

DOI: 10.3724/SP.J.1005.2011.00785

# 马铃薯晚疫病菌全基因组分泌蛋白的初步分析

周晓罡<sup>1,2</sup>, 侯思名<sup>3</sup>, 陈铎文<sup>1,2,4</sup>, 陶南<sup>1,2</sup>, 丁玉梅<sup>1,2</sup>, 孙茂林<sup>1,2</sup>, 张绍松<sup>1,2</sup>

1. 云南省农业科学院生物技术与种质资源研究所, 昆明 650223;
2. 云南省农业生物技术重点实验室, 昆明 650223;
3. 昆明学院生命科学与技术系, 昆明 650031;
4. 云南大学生命科学学院, 昆明 650091

**摘要:** 利用马铃薯晚疫病菌全基因组测序结果, 结合计算机技术和生物信息学的方法, 对马铃薯晚疫病菌的蛋白进行分析, 为明确该病原菌与寄主互作的分子机制奠定基础。文章应用信号肽预测软件 SignalP v3.0 和 PSORT, 跨膜螺旋结构预测软件 TMHMM-2.0 和 THUMBUP, GPI 锚定位点预测软件 big-PI Predictor, 亚细胞器中蛋白定位分布预测软件 TargetP v1.01, 对已经公布的马铃薯晚疫病菌全基因组 22 658 个蛋白质氨基酸序列进行分析。结果发现, 晚疫病菌全基因组编码蛋白中有 671 个为潜在的分泌型蛋白, 占编码蛋白总数的 3.0%。其中有 45 个分泌蛋白有功能方面的描述, 其功能涉及细胞代谢、信号转导等方面; 此外, 还有一些与激发子类似的分泌蛋白, 它们可能与晚疫病菌的毒性有关。

**关键词:** 晚疫病菌; 信号肽; 分泌蛋白

## Genome-wide analysis of the secreted proteins of *phytophthora infestans*

ZHOU Xiao-Gang<sup>1,2</sup>, HOU Si-Ming<sup>3</sup>, CHEN Duo-Wen<sup>1,2,4</sup>, TAO Nan<sup>1,2</sup>, DING Yu-Mei<sup>1,2</sup>, SUN Mao-Lin<sup>1,2</sup>, ZHANG Shao-Song<sup>1,2</sup>

1. Biotechnology and Genetic Germplasm Institute, Yunnan Academy of Agricultural Sciences, Kunming 650223, China;
2. Key Laboratory of Agricultural Biotechnology of Yunnan Province, Kunming 650223, China;
3. College of Life Science and Technology, Kunming University, Kunming 650031, China;
4. School of Life Sciences, Yunnan University, Kunming 650091, China

**Abstract:** Based on the *Phytophthora infestans* genome sequence, we used bioinformatics and computer-based prediction algorithms to predict the secreted proteins of *P. infestans* in detail, which would help us to elucidate the molecular mechanism underlying the interaction between the host plants and the *P. infestans*. In this study, the signal peptide prediction algorithms SignalP v3.0 and PSORT, transmembrane helix prediction algorithms TMHMM-2.0 and THUMBUP, GPI-anchoring site prediction algorithm big-PI Predictor, and subcellular protein location distribution algorithm TargetP v1.01 were used to analyze the 22658 protein sequences of *P. infestans* published. Our results suggested that there might be 671 secreted pro-

收稿日期: 2010-10-12; 修回日期: 2011-04-02

基金项目: 云南省十一五科技攻关项目(编号: 2006NG08), 国家自然科学基金项目(编号: 31060021)和云南省自然科学基金项目(编号: 2006C0062M, 2008ZC100M)资助

作者简介: 周晓罡, 硕士, 副研究员, 研究方向: 植物病原微生物生物技术。E-mail: zxcg88@163.com

侯思名, 博士后, 研究方向: 植物遗传育种与基因工程。E-mail: guohuih699@163.com

周晓罡和侯思名同为第一作者。

通讯作者: 孙茂林, 研究员, 研究方向: 植物病害。E-mail: sun\_maolin@tom.com

teins, accounting for 3.0% of the total proteins. Among them, the functions of the 45 secreted proteins had been described previously. Their functions involved cellular metabolism and signal transduction etc. In addition, some of the secreted proteins were functionally similar to elicitor, which were likely to be associated with the virulence of *P. infestans*.

**Keywords:** *Phytophthora infestans*; signal peptides; secreted proteins

马铃薯晚疫病(*Phytophthora infestans*)是引起马铃薯和西红柿晚疫病的致病菌, 每年由 *P. infestans* 引起的马铃薯减产带来的损失超过 5 亿美元<sup>[1]</sup>。 *P. infestans* 属于卵菌, 卵菌包括很多侵染植物、昆虫、甲壳类动物、鱼类、脊椎动物、微生物的致病菌<sup>[2, 3]</sup>, 其中侵染植物的致病菌每年都要造成全世界多种粮食作物和观赏植物的大量损失<sup>[4]</sup>。卵菌具有鞭毛, 在传统分类学上属于真菌。现代分子生物学研究发现, 卵菌在分类地位上与鞭毛菌差距较远, 而与原生藻菌界(*Stramenopiles*)中的蓝藻(异鞭型)和硅藻亲缘关系较为接近<sup>[3, 5, 6]</sup>。

2007 年 3 月公布了晚疫病菌 T30-4 菌株全基因组测序结果, 该菌基因组大小为 237 Mb, GC 含量为 52%, 分布在 8~10 条染色体上<sup>[7]</sup>。在晚疫病菌基因组编码的分泌蛋白中发现一个特异性的与糖基结合的结构域<sup>[7]</sup>。此外, 在其基因组中发现一个与 G 蛋白受体结合相关的基因家族以及几个新的与毒性相关的基因。随着晚疫病菌基因组计划的完成, 标志着对该物种致病机制的解析进入了功能基因组时代, 并为该病原菌的功能基因组学研究提供全面的信息。

1971 年 Blobel 等<sup>[8]</sup>提出的信号肽假说解释了蛋白质合成后的转运和定位, 分泌蛋白通常是指由信号肽引导穿过细胞膜被运输到达胞外的蛋白质。位于分泌蛋白 N-端的起引导作用的 15~60 个氨基酸通常就被称作信号肽。信号肽一般包括以下 3 个结构: 亲水的 N 区、疏水的 H 区和存在信号肽酶切割位点的 C 区<sup>[9]</sup>。其中-3 和-1 位上的氨基酸残基决定着信号肽酶切位点, 正在翻译的核糖体由信号肽引导附着到内质网膜上。当信号肽被内质网膜上的受体识别并结合后进入内质网腔后, 被腔内的信号肽酶识别并降解, 此时, 被信号肽引导进入腔内的新合成的多肽就可以通过内质网的膜系统分泌到胞外。

分泌蛋白是在病原菌细胞中合成后, 通过各种

分泌途径转运到细胞表面、质膜外空间感知周围环境并做出各种适应性反应的主要感应蛋白, 其中许多是细胞分泌到胞外的酶类, 这些酶类对于微生物与周围环境的物质、能量和信息交换具有重要的意义<sup>[10]</sup>。此外, 近年来的研究表明病原物分泌蛋白与寄主抗性基因蛋白产物的分子互作是病原物与寄主互作机制的核心<sup>[11]</sup>。因此, 笔者通过对蛋白质氨基酸序列中 N 端信号肽的预测分析初步判断该蛋白是否为潜在的分泌蛋白及其特性, 将有助于深入研究该病原菌与寄主互作的分子机制, 为持续有效控制马铃薯晚疫病提供新的思路。本研究根据已公布的晚疫病菌基因组信息对其中的分泌蛋白及其功能进行了初步预测分析。

## 1 材料和方法

### 1.1 材料

马铃薯晚疫病菌全基因组 22 658 个蛋白质的氨基酸序列来自 [http://www.broad.mit.edu/annotation/genome/phytophthora\\_infestans](http://www.broad.mit.edu/annotation/genome/phytophthora_infestans)。

### 1.2 方法

SignalP v3.0 (<http://genome.cbs.dtu.dk/services/SignalP-3.0/>)<sup>[12]</sup>软件用于分析马铃薯晚疫病菌全基因组编码氨基酸序列中是否具有信号肽以及该信号肽的切割位点; TMHMM-2.0(<http://genome.cbs.dtu.dk/services/TMHMM-2.0/>)<sup>[13]</sup>软件和 THUMBUP([http://sparks.informatics.iupui.edu/Softwares-Services\\_files/thumbup.htm](http://sparks.informatics.iupui.edu/Softwares-Services_files/thumbup.htm))软件对该序列的跨膜区进行分析<sup>[14]</sup>; PSORT(<http://psort.nibb.ac.jp/form.html>)<sup>[15]</sup>软件进一步预测该信号肽切割位点是否可被识别并被切割; TargetPv1.01(<http://www.cbs.dtu.dk/services/TargetP/>)<sup>[16]</sup>软件预测靶标肽段在亚细胞器中的分布和定位, 进一步确定该信号序列是否为穿膜信号肽。big-PI Predictor([http://mendel.imp.univie.ac.at/gpi/gpi\\_server](http://mendel.imp.univie.ac.at/gpi/gpi_server))<sup>[17]</sup>软件用于鉴定蛋白质的 GPI-锚定位点。

## 2 结果与分析

### 2.1 马铃薯晚疫病病菌全基因组分泌蛋白信号肽预测

SignalP v3.0 是用于预测原核及真核生物氨基酸序列中信号肽切割位点是否存在及其存在的具体位置, 该软件综合了人工神经网络(Artificial Neural networks)算法和隐马可夫模型(Hidden Markov Models)的功能<sup>[18]</sup>。神经网络算法能够对信号肽的分泌途径以及切割位点进行预测。预测值中具有 S-mean 值和 Sprob 值。其中 S-mean 值表示该蛋白质中单个氨基酸信号肽位置的平均预测值, 用来衡量该蛋白信号肽是否进入到正常的分泌途径。Sprob 值主要是用来分析和预测该氨基酸序列是否具有信号肽以及信号肽切割位点和分布。通过人工神经网络算法(NN)预测出有 3 029 个蛋白序列具有分泌型信号肽, 占 13.4%。通过隐马可夫模型(HMM)预测出有 3 031 个蛋白具有分泌型信号肽, 占 13.4%。两种算法的预测结果仅相差 2 个蛋白。对上述结果进一步使用分泌信号肽判别式<sup>[19]</sup>:  $L = -918.235 - 123.455 \times (\text{S-mean 值}) + 1983.44 \times (\text{HMM 值})$  对上述预测值进行判定。如果 L 值 > 0, 则判定该蛋白序列具有分泌型信号肽。经计算分析后表明该物种具有分泌型信号肽的蛋白质序列有 1 626 个。

在含有分泌型信号肽的蛋白质序列中, 如果含有跨膜区则表明该蛋白可能为膜受体, 也可能是膜上的锚定蛋白或者离子通道蛋白。因此, 我们用 TMHMM v2.0 软件来预测蛋白序列的跨膜螺旋结构, 排除具有跨膜结构域的蛋白序列。结果发现在 1 626 个蛋白序列中, 1 125 个蛋白序列跨膜螺旋为 0, 282 个蛋白序列有 1 个跨膜螺旋区, 其余 219 个蛋白序列均有两个以上的跨膜区, 因此既符合跨膜螺旋为 0 又具有分泌型信号肽的蛋白序列有 1 125 个。

采用预测跨膜螺旋软件 THUMBUP 对上述 1 125 个蛋白序列进行预测分析。结果表明有 841 个蛋白序列没有跨膜螺旋区, 220 个蛋白序列有 1 个跨膜螺旋区, 其余 64 个蛋白序列均有两个以上的跨膜区, 因此符合条件的蛋白序列有 841 个。

为了确定这些分泌型信号肽的准确性以及该信号肽切割位点是否可被识别并切割, 我们又采用 PSORT 软件对这 841 个蛋白序列进行补充分析。PSORT 软件的预测结果分为两部分: 首先预测该蛋

白序列是否具有信号肽, 结果以 MCG 显示, 当 MCG 值 > 0 即表明该蛋白具有信号肽; 其次, 采用切割位点模式(即-3 位到-1 位)对该信号肽是否可被信号肽酶切割进一步预测。结果表明在 841 个蛋白序列中有 163 个蛋白为没有信号肽或者是不被切割的信号肽, 具有切割位点并且被信号肽酶识别并切割的信号肽有 678 个。

采用 TargetP v1.01 对晚疫病病菌的亚细胞结构中真核蛋白序列的位置与分布进行预测。该软件主要是通过通过对已知序列 N 端前导链中叶绿体转运肽段(cTP)、线粒体的靶标肽段(mTP)以及具有分泌途径的信号肽(SP)进行预测分析, 从而确定该蛋白序列是否具有信号肽分泌途径。对上述 678 蛋白序列进行分析, 发现除 7 个蛋白序列分布于其他细胞器外, 剩余的 671 个都具有分泌途径的信号肽(SP)。GPI 锚定在真核生物蛋白质转录后修饰中是比较常见的。在多肽的 C 端切割后 GPI 基团结合在羧基端的  $\gamma$  位点, 使蛋白锚定在质膜上, 所以在预测分泌蛋白时要将这一部分蛋白质排除。我们采用 GPI 锚定位点预测软件 big-PI predictor 对 671 个氨基酸序列进行分析, 预测该蛋白是否锚定在细胞膜上。

结果表明这 671 个氨基酸序列均不具锚定位点。

综合以上 6 种算法程序(图 1)对晚疫病病菌全基因组分泌蛋白的测算结果, 发现在该物种 22 658 个蛋白序列中同时满足这 6 种软件具可溶性分泌型信号肽的蛋白序列有 671 个, 占总数的 3.0%。在这 671 个具有分泌信号肽的蛋白序列中, 编码蛋白序列的最小“开放阅读框”(Open reading frame, ORFs)长度是 455 bp, 最大的长度为 9 050 bp, 平均长度为 1 013 bp。大多集中在 600~899 bp(图 2)。分泌型信号肽的长度分布范围为 14~57aa, 平均长度为 22aa。其中长度为 21aa 的分泌型信号肽数量最多, 达到 106 个。其次是长度为 20aa 的信号肽, 有 93 个(图 3)。

### 2.2 马铃薯晚疫病病菌全基因组分泌蛋白信号肽特征分析

目前的研究表明, 大多数物种的信号肽主要是通过 4 种信号肽酶识别位点被信号肽酶所识别并被切割, 从而使成熟蛋白穿过膜转运到细胞不同的部位<sup>[20]</sup>。通过对晚疫病病菌分泌蛋白信号肽切割位点研究分析表明, 在该信号肽与成熟蛋白切割位点-3 位~+3 位

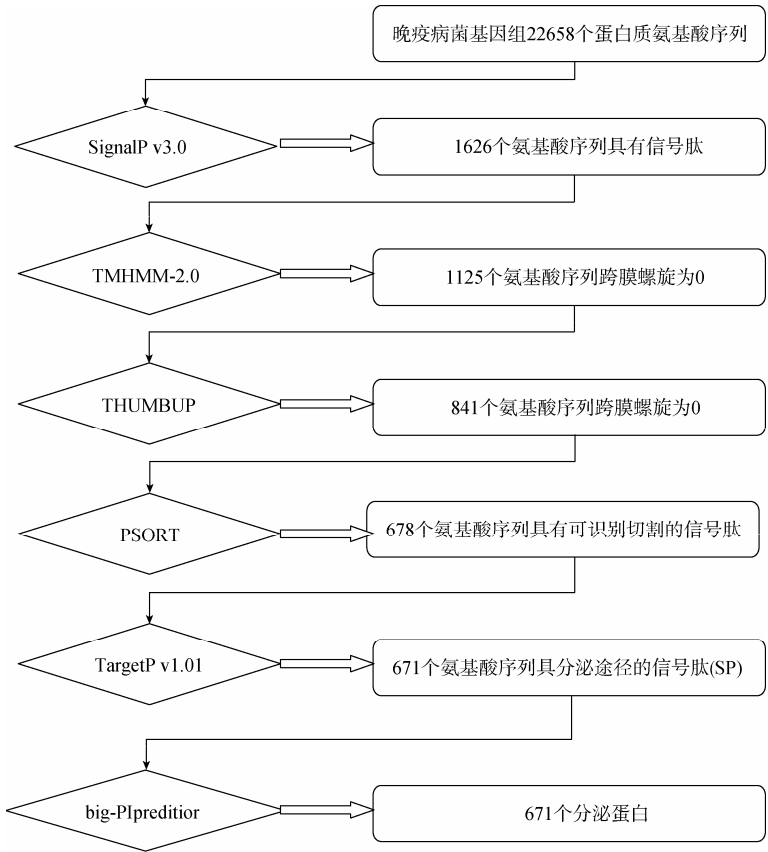


图 1 分泌蛋白预测流程图

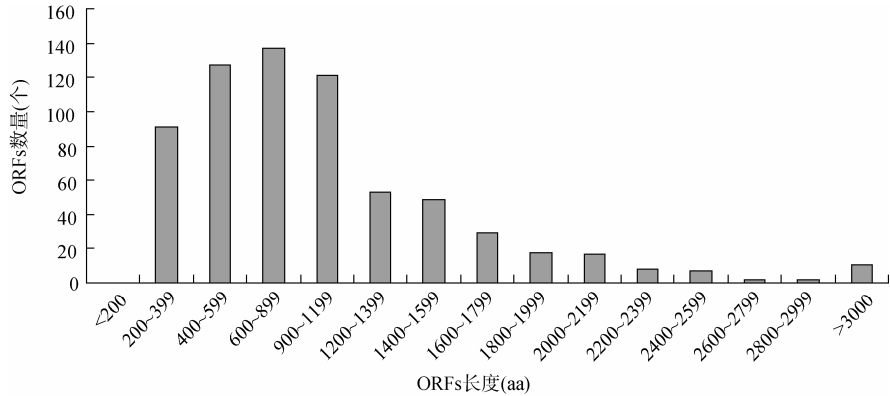


图 2 分泌蛋白 ORF 长度分布

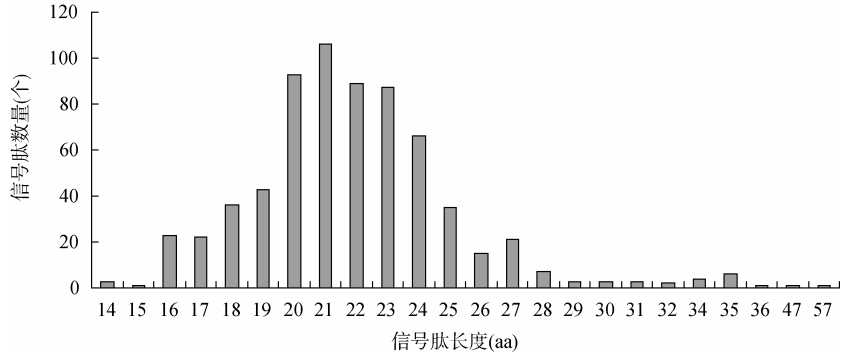


图 3 信号肽长度分布

分布的氨基酸数量最多的都是丙氨酸(A)(表 1)。从该物种信号肽切割位点氨基酸种类分布情况可以看出, 该物种的信号肽切割位点是比较典型的, 可被 SPaseI 型信号肽酶识别的结构, 即-3 到-1 位氨基酸组成为 A-X-A。

我们还对 20 种氨基酸在 671 个分泌蛋白信号肽中

出现频率进行了分析(图 4)。结果表明, 在组成信号肽的氨基酸中, 非极性、疏水的氨基酸出现频率最高(A、V、L、G、I、P), 占 46.6%。其次是有极性、不带电荷的氨基酸(S、T、C、M、N、Q), 占 30.8%, 带负电荷的酸性氨基酸(A、E)占 11.7%, 带正电荷的碱性氨基酸(K、R、H)占 6.2%, 芳香族氨基酸(W、F、Y)占 4.6%。

表 1 马铃薯晚疫病病菌信号肽切割位点氨基酸组成与分布

氨基酸种类	信号肽中 20 种氨基酸在酶切位点-3~+3 使用频率											
	-3		-2		-1		1		2		3	
	数量	百分比 (%)	数量	百分比 (%)	数量	百分比 (%)	数量	百分比 (%)	数量	百分比 (%)	数量	百分比 (%)
A	135	20.1	145	21.6	150	22.4	114	17	97	14.5	89	13.3
C	15	2.2	14	2.1	13	1.9	13	1.9	11	1.6	15	2.2
D	32	4.8	39	5.8	45	6.7	57	8.5	51	7.6	60	8.9
E	21	3.1	26	3.9	23	3.4	44	6.6	35	5.2	40	6
F	19	2.8	21	3.1	20	3	15	2.2	16	2.4	19	2.8
G	44	6.6	25	3.7	36	5.4	26	3.9	37	5.5	47	7
H	8	1.2	11	1.6	7	1	10	1.5	13	1.9	7	1
I	27	4	26	3.9	25	3.7	21	3.1	14	2.1	27	4
K	8	1.2	11	1.6	14	2.1	10	1.5	21	3.1	31	4.6
L	61	9.1	50	7.5	65	9.7	46	6.9	33	4.9	39	5.8
M	6	0.9	9	1.3	6	0.9	8	1.2	8	1.2	7	1
N	19	2.8	26	3.9	26	3.9	22	3.3	31	4.6	34	5.1
P	17	2.5	22	3.3	21	3.1	22	3.3	35	5.2	39	5.8
Q	19	2.8	15	2.2	18	2.7	23	3.4	28	4.2	25	3.7
R	10	1.5	9	1.3	16	2.4	23	3.4	14	2.1	25	3.7
S	88	13.1	84	12.5	72	10.7	87	13	89	13.3	60	8.9
T	65	9.7	47	7	47	7	56	8.3	77	11.5	58	8.6
V	70	10.4	80	11.9	51	7.6	63	9.4	44	6.6	35	5.2
W	3	0.4	3	0.4	6	0.9	1	0.1	2	0.3	5	0.7
Y	4	0.60%	8	1.2	10	1.5	10	1.5	15	2.2	9	1.3

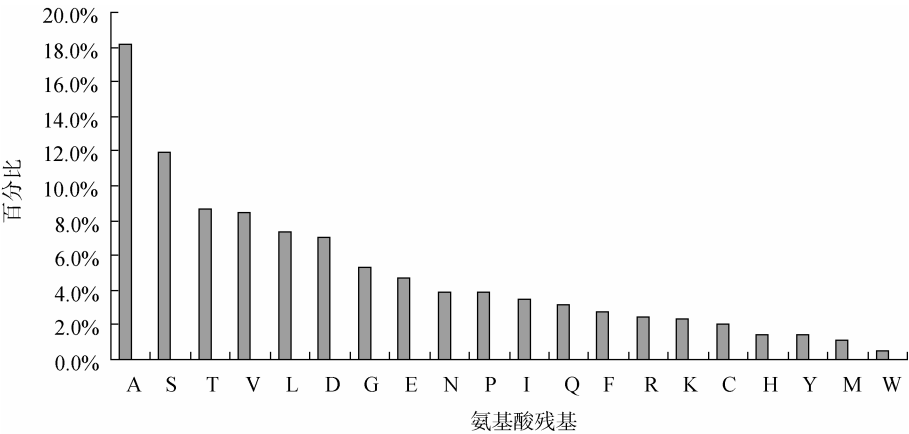


图 4 20 种氨基酸在分泌蛋白信号肽中所占比例

对比目前已经进行分泌蛋白预测的物种(表 2)可以看出,晚疫病菌的切割位点从-3~+3 使用频率最高的氨基酸都是 A(丙氨酸),这说明晚疫病菌的切割位点相对保守,这一现象在其他物种中是比较少见的<sup>[21~23]</sup>,另一方面,随着进化程度的增高,在切割位点上氨基酸使用的种类随之增多<sup>[21~23]</sup>。通过对比也发现在晚疫病菌切割位点的-3~+3 每个位置上 20 种氨基酸都有使用,这一现象即使在线虫中也是没有的<sup>[23]</sup>。这说明晚疫病菌在氨基酸的使用上又具有很高的多态性。这会不会和晚疫病菌基因组的可塑性有关?这些都需要进一步的研究。

对马铃薯晚疫病菌(*P. infestans*)、粗糙脉孢霉(*Neurospora crassa*)<sup>[21]</sup>、根癌土壤杆菌(*Agrobacterium tumefaciens*)<sup>[22]</sup>和秀丽小杆线虫(*Caenorhadtis elegans*)<sup>[23]</sup>的切割位点处氨基酸使用频率的变化情况进行分析(表 3)。从中可以看出, *A. tumefaciens* 切割位点氨基酸使用频率的变化最为明显。如在-1 位置上, A(丙氨酸)使用频率最高,为 93.7%。而 G(甘氨酸)、S(丝氨酸)使用频率分别为 1.9%和 4.4%,其它种类氨基酸都未使用。而 *C. elegans* 和 *P. infestans* 的变化频率相对来说变化较小。如 *P. infestans*, 它的氨基酸使用频率变化范围最大的在-1 位,范围从 0.9%(M, 甲硫氨酸)到 22.4%(A, 丙氨酸)。*Caenorhadtis elegans* 氨基酸使用频率变化范围最大的在-1 位,范围从 0%(C、L、

K、M、V、Y)到 49.9%(A)。值得注意的是,在这几个物种的切割位点的-1 处氨基酸使用频率的变化范围都是最大的。

显示出这一位置的多样性。

2.3 马铃薯晚疫病菌全基因组分泌蛋白功能预测分析

在该物种 671 个预测的潜在分泌蛋白中, 45 个有功能描述, 其余 626 个则描述为预测蛋白(Predicted protein)、假定蛋白(Hypothetical protein)、保守假定蛋白(Conserved hypothetical protein)。在已有功能描述的这些蛋白中, 其功能主要集中在细胞信号识别与传导(如基质细胞衍生因子 2 前体)、细胞代谢(如中性  $\alpha$ -葡萄糖苷酶 AB 前体等)、能量形成及转运等方面。还有些降解细胞壁组分的酶类, 如葡萄糖苷酶等。还有一些与激发子相似的假定蛋白(如与激发素蛋白相似的假定蛋白等), 这些蛋白与晚疫病菌的毒性有关(表 4)。

目前, 已有多个物种开展了分泌蛋白功能预测分析。吴毅歆等<sup>[24]</sup>发现在水稻的 4 433 个分泌蛋白中, 酶类有 1 102 个, 而且以激酶数量最多, 有 309 个, 占 28%。其次是水解酶类和氧化酶类, 分别有 157 个和 136 个。于钦亮等<sup>[25]</sup>利用 MEME 软件对禾谷镰刀菌(*Fusarium graminearum*)的分泌蛋白进行

表 2 信号肽切割位点的氨基酸组成

物种	信号肽中 20 种氨基酸在酶切位点-3~+3 使用频率											
	-3		-2		-1		1		2		3	
	种类	频率最高	种类	频率最高	种类	频率最高	种类	频率最高	种类	频率最高	种类	频率最高
<i>Phytophthora infestans</i>	20	A	20	A	20	A	20	A	20	A	20	A
<i>Neurospora crassa</i>	11	A	20	L	10	A	20	A	20	P	-	-
<i>Agrobacterium tumefaciens</i>	8	A	18	L	3	A	18	A	15	D	20	L
<i>Caenorhadtis elegans</i>	18	V	20	S	13	A	19	Q	20	P	20	I

表 3 信号肽切割位点氨基酸使用频率变化

物种	信号肽中 20 种氨基酸在酶切位点-3~+3 使用频率						
	-3	-2	-1	1	2	3	
<i>Phytophthora infestans</i>	0.6%~20.1%	0.4%~21.6%	0.9%~22.4%	0.1%~17.0%	0.3%~14.5%	0.7%~13.3%	
<i>Neurospora crassa</i>	0%~38.7%	0.5%~2.2%	0%~77.3%	0.5%~2.6%	1.1%~33.4%	-	
<i>Agrobacterium tumefaciens</i>	0%~81.0%	0%~17.7%	0%~93.7%	0%~25.9%	0%~28.5%	0.6%~11.4%	
<i>Caenorhadtis elegans</i>	0%~30.4%	0.6%~15.2%	0%~49.9%	0%~18.3%	0.8%~11.5%	0.8%~8.7%	

表 4 部分晚疫病病菌分泌蛋白功能描述

基因编号	功能描述	蛋白分子量(Da)	等电点(pI)	脂肪族氨基酸指数*	ORF 长度(aa)
PITG_06984.1	基质细胞衍生因子 2 前体	33050.7	5.75	62.27	957
PITG_09817.1	磷酸核糖焦磷酸激酶	40693	9.11	97.51	1247
PITG_09467.1	果胶裂解酶前体	20406.2	8.98	87.74	696
PITG_08650.1	中性 $\alpha$ -葡萄糖苷酶 AB 前体	48928.2	4.32	79.36	987
PITG_11112.1	中性 $\alpha$ -葡萄糖苷酶 AB 前体	37131.2	5.84	83.78	987
PITG_00527.1	胞内结合蛋白 3 前体	72492.7	4.72	87.55	2088
PITG_05339.1	与转谷氨酰胺酶激活子前体相似的假定蛋白	63774.7	4.49	59.23	1843
PITG_16981.1	与血小板反应素-1 相似的假定蛋白	258922.6	5.97	56.56	7212
PITG_12878.1	与性别诱导蛋白 3 相似的假定蛋白	24366.4	5.22	54.87	753
PITG_05156.1	与分泌蛋白 OPEL 相似的假定蛋白	52162.3	4.78	66.74	1892
PITG_03694.1	与蛋白二硫化物异构酶 6 相似的假定蛋白	48756	4.8	77.59	1427
PITG_16822.1	与多聚半乳糖醛酸酶 2 相似的假定蛋白	34537.8	9.42	80.81	1083
PITG_10255.1	与多聚半乳糖醛酸酶相似的假定蛋白	39484.5	8.93	83.8	1227
PITG_19620.1	与多聚半乳糖醛酸酶相似的假定蛋白	39047.9	9.34	81.14	1722
PITG_19455.1	与多聚半乳糖醛酸酶相似的假定蛋白	34925.4	9.38	83.02	1157
PITG_21245.1	与多聚半乳糖醛酸酶相似的假定蛋白	39498.3	9.26	79.76	2171
PITG_14669.1	与磷脂酶 A2 相似的假定蛋白	27432.5	5.89	61.47	938
PITG_07638.1	与磷脂酶 A2 相似的假定蛋白	27432.5	5.89	61.47	933
PITG_15299.1	与果胶酸盐裂解酶相似的假定蛋白	28178	8.09	86.98	836
PITG_12828.1	与果胶酸盐裂解酶相似的假定蛋白	27873.2	5.56	66.18	789
PITG_12829.1	与果胶酸盐裂解酶相似的假定蛋白	27873.2	5.56	66.18	789
PITG_16866.1	与坏死诱导蛋白相似的假定蛋白	25518.2	5.27	74.24	846
PITG_15418.1	与坏死和乙烯诱导蛋白 7 相似的假定蛋白	38439.2	6.32	66.05	2191
PITG_16867.1	与坏死和乙烯诱导蛋白 4-2 相似的假定蛋白	26377.4	5.83	67.09	746
PITG_18453.1	与坏死和乙烯诱导蛋白 3 相似的假定蛋白	28114.7	5.25	84.4	794
PITG_13655.1	与葡聚糖酶抑制蛋白 3 相似的假定蛋白	26505.2	9.19	78.95	774
PITG_13638.1	与葡聚糖酶抑制蛋白 2 相似的假定蛋白	26716.9	4.95	83.14	777
PITG_13641.1	与葡聚糖酶抑制蛋白 2 相似的假定蛋白	23792.6	5.76	73.04	771
PITG_13671.1	与葡聚糖酶抑制蛋白 2 相似的假定蛋白	27124.6	5.72	82.67	839
PITG_12562.1	与 $\gamma$ -巨精胺相似的假定蛋白	17734.7	4.29	68.14	745
PITG_12556.1	与激发素蛋白相似的假定蛋白	18463.5	3.53	72.73	779
PITG_12561.1	与激发素蛋白相似的假定蛋白	18710.3	3.73	63.56	1792
PITG_03020.1	与半胱氨酸蛋白酶相似的假定蛋白	58419.5	4.88	73.64	695
PITG_12832.1	与角质酶相似的假定蛋白	22083.5	8.96	92.21	686
PITG_12361.1	与角质酶相似的假定蛋白	21885	5.53	83.91	754
PITG_12401.1	与角质酶相似的假定蛋白	22030.2	5.28	83.91	686
PITG_12422.1	与角质酶相似的假定蛋白	21849	5.53	84.84	1781
PITG_21378.1	与重链和轻链相似的假定蛋白	57322.9	4.61	82.28	720
PITG_16992.1	与细胞 12A 内切葡聚糖酶相似的假定蛋白	25858	4.87	82.05	723
PITG_08944.1	与细胞 12A 内切葡聚糖酶相似的假定蛋白	26493	8.32	75.54	995
PITG_08518.1	与 25 kDa 蛋白质激酶相似的假定蛋白	31305.1	9.43	53.49	623
PITG_12551.1	激素前体	12312	4.86	83.56	1415
PITG_18934.1	钙网蛋白前体	47290.1	4.44	62.53	1415
PITG_15358.1	钙网蛋白前体	47292.1	4.45	61.6	1104
PITG_09470.1	阿拉伯糖胶内切-1, 5- $\alpha$ -L-阿拉伯糖苷酶 A 前体	34747.8	6.08	70.38	493

注: \*脂肪族氨基酸指数: 所有氨基酸中脂肪族氨基酸所占比例。

分析后发现 157 个分泌蛋白 N 端信号肽下游 120 个氨基酸内存在一个有四个碱基组成的 RXLX 模体, 该模体存在于卵菌和疟原虫的致病蛋白中<sup>[26]</sup>, 进一步的分析表明, 在 157 个具有 RXLX 模体的蛋白中, 有 79 个具有可预测的功能性描述, 其功能大多是与代谢、蛋白折叠、修饰有关, 一些则与能量、细胞循环和处理有关。由此可见, 具有该模体结构的分泌蛋白可在不同的生理活动中发挥作用。

周晓罡等<sup>[21]</sup>报道在 *N.crassa* 437 个分泌蛋白中, 有 205 个具有可预测的功能。已知功能的蛋白主要为参与细胞代谢的各种酶和维持细胞自身能量形成、转运以及结构定位和自身修复防卫等非酶蛋白类。苏源等<sup>[27]</sup>报道了稻瘟病菌(*Magnaporthe grisea*)分泌蛋白的功能有参与细胞代谢、细胞修复与防卫机制和降解细胞壁。而且稻瘟病菌产生的各种降解细胞壁的酶类可能与它在侵染水稻时形成附着孢穿过细胞壁的机制有关。

### 3 讨论

具有信号肽的分泌蛋白往往在生物体的生长、发育以及细胞凋亡过程中发挥着重要的作用。这类蛋白质通常在细胞内合成, 然后通过内质网-高尔基体-质膜的分泌途径分泌到细胞外<sup>[28]</sup>。尽管马铃薯晚疫病菌(*P. infestans*)、柃树猝死病菌(*Phytophthora ramorum*)、大豆疫霉菌(*Phytophthora sojae*)这 3 种危害严重的植物致病卵菌全基因组序列已经公布, 但是对于它们具有重要生理功能的分泌蛋白种类、分泌途径及功能还未见相关报道。

本文通过对 *P. infestans* 全基因组预测得出的 671 个分泌蛋白功能进行分析, 发现其中主要是有关于细胞代谢(如中性  $\alpha$ -葡萄糖苷酶 AB 前体、与内切葡聚糖酶相似的假定蛋白等)、细胞信号转导(如基质细胞衍生因子 2 前体、磷酸核糖焦磷酸激酶等)和降解细胞壁(如果胶裂解酶前体、与果胶酸盐裂解酶相似的假定蛋白、与磷脂酶 A2 相似的假定蛋白等)的酶类, 这一点与 *M.grisea* 分泌蛋白的功能较为相似<sup>[27]</sup>。另外还有类似激发子的假定蛋白(如与激发素蛋白相似的假定蛋白、激发素前体等), 这些蛋白可能与 *P. infestans* 的致病性密切相关<sup>[29]</sup>。从分析结果发现致病菌与非致病菌在分泌蛋白的功能方面存在

一定的差别, 如 *P. infestans* 和 *M. grisea* 分泌蛋白中有一部分是降解细胞壁的酶类, 而这一类酶在 *N. crassa* 中却未发现。此外, 还包含了一些病程相关蛋白。由此可见, 分泌蛋白在各种生物中的正常生命活动中起着非常重要的作用, 而且不同物种之间分泌蛋白功能差别也较大。

随着 *P. ramorum*、*P. sojae* 等物种全基因组序列的公布, 为我们进一步利用生物信息学手段对这些卵菌基因组中有信号肽的分泌蛋白进行对比分析提供了条件, 从而可对该属病原菌分泌蛋白结构、功能等特征进行差异性比较, 为今后利用实验开展该物种分泌蛋白组研究提供依据。

### 参考文献(References):

- [1] Duncan JM. *Phytophthora*—an abiding threat to our crops. *Microbiol Today*, 1999, 26(99): 114–116.
- [2] Kamoun S. Molecular genetics of pathogenic oomycetes. *Eukaryotic Cell*, 2003, 2(2): 191–199.
- [3] Margulis L, Schwartz KV. Five Kingdoms—An Illustrated Guide to the Phyla of Life on Earth. San Francisco: Freeman, 1982.
- [4] Erwin DC, Ribeiro OK. *Phytophthora: Diseases Worldwide*. St. Paul: APS Press, 1996.
- [5] Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, 2000, 290(5493): 972–977.
- [6] Sogin ML, Silberman JD. Evolution of the protists and protistan parasites from the perspective of molecular systematic. *Int J Parasitol*, 1998, 28(1): 11–20.
- [7] Sansome E, Brasier CM. Diploidy and chromosomal structural hybridity in *Phytophthora infestans*. *Nature*, 1973, 241(5388): 344–345.
- [8] Blobel G, Sabatini DD. Ribosome-membrane interaction in eukaryotic cells. *Biomembranes*, 1971, 2(2): 193–195.
- [9] Martoglio B, Dobberstein B. Signal sequence: more than just greasy peptides. *Trends Cell Biol*, 1998, 8(10): 410–415.
- [10] Greenbaum D, Luscombe NM, Jansen R, Qian J, Gerstein M. Interrelating different types of genomic data from proteome to secretome: 'oming in on function. *Genome Res*, 2001, 11(9): 1463–1468.
- [11] Nimchuk Z, Eulgem T, Holt BF III, Dangl JL. Recognition and response in the plant immune system. *Annu Rev Genet*, 2003, 37(1): 579–609.
- [12] Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identi-

- fication of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*, 1997, 10(1): 1–6.
- [13] Möller S, Croning MDR, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, 2001, 17(7): 646–653.
- [14] Zhou HY, Zhou YQ. Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model-based method. *Protein Sci*, 2003, 12(7): 1547–1555.
- [15] Nakai K. Protein sorting signals and prediction of subcellular localization. *Adv Protein Chem*, 2000, 54(9): 277–344.
- [16] Eisenhaber B, Bork P, Eisenhaber F. Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein Engineering*, 1998, 11(12): 1155–1161.
- [17] Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol*, 1999, 292(3): 741–758.
- [18] Barash S, Wang W, Shi YG. Human secretory signal peptide description by hidden Markov model and generation of a strong artificial signal peptide for secreted protein expression. *Biochem Biophys Res Commun*, 2002, 294(4): 835–842.
- [19] Lee SA, Wormsley S, Kamoun S, Lee AFS, Joiner K, Wong B. An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms. *Yeast*, 2003, 20(7): 595–610.
- [20] Tjalsma H, Bolhuis A, Jongbloed JDH, Bron S, van Dijk JM. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microb Mol Biol Rev*, 2000, 64(3): 515–547.
- [21] 周晓昱, 李成云, 赵之伟, 苏源, 李进斌, 杨静, 刘林, 业艳芬, 张绍松. 粗糙脉孢菌基因组分泌蛋白的初步分析. *遗传*, 2006, 28(2): 200–207.
- [22] 范成明, 李成云, 赵明富, 何月秋. 根癌土壤杆菌 C58 中分泌蛋白信号肽分泌. *微生物学报*, 2005, 45(4): 561–566.
- [23] 吴红芝, 李成云, 朱有勇, 毕玉芬. 秀丽小杆线虫分泌蛋白组的计算机分析. *遗传*, 2006, 28(4): 470–478.
- [24] 吴毅歆, 熊国如, 袁远, 罗灯涛, 何月秋. 水稻基因组中分泌蛋白的初步分析. *分子植物育种*, 2008, 6(5): 1011–1014.
- [25] 于钦亮, 马莉, 刘林, 杨静, 苏源, 王云月, 朱有勇, 李成云. 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. *生物技术通报*, 2008, (1): 160–165.
- [26] Bhattacharjee S, Hiller NL, Liolios K, Win J, Kanneganti TD, Young C, Kamoun S, Haldar K. The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog*, 2006, 2(5): 453–465.
- [27] 苏源, 李成云, 赵之伟, 周晓昱, 李进斌, 杨静, 刘林, 业艳芬. 稻瘟菌基因组规模分泌蛋白的预测分析. *云南农业大学学报*, 2006, 21(3): 271–292.
- [28] 杨福榆. 蛋白质跨膜运送研究的新进展. *生物化学与生物物理进展*, 1991, 18(2): 81–87.
- [29] Yu LM. Elicitins from *Phytophthora* and basic resistance in tobacco. *Proc Natl Acad Sci USA*, 1995, 92(10): 4088–4094.

## • 综合信息 •

### 第十二届全国植物基因组学大会即将在安阳召开

为充分展示植物基因组研究领域的重大进展, 推动我国植物基因组学研究的深入和农业生物技术产业的快速发展, 定于 2011 年 8 月 19 - 21 日(18 日报到, 22 日考察)在河南省安阳市召开第十二届全国植物基因组学大会。会议网址 <http://www.plantgenomics.cn/>

主办单位: 中国遗传学会植物遗传与基因组学专业委员会

承办单位: 中国农业科学院棉花研究所

大会的会议语言是英文, 会议设置 6 个专题, 按专题组织报告。大会报告人有:

Catherine Feuillet; Mitchell Olds; Benfey Phil; J. Antoni Rafalski; Dan Voytas; Dianjing Guo; Honming Lam; 曹晓风; 郭惠珊; 何祖华; 贾继增; 李传友; 凌宏清; 刘宝; 马红; 孟金陵; 王坤波; 王石平; 王秀杰; 吴昌银; 薛红卫; 薛勇彪; 严建兵; 喻树迅; 张大兵; 张启发; 张天真; 张宪省; 张献龙; 朱 祯等。

会议联系人: 王晓辉, 电话: 0372-2562208, E-mail: wangxh@cricaas.com.cn; wywxhwqq@126.com