

基于全基因组测序的粘绿木霉 Gv29-8 中分泌蛋白预测

韩长志

(西南林业大学 林学院;云南省森林灾害预警与控制重点实验室,昆明 650224)

摘要:粘绿木霉 Gv29-8 作为木霉属中重要的生防菌之一,对该菌分泌蛋白进行预测及其特征进行明确具有重要的理论意义。利用 SignalP、ProtComp 等预测程序对该菌中 12427 条蛋白质序列进行分泌蛋白找寻,并对上述分泌蛋白的氨基酸分布、信号肽长度大小及切割位点等性质进行分析。粘绿木霉含有分泌蛋白为 377 个,其氨基酸长度、信号肽长度与植物病原菌不同;信号肽切割位点属于 A-X-A 类型,与其他已经报道的植物病原真菌、卵菌中分泌蛋白信号肽切割位点一致。

关键词:粘绿木霉;分泌蛋白;信号肽;预测程序

中图分类号:S435.1

文献标志码:A

一般而言,真菌的分泌蛋白具有 N-端含有信号肽、无跨膜结构域、无 GPI 锚定位点、不具有将蛋白输送至线粒体或其他胞内细胞器的预测定位信号等四个基本特征^[1]。目前,学术界对于化学药剂的替代品生防菌的研究逐年增加,主要涉及木霉属(*Trichoderma* spp.)、芽孢杆菌属(*Bacillus* spp.)等。就木霉种类而言,已经报道的超过了 140 种,多数木霉可产生抗性代谢产物,如粘绿木霉(*T. virens*)、里氏木霉(*T. reesei*)、绿色木霉(*T. viride*)、哈茨木霉(*T. harzianum*)等。

学术界关于稻瘟菌^[1]、大丽轮枝菌^[2]、致病疫霉^[3]、希金斯炭疽菌^[4]、禾谷炭疽菌^[5]等病原菌以及 *T. reesei*^[6]拮抗菌等植物病原菌的分泌蛋白已经明确。然而,尚未见到对重要植物病原菌具有生防作用的粘绿木霉 Gv29-8 中分泌蛋白的系统性预测报道^[7],因此,开展该菌分泌蛋白的预测具有重要的理论意义。

本研究以粘绿木霉 Gv29-8 中蛋白序列为基础,利用 SignalP、ProtComp 等在线分析程序对分泌蛋白进行找寻,为进一步解析分泌蛋白在该菌发挥拮抗植物病原菌的功能研究打下坚实的理论基础。

1 材料与方法

1.1 粘绿木霉 Gv29-8 蛋白序列来源

T. virens Gv29-8^[7]全基因组序列均来源于联合基因组研究所(http://genome.jgi.doe.gov/TriviGv29_8_2/TriviGv29_8_2.home.html),其 BioProject ID 为 PRJNA183001。

1.2 分泌蛋白确定方法

涉及利用 SignalP v4.1 在线分析实现 N-端信号肽预测^[8],利用 ProtComp 亚细胞定位预测实现亚细胞定位预测,以及利用 TMHMM-v2.0 对蛋白的跨膜结构域进行预测^[9],同时,利用 big-PI Fungal predictor 在线分析实现蛋白脂质锚定修饰的预测^[10],最终利用 TargetP-v1.1^[11]进一步确定预测蛋白的亚细胞定位,获得分泌蛋白。

收稿日期:2016-05-07;修回日期:2016-10-10。

基金项目:国家自然科学基金项目(31560211);云南省森林灾害预警与控制重点实验室开放基金项目(ZK150004);云南省林学一流学科建设项目(51600625);云南省优势特色重点学科生物学一级学科建设项目(50097505);云南省高校林下生物资源保护及利用科技创新团队(2014015)。

作者简介(通信作者):韩长志(1981-),男,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。

1.3 分泌蛋白理化性质预测

利用蛋白质数据库^[12]在线进行分析分泌蛋白理化性质预测。

2 结果与分析

2.1 粘绿木霉 Gv29-8 中含有 337 个分泌蛋白

在 12 427 个蛋白序列中,有 7 条序列氨基酸大小超过预测阈值,未能进行信号肽预测,其余蛋白信号肽预测表明,仅有 681 个序列在 N 端含有典型的信号肽序列,进一步开展细胞定位情况分析,发现 378 个蛋白质分泌至胞外,303 个蛋白分泌到细胞内,其中,转运至细胞质膜、线粒体、内质网膜、细胞质、高尔基体、溶酶体、核、液泡、过氧化物酶体的蛋白数量分别为 127、77、38、24、14、12、5、4、2。然后,对上述转运到胞外的蛋白进行跨膜结构域分析,明确 351 个蛋白不含有跨膜结构域,25 个蛋白含有 1 个跨膜结构域,2 个蛋白含有 2 个跨膜结构域。进一步选择 376 个蛋白质序列进行后续分析。同时,基于 big-PI Predictor 对上述序列进行 GPI 锚定蛋白预测,明确 29 个是 GPI 锚定蛋白,347 个蛋白序列不具有 GPI 锚定位点。最后,利用 TargetP 程序对上述非 GPI 锚定的蛋白进行分析,明确含有胞外定位信号、线粒体目标肽、其他定位信号分别为 337、8、2 个。通过上述分析,明确 337 个分泌蛋白。另外,通过对上述分泌蛋白氨基酸大小进行分析,明确上述分泌蛋白多集中于 101~600 aa 之间,所占比例为 80.12%,尤以 101~200 aa 之间最多,所占比例为 22.55% (图 1)。

2.2 粘绿木霉分泌蛋白的信号肽特征

通过对上述分泌蛋白所含的信号肽长度进行分析,明确信号肽长度为 19 个 aa 的蛋白居多,所占比例为 17.24%,信号肽长度多集中于 17~20 个 aa,上述蛋白质数量最多,所占比例为 62.02% (图 2)。

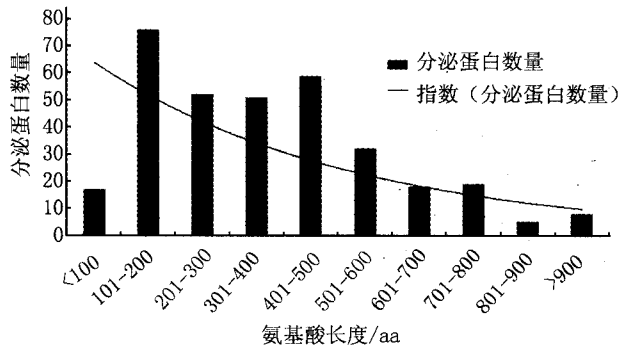


图1 粘绿木霉Gv29-8分泌蛋白氨基酸长度分析

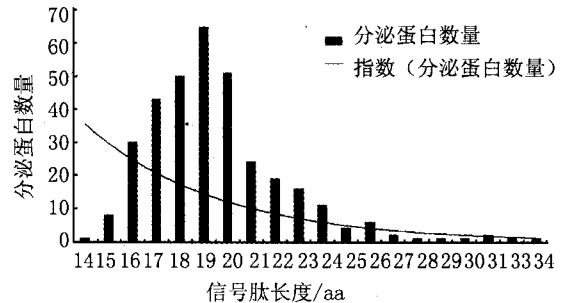


图2 粘绿木霉Gv29-8分泌蛋白信号肽长度分析

此外,通过对 *T. virens* Gv29-8 分泌蛋白信号肽的切割位点-3 位到+3 位进行分析,明确位于-3 位、-2 位、-1 位、1 位、2 位、3 位最多的氨基酸分别为 A, S, A, A, P, T (表 1)。切割位点属于 A-X-A 类型,为 SP I 型信号肽识别位点,与前人通过对禾谷炭疽菌^[5]、粗糙链孢菌^[13]、致病疫霉^[3]等分泌蛋白具有的信号肽酶切位点类型相同^[3,13]。

同时,对组成蛋白质的 20 种氨基酸在信号肽中的分布情况进行统计分析,明确 A 所占比例最大,其次为 L,再次为 S, V, M, T, F, I, G, P, K, R, Q, N, C, Y, H, W, E, D (图 3)。

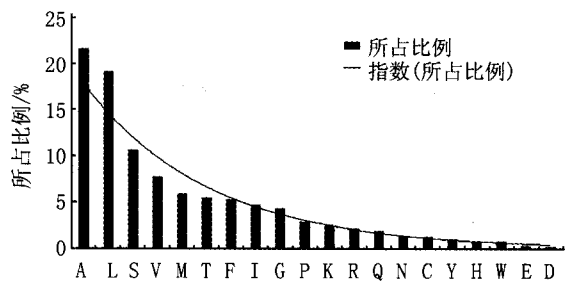


图3 20种氨基酸在粘绿木霉中分泌蛋白信号肽中所占比例情况

3 讨 论

近些年,学术界对植物病原真菌^[1-2,5,13-14]、卵菌^[3,15]等真核生物分泌蛋白的预测较多,而就植物病原菌生防菌分泌蛋白的研究较少^[6]。就生防菌所作用的植物病害而言,则多集中于农作物、蔬菜等方面,而对于木本油料作物生防菌的报道较为少见^[16]。

同时,前期研究发现,植物病原真菌、卵菌所含分泌蛋白的数量不等^[5],所占比例也不同,其中,植物病原真菌分泌蛋白在全蛋白中的比例为 3.65%~9.58%^[5,17-18],卵菌所含分泌蛋白所占比例为 2.96%~4.01%^[3,15]。与植物病原真菌、卵菌分泌蛋白所占比例不同,本研究发现的 337 个分泌蛋白,仅占全蛋白数量的 2.71%,显然低于植物病原真菌、卵菌。推测与益生菌、植物外生菌根菌不同,植物病原菌在与植物长期的互作过程中,进化出较多的分泌蛋白,共同操控植物。此外,前人对于里氏木霉开展的分泌蛋白的预测,明确该木霉具有 294 个分泌蛋白,占全蛋白中的比例为 2.94%^[6],也低于上述植物病原真菌、卵菌,推测木霉分泌蛋白较少,与其具有的功能有关。该研究是对于前期所开展植物病原真菌分泌蛋白的预测以及 G 蛋白信号途径上相关蛋白生物信息学分析研究的重要延伸^[19-20],有助于进一步明确植物病原真菌与生防菌之间在分泌蛋白方面具有的差异性,更好地理解上述两者之间的互作关系。

表 1 粘绿木霉中分泌蛋白信号肽切割位点氨基酸分布情况

氨基酸 种类	-3		-2		-1		1		2		3	
	数量	所占比例/%	数量	所占比例/%	数量	所占比例/%	数量	所占比例/%	数量	所占比例/%	数量	所占比例/%
A	139	41.25	42	12.46	274	81.31	83	24.6	14	4.15	30	8.90
G	11	3.26	5	1.48	28	8.31	13	3.86	11	3.26	9	2.67
S	37	10.98	78	23.15	19	5.64	33	9.79	26	7.72	32	9.50
C	8	2.37	3	0.89	9	2.67	5	1.48	8	2.37	11	3.26
Q	0	0.00	21	6.23	2	0.59	43	12.76	11	3.26	10	2.97
L	2	0.59	52	15.43	1	0.30	31	9.20	16	4.75	38	11.28
T	35	10.39	22	6.53	1	0.30	22	6.53	30	8.90	39	11.57
M	0	0.00	12	3.56	1	0.30	1	0.30	4	1.19	2	0.59
D	0	0.00	9	2.67	1	0.30	14	4.15	15	4.45	16	4.75
P	0	0.00	2	0.59	1	0.30	0	0.00	98	29.08	26	7.72
F	0	0.00	18	5.34	0	0.00	8	2.37	5	1.48	11	3.26
N	0	0.00	14	4.15	0	0.00	7	2.08	17	5.04	11	3.26
V	97	28.78	11	3.26	0	0.00	19	5.64	24	7.12	38	11.28
R	0	0.00	10	2.97	0	0.00	6	1.78	8	2.37	8	2.37
I	8	2.37	8	2.37	0	0.00	10	2.97	10	2.97	26	7.72
H	0	0.00	8	2.37	0	0.00	14	4.15	5	1.48	3	0.89
W	0	0.00	8	2.37	0	0.00	5	1.48	3	0.89	2	0.59
E	0	0.00	7	2.08	0	0.00	10	2.97	12	3.56	10	2.97
Y	0	0.00	6	1.78	0	0.00	5	1.48	7	2.08	5	1.48
K	0	0.00	1	0.30	0	0.00	8	2.37	13	3.86	10	2.97

4 结 论

本研究明确粘绿木霉中含有 337 个分泌蛋白,明显低于植物病原真菌、细菌和卵菌中分泌蛋白的数量和比例,同时,上述分泌蛋白在氨基酸长度、信号肽长度方面也与前期所研究的植物病原菌报道不相同,而无论是植物病原真菌,还是植物病原细菌、卵菌,抑或是本研究的生防菌,其信号肽酶切位点类型均为 AXA,表明信号肽的酶切位点类型具有物种保守性特点。

参 考 文 献

- [1] 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.
- [2] 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.

- [3] 周晓罡,侯思名,陈铎文,等. 马铃薯晚疫病病菌全基因组分泌蛋白的初步分析[J]. 遗传, 2011, 33(7): 125-133.
- [4] 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.
- [5] 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.
- [6] 唐雯,严明. 里氏木霉(*Trichoderma reesei*)分泌组的预测及分析[J]. 微生物学报, 2008(4): 473-479.
- [7] Kubicek C P, Herrera-Estrella A, Seidl-Seiboth V, et al. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma* [J]. *Genome biology*, 2011, 12(4): R40.
- [8] Petersen T N, Brunak S, von Heijne G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions[J]. *Nat Methods*, 2011, 8(10): 785-786.
- [9] Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes[J]. *Journal of molecular biology*, 2001, 305(3): 567-580.
- [10] 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.
- [11] Emanuelsson O, Brunak S, von Heijne G, et al. Locating proteins in the cell using TargetP, SignalP and related tools[J]. *Nature protocols*, 2007, 2(4): 953-971.
- [12] Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server[M]. *The proteomics protocols handbook*: Springer; 2005: 571-607.
- [13] 周晓罡,李成云,赵之伟,等. 粗糙脉孢菌基因组分泌蛋白的初步分析[J]. 遗传, 2006, 38(2): 200-207.
- [14] 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.
- [15] 韩长志. 全基因组预测棒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.
- [16] 韩长志. 植物病害生防菌的研究现状及发展趋势[J]. 中国森林病虫, 2015, 34(1): 33-37, 25.
- [17] 陈琦光,王陈骄子,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.
- [18] 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.
- [19] 韩长志. 胶孢炭疽菌 RGS 蛋白生物信息学分析[J]. 河南师范大学学报(自然科学版), 2015, 43(1): 116-122.
- [20] 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.

Prediction for Secreted Proteins from *Trichoderma Viride* Gv29-8 Genome

HAN Changzhi

(College of Forestry; The Key Laboratory of Forest Disaster

Warning and Control of Yunnan Province, Southwest Forestry University Kunming 650224, China)

Abstract: *Trichoderma virens* Gv29-8 is the important bio-control agent of *Trichoderma* ssp., which can inhibit many plant pathogens. To identify the secreted protein from *T. virens* and clearly its characteristic. 12,427 protein sequences in *T. virens* was analyzed to find the secreted protein with the program including SignalP, ProtComp, TMHMM, big-PI Fungal predictor and TargetP. Meanwhile, the distribution of amino acids, the length of signal peptide as well as the signal peptide cleavage site of secreted protein were analyzed. 377 secreted proteins were found in *T. virens*. And the length of amino acids and the signal peptide were different plant pathogens. The signal peptide cleavage site belongs to AXA type, which was same as other plant pathogenic fungi, bacteria and oomycete.

Keywords: *Trichoderma virens*; secreted protein; signal peptide; prediction algorithm