

# 基于全基因组序列预测里氏木霉 QM6a 的分泌蛋白

韩长志<sup>1</sup> 王娟<sup>2</sup>

1.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;

2.云南省林业科学院/云南省森林植物培育与开发利用重点实验室/国家林业局  
云南珍稀濒危森林植物保护和繁育重点实验室,昆明 650201

**摘要** 里氏木霉(*Trichoderma reesei*)QM6a为重要的产纤维素酶工业生产菌株,为明确其内存在的分泌蛋白,对该菌分泌蛋白进行预测,并明确其特征。利用 SignalP、ProtComp 等 5 个软件对该菌中 9 143 条蛋白质序列进行分泌蛋白预测,并对上述分泌蛋白的氨基酸大小分布情况、信号肽长度及其切割位点等性质进行分析。里氏木霉中含有分泌蛋白 356 个,其氨基酸长度、信号肽长度与其他植物病原菌不同;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、细菌以及卵菌中分泌蛋白信号肽切割位点一致。通过上述生物信息学分析方法有效地实现了里氏木霉分泌蛋白的预测,也证明了分泌蛋白的信号肽切割位点具有物种保守性特点。

**关键词** 里氏木霉;分泌蛋白;信号肽;预测软件

**中图分类号** S 436.341 **文献标识码** A **文章编号** 1000-2421(2017)02-0028-05

分泌蛋白在植物病原真菌感染植物、操控植物防卫反应方面发挥着重要的作用<sup>[1]</sup>,其特征表现为:N-端含有信号肽(signal peptide, SP)、没有跨膜结构域(transmembrane domain, TMD)和糖基磷脂酰肌醇化蛋白(glycosylphosphatidylinositol, GPI)锚定位点以及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<sup>[2]</sup>。目前,学术界对稻瘟菌<sup>[2]</sup>、大丽轮枝菌<sup>[3]</sup>、禾谷炭疽菌<sup>[4]</sup>、希金斯炭疽菌<sup>[5]</sup>等植物病原真菌中的分泌蛋白得以明确。里氏木霉(*Trichoderma reesei*)是一种重要的产纤维素酶工业用菌种之一,与其他木霉作为传统意义的生防菌不同,明确其分泌蛋白特殊性对于后续开展应用研究具有现实意义。目前对于植物非病原菌分泌蛋白的研究尚有待进一步加强,同时,伴随着 *T. reesei* QM6a 中全基因组序列的释放<sup>[6]</sup>,为进一步开展其分泌蛋白的预测即功能解析提供了重要的数据支撑。

目前,对于全基因组序列已经公布的真菌分泌蛋白的预测,则是根据上述特性采用生物信息学分析软件的方法进行。如利用 WOLFP SORT、TM-

HMM 2.0、SignalP 4.1、Phobius 依次进行筛选分泌蛋白,或是仅选用 SignalP 4.1、TargetP、TMHMM 2.0 等软件进行真菌分泌蛋白的筛选<sup>[7]</sup>。

本研究以里氏木霉 QM6a 所具有的蛋白序列为基础,依据前人所报道的有关真菌分泌蛋白所具有的主要特征,利用 SignalP、ProtComp 等 5 个软件对分泌蛋白进行筛选,旨在为进一步开展分泌蛋白后续研究提供重要的理论基础。

## 1 材料与方法

### 1.1 里氏木霉 QM6a 序列来源

*T. reesei* QM6a<sup>[6]</sup>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](http://genome.jgi.doe.gov/TriviGv29_8_2/TriviGv29_8_2.home.html))。

### 1.2 里氏木霉 QM6a 分泌蛋白预测

1)N-端含有信号肽的蛋白筛选。通过 SignalP 4.1 软件进行明确<sup>[8]</sup>。

2)亚细胞定位在胞内的蛋白筛选。利用 ProtComp v9.0 在线明确对 N-端含有信号的蛋白亚细胞定位在胞内的蛋白<sup>[9]</sup>。

收稿日期:2016-09-13

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);云南省教育厅科学研究基金项目(2014Y330);云南省科技计划项目(2015IA005,2015IA004)

韩长志,博士,副教授。研究方向:经济林木病害生物防治与真菌分子生物学。E-mail: hanchangzhi2010@163.com

通信作者:王娟,博士,教授。研究方向:生物多样性、植物生态学和分子生物学。E-mail: schima@163.com

3)排除具有跨膜结构域的蛋白。对亚细胞定位在胞内的蛋白,利用 TMHMM v2.0<sup>[10]</sup> 排除具有跨膜结构域的蛋白。

4)排除具有 GPI 锚定位点的蛋白。利用 big-PI Fungal predictor 在线分析,排除具有 GPI 位点的蛋白<sup>[11]</sup>。

5)分泌蛋白的确定。利用 TargetP-v1.1<sup>[7]</sup> 进一步确定上述蛋白的亚细胞定位情况,最终获得分泌蛋白。

### 1.3 里氏木霉 QM6a 中分泌蛋白的理化性质分析

利用蛋白质数据库<sup>[12]</sup>,对里氏木霉 QM6a 中分泌蛋白的理化性质进行分析。

## 2 结果与分析

### 2.1 里氏木霉 QM6a 中含有 356 个分泌蛋白

在 9 143 个蛋白序列中,有 2 条序列氨基酸大小超过预测阈值,未能进行信号肽预测,对其他蛋白进行信号肽预测,发现 739 个蛋白具有典型的信号肽序列,亚细胞定位分析,则发现 406 个蛋白分泌至细胞外,另有 333 个蛋白则分泌至细胞内不同细胞器中,其中,转运至过氧化物酶体、核、液泡、高尔基体、溶酶体、细胞质、内质网膜、线粒体、细胞质膜的蛋白数量分别为 7、7、8、18、21、24、42、85、121。随后,对 406 个转运到胞外的蛋白进行跨膜结构域分析,明确 377 个蛋白不含有 TMD,26 个蛋白含有 1 个 TMD,3 个蛋白含有 2 个 TMD。由于 TMHMM v2.0 软件所存在的弊端——不能完全对信号肽序列和跨膜结构域的序列进行完全区分,因此,对 403 个蛋白质序列进行进一步分析,包括无跨膜结构域的蛋白 377 个,以及 26 个含有 1 个跨膜结构域的蛋白。

然后,利用 big-PI Predictor 对上述序列进行 GPI 锚定蛋白预测,明确 362 个蛋白不具有 GPI 锚定位点,将其余具有 GPI 锚定位点的 41 个蛋白进行排除。最后,利用 TargetP 软件对上述非 GPI 锚定的蛋白进行分析,结果显示,含有胞外定位信号、线粒体目标肽分别为 356、6 个。

通过上述 SignalP、ProtComp、TMHMM、big-PI predictor、TargetP 等生物信息学软件对里氏木霉 QM6a 中 9 143 条蛋白序列进行预测分析,最终获得 356 个分泌蛋白。对分泌蛋白中氨基酸大小的分布情况进行进一步分析,明确其多集中于 101~600 aa 之间,所占比例高达 77.81%,尤以 101~200

aa 之间最多,所占比例为 20.51%(图 1)。

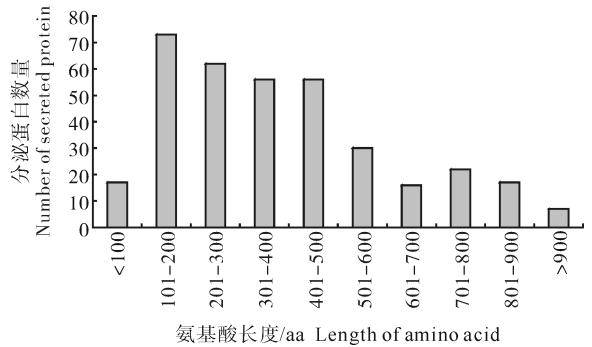


图 1 里氏木霉 QM6a 中分泌蛋白中氨基酸长度分布

Fig.1 The distribution of secreted protein with different length of amino acid in *T. reesei* QM6a

### 2.2 里氏木霉 QM6a 分泌蛋白的信号肽特征

通过对上述 356 个分泌蛋白所含的信号肽长度进行分析,结果显示,信号肽长度为 19 个 aa 的蛋白居多,所占比例为 20.22%,信号肽长度多集中于 17~20 aa,上述蛋白质数量最多,所占比例为 60.39%(图 2)。

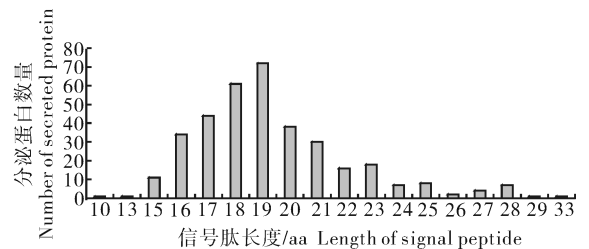


图 2 里氏木霉 QM6a 中分泌蛋白中信号肽长度分布

Fig.2 The distribution of secreted protein with different length of SP in *T. reesei* QM6a

此外,通过对 *T. reesei* QM6a 中分泌蛋白所含有的 SP 的切割位点分析,发现在 -3、-2、-1、1、2 和 3 位中存在最多的氨基酸残基分别为:丙氨酸(A)、丝氨酸(S)、A、A、脯氨酸(P)和苏氨酸(T),所占比例分别为 49.72%、20.51%、83.43%、25.84%、31.46%、12.92%(表 1)。同时,对位于信号肽切割位点之前的氨基酸组成进行分析,结构为 A-S-A,系 A-X-A 类型,与前人对禾谷炭疽菌<sup>[4]</sup>、希金斯炭疽菌<sup>[5]</sup>、粗糙脉孢菌<sup>[13]</sup>、致病疫霉<sup>[14]</sup> 等中分泌蛋白报道相同。

此外,对上述分泌蛋白的 SP 中氨基酸组成进行分析,发现 A 所占比例最大,其次为 L,再次为 S、V、M、T、G、F、I、P、R、K、Q、N、C、H、W、Y、E、D,所占比例分别为 10.18%、7.93%、5.85%、5.71%、4.60%、

表 1 里氏木霉 QM6a 分泌蛋白 SP 切割位点氨基酸情况

Table 1 The amino acids distribution in cleavage sites of SP of secreted proteins in *T. reesei* QM6a

常见氨基酸种类 Types of amino acids	-3		-2		-1		1		2		3	
	数量 No.	所占比例/% Ratio	数量 No.	所占比例/% Ratio	数量 No.	所占比例/% Ratio	数量 No.	所占比例/% Ratio	数量 No.	所占比例/% Ratio	数量 No.	所占比例/% Ratio
A	177	49.72	52	14.61	297	83.43	92	25.84	21	5.90	43	12.08
G	11	3.09	4	1.12	36	10.11	16	4.49	18	5.06	18	5.06
S	24	6.74	73	20.51	18	5.06	30	8.43	27	7.58	31	8.71
L	3	0.84	71	19.94	2	0.56	36	10.11	15	4.21	27	7.58
P	1	0.28	1	0.28	2	0.56	1	0.28	112	31.46	31	8.71
C	8	2.25	1	0.28	1	0.28	3	0.84	3	0.84	8	2.25
T	22	6.18	31	8.71	0	0.00	22	6.18	25	7.02	46	12.92
V	100	28.09	21	5.90	0	0.00	19	5.34	20	5.62	31	8.71
Q	0	0.00	17	4.78	0	0.00	42	11.80	15	4.21	14	3.93
E	0	0.00	13	3.65	0	0.00	14	3.93	14	3.93	7	1.97
H	0	0.00	12	3.37	0	0.00	12	3.37	2	0.56	13	3.65
N	0	0.00	11	3.09	0	0.00	11	3.09	14	3.93	15	4.21
D	0	0.00	10	2.81	0	0.00	9	2.53	21	5.90	12	3.37
F	0	0.00	9	2.53	0	0.00	6	1.69	9	2.53	8	2.25
I	10	2.81	7	1.97	0	0.00	11	3.09	8	2.25	28	7.87
R	0	0.00	7	1.97	0	0.00	9	2.53	11	3.09	6	1.69
W	0	0.00	7	1.97	0	0.00	7	1.97	3	0.84	1	0.28
M	0	0.00	4	1.12	0	0.00	3	0.84	2	0.56	5	1.40
Y	0	0.00	3	0.84	0	0.00	8	2.25	3	0.84	5	1.40
K	0	0.00	2	0.56	0	0.00	5	1.40	13	3.65	7	1.97

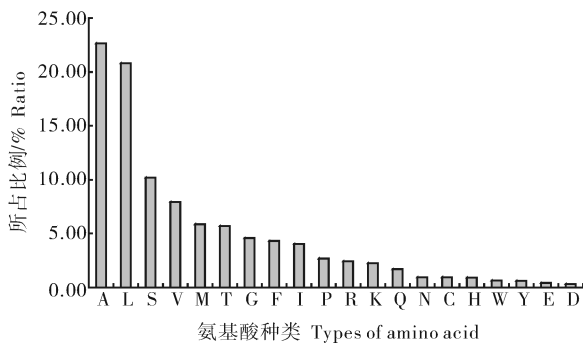


图 3 里氏木霉 QM6a 的分泌蛋白信号肽中氨基酸残基所占比例

Fig.3 The frequency of amino acid residues of signal peptides of secreted proteins in *T. reesei* QM6a

4.33%、4.04%、2.69%、2.44%、2.25%、1.70%、0.96%、0.94%、0.91%、0.67%、0.61%、0.44%、0.32%(图 3)。对上述氨基酸进行进一步分析,发现

非极性氨基酸中的 A、L、V、M 的出现频率相对较高,所占比例在 5%以上;而有带电侧链的氨基酸 E 和 D 的出现频率最低。

### 3 讨论

目前,诸多学者对植物病原真菌<sup>[2-4,13-15]</sup>、卵菌<sup>[14,16]</sup>等真核生物分泌蛋白开展了预测,同时,对植物病原菌的拮抗菌研究较少。尽管前人应用 SignalP 3.0、Proteome Analyst、Pratt、LipoP 1.0 等生物信息学分析工具对里氏木霉中 9 997 条蛋白序列进行信号肽和跨膜区的分析,预测出 294 条分泌蛋白<sup>[17]</sup>。然而,本研究通过 SignalP、ProtComp、TM-HMM、big-PI predictor、TargetP 等生物信息学软件对里氏木霉 QM6a 中 9 143 条蛋白序列进行系统性预测分析,结果明确该菌中存在最终获得 356 个分泌蛋白。

目前,学术界对于化学药剂替代品生防菌的研究越来越多,就生防菌的种类而言,主要涉及木霉属(*Trichoderma* spp.)、芽孢杆菌属(*Bacillus* spp.)等。已经报道的木霉种类超过了140种,多数木霉可产生抗生素代谢产物,如粘绿木霉(*T. virens*)、绿色木霉(*T. viride*)、哈茨木霉(*T. harzianum*)等。重寄生作为木霉侵染植物病原菌的重要作用机制已经明确,在此过程中,木霉分泌一系列诸如纤维素酶、葡聚糖酶、几丁质酶等代谢产物,其抗生素代谢产物也成为植物病害生防菌筛选的重要指标。

此外,就植物病原真菌、拮抗菌中分泌蛋白数量而言,前期研究发现植物病原真菌、卵菌、细菌中所含的分泌蛋白数量不尽相等<sup>[4]</sup>,真菌分泌蛋白在全蛋白中的比例为3.65%~9.58%<sup>[4,18-19]</sup>,卵菌所含分泌蛋白所占比例为2.96%~4.01%<sup>[14,16]</sup>,本研究中分泌蛋白仅占全蛋白数量的比例为3.89%,与前述植物病原菌中分泌蛋白所占比例不同。因此,与前期对于生防菌 XF-1 的研究相一致<sup>[20]</sup>,说明与植物益生菌、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。同时,本研究仅对里氏木霉 QM6a 中所存在的全分泌蛋白开展了生物信息学预测,尚缺乏就某种具体类型分泌蛋白的深入解析<sup>[21]</sup>,今后有待于进一步开展相关研究。

无论是植物病原真菌、细菌以及卵菌,还是工业化纤维素酶生产菌,其分泌蛋白的信号肽酶切位点类型具有物种保守性特点,均为A-X-A。该研究为深入开展里氏木霉 QM6a 生产纤维素酶机制的解析打下了坚实的理论基础。

## 参 考 文 献

- [1] 韩长志.植物与病原菌互作理论研究进展[J].河南农业科学,2012(11):5-8.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] MARTINEZ D, BERKA R M, HENRISSAT B, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*)[J]. Nat Biotechnol, 2008, 26(5):553-560.
- [7] 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.
- [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] EMANUELSSON O, NIELSEN H, BRUNAK S, et al. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence[J]. Journal of molecular biology, 2000, 300(4):1005-1016.
- [10] KROGH A, LARSSON B, VON HEIJNE G, 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.
- [11] 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.
- [12] GASTEIGER E, HOOGLAND C, GATTIKER A, et al. Protein identification and analysis tools on the ExpASY server[M]. Berlin: 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] 周晓罡,侯思名,陈译文,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.
- [15] 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):160-165,180.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.

## Predicting forsecreted proteins from *Trichoderma reesei* QM6a genome

HAN Changzhi<sup>1</sup> WANG Juan<sup>2</sup>

1. *The Key Laboratory of Forest Disaster Warning and Control of Yunnan Province/  
College of Forestry, Southwest Forestry University, Kunming 650224, China;*

2. *Yunnan Academy of Forestry/Yunnan Provincial Key Laboratory of Cultivation  
and Exploitation of Forest Plants/Key Laboratory for Conservation of Rare,  
Endangered & Endemic Forest Plants, State Forestry Administration, Kunming 650201, China*

**Abstract** *Trichoderma reesei* QM6a was an important industrial production strains for cellulase production. To identify the secreted protein from *T. reesei* and clear its characteristic, 9 143 protein sequences in *T. reesei* was analyzed to predict the secreted protein using 5 programs 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. A total of 356 secrete proteins were found in *T. reesei*, and the length of amino acids and the signal peptide were different from the plant pathogens. The signal peptide cleavage site belonged to AXA type, which was consistent with other plant pathogenic fungi, bacteria and oomycete. Through the above bioinformatics analysis, the prediction of secreted proteins can be effectively achieved in *T. reesei*, and the type of signal peptide cleavage site was consistent with other secreted proteins from different species.

**Keywords** *Trichoderma reesei*; secreted protein; signal peptide; prediction algorithm

(责任编辑:边书京)