

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

祝友朋,蔡旺芸,韩长志

(西南林业大学 生物多样性保护与利用学院;云南省森林灾害预警与控制重点实验室,昆明 650224)

摘要:尖孢镰刀菌是一种以土壤习居为主要特征的植物病原菌,广泛分布于世界各地,且寄主广泛,能引起豆科、茄科、葫芦科等科 100 多种植物根腐、茎腐、茎基腐等病害,严重时造成植株萎蔫死亡,严重影响着植物的产量和品质.同时,目前生产上对于该菌的防治多使用甲霜灵·锰锌、多菌灵、百菌清等化学药剂,然而,防治效果不佳,急需开发针对该菌新的作用靶标的药剂.前人已经明确植物病原真菌分泌蛋白在侵染、操控植物等过程中发挥着重要功能,然而,学术界尚未见有关该菌中分泌蛋白的报道.本研究以尖孢镰刀菌菌株(*Fusarium oxysporum* f. sp. lycopersici 4287)的蛋白序列为基础数据,以分泌蛋白所具有的 N-端含有信号肽、不含有跨膜结构域、没有 GPI 锚定位点、将蛋白分泌在胞外 4 大特征为依据,利用生物信息学在线分析程序明确该菌中含有 778 个分泌蛋白,并对上述蛋白开展特征分析,明确上述蛋白的氨基酸长度主要集中在 101~400 aa、氨基酸组成中以 G, T, S, A 较多,信号肽长度主要集中在 16~21 aa、信号肽氨基酸组成中以 P, S, A, T, G 较多,信号肽切割位点为 T-X-T 类型.为今后进一步开展尖孢镰刀菌的分泌蛋白功能解析和开发新型药剂靶标提供了重要的理论基础.

关键词:尖孢镰刀菌;分泌蛋白;预测;特征;生物信息学分析

中图分类号:S435.1

文献标志码:A

尖孢镰刀菌(*Fusarium oxysporum*),属于半知菌类镰刀菌属真菌,是一种以土壤习居为主要特征的植物病原真菌,常引起豆科、茄科、葫芦科等 100 多种植物根腐、茎腐、茎基腐等病害,严重时甚至造成植株萎蔫枯死.前人研究发现该菌广泛分布于世界各地,可在土壤、空气中存活 10 a 以上仍具有较强的致病性^[1].*F. oxysporum* 可侵染并破坏寄主植物的维管束组织,从而导致植物不能正常的进行营养和水分等物质的运输,同时,在植物的生长发育过程中分泌毒素,引起根腐、茎腐、茎基腐等,造成植株萎蔫枯死,严重影响着植物的产量和品质.

前期本研究小组发现,植物病原真菌(禾谷炭疽菌^[2]、希金斯炭疽菌^[3])、卵菌(樟疫霉^[4])以及细菌(黄单胞菌^[5])中存在着大量的分泌蛋白,其在侵染植物的过程中发挥着重要的作用,不仅有助于实现其对植物的侵染、定植和扩散^[6],而且有助于形成对抗植物免疫防卫反应的效应分子、信号分子和诱饵蛋白^[7].然而,尚未见到有关土壤习居菌——尖孢镰刀菌中分泌蛋白的报道.因此,本研究以全基因组序列中已经公布的尖孢镰刀菌菌株(*Fusarium oxysporum* f. sp. lycopersici 4287)的蛋白序列为基础数据,以真菌分泌蛋白所具有的特征(N-端含有信号肽、不含有跨膜结构域、没有 GPI 锚定位点、将蛋白分泌在胞外)为依据,利用生物信息学分析程序 SignalP v4.1, ProtCompB v9.0, TMHMM v2.0, big-PI Fungal predictor, TargetP v1.1 获得分泌蛋白,同时,对上述分泌蛋白所具有的氨基酸长度及组成、信号肽长度及组成等特征进行深入分析,以期为进一步开展防治土壤习居菌药剂靶标开发提供重要的理论基础,也为进一步解析尖孢镰刀菌的分泌蛋白在侵染、操控植物等过程中的功能打下坚实的理论基础.

收稿日期:2018-07-08;**修回日期:**2018-11-10.

基金项目:国家自然科学基金(31560211);云南省研发经费投入补助项目(2016YB462);国家级大学生创新创业训练计划项目(201710677013);西南林业大学大学生科技创新项目(C16094).

作者简介:祝友朋(1996-),男,云南曲靖人,西南林业大学在读研究生,研究方向为经济林木病害生物防治, E-mail: 3420204485@qq.com.

通信作者:韩长志(1981-),河北石家庄人,西南林业大学副教授,博士,研究方向为经济林木病害生物防治与真菌分子生物学, E-mail: hanchangzhi2010@163.com.

1 材料与方 法

1.1 尖孢镰刀菌序列来源

全蛋白组序列来源 NCBI ([https://www.ncbi.nlm.nih.gov/genome/?term = Fusarium%20oxysporum](https://www.ncbi.nlm.nih.gov/genome/?term=Fusarium%20oxysporum)).

1.2 用于分泌蛋白的确定规则

基于前期关于真菌分泌蛋白的预测程序等获得尖孢镰刀菌的分泌蛋白^[8].

1.3 分泌蛋白中的氨基酸残基和信号肽特征分析

运用 Microsoft office 2010 软件对分泌蛋白中的氨基酸残基和信号肽特征进行统计和分析.

2 结果与分析

2.1 分泌蛋白数量为 778 个

F. oxysporum f. sp. *lycopersici* 4 287 共有 27 347 条蛋白质序列,通过 SignalP v4.1 对上述序列进行信号肽预测,明确 2053 个蛋白在 N 端含有典型的信号肽序列,所占比例为 7.51%;进一步利用 ProtComp v9.0 对上述蛋白进行亚细胞定位情况分析,明确 941 个蛋白分泌至胞外,属于胞外分泌蛋白,其余 1 112 个蛋白转运至细胞质膜、线粒体、细胞质、内质网、细胞核、高尔基体中,所占比例分别为 24.73%、30.04%、20.77%、5.49%、6.47%、5.85%,而转运至溶酶体、过氧化物酶体、液泡的蛋白质总的有 74 个,所占比例为 6.65%(图 1).

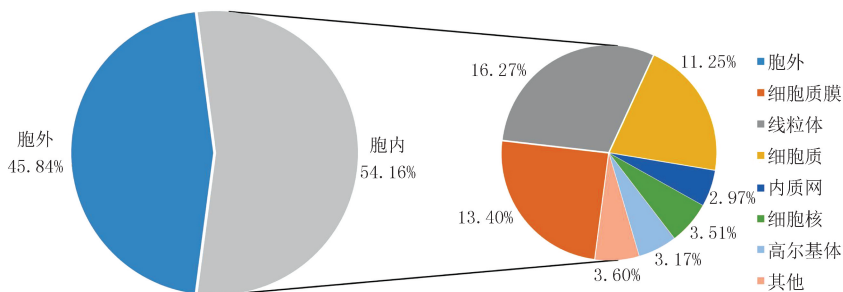


图 1 2053 个具有信号肽的蛋白质的亚细胞定位情况

Fig.1 Subcellular localization of 2053 proteins with signal peptides

对上述所获得的 941 个胞外分泌蛋白进行跨膜结构域分析,明确 903 个蛋白不含有跨膜结构域,所占比例为 95.96%;有 38 个蛋白含有不同数量的跨膜结构域,其中,含有 1 个跨膜结构域的有 32 个,2 个跨膜结构域的有 6 个.鉴于 TMHMM v2.0 程序并不能完全对信号肽序列和所属跨膜结构域序列进行区分,因此,选择 935 个蛋白(不含有跨膜结构域的 903 个蛋白和含有 1 个跨膜结构域的 32 个蛋白)进行 big-PI Fungal predictor 预测,明确 92 个是 GPI 锚定蛋白,823 个蛋白序列不具有 GPI 锚定位点,另外有 20 个蛋白序列因少于 55 个氨基酸未能进行预测.同时,通过对 823 个非 GPI 锚定的蛋白进行分析,明确 42 个含有线粒体目标肽,778 个含有胞外定位信号,另有 3 个蛋白序列超过 4 000 个氨基酸未能进行预测.因此,明确尖孢镰刀菌中含有 778 个分泌蛋白,占总蛋白数量比例 2.84%(图 2).

2.2 分泌蛋白长度集中在 101~400 aa

通过对尖孢镰刀菌 778 个分泌蛋白的氨基酸长度进行统计分析,明确其分泌蛋白长度多集中于 101~

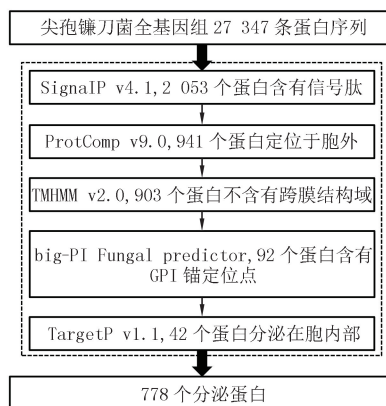


图 2 分泌蛋白确定过程

Fig.2 Secreted protein determination process

400 aa,属于小分子蛋白(图3),所占比例为57.07%,其中,以101~200 aa最多,所占比例为21.34%。

2.2 分泌蛋白的氨基酸组成以G居多

通过对分泌蛋白的氨基酸种类和数量进行统计分析,明确其分泌蛋白中,氨基酸最多的为G(甘氨酸),所占比例为9.20%,其次为T(苏氨酸)、S(丝氨酸)、A(丙氨酸),所占比例分别为8.95%、8.54%、8.20%;然而,尤以H(组氨酸)、W(色氨酸)、M(蛋氨酸)的数量较少,所占比例分别为1.82%、1.80%、1.24%(图4)。

2.3 分泌蛋白的信号肽长度集中在16~21 aa

通过对分泌蛋白的信号肽长度进行统计分析,明确其长度主要集中于16~21 aa(图5),所占比例为81.49%,其中数量最多的是18 aa,所占比例为21.72%。

2.4 分泌蛋白信号肽的氨基酸组成以P居多

通过对分泌蛋白信号肽的氨基酸种类和数量进行统计分析,明确该菌分泌蛋白的信号肽组成中,以P(脯氨酸)最多,所占比例为9.12%,其次为S,A,T,G,所占比例分别为9.05%、8.95%、8.74%、7.69%;然而,F(苯丙氨酸)、H,W,M的数量较少,所占比例分别为2.59%、2.36%、1.39%、0.91%(图6)。

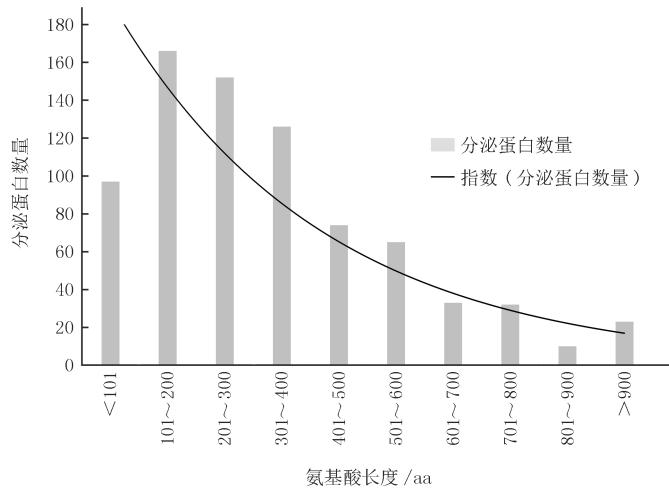


图3 分泌蛋白氨基酸长度分析

Fig.3 Analysis of amino acid length of secreted protein

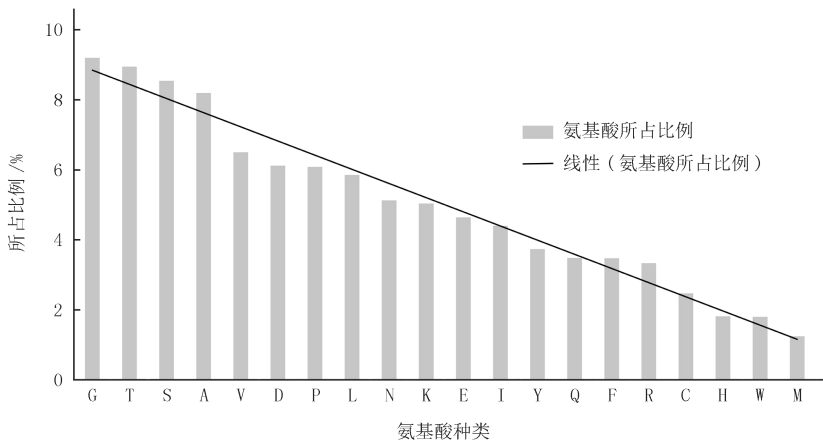


图4 分泌蛋白氨基酸组成情况分析

Fig.4 Analysis of amino acid composition of secreted protein

2.5 分泌蛋白中信号肽切割位点属于T-X-T型

对分泌蛋白中信号肽切割位点的氨基酸组成情况进行分析,明确位于-3位、-2位、-1位、1位、2位、3位最多的氨基酸分别是T,S,T,A,S,S(图7),为T-X-T类型,属于SP I型信号肽识别位点,与LipoP预测相一致.进一步对“-1位、1位”氨基酸组成情况进行分析,明确-1位氨基酸最多的是T,所占比例为11.57%,其次为S,A,G,所占比例分别为10.03%、8.48%、8.23%;1位氨基酸最多的是A,所占比例为10.28%,其次为T,G,S,所占比例分别为10.03%、8.74%、7.58%(图8)。

3 讨 论

近些年,国内外学者就植物病原真菌、卵菌等真核生物以及植物病原细菌、植物病原生防菌中的分泌蛋白及其特征分析研究较多^[7],同时,前期本研究小组研究明确,植物病原真菌、卵菌以及细菌中所含分泌蛋白的数量及其比例均不相同^[4,8,9],所占数量分别为 294~1235、671~3 439、102~218,其所占比例分别为 2.57%~11.12%^[6,8,10-18]、1.05%~2.96%^[4,19]、2.29%~5.41%^[9,20-23]。其中尖孢镰刀菌甜瓜专化性的分泌蛋白数量为 1 145 个,所占比例为 4.34%^[24];尖孢镰刀菌古巴专化性的分泌蛋白数量为 1 054 个,所占比例为 4.69%^[25],然而,本研究发现尖孢镰刀菌中分泌蛋白数量为 778 个,所占比例为 2.84%,此外,尖孢镰刀菌、尖孢镰刀菌古巴专化性和尖孢镰刀菌甜瓜专化性 3 菌相比较,其含有 N 端信号肽的蛋白所占比例分别为 7.51%、7.97%、8.00%,其分泌蛋白大多为小分子蛋白,长度集中在 101~400 aa,信号肽长度集中在 17~21 aa,说明尖孢镰刀菌及其不同的专化型菌株在分泌蛋白的特征方面存在着相似性。但其分泌蛋白数量变化较大,究其原因,认为该菌作为土壤习居菌,其在进化过程中,不断增加分泌蛋白数量,从而获得较强的寄生及腐生在植物上的“活性”,从而保证该菌具有营腐生生活的特性。

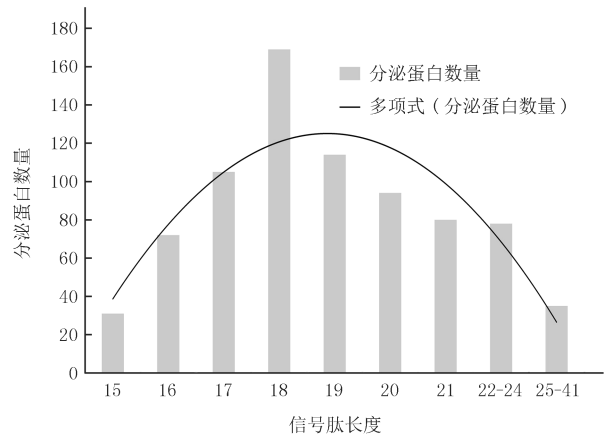


图5 分泌蛋白氨基酸信号肽长度分析

Fig.5 Length analysis of secreted protein amino acid signal peptide

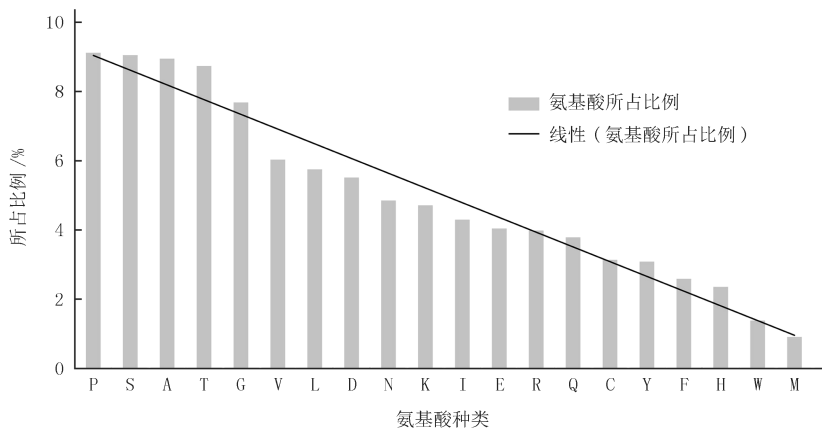


图6 分泌蛋白信号肽的氨基酸组成分析

Fig.6 Analysis of amino acid composition of secreted protein signal peptide

就尖孢镰刀菌中分泌蛋白的氨基酸组成中氨基酸残基的种类及数量和信号肽的氨基酸组成中氨基酸残基的种类及数量存在着相识性,氨基酸残基数量较多的都为 G,T,S,A,P,较少的都为 F,H,W,M.较多的氨基酸残基属于脂肪族氨基酸、羟基类氨基酸及酸性氨基酸等,较少的氨基酸残基属于芳香族类氨基酸、含硫类氨基酸及碱性氨基酸等。推测可能与分泌蛋白的属性相关,使信号肽更容易穿过细胞质膜,从而行使信号指导功能,使分泌蛋白更容易实现其对植物的侵染、定植和扩散。

就尖孢镰刀菌中分泌蛋白所具有的信号肽切割位点类型而言,其属于 T-X-T 型,该类型为首次报道,其与禾谷炭疽菌、希金斯炭疽菌、尖孢镰刀菌古巴专化性等植物病原真菌以及枯草芽孢杆菌、里氏木霉^[15]等不同真菌、细菌中分泌蛋白的信号肽切割位点类型 A-X-A 均不同.T 和 A 在氨基酸归类方面分属于羟基类和脂肪族类,虽然都为 I 型,具有相对保守性特点,但其功能之间的差异性有待于进一步研究。

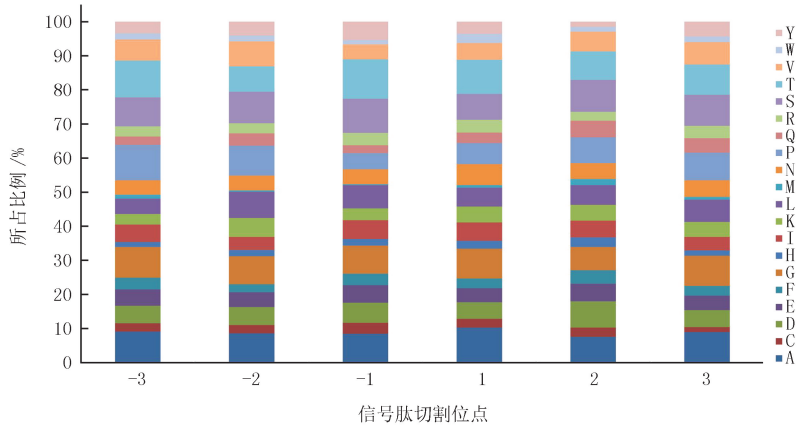


图7 分泌蛋白中信号肽切割位点-3~3氨基酸分布情况

Fig.7 Distribution of signal peptide cleavage site -3~3 amino acids in secreted proteins

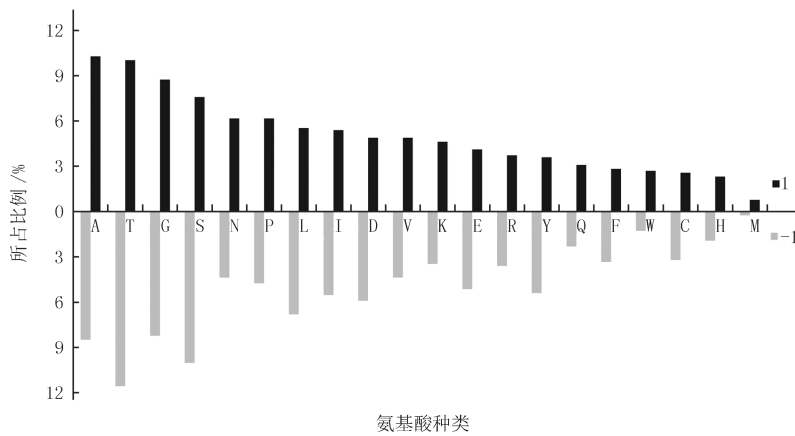


图8 分泌蛋白中信号肽切割位点-1~1氨基酸分布情况

Fig.8 Distribution of signal peptide cleavage site -1~1 amino acid in secreted protein

4 结论

通过生物信息学分析在线分析程序对尖孢镰刀菌菌株的蛋白质序列进行预测分析,最终得到 778 个分泌蛋白,在全蛋白中的比例为 2.84%,并发现分泌蛋白的氨基酸长度集中在 101~400 aa,其组成氨基酸较多的为 G,T,S,A;分泌蛋白信号肽长度集中于 16~21 aa,数量最多的是 18 aa,其组成氨基酸较多的是 P,S,A,T,G;位于信号切割位点前的-3位、-2位、-1位的氨基酸组成中较多的是 T,S,T,属于 T-X-T 类型,属于 SP I 型信号肽识别位点,具有相对保守性特点。

参 考 文 献

- [1] 高晓敏,王璐钢,马立国,等.尖孢镰刀菌致病机理和化感作用研究进展[J].微生物学通报,2014,41(10):2143-2148.
- [2] 韩长志.基于全基因组测序的禾谷炭疽菌中碳水化合物酶类蛋白预测[J].河南师范大学学报(自然科学版),2016,44(4):118-124.
- [3] 韩长志,祝友朋.希金斯炭疽菌中碳水化合物酶类蛋白的理化性质分析[J].华中农业大学学报,2018,37(03):46-51.
- [4] 韩长志.全基因组预测樟疫霉的候选效应分子[J].南京林业大学学报(自然科学版),2015,39(2):69-74.
- [5] 韩长志,王娟.黄单胞菌 *Xanthomonas campestris* pv. *raphani* 756C 中 VI 型分泌蛋白的生物信息学分析[J].华中农业大学学报,2016,35(4):42-48.
- [6] 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.
- [7] 韩长志,许僖.植物病原丝状真菌分泌蛋白及 CAZymes 的研究进展[J].南京林业大学学报(自然科学版),2017,41(5):152-160.
- [8] 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.
- [9] 韩长志.全基因组预测枯草芽孢杆菌 XF-1 的分泌蛋白[J].福建农业学报,2016,31(4):384-388.

- [10] 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(12):2474-2482.
- [11] 陈琦光,王陈骛子,杨媚,等.希金斯刺盘孢全基因组候选效应分子的预测[J].热带作物学报,2015,36(6):1105-1111.
- [12] 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(6):869-876.
- [13] 韩长志.全基因组预测希金斯炭疽菌的候选效应分子[J].生物技术,2015,25(6):546-551.
- [14] 韩长志.基于全基因组测序的粘绿木霉 Gv29-8 中分泌蛋白预测[J].河南师范大学学报(自然科学版),2016,44(6):145-148.
- [15] 韩长志,王娟.基于全基因组序列预测里氏木霉 QM6a 的分泌蛋白[J].华中农业大学学报,2017,36(2):28-32.
- [16] 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(3):271-275.
- [17] 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.
- [18] 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.
- [19] 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):125-133.
- [20] 范成明,李成云,赵明富,等.根瘤土壤杆菌 C58 Cereon 中分泌蛋白信号肽分析[J].微生物学报,2005(4):561-566.
- [21] 黄俊丽,吴金钟,肖崇刚,等.植物病原细菌 *Ralstonia solanacearum* GM11000 中分泌蛋白信号肽分析[J].遗传,2007(11):1409-1416.
- [22] 任文来,韩长志,刘通,等.全基因组预测枯萎芽孢杆菌 BEST195 的分泌蛋白[J].西南林业大学学报,2016,36(6):106-111.
- [23] 肖文超,范会云,白亭亭,等.香蕉细菌性软腐病菌 XJ8-3-3 基因组中 ORFs 的信号肽及分泌蛋白功能预测分析[J].果树学报,2014,31(6):1057-1064.
- [24] 聂燕芳,周淦,黄嘉瑶,等.尖孢镰刀菌甜瓜专业化型基因组规模分泌蛋白的预测与分析[J].华中农业大学学报,2016,35(3):24-29.
- [25] 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(2):288-294.

Prediction of secreted proteins of *Fusarium oxysporum* based on whole genome sequence

Zhu Youpeng, Cai Wangyun, Han Changzhi

(The Key Laboratory of Forest Disaster Warning and Control of Yunnan Province,

College of biodiversity conservation and utilization, Southwest Forestry University, Kunming 650224, China)

Abstract: *Fusarium oxysporum* is a plant pathogen with soil habits as the main feature. It is widely distributed all over the world and has a wide range of hosts. It can cause more than 100 kinds of root rot and stem rot of legumes, Solanaceae and Cucurbitaceae. In some cases, even the plants will die, seriously affecting the yield and quality of the plants. At the same time, at present, chemical agents such as metalaxyl, manganese zinc, carbendazim, and chlorothalonil are often used for the prevention and treatment of the fungus. However, the control effect is not good, and it is urgent to develop a drug for the new target of the fungus. Predecessors have already made clear that the secret proteins of plant pathogenic fungi play important roles in the process of infecting and manipulating plants. However, there has been no report on secreted proteins in this bacterium. In this study, based on the protein sequence of *Fusarium oxysporum* f. sp. *Lycopersici* 4287, the secreted protein has a N-terminal signal peptide, no transmembrane domain, and no GPI anchor site. Based on the four major characteristics of protein secretion in the extracellular, the bioinformatics online analysis program was used to identify 778 secreted proteins in the fungus, and the characteristics of the above proteins were analyzed. The amino acid length of the above proteins was mainly concentrated in 101-400 aa, amino acid composition with G, T, S, A, signal peptide length is mainly concentrated in 16~21 aa, signal peptide amino acid composition, especially P, S, A, T, G, signal The peptide cleavage site is of the TXT type. This study provides an important theoretical basis for the further analysis of the secreted protein function of *F. oxysporum* and the development of new pharmaceutical targets.

Keywords: *Fusarium oxysporum*; secreted protein; prediction; characteristics; bioinformatics analysis

[责任编辑 王凤产]