

南方和北方根结线虫分泌蛋白组预测与比较分析

李午佼, 颜瑾, 王运生*

(湖南农业大学 生物安全科学技术学院, 湖南 长沙 410128)

摘要: 通过 SignalP 3.0、TargetP 1.1、KohGPI、TMHMM 2.0 等系列软件对南方和北方根结线虫分泌蛋白组进行预测。结果表明, 南方和北方根结线虫预测分泌蛋白数为 2 622 和 1 696 个, 分别占总基因数的 13.65% 和 12.95%。经 BlastP($E < 1e-10$) 分析, 分别有 754 个和 466 个能在 KOG(eukaryotic orthologous groups) 库中找到相似蛋白, 其功能主要包括能量代谢与转运、信号传导、转录活性调控等。基因本体(GO) 分析结果表明, 南方和北方根结线虫的预测分泌蛋白数目在核定位信号和转录活性调控上存在显著差异。用 MEME/MAST 搜索并分析根结线虫分泌信号肽保守的模体(Motif), 得到 54 个南、北方根结线虫分泌信号肽共有的模体。

关键词: 根结线虫; 分泌蛋白; 模体

中图分类号: S432.4⁺5 文献标志码: A 文章编号: 1007-1032(2011)04-0376-05

Prediction and comparative analysis of secretory proteins of *Meloidogyne incognita* and *Meloidogyne hapla*

LI Wu-jiao, YAN Jin, WANG Yun-sheng*

(College of Bio-Safety Science and Technology, Hunan Agricultural University, Changsha 410128, China)

Abstract: Secretory proteins play an important role in the process of molecular interaction between pathogen and host plants. In this paper, we employed SignalP 3.0, TargetP 1.1, KohGPI and TMHMM 2.0 to predict the secretory proteins of root-knot nematode and finally we got 2 622 and 1 696 secretory proteins for *Meloidogyne incognita* and *Meloidogyne hapla* respectively, correspondingly accounted for 13.65% and 12.95% of the total proteins. 754 and 466 secretory proteins from *Meloidogyne incognita* and *Meloidogyne hapla* respectively showed homology to proteins in KOG (eukaryotic orthologous groups) database with blastp ($E < 1e-10$). The main functions of those homologous proteins in KOG database were related to energy metabolism, transport, signal transduction and regulation of transcriptional activity. GO(gene ontology) analysis showed that *M. incognita* and *M. hapla* differed significantly in the number of secretory proteins for nuclear localization signal and transcription regulator activity. Finally, 54 motifs were found to be shared by the signal peptides of *M. incognita* and *M. hapla* using MEME / MAST.

Key words: root-knot nematode; secretory protein; motif

根结线虫(*Meloidogyne* spp.)引起的根结线虫病是一种世界性分布的土传病害, 侵害植物根部形成根结, 给农林、园艺等产业造成巨大损失^[1-2]。根结线虫的分泌物和寄主防御基因的表达产物在信号识别中起着重要的作用^[3], 这种分子互作决定着寄主植物的抗、感性, 对抗线虫育种和防治根

结线虫病具有重要意义。2002 年, Jaubert 等^[4]用神经刺激剂间苯二酚处理根结线虫寄生前 2 龄幼虫, 模拟寄主植物存在的环境, 诱导线虫表达寄生相关蛋白, 通过双向电泳分离差异蛋白, 获得包括钙网蛋白(Mi-CRT)在内的部分上调表达基因。2008 年, Bellafiore 等^[5]采用类似的方法鉴定

收稿日期: 2010-12-28

基金项目: 国家自然科学基金项目(30900925); 湖南农业大学稳定人才基金项目(09WD22)

作者简介: 李午佼(1989—), 男, 湖南新邵人, 硕士研究生, 主要从事生物信息学研究; *通信作者, wyunsheng@gmail.com

出了南方根结线虫 486 个分泌蛋白,其中包括与寄主植物、动物寄生线虫同源的基因,表明根结线虫可能主动参与调节植物的代谢与分化,推测动植物寄生线虫中有着相似的致病分子机制。

分泌蛋白质组是组织、细胞等分泌的全部蛋白质的总称。分泌蛋白实验鉴定方法主要涉及分泌蛋白的制备、多维色谱或二维凝胶电泳分离、质谱鉴定等^[6],这些方法耗时费力。分泌蛋白在 N 端往往具有信号肽结构,一般由 10~40 个氨基酸残基组成。基于信号肽学说,当内质网结合核糖体仍处于游离状态时,即已开始翻译多肽链,并在分泌蛋白 N 端形成信号肽,这个信号肽被识别后将游离核糖体引导至内质网膜。信号肽通常含 3 个结构域,即 N 端结构域、H 结构域和 C 端结构域。N 端结构域是富含丙氨酸的片断,疏水中心区包含有 6~15 个氨基酸残基,对肽段的穿膜和正确定位具有重要作用^[7]。基于信号肽保守的结构域特征,可以通过生物信息学方法预测信号肽的有无来推测是否为分泌蛋白,这类软件以 SignalP 为代表^[8]。由于跨膜蛋白和靶向细胞器如线粒体、叶绿体的蛋白有时也含有信号肽,一般在预测分泌蛋白时应将该部分蛋白去除,可用 TargetP 进行亚细胞定位,预测靶向细胞器的蛋白^[9],用 TMHMM 预测是否存在跨膜结构域(膜蛋白)^[10],用 KohGPI 预测是否为 GPI (Glycosylphosphatidylinositol) 蛋白^[11]。

利用生物信息学软件已对多个物种进行了分泌蛋白质组的预测,如秀丽小杆线虫^[12]、酿酒酵母^[13]、稻瘟菌^[14]、里氏木霉^[15]等。南、北方根结线虫全基因组序列^[16-17]的公布,为从基因组水平上挖掘和比较其分泌蛋白质组提供了基础。研究寄生线虫的分泌蛋白和寄主植物受体蛋白两者之间的关系,是解析线虫的寄主选择性及植物应答线虫入侵后一系列复杂信号传导通路的关键^[18]。笔者在基因组水平上对南、北方根结线虫分泌蛋白组进行预测和比较分析,研究根结线虫分泌蛋白的特征,以期为进一步研究根结线虫与寄主植物间的分子互作提供基础。

1 材料和方法

1.1 序列下载

南、北方根结线虫基因组序列从 wormbase ftp 下载;南方根结线虫蛋白序列从 [http://www.inra.](http://www.inra.fr/meloidogyne_incognita/content/download/3010/29690/version/2/file/MincV1A1.fas)

[fr/meloidogyne_incognita/content/download/3010/29690/version/2/file/MincV1A1.fas](http://www.inra.fr/meloidogyne_incognita/content/download/3010/29690/version/2/file/MincV1A1.fas) 下载;北方根结线虫蛋白序列根据基因组注释结果文件提取。

1.2 方法

1.2.1 分泌蛋白组预测

首先用 SignalP 3.0 预测南、北方根结线虫蛋白的信号肽,选用真核生物模型,其他参数默认,含有信号肽的蛋白用 TargetP 1.1 预测其亚细胞定位,参数选择 N(非植物),过滤掉细胞器蛋白后,用 TMHMM 2.0 预测跨膜蛋白和用 KohGPI 预测 GPI 蛋白,将预测为跨膜蛋白和 GPI 蛋白过滤后得到候选分泌蛋白。

1.2.2 基因本体(GO)分析

预测得到的分泌蛋白用 Blast2GO^[19-20]进行 GO 注释,用 WEGO^[21](Web gene ontology annotation plot. <http://wego.genomics.org.cn>)^[21]进行 GO 比较分析。

1.2.3 分泌蛋白组的功能注释

将预测的分泌蛋白与 KOG 库^[22]进行 BlastP^[23]($E < 1e-10$)比对,选取匹配最好的目标序列作为查询序列的功能注释,统计各功能类别蛋白数。

1.2.4 信号肽模体预测

用 MEME/MAST^[24-25]软件包对根结线虫的信号肽模体(motif)进行分析。用 MEME/MAST 从预测为分泌蛋白序列中寻找潜在的模体,返回前 100 个模体。模体对应的分泌蛋白序列用 BlastP($E < 1e-10$)搜索 KOG 库,预测模体的功能并依此推测分泌蛋白可能的功能和分泌途径。

2 结果

2.1 南、北方根结线虫分泌蛋白组预测结果

北方根结线虫蛋白组序列有 13 072 个,经 SignalP 3.0 预测,含有信号肽的蛋白序列有 2 632 个,经 TargetP 1.1 预测为非细胞器蛋白的为 2 460 个,经 TMHMM 2.0 预测,1 793 个蛋白序列无跨膜区,经 KohGPI 预测,97 个蛋白序列含 GPI 锚定位点,最后得到 1 696 个分泌蛋白。类似地,南方根结线虫蛋白组序列有 19 212 个,经 Signal 3.0 和 TargetP 预测含信号肽的非细胞器蛋白有 3 835 个,2 752 个蛋白序列经 TMHMM2.0 验证无跨膜区,经

KohGPI 预测, 有 130 个含 GPI 锚定位点, 最后得到 2 622 个分泌蛋白。南、北方根结线虫分泌蛋白占总蛋白比例相近, 分别为 13.65%和 12.97%。

2.2 信号肽长度与氨基酸组成分布

信号肽长度统计结果表明, 南、北方根结线虫信号肽长度分布相似, 集中分布于[15~30]区间, 分别占 84.82%和 80.08%(图 1-A)。两个根结线虫信号肽的氨基酸组成分布相似, 但氨基酸分布不均匀(图 1-B), 氨基酸出现频率较高的有 L(亮

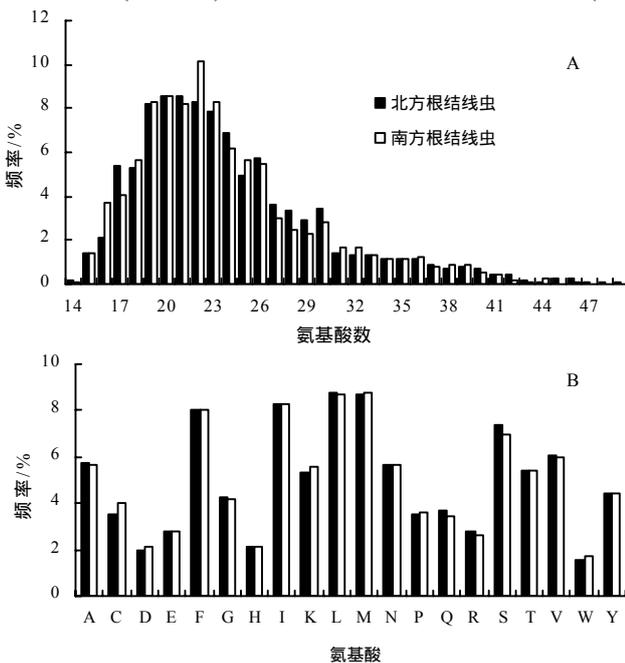


图 1 信号肽长度(A)及氨基酸组成(B)的频率

Fig. 1 Frequency distribution of length(A) and amino acids of signal peptides(B)

氨酸)、M(甲硫氨酸)、I(异亮氨酸)、F(苯丙氨酸)、S(丝氨酸)和 V(缬氨酸)。

2.3 分泌蛋白组 GO 分类

用 Blast2GO 对南、北方根结线虫的分泌蛋白进行 GO 分类, 再用 WEGO 进行比较分析。结果表明, 根结线虫分泌蛋白主要包括生物调控、色素沉着、发育过程、多细胞生物体过程、催化活动、绑定和细胞组成等方面。比较南、北方根结线虫的分泌蛋白, 两者在含氮化合物代谢(GO: 0006807)、细胞定位(GO: 0051641)、定位调控(GO: 0032879)、蛋白复合体定位(GO: 0031503)、核酸结合(GO: 0003676)、转录活性调控(GO: 0030528)、膜(GO: 0016020)等差异显著($P < 0.05$)。

2.4 分泌蛋白组功能注释结果

1 670 个根结线虫分泌蛋白序列在 KOG 中能找到对应的功能分类号(BlastP $E < 1e-10$), 南、北方根结线虫分别为 754 个和 446 个, 主要包括蛋白质翻译后修饰、信号转导、碳水化合物运输和代谢等(表 1)。

表 1 根结线虫分泌蛋白 KOG 分类

Table 1 KOG classification of secretory proteins of root-knot nematode

KOG 类	蛋白数/个		功能描述
	南方根结线虫	北方根结线虫	
A	20	10	RNA 加工和修饰
B	2	0	染色体结构与动力学
C	27	14	能量代谢与转换
D	15	8	细胞周期调控、细胞分裂和染色体分离
E	24	19	氨基酸运输和代谢
F	6	4	核苷酸运输和代谢
G	81	38	碳水化合物运输和代谢
H	7	4	辅酶运输和代谢
I	41	29	脂质运输和代谢
J	6	6	翻译、核糖体结构与合成
K	32	5	转录
L	4	5	复制、重组和修复
M	2	5	细胞壁/膜/荚膜合成
O	144	84	翻译后修饰、蛋白质折叠、分子伴侣
P	23	19	无机离子转运和代谢

续表

KOG 类	蛋白数/个		功能描述
	南方根结线虫	北方根结线虫	
Q	23	11	次生代谢物的生物合成、转运和分解
R	94	59	预测功能
S	56	35	功能未知
T	117	88	信号传导机制
U	26	20	胞内分泌和膜泡运输
V	22	8	防御机制
W	70	28	胞外结构
Y	2	2	核结构
Z	11	5	细胞骨架

2.5 信号肽模体预测结果
 使用 MEME/MAST 预测根结线虫分泌蛋白信号肽前 100 个模体，南、北方根结线虫中共有模体 54 个，但分布频率存在显著差异，如 IL[LI] YL[IV]I[LF]L[NK]I[QK]C[NA]G 模体在南、北方根结线虫中分别有 42 和 2 个分泌蛋白信号肽含有该模体，且大部分功能预测为转录因子(表 2)。

表 2 南、北方根结线虫分泌信号肽共有的前 10 个模体及其分布

Table 2 The distribution of the top 10 motifs of secretory proteins shared by both *M. incognita* and *M. hapla*

序号	模体	蛋白数/个		KOG 类别
		南方根结线虫	北方根结线虫	
1	IL[LI]YL[IV]I[LF]L[NK]I[QK]C[NA]G	42	2	K
2	MKL[IA]S[VFI]LI[FV]LIFN[SW]IL	38	3	/
3	M[SL]K[KF]F[NI][AV][ILF]VI[LF][TF][FI][VI][LF][KI][NEH][YC][CS][EVD]G	24	1	R
4	MWLFLF[IF][FV]LN[SIT]F[YF][NK]	15	1	D
5	TG[PI][LWT][ATI]R[KR][SR]ST[LI][CF][SK]LL[LI]L[LI][TP][AV]	13	8	O
6	MYLKLIF[FY][FL]FLLIFINIS[IV]G	7	4	G
7	M[RK][IF][SN][SQT]L[FL][LV]L[LA][VAI][IV][GAL]L[LAQ][FGV][LP][VS][ANQ][LQ][VG]	7	4	I
8	RGKAKM[LF]LLPFNL[VF][CW]LA[LW]ILA	6	3	/
9	[VAS][AVT][IL][MV]NNE[FL]T[HNQ]A	6	2	S
10	[QN]ETK[VA][VI][IL]GIAC[LF]SS[LF]LA[VI][LMF][AS]TL[VI]V[ILM]PQLY	5	4	W

3 讨论

Bellafore 等^[5]直接分离南方根结线虫食道腺分泌蛋白并进行质谱鉴定，共鉴定出 486 个分泌蛋白，根据 *M. incognita* 的 EST 和 mRNA 序列预测得到 289 个分泌蛋白序列，经 BlastP 分析($E < 1e-10$)，其中包含在南方根结线虫蛋白组的有 263 个序列，占 91%，而包含在本研究预测的分泌组中只有 66 个序列，还有约 77% 的分泌蛋白按本研究方法并未预测出来，可能的原因：①南方根结线虫中可能存在非传统的分泌途径；②基因组注释结果中基因结构的 5'端不完整；根结线虫分泌信号肽可能存在特异性。这意味着需通过实验方法直接鉴定根结线虫的分泌蛋白，才有可能发现新的、特异的分泌蛋白。

根结线虫分泌蛋白中存在大量与碳、脂代谢、能量合成与转换的相关蛋白，这些分泌蛋白可能与寄生线虫从寄主植物中吸取营养相关。有意思的是，在预测的根结线虫分泌蛋白组中含有大量具转录活性和核酸绑定结构域的蛋白，这表明，根结线虫分泌的转录因子可能参与寄主植物基因的转录调控，参与寄主根部巨形细胞的形成。Huang 等^[26]报道，蛋白 7H08 含有 1 个典型的核定位信号 SV40，另外在 6 个分泌蛋白中存在卷曲螺旋(coiled-coil domain)，3 个蛋白中发现有内部重复(internal repeats)，推测与转录调控有关。Bellafore 等^[5]也预测出 66 个分泌蛋白含有核定位信号、DNA 绑定、染色体修饰结构域。笔者在比较南、北方根结线虫分泌蛋白 GO 分布时发现，两者在转录活性调节(蛋白数目比 21 : 5, $P=0.047$)和

核酸绑定(蛋白数目比 36 : 11, $P=0.036$)上存在显著差异,可能与南方线虫具有更大的寄主范围相关,其中 Minc18551、Minc06060、Minc10010 同时含有转录活性调节和核酸绑定结构域, BlastP 结果表明,预测 Minc18551、Minc06060 为锌指蛋白,而 Minc10010 为转录起始因子。

分泌信号肽模体分析结果表明,根结线虫的信号肽中存在一定保守的模体,且含有同一模体的成员大部分预测功能类似,但这些保守的模体是否与真实功能相关,有待进一步研究。

参考文献:

- [1] Hesamedin R. Effect of different initial population densities of *Meloidogyne incognita* on the growth of *Cicer arietinum*[J]. Commun Agric Appl Biol Sci, 2006, 71(3): 669–674.
- [2] Jain A, Mohan J, Singh M, et al. Potentiality of different isolates of wilt fungus *Fusarium oxysporum* collected from rhizosphere of tomato against root-knot nematode *Meloidogyne incognita*[J]. J Environ Sci Health B, 2008, 43(8): 686–691.
- [3] 茆振川, 谢丙炎, 杨之为, 等. 根结线虫与植物的分子互作[J]. 园艺学报, 2006, 33(4): 901–907.
- [4] Jaubert S, Ledger T N, Laffaire J B, et al. Direct identification of stylet secreted proteins from root-knot nematodes by a proteomic approach[J]. Mol Biochem Parasitol, 2002, 121(2): 205–211.
- [5] Bellafiore S, Shen Z, Rosso M N, et al. Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential[J]. PLoS Pathog, 2008, 4(10): e1000192.
- [6] 张海婧, 刘辉, 吕丽艳, 等. 分泌蛋白质组研究进展[J]. 分析化学评述与进展, 2007, 35(6): 912–918.
- [7] Martoglio B, Dobberstein B. Signal sequences: More than just greasy peptides[J]. Trends Cell Biol, 1998, 8(10): 410–415.
- [8] Nielsen H, Engelbrecht J, Brunak S, et al. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites [J]. Protein Eng, 1997, 10(1): 1–6.
- [9] Emanuelsson O, Nielsen H, Brunak S, et al. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence[J]. J Mol Biol, 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]. J Mol Biol, 2001, 305(3): 567–580.
- [11] Fankhauser N, Maser P. Identification of GPI anchor attachment signals by a Kohonen self-organizing map[J]. Bioinformatics, 2005, 21(9): 1846–1852.
- [12] 吴红芝, 李成云, 朱有勇, 等. 秀丽小杆线虫分泌蛋白组的计算机分析[J]. 遗传, 2006, 28(4): 470–478.
- [13] 杨静, 李成云. 酿酒酵母分泌蛋白组的计算机分析[J]. 中国农业科学, 2005, 38(3): 516–522.
- [14] 陈继圣, 郑士琴. 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.
- [15] 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, 48(4): 473–479.
- [16] Abad P, Gouzy J, Aury J M, et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*[J]. Nat Biotechnol, 2008, 26(8): 909–915.
- [17] Opperman C H, Bird D M, Williamson V M, et al. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism[J]. Proc Natl Acad Sci USA, 2008, 105(39): 14802–14807.
- [18] 牛俊海, 卜祥霞, 薛慧, 等. 植物根结线虫基因组学研究进展[J]. 植物病理学报, 2010, 40(3): 225–234.
- [19] Ashburner M, Ball C A, Blake J A, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium[J]. Nat Genet, 2000, 25(1): 25–29.
- [20] Conesa A, Gotz S, Garcia-Gomez J M, et al. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research[J]. Bioinformatics, 2005, 21(18): 3674–3676.
- [21] Ye J, Fang L, Zheng H, et al. WEGO: A web tool for plotting GO annotations[J]. Nucleic Acids Res, 2006, 34: 293–297.
- [22] Tatusov R L, Fedorova N D, Jackson J D, et al. The COG database: An updated version includes eukaryotes [J]. BMC Bioinformatics, 2003, 4: 41.
- [23] Altschul S F, Madden T L, Schaffer A A, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs[J]. Nucleic Acids Res, 1997, 25(17): 3389–3402.
- [24] Bailey T L, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers[J]. Proc Int Conf Intell Syst Mol Biol, 1994(2): 28–36.
- [25] Bailey T L, Gribskov M. Combining evidence using p-values: Application to sequence homology searches [J]. Bioinformatics, 1998, 14(1): 48–54.
- [26] Huang G, Dong R, Allen R, et al. A root-knot nematode secretory peptide functions as a ligand for a plant transcription factor[J]. Mol Plant Microbe Interact, 2006, 19(5): 463–470.

责任编辑: 罗慧敏