Analysis of Secretary Proteins in the Genome of the Plant Pathogenic fungus *Botrytis cinerea*

ZHANG Yue, YANG Jing, LIU Lin, SU Yuan, XU Ling, ZHU Youyong, LI Chengyun*

Key Laboratory of Agro-biodiversity and Pest Management of the Education Ministry of China, Yunnan Agricultural University, Kunming, 650201, China.

*Corresponding author, E-mail: li.chengyun@gmail.com

Abstract: The signal peptides prediction algorithm SignalP v3.0, subcellular protein location prediction algorithm TargetP.v1.1, potential GPI-anchor sites prediction algorithm big-PI predictor, trans-membrane domains prediction algorithm TMHMM v2.0 and bioinformatics algorithm MEME were used to analyze 16446 protein sequences of *Botrytis cinerea*. The results showed that there were 579 deduced secretary proteins. Among these proteins, the minimum and maximum of open read frame were 102 bp and 4848 bps respectively and mean score was 1271 bps. The signal peptides' length was concentrated to 16~39 amino acids and the average length was 21. 122 of these proteins contain the highly conserved host-targeting-motif RxLx within 100 residues adjacent to the signal peptide cleavage site. According to PEDNAT and COG of GenBank database, this motif's functions include metabolism modification and cell secretion etc. We blast those putative secretary proteins with RxLx motif in GenBenk and found 47.54% of them have highly conserved homologues in other species, among them 74.1% have putative protein domains. This means these proteins may be more stable or earlier origin. We suppose these proteins are candidate participating in the pathogenesis of *Botrytis cinerea* but we still need more experimental evidence to confirm their definite functions.

Key words: *Botrytis cinerea*; signal peptide; secretary protein; host-targeting-motif

Botrytis cinerea belongs to Deuteromycotina and is a widespread phytopathogenic fungus causing disease in a substantial number of economically important crops [1]. It causes Gray-mold rot or Botrytis blight and affects most vegetable and fruit crops, as well as a large number of shrubs, trees, flowers, and weeds. It also has a beneficial role in the production of rare dessert wines. The genome sequence information of *Botrytis cinerea* was released in 2005 and was helpful for us to understand this ascomycete's complex developmental life cycle, Pathogenesis mechanisms and interactions with its different host plants.

Protein is the basic function element of living organism. Many pathogenic microbes could secrete kinds of proteins into the host cell to facilitate its infection process [2]. So analysis of secreted proteins in the pathogen's genome will be

helpful to reveal its pathogenesis mechanisms. The secreted proteins used to be synthesized by ribosome and need a transport process to secrete outside the cells. There are two mechanisms for peptides transportation. The fist is cotranslational transfer. In this way, synthesized partial signal peptide combined to endoplasmic reticulum and the secreted proteins were synthesized meanwhile entered the endoplasmic reticulum, after moderated by endoplasmic reticulum and Golgi complex they were secreted outside. The second is posttranslational translocation. By this mechanism the complete proteins were synthesized and then were transported for modifying with the help of leader peptide [3]. In both mechanisms, the signal peptide has played the fundamental role. The signal peptide was usually composed by 10 to 60 amino acids. It contained a hydrophobic region (H-region) which was constituted by 6 to 15 amino-acid residues in the center and hydrophilic N terminal and C terminal at the both sides [4]. According to Gunter Blobel's signal peptide hypothesis, secretary protein's destiny was decided by its signal peptide and this peptide will be cut off when the protein arrive its destination. So we can decide whether a protein is a secretary protein by analysis of its signal peptide of N terminal [5]. Several software had been developed to indentify the signal peptide in the protein. Lee used SignalP(v2.0) analyzed 47 secretary protein and 47 other proteins of *Candida albicans,* it shows that the putative results of this software is credible[6].

The interactions between pathogens and their hosts is a hot spot for scientific research recently. How the secretary proteins entered the plant cells and play their function is still not clear now. Bacteria's type III secretary system have been illustrated by many researchers, but the pathway of the eukaryotic pathogen's secretary proteins is still unclear [7, 8]. There is a report revealed that during the process of *Plasmodium Falciparum* infected erythrocyte, most of secretary proteins which will be injected into erythrocyte contain an RxLxE/D/Q motif at 60 amino-acid residues downstream the cutting site of the signal peptide [9]. Souvik Bhattacharjee also indicated that RxLx motif also existed in hundreds of pathogenic secretary proteins of *Phytophthora infestans* which play the same function as RxLxE/D/Q motifs of *Plasmodium Falciparum. Plasmodium Falciparum* and *Phytophthora infestans* are far related and infect animal and plant respectively; they should have different pathogenic process and mechanism, so RxLx motif may be a conserved signal recognition motif of eukaryotic pathogen [10]. In this study, we try to make use of the genome data to indicate how many secretary proteins contained by *Botrytis cinerea* and whether RxLx motif exist in this saprophytic fungi's secretary proteins and play pathogenic function.

Materials and methods

The sequence data of *Botrytis cinerea* was downloaded from the database of BROAD institute（*Botrytis cinerea* strain B05.10. It totally contained 16446 putative proteins in the database (http://www.broad.mit.edu/annotation/genome/botrytis_cinerea/Home.html), by the use of signal putative software SignalP (http://www.cbs.dtu.dk/services/SignalP/) [11], subcellular organelle located software TargetP (http: //www.cbs.dtu.dk/services/TargetP1.1) [12, 13], anchoring protein analysis software big-PIPredicto（ http://mendel.imp.ac.at/sat/gpi/gpi_server ）[14] and transmembrane helix analysis software TMHMMServer (http://www.cbs.dtu. dk/services /TMHMM）[15], we selected those proteins with signal peptide (satisfied L=-918.235-123.455*(Mean.S.score)+1983.44*(HMM scores) and L>0) [6], secreted outside the cells not to other subcellular organelle, not anchoring protein and didn't contain transmembrane helix as secretary proteins. Then we use MEME(Multiple Expectation Maximization for Motif Elicitaion）[16] to test whether these putative secretary proteins contain RxLx motif and whether there exist conserved amino-acid residue at the two sides of this motif. Then the graph was produced by software LOGO[17] according to the result of MEME. We also compared the sequence of these putative secretary proteins with RxLx motif to PEDNAT database（http://pedant.gsf.de/index.jsp）and searched them in the COG database from GeBank （http://www.ncbi.nlm.nih.gov/COG/old/xognitor.html）to find and categorized the putative functions of these proteins. At last, we blast these sequence in the GeBank （ BLASTP 2.2.17 (Jun-24-2007)（http://www.ncbi.nlm.nih.gov/）try to find the homologues of these proteins and conjectured their functions.

Results and analysis

As figure 1 showed, we get 868 proteins contained the signal peptide in the *Botrytis cinerea's* genome, among them 579 proteins which account for 3.52% of the *Botrytis cinerea's* genes were predicted to be secretary proteins by the method have been mentioned.

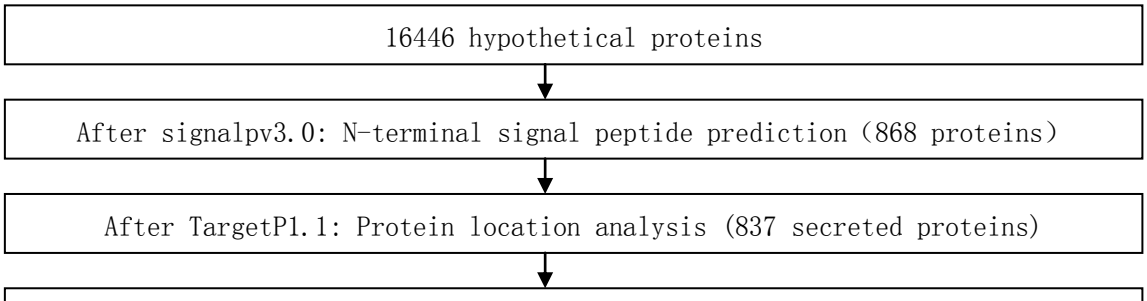

Fig 1. Analysis flow chart of *Botrytis cinerea* secretary proteins

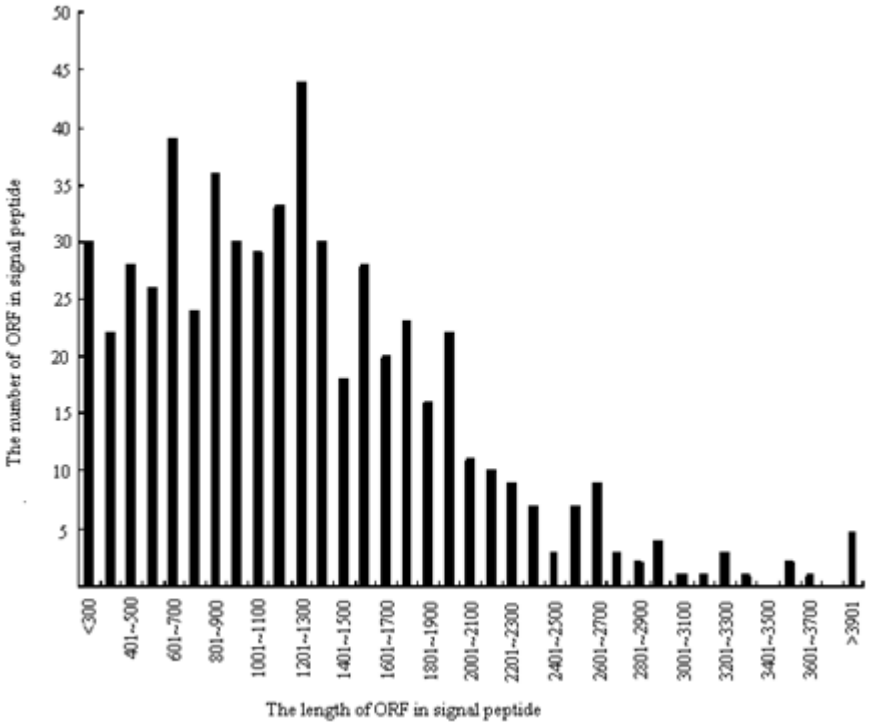

Fig 2. The ORF length distribution of 284 secretary proteins in *Botrytis cinerea*

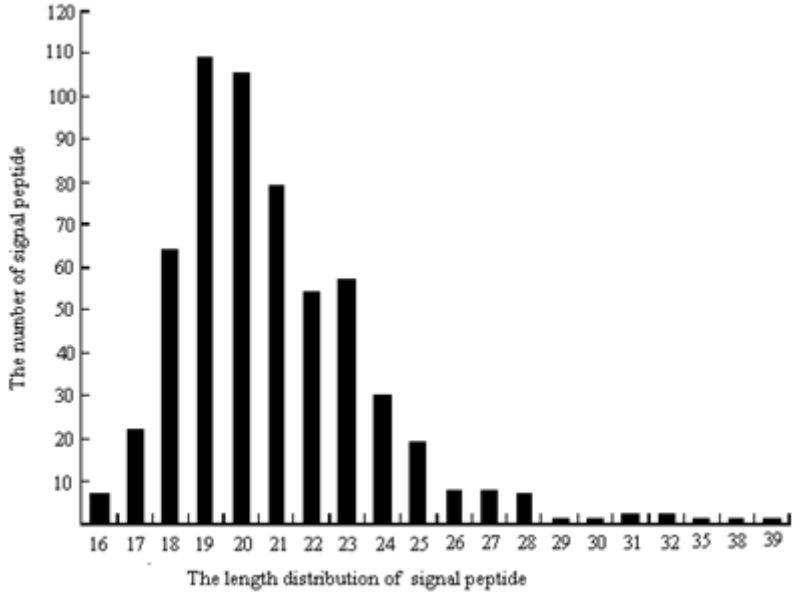

Fig 3. The signal peptide length distribution of 284 secretary proteins in *Botrytis cinerea*

The average length of the predicted secretary proteins of *Botrytis cinerea* is 1271bp; the longest one and the shortest one are 4848bp and 102bp respectively. From figure 2, we can find that the most of them are 500-2000bp, which account for 72.19% of all the genes. The average length of these secretary proteins' peptide is 21 amino-acid residues, the longest one is 39 and the shortest one is 16. The figure 3 showed the length distribution of these proteins' peptide. 540 of them are 17-25 amino-acid residues which account for 93.3% of all and proteins with 19 amino-acid residues come up to 109 proteins (18.8%).

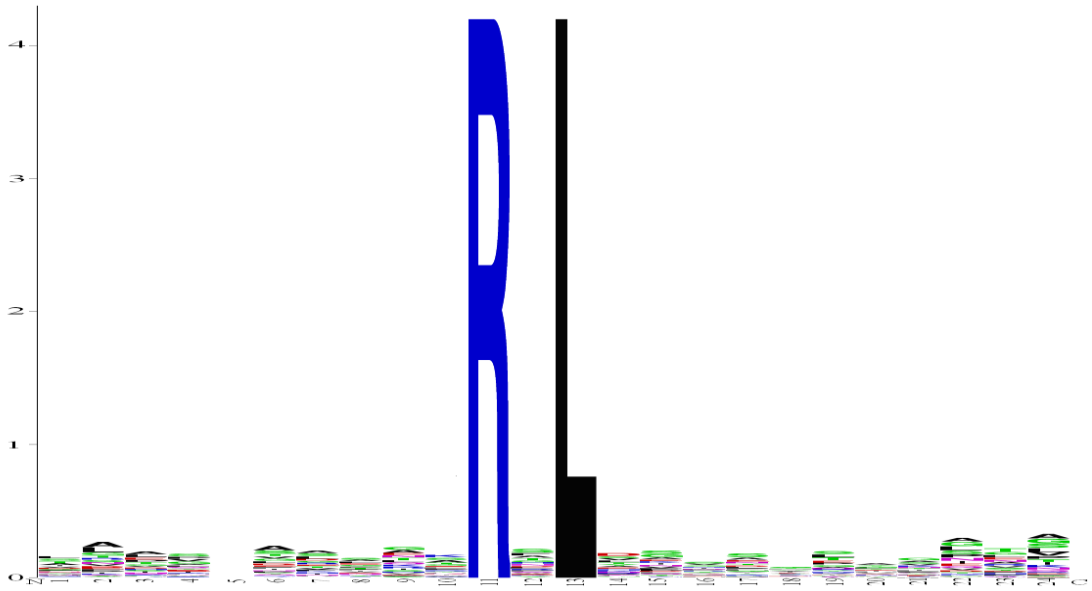

Fig 4. Logo shows the conservation of the RxLx motif from predicted *Botrytis cinerea* secretary proteins

The motifs are highlighted in board letters in the table. Abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu;F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P,Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

We analyzed the all these putative secretary proteins by MEME software and found that there are 122 proteins (21% of total) contain RxLx motif within the 100 amino-acid residues downstream of the cutting site of signal peptide. The report had showed that *Plasmodium Falciparum*'s pathogenic secretary proteins had conserved E, D or Q fowled the RxLx motif while *Phytophthora infestans* didn't. But from the figure 4, we can found that there more A, G, L and S appeared at the downstream and upstream of the RxLx motif in *Botrytis cinerea*. A, G and L are all nonpolar amino acid and S is polar neutral amino acid. The most conserved amino-acid residue followed the RxLx motif is D. This is just the same as *Plasmodium Falciparum*.

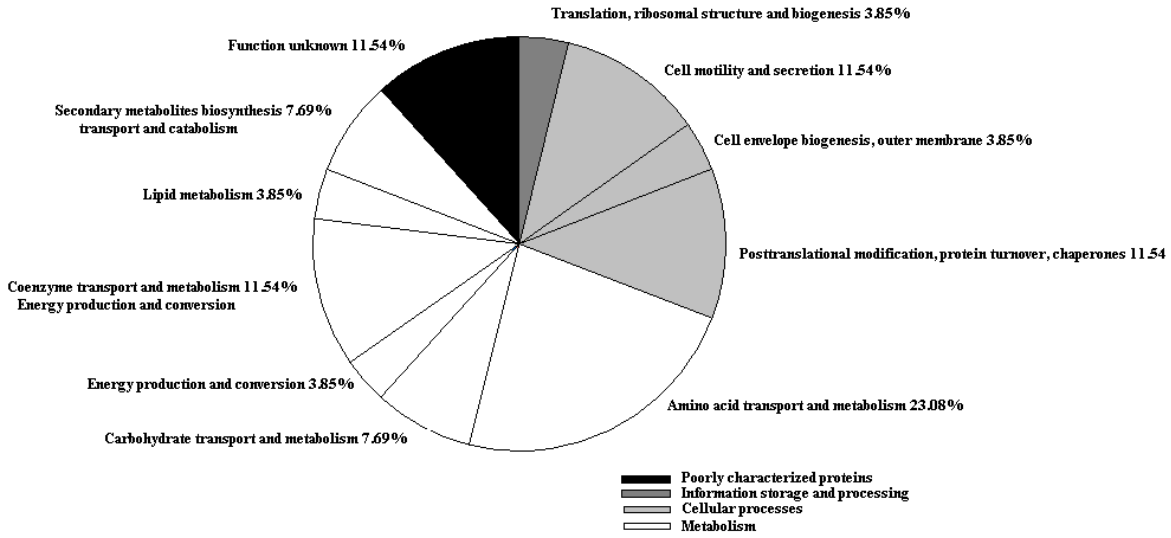

Fig 5. Functional categorization of putative secretary proteins containing RxLx motif which have function descriptions in COG database

Then we categorized the putative functions of these proteins by COG of GenBank (figure 5). There are only 26 (21.3%) proteins found in COG and by categorized into 11 different kinds of functions. Most of them are related to amino-acid metabolism (23.08%). For further predicted the functions of our putative secretary proteins, we compared them in the PEDANT database but only 6 of them have the specific function description (Table 1). The most of them are

Tab 1. The functional description of secretary proteins containing the RxLx motif in *Botrytis cinerea*

Gene name	start	end	Functional description
BC1G_15580	29028	29981	may act as a sorting receptor in the delivery of vacuolar hydrolases, partial
BC1G_05799	27623	28429	peptidylprolyl isomerase
BC1G_12776	79340	77335	tripeptidyl peptidase
BC1G_04994	285638	283980	hypothetical protein similar to alpha-L-arabinofuranosidase
BC1G_06540	150762	152263	hypothetical protein similar to aspartic proteinase precursor
BC1G_03579	412126	410768	hypothetical protein similar to aspartyl protease

related to cell metabolism and some of them also appeared in the *Phytophthora infestans's* pathogenic secretary proteins.

Table 2. The blast results of putative secretary proteins include RxLx motif

Gene name	number of homologs	No. of the species containing	No. of Pathogenic species	Conserved protein domain

		homolog		
BC1G_00639	5	2	1	Tannase
BC1G_00978	2	1	1	Pro-kumamolisin
BC1G_01009	9	5	4	N
BC1G_01027	18	10	7	Peptidase_S10
BC1G_01073	4	2	1	Pro-kumamolisin
BC1G_01628	6	6	4	Peroxidase
BC1G_01874	7	6	3	Glycine-rich protein domain
BC1G_02021	19	11	8	GMC oxidoreductase; choline dehydrogenase
BC1G_02163	30	24	14	Cerato-platanin
BC1G_02492	2	2	2	N
BC1G_02944	1	1	1	Pro-kumamolisin
BC1G_03275	14	13	8	N
BC1G_03557	24	22	9	N
BC1G_03560	4	4	2	N
BC1G_03579	28	22	7	Asp, Eukaryotic aspartyl protease
BC1G_04705	7	6	3	Peroxidase
BC1G_04994	26	18	9	Alpha-L-arabinofuranosidase B
BC1G_05488	4	4	2	N
BC1G_05765	7	4	3	Pro-kumamolisin
BC1G_05799	58	39	8	FKBP-type peptidyl-prolyl cis-trans isomerase
BC1G_05885	5	4	3	N
BC1G_06035	100	43	19	Glycosyl hydrolase family 7; Fungal cellulose binding domain
BC1G_06328	6	6	4	Bacterial alpha-L-rhamnosidase
BC1G_06540	14	13	11	Asp, Eukaryotic aspartyl protease
BC1G_07149	24	18	11	Peptidase_S10
BC1G_07160	8	8	6	Phospholipase C
BC1G_07483	3	3	2	Esterase_lipase
BC1G_07899	2	1	0	Cutinase
BC1G_08048	5	4	3	Amidase
BC1G_08735	27	21	12	Cerato-platanin
BC1G_08755	50	21	7	Glycosyl hydrolase family 15; The family 20 carbohydrate-binding module (CBM20)
BC1G_09129	13	13	10	DnaJ domain
BC1G_09495	2	2	2	Tannase
BC1G_09611	5	5	4	DadA, Glycine/D-amino acid oxidases
BC1G_10333	4	4	1	Rossmann-fold NAD(P)(+)-binding proteins
BC1G_10397	10	10	6	N
BC1G_10482	2	2	1	N
BC1G_10768	12	8	5	Intradiol_dioxygense_like domain
BC1G_11019	8	8	5	Salicylate hydroxylase

BC1G_11134	9	9	6	Survival protein SurE
BC1G_12138	8	6	3	Alpha-L-arabinofuranosidase C-terminus
BC1G_12157	58	29	15	Arginase family
BC1G_12171	11	7	4	N
BC1G_12200	18	14	5	Peptidase family M20/M25/M40; Acetylornithine deacetylase
BC1G_12456	9	7	6	Glyco_hydrolase_16
BC1G_12525	4	3	3	Peroxidase
BC1G_12619	2	2	2	N
BC1G_12776	3	2	2	Pro-kumamolisin
BC1G_12932	25	7	2	Tannase
BC1G_13158	2	2	1	N
BC1G_13581	10	6	6	N
BC1G_13855	7	6	3	BglC, Endoglucanase
BC1G_14244	86	62	13	N
BC1G_14398	15	12	8	DadA, Glycine/D-amino acid oxidases
BC1G_14702	100	36	11	Glycosyl hydrolase family 7
BC1G_15580	9	9	6	N
BC1G_15641	5	3	1	tol-pal system beta propeller repeat protein TolB
BC1G_16238	8	6	3	BglC, Endoglucanase

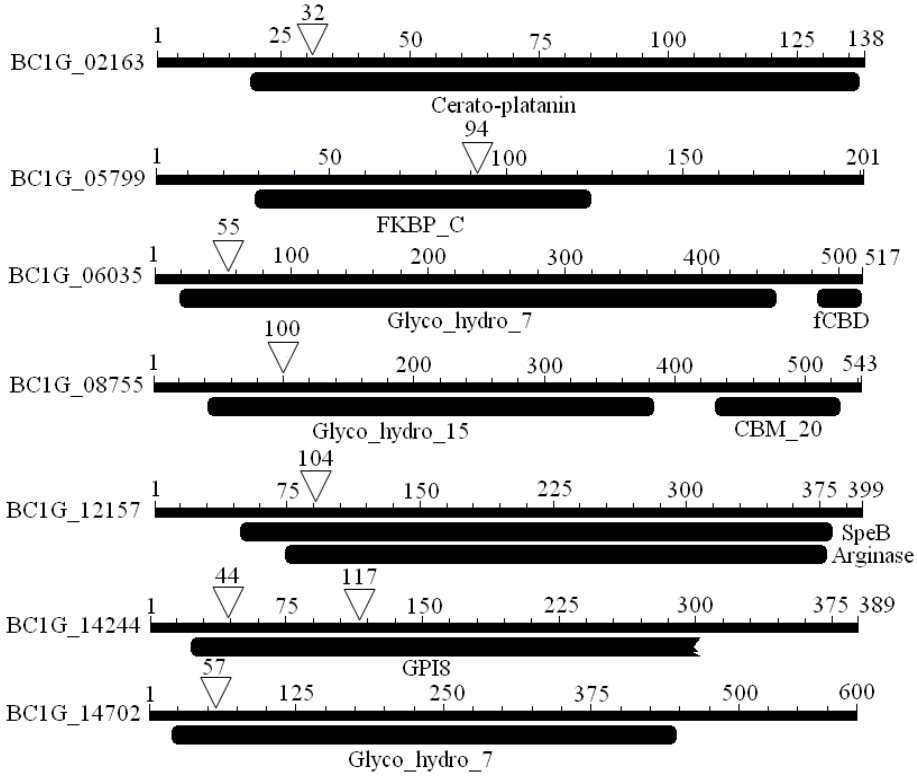

Fig 6. The location of RxLx motif and putative protein domain of 7 proteins which contain over 50 homologs in other species

Then we blast all these proteins in the GenBank and found that 58 proteins have high conserved homologues (E-value〈1×10-20 and identities〉40%) in other species (accounted for 47.54% of all predicted proteins) and most of them have putative conserved protein domains (Table 2). Then we selected 7 the most conserved (have more than 50 homologues) proteins and marked out the location of their RxLx motif and protein domains (Figure 6). We can see that in 6 proteins, the motif appeared within the first 10 amino-acid residues of the C terminals of their protein domains.

Discussion

By application of software of bioinformatics' analysis, we found 579 putative secretary proteins and 122 (21%) of them contained Host-Targeting-motif RxLx within the 100 amino-acid residues downstream of their cutting sites in the genome of *Botrytis cinerea*. The length of most of these proteins and their signal peptide are moderate. RxLx motif spread in the secretary proteins of this kind of saprophytic fungi indicate they maybe play functions in this fungus's pathogenic secretary pathways. But we still need experiment evidence to conform this hypothesis.

By compared these proteins contain RxLx motif in the PEDANT database and COG of GenBank, we seldom found the definite description to them. For those proteins with description in COG, most of them are related to cell metabolism and cellular process. Interestingly, among these proteins, BC1G_06540 is described as an aspartic proteinase precursor. It is had already been revealed that *Botrytis cinerea* could secreted aspartic proteinase into its host cell during its pathogenic process [18]. Researchers also found that the a pathogens infected the plant, it must break through the plant cell's physical outline such as cell wall as well as changed the inner environment of host cells to fit themselves. In this process, their secretions must play the most important roles [19]. Considered other 5 RxLx motif containing proteins with definite descriptions in PEDANT database are all enzymes involved in the cell metabolism and the categorization with most proteins by COG is related to amino-acid metabolism, we supposed that maybe *Botrytis cinerea* need secreted proteinase into the plant cells in its pathogenic process.

Blast these RxLx motif containing proteins in the GenBank, we find that 47.54% of them have high conserved homologues in other species and most have

the conserved protein domains. This indicates these proteins with RxLx motif are conserved during the evolution process or early originated in the history. Most of the homologues contained in fungi but still some appeared in higher eukaryotes. Among them, BC1G_14702, BC1G_14244, BC1G_05799 and BC1G_06035 contained huge number of homologues distributed in so many different kinds of plants and animals. This indicated that maybe these genes played the irreplaceable role in organisms as well as *Botrytis cinerea*. BC1G_05799 had been depicted as peptidylprolyl isomerase in both COG and PEDANT database. This kind of protein interconverts the *cis* and *trans* isomers of peptide bonds with the amino acid proline and involved in many cellular process. The 6 of 7 RxLx motifs in the highest conserved putative secretary proteins are all located at the N terminal of their protein domains. Interestingly, the description of protein domains contained by these 7 proteins indicated they all maybe involved in the secretion process even in the pathogenic process. Cerato-platanin had been reported as a kind of phytotoxic protein which was secreted by *Ceratocystis fimbriata* f.sp. *platani* [20]. FKBP_C is related to protein synthesis and locate of plant plasmids [21]. Glyco_hydro_7 is a kind of glycosyl hydrolase which was secreted to destroy cellulose in the plant cell [22]. SpeB have been reported as a kind of virulent effectors secreted by Group A Streptococcal [23]. GPI8 is related to the synthesis of phosphatidylinositol anchor protein which could anchor it to the cell surface [24]. So these proteins may play their functions in the *Botrytis cinerea's* secretion pathway and contributed to its pathogenic process.

Most of pathogenic proteins reported now are secretary proteins. In this article we predicted 579 secretary proteins and 122 of them contain Host-Targeting-motif RxLx which could be treated as candidate pathogenic proteins of *Botrytis cinerea*. Although we still need experiment to prove whether these proteins contributed to pathogeneses, find these candidate proteins will accelerate our understandings of pathogenic mechanism of *Botrytis cinerea*. Many software used to analyze the protein had been proved effective and it is conveniently for us to understand the information lied in the genome by the help of them.

Reference

[1] Hausbeck MK and Moorman GW. Managing Botrytis in greenhouse-grown flower crops. Plant Dis. 1996, 80:1212-1219.

[2] Ellis, J. Catanzariti, A. M. and Dodds, P. The problem of how fungal and oomycete avirulence proteins enter

plant cells, Trends Plant Sci. 2006, 11:61–63.

[3] Meyer, D. I. The signal hypothesis — a working model. 1982, 7:320-321.

[4] Martoglio, B. and Dobberstein, B. Signal sequences: more than just greasy peptides. Trends in cell biology. 1998, 8:410-415.

[5] Cheng-Gang, Z. and Fu-Chu, H. E. Bioinfornation methods and practice. Beijing: Science Press. 2002, 67～69.

[6] Lee, S. A. Wormsley, S. and Kamoun, S. Ananalysis of the Candida albicans genome database for soluble secreted proteins using computer-based prediction algorithms. Yeast. 2003, 20: 595-610.

[7] Christie, P. J. and Atmakuri, K. Biogenesis, architecture, and function of bacterial type IV secretion systems. Annu Rev Microbiol. 2005, 59: 451–485.

[8] Journet, L. and Hughes, K. T. Type III secretion: A secretory pathway serving both motility and virulence (review). Mol Membr Biol. 2005, 22:41-50.

[9] Hiller, N. L. and Bhattacharjee, S. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science. 2004, 306: 1934–1937.

[10]Bhattacharjee, S. and Hiller, N. L. The malarial host-targeting signal is conserved in the Irish potato famine pathogen. PLoS Pathogens. 2006, 2:e50

[11]Jannick D. B. and Henrik N. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 2004, 340:783-795.

[12]Emanuelsson, O. and Nielsen, H. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 200: 300: 1005-1016.

[13]Nielsen, H. and Engelbrecht, J. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Design and Selection. 1997, 10:1-6.

[14]Gpi, G. P. I. Automated annotation of GPI anchor sites: case study C. elegans. TIBS. 2000, 25: 340-341.

[15]Krogh, A. and Larsson, B è Predicting transmembrane protein topology with a hiddenMarkov model : application to complete genomes. Journal of Molecular Biology. 2001, 305:567-580.

[16]Bailey, T. L. and Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. 1994, 2:28-36.

[17]Crooks GE, Hon G, Chandonia JM, Brenner SE Web Logo: A sequence logo generator Genome Research, 14:1188-1190, (2004)

[18]ten Have, A. and Dekkers, E. An aspartic proteinase gene family in the filamentous fungus Botrytis cinerea contains members with novel features. Microbiology. 2004, 150:2475-2489.

[19]Haldar, K. and Kamoun, S. Common infection strategies of pathogenic eukaryotes. Nat Rev Microbiol. 2006, 4: 922–931.

[20]Pazzagli, L. and Cappugi, G. Purification, Characterization, and Amino acid sequence of cerato-platanin, a new phytotoxic protein. from *Ceratocystis fimbriata f.sp. platani*. The Journal of Biological Chemistry. 1999, 274: 24959-24964.

[21]Li, S. and Nosenko, T. Phylogenomic Analysis Identifies Red Algal Genes of Endosymbiotic Origin in the Chromalveolates. Mol. Biol. Evol. 2005, 23: 663-674.

[22]Sulzenbacher, G. and Driguez, H. Structure of the Fusarium oxysporum endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. Biochemistry. 1996, 35:15280-15287.

[23]Kansal, R. G. and McGeer, A. Inverse Relation between Disease Severity and Expression of the Streptococcal Cysteine Protease, SpeB, among Clonal M1T1 Isolates Recovered from Invasive Group A Streptococcal Infection Cases. Infect and Immun. 2000, 68: 6362-6369

[24]Vidugiriene, J. and Menon, A. K. The GPI anchor of cell-surface proteins is synthesized on the cytoplasmic face of the endoplasmic reticulum. J. Cell Biol. 1994, 127:333-341.