

# 禾谷炭疽菌碳水化合物酶类蛋白注释及其理化性质分析

韩长志, 许僖

(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:** 禾谷炭疽菌侵染玉米、小麦等农作物引起的炭疽病, 给各国农业生产造成了巨大的经济损失. 基于前期研究结果, 以 267 个碳水化合物酶类蛋白(CAZymes)序列为基础, 对其进行理化性质分析, 结果表明, 近 2/3 的蛋白理论等电点小于 6, 属于酸性蛋白, 同时, 通过随机数软件对上述蛋白进行 10% 抽样分析, 明确了不同理论等电点类别的分泌蛋白所具有的分子质量、分子式、原子数量、半衰期、不稳定性系数、脂肪族氨基酸指数、总平均亲水性等性质. 该研究为深入解析禾谷炭疽菌 CAZymes 蛋白的功能奠定了坚实的理论基础.

**关键词:** 禾谷炭疽菌; 分泌蛋白; 理化性质; 预测程序

**中图分类号:** S435.1

**文献标志码:** A

禾谷炭疽菌(*Colletotrichum graminicola* (Cesati) Wilson)作为炭疽菌属重要的植物病原菌, 可以侵染玉米、小麦、高粱等禾本科作物而引起炭疽病, 给各国农业生产造成巨大的经济损失<sup>[1-5]</sup>. 自 20 世纪 70 年代以来, 由该菌引起的玉米炭疽病在美国、印度等国就非常普遍<sup>[1,6,7]</sup>. 随着该菌全基因组序列的释放<sup>[8]</sup>, 目前, 关于该菌 MAPK 途径蛋白预测<sup>[9]</sup>、分泌蛋白预测<sup>[10]</sup>以及 14-3-3 蛋白<sup>[11]</sup>和 RGS<sup>[12]</sup>生物信息学分析等已见报道. 前期本研究小组利用 SignalP, ProtComp, TMHMM, big-PI Fungal Predictor 和 TargetP 等预测程序对 *C. graminicola* 蛋白数据库中的 12006 条序列进行分析预测, 明确该菌中含有 630 个分泌蛋白<sup>[10]</sup>. 后续通过 CAT 分析, 明确上述蛋白中含有 267 个碳水化合物酶类蛋白(Carbohydrate-Active Enzymes, CAZymes)<sup>[13]</sup>, 然而, 对于上述蛋白的性质和功能一直缺少较为深入的后续研究报道.

CAZymes 作为分泌蛋白中的一大类, 是病原菌侵染过程中突破寄主细胞第一道屏障——植物细胞壁的关键因素. 对其开展深入研究, 有助于解析植物病原菌侵入、操控植物的作用机制. 目前学术界关于禾谷炭疽菌中 CAZymes 的数量有相关报道<sup>[14]</sup>, 但尚未见有关该菌中 CAZymes 理化性质分析报道.

本研究基于前期所获得的 267 个 CAZymes 蛋白序列, 通过生物信息学分析方法解析该菌上述蛋白的基本理化性质, 以为深入解析该病原菌 CAZymes 蛋白在侵染操控寄主机制方面提供重要的理论基础.

## 1 材料与方法

### 1.1 材料

前期所获得的 267 个 CAZymes 蛋白序列<sup>[13]</sup>.

### 1.2 后续蛋白选择

利用随机数形成软件, 对所获得的 CAZymes 蛋白中不同类别蛋白进行 10% 的随机抽样选择, 用于后续蛋白分析.

### 1.3 CAZymes 理化性质预测

利用蛋白质数据库(<http://www.expasy.ch/tools/protparam.html>)<sup>[15]</sup>在线进行分析预测分泌蛋白的

收稿日期: 2016-12-12; 修回日期: 2017-03-10.

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资助(51600625).

作者简介(通信作者): 韩长志(1981—), 男, 河北石家庄人, 西南林业大学副教授, 博士, 研究方向为经济林木病害生物防治与真菌分子生物学, E-mail: hanchangzhi2010@163.com.

等电点、分子质量及氨基酸组成、脂肪族氨基酸指数等特征.

## 2 结果与分析

### 2.1 2/3 以上的 CAZymes 蛋白等电点小于 6.0

通过对 267 个 CAZymes 蛋白所具有的理论等电点数值进行分析,发现最低值为 4.03,最高值为 9.45,以 0.50 为一个间隔单位,对上述分泌蛋白进行归类分析,结果表明,位于 4.51~5.00 之间的分泌蛋白数量最多,为 78 个,所占比例为 29.21%(图 1),对上述理论等电点小于 6.0 的 CAZymes 蛋白进行统计,数量为 175 个,所占比例为 65.54%(图 2),理论等电点高于 8.0 的 CAZymes 蛋白,数量为 39 个,所占比例为 14.61%,而位于 6.0~8.0 的 CAZymes 蛋白的数量为 53 个,所占比例为 19.85%(图 2).

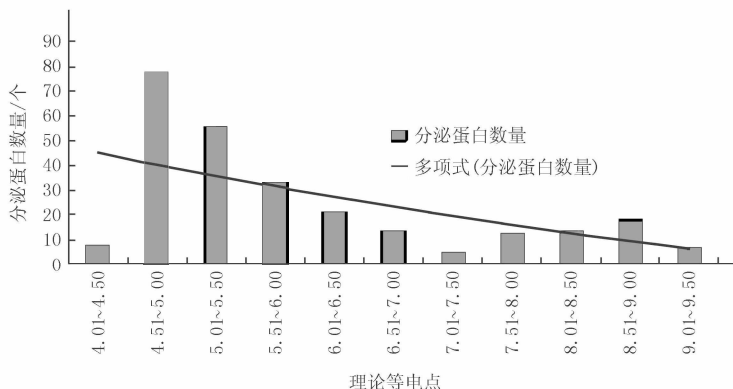


图1 禾谷炭疽菌中267个CAZymes蛋白理论等电点的分布情况

### 2.2 CAZymes 蛋白理化性质分析

为了更好地分析禾谷炭疽菌中 CAZymes 蛋白所具有的基本理化性质,本研究按照理论等电点所处的酸碱度范围不同,利用随机数形成软件,对上述蛋白进行 10% 的随机抽样选择,获得用于后续研究的蛋白基本信息(表 1).

对上述蛋白序列进行功能预测以及基本理化性质等生物信息学分析,结果显示,在 18 个酸性蛋白中,仅有 2 个为假定蛋白(hypothetical protein),所占比例为 11.11%,其余 16 个蛋白均有所预测的功能,分别为多铜氧化酶(multicopper oxidase)、GPI 锚定细胞壁内切葡聚糖酶 EglC(GPI-anchored cell wall beta-1,3-endoglucanase EglC)、kelch 结构域包含蛋白(kelch domain-containing protein)、糖基水解酶家族 20 蛋白(glycosyl hydrolase family 20)、糖苷水解酶家族 45 蛋白(glycoside hydrolase family 45)等;在 5 个中性蛋白所预测的功能则涉及假定蛋白、WSC 结构域包含蛋白(WSC domain-containing protein)、多糖脱乙酰酶(polysaccharide deacetylase)、角质酶(cutinase)以及糖基水解酶家族 12 蛋白(glycosyl hydrolase family 12);在 4 个碱性蛋白中所预测的功能涉及 PHB 解聚酶家族酯酶(PHB depolymerase family esterase)、糖基水解酶家族 28、53 蛋白(glycosyl hydrolase family 28/53)、果胶酸裂解酶(pectate lyase)(表 1). 上述 CAZymes 无论是酸性蛋白还是中性蛋白、碱性蛋白,均分属于 AA,CE,PL,GH,GT,CBM 等家族中.

同时,对上述 CAZymes 蛋白的相对分子质量、分子式、原子数量、半衰期、不稳定性系数、脂肪族氨基酸指数、总平均亲水性等性质进行分析,结果表明,相对分子质量较大的蛋白,其原子数量也多;而就所预测的不稳定系数而言,上述 27 个蛋白中除 G1 号为 GLRG\_00636 的蛋白为不稳定性蛋白外,其他均为稳定性蛋白;就上述蛋白的疏水性而言,除 GLRG\_11936,GLRG\_02120,GLRG\_08340 等 3 个蛋白总平均亲水性大于

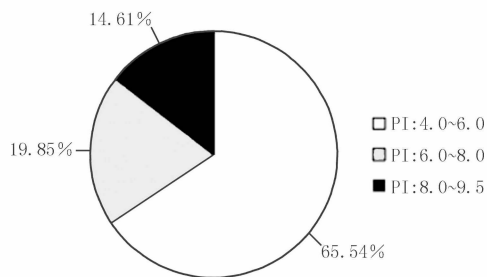


图2 禾谷炭疽菌中所具有的不同理论等电点 CAZymes 蛋白的分布情况

0,属于疏水性蛋白外,其他均属于亲水性蛋白(表 2).

表 1 禾谷炭疽菌中不同酸碱度 CAZymes 蛋白的基本信息

分泌蛋白 类型	数量	序号	蛋白 GI/ pI 值	理论等电点 pI 值	预测功能	CAZymes 分类
酸性蛋白	18	1	GLRG_04766	4.42	multicopper oxidase	AA1
		2	GLRG_01144	4.58	hypothetical protein	GII93
		3	GLRG_10490	4.65	hypothetical protein	GII32
		4	GLRG_09366	4.66	GPI-anchored cell wall beta-1,3-endoglucanase EglC	GII17
		5	GLRG_02805	4.78	kelch domain-containing protein	CBM32, AA5
		6	GLRG_06280	4.82	glycosyl hydrolase family 20	GII20
		7	GLRG_10476	4.89	endonuclease/Exonuclease/phosphatase	CBM13
		8	GLRG_04872	4.93	glycosyl hydrolase family 3 N terminal domain-containing protein	GII3
		9	GLRG_00636	4.94	glycosyl hydrolase family 45	GII45
		10	GLRG_11936	4.96	trypsin	CBM12
		11	GLRG_11861	5.05	kelch domain-containing protein	CBM57
		12	GLRG_06643	5.12	phosphoesterase	CBM32
		13	GLRG_06555	5.22	pectate lyase	PL1
		14	GLRG_08165	5.38	glycosyl hydrolase family 10	GII10
		15	GLRG_06582	5.51	phosphoglycerate mutase	GT2
		16	GLRG_08949	5.65	multicopper oxidase	AA1
		17	GLRG_02120	5.67	cutinase	CE5
		18	GLRG_09012	5.85	pectate lyase	PL1
中性蛋白	5	19	GLRG_03559	6.08	WSC domain-containing protein	CBM1
		20	GLRG_04776	6.17	polysaccharide deacetylase	CE4
		21	GLRG_08340	6.42	cutinase	CE5
		22	GLRG_11267	6.87	hypothetical protein	CBM1
		23	GLRG_11951	7.67	glycosyl hydrolase family 12	GII12
碱性蛋白	4	24	GLRG_10654	8.18	PIIB depolymerase family esterase	CE1
		25	GLRG_01092	8.37	glycosyl hydrolase family 28	GII28
		26	GLRG_04044	8.83	glycosyl hydrolase family 53	GII53
		27	GLRG_00973	9.01	pectate lyase	PL1

### 3 讨 论

前期,本研究小组对樟疫霉<sup>[16]</sup>、禾谷炭疽菌<sup>[10]</sup>、褐环乳牛肝菌以及希金斯炭疽菌的分泌蛋白开展的全基因组预测工作,同时,前人对一些其他植物病原卵菌、真菌、细菌的分泌蛋白均进行了相关预测以及分析<sup>[17-26]</sup>等.然而,上述分析均停留在对于上述生物分泌蛋白的预测上,尚缺乏深入而系统的功能解析.此外,前还开展了植物病原菌的寄生性与 CAZymes 之间的关系研究,明确半活体营养型的禾谷炭疽菌 CAZymes 蛋白占全部分泌蛋白的 41.04%<sup>[14]</sup>,上述研究尚缺乏对于这些蛋白理化性质的深入解析.本研究对上述 CAZymes 蛋白进行深入探索,明确禾谷炭疽菌中部分 CAZymes 蛋白的理化性质,尽管如此,上述研究均有待于今后进一步得以明确.

本研究对前期所获得的 267 个 CAZymes 蛋白进行理化性质分析,其中 2/3 的蛋白的理论等电点均小于 6,为什么这些蛋白的理论等电点会小于 6 或者这些蛋白多数属于酸性蛋白,在侵染植物以及操控植物方面上具有什么特殊功能,均有待于今后进一步通过生物学试验进行验证.

表 2 禾谷炭疽菌中 CAZymes 蛋白的理化性质

序号	蛋白 GI	理论等电点 pI 值	相对分子质量/Da	负电荷氨基酸残基数	正电荷氨基酸残基数	分子式	原子数量	半衰期/h	不稳定性系数	脂肪族氨基酸指数	总平均亲水性(GRAVY)
1	GLRG_04766	4.42	62 258.5	59	21	C <sub>2746</sub> II <sub>4122</sub> N <sub>730</sub> O <sub>889</sub> S <sub>21</sub>	8508	30	24.46	71.96	-0.232
2	GLRG_01144	4.58	41 513.4	36	19	C <sub>1892</sub> II <sub>2809</sub> N <sub>471</sub> O <sub>567</sub> S <sub>9</sub>	5748	30	35.85	77.49	-0.123
3	GLRG_10490	4.65	71 917.6	71	35	C <sub>3242</sub> II <sub>4810</sub> N <sub>850</sub> O <sub>992</sub> S <sub>11</sub>	9905	30	35.73	72.28	-0.283
4	GLRG_09366	4.66	33 260.7	31	19	C <sub>1485</sub> II <sub>2228</sub> N <sub>404</sub> O <sub>460</sub> S <sub>5</sub>	4582	30	25.22	74.73	-0.115
5	GLRG_02805	4.78	76 773.7	80	52	C <sub>3376</sub> II <sub>5188</sub> N <sub>922</sub> O <sub>1068</sub> S <sub>31</sub>	10585	30	30.23	73.56	-0.221
6	GLRG_06280	4.82	79 001.5	80	48	C <sub>3568</sub> II <sub>5371</sub> N <sub>941</sub> O <sub>1063</sub> S <sub>17</sub>	10960	30	39.92	80.82	-0.212
7	GLRG_10476	4.89	47 503.4	45	32	C <sub>2086</sub> II <sub>3197</sub> N <sub>573</sub> O <sub>678</sub> S <sub>11</sub>	6545	30	24.51	70.18	-0.338
8	GLRG_04872	4.93	93 242.5	83	60	C <sub>4176</sub> II <sub>6366</sub> N <sub>1094</sub> O <sub>1284</sub> S <sub>25</sub>	12945	30	29.64	76.89	-0.162
9	GLRG_00636	4.94	24 772.4	20	16	C <sub>1079</sub> II <sub>1607</sub> N <sub>297</sub> O <sub>343</sub> S <sub>17</sub>	3343	30	48.84	48.26	-0.32
10	GLRG_11936	4.96	38 101.2	30	21	C <sub>1641</sub> II <sub>2617</sub> N <sub>461</sub> O <sub>559</sub> S <sub>11</sub>	5289	30	32.06	81.48	0.141
11	GLRG_11861	5.05	35 853.5	28	19	C <sub>1606</sub> II <sub>2459</sub> N <sub>431</sub> O <sub>479</sub> S <sub>12</sub>	4987	30	24.86	81.33	-0.036
12	GLRG_06643	5.12	47 061.2	44	34	C <sub>2086</sub> II <sub>3171</sub> N <sub>581</sub> O <sub>647</sub> S <sub>10</sub>	6495	30	32.57	72.99	-0.294
13	GLRG_06555	5.22	34 212.9	33	25	C <sub>1492</sub> II <sub>2361</sub> N <sub>413</sub> O <sub>496</sub> S <sub>6</sub>	4768	30	22.47	83.06	-0.027
14	GLRG_08165	5.38	65 582	42	32	C <sub>2903</sub> II <sub>4475</sub> N <sub>781</sub> O <sub>924</sub> S <sub>15</sub>	9098	30	38.37	72.15	-0.046
15	GLRG_06582	5.51	40 577.5	44	34	C <sub>1832</sub> II <sub>2744</sub> N <sub>482</sub> O <sub>544</sub> S <sub>11</sub>	5613	30	38.33	67.26	-0.366
16	GLRG_08949	5.65	66 621.6	65	52	C <sub>2946</sub> II <sub>4452</sub> N <sub>820</sub> O <sub>924</sub> S <sub>15</sub>	9157	30	34.62	65.85	-0.604
17	GLRG_02120	5.67	22 790.8	12	9	C <sub>1006</sub> II <sub>1591</sub> N <sub>273</sub> O <sub>312</sub> S <sub>9</sub>	3191	30	25.09	89.54	0.135
18	GLRG_09012	5.85	36 797.4	34	27	C <sub>1654</sub> II <sub>2524</sub> N <sub>448</sub> O <sub>491</sub> S <sub>8</sub>	5125	30	24.8	86.49	-0.168
19	GLRG_03559	6.08	50 591.9	44	40	C <sub>2218</sub> II <sub>3369</sub> N <sub>613</sub> O <sub>678</sub> S <sub>35</sub>	6913	30	36.11	59.35	-0.38
20	GLRG_04776	6.17	81 837.7	58	52	C <sub>3628</sub> II <sub>5590</sub> N <sub>978</sub> O <sub>1131</sub> S <sub>26</sub>	11353	30	35.59	75.57	-0.105
21	GLRG_08340	6.42	32 710.2	25	24	C <sub>1417</sub> II <sub>2311</sub> N <sub>409</sub> O <sub>449</sub> S <sub>14</sub>	4600	30	32.94	93.82	0.084
22	GLRG_11267	6.87	27 591.8	11	11	C <sub>1231</sub> II <sub>1891</sub> N <sub>319</sub> O <sub>388</sub> S <sub>7</sub>	3836	30	20.32	72.57	-0.054
23	GLRG_11951	7.67	26 459.3	13	14	C <sub>1203</sub> II <sub>1750</sub> N <sub>310</sub> O <sub>357</sub> S <sub>6</sub>	3626	30	36.69	63.49	-0.326
24	GLRG_10654	8.18	30 496.2	17	19	C <sub>1339</sub> II <sub>2070</sub> N <sub>374</sub> O <sub>415</sub> S <sub>14</sub>	4212	30	28.64	73.23	-0.103
25	GLRG_01092	8.37	49 496.9	44	48	C <sub>2145</sub> II <sub>3447</sub> N <sub>617</sub> O <sub>684</sub> S <sub>21</sub>	6914	30	26.04	74.73	-0.257
26	GLRG_04044	8.83	38 088	24	28	C <sub>1722</sub> II <sub>2635</sub> N <sub>451</sub> O <sub>513</sub> S <sub>7</sub>	5328	30	27.42	85.34	-0.122
27	GLRG_00973	9.01	33 722.8	23	28	C <sub>1476</sub> II <sub>2359</sub> N <sub>425</sub> O <sub>469</sub> S <sub>5</sub>	4734	30	20.81	88.86	-0.058

## 4 结 论

本研究基于前期所获得的 267 个 CAZymes 蛋白,对其理化性质进行分析,同时,根据预测的理论等电点进行分类,明确 2/3 以上的分泌蛋白属于酸性蛋白.另外,通过随机选择 10% 的分泌蛋白进行理化性质分析,明确上述蛋白的相对分子质量、分子式、原子数量、半衰期、不稳定性系数、脂肪族氨基酸指数、总平均亲水性等基本性质.该研究为进一步解析禾谷炭疽菌 CAZymes 蛋白在该菌侵染和操控植物提供重要的理论基础.

## 参 考 文 献

- [1] Bergstrom GC, Nicholson RL. The biology of corn anthracnose: knowledge to exploit for improved management[J]. Plant Disease, 1999, 83(7): 596-608.
- [2] Callaway M, Smith M, Collman W. Effect of anthracnose stalk rot on grain yield and related traits of maize adapted to the northeastern United States[J]. Canadian Journal of Plant Science, 1992, 72(4): 1031-1036.

- [3] Keller N, Bergstrom G. Developmental predisposition of maize to anthracnose stalk rot[J]. *Plant Disease*, 1988(72):972-980.
- [4] Leonard K, Thompson D. Effects of temperature and host maturity on lesion development of *Colletotrichum graminicola* [stalk rot] on corn [Fungus diseases][J]. *Phytopathology*, 1976(66):635-639.
- [5] Warren II, Nicholson R, Ullstrup A, et al. Observations of *Colletotrichum graminicola* on sweet corn in Indiana[J]. *Plant Dis Rep*, 1973(57):143-144.
- [6] Anderson B, White D. Fungi associated with cornstalks in Illinois in 1982 and 1983[J]. *Plant Disease*, 1987, 71(2):135-137.
- [7] Byrnes K, Carroll R. Fungi causing stalk rot of conventional-tillage and no-tillage corn in Delaware[J]. *Plant Disease*, 1986, 70(3):238-239.
- [8] O'Connell R J, Thon M R, Hacquard S, et al. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses[J]. *Nat Genet*, 2012, 44(9):1060-1065.
- [9] 林春花, 蔡志英, 黄贵修. 全基因组法绘制禾谷炭疽菌和希金斯炭疽菌中 MAPK 级联信号途径简图[J]. *热带作物学报*, 2012(4):674-680.
- [10] 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(2):36-41.
- [11] 韩长志. 禾谷炭疽菌 14-3-3 蛋白生物信息学分析[J]. *河南师范大学学报(自然科学版)*, 2014, 42(3):109-114, 118.
- [12] 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.
- [13] 韩长志. 基于全基因组测序的禾谷炭疽菌中碳水化合物酶类蛋白预测[J]. *河南师范大学学报(自然科学版)*, 2016, 44(4):118-124.
- [14] 陈相永, 陈捷胤, 肖红利, 桂月靖, 李蕾, 戴小枫. 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.
- [15] Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server [C]. *GY wiederschain The proteomics protocols handbook*[M]. New York: Humanapress, 2005:571-607.
- [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(2):69-74.
- [17] 周晓罡, 侯思名, 陈铎文, 等. 马铃薯晚疫病菌全基因组分泌蛋白的初步分析[J]. *遗传*, 2011(7):125-133.
- [18] 杨静, 李成云, 王云月, 等. 酿酒酵母分泌蛋白组的计算机分析[J]. *中国农业科学*, 2005(3):516-522.
- [19] Wymelenberg AV, Sabat G, Martinez D, Rajangam AS, Teeri TT, Gaskell J, et al. The *Phanerochaete chrysosporium* secretome: Database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium[J]. *Journal of biotechnology*, 2005, 118(1):17-34.
- [20] 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.
- [21] 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(2):200-207.
- [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(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(15):3142-3153.
- [24] 任楠, 李俊星, 沈徐凯. 米曲霉分泌组的预测及分析[J]. *安徽农业科学*, 2010(25):13622-13625.
- [25] 范成明, 李成云, 赵明富, 等. 根瘤土壤杆菌 C58 Cereon 中分泌蛋白信号肽分析[J]. *微生物学报*, 2005(4):561-566.
- [26] Gomez M, Johnson S, Gennaro M L. Identification of Secreted Proteins of *Mycobacterium tuberculosis* by a Bioinformatic Approach[J]. *Infection and immunity*, 2000, 68(4):2323-2327.

## Prediction for CAZymes Protein from *Colletotrichum graminicola* Genome

Han Changzhi, Xu Xi

(College of Forestry; The Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, Southwest Forestry University, KunMing 650224, China)

**Abstract:** *Colletotrichum graminicola* can infect corn, wheat and other food crops, and cause tremendous economic losses in agricultural production in many countries. Based on preliminary 267 CAZymes proteins sequence the physical and chemical properties were analyzed, the results showed that the protein isoelectric point was less than 6 in nearly two-thirds proteins, which are acidic protein; while the 10 percent proteins of the samples were screened to analyze by the random number software, and it was defined that the molecular mass, the molecular formula, atomic number, half-life, the instability factor, aliphatic amino acid index, total average hydrophilicity properties of different theoretical isoelectric point category secreted protein. The study in-depth analysis of CAZymes protein function in *C. graminicola* has laid a solid theoretical foundation.

**Keywords:** *Colletotrichum graminicola*; secreted protein; physicochemical property; prediction algorithm