Whole-genome Sequencing of 37 *Bombyx mori* Strains

MGI DNBSEQ sequencing platform enables genome information decryption and germplasm resource protection of *Bombyx mori*

Professor Iksoo Kim's team at Chonnam National University in Korea performed the WGS library construction and sequencing of multiple *Bombyx mori* strains using the MGI library preparation kit and DNBSEQ-G400 Sequencer., The researchers completed the genome variants analysis and germplasm resource survey of 37 breeding line *Bombyx mori* strains established in the 20th century in Korea, which provided data support for the subsequent germplasm resource preservation and directional selection, and also provided reference for other economic species.

This work was published online on April 26, 2022 in *Scientific Data*, a sub-journal of *Nature*, with the title *"Whole-genome sequences of 37 breeding line Bombyx mori strains and their phenotypes established since 1960s"* [1].

Recommended application: Molecular breeding

Recommended model: DNBSEQ-G400RS

- ### Data output is efficient and high-quality

  DNBSEQ sequencing technology has significant features such as high accuracy, low duplication rate, and low index hopping rate, etc.

- ### Providing a full portfolio of experimental processes

  Based on self-developed automated library preparation and efficient automated data analysis software, the whole process portfolio from sample input to report output is realized by optimizing the experimental process and simplifying the manual operation steps of customers.

Background

The *Bombyx mori* was domesticated more than 5,000 years ago[2]. It is a key silk-producing insect and one of the most important economic animals on which many farmers around the world depend for their livelihood. *Bombyx mori* rearing and harvesting silk is a labor-intensive industry. Global silk production keeps declining due to decreasing production in China, so the vast majority of raw silk in the world is produced in India now (https://inserco.org/en/statistics). However, silkworm remains an economically important animal, and some developing countries use it as a new income resource. Beyond textiles, *Bombyx mori* and its by-products play a role in the manufacture of pharmaceuticals, tissue engineering, medical textiles, drug delivery systems, cosmeceuticals, food additives and valuable biomaterials. Therefore, the importance of the *Bombyx mori* as one of the important animal resources is growing[3,4].

Although silk production is declining on general farms in Korea, national research institutions continue to make efforts to ensure available genetic resources by establishing breeding lines for various *Bombyx mori* strains due to the importance of *Bombyx mori* resources. The National Institute of Agricultural Sciences of the Rural Development Administration of Korea (NIAS, RDA, Korea) collected *Bombyx mori* resources with multiple expression traits from the 1960s and developed breeding lines to be used as genetic resources for F1 crosses. According to the purpose of use, lines with different phenotypes can be effectively used to enhance specific phenotypes through artificial selection. These breeding lines are valuable biological resources that can be prepared for unexpected environmental changes (e.g., feeding). Furthermore, the whole-genome sequences of these strains associated with their phenotypes can be used as an important research resource to expand our understanding of the molecular mechanisms of various phenotypes.

Study description

In this study, the authors report the whole-genome sequences of 37 breeding lines of *Bombyx mori* developed in the last 60 years, as well as a description of their phenotypic features. The whole-genome sequencing of some strains was performed on the DNBSEQ-G400 sequencing platform. These whole-genome sequences associated with the phenotypes of established breeding lines could help us understand the *Bombyx mori* genome and provide novel insights into the molecular mechanisms of various phenotypes.

Materials and Methods

A. Sample collection

Thirty-seven *Bombyx mori* strains produced by two-way or three-way crosses were collected from Korea. All strains were selected by self-crossing for at least 10 generations to maintain the specific phenotype. The breeding lines propagates one generation by incubating eggs every year starting in spring and retain the eggs by self-propagation under suitable conditions. Representative males for each breeding line strain were randomly selected and the epidermal tissue was separated during the pupal stage.

B. Library preparation and sequencing

DNA from the epidermal tissue was extracted by the QIAGEN DNeasy Blood and Tissue Kit. Gel electrophoresis was performed to detect DNA fragment size. Trianan, picogreen, and bioanalyzer were used for measuring DNA quality. Sequencing libraries of five triple molt mutant strains (KRSM, SH, HS, S7 and SD) were generated using MGI's MGIEasy DNA Library Preparation Kit following the manufacturer's instructions. The major fragment size of the prepared library was 500 bp, and the paired-end 150 bp (PE150) sequencing was performed using DNBSEQ-G400 sequencing platform. The rest of the strains were sequenced using sequencers from other vendors.

C. Date analysis

The reads filtered by Trimmomatic[5] were aligned to the p50T reference genome from the NCBI Reference Sequence Database by bwa-mem2[6,7] with default settings. Samtools[8] was used for removing PCR duplicates and calling variants. VCFtools[9] was employed to detect biallelic Single Nucleotide Variant (SNV) locus without missing from 38 samples including p50T. SvABA[10] was used to identify InDel and structural variants for each strain. SnpEff was employed to annotate identified SNVs using the custom database information annotated by Refseq. Tassel 5[11] was used to construct the phylogenetic trees with the Neighbor- joining method.

Sample collection	Library preparation and sequencing	Bioinformatics analysis	Result analysis
Thirty-seven *Bombyx mori* strains produced by two-way or three-way crosses and selected for more than 10 generations	 MGIEasy DNA Library Preparation Kit  DNBSEQ-G400 gene sequencer	Trimmomatic bwa-mem2 Samtools SvABA snpEff Tassel5	Phenotypic data collection, variation detection and phylogenetic analysis

Results

Whole-genome sequencing data of *Bombyx mori* strains

The statistical information of whole-genome sequencing data for each *Bombyx mori* strains is shown in Table 1, and the depth of coverage of the sequencing data was at least 30X according to the genome size of the *Bombyx mori* strains (~450 Mb).

Phenotypes and pictures of 37 *Bombyx mori* strains

In answer to the shortages of labor and changes in environment since 1990s after rapid industrialization in Korea, sericulture concerns the breeding strains that could use artificial feed and need less labor, and are easily distinguished by gender through the use of larval markers and cocoon colors. The 37 strains in this study have great values as seed lines applied to establish customized hybrid strains to respond to the environmental changes and the needs of local farmers. pictures of eggs, larvae, cocoons, pupae and adults of the 37 *Bombyx mori* strains covering the entire life cycle were shown in Figure 1.

Variation detection and phylogenetic analysis

Samtools and SvABA were used to identify genomic variation of each strain. A total of 23,478,741 SNVs were detected and 1,506,850 SNVs with variant quality under Q30 and multiallelic loci filtered. Among the filtered 21,971,891 SNVs, 1,327,196 SNVs were located in the CDS region; 1,002,715 (75.551%) SNVs were synonymous variants and 324481 (24.449%) SNVs were non-synonymous variants. SvABA was employed to discover InDel and structural variant on individual strains, and average of 622,531 InDels and 41,348 structural variants were detected. To clarify the evolutionary relationships of 37 breeding lines including P50T, genome-wide variants from sequencing data were used to perform phylogenetic analysis. The phylogenetic relationships between the 37 *Bombyx mori* strains studied in this research and the p50T reference strain were shown in Figure 2. Among the five tri-molt mutant strains, four strains showed close evolutionary relationships except for SH, and some strains showed more closely evolutionary relationships in spite of the external differences.

Strain	Instrument	Read Type	Read Count	Length(bp)	Total Bases (bp)	SRA
KRSM	MGIseq-2000	Paired	199,692,448	150	59,907,734,400	SRR15525308
SH	MGIseq-2000	Paired	57,272,074	150	17,181,622,200	SRR15458431
HS	MGIseq-2000	Paired	52,774,714	150	15,832,414,200	SRR15458432
S7	MGIseq-2000	Paired	59,371,675	150	17,811,502,500	SRR15458433
SD	MGIseq-2000	Paired	49,763,805	150	14,929,141,500	SRR15458430

Table 1. Statistical table of genome sequencing data for the five tri-molt strains of *Bombyx mori*.

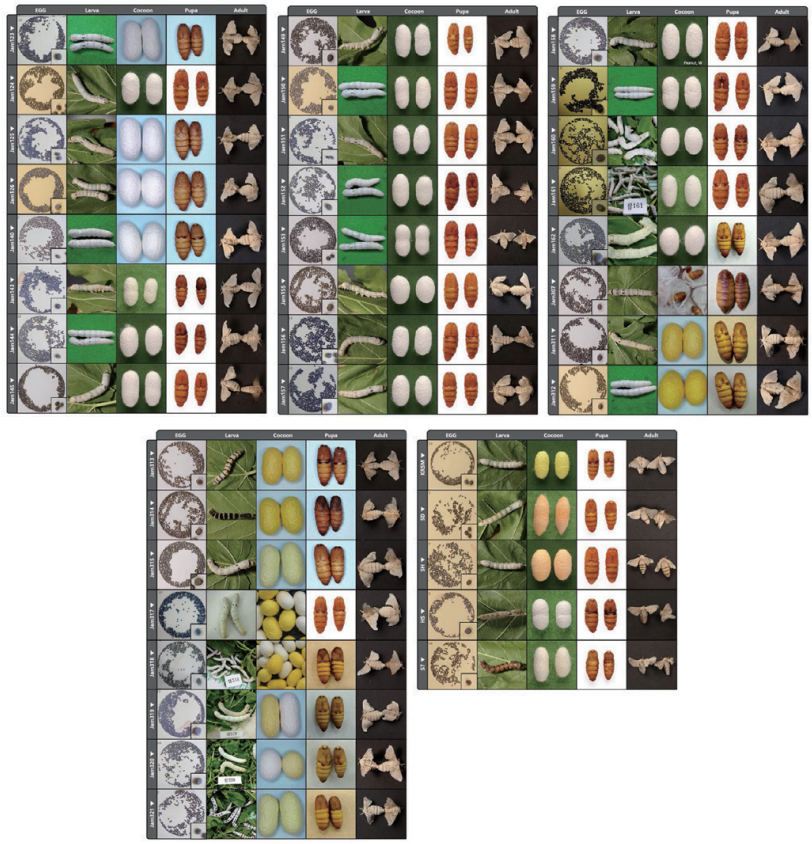

Figure 1. Pictures of eggs, larvae, cocoons, pupae and adults of the 37 *Bombyx mori* strains.

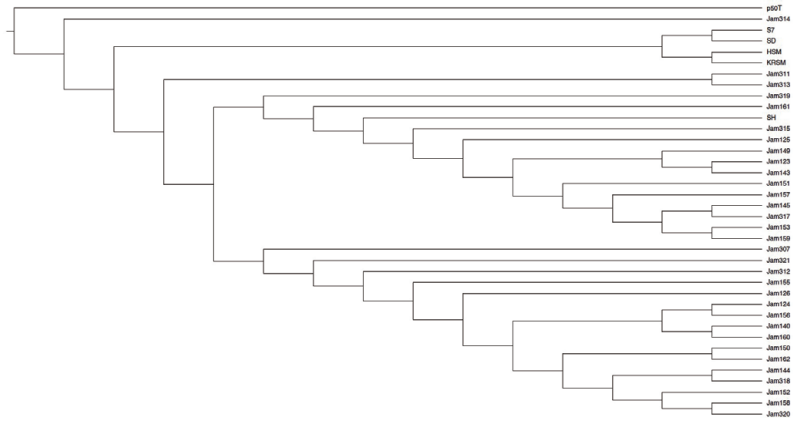

Figure 2. Evolutionary tree of 37 *Bombyx mori* strains

Conclusion

This study released the whole-genome sequences and described the phenotypes of 37 breeding lines of *Bombyx mori* strains constructed in the past 60 years in Korea. The data could provide a valuable resource for further studying the genome and molecular mechanism of phenotypes of the *Bombyx mori*. It provides new ideas and strategies for germplasm resource investigation, and could use as an example for germplasm preservation and utilization of other economics species.

In this study, whole-genome sequencing of some mutant strains was performed based on the DNBSEQ-G400 sequencing platform and related kits developed by MGI. Based on the unique DNBSEQ™ technology, it has the advantages of high accuracy, low duplicate sequence rate and low label jumping rate.

DNBSEQ-G400 Genetic Sequencer

References

1. Kim, SW. Whole-genome sequences of 37 breeding line Bombyx mori strains and their phenotypes established since 1960s. *Sci Data*, 189 (2022).
2. Brock&H. E. J. W. Barber. Prehistoric textiles: The development of cloth in the Neolithic and Bronze Ages.with special reference to the Aegean. *Antiquity*, 66(250), 271-272. (1992).
3. Holland, C. et al. The biomedical use of silk: past, present, future. *Advanced healthcare materials*, 1800465 (2019).
4. Jaiswal, K.K., Banerjee, I. & Mayookha, V. Recent trends in the development and diversification of sericulture natural products for innovative and sustainable applications. *Bioresource Technology Reports*, p. 100614 (2020).
5. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15), 2114–2120 (2014).
6. Kawamoto, M. et al. High-quality genome assembly of the silkworm, Bombyx mori. *Insect Biochemistry and Molecular Biology* 107, 53–62 (2019).
7. Vasimuddin, M. et al. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). *IEEE* 2019.
8. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16), 2078–2079 (2009).
9. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* 27(15), 2156–2158 (2011).
10. Wala, J. A. et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome research* 28(4), 581–591 (2018).
11. Glaubitz, J. C. et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PloS one* 9(2), e90346 (2014).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G400RS	900-000170-00
	MGISP-100RS Automated Sample Preparation System	900-000206-00
	MGISP-960RS Automated Sample Preparation System	900-000146-00
Software	MegaBOLT Bioinformatics analysis accelerator	900-000555-00
Library Prep	MGIEasy Universal DNA Library Prep Set（16 RXN）	1000006985
Sequencing Reagents	DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150)	1000016952

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

Version: August 2023

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Author：Jie Chen, Hao Zhang

Editor-in-Charge：Qiwei Wang

Reviewer：Yao Jiang

1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.
2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.
3. For HotMPS reagents: This sequencing reagent is only available in selected countries.