

粘绿木霉 Gv29-8 的碳水化合物活性酶类蛋白 预测及遗传关系分析

韩长志 许 僊

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 利用 CAT(CAZymes analysis toolkit)在线分析程序对 *Trichoderma virens* Gv29-8 中碳水化合物活性酶类蛋白(carbohydrate-active enzymes, CAZymes)不同亚家族的分布情况进行分析,结果表明:粘绿木霉 Gv29-8 中含有 185 个 CAZymes 蛋白,分为主要类别和复合类别两种类型;其中,主要类别含有 71 个 GH 蛋白、54 个 CBM 蛋白及 12 个 AA 蛋白、25 个 CE 蛋白、3 个 PL 蛋白、3 个 GT 蛋白;复合类型含有 17 个蛋白。遗传关系分析表明,CAZymes 分类应进一步细化到更小的类别。

关键词 碳水化合物活性酶类蛋白;粘绿木霉;遗传关系;预测程序

中图分类号 S 436.341 **文献标识码** A **文章编号** 1000-2421(2017)05-0039-06

碳水化合物活性酶类(carbohydrate-active enzymes, CAZymes)在植物病原菌侵染植物过程中,特别是在突破寄主植物细胞壁过程中发挥着关键作用^[1]。作为描述碳水化合物绑定模块(功能域)结构相关催化酶的专业数据库^[2],CAZymes 数据库中包含了诸多涉及水解、修饰、转移、断裂糖苷键等基因功能的 CAZymes 蛋白,极大地促进了国内外学者对植物病原菌细胞壁降解酶活性基因的研究^[3-4]。特别是最近几年,学术界关于 CAZymes 的研究呈现出迅速发展的趋势^[2,5-6],按照最新研究成果,特别是 CAZymes 数据库的分类方式,主要有以下六大类:GHs (glycoside hydrolases,糖苷水解酶)^[7]、GTs (glycosyl transferases,糖基转移酶)^[8]、PLs (polysaccharide lyases,多糖裂解酶)^[9]、CEs(carbohydrate esterases,碳水化合物酯酶)^[9]、AAs(auxiliary activities,辅助酶类家族)^[10]以及 CBMs(carbohydrate-binding modules,碳水化合物绑定结构)^[11]。

近年来国内外学者的研究重点由防治植物病原菌的化学药剂开发转向生防菌的研究利用^[12-13],防治植物病原菌的生防菌主要包含两大类:真菌中的木霉属(*Trichoderma* spp.)和细菌中的芽孢杆菌属(*Bacillus* spp.)等^[14]。就木霉属真菌而言,已经报道的种类超过了 140 种,如粘绿木霉(*T. virens*)、绿色木霉(*T. viride*)、哈茨木霉(*T. harzianum*)、里氏木霉(*T. reesei*)等^[15]可产生抗生素代谢产物。然而,尚未见到对植物病原菌具有重要生防作用的 *T. virens* Gv29-8 中 CAZymes 蛋白的预测报道,开展该菌 CAZymes 蛋白的预测及遗传关系分析具有重要的意义。

基于其全基因组序列^[16],笔者前期通过生物信息学分析软件 SignalP、ProtComp、TMHMM、big-PI Fungal predictor 等对真菌中所含有的分泌蛋白进行筛选,明确其存在 377 个分泌蛋白^[17],本研究利用 CAT(CAZymes analysis toolkit)^[4]在线分析程序对 *T. virens* Gv29-8 中 CAZymes 不同亚家族的分布情况及其遗传关系进行分析,以期为深入研究该菌 CAZymes 的生物学功能提供重要参考。

1 材料与方法

1.1 粘绿木霉分泌蛋白序列来源

377 个分泌蛋白来自笔者前期研究^[17]。

1.2 CAZymes 预测

利用 CAT 在线分析程序对 *T. virens* Gv29-8 中碳水化合物活性酶类蛋白(carbohydrate-active enzymes, CAZymes)不同亚家族的分布情况进行预

收稿日期: 2017-05-01
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)
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

测分析。

1.3 CAZymes 遗传关系分析

利用 Clustal X 对所获得的 CAZymes 蛋白序列进行多重比对分析,利用 MEGA 7.0 软件^[18]采用邻近法构建系统发育树,各分支之间的距离计算采用 p-distance 模型,系统可信度检测采用自举法重复 1 000 次进行^[19]。

1.4 CAZymes 遗传关系分析相互作用预测

利用 STRING v.10^[20] 在线数据库对上述所获得的 CAZymes 蛋白序列开展相互作用预测,明确蛋白-蛋白相互作用关系。

2 结果与分析

2.1 粘绿木霉 Gv29-8 中 CAZymes 分布

通过 CAT 预测到粘绿木霉 Gv29-8 中含有 185 个 CAZymes,进行统计分析发现,这些蛋白分为主要类别和复合类别两类;前者涉及 GH、CBM、CE、AA、GT 以及 PL,尤以 GH 数量较多,达 71 个,占比为 38.38%;其次为 CBM,有 54 个,占比为 29.19%;复合类型蛋白有 17 个,包括 GH/CBM、CE/CBM、AA/CBM、GH/CBM/CBM(图 1)。

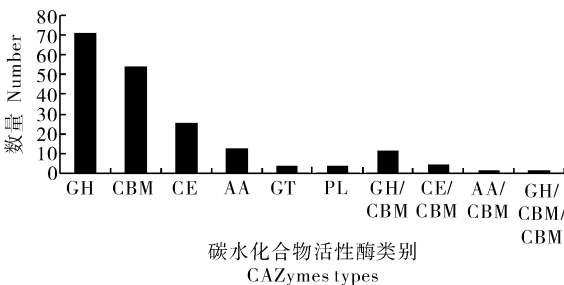


图 1 粘绿木霉 Gv29-8 中不同碳水化合物活性酶分布情况
Fig.1 The distribution of CAZymes from *T. virens* Gv29-8

2.2 粘绿木霉 Gv29-8 中的 GH 分布

通过对 GH 蛋白进行分析,发现 GH18 数量较多,有 7 个;其次 GH3、GH55 较多,有 6 个;GH27 有 5 个;GH28 有 4 个;这类蛋白所占比例为 30.99%(表 1)。此外,还涉及 GH1、GH2、GH5 等类别。

2.3 粘绿木霉 Gv29-8 中 CBM 分布

通过对 CBM 蛋白进行分析,发现尤以 CBM50 数量最多,为 15 个;其次是 CBM1,数量为 12 个;再次为 CBM13,数量为 7 个;这类蛋白所占比例为 62.96%(表 2)。此外,还涉及 CBM5、CBM6、CBM12

等类别。值得关注的是jgi|TriviGv29_8_2|28185具有 CBM6/CBM32 两种不同类别,推测其在发挥 CBM 功能方面具有特殊性。

表 1 粘绿木霉 Gv29-8 中糖基水解酶分布情况
Table 1 The distribution of GH from *T. virens* Gv29-8

类型 Type	ID/jgi TriviGv29_8_2	数量 Number
GH1	193599	1
GH2	48076、181642、28953	3
GH3	29366、222727、41347、151663、 90223、30124	6
GH5	35701	1
GH7	182161、90504	2
GH10	53264	1
GH11	8282	1
GH12	28149、224618、42536	3
GH16	212924、183603	2
GH17	16072	1
GH18	178019、213202、52337、89999、 201336、85551、42107	7
GH20	49762、180069	2
GH24	215230	1
GH25	45518	1
GH27	41607、46494、39936、33738、27922	5
GH28	14392、33554、48906、51095	4
GH30	32109、53160、51554	3
GH32	35658	1
GH35	42143	1
GH36	46577	1
GH39	51209	1
GH43	87809	1
GH55	45064、175914、46121、 46692、32683、37929	6
GH62	48966、152027	2
GH64	50439	1
GH65	51014	1
GH71	27940	1
GH75	87418、174923	2
GH78	192768	1
GH79	48997	1
GH81	11067	1
GH92	192950、49694	2
GH95	52642	1
GH103	87809	1
GH127	209784	1
GH128	230914	1
合计 Total		71

表 2 粘绿木霉 Gv29-8 中碳水化合物绑定结构蛋白分布情况

Table 2 The distribution of CBM from *T. vires* Gv29-8

类型 Type	ID/jgi TriviGv29_8_2	数量 Number
CBM1	200371,46583,49838,42107,32357,51211,8282,85551,35701,152027,90504,182161	12
CBM5	217176,47273,90003	3
CBM6	28185	1
CBM12	89230,231073	2
CBM13	33902,41430,88460,143434,46830,40892,209784	7
CBM18	27891,201336,51784,49798	4
CBM32	173869,87714,28185	3
CBM35	46494,39936,33738,27922	4
CBM50	40262,131286,52825,156047,200487,124493,1900,128337,51902,222643,193212,220308,86763、 84937,201336	15
CBM6/CBM32	28185	1
CBM63	82045,29842	2
合计 Total		54

2.4 粘绿木霉 Gv29-8 中 AA 分布

通过对辅助酶家族进行分析,发现 AA7 较多,有 6 个;其次是 AA1,有 3 个;这类蛋白所占比例为 75%(表 3)。此外,还涉及 AA2、AA3、AA9 等类型,其数量均为 1 个。

表 3 粘绿木霉 Gv29-8 中辅酶家族分布情况

Table 3 The distribution of AA from *T. vires* Gv29-8

不同类型 Type	ID/jgi TriviGv29_8_2	数量 Number
AA1	91146,29729,48916	3
AA2	140278	1
AA3	46709	1
AA7	154489,29663,36910,182568、 227182,170440	6
AA9	27891	1
合计 Total		12

2.5 粘绿木霉 Gv29-8 中 CE 分布

通过对 CE 蛋白进行分析,发现 CE10 数量较多,有 12 个;其次是 CE5、CE1,分别有 4、3 个;这类蛋白所占比例为 76%(表 4)。另外,还涉及 CE3、CE4、CE15 等类别。

表 4 粘绿木霉 Gv29-8 中碳水化合物酯酶蛋白分布情况

Table 4 The distribution of CE from *T. vires* Gv29-8

类型 Type	ID/jgi TriviGv29_8_2	数量 Number
CE1	41882,51690,39449	3
CE3	223478	1
CE4	51784,49798	2
CE5	39549,47222,129988,51211	4
CE8	49025,31131	2
CE15	32357	1
CE10	193306,147072,35347,52507,40764,84609、 37227,150720,52168,227371,151632,28159	12
合计 Total		25

2.6 粘绿木霉 Gv29-8 中 PL 及 GT 分布

通过对 PL 蛋白进行分析,发现 PL7 有 2 个,其 ID 分别为 jgi | TriviGv29_8_2 | 52707、jgi | TriviGv29_8_2|41602;其次 PL20 有 1 个,其 ID 为 jgi|TriviGv29_8_2|215593。同时,通过对 GT 蛋白进行分析,发现涉及 GT2、GT4、GT20 等类别,数量均为 1 个,其 ID 分别为 jgi | TriviGv29_8_2 | 208383、jgi| TriviGv29_8_2| 228094、jgi| TriviGv29_8_2|10277。

2.7 粘绿木霉 Gv29-8 中的 CAZymes 复合类型

通过对 CAZymes 复合类型蛋白进行分析,发现 CBM35/GH27 较多,有 4 个;其次为 CE4/CBM18、GH18/CBM1、GH7/CBM1,均有 2 个(表 5);这类蛋白所占比例为 58.82%。

表 5 粘绿木霉 Gv29-8 中复合类型分布情况

Table 5 The distribution of composite category from *T. vires* Gv29-8

类型 Type	ID/jgi TriviGv29_8_2	数量 Number
AA9/CBM18	27891	1
CBM35/GH27	46494,39936,33738,27922	4
CBM50/CBM18/GH18	201336	1
CE15CBM1	32357	1
CE4/CBM18	51784,49798	2
CE5/CBM1	51211	1
GH11/CBM1	8282	1
GH18/CBM1	85551,42107	2
GH5/CBM1	35701	1
GH62/CBM1	152027	1
GH7/CBM1	90504,182161	2
合计 Total		17

2.8 粘绿木霉中 CAZymes 遗传关系分析

对 CAZymes 进行遗传关系分析,可知粘绿木

霉 Gv29-8 中存在 GH/CBM、CE/GH、CE/CBM、CE/AA/GH/CBM/GT 等类型,而在前期所预测的复合类型中仅涉及 GH/CBM、CE/CBM 两类及 CE/GH、CE/AA/GH/CBM/GT 两类。本研究涉及 CE、GH、AA、CBM 以及 GT 等类型(图 2),CE10 存在上述几个主要类群中(图 2 红色方框),表明 CE 的分类有待进一步研究。CAZymes 分类系统的理论基础是氨基酸序列的相似性,而氨基酸序列同源

性并不意味着蛋白质具有相同的功能^[21],因此,CAZymes 分类系统有待于进一步进行小类划分,更好地开展后续研究^[2]。

2.9 粘绿木霉中 CAZymes 蛋白相互关系分析

通过 STRING v.10 在线数据分析,发现大部分 CAZymes 蛋白结构特征尚未知(图 3 中小球),因此,未能预测出其相互关系。另外,EHK19480、16950、17389、22789、16564、20518、229813、26379、

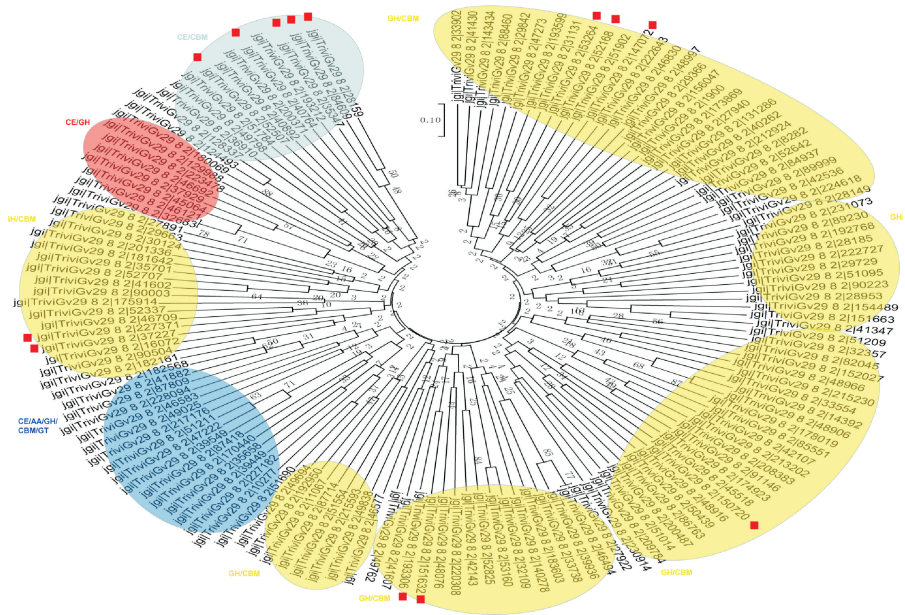


图 2 粘绿木霉 Gv29-8 中 CAZymes 蛋白遗传关系分析

Fig.2 The genetic relationship of CAZymes from *T. virens* Gv29-8

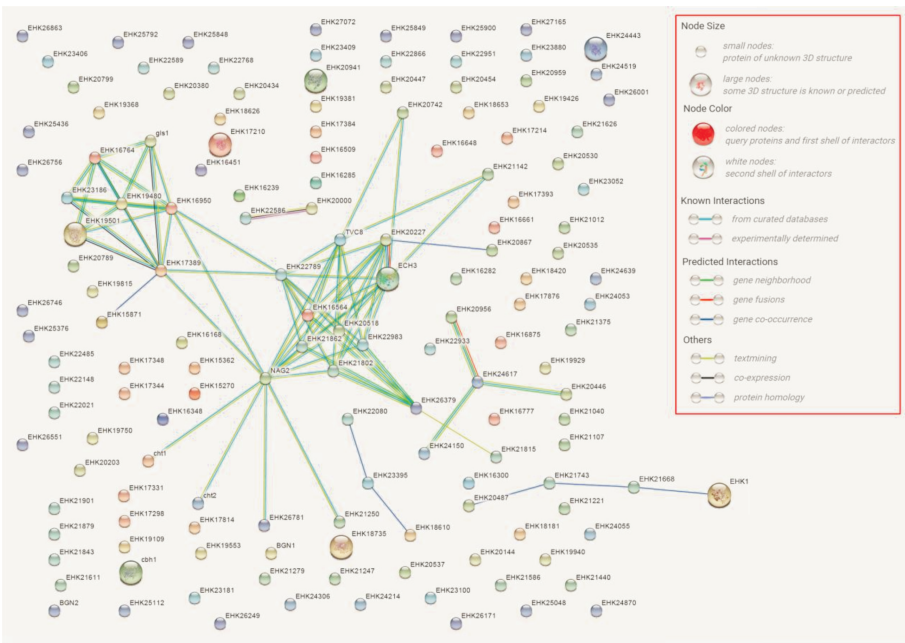


图 3 粘绿木霉 Gv29-8 中 CAZymes 蛋白相互关系分析

Fig.3 Protein-protein interaction (PPI) analysis of CAZymes from *T. virens* Gv29-8

24617、21862、21802 以及 NAG2 等蛋白作为重要的节点蛋白(图 3),在与其他已知关系蛋白中,发挥着重要的作用。

3 讨论

木霉属真菌可以侵染植物病原菌的重要作用机制之一为重寄生,在此过程中,木霉属真菌往往会分泌一系列诸如纤维素酶、葡聚糖酶、几丁质酶等代谢产物,在侵染病原菌过程中发挥重要生防作用;同时,其抗生性代谢产物性质的稳定程度也成为开展植物病害生防菌筛选的重要指标之一。本研究通过生物信息学相关技术,分析了生防真菌粘绿木霉 *T. virens* Gv29-8 中的分泌蛋白,预测该生防菌中含有 185 个 CAZymes 蛋白,这些蛋白是否包含了粘绿木霉 29-8 中的全部蛋白类型,有待于进一步开展试验进行验证。

随着大量植物病原真菌、卵菌基因组测序工作的完成,稻瘟病菌^[22]、大丽轮枝菌^[23]、希金斯炭疽菌^[24]、禾谷炭疽菌^[25]、致病疫霉^[26]等分泌蛋白已经得到确认;对于 *T. reesei* 拮抗菌中分泌蛋白的研究^[27]也有报道,通过比较基因组学分析病原菌的寄生性与基因组变异的关系、病原菌与拮抗菌之间关系已经成为研究热点^[28]。基于不同营养类型的植物病原真菌 CAZymes 的分析,发现属于非活体营养型植物病原菌(不仅涉及死体营养型植物病原菌,也包括半活体营养型的植物病原菌)的基因组较那些属于活体营养型植物病原菌具有更多的 CAZymes 蛋白^[29-31],对于植物病原菌生防菌是否存在对于植物病原菌侵染过程中的活体营养型、非活体营养型等阶段,有待进一步开展试验验证。此外, Martinez 等^[32]报道里氏木霉 QM6a 中 CAZymes 蛋白种类为 GH、CBM、CE、PL、GT,数量分别为 200、36、16、3、103 个;与此不同,本研究所获得粘绿木霉 29-8 中数量分别为 71、54、25、3、3,远低于同属的里氏木霉 QM6a,造成差异的原因有待进一步研究。

参考文献

- [1] VAN DEN BRINK J, DE VRIES R P. Fungal enzyme sets for plant polysaccharide degradation[J]. Appl Microbiol Biotechnol, 2011, 91(6): 1477-1492.
- [2] 王帅, 陈冠军, 张怀强, 等. 碳水化合物活性酶数据库(CAZy)及其研究趋势[J]. 生物加工过程, 2014, 12(1): 102-108.
- [3] CANTAREL B L, COUTINHO P M, RANCUREL C, et al. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics[J]. Nucleic Acids Res, 2009, 37: D233-238.
- [4] PARK B H, KARPINETS T V, SYED M H, et al. CAZymes analysis toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database[J]. Glycobiology, 2010, 20(12): 1574-1584.
- [5] LOMBARD V, GOLACONDA- RAMULU H, DRULA E, et al. The carbohydrate-active enzymes database (CAZy) in 2013[J]. Nucleic Acids Res, 2014, 42: D490-495. doi:10.1093/nar/gkt1178.
- [6] 聂燕芳, 周淦, 黄嘉瑶, 等. 尖孢镰刀菌甜瓜专化型基因组规模分泌蛋白的预测与分析[J]. 华中农业大学学报, 2016, 35(3): 24-29.
- [7] BOURNE Y, HENRISSAT B. Glycoside hydrolases and glycosyltransferases: families and functional modules[J]. Current opinion in structural biology, 2001, 11(5): 593-600.
- [8] CAMPBELL J A, DAVIES G J, BULONE V, et al. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities[J]. Biochem J, 1997, 326: 929-939.
- [9] LOMBARD V, BERNARD T, RANCUREL C, et al. A hierarchical classification of polysaccharide lyases for glycogenomics[J]. Biochem J, 2010, 432(3): 437-444.
- [10] LEVASSEUR A, DRULA E, LOMBARD V, et al. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes[J]. Biotechnol Biofuels, 2013, 6(1): 41.
- [11] BORASTON A B, BOLAM D N, GILBERT H J, et al. Carbohydrate-binding modules: fine-tuning polysaccharide recognition[J]. Biochem J, 2004, 382: 769-781.
- [12] 韩长志, 霍超. 核桃炭疽病生物防治菌株 YB-15 的筛选和鉴定[J]. 经济林研究, 2016, 34(1): 83-89.
- [13] 韩长志, 霍超. 核桃炭疽病生防菌 yb33 的鉴定及其次生代谢物特性分析[J]. 经济林研究, 2015, 33(3): 63-67, 74.
- [14] 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.
- [15] 孙瑞艳. 中国南方地区木霉菌资源收集、鉴定与生防功能评价研究[D]. 上海: 上海交通大学, 2013.
- [16] 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): 40.
- [17] 韩长志. 基于全基因组测序的粘绿木霉 Gv29-8 中分泌蛋白预测[J]. 河南师范大学学报(自然科学版), 2016, 44(6): 145-148.
- [18] TAMURA K, STECHER G, PETERSON D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0[J]. Molecular biology and evolution, 2013, 30(12): 2725-2729.
- [19] 韩长志. 禾谷炭疽菌 RGS 蛋白生物信息学分析[J]. 微生物学

- 通报, 2014, 41(8):1582-1594.
- [20] SZKLARCZYK D, FRANCESCHINI A, WYDER S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life[J]. Nucleic acids research, 2015, 43:D447-452. <https://doi.org/10.1093/nar/gku1003>.
- [21] LEE D, REDFERN O, ORENGO C. Predicting protein function from sequence and structure[J]. Nature reviews molecular cell biology, 2007, 8(12):995-1005.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] SCHMIDT S M, PANSTRUGA R. Pathogenomics of fungal plant parasites: what have we learnt about pathogenesis? [J]. Curr Opin Plant Biol, 2011, 14(4):392-399.
- [29] 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.
- [30] ZHAO Z, LIU H, WANG C, et al. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi[J]. BMC genomics, 2013, 14(1):274.
- [31] 赵中涛. 真菌生长发育和致病相关基因家族的比较和进化基因组学研究[D]. 杨凌:西北农林科技大学, 2014.
- [32] 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.

Prediction and genetic relationship of CAZymes proteins from *Trichoderma virens* Gv29-8 genome

HAN Changzhi XU Xi

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

Abstract Carbohydrate active enzymes protein (CAZymes), one large kind of important proteins of plant pathogenic fungi and bacteria, plays an important role in its growth and development. *Trichoderma virens* Gv29-8 is the important bio-control agent of *Trichoderma* ssp., inhibiting many plant pathogens. Predicting and analyzing the characteristics of the CAZymes is of great significance. Based on 377 secreted proteins obtained preliminary data, CAZymes protein was predicted with CAT forecasting procedures. 185 CAZymes proteins were found in *T. virens* Gv29-8 and divided into major categories and composite category. The former includes 71 GHs, 54 CBMs, 12 AAs, 25 CEs, 3 PLs and 3 GTs; the latter includes 17 CAZymes proteins. Genetic analyses showed that CAZymes classification should be further refined to a smaller category. It will lay a solid theoretical foundation for further studying its function.

Keywords CAZymes; *Trichoderma virens*; genetic relationship; prediction algorithm

(责任编辑:张志钰)