

尖孢镰刀菌甜瓜专化型基因组规模分泌蛋白的预测与分析

聂燕芳¹ 周 淦² 黄嘉瑶² 王振中² 李云锋²

1. 华南农业大学材料与能源学院, 广州 510642;

2. 华南农业大学农学院/广东省微生物信号与作物病害重点实验室, 广州 510642

摘要 利用 SignalP、WoLF PSORT、TMHMM、big-PI Predictor、TargetP 和 SecretomeP 等软件, 对尖孢镰刀菌甜瓜专化型(*Fusarium oxysporum* f. sp. *melonis*, FOM)全基因组 26 811 条蛋白质氨基酸序列进行了分泌蛋白的预测分析。结果表明, FOM 全基因组编码蛋白中有 1 145 个经典分泌蛋白, 占编码蛋白总数的 4.3%; 有 8 471 个非经典分泌蛋白, 占编码蛋白总数的 31.6%。对经典分泌蛋白特征进行分析, 发现其氨基酸长度集中在 100~600 个氨基酸, 信号肽长度集中在 17~22 个氨基酸。对经典分泌蛋白的功能预测分析表明, 有 605 个分泌蛋白获得了注释, 主要涉及糖代谢、转运过程、过氧化氢代谢和生物合成等。对碳水化合物酶类(CAZymes)进行分析, 结果表明有 277 个分泌蛋白为 CAZymes, 占经典分泌蛋白总数的 24.2%。

关键词 尖孢镰刀菌甜瓜专化型; 分泌蛋白; 信号肽; 碳水化合物酶类; 蛋白功能预测

中图分类号 S 435.111.41 **文献标识码** A **文章编号** 1000-2421(2016)03-0024-06

由尖孢镰刀菌甜瓜专化型(*Fusarium oxysporum* f. sp. *melonis*, 简称 FOM)引起甜瓜枯萎病是一种土传维管束病害, 在甜瓜全生育期均可发生。随着我国甜瓜种植面积的不断扩大, 以及设施栽培的迅速发展, 甜瓜枯萎病的发生也日趋严重, 已成为我国甜瓜主产区最严重的病害之一^[1]。FOM 共有 4 个生理小种(0、1、2 和 1.2), 其中我国台湾地区存在小种 2 号, 在我国其他地区则仅有 FOM 致病性分化的相关报道^[2-3]。尽管国内外学者对甜瓜枯萎病开展了大量研究, 但由于 FOM 遗传背景复杂, 目前关于 FOM 的致病机理等仍不清楚。分泌蛋白作为一类非常重要的生物活性分子, 在病原菌对植物的侵入、定殖和扩展等致病过程中发挥着重要作用^[4-5]。因此, 开展病原菌分泌蛋白质的研究已成为解析病原真菌致病分子机理的重要目标^[6]。

分泌蛋白(secreted proteins)是指在胞内合成分泌到胞外发挥功能的蛋白质, 其在细胞凋亡、信号转导、以及生物体的发育等许多生理和病理过程中扮演着重要角色^[7]。大多数分泌蛋白质含有信号肽, 通过经典的内质网/高尔基体途径分泌至胞外;

少数分泌蛋白质由于缺少常规信号肽等, 而通过非经典的分泌途径(non-classical secretory pathway)进行分泌^[8]。近年来, 很多重要病原真菌的基因组测序已经完成, 利用生物信息学技术对其全基因组数据进行挖掘, 预测分析分泌蛋白所具有的结构及功能越来越受到重视。目前, 已有许多学者开展了病原真菌分泌蛋白的预测分析^[9-11]。但迄今为止, 我们对 FOM 分泌蛋白的组成等仍知之甚少。FOM 基因组测序的完成^[12]为 FOM 分泌蛋白的研究等提供了便利。本研究结合 SignalP、WoLF PSORT、TMHMM、big-PI Predictor、TargetP 和 SecretomeP 等软件, 对 FOM 分泌蛋白进行预测分析, 拟为进一步明确 FOM 与寄主互作的分子机制等奠定基础。

1 材料与方法

1.1 FOM 基因组编码蛋白质数据来源

FOM 基因组编码的 26 811 个蛋白质氨基酸序列下载自数据库(http://www.broadinstitute.org/annotation/genome/fusarium_group/MultiDown-

收稿日期: 2015-09-20

基金项目: 国家自然科学基金项目(30971887); 广东省普通高校青年创新人才项目(粤教科函[2015]3 号)

聂燕芳, 博士, 讲师. 研究方向: 蛋白质组学. E-mail: yanfangnie@scau.edu.cn

通信作者: 李云锋, 博士, 教授. 研究方向: 病原菌与植物相互作用. E-mail: yunfengli@scau.edu.cn

loads.html)。

1.2 FOM 经典分泌蛋白的预测标准

将同时符合以下4个特征的FOM蛋白质定义为经典分泌蛋白:(1)具有N-端信号肽;(2)亚细胞定位为胞外分泌类型,不具有细胞质、细胞质膜及其他亚细胞预测定位信号;(3)不含有跨膜结构域;(4)没有GPI锚定位点^[13]。

1.3 分泌蛋白 N-端信号肽的预测

利用 SignalP 4.1 Server(<http://www.cbs.dtu.dk/services/SignalP/>)对 FOM 基因组编码的所有蛋白质 N-端信号肽进行在线预测分析^[14]。分泌信号肽的判别式为: $L = -918.235 - 123.455(S\text{-mean score}) + 1983.44$ (HMM score);其中,S-mean 值为蛋白质中单个氨基酸信号肽位置的平均预测值;HMM 值是隐马可夫模型预测出的值;若 L 值 > 0 , 则判定该蛋白质具有分泌型的信号肽^[15]。

1.4 分泌蛋白的亚细胞定位预测

利用 WoLF PSORT (http://www.genscript.com/psort/wolf_psort.html)^[16] 和 TargetP 1.1 Server (<http://www.cbs.dtu.dk/services/TargetP/>)^[17] 对 FOM 中含有 N-端信号肽的蛋白质进行亚细胞定位预测分析。

1.5 分泌蛋白跨膜结构域的预测

利用 TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>)对 FOM 中具有 N-端信号肽且定位于胞外的蛋白质跨膜结构域进行预测分析^[18]。

1.6 GPI 锚定位点预测

选取 FOM 中含有 N-端信号肽、定位于胞外、且具有跨膜结构域的蛋白质,利用 big-PI Predictor (http://mendel.imp.ac.at/sat/gpi/gpi_server.html)^[19] 对其 GPI 锚定位点进行分析。

1.7 FOM 非经典分泌蛋白质的预测分析

选取不含有 N-端信号肽的 FOM 蛋白质氨基酸序列,利用 SecretomeP 2.0 Server (<http://www.cbs.dtu.dk/services/SecretomeP/>)对非经典分泌蛋白进行在线预测分析;利用人工神经网络算法(neural network, NN)进行预测,并将不含有 N-端信号肽、且氨基酸序列 NN-score > 0.5 的蛋白质定义为非经典分泌蛋白^[20]。

1.8 FOM 经典分泌蛋白的功能分析及 CAZymes 的注释

利用 UniProt 数据库(<http://www.uniprot.org/>),对 FOM 所有经典分泌蛋白的功能进行预测。利用 BLASTP 软件比较 FOM 经典分泌蛋白与 CAZymes 序列(下载自 <http://www.cazy.org/>)的同源性(E 值 $\leq 10^{-5}$ 、且序列一致性 $\geq 30\%$),输出结果去除冗余,保留最佳匹配结果,得到 FOM 分泌蛋白的 CAZymes^[21];参与植物细胞壁组分降解的 CAZymes 分类按 CAZy 数据库提供的家族功能进行分析^[22]。

org/),对 FOM 所有经典分泌蛋白的功能进行预测。利用 BLASTP 软件比较 FOM 经典分泌蛋白与 CAZymes 序列(下载自 <http://www.cazy.org/>)的同源性(E 值 $\leq 10^{-5}$ 、且序列一致性 $\geq 30\%$),输出结果去除冗余,保留最佳匹配结果,得到 FOM 分泌蛋白的 CAZymes^[21];参与植物细胞壁组分降解的 CAZymes 分类按 CAZy 数据库提供的家族功能进行分析^[22]。

2 结果与分析

2.1 FOM 全基因组经典分泌蛋白的预测分析结果

1)基于 N-端信号肽的经典分泌蛋白预测分析结果。SignalP 软件常用于预测真核生物蛋白质中信号肽的切割位点及存在的具体位置。我们利用 SignalP 4.1 Server 对 FOM 全基因组 26 811 条蛋白质氨基酸序列的信号肽进行在线预测分析,结果发现具有分泌型信号肽的蛋白质氨基酸序列有 2 146 个,占总数的 8.0%。

2)基于亚细胞定位的分泌蛋白预测分析结果。蛋白质是否具有 N-端信号肽,是明确其是否属于经典分泌蛋白的必要条件;但对于该蛋白是否分泌到胞外并不清楚。因此,进一步利用 WoLF PSORT 软件,对上述具有信号肽的 2 146 个分泌型蛋白质进行亚细胞定位预测分析;结果表明有 1 442 个蛋白质分泌到胞外,属于胞外分泌蛋白类型;有 704 个蛋白质不分泌到胞外,而是分别转运至细胞质(142 个)、线粒体(186 个)、细胞核(80 个)和细胞质膜(293 个)中;此外,有 3 个蛋白质未能成功预测。

3)基于跨膜结构域的分泌蛋白预测分析结果。具有胞外分泌型信号肽的蛋白质如果同时含有跨膜区,则表明其可能为离子通道蛋白、膜锚定蛋白或膜受体,而不是典型的分泌蛋白^[23]。因此,进一步利用 TMHMM Server v2.0 对上述属于胞外分泌蛋白类型的 1 442 个蛋白质的跨膜结构域进行分析;结果表明有 1 208 个蛋白质没有跨膜螺旋,具有典型的分泌蛋白特征;有 216 个蛋白质氨基酸序列含有 1 个跨膜螺旋区,18 个蛋白质氨基酸序列都有 2 个以上的跨膜螺旋区,不属于典型的分泌蛋白。

4)基于 GPI 锚定位点的分泌蛋白预测分析结果。GPI 锚定是真核生物中一种常见的蛋白质翻译后修饰,其通过羧基末端的糖基化磷脂酰肌醇(glycosylphosphatidylinositol-anchored protein)结构锚定于细胞膜表面,因此在预测分泌蛋白时需要将

这一部分蛋白质进行排除^[9]。利用 big-PI Predictor 软件对上述 1 208 个没有跨膜结构域的蛋白质进行分析,结果表明有 12 个蛋白质具有 GPI 锚定位点,不符合经典的分泌蛋白特征;有 1 196 个蛋白质氨基酸序列不具有 GPI 锚定位点,具有经典分泌蛋白特征。

5) 基于亚细胞定位的分泌蛋白验证分析结果。为进一步精确分析上述 1 196 个蛋白质是否属于经典分泌蛋白,进一步结合 TargetP 1.1 Server 对这些蛋白质的亚细胞定位进行了分析。TargetP 1.1 Server 可以分析蛋白质是否含有线粒体靶向肽(mitochondrial targeting peptide)、叶绿体转运肽(chloroplast transit peptide)、分泌型信号肽(secretory pathway signal peptide)等;与 WoLF PSORT 分析结果相结合,能进一步提高经典分泌蛋白预测的精确性。结果表明,含有线粒体靶向肽和叶绿体转运肽的蛋白质氨基酸序列分别有 43 个和 8 个,含有分泌型信号肽的有 1 145 个。综合上述结果,基于 FOM 全基因组预测分析的 26 811 个蛋白质氨基酸序列中,有 1 145 个蛋白质具有经典分泌蛋白特征,所占比例为 4.3%。

2.2 FOM 经典分泌蛋白的特征分析

1) FOM 经典分泌蛋白的氨基酸长度分析结果。对 1 145 个 FOM 经典分泌蛋白的氨基酸长度分布特征进行分析,结果表明其大多集中在 100~600 个氨基酸之间(图 1);其中,以 300~399 个氨基酸长度为最多,占经典分泌蛋白总数的 19.9%;以 200~299 个氨基酸长度次之,占经典分泌蛋白总数的 16.7%;以 400~499 个氨基酸长度第三,占经典分泌蛋白总数的 15.1%。

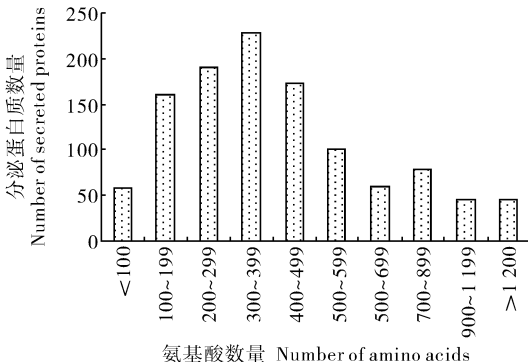


图 1 FOM 经典分泌蛋白的氨基酸长度分布

Fig.1 The amino acid length distribution of classically secreted proteins in FOM

对 1 145 个 FOM 经典分泌蛋白所含信号肽的氨基酸长度进行分析,结果表明以信号肽长度为 17~22 个氨基酸的序列最为集中,占经典分泌蛋白总数的 76.8%(图 2)。其中,以信号肽长度为 19 个氨基酸的分泌蛋白最多,占经典分泌蛋白总数的 16.8%;以信号肽长度为 20 个氨基酸的分泌蛋白次之,占经典分泌蛋白总数的 15.5%。

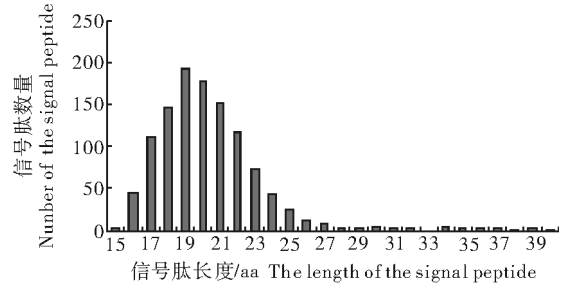


图 2 FOM 经典分泌蛋白的信号肽长度分布

Fig.2 The signal peptides length distribution of classically secreted proteins in FOM

2.3 FOM 中非经典分泌蛋白的预测分析

对 FOM 全基因组 26 811 个蛋白质氨基酸序列的信号肽分析中,仅有 2 146 个蛋白质具有分泌型信号肽。在此基础上,我们利用 SecretomeP 2.0 Server 对 FOM 中不具有分泌型信号肽的 24 665 个蛋白质氨基酸序列进行了分析。结果表明,有 8 471 个蛋白质氨基酸序列符合 SecretomeP 2.0 的非典型分泌蛋白特征,占全基因组编码蛋白总数的 31.6%。

2.4 FOM 经典分泌蛋白的功能预测分析

利用 UniProt 数据库 (<http://www.uniprot.org/>),对 1 145 个 FOM 经典分泌蛋白质的功能进行预测分析。结果表明,在 1 145 个蛋白质中,有 605 个分泌蛋白获得了注释,占经典分泌蛋白总数的 52.8%(表 1)。其中,232 个蛋白具有生物学过程(biological process)注释,以糖代谢过程相关分泌蛋白最多(105 个);465 个蛋白具有分子功能(molecular function)注释,以水解酶活性分泌蛋白最多(134 个)。

2.5 FOM 经典分泌蛋白 CAZymes 的注释与分析

碳水化合物酶类(carbohydrate-active enzymes, CAZymes)是参与碳水化合物合成和降解的重要酶类,主要包括糖基转移酶(glycosyltransferases, GT)、多糖裂解酶(polysaccharide lyases, PL)、糖基水解酶(glycoside hydrolases, GH)、碳水

2) FOM 经典分泌蛋白信号肽长度分析结果。

表 1 FOM 经典分泌蛋白的功能注释分析

Table 1 Functional annotation and categories of classically secreted proteins in FOM

分子功能 Molecular function	分泌蛋白数量 Number of secreted proteins	生物学功能 Biological process	分泌蛋白数量 Number of secreted proteins
水解酶活性 Hydrolase activity	134	糖代谢 Carbohydrate metabolic process	105
氧化还原酶活性 Oxidoreductase activity	111	过氧化氢代谢 Hydrogen peroxide catabolic process	17
激酶活性 Kinase activity	36	生物合成 Biosynthetic process	24
ATP 结合 ATP binding	28	转运过程 Transport process	28
核酸结合 Nucleic acid binding	23	细胞壁大分子代谢 Cell wall macromolecule catabolic process	12
裂解酶活性 Lyase activity	12	其他 Others	46
催化活性 Catalytic activity	46	—	—
其他 Others	44	—	—
总计 Total	465		232

注：经典分泌蛋白的功能注释与分类均来自 UniProt 数据库。Note: Functional annotation and categories of each classically secreted protein was from UniProt database.

化合物酯酶(carbohydrate esterases, CE)、碳水化合物结合模块(carbohydrate-binding modules, CBM)及辅助模块酶类(auxiliary activities, AA)等 6 类^[24]。对 FOM 中具有功能注释的 605 个经典分泌蛋白进行了分析,结果表明有 277 个分泌蛋白为 CAZymes,占经典分泌蛋白总数的 24.2%(表 2)。其中, GH 家族最多,占 33.6%;GT 家族次之,占 27.4%;PL 家族最少,仅有 2 个蛋白。进一步对直接参与寄主植物细胞壁组分降解的不同种类 CAZymes 的亚家族进行分析,结果表明,参与植物细胞壁组分纤维素降解的 GH 亚家族共有 8 个(含 5 个 GH3、2 个 GH5 和 1 个 GH7),参与植物细胞壁组分果胶降解的 GH 和 PL 亚家族共有 22 个(含 6 个 GH28、2 个 GH35、6 个 GH43、2 个 GH51、1 个 GH53、1 个 GH54、1 个 PL1、1 个 PL3 和 2 个 CE8),参与植物细胞壁组分木聚糖降解的 GH 亚

家族共有 12 个(含 1 个 GH10、2 个 GH11、6 个 GH43、1 个 GH62 和 2 个 GH67)。

3 讨论

随着大量病原真菌基因组测序的完成,许多学者先后开展了真菌基因组规模的分泌蛋白预测分析。例如,于钦亮等^[25]结合 4 个分析软件(SignalP、TargetP、Big-PI Predictor 和 TMHMM)对小麦赤霉病菌(*F. graminearum*)全基因组 11 640 个蛋白编码基因的 N-端氨基酸序列进行了分析,预测出 606 个经典分泌蛋白编码基因,占基因组中编码蛋白基因总数的 5.4%。也有学者先后开展了基于全基因组的稻瘟病菌(*Magnaporthe oryzae*)、马铃薯晚疫病菌(*Phytophthora infestans*)、禾谷炭疽菌(*Colletotrichum graminicola*)和小麦禾生球腔菌(*Mycosphaerella graminicola*)等的分泌蛋白预测分析,结果表明分泌蛋白分别占编码蛋白总数的 9.0%、3.0%、5.2% 和 4.5%^[9-11,26]。综合应用 SignalP、WoLF PSORT、TargetP、TMHMM 和 Big-PI Predictor 等软件,本文首次对 FOM 全基因组 26 811 个蛋白质氨基酸序列进行了分泌蛋白的预测分析;结果表明有 1 145 个蛋白质为经典分泌蛋白,占编码蛋白总数的 4.3%。与其他病原真菌相比, FOM 经典分泌蛋白占全基因组编码蛋白比例相似。与经典分泌蛋白预测相比,病原真菌非经典分泌蛋白的预测分析则开展较少。业艳芬等^[23]用 SecretomeP 对已公布的 1 486 个稻瘟菌小蛋白基因的 N-端氨基酸序列进行了信号肽分析,发现有 949 个为非经典分泌蛋白。本文应用 SecretomeP 对 FOM

表 2 FOM 经典分泌蛋白 CAZymes 的注释与分析

Table 2 Annotation and analysis of the CAZymes of classically secreted proteins in FOM

CAZymes 家族 CAZymes families	CAZymes 亚家族数量 Number of CAZymes subfamilies	CAZymes 数量 Number of CAZymes
GH	36	93
GT	14	76
PL	2	2
CE	6	23
CBM	14	27
AA	6	56
总计 Total	78	277

注：蛋白质的 CAZymes 注释来自 CAZY 数据库。Note: Functional annotation of each protein was from CAZY database.

的非经典分泌蛋白进行了预测分析,结果表明有 8 471 个蛋白质为非经典分泌蛋白,占编码蛋白总数的 31.6%。非经典分泌途径是经典分泌途径的有益补充,有利于蛋白正确构象的形成和正常功能的行使,与外界的胁迫反应、自噬等生命活动密切相关^[27-28]。但目前对于非经典分泌途径的分子机制等仍不是很清楚,有待进一步阐明^[29]。

CAZymes 是病原菌重要的致病因素,具有降解寄主植物细胞壁的作用,是保证病原菌成功感染寄主的关键因素之一^[30]。对 FOM 经典分泌蛋白的分析结果表明,有 277 个分泌蛋白为 CAZymes,占分泌蛋白总数的 24.2%。陈相永等^[24]对核盘菌(*Sclerotinia sclerotiorum*)、灰霉菌(*Botrytis cinerea*)、小麦叶斑病菌(*Stagonospora nodorum*)、玉米小斑病菌(*Cochliobolus heterotrophus*)和大丽轮枝菌(*Verticillium dahliae*) 5 种死体营养型真菌进行了分析,结果表明其 CAZymes 数量为 176~289 个,约占分泌蛋白总数的 26.0%。与其他死体营养型真菌相比,FOM 所分泌的 CAZymes 数量和占分泌蛋白的比例基本相当,并都远高于活体营养型真菌(CAZymes 约占分泌蛋白总数的 15.0%)^[24]。对 CAZymes 亚家族的分析结果表明,与其他死体营养型真菌相比(上述 5 种死体营养型真菌中参与纤维素降解的 GH 亚家族数量为 26~51 个、参与果胶降解的 GH 和 PL 亚家族数量为 36~84 个、参与木聚糖降解的 GH 亚家族为 11~37 个),FOM 所分泌的细胞壁降解酶数量较少,其原因仍有待进一步研究。

应用生物信息学技术,本文从全基因组水平对 FOM 分泌蛋白质组进行了预测,并对经典分泌蛋白的功能注释及 CAZymes 等进行了分析,有助于进一步获得感兴趣的重要功能蛋白。但我们对于 FOM 与寄主互作过程中蛋白质的真实分泌情况等并不清楚。分泌蛋白质组学(secretomics)的兴起和发展,使得大规模地分析 FOM 分泌蛋白质组成为可能。因此,进一步应用蛋白质组学技术,分析不同生理和环境条件、以及病害侵染过程中 FOM 的分泌蛋白质组表达变化,并开展重要分泌蛋白质的基因功能研究,将成为了解分析 FOM 致病机制的重要手段。

参 考 文 献

[1] 李旭,赵娟,徐帅,等.甜瓜枯萎病及其综合防治研究进展[J].

中国植保导刊,2014,34(12):17-21.

- [2] SESTILI S,POLVERARI A,LUONGO L,et al.Distinct colonization patterns and cDNA-AFLP transcriptome profiles in compatible and incompatible interactions between melon and different races of *Fusarium oxysporum* f.sp.*melonis* [J].BMC genomics,2011,12:122.
- [3] 周小林,陈功友,邵元建,等.甜瓜枯萎病菌(*Fusarium oxysporum*)专化型的初步研究[J].植物保护,2008,34(6):80-84.
- [4] WANG Y,WU J,PARK Z Y,et al.Comparative secretome investigation of *Magnaporthe oryzae* proteins responsive to nitrogen starvation[J].Journal of proteome research,2011,10(7):3136-3148.
- [5] XIAO X,XIE J,CHENG J,et al.Novel secretory protein Ss-Caf1 of the plant-pathogenic fungus *Sclerotinia sclerotiorum* is required for host penetration and normal sclerotial development[J].Molecular plant-microbe interactions,2014,27(1):40-55.
- [6] 李云锋,聂燕芳,王振中.植物病原真菌分泌蛋白质组学研究进展[J].微生物学通报,2015,42(6):1101-1107.
- [7] KRAUSE C,RICHTER S,KNÖLL C,et al.Plant secretome—from cellular process to biological activity[J].Biochimica et biophysica acta,2013,1834(11):2429-2441.
- [8] PRUDOVSKY I,TARANTINI F,LANDRISCINA M,et al.Secretion without Golgi[J].Journal of cellular biochemistry,2008,103(5):1327-1343.
- [9] 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(7):785-793.
- [10] 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(2):36-41.
- [11] MORAIS DO AMARAL A,ANTONIOW J,RUDD J J,et al.Defining the predicted protein secretome of the fungal wheat leaf pathogen *Mycosphaerella graminicola* [J].PLoS One,2012,7(12):e49904.
- [12] MA L J,SHEA T,YOUNG S,et al.Genome sequence of *Fusarium oxysporum* f.sp.*melonis* strain NRRL 26406, a fungus causing wilt disease on melon [J].Genome Announc,2014,2(4):e00730-14.
- [13] 陈继圣,郑士琴,郑武,等.全基因组预测稻瘟菌的分泌蛋白[J].中国农业科学,2006,39(12):2474-2482.
- [14] BENDTSEN J D,NIELSEN H,HEIJNE G V,et al.Improved prediction of signal peptides;SignalP 3.0[J].Journal of molecular biology,2004,340(4):783-795.
- [15] LEE S A,WORMSLEY S,KAMOUN S,et al.An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms [J].Yeast,2003,20(7):595-610.
- [16] HORTON P,PARK K J,OBAYASHI T,et al.WoLF PSORT: protein localization predictor[J].Nucleic acids research,2007,35:585-587.
- [17] EMANUELSSON O,BRUNAK S,HEIJNE G V,et al.Locating proteins in the cell using TargetP, SignalP, and related

- tools[J].Nature protocols,2007,2:953-971.
- [18] KROGH A, LARSSON B, HEIJNE G V, et al. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes[J]. Journal of molecular biology, 2001, 305(3): 567-580.
- [19] EISENHABER B, SCHNEIDER G, WILDPANER M, et al. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*[J]. Journal of molecular biology, 2004, 337(2): 243-253.
- [20] BENDTSEN J D, JENSEN L J, BLOM N, et al. Feature based prediction of non-classical and leaderless protein secretion[J]. Protein Eng Des Sel, 2004, 17(4): 349-356.
- [21] LOMBARD V, RAMULU H G, DRULA E, et al. The carbohydrate-active enzymes database (CAZy) in 2013 [J]. Nucleic Acids Res, 2014, 42: D490-D495.
- [22] PARK B H, KARPINETS T V, SYED M H, et al. CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database [J]. Glycobiology, 2010, 20(12): 1574-1584.
- [23] 业艳芬, 李成云, 杨静, 等. 稻瘟病菌含信号肽小蛋白的计算机分析[J]. 生物技术通报, 2006(增刊): 276-282.
- [24] 陈相永, 陈捷胤, 肖红利, 等. 植物病原真菌寄生性与分泌蛋白组 CAZymes 的比较分析[J]. 植物病理学报, 2014, 44(2): 163-172.
- [25] 于钦亮, 马莉, 刘林, 等. 禾谷镰刀菌基因组中含寄主靶向模体分泌蛋白功能的初步分析[J]. 生物技术通报, 2008(1): 160-165.
- [26] 苏源, 李成云, 赵之伟, 等. 稻瘟菌基因组规模分泌蛋白的预测分析[J]. 云南农业大学学报, 2006, 21(3): 271-275, 292.
- [27] AGRAWAL G K, JWA N S, LEBRUN M H, et al. Plant secretome: unlocking secrets of the secreted proteins [J]. Proteomics, 2010, 10(4): 799-827.
- [28] DING Y, WANG J, WANG J, et al. Unconventional protein secretion [J]. Trends Plant Sci, 2012, 17(10): 606-615.
- [29] 张楠楠, 刘欣, 孙晶, 等. 真核细胞非经典蛋白分泌途径 [J]. 遗传, 2009, 31(1): 29-35.
- [30] VAN DEN BRINK J, DE VRIES R P. Fungal enzyme sets for plant polysaccharide degradation [J]. Applied microbiology and biotechnology, 2011, 91(6): 1477-1492.

Genome-scale prediction and analysis of secreted proteins of *Fusarium oxysporum* f. sp. *melonis*

NIE Yanfang¹ ZHOU Gan² HUANG Jiayao² WANG Zhenzhong² LI Yunfeng²

1. College of Materials and Energy, South China Agricultural University, Guangzhou 510642, China;
2. Guangdong Province Key Laboratory of Microbial Signals and Disease Control / College of Agriculture, South China Agricultural University, Guangzhou 510642, China

Abstract Genome-scale prediction of FOM (*Fusarium oxysporum* f. sp. *melonis*) secretome was firstly performed by combining different bioinformatic softwares, including SignalP, TMHMM, WOLFP-RT, TargetP, big-PI Fungal Predictor, and SecretomeP in this study. Of the 26 811 FOM protein sequences, 1 145 sequences (4.3%) were predicted to be classically secreted proteins and 8 471 sequences (31.6%) might be secreted in a non-classical manner. The 1 145 secreted proteins were characterized with 100 to 600 amino acids, and the length of the signal peptides was concentrated between 17 to 22 amino acids. Of the 1 145 classically secreted proteins, a total of 605 proteins were functionally annotated to be involved in carbohydrate metabolic process, transport process, hydrogen peroxide catabolic process and biosynthetic process. Among the 1 145 classically secreted proteins, 277 proteins (24.2%) were predicted to be CAZymes.

Keywords *Fusarium oxysporum* f. sp. *melonis*; secreted protein; signal peptide; carbohydrate-active enzyme; protein function prediction