JOURNAL OF PURE AND APPLIED MICROBIOLOGY, August 2014. Vol. **8**(4), p. 2645-2654

Genome-wide Analysis for Secreted Proteins of *Cochliobolus heterostrophus*

Liqin Zhao[1,2]* and Tong Liu[3]

[1]Agricultural College, Northeast Agricultural University, Harbin 150030, P. R. China.
[2]Test Center, Heilongjiang Bayi Agricultural University, Daqing 163319, P. R. China.
[3]Institute of Plant Pathology and Applied Microbiology, Heilongjiang Bayi Agricultural University, Daqing 163319, P. R. China.

(Received: 23 January 2014; accepted: 02 April 2014)

Secreted proteins of *Cochliobolus heterostrophus* were first identified in genome wide using bioinformatics and computer-based prediction algorithms, which would contribute to elucidate the molecular mechanism of the pathogen-plant interaction. In this study, several bioinformatics softwares were used to identify secreted proteins from 13 316 gene encoding proteins of *C. heterostrophus*. Subsequently, characteristics of N-terminal signal peptides of secreted proteins were analyzed. Predicted secreted proteins were annotated using pathogen-host interaction (PHI) database and carbohydrate-active enzymes (CAZYmes) database respectively to screen putative PHI associated proteins, cellulases, pectinase and cutinases. Potential fungal effectors were predicted based on their conserved domains. As a result, *C. heterostrophus* secretome contained 886 secreted proteins. Secreted proteins with 18 amino acid residues of signal peptides accounted for the largest percentage. The frequency of nonpolar amino acids used in signal peptides was the highest, while that of polar and negatively charged amino acids the lowest. Amino acids used at -3 and -1 sites of signal peptide cleavage sites were relatively conserved. *C. heterostrophus* secretome contained 164 putative PHI-associated proteins, 162 CAZYmes (32 cellulases, 15 pectinases and 14 cutinases), 147 small cytein-rich secreted proteins, which help us understand the pathogenesis of *C. heterostrophus*.

Key words: *Cochliobolus heterostrophus*, Secreted protein, Signal peptide, Secretome, CAZYmes.

C. heterostrophus is a plant pathogenic fungus, which can cause Southern corn leaf blight in maize, an important leaf disease in maize. *C. heterstrophus* is found in many tropical regions and in the southern part of the US. *C. heterstrophus*, although not currently the most economically serious disease, can be a very serious crop disease. As a plant pathogen, *C. heterstrophus* can cause host plant disease by undergoing a series of crucial processes including attachment to the plant surface, germination, formation of infection structures, penetration and colonization[1]. Many virulence gene encoding proteins such as effectors and several pathogenic factors such as toxin, melanin and extracellular enzymes together play important roles in these processes, and some proteins of which are secreted proteins including effectors and extracellular enzymes essential for pathogen attachment, penetration and colonization. Therefore, understanding the secreted proteins of the pathogen will contribute to further realization of its pathogenic mechanism.

Secreted proteins are a group of proteins that can be secreted out of the cell by use of their own signal peptides, which usually are located at the 15-60 N-terminal amino acids and guide the transshipment and localization of these proteins after production[2]. Signal peptide often contains the hydrophilic N section, lyophobic H section and C section which has a cleavage site of

* To whom all correspondence should be addressed.
E-mail: lqzhao2000@126.com

restriction endonuclease[3]. Secreted proteins are synthesized in endoplasmic reticulum. Firstly, translating ribosome is attached to endoplasmic reticulum by the guide of signal peptide. Secondly, signal peptide is identified and combined by the receptor precursor in the endoplasmic reticulum. Thirdly, signal peptide is identified and degraded by signal peptidase, in which two amino acid residues at -3 and -1 sites in font of the cleavage site are crucial. At last, the polypeptide is secreted out of the cell.

After being synthesized, secreted proteins can be transported to cell surface, and periplasmic space to perceive its environment and do adaptive reaction. They contain lots of enzymes which are important for the exchange of material, energy and information[4]. In recent years, studies have shown that the interaction between secreted proteins from pathogen such as effectors and peptide products of resistance genes from host is central to the pathogen-host interaction. Additionally, extracellular enzymes including cellulase, pectinase and cutinase also are secreted proteins, and can degrade plant cell wall and cutin to facilitate pathogen penetration. So far, we have entered the era of omics such as secretome that contains all secreted proteins of organism as the development of molecular biology and DNA sequencing technology. Thus, it is necessary to study the scretome of *C. heterstrophus* for fully revealing its interaction with host plant, which can be achieved by the genome sequencing of *C. heterstrophus*.

MATERIALS AND METHODS

Secretome prediction

Secreted proteins should contain following four characteristics: N-terminal signal peptide, no transmembrane domains, no GPI-anchor site, no localization peptides targeting the protein to mitochondria or other intracellular organelles. N-terminal signal peptides, transmembrane helices and GPI-anchor sites were predicted using SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/)[5], TMHMM v 2.0 (http://www.cbs.dtu.dk/services/TMHMM/)[6] and Big-PI Predictor (http://mendel.imp.ac.at/gpi/gpi_server)[7-8], respectively. TargetP 1.1 was used to predict the subcellular location of eukaryotic proteins (http://www.cbs.dtu.dk/services/TargetP/)[9]. The location assignment was based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP), thus SPs were predicted by TargetP 1.1.

Prediction of pathogenicity associated secreted proteins

Putative secreted proteins related to pathogen-host interaction were identified by blast searches with a cutoff of 10^{-5} in predicted secreted proteins against pathogen-host interaction (PHI) database (http://www.phi-base.org/)[10]. Cellulase, pectinase and cutinase were screened using CAZYmes Analysis Toolkit (http://mothra.ornl.gov/cgi-bin/cat.cgi)[11]. Small cytein-rich secreted proteins (SCRSPs) were predicted based on their expected sequence characteristics[12], and usually had 20 to 200 amino acid residues of open reading frame (ORF) containing a N-terminal signal peptide and at least four cystein residues[13]. Secreted proteins of *C. heterstrophus* with these characteristics were considered as the putative SCRSPs. Searches for conserved domains of SCRSPs were performed using an online tool CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)[14].

Data set

Protein sequences of *C. heterstrophus* were downloaded from NCBI (Accession No. EMD84629-EMD97944)[15].

RESULTS

Genome-wide prediction of secreted proteins

1179 ORFs containing N-terminal signal peptides were screened out from 13316 ORFs of *C. heterstrophus* C5 using SignalP 4.1. Transmembrane helices of the 1179 ORFs were analyzed using TMHMM Server v. 2.0. As a result, a total of 953 ORFs were with no transmembrane helices after excluding 226 ORFs with at least one transmembrane helice. Subsequently, the 953 ORFs without transmembrane helices were analyzed by Big-PI Predictor to predict GPI-anchored sites, and consequently 917 ORFs without GPI-anchored sites were obtained after excluding 36 ORFs with GPI-anchored sites. Lastly, 886 proteins with secretory pathway signal peptides (about 6.7% of

Table 1. Frequencies of amino acid residues around the cleavage site of signal peptides in *C. heterostrophus*

Amino acid	-3		-2		-1		+1		+2		+3	
	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)
A	394	44.5*	121	13.7	689	77.9*	193	44.5*	29	3.3	89	10.1
C	11	1.2	7	0.8	11	1.2	8	1.2	14	1.6	26	2.9
D	1	0.1	14	1.6	0	0.0	25	0.1	50	5.6	27	3.1
E	0	0.0	18	2.0	1	0.1	24	0.0	33	3.7	21	2.4
F	1	0.1	43	4.9	2	0.2	35	0.1	23	2.6	40	4.5
G	21	2.4	8	0.9	75	8.5	34	2.4	38	4.3	41	4.6
H	0	0.0	29	3.3	0	0.0	48	0.0	8	0.9	26	2.9
I	29	3.3	31	3.5	0	0.0	26	3.3	23	2.6	81	9.2
K	4	0.5	10	1.1	1	0.1	22	0.5	15	1.7	15	1.7
L	12	1.4	150	16.9	3	0.3	88	1.4	36	4.1	76	8.6
M	2	0.2	11	1.2	1	0.1	2	0.2	5	0.6	10	1.1
N	2	0.2	37	4.2	0	0.0	12	0.2	38	4.3	38	4.3
P	1	0.1	3	0.3	11	1.2	1	0.1	280	31.6*	64	7.2
Q	1	0.1	46	5.2	2	0.2	140	0.1	36	4.1	21	2.4
R	1	0.1	17	1.9	5	0.6	28	0.1	20	2.3	19	2.1
S	68	7.7	184	20.8*	65	7.3	84	7.7	88	9.9	74	8.4
T	91	10.3	67	7.6	18	2.0	34	10.3	69	7.8	103	11.6*
V	246	27.8	52	5.9	0	0.0	48	27.8	46	5.2	83	9.4
W	0	0.0	8	0.9	0	0.0	8	0.0	10	1.1	10	1.1
Y	0	0.0	29	3.3	1	0.1	25	0.0	24	2.7	21	2.4

*The highest frequency of amino acid residue. -3 to -1 mean the third, second and first amino acid sites at upstream of cleavage, respectively. +1 to +3 mean the first, second and third amino acid sites at downstream of cleavage, respectively.

13316 gene encoding proteins) identified from the 917 ORFs using Target P 1.1 were considered as secreted proteins, showing that secreted proteins were well present in the genome of *C. heterstrophus*.

The sizes of these 886 secreted proteins of *C. heterstrophus* were 59 to 1864 amino acids (aa), and the average size was 378 aa. These secretory proteins were classified on the basis of their sizes (Fig. 1A). It was showed that the amount of secretory proteins with 201-400 aa accounted for the highest percentage (305), followed by 401-600 aa (228), less 200 aa (221), 601-800 aa (90), 801-1000 aa (31) and 1001-1200 aa (7) in turn, showing that the sizes of secreted proteins were mainly rich in less $\leq$ 600 aa (85%).

Fig. 1B showed the length distribution of signal peptides of 886 secreted proteins. The length

Table 2. The highest frequency of amino acid residue around the cleavage site of signal peptide in *C. heterostrophus* and other organisms

Organisms	-3	-2	-1	+1	+2	+3
Cochliobolus heterostrophus	A	S	A	A	P	T
Agrobacterium tumefaciens	A	L	A	A	D	L
Phytophthora infestans	A	A	A	A	A	A
Neurospora crassa	A	L	A	A	P	
Verticillium dahliae	A	S	A	A	P	L
Ctenopharyngodon idellus	A	N	A	A	S	S/W
Caenorthaditis elegans	V	S	A	Q	P	I

-3 to -1 mean the third, second and first amino acid sites at upstream of cleavage, respectively. +1 to +3 mean the first, second and third amino acid sites at downstream of cleavage, respectively.

Table 3. Conserved domains search of small cytein-rich secreted proteins against CDB database

Proteins accession	Hit type	Match site	E-Value	Accession	Conserved domains
EMD93850.1	specific	1-131	1.64E-46	cd00833	PKS
EMD89003.1	specific	21-87	8.82E-09	pfam05730	CFEM
EMD85494.1	superfamily	24-138	6.61E-25	cl17157	Alt_A1 superfamily
EMD85503.1	superfamily	18-126	0.000926	cl07470	CVNH superfamily
EMD85360.1	superfamily	121-165	0.000693	cl18857	frhA superfamily
EMD92796.1	specific	21-138	8.79E-63	pfam07249	Cerato-platanin
EMD85692.1	superfamily	24-164	5.15E-08	cl16763	DUF4360 superfamily
EMD85589.1	superfamily	22-199	1.90E-40	cl16763	DUF4360 superfamily
EMD87933.1	specific	43-140	9.09E-38	cd00606	fungal_RNase
EMD92088.1	superfamily	74-114	0.009332	cl09109	NTF2_like superfamily
EMD94761.1	specific	66-142	6.92E-11	pfam01822	WSC
EMD91522.1	specific	19-84	3.37E-09	pfam05730	CFEM
EMD93471.1	superfamily	113-155	0.00225	cl06022	Hydrophobin_2 superfamily
EMD96727.1	superfamily	18-138	2.17E-06	cl17157	Alt_A1 superfamily
EMD92194.1	specific	26-153	1.11E-09	cd00161	RICIN
EMD89918.1	superfamily	29-136	8.92E-22	cl17157	Alt_A1 superfamily
EMD90124.1	superfamily	40-89	1.56E-05	cl00112	PAN_APPLE superfamily
EMD91820.1	superfamily	30-145	1.22E-15	cl00212	microbial_RNases superfamily
EMD88914.1	superfamily	19-51	0.001	cl16916	ChtBD1 superfamily
EMD90271.1	specific	48-116	0.000106	pfam05730	CFEM
EMD97201.1	specific	28-89	5.15E-10	pfam06766	Hydrophobin_2
EMD96258.1	specific	115-147	3.17E-06	pfam00734	CBM_1
EMD90447.1	specific	24-88	1.27E-25	pfam06766	Hydrophobin_2

range of signal peptides was 10 to 34 aa, and the average was 19 aa. The distribution of signal peptides was close to normal distribution by quantity, and the number of signal peptides with 18 aa accounted for the highest percentage (169), followed by 19 aa (153) and 17 aa (127) in turn (Fig. 1B), which was similar to that in *Caenorthaditis elegans*[16] and *Verticillium dahliae*[17]. The length variation of the signal peptides should be related to the variation of protein function.

Characteristics analysis for signal peptides of secreted proteins

The compositions of amino acid residues in signal peptides of secreted proteins were analyzed. 20 amino acid residues were in following order based on the frequencies in signal peptides of secreted proteins from high to low: alanine acid (A), Leucine (L), Serine (S), Threonine (T), Valine (V), Methionine (M), Phenylalanine (F), Isoleucine (I), Glycine (G), Proline (P), Arginine (R), Lysine (K), Glutamine (Q), Tyrosine (Y), Cysteine (C), Histidine (H), Asparagine (N), Tryptophan (W), Aspartic acid (D) and Glutamic acid (E) (Fig. 1C). Nonpolar amino acids (A, L, V, M, F, I, P and W) had higher frequencies (66.7%), of which the frequency of A was the highest (20.4%), followed by polar amino acids without charge (S, T, G, Q, C, Y and N) (26.7%) and polar and positively charged

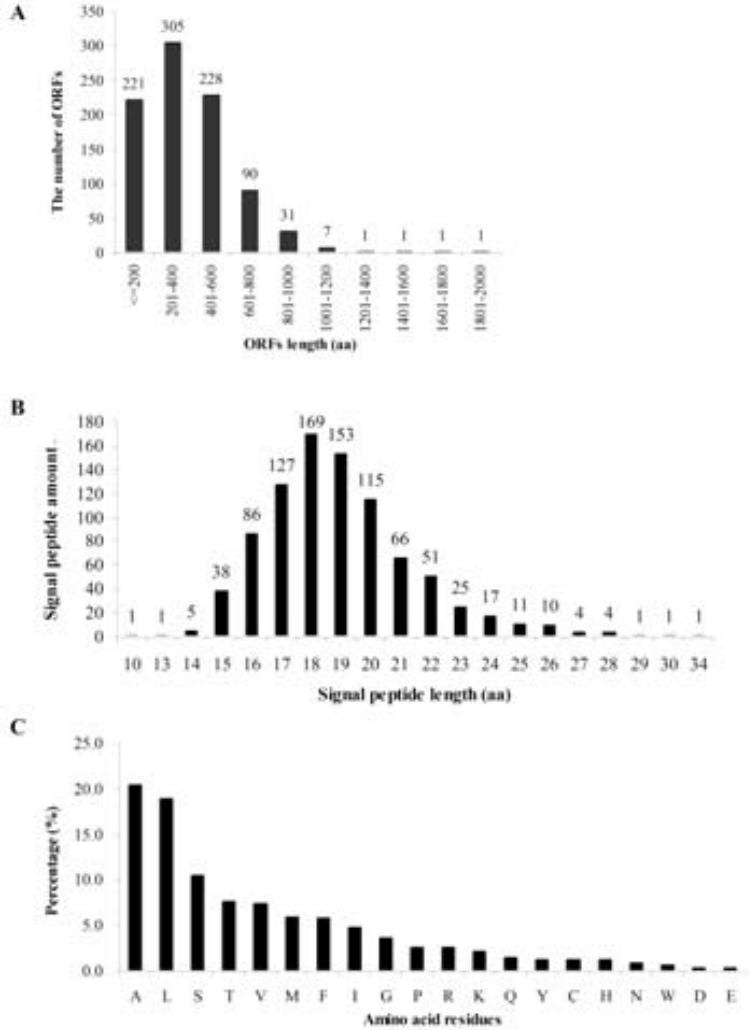

Fig. 1. Analysis of secreted proteins and signal peptides of *C. heterostrophus*. (A). The length distribution of ORFs for 886 secreted proteins of *C. heterostrophus*. (B) The signal peptide length distributions of of 886 secreted proteins in *C. heterostrophus*. (C) The frequencies of 20 amino acid residues in the signal peptides of secreted proteins of *C. heterostrophus*. aa, amino acid.

amino acids (R, K and H) (6.0%). The lowest frequencies (0.3%) of amino acids were two polar and negatively charged amino acids (D and E). Additionally, the frequency of aliphatic amino acids (A, L, V, I and G) was over half (55.1%), which probably was relevant to signal peptides through the plasma membrane.

It was reported that signal peptides of secreted proteins could be recognized and cleaved by four signal peptidase in most species, and then mature protein could be transported to different places through membranes[18]. Therefore, the frequencies of amino acid residues around the cleavage sites were analyzed (Table 1). The highest frequencies of amino acid residue both at -3 and -1 sites was A (45.5% at -3 site, 77.9% at -1 site). In addition to A, V (27.8%) and T (10.3%) were also well present at -3 site, and G (8.5%) and S (7.3%) at -1 site, which were similar to that in *Verticillium dahliae*[17], thus speculating that amino acid residues used at -3 and -1 sites were conservative in different organisms. Among 20 amino acid residues, D, E, F, H, M, N, P, Q, R, W, Y were hardly ever (<0.5%) or not used at -3 site, and D, E, F, H, I, K, L, M, N, Q, V, W and Y at -1 site. However, all 20 amino acid residues were used at -2 and +1 to +3 sites, and there were no relatively significant differences in the frequencies of 20 amino acid residues at the four sites compared to -3 and -1 sites. The amino acid residues at -3 to +3 sites were compared between *C. heterostrophus* and other organisms (Table 2). Alanine acid (A) was the highest frequency of amino acid residue at -3, -1 and +1 sites in *C. heterostrophus*, *Agrobacterium tumefaciens*[19], *Phytophthora infestans*[20], *Neurospora crassa*[21], *Verticillium dahliae*[17] and *Ctenopharyngodon idellus*[22], but not in *Caenorthaditis elegans*[16], proving above-mentioned speculation that amino acid residues at -3 and -1 sites were relatively conservative in most species. This conservation was important in the recognition and cleavage of signal peptidase.

Pathogenicity associated secreted proteins

Secreted proteins of pathogenic fungi play an important role in pathogen-host interaction, in which plant pathogenic fungi secrete many extracellular enzymes (cellulase, pectinase and cutinase) outside the cell to degrade cell walls and cuticles of host plants. By blast search against PHI database, 164 of 886 secreted proteins were

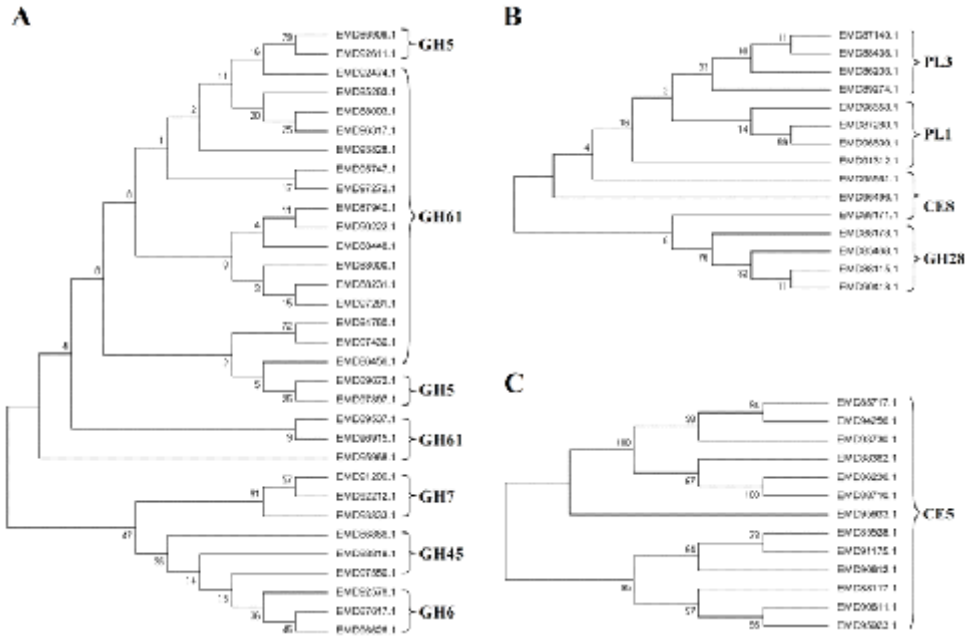

Fig. 2. Cellulases, pectinases and cutinases in secretome of *C. heterostrophus*. (A) Phylogenetic trees for cellulase containing GH5, GH6, GH7, GH45 and GH61 families of glucoside hydrolase. (B) Phylogenetic trees for pectinase containing GH28, CE8, PL1 and PL3 families. (C) Phylogenetic trees for cutinase containing only CE5 family.

involved in the pathogen-host interaction. Using CAZYmes Analysis Toolkit, we found that *C. heterostrophus* secretome contained 162 CAZYmes, over one third of which (61) were extracellular enzymes including 32 cellulases, 15 pectinases and 14 cutinases (Fig. 2). 32 cellulases contained 5 CAZYmes families: GH5 (4), GH6 (3), GH7 (3), GH45 (3) and GH61 (19) (Fig. 2A). 15 pectinases contained 4 CAZYmes families: GH28 (4), CE8 (3), PL1 (4) and PL3 (4) (Fig. 2B). 14 cutinases only contained one CAZYmes family CE5 (14) (Fig. 2C). The amounts of cellulases and pectinases in *C. heterostrophus* as well as *Magnaporthe grisea* (33 cellulases, 7 pectinases) and *V. dahliae* (29 cellulases, 28 pectinases) were much more than that in *S. cerevisiae* (no cellulase, 1 pecinase) and *U. maydis* (1 cellulase, 3 pectinases).

Small cytein-rich secreted proteins

Small cytein-rich secreted proteins (SCRSPs) of many plant pathogenic fungi had a variety of biological functions including pathogenicity[23], colonization[24] and host hypersensitive response (HR) induction[25-26]. It was noted that some SCRSPs as effectors interfered host recognizing pathogen and induce effector-triggered immunity of plant[27-28]. 147 SCRSPs were identified in *C. heterostrophus* secretome (Table S1), which would contribute to clarify the pathogenesis of *C. heterostrophus*. By search for conserved domain in 147 SCRSPs, 23 SCRSPs were mapped to conserved domains of CDB database (Table 3), of which three (EMD93471.1, EMD97201.1 and EMD90447.1) contained hydrophobin domains, three (EMD89003.1, EMD91522.1 and EMD90271.1) contained CFEM domains, and one contained cerato-platanin domain (EMD92796.1).

DISCUSSION

Secreted proteins containing signal peptides are synthesized within the cell, secreted out of the cell, and then play key roles in pathogenic process of pathogen. *C. heterostrophus* is an important corn pathogen. Thus, in this study, we analyzed secretome of *C. heterostrophus* by bioinformatics, which is facilitated by the genome sequencing of *C. heterostrophus*. Our analysis about the characteristics of secretome and pathogenicity associated secreted proteins contributed to further reveal pathogenicity mechanism of *C. heterostrophus*.

Although predictions of secreted proteins were not difficult, whole process was complex and needed several tools such as SignalP 4.1 Server, TMHMM v 2.0, Big-PI Predictor and TargetP 1.1. These tools were widely used in secreted proteins predictions of other organisms[17,29]. Thus, these tools were used together to predict secreted proteins of *C. heterostrophus* for the prediction veracity.

As a plant pathogenic fungus, *C. heterostrophus* secretes extracellular enzymes to degrade cuticles and cell walls of plant in order to overcome host passive defenses or absorb degraded cells for growth and development. Cellulose and pectin are main components, which can be degraded by cellulases and pectinase, respectively. In this study, we identified a lot of cutinases, cellulases and pectinase from secretome of *C. heterostrophus*, which was similar to another two pathogenic fungi (*M. grisea* and *V. dahliae*). *S. cerevisiae* hardly contained cellulase and petcinase[30], which could be explained that *S. cerevisiae* was not a plant pathogen. *U. maydis* kept symbiotic relationship with host plant after invading into host plant[31], which account for much less cellulases and petcinases in *U. maydis* than in *C. heterostrophus*[31], although *U. maydis* was a maize pathogen. These proved the important roles of cellulases and pectinases in the pathogenicity of *C. heterostrophus*.

It was well known that pathogens of plants directly secreted proteins such as effectors into host plant cells in order to suppress the plant innate immune system[32-33]. Effector protein was usually small in size and secretory, and had key function in the pathogen-host interaction[27-28]. Hydrophobins, small and cysteine-rich hydrophobic proteins, were assembled on the surface of hyphae and required as effectors in pathogen attaching to hydrophobic surfaces[34]. Cerato-platanin as elicitor was secreted into host cell to induce various defense responses of host including production of phytoalexins and cell death, so it was considered a pathogen-associated molecular pattern[25]. The CFEM domain contained eight cysteines and was fungal-specific extracellular membrane proteins[35] such as Pth11 of

Magnaporthe grisea, which played important roles in appressorium formation and fungal pathogenesis[36]. Therefore, hydrophobin, CFEM or cerato-platanin domains contained SCRSPs identified in *C. heterostrophus* also had key functions in pathogenesis and were served as important candidate proteins for the search of pathogen-host interaction mechanism. Additionally, the other SCRSPs without annotation were also putative candidate proteins involving in pathogenicity of *C. heterostrophus*, and thus it was necessary to further analyze them to determine their relationships with the pathogenicity of *C. heterostrophus*.

In conclusion, *C. heterostrophus* contained a large secretome, which were mainly rich in less 600 aa in sequence length. Although three was variation in the length of signal peptides, the lengths were normally distributed by quantity. Amino acid residues at C terminal of signal peptides were relatively conservative. Secretome of *C. heterostrophus* contained a great quantity of pathogen-host interaction associated gene encoding proteins, pathogenic extracellular enzymes and small cytein-rich proteins related to pathogenicity.

REFFERENCES

1. Schafer W. Molecular mechanisms of fungal pathogenicity to plants. *Annual review of Phytopathology*, 1994; **32**(1):461-477. doi:10.1146/annurev.py.32.090194.002333
2. Blobel G, Sabatini DD. Ribosome-membrane interaction in eukaryotic cells. *In: Biomembranes.* Springer, 1971; pp 193-195
3. Martoglio B, Dobberstein B. Signal sequences: more than just greasy peptides. *Trends Cell Biol*, 1998; **8**(10):410-415. doi:10.1016/S0962-8924(98)01360-9
4. Greenbaum D, Luscombe NM, Jansen R, Qian J, Gerstein M. Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. *Genome Res*, 2001; **11**(9): 1463-1468. doi:10.1101/gr.207401
5. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 2011; **8**(10):785-786. doi:10.1038/nmeth.1701
6. Chen Y, Yu P, Luo J, Jiang Y. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. Mammalian Genome, 2003; **14**(12):859-865. doi:10.1007/s00335-003-2296-6
7. Eisenhaber B, Bork P, Eisenhaber F. Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein engineering*, 1998; **11**(12):1155-1161. doi:10.1093/protein/11.12.1155
8. Eisenhaber B, Bork P, Yuan Y, Löffler G, Eisenhaber F. Automated annotation of GPI anchor sites: case study *C. elegans*. *Trends in biochemical sciences*, 2000; **25**(7):340-341. doi:10.1016/S0968-0004(00)01601-7
9. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, 2007; **2**(4): 953-971. doi:10.1038/nprot.2007.131
10. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Kohler J, Hammond-Kosack KE. PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res*, 2006; **34**: D459-464. doi:10.1093/nar/gkj047
11. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC. CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology*, 2010; **20**(12):1574-1584. doi:10.1093/glycob/cwq106
12. Graham MA, Silverstein KA, Cannon SB, VandenBosch KA. Computational identification and characterization of novel genes from legumes. *Plant Physiology*, 2004; **135** (3):1179-1197. doi:10.1104/pp.104.037531
13. Marcet-Houben M, Ballester AR, de la Fuente B, Harries E, Marcos JF, Gonzalez-Candelas L, Gabaldon T. Genome sequence of the necrotrophic fungus *Penicillium digitatum*, the main postharvest pathogen of citrus. *Bmc Genomics*, 2012; **13**: 646. doi:10.1186/1471-2164-13-646
14. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*, 2011; **39**: D225-229. doi:10.1093/nar/gkq1189
15. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, Hesse CN, Kosti I,

LaButti K, Lindquist EA, Lucas S, Salamov AA, Bradshaw RE, Ciuffetti L, Hamelin RC, Kema GH, Lawrence C, Scott JA, Spatafora JW, Turgeon BG, de Wit PJ, Zhong S, Goodwin SB, Grigoriev IV. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *Plos Pathog*, 2012; **8**(12): e1003037. doi:10.1371/journal.ppat.1003037

16. Hongzhi W, Chengyun L, Youyong Z. Computational analysis of signal peptide-dependent secreted protein in Caenorthaditis elegans ws123. *Hereditas*, 2006; 28 (4):470-478
17. Li T, Jieyin C, Xiangyong C, Jiani W, Xiaofeng D. Prediction and analysis of *Verticillium dahliae* VdLs.17 secretome. *Scientia Agricultura Sinica*, 2011; **44**(15):3142-3153. doi:10.3864/j.issn.0578-1752.2011.15.009
18. Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev*, 2000; **64**(3):515-547. doi:10.1128/MMBR.64.3.515-547
19. Chengming F, Chengyun L, Mingfu Z, Yueqiu H. Analysis of signal peptides of the secreted proteins in Agrobacterium tumefaciens C58. *Acta Microbiologica Sinica*, 2005; **45**(4):561-566
20. Zhou XG, Hou SM, Chen DW, Tao N, Ding YM, Sun ML, Zhang SS. Genome-wide analysis of the secreted proteins of *phytophthora infestans*. *Yi Chuan*, 2011; **33**(7):785-793.
21. Xiaogang Z, Chengyun L, Zhiwei Z, Yuan S, Shaosong Z, Jinbin L, Jing Y, Lin L, Yanfen Y. Analysis of the secreted proteins encoded by genes in genome of Filamental Fungus (*Neurospora crassa*). *Hereditas*, 2006; **28**(2): 200-207
22. Hanchang S, Fan Y, Jingming X, Wanan Y. Analysis of the secreted proteins in genome of signal peptides in grass carp *Ctenopharyngodon idellus*. *Fisheries Science*, 2011; 30 (3):164-167
23. Stephenson SA, Hatfield J, Rusu AG, Maclean DJ, Manners JM. CgDN3: an essential pathogenicity gene of colletotrichum gloeosporioides necessary to avert a hypersensitive-like response in the host Stylosanthes guianensis. *Mol Plant Microbe Interact*, 2000; **13**(9):929-941. doi:10.1094/MPMI
24. Rep M, van der Does HC, Meijer M, van Wijk R, Houterman PM, Dekker HL, de Koster CG, Cornelissen BJ. A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Mol Microbiol*, 2004; **53**(5):1373-1383. doi:10.1111/j.1365-2958.2004.04177.x
25. De Oliveira AL, Gallo M, Pazzagli L, Benedetti CE, Cappugi G, Scala A, Pantera B, Spisni A, Pertinhez TA, Cicero DO. The structure of the elicitor Cerato-platanin (CP), the first member of the CP fungal protein family, reveals a double psibeta-barrel fold and carbohydrate binding. *J Biol Chem*, 2011; **286**(20):17560-17568. doi:10.1074/jbc.M111.223644
26. Westerink N, Brandwagt BF, de Wit PJ, Joosten MH. *Cladosporium fulvum* circumvents the second functional resistance gene homologue at the Cf-4 locus (Hcr9-4E) by secretion of a stable avr4E isoform. *Mol Microbiol*, 2004; **54** (2):533-545. doi:10.1111/j.1365-2958.2004.04288.x
27. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeyer C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan S, Butler J, Calvo S, Ma LJ, Nicol R, Purcell S, Nusbaum C, Galagan JE, Birren BW. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, 2005; **434** (7036):980-986. doi:10.1038/nature03449
28. De Wit PJ, Mehrabi R, Van den Burg HA, Stergiopoulos I. Fungal effector proteins: past, present and future. *Mol Plant Pathol*, 2009; 10 (6):735-747. doi:10.1111/j.1364-3703.2009.00591.x
29. Lee SA, Wormsley S, Kamoun S, Lee AF, Joiner K, Wong B. An analysis of the Candida albicans genome database for soluble secreted proteins using computer-based prediction algorithms. *Yeast*, 2003; **20**(7): 595-610. doi:10.1002/yea.988
30. Jing Y, Chengyun L, Yunyue W, Youyong Z, Jinbin L, Xiahong H, Lin L, Yanfe Y, Xiaogang Z, Ping T. Computational analysis of signal peptide-dependent secreted proteins in *Saccaromyces cerevisiae*. *Scientia Agricultura Sinica*, 2005; **38**(3):516-522. doi:10.1016/S1671-2927(06)60042-4
31. Mueller O, Kahmann R, Aguilar G, Trejo-Aguilar B, Wu A, de Vries RP. The secretome of the maize pathogen *Ustilago maydis*. *Fungal Genet Biol*, 2008; **45**Suppl 1:S63-70. doi:10.1016/j.fgb.2008.03.012
32. Nimchuk Z, Eulgem T, Holt BF, 3rd, Dangl JL. Recognition and response in the plant immune system. *Annu Rev Genet*, 2003; **37**: 579-609. doi:10.1146/annurev.genet.37.110801.142628
33. Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. Direct interaction of resistance gene and avirulence gene products confers rice blast

resistance. *EMBO J*, 2000; **19** (15):4004-4014. doi:10.1093/emboj/19.15.4004

34. Elliot MA, Talbot NJ. Building filaments in the air: aerial morphogenesis in bacteria and fungi. *Curr Opin Microbiol*, 2004; **7**(6):594-601. doi:10.1016/j.mib.2004.10.013
35. Kulkarni RD, Kelkar HS, Dean RA. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. *Trends Biochem Sci*, 2003; **28**(3):118-121. doi:10.1016/S0968-0004(03)00025-2
36. DeZwaan TM, Carroll AM, Valent B, Sweigard JA. Magnaporthe grisea pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. *The Plant Cell Online*, 1999; **11**(10):2013-2030. doi:10.1105/tpc.11.10.2013.