Identification of laccase genes in *Ganoderma boninense* draft genome assembly

Létizia Camus-Kulandaivelu[(1)], Mercière Maxime[(1)], Tan Joon Sheong[(2)], Klopp Christophe[(3)], Tristan Durant-Gasselin[(4)], Sharifah Shahrul Rabiah Syed Alwee[(2)], Breton Frédéric[(1)]

Cirad(1), Felda(2), Palmelit(3), INRA(4)

(1) CIRAD, UMR 108 AGAP, Avenue Agropolis, 34398 Montpellier, Cedex 5 FRANCE

(2) FELDA GLOBAL VENTURES RESEARCH & DEVELOPMENT SDN BHD, Level 42, Menara Felda, Platinum Park, No. 11 Persiaran KLCC, 50088 Kuala Lumpur, MALAYSIA

(3) INRA Toulouse, CS 52627, 31326 Castanet-Tolosan, FRANCE

(4) PalmElit SAS, Bat 14, Parc Agropolis, 2214 Bd de la Lironde, 34980 Montferrier sur Lez, FRANCE

ABSTRACT: *Ganoderma boninense*, a soil born fungus, is the main agent of basal stem rot, one of the most devastating diseases of oil palm (*Eleais guinensis*). Observation of oil palm infection by *G. boninense* in natural conditions has shown the fungus capacity to quickly degrade the stem base, leading to important cracks and finally to tree fall. This important degradation of host tissue likely implies lignolitic enzymes, in particular laccase activity. Those enzymes have been well described in several rot disease involving fungus and their role in the pathogenicity of some fungus like the honey mushroom (*Armillaria mellea*) is well established. In this context, the study of *G. boninense* wood degradation genes, and in particular of laccase genes, seems to be of key interest to a better understanding of basal stem rot disease. We produced and assembled a draft sequence genome of an Indonesian *G. boninense* isolate and the draft sequence of a Malaysian isolate transcriptom. The draft genome assembly was annotated *ab initio* with Augustus software that predicts genes models from genomic sequence. We obtained 22228 gene models, among which 33 showed similarity with laccase. Among these 33 gene models, 25 exhibited the 4 domain laccase signature sequence and seven showed matches with expressed transcripts. Their length, intron number, subcellular addressing and peptide signal are classical of fungal laccases. A phylogenic analysis of *G. boninense* laccase predicted gene models along with other fungal laccases suggest recent and extensive gene duplication in *G. boninense* for a laccase clade specific to some polyporales white rot fungi.

Keywords: *Ganoderma boninense*, oil palm, basal stem root disease, genome sequence, *ab initio* gene prediction, laccase genes

1. INTRODUCTION

Ganoderma boninense, a soil born fungus, is the main agent of basal stem rot, one of the most devastating diseases of oil palm (*Eleais guinensis*). Observation of oil palm infection by *G. boninense* in natural conditions has shown the fungus capacity to quickly degrade the stem base, leading to important cracks and finally to tree fall. This important degradation of host tissue likely implies lignolitic enzymes, in particular laccase activity. Laccases are blue copper oxidase and catalyze the one-electron oxidation of phenols, aromatic amines and other electron-rich substrates by reducing molecular oxygen to water through an oxidoreductive multi-copper system (Thurston 1994). Fungal laccase are typically around 500 to t 550 amino acid long, bears a signal peptide and exhibit a four domain sequence signature encompassing cysteine and histidine residues involved in copper binding (Kumar et al, 2003). Fungal laccases have been well described in several rot disease involving fungus and their role in the pathogenicity of some fungus like the honey mushroom (*Armillaria mellea*) is well established (Yaver et al., 1996, Schwarze & Baum 2000). In this context, the identification and study of *G. boninense* laccase genes seems to be of key interest to a better understanding of basal stem rot disease

2. MATERIAL AND METHOD

2.1 DNA extraction, genome sequencing, preliminary assembly and gene prediction

The draft genome sequence was produced with the DNA extracted from mycelium of an Indonesian isolate (NJ3), grown in petri dishes on a PDA-chloramphenicol medium and provided by PT Socfindo. Mycelium was crushed in powder and extracted with a MATAB protocol (Tris HCl 1M ph 8,NaCl 5M, EDTA 500mM, MATAB, PEG 6000 and Sodium sulfite) with some adaptation to improve DNA extraction yield. Notably, a first step of tissue digestion with GLUCANEX buffer (GLUCANEX 0.5g, NaCl 1M 70ml, H2O 30ml, buffer at ph 6) allows the extraction of a higher DNA amount and removes an important part of polysaccharide compounds. G. boninense genomic DNA was sequenced by GATC biotech i) in a 454 GS-FLX plateform for shotgun sequencing of 400 bp fragments and ii) in a Illumina Hiseq 2000 plateform for sequencing of a 3 kb mate pair library with fragment length of 50 bp.
After quality check and reads trimming, 454-GS-FLX and Illumina Hiseq 2000 reads were assembled together into contigs with CLC Genomics Workbench 4.6.1 (CLC bio). Concerning general algorithm parameters, conflict resolution was set to “Vote (A, C, G, T)” and nonspecific matches was set to “random”. 454-GS-FLX reads were treated as “long” and “single” reads, with all other parameters kept to default values. Illumina reads were treated as “short” and “paired” data with Mindistance=1900 and Maxdistance=3500 for expected distances between paired reads. Scaffolding was done based on Illumina Hiseq 2000 mate pairs reads using SSPACE 2.0 software (Boetzer et al. 2011) using standard parameter values. A*b initio* gene model prediction on scaffolds was performed with Augustus 2.6.1 (Stanke and Waack 2003) using *Phanerochaete chrysosporium* as training organism.

2.2 RNA extraction, sequencing and preliminary assembly

The draft transcriptome *sequence of Ganoderma boninense* was obtained from an isolate collected in Felda’s plantation in Malaysia. The pure isolate was grown on a Ganoderma Selective Medium (GSM) and subsequently sub-cultured and grown on an oil palm extract medium according to Kok et al. (2013). Fresh mycelium was used to extract total RNA according to Saïdi et al. (2009). The RNA quality and quantity were measured using NanoDrop 2000

(Thermo Fisher Scientific Inc., USA) and Qubit 2.0 RNA Broad Range Assay (Invitrogen, USA). The samples were also run on Agilent Bioanalyzer RNA Nano chip to determine their RNA integrity number (RIN). Subsequently, messenger RNA isolation and cDNA synthesis were performed using TruSeq RNA Sample Preparation Kit (Illumina, USA) and SuperScript II Reverse Transcriptase (Invitrogen, USA) according to the manufacturers' protocol. The synthesised cDNA was later quantitated using Qubit 2.0 DNA Broad Range Assay (Invitrogen, USA). In this project, approximately 15ng cDNA was fragmented using Covaris S220 (Covaris Inc, USA) to a targeted size of 200 – 300bp. The fragmented cDNA was then end repaired, ligated to Illumina TruSeq adapters, and PCR-enriched using TruSeq RNA Sample Preparation Kit (Illumina, USA) according to the manufacturer's protocol. Sequencing was performed on an Illumina Hiseq 2000 machine. After quality check and adaptor trimming, reads were assembled using "Trinity**"**.

2.3 Laccase mining in public database and G. boninense genome

To retrieve known fungi laccase, we queried NCBI "nr" protein database using keywords "laccase" and "fungi". To keep only complete laccase sequences, we discarded 1) protein sequences shorter than 100 amino acid 2) protein sequences not starting by methionine amino acid and 3) protein sequences with name not containing 'laccase'. We aligned protein sequences with MAFFT software (Katoh and Standley 2013) using "-einsi" option to allow for large unalignable regions. We then used HMMER3 software (Mistry et al., 2013) to synthetize sequence profile common to laccase alignment ("hmmbuild" module) and mined G. boninense predicted gene model database for putative laccase ("hmmsearch" module).

2.4. G. boninense predicted laccase analysis

We aligned all *G. boninense* gene model matching HMMER laccase pattern using MAFFT software (Katoh and Standley 2013). We identified laccase signature sequences (Kumar et al. 2003) and kept for further analysis only those exhibiting the full 4 domain fungal laccase signature. We blasted predicted laccase amino acid sequences against transcriptome assembly contig in order to determine those possibly expressed in oil palm extract medium. We determined potential signal peptides with SIGNALP 3.0 (http://www.cbs.dtu.dk/services/SignalP) and predicted subcellular location with TARGETP 1.1 (http://www.cbs.dtu.dk/services/TargetP). We aligned putative *G. boninense* laccases with other known fungal laccases from *Laccaria bicolor*, *Coprinopsis cinerea*, *Trametes villosa* , *Ganoderma lucidum* and *Phanerochaete chrysosporum* (in this latter case Multi-Copper Oxidase (MCO)) retrieved from NCBI with MAFFT software (Katoh and Standley 2013). These four species are basidiomycota, either from the polyporales order (*Trametes villosa* , *Ganoderma lucidum*, *Phanerochaete chrysosporum*) or from the agaricales order (*L. bicolor*, *C. cinerea*,). We removed poorly aligned regions with Gblocks software (Castresana 2000) and produced a phylogenetic tree with PhyML (Guindon et al. 2010) using default values for amino acid analysis.

3. RESULTS

3.1 Preliminary genome assembly characteristics and gene prediction

We obtained 2007494 reads from 454-GS-FLX platform with a total sequence length of 760556282 bp and 276922790 reads from Illumina Hiseq 2000 platform with total raw sequence length of 14123062290 bp and average insert size of 2700 bp. After cleaning, trimming, assembly and scaffolding, 5754 scaffolds were built, for a total sequence length of 70885986 bp and N50

scaffold length of 11474 bp. Missing data ('N') represents 14.73 % of total scaffold length. A total of 22228 gene models were predicted.

3.2 Preliminary transcriptome assembly

We obtained 229240656 raw reads from Illumina Hiseq 2000 machine, with total sequence length of 23153306256bp. After cleaning, trimming and assembling, a total of 66 807 contigs were obtained with average length of 1796 bp.

3.3 Laccase mining in public database and G. boninense genome

Requesting "laccase" and "fungi" in the NCBI nr protein database cropped a total of 4349 entries. After applying the different filters, we retained 841 protein sequences. Based on full sequence score and E-value, HMMER software retrieved 33 *G. boninense* gene models showing meaningful similarities to the inputed 841 fungus laccase alignment.

3.4 *G. boninense* predicted laccase characteristics

Among these 33 gene models fitting HMMER laccase pattern, 25 bore the 4 domain laccase signature These 25 gene models display a predicted coding region length between 1473 bp and 1926 bp (between 491 and 642 amino acids). These gene models are distributed on seventeen scaffolds. Scaffold 1 bears 3 gene models with the 4 domain laccase signature, scaffold 49 bears 3 gene models with the 4 domain laccase signature, scaffold 46, 149 and 197 bears 2 gene models each with the 4 domain laccase signature. Predicted intron number varies between 3 and 12. Twenty four gene models are predicted to be secreted and 23 are displaying a predicted peptide signal. Out of 25 gene models with four domain laccase signature, seven showed strong similarity with contigs from assembled transcript (sequence similarity between 93 to 99 % over more than 1300 bp)

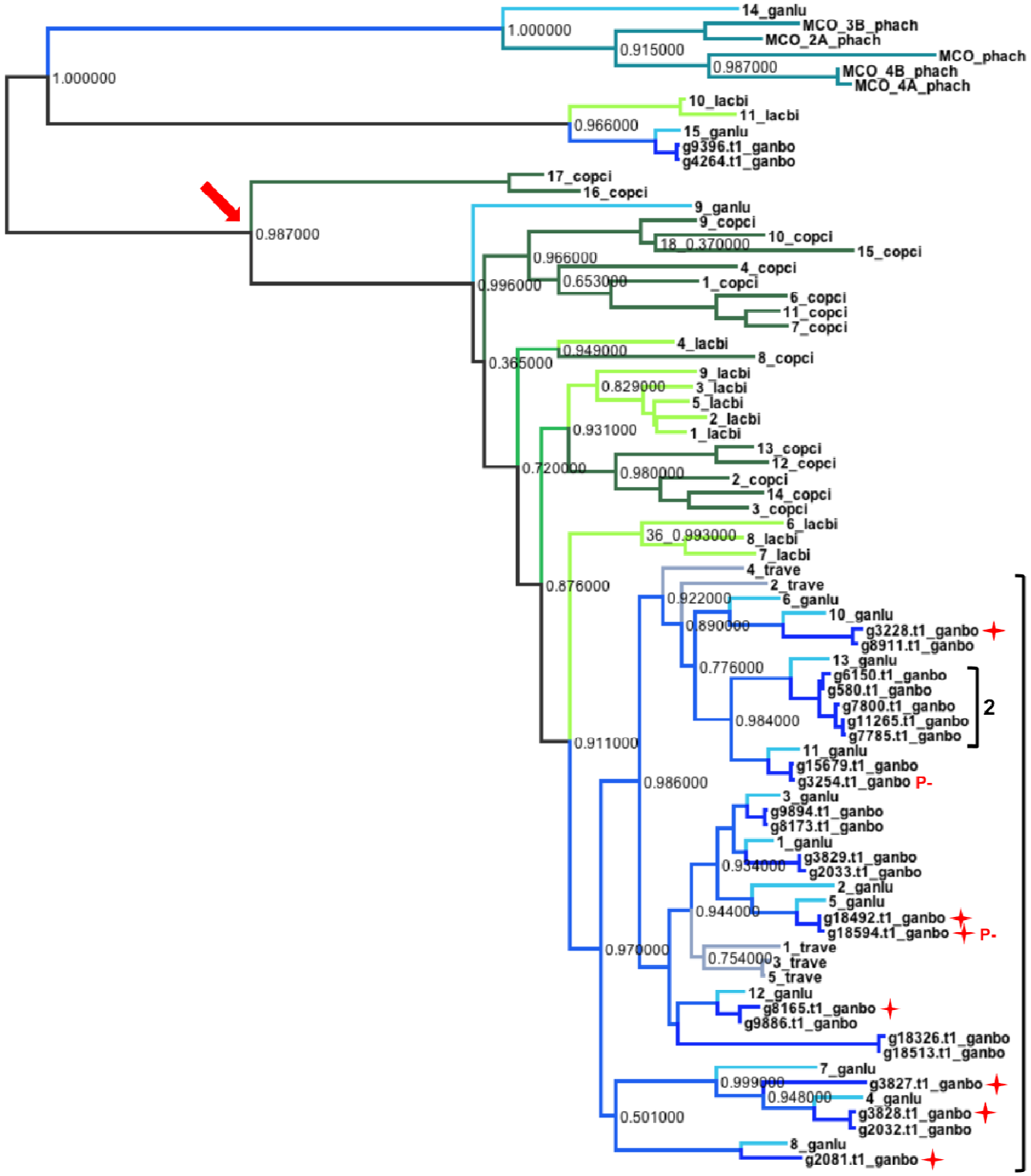
3.4. *G. boninense* predicted laccase among fungal laccases

We selected for phylogenetic analysis the set of 25 *G. boninense* predicted laccases (e.g. Table 1) displaying the full four domain laccase signature and use them with other laccases from four other fungi species to build a phylogenetic tree (eg fig 1). *G. boninense* predicted laccases are positioned in two groups. Gene model g9396 and g 4264 are gathered with a *G. lucidum* and two *L bicolor* laccases belonging to laccase *sensu lato* (Hoegger *et al.* 2006). More specifically, these two *L. bicolor* laccases are ferrodoxidase. The other 23 *G. boninense* predicted gene models gathers in a large group solely composed of polyporales laccases, including all five *T. versicolor* laccases and 12 *G. lucidum* laccases. Note that the seven laccase predicted gene models showing transcript evidence are located in the polyporales gene cluster.

Figure 1: Phylogenetic tree for G boninenese predicted laccases and laccases of four other basidiomycetes species

Phylogenetic tree was obtained with PhyML (Guindon et al. 2010) using default values for amino acid analysis. It is rooted on the median point. ganbo: *G. boninense* predicted laccase gene models (electric blue), ganlu: *G. lucidum* (cyan), trave: *T. versicolor* (grey), phach: *P. chrysosporium* (Turkish green), copci: *C. cinerea* (dark green), lacbi: *L. bicolor* (light green). Plain blue indicates branches specific to polyporales and plain green indicates branchs specific MCO: Multi Copper Oxidase. Branch support values were obtained with LRT SH-like test. A red arrow indicates sensu stricto laccases as defined by Hoegger *et al.* (2006). Bracket one underline a laccase clade specific to polyporales order and bracket two pinpoint a possible copy expansion

in G. lucidum. Red star indicates *G. boninense* gene model with transcript match and red "P-" indicates *G. boninense* gene models with no peptide signal.



4. DISCUSSION

We identified 25 gene models exhibiting the 4 domain fungal laccase signature and the main characteristic of fungal laccases in term of length, cellular addressing and signaling and intron number. For 7 of these 25 laccase predicted gene models, we found transcript evidence, supporting gene prediction. Our results suggest the existence of a specific group of polyporales *sensu stricto* laccases (Hoegger *et al.* 2006) for which ancient duplications occurred (e.g. fig 1). This group includes laccases from the white rot *G. lucidum*, *G. boninense* and *T. versicolor* but not *P. chrysosporium*. Whithin the polyporales *sensu stricto* laccase group, the phytopathogenic *G. boninense* exhibit a noticeable behavior, with the specific expansion of some gene sub-familly (for example gene model g6150, g580, g7800, g11265, g785), leading to a potentially doubled laccase copy number compared to its relative *G. lucidum.*

5. ACKNOLEDGEMENTS

We are grateful to PT sofindo who provided NJ3 isolate. We also want to express our thanks to Jean-François Dufayard for his helpful advice in phylogeny.

6. REFERENCES

Boetzer M., Henkel C. V., Jansen H.J., Butler D, Pirovano W. (2011). Scaffolding preassembled contigs using SSPACE. Bioinformatics 27 (4): 578-579.

Burke R. M. , Cairney J.W.G. (2002). Laccases and other polyphenol oxidases in ecto- and ericoid mycorrhizal fungi. Mycorrhiza 12. 105-116.

Chen S. et al. (2012). Genome sequence of the model medicinal mushroom Ganoderma lucidum. Nature communications 3:913 | DOI: 10.1038.

Courty P. E., Hoegger P. J., Kilaru S., Kohler A., Buée M., Garbaye J., Martin F., Kües U. (2009). Phylogenetic analysis, genomic organization, and expression analysis of multi-copper oxidases in the ectomycorrhizal basidiomycete Laccaria bicolor. New Phytologist 182: 736-750.

Guindon S., Dufayard J. F., Lefort V., Anisimova M., Hordijk W., Gascuel O. (2010). New algorithms and methods to estimate Maximum Likelyhood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59 (3): 307-310.

Hoegger P.J., Kilaru S., James T.Y., Thacker J.R., Kües U. 2006. Phylogenetic comparison and classification of laccase and related multicopper oxidase protein sequences. FEBS Journal 273: 2308–2326.

Katoh K. , Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements and usability. Molecular Biology and Evolution 30 (4) :772-780.

Kok SM, Goh YK, Tung HJ, Goh KJ, Wong WC, Goh YK (2012). *In vitro* growth of *G. boninense* isolates on a novel Plam extract medium and virulence on oil palm seedlings (*E. guineensis*)

Kumar S. V., Phale P.S., Durani S., Wangikar P. P. (2003). Combined sequence and structure analysis of the fungal laccase family. Biotechnoloy and bioingeneering 83: 386-394.

Mistry J., Finn R .D., Eddy S. R., Bateman A., Punta M. (2013). Challenge in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic acid research 41 (12): e121.

Saidi N.S., Gargouri-Bouzid R., Rayanni M., Drira N. (2009). Optimization of RNA isolation from Brittle Leaf affected date palm leaves and construction of a subtractive cDNA library. Molecular biotechnology 41: 63-68.

Schwarze, F. W. M. R. and S. Baum (2000). Mechanisms of reaction zone penetration by decay fungi in wood of beech (Fagus sylvatica). New Phytologist 146(1): 129-140.

Stanke M. , Waack S. (2003). Gene prediction with Hidden-Markov Model and a new intron submodel. Bioinformatics 19 (suppl 2):ii215-ii225.

Thurston CF (1994). The stucture and function of fungal laccases. Microbiology 140: 19-26.

Yaver, D. S., F. Xu, E. J. Golightly, K. M. Brown, S. H. Brown, M. W. Rey, P. Schneider, T. Halkier, K. Mondorf and H. Dalboge (1996). Purification, characterization, molecular cloning, and expression of two laccase genes from the white rot basidiomycete Trametes villosa. Appl Environ Microbiol 62(3): 834-841.