Genome-wide identification of oil palm *(Elaeis guineensis* Jacq.) chitinases and their response to *Ganoderma boninense* Pat. infection

IOP Conf. Series: Earth and Environmental Science **293** (2019) 012033 doi:10.1088/1755-1315/293/1/012033

Genome-wide identification of oil palm (*Elaeis guineensis* Jacq.) chitinases and their response to *Ganoderma boninense* Pat. infection

R Aditama[1], R Tryono[2], Z A Tanjung[1], C Utomo[3,*] and T Liwang[4]

[1]Section of Bioinformatics, Plant Production and Biotechnology Division, PT. SMART Tbk, Bogor, 16810, Indonesia
[2]Section of Genetic Engineering, Plant Production and Biotechnology Division, PT. SMART Tbk, Bogor, 16810, Indonesia
[3]Department of Biotechnology, Plant Production and Biotechnology Division, PT. SMART Tbk, Bogor, 16810, Indonesia
[4]Plant Production and Biotechnology Division, PT SMART Tbk, Bogor, 16810, Indonesia

*Corresponding author: biotechnology@sinarmas-agri.com

Abstract. The basal stem rot disease caused by *Ganoderma boninense* Pat. is a serious problem in oil palm plantation. Naturally, plant cells produce chitinases to protect themselves from chitin-containing parasites, particularly fungi. This study employed the advance of sequencing technology and availability of oil palm genomic and transcriptomic data to identify all chitinases in oil palm genome. Homology study using the combination of BLAST and HMM methods successfully identified 35 chitinases in oil palm which comprised of 22 GH18 and 13 GH19 family members. A multiple sequence alignment method classified those chitinases into class I, II, III, IV, V, VII and an unknown class. Chitinase domains analysis against PFAM databases showed more than half of chitinases possess partial domains of GH 18 or GH 19 in their sequences. Transcriptome analysis revealed that most of GH 18 family members were expressed specifically in roots under *G. boninense* infection treatment. Differently, most of GH 19 members were expressed constitutively in any tissues and under biotic stress conditions. This suggests that some GH 18 members might play an important role during host defence mechanism
Keywords: Basal stem rot, defence mechanism, fungal parasite, oil palm genome, protein homology, RNA-sequencing.

1. Introduction

African origin oil palm (*Elaeis guineensis* Jacq.) is one of the most important oil-bearing crops in the world with a capability to produce up to 12 t oil per hectare annually [1]. However, fungal disorders may impede this productivity. Basal stem rot, caused by *Ganoderma boninense* Pat. and Fusarium wilt, caused by *Fusarium oxysporum* f. sp. *elaedis* have become serious problems on oil palm in Southeast Asia and Africa region respectively [2]. Many efforts were performed to control the

diseases, including chemical treatments, biological controls and screening for resistant planting materials, but all harboured varies results [3].

Naturally, plant cells produce chitinases to protect themselves from chitin-containing parasites, particularly fungi [4]. The chitinase enzyme (poly[1,4-N-acetyl-β-D-glucosaminide] glycanhydrolase, E.C. 3.2.1.14) hydrolyses the chitin polymer to release N-acetyl glucosamine oligomers, following either endo or exo cleavages of the β-1,4 bonds. The presence of chitinases have been reported in various monocotyledonous and dicotyledonous plant species from widely different tissues [5]. Various stress conditions, such as biotic and abiotic inducing agents, induce their expressions in the cell [6].

Whole-genome analysis of carbohydrate-related enzymes including chitinases are available for *Arabidopsis thaliana* (L.) Heynh and *Oryza sativa* L. (rice) [7]. The first mentioned species genome contains 25 chitinases encoding genes, eleven of them belong to GH18 while the remaining members belong to GH19 family. Gene expression profile derived from Geneinvestigator databases (https://www.geneinvestigator.com) showed two of them (AT4G19829 from GH18 and AT1G05850 from GH19) were expressed constitutively at various stages of plant development. In rice, the genome contains 49 chitinase genes consisting of 33 GH18 and 16 GH19 family members. Remarkably, almost of rice chitinases were expressed constitutively at various stages of development [8]. In oil palm, genomics and proteomics profiles of several chitinases were studied. Increased activity of oil palm chitinases was detected after the pathogenic as well as non-pathogenic fungi treatments [9].

A quantitative PCR study of three oil palm chitinases genes (*EgCHI1*, *EgCHI2* and *EgCHI3*) following *G. boninense* and *Trichoderma harzianum* Rifai treatments showed various expression patterns on different tissue and treatment types [10]. At 5 wk post infection, *G. boninense* treatment increased the relative abundance of *EgCHI1*, *EgCHI2* and *EgCHI3* chitinases in the root up to eight folds change. Another study sequenced three oil palm chitinases full length cDNA (*EgChit3-1*, *EgChit1-1* and *EgChit5-1*) and quantified their expressions using quantitative reverse-transcription (qRT)-PCR. The result showed that the expression of *EgChit3-1* and *EgChit5-1* were increased (up to 4.5 folds change) in root samples after treated with *G. boninense* [11].

In author knowledge to date, there is no study reports the comprehensive genome-wide insight of chitinases in oil palm due to sequence information, protein structure, phylogenetics and gene expression profile. Fortunately, complete genome sequence of oil palm is available in the National Center for Biotechnology Information (NCBI) GenBank database [12]. Recently, RNA sequencing analysis were conducted to study the transcriptome profile of host-fungal interaction in oil palm at the nursery level [13]. It provides an insight of the molecular interaction between oil palm and its pathogen, *G. boninense*. *In vitro* infection *G. boninense* in oil palm culture can be used to determine the detailed molecular mechanism of oil palm host-fungal interaction without interference of other microbiomes. This study aims to characterize the protein sequence of oil palm chitinases and their transcript expression during infection of *G. boninense*.

2. Materials and methods

2.1. Chitinases identification

An oil palm protein RefSeq database was downloaded from NCBI GenBank under accession no GCF_000442705.1. This was used as the reference to identify all possible chitinases using two approaches. The first approach utilized chitinase sequences of *A. thaliana* and *O. sativa* as query for sequence similarity search employing BLAST+ tool from the NCBI [14]. The second approach was based on the probabilistic inference to find chitinases in oil palm protein database. The protein sequences alignment of Glycosyl hydrolases family 18 and 19 were downloaded from Pfam databases (https://pfam.xfam.org) and used as profile to perform Hidden Markov Model (HMM) search method employing HMMER software [15].

2.2. Protein domain and phylogenetic analysis

Pfam server was used to identify all conserved domains in all analysed protein using sequence alignment and hidden Markov models. A sequence was considered as chitinase if it possesses either GH18 or GH19 family domain. Further, the alignment process of chitinases were performed using Geneious software version 10.2 [16] based on the cost matrix Blosum62. Phylogenetic tree was built based on these alignments using the Neighbour-Joining method [17] with Jukes-Cantor genetic distance model [18].

2.3. Transcript expression analysis

The transcription level of all chitinases in oil palm was analysed using open access RNA-seq raw datasets that were deposited in the NCBI Sequence Reads Archive (SRA) (table 1). All RNA-seq reads were mapped to oil palm nucleotide RefSeq using BWA MEM software [19]. The mapped reads were extracted and sorted using SAMtools [20]. Transcript abundance of each chitinases was calculated using eXpress high throughput quantification software (http://bio.math.berkeley.edu/eXpress).

Table 1. Oil palm RNA-seq datasets derived from SRA NCBI.

SRA ID	Tissue	Treatment	Platform
ERS542767	Root seedling	Normal	Illumina HiSeq 2000 PE
ERS542764	Root seedling	Ganoderma	Illumina HiSeq 2000 PE
ERS542765	Root seedling	Trichoderma	Illumina HiSeq 2000 PE
ERS542766	Root seedling	Ganoderma + Trichoderma	Illumina HiSeq 2000 PE
SRS420943	Root seedling	Normal	454 GS FLX
SRS420978	Root mature	Normal	454 GS FLX
SRS420977	Shoot apex	Normal	454 GS FLX
SRS190309	Leaf young	Normal	454 GS FLX
DRS039961	Leaf dura	Normal	Illumina HiSeq 2000 PE
DRS039962	Flower male	Normal	Illumina HiSeq 2000 PE
DRS039960	Flower female	Normal	Illumina HiSeq 2000 PE
DRS039971	Whole Fruit	Normal	Illumina HiSeq 2000 PE

3. Results and discussion

3.1. Chitinase identification

Protein sequence of oil palm chitinases were identified using a combination of two approaches. The first approach utilized chitinase protein sequences from *A. thaliana* and *O. sativa* to perform sequence similarity search against oil palm RefSeq database employing BLASTP software. Previously, 26 and 47 chitinase protein sequences of both *A. thaliana* and *O. sativa* respectively was downloaded based on the study of Grover [8]. This approach resulted 19 homologous protein sequences in oil palm with sequence identity ranged between 37 % to 75 %. All proteins exceeded a minimum homology threshold score at 30 % [21]. An amount 14 detected chitinases were homolog with *O. sativa* chitinases while only nine were homolog with *A. thaliana*.

The second approach identified oil palm chitinase protein sequences using probabilistic method employing HMMER software version 3.1b2. Plant chitinases comprise of two families i.e. 18 and 19 glycosyl hydrolase [22]. Protein sequence alignments of Glycosyl hydrolases family 18 and 19 domains were downloaded from Pfam database under protein ID number of PF00704 and PF00182, respectively. These sequence alignments were used to build HMM profile for chitinases identification. Both hmmsearch and hmmscan algorithm were used in this approach with an e-value threshold of 10^{-5} and the results were merged. In hmmsearch, protein sequences acted as a query and HMM profile as a

subject. Different with hmmsearch, hmmscan method employing HMM profile as query and protein sequences as a subject. Combination of these methods identified 24 and 13 chitinases from family GH18 and GH19, respectively.

The data from first and second approaches were combined and used to further analysis. All of chitinases identified using first approach were also detected in second approach but not *vice-versa* (figure 1). Most of chitinases identified in oil palm were GH18 family and derived from HMM approach. Only less than half (11 of 24) of GH18 family that identified by HMM were also detected by BLAST method. Different with GH18 family, more than half (8 of 13) GH19 family chitinases that identified by HMM were also detected by BLAST method. Surprisingly, only 1 out of 13 GH18 family chitinase detected by BLAST method were derived from both *O. sativa* and *A. thaliana*. This indicate that GH18 family members are more diverse in sequence compared to GH19.

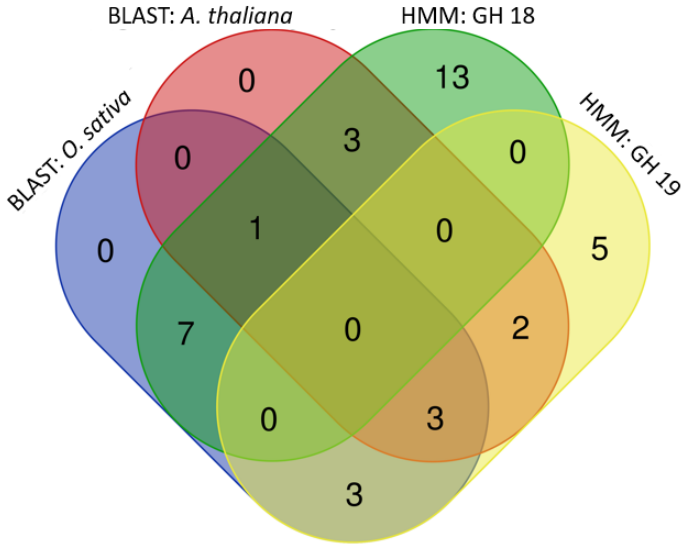

Figure 1. A venn diagram of oil palm chitinases identification.

3.2 Classification of oil palm chitinases

Chitinases are classified into seven classes (class I to VII) based on their amino acid similarities, substrate specificities, catalytic mechanisms and sensitivity to inhibitor [23]. This study utilized chitinases sequence from other plant species as references to identify classification of oil palm chitinases based on sequence similarity. Previously, multiple sequence alignment was performed to chitinases of oil palm combined with other plant species (table 2). Subsequently, the distance matrix was created from the alignment and visualized using multi-dimensional scaling (MDS) method (figure 2).

This analysis separated all chitinases into four clusters. First cluster contains chitinases from GH 19 family (class I, II, IV and VII) while second and third cluster contain GH 18 family chitinases (class V and III, respectively). Surprisingly, the fourth cluster which located between the second and third cluster contains 11 oil palm chitinases without any counterparts in reference chitinases. From whole oil palm chitinases, three were identified as class I, one as class II, eight as class III, five as class IV, three as class V and four as class VII. Eleven oil palm sequences were not identified as any class of chitinases since there were no significant similarity found from their sequences. There was no sequence identified as class VI chitinases of oil palm identified from this analysis.

Table 2. List of chitinases from other plant species.

ID	Species	Class
3IWR_A	*O. sativa* L. japonica	I
AAA56786	*Hordeum vulgare* L.	I
ABD47583	*Musa x paradisiaca* L.	I
AAB96341	*Solanum tuberosum* L.	II
AAC36359	*Capsicum annuum* L.	II
AAC37395	*Cucumis sativus* L.	III
AF082284_1	*Cucurbita moschata* Duch.	III
CAA49998	*Cicer arietinum* L.	III
CAA77656	*Nicotiana tobacum* L.	III
CAC14016	*Vitis vinifera* L.	III
AAB65776	*V. vinifera* L.	IV
AAP03085	*Galega orientalis* Lam.	IV
CHI4_BRANA	*Brassica napus* L.	IV
AF498100_1	*Momordica charantia* L.	V
AGI_URTDI	*Urtica dioica* L.	VI
AAP80801	*Gossypium hirsutum* L.	VII
AAQ56598	*G. hirustum* L.	VII
AAQ56599	*G. hirustum* L.	VII
ACD93719	*Mikania micrantha* Kunth	VII

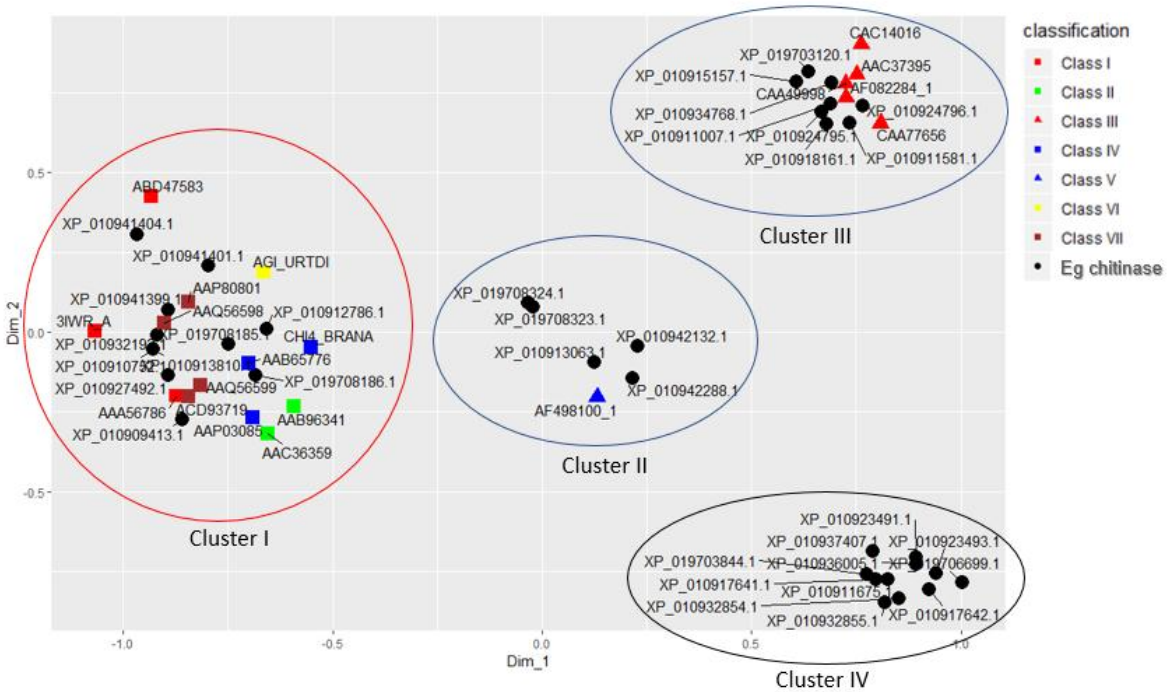

Figure 2. Multidimensional scaling of oil palm chitinases classification based on protein sequence similarity. This analysis separated all chitinases into four clusters. Cluster I contained chitinases from GH 19 family, cluster II contained class V chitinases, cluster III contained class III chitinases, cluster IV contained oil palm chitinases with no reference contained.

3.3. Phylogenetic and protein domain analysis

Phylogenetic analysis using the neighbour-joining method with Jukes-Cantor genetic distance model revealed the evolutionary relationship of oil palm chitinases, resulted four main branches of phylogenetic tree that separated GH18 from GH19 family (figure 3). This result was consistent with previous analysis using sequence similarity method. Oil palm chitinases from GH19 family (class I, class II, class IV and class VII) were clustered in a single branch while class III and V were separated into two distinct branches. Chitinases with no matching classification were clustered in a single branch with the class III chitinases as nearest cluster.

Further analysis identified the position and completeness of conserved domains from each sequence using Pfam server. All of class I and class II chitinases possess the complete domain of GH19 family while four chitinases from class IV and two from class VII possess the partial domain. The sequences with partial domain were also observed for GH18 family which occurred in two of class III chitinases and all of unclassified chitinases. Based on the sequence context, partial domains can be divided into three types: (i) split domain—single domains that have been broken into several parts by the HMM alignment process; (ii) bounded partials—domains that are bounded by other non-homologous domains or the ends of the proteins; (iii) unbounded partials—those that appear to be partial but are found in a region of protein that could contain a more complete domain [24].

Based on above definition, two partial domains in class IV chitinases are unbounded partials while the two others are bounded partials. Class VII chitinases possess two partial domains, which all of them were unbounded. All partials observed in class III and unclassified chitinases are unbounded. It was found that bounded partial domains are over-represented in eukaryotes and in lower quality protein predictions. This suggest that they often result from inaccurate genome assembly or gene models. It was also found that a large percentage of unbounded partial domains produce long alignments, which indicate that their annotation as a partial is an alignment artefact [24]. This indicate that partial domains occurred in oil palm chitinases should be viewed with cautions.

Domain analysis also identified chitin binding domain (CBD) as carbohydrate-binding module (PFAM: PF00187) in class I and class IV oil palm chitinases. This domain binds to insoluble chitin, soluble chitin, cellulose and *N*-acetylchitohexaose to increase chitinase activity. A deleted CBD results about 50 % reduction of the hydrolysing activity towards the insoluble chitin substrates [25]. Another structural features of chitinases analysed in this study was the secretory signal peptide. It is a protein-sorting signal that targets its passenger protein for translocation across the endoplasmic reticulum membrane in eukaryotes and the cytoplasmic membrane in prokaryotes [26]. Computational analysis employing SignalP server [27] revealed that almost all oil palm chitinases possess this signal peptide except for XP_010924796.1 and XP_010911675.1. This suggest that almost all chitinases in oil palm act as extracellular enzyme and do not have intracellular function.

Sequence similarity showed that EgCHI1, EgCHI2 and EgCHI3 from previous study matched with chitinases XP_010927492.1, XP_010909413.1 and XP_010924796.1 respectively. Although previously EgCHI1 was identified as member of class I chitinases [10], this study surprisingly found that this protein belongs to class VII chitinases. In fact, there was no chitin binding domain found in XP_010927492.1 sequences as required for classification of class I chitinases. However, classification of EgCHI2 and EgCHI3 were consistent with current study. Another previous study cloned full-length cDNA sequences encoding a putative chitinases (EgChit3-1) and two chitinase-like proteins (EgChit1-1 and EgChit5-1) [11]. Sequence alignment identified EgChit1-1, EgChit3-1 and EgChit5-1 as XP_010910752.1, XP_010924796.1 and XP_010913063.1, respectively. EgChit1-1 previously identified as chitinase-like protein because it lacks of both the highly conserved N-terminal cysteine-rich CBD and the short proline-rich variable hinge region. However, current study revealed that this sequence belongs to class VII chitinases because it more similar with AAP80801 (class VII chitinase of *Gossypium hirustum* L.). Classification of EgChit3-1 and EgChit5-1 from study by Yeoh et al. [11] were similar with current result.

3.4. Transcript expression analysis

RNA-seq is a developed approach to profile DNA transcription using deep sequencing technology that provides a far more precise measurement of level of transcripts than any other methods [28]. The analysis results provided detailed insight of chitinases gene expressions from oil palm roots under *G. boninense* infections and from oil palm tissues under normal conditions (figure 4). This analysis revealed the different expression pattern between GH18 and 19 families of chitinases. GH 18 family chitinases were expressed mostly in root tissues both on normal and stress conditions. XP_010924796.1 was the most expressed class III chitinases. It was expressed in any conditions of root tissues but was not significantly expressed in any other tissues. There were no significant expressions observed from class V chitinases except the expression of XP_010913063.1 on root seedling under normal condition.

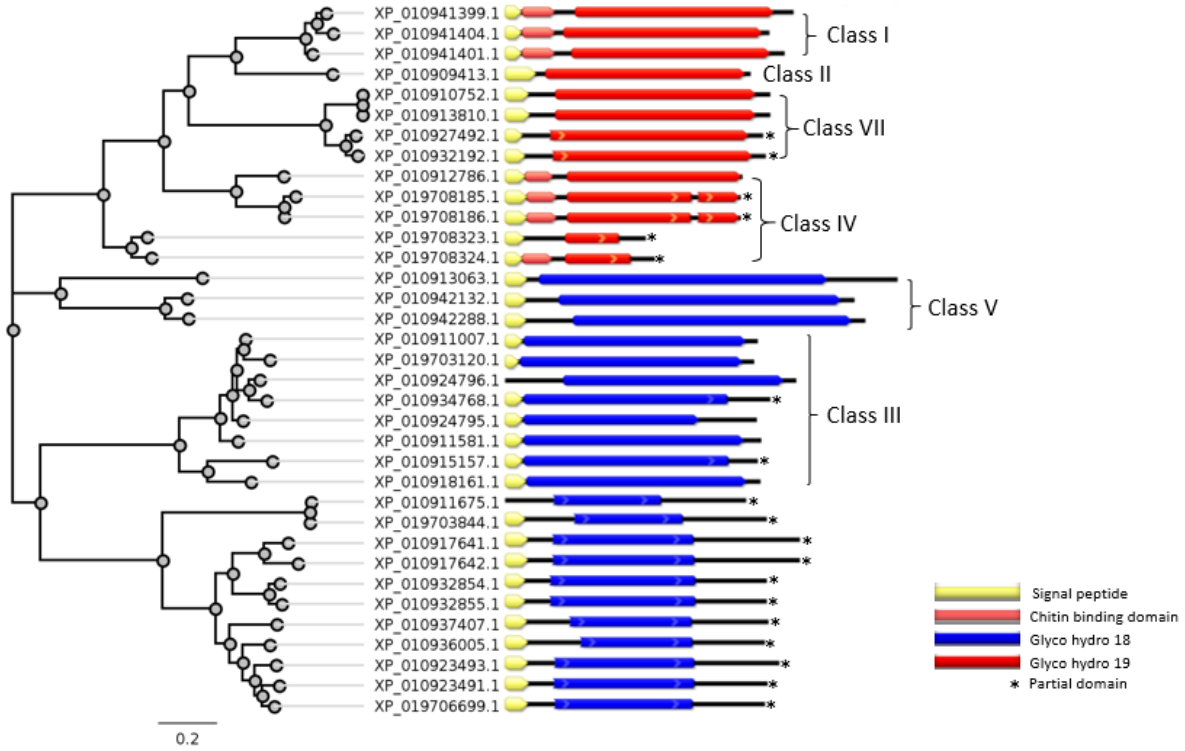

Figure 3. Phylogenetic tree, domain structure and classification of oil palm chitinases.

Unlike chitinases from GH18 family, most of chitinases from GH19 family were constitutively expressed in any tissues under normal or stress conditions. Chitinases from class I were low expressed in root tissues but high expressed in seedling, shoot, leaves and flowers. Unlike class I, class II chitinases showed no significant expressions under any tissues and conditions. XP_019708324.1 from class IV chitinases showed gradually increased expression in root tissues under artificial infection of *G. boninense* after one- and three-weeks post inoculation. Two other class IV chitinases, XP_019708185.1 and XP_019708186.1 showed similar expression pattern and were highly expressed in female flower tissue. Surprisingly, unlike other class IV which possessed partial GH19 domain, XP_010912786.1 which possess the complete domain of GH19 showed no significant expression under any tissues and conditions. Chitinases from class VII showed different expression pattern regarding completeness of GH 19 domains. Class VII chitinases which have complete domains were only expressed in root tissues while chitinases with partial domain were expressed in almost all tissues and conditions. Most of unclassified chitinases showed no significant expression under any conditions

except XP_010923491.1 and XP_0109706699.1 which were expressed in root tissues under normal and stress conditions.

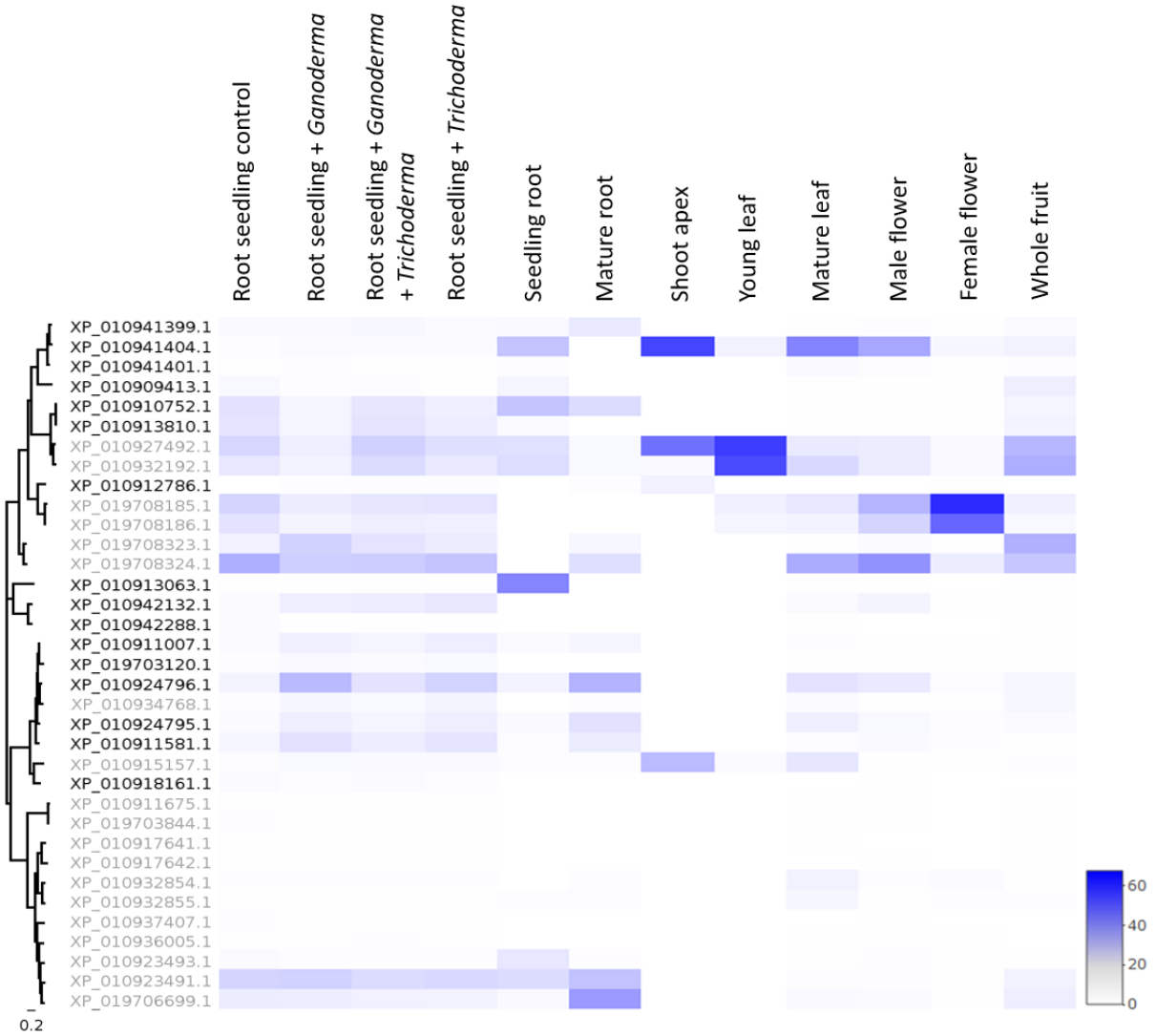

Figure 4. Chitinases relative expression of oil palm root under *G. boninense* infection and other tissues under normal conditions. Grey text indicating partial domain of GH18 or 19.

4. Conclusion

Employing the advance of DNA sequencing technology and the availability of oil palm genomic and genomic databases, this study successfully identified 35 chitinases which comprised of 22 and 13 GH18 and 19 family members, respectively. These chitinases were classified into class I, II, III, IV, V, VII and a group with unknown classifications. Detailed analysis of protein sequences revealed the position and completeness of conserve domains in each chitinase sequence. The 19 chitinases were identified possess only partial conserve domains and should be viewed with cautions. Transcript analysis revealed the difference between GH18 and GH19 chitinases. Most of GH18 family members were only expressed in roots under normal condition and *G. boninense* infection. In other hand, almost all of chitinases from GH19 family were expressed constitutively in any tissues and stress conditions.

This suggest that different functions of GH18 and GH19 family of oil palm chitinases in term of defence mechanism.

Reference

[1] Woittiez L S, van Wijk M T, Slingerland M, van Noordwijk M and Giller K E 2017 Yield gaps in oil palm: A quantitative review of contributing factors *Eur. J. Agron.* **83** 57–77. https://www.sciencedirect.com/science/article/pii/S1161030116302131

[2] Corley R H V and Tinker P B 2015 *The Oil Palm* (Chichester, UK: John Wiley & Sons) https://onlinelibrary.wiley.com/doi/book/10.1002/9781118953297

[3] Hushiarian R, Yusof N A and Dutse S W 2013 Detection and control of *Ganoderma boninense*: Strategies and perspectives *Springerplus* 2(555) 1–12 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3824713/

[4] Funkhouser J D and Aronson Jr N N 2007 Chitinase family GH18: Evolutionary insights from the genomic history of a diverse protein family *BMC Evol. Biol.* **16** 1–16 https://www.ncbi.nlm.nih.gov/pubmed/17594485

[5] Sharma N, Sharma P K, Gaur R K and Gupta V K 2011 Role of chitinase in plant defence *Asian J. Biochem.* **6**(1) 29–37 https://scialert.net/fulltextmobile/?doi=ajb.2011.29.37

[6] Kasprzewska A 2003 Plant chitinases–regulation and function *Cell. Mol. Biol. Lett.* **8**(3) 809–24 https://www.ncbi.nlm.nih.gov/pubmed/12949620

[7] Yokoyama R and Nishitani K 2004 Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and Arabidopsis *Plant Cell Physiol.* **45**(9) 1111–21 https://www.ncbi.nlm.nih.gov/pubmed/15509833

[8] Grover A 2012 Plant chitinases: Genetic diversity and physiological roles *CRC. Crit. Rev. Plant Sci.* **31**(1) 57–73 https://www.tandfonline.com/doi/abs/10.1080/07352689.2011.616043

[9] Naher L, Tan S G, Yusuf U K, Ho C L and Siddiquee S 2012 Activities of chitinase enzymes in the oil palm (*Elaeis guineensis* Jacq.) In interactions with pathogenic and non-pathogenic fungi *Plant Omics.* **5**(4) 333–36 https://www.semanticscholar.org/paper/Activities-of-chitinase-enzymes-in-the-oil-palm-in-Naher-Tan/78a6bb5ac0352a58a9859d12834ac508a3c57123

[10] Naher L, Ho C L, Tan S G, Yusuf U K and Abdullah F 2011 Cloning of transcripts encoding chitinases from *Elaeis guineensis* Jacq. and their expression profiles in response to fungal infections *Physiol. Mol. Plant Pathol.* **76**(2) 96–103 https://www.sciencedirect.com/science/article/pii/S0885576511000609

[11] Yeoh K A, Othman A, Meon S, Abdullah F and Ho C L 2013 Sequence analysis and gene expression of putative oil palm chitinase and chitinase-like proteins in response to colonization of *Ganoderma boninense* and *Trichoderma harzianum Mol. Biol. Rep.* **40**(1) 147–58 https://www.ncbi.nlm.nih.gov/pubmed/23065213

[12] Singh R et al. 2013 Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds *Nature* **500**(7462) 335–39 https://www.nature.com/articles/nature12309

[13] Ho C L, Tan Y C, Yeoh K A, Ghazali A K, Yee W Y and Hoh C C 2016 De novo transcriptome analyses of host-fungal interactions in oil palm (*Elaeis guineensis* Jacq.) *BMC Genomics* **17**(66) 1–19 https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2368-0

[14] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden T L 2009 BLAST+: architecture and applications *BMC Bioinformatics* **10** 421 https://www.ncbi.nlm.nih.gov/pubmed/20003500
[15] Finn R D, Clements J and Eddy S R 2011 HMMER web server: Interactive sequence similarity searching *Nucleic Acids Res.* **39**(2) 1–9 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3125773/
[16] M. Kearse et al. 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data *Bioinformatics* **28**(12) 1647–49 https://www.ncbi.nlm.nih.gov/pubmed/22543367
[17] Saitou N and Nei M 1987 The neighbour-joining method: a new method for reconstructing phylogenetic trees *Mol. Biol. Evo.* **4**(4) 406–25 https://www.ncbi.nlm.nih.gov/pubmed/3447015
[18] Jukes T H and Cantor C 1969 *Evolution of Protein Molecules Mammalian Protein Metabolism* Vol. III ed H N Munro (Cambridge: Academic Press) pp 21–132 https://www.sciencedirect.com/book/9781483232119/mammalian-protein-metabolism
[19] Li H and Durbin R 2009 Fast and accurate short read alignment with Burrows-Wheeler transform *Bioinformatics* **25**(14) 1754–60 https://www.ncbi.nlm.nih.gov/pubmed/19451168
[20] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R 2009 The sequence alignment/map format and SAMtools *Bioinformatics* **25**(16) 2078–79 https://www.ncbi.nlm.nih.gov/pubmed/19505943
[21] Pearson W R 2013 An introduction to sequence similarity ('homology') searching *Curr. Protoc. Bioinforma.* **03**(10) 1286–92 https://www.ncbi.nlm.nih.gov/pubmed/19728288
[22] Li H and Greene L H 2010 Sequence and structural analysis of the chitinase insertion domain reveals two conserved motifs involved in chitin-binding *PLoS One* **5**(1) e8654 https://www.ncbi.nlm.nih.gov/pubmed/20084296
[23] Singh A, Kirubakaran S I and Sakthivel N 2007 Heterologous expression of new antifungal chitinase from wheat *Protein Expr. Purif.* **56**(1) 100–09 https://www.ncbi.nlm.nih.gov/pubmed/17697785
[24] Triant D A and Pearson W R 2015 Most partial domains in proteins are alignment and annotation artifacts *Genome Biol.***16**(1) 1–12 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4443539/
[25] Itoh Y, Kawase T, Nikaidou N, Fukada H, Mitsutomi M, Watanabe T and Itoh Y 2002 Functional analysis of the chitin-binding domain of a family 19 chitinase from Streptomyces griseus HUT6037: Substrate-binding affinity and cis-dominant increase of antifungal function *Biosci. Biotechnol. Biochem.* **66**(5) 1084–92 https://www.ncbi.nlm.nih.gov/pubmed/12092819
[26] Redei G P 2008 *Encyclopedia of Genetics, Genomics, Proteomics, and Informatics* (Berlin, Germany: Springer) https://www.springer.com/la/book/9781402067532
[27] Petersen T N, Brunak S, von Heijne G and Nielsen H 2011 SignalP 4.0: Discriminating signal peptides from transmembrane regions *Nat. Methods* **8**(10) 785–86 https://www.nature.com/articles/nmeth.1701
[28] Wang Z, Gerstein M and Snyder M 2010 RNA-Seq: A revolutionary tool for transcriptomics *Nat. Rev. Genet.* **10**(1) 57–63 https://www.ncbi.nlm.nih.gov/pubmed/19015660