

37种家蚕品系的全基因组测序

华大智造DNBSEQ测序平台助力家蚕基因组信息破解及种质资源保护

韩国全南大学 Iksoo Kim 教授团队使用华大智造建库试剂盒、MGISEQ-2000测序仪及配套试剂套装完成家蚕品系WGS文库制备、测序及多品系间的变异分析，完成了20世纪韩国育成的家蚕种质资源调查，为后续种质资源保存及定向选育特定品系提供数据支持，也为其他经济物种的保存提供参考。

该工作的成果以题目“Whole-genome sequences of 37 breeding line *Bombyx mori* strains and their phenotypes established since 1960s”于2022年4月26日在线发表在Nature旗下的*Scientific Data*杂志¹。

推荐应用：分子育种

推荐机型：MGISEQ-2000RS

• 高效准确的测序

独有的DNBSEQ™技术具有高准确性、低数据冗余、低标签跳跃，基于此的WGS测序产品能获取更全面的基因组信息。

• 提供实验流程全套产品组合

基于自主研发的自动化建库和高效的自动化数据分析软件，通过优化实验流程简化客户手动操作步骤，实现了从样本输入到报告输出的全流程产品组合。



背景介绍

家蚕在 5000 多年前被驯化²，它是重要的产丝昆虫，也是国际上许多农民维持生计所依赖的重要的经济动物之一。养蚕并获得蚕丝是一个劳动密集型产业，由于中国的蚕丝产量下降导致全球产量持续下降，现印度生丝产量占世界绝大部分 (<https://inserco.org/en/statistics>)。然而，目前家蚕仍然是最重要的经济动物之一，在一些发展中国家被用作新的收入来源。除了在纺织工业中简单地使用家蚕丝作为丝源外，蚕和蚕副产品在药物，组织工程，医用纺织品，药物输送系统，药妆，食品添加剂和有价值的生物材料制造等领域的也发挥着一定的作用。因此，家蚕作为重要动物资源的重要性正在增加^{3,4}。

尽管在韩国一般农场的蚕丝产量正在减少，但由于家蚕资源的重要性，国家研究机构仍不断努力通过为各种家蚕品系建立育种系来确保有可用的遗传资源。韩国农村发展管理局国家农业科学研究所从 20 世纪 60 年代开始收集具有各种表现性状的蚕资源，并建立了将其用作 F1 杂交遗传资源的育种系。根据使用目的，通过额外的选择性育种和杂交，可以有效地利用具有各种表型的品系来增强特定表型。这些育种系是宝贵的生物资源，可以为意外的环境变化(如喂养)做好准备。此外，这些品系与其表型相关的全基因组序列可以用作主要的研究资源，以扩展对家蚕的分子背景的了解。

研究描述

在这项研究中，作者团队报告了过去 60 年中建立的 37 个育种系家蚕的全基因组序列，以及表型特征的描述。其中，部分种系的全基因组测序是在华大智造 MGISEQ-2000 测序平台上完成的。这些与已建立的育种系的表型特征相关的全基因组序列可能是理解家蚕基因组的宝贵资源，并为各种表型的分子背景提供更多的见解。

材料和方法

样品准备

收集韩国国内通过双向或三项杂交培育的37个家蚕品系，所有品系都经历了至少10代以上选育以保证性状的稳定遗传，每年在春季饲养自我繁殖保存品系的延续，在适宜的条件下孵化饲养保证正常发育。

文库制备与测序

在蛹期随机选择每个品系的代表性雄性个体，并在蛹期分离表皮组织，并使用 QIAGEN DNesay 血液和组织试剂盒提取DNA。对提取的DNA进行凝胶电泳以确认DNA片段化，并使用trinean, picogreen, 生物分析仪来检查DNA的质量。其中的5种三蜕皮突变株 (KRSM、SH、HS、S7和SD) 使用华大智造的MGIEasy DNA文库制备试剂盒构建测序文库。

详细操作过程可参考相关说明书。所构建的文库的主峰大小为500bp，利用MGISEQ-2000测序平台进行双端150 bp (PE150) 的测序。其余品系使用的其他厂商测序仪进行测序。

生信流程及亲缘关系分析

使用Trimmomatic⁵去除接头序列和低质量碱基的reads，并将过滤后的reads通过 bwa-mem2程序比对到来自于 NCBI Refseq 数据库的 p50T参考基因组^{6,7}，该步骤使用默认参数。使用samtools删除PCR重复reads和鉴定变异⁸，使用VCFtools提取了包括p50T在内的38个样品中未缺失的双列单核苷酸变体 (SNV) 位点⁹。使用SvABA鉴定每个菌株的InDel和结构变异¹⁰。通过Snpeff使用Refseq注释的自定义数据库信息对已识别的SNP进行注释。通过使用Tassel 5的邻接算法构建了系统发生树¹¹。

结果

基因组测序下机数据

通过对这些家蚕品系进行基因组测序组装，每个家蚕品系的下机数据情况如表 1 所示，基于家蚕 (约 450 Mb) 的基因组大小，生成数据的最小深度覆盖率超过 30X。

样本采集与饲养	文库制备和测序	生信分析	结果分析
双向或三项杂交， 10代以上选育	 MGIEasy DNA 文库制备试剂盒  MGISEQ-2000 测序仪	Trimmomatic bwa-mem2 Samtools SvABA Snpeff Tassel5	形状表型数据、 变异检测及 系统发育分析

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

表1.5个三蜕皮突变家蚕品系的下机数据统计表

性状表型数据与图片

从 20 世纪 90 年代开始，在韩国快速工业化之后，为了应对劳动力短缺和环境变化，家蚕培育的工作重点是培育可以使用人工饲料，需要较少劳动力的品系，并且通过使用标记和茧颜色很容易区分幼虫性别。本研究中报告的 37 种品系作为种质品系具有重要价值，用于开发定制的杂交品系，以应对养蚕环境的变化和当地农民的要求。图 1 显示了来自 37 个家蚕品系的卵，幼虫，茧，蛹和成虫的每张图片。显示了来自 37 个家蚕品系全龄期的卵，幼虫，茧，蛹和成虫的图片。

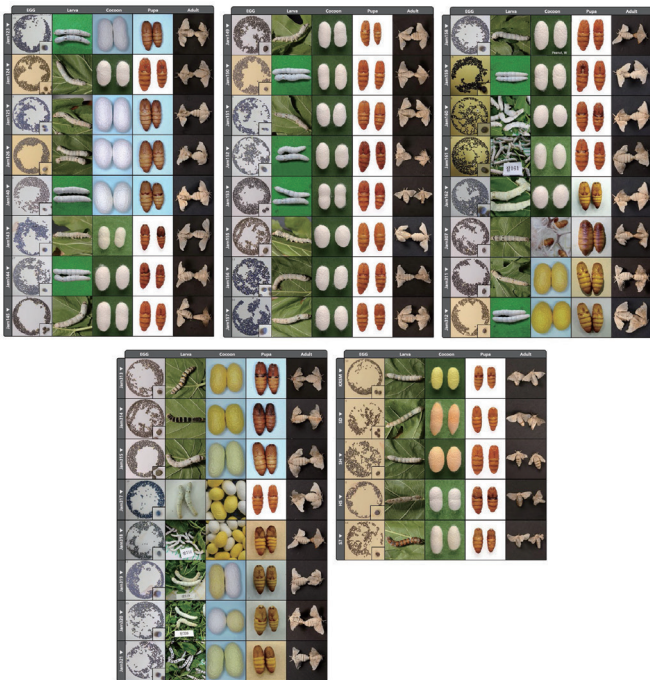


图1.37个家蚕品系全龄期的卵，幼虫，茧，蛹和成虫的图片

变异检测及系统发育分析

使用 samtools 和 SvABA 鉴定每个品系的基因组变异。共鉴定出 23478741 个 SNV，过滤了 1506850 个 SNV (Q30 和多 allelic 位点下的变异质量)。在过滤后的 21971891 个 SNV 中，有 1327196 个 SNV 位于 CDS 区域；1002715 (75.551%) SNV 为同义变异，324481 (24.449%) 为非同义变体。在对单个菌株使用 SvABA 进行 InDel 和结构变体调用中，平均确定了 622531 个 InDel 和 41348 个结构变体。所有变体调用信息都可以在原文方法的链接部分中找到。为了弄清楚包括 p50T 在内的 37 个育种品系的进化关系，使用来自测序数据的全基因组变异进行系统发育分析。图 2 显示了本研究中报道的 37 个家蚕品系与 p50T 参考品系之间的系统发育关系。在 5 个表现出三蜕皮特征的品系中，除 SH 外，4 个品系表现出密切的进化关系，尽管存在一些差异，但部分品系具有更密切的进化关系。

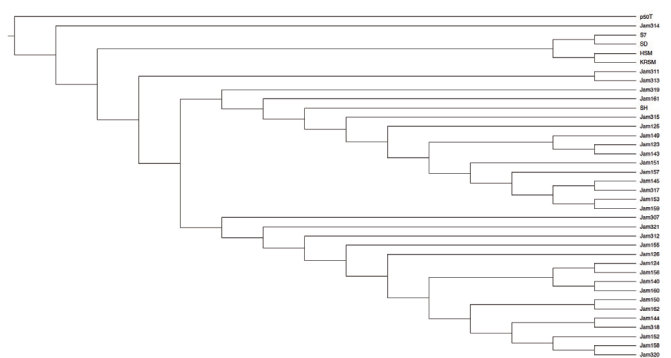


图2.37个家蚕品系的进化树

总结

本研究报告了韩国过去 60 年中建立的 37 个育种系家蚕品系的全基因组序列组装，以及表型特征的描述。这些与已建立的育种系的表型特征相关的全基因组序列可能是理解家蚕基因组的宝贵资源，并为各种表型的分子背景研究提供更多的参考。这为种质资源研究调查提供了新的指导思路与方案，通过家蚕的样本实例为其他经济物种的保存研究与利用提供了完整的方案。

本研究基于华大智造自主研发的 MGISEQ-2000 测序平台及配套试剂盒套装对突变品系进行了全基因组测序，基于独有的 DNBSEQ™ 技术，具有高准确性、低重复序列率、低标签跳跃的优势。



基因测序仪 MGISEQ-2000RS

参考文献

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).

推荐订购信息

产品类型	产品名称	产品货号
仪器	基因测序仪 MGISEQ-2000RS	900-000035-00
	MGISP-100RS 自动化样本制备系统	900-000070-00
	MGISP-960RS 自动化样本制备系统	900-000100-00
软件	MegaBOLT 生信分析加速器(工作站式服务器)	970-000085-00
建库试剂	MGI Easy 通用 DNA 文库制备试剂套装(16 RXN)	1000006985
测序试剂	MGISEQ-2000RS 高通量测序试剂套装(FCL PE150)	1000012555

深圳华大智造科技股份有限公司

深圳市盐田区北山工业区综合楼11栋

4000-688-114

www.mgi-tech.com

MGI-service@mgi-tech.com

股票简称: 华大智造

股票代码: 688114



仅供研究使用

版权声明: 本手册版权属于深圳华大智造科技股份有限公司所有, 未经本公司书面许可, 任何其他个人或组织不得以任何形式将本手册中的各项内容进行复制拷贝、编辑或翻译为其他语言。本手册中所有商标或标识均属于深圳华大智造科技股份有限公司及其提供者所有。

版本: 2023年8月版

撰稿: 陈洁 张镐

责任编辑: 王其伟

审稿: 江遥