

Chapter 15

Genome-Wide Identification and Analysis of Protease-Coding Genes in Pineapple



Xiang Jia Min, Ching Man Wai, Brian Powell, and Ray Ming

Abbreviations

FPKM Fragments per kilobase exon per million reads mapped
rpsBLAST Reversed position-specific BLAST

Introduction

Pineapple, *Ananas comosus* L., is an important tropical and subtropical fruit plant. It possesses a crassulacean acid metabolism (CAM) photosynthesis pathway, i.e., carbon dioxide is fixed as malate during the night, and the carbon is refixed using C3 pathway during the day (Bartholomew and Kadzimin 1997; Bartholomew and Malézieux 1994). Its fruit ripening process is known to be non-climacteric, i.e., ripening without ethylene and respiration bursts. Its fruits and juice can be consumed directly. Its products also include bromelain which was used as folk medicine and a culinary ingredient as a meat tenderizer (Taussig and Batkin 1988). Bromelain belongs to a group of protein-digesting enzymes obtained commercially from the fruit or stem of pineapple ((Pavan et al. 2012) for review). Stem bromelain is a mixture of different proteolytic enzymes and also contains other components, in small quantity, like phosphatase, glucosidase, peroxidase, cellulase, escharase, and several protease inhibitors (Heinicke and Gortner 1957). In vitro and in vivo studies

X. J. Min (✉)

Center for Applied Chemical Biology, Department of Biological Sciences, Youngstown State University, Youngstown, OH, USA

e-mail: xmin@ysu.edu

C. M. Wai · R. Ming

Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

B. Powell

Department of Computer Science and Information Systems, Youngstown State University, Youngstown, OH, USA

demonstrate that bromelain exhibits various fibrinolytic, antiedematous, antithrombotic, and anti-inflammatory activities (Pavan et al. 2012; Meiser et al. 2014).

Stem bromelain (EC 3.4.22.32) was isolated from the stem of pineapples and was characterized as a complex of cysteine proteases (Taussig and Batkin 1988; Arshad et al. 2014). Cysteine proteases are involved in various physiological and developmental processes in plants including programmed cell death during organ senescence and tracheary element differentiation ((van der Hoorn 2008) for review). Pineapple stem bromelain was found to inhibit fungal plant pathogens (López-García et al. 2012). Overexpression of the pineapple fruit bromelain gene (Accession number: BAA21849) in transgenic Chinese cabbage (*Brassica rapa*) results in enhanced resistance to bacterial soft rot (Jung et al. 2008). Another cysteine proteinase (AcCP2) gene from pineapple fruit was found to be gradually increased in its expression during fruit development until maturity, and the overexpression of the gene in *Arabidopsis thaliana* improved the resistance to fungal pathogen of *Botrytis cinerea* (Wang et al. 2014). The interaction of fruit bromelain (EC 3.4.22.33) and cystatins, inhibitors of cysteine proteinases, regulates the fruit development and ripening and directly correlates to the resistance to blackheart development in pineapple fruits during postharvest chilling stress (Neuteboom et al. 2009; Raimbault et al. 2013a). A minor form of pineapple stem cysteine protease, named as ananain (EC 3.4.22.31), was isolated and found that it differs from stem and fruit bromelains in being inhibited by chicken cystatin (Carter et al. 2000; Lee et al. 1997). Recently a novel aspartic acid protease gene (AcAP1) from pineapple fruit was characterized and found the gene was involved in postharvest chilling stress resistance in pineapple fruits (Raimbault et al. 2013b). One common feature among all these characterized pineapple enzymes including bromelains, cystatins, and aspartic acid protease, we noticed, is that they are all classic secreted enzymes, i.e., possessing a secretory signal peptide.

Recently the complete pineapple genome was sequenced and annotated (Ming et al. 2015). A total of 27,024 protein-coding genes from the genome were predicted. The complete gene set provides an unprecedented opportunity for genome-wide identification of all genes encoding proteolytic enzymes from pineapple. Identifying these genes and proteins will facilitate further characterization of these genes and their products in the processes of pineapple growth, development, and fruit storage, as well as expression and characterization of recombinant proteins for commercial applications. The effort for expressing recombinant protein has been made with pineapple stem bromelain (Amid et al. 2011). Recently the pineapple stem bromelain was successfully expressed in a prokaryotic system and proved that the recombinant bromelain protein had the bactericidal property (George et al. 2014).

Knowing the subcellular locations of proteins would help to elucidate the protein cellular functions. Recently we have constructed the PlantSecKB, the plant secretome and subcellular proteome knowledgebase (<http://proteomics.yzu.edu/secretomes/plant/index.php>), using all available plant protein sequences (over 1.4 million entries) in UniProtKB (Lum et al. 2014) and the protein sequences predicted from the newly sequenced genome *Nelumbo nucifera* (Gaertn.) (sacred lotus) and pineapple (Ming et al. 2015; Lum et al. 2013). The secreted proteolytic enzymes are

potential targets for commercial applications (González-Rábade et al. 2011). The work described in this chapter was primarily based on our recent report (Wai et al. 2016a) with additional analysis and mapping to the MEROPS classification of proteases (Rawlings et al. 2015).

Identification of Genes Encoding Proteolytic Enzymes

To identify protease sequences in pineapple and perform comparative analysis with other plant species, we used the whole set of 27,024 predicted pineapple protein sequences and the complete proteome data in 3 other plant species including *Oryza sativa* (subsp. *japonica*) (rice), *Sorghum bicolor* (sorghum), and *Brachypodium distachyon* (purple false brome). The proteome data were downloaded from the UniProt database (UniProt (UniProt Consortium 2015)). The protein sequences were used to search against the Conserved Domain Database (CDD) using locally installed rps-BLAST program with a cutoff E-value of $1e-5$. The conserved domain database contains multiple datasets including PFAM, SMART, COGs collection, TIGRFAMs, NCBI's in-house data curation effort, etc. (Marchler-Bauer et al. 2015). We retrieved the protein domain families having the terms of peptidase, protease, or proteinase (Barrett and McDonald 1986) and then excluded entries having a term of "inhibitor."

The total genes encoding proteases vary in different monocot species. The total numbers of identified protease-coding genes were 512 in pineapple, 546 in rice, 535 in sorghum, and 646 in *Brachypodium* (Wai et al. 2016a). Except pineapple, there are proteases identified for the 3 monocot plants in the MEROPS database, which identified 1,395 in rice, 549 in sorghum, and 794 in *B. distachyon* ((Rawlings et al. 2015); <https://merops.sanger.ac.uk/index.shtml>). The proteases in rice were identified from both *japonica* and *indica* subspecies; thus the number of proteases was twice more than those we identified as we only used the data from *japonica* rice. The slight differences in protease numbers identified in sorghum and *B. distachyon* might be caused by differences in the datasets or the identification methods. We used the batch BLAST server in the MEROPS website to map the proteases we identified to the protease categories classified by the MEROPS database (Rawlings and Morton 2008). The classified numbers of proteases were 448 in pineapple, 475 in rice, 469 in sorghum, and 563 in *Brachypodium*. Thus we regenerated the protease files by adding the mapped MEROPS protease family or subfamily category to the protease list. The raw MEROPS BLAST results with predicted active site residues and some other intermediate data files can be downloaded (http://proteomics.yasu.edu/publication/data/Pineapple/book_chapters/protease/).

The results for MEROPS protease families or subfamilies having at least five members or more in at least one species were summarized in Table 15.1. The species associated with each enzyme category was obtained using BLASTP search. The enriched protease families in these monocot plants were nepenthesin in aspartic peptidase clan, papain family in cysteine peptidase clan, and subtilisin family and

Table 15.1 Distributions of protease families in pineapple and other three monocot plants

Clan	(Sub) family	Pineapple	Rice	Sorghum	<i>Brachypodium</i>	Enzyme name
Clans of aspartic peptidases						
AA	A1B	63	43	58	43	Nepenthesin (<i>Nepenthes gracilis</i>)
AD	A22B	5	1	7	7	Impas 1 peptidase (<i>Homo sapiens</i>)
Clans of cysteine peptidases						
CA	C1A	62	54	43	42	Papain (<i>Carica papaya</i>)
CA	C12	2	5	5	6	Ubiquitinyl hydrolase-L1 (<i>Homo sapiens</i>)
CA	C19	22	31	22	29	Ubiquitin-specific peptidase 14 (<i>Homo sapiens</i>)
CA	C65	6	6	6	7	Otubain-1 (<i>Homo sapiens</i>)
CA	C85A	10	12	9	19	OTLD1 deubiquitylating enzyme (<i>Homo sapiens</i>)
CD	C14B	5	9	12	12	Metacaspase Yca1 (<i>Saccharomyces cerevisiae</i>)
CE	C48	14	16	32	22	Ulp1 peptidase (<i>Saccharomyces cerevisiae</i>)
CP	C97	8	10	10	14	DeSI-1 peptidase (<i>Mus musculus</i>)
Clans of metallo peptidases						
MA	M1	5	4	4	8	Aminopeptidase N (<i>Homo sapiens</i>)
MA	M10A	2	6	4	4	Matrix metallopeptidase-1 (<i>Homo sapiens</i>)
MA	M41	12	5	6	12	FtsH peptidase (<i>Escherichia coli</i>)
ME	M16B	10	9	7	12	Mitochondrial processing peptidase beta-subunit (<i>S. cerevisiae</i>)
ME	M16C	6	1	1	1	Eupitrylsin (<i>Homo sapiens</i>)
MG	M24A	8	8	5	14	Methionyl aminopeptidase 1 (<i>Escherichia coli</i>)
MG	M24B	2	5	3	3	Aminopeptidase P (<i>Escherichia coli</i>)
MH	M20D	3	7	9	8	Carboxypeptidase Ss1 (<i>Sulfolobus solfataricus</i>)
MH	M28B	2	3	3	5	Glutamate carboxypeptidase II (<i>Homo sapiens</i>)
unassigned	M79	5	5	5	7	RCE1 peptidase (<i>Saccharomyces cerevisiae</i>)
Clans of serine peptidases						
SB	S8A	58	67	62	63	Subtilisin Carlsberg (<i>Bacillus licheniformis</i>)

(continued)

Table 15.1 (continued)

Clan	(Sub) family	Pineapple	Rice	Sorghum	<i>Brachypodium</i>	Enzyme name
SC	S10	33	53	55	77	Carboxypeptidase Y (<i>Saccharomyces cerevisiae</i>)
SC	S28	4	6	8	7	Lysosomal Pro-Xaa carboxypeptidase (<i>Homo sapiens</i>)
SC	S9A	5	5	4	6	Prolyl oligopeptidase (<i>Sus scrofa</i>)
SC	S9C	2	4	2	6	Acylaminoacyl-peptidase (<i>Homo sapiens</i>)
SF	S26A	5	8	9	11	Signal peptidase I (<i>Escherichia coli</i>)
SF	S26B	3	3	3	6	Signalase 21 kDa component (<i>Saccharomyces cerevisiae</i>)
SK	S14	19	10	9	15	Peptidase Clp (<i>Escherichia coli</i>)
Clans of mixed (C, S, T) catalytic type						
PA	S1C	10	23	18	29	DegP peptidase (<i>Escherichia coli</i>)

carboxypeptidase family in serine peptidase clan (Table 15.1). Particularly, pineapple proteome has more nepenthesin and papain peptidases than other monocot species. Pineapple bromelain was characterized as a complex of cysteine proteases ((Taussig and Batkin 1988; de Lencastre Novaes et al. 2016) for review). One pineapple aspartic acid protease gene (AcAP1) was recently examined and found that the gene was involved in postharvest chilling stress resistance in pineapple fruits (Raimbault et al. 2013b). It should be noticed that when we used PFAM data, we found pineapple proteome has 15 members of RVP_2 (single domain retroviral aspartyl protease), while there is no protein having this domain in other species. Thus the origin of RVP_2 proteins and their functional significance in pineapple is worthy further examined. Pineapple has only 1 member of Trypsin_2 (trypsin-like peptidase domain), while there are 17–22 members in other species (see Wai et al. (2016a)).

We would like to point out that using MEROPS family classification and Conserved Domain Database (CDD) including PFAM generated slightly different results for the proteases. For example, entries which were classified as the cd05472 (cnd41-like; chloroplast nucleoids DNA-binding protease) using CDD were classified as A1B and nepenthesin in the clan of aspartic peptidases using MEROPS classification. However, even the naming is different in the two classification methods; the functionality of the protease remains similar. Chloroplast nucleoids DNA-binding protease, catalyzing the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase, is classified as a pepsin-like aspartic protease. Thus the protease lists, available for downloading, including both CDD rpsBLAST results and the MEROPS classification, will help to predict the functionality of the identified proteases.

Proteases are located in different subcellular locations to play their biological roles. Among the proteases, secreted proteases accounted for a large proportion. There were 152 in pineapple-, 161 in rice-, 218 in sorghum-, and 222 in *Brachypodium*-secreted proteases. The large secreted protease families are papain family cysteine protease (Peptidase_C1), peptidase S8 family (Peptidases_S8_3), provisional aspartyl protease family (PLN03146), serine carboxypeptidase (Peptidase_S10), and chloroplast nucleoids DNA-binding protease (cnd41_like) (Wai et al. 2016a). The most noticeable difference is the number of secreted aspartyl protease family; there are 29 members in pineapple and only 6–7 members in other species. Pineapple proteases were also localized in other subcellular compartments including 74 in cytosol, 60 in chloroplast, 67 in nucleus, and 18 in mitochondria, and a small number of proteases in other various locations.

RNA-seq Expression Analyses of Genes Encoding Proteases

RNA-seq technology is replacing microarray technology for genome-wide gene expression profiling. We presented some detailed analysis of gene expression in a previous report (Wai et al. 2016b) and also a chapter in the book (Wai et al. 2018). In this section we focus on the expression analysis of protease-coding genes. There were a total of 12,768 genes expressed having ≥ 10 FPKM (fragments per kilobase exon model per million reads mapped) in at least 1 tissue. Based on at least twofold changes of the $\log_2(\text{FPKM})$ values, pairwise tissue gene expression comparison revealed that 3,502 of them were differentially expressed in different tissues and 532 of them were predicted to encode secreted proteins. Comparing gene expression in young fruits with ripening fruits revealed that in the ripening fruits, 931 genes were upregulated including 167 of them encoding secreted proteins, and 189 were downregulated with 20 of them encoding secreted proteins. These upregulated genes encoded diverse families of enzymes including lipase, glycosyl hydrolases, proteases, peroxidases, and others, which were involved in fruit ripening process (Wai et al. (2016a, b); Wai et al. 2018).

Among 512 genes encoding putative proteases in pineapple, 159 were lowly expressed with < 10 FPKM in all tissues, 273 were moderately or highly expressed (having ≥ 10 FPKM in at least 1 tissue), and 80 genes were not detected. Within the 273 moderately or highly expressed genes encoding proteases, there are 71 secreted proteins and 42 proteins localized in cytosol, 42 proteins in chloroplast, 32 proteins in nucleus, 10 proteins in mitochondria, 7 proteins in ER, and the remaining in other subcellular compartments or the locations not predicted (Wai et al. 2016a). In addition, we also found that 72 protease-coding genes were differentially expressed in different tissues. We also found that among the proteins encoded by the 273 expressed protease genes, there were 25 papain family cysteine proteases (pfam00112, Peptidase_C1; C1A family), 18 serine carboxypeptidase (pfam00450, Peptidase_S10), 11 having peptidase S8 family domain (cd04852, Peptidases_S8_3), 10 chloroplast nucleoids DNA-binding proteases (cd05472, cnd41_like), 9 OTU-like cysteine proteases (pfam02338, OUT), etc. (Wai et al. 2016a).

The pineapple protease genes having >100 FPKM were listed with information of predicted subcellular locations and MEROPS protease classification (Table 15.2). The top 9 highly expressed genes, having >400 FPKM, included seven genes encoding papain family cysteine protease (C1A subfamily), i.e., bromelain, with six of them encoding secreted cysteine proteases (Table 15.2). Even among these seven bromelain genes, they had different expression levels, and some showed differential expression among different tissues. For example, Aco004356 gene encoding a cysteine protease was extremely highly expressed in both young fruit and ripening fruit with >7000 FPKM while also highly expressed in flower and leaf tissues (>4000 FPKM), relatively modestly expressed in root tissue (292 FPKM). Aco017317 gene was expressed lowly in young fruit (18 FPKM), moderately expressed in root (190 FPKM), and highly expressed (>1000 FPKM) in leaf, flower, and ripening fruit tissues. Aco004355 gene was highly expressed (>1000 FPKM) in all tissues except root tissues (only 2 FPKM), while Aco013510 was highly expressed (>1000 FPKM) in all tissues sampled (Table 15.2).

Molecular Phylogenetic Analyses of Papain Family Proteins

Pineapple bromelain proteins belong to papain family (C1A) within the cysteine protease clan (CA). We identified 62 members in pineapple genome belonging to papain family. We further performed phylogenetic analysis of the papain family protein sequences. Two protein sequences (Aco031698 and Aco009865) were removed as they were too short to be aligned properly. Plant papain-like cysteine proteases had been subclassified into nine subfamilies (Richau et al. 2012). We obtained the plant papain-like cysteine protease sequences collected by Richau et al. (2012) (from Dr. R. A. L. van der Hoorn, corresponding author) and added pineapple papain protease sequences to the dataset and reconstructed the phylogenetic tree using neighbor-joining method with MEGA6 (Tamura et al. 2013). The bootstrap consensus tree is shown as Fig. 15.1. The subclassification of pineapple papain proteins was based on the phylogenetic analysis with other plant papain-like protease protein sequences that were subclassified previously (Richau et al. 2012) (Fig. 15.1). Our reconstructed phylogenetic tree and subclassification were overall consistent with the tree and subclassification as reported by Richau et al. (2012), except CEPs and CatBs that were mixed together in our tree (Fig. 15.1). To provide a clearer picture of their phylogenetic relationships among pineapple cysteine proteases, we aligned these 60 protein sequences using MUSCLE program, and the evolutionary history was inferred by using the maximum likelihood method based on the JTT matrix-based model implemented in the MEGA6 tool (Tamura et al. 2013; Edgar 2004; Jones et al. 1992) (Fig. 15.1). Pineapple papain proteases formed four larger clusters (Fig. 15.2). Three clusters belong to AALPs/RD19s, TH1s, and PAPS/SAG12s subfamilies, respectively (Fig. 15.2). The fourth cluster contains the rest subfamilies which are distributed into different smaller clusters (Figs. 15.1 and 15.2). The biochemical properties and physiological roles of these papain proteases in each subfamily need to be examined in the future.

Table 15.2 Gene expression levels of proteases in different tissues of pineapple

ID	Leaf	Flower	Root	Fruit1_5	Fruit6_8	SubLoc	MEROPS	PFAM
Aco004356	4917	4756	292	9749	7462	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco017317*	2748	2746	190	18	1840	E	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco013510	1777	1676	1236	1444	1968	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco004355*	1046	1164	2	2301	1616	-	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco028299*	649	747	0	119	551	N	-	cd08012, M20_ArgE-related, M20 peptidases with similarity
Aco011776	602	612	167	121	443	S	S10	pfam00450, Peptidase_S10, serine carboxypeptidase
Aco011478	442	403	529	467	419	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco021005	408	403	372	187	312	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco017299*	402	399	0	2	271	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco024880*	395	369	82	12	259	S	A1B	cd05472, cnd41_like, chloroplast nucleoids DNA-binding protease
Aco007211*	382	314	23	3	233	Y	C14B	pfam00656, Peptidase_C14, caspase domain
Aco018123	365	433	378	468	406	Y	S33	TIGR01249, Putative_proline_iminopeptidase, proline iminopeptidase
Aco018849*	322	355	0	72	263	-	M20A	cd08012, M20_ArgE-related, M20 peptidases with similarity to
Aco019014	296	298	378	876	859	-	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco028298*	272	262	1	42	218	-	M20A	cd08012, M20_ArgE-related, M20 peptidases with similarity to
Aco017558*	262	244	2	0	169	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco002588	218	224	345	239	215	Y	C15	cd00501, Peptidase_C15, pyroglutamyl peptidase (PGP) type I
Aco013996*	206	220	34	3	146	S	S10	pfam00450, Peptidase_S10, serine carboxypeptidase
Aco003873	197	206	118	120	164	M	M16B	COG0612, Pqql, predicted Zn-dependent peptidases
Aco021295	183	187	14	23	133	S	S10	pfam00450, Peptidase_S10, serine carboxypeptidase
Aco000819*	178	173	0	3	117	S	A1B	cd05472, cnd41_like, chloroplast nucleoids DNA-binding protease
Aco008566	171	159	206	129	135	-	S26A	TIGR02227, Inactive_signal_peptidase_IA, signal peptidase I
Aco012022*	164	156	0	0	107	S	S8A	cd04852, Peptidases_S8_3, peptidase S8 family domain
Aco012896	159	163	61	28	115	Y	M20A	cd08012, M20_ArgE-related, M20 peptidases
Aco005969	148	151	110	121	128	M	M16B	COG0612, Pqql, predicted Zn-dependent peptidases

ID	Leaf	Flower	Root	Fruit1_5	Fruit6_8	SubLoc	MEROPS	PFAM
Aco012854	143	142	45	33	104	Y	C97	pfam05903, Peptidase_C97, PPPDE putative peptidase domain
Aco005846*	143	149	604	21	142	S	C13	pfam01650, Peptidase_C13, peptidase C13 family
Aco004358*	137	131	3	39	92	Y	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco004359	133	123	92	12	85	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco019689	126	130	83	84	108	-	M17	PRK00913, PRK00913, multifunctional aminopeptidase A; provisional
Aco010227	126	148	103	154	123	Y	M1	cd09600, M1_APN_1, peptidase M1 family containing aminopeptidase
Aco014573*	123	113	21	10	83	N	C97	pfam05903, Peptidase_C97, PPPDE putative peptidase domain
Aco003144	122	132	30	28	92	P	S41A	PLN00049, PLN00049, carboxyl-terminal processing protease
Aco009319	121	123	165	130	102	T	A22B	pfam04258, Peptidase_A22B, signal peptide peptidase.
Aco011810	120	134	84	37	92	C	S9A	COG1505, COG1505, serine proteases of the peptidase family S9A
Aco012864	117	121	145	143	134	-	C12	cd09616, Peptidase_C12_UCH_L1_L3, cysteine peptidase C12
Aco027767	111	124	366	222	216	T	C1A	cd02620, Peptidase_C1A_CathepsinB, cathepsin B group
Aco012310	110	123	78	145	116	Y	M24B	cd01085, APP, X-Prolyl Aminopeptidase 2.E.C. 3.4.11.9
Aco018144*	110	108	12	6	74	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco016160	109	120	15	30	85	Y	-	COG0542, clpA, ATP-binding subunits of Clp protease, and DnaK/DnaJ
Aco011447	107	117	15	92	77	-	S8A	cd04852, Peptidases_S8_3, peptidase S8 family domain
Aco010611*	106	118	33	2	75	S	S10	pfam00450, Peptidase_S10, serine carboxypeptidase
Aco006954*	101	101	3	1	68	N	C97	pfam05903, Peptidase_C97, PPPDE putative peptidase domain
Aco028297*	101	109	3	27	75	-	M20A	cd08012, M20_ArgE-related, M20 peptidases
Aco005269	96	103	45	37	77	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco008796	95	103	93	56	74	T	A22B	pfam04258, Peptidase_A22B, Signal peptide peptidase
Aco012083	97	102	49	70	85	C	S10	pfam00450, Peptidase_S10, serine carboxypeptidase
Aco025188*	1	1	483	0	1	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco001368*	1	0	411	0	0	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco006487*	5	5	121	198	21	S	A1B	cd05472, cnd41_like, chloroplast nucleoids DNA-binding protease

(continued)

Table 15.2 (continued)

ID	Leaf	Flower	Root	Fruit1_5	Fruit6_8	SubLoc	MEROPS	PFAM
Aco014769	27	25	118	13	20	–	A1B	PLN03146, PLN03146, aspartyl protease family protein; provisional
Aco008415	27	22	116	151	85	N	C85A	pfam02338, OTU, OTU-like cysteine protease.
Aco005786	64	69	113	79	65	C	C19	cd02657, Peptidase_C19A, a subfamily of peptidase C19.
Aco026097	84	76	101	57	76	N	C14B	pfam00656, Peptidase_C14, caspase domain
Aco007611	56	60	101	80	58	Y	M24A	PLN03158, PLN03158, methionine aminopeptidase; provisional
Aco018520	0	0	100	2	0	Y	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco004901	92	71	62	294	112	S	S8A	cd04852, Peptidases_S8_3, peptidase S8 family domain
Aco012554	70	77	88	190	84	T	–	cd02125, PA_VSR, PA_VSR; protease-associated (PA) domain
Aco010208	80	91	60	161	81	N	C85A	pfam02338, OTU, OTU-like cysteine protease.
Aco004360	73	68	15	160	91	S	C1A	pfam00112, Peptidase_C1, papain family cysteine protease
Aco013115	60	65	87	129	76	Y	S16	TIGR00763, Lon_protease, endopeptidase La.
Aco023270	56	59	73	106	81	P	–	pfam06703, SPC25, microsomal signal peptidase 25 kDa subunit
Aco011236	54	54	67	104	68	S	–	cd02123, PA_C_RZF_like, PA_C_RZF_like; protease-associated (PA)
Aco018628	36	36	43	104	54	C	S14	cd07017, S14_ClpP_2, Caseinolytic protease (ClpP)

The list contains the protease-coding genes having an expression level of >100 FPKM. The numbers are the normalized FPKM (fragments per kilobase exon per million reads mapped) values of each sample. Differentially expressed genes are labelled with * symbol. Subcellular location notation – C chloroplast, E ER, G Golgi, K cytoskeleton, M mitochondrial, N nuclear, O peroxisome, P plasma membrane, S secreted, T membrane, V vacuole, Y cytoplasm

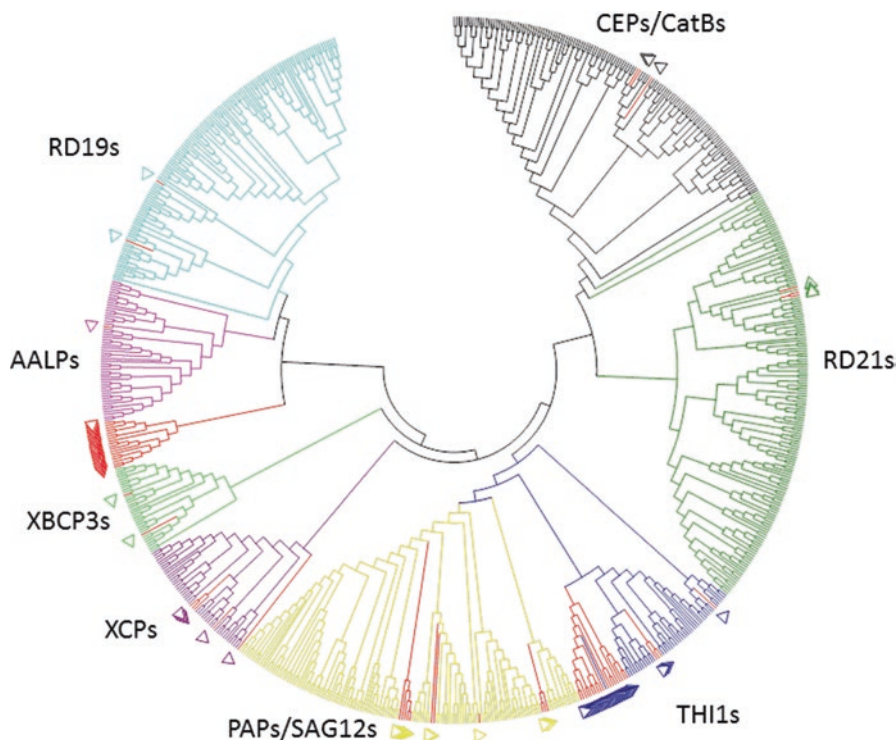


Fig. 15.1 Molecular phylogenetic analyses of plant papain-like protease sequences by neighbor-joining method. Eight subfamilies of plant papain-like proteins are labelled in different colors. Pineapple proteins were highlighted in red color, and their subtypes were indicated with the colored triangle. The pineapple protein clusters with red triangles were similar to both AALPs and RD19s

We would like to point out that there were a total of 14 bromelain or bromelain-like pineapple protein sequences collected in the UniProt database, prior to the pineapple genome sequences available; 2 fragment sequences (only 20 amino acids) were not further analyzed. Among the remaining 12 bromelain or bromelain-like sequences, 5 of them were annotated as “bromelain,” and 7 of them were annotated with other terms. Among the five bromelain proteins, two of them (F1KD58 and P14518) were annotated as stem bromelain, one (O23791) as fruit bromelain, and other two (O24641, Q7DNA3) only annotated as bromelain. F1KD58, O23791, and O24641 were predicted to be secreted, Q7DNA3 to be located in nucleus, and P14518 having no subcellular location predicted, probably due to incompleteness in its sequence. The other bromelain-like sequences with their annotations are O23800 (FB22, precursor), F8UN02 (cysteine peptidase), O81085 (cysteine proteinase, AN11, precursor), O23801 (FB1035), O23799 (FBSB, precursor), O81084 (cysteine proteinase, AN8, precursor), and P80884 (ananain, AN1). All of them were predicted to be secreted proteins, except the subcellular location of O23801 that was

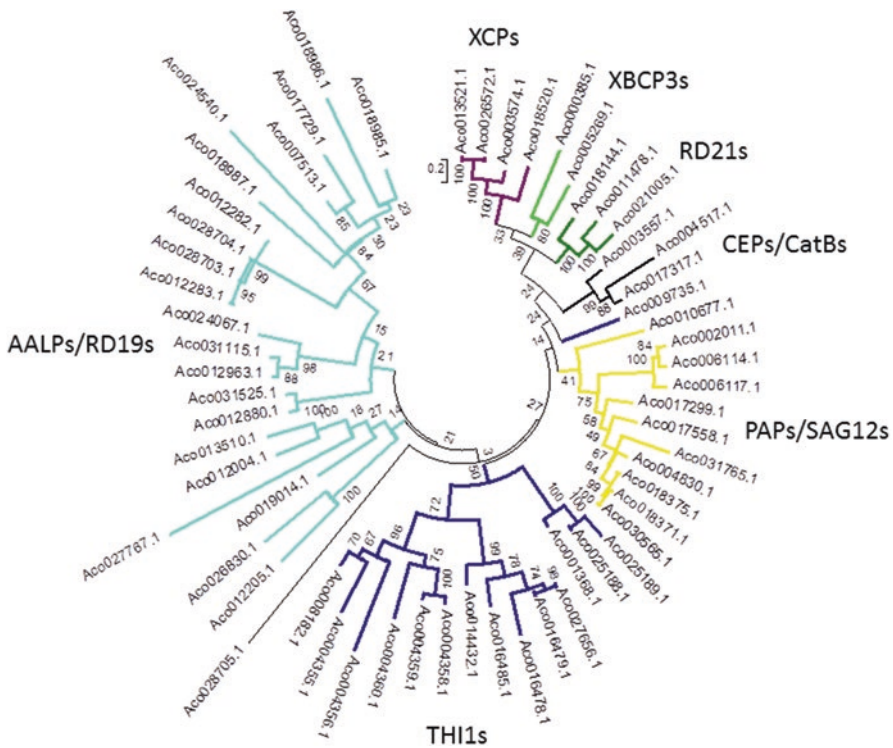


Fig. 15.2 Molecular phylogenetic analyses of pineapple papain (bromelain) family proteins of cysteine proteases by maximum likelihood method. The numbers are the bootstrap values. The subfamilies are labelled in different colors

not predicted due to incompleteness in its sequence. The information of the predicted subcellular locations of the pineapple proteins can be found in PlantSecKB (Lum et al. 2014).

We performed a phylogenetic analysis of the 12 previously characterized bromelain sequences with 60 cysteine proteases which we identified from the genome. These 12 bromelain sequences formed a cluster with Aco004356, Aco004355, and Aco008182, belonging to the subfamily of THI1s (Fig. 15.3). Both Aco004356 and Aco004355 were highly expressed in fruit tissues, leaf, and flower but only weakly or moderately expressed in roots (Table 15.2). Aco008182 were not expressed in all tissues. Though there was no distinct structural feature in this subfamily noted by Richau et al. (2012), however, they represented the previously characterized fruit and stem bromelain proteins before the genome was sequenced; thus this subfamily of cysteine proteases may be worthy of further detailed examination for exploring commercial applications.

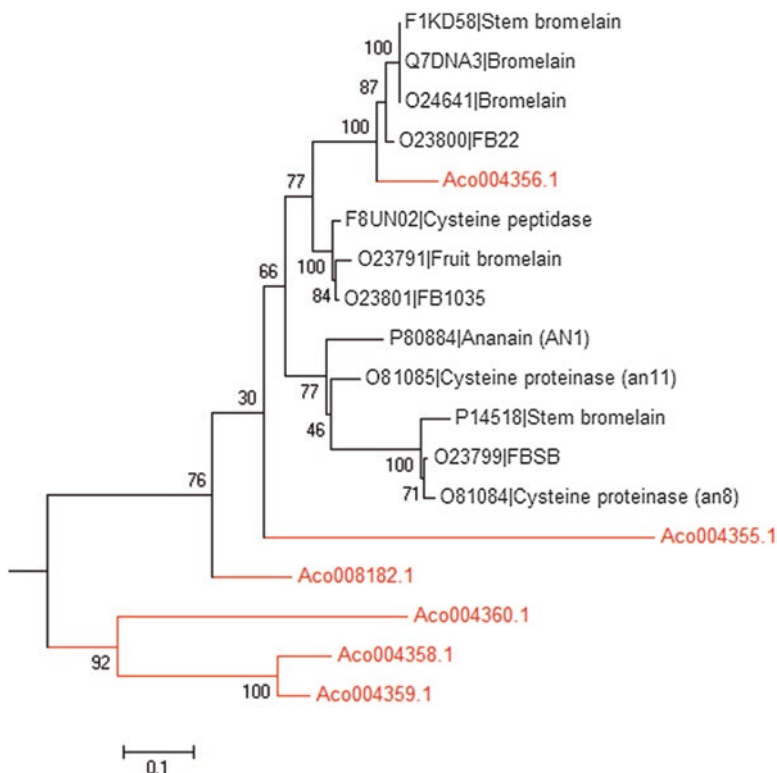


Fig. 15.3 A subtree shows the phylogenetic relationship of characterized bromelain proteins with cysteine protease sequences predicted from pineapple genome by maximum likelihood method. The numbers are the bootstrap values. Predicted pineapple protease sequences from the genome are labelled in red color

Concluding Remarks

Plant proteases are key regulators of a variety of biological processes, including meiosis, gametophyte survival, embryogenesis, seed coat formation, cuticle deposition, epidermal cell fate, stomata development, chloroplast biogenesis, and local and systemic defense responses (van der Hoorn 2008). Pineapple bromelains can be used for medical treatments (Meiser et al. 2014); also proteases were involved in fruit ripening (Neuteboom et al. 2009; Raimbault et al. 2013a, b). Comparative analysis with other monocot species including rice, sorghum, and *Brachypodium* revealed that, though pineapple had a relatively smaller proteome, several protease families including papain family (C1A) and aspartyl protease family (A1B) were expanded (Table 15.1). Plastids, mitochondria, and peroxisomes are key organelles in plants, and the plant organellar proteases play important roles in degrading unwanted proteins within organelles (van Wijk 2015). Secreted proteases are likely involved in cell wall structure remodeling and plant defense (Lum and Min 2011).

The predicted secreted and organellar proteases in our analysis will be useful for further elucidating their biological roles (Wai et al. 2016a). Thus the work along with the protein subcellular location information, which is available in PlantSecKB (Lum et al. 2104), is expected to provide a resource to facilitate further designing experiments to characterize the protease genes and proteins to understand their biological roles in pineapple and explore the potential for commercial applications.

Acknowledgments The work was supported by the University of Illinois at Urbana-Champaign to RM and the Youngstown State University (YSU) Research Professorship award to XJM.

References

- Amid A, Ismail NA, Yusof F, Salleh HM (2011) Expression, purification, and characterization of a recombinant stem bromelain from *Ananas comosus*. *Process Biochem* 46:2232–2239
- Arshad ZIM, Amid A, Yusof F et al (2014) Bromelain: an overview of industrial application and purification strategies. *Appl Microbiol Biotech* 98:7283–7297
- Barrett AJ, McDonald JK (1986) Nomenclature: protease, proteinase and peptidase. *Biochem J* 237:935
- Bartholomew DP, Kadzimin SB (1997) Pineapple. In: Alvin PT, Kozeowski TT (eds) *Ecophysiology of tropical crops*. Academic Press, New York, NY, pp 113–156
- Bartholomew DP, Malézieux EP (1994) Pineapple. In: Schaffer B, Andersen PC (eds) *Handbook of environmental physiology of fruit crops*, vol 2. CRC Press, Boca Raton, pp 243–291
- Carter CE, Marriage H, Goodenough PW (2000) Mutagenesis and kinetic studies of a plant cysteine proteinase with an unusual arrangement of acidic amino acids in and around the active site. *Biochemistry* 39:11005–11013
- de Lencastre Novaes LC, Jozala AF, Lopes AM et al (2016) Stability, purification, and applications of bromelain: a review. *Biotechnol Prog* 32:5–13
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5(1):1
- George S, Bhasker S, Madhav H, Nair A, Chinnamma M (2014) Functional characterization of recombinant bromelain of *Ananas comosus* expressed in a prokaryotic system. *Mol Biotechnol* 56:166–174
- González-Rábade N, Badillo-Corona JA, Aranda-Barradas JS et al (2011) Production of plant proteases in vivo and in vitro—a review. *Biotechnol Adv* 29:983–996
- Heinicke RM, Gortner WA (1957) Stem bromelain - a new protease preparation from pineapple plants. *Econ Bot* 11:225–234
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282
- Jung YJ, Choi CS, Park JH et al (2008) Overexpression of the pineapple fruit bromelain gene (BAA) in transgenic Chinese cabbage (*Brassica rapa*) results in enhanced resistance to bacterial soft rot. *Electron J Biotechnol* 11:71–79
- Lee KL, Albee KL, Bernasconi RJ, Edmunds T (1997) Complete amino acid sequence of ananain and a comparison with stem bromelain and other plant cysteine proteases. *Biochem J* 327(Pt 1):199–202
- López-García B, Hernández M, Segundo BS (2012) Bromelain, a cysteine protease from pineapple (*Ananas comosus*) stem, is an inhibitor of fungal plant pathogens. *Lett Appl Microbiol* 55:62–67
- Lum G, Min XJ (2011) Plant secretomes: current status and future perspectives. *Plant Omics* 4:114–119

- Lum G, Vanburen R, Ming R, Min XJ (2013) Secretome prediction and analysis in sacred lotus (*Nelumbo nucifera* Gaertn.). *Trop Plant Biol* 6:131–137
- Lum G, Meinken J, Orr J, Frazier S, Min XJ (2014) PlantSecKB: the plant secretome and subcellular proteome knowledgebase. *Comput Mol Biol* 4(1):1–17
- Marchler-Bauer A, Derbyshire MK, Gonzales NR et al (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue):222–226
- Meiser P, Xu Z, Kirsch G, Jacob C (2014) Systemic enzyme therapy: fact or fiction? a review with focus on bromelains, proteolytic enzymes from the pineapple plant. In: *Recent advances in redox active plant and microbial products*. Springer, The Netherlands, pp 449–467
- Ming R, VanBuren R, Wai CM et al (2015) The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet* 47:1435–1442
- Neuteboom LW, Matsumoto KO, Christopher DA (2009) An extended AE-rich N-terminal trunk in secreted pineapple cystatin enhances inhibition of fruit bromelain and is post-translationally removed during ripening. *Plant Physiol* 151:515–527
- Pavan R, Jain S, Kumar A (2012) Properties and therapeutic application of bromelain: a review. *Biotech Res Int* 2012:6. <https://doi.org/10.1155/2012/976203>
- Raimbault AK, Zuily-Fodil Y, Soler A et al (2013a) The expression patterns of bromelain and AcCYS1 correlate with blackheart resistance in pineapple fruits submitted to postharvest chilling stress. *J Plant Physiol* 170:1442–1426
- Raimbault AK, Zuily-Fodil Y, Soler A et al (2013b) A novel aspartic acid protease gene from pineapple fruit (*Ananas comosus*): cloning, characterization and relation to postharvest chilling stress resistance. *J Plant Physiol* 170:1536–1540
- Rawlings ND, Morton FR (2008) The MEROPS batch Blast: a tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie* 90:243–259
- Rawlings ND, Barrett AJ, Finn R (2015) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 44(Database):343–350
- Richau KH, Kaschani F, Verdoes M et al (2012) Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant Physiol* 158:1583–1599
- Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evo* 30:2725–2729
- Taussig SJ, Batkin S (1988) Bromelain, the enzyme complex of pineapple (*Ananas comosus*) and its clinical application: an update. *J Ethnopharmacol* 22:191–203
- UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res* 43(Database issue):204–212
- van der Hoorn RA (2008) Plant proteases: from phenotypes to molecular mechanisms. *Annu Rev Plant Biol* 59:191–223
- van Wijk KJ (2015) Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol* 66:75–111
- Wai CM, Powell B, Ming R, Min XJ (2016a) Genome-wide identification and analysis of genes encoding proteolytic enzymes in pineapple. *Trop Plant Biol* 9(3):161–175
- Wai CM, Powell B, Ming R, Min XJ (2016b) Analysis of alternative splicing landscape in pineapple (*Ananas comosus*). *Trop Plant Biol* 9(3):150–160
- Wai CM, Powell B, Ming R, Min XJ (2018) Analysis of transcriptome and alternative splicing landscape in pineapple. In *Plant Genetics, Genomics, Vol. 22*, Ray Ming (Eds): *Genetics and Genomics of Pineapple*. Springer Nature.
- Wang W, Zhang L, Guo N et al (2014) Functional properties of a cysteine proteinase from pineapple fruit with improved resistance to fungal pathogens in *Arabidopsis thaliana*. *Molecules* 19:2374–2389