

家蚕十五个产丝相关基因的生物信息学分析*

潘尧^{1**} 胡凯文³ 王作奇² 李飞² 梅洋^{2***}

(1. 浙江大学动物科学学院, 动物医学系, 杭州 310058; 2. 浙江大学昆虫科学研究所, 浙江省作物病虫害生物学重点实验室, 杭州 310058; 3. 浙江大学动物科学学院, 动物科学系, 杭州 310058)

摘要 【目的】家蚕 *Bombyx mori* 是重要的资源昆虫, 其主产品蚕丝具有极高的生物学价值与经济价值。目前, 与家蚕产丝调控相关的基因研究较少。为丰富家蚕产丝相关基因的基础数据, 深入挖掘家蚕产丝机制, 本研究开展了家蚕产丝相关基因的生物信息学分析。【方法】本研究整合了前期对 15 个蚕丝产量相关的改良相关基因的注释, 采用转录组分析方法筛选出显著差异表达的家蚕产丝相关基因, 对其进行了染色体定位分析、理化性质分析及亚细胞定位分析、富集分析; 并针对关键家蚕产丝相关基因 *agtpbp1* 与 *lov* 进行了 60 个鳞翅目物种范围的种间基因家族鉴定与进化分析。【结果】结果表明, 蚕丝产量相关的改良基因 *NDRG3*、*PDSS1*、*beta-Spec*、*KNCK1*、*unc-22*、*BR3*、*ZMYND8*、*XNP_1*、*Bin1*、*lov* 在家蚕蛹前期与 5 龄幼虫的前丝腺部位具有显著表达差异, *P* 值分别为 4.523×10^{-6} 、 1.586×10^{-4} 、 5.261×10^{-7} 、 6.210×10^{-3} 、 8.113×10^{-10} 、 2.014×10^{-19} 、 4.858×10^{-2} 、 7.767×10^{-23} 、 3.361×10^{-17} 、 3.896×10^{-4} ; 蚕丝产量相关的改良基因 *beta-Spec*、*XNP_2*、*KNCK1*、*agtpbp1*、*BR3*、*XNP_1*、*Bin1*、*dpyd*、*lov* 在家蚕蛹前期与 5 龄幼虫的中丝腺部位具有显著表达差异, *P* 值分别为 9.518×10^{-8} 、 1.984×10^{-3} 、 4.451×10^{-3} 、 1.536×10^{-2} 、 3.185×10^{-8} 、 1.111×10^{-9} 、 6.770×10^{-30} 、 9.297×10^{-3} 、 4.735×10^{-2} ; 蚕丝产量相关的改良基因 *NDRG3*、*beta-Spec*、*XNP_2*、*KNCK1*、*agtpbp1*、*unc-22*、*BR3*、*lov* 在家蚕蛹前期与 5 龄幼虫的后丝腺部位具有显著表达差异, *P* 值分别为 8.433×10^{-3} 、 1.372×10^{-46} 、 1.059×10^{-10} 、 6.215×10^{-5} 、 1.235×10^{-11} 、 2.491×10^{-3} 、 8.664×10^{-25} 、 9.349×10^{-4} 。这些基因在转录、转录后修饰、翻译及翻译后修饰等多个层面直接或间接影响蚕丝蛋白的合成与储存。分析显示, *agtpbp1* 和 *lov* 种间基因家族分别具有 34 和 37 个基因, 且两个家族分别有 26 和 1 个正选择基因, *agtpbp1* 家族的 *Bbet001799* 等 8 个基因存在正选择位点。【结论】本研究分析了 15 个蚕丝产量相关的改良相关基因, 鉴定了 *agtpbp1* 和 *lov* 两个关键基因的种间基因家族, 完成了其进化分析。

关键词 家蚕; 产丝相关基因; 产丝机制; 进化分析; 基因家族分析

Bioinformatics analysis of fifteen *Bombyx mori* genes related to silk production

PAN Yao^{1**} HU Kai-Wen³ WANG Zuo-Qi² LI Fei² MEI Yang^{2***}

(1. Department of Veterinary Medicine, College of Animal Sciences, Zhejiang University, Hangzhou 310058, China; 2. Zhejiang Provincial Key Laboratory of Crop Disease and Insect Biology, Institute of Insect Sciences, Zhejiang University, Hangzhou 310058, China; 3. Department of Animal Science, College of Animal Sciences, Zhejiang University, Hangzhou 310058, China)

Abstract 【Aim】To provide baseline data on the mechanisms underlying silk production in the silkworm, *Bombyx mori*. 【Methods】This study integrated previous annotations of 15 *B. mori* genes related to silk production and employed transcriptome analysis to identify those that were significantly differentially expressed. Chromosome mapping, physicochemical property analysis, subcellular localization, and enrichment analysis were then used to elucidate the function

*资助项目 Supported project: 浙江省自然科学基金重点项目 (LZ23C140002)

**第一作者 First author, E-mail: 3200101597@zju.edu.cn

***通讯作者 Corresponding author, E-mail: meiyang12@zju.edu.cn

收稿日期 Received: 2024-06-03; 接受日期 Accepted: 2024-10-25

of these genes. In addition, interspecies gene family identification and a phylogenetic analysis were performed on the key silk production related genes *agtpbp1* and *lov* of 60 lepidopteran species. **[Results]** Silk production related genes *NDRG3*, *PDSS1*, *beta-Spec*, *KNCK1*, *unc-22*, *BR3*, *ZMYND8*, *XNP_1*, *Bin1* and *lov*, were significantly differentially expressed in anterior silk gland of *B. mori* during the dispersal stage compared to 5th instar larva, respectively, with the *P*-values of 4.523×10^{-6} , 1.586×10^{-4} , 5.261×10^{-7} , 6.210×10^{-3} , 8.113×10^{-10} , 2.014×10^{-19} , 4.858×10^{-2} , 7.767×10^{-23} , 3.361×10^{-17} , 3.896×10^{-4} ; *beta-Spec*, *XNP_2*, *KNCK1*, *agtpbp1*, *BR3*, *XNP_1*, *Bin1*, *dpyd* and *lov*, were significantly differentially expressed in middle silk gland of *B. mori* during the dispersal stage compared to 5th instar larva, respectively, with the *P*-values of 9.518×10^{-8} , 1.984×10^{-3} , 4.451×10^{-3} , 1.536×10^{-2} , 3.185×10^{-8} , 1.111×10^{-9} , 6.770×10^{-30} , 9.297×10^{-3} , 4.735×10^{-2} ; *NDRG3*, *beta-Spec*, *XNP_2*, *KNCK1*, *agtpbp1*, *unc-22*, *BR3* and *lov*, were significantly differentially expressed in posterior silk gland of *B. mori* during the dispersal stage compared to 5th instar larva, respectively, with the *P*-values of 8.433×10^{-3} , 1.372×10^{-46} , 1.059×10^{-10} , 6.215×10^{-5} , 1.235×10^{-11} , 2.491×10^{-3} , 8.664×10^{-25} , 9.349×10^{-4} . These genes directly, or indirectly, influence silk protein synthesis and storage at multiple levels, including transcription, post-transcriptional modification, translation and post-translational modification. The interspecies gene families *agtpbp1* and *lov* contain 34 and 37 genes, respectively, with 26 and 1 positively selected gene in each family. Eight genes, including *Bbet001799* in the *agtpbp1* family have positively selected sites. **[Conclusion]** These results shed light on the function of 15 silk production related genes and the interspecies gene families *agtpbp1* and *lov*.

Key words *Bombyx mori*; silk production related gene; silk production mechanism; phylogenetic analysis; gene family analysis

家蚕 *Bombyx mori* 是具有极高生物学价值与经济价值的农业昆虫。在生命科学领域, 家蚕以其低廉的繁殖成本、丰富的后代数量、短暂的世代周期以及清晰的遗传背景等特点脱颖而出(常怀普, 2017)。家蚕的主产品蚕丝是一种卓越天然生物材料, 广泛应用于高档消费品如衣物、棉被和饰品的制造。此外, 蚕丝也可以作为缝合线等医用耗材, 在医学领域发挥重要作用(王少华, 2017)。

目前, 对家蚕产丝相关基因的鉴定与功能研究已取得进展。研究表明, 蚕丝的核心成分是丝蛋白和丝胶蛋白, 其合成部位为家蚕的丝腺(Aramwit *et al.*, 2012), 而丝蛋白、丝胶蛋白的合成与分泌受到多个基因的调控与协同作用(王少华, 2017; Dong *et al.*, 2019)。近 15 年来, *Ras1CA*、*Bmovo-1*、*beta-1, 4-N-acetylglucosaminidase 1* 和 *Vps13d* 等蚕丝蛋白合成调控基因的功能及其表达调控机制已被揭示(李静芝, 2013; 宋作伟, 2013; 李娟, 2017; Li *et al.*, 2020; 龙安星, 2022; Zhao *et al.*, 2022); 且 *Antennapedia*、*BmSPI* 和 *BmEckL1* 等发育调控基因被证明与产丝相关(Tsubota *et al.*, 2016; Liu *et al.*, 2023; Li *et al.*, 2024)。此外, 稳定高效的家蚕转基因技术体系的发展(Wang *et al.*, 2013; Long

et al., 2016; Wei *et al.*, 2022; Long *et al.*, 2024) 显著推动了外源基因或内源编辑基因对产丝性能影响的研究与家蚕生物反应器的开发(Xu *et al.*, 2019; Zhang *et al.*, 2022)。随着测序技术的发展, 通过多组学与单细胞技术解析产丝基因表达动态的研究涌现(Ruan *et al.*, 2021)。2022 年, 单细胞分辨率的家蚕丝腺转录图谱建立, 完成了家蚕丝腺中的多种主要细胞类群的鉴定, 并揭示了其在蚕丝蛋白合成与液态丝纤维化等过程中的基因表达差异(Ma *et al.*, 2022)。上述研究加深了对家蚕产丝机制的理解, 但当前针对家蚕产丝关键基因的研究仍显不足。近 20 年来, 由于缺乏对产丝关键基因的深入研究, 我国对家蚕的品种改良与品种选育进展相对缓慢(李娟, 2017)。因此, 深入研究家蚕基因组以及与蚕丝生产相关基因, 对于筛选优良家蚕品种, 并推动家蚕育种工作的顺利进展具有重大意义。

对于家蚕基因组的研究已经持续近 30 年。1995 年, Nagaraja 和 Nagaraju (1995) 利用随机扩增多态性 DNA (The random amplified polymorphic DNA, RAPD) 技术对 13 个家蚕基因型进行了 DNA 图谱分析。2004 年, 首个家蚕品系 p50T 完成了基因组组装(Mita *et al.*, 2004), 为家蚕的遗传分析提供了数据支持(Xia *et al.*,

2004)。Kawamoto 等(2019)使用 PacBio 长读数和 BAC 序列方法组装了 p50T 的基因组序列,提供了一个高度连续的家蚕基因组。2019 年,家蚕组学数据库 SilkDB 更新为 SilkDB 3.0 版本,提供了家蚕基因组相关的全面信息(Wang *et al.*, 2005; Lu *et al.*, 2020)。2022 年,InsectBase 2.0 数据库建立,系统性总结了家蚕及其亲缘物种的基因组信息(Mei *et al.*, 2022)。上述研究为家蚕基因组研究提供了广泛可靠的数据基础。随着家蚕基因组研究的不断深入,通过基因家族分析,并进一步探究家蚕产丝相关基因的生理功能成为可能。

本研究利用公共数据资源,筛选家蚕产丝相关基因,并深入开展这些基因的染色体定位分析、理化性质分析、亚细胞定位分析、功能预测以及跨物种的基因家族分析。旨在为家蚕产丝相关基因的分子机制研究提供扎实的基础数据,有助于更深入地理解家蚕产丝的分子机制,为培育出更高产、更优质的家蚕品种奠定坚实的基础,从而为家蚕育种工作提供有力的理论依据和新的研究方向。

1 材料与方法

1.1 15 个家蚕产丝相关的改良相关基因(IAGs)的注释整理

龙安星(2022)通过选择性消除分析,在中、日系家蚕实用种品系中,鉴定了 15 个与蚕丝产量相关的改良相关基因(Improvement-associated genes, IAGs)。基于该研究数据,利用 15 个 IAGs 的基因 ID,在 SilkDB 3.0 数据库(Lu *et al.*, 2020)中检索获取 15 个 IAGs 的全长核酸序列与蛋白质编码序列,整合龙安星(2022)与 SilkDB 3.0 数据库对 15 个 IAGs 的所有可变剪接体的注释。

1.2 15 个 IAGs 的差异表达分析

在 NCBI 数据库(Benson *et al.*, 2018)中,检索获得 5 龄第 3 日与蛹前期(Dispersal stage 或 wandering stage)家蚕幼虫的前丝腺(Anterior silk gland)、中丝腺(Middle silk gland)、后丝腺(Posterior silk gland)的转录组数据,共 18 个样本,其中 2 个龄期的 3 种丝腺转录组数据均

包含 3 个重复(表 1);在 SilkDB 3.0 数据库获取家蚕基因组注释文件“BmDZ.v3.6.gff”。基于上述转录组数据与基因组注释文件,利用 RSEM 软件(Version 1.3.3)(Li and Dewey, 2011)对 15 个 IAGs 进行表达量计算,获得 2 个龄期的 3 种丝腺内的家蚕基因表达矩阵,共 6 个。利用 DESeq2(Love *et al.*, 2014),基于同种丝腺在 2 个龄期时的表达矩阵进行差异表达分析,分别筛选 3 种丝腺在 2 个龄期间显著差异表达的 IAGs;对 3 种丝腺内存在龄期间显著表达差异的 IAGs 取并集,最终获得 12 个 IAGs。

1.3 12 个 IAGs 的理化分析与功能分析

在 SilkDB 3.0 数据库中检索获取 12 个经筛选 IAGs 的染色体定位信息,并制作染色体定位信息文件;利用 RIdeogram 包(Hao *et al.*, 2020)进行可视化,绘制染色体定位图谱。其次,利用 12 个 IAGs 的蛋白质编码序列,分别使用 Expasy(Artimo *et al.*, 2012)、Cell-PLoc 2.0(Chou and Shen, 2008)和 WoLF PSORT(Horton *et al.*, 2007)平台分析其理化性质和亚细胞定位情况。最后,在 SilkDB 3.0 数据库获取家蚕基因组功能注释文件“annotation.txt”。以此为基础,利用 Python 脚本分别制作家蚕基因组的 KEGG 与 GO 背景注释文件;汇总 12 个 IAGs 的基因 ID,制作 KEGG 与 GO 前景注释文件。利用 TBtools 软件(Version 2.042)(Chen *et al.*, 2023)的“GO enrichment”和“KEGG enrichment analyse”功能分别完成 KEGG 分析与 GO 分析。

1.4 2 个关键 IAGs 的种间基因家族鉴定

在 InsectBase 2.0 数据库(Mei *et al.*, 2022)中,筛选获得隐金夜蛾 *Abrostola tripartita* 等 60 个高质量的鳞翅目物种基因组。利用上述基因组数据,通过 BLAST 软件(Version 2.15.0)的 makeblastdb 功能建立 60 个物种的 CDS 数据库和蛋白数据库。对 12 个 IAGs 的差异表达分析与理化、功能分析结果表明, *agtpbp1* 和 *lov* 两个 IAGs 分别在 2 个、3 个丝腺部位存在显著性表达差异,深入分析后发现其可能在产丝过程中发挥关键功能;同时,当前与 *agtpbp1* 和 *lov* 相

表 1 18 个转录组样本明细
Table 1 Details of 18 transcriptome samples

转录组号 Transcriptome ID	组织类型 Organizational type	取样龄期 Sampling instar
SRR10035756	前丝腺 Anterior silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035757	前丝腺 Anterior silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035758	前丝腺 Anterior silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035669	中丝腺 Middle silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035670	中丝腺 Middle silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035671	中丝腺 Middle silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035619	后丝腺 Posterior silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035620	后丝腺 Posterior silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035621	后丝腺 Posterior silk gland	5 龄第 3 日幼虫 3rd day of 5th instar larva
SRR10035749	前丝腺 Anterior silk gland	蛹前期 Dispersal stage
SRR10035750	前丝腺 Anterior silk gland	蛹前期 Dispersal stage
SRR10035751	前丝腺 Anterior silk gland	蛹前期 Dispersal stage
SRR10035662	中丝腺 Middle silk gland	蛹前期 Dispersal stage
SRR10035663	中丝腺 Middle silk gland	蛹前期 Dispersal stage
SRR10035664	中丝腺 Middle silk gland	蛹前期 Dispersal stage
SRR10035612	后丝腺 Posterior silk gland	蛹前期 Dispersal stage
SRR10035614	后丝腺 Posterior silk gland	蛹前期 Dispersal stage
SRR10035615	后丝腺 Posterior silk gland	蛹前期 Dispersal stage

关的研究较少,具有较高的研究价值。因此,利用 *agtpbp1* 和 *lov* 的核酸序列与蛋白序列,分别通过 BLAST 软件的 blastn 与 blastp 功能检索 $e\text{-value} < 10^{-5}$ 的核酸序列与蛋白序列(其他参数设置为默认),两者取交集,筛选出 *agtpbp1* 和 *lov* 基因家族的物种间初步候选基因。

利用 Interpro 数据库 (Paysan-Lafosse *et al.*, 2022) 筛选上述候选基因中具有相似结构域的序列,随后利用 TBTools 软件 (Chen *et al.*, 2023) 的 “Simple HMM search” 功能对上述序列进行验证。最后利用 MAFFT 软件 (Version 7.402) (Katoh and Standley, 2013) 进行多序列比对,去除假阳性基因,获得最终候选基因。

1.5 2 个关键 IAGs 基因家族的系统发育、保守基序及保守结构域分析

1.5.1 系统发育分析 使用 MAFFT 软件 (Katoh and Standley, 2013) 分别对两个基因家族的候选

基因蛋白序列进行多序列比对 (参数设置为 - auto), 随后利用 trimAl 软件 (version 3) (Capella-Gutiérrez *et al.*, 2009) 对多序列比对结果进行裁剪 (参数设置为 -automated1)。随后,依托 TBtools 软件 (Chen *et al.*, 2023) 的 “IQ-Tree Wrapper” 功能,利用最大似然法 (Maximum likelihood, ML) 构建系统发育树。最终,利用 iTOL 网站 (<https://itol.embl.de/>) (Letunic and Bork, 2021) 完成系统发育树的可视化。

1.5.2 保守基序与保守结构域分析 利用 MEME 平台 (Version 5.5.5) (Bailey *et al.*, 2015) 分析 *agtpbp1* 和 *lov* 基因家族的保守基序 (保守基序个数设置为 10, 其他参数设置为默认), 获得两个基因家族的保守基序分布信息与序列保守性标识; 利用 NCBI 的 Batch CD-Search 功能分析两个基因家族的保守结构域, 获得其保守结构域分布信息。利用 TBtools 软件 (Chen *et al.*, 2023) 的 “BioSequence structure illustrator” 功

能可视化保守基序与保守结构域。

1.6 2 个关键 IAGs 基因家族的选择压力分析与正选择位点分析

利用两个家族基因的多序列联配结果, 使用 Codeml 程序以及 PAML 程序包 (Version 4.9j) (Yang, 2007) 对每个同源基因的氨基酸非同义替换 (K_a) 与同义替换 (K_s) 的平均比值 (ω) 进行估算 (参数 Model=0, 其余参数设置为默认), K_a/K_s 值大于 1 的基因被认为是正选择基因。随后, 利用 Datamonkey 平台 (Weaver *et al.*, 2018) 的 RELAX 功能计算两个基因家族中多拷贝基因的 K 值及显著性, 以评估基因拷贝数对自然选择偏好性的影响。RELAX 功能可估算多拷贝基因所受自然选择严格程度的变化趋势, $K>1$ 表示选择强度加强, $K<1$ 表示选择强度降低。最后, 筛选 *agtpbp1* 和 *lov* 基因家族中 K_a/K_s 值大于 1 的基因, 使用 Codeml 程序以及 PAML 程序包 (Yang, 2007) 分析其正选择位点 (参数 Model=2, 其余参数设置为默认)。

2 结果与分析

2.1 15 个 IAGs 的注释整理结果

基于前期研究, 分别对 15 个家蚕产丝相关的 IAGs 进行注释 (表 2)。特别地, *agtpbp1* 为胞质羧肽酶 2 或逆转录酶的编码基因, 其羧肽酶结构域具有酶学活性; *BR3* 为 balbiani 环状蛋白 3 异构体 X2 的编码基因, 具有 Fe^{2+} 代谢功能; *lov* 为含锌指和 BTB 结构域的蛋白 42 的编码基因, 其锌指结构与 BTB 结构域与发育调控相关; *XNP* 为转录因子 ARTX 同源物异构体 X4 的编码基因, 具有转录调节功能 (*XNP_1* 与 *XNP_2* 为具有显著性表达差异的 *XNP* 基因可变剪接体)。15 个 IAGs 在 SlikDB 数据库中的基因 ID 见表 2。

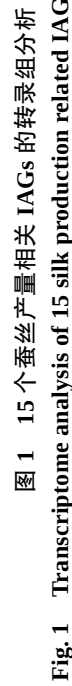
2.2 15 个 IAGs 的差异表达分析结果

利用 RSEM 软件与家蚕幼虫的转录组数据, 对 15 个 IAGs 进行差异表达分析。结果表明, 在家蚕的蛹前期和 5 龄之间, 12 个 IAGs 在丝腺部位具有显著表达差异 (图 1: A1-C2)。12 个

IAGs 在 SlikDB 数据库中的基因 ID 如下: KWMTBOMO02872、KWMTBOMO11045、KWMTBOMO15351、KWMTBOMO13706、KWMTBOMO10489、KWMTBOMO13565、KWMTBOMO13404、KWMTBOMO00182、KWMTBOMO05047、KWMTBOMO16029、KWMTBOMO11059 和 KWMTBOMO09152。如图 1, 蚕丝产量相关的改良基因 *NDRG3*、*PDSS1*、*beta-Spec*、*KNCK1*、*unc-22*、*BR3*、*ZMYND8*、*XNP_1*、*Bin1*、*lov* 在家蚕蛹前期与 5 龄幼虫的前丝腺部位具有显著表达差异: 相对于蛹前期, 上述基因在 5 龄第 3 日的表达情况分别为上调、上调、上调、上调、上调、上调、下调、下调、下调、下调, P 值分别为 $4.523\ 2\times 10^{-6}$ 、 $1.586\ 6\times 10^{-4}$ 、 $5.261\ 5\times 10^{-7}$ 、 $6.210\ 8\times 10^{-3}$ 、 $8.113\ 4\times 10^{-10}$ 、 $2.014\ 5\times 10^{-19}$ 、 $4.858\ 2\times 10^{-2}$ 、 $7.767\ 6\times 10^{-23}$ 、 $3.361\ 3\times 10^{-17}$ 、 $3.896\ 7\times 10^{-4}$; 蚕丝产量相关的改良基因 *beta-Spec*、*XNP_2*、*KNCK1*、*agtpbp1*、*BR3*、*XNP_1*、*Bin1*、*dpyd*、*lov* 在家蚕蛹前期与 5 龄幼虫的中丝腺部位具有显著表达差异: 相对于蛹前期, 上述基因在 5 龄第 3 日的表达情况分别为上调、上调、上调、上调、上调、上调、下调、下调、下调、下调, P 值分别为 $9.518\ 4\times 10^{-8}$ 、 $1.984\ 3\times 10^{-3}$ 、 $4.451\ 3\times 10^{-3}$ 、 $1.536\ 7\times 10^{-2}$ 、 $3.185\ 9\times 10^{-8}$ 、 $1.111\ 6\times 10^{-9}$ 、 $6.770\ 7\times 10^{-30}$ 、 $9.297\ 5\times 10^{-3}$ 、 $4.735\ 3\times 10^{-2}$; 蚕丝产量相关的改良基因 *NDRG3*、*beta-Spec*、*XNP_2*、*KNCK1*、*agtpbp1*、*unc-22*、*BR3*、*lov* 在家蚕蛹前期与 5 龄幼虫的后丝腺部位具有显著表达差异: 相对于蛹前期, 上述基因在 5 龄第 3 日的表达情况分别为上调、上调、上调、上调、上调、上调、上调、下调, P 值分别为 $8.433\ 2\times 10^{-3}$ 、 $1.372\ 5\times 10^{-46}$ 、 $1.059\ 8\times 10^{-10}$ 、 $6.215\ 6\times 10^{-5}$ 、 $1.235\ 0\times 10^{-11}$ 、 $2.491\ 7\times 10^{-3}$ 、 $8.664\ 0\times 10^{-25}$ 、 $9.349\ 1\times 10^{-4}$ 。特别地, 鉴于 *lov* 基因内嵌的 BTB/POZ 结构域具备介导泛素化过程的能力 (Geyer *et al.*, 2003), 该基因可能参与丝蛋白的翻译后修饰与降解过程, 调控丝蛋白的稳态。*XNP_1* 和 *XNP_2* 是转录因子 ATRX 的同源物 (Lu *et al.*, 2020), 在家蚕不同的丝腺部位展现出不同的表达模式。

表 2 15 个蚕丝产量相关 IAGs 的注释整合
Table 2 Annotation integration of 15 silk production related IAGs in *Bombyx mori*

基因名称 Gene name	基因 ID Gene ID	染色体定位 Chromosomal localization	起始位置 (bp) Start position (bp)	终止位置 (bp) End position (bp)	结构与功能注释 Structure and function annotation
<i>ABCF3</i>	KWMTBOMO05851	Chr10	12 685 275	1 2700 530	ATP-binding cassette sub-family F member 3
<i>agtpbp1</i>	KWMTBOMO02872	Chr5	13 283 692	13 293 373	Reverse transcriptase
<i>beta-Spec</i>	KWMTBOMO11045	Chr18	11 561 068	11 591 244	Spectrin beta chain isoform X3
<i>BR3</i>	KWMTBOMO15351	Chr25	10 010 188	10 029 127	Balbani ring protein 3 isoform X2
<i>Bin1</i>	KWMTBOMO13706	Chr23	7 302 306	7 303 671	Histone deacetylase complex subunit SAP18 [<i>Papilio machaon</i>]
<i>ct</i>	KWMTBOMO09977	Chr16	14 329 754	14 335 297	Homeobox protein cut isoform X2
<i>dpyd</i>	KWMTBOMO10489	Chr17	14 372 520	14 396 064	Dihydropyrimidine dehydrogenase [DP(+)] [<i>Papilio machaon</i>]
<i>KNCK1</i>	KWMTBOMO13565	Chr23	1 601 018	1 610 064	Potassium channel subfamily K member 1-like [<i>Amyelois transitella</i>]
<i>lov</i>	KWMTBOMO13404	Chr22	16 143 339	16 158 347	XP 004924336.2 uncharacterized protein LOC101736836 isoform X1
<i>NDRG3</i>	KWMTBOMO00182	Chr1	6 064 244	6 065 275	Ndr family
<i>PDSSI</i>	KWMTBOMO05047	Chr9	6 507 961	6 515 395	Decaprenyl-diphosphate synthase subunit 1 isoform X1 [<i>Papilio xuthus</i>]
<i>unc-22</i>	KWMTBOMO16029	Chr27	6 751 739	6 805 883	Hypothetical protein KGM 12177 [<i>Danaus plexippus</i>]
<i>XNP_1</i>	KWMTBOMO11059	Chr18	12 177 130	12 209 076	Transcriptional regulator ATRX homolog [<i>Amyelois transitella</i>]
<i>XNP_2</i>	KWMTBOMO11059	Chr18	12 209 211	12 212 087	-
<i>ZMYND8</i>	KWMTBOMO09152	Chr15	12 272 765	12 298 194	Uncharacterized protein LOC101739573 isoform X2
-	KWMTBOMO00439	Chr1	13 198 399	13 200 280	Uncharacterized protein LOC101737920



A1、A2. 家蚕前丝腺部位蚕丝产量相关 IAGs 在蛹前期与 5 龄幼虫两龄期间的表达差异; B1、B2. 家蚕中丝腺部位蚕丝产量相关 IAGs 在蛹前期与 5 龄幼虫两龄期间的表达差异; C1、C2. 家蚕后丝腺部位蚕丝产量相关 IAGs 在蛹前期与 5 龄幼虫两龄期间的表达差异。stage 5_1-3 表示 5 龄幼虫第 1-3 个样本。

A1、A2. Expression significance of silk production related IAGs between dispersal stage and the 5th instar larvae of *B. mori* in anterior silk gland; B1, B2. Expression significance of silk production related IAGs between dispersal stage and 5th instar larvae of *B. mori* in middle silk gland; C1, C2. Expression significance of silk production related IAGs between dispersal stage and 5th instar larvae of *B. mori* in posterior silk gland. stage 5_1-3 indicates 1st-3rd sample of 5th larva.

XNP_1 在前、中丝腺显著下调表达 (P 值分别为 $7.767\ 6\times10^{-23}$ 、 $1.111\ 6\times10^{-9}$)，*XNP_2* 则在中、后丝腺显著上调表达 (P 值分别为 $1.984\ 3\times10^{-3}$ 、 $1.059\ 8\times10^{-10}$)。此外，根据 SilkDB 3.0 数据库的预测，*agtpbp1* 具有胞质羧肽酶结构 (Lu *et al.*, 2020)，结合该基因在中、后丝腺上调表达的观测结果，*agtpbp1* 可能参与丝蛋白与丝胶蛋白一级结构的形成。SilkDB 3.0 数据库预测 *BR3* 具有 Fe^{2+} 代谢相关的缺氧诱导因子脯氨酰羟化酶结构 (Lu *et al.*, 2020)。鉴于离子代谢对丝蛋白

结构转化的影响已被广泛研究 (Chang *et al.*, 2015)，*BR3* 可能通过调控 Fe^{2+} 代谢参与丝蛋白高级结构的形成。

2.3 12 个 IAGs 的染色体定位、理化性质与亚细胞定位分析

利用 RIdrogram 进行基因定位可视化，显示 12 个 IAGs 无规则分布在基因组中 (图 2)。分别利用 Expasy、Cell-PLoc 2.0 和 WoLF PSORT 完成理化性质分析与亚细胞定位 (表 3)。其结

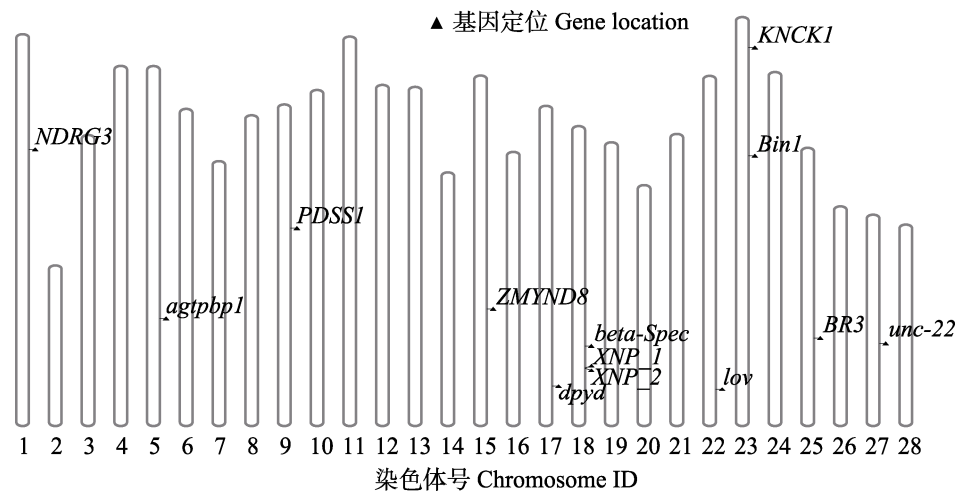


图 2 12 个差异表达 IAGs 的染色体定位图
Fig. 2 Chromosomal localization of 12 differentially expressed IAGs

表 3 12 个 IAGs 的理化性质与亚细胞定位
Table 3 Physical, chemical properties and sub-cellular localization of 12 IAGs

基因名称 Gene name	氨基酸个数 Number of amino acids	分子质量 (D) Molecular weight (D)	等电点 Isoelectric point	亚细胞定位 Sub-cellular localization
<i>agtpbp1</i>	706	79 522.34	8.12	细胞核 Nucleus
<i>beta-Spec</i>	63	7 248.19	6.54	细胞膜/细胞骨架 Cell membrane / Cytoskeleton
<i>BR3</i>	456	51 269.21	8.99	细胞质 Cytoplasm
<i>Bin1</i>	157	18 100.71	9.03	细胞核 Nucleus
<i>dpyd</i>	1 059	115 318.02	6.66	细胞质 Cytoplasm
<i>KNCK1</i>	311	34 262.68	7.06	细胞质 Cytoplasm
<i>lov</i>	705	77 363.82	6.55	细胞核 Nucleus
<i>NDRG3</i>	54	6 149.92	4.49	细胞质/细胞外 Cytoplasm / Extracellular
<i>PDSS1</i>	238	26 261.28	8.22	细胞质 Cytoplasm
<i>unc-22</i>	2 205	251 331.74	5.78	细胞质 Cytoplasm
<i>XNP_1</i>	1 208	138 147.74	6.70	细胞核 Nucleus
<i>XNP_2</i>	683	77 197.01	6.39	细胞核 Nucleus
<i>ZMYND8</i>	1 499	164 949.26	8.61	细胞核 Nucleus

果展现了 IAGs 的作用功能定位: 10 个 IAGs 编码的蛋白定位于胞内, 其中 *agtpbp1*、*Bin1*、*lov*、*XNP_1*、*XNP_2* 和 *ZMYND8* 定位于细胞核, 参与丝蛋白和丝胶蛋白的转录与转录后修饰; *BR3*、*dpyd*、*KNCK1*、*PDSS1* 和 *unc-22* 定位于细胞质, 调控丝蛋白与丝胶蛋白的翻译与翻译后修饰。特别地, *NDRG3* 定位于细胞质与细胞外, 属于 NDR 蛋白激酶家族。鉴于 NDR 家族在细胞周期调控中的重要作用 (Yoon *et al.*, 2021) 以及 *NDRG3* 在后丝腺表达上调的现象, *NDRG3* 可能作为胞外环境感受器, 通过感知产丝相关信号分子来调控后丝腺组织的内环境, 进而影响细胞周期和丝腺组织发育速率。此外, *beta-Spec* 定位于细胞膜和细胞骨架, 其表达产物为短链多肽, 由于短链多肽在信号传递和生物过程调控中的关键作用 (Kondo *et al.*, 2007), *beta-Spec* 可能参与产丝信号的识别与跨膜传递以及丝腺细胞骨架的合成过程。

2.4 12 个 IAGs 的功能富集分析

通过 TBtools 软件进行 12 个 IAGs 的功能富集分析 (图 3: A, B)。KEGG 富集分析显示, IAGs 的功能在泛酸、辅酶 A 的合成以及多种酶类的合成通路中高度富集, 同时也富集于氨基酸、核苷酸等小分子代谢过程, 这表明其在能量代谢与生物合成过程中具有关键作用。富集最为显著的 KEGG 通路是泛酸和辅酶 A 的生物合成 (P 值为 $1.908\ 8 \times 10^{-2}$), 该结果表明泛酸和辅酶 A 的合成可能是家蚕产丝过程中的关键步骤, 对丝蛋白的合成与代谢具有重要影响。GO 富集分析显示, IAGs 的功能在组织发育过程中显著富集 (P 值为 $5.004\ 3 \times 10^{-3}$), 这进一步揭示了 12 个 IAGs 在细胞信号传递和生长发育调控中的重要作用。富集最为显著的 GO 通路是蛋白结合 (P 值为 $4.140\ 0 \times 10^{-4}$), 表明了蛋白质互作过程在家蚕产丝过程中的重要作用。

2.5 2 个关键 IAGs 的种间基因家族鉴定结果

在 12 个 IAGs 中, *agtpbp1* 和 *lov* 基因在至少两个丝腺部位展现出了显著性表达差异: *agtpbp1* 基因在中、后丝腺具有显著表达差异 (P 值分别

为 $1.536\ 7 \times 10^{-2}$ 、 $1.235\ 0 \times 10^{-11}$); *lov* 基因在前、中、后丝腺具有显著表达差异 (P 值分别为 $3.896\ 7 \times 10^{-4}$ 、 $4.735\ 3 \times 10^{-2}$ 、 $9.349\ 1 \times 10^{-4}$)。

本研究基于 60 个鳞翅目物种的基因组注释信息, 利用 BLAST、MAFFT 和 TBtools 软件成功筛选出了 *agtpbp1* 和 *lov* 基因家族的候选基因, 分别为 34 和 37 条 (表 4)。

2.6 2 个关键 IAGs 基因家族的系统发育、保守基序及保守结构域分析结果

通过 MAFFT, trimAl 和 TBtools 软件进行 *agtpbp1*、*lov* 两个 IAGs 的系统发育分析, 并分别利用 MEME 和 Batch-CD 分析其保守基序及保守结构域信息 (图 4-图 6)。从保守结构域的数量角度看, *agtpbp1* 家族拥有 4 个保守结构域, *lov* 家族拥有 9 个保守结构域。进一步分析显示, 基因间系统发育的亲缘关系越近, 它们的保守基序与保守结构域的分布与排列方式越趋近于相似。

系统发育分析结果表明, 尽管烟草天蛾 *Manduca sexta* 与小象鹰蛾 *Deilephila porcellus* 的亲缘关系较为接近 (Mei *et al.*, 2022), 但在 *agtpbp1* 家族中, 烟草天蛾却与宽边蜂鹰蛾 *Hemaris fuciformis* 展现出更为紧密的亲缘关系。该现象在多个物种中出现, 且同样存在于 *lov* 基因家族。这种进化模式表明, 基因家族成员可能在环境压力的约束下保留某些对特定生理功能至关重要的序列和结构。保守基序与保守结构域分析结果表明, *agtpbp1* 家族存在保守的羧肽酶结构域, 印证了其肽酶相关功能, 这为其参与形成丝蛋白和丝胶蛋白一级结构的观点提供了进一步的证据。*lov* 家族中, 除了 BTB/POZ 结构域外, 部分家族成员还包含半胱氨酸-2-组氨酸-2-锌指蛋白 (zf-C2H2) 结构域。研究表明 zf-C2H2 是一类能够调控组织发育的转录因子 (Wu *et al.*, 2019)。因此, 除了调控丝蛋白的翻译后修饰与降解过程, 部分 *lov* 家族成员可能具备丝腺发育调控功能。

2.7 2 个产丝关键 IAGs 基因家族的选择压力与正选择位点分析结果

利用 codeml 与 PAML 包对涵盖 *agtpbp1* 和 *lov* 两个家族的基因进行了选择压力分析, 并利

用 Datamonkey 平台进行了多拷贝基因的自然选择偏好性分析 (表 4, 表 5)。枝模型 One ratio 模型分析结果显示, *agtpbp1* 家族整体 $\omega=1.615\ 0$ ($P\text{-value}=0.049\ 6$), *lov* 家族整体 $\omega=0.076\ 6$ ($P\text{-value}=2.250\ 5\times 10^{-10}$)。对各个多拷贝基因的偏好性分析表明, 虽然其 K 值均偏离 1, 但 P 值均大于 0.05, 不位于置信区间内, 表明多拷贝

基因的选择压力分析偏好性不显著。利用 codeml 与 PAML 包进行枝位点模型 Two ratio 模型分析与正选择位点分析 (表 5, 表 6)。结果显示, 仅在 *agtpbp1* 家族的 *Bbet001799*、*Bfus012502*、*Dpor018892*、*Hfuc009526*、*Msex012828*、*Ndro005471*、*Pdar001813* 和 *Sexi005476* 基因中存在显著性正选择位点。

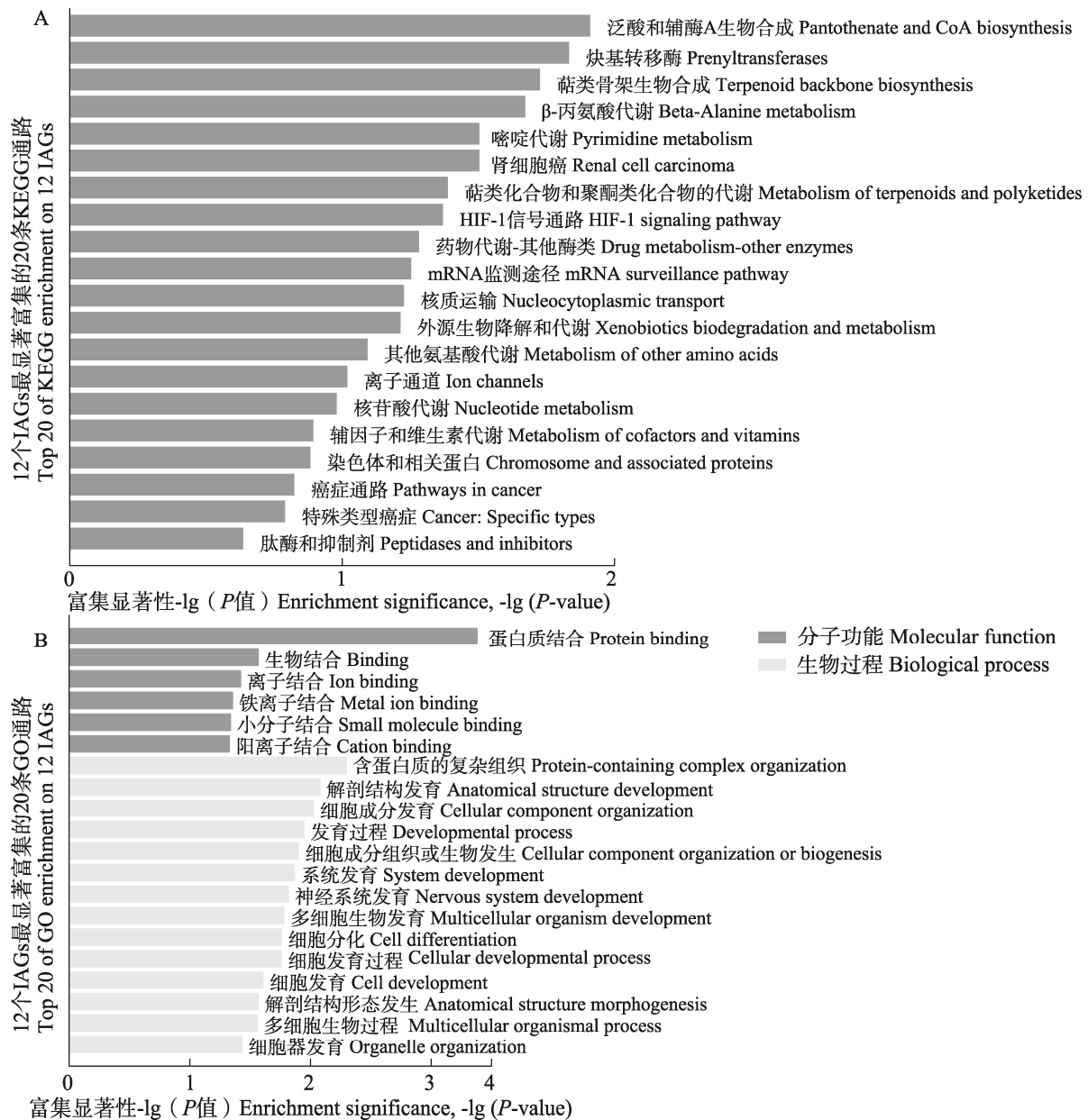


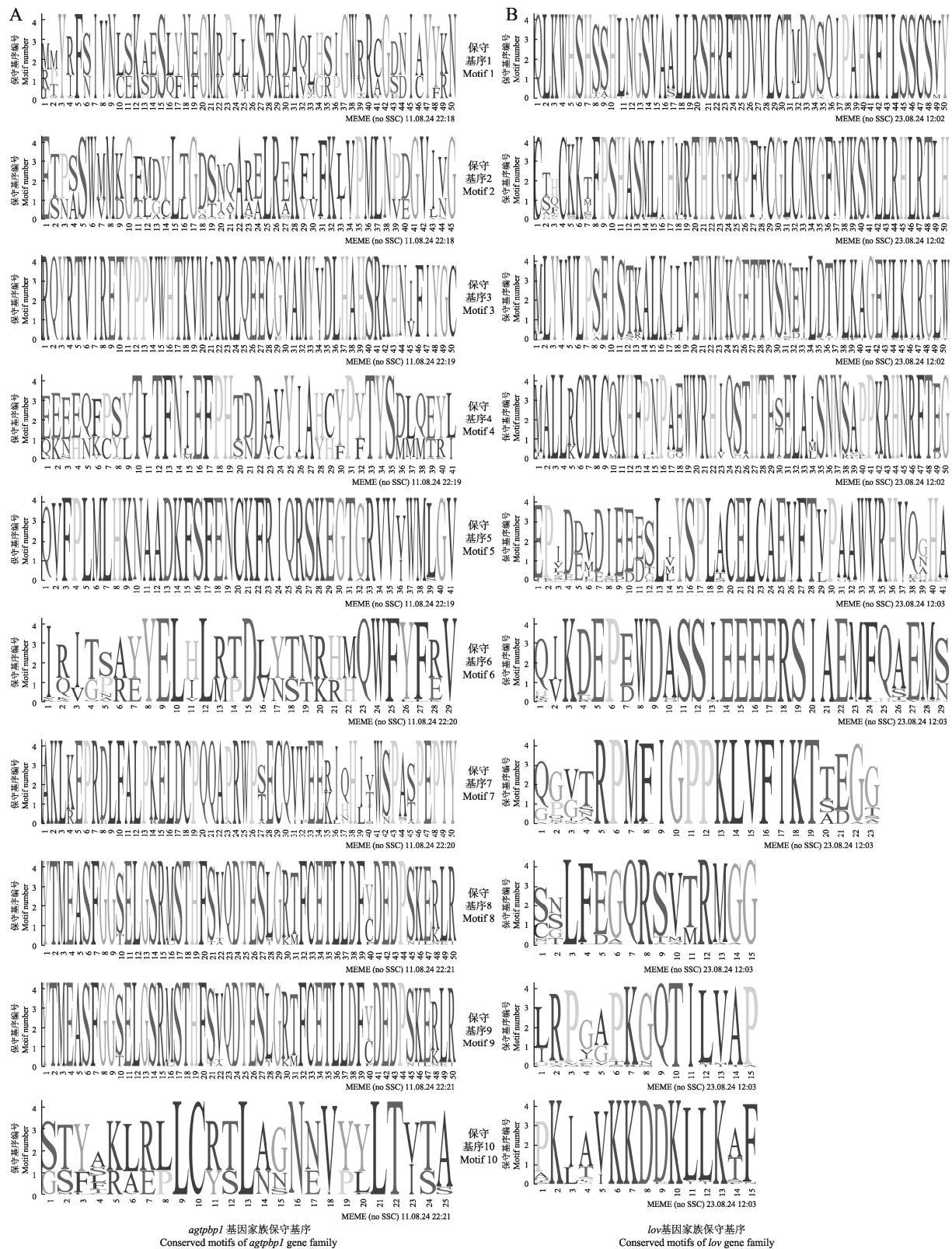
图 3 12 个差异表达 IAGs 的 KEGG 与 GO 通路富集分析结果

Fig. 3 KEGG and GO pathway analysis results of 12 differentially expressed IAGs

A. 家蚕产丝相关 IAGs 的 KEGG 通路富集分析结果; B. 家蚕产丝相关 IAGs 的 GO 通路富集分析结果。

A. KEGG pathway enrichment analysis results of silk production related IAGs in *B. mori*;

B. GO pathway enrichment analysis results of silk production related IAGs in *B. mori*.

图 4 *agtbp1* 与 *lov* 基因家族保守基序Fig. 4 Conserved motifs of *agtbp1* and *lov* gene familiesA. *agtbp1* 基因家族保守基序; B. *lov* 基因家族保守基序。A. Conservative motifs of *agtbp1* gene family; B. Conservative motifs of *lov* gene family.

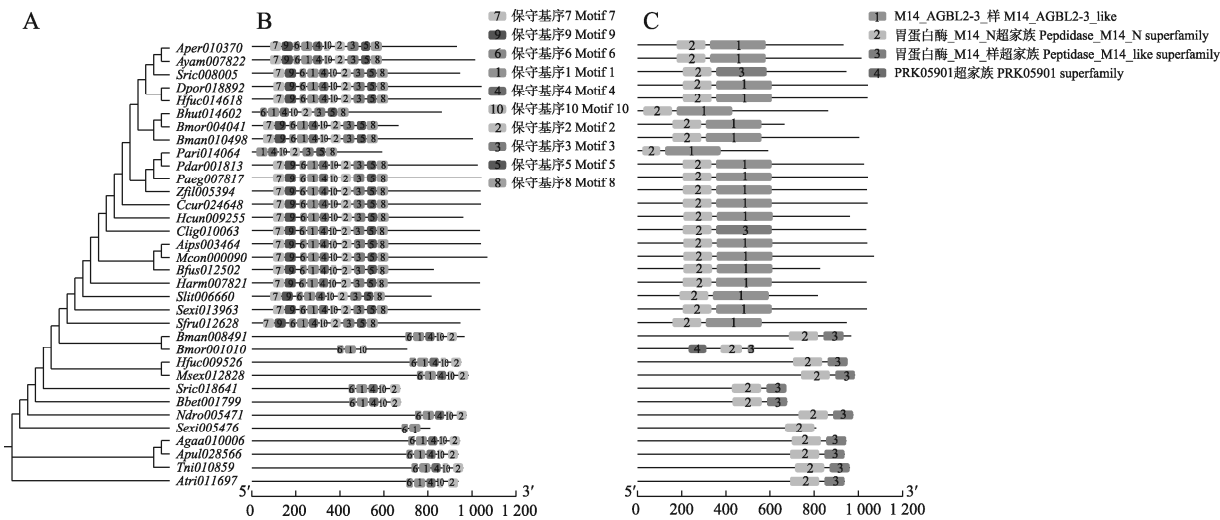


图 5 *agtbp1* 基因家族系统进化树、保守基序与保守结构域

Fig. 5 Phylogenetic tree, conserved motifs, and conserved domains of *agtbp1* gene family

A. *agtbp1* 基因家族系统进化树; B. *agtbp1* 基因家族保守基序; C. *agtbp1* 基因家族保守结构域。

A. Phylogenetic tree of *agtbp1* gene family; B. Conservative motifs of *agtbp1* gene family;
C. Conserved protein domain of *agtbp1* gene family.

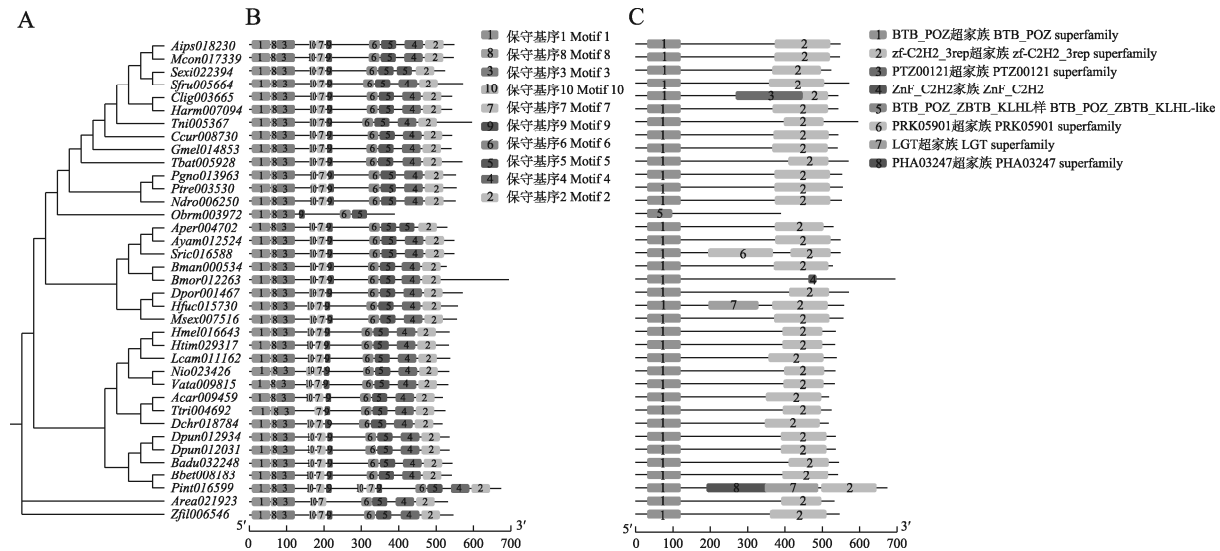


图 6 *lov* 基因家族系统进化树、保守基序与保守结构域

Fig. 6 Phylogenetic tree, conserved motifs, and conserved domains of *lov* gene family

A. *lov* 基因家族系统进化树; B. *lov* 基因家族保守基序; C. *lov* 基因家族保守结构域。

A. Phylogenetic tree of *lov* gene family; B. Conservative motifs of *lov* gene family;
C. Conserved protein domain of *lov* gene family.

表 4 *agtbp1* 与 *lov* 基因家族成员在 60 个物种中的分布

Table 4 Distribution of *agtbp1* and *lov* gene family members in 60 species

物种名称 Specie name	基因家族成员的物种内数量 The number of gene family members within species	
	<i>agtbp1</i>	<i>lov</i>
隐金夜蛾 <i>Abrostola tripartita</i>	1	0
月形天蚕蛾 <i>Actias luna</i>	0	0
雷氏长角蛾 <i>Adela reaumurella</i>	0	1

续表 4 (Table 4 continued)

物种名称 Specie name	基因家族成员的物种内数量 The number of gene family members within species	
	<i>agtpbp1</i>	<i>lov</i>
小地老虎 <i>Agrotis ipsilon</i>	1	1
柞蚕 <i>Antheraea pernyi</i>	1	1
天蚕 <i>Antheraea yamamai</i>	1	1
橙尖襟粉蝶 <i>Anthocharis cardamines</i>	0	1
伽马丫纹夜蛾 <i>Autographa gamma</i>	1	0
美丫纹夜蛾 <i>Autographa pulchrina</i>	1	0
桦尺蛾 <i>Biston betularia</i>	1	1
焦斑织蛾 <i>Blastobasis adustella</i>	0	1
直线野蚕蛾 <i>Bombyx huttoni</i>	1	0
野蚕 <i>Bombyx mandarina</i>	2 ($K=0.84$, $P=0.546$)	1
家蚕 <i>Bombyx mori</i>	2 ($K=0.86$, $P=0.591$)	1
非洲蛀茎夜蛾 <i>Busseola fusca</i>	1	0
桃小食心虫 <i>Carposina sasakii</i>	0	0
琉璃灰蝶 <i>Celastrina argiolus</i>	0	0
短扇舟蛾 <i>Clostera curtula</i>	1	1
可可荚蛾 <i>Conopomorpha cramerella</i>	0	0
女贞首夜蛾 <i>Craniophora ligustri</i>	1	1
苹果蠹蛾 <i>Cydia pomonella</i>	0	0
金斑蝶 <i>Danaus chrysippus</i>	0	1
小象鹰蛾 <i>Deilephila porcellus</i>	1	1
报喜绢粉蝶 <i>Delias pasithoe</i>	0	0
马尾松毛虫 <i>Dendrolimus punctatus</i>	0	2 ($K=50.00$, $P=0.203$)
弧眉钩蛾 <i>Drepana arcuata</i>	0	0
大蜡螟 <i>Galleria mellonella</i>	0	1
红带袖蝶 <i>Heliconius melpomene</i>	0	1
提马袖蝶 <i>Heliconius timareta</i>	0	1
棉铃虫 <i>Helicoverpa armigera</i>	1	1
烟芽夜蛾 <i>Heliothis virescens</i>	0	0
宽边蜂鹰蛾 <i>Hemaris fuciformis</i>	2 ($K=0.97$, $P=0.891$)	1
带纹青尺蛾 <i>Hylaea fasciaria</i>	0	0
美国白蛾 <i>Hyphantria cunea</i>	1	0
枯叶蛱蝶 <i>Kallima inachus</i>	0	0
红线蛱蝶 <i>Limenitis camilla</i>	0	1
蓓带夜蛾 <i>Mamestra configurata</i>	1	1
烟草天蛾 <i>Manduca sexta</i>	1	1
椴六点天蛾 <i>Mimas tiliae</i>	0	0

续表 4 (Table 4 continued)

物种名称 Specie name	基因家族成员的物种内数量 The number of gene family members within species	
	<i>agtpbp1</i>	<i>lov</i>
驼背舟蛾 <i>Notodonta dromedarius</i>	1	1
孔雀蛱蝶 <i>Nymphalis io</i>	0	1
冬尺蛾 <i>Operophtera brumata</i>	0	1
亚洲玉米螟 <i>Ostrinia furnacalis</i>	0	0
海岛虎凤蝶 <i>Papilio aristodemus</i>	1	0
非洲凤蝶 <i>Papilio dardanus</i>	1	0
虎纹凤蝶 <i>Papilio glaucus</i>	0	0
林眼蝶 <i>Pararge aegeria</i>	1	0
小光舟蛾 <i>Pheosia gnoma</i>	0	1
杨光舟蛾 <i>Pheosia tremula</i>	0	1
印度谷螟 <i>Plodia interpunctella</i>	0	1
小菜蛾 <i>Plutella xylostella</i>	0	0
蓖麻蚕 <i>Samia ricini</i>	2 ($K=0.75, P=0.467$)	1
甜菜夜蛾 <i>Spodoptera exigua</i>	2 ($K=1.12, P=0.815$)	1
斜纹夜蛾 <i>Spodoptera litura</i>	1	0
草地贪夜蛾 <i>Spodoptera frugiperda</i>	1	1
桦太波夜蛾 <i>Thyatira batis</i>	0	1
三点衣蛾 <i>Tinea trinitella</i>	0	1
粉纹夜蛾 <i>Trichoplusia ni</i>	1	1
小红蛱蝶 <i>Vanessa atalanta</i>	0	1
六斑斑蛾 <i>Zygaena filipendulae</i>	1	1
总计 Total	34	37

3 讨论

本研究整合了前期对 15 个家蚕 IAGs 的注释。家蚕产丝相关 IAGs 由家蚕中日系实用种的改良相关基因筛选而来,可能与家蚕重要经济性状的提升有关(龙安星, 2022)。研究表明,丝腺在丝蛋白和丝胶蛋白的合成与分泌调控过程中发挥重要作用。其中,中丝腺主要负责丝胶蛋白的合成,后丝腺主要负责丝蛋白的合成,前丝腺则协助丝蛋白和丝胶蛋白的分泌(Inoue *et al.*, 2000)。差异表达分析结果表明,12 个家蚕 IAGs 中有 8 个(66.7%)在后丝腺发挥功能,即后丝腺在产丝过程中发挥主导作用,这与 Inoue 等(2000)的研究结果一致。通过差异表达分析筛

选获得的 12 个 IAGs 在产丝期影响产丝过程;虽然 *ABCF3*、*ct* 和蛋白 LOC101737920 编码基因未展现显著表达差异,但它们对产丝过程的影响仍可能存在。*ABCF3* 为 ATP 结合盒 F 亚家族(ATP-binding cassette subfamily F)成员,是已知最大的外排泵组(Peterson *et al.*, 2019; Pan *et al.*, 2020)。*ABCF3* 表达差异不显著的现象可能表明膜上 *ABCF3* 蛋白的总量保持稳定,这些蛋白作用于整个家蚕发育过程。*ct* 在不同物种的神经系统中呈现一致性表达(Leyva-Díaz, 2023),鉴于家蚕神经系统的活动是其各种生命活动如结茧、交配等的基础(于奇等, 2011), *ct* 基因可能通过调控家蚕神经系统的发育间接影响产丝。蛋白 LOC101737920 编码基因被证实参与了

表 5 *agtpbp1* 与 *lov* 基因家族选择压力
Table 5 Selective pressure of *agtpbp1* and *lov* gene families

基因家族 Gene family	模型类型 Model types	基因 ID Gene ID	比值 $\omega = dN / dS$	基因家族 Gene family	模型类型 Model types	基因 ID Gene ID	比值 $\omega = dN / dS$
<i>agtpbp1</i>	One ratio		1.615 0			<i>Aips018230</i>	0.062 1
	Two ratio	<i>Agaa010006</i>	0.770 5			<i>Aper004702</i>	0.056 5
		<i>Aips003464</i>	0.829 4			<i>Area021923</i>	0.112 7
		<i>Aper010370</i>	0.838 2			<i>Ayam012524</i>	0.071 4
		<i>Apul028566</i>	0.458 5			<i>Badu032248</i>	0.058 3
		<i>Atri011697</i>	1.317 4			<i>Bbet008183</i>	0.086 6
		<i>Ayam007822</i>	1.968 1			<i>Bman000534</i>	0.000 1
		<i>Bbet001799</i>	1.830 2			<i>Bmor012263</i>	0.000 1
		<i>Bfus012502</i>	10.110 1			<i>Ccur008730</i>	0.068 9
		<i>Bhut014602</i>	1.127 1			<i>Clig003665</i>	0.046 1
		<i>Bman008491</i>	999			<i>Dchr018784</i>	0.027 3
		<i>Bman010498</i>	0.000 1			<i>Dpor001467</i>	0.059 3
		<i>Bmor001010</i>	1.726 2			<i>Dpun012031</i>	0.000 1
		<i>Bmor004041</i>	1.481 5			<i>Dpun012934</i>	2.323 2
		<i>Ccur024648</i>	1.081 4			<i>Gmel014853</i>	0.040 8
		<i>Clig010063</i>	1.432 7			<i>Harm007094</i>	0.162 0
		<i>Dpor018892</i>	5.445 6			<i>Hfuc015730</i>	0.037 5
		<i>Harm007821</i>	0.519 6			<i>Hmel016643</i>	0.000 1
		<i>Hcun009255</i>	0.873 5			<i>Htim029317</i>	0.000 1
		<i>Hfuc009526</i>	3.964 0			<i>Lcam011162</i>	0.106 3
		<i>Hfuc014618</i>	0.690 4			<i>Mcon017339</i>	0.040 7
		<i>Mcon000090</i>	2.447 2			<i>Msex007516</i>	0.042 5
		<i>Msex012828</i>	3.420 3			<i>Ndro006250</i>	0.092 7
		<i>Ndro005471</i>	3.054 7			<i>Nio023426</i>	0.032 9
		<i>Paeg007817</i>	1.244 3			<i>Obrm003972</i>	0.179 9
		<i>Pari014064</i>	1.720 6			<i>Pgno013963</i>	0.000 1
		<i>Pdar001813</i>	5.667 5			<i>Pint016599</i>	0.047 2
		<i>Sexi005476</i>	2.416 3			<i>Ptre003530</i>	0.000 1
		<i>Sexi013963</i>	3.333 8			<i>Sexi022394</i>	0.087 0
		<i>Sfru012628</i>	2.212 2			<i>Sfru005664</i>	0.000 1
		<i>Slit006660</i>	1.582 0			<i>Sric016588</i>	0.127 9
		<i>Sric008005</i>	1.720 0			<i>Tbat005928</i>	0.076 9
		<i>Sric018641</i>	1.541 0			<i>Tni005367</i>	0.214 5
		<i>Tni010859</i>	2.999 7			<i>Ttri004692</i>	0.115 1
		<i>Zfil005394</i>	1.348 8			<i>Vata009815</i>	0.087 7
<i>lov</i>	One ratio		0.076 6			<i>Zfil006546</i>	0.022 9
	Two ratio	<i>Acar009459</i>	0.139 8				

家蚕从幼虫到蛹的变态过程 (Ou *et al.*, 2014), 表明该基因可能影响家蚕化蛹过程, 从而间接影响产丝。可聚焦于 12 个 IAGs, 通过实验探究每个基因对家蚕产丝性能的贡献, 并筛选出具有优良产丝性能的基因组合; 同时, 可针对 *ABCF3*、*ct* 和蛋白 LOC101737920 编码基因进行功能验证实验, 探索这些基因在特定环境下的调控机制。

KEGG 富集分析表明泛酸和辅酶 A 的生物合成通路可能为家蚕产丝核心环节, 直接参与调控丝蛋白的合成与代谢。因此, 通过调控泛酸和辅酶 A 的生物合成通路, 有望实现对家蚕产丝过程的优化和提升。GO 富集分析结果细化了 IAGs 的功能分布情况, 表明这些基因在家蚕产丝信号传递和发育调控中发挥关键作用, 这与李娟 (2017) 研究结果一致。12 个 IAGs 可能通过调控细胞增殖、分化、凋亡等生物学过程来影响家蚕的整体发育和产丝能力。因此, 深入研究 IAGs 在发育调控中的具体作用机制将有助于更好地理解家蚕产丝过程, 并为产丝优化提供新的思路和方法。

目前, 关于 *agtpbp1* 与 *lov* 这两个 IAGs 的基础研究较为空白, 其分子功能与作用机制仍待探究。根据上文的分析, 对这 3 个基因的进一步深入研究, 将有助于更全面地理解家蚕产丝相关过程的分子调控机理, 为家蚕产丝性能的优化提供基础性数据。

基因家族鉴定聚焦于在丝蛋白合成调控和稳态维持中具有重要作用的关键基因——*agtpbp1* 和 *lov*。基因家族成员的结构差异表明家族成员具有不同的功能 (时小东等, 2019; 马淑娟等, 2023)。基因的外显子与内含子结构与其进化分支有着密切的联系, 具有相似外显子与内含子结构的基因往往归属于相同的进化分支 (郝爱平等, 2023)。因此, 结构上的差异性可能导致 *agtpbp1* 和 *lov* 家族产生功能各异且在产丝的特定环节发挥作用的旁系同源基因。*agtpbp1* 家族成员均具备连续的 M₁₄ 羧肽酶结构域, 且 *agtpbp1* 家族 30 个成员具有连续的 motif6、motif1、motif4、motif10 和 motif2 保守基序分布,

以此构成 M₁₄ 结构域的 N 端与主体。该现象表明上述保守基序对于维持 M₁₄ 结构域的构象具有重要意义。该家族 20 个成员存在 AGBL2-3_like 结构域, 表明其可能存在类似于 AGBL 转录因子的功能 (Wang *et al.*, 2018)。*lov* 家族基因的中段均存在长度为 50-100 个氨基酸的非保守基序与非保守结构域区域, 并将 BTB/POZ 与 zf-C2H2 两结构域与众多保守基序分离, 这意味着 *lov* 家族基因的表达产物可能会被切割为两部分, 分别执行泛素化和转录调控的功能。此外, 研究表明 BTB/POZ 与 zf-C2H2 结构域组合在某些真核生物中展现出抗菌活性 (Yin *et al.*, 2022), 因此 *lov* 家族可能是一类广泛存在于家蚕及其亲缘物种中的昆虫免疫因子, 在家蚕的免疫防御机制中发挥重要作用。

对 *agtpbp1* 家族的选择压力分析和正选择位点分析结果表明, 该家族整体受到正选择作用, 且各系统发育分支的 ω 值计算结果佐证了此结论。与此同时, 该家族存在 18 个 ω 值大于 1 但无显著性正选择位点的假阳性基因, 这可能与多序列联配准确性、模型选择、算法等多个因素有关 (Yang, 2007; Haasdijk and Heinerman, 2018; Chao *et al.*, 2022)。

家蚕作为一种历史悠久的农业昆虫, 对于人类社会具有重要价值。尽管学界已初步探索了家蚕的产丝机制并挖掘了部分产丝相关 IAGs, 但对于这些基因的深入研究仍显不足。本研究基于数据库中的家蚕蛹前期与 5 龄幼虫转录组数据, 整合了前期对 15 个 IAGs 的注释, 对存在龄期间显著表达差异的 12 个 IAGs 的分子机制进行了初步探讨。此外, 本研究鉴定并分析了 2 个家蚕产丝过程的关键 IAGs 在 60 个鳞翅目物种中的种间基因家族, 深入挖掘了这些家族基因的结构特征、分子功能与进化情况。这一研究不仅为家蚕产丝相关基因的深入挖掘提供了宝贵的基础性数据与见解, 还为家蚕育种工作提供了新的思路与方向。综上所述, 本研究通过综合分析 with 深入挖掘, 为家蚕产丝相关基因的研究与家蚕育种工作奠定了一定基础, 有望推动家蚕产业的持续发展与创新。

表 6 *agtpbp1* 与 *lov* 基因家族正选择位点
Table 6 Positively selected sites of *agtpbp1* and *lov* gene families

基因 ID Gene ID	氨基酸同义替换与非同义 替换比值 $\omega=K_a/K_s$ Ratio of synonymous to non- synonymous substitutions	正选择位点 Positively selected sites	P 值 P-value
<i>Bbet001799</i>	1.830 2	47H 0.742, 72Q 0.642, 96V 0.848, 132I 0.824	0.004 6
<i>Bfus012502</i>	10.110 1	1C 0.768, 2C 0.768, 3M 0.768, 4E 0.768, 5D 0.768, 6A 0.989*, 7S 0.768, 8V 0.768, 9A 0.768, 10C 0.768, 11F 0.768, 12G 0.768, 13C 0.768, 14S 0.768, 15T 0.768, 16V 0.768, 17G 0.768, 18P 0.768, 19L 0.767, 20R 0.768, 21H 0.768, 22H 0.768, 23P 0.768, 24S 0.968*, 25D 0.768, 26F 0.767, 27A 0.768, 28A 0.769, 29P 0.768, 30L 0.768, 31R 0.768, 32S 0.768, 33G 0.768, 34E 0.768, 35H 0.968*, 36C 0.949, 37V 0.768, 38R 0.768, 39M 0.768, 40S 0.768, 41R 0.770, 42F 0.862, 43T 0.768, 44G 0.768, 45S 0.768, 46A 0.768, 47H 0.768, 48S 0.768, 49V 0.768, 50Q 0.768, 51M 0.768, 52N 0.769, 53Q 0.768, 54N 0.768, 55M 0.768, 56G 0.968*, 57Y 0.769, 58Q 0.768, 59E 0.768, 60Y 0.768, 61V 0.768, 62P 0.768, 63H 0.768, 64Y 0.768, 65V 0.767, 66N 0.768, 67K 0.768, 68P 0.768, 69Y 0.768, 70L 0.768, 71Q 0.768, 72Q 0.768, 73V 0.771, 74Q 0.768, 75R 0.767, 76R 0.957*, 77L 0.768, 78T 0.768, 79R 0.768, 80H 0.940, 81H 0.768, 82G 0.768, 83V 0.768, 84R 0.768, 85K 0.768, 86K 0.768, 87T 0.768, 88Q 0.768, 89I 0.768, 90Q 0.968*, 91Q 0.769, 92I 0.768, 93Q 0.768, 94T 0.768, 95T 0.768, 96V 0.768, 97Q 0.768, 98I 0.768, 99I 0.767, 100K 0.768, 101A 0.768, 102T 0.767, 103S 0.768, 104F 0.768, 105L 0.768, 106T 0.768, 107I 0.768, 108L 0.768, 109K 0.768, 110L 0.768, 111V 0.768, 112R 0.768, 113A 0.768, 114V 0.768, 115T 0.768, 116I 0.768, 117L 0.768, 118M 0.767, 119R 0.768, 120K 0.768, 121K 0.768, 122L 0.976*, 123L 0.768, 124D 0.785, 125I 0.768, 126C 0.768, 127P 0.768, 128P 0.768, 129W 0.768, 130S 0.768, 131K 0.768, 132I 0.768, 133C 0.768, 134L 0.768, 135T 0.876, 136N 0.768, 137I 0.768, 138T 0.768, 139K 0.768, 140V 0.768, 141N 0.768, 142S 0.768, 143F 0.768, 144S 0.768	0.023 4
<i>Dpor018892</i>	5.445 6	1C 0.545, 2C 0.553, 3M 0.565, 4E 0.580, 5D 0.558, 6A 0.553, 7S 0.558, 8V 0.556, 9A 0.858, 10C 0.575, 11F 0.567, 12G 0.953*, 13C 0.572, 14S 0.713, 15T 0.572, 16V 0.725, 17G 0.557, 18P 0.570, 19L 0.956*, 20R 0.585, 21H 0.557, 22H 0.562, 23P 0.555, 24S 0.553, 25D 0.556, 26F 0.523, 27A 0.869, 28A 0.591, 29P 0.592, 30L 0.727, 31R 0.552, 32S 0.549, 33G 0.543, 34E 0.566, 35H 0.561, 36C 0.575, 37V 0.568, 38R 0.573, 39M 0.580, 40S 0.562, 41R 0.623, 42F 0.563, 43T 0.555, 44G 0.557, 45S 0.555, 46A 0.564, 47H 0.567, 48S 0.578, 49V 0.550, 50Q 0.573, 51M 0.581, 52N 0.571, 53Q 0.553, 54N 0.544, 55M 0.893, 56G 0.575, 57Y 0.578, 58Q 0.884, 59E 0.745, 60Y 0.897, 61V 0.550, 62P 0.573, 63H 0.578, 64Y 0.575, 65V 0.582, 66N 0.547, 67K 0.557, 68P 0.559, 69Y 0.572, 70L 0.579, 71Q 0.593, 72Q 0.903, 73V 0.903, 74Q 0.543, 75R 0.524, 76R 0.576, 77L 0.701, 78T 0.553, 79R 0.593, 80H 0.907, 81H 0.563, 82G 0.568, 83V 0.547, 84R 0.703, 85K 0.562, 86K 0.557, 87T 0.565, 88Q 0.559, 89I 0.823, 90Q 0.896, 91Q 0.573, 92I 0.561, 93Q 0.868, 94T 0.561, 95T 0.551, 96V 0.576, 97Q 0.571, 98I 0.577, 99I 0.515, 100K 0.557, 101A 0.573, 102T 0.515, 103S 0.556, 104F 0.567, 105L 0.584, 106T 0.563, 107I 0.554, 108L 0.576, 109K 0.898, 110L 0.653, 111V 0.554, 112R 0.556, 113A 0.584, 114V 0.556, 115T 0.567, 116I 0.549, 117L 0.578, 118M 0.515, 119R 0.561, 120K 0.575, 121K 0.544, 122L 0.577, 123L 0.582, 124D 0.560, 125I 0.888, 126C 0.590, 127P 0.887, 128P 0.747, 129W 0.576, 130S 0.577, 131K 0.585, 132I 0.604, 133C 0.558, 134L 0.560, 135T 0.810, 136N 0.550, 137I 0.554, 138T 0.574, 139K 0.556, 140V 0.551, 141N 0.559, 142S 0.566, 143F 0.883, 144S 0.899	0.022 0

续表 6 (Table 6 continued)

基因 ID Gene ID	氨基酸同义替换与非同义 替换比值 $\omega=K_a/K_s$ Ratio of synonymous to non- synonymous substitutions	正选择位点 Positively selected sites	P 值 P-value
<i>Hfuc009526</i>	3.964 0	1C 0.606, 2C 0.606, 3M 0.533, 4E 0.596, 5D 0.582, 6A 0.588, 7S 0.607, 8V 0.599, 9A 0.595, 10C 0.609, 11F 0.605, 12G 0.604, 13C 0.604, 14S 0.609, 15T 0.587, 16V 0.876, 17G 0.614, 18P 0.602, 19L 0.618, 20R 0.624, 21H 0.585, 22H 0.591, 23P 0.616, 24S 0.869, 25D 0.582, 26F 0.975*, 27A 0.606, 28A 0.610, 29P 0.603, 30L 0.622, 31R 0.616, 32S 0.606, 33G 0.718, 34E 0.614, 35H 0.652, 36C 0.812, 37V 0.912, 38R 0.623, 39M 0.590, 40S 0.638, 41R 0.590, 42F 0.765, 43T 0.595, 44G 0.601, 45S 0.908, 46A 0.597, 47H 0.900, 48S 0.576, 49V 0.920, 50Q 0.606, 51M 0.573, 52N 0.597, 53Q 0.922, 54N 0.572, 55M 0.761, 56G 0.899, 57Y 0.607, 58Q 0.607, 59E 0.590, 60Y 0.633, 61V 0.604, 62P 0.616, 63H 0.825, 64Y 0.873, 65V 0.844, 66N 0.572, 67K 0.609, 68P 0.767, 69Y 0.617, 70L 0.607, 71Q 0.628, 72Q 0.713, 73V 0.608, 74Q 0.772, 75R 0.618, 76R 0.613, 77L 0.633, 78T 0.839, 79R 0.617, 80H 0.627, 81H 0.605, 82G 0.607, 83V 0.584, 84R 0.581, 85K 0.578, 86K 0.577, 87T 0.852, 88Q 0.624, 89I 0.577, 90Q 0.615, 91Q 0.605, 92I 0.603, 93Q 0.605, 94T 0.604, 95T 0.952*, 96V 0.589, 97Q 0.606, 98I 0.943, 99I 0.561, 100K 0.940, 101A 0.604, 102T 0.602, 103S 0.897, 104F 0.586, 105L 0.613, 106T 0.591, 107I 0.586, 108L 0.613, 109K 0.589, 110L 0.544, 111V 0.637, 112R 0.584, 113A 0.587, 114V 0.938, 115T 0.769, 116I 0.576, 117L 0.916, 118M 0.626, 119R 0.594, 120K 0.580, 121K 0.584, 122L 0.613, 123L 0.609, 124D 0.724, 125I 0.890, 126C 0.702, 127P 0.597, 128P 0.890, 129W 0.869, 130S 0.592, 131K 0.581, 132I 0.597, 133C 0.590, 134L 0.598, 135T 0.584, 136N 0.579, 137I 0.563, 138T 0.593, 139K 0.589, 140V 0.586, 141N 0.572, 142S 0.583, 143F 0.587, 144S 0.605	0.001 9
<i>Msex012828</i>	3.420 3	30L 0.718, 33G 0.715, 35H 0.767, 36C 0.537, 40S 0.748, 42F 0.645, 51M 0.866, 55M 0.584, 60Y 0.799, 63H 0.510, 67K 0.824, 68P 0.759, 71Q 0.804, 72Q 0.681, 74Q 0.694, 75R 0.773, 76R 0.784, 79R 0.818, 80H 0.749, 88Q 0.833, 92I 0.768, 94T 0.728, 102T 0.916, 103S 0.545, 111V 0.803, 114V 0.754, 115T 0.613, 118M 0.915, 124D 0.710, 126C 0.985*, 132I 0.748	0.010 2
<i>Ndro005471</i>	3.054 7	8V 0.866, 22H 0.792, 24S 0.732, 25D 0.885, 26F 0.786, 29P 0.672, 33G 0.868, 35H 0.513, 36C 0.672, 37V 0.628, 38R 0.782, 40S 0.764, 42F 0.807, 44G 0.804, 47H 0.829, 50Q 0.807, 53Q 0.509, 57Y 0.814, 58Q 0.838, 60Y 0.759, 61V 0.803, 62P 0.712, 63H 0.743, 71Q 0.823, 72Q 0.847, 74Q 0.500, 77L 0.715, 78T 0.663, 81H 0.798, 86K 0.825, 87T 0.797, 90Q 0.795, 94T 0.773, 95T 0.854, 97Q 0.829, 98I 0.972*, 99I 0.898, 100K 0.559, 101A 0.961*, 102T 0.521, 104F 0.859, 105L 0.823, 107I 0.609, 108L 0.835, 109K 0.801, 111V 0.781, 113A 0.847, 114V 0.916, 115T 0.665, 124D 0.803, 126C 0.922, 128P 0.893, 129W 0.741, 130S 0.508, 135T 0.814, 141N 0.895	0.002 7

续表 6 (Table 6 continued)

基因 ID Gene ID	氨基酸同义替换与非同义 替换比值 $\omega=K_a/K_s$ Ratio of synonymous to non- synonymous substitutions	正选择位点 Positively selected sites	P 值 P-value
<i>Pdar001813</i>	5.667 5	1C 0.748, 2C 0.747, 3M 0.750, 4E 0.746, 5D 0.953*, 6A 0.746, 7S 0.750, 8V 0.748, 9A 0.743, 10C 0.749, 11F 0.784, 12G 0.975*, 13C 0.973*, 14S 0.844, 15T 0.811, 16V 0.761, 17G 0.746, 18P 0.748, 19L 0.738, 20R 0.859, 21H 0.750, 22H 0.750, 23P 0.746, 24S 0.760, 25D 0.746, 26F 0.746, 27A 0.769, 28A 0.794, 29P 0.747, 30L 0.749, 31R 0.745, 32S 0.745, 33G 0.748, 34E 0.745, 35H 0.807, 36C 0.949, 37V 0.748, 38R 0.749, 39M 0.755, 40S 0.747, 41R 0.748, 42F 0.749, 43T 0.747, 44G 0.749, 45S 0.749, 46A 0.750, 47H 0.745, 48S 0.747, 49V 0.749, 50Q 0.892, 51M 0.744, 52N 0.749, 53Q 0.748, 54N 0.744, 55M 0.749, 56G 0.750, 57Y 0.791, 58Q 0.752, 59E 0.753, 60Y 0.837, 61V 0.745, 62P 0.749, 63H 0.753, 64Y 0.748, 65V 0.749, 66N 0.744, 67K 0.746, 68P 0.804, 69Y 0.750, 70L 0.747, 71Q 0.824, 72Q 0.754, 73V 0.748, 74Q 0.745, 75R 0.756, 76R 0.747, 77L 0.792, 78T 0.790, 79R 0.753, 80H 0.752, 81H 0.749, 82G 0.753, 83V 0.742, 84R 0.748, 85K 0.745, 86K 0.752, 87T 0.748, 88Q 0.751, 89I 0.748, 90Q 0.852, 91Q 0.749, 92I 0.747, 93Q 0.750, 94T 0.749, 95T 0.747, 96V 0.748, 97Q 0.749, 98I 0.751, 99I 0.738, 100K 0.746, 101A 0.749, 102T 0.738, 103S 0.746, 104F 0.748, 105L 0.750, 106T 0.746, 107I 0.746, 108L 0.746, 109K 0.746, 110L 0.747, 111V 0.747, 112R 0.746, 113A 0.751, 114V 0.748, 115T 0.750, 116I 0.745, 117L 0.750, 118M 0.738, 119R 0.747, 120K 0.750, 121K 0.744, 122L 0.750, 123L 0.751, 124D 0.751, 125I 0.749, 126C 0.746, 127P 0.900, 128P 0.746, 129W 0.752, 130S 0.746, 131K 0.913, 132I 0.755, 133C 0.910, 134L 0.747, 135T 0.821, 136N 0.841, 137I 0.747, 138T 0.750, 139K 0.807, 140V 0.751, 141N 0.745, 142S 0.748, 143F 0.788, 144S 0.895	0.0084
<i>Sexi005476</i>	2.416 3	10C 0.502, 11F 0.717, 14S 0.502, 17G 0.504, 19L 0.512, 20R 0.525, 24S 0.520, 26F 0.775, 27A 0.500, 28A 0.501, 29P 0.591, 30L 0.509, 31R 0.836, 33G 0.580, 35H 0.502, 36C 0.503, 37V 0.627, 38R 0.527, 39M 0.727, 40S 0.615, 41R 0.524, 42F 0.517, 47H 0.502, 50Q 0.803, 53Q 0.802, 57Y 0.524, 58Q 0.791, 61V 0.804, 62P 0.702, 63H 0.556, 64Y 0.518, 66N 0.887, 67K 0.832, 68P 0.790, 69Y 0.791, 70L 0.814, 71Q 0.562, 72Q 0.511, 74Q 0.522, 75R 0.720, 76R 0.806, 77L 0.613, 78T 0.684, 79R 0.507, 86K 0.826, 87T 0.733, 89I 0.802, 90Q 0.503, 91Q 0.784, 92I 0.730, 94T 0.758, 95T 0.853, 96V 0.810, 101A 0.748, 102T 0.540, 103S 0.832, 104F 0.853, 105L 0.619, 107I 0.646, 108L 0.507, 109K 0.783, 110L 0.811, 111V 0.768, 112R 0.793, 115T 0.828, 117L 0.519, 122L 0.515, 124D 0.712, 126C 0.793, 127P 0.511, 128P 0.610, 129W 0.597, 130S 0.517, 133C 0.810, 134L 0.751, 142S 0.797, 144S 0.516	0.0253

参考文献 (References)

Aramwit P, Siritientong T, Srichana T, 2012. Potential applications of silk sericin, a natural protein from textile industry by-products. *Waste Management & Research*, 30(3): 217–224.

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G,

de-Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H, 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40: W597–W603.

Bailey TL, Johnson J, Grant CE, Noble WS, 2015. The MEME suite.

- Nucleic Acids Research*, 43: W39–W49.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW, 2018. GenBank. *Nucleic Acids Research*, 41: D36–D42.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T, 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15): 1972–1973.
- Chang HP, Cheng TC, Wu YQ, Hu WB, Long RW, Liu C, Zhao P, Xia QY, 2015. Transcriptomic analysis of the anterior silk gland in the domestic silkworm (*Bombyx mori*)-insight into the mechanism of silk formation and spinning. *PLoS ONE*, 10(9): e0139424.
- Chang HP, 2017. Omics Analysis of the silkworm silk gland and research on the functions of cuticular proteins in the silk gland. Doctor dissertation. Chongqing: Southwest University. [常怀普, 2017. 家蚕丝腺组学分析及丝腺表皮蛋白功能研究. 博士学位论文. 重庆: 西南大学.]
- Chao JN, Tang FR, Xu L, 2022. Developments in algorithms for sequence alignment: A review. *Biomolecules*, 12(4): 546.
- Chen CJ, Wu Y, Li JW, Wang X, Zeng ZH, Xu J, Liu YL, Feng JT, Chen H, He YH, Xia R, 2023. TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Molecular Plant*, 16(11): 1733–1742.
- Chou KC, Shen HB, 2008. Cell-PLoc: A package of Web servers for predicting subcellular localization of proteins in various organisms. *Nature Protocols*, 3(2): 153–162.
- Dong ZM, Guo KY, Zhang XL, Zhang T, Zhang Y, Ma SY, Chang HP, Tang MY, An LN, Xia QY, Zhao P, 2019. Identification of *Bombyx mori* sericin 4 protein as a new biological adhesive. *International Journal of Biological Macromolecules*, 132: 1121–1130.
- Geyer R, Wee S, Anderson S, Yates J, Wolf DA, 2003. BTB/POZ domain proteins are putative substrate adaptors for cullin 3 ubiquitin ligases. *Molecular Cell*, 12(3): 783–790.
- Haasdijk E, Heinerman J, 2018. Quantifying selection pressure. *Evolutionary Computation*, 26(2): 213–235.
- Hao AP, Han XY, Xue JK, Wei JC, Ren RY, 2023. Bioinformatics analysis of DBB gene family in rice. *Southwest China Journal of Agricultural Sciences*, 36(10): 2135–2143. [郝爱平, 韩雪杨, 薛巨坤, 魏继承, 任如意, 2023. 水稻 DBB 基因家族的生物信息学分析. 西南农业学报, 36(10): 2135–2143.]
- Hao ZD, Lü DK, Ge Y, Shi JS, Weijers D, Yu GC, Chen JH, 2020. *Rideogram*: Drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science*, 6: e251.
- Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai KT, 2007. WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, 35: W585–W587.
- Inoue S, Tanaka K, Arisaka F, Kimura S, Ohtomo K, Mizuno S, 2000. Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6 : 6 : 1 molar ratio. *Journal of Biological Chemistry*, 275(51): 40517–40528.
- Katoh K, Standley DM, 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4): 772–780.
- Kawamoto M, Jouraku A, Toyoda A, Yokoi K, Minakuchi Y, Katsuma S, Fujiyama A, Kiuchi T, Yamamoto K, Shimada T, 2019. High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 107: 53–62.
- Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y, 2007. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nature Cell Biology*, 9(6): 660–665.
- Letunic I, Bork P, 2021. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49: W293–W296.
- Leyva-Díaz E, 2023. CUT homeobox genes: Transcriptional regulation of neuronal specification and beyond. *Frontiers in Cellular Neuroscience*, 17: 1233830.
- Li B, Dewey CN, 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12: 323.
- Li CL, Tong XL, Zuo WD, Hu H, Xiong G, Han MJ, Gao R, Luan Y, Lu KP, Gai TT, Xiang ZH, Lu C, Dai FY, 2020. The beta-1, 4-N-acetylglucosaminidase 1 gene, selected by domestication and breeding, is involved in cocoon construction of *Bombyx mori*. *PLoS Genetics*, 16(7): e1008907.
- Li J, 2017. Screening and analysis of genes related with the characters of cocoon and silk production in silkworm. Master dissertation. Zhenjiang: Jiangsu University of Science and Technology. [李娟, 2017. 家蚕茧丝产量相关基因的挖掘与分析. 硕士学位论文. 镇江: 江苏科技大学.]
- Li JZ, 2013. Molecular design of silk protein gene and study on improving silk yield of practical varieties by overexpressing RasICA gene. Master dissertation. Suzhou: Soochow University. [李静芝, 2013. 丝蛋白基因的分子设计及过表达 RasICA 基因提高实用品种产丝量研究. 苏州: 苏州大学.]
- Li SM, Lao JJ, Sun Y, Hua XT, Lin P, Wang F, Shen GW, Zhao P, Xia QY, 2024. CRISPR/Cas9-mediated editing of *BmEckL1* gene sequence affected silk gland development of silkworms (*Bombyx mori*). *International Journal on Multicultural Societies*,

- 25(3): 1907.
- Liu SQ, Tian H, Xu YS, Wang HB, 2023. Juvenile hormone regulates silk gene expression by m⁶A RNA methylation. *Cellular and Molecular Life Sciences*, 80(11): 331.
- Long AX, 2022. Identification of improvement-associated genes of *Bombyx mori* and development of SV molecular markers related to silk yield. Master dissertation. Chongqing: Southwest University. [龙安星, 2022. 家蚕实用种改良相关基因鉴定与蚕丝产量相关 SV 分子标记开发. 硕士学位论文. 重庆: 西南大学.]
- Long DP, Liu RP, Huang Y, Fu AY, Zhang YL, Hao ZZ, Li Q, Xu HF, Xiang ZH, Zhao AC, 2024. An efficient and safe strategy for germ cell-specific automatic excision of foreign DNA in F1 hybrid transgenic silkworms. *Insect Science*, 31(1): 28–46.
- Long DP, Lu WJ, Hao ZZ, Xiang ZH, Zhao AC, 2016. Highly efficient and inducible DNA excision in transgenic silkworms using the FLP/FRT site-specific recombination system. *Transgenic Research*, 25(6): 795–811.
- Love MI, Huber W, Anders S, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12): 550.
- Lu F, Wei ZY, Luo YJ, Guo HL, Zhang GQ, Xia QY, Wang Y, 2020. SilkDB 3.0: Visualizing and exploring multiple levels of data for silkworm. *Nucleic Acids Research*, 48: D749–D755.
- Ma SJ, Xu YJ, He K, Ma RF, Zhu Y, 2023. Molecular evolution and expression patterns of a multigene family of toll-like receptors in ruminants. *Acta Veterinaria et Zootechnica Sinica*, 54(9): 3722–3734. [马淑娟, 徐伟洁, 何珂, 马瑞丰, 朱英, 2023. 反刍动物 Toll 样受体多基因家族的分子进化及表达模式分析. 畜牧兽医学报, 54(9): 3722–3734.]
- Ma Y, Zeng WH, Ba YB, Luo Q, Ou Y, Liu RP, Ma JW, Tang YY, Hu J, Wang HM, Tang X, Mu YY, Li QJ, Chen YQ, Ran YT, Xiang ZH, Xu HF, 2022. A single-cell transcriptomic atlas characterizes the silk-producing organ in the silkworm. *Nature Communication*, 13(1): 3316.
- Mei Y, Jing D, Tang SY, Chen X, Chen H, Duanmu HN, Cong YY, Chen MY, Ye XH, Zhou H, He K, Li F, 2022. InsectBase 2.0: A comprehensive gene resource for insects. *Nucleic Acids Research*, 50: D1040–D1045.
- Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-I T, Abe H, Shimada T, Morishita S, Sasaki T, 2004. The genome sequence of silkworm, *Bombyx mori*. *DNA Research*, 11(1): 27–35.
- Nagaraja GM, Nagaraju J, 1995. Genome fingerprinting of the silkworm, *Bombyx mori*, using random arbitrary primers. *Electrophoresis*, 16(9): 1633–1638.
- Ou J, Deng HM, Zheng SC, Huang LH, Feng QL, Liu L, 2014. Transcriptomic analysis of developmental features of *Bombyx mori* wing disc during metamorphosis. *BMC Genomics*, 15(1): 820.
- Pan YO, Zeng XC, Wen SY, Gao XW, Liu XM, Tian FY, Shang QL, 2020. Multiple ATP-binding cassette transporters genes are involved in thiamethoxam resistance in *Aphis gossypii* glover. *Pesticide Biochemistry and Physiology*, 167: 104558.
- Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi HY, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, 2023. InterPro in 2022. *Nucleic Acids Research*, 51: D418–D427.
- Peterson E, Shippee E, Brinton MA, Kaur P, 2019. Biochemical characterization of the mouse ABCF3 protein, a partner of the flavivirus-resistance protein OAS1B. *Journal of Biological Chemistry*, 294(41): 14937–14952.
- Ruan JH, Wu MY, Ye XG, Zhao S, Liang JS, Ye LP, You ZY, Zhong BX, 2021. Comparative mRNA and lncRNA analysis of the molecular mechanisms associated with low silk production in *Bombyx mori*. *Frontiers in Genetics*, 11: 592128.
- Shi XD, Wu Q, Xiang DB, Wan Y, Zhao G, 2019. Bioinformatics analysis of GIS gene family in *Fagopyrum tataricum*. *Southwest China Journal of Agricultural Sciences*, 32(8): 1717–1722. [时小东, 吴琪, 向达兵, 万燕, 赵钢, 2019. 苦荞 GIS 基因家族生物信息学分析. 西南农业学报, 32(8): 1717–1722.]
- Song ZW, 2013. Effect of overexpression of Bmovo-1 on ovarian development of silkworm and secretion and expression of hIGF-I in silk gland biological factory. Master dissertation. Suzhou: Soochow University. [宋作伟, 2013. 过表达 Bmovo-1 对家蚕卵巢发育的影响及丝腺生物工厂分泌表达 hIGF-I. 硕士学位论文. 苏州: 苏州大学.]
- Tsubota T, Tomita S, Uchino K, Kimoto M, Takiya S, Kajiwarra H, Yamazaki T, Sezutsu H, 2016. A hox gene, antennapedia, regulates expression of multiple major silk protein genes in the silkworm *Bombyx mori*. *Journal of Biological Chemistry*, 291(13): 7087–7096.
- Wang F, Xu HF, Yuan L, Ma SY, Wang YC, Duan XL, Duan JP, Xiang ZH, Xia QY, 2013. An optimized sericin-1 expression system for mass-producing recombinant proteins in the middle silk glands of transgenic silkworms. *Transgenic Research*, 22(5): 925–938.
- Wang J, Xia Q, He X, Dai M, Ruan J, Chen J, Yu G, Yuan H, Hu Y,

- Li R, Feng T, Ye C, Lu C, Wang J, Li S, Wong GK, Yang H, Wang J, Xiang Z, Zhou Z, Yu J, 2005. SilkDB: A knowledgebase for silkworm biology and genomics. *Nucleic Acids Research*, 33: D399–D402.
- Wang LL, Jin XH, Cai MY, Li HG, Chen JW, Wang FW, Wang CY, Hu WW, Liu F, Xie D, 2018. AGBL2 promotes cancer cell growth through IRGM-regulated autophagy and enhanced Aurora A activity in hepatocellular carcinoma. *Cancer Letters*, 414: 71–80.
- Wang SH, 2017. The investigation of molecular mechanism underlying silk production and the modification of silk property in silkworm, *Bombyx mori*. Doctor dissertation. Hangzhou: Zhejiang University. [王少华, 2017. 家蚕丝产量性状分子调控机理的研究及蚕丝性能的改造. 博士学位论文. 杭州: 浙江大学.]
- Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL, 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, 35(3): 773–777.
- Wei JH, Fan YP, Jing XL, Fei ZH, Li CF, Pan GQ, Bao JL, Zhou ZY, 2022. Establishment of a novel baculovirus-silkworm expression system. *Microorganisms*, 10(5): 1013.
- Wu SY, Tong XL, Li CL, Lu KP, Tan D, Hu H, Liu H, Dai FY, 2019. Genome-wide identification and expression profiling of the C₂H₂-type zinc finger protein genes in the silkworm *Bombyx mori*. *PeerJ*, 7: e7222.
- Xia QY, Zhou ZY, Lu C, Cheng DJ, Dai FY, Li B, Zhao P, Zha XF, Cheng TC, Chai CL, Pan GQ, Xu JS, Liu C, Lin Y, Qian JF, Hou Y, Wu ZL, Li GR, Pan MH, Li CF, Shen YH, Lan XQ, Yuan LW, Li T, Xu HF, Yang GW, Wan YJ, Zhu Y, Yu MD, Shen WD, Wu DY, Xiang ZH, Yu J, Wang J, Li RQ, Shi JP, Li H, Li GY, Su JN, Wang XL, Li GQ, Zhang ZJ, Wu QF, Li J, Zhang QP, Wei N, Xu JZ, Sun HB, Dong L, Liu DY, Zhao SL, Zhao XL, Meng QS, Lan FD, Huang XG, Li YZ, Fang L, Li CF, Li DW, Sun YQ, Zhang ZP, Yang Z, Huang YQ, Xi Y, Qi QH, He DD, Huang HY, Zhang XW, Wang ZQ, Li WJ, Cao YZ, Yu YP, Yu H, Li JH, Ye JH, Chen H, Zhou Y, Liu B, Wang J, Ye J, Ji H, Li ST, Ni PX, Zhang JG, Zhang Y, Zheng HK, Mao BY, Wang W, Ye C, Li SG, Wang J, Wong GK, Yang HM, Group BA, 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 306(5703): 1937–1940.
- Xu S, Wang F, Wang YC, Wang RY, Hou K, Tian C, Ji YT, Yang QQ, Zhao P, Xia QY, 2019. A silkworm based silk gland bioreactor for high-efficiency production of recombinant human lactoferrin with antibacterial and anti-inflammatory activities. *Journal of Biological Engineering*, 13: 61.
- Yang ZH, 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8): 1586–1591.
- Yin ZH, Nie HT, Jiang KY, Yan XW, 2022. Molecular mechanisms underlying *Vibrio* tolerance in *Ruditapes philippinarum* revealed by comparative transcriptome profiling. *Frontiers in Immunology*, 13: 879337.
- Yoon HS, Fujino K, Liu SK, Takano T, Tsugama D, 2021. NDR/LATS-family protein kinase genes are indispensable for embryogenesis in *Arabidopsis*. *FEBS Open Bio*, 11(9): 2600–2606.
- Yu Q, Zhou QS, Liu QX, 2011. Advances in the research on the nervous system of the silkworm, *Bombyx mori*. *Acta Entomologica Sinica*, 54(10): 1172–1180. [于奇, 周启升, 刘庆信, 2011. 家蚕神经系统研究进展. 昆虫学报, 54(10): 1172–1180.]
- Zhang W, Li Z, Lan W, Guo H, Chen F, Wang F, Shen G, Xia Q, Zhao P, 2022. Bioengineered silkworm model for expressing human neurotrophin-4 with potential biomedical application. *Frontiers in Physiology*, 13: 1104929.
- Zhao LC, Sun X, Wang XY, Qin S, Kong YH, Li MW, 2022. *Bombyx mori* Vps13d is a key gene affecting silk yield. *PLoS ONE*, 17(7): e0270840.