

山核桃茶蔗子葡萄座腔菌全基因组的分泌蛋白预测和功能分析

吴怡豪¹, 诸葛君昊¹, 张立华², 周乐³, 杜盛楠¹, 杜世平¹, 苏秀¹

(1. 浙江农林大学 林业与生物技术学院, 浙江 杭州 311300; 2. 泰顺县自然资源和规划局, 浙江 温州 325500;
3. 临海市河头镇农业综合服务中心, 浙江 临海 317034)

摘要: 茶蔗子葡萄座腔菌 *Botryosphaeria dothidea* 是一种世界性分布真菌, 能引起山核桃干腐病, 导致严重经济损失。为研究该病害的致病机理及筛选致病基因, 通过各类预测与分析工具, 对 13 130 条蛋白序列的山核桃 *Carya cathayensis* 茶蔗子葡萄座腔菌 BDLA16-7 全基因组进行了分泌蛋白预测及其功能分析。结果表明, 在 BDLA16-7 全基因组的编码蛋白序列中, 有 898 条蛋白序列符合典型分泌蛋白的特征, 占总蛋白数的 6.84%; 分泌蛋白组的氨基酸组成呈现以下特征: 非极性、疏水性氨基酸的含量明显高于极性、亲水性氨基酸, 分别占 43.6% 和 27.7%; 蛋白质长度主要分布在 100 至 600 个氨基酸之间; 分泌蛋白序列中通常以 19 个氨基酸构成信号肽, 丙氨酸 (A) 是最常见的非极性氨基酸, 而带电侧链的天冬氨酸 (D) 和谷氨酸 (E) 的出现频率较低; 信号肽的切割位点 -3 和 -1 位置上的氨基酸相对保守, 典型的氨基酸组合为 A-S-A, 符合 A-X-A 型的特征。此外, 分泌蛋白组中包含 600 个序列可以完整注释其功能, 其中 229 个分泌蛋白主要涉及碳水化合物的运输和代谢。在 BDLA16-7 分泌蛋白组中, 共存在 271 个潜在的效应蛋白, 其中 14 个与其他病原菌致病相关的效应蛋白具有较高的相似性。此外, BDLA16-7 分泌蛋白组中包含 293 个碳水化合物活性酶, 以糖苷水解酶家族成员为最多。明确了茶蔗子葡萄座腔菌 BDLA16-7 分泌蛋白的长度分布特征, 功能分类、信号肽切割位点的氨基酸组成、切割位点的类型 and CAZymes 数目, 以及 CAZymes 中纤维素酶、果胶酶、木聚糖酶的亚家族组成, 并筛选到 14 个与致病性紧密联系的潜在效应蛋白。

关键词: 茶蔗子葡萄座腔菌; 分泌蛋白; 效应蛋白; 功能

中图分类号: S436.64; S664.1 文献标识码: A 文章编号: 1001-3776 (2024) 01-0008-011

Prediction and Functional Analysis on Secretory Proteome from the Whole Genome of *Botryosphaeria dothidea*

WU Yihao¹, ZHUGE Junhao¹, ZHANG Lihua², ZHOU Le³, DU Shengnan¹, DU Shiping¹, SU Xiu¹

(1. College of Forestry and Biotechnology, Zhejiang A & F University, Hangzhou 311300, China; 2. Taishun Natural Resources and Planning Bureau of Zhejiang, Wenzhou 325500, China; 3. Linhai Hetou Township Agricultural Service of Zhejiang, Linhai 317034, China)

Abstract: The full genome of *Botryosphaeria dothidea* strain BDLA16-7, a pathogenic fungus from an infected *Carya cathayensis* tree, comprising 13 130 protein sequences, which were analyzed to predict and functionally characterize secretory proteins by various prediction and analysis tools. The results indicated that in the encoded protein sequences of the BDLA16-7 whole genome, there were 898 sequences typical characteristics of

收稿日期: 2023-09-06; 修回日期: 2023-12-13

基金项目: 浙江省重点研发项目 (2019C02024)

作者简介: 吴怡豪, 硕士研究生, 从事森林保护学研究; E-mail: 1505817181@qq.com。通信作者: 苏秀, 副教授, 博士, 从事林木病理学研究; E-mail: suxiu@zafu.edu.cn。

secretory proteins, accounting for 6.84% of the total ones. The amino acid composition of the secretory protein group showed the following characteristics: the content of non-polar and hydrophobic amino acids was significantly higher than that of polar and hydrophilic amino acids, accounting for 43.6 % and 27.7 %, respectively. The length of the protein is mainly distributed between 100 and 600 amino acids. Alanine (A) is the most common non-polar amino acid, while aspartic acid (D) and glutamic acid (E) with charged side chains appear less frequently. The amino acids at the cleavage sites-3 and-1 of the signal peptide are relatively conserved, and the typical amino acid combination is A-S-A, which conforms to the characteristics of A-X-A type. Within the secretory protein group, 600 sequences could be fully annotated for their functions, among them, 229 involved in the transport and metabolism of carbohydrates. Among the BDLA16-7 secretory proteins, there were a total of 271 potential effector proteins, with 14 of them showing high similarity to effectors associated with other pathogenic organisms. BDLA16-7 secretory protein group contains 293 CAZymes, dominated by GHs family members. Analysis were made on the length distribution characteristics, functional classification, amino acid composition of signal peptide cleavage site, type of cleavage site and number of CAZymes of secretory proteome of *B. dothidea*, as well as on the subfamily composition of cellulase, pectinase and xylanase in CAZymes. 14 potential effectors closely related to pathogenicity were screened.

Key words: *Botryosphaeria dothidea*; secretory proteome; effector proteins; function

茶蔗子葡萄座腔菌 *Botryosphaeria dothidea* 属子囊菌门 Ascomycota、座囊菌纲 Dothideomycete、葡萄座腔菌目 Botryosphaeriales、葡萄座腔菌科 Botryosphaeriaceae、葡萄座腔菌属 *Botryosphaeria* 真菌^[1-2], 其分布极为广泛, 可以侵染多种经济林、用材林、园林绿化林木, 如苹果 *Malus pumila*、梨 *Pyrus sp.*、桃 *Prunus persica*、核桃 *Juglans regia*、葡萄 *Vitis vinifera*、杨 *Populus sp.*、桉 *Eucalyptus sp.*^[3-10]等, 导致严重的果实损失和生产减产。该菌可引起山核桃干腐病 *Macrophoma caryae*, 对山核桃 *Carya cathayensis* 的产量和品质产生了显著的负面影响^[11]。因此, 对于了解茶蔗子葡萄座腔菌的致病机理以及筛选抗病品种的研究具有重要的实践意义, 可以为山核桃生产提供有力的支持, 减少果实损失, 维护生产稳定性和生态平衡。

病原菌产生的分泌蛋白与寄主抗性基因的蛋白产物之间的分子互作是病原菌与寄主之间互作的核心^[12-13], 分泌蛋白在其中起着重要的作用。研究表明, 黄曲霉 *Aspergillus flavus* 通过分泌一些细胞外酶和毒素来侵染寄主, 这些酶可以破坏宿主细胞壁和木质素, 可以诱导宿主细胞凋亡和坏死, 从而使得该菌对宿主的感染更容易发生^[14]; 果树炭疽菌 *Colletotrichum gloeosporioides* 分泌的氧化还原酶可以降低植物的抗氧化能力, 导致宿主细胞的损伤和死亡^[15]; 稻瘟菌 *Magnaporthe oryzae* 利用分泌外源蛋白, 包括纤维素酶、蛋白酶和细胞壁水解酶等, 破坏植物细胞壁和膜结构, 从而增加了其侵染的能力。此外, 该病原体还释放毒素, 如稻瘟菌素和谷氨酰胺酰胺酶, 导致植物细胞的死亡和坏死^[16]。分泌蛋白不仅对真菌自身的生长和代谢有重要的调节作用, 还能够通过对植物细胞的破坏、毒素的作用等方式, 导致植物的损伤和病害发生。因此, 研究茶蔗子葡萄座腔菌分泌蛋白的功能和作用机制, 对研究该病害的发病机理与防控具有重要意义和应用价值。

茶蔗子葡萄座腔菌全基因组测序数据的公开, 为开展该菌分泌蛋白的研究提供了坚实的基础。本研究以山核桃茶蔗子葡萄座腔菌 BDLA16-7 的全基因组蛋白序列为基础, 以经典分泌蛋白特征为依据, 结合 Augustus、SignalP、Funannotate、WoLF PSORT、Ensembl Fungi、GenomeTools 等网站和软件, 对 BDLA16-7 的分泌蛋白组进行了预测和功能分析, 为进一步深入研究该病原菌与寄主互作的分子机制提供了基础, 并为挖掘病原菌关键致病基因提供依据。

1 材料和方法

1.1 材料

山核桃茶蔗子葡萄座腔菌 BDLA16-7 是从浙江省杭州市临安区的一棵患有溃疡病的山核桃树的树干上分离出的山核桃茶蔗子葡萄座腔菌菌株。最初, 该真菌在马铃薯葡萄糖琼脂培养基上的菌落呈淡白色, 随后逐渐转变为黑色。子囊座 *pycnidia* 通常是单独存在的, 呈灰黑色、球形, 并被菌丝覆盖。该菌株的基因组 DNA 和 mRNA

是从培养了 10 天的菌丝体中提取而得。山核桃茶蔗子葡萄座腔菌 BDLA16-7 全基因组蛋白质氨基酸序列, 来自中国生物信息中心 (CNCB-NGDC 成员和合作伙伴 2021) 的基因组仓库 (GWH) (<https://ngdc.cncb.ac.cn/gwh/>), 注册号为 GWHBEO00000000, 原始序列来自 CNCBNGDC 的基因组序列档案 (<https://ngdc.cncb.ac.cn/gsa/>) 上公开访问, 登录号为 CRA004612 (生物项目 PRJCA005744) [17]。

1.2 预测方法

1.2.1 茶蔗子葡萄座腔菌全基因组蛋白 N-端信号肽预测 使用 SignalP 5.0 在线工具 (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>) 和 Phobius (<https://phobius.sbc.su.se/>) 来预测基因组编码的所有蛋白质是否带有 N-端信号肽 [18]。

1.2.2 茶蔗子葡萄座腔菌蛋白的亚细胞定位 使用蛋白亚细胞定位在线预测软件 Cell-PLoc 2.0 [19] (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>) 来确定带有 N-端信号肽的蛋白质的亚细胞定位。

1.2.3 茶蔗子葡萄座腔菌分泌型蛋白的跨膜结构预测 使用蛋白跨膜结构预测软件 TMHMM [20] (<https://dtu.biolib.com/DeepTMHMM>), 对全蛋白序列中具有信号肽且主要定位于细胞外的蛋白进行是否含有跨膜结构预测。

1.2.4 茶蔗子葡萄座腔菌分泌型蛋白的锚定位点预测 使用 GPI 锚定位点预测软件 GPI-SOM (<http://gpi.unibe.ch>) [21], 对全蛋白序列中具有信号肽、主要定位于细胞外且不包含跨膜结构域的蛋白序列进行 GPI 锚定位点的预测。

1.2.5 分泌蛋白中效应蛋白的预测和功能分析 使用 EffectorP3.0 (<http://effectorp.csiro.au>) [22] 软件对全蛋白序列中具有信号肽、主要定位于细胞外、不包含跨膜结构域且不包含 GPI 锚定位点的蛋白序列进行效应蛋白的预测。同时, 通过 eggno-mapper [23] 数据库和 Pathogen Host Interactions [24] 数据库对预测出的效应蛋白的组成和功能进行分析。

1.2.6 分泌蛋白中碳水化合物活性酶类预测和功能分析 使用 CAZymes 在线预测工具 dbcan [25] (<http://bcb.unl.edu/dbCAN2/>) 中的 HMMER、DIAMOND 和 HOTPEP 程序, 对分泌蛋白中的碳水化合物活性酶进行预测和功能注释。

2 结果与分析

2.1 山核桃茶蔗子葡萄座腔菌全基因组蛋白 N-端信号肽预测

BDLA16-7 基因组编号为 GWHPBEO000001 到 GWHPBEO013130, 总共编码了 13 130 条蛋白质序列。通过使用 SignalP 5.0 和 Phobius 软件进行分析, 结果表明其中含有 N-端信号肽序列的有 1 402 条蛋白质序列, 占总蛋白质序列数量的 10.67%。这些具有 N-端信号肽序列的蛋白质全部属于标准的 Sec/SPI 型分泌蛋白质, 它们会在信号肽序列的引导下进入细胞的内质网腔中。

2.2 山核桃茶蔗子葡萄座腔菌全基因组蛋白的亚细胞定位

在 BDLA16-7 全基因蛋白序列中, 将含有信号肽的 1 402 条蛋白序列经 Cell-PLoc 2.0 软件分析发现, 其中有 1 188 条蛋白序列属于胞外分泌型, 而其余 214 条序列虽然具有信号肽, 但其并不分泌到细胞外, 而是转运到了细胞中不同的细胞器或质膜上。

2.3 山核桃茶蔗子葡萄座腔菌分泌型蛋白的跨膜结构预测

在 BDLA16-7 全基因蛋白序列中, 有一些蛋白具有信号肽序列, 并且属于胞外分泌型。经过跨膜结构域预测软件 TMHMM 分析, 发现 1 188 条胞外分泌型蛋白序列中, 有 97 条蛋白序列具有跨膜区。因此, 这些序列有可能是离子通道蛋白、膜锚定蛋白或膜受体蛋白, 而无法分泌到细胞膜外。除此之外的 1 091 条蛋白序列没有跨膜结构域, 具有典型的分泌蛋白特征。

2.4 山核桃茶蔗子葡萄座腔菌分泌型蛋白的 GPI 锚定位点预测

在 BDLA16-7 全基因蛋白序列中, 有一些蛋白具有信号肽序列, 属于胞外分泌型蛋白, 并且没有跨膜结构域。经过 GPI-SOM 软件预测, 发现 1 091 条蛋白序列中有 193 条含有 GPI 锚定位点, 不符合经典的分泌型蛋白

特征, 因此将其去除。除此之外, 有 898 条分泌型蛋白序列不含 GPI 锚定位点, 符合经典的分泌蛋白特征, 因此将其认定为 BDLA16-7 的分泌蛋白组 (表 1)。

表 1 山核桃茶蔗子葡萄座腔菌分泌蛋白质组和致病相关效应因子的预测结果
Tab. 1 Prediction of secretory proteome and pathogenicity-related effectors of *B. dothidea*

预测程序或数据库	项目	数量/个	预测程序或数据库	项目	数量/个
NCBI	全部蛋白	13 130	GPI-SOM	不含 GPI 锚定位点的蛋白	898
SignalP 5.0、Phobius	含信号肽的蛋白	1 402	EggNOG	注释到完整功能的蛋白	600
Cell-PLoc 2.0	胞外蛋白	1 188	EffectorP3.0	预测为效应蛋白	271
TMHMM	无跨膜结构域的蛋白	1 091	PHI	可比对到致病功能的蛋白	14

2.5 山核桃茶蔗子葡萄座腔菌分泌蛋白组特征分析

2.5.1 山核桃茶蔗子葡萄座腔菌分泌蛋白组的氨基酸长度分析 对 BDLA16-7 分泌蛋白组的序列长度进行分析, 发现分泌蛋白的长度分布跨度较大。其中, 存在长度小于 100 个氨基酸的小分子蛋白, 最短的为 66 个氨基酸; 同时也存在长度超过 1 000 个氨基酸的大分子蛋白, 最长的蛋白为 2 028 个氨基酸。分泌蛋白的氨基酸长度主要分布在 100 ~ 600 个氨基酸范围内, 共有 721 条序列, 占分泌蛋白序列总数的 80.29% (图 1)。

2.5.2 山核桃茶蔗子葡萄座腔菌分泌蛋白组的氨基酸组成分析 在 BDLA16-7 分泌蛋白组的氨基酸组成分析中, 发现不同的氨基酸含量有显著差异, 按照含量从高到低的顺序排列如下: A 丙氨酸、G 甘氨酸、S 丝氨酸、L 亮氨酸、T 苏氨酸、V 缬氨酸、P 脯氨酸、D 天冬氨酸、N 天冬酰胺、E 谷氨酸、I 异亮氨酸、F 苯丙氨酸、Y 酪氨酸、Q 谷氨酰胺、R 精氨酸、K 赖氨酸、W 色氨酸、H 组氨酸、C 半胱氨酸、M 蛋氨酸, 氨基酸可以被分为不同的类别。其中, 非极性疏水氨基酸 (A、G、L、V、P 和 I) 的含量最高, 占总含量的 43.6%, 显示出明显的优势, 亲水的极性氨基酸 (S、T、N、Q、M 和 C) 的含量相对较低, 为 27.7%; 酸性氨基酸 (A 和 E) 的含量为 14.8%, 碱性氨基酸 (K、R、H) 的含量为 8.7%, 芳香族氨基酸 (F、Y 和 W) 的含量为 9.7% (图 2)。

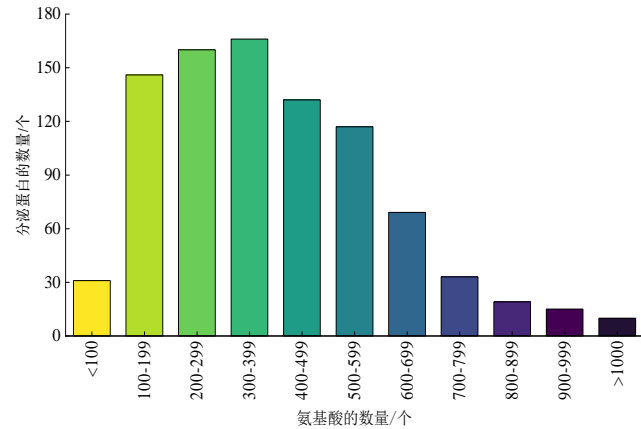


图 1 山核桃茶蔗子葡萄座腔菌分泌蛋白质组的长度分布
Fig. 1 Length distribution of secretory proteome of *B. dothidea*

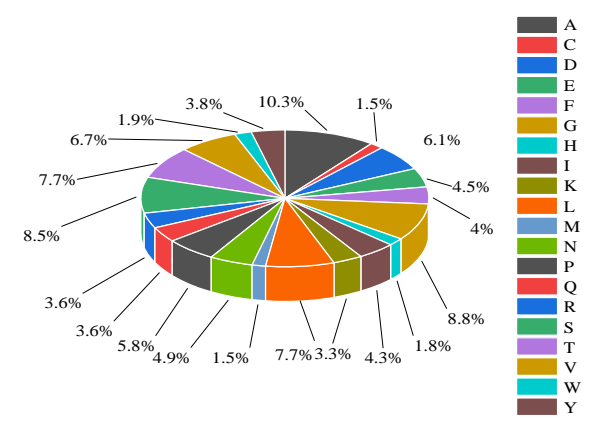


图 2 山核桃茶蔗子葡萄座腔菌分泌蛋白质组中氨基酸组成比例
Fig. 2 Composition of amino acids of secretory proteome of *B. dothidea*

2.5.3 山核桃茶蔗子葡萄座腔菌分泌蛋白组信号肽的氨基酸长度分析 BDLA16-7 分泌蛋白组中的信号肽序列具有不同的氨基酸长度, 这个长度范围从 14 到 41 个氨基酸。在这些信号肽序列中, 长度为 16 到 24 个氨基酸的序列最为常见, 总共有 806 个, 占了分泌蛋白组总数的 89.76%。其中, 信号肽长度为 19 个氨基酸的序列最多, 共有 169 个, 占了分泌蛋白组总数的 18.82%。其次, 长度为 18 个氨基酸的信号肽序列有 132 个, 占了分泌蛋白组总数的 14.70%。相反, 信号肽长度为 31、35、39 和 41 个氨基酸的序列最为罕见, 每个长度仅有 1 个序列 (图 3)。

2.5.4 山核桃茶蔗子葡萄座腔菌分泌蛋白组信号肽的氨基酸组成分析 根据对 BDLA16-7 分泌蛋白组信号肽的氨基酸组成进行的分析, 发现 20 种氨基酸在信号肽中的出现频率存在显著差异, 按频率从高到低排列如下: A、

L、S、V、T、M、F、I、G、P、R、K、Q、H、C、N、Y、W、D 和 E。其中，非极性氨基酸丙氨酸（A）的出现频率最高，占总出现频率的 22.40%，亮氨酸（L）的出现频率次之，为 19.29%。相反地，天冬氨酸（D）和酸性氨基酸谷氨酸（E）的出现频率最低，分别仅为 0.29% 和 0.27%（图 4）。

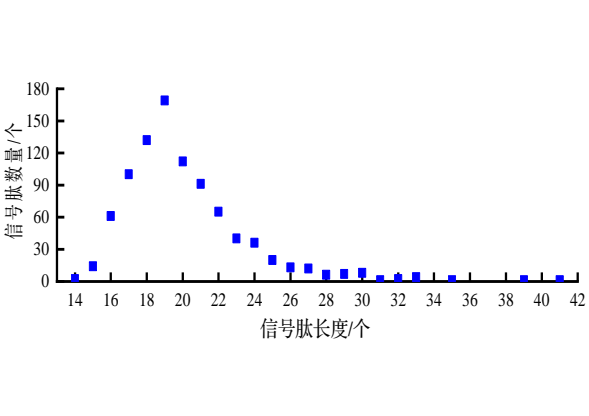


图 3 山核桃茶蔗子葡萄座腔菌分泌蛋白质组信号肽长度分布

Fig. 3 Length distribution of signal peptides of secretory proteome *B. dothidea*

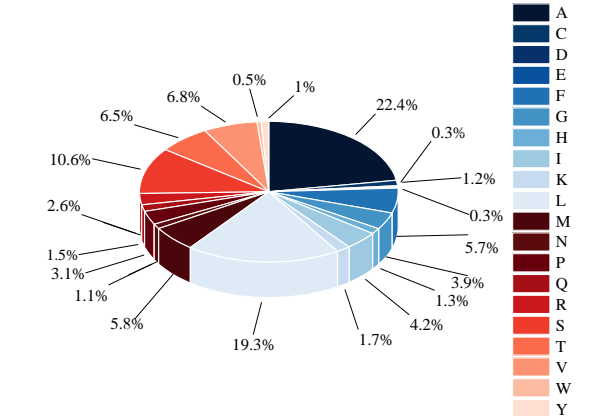


图 4 山核桃茶蔗子葡萄座腔菌分泌蛋白质组信号肽中氨基酸组成比例

Fig. 4 Composition of amino acids of signal peptides of secretory proteome in *B. dothidea*

2.5.5 山核桃茶蔗子葡萄座腔菌分泌蛋白组信号肽切割位点分析 通过对 BDLA16-7 分泌蛋白组信号肽的切割位点（-3 到+2 位点）进行统计分析，得到以下结果：在-3、-2、-1、+1 和+2 位点出现频率最高的氨基酸分别为 A、S、A、A 和 P，它们分别占据了 50.56%、18.93%、81.74%、25.61%和 37.08%的比例。具体来说：在-3 位点，除了 A 之外，V（24.28%）和 S（8.91%）也比较常见，而 E、F、H、K、M、N、W 和 Y 没有被使用。在-1 位点，除了 A 之外，G（8.24%）和 S（5.90%）也相对常见，而 D、E、H、I、L、M、N、V、W 和 Y 没有出现。在-2、+1 和+2 位点，所有的氨基酸都被使用。这些结果表明，在 BDLA16-7 分泌蛋白组信号肽中，切割位点-3 和-1 位置上的氨基酸相对来说比较保守，可能在信号肽酶的识别过程中起到关键作用。其他位点则更加多样化，容纳了各种不同的氨基酸（表 2）。

表 2 信号肽酶切割位点周围 20 个氨基酸残基的频率 Tab. 2 The frequency of 20 amino acid residues around the signal peptidase cleavage sites										
氨基酸	-3 位点		-2 位点		-1 位点		+1 位点		+2 位点	
	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%
A	454	50.56	158	17.59	734	81.74	230	25.61	32	3.56
C	14	1.56	8	0.89	23	2.56	8	0.89	13	1.45
D	1	0.11	20	2.23	—	—	35	3.90	57	6.35
E	—	—	21	2.34	—	—	23	2.56	32	3.56
F	—	—	41	4.57	1	0.11	30	3.34	22	2.45
G	31	3.45	12	1.34	74	8.24	35	3.90	21	2.34
H	—	—	46	5.12	—	—	34	3.79	11	1.22
I	24	2.67	25	2.78	—	—	36	4.01	22	2.45
K	—	—	6	0.67	1	0.11	16	1.78	20	2.23
L	7	0.78	153	17.04	—	—	73	8.13	34	3.79
M	—	—	9	1.00	—	—	7	0.78	12	1.34
N	—	—	45	5.01	—	—	30	3.34	32	3.56
P	1	0.11	1	0.11	3	0.33	1	0.11	333	37.08

表 2 (续)										
氨基酸	-3 位点		-2 位点		-1 位点		+1 位点		+2 位点	
	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%
Q	1	0.11	38	4.23	1	0.11	107	11.92	34	3.79
R	1	0.11	25	2.78	1	0.11	25	2.78	15	1.67
S	80	8.91	170	18.93	53	5.90	85	9.47	70	7.80
T	66	7.35	45	5.01	7	0.78	46	5.12	65	7.24
V	218	24.28	39	4.34	—	—	47	5.23	47	5.23
W	—	—	7	0.78	—	—	11	1.22	5	0.56
Y	—	—	29	3.23	—	—	19	2.12	21	2.34

注: “—” 表示无此类型氨基酸残基。

2.5.6 山核桃茶蔗子葡萄座腔菌分泌蛋白组的功能注释

通过使用 eggnog-mapper 数据库、PHI 数据库和 Geneious Prim Trial 数据库对 BDLA16-7 分泌蛋白组的 898 条序列进行功能分析。结果表明,共有 600 条序列可以被注释到具有完整功能。通过 COG (Clusters of Orthologous Genes) 数据库对这些功能进行分类,发现最多的分泌蛋白被归类为 G 类,即碳水化合物的运输和代谢 (Carbohydrate transport and metabolism),总共有 229 条序列归入这一类别。其次是 S 类,即功能未知类 (Function unknown),其中包括一些蛋白的结构域和功能未知的蛋白,如酪氨酸酶中心结构域 (Comon central domain of tyrosinase)、泛素 3 绑定蛋白 But2 的 C-端结构域 (Ubiquitin 3 binding protein But2 C-terminal domain)、DUF3129 和 DUF3455 等,总共有 152 条序列归入这一类别。其他类别: O 类蛋白翻译后修饰和分子伴侣 (Post-translational modification, protein turnover, chaperones) 86 条序列, E 类氨基酸运输和代谢相关 (Amino acid transport and metabolism) 44 条序列, C 类能量产生和转化相关 (Energy production and conversion) 24 条序列,以及 Q 类次生代谢物的生物合成、转运和分解相关 (Secondary metabolites biosynthesis transport and catabolism) 23 条序列。其余的功能分类包含的分泌蛋白数目较少,没有超过 10 个。这些分析结果揭示了 BDLA16-7 分泌蛋白组的功能多样性,其中大部分分泌蛋白与碳水化合物代谢相关,但也包含一些功能未知的蛋白以及其他生物学功能。这种功能分类有助于理解这些蛋白在生物体内的作用和生物过程中的角色 (图 5)。

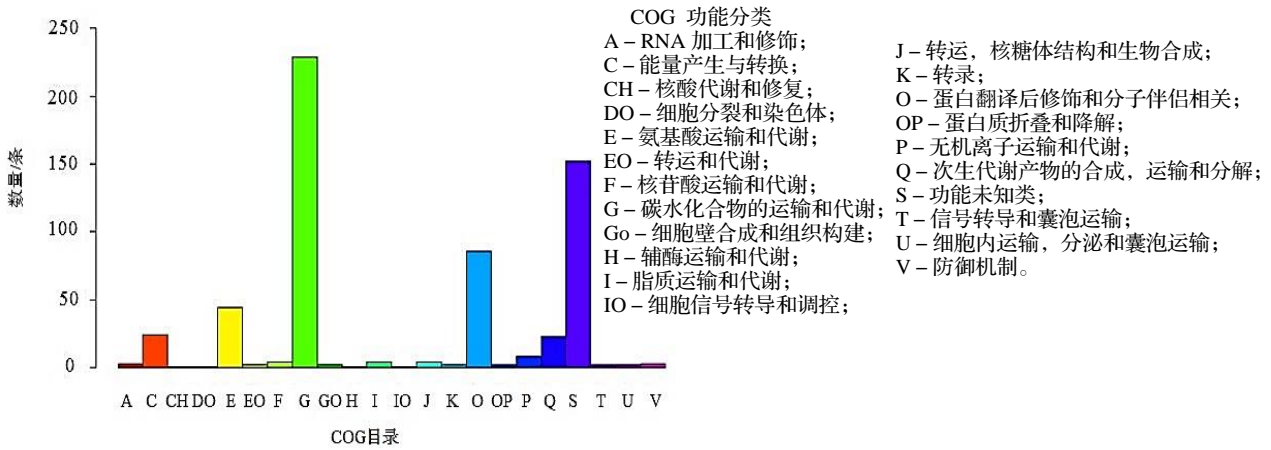


图 5 山核桃茶蔗子葡萄座腔菌分泌蛋白质组的 COG 目录

Fig. 5 COG catalog of the secretory proteome of *B. dothidea*

2.6 山核桃茶蔗子葡萄座腔菌分泌蛋白组中效应蛋白的预测和功能分析

BDLA16-7 分泌蛋白组经 TargetP 2.0 Server 软件分析,再次确认 898 条蛋白序列全部含有信号肽,并可以分泌到细胞外。继而通过效应蛋白预测软件 EffectorP 3.0 对潜在的效应蛋白进行预测,结果发现 271 条蛋白序

列可被预测为效应蛋白, 占分泌蛋白总数的 30.18%。其中 101 条序列可以被 eggno-map 数据库注释到具体的功能, 其中 42 条序列在 COG 分类中属于 S 类, 主要包括蛋白结构域和功能未知的蛋白, 如: IgE 结合蛋白 (IgE binding protein)、Cupin 家族蛋白、成束蛋白结构域 (Fasciclin-domain)、CFEM 结构域 (CFEM-domain) 等, 41 条序列属 G 类, 14 条序列属于 O 类, 2 条序列属于 C 类, 1 条序列属于 A 类 RNA 加工和修饰相关 (RNA processing and modification), 1 条序列属于 E 类 (图 6)。

对预测为潜在效应蛋白的 271 条序列, 经 PHI 数据库 Blastp 比对, 有 14 条序列可以比对到相应的致病功能注释。其中 1 条序列 (GWHPBEO006248) 可以比对到黄曲霉的致病效应蛋白序列, 1 条蛋白序列 (GWHPBEO004805) 可以比对到禾谷镰刀菌 *Fusarium graminearum* 和黄单胞菌 *Xanthomonas axonopodis* 的致病效应蛋白序列, 1 条蛋白序列 (GWHPBEO011899 和 GWHPBEO003128) 可以比对到寄生疫霉 *Phytophthora parasitica* 的致病效应蛋白序列, 1 条蛋白序列 (GWHPBEO008334) 可以比对到稻瘟菌的致病效应蛋白序列, 2 条序列 (GWHPBEO008083 和 GWHPBEO000994) 可以比对到禾谷镰刀菌的致病效应蛋白序列, 2 条蛋白序列 (GWHPBEO006043 和 GWHPBEO006222) 可以比对到扩展青霉 *Penicillium expansum* 的致病效应蛋白序列, 5 条蛋白序列 (GWHPBEO010105、GWHPBEO006002、GWHPBEO003061、GWHPBEO003727 和 GWHPBEO001912) 与褐腐菌 *Monilinia fructicola* 致病性增强的效应蛋白序列有较高的相似性, 其同时也与柑橘褐斑菌 *Alternaria alternata* 的致病效应蛋白序列存在较高的相似性。

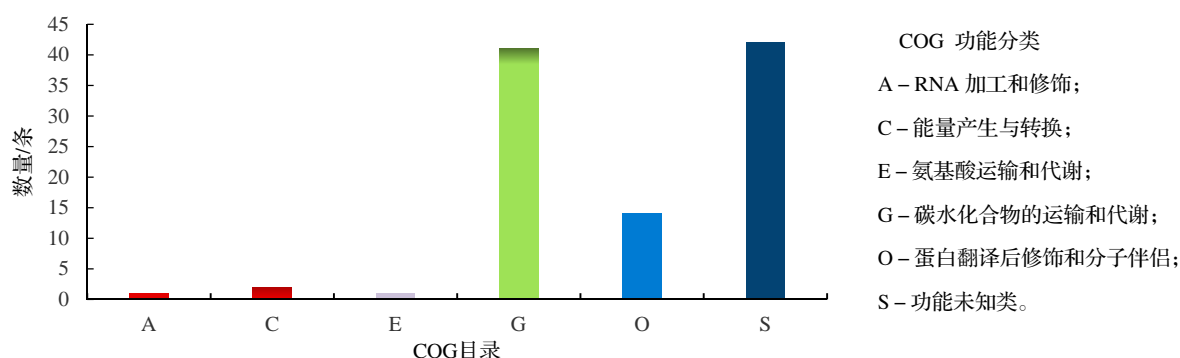


图 6 山核桃茶蔗子葡萄座腔菌的效应蛋白质组 COG 目录

Fig. 6 COG catalog of the effector proteome of *B. dothidea*

2.7 山核桃茶蔗子葡萄座腔菌分泌蛋白组中碳水化合物活性酶类 (CAZymes) 预测

碳水化合物活性酶类 (CAZymes) 是病原菌分泌蛋白组的重要组成部分, 其主要参与对寄主细胞壁的降解作用, 与病原菌的致病性有紧密的联系。利用 HMMER、DIAMOND 和 HOTPEP 三个软件, 在 BDLA16-7 分泌蛋白组 898 条蛋白序列, 共预测到 293 个分泌蛋白属于 CAZymes 酶类, 分属于糖苷水解酶 (Glycoside Hydrolases, GHs)、氧化还原酶 (Auxiliary Activities, AAs)、糖酯酶 (Carbohydrate Esterases, CEs) 和多糖裂解酶 (Polysaccharide Lyases, PLs) 四个分类, 占总分泌蛋白数的 32.26%。其中基于序列相似性的 HMMER 软件预测发现 282 个分泌蛋白属于 CAZymes, 基于结构域的 DIAMOND 和 HOTPEP 软件分别预测到了 188 个和 275 个分泌蛋白属于 CAZymes。三个软件都预测到 CAZymes 的数量为 169 个 (图 7)。

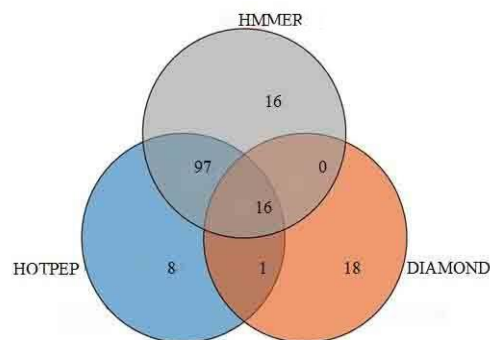


图 7 Hmmer Diamond 和 Hotpep 对山核桃茶蔗子葡萄座腔菌分泌蛋白组碳水化合物酶类预测结果

Fig. 7 CAZymes prediction on secretory proteome of *B. dothidea* by Hmmer diamond and Hotpep

在预测到的 169 个 CAZymes 中, GHs 数量最多为 99 个, 主要包括 GH3、GH28、GH92、GH1 和 GH31 等多个亚家族; 其次为 AAs, 有 42 个, 主要包括 AA3_2、AA1_3、AA3 和 AA9 等亚家族; CBMs 为 12 个, 主要包括 CBM1、CBM35 和 CBM18 等亚家族; PLs 为 9 个, 包括 PL1_4、PL3_2 等亚家族; CEs 为 6 个, 包括 CE0、CE1、CE16、CE4 和 CE8 亚家族; CTs 的数量最少, 为 1 个, 属于 GT32 亚家族(图 8)。

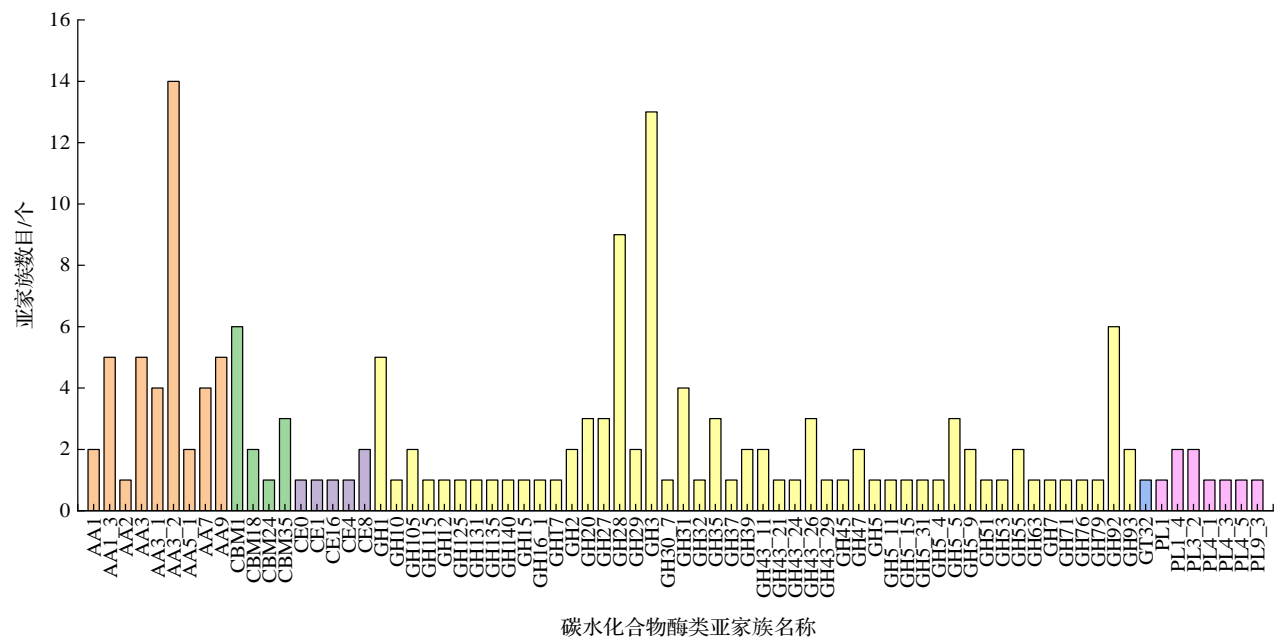


图 8 山核桃茶蔗子葡萄座腔菌分泌蛋白质组碳水化合物酶类亚家族分布预测

Fig. 8 Prediction of distribution of carbohydrate enzyme subfamily in secretory proteome of *B. dothidea*

在 BDLA16-7 分泌蛋白组的 169 个 CAZymes 中, 参与寄主细胞壁纤维素、木聚糖和果胶降解的 CAZymes 数量有 85 个, 而在丝孢堆黑粉菌 *Sporisorium reilianum*、玉蜀黍黑粉菌 *Ustilago maydis*、水稻稻瘟病菌 *Magnaporthe grisea*、禾谷镰刀菌、灰葡萄孢 *Botrytis cinerea*、大丽轮枝菌 *Verticillium dahliae* 分别有 26、30、128、127、87、165 个(表 3)^[32]。

表 3 不同营养类型真菌参与寄主细胞壁纤维素、木聚糖和果胶降解的 CAZymes 数量

Tab. 3 The number of CAZymes involved in the degradation of host cell wall cellulose, xylan and pectin by different trophic fungi

营养类型	真菌名称	参与降解纤维素酶/个	参与降解木聚糖酶/个	参与降解果胶酶/个	总计/个
活体营养型	丝孢堆黑粉菌	8	7	11	26
活体营养型	玉蜀黍黑粉菌	10	10	10	30
半活体营养型	水稻稻瘟病菌	42	47	39	128
半活体营养型	禾谷镰刀菌	32	36	59	127
半活体营养型	茶蔗子葡萄座腔菌	21	24	40	85
死体营养型	灰葡萄孢	26	17	44	87
死体营养型	大丽轮枝菌	46	35	84	165

3 讨论

病原真菌分泌蛋白的预测最为重要的是对信号肽进行预测。为了提高信号肽预测的准确性, 本文通过 SignalP 和 Phobius 相结合的方法, 排除了与信号肽具有相似结构导肽(transit peptide)的影响, 保证了含有信号肽分泌蛋白预测的准确性。通过 Cell-PLoc 2.0、TMHMM、GPI-SOM、EffectorP3.0、dbcan 等软件的分析, 本文成功预测了山核桃茶蔗子葡萄座腔菌 BDLA16-7 分泌蛋白组 898 条蛋白序列, 占其蛋白序列总条数的 6.84%;

与尖孢镰刀菌 *Fusarium oxysporum* (7.5%)^[26]、假禾谷镰孢菌 *Fusarium pseudograminearum* (5.5%)^[27]、新月弯孢菌 *Curvularia lunata* (7.7%)^[28]、禾谷炭疽菌 *Colletotrichum graminicola* (5.2%)^[29]分泌蛋白在全基因组蛋白序列中所占的比例相似。

分析 BDLA16-7 分泌蛋白组的氨基酸构成显示,非极性疏水氨基酸(包括 A、G、L、V、P 和 I)的含量为 43.6%,明显高于亲水极性氨基酸(S、T、N、Q、M 和 C)的含量(27.7%)。这表明,分泌蛋白通常需要在其折叠和稳定化过程中形成一个疏水核心结构。这个核心结构有助于保护蛋白质免受水分子和其他极性分子的干扰,同时有助于维持蛋白质的结构完整性,从而使其能够正确履行其功能。特别值得注意的是,信号肽长度为 19 个氨基酸的分泌蛋白数量最多,这可能意味着这个长度的信号肽对于茶蔗子葡萄座腔菌分泌蛋白质的正确转运和定位至关重要。这也可能是生物进化过程中的结果,即这个长度的信号肽的功能在生物进化中被保留下来并广泛应用。需要强调的是,这个结论是基于统计学推断的,不一定适用于所有分泌蛋白。此外,需要进一步的研究来验证这一结论,并探索不同长度信号肽的功能和进化意义。对于 BDLA16-7 分泌蛋白组的信号肽氨基酸构成,发现 20 种氨基酸在信号肽中的出现频率存在较大差异。前六位出现频率最高的氨基酸均为脂肪族氨基酸,这可能与信号肽需要穿越细胞质膜有关。此外,分泌蛋白组的信号肽切割位点在-3 和-1 位置的氨基酸相对保守,其氨基酸组成为 A-S-A,属于 A-X-A 型,这是真核生物中相对典型的信号肽切割位点,可被 SpaseI 型信号肽酶所识别。

效应蛋白(effector)是一种在植物病原微生物中广泛存在的重要致病因子,它们通过干扰寄主植物的免疫反应来感染宿主^[30]。因此,效应蛋白既是病原微生物致病的关键分子,也是用于筛选植物抗病性基因的重要工具^[31]。研究表明,在对葡萄座腔菌效应蛋白预测中,预测为潜在效应蛋白的 119 条序列,有 11 条序列可以比对到相应的致病功能注释^[32]。当在对 BDLA16-7 分泌蛋白组效应蛋白的预测中,我们发现 271 条序列(30.18%)符合效应蛋白的定义,其中 101 条序列可以被 eggno-map 数据库注释到具体的功能,经 PHI 数据库 Blastp 比对,有 14 条序列可以与黄曲霉、禾谷镰刀菌、黄单胞菌、寄生疫霉、稻瘟菌、扩展青霉、褐腐菌、柑橘褐斑菌比对到相应的致病功能注释,其中 5 条蛋白序列与褐腐菌 *M. fructicola* 致病性增强的效应蛋白序列有较高的相似性,但其同时也与柑橘褐斑菌 *A. alternata* 致病性降低的致病效应蛋白序列存在较高的相似性。相比较而言,我们研究比对的潜在效应蛋白序列更多(271 条),这意味着它在研究潜在效应蛋白与致病菌之间的关系时有更多的数据支持,可以提供更全面的信息;比对到的致病菌种类更多,从细菌到真菌,从黄曲霉到寄生疫霉等,这增加了研究的多样性和广泛性。但目前尚不清楚这些潜在的效应蛋白如何具体影响茶蔗子葡萄座腔菌的致病性,需要进一步通过实验手段来验证和研究其功能。

CAZymes(碳水化合物活性酶类)在真菌中指的是一组各种酶类,参与碳水化合物的分解和代谢,这些酶在生物质降解、病原菌感染宿主、形态变化、寄生植物侵染以及真菌与寄主的相互作用等方面发挥着作用^[33]。为了准确预测 BDLA16-7 中的 CAZymes 数量,我们使用了三种不同的软件,包括 Diamond、Hmmer 和 Hotpep,通过对这三种软件的预测结果求交集的方法来进行预测。在众多 CAZymes 中,特别是在直接参与植物细胞壁成分降解的 GH、AA、PL 和 CE 亚家族中,这些酶发挥着关键作用^[34]。在三种软件都预测到的 169 个 CAZymes 中,GHs 数量最多,共有 99 个,这些亚家族主要负责果胶和纤维素的降解。此外,BDLA16-7 的 CAZymes 中还包括大量的辅酶模块,其中 AAs 有 42 个,PLs 有 9 个,CEs 有 6 个,这进一步增强了茶蔗子葡萄座腔菌对细胞壁的降解能力,有助于病原菌成功侵染寄主植物。许多研究表明,植物病原真菌分泌的 CAZymes 与其寄生性之间存在一定的相关性^[35]。通常认为,寄生性更强的植物病原真菌编码的胞外 CAZymes 数量较少^[36]。通过比较不同营养型真菌的 CAZymes 数量,本研究发现,半活体营养型的茶蔗子葡萄座腔菌中,参与寄主细胞壁纤维素、木聚糖和果胶降解的 CAZymes 数量与活体营养型真菌相比显著增加,但与死体营养型真菌相比,CAZymes 数量基本相同或更少。这一研究结果与对终极腐霉 *Pythium ultimum*^[37]、颖枯壳多孢 *Stagonospora nodorum*^[38]、大丽轮枝菌^[39]等非活体营养型病原菌 CAZymes 数量的研究结果相一致。

4 结论

本文以山核桃茶藨子葡萄座腔菌 BDLA16-7 全基因组蛋白序列为基础,采用生物信息学的方法,对 BDLA16-7 分泌蛋白组进行了预测,明确了 BDLA16-7 分泌蛋白的长度分布特征,功能分类、信号肽切割位点的氨基酸组成、切割位点的类型和 CAZymes 数目,以及 CAZymes 中纤维素酶、果胶酶、木聚糖酶的亚家族组成,并筛选到 14 个 GWHPBEO0(06248、04805、11899、03128、08334、08083、00994、06043、06222、10105、06002、03061、03727、01912)与致病性紧密联系的潜在效应蛋白,为从分子水平上解析茶藨子葡萄座腔菌的致病机理奠定了基础。

参考文献

- [1] MARSBERG A, KEMLER M, JAMI F, et al. *Botryosphaeria dothidea*: a latent pathogen of global importance to woody plant health[J]. Molec Plant Pathol, 2017, 18 (4): 477–488.
- [2] PHILLIPS A J L, ALVES A, ABDOLLAHZADEH J, et al. The Botryosphaeriaceae: genera and species known from culture[J]. Stud Mycol, 2013, 76 (1): 51–167.
- [3] LI W Y, ZHUANG W Y. Taxonomy studies on the Botryosphaeria from China[J]. Mycosystema, 2013, 32: 108–114.
- [4] TANG W, DING Z, ZHOU Q Q, et al. Phylogenetic and pathogenic analyses show that the causal agent of apple ring rot in China is *Botryosphaeria dothidea*[J]. Plant Disease, 2012, 96 (4): 486–496.
- [5] ZHAI L F, ZHANG M X, LV G, et al. Biological and molecular characterization of four *Botryosphaeria* species isolated from pear plants showing stem wart and stem canker in China[J]. Plant Diseases, 2014, 98 (6): 716–726.
- [6] WANG F, ZHAO L N, LI G H. Identification and characterization of *Botryosphaeria* spp causing gummosis of peach trees in Hubei province central China[J]. Plant Diseases, 2011, 95 (11): 1378–1384.
- [7] CHEN S F, MORGAN D P, HASEY J K, et al. Phylogeny, morphology, distribution and pathogenicity of Botryosphaeriaceae and Diaporthaceae from English walnut in California[J]. Plant Diseases, 2014, 98 (5): 636–652.
- [8] YAN J Y, XIE Y, ZHANG W, et al. Species of Botryosphaeriaceae involved in grapevine dieback in China[J]. Fung Divers, 2013, 61: 221–236.
- [9] 余仲东, 赵官成, 淡静雅, 等. 葡萄座腔菌属 ITS-nrDNA 的分子系统学分析[J]. 菌物学报, 2010, 29 (02): 285–293.
- [10] DRAGINJA P Z, MALEME H M, BARBARA P, et al. Diversity, phylogeny and pathogenicity of Botryosphaeriaceae on nonnative *Eucalyptus* grown in an urban environment: A case study[J]. Urban ForUrban Green, 2017, 26: 139–148.
- [11] 刘锦, 于炜. 山核桃干腐病研究进展[J]. 浙江林业科技, 2017, 37 (06): 99–102.
- [12] 周晓罡, 侯思名, 陈铎文, 等. 马铃薯晚疫病菌全基因组分泌蛋白的初步分析[J]. 遗传, 2011, 33 (07): 125–133.
- [13] NIMCHUK Z, EULGEM T, HOLT B F III, et al. Recognition and response in the plant immune system[J]. Ann Rev Genet, 2003, 37 (1): 579–609.
- [14] SWEANY R R, DEROBERTIS C D, KALLER M D, et al. Intraspecific growth and aflatoxin inhibition responses to atoxigenic *Aspergillus flavus*: evidence of secreted, inhibitory substances in biocontrol[J]. Phytopathology, 2022, 112 (10): 2084–2098.
- [15] YANG K, WU J, WANG X, et al. Genome wide characterization of the mitogen activated protein kinase gene family and their expression patterns in response to drought and *Colletotrichum gloeosporioides* in walnut (*Juglans regia*)[J]. Plants, 2023, 12 (3): 586.
- [16] MARTINEZ-DALTO A, YAN X, DETOMASI T C, et al. Characterization of a unique polysaccharide monooxygenase from the plant pathogen *Magnaporthe oryzae*[J]. Proc Natl Acad Sci, 2023, 120 (8): e2215426120–e2215426120.
- [17] BAO J, WU Q, HUANG J, et al. High quality genome assembly and annotation resource of *Botryosphaeria dothidea* strain BDLA16-7, causing trunk canker disease on Chinese hickory[J]. Plant Disease, 2022, 106 (3): 1023–1026.
- [18] PETERSEN T N, BRUNAK S, VON HEIJNE G, et al. SignalP4.0: discriminating signal peptides from transmembrane regions[J]. Nat Meth, 2011, 8 (10): 785–786.
- [19] HORTON P, PARK K J, OBAYASHI T, et al. WoLF PSORTRT: protein localization predictor[J]. Nucl Acid Res, 2007, 35 (S2): 585–587.
- [20] KROCH A, LARSSON B, HEIJNE G V, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes[J]. J Molecul Biol, 2001, 305 (3): 567–580.

- [21] NIKLAUS F, PASCAL M. Identification of GPI anchor attachment signals by a Kohonen self-organizing map[J]. *Bioinformatics*, 2005, 21(9): 1846–1852.
- [22] JANA S, PETER N D, DONALD M, et al. Improved prediction of fungal effector proteins from secretomes with EffectorP2.0[J]. *Molecular Plant Pathology*, 2018, 19(9): 2094–2110.
- [23] HUERTA C J, SZKLARCZYK D, HELLER D, et al. EggNOG5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses[J]. *Nucleic Acid Research*, 2019, 47(D1): 309–314.
- [24] RAINER W, MARTIN U, ANDREW B, et al. PHI base update: additions to the pathogen host interaction database[J]. *Nucleic Acid Research*, 2008, 36(S1): 572–576.
- [25] ZHANG H, YOHE T, HUANG L, et al. dbCAN2: a meta server for automated carbohydrate active enzyme annotation[J]. *Nucleic Acid Research*, 2018, 46(W1): 95–101.
- [26] 祝友朋, 蔡旺芸, 韩长志. 基于全基因组序列的尖孢镰刀菌分泌蛋白预测及其特征分析[J]. *河南师范大学学报(自然科学版)*, 2019, 47(02): 92–97.
- [27] 王利民, 张银山, 张梦娟, 等. 假禾谷镰孢菌全基因组分泌蛋白的预测[C]//中国植物病理学会: 中国植物病理学会 2018 年学术年会论文集. 北京: 中国农业科学技术出版社, 2018: 172.
- [28] 高金欣, 高士刚, 李雅乾, 等. 玉米弯孢叶斑病菌全基因组分泌蛋白的预测与分析[J]. *植物保护学报*, 2015, 42(06): 869–876.
- [29] 韩长志. 全基因组预测禾谷炭疽菌的分泌蛋白[J]. *生物技术*, 2014, 24(02): 36–41.
- [30] WIT P J G M, MEHRABI R, VANDEN BURG H A, et al. Fungal effector proteins: past, present and future[J]. *Molecular Plant Pathology*, 2009, 10(6): 735–747.
- [31] STERGIOPOULOS I, WIT P J G M. Fungal effector proteins[J]. *Annual Review of Phytopathology*, 2009, 47: 233–263.
- [32] 臧睿, 宋璐璐, 尹新明, 等. 葡萄座腔菌全基因组分泌蛋白的预测及功能分析[J]. *植物病理学报*, 2021, 51(04): 559–571.
- [33] 聂燕芳, 黄嘉瑶, 周玲苑, 等. 香蕉枯萎病菌热带 4 号小种基因组规模分泌蛋白的预测与分析[J]. *江苏农业学报*, 2017, 33(02): 288–294.
- [34] 康振生, 王晓杰, 赵杰, 等. 小麦条锈菌致病性及其变异研究进展[J]. *中国农业科学*, 2015, 48(17): 3439–3453.
- [35] OSPINA G M D, GRIFFITH J G, LAIRD E W, et al. The CAZyme of *Phytophthora* spp.: a comprehensive analysis of the gene complement coding for carbohydrate active enzymes in species of the genus *Phytophthora*[J]. *BMC Genomics*, 2010, 11: 525.
- [36] ELLWOOD S R, LIU Z H, SYME R A, et al. A first genome assembly of the barley fungal pathogen *pyrenophora teres f. teres*[J]. *Genome Biology*, 2010, 11(11): R109.
- [37] OLIVER R P, IPCHO S V S. Arabidopsis pathology breathes new life into the necrotrophs-vs.-biotrophs classification of fungal pathogens[J]. *Molecular Plant Pathology*, 2004, 5(4): 347–352.
- [38] LEVESQUE C A, BROUWER H, CANO L, et al. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire[J]. *Genome Biology*, 2010, 11(7): R73.
- [39] 田李, 陈捷胤, 陈相永, 等. 大丽轮枝菌 (*Verticillium dahliae* VdLs.17) 分泌组预测及分析[J]. *中国农业科学*, 2011, 44(15): 3142–3153.