

钟 静,胡 颖,陈亚波,等. 玉米 FLA 蛋白家族的生物信息学分析[J]. 江苏农业科学,2017,45(7):23-28.
doi:10.15889/j.issn.1002-1302.2017.07.005

玉米 FLA 蛋白家族的生物信息学分析

钟 静^{1,2}, 胡 颖², 陈亚波¹, 宋群星¹

(1. 湖北第二师范学院植物抗癌活性物质提纯与应用湖北省重点实验室,湖北武汉 430205;

2. 武汉大学杂交水稻国家重点实验室,湖北武汉 430072)

摘要:类成束阿拉伯半乳糖蛋白(fasciclin-like arabinogalactan proteins,简称 FLAs)是一类广泛分布于植物体内的富含羟脯氨酸的糖蛋白,在植物生长发育和形态构建中发挥着重要作用。基于已公布的玉米蛋白数据库确定了 26 个 FLA 蛋白,并对其理化性质、系统发生树、蛋白结构和功能域等进行了分析。结果表明,FLA 蛋白氨基酸长度在 249~682 个之间,理论等电点在 4.99~10.76 之间,主要定位在质膜上,多数为疏水性蛋白;二级结构由 α -螺旋、无规则卷曲和延伸链等元件组成;部分蛋白空间结构具有较强保守性。

关键词:玉米;FLA;蛋白家族;基因家族;生物信息学;理化性质;结构特征

中图分类号: Q943.2 **文献标志码:** A **文章编号:** 1002-1302(2017)07-0023-05

阿拉伯半乳糖蛋白(arabinogalactan proteins,简称 AGPs)是一类富含羟脯氨酸的糖蛋白^[1]。根据核心蛋白骨架差异,AGPs 被分为经典 AGPs(classical AGPs)、非经典 AGPs(non-classical AGPs)、赖氨酸富集 AGPs(lysine-rich AGPs)、AG 多肽(AG peptides)、类成束蛋白 AGPs(fasciclin-like arabinogalactan proteins,简称 FLAs)、类木质形成素 AGPs(xylogen-like AGPs,简称 XLAs)和类植物蓝素 AGPs(phytoeyanin-like AGPs,简称 PLAs)等 7 类^[2-4]。AGPs 的蛋白骨架一般包括 N 端信号肽和 1 段长度可变且富含脯氨酸(Pro)、丙氨酸(Ala)、丝氨酸(Ser)、苏氨酸(Thr)的区域(PAST)^[1]。经典 AGPs 还包含 1 个 C 端 GPI 锚定信号。FLAs 作为 AGPs 的 1 个亚类,除具有 AGPs 的典型结构之外,还含有 1~2 个成束蛋白结构域(fasciclin domains,简称 FAS)^[5]。

FLAs 在植物生长发育过程中发挥重要作用。有研究显示,FLAs 参与了茎、侧根、花粉小孢子发育,以及棉花纤维细胞的伸长等过程^[6-8],此外还可影响植株茎的强度和弹性^[9]。对其作用机制的研究表明,FLAs 可能是通过影响细胞壁的形成从而发挥其生物学功能^[10-13]。目前已经在拟南芥、油菜、水稻、小麦中分别鉴定出了 21、33、27、34 个 FLAs^[3,5,14-15],但是在重要的粮食和经济作物玉米中还未见关于 FLAs 的报道。因此,本研究采用生物信息学方法,对玉米 FLAs 蛋白家族成员进行全面的鉴定和系统进化分析,并进一步分析其蛋白质产物的理化性质、结构特征和功能域等特点,以期为进一步揭示植物 FLAs 的结构和功能奠定基础。

1 材料与方法

1.1 序列来源

收稿日期:2016-05-06

基金项目:湖北省自然科学基金(编号:2015CFC884)。

作者简介:钟 静(1979—),女,湖北荆州人,博士,讲师,主要从事发育生物学研究。Tel:027(87943931);E-mail:jjing2003_1@163.com。

分别从 Pfam(<http://pfam.sanger.ac.uk/>)、Interpro(<http://www.ebi.ac.uk/interpro/>)数据库中下载玉米中含有 FAS 结构域的蛋白序列^[16-17]。将获得的序列提交到 maizeG-DB(<http://www.maizegdb.org/>)数据库中进行比对,获得目标蛋白质的全长序列。从拟南芥官方数据库 TAIR(<http://www.arabidopsis.org/>)中下载获得拟南芥 FLAs 基因和蛋白序列数据。

1.2 序列分析

利用 Compute pI/Mw tool(http://web.expasy.org/compute_pi/)计算相对分子量及理论等电点。利用 SignalP 4.1(<http://www.cbs.dtu.dk/services/SignalP/>)、Plant-mPLoc(<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>)分析蛋白信号肽和亚细胞定位^[18]。利用 ProtScale(<http://web.expasy.org/cgi-bin/protscale/protscale.pl>)分析蛋白亲水性/疏水性。利用 NPS(http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.plpage=/NPSA/npsa_hnn.html)分析蛋白质的二级结构。利用 SWISS-MODEL(<http://swissmodel.expasy.org/>)分析蛋白质的三级结构。

1.3 模体识别

利用 MEME program3(<http://meme-suite.org/>)模体搜索工具识别 FLAs 家族相关蛋白质所共有的模体,并对相关参数进行修改,将可找到的模体数最大值调整为 15 个,每个模体的最大宽度调整为 100,其他均为默认值^[19]。

1.4 系统进化树的构建

应用多序列比对工具 ClustalX 1.83 对玉米、拟南芥 FLAs 氨基酸全序列进行比对^[20]。采用 MEGA7.0 软件中的邻接法(neighbor-joining,简称 NJ)构建系统发育树^[21]。通过随机逐步比较的方法搜索最佳系统进化树,对生成的系统树进行 Bootstrap 校正。

2 结果与分析

2.1 玉米 FLAs 家族成员的鉴定和理化性质分析

分别从 Pfam、Interpro 数据库下载玉米中含有 FAS 结构

域的蛋白序列,并将获得的序列分别提交到 maizeGDB 数据库中进行比对,获得其全长序列。对来源于 2 个数据库中的蛋白质全长进行交叉比对,去除重复序列。进一步将获得的蛋白序列提交到 SignalP 4.1 检测 N 端信号肽,不含有信号肽的不属于 AGPs 家族。对含有信号肽的序列手动查找其中富含 PAST 的区域,并将蛋白序列提交到 PSORT 检测 C 端糖基磷脂酰肌醇(GPI)锚定信号,最终获得了 26 个玉米 FLAs 并分别命名为 ZmFLA1 ~ ZmFLA26(表 1)。

对这 26 个 ZmFLAs 进行氨基酸理化性质分析表明,氨基酸序列长度在 249 ~ 682 个之间,分子量在 25 175. 47 ~

72 410. 16 u 之间,理论等电点在 4. 99 ~ 10. 76 之间(表 1)。亚细胞定位分析显示,绝大多数 ZmFLAs 都定位在质膜上,其中 ZmFLA8、ZmFLA14 除定位在质膜上外,还可能定位在核中,而 ZmFLA5 则仅定位在核上(表 1)。亲水性/疏水性分析显示,ZmFLA2/6/24 亲水区域明显大于疏水区域,说明这 3 个 ZmFLA 蛋白亲水性较强,为亲水性蛋白;其余 23 个 ZmFLA 蛋白为疏水性蛋白(表 1)。N-糖基化位点检测结果显示,所有 ZmFLAs 都至少含有 1 个 N-糖基化位点,ZmFLA22 甚至包含 7 个 N-糖基化位点(表 1),说明 ZmFLAs 蛋白骨架可能具有较高的糖基化修饰率。

表 1 玉米 FLAs 蛋白的理化性质分析结果

蛋白名称	基因座位号	氨基酸数量 (个)	分子量 (u)	等电点	亚细胞定位	亲疏水性	N-糖基化位点数 (个)
ZmFLA1	GRMZM2G065718	392	41 294. 19	6. 95	质膜	疏水	3
ZmFLA2	GRMZM2G329181	465	50 042. 17	6. 08	质膜	亲水	2
ZmFLA3	GRMZM2G003752	429	44 033. 16	6. 71	质膜	疏水	5
ZmFLA4	GRMZM2G403276	271	27 372. 26	6. 83	质膜	疏水	3
ZmFLA5	GRMZM5G862219	270	27 711. 47	4. 99	质膜	疏水	4
ZmFLA6	GRMZM2G177242	272	27 404. 05	9. 30	质膜	亲水	3
ZmFLA7	GRMZM2G003165	410	43 923. 42	8. 50	质膜	疏水	4
ZmFLA8	AC209784. 3_FG011	682	72 410. 16	9. 61	质膜、核	疏水	6
ZmFLA9	GRMZM2G022931	249	25 175. 47	9. 07	质膜	疏水	4
ZmFLA10	GRMZM2G301908	249	25 175. 47	9. 07	质膜	疏水	3
ZmFLA11	AC234156. 1_FG005	269	28 212. 65	9. 34	质膜	疏水	3
ZmFLA12	AC184130. 4_FG005	297	30 427. 41	10. 76	质膜	疏水	4
ZmFLA13	GRMZM2G027825	473	50 469. 57	6. 13	质膜	疏水	2
ZmFLA14	GRMZM2G109291	296	29 933. 80	6. 20	质膜、核	疏水	3
ZmFLA15	GRMZM2G329710	675	71 960. 31	5. 99	核	疏水	1
ZmFLA16	GRMZM2G177142	276	28 778. 44	10. 75	质膜	疏水	4
ZmFLA17	AC213621. 5_FG004	430	43 982. 14	5. 16	质膜	疏水	6
ZmFLA18	GRMZM2G021794	273	28 108. 07	9. 04	质膜	疏水	1
ZmFLA19	GRMZM2G133053	268	26 696. 55	8. 43	质膜	疏水	3
ZmFLA20	GRMZM2G174799	265	27 451. 70	8. 56	质膜	疏水	4
ZmFLA21	GRMZM2G001514	282	28 748. 74	8. 99	质膜	疏水	1
ZmFLA22	GRMZM2G421415	430	44 120. 48	6. 53	质膜	疏水	7
ZmFLA23	GRMZM2G035933	426	43 630. 80	5. 50	质膜	疏水	4
ZmFLA24	GRMZM2G144610	472	48 269. 44	5. 58	质膜	亲水	3
ZmFLA25	GRMZM2G341698	278	28 740. 49	10. 65	质膜	疏水	3
ZmFLA26	GRMZM2G003917	275	29 270. 61	5. 51	质膜	疏水	3

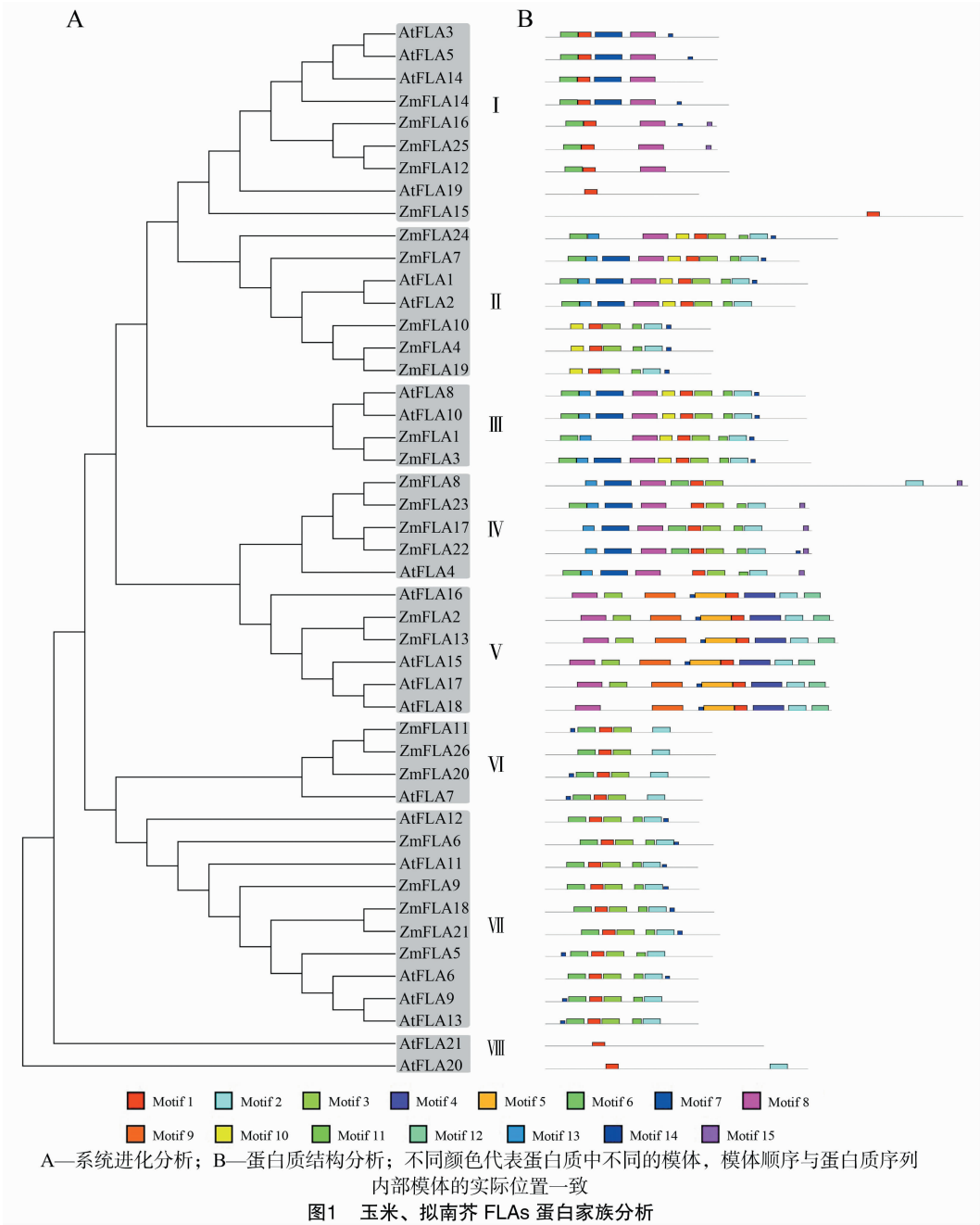
2.2 ZmFLAs 系统进化和序列特异性分析

对玉米、拟南芥中的 FLAs 进行系统进化分析显示,26 个 ZmFLAs、21 个 AtFLAs 被聚类在 8 个独立的进化分支中(图 1-A)。分支Ⅷ仅含有 2 个 AtFLAs,其余 7 个分支都同时有 ZmFLAs、AtFLAs 的分布。但是在同一分支中的 ZmFLAs、AtFLAs 并无明显交叉分布,来源于不同物种的 FLAs 通常分别聚类在小的分支中。例如分支Ⅱ含有 2 个拟南芥 FLAs、5 个玉米 FLAs,其中 AtFLA1、AtFLA2 聚类在 1 个小分支中,而 ZmFLA4/10/19 聚类在另 1 个小分支中。进一步的序列特异性分析显示,尽管玉米、拟南芥的 FLAs 都含有模体 1,但是不同的 FLAs 含有其他不同的模体。同时,处于同一个进化分支中的 FLAs 往往具有类似的模体。例如进化分支Ⅰ中,AtFLA3/5/14、ZmFLA14 都含有模体 1、模体 7、模体 8、模体 11,ZmFLA12/16/25 则含有模体 1、模体 8、模体 11;模体 10 仅出现在分支Ⅱ、Ⅲ的蛋白质序列中,而模体 4、模体 5、模

体 9 仅出现在分支Ⅴ中(图 1-B)。

2.3 ZmFLAs 蛋白结构分析

通过 NPS 程序对这 26 个 ZmFLAs 蛋白序列进行二级结构分析表明,蛋白质均由 α-螺旋、无规则卷曲和延伸链等结构元件组成,但是这 3 种元件的比例和分布存在差异。其中无规则卷曲所占比例最高,在 38. 05% ~ 58. 97% 之间;其次为 α-螺旋,在 25. 93% ~ 46. 83% 之间;延伸链所占比例最低,在 8. 86% ~ 21. 63% 之间(表 2)。进一步利用 Swiss-Model 同源建模的方法预测这 26 个 ZmFLAs 蛋白的三级结构。从图 2 可以看出,三级结构主要是由 α-螺旋、无规则卷曲和延伸链等元件组成,其中 ZmFLA1/2/3/7/13/17/22/23/24 空间结构较为保守;ZmFLA8/12/14/16/18/19/25 与这些蛋白类似,但是空间结构明显较为松散;其余 ZmFLAs 的结构则与上述蛋白存在明显差异,其中 ZmFLA5/10/11/20 结构较类似。

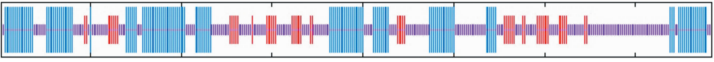
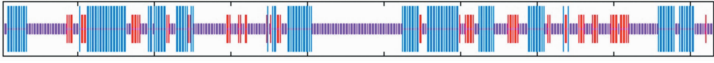
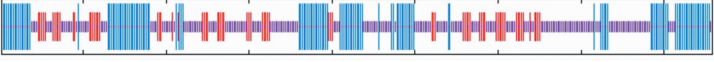
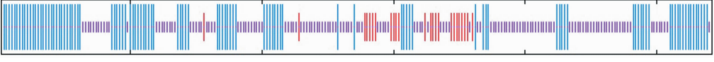
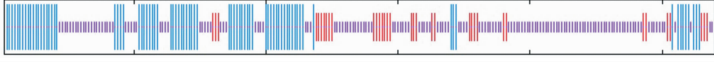
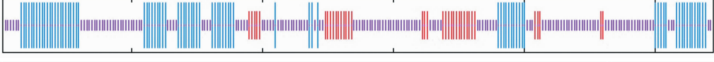
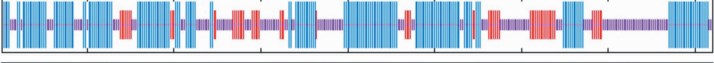
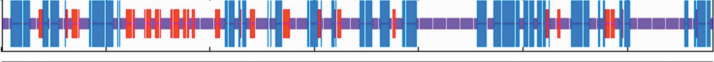
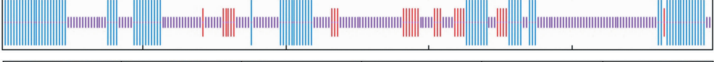
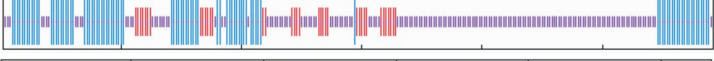
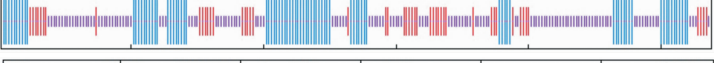
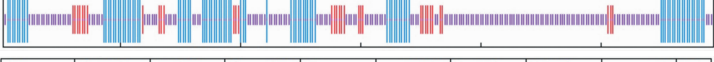
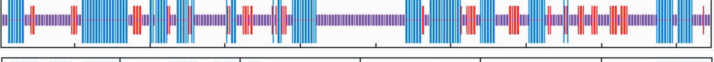
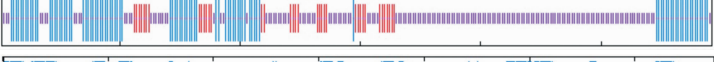

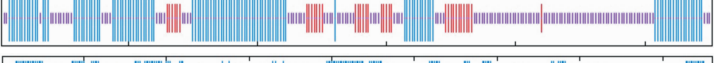
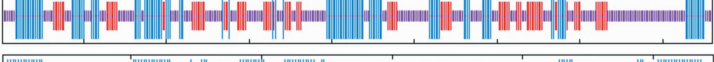
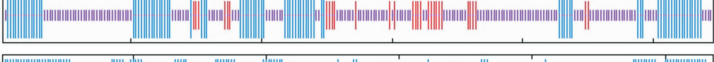
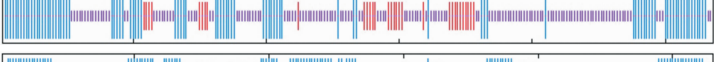
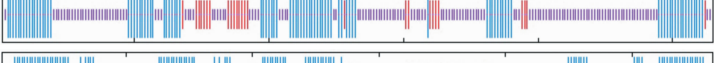
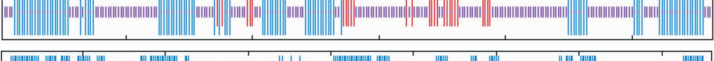
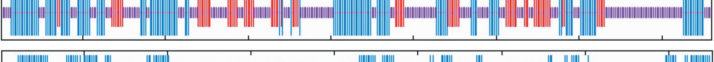
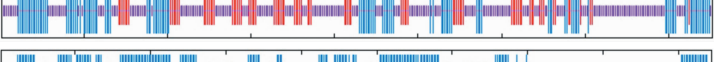
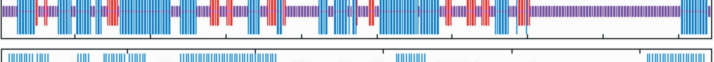
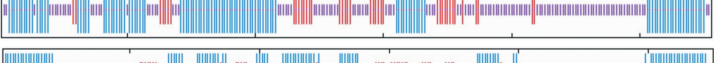
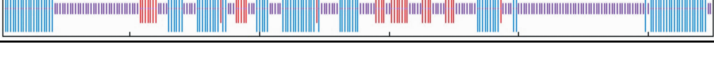


3 结论与讨论

本研究通过多种数据库和在线软件,对玉米 FLA 蛋白基因家族进行了全面的鉴定和分析。结果显示,从玉米全基因组中共鉴定出了 26 个 FLAs,对其理化性质的研究显示,大多数 ZmFLAs 都是疏水性氨基酸。亚细胞定位预测结果显示,绝大多数 ZmFLAs 定位在质膜上,这可能与 FLAs 参与细胞间信号转导有关^[13]。系统进化分析结果显示,玉米、拟南芥 FLAs 被聚类在 8 个独立的进化分支中,每个进化分支中都有玉米、拟南芥 FLAs 蛋白的分布,暗示在单子叶、双子叶植物中,该家族成员可能拥有共同的祖先。但是,在每个进化分支中,不同物种来源的 FLAs 蛋白并没有呈现交叉分布,这说明该蛋白基因家族的扩张可能是发生在单子叶、双子叶植物进

化之后,并且其扩张在单子叶、双子叶植物中是独立进行的。目前认为,蛋白质的一级结构决定二级结构,二级结构又决定着三级结构,而蛋白质的功能往往取决于其空间结构。本研究对 26 个 ZmFLAs 的结构进行了分析。二级结构分析结果表明,ZmFLAs 蛋白包含 α -螺旋、无规则卷曲和延伸链等元件。其中无规则卷曲所占比例最高, α -螺旋次之,延伸链比例最低。同时,各个 ZmFLAs 中不同元件的比例和分布也存在差异。进一步的三级结构分析表明,尽管部分 ZmFLAs 的空间结构十分保守或者类似,但是其拓扑结构和聚合角度也存在明显差异,这暗示不同的 ZmFLAs 可能具有各自独特的功能。已有关于 FLAs 的报道主要集中在双子叶模式植物拟南芥中,鲜见对单子叶经济作物中 FLAs 生物学功能的相关报

表 2 玉米 FLA 蛋白二级结构分析结果

蛋白名称	二级结构元件比例(%)			二级结构元件分布
	α -螺旋	无规则卷曲	延伸链	
ZmFLA1	35.20	50.77	14.03	
ZmFLA2	33.76	51.18	15.05	
ZmFLA3	31.24	51.98	16.78	
ZmFLA4	39.11	52.03	8.86	
ZmFLA5	30.00	57.04	12.96	
ZmFLA6	30.15	56.25	13.06	
ZmFLA7	46.83	38.05	15.12	
ZmFLA8	35.04	51.17	13.78	
ZmFLA9	32.93	56.22	10.84	
ZmFLA10	37.08	52.43	10.49	
ZmFLA11	31.23	50.93	17.84	
ZmFLA12	29.63	58.92	11.45	
ZmFLA13	34.67	50.74	14.59	
ZmFLA14	30.41	57.43	12.16	
ZmFLA15	25.93	58.96	15.11	
ZmFLA16	40.58	46.38	13.04	
ZmFLA17	28.60	49.77	21.63	
ZmFLA18	30.40	58.97	10.62	
ZmFLA19	33.58	54.85	11.57	
ZmFLA20	34.34	55.85	9.81	
ZmFLA21	35.46	55.32	9.22	
ZmFLA22	34.65	46.98	18.37	
ZmFLA23	26.76	53.52	19.72	
ZmFLA24	37.92	50.21	11.86	
ZmFLA25	39.21	46.76	14.03	
ZmFLA26	35.27	52.36	12.36	

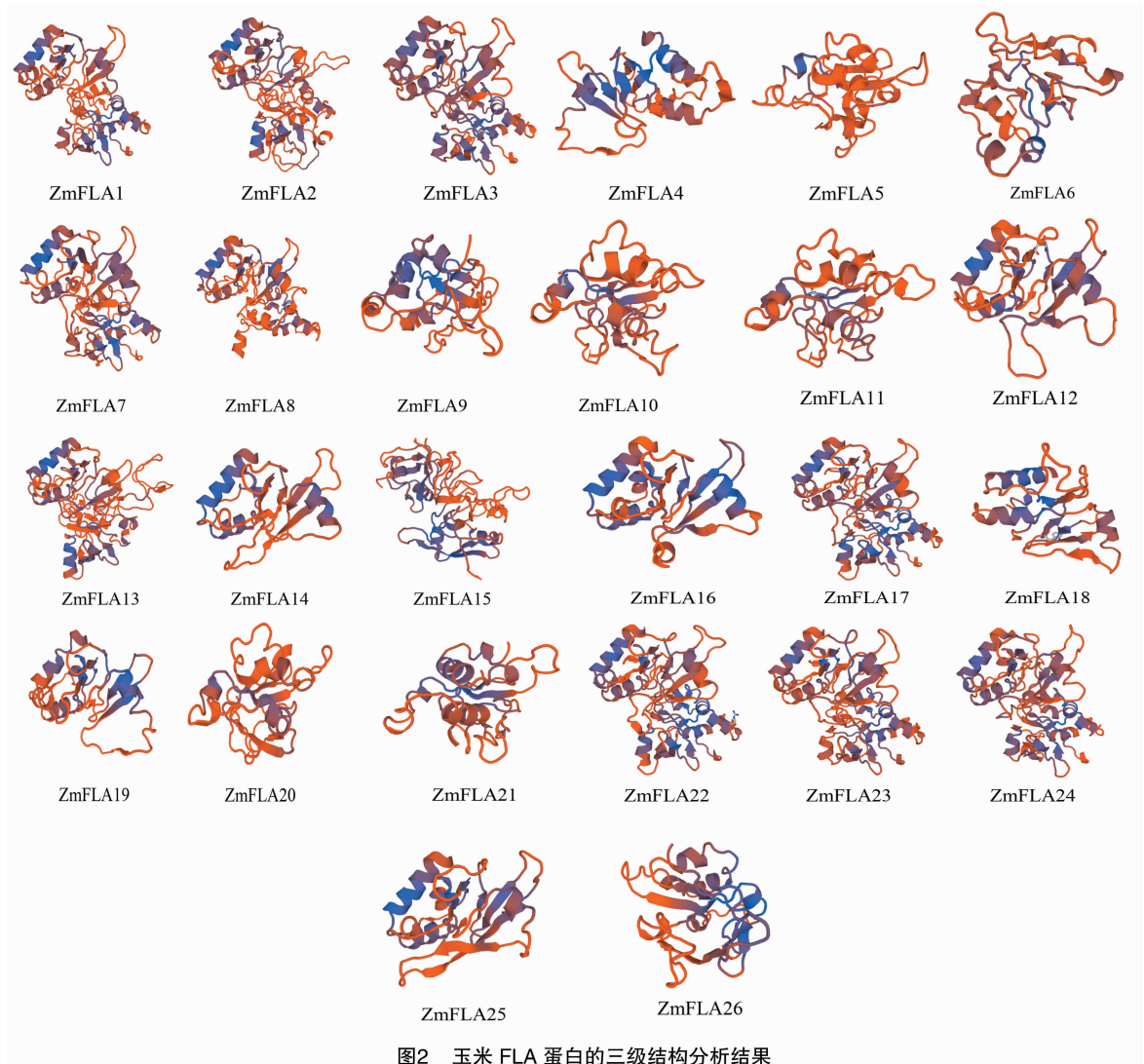


图2 玉米 FLA 蛋白的三级结构分析结果

道。本研究应用生物信息学的方法鉴定了玉米 FLAs 蛋白家族成员,并对 FLAs 的基本理化性质、亚细胞定位、系统进化、二级结构和三级结构等进行了详细分析,以期为进一步探讨 ZmFLAs 的生物学功能提供理论基础。

参考文献:

- [1] Ellis M, Egelund J, Schultz C, et al. Arabinogalactan - proteins: key regulators at the cell surface? [J]. *Plant Physiology*, 2010, 153 (2): 403 - 419.
- [2] Schultz C, Johnson K, Currie G, et al. The classical arabinogalactan protein gene family of *Arabidopsis* [J]. *The Plant Cell*, 2000, 12 (9): 1751 - 1768.
- [3] Ma H, Zhao J. Genome - wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.) [J]. *Journal of Experimental Botany*, 2010, 61 (10): 2647 - 2668.
- [4] 马浩力, 余礼, 梁荣洪, 等. 高等植物阿拉伯半乳糖蛋白的功能研究 [J]. *中国科学: 生命科学*, 2015, 45 (2): 113 - 123.
- [5] Johnson K L, Jones B J, Bacic A, et al. The fasciclin - like arabinogalactan proteins of *Arabidopsis*. A multigene family of putative cell adhesion molecules [J]. *Plant Physiology*, 2003, 133 (4): 1911 - 1925.

- [6] Li J, Yu M, Geng L, et al. The fasciclin - like arabinogalactan protein gene, FLA3, is involved in microspore development of *Arabidopsis* [J]. *The Plant Journal: for Cell and Molecular Biology*, 2010, 64 (3): 482 - 497.
- [7] Johnson K, Kibble N, Bacic A, et al. A fasciclin - like arabinogalactan - protein (FLA) mutant of *Arabidopsis thaliana*, flal1, shows defects in shoot regeneration [J]. *PLoS One*, 2011, 6 (9): e25154.
- [8] Huang G, Gong S, Xu W, et al. A fasciclin - like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton [J]. *Plant Physiology*, 2013, 161 (3): 1278 - 1290.
- [9] Macmillan C, Mansfield S, Stachurski Z, et al. Fasciclin - like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in *Arabidopsis* and *Eucalyptus* [J]. *The Plant Journal: for Cell and Molecular Biology*, 2010, 62 (4): 689 - 703.
- [10] Shi H, Kim Y, Guo Y, et al. The *Arabidopsis* SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion [J]. *The Plant Cell*, 2003, 15 (1): 19 - 32.
- [11] Persson S, Wei H, Milne J, et al. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102 (24): 8633 - 8638.

王慧丽, 江 娟. 锯末牛粪堆肥微生物多样性的宏基因组学分析[J]. 江苏农业科学, 2017, 45(7): 28–32.
doi:10.15889/j.issn.1002-1302.2017.07.006

锯末牛粪堆肥微生物多样性的宏基因组学分析

王慧丽, 江 娟

(文华学院环境工程系, 湖北武汉 430074)

摘要:运用宏基因组技术研究不同比例锯末牛粪堆肥产物中微生物的组成和分布。结果表明, 牛粪与锯末质量比为 2 : 1、3 : 1、4 : 1 的 3 个堆肥样本中微生物可操作分类单元(operational taxonomic units, OTU)有较大的差异, 其中含量最高的 3 个门的微生物为变形菌门、拟杆菌门和厚壁菌门, 在不同样本中它们的含量也有较大差异。从属水平分析, 含量最高的是甲基单胞菌属、噬氢菌属、固氮菌属和寡养单胞菌等营养转化和合成的菌属。降解纤维素和木质素的也有一定的含量, 特别是假单胞菌属含量较高。经过对堆肥产物基本性质及其微生物的分析发现, 当牛粪与锯末的比例为 2 : 1 时, 堆肥样品中的营养转化和合成菌属以及去除污染菌属都高于其他 2 个样品, 堆肥效果较好。

关键词:牛粪堆肥; 宏基因组; 微生物; 锯末; 可操作分类单元(OUT); 分子生态学依据

中图分类号: S141.4 **文献标志码:** A **文章编号:** 1002-1302(2017)07-0028-05

近年来, 由于产业结构调整, 我国规模化养殖业发展迅速。以生产牛奶为主的奶牛养殖业产生了可观的经济效益, 但是同时也带来了大量的牛粪等固体有机废弃物, 为了保护环境、实现养殖业的可持续发展, 必须进行牛粪的无害化处理和资源化利用, 堆肥处理是目前有机固体废物处理的重要方式。堆肥的本质是微生物群落结构演替协同作用于堆肥基质的生化过程^[1-2]。在细菌和真菌等微生物的作用下, 降解有机物, 杀死病原菌, 促进有机质稳定化和腐殖化, 最终达到无

害化和资源化的目的, 并生成大量可被植物吸收利用的有效态氮、有效态磷、有效态钾化合物和其他有机物质作为土壤肥力活性物质。因此, 研究微生物的群落结构对于了解微生物的协同关系、调控堆肥以及研究高效稳定的复合菌系都有重要意义。堆肥过程中的微生物菌系复杂, 变化较大, 而且其中大多数菌系不能分离培养, 因而常规的试验方法和手段难以完全认识和真实反映微生物的群落组成^[3-4]。宏基因组学技术克服了传统微生物分析仅限于可分离培养组分的缺陷, 通过对样品中微生物基因序列的分析, 可以快速确定微生物的种类和丰度, 主要采用 16S rDNA 序列分析进行微生物的群落分析检测出样品中大量的低丰度生物。目前, 宏基因组学技术已经应用于海洋、土壤、石油以及人体胃肠道等多个方面^[5-7], 但用于堆肥研究的较少。在牛粪堆肥中, 通常以农作物秸秆、木质、锯末废弃物等作为辅料, 起到加速堆肥过程和保证堆肥效果的作用^[8]。本研究就地取材, 在牛粪中添加不同比例的锯末, 采用宏基因组学技术对不同比例锯末添加产

收稿日期: 2016-01-14

基金项目: 湖北省省级战略新兴(支柱)产业人才培养计划——环境工程(编号: j0900760102)。

作者简介: 王慧丽(1980—), 女, 陕西渭南人, 硕士, 讲师, 从事固体废物资源化利用、环境污染评价等研究。Tel: (027) 87583433; E-mail: lypp1229@126.com。

通信作者: 江 娟, 工程师, 从事环境微生物和固体废物资源化研究。Tel: (027) 87583433; E-mail: 14784021@qq.com。

- [12] Seifert G, Xue H, Acet T. The *Arabidopsis thaliana* fasciclin like arabinogalactan protein 4 gene acts synergistically with abscisic acid signalling to control root growth[J]. *Annals of Botany*, 2014, 114(6): 1125–1133.
- [13] Wang H, Jiang C, Wang C, et al. Antisense expression of the fasciclin-like arabinogalactan protein *FLA6* gene in *Populus* inhibits expression of its homologous genes and alters stem biomechanics and cell wall composition in transgenic trees[J]. *Journal of Experimental Botany*, 2015, 66(5): 1291–1302.
- [14] Faik A, Abouzouhair J, Sarhan F. Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (*Triticum aestivum*) and rice (*Oryza sativa*): identification and bioinformatic analyses[J]. *Molecular Genetics and Genomics*, 2006, 276(5): 478–494.
- [15] Showalter A, Keppler B, Lichtenberg J, et al. A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins[J]. *Plant Physiology*, 2010, 153(2): 485–513.

- [16] Finn R, Mistry J, Schuster-Böckler B, et al. Pfam: clans, web tools and services[J]. *Nucleic Acids Research*, 2006, 34: D247–D251.
- [17] Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: protein domains identifier[J]. *Nucleic Acids Research*, 2005, 33: W116–W120.
- [18] Petersen T N, Brunak S, von Heijne G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions[J]. *Nature Methods*, 2011, 8(10): 785–786.
- [19] Bailey T L, Williams N, Misleh C, et al. MEME: discovering and analyzing DNA and protein sequence motifs[J]. *Nucleic Acids Research*, 2006, 34: W369–W373.
- [20] Thompson J D, Gibson T J, Plewniak F, et al. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools[J]. *Nucleic Acids Research*, 1997, 25(24): 4876–4882.
- [21] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Molecular Biology and Evolution*, 2016, 33(7): 1870–1874.