

全基因组预测稻瘟菌的分泌蛋白

陈继圣, 郑士琴, 郑武, 周洁, 鲁国东, 王宗华

(福建农林大学功能基因组学研究中心 / 福建农林大学生物农药与化学生物学教育部重点实验室, 福州 350002)

摘要: 【目的】分泌蛋白多为病原微生物与植物受体蛋白起作用的激发子和其它致病因子, 深入研究分泌蛋白将有助于明确植物与病原微生物互作的分子机制。利用稻瘟菌基因组学研究成果, 结合计算机技术和生物信息学的方法, 分析其分泌蛋白组学, 将有助于全面掌握其致病因子的结构与功能。【方法】利用 SignalP 对稻瘟菌基因组中所有 ORF 的 N-端信号肽存在与否进行预测, 再依次通过 Protcomp、TMHMM、big-PI Predictor 和 TargetP 预测程序进行验证, 寻找出所有可编码信号肽的基因。【结果】对 11 108 个稻瘟菌的 ORF 进行分析, 最终预测出共有 1 235 个 ORF 可编码分泌蛋白。【结论】经验证此预测方法之可靠性较高, 这为深入研究分泌蛋白组学奠定了基础。

关键词: 稻瘟菌; 分泌蛋白; 信号肽; 预测程序

Prediction for Secreted Proteins from *Magnaporthe grisea* Genome

CHEN Ji-sheng, ZHENG Shi-qin, ZHENG Wu, ZHOU Jie, LU Guo-dong, WANG Zong-hua

(The Ministry of Education Key Laboratory of Bio-Pesticide and Chemistry Biology, Functional Genomics Center,
Fujian Agriculture and Forestry University, Fuzhou 350002)

Abstract: 【Objective】Many secreted proteins of plant pathogens have been shown to be the elicitor and the pathogenetic factors in interacting with plant receptors. Pathogen were analyzed by utilizing genomic database information and computer prediction algorithms. This facilitated clarification of the molecular mechanism in the interaction between plant and plant pathogens.

【Method】To investigate the function of secreted proteins in *Magnaporthe grisea*, a set of predicted algorithms were used to predict the secreted proteins from the *M.grisea* genome. First, the presence or absence of an N-terminal signal peptide for all 11 108 ORFs from *M.grisea* were predicted by the SignalP program. In addition, all the predicted ORFs were tested by the Protcomp, TMHMM, big-PI Predictor, and TargetP programs step by step. 【Result】Finally, 1235 ORFs were predicted to be secreted proteins from *M.grisea* genome. 【Conclusion】The reliability of these prediction algorithms was relatively high. The results reported in this paper give the basis for further studies of the secretome of *M. grisea*.

Key words: *Magnaporthe grisea*; Secreted protein; Signal peptide; Prediction algorithm

0 引言

【本研究的重要意义】稻瘟菌 [*Magnaporthe grisea* (Hebert) Barr.] 在世界范围内危害水稻生产。该菌除为害水稻造成严重减产外, 还为害小麦、大麦、草坪等多种禾本科植物, 造成一定的经济损失^[1]。同时, 稻瘟菌也是研究真菌遗传以及研究植物与病原微

生物相互作用的重要模式生物^[2]。稻瘟菌全基因组测序的完成^[3], 使得从全基因组水平分析和研究该菌某一类型基因功能成为可能。水稻全基因组序列的公布^[4,5], 进一步促进了稻瘟菌与水稻相互作用的分子机制的研究。在植物与病原菌相互作用的过程中, 涉及双方许多信号分子的识别与传导。在病原菌方面, 这些信号分子往往就是与植物受体蛋白起作用的激发子和

收稿日期: 2006-01-15; 接受日期: 2006-09-19

基金项目: 国家自然科学基金 (30471132), 福建自然科学基金重点项目 (B0520002)

作者简介: 陈继圣 (1981-), 男, 福建福州人, 硕士研究生, 研究方向为生化与分子生物学。Tel: 0591-83758395; E-mail: sishenbb@163.com。通讯作者鲁国东 (1967-), 男, 福建上杭人, 教授, E-mail: guodonglu@yahoo.com; 并列通讯作者王宗华 (1962-), 男, 福建尤溪人, 研究员, 研究方向为分子植物病理学。E-mail: zonghuaw@163.com

其它致病因子。大量研究表明, 病菌之激发子与致病相关因子首先要分泌到体外, 才可被植物受体蛋白所感知。已报道马铃薯晚疫病菌 (*Phytophthora infestans*) Inf1^[6]及豌豆锈病病菌 (*Uromyces vignae*) 小种 1 号产生的特异性激发子^[7]为分泌蛋白; 许多无毒基因和致病基因编码产物就是分泌蛋白, 如稻瘟菌无毒基因 Avr-Pita 和 Pwl2 所编码的蛋白^[5,8]等; 众多病原菌细胞壁降解酶类致病因子也多为分泌蛋白, 如稻瘟菌的木聚糖酶 Xyn22 和 Xyn33^[9]和感染拟南芥的白粉病菌 (*Erysiphe cichoracearum*) 果胶裂解酶 Pmr6^[10]等; 还有其它致病相关基因, 如稻瘟菌致病基因 *Gas1* 和 *Gas2*^[11]、疏水性蛋白基因 *Mpg1*^[12]、类似 tetraspanin 蛋白基因 *Pls1*^[13]以及编码几丁质结合蛋白的基因 *Cbp1*^[14]等所编码的产物, 也都是分泌蛋白。将蛋白从合成的场所转运至胞外是通过形成含有 N 端信号肽 (15~30 个疏水氨基酸残基) 的蛋白前体来实现的^[15-17]。在分泌蛋白跨膜及最终转运至膜外过程中, 信号肽起着极为重要的作用。几乎各种分泌蛋白 (如真核细胞的前清蛋白、免疫球蛋白轻链、催乳素等, 原核细胞的脂蛋白、青霉素酶等) 均含有信号肽。因此, 本研究就是基于分泌蛋白具有氨基末端信号肽特征, 而且最有可能是病原菌与宿主受体蛋白起作用的激发子或致病因子的特点, 从全基因组水平预测和分析稻瘟菌的分泌蛋白。【前人研究进展】运用计算机程序分析蛋白质 N 端氨基酸的信号肽序列, 可以预测 1 个蛋白是否为分泌蛋白。迄今为止 SignalP、ProtComp、TargetP、TMHMM 等, 已被成功用于分析蛋白的剪切位点。Lee 等^[18]首先用 SignalP-v2.0 分析了假丝酵母 (*Candida albicans*) 中 47 个已知的分泌蛋白和 47 个已知的非分泌蛋白, 其预测的结果与实际一致。继而他们用 SignalP-v2.0、TargetP 等程序分析了酵母全基因组所有 6 165 个 ORF (阅读框) 的编码产物, 其中 283 个 ORF 编码分泌蛋白。Torto 等^[19]用 SignalP-v2.0 分析了马铃薯晚疫病菌的 2 147 个表达标签序列 (EST), 结果 261 个 EST 预测为分泌蛋白编码基因。进一步分析后, 发现两个是能诱导番茄抗病反应的分泌蛋白 Cnm1 和 Cnm2。【本研究的切入点】稻瘟菌全基因组序列的测定使系统地分析其激发子和致病因子成为可能。【拟解决的关键问题】本研究基于生物信息学方法, 以 SignalP、Protcomp、TargetP、big-PI predictor 和 TMHMM 等程序, 对稻瘟菌的全基因组 11 108 个 ORF 进行预测, 以期获得所有可编码分泌蛋白的相关基因, 为系统分析稻瘟菌的激发子和

致病因子, 揭示水稻与稻瘟菌相互作用的分子机制奠定基础。

1 材料与方法

1.1 材料

稻瘟菌 11 108 个蛋白序列来源于稻瘟菌全基因组数据库 (<http://www.genome.wi.mit.edu/annotation/fungi/magnaporthe>)。用于软件可靠性分析之稻瘟菌蛋白及其登录号: 分泌蛋白 GAS1 (AF363065)、GAS2 (AF264035)、MPG1 (L20685)、AVR-Pita (AF207841)、PWL2 (U26313)、XYN22 (L37529)、XYN33 (L37530)、ACE1 (AJ704622)、PLS1 (AX058235)、CBP1 (AB064264)、NUT1 (U60290) 等; 非分泌蛋白 TRE1 (AF543465)、MGB1 (AB086901)、MVP1 (AY044846)、CDC42 (AF250928)、CHM1 (AY057371)、SPM1 (AB070268)、MST20 (AY332225)、CBS1 (AF422799)、CUT1 (X61500)、PDE1 (AF408395)、PTH2 (AF027979)、PTH11 (AF119672)、CPKA (MGU12335)、ABC1 (AF420471)、CYP1 (AF293848) 等。

1.2 分析方法

分泌蛋白应该具有以下特征: (1) N-端信号肽; (2) 无跨膜结构域; (3) 无 GPI 锚定位点; (4) 没有将蛋白输送至线粒体或其它胞内细胞器的预测定位信号。满足以上 4 个标准的便可定义为分泌蛋白编码基因。其分析过程如下:

以 SignalP-v3.0 分析获得预测蛋白之 C、S 和 Y 的最大值, 以及位于 N 端和被预测的剪切位点间的 S 曲线的中间值, 用以区分信号肽和非信号肽。信号肽剪切位点位于所预测的含信号肽蛋白之 Y 曲线的最大值处。当预测蛋白之 S 值大于 0.6 及 HMM 值大于 0.8, 则预测其为分泌蛋白 (SP⁺); 当预测蛋白之 S 值小于 0.25 及 HMM 值小于 0.1 时, 则为非分泌蛋白 (SP⁻)^[20-22]。

在 SignalP-v3.0 程序进行信号肽预测的基础上, 进一步通过 Protcomp-v6.0、TMHMM-v2.0、big-PI Predictor、TargetP-v1.1 进行验证。

Protcomp-v6.0 用于对动物或真菌中的蛋白进行亚细胞定位, 它可将蛋白按以下归属进行划分: 细胞核、质膜、胞外分泌、细胞质、线粒体、内质网、过氧化物酶体、溶酶体和高尔基体; TMHMM-v2.0 用于预测蛋白的跨膜结构域^[23]; big-PI predictor 程序则用于预测蛋白是否有脂质锚定修饰^[24,25]; 而 TargetP-v1.1

用于进一步确定预测蛋白的亚细胞定位, 可将蛋白的定位归属于线粒体、叶绿体、胞外分泌以及其它亚细胞定位^[26] (表 1)。

表 1 分泌蛋白的预测程序

Table 1 The programs for the prediction of secreted protein

预测程序 Prediction algorithms	预测对象 Objects predicted	精确度 Accuracy (%)	适用范围 Applicable organisms	参考 References
SignalP-v3.0	N 端信号肽 N-terminal signal peptides	97	真核生物 Eukaryotes	Nielsen 1997 http://www.cbs.dtu.dk/services/SignalP-3.0/
ProtComp-v6.0	定位序列 (胞外、脂膜、线粒体、高尔基体) Localization sequences such as extracellular, plasma membrane, mitochondrial and golgi	95	真菌 Fungi	http://www.softberry.com/protein_location/protcomp/predict_for_animal/fungi/
TMHMM-v2.0	跨膜域 Transmembrane domains	97-98	除植物外 Non-plant	Krogh 2001 http://www.cbs.dtu.dk/services/TMHMM/
Big-PI predictor	GPI 锚定位点 GPI-anchor site	>80	真菌 Fungi	Eisenhaber 1999,2001 http://mendel.imp.univie.ac.at/sat/gpi/fungi_server.html
TargetP-v1.1	线粒体或其它定位序列 Mitochondrial or other localization sequence	90	除植物外 Non-plant	Emanuelsson 2000 http://www.cbs.dtu.dk/services/TargetP/

2 结果与分析

2.1 预测软件的可靠性分析

首先以 SignalP-v3.0、Protcomp-v6.0、TargetP-v1.1、TMHMM-v2.0 和 big-PI Predictor 等 5 个分泌蛋白预测软件分别对稻瘟菌中已报道的分泌蛋白和非分泌蛋白进行分析, 以验证这些程序在分析稻瘟菌分泌蛋白中的可靠性。以稻瘟菌类似 tetraspanin 蛋白-PLS1 为例, 用 SignalP-v3.0 程序进行分析, SignalP-NN 预测其 S 中间值为 0.718 (>0.6), SignalP-HMM 预测其 HMM 值为 0.942 (>0.8), 信号肽剪切位点位于第 15~16 个碱基之间 (ATT-CT), 符合预测标准 (图 1)。再经 Protcomp-v6.0 分析, 判定为胞外蛋白 (Extracellular); 继而经 TMHMM-v2.0 预测, 不含跨膜结构; big-PI Predictor 分析表明, PLS1 蛋白无 GPI 锚定修饰; 最后通过 TargetP-v1.1 证实, 含胞外分泌信号, 无线粒体分泌信号。推测的结果表明其为分泌到胞外的蛋白, 与文献报道一致^[13]。

同样方法推测稻瘟菌已报道的几个相关蛋白, 其结果表明, GAS1、GAS2、MPG1、AVR-pita、PWL2、XYN22、XYN33、ACE1、CBP1、NUT1 等均为分泌蛋白; 而 TRE1、MGB1、MVP1、CDC42、CHM1、SPM1、MST20、CBS1、CUT1、PDE1、PTH2、PTH11、CPKA、ABC1、CYP1 等均为非分泌蛋白 (表 2、3), 与文献报道相符。因而, 这些分泌蛋白预测软件可靠性较强, 可用于对稻瘟菌全基因组数据库的分析。

2.2 稻瘟菌全基因组分析分泌蛋白

SignalP-v3.0 分析稻瘟菌 11 108 个 ORF 编码产物, 其结果表明 22.2% 推测蛋白的 S 值大于 0.6, HMM 值大于 0.8。用 3 个 SignalP 预测标准 (针对 SP⁺ 的 S 值、针对 SP⁺ 的 HMM 值及 N 端第 10~40 氨基酸内信号肽的剪切位点), 推测在 11 108 个 ORF 中, 共有 2 468 个编码含 N 端信号肽的蛋白。

进一步以 Protcomp-v6.0 程序分析含信号肽的 2 468 个蛋白结果表明, 共有 1 858 个信号肽分泌至胞外, 337 个转运至细胞质膜, 149 个传输至线粒体, 124 个转运至细胞核 (图 2)。从 TMHMM-v2.0 程序蛋白跨膜结构域分析结果可以看出, 该 1 858 个蛋白中, 有 157 个信号肽含有两个或多个跨膜域, 115 个只含有 1 个跨膜域, 而有 1 586 个信号肽则不含跨膜域; 其中所预测的 115 个只含 1 个跨膜域的信号肽中, 其跨膜域都位于 N 端开始 40 个氨基酸以内。由于 TMHMM 程序可能无法将信号肽序列和跨膜域区分开来, 因此, 所预测 1 586 个不含跨膜域的推测蛋白和 115 个含有单个跨膜域的蛋白 (共 1 701 个) 都可初步推断为分泌蛋白 (图 2)。

继而进行 GPI 锚定蛋白的预测, 以判断这些被初步推断为分泌蛋白是否为胞外蛋白。将 1 701 个分泌蛋白用 big-PI Predictor 程序进行分析, 发现有 206 个为 GPI 锚定蛋白, 而 1 495 个为非 GPI 锚定蛋白 (图 2)。

由于在真核细胞中, 分泌蛋白的分泌目标有可能

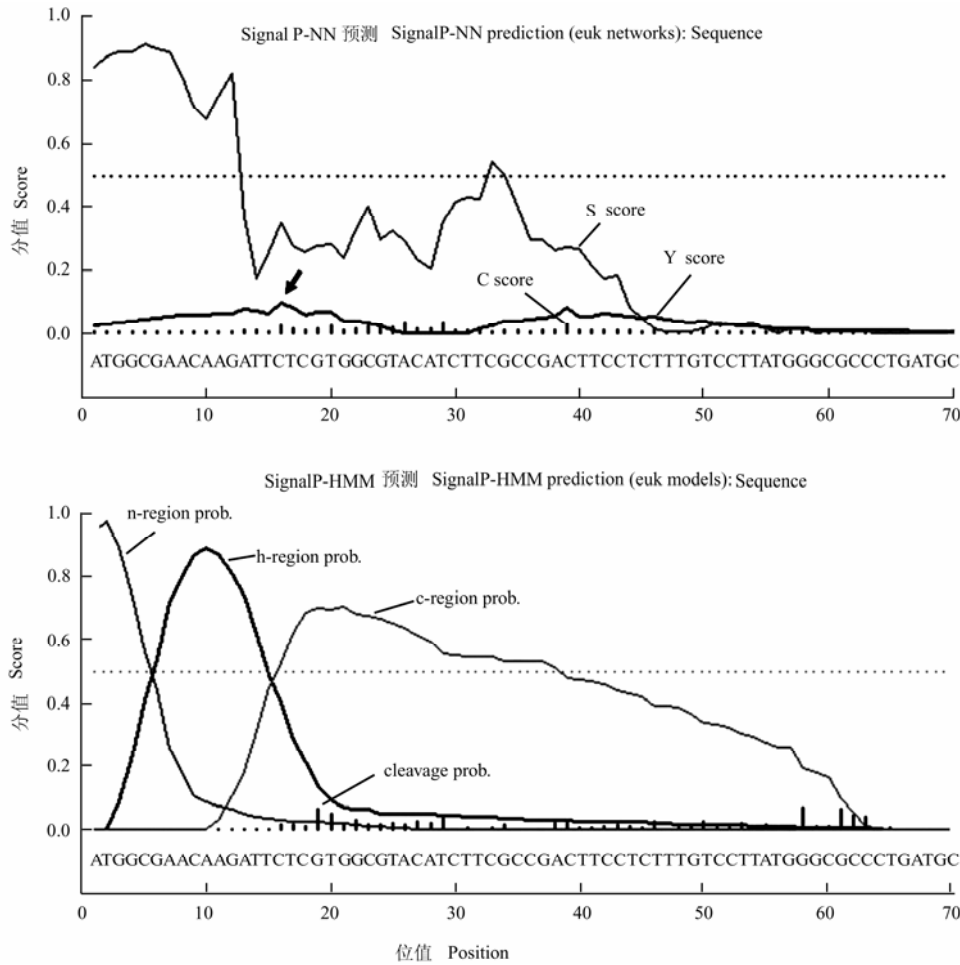


图 1 SignalP-v3.0 对稻瘟菌 PLS1 蛋白的分泌信号 S 值和 HMM 值及其剪切位点的预测 (箭头标示剪切位点)

Fig. 1 Mean S and HMM scores of protein PLS1 in *M.grisea* and its cleavage site predicted by the algorithm SignalP-v3.0 (Arrow shows the cleavage site)

是胞内细胞器,而非胞外。因而进一步以 TargetP-v1.1 程序以排除非胞外分泌蛋白。对 1 495 个非 GPI 锚定的分泌蛋白的分析结果发现,其中有 188 个含有线粒体定位信号,72 个含有其它定位信号,剩下 1 235 个都含有胞外定位信号(图 2)。

总之,11 108 个稻瘟菌 ORF 所编码的产物通过 SignalP、Protcomp、TMHMM、big-PI Predictor 和 TargetP 等程序分析后,只有 1 235 个满足分泌蛋白的 4 个特征。因而,推测稻瘟菌全基因组共有 1 235 个 ORF 编码的产物为分泌蛋白。

3 讨论

在已报道的真菌分泌蛋白中,多数与其致病关系密切。因此,进一步分析分泌蛋白的分泌转运特点及其功能将有助于加深了解植物病原菌的致病机理。对

相关分泌蛋白深入研究是必要的,但全基因组水平分析植物病原菌分泌蛋白的特点,将有助于全面了解其致病因子的整体概况。

单纯地以试验手段去捕获生物中重要蛋白并进一步深入研究是十分有限的。通过生物信息学的手段可以高通量地获悉生物中某一类型的重要蛋白,并在此基础上进一步分选以获得感兴趣的蛋白,因此是一种快速而有效的方法。有关预测蛋白结构和功能的众多程序不仅能在计算机上运行,而且可以通过互联网共享,众多的软件已经被用于最初的对所有预测蛋白的功能分析,并已证实十分有效。最近对整个酵母基因组数据库的分析包括:酿酒酵母中 GPI-锚定蛋白的确定^[43]、对假丝酵母的有性世代基因的综合 BLAST 分析^[44],以及对整个酿酒酵母蛋白库的亚细胞定位预测^[45]。Harcus 等^[46]分析了巴西钩虫 1 234 个 EST 的信

表 2 部分已报道的稻瘟菌分泌蛋白的验证

Table 2 The validation for some reported secreted-proteins in *M.grisea*

蛋白 Protein	蛋白注释与文献来源 Annotation and reference	S中间值 Mean S score	HMM值 HMM score	亚细胞定位 Subcellular localization	跨膜域 Transmembrane domains	GPI 锚定位点 GPI-anchored site	线粒体定位信号 Mitochondrial SS	预测结果 Prediction result
GAS1	糖脂锚定表面蛋白 Glycolipid anchored surface protein ^[111]	0.645	0.902	胞外 Extracellular	0	N	N	SP ⁺
GAS2	类似于GAS1的假定蛋白 Hypothetical protein similar to GAS1 ^[111]	0.770	0.965	胞外 Extracellular	0	N	N	SP ⁺
MPG1	疏水蛋白 Hydrophobin-like protein ^[112]	0.826	0.857	胞外 Extracellular	0	N	N	SP ⁺
AVR-pita	无毒蛋白 Avirulence protein ^[5]	0.691	0.804	胞外 Extracellular	0	N	N	SP ⁺
PWL2	寄主种特异性蛋白 Host-species specificity protein ^[8]	0.778	0.998	胞外 Extracellular	0	N	N	SP ⁺
XYN22	内源 β -1,4-D木聚糖酶 Endo-beta-1,4-D-xylanase ^[9]	0.732	0.831	胞外 Extracellular	0	N	N	SP ⁺
XYN33	内源 β -1,4-D木聚糖酶 Endo-beta-1,4-D-xylanase ^[9]	0.616	0.866	胞外 Extracellular	0	N	N	SP ⁺
ACE1	聚肽或肽合成酶 Polyketide synthase/peptide synthetase ^[27]	0.842	0.903	胞外 Extracellular	0	N	N	SP ⁺
PLS1	水稻致病必须蛋白 Indispensable protein for the fungal pathogenicity on rice ^[113]	0.718	0.942	胞外 Extracellular	0	N	N	SP ⁺
CBP1	几丁质结合蛋白 Chitin binding protein ^[14]	0.647	0.870	胞外 Extracellular	0	N	N	SP ⁺
NUT1	氮调节蛋白 Nitrogen regulatory protein ^[28]	0.901	0.994	胞外 Extracellular	0	N	N	SP ⁺

表 3 部分已报道的稻瘟菌非分泌蛋白的验证

Table 3 The validation for some reported non-secreted proteins in *M.grisea*

蛋白 Protein	蛋白注释与文献来源 Annotation and reference	S中间值 Mean S score	HMM值 HMM score	亚细胞定位 Subcellular localization	跨膜域 Transmembrane domains	GPI 锚定位点 GPI-anchored site	线粒体定位信号 Mitochondrial SS	预测结果 Prediction result
TRE1	海藻糖酶前体 Trehalase precursor ^[29]	0.612	0.861	线粒体 Mitochondrial	0	N	Y	SP ⁻
MGB1	G蛋白 β 亚基 G-protein beta subunit ^[30]	0.631	0.993	线粒体 Mitochondrial	0	N	Y	SP ⁻
MVP1	假定的液泡型ATP酶 Putative vacuolar ATPase ^[31]	0.590	0.981	线粒体 Mitochondrial	0	N	Y	SP ⁻
CDC42	GTIP结合蛋白 GTP-binding protein ^[32]	0.556	0.923	质膜 Plasma membrane	0	N	N	SP ⁻
CHM1	PAK激酶 PAK kinase ^[33]	0.608	0.886	线粒体 Mitochondrial	0	N	Y	SP ⁻
SPM1	液泡型林草菌素类丝氨酸蛋白酶 Vacuolar subtilisin-like serine protease ^[34]	0.737	0.911	质膜 Plasma membrane	0	N	N	SP ⁻
MST20	PAK激酶 PAK kinase ^[33]	0.723	0.916	线粒体 Mitochondrial	0	N	Y	SP ⁻
CBS1	胱硫醚 β -合成酶 Cystathionine beta synthase ^[35]	0.539	0.963	细胞核 Nuclear	0	N	N	SP ⁻
CUT1	角质酶 Cutinase ^[36]	0.798	0.690	质膜 Plasma membrane	0	N	N	SP ⁻
PDE1	P型ATP酶 P-type ATPase ^[37]	0.670	0.988	质膜 Plasma membrane	0	N	N	SP ⁻
PTH2	肉毒碱乙酰转移酶 Carnitine acetyl transferase ^[38]	0.519	0.949	细胞核 Nuclear	0	N	N	SP ⁻
PTH11	跨膜蛋白 Integral membrane protein ^[39]	0.535	0.591	线粒体 Mitochondrial	0	N	Y	SP ⁻
CPKA	依赖cAMP的蛋白激酶 cAMP-dependent protein kinase ^[40]	0.453	0.986	线粒体 Mitochondrial	0	N	Y	SP ⁻
ABC1	ABC转运蛋白 ABC transporter protein ^[41]	0.724	0.821	线粒体 Mitochondrial	0	N	Y	SP ⁻
CYP1	亲环素 Cyclophilin ^[42]	0.689	0.824	线粒体 Mitochondrial	0	N	Y	SP ⁻

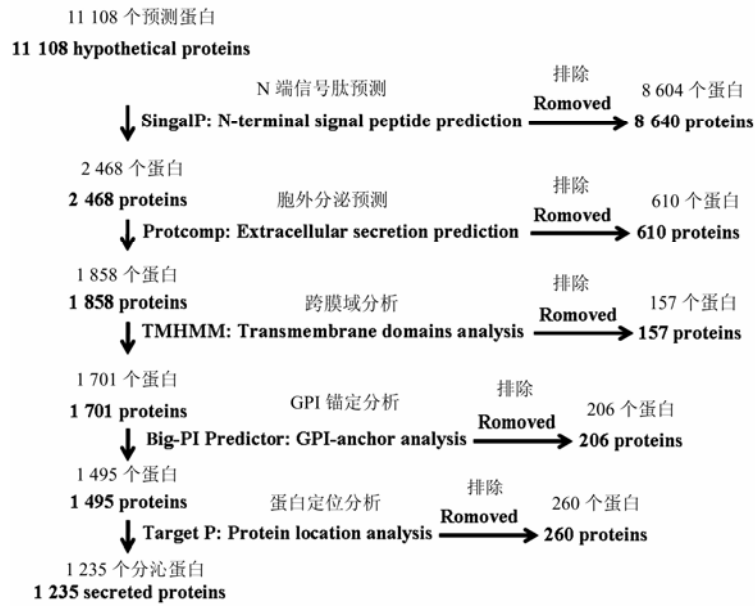


图 2 经预测程序确定稻瘟菌分泌蛋白的流程图

Fig. 2 Flowchart of strategy used to identify *M. grisea* secreted proteins using a series of prediction algorithms

号序列。Wymelenberg 等^[47]对土壤半知菌 *Phanerochaete chrysosporium* 的 268 个分泌蛋白进行了计算机预测。李成云等^[48,49]分别全基因组预测了酿酒酵母和根癌土壤杆菌 C58 Cereon 的分泌蛋白。Manuel 等^[50]用 SignalP、SPScan 和 Tmpred 等软件分析了 *Mycobacterium tuberculosis* 基因组中 3 924 个蛋白, 获得了 52 个分泌蛋白, 其研究结果进一步证实了综合利用计算机手段分析 *M. tuberculosis* 蛋白是高效而又精确的。

本研究采用多种分析软件推测出稻瘟菌基因组中一共存在 1 235 个分泌蛋白, 占全基因组编码产物总数的 11.12%。我们将所预测部分编码分泌蛋白的基因 (110 个), 接上 RGSH₆ tag 后, 置于稻瘟菌中过量表达。用 Ni-NTA 亲和层析纯化表达蛋白, 然后用 RGSH₆ tag 特异性抗体, 通过 Western blot 的方法, 检测目标蛋白的分泌情况。结果发现将近 60% 的候选蛋白可以在胞外得以检测, 这在一定程度上印证了预测的可行性 (结果另行发表)。至于为什么还有 40% 的候选蛋白产物未能在胞外检测到, 可能是由于它们的表达量太低, 或者由于基因转入稻瘟菌后阅读框受到破坏、基因整合位点不恰当、表达产物有毒性等因素造成。因为, 对于这些没有在胞外检测到的蛋白, 在胞内同样无法检测到。

然而, 受到这些预测程序固有的局限性影响, 大

量的 ORF 可能分配不当。SignalP-v3.0 程序将稻瘟菌 ORF 的编码产物划分为分泌蛋白和非分泌蛋白两大类, 但仍有一小部分 ORF 的编码产物属于中间类型。作者在用该程序进行分析时, 将这些中间类型排除在外。因此, 稻瘟菌实际的分泌蛋白有可能超过 1 235 个。

而且, 由于这些预测程序的精确度无法达到 100%, 就存在判断偏差的可能性, 例如, 酵母的 STE12 蛋白为非分泌蛋白^[51], 笔者用这 5 个程序分析其稻瘟菌中的同源蛋白 MST12, 推测结果却为分泌蛋白。再如编码腺苷酸环化酶的基因 MAC1 定位于细胞膜上, 将 ATP 转换成 cAMP^[52], 可推测的结果仍是分泌蛋白, 分泌至胞外行使功能。所以单靠计算机预测的方法进行判定是远远不够的, 还要结合实验加以佐证。

在接下来的工作中, 笔者将对稻瘟菌分泌蛋白开展以下几方面的研究: (1) 对于程序分析出来的分泌蛋白, 需进一步通过实验验证其预测的可靠性; (2) 对分泌蛋白进行功能分析, 特别是与激发子和致病因子有关的分泌蛋白的功能; (3) 分析是否存在大量的不需要信号肽, 而仍行使分泌功能的蛋白, 它们是如何到达胞外空间的; (4) 稻瘟菌中各类蛋白定位到合适的胞内位置所需的特殊目标信号是什么。幸运的是, 对酵母类似方面的研究将为我们对稻瘟菌这些问题的探究提供了极大的便利。

4 结论

已知分泌蛋白多为病原微生物与植物受体蛋白起作用的激发子和其它致病因子。本研究结合计算机技术和生物信息学的方法,推测出稻瘟菌全基因组所编码的所有分泌蛋白,其中包括多个已报道的分泌蛋白,这说明预测方法之可靠性较高。因而,在此基础上深入研究分泌蛋白组学,将有助于全面掌握稻瘟菌的激发子和致病因子的结构与功能,为揭示水稻与稻瘟菌相互作用的分子机制奠定基础。

References

- [1] Zeigler R S, Leong S A, Teng P S. *Rice Blast Disease*. Commonwealth Agricultural Bureaux, Wallingford, UK, 1994: 626.
- [2] Valent B. Rice blast as a model system for plant pathology. *Phytopathology*, 1990, 80: 33-36.
- [3] Dean R A, Talbot N J, Ebbole D J, Farman M L, Mitchell T K, Orbach M J, Thon M, Kulkarni R, Xu J R, Pan H, Read N D, Lee Y H, Carbone I, Brown D, Oh Y Y, Donofrio N, Jeong J S, Soanes D M, Djonovic S, Kolomiets E, Rehmeier C, Li W, Harding M, Kim S, Lebrun M H, Bohnert H, Coughlan S, Butler J, Calvo S, Ma L J, Nicol R, Purcell S, Nusbaum C, Galagan J E, Birren B W. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, 2005, 434(7036): 980-986.
- [4] Goff S A, Ricke D, Lan T H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange B M, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun W L, Chen L, Cooper B, Park S, Wood T C, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller R M, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalima T, Oliphant A, Briggs S. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 2002, 296(5565): 92-100.
- [5] Yu J, Hu S, Wang J, Wong G K, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002, 296(5565): 79-92.
- [6] Kamoun S, van West P, Vleeshouwers V G, de Groot K E, Govers F. Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. *Plant Cell*, 1998, 10: 1413-1426.
- [7] D'Silva I, Heath M C. Purification and characterization of two novel hypersensitive response-inducing specific elicitors produced by the cowpea rust fungus. *Journal of Biological Chemistry*, 1997, 272: 3924-3927.
- [8] Sweigard J A, Carroll A M, Kang S, Farrall L, Chumley F G, Valent B. Identification, cloning, and characterization of PWL2, a gene for host species specificity in the rice blast fungus. *Plant Cell*, 1995, 7: 1221-1233.
- [9] Wu S C, Kauffmann S, Darvill A G, Albersheim P. Purification, cloning and characterization of two xylanases from *Magnaporthe grisea*, the rice blast fungus. *Molecular Plant-Microbe Interactions*, 1995, 8: 506-514.
- [10] Vogel J P, Raab T K, Schiff C, Somerville S C. PMR6, a pectate lyase-like gene required for powdery mildew susceptibility in *Arabidopsis*. *Plant Cell*, 2002, 14: 2095-2106.
- [11] Xue C, Park G, Choi W, Zheng L, Dean R A, Xu J R. Two novel fungal virulence genes specifically expressed in appressoria of the rice blast fungus. *Plant Cell*, 2002, 14: 2107-2119.
- [12] Talbot N J, Ebbole D J, Hamer J E. Identification and characterization of MPG1, a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. *Plant Cell*, 1993, 5: 1575-1590.
- [13] Clergeot P H, Gourgues M, Cots J, Laurans F, Latorse M P, Pepin R, Tharreau D, Nottoghem J L, Lebrun M H. PLS1, a gene encoding a tetraspanin-like protein, is required for penetration of rice leaf by the fungal pathogen *Magnaporthe grisea*. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98: 6963-6968.
- [14] Kamakura T, Yamaguchi S, Saitoh K, Teraoka T, Yamaguchi I. A novel gene, CBP1, encoding a putative extracellular chitin-binding protein, may play an important role in the hydrophobic surface sensing of *Magnaporthe grisea* during appressorium differentiation. *Molecular Plant-Microbe Interactions*, 2002, 15: 437-444.
- [15] Von Heijne G. Protein targeting signals. *Current Opinion in Cell*

- Biology*, 1990, 2: 604-608.
- [16] Von Heijne G. The signal peptide. *Journal Membrane Biology*, 1990, 115: 195-201.
- [17] Von Heijne G. Life and death of a signal peptide. *Nature*, 1998, 396: 111-113.
- [18] Lee S A, Wormsley S, Kamoun S, Lee A F, Joiner K, Wong B. An analysis of the *Candida albicans* genome database for soluble secreted proteins using computer-based prediction algorithms. *Yeast*, 2003, 20: 595-610.
- [19] Torto T A, Li S, Styer A, Huitema E, Testa A, Gow N A, van West P, Kamoun S. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Research*, 2003, 13: 1675-1685.
- [20] Nielsen H, Engelbrecht J, Brunak S, Von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*, 1997, 10: 1-6.
- [21] Nielsen H, Krogh A. Prediction of signal peptides and signal anchors by a hidden Markov model. In *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB 6)*: AAAI Press, 1998: 122-130.
- [22] Nielsen H, Brunak S, Von Heijne G. Machine learning approaches to the prediction of signal peptides and other protein sorting signals. *Protein Engineering*, 1999, 12: 3-9.
- [23] Krogh A, Larsson B, Von Heijne G, Eriik L, Sonnhammer L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, 2001, 305: 567-580.
- [24] Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in protein sequences. *Journal of Molecular Biology*, 1999, 292: 741-758.
- [25] Eisenhaber B, Bork P, Eisenhaber F. Post-translational GPI lipid anchor modification of proteins in kingdoms of life: analysis of protein sequence data from complete genomes. *Protein Engineering*, 2001, 14: 17-25.
- [26] Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, 2000, 300: 1005-1016.
- [27] Bohnert H U, Fudal I, Dioh W, Tharreau D, Nottoghem J L, Lebrun M H. A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. *Plant Cell*, 2004, 16: 2499-2513.
- [28] Froeliger E H, Carpenter B E. NUT1, a major nitrogen regulatory gene in *Magnaporthe grisea*, is dispensable for pathogenicity. *Molecular & General Genetics*, 1996, 251: 647-656.
- [29] Foster A J, Jenkinson J M, Talbot N J. Trehalose synthesis and metabolism are required at different stages of plant infection by *Magnaporthe grisea*. *The EMBO Journal*, 2003, 22: 225-235.
- [30] Nishimura M, Park G, Xu J R. The G-beta subunit MGB1 is involved in regulating multiple steps of infection-related morphogenesis in *Magnaporthe grisea*. *Molecular Microbiology*, 2003, 50(1):231-243.
- [31] Asiegbu F O, Choi W, Jeong J S, Dean R A. Cloning, sequencing and functional analysis of *Magnaporthe grisea* MVP1 gene, a hex-1 homolog encoding a putative 'woronin body' protein. *FEMS Microbiology Letters*, 2004, 230(1): 85-90.
- [32] 郑武, 陈继圣, 郑士琴, 鲁国东, 王宗华. 稻瘟菌 Cdc42 若干推测互作蛋白的结构和表达特点. *中国农业科学*, 2006, 39: 709-714. Zheng W, Chen J S, Zheng S Q, Lu G D, Wang Z H. Structure and expression pattern of several putative Cdc42-interacting proteins in *Magnaporthe grisea*. *Scientia Agricultura Sinica*, 2006, 39(4): 709-714. (in Chinese)
- [33] Li L, Xue C, Bruno K, Nishimura M, Xu J R. Two PAK kinase genes, CHM1 and MST20, have distinct functions in *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions*, 2004, 17: 547-556.
- [34] Fukiya S, Kuge T, Tanishima T, Sone T, Kamakura T, Yamaguchi I, Tomita F. Identification of a putative vacuolar serine protease gene in the rice blast fungus, *Magnaporthe grisea*. *Bioscience, Biotechnology, and Biochemistry*, 2002, 66: 663-666.
- [35] Lo S C, Hamer L, Hamer J E. Molecular characterization of a cystathionine beta-synthase gene, CBS1, in *Magnaporthe grisea*. *Eukaryotic Cell*, 2002, 1: 311-314.
- [36] Sweigard J A, Chumley F G, Valent B. Cloning and analysis of CUT1, a cutinase gene from *Magnaporthe grisea*. *Molecular & General Genetics*, 1992, 232(2): 174-182.
- [37] Balhadere P V, Talbot N J. PDE1 encodes a P-type ATPase involved in appressorium-mediated plant infection by the rice blast fungus *Magnaporthe grisea*. *Plant Cell*, 2001, 13: 1987-2004.
- [38] Sweigard J A, Carroll A M, Farrall L, Chumley F G, Valent B. *Magnaporthe grisea* pathogenicity genes obtained through insertional mutagenesis. *Molecular Plant-Microbe Interactions*, 1998, 11: 404-412.
- [39] Dezwaan T M, Carroll A M, Valent B, Sweigard J A. *Magnaporthe grisea* pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. *Plant Cell*, 1999, 11: 2013-2030.
- [40] Mitchell T K, Dean R A. The cAMP-dependent protein kinase catalytic subunit is required for appressorium formation and

- pathogenesis by the rice blast pathogen *Magnaporthe grisea*. *Plant Cell*, 1995, 7: 1869-1878.
- [41] Avila-Adame C, Koller W. Insertional mutagenesis of *Magnaporthe grisea* toward decreased responsiveness of alternative respiration to inhibition by azoxystrobin. *Journal of Genetic Plant Pathology*, 2003, 69: 126-130.
- [42] Viaud M C, Balhadere P V, Talbot N J. A *Magnaporthe grisea* cyclophilin acts as a virulence determinant during plant infection. *Plant Cell*, 2002, 14: 917-930.
- [43] Sundstrom P. Adhesion in *Candida* spp. *Cell Microbiology*, 2002, 4: 461-469.
- [44] Tzung K W, Williams R M, Scherer S, Federspiel N, Jones T, Hansen N, Bivolarevic V, Huizar L, Komp C, Surzycki R, Tamse R, Davis R W, Agabian N. Genomic evidence for a complete sexual cycle in *Candida albicans*. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98: 3249-3253.
- [45] Kumar A, Agarwal S, Heyman J A, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, Cheung K H, Miller P, Gerstein M, Roeder G S, Snyder M. Subcellular localization of the yeast proteome. *Genes & Development*, 2002, 16: 707-719.
- [46] H Marcus Y M, Parkinson J, Fernandez C, Daub J, Selkirk M E, Blaxter M L, Maizels R M. Signal sequence analysis of expressed sequence tags from the nematode *Nippostrongylus brasiliensis* and the evolution of secreted proteins in parasites. *Genome Biology*, 2004, 5(6): 39.
- [47] Wymelenberg A V, Sabat G, Martinez D, Rajangam A S, Teeri T T, Gaskell J, Kersten P J, Cullen D. The *Phanerochaete chrysosporium* secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. *Journal of Biotechnology*, 2005, 118(1): 17-34.
- [48] 范成明, 李成云, 赵明富, 何月秋. 根癌土壤杆菌 C58 *Cereon* 中分泌蛋白信号肽分析. *微生物学报*, 2005, 45: 561-566.
- Fan C M, Li C Y, Zhao M F, He Y Q. Analysis of signal peptides of the secreted proteins in *Agrobacterium tumefaciens* C58. *Acta Microbiologica Sinica*, 2005, 45: 561-566. (in Chinese)
- [49] 杨 静, 李成云, 王云月, 朱有勇, 李进斌, 何霞红, 刘 林, 业艳芬, 周晓罡, 唐 萍. 酿酒酵母分泌蛋白组的计算机分析. *中国农业科学*, 2005, 38: 516-522.
- Yang J, Li C Y, Wang Y Y, Zhu Y Y, Li J B, He X H, Liu L, Ye Y F, Zhou X G, Tang P. Computational analysis of signal peptide-dependent secreted proteins in *Saccaromyces cerevisiae*. *Scientia Agricultura Sinica*, 2005, 38: 516-522. (in Chinese)
- [50] Gomez M, Johnson S, Gennaro M L. Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infection and Immunity*, 2000, 68: 2323-2327.
- [51] Calcagno A M, Bignell E, Warn P, Jones M D, Denning D W, Muhlschlegel F A, Rogers T R, Haynes K. *Candida glabrata* STE12 is required for wild-type levels of virulence and nitrogen starvation induced filamentation. *Molecular Microbiology*, 2003, 50: 1309-1318.
- [52] Choi W, Dean R A. The adenylate cyclase gene MAC1 of *Magnaporthe grisea* controls appressorium formation and other aspects of growth and development. *Plant Cell*, 1997, 9: 1973-1983.

(责任编辑 王红艳)