230	501	2.84	10.72	23.35
24	17.35	646	37.23	56	238	369	3.23	13.72	21.27
25	14.54	560	38.51	56	196	349	3.85	13.48	24.00
26	11.47	378	32.96	45	164	242	3.92	14.30	21.10
27	10.93	264	24.15	47	143	230	4.30	13.08	21.04
28	10.60	337	31.79	33	169	190	3.11	15.94	17.92

Table 2. MAR-Associated features. The MAR associated motifs were identified in the SG 1, SG 5, and SG 7 datasets using RSAT. The number of times the signal is detected for these motifs in each dataset is shown.

S. No.	Motif Name	Motif Index	Sequence	SG 1	SG 5	SG 7
1	ORI Signal	m_1	ATTA	12611	40445	48448
2	ORI Signal	m_2	ATTTA	4934	16748	18480
3	ORI Signal	m_3	ATTTTA	1670	5082	5623
4	TG Rich Signal	m_4	TGTTTTG	75	307	386
5	TG Rich Signal	m_5	TGTTTTTTG	37	48	36
6	TG Rich Signal	m_6	TTTTGGGG	9	18	81
7	Curved DNA Signal	m_7	$AAAAn_7AAAAn_7AAAA$	75	183	90
8	Curved DNA Signal	m_8	$TTTTn_7TTTTn_7TTTT$	75	183	90
9	Curved DNA Signal	m_9	TTTAAA	1161	4194	4202
10	Kinked DNA Signal	m_{10}	Tan_3TGn_3CA	107	262	847
11	Kinked DNA Signal	m_{11}	TAn_3CAn_3TG	589	694	3796
12	Kinked DNA Signal	m_{12}	TGn_3TAn_3CA	72	160	215
13	Kinked DNA Signal	m_{13}	TGn_3 CA n_3TA	107	262	847
14	Kinked DNA Signal	m_{14}	CA n_3TAn_3TG	111	470	468
15	Kinked DNA Signal	m_{15}	CA n_3TGn_3TA	589	694	3796
16	mtopo-II Signal	m_{16}	RnYnnCnnGYnGKTnYnY	18	27	97
17	dtopo-II Signal	m_{17}	GTnWAYATTnATnnR	11	30	30
18	AT Rich Signal	m_{18}	WWWWWW	22009	37915	77404
19	MRS	MRS		160	560	640

Table 3. Transposable elements associated with posterior silk gland NuMat: Repeat Masker was used with rmblast by masking interspersed and simple repeats against Drosophila TE database. L1Bm and R1Bmks were identified in the three datasets using BLAST2 (hits with >=90% percent identity and >=100bp were considered).

S. No.	Type	Number of elements			Length occupied (bp)			Percentage of sequence (%)		
		SG 1	SG 5	SG 7	SG 1	SG 5	SG 7	SG 1	SG 5	SG 7
1.	LINEs									
	L1Bm	12	2	84	3108	427	20512	0.56	0.0261	0.7997
	R1Bmks	2	-	2	1418	-	2124	0.256		0.082
	R1/LOA/Jockey	0	0	2	0	0	124	0.00	0.00	0.00
2	LTR elements									
	Ty1/Copia	0		2	0		262	0.00		0.01
	Gypsy/DIRS1	0	1	1	0	111	168	0.00	0.01	0.01
3	Total interspersed repeats:	-	-	-	1360	196	11060	0.25	0.01	0.43
4	Small RNA	39	1	43	12045	107	12993	2.18	0.01	0.51
5	Simple repeats	396	877	745	24797	55758	48261	4.48	3.41	1.88
6	Low complexity	61	209	153	3317	10250	7744	0.60	0.63	0.30

Figures

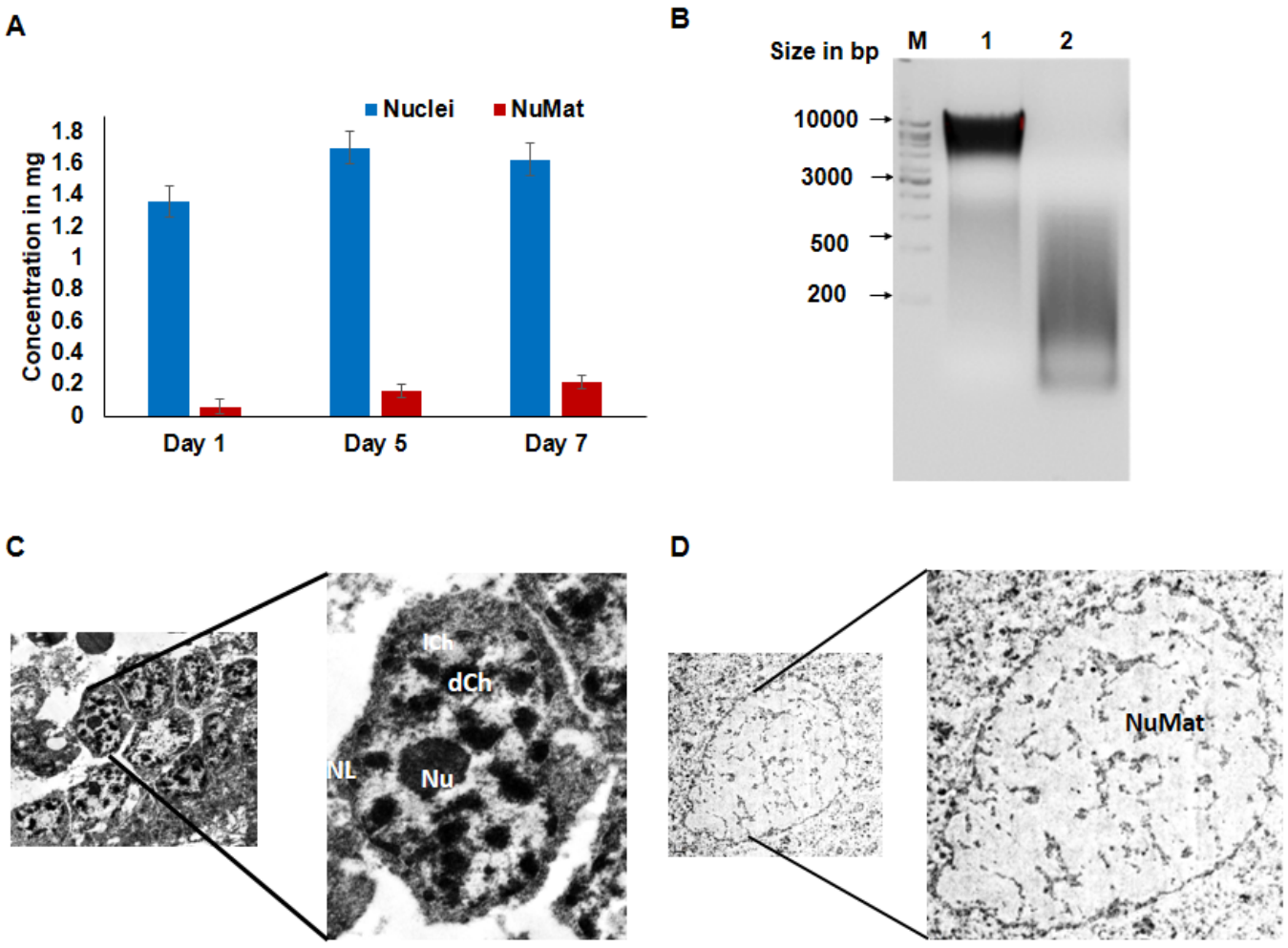

Figure 1

Isolation of MAR DNA from 5th instar posterior silk glands of *B. mori*. **(A)** The nuclear and MAR DNA of day 1, day 5 and day 7 were quantified using U.V spectrophotometry (n=5). Student's T-test (two sample using unequal variances) used for statistical analysis and no significant enrichment was found in MAR DNA from day 1 to day 7 was observed ($p > 0.05$) while significant difference was observed between concentrations of nuclei and nuclear matrix (NuMat) DNA ($p < 0.005$). **(B)** A 1kb DNA ladder (labelled 'M'), the isolated nuclear DNA and MAR DNA from 5th instar day 5 PSGs were run on lane 1, lane 2 and lane 3 respectively on a 1.2% agarose gel.**(C)** Transmission electron microscopy imaging of nucleus (left), Scale: 1µm, showing nucleolus (Nu), peripheral condensed chromatin (dCh) and nuclear lamina (NL) and **(D)** Transmission electron microscopy image of nuclear matrix (right) revealed the continuation of the inner network towards the inner periphery of the nuclear lamina. The residual nucleolus was not observed in the NuMat preparation of this study. Scale: 1µm.

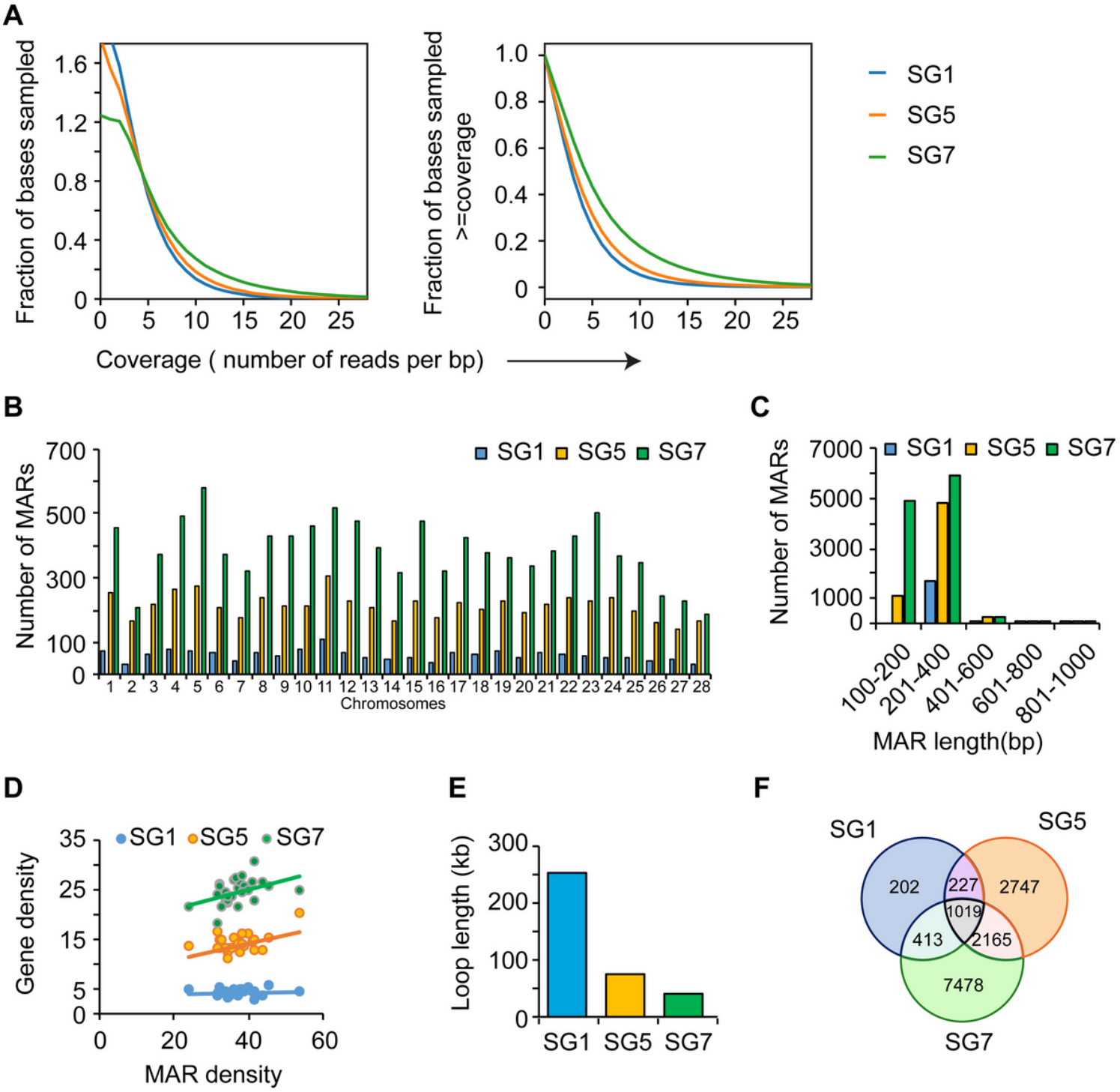

Figure 2

Analysis of Chromosome – wise distributed MARs. The MAR DNA from posterior silk glands extracted from day 1, day 5 and day 7 were sequenced using Illumina platform and mapped against the *Bombyx mori* annotated genome downloaded from Silkbase. **(A)** The depth (left) and coverage (right) of the mapped sequenced reads was analysed using PlotCoverage. **(B)** Chromosome wise distribution of MARs **(C)** Length of the MARs in the three datasets was derived from the enriched MAR peaks data. **(D)** Scatter plot of the density of MARs plotted against the gene densities for each chromosome. **(E)** The average loop length in

the three datasets derived from the distance between MAR regions from enriched MAR peaks data. **(F)** Common and unique MARs among the three datasets.

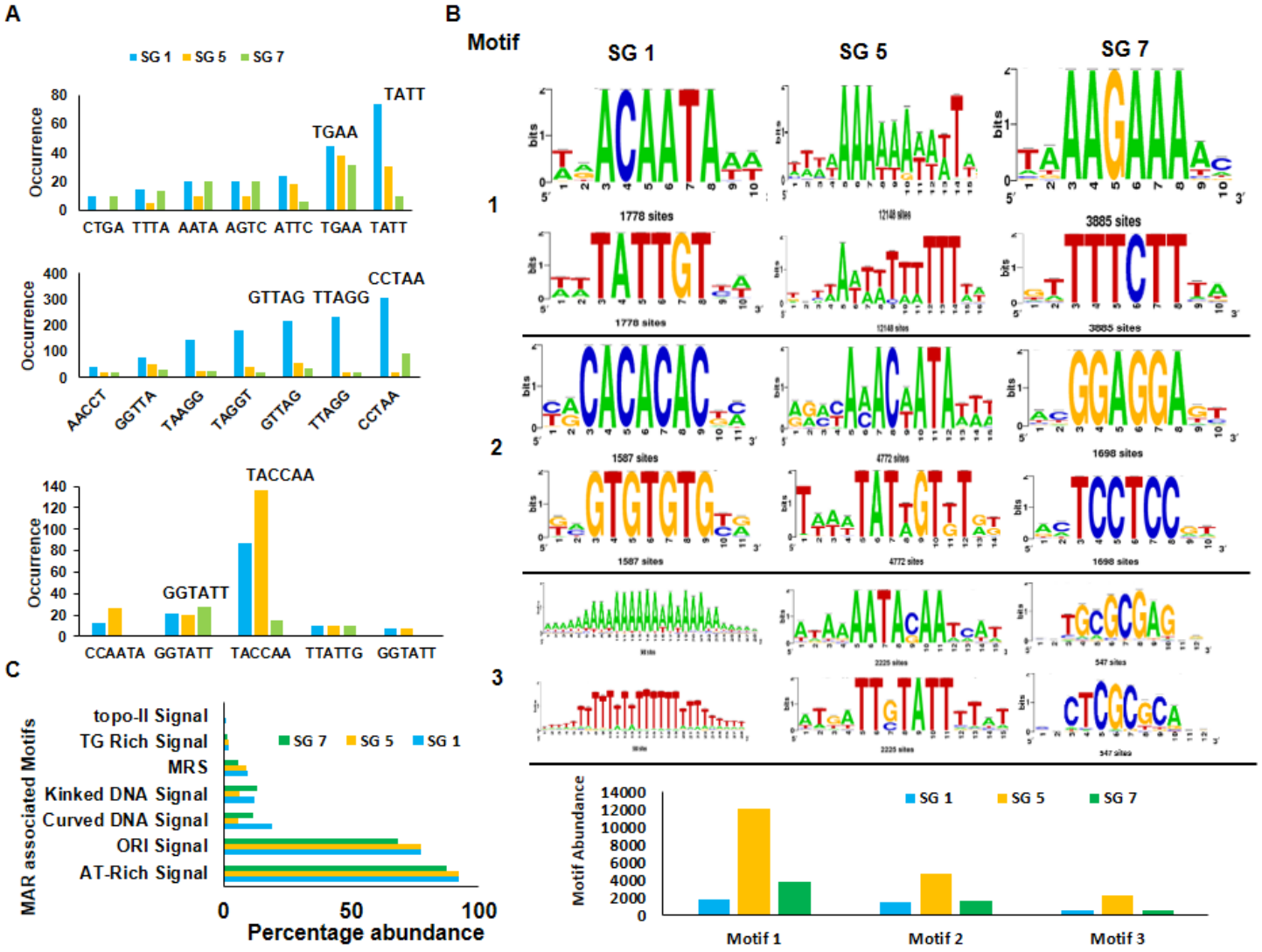

Figure 3

Enrichment of Simple tandem repeats and motifs. **(A)** Tetra, penta, hexa--nucleotide repeats enriched in the datasets. The y-axis shows the occurrence of these repeat signals in the datasets and the x-axis shows the repeat sequences. **(B)** The three most abundantly occurring motifs in the three datasets (above) and their abundance (below). The y-axis shows the number of sequences detected with reference to the specific motifs and the x-axis represents the specific motifs that were found in abundance in the datasets. **(C)** Percentage abundance of each the seven MAR associated motifs in the datasets.

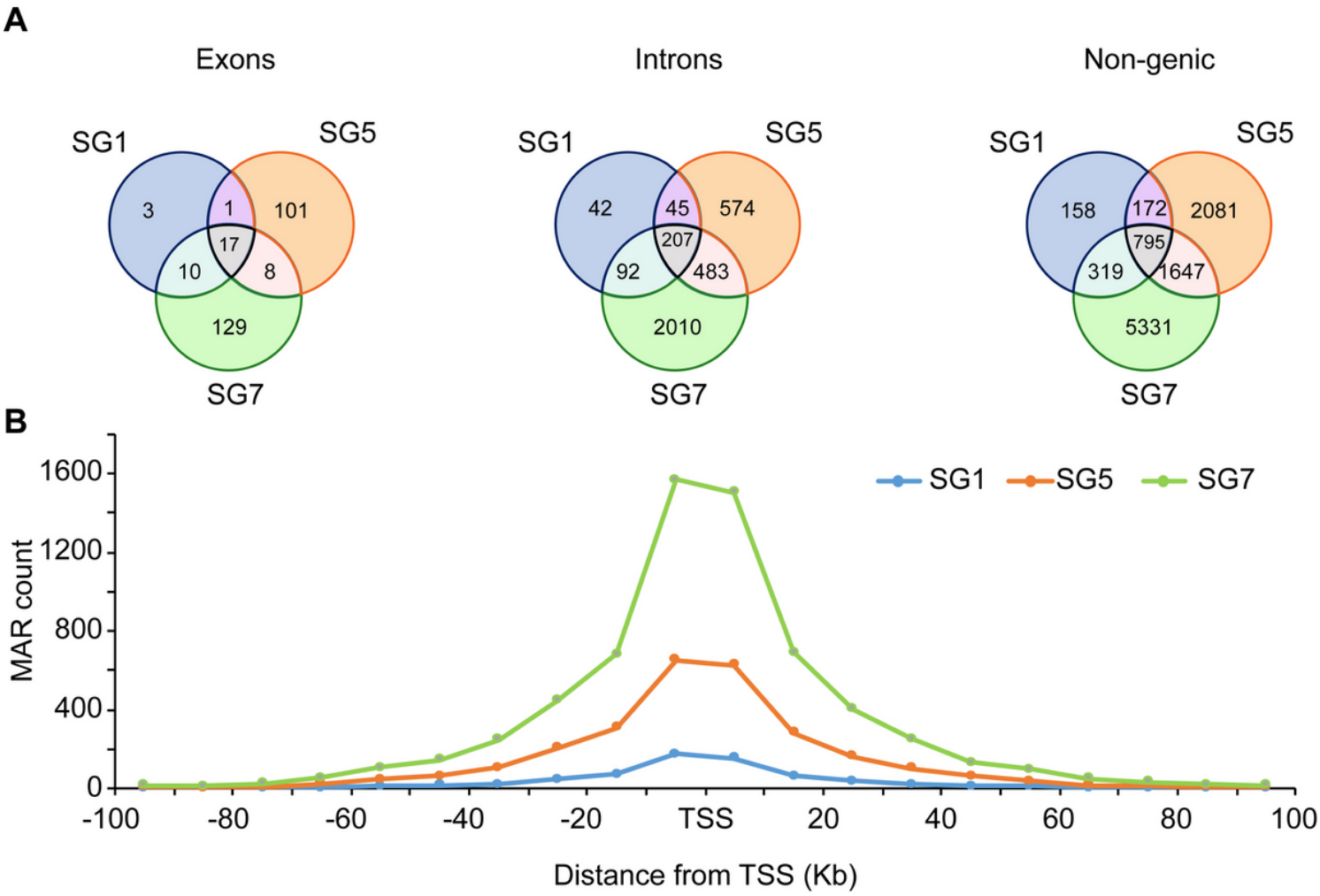

Figure 4

MAR localization in Genomic Elements: The MARs obtained on day 1, day 5 and day 7 were checked for their distribution in *B. mori* genome. **(A)** MARs present in exons, introns and non-genic regions (from left to right) in the three datasets **(B)** Distance of MARs from transcription start site (in the -100kb to +100kb region) is plotted against the MAR count found in these regions among the three datasets.

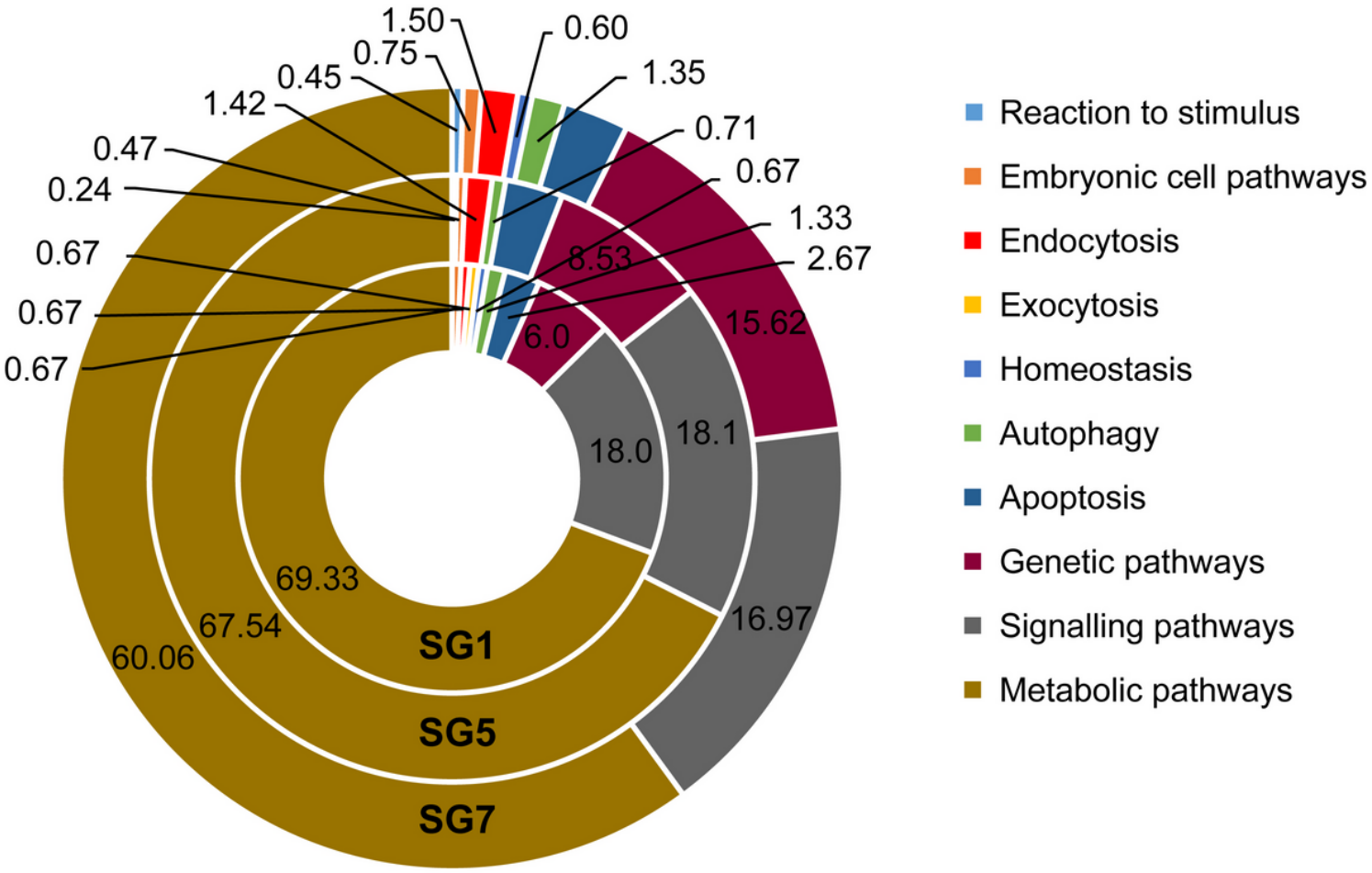

Figure 5

Biological Pathways: The gene list of the annotated reference genome from Silkbase was used to annotate the datasets through gene IDs which were matched against those of *Bombyx mori* database in SGID. Further, the genes found to be associated with the MAR datasets were used to draw a comparison of various biological pathways associated with these genes. The shift in the number of genes associated with the biological pathways during development are possibly due to the chromatin dynamics and MARs responsible for chromatin loop formation.

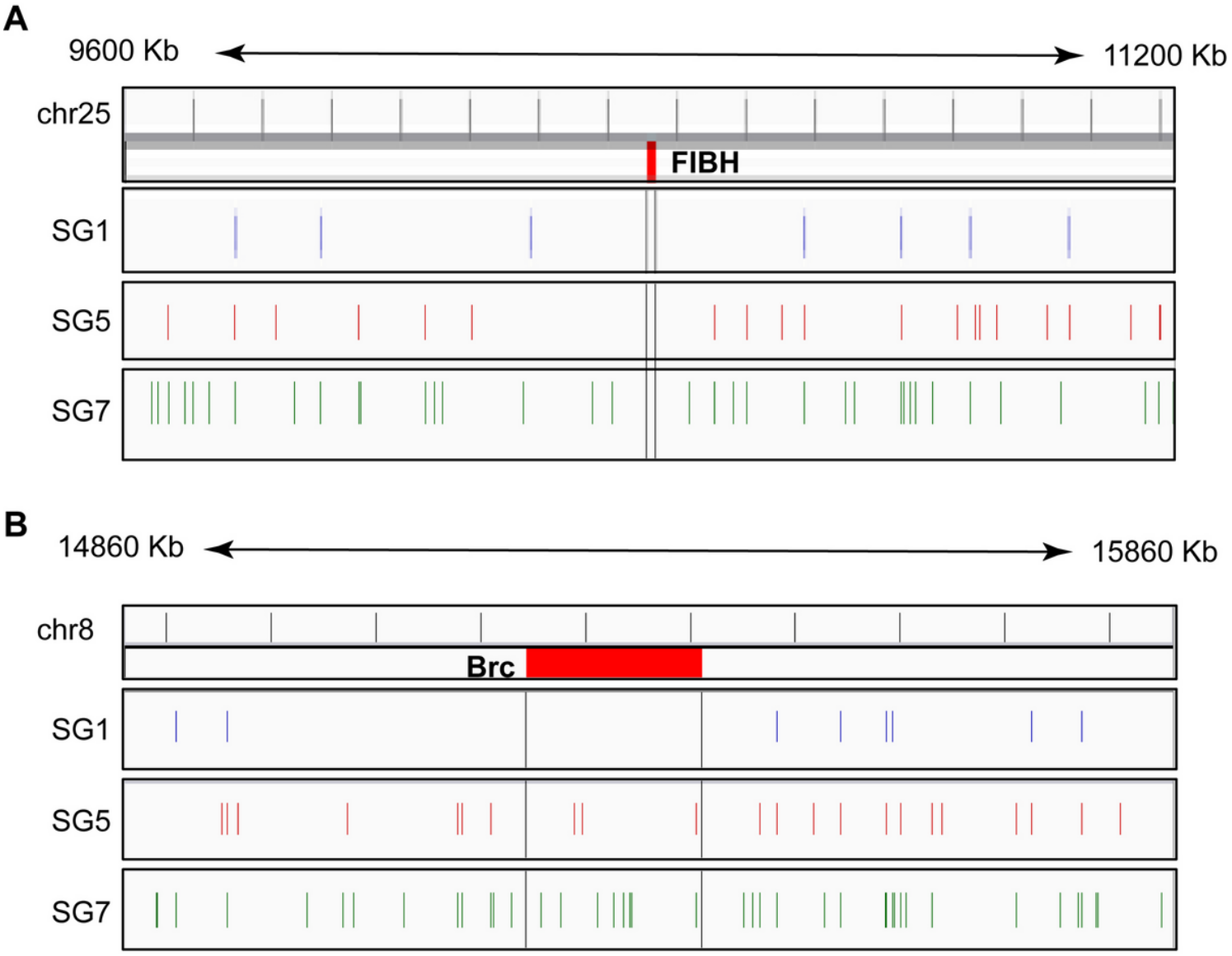

Figure 6

Visualization of MARs flanking FIBH and Br-c across the datasets. (A) The 1600Kb flanking region of FIBH gene (Chr 25). **(B)** The 1000Kb flanking region of Br-c gene (Chr 8).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.pdf