LACCASES REPERTOIRE OF A SUBTERRANEAN TERMITE *Coptotermes curvignathus* HOLMGREN (BLATTODEA: RHINOTERMITIDAE)

Hoe Pik Kheng[1], King Jie Hung, Patricia[2*], Ong Kian Huat[2], Bong Choon Fah, Joseph[1] & Nor Muhammad Mahadi[3]

[1]Department of Crop Science, Faculty of Agriculture and Food Sciences, Universiti Putra Malaysia Bintulu Sarawak Campus, Nyabau Road, 97000 Bintulu, Sarawak, Malaysia.

[2]Centre of Excellence in Borneo Ecosystem Science Research, Faculty of Agriculture and Food Sciences, Universiti Putra Malaysia Bintulu Sarawak Campus, Nyabau Road, 97000 Bintulu, Sarawak, Malaysia.

[3]Malaysia Genome Institute, 43600 Bangi, Selangor, Malaysia.

*Corresponding author: *patricia@upm.edu.my*

ABSTRACT

Coptotermes curvignathus is a subterranean termite species that vigorously feed on living-tree. The cellulose, hemicelluloses, lignin components in the wood are too robust for many organisms to break down and extract energy from but they serve as main carbon and energy sources for *C. curvignathus*. The ability of this subterranean termite thrive on wood diet may attribute by its array of laccase repertoire. Laccases are known for many functions including detoxification of plant tissues via xenobiotic pathway, and most importantly for termite is lignin modification. This paper highlighted the type of laccases produced by *C. curvignathus* based on transcriptomic data that were generated from 500 termites' digestive system using Illumina HiSeq 2000. Raw data was trimmed and assembled by SOLEXAQA and Bowtie before loaded into Gene Ontology based data mining software, Blast2GO (B2G). The result showed that, *C. curvignathus* expressed diverse laccase genes that were phylogenetically similar to other termites' laccases and distinctly related to fungal or bacterial laccases. Other than providing laccase genes sequences for further gene and enzyme characterization, the result of this study is the first insight into *C. curvignathus* laccase repertoire that is important to elucidate how *C. curvignathus* could digest wood efficiently from either intact or partially hydrolyzed wood.

Keywords: *Coptotermes curvignathus*, laccase, lignin modification

ABSTRAK

Coptotermes curvignathus merupakan satu species anai-anai bawah tanah yang aktif memakan kayu. Kayu terbentuk daripada komponen selulosa, hemiselulosa dan lignin adalah sangat tegap dan sukar dicerna oleh kebanyakan organisma hidup. Namun bagi anai-anai *C. curvignathus* ia adalah sumber utama karbon dan tenaga. Kebolehan anai-anai bawah tanah ini bermandiri dengan pemakanan kayu ini adalah disebabkan mereka mempunyai sekumpulan enzim lakase yang berbagai. Lakase adalah diketahui memainkan pelbagai peranan dalam

proses biologi dan ini termasuklah proses penyahtoksikan toksik daripada tisu-tisu tumbuhan melalui laluan xenobiotik. Kajian ini melaporkan jenis-jenis lakase yang dihasilkan oleh *C. curvignathus* berdasarkan data transkriptomik yang dijana daripada sistem pencernaan 500 ekor anai-anai dengan menggunakan Illumina HISeq 2000. Data mentah dipangkas dan dihimpun semula dengan SOLEXAQA dan dianalisis dengan Bowtie sebelum data-data ini dimasukkan ke dalam perisisan pelombongan data Gene Ontology, Blast2GO (B2G). Hasil kajian menunjukkan bahawa *C. curvignathus* mengekspres pelbagai gen lakase yang spesifik kepada anai-anai dan berbeza daripada lakase kulat dan bakteria. Jujutan gen lakase yang diperolehi daripada hasil kajian ini adalah penting untuk pencirian lanjutan, dan ia adalah laporan pertama yang menfokuskan enzim lakase *C. curvignathus.* Informasi ini adalah penting untuk menerangkan bagaimana *C. curvignathus* dapat bermandiri berdasarkan pemakanan kayu.

Kata kunci: *Coptotermes curvignathus*, lakase, modifikasi lignin

INTRODUCTION

Laccases (benzenediol: oxygen oxidoreductase, EC 1.10.3.2) is a group of enzymes (blue copper proteins or blue copper oxidases) that are characterized by containing 4 catalytic copper atoms. They are widely known for their ability to catalyze the oxidation of a large range phenolic, amines and ascorbate compounds and to reduce molecular oxygen to water (Madhavi & Lele 2009; Messerschmidt 1997; Solomon et al. 1996; Solomon et al. 2001; Thurston 1994; Xu 1996). Laccases are widely distributed in many organisms ranging from plants, insects, fungi and bacteria (Assavanig et al. 1992; Benfield et al. 1964; Diamantidis et al. 2000; Levine,1965; Morozova et al. 2007; Octavio et al. 2006; Xu 1996). The functions of laccases in these organisms vary, which include roles in 1) morphogenesis and differentiation of sporulating and resting structure in basidiomyces, 2) lignin biodegradation (white-rot fungi), 3) pigment formation in mycelia and fruiting body, 4) improves cell to cell adhesion, 5) assists in the formation of polyphenolic glue that binds hyphae together, 6) enable fungi to overcome the immune response of the host, 7) detoxification of plant tissues via the oxidation of antifungal phenols or deactivation of phytoalexins, 8) plant cell biosynthesis, 9) phytopathogenesis, 10) woody material degradation, 11) insect sclerotization, 12) participating in the cuticle melanization pathway, and 13) bacterial melanination. The potential function and application of laccases are extended to detoxify a range of environmental pollutants. They have been used for several purposes in many industries including paper, pulp, textile, petrochemical industries, food processing, medical and health care.

Thus, the objectives of our research were to study 1) the laccase genes expressed in *C. curvignathus* when they were fed on wood diet and partially hydrolyzed empty oil palm fruit bunches and 2) the structural similarities and evolutionary relationships of different organisms laccases from plants, insects and microbes

MATERIALS AND METHODS

Sample Collection of Termites *C. curvignathus*

The termites *C. curvignathus* were collected from a baiting station at the botanical garden in Universiti Putra Malaysia Bintulu Sarawak Campus (UPMKB). Fresh rubber wood stakes (30–50 cm length) obtained from the rubber plantation in UPMKB were used as baits. These infested baits were removed, placed in black plastic containers and transported back to the laboratory. The termite samples were identified as *C. curvignathus* using soldier morphometric

measurement based on Thapa (1982) and Tho (1992). After transportation, the termite samples were left in the container without disturbance in total darkness at 28ºC±2ºC and 90% humidity for at least one week prior to any feeding assays.

Force Feeding of Termites *C. curvignathus* with Different Food Materials

Termites collected from the field were subjected to force-feeding assays with three types of diet: 1) the rubber wood planks (coded as DRW) 2) partially hydrolysed oil palm empty fruit bunches with commercial cellulases, Cellic CTEC (coded as CELL) and 3) partially hydrolysed oil palm empty fruit bunches with commercial cellulases, Accelerase 1500 (coded as ACCE). Force-feeding assay is a feeding test where only one type of food sources was given to the termites. A 15cm Petri dish was used as termite harbourage and food chamber. 190 workers and 10 soldiers were used in each set of experiment. For each treatment, there were five replicates. The termites fed with different food sources were left in the incubator at 28ºC±1ºC and 100% humidity for a period of two weeks prior to dissection and extraction of total RNA.

Dissection of Termite Digestome and Total RNA Isolation

Approximately 500 worker termites were dissected and used for total RNA extraction after the end of the feeding bioassays. The digestomes, which includes the salivary glands, foregut, midgut, hindgut and rectum were removed from each termite in a RNase-free condition. Termites' digestome were immersed in RNAlater (Ambion, USA) and stored in –80ºC freezer till total RNA extraction.

Total RNA extraction was executed using Sepasol RNA 1 Super G (Nacalai Tesque, Japan) according to the manufacturer protocol. The termites' samples were grinded and homogenized using a micropestle in the Sepasol RNA 1 Super G solution. The Sepasol was prior heated in 65ºC before pipetting into the termite digestome samples. During the separation phase, 200–400µL of chloroform was used for each tube depending on the contaminants in the samples. The total RNA was precipitated with 2-propanol and washed with 75% ethanol and dissolved in nuclease free water. The quality of the total RNA was checked using Nanodrop spectrophotometer (Thermo Scientific, USA) and native gel electrophoresis.

cDNA Library Construction and Sequencing of Termite Digestome Library

Total RNA of 4µg was used to construct the cDNA library. The cDNA library construction was carried out using TruSeq RNA kit (Illumina, USA) according the manufacturer protocols that included end repair, ligation and PCR enrichment of the double strand cDNA. Paired-end sequencing cDNA library were sequenced using Illumina HiSeq 2000 platform (Illumina, USA). Three libraries were constructed and coded as HPK2EODRW (cDNA from *C. curvignathus* fed on rubber wood planks), HPK2EOCEL (cDNA from *C. curvignathus* fed on partial hydrolysed oil palm empty fruit bunches using Cellic CTEC) and HPK2EOACC (cDNA from *C. curvignathus* fed on partial hydrolysed oil palm empty fruit bunches using Accellerase 1500). The quality of the subjected libraries was verified by microfluid analysis (2100 Bioanalyser, Agilent, Santa Clara, Ca, USA). The reads from the sequencer were pre-processed by trimming the reads based on quality and length using SolexaQA v2 (Cox et al. 2010). The Phred value used was QV>20, which is 99% accuracy to ensure the sequence quality and all reads less than 50% of the read length (50bp) were trimmed. The spiked control PhiX-174 reads were screened and removed from the trimmed sequence pool using Bowtie2 by mapping the PhiX-174 known sequences to the libraries. The pairwise sequence and singleton files were run through an array of single-k assemblies with hash length value of 51–75 using Velvet v2.0 (Zerbino & Birney 2008) and Oases v0.2.08 (Schultz et al. 2012) to *de novo* assembly the reads. BLAST search were performed for both paired and singletons transcripts obtained from the

assembly process against the Genbank "nr" database available at public domain www.ncbi.nlm.nih.gov at E value of e-5. The blast results were subjected to BLAST TO GO for functional, gene ontology and annotation analysis (Götz et al. 2008).

Sequence Alignment and Phylogenetic Analysis

ESTs were organized into contiguous sequences (contigs) using ClustalW with default parameters in Mega v5.02 (Tamura et al. 2011; Thompson et al. 1994). Gene discovery rate for each library was estimated by dividing the number of contiguous sequences by the total number of singleton sequences. Minimum average sequences insert sizes of cDNA clones were estimated by dividing the total number of base pairs sequenced by the number of ESTs in each library. Signal peptides were predicted using BLAST2GO: SignalP and HMM that are available with the BLAST2GO programme (Götz et al. 2008).

In order to identify putative metal ion binding residues, amino acid sequences for representative laccases from plant, bacteria, fungi, and insect were obtained from the NCBI protein databases and aligned with the laccases from cDNA library MEGA v5.02 software under default parameters of ClustalW algorithms. The sequences used in this analysis were listed in Table 1. Phylogenetic analyses were conducted to examine the relationship of termite laccases to those of other insects, fungi, bacteria and plants. A *Streptomyces lavendulae* laccase was used as an out group. Two types of phylogenetic analysis were used for tree construction in this study; 1) maximum likelihood, a neighbour-joining phylogenetic network was generated from the distance matrix and 2) maximum parsimony trees, a parsimony phylogenetic network that was generated from evolution steps matrix. Both trees constructions were subjected to 1000 bootstrap replicates using default settings from Mega v5.02.

Table 1. Summary statistics for *C. curvignathus* metatranscriptomics quality

cDNA libraries	HPK2EODRW (rubber wood)	HPK2EOACC (enzymatic treated EFB with Accellerase 1500 commercial enzyme cocktails)	HPK2EOCEL (enzymatic treated EFB with Cellic CTEC commercial enzyme cocktails)
Total reads	37,285,900	27,712,358	47,449,836
Reads length (bp)	101	101	101
Total sequence data (bp)	3,765,875,900	2,798,948,158	4,792,433,436
Total high quality reads	31,731,481	23,545,594	39,762,194
Total transcripts	123,479	94,920	105,490
Total transcript size (bp)	104,619,854	68,229,579	77,585,064
Mean transcript size (bp)	847	719	735

RESULTS AND DISCUSSION

Sequencing Data Summary

In the present study, laccases transcripts were obtained from three cDNA libraries, HPK2EODRW, HPK2EOCEL and HPK2EOACC generated from the mRNA of *C. curvignathus* digestome and sequenced using the Illumina HiSeq 2000 platform. This approach resulted 37285358, 47449836 and 27712358 reads with an average read length of 101 nucleotides (Table 1). A total of 3.7, 4.7 and 2.7 gigabases of sequence data were generated from the three cDNA libraries. High quality reads were generated by removing any sequences that had a lower than the value of QV= 20 and 50 base pair sequences. This resulted 31731481, 39762194 and 23545594 high quality reads, respectively, with an average read length of 93 nucleotides for HPK2EODRW and 92 nucleotides for both HPK2EOCEL and HPK2EOACC. The high quality reads were *de novo* assembled into a total of 123493, 105490 and 94920 total transcripts from 51844, 48799 and 46429 loci, respectively, with an average length of 847, 735 and 719 base pairs. Using BLASTX searches against the non-redundant database from Genbank indicated 73,449 (59.48%), 54,479 (51.64%) and 78,534 (82.73%) transcripts had at least one significant match (E-value$\leq 10^{-5}$).

Transcripts that had a match for laccase were selected from the three cDNA libraries for focus study. The gene discovery for laccases in each cDNA libraries were 2.037, 2.195 and 4.147 of locus discovery, respectively for transcriptomes HPK2EODRW, HPK2EOCEL, and HPK2EOACC. The *C. curvignathus* that fed on partially hydrolysed wood diet expressed higher number of laccase genes (Table 2). A total of 10 contigs and 13 singletons that were predicted to be laccase were found in the three cDNA libraries, with 3 contigs and 3 singletons for HPK2EODRW, 3 contigs and 5 singletons for HPK2EOCEL and 4 contigs and 5 singletons for HPK2EOACC after the assembling of transcripts for each cDNA libraries (Table 3). The term contig used here is defined as a contiguous segment of the genome made by joining overlapping sequences and the term singleton is a read (transcript) that could not be assemble as it is a unique sequence (transcript) with no overlapping with other sequence (transcript).

Table 2. Gene discovery, total contigs and total singletons for *C. curvignathus* eukaryotes metatranscriptomic analysis under different diet condition

cDNA libraries	Total locus	Locus discovery	Total laccase transcripts/ Total transcripts	Transcript discovery	Total gene	Gene discovery	Total contigs	Total singletons
HPK2EODRW	6	2.037	10/16	3.394	6	2.037	3	3
HPK2EOCEL	8	2.195	12/13	3.292	8	2.195	3	5
HPK2EOACC	9	4.147	14/15	6.452	9	4.147	4	5

All the contigs and singletons from each cDNA libraries were then shuffled and reassembled to search for highly similar contigs and singletons among the three types of diet tested. The results showed that some contigs and singletons among the three cDNA libraries were highly similar. These laccases gene were regarded as core laccase genes in *C. curvignathus*. A total of 2 contigs CC-Laccase_001 (958bp) and CC-Laccase_007 (1276bp) were regarded as core laccase genes in *C. curvignathus* (Table 4). Other contigs such as CC-Laccase_002, CC-Laccase_003, CC-Laccase_004, CC-Laccase_005 and CC-Laccase_006 were common in at least two libraries. CC-Laccase_008, CC-Laccase_009 and CC-

Laccase_010 were only expressed in HPK2EOCELL cDNA library. On the other hand, CC-Laccase_011, CC-Laccase_012, CC-Laccase_013 and CC-Laccase_014 can be found only in HPK2EOACC cDNA library. The transcriptomic sequences for CC-Laccase 001-014 were shown in Table 5. The result of this study shows *C. curvignathus* were able to express diverse laccases when force-fed with either intact wood (DRW) or partially hydrolyzed wood (CELL and ACC).

Table 3. Total contigs and the variables within each *C. curvignathus* transcriptomics library

cDNA libraries	Contig/Singleton name	Locus within flow cell	Contig/Singleton	Number of total transcripts	Number of transcript involve in contig assembly	Number of contigs/singletons	Number of isoforms	Changes (SNP or variable region)	Site of SNP or variable region	Length of contig	GC content
HPK2EODRW	DRW001	25809	Singleton	1	NA	1	1	-	-	497	0.425
HPK2EODRW	DRW002	34295	Singleton	1	NA	1	1	-	-	701	0.431
HPK2EODRW	DRW003	44580	Singleton	1	NA	1	1	-	-	386	0.49
HPK2EODRW	DRW004	26123	Contig	3	3	1	2	SNP substituition	220	958	0.461
HPK2EODRW	DRW005	5855	Contig	2	2	1	1	-	-	1530	0.424
HPK2EODRW	DRW006	30051	Contig	8	3	1	1	-	-	3696	0.456
HPK2EOCEL	CEL001	12666	Singleton	1	NA	1	1	-	-	312	0.5
HPK2EOCEL	CEL002	24583	Singleton	1	NA	1	1	-	-	1099	0.424
HPK2EOCEL	CEL003	18308	Singleton	1	NA	1	1	-	-	267	0.509
HPK2EOCEL	CEL004	18882	Singleton	1	NA	1	1	-	-	418	0.545
HPK2EOCEL	CEL005	19513	Singleton	1	NA	1	1	-	-	949	0.501
HPK2EOCEL	CEL006	18142	Contig	2	2	1	1	-	-	962	0.475
HPK2EOCEL	CEL007	23792	Contig	2	2	1	1	-	-	1982	0.417
HPK2EOCEL	CEL008	30379	Contig	4	4	1	1	-	-	2634	0.479
HPK2EOACC	ACC001	9687	Singleton	1	1	1	1	-	-	345	0.383
HPK2EOACC	ACC002	19112	Singleton	1	1	1	1	-	-	747	0.466
HPK2EOACC	ACC003	20620	Singleton	1	1	1	1	-	-	200	0.41
HPK2EOACC	ACC004	24873	Singleton	1	1	1	1	-	-	346	0.436
HPK2EOACC	ACC005	29109	Singleton	1	1	1	1	-	-	184	0.527
HPK2EOACC	ACC006	15422	Contig	2	3	1	1	-	-	449	0.461
HPK2EOACC	ACC007	36170	Contig	3	3	1	2	SNP	3081	3116	0.465
HPK2EOACC	ACC008	29568	Contig	2	2	1	1	-	-	849	0.468
HPK2EOACC	ACC009	15887	Contig	2	1	1	1	-	-	240	0.468

Table 4. Total contigs, total singletons and the variables in all three *C. curvignathus* eukaryotes cDNA libraries pool

New Contig/ Singleton name	Contig/ Singleton	Contig/ singleton used	Total transcript used	Total number of isoforms	Changes (SNP or variable region)	Site of SNP or variable region	Isoform name	Number of transcripts involved in the isoform	Origin of isoform from which library	Length of contig
CC-Laccase_001	Contig	DRW004	3/5	2	SNP	220	a	4/5	HPK2EODRW	958
		CEL001	1/5						HPK2EOACC	
		ACC002	1/5						HPK2EOCEL	
							b	1/5	HPK2EODRW	
CC-Laccase_002	Contig	DRW001	1/2	1	-	-	-	-	-	1099
		CEL002	1/2							
CC-Laccase_003	Contig	DRW003	1/3	1	-	-	-	-	-	849
		ACC008	2/3							
CC-Laccase_004	Contig	DRW005	2/5	1	-	-	-	-	-	1530
		ACC006	3/5							
CC-Laccase_005	Contig	DRW006	3/7	2	Variable region *	113 - 230	a	3/7	HPK2EODRW	3811
		CEL008	4/7				b	4/7	HPK2EOCEL	
CC-Laccase_006	Contig	ACC004	1/3	1	-	-	-	-	-	1982
		CEL007	2/3							
CC-Laccase_007	Contig	ACC003	2/4	1	-	-	-	-	-	1276
		DRW002	1/4							
		CEL006	1/4							
CC-Laccase_008	Singleton	CEL003	1/1	1	-	-	-	-	-	267
CC-Laccase_009	Singleton	CEL004	1/1	1	-	-	-	-	-	418
CC-Laccase_010	Singleton	CEL005	1/1	1	-	-	-	-	-	949
CC-Laccase_011	Singleton	ACC001	1/1	1	-	-	-	-	-	345

CC-Laccase_012	Singleton	ACC005	1/1	1	-	-	-	-	-	184
CC-Laccase_013	Contig	ACC007	3/3	2	SNP	3081	a b	2/3 1/3	HPK2EOACC HPK2EOACC	3116
CC-Laccase_014	Contig	ACC009	2/2	1	-	-	-	-	-	240

Note (*): Variable region may have several base pairs within the regions that are similar among all the contigs during assembly

Table 5. Transcriptomic sequences of *C. curvignathus* for laccases gene

Contig	Consensus sequence
CC_001	GTTTTGAAGAGGCAGATGGCATGTTTGGAACCATGATTGTCCGACGTCCTGTCAGTAAAGAACCACACAGCGACCTGTA TGATGAGGATAGGTCTGAACATTCAATGATTGTATGGCACTGGTTTGGGAGTTCTGCTAGGGAAGTGCTGACCATCTCG AAGTACACTGGGGCACGCTCTAGAGGAGAAGGGCTCATCATTAATGGCCTGGGAGGCCTTGRGGCTTTTGAACTACCAG TGGAAAACACGTTTGATACAATACCAAGAGAAGTGTTCAGAGTTCAGCAGGGCCGACGATACCGCTTCCGAGTGATCT ACAACAACCAGATTCCTTGTCCAGTGCAGCTGTCTGTGCAGAATCACAGTCTCCTGGTTATTGCCAGTGATGGTGCGAG CTTCCAACCTGTTGAAGCAAATTCCATAATGCTTAATGGAGGTGAGAGGTATGACTTTGTACTGAAAGCAGATCAGCCA GATAACAATTACTGGATACGGTTTCGGGGGTTGGTAAGTTGTAACAGTGGTGAGAGGAAAGTACATCAGGAGGCAGTG CTGCATTATGAAGGTGCTGATGAAGCTTTGCCTGAAGGGGAGTCACAATATGACGACGCTATTGCCACAGGGACGTTGG TGAATCCAATTGGAGCAATTGCTTTCAACTATTCTGCCAACGAACTTATATATGTGTCAGATCTGGAAAATGTGGATAC AGAAAGAGCCCAAAACATCAGTGGTGATGCAGATCAGATCATTTACGTTGACTTCCAGTTCAACACATATGAGTCAGAG GATATTTCAGGTTCTTGGCCACAGGTGAACTCTCGCACATTCAGTTACCCACCTTTCCCCCTTCTGACACAACGTTATGA CATAACTCGAGACATGTACTGCACAGATGAAGACATTTGTGCTGATGGCTTGTTCTGTGCATGCCCATATTTGTACAATG TGGAATTGGG
CC_002	CATACATACAGATTTATTTCCAGAACTTTTTAAAAAAAGAAGTTACAGGCACCACACTGTGATGTTAAACACCACAGTT TGCAACACTGAGCAGAAACCAACTCAGTAGTACCAACATGATGGTACCCACATGTGAACCACTAGCTGCACCACTAGG AATTTCAGTACTTTCCACCCAGTTTCCACAACGAGGAAAGTCATATGGTGTGTTGGGCATATCAGTTAGGTTACCAACTT GCAAAACAAACCCCATACCCATATCTGCATGACTGGATATATGACAGTGGAAAAACCAAAAACCTGGATTGTCAGCAA CAAATCTAATCACAGCATATCCTGAGCTTGGAACAGACACTGTATCTTTTATAGGTCCATCAAAGTTCTTAGAGATGTTG TTTTGTTCATTTAGTTCCCTCACAATATCCGTGGATATGTCACTTGCAATTACGCCAGCAGCCACAACATGGAAACCATA GCCATGTAAATGGAATGGGTGATCTCCACCTCCTTGCGCGAGGTCAATAAACACAATTTCAACCAAGTTTCCCAATTCC

ACATTGTACAAATATGGGCATGCACAGAACAAGCCATCAGCACAAATGTCTTCATCTGTGCAGTACATGTCTCGAGTTA
TGTCATAACGTTGTGTCAGAAGGGGGAAAGGTGGGTAACTGAATGTGCGAGAGTTCACCTGTGGCCAAGAACCTGAAA
TATCCTCTGACTCATATGTGTTGAACTGGAAGTCAACGTAAATGATCTGATCTGCATCACCACTGATGTTTTGGGCTCTT
TCTGTATCCACATTTTCCAGATCTGACACATATATAAGTTCGTTGGCAGAATAGTTGAAAGCAATTGCTCCAATTGGATT
CACCAACGTCCCTGTGGCAATAGCGTCGTCATATTGTGACTCCCCTTCAGGCAAAGCTTCATCAGCACCTTCATAATGCA
GCACTGCCTCCTGATGTACTTTCCTCTCACCACTGTTACAACTTACCAACCCCCGAAACCGTATCCAGTAATTGTTATCT
GGCTGATCTGCTTTCAGTACAAAGTCATACCTCTCACCTCCATTAAGCATTATGGAATTTGCTTCAAC

CC_003
GTTCTTCATTCACAGGGGCACCAACATAATGAAGCACAGCAACCTGGTGAGCTTTAGTGAAACGTTCATCACAATCCAT
TAGCCCTCGGAATCTCATCCAGTAGTTGCCAACTTTTTTATTTGCATGCAGCACAAAATCAAATCTTTCTCCAGCATAGC
TCACCAGAGAGTCAGCTTGCACTGGCTGTATATCACTTCCATCACTGCTGATCACTAACAGGGTGTGATTATCCACTGAT
AGTTCAATAGGACAGTTGAGGAAGCCTGCATTTATAAGGCGGAACCTGTAGCGGAAGCCCTGCTGAACAGTGAATGTT
GCAGTTGGTGAATGTGTCACATTTACAGTGATAGAATTTACATCCTGAAATTGGTGGAAGCGACCCCTGCCATTAACCA
ACAGGGTCTTTGGTTTGTTGTTGCCATTGGCATGATGGTGAGCAAGGAACTTGTCCACTCCCAGCTTGTCATCCCAGTCC
AGGATTTGCAAAGTGTGTTCAGACAGATCTTGGTCATACAGCTGTCTGTGAGGATCACTGGATGATGGCACACGGATGA
TCAGTGCTCCAAATACTCCATCTGCCCGCTGACAACCAGAATGTGAATGCCAGAAGTGGGTTCCAGGTGTGTCTGCTAT
GTAGTGGTACCGGAATGATGACTTTGGTGGGATGGGACACTGTGTGACATATGGAACTCCATCCATGTAAGGCGTTCCA
CGTTGGTGATGTCCATGCCAATGTATACTTGTGCTCTCCCCCATCAAGTGATTCTCCATATCCACTATTATGTGATCACCC
TTGCACACCTCGACACTAGGACCAGGCATCTGTCTGTTGACCACTATTATAGTGG

CC_004
GTCATTATTATATTTACTTTTTTATTCAAATTCTCATTTTCTATTCCCTCCTGCCTAAAATTCAGGGGTGAGGCAAATACC
TGTTTGCCTTGCCTGCCTTGCCAGTGTCCATGAACATACAGACACTCATTAATATCTGCAAATCCAGTAACATATTAAGA
GCAAGGGAACTGAAATGTGCAGACAAGATTGCATAGAATTAGTACATACTTGACAACTCTCAACCAGAGGCACACGCC
ATAAAACAGCAGCTTAACGTATCTGCTGGTGCTGTCTGTTACAAGTTCATAACAAAACAAAATACTTACTGAGTTTATCT
CTTGTTGAATAGTCACAGTTTTTCCAACTTGGCCTCTGAGCTTAATTAAAGAGCACAGGGGTTGTTCAAGAACAGGGTA
CGGAGCAGAATATTTGAACATAAGGCAGGGGACATGATAGGAGAATGGAGAAAACTGCATGACAAAGAACTTCATAA
TTTGTACTCTCACAAAATATCATTATGGTGATGACCCTATGATGAACTGAGGAGATGAGAAATGCACACAAAAGTTCAG
TTGGAAAACTTGCAGGGAAGAGCTTACTGGGGGGACCTAGACATAGATAGGCAGGTAATATTAAAACGGATCTTAAGG
GAATACATTGCGAGGATGATGGGATTCATCTGGCTCAAATATGATCTAGTGGAAGATACTTGTGACCACTGTAATGAAT
TGTCAGTTTTCATGAAAGGCAGAATGTTTGTCATGTGGGTAACTACAACCTTTCCCACAAGAGCAATGCTCCATGTAGT
ACTGCCACTTTCAGGTCATTTGGAATATGGACAGGCAGACAGGCTGACATAGAGCTTGCTAATACATTTCCAGACAATG
AACTCTTGTTTCTCCGTAAACCTGGGCAAGAATGCATGAAAGGAATACCTGTCACCTGGAAATTTTTCTGCTTGCAATTT

TTAGGGAGAGAAATTGCAGTACACCTTATGTGTCTGTGGAGCAGATGTGGTCACCGAGCAGCTGCACAGAGCAATAAC
AGGGGGATAAGGTTTGGTAACACAGTTATCTTGAATGATCCACTGCCAGAAGTTGATTTGAGAGGCCACCAGTGTCGAG
TAGAGCTGATATGGATGTCAACAGCATTGTCCACGTCACTGGTTGCTTGTTCAAGGTCTTGTTCAGATGGCATGTAGTTG
TTGCAGCGAGGAAAGCCAGGTGGAGAGGATGCAAATTCTTCATGTTGACCCACCTTGAAAACAACAGCCATGCCAAGT
TCCACATGGAATTCAATGTGACAGTGAAACAACCAGTAACCTGGATTACTTGCATGGACACGCACGATTGTAAAACCTC
CATCCGGAACGGTGATAGTGTCTTTGATTGGTGCGTTTTTAAGCTTCCGATGGATCAGACCTCTTTTATCTCTGGCTTCA
ACTTCTTCTACTGTTGTGGTGCTGTTAAGACGTTGCATGGCTATCACGCGGAATGCATGGCCATGCAGGTGGAAAGGAT
GATTTGCATCAAATGCCTTGCCTTTGTC

CC_005 CCAGGGTATGTTAACAACTCTCGTTAAAGTTTTCAGTACCAAAGGTCCTCTAAGGTACAAATGTATCAAAGGACGGTCG
ACATTTTAGAAGAATGCTGCAATAATGCCTTTCSYCKTTTRWKRWKSRWYTTAWYWTMTAAAGWYAGMSWMCRYRM
MTRYWTAYTMACAYGGCCCWAMSAGSKTWTRTYRCACAAYTYTMKTYAARGWWWYMWGTRCCMWWRGWCYGAT
MYAAGACTGAACGTCTGCAGTATGTGGCTCTGCGTCCTGCTTGCCTGCGCAATTGGAATGGCTTCTGCAGCGACATCAG
TGCTCCTGAATTCATACCTTCAGCCCAACGATGACATTGATAGAAACGCGTACCTCCTAAAAGCACAGAGCAGCAACTG
TGCCCGTATGTGCAACGCGAGAGAGGCACCCAAAATCTGCTACTACCGATGGATAGTTGAGAACTACGTAACAGTGTC
AGGCGCGTGTGGCAACTGTCCCGTGAATACGACGGCCTGTTACAACCCACAGTGCATCACTGCCGATGGATATGAGCGC
AGTATTCTTTCGATAAACAGGAAGTTACCGGGACCTTCCATTGAGGTGTGCCTCAGAGACAGAGTAATTGTGGATATAA
CCAACAACATGGCTGGGCGGACTACAGCCATCCACTGGCATGGGTTATTTCAGAAAGGGTCCCAGTACATGGACGGCG
TTCCCATGGTAACCCAGTGCTCTATACATGAGGGTGACACATTTCGGTACGACTTTATCGCTAACAACGAGGGAACTCA
TTTCTGGCATTCCCATGACGGTTTGCAGAAGCTCGATGGCTTGGCAGGTAACTTGGTGGTTAGGGTGGATGTAAATGAC
GACTCGAACGGAAAAATGTACGACTTTGACCTACCAGAACACAAAATTTTCATCAGCGACTGGCTACGTCTTCCTGCAG
ATGACCACTTTCCCGGACTCCGAGGGGCAAATTCAGGACAAGATGCCGACTCCTTTCTCATTAACGGGAGAGGAAGTAC
CTCGATTGGAACTCAGTCCACCAACACACCTTATGCACAGATAAAGGTGCAGTGGGGCAGGAGGTATCGGCTTCGCATT
GTAGGAGCCTTGTGCACTGTGTGCCCCACACAGCTCACAATTGATGGGCACAAAATGACCGTCATAGCTACCGACGGCA
ACCCTGTGGCTCCTACCAGAATCGACTCCCTTATCATTTACTCTGGTGAAAGATATGATGTCGTGTTAGAAGCAACTAAG
ACGGAAGGATCTTACTGGATCCACCTAAAAGGCCTCGCCACTTGTGTTATAACCAGAGTTTACCAGCTGGGAGTGCTGC
AGTATGAAAATACAACAACCAATAACCTCCATGCTCTGACCCCTGATCCAGGGTACGACGGATTCCCGCAACCAGCAGC
ATACCGGGTCCTGAACCCAGAGAACGCAAGCTGCAGCGCCGGCTCGACAGGCTTATGCGTCACGCAACTGGCGAACTC
GGACCCCGTGCCACGGGACATCCTAACCCAGCTCCCGGACGTTAACTATCTTCTCAACTTTGGATTTGAAACTTTCGAGT
CCAAAACTTTCTTCACTGGTTACGACAGATATTTTGTCAGCCCCCTTCTCGAGTTACTCAGCAGTACCGTCAACAACATT
TCTTTCGTTGCGCCCCCATCTCCGCTCCTCTCACAAAGGGGAGATATACCAGACGACATCATATGCCCGACGGCGGCTG
ATGGCCTACCCCAGTGTCCCGGAGGAAAATCCTACTGTACATGTGTCCATGTCATCAAAATCAAACTGGGTGCTTTGGT

ACAGATCATCCTGTCAGACCAGACACCCAAATCCGGCCTGAACCACCCGTTCCATCTACATGGACATGCGTTTTACGTC
ATGGACATGGGGCAATATGCCTCAGGACAGACAGCGGAAAGCCTCCTTAACTCTTTGAAGAGTAATGTGACGAGAGTG
TCCCCTGCGCCTGCTCTTAAAGATACCATCGCGGTTCCATCTGGCGGCTACGCGGTCATCAAGTTCAGAGCAAGAAACC
CTGGTTTCTGGTTCCTTCACTGTCACTTCCTCTACCATATAGCGACCGGGATGGCTGCTGTATTCCAGGTGGGAGAAACA
AGCGACTATCCCCCTGCACCAGACGGTTTTCCGAAGTGTGGAAACTTCACACCTCCAGTGACCACCAATTGAAGTCTGA
ACGTCATCTGACGAATTGTGAAAACCTAAAAAGGTTTTTATTACGACGTCCCTCGGTGGAAAATGTAGTGACTTAGTAT
GAAAGAACATACTGTAAAAGGAATTAAATAGTTTCACTCTGTTAAGCCGAACGTCGTGGTGTTGTGGTTCGCAAAATTG
AATAGAAATTGGGAGGTCCCGGATAAGGGGCTCGACCCAGACTGGGGGTATACGGAGATGTTTAACGGTTGTCTTGAG
TGTATCCACTGATTCTGTAATTCGGTGCTTCAAGTATATGCAGATAGCGTTAATACCCTTATATAAACGAACAAATAAAT
AAACAAATAAAACTTGTCCTAAATTTCAAGGAAGCTTTCTAACAATACGGCTGTCAAGAAAGGTGTACATATATACTCA
GCTTCCACGCTGATGGTGATATCTAAATGACGAATCGAAGTAGCATGTGAAGTGTATGGAGACAGCGACAGAGGGAGA
GAGGTGCTTTATTTAATGAAGTTGTCAATTGTTAAATTTATACAGCATCAATGAAAATGAATGAACTATGTGCATAGAG
CGTTGTTAGAATGATACTGATAGAGGAAGGCCGACGTATTCAAAGAAAACCCTGTCACATTTACTCATAAAAATCACAC
ACGGACTGGAATGTTATCGAATTCAAGCCTCCGCAGTGACAGACCGGTGACTAATCCTTTGAAATATTACATCCTGTTCT
TTTAAAGGTAGCTAATAACCGTAGCTATAAATTCGGTATCAAATACGATATTGAGTATTGTTGTCACGCCAGCAGTTAC
AAGAATGGCGACCGTGCGAAAATGTCGGCTTATACCTGTCAAAATTAAGCTAGCTGCAGCGACAATTTTTACCAGCTTA
AAGATGCACATAAATAAATCTCGTGGGGTAGAGTCGTTTTTAAAAGGAAAACAGTCTCACATCCGGTCAACATCCACCG
CCTTCAGTGGAATCCGCCGGGTCATTACTGTGTATACAGAAGGCCACCACCGGCCCTTACTCTGAGCCAGCTGCATCTA
GTCGAGCCACAAGGGCCAAAAGACTGTTCCATGAAATCTCTTTGTAACATCCTTCTCCAGCCACAGTCGTCTCTTCCTTG
TAAACTCAACCGAAAGTTTGTACTTTGTTTGGTACTTTTTGATGCACGTTTTAGGATCCGTCCATCTTATCTTTCGTAATT
TGATCGCCGTAATAATATGGTGTAGGGAATAAATTACTAAAATCATCAGTGTGAGAACTTCACCATTCCCCTGTCGGTT
CTTCTTTCTTAGGTTCACACATTCTCCTCAAAACCCTCTTTTCATACAGGCATTCCTTATGGTAAGGGACCAACTTATACA
CGTTGTTGTTTAAAAAGTTCGTAATTACTAATTTATGCTAAACTCACAGTACAGATTAAGAAATCTAGGGCACGTTGTCG
TCATCAGTTTTTCCCCGGACTTATTTGTAGACTACAGACTTCTCTAGTCCCAACAAAGCGCTTAATTTTCTTGTCATCTCT
TACTAATAAGTAATTATTTCGTGGGTTCATCGATATTTTACGTTCCTTAAGATACGATACGTTTTGATCTTCTTGTTACTA
TTCTTCTAAACACTGAGT

CC_006 CAGGAAAGGACTTCAAGTAAATTCACTGAACATGGGGACCGGCACTATTGGATCATTGGTCATTCCTGAATTGAATGCA
ATAGCCAATGACGATGAAAGTCTAAAGGAAGAACCAGACTTAAAATTTTATTTGTCGTATGATTTCTACGAAATAGATA
ATCCTCATTACCACCGTTCACCATTCTATGGCTTCCACCAAGTTTCCAGCAAATTACAACGTCTGCTCACACCTCAGCTG
AACCATATATCATTAAGGCTGCCACCATTTCCACTTCTGTCACAGCATAGTGACCTTGGAACAAAATACCCTTTCTGTAA
TGAGACTACTGTAAGTGACTGTGAAAACAAGTACTGCGAGTGCACACATGTAATTGAGGTGCCACTGGGGACTGTTCTG

GAACTCATTATTATTGACAAAGGCAAGGCATTTGATGCAAATCATCCTTTCCACCTGCATGGCCATGCATTCCGCGTGAT
AGCCATGCAACGTCTTAACAGCACCACAACAGTAGAAGAAGTTGAAGCCAGAGATAAAAGAGGTCTGATCCATCGGAA
GCTTAAAAACGCACCAATCAAAGACACTATCACCGTTCCGGATGGAGGTTTTACAATCGTGCGTGTCCATGCAAGTAAT
CCAGGTTACTGGTTGTTTCACTGTCACATTGAATTCCATGTGGAACTTGGCATGGCTGTTGTTTTCAAGGTGGGTCAACA
TGAAGAATTTGCATCCTCTCCACCTGGCTTTCCTCGCTGCAACAACTACATGCCATCTGAACAAGACCTTGAACAAGCA
ACCAGTGACGTGGACAATGCTGTTGACATCCATATCAGCTCTACTCGACACTGGTGGCCTCTCAAATCAACTTCTGGCA
GTGGATCATTCAAGATAACTGTGTTACCAAACCTTATCCCCCTGTTATTGCTCTGTGCAGCTGCTCGGTGACCACATCTG
CTCCACAGACACATAAGGTGTACTGCAATTTCTCTCCCTAAAAATTGCAAGCAGAAAAATTTCCAGGTGACAGGTATTC
CTTTCATGCATTCTTGCCCAGGTTTACGGAGAAACAAGAGTTCATTGTCTGGAAATGTATTAGCAAGCTCTATGTCAGCC
TGTCTGCCTGTCCATATTCCAAATGACCTGAAAGTGGCAGTACTACATGGAGCATTGCTCTTGTGGGAAAGGTTGTAGTT
ACCCACATGACAAACATTCTGCCTTTCATGAAAACTGACAATTCATTACAGTGGTCACAAGTATCTTCCACTAGATCATA
TTTGAGCCAGATGAATCCCATCATCCTCGCAATGTATTCCCTTAAGATCCGTTTTAATATTACCTGCCTATCTATGTCTAG
GTCCCCCCAGTAAGCTCTTCCCTGCAAGTTTTCCAACTGAACTTTTGTGTGCATTTCTCATCTCCTCAGTTCATCATAGGG
TCATCACCATAATGATATTTTGTGAGAGTACAAATTATGAAGTTCTTTGTCATGCAGTTTTCTCCATTCTCCTATCATGTC
CCCTGCCTTATGTTCAAATATTCTGCTCCGTACCCTGTTCTTGAACAACCCCTGTGCTCTTTAATTAAGCTCAGAGGCCA
AGTTGGAAAAACTGTGACTATTCAACAAGAGATAAACTCAGTAAGTATTTTGTTTTGTTATGAACTTGTAACAGACAGC
ACCAGCAGATACGTTAAGCTGCTGTTTTATGGCGTGTGCCTCTGGTTGAGAGTTGTCAAGTATGTACTAATTCTATGCAA
TCTTGTCTGCACATTTCAGTTCCCTTGCTCTTAATATGTTACTGGATTTGCAGATATTAATGAGTGTCTGTATGTTCATGG
ACACTGGCAAGGCAGGCAAGGCAAACAGGTATTTGCCTCACCCCTGAATTTTAGGCAGGAGGGAATAGAAAATGAGAA
TTTGAATAAAAAAGTAAATATAATAATGACAAAAAATACATTTGTTTTTAAAAGATTTCATTCCACCGAG

CC_007 CTTTCGAAAAGATTGTGTTGTGGCGGACGGTGTAAGGCGTCCTATAATAGTGGTCAACAGACAGATGCCTGGTCCTAGT
GTCGAGGTGTGCAAGGGTGATCACATAATAGTGGATATGGAGAATCACTTGATGGGGGAGAGCACAAGTATACATTGG
CATGGACATCACCAACGTGGAACGCCTTACATGGATGGAGTTCCATATGTCACACAGTGTCCCATCCCACCAAAGTCAT
CATTCCGGTACCACTACATAGCAGACACACCTGGAACCCACTTCTGGCATTCACATTCTGGTTGTCAGCGGGCAGATGG
AGTATTTGGAGCACTGATCATCCGTGTGCCATCATCCAGTGATCCTCACAGACAGCTGTATGACCAAGATCTGTCTGAA
CACACTTTGCAAATCCTGGACTGGGATGACAAGCTGGGAGTGGACAAGTTCCTTGCTCACCATCATGCCAATGGCAACA
ACAAACCAAAGACCCTGTTGGTTAATGGCAGGGGTCGCTTCCACCAATTTCAGGATGTAAATTCTATCACTGTAAATGT
GACACATTCACCAACTGCAACATTCACTGTTCAGCagggcttccgctacAGGTTCCGCCTTATAAATGCAGGCTTCCTCAACTGT
CCTATTGAACTATCAGTGGATAATCACACCCTGTTAGTGATCAGCAGTGATGGAAGTGATATACAGCCAGTGCAAGCTG
ACTCTCTGGTGAGCTATGCTGGAGAAAGATTTGATTTTGTGCTGCATGCAAATAAAAAAGTTGGCAACTACTGGATGAG
ATTCCGAGGGCTAATGGATTGTGATGAACGTTTCACTAAAGCTCACCAGGTTGCTGTGCTTCATTATGTTGGTGCCCCTG

	TGAATGAAGAACCACCTGGAGAGGTGAGCTACAAGGAAGCTCGCAGGAAAGGACTGCAAGTAAATTCACTGAACATG GGGACCGGCaCTATTGGATCATTGGTCATTCCTGAATTGAATGCAATAGCCAATGACGATGAAAGTCTAAAGGAAGAAC CAGACTTAAAATTTTATTTGTCGTATGATTTCTACGAAATAGATAATCCTCATTACCACCGTTCACCATTCTATGGCTTCC ACCAAGTTTCCAGCAAATTACAACGTCTGCTCACACCTCAGCTGAACCATATATCATTAAGGCTGCCACCATTTCCACTT CTGTCACAGCATAGTGACCTTGGAACAAAATACCCTTTCTGTAATGAGACTACTGTAAGTGACTGTGAAAACAAGTACT GCGAG
CC_008	GTCCAAGCGTGACATAGTCTTCGGCGAACCATTTGTAGTAGCATATCTTTGGGGCTGCTCCAGCTACGCACTGCCTGGC GCACTGGTTGCTGTTGGTTGGTGTTAAGAATTGCTGGCGTGTTAGACGATCTGGCTGCAGGTAATTCTCTAGGTTAAGTC GTGTTGCAGCCGACACTAGACGGGTGTCAGTGATGAGCAATACCAAAAGAAGACATAGACACATTTCTGCGGAAACCA GGGTTTGCCTTCCTTGGCCTGCCATCTCAG
CC_009	CAAAGACTGGGCTTCCTGTCCCCTACAAGGAGGTGTGGGTGGAGCGTGGAAAGAGATTCCGTCTACGTCTTGTGGGTGG CCTGTGCACAGTGTGTGGTGTCCAGTTCTCCATAGAAGGTCATGACATGACCCTCATCGCAACAGACGGCACGCCCGTC AACCCCGTCACTGTCAGATCCGTCGTCATATTCTCAGGCGAAAGATACGATGTCGTGGTGAATGCGAACCAGAATCCGG GCACATATTGGATTCACATTAAGGGACTGTCAGAATGCGCCTTTCCTGAAGAACAAGTGTATCAGCTTGCAGTTCTCAG GTACCTGGGCACCAACCAAGCGCGTAACGACTTCCCTACCTACAATGGTGGCTTCGGAAGAGGCGGGACGGTTGTGAA CCCCCAGAACGCCACGTGTGCCAG
CC_010	GTGGACAAGGCGGTGTGTGCGTTTCTCAGTTGGTGGGTTCAATCCCCGACAATCAGAACGTCCTCAACAGTTTTCCTGAT GTGAACATTGTGCTCAAGTTCGGATTCTACCTTTTCCGTGATCTGCAGCGGTCTTTCAACAACGGAAGATACGACAGGTT CCTTGTGGCCCCTACTAGGGACGTGGTGACCAGCTGGATGAACAACGTCTCCTTCGTTTCGCCACCCTCACCCCTCCTAA GCCAAGGTCCCGACATACCCACAGAATTGTTCTGCCCCATCGGAAGAAATGGTTTCCCTCAATGCCCGGCCAACAAACC CTCGGATGGCTATTGTGAGTGTGTGCACGTCATCAAGGTGCCACGCGGAGCAGTTGTGCAGATGGTCATCGGTGACGCG ACACCAGCAGTAGACCTTCACCATCCTTTCCATCTCCATGGATATGACTTCTTCGTCATGGCCATGGATCAGTTAAGAGG TGGGGATACCCTGGAGAGCATTAGTGCAACGCTCTTGGCAACGAACTTCAGACGGTCAACACTGCCTGCCCGCAAGGA CACCATAACTGTCCCCTCCAACGGGTACGCTGCTATAAGGTTCAGGGCAGATAACCCAGGCCTCTGGTTCTTCCACTGC CACTTCATGTATCATCTGGCCACTGGCATGTCTTTAGTTCTTCAAGTAGGAGAGGACGGTGACTGGCCACCAGTTCCAA GGAATTTCCCTAGATGTGGCAACTATCTGCCCACGCCATATTTGGAAGACTATGGCAAGTAGTTCAGGACTGACGTCGA ACTATCTGCTTTCCCCTTTTTCGCCCACTTTGAACGGCCACAAATCTGAGTCCGGAACTTATATCAATGCCTGTGGTTAC AGCAATTGCACTTGCAATGAAGAAAACAAAAAGATGTAATATATTTTCATTGTCAAAAAAGAGGCAGAATTAAAAC

CC_011 GTACACGTCTAGTTCTATCATGACAGTGATTTATTGCGTATTACAGCACCATTTTCGAGAGTACCGGCCAACCTTCATAT
TACAACACCATGGGTTGATAAAACTGTATCATACACACATAAAAATCTGATGTATGTATTCCTGTAAAAAACACAGTGC
CATGAGGAAACCTGCAAACAGAAATACACTCAAATTTTCCTGCATATAAAGTCATGACACCATAGACATCCACAAACTG
GTGACACCGTCATGGTTATACCAATGTGTTTTCTGTGAAACCTGAATTGGTATTTGGATCTGTGTCCATTACCTTGAACA
CATATCCACAGCTGGGTATTTCTTAAG

CC_012 GGGACCTTGGCTTAGGAGGGGTGAGGGTGGCGAAACGAAGGAGACGTTGTTCATCCAGCTGGTCACCACGTCCCTAGT
AGGGGCCACAAGGAACCTGTCGTATCTTCCGTTGTTGAAAGACCGCTGCAGATCACGGAAAAGGTAGAATCCGAACTT
GAGCACAATGTTCACATCAGGAAAACTG

CC_013 GCGAGATTTATTTATGTGCATCTTTAAGCTGGTAAAAATTGTCGCTGCAGCTAGCTTAATTTTGACAGGTATAAGCCGAC
ATTTTCGCACGGTCGCCATTCTTGTAACTGCTGGCGTGACAACAATACTCAATATCGTATTTGATACCGAATTTATAGCT
ACGGTTATTAGCTACCTTTAAAAGAACAGGATGTAATATTTCAAAGGATTAGTCACCGGTCTGTCACTGCGGAGGCTTG
AATTCGATAACATTCCAGTCCGTGTGTGATTTTTATGAGCAAATGTGACAGGGTTTTCTTTGAATACGTCGGCCTTCCTC
TATCAGTATCATTCTAACAACGCTCTATGCACATAGTTCATTCATTTTCATTGATGCTGTATAAATTTAACAATTGACAA
CTTCATTAAATAAAGCACCTCTCTCCCTCTGTCGCTGTCTCCATACACTTCACATGCTACTTCGATTCGTCATTTAGATAT
CACCATCAGCGTGGAAGCTGAGTATATATGTACACCTTTCTTGACAGCCGTATTGTTAGAAAGCTTCCTTGAAATTTAGG
ACAAGTTTTATTTGTTTATTTATTTGTTCGTTTATATAAGGGTATTAACGCTATCTGCATATACTTGAAGCACCGAATTAC
AGAATCAGTGGATACACTCAAGACAACCGTTAAACATCTCCGTATACCCCCAGTCTGGGTCGAGCCCCTTATCCGGGAC
CTCCCAATTTCTATTCAATTTTGCGAACCACAACACCACGACGTTCGGCTTAACAGAGTGAAACTATTTAATTCCTTTTA
CAGTATGTTCTTTCATACTAAGTCACTACATTTTCCACCGAGGGACGTCGTAATAAAAACCTTTTTAGGTTTTCACAATT
CGTCAGATGACGTTCAGACTTCAATTGGTGGTCACTGGAGGTGTGAAGTTTCCACACTTCGGAAAACCGTCTGGTGCAG
GGGGATAGTCGCTTGTTTCTCCCACCTGGAATACAGCAGCCATCCCGGTCGCTATATGGTAGAGGAAGTGACAGTGAAG
GAACCAGAAACCAGGGTTTCTTGCTCTGAACTTGATGACCGCGTAGCCGCCAGATGGAACCGCGATGGTATCTTTAAGA
GCAGGCGCAGGGGACACTCTCGTCACATTACTCTTCAAAGAGTTAAGGAGGCTTTCCGCTGTCTGTCCTGAGGCATATT
GCCCCATGTCCATGACGTAAAACGCATGTCCATGTAGATGGAACGGGTGGTTCAGGCCGGATTTGGGTGTCTGGTCTGA
CAGGATGATCTGTACCAAAGCACCCAGTTTGATTTTGATGACATGGACACATGTACAGTAGGATTTTCCTCCGGGACAC
TGGGGTAGGCCATCAGCCGCCGTCGGGCATATGATGTCGTCTGGTATATCTCCCCTTTGTGAGAGGAGCGGAGATGGGG
GCGCAACGAAAGAAATGTTGTTGACGGTACTGCTGAGTAACTCGAGAAGGGGGCTGACAAAATATCTGTCGTAACCAG
TGAAGAAAGTTTTGGACTCGAAAGTTTCAAATCCAAAGTTGAGAAGATAGTTAACGTCCGGGAGCTGGGTTAGGATGTC
CCGTGGCACGGGGTCCGAGTTCGCCAGTTGCGTGACGCATAAGCCTGTCGAGCCGGCGCTGCAGCTTGCGTTCTCTGGG
TTCAGGACCCGGTATGCTGCTGGTTGCGGGAATCCGTCGTACCCTGGATCAGGGGTCAGAGCATGGAGGTTATTGGTTG

	TTGTATTTTCATACTGCAGCACTCCCAGCTGGTAAACTCTGGTTATAACACAAGTGGCGAGGCCTTTTAGGTGGATCCAG TAAGATCCTTCCGTCTTAGTTGCTTCTAACACGACATCATATCTTTCACCAGAGTAAATGATAAGGGAGTCGATTCTGGT AGGAGCCACAGGGTTGCCGTCGGTAGCTATGACGGTCATTTTGTGCCCATCAATTGTGAGCTGTGTGGGGCACACAGTG CACAAGGCTCCTACAATGCGAAGCCGATACCTCCTGCCCCACTGCACCTTTATCTGTGCATAAGGTGTGTTGGTGGACT GAGTTCCAATCGAGGTACTTCCTCTCCCGTTAATGAGAAAGGAGTCGGCATCTTGTCCTGAATTTGCCCCTCGGAGTCCG GGAAAGTGGTCATCTGCAGGAAGACGTAGCCAGTCGCTGATGAAAATTTTGTGTTCTGGTAGGTCAAAGTCGTACATTT TTCCGTTCGAGTCGTCATTTACATCCACCCTAACCACCAAGTTACCTGCCAAGCCATCGAGCTTCTGCAAACCGTCATGG GAATGCCAGAAATGAGTTCCCTCGTTGTTAGCGATAAAGTCGTACCGAAATGTGTCACCCTCATGTATAGAGCACTGGG TTACCATGGGAACGCCGTCCATGTACTGGGACCCTTTCTGAAATAACCCATGCCAGTGGATGGCTGTAGTCCGCCCAGC CATGTTGTTGGTTATATCCACAATTACTCTGTCTCTGAGGCACACCTCAATGGAAGGTCCCGGTAACTTCCTGTTTATCG AAAGAATACTGCGCTCATATCCATCGGCAGTGATGCACTGTGGGTTGTAACAGGCCGTCGTATTCACGGGACAGTTGCC ACACGCGCCTGACACTGTTACGTAGTTCTCAACTATCCATCGGTAGTAGCAGATTTTGGGTGCCTCTCTCGCGTTGCACA TACGGGCACAGTTGCTGCTCTGTGCTTTTAGGAGGTACGCGTTTCTATCAATGTCATCGTTGGGCTGAAGGTATGAATTC AGGAGCACTGATGTCGCTGCAGAAGCCATTCCAATTGCGCAGGCAAGCAGGACGCAGAGCCACATACTGCAGACGTTC AGTCTTGTATCAGTCTAAGGGCACATGTTTCCTTGAATAAAATTGTGCAATAAAAGCTCTTTGATGTTAGTATGCAGGCA CGGTGGCTGTCTTTAGAAGATAAGACGAATAATAAAAGACGAAAGGCATTATTGCAGCATTCTTCTAAAATGTCGACCG TCCTTTGATACATTTGTACCTTAGAGGACCTTTGGTACTGAAAACTTTAACGAGAGTTGTTRACATACCCTGGTAGGGCC GTGTGAATAAATATTTG
CC_014	GGTGGTGATAATGGTGTTAATTTGGGTTTTGATAACTTGTCGGTGGACAAACAATGTTTTTGGAAAGACTGCTACTGCCA TTTTCAGTGTACAATTGTTACTTATGCATTTAAATGGAGGTAGTGTGTGTCTCTGAAATATTGGTTGCAAATCTCAAACT ACACAGTGACAAAACCAAATCTTAAGTAGAAACAGTTTGCAGGTGATTGCCTGTACTTTAATATCTGTGCTTTTGCTGTG

In addition, two isoforms were found for contig CC-Laccase_001. The isoforms were different by one SNP (single nucleotide polymorphism) at site 220 where there were substitution between purines nucleotides, nucleoside triphosphates, adenosine triphosphate (ATP), and guanosine triphosphate (GTP). The SNP located within the coding region and three ORF (open reading frame) (Figure 1 & Table 6). This SNP leads to changes in amino acid where CC_Laccase_001_a had an arginine while CC_Laccase_002_b had a glycine at the SNP site (Figure 2 & Table 7).

Laccases are believed to play a vital role in lignin modification, which is important for cellulosic biomass degradation (Sigoillot et al. 2012). Lignin is known to contribute to the recalcitrance in enzymatic saccharification of cellulosic biomass (Floudas et al. 2012). Lignin structure modification is crucial to facilitate the biomass degradation in both the natural biomass utilization system and the biofuels industry (Sun & Scharf 2010). Therefore, laccases are deemed as very important enzymes in termites that feed primary on wood. From this study, it can be elucidated that *C. curvignathus* possesses a range of laccases repertoire that allow it to digest wood from either intact or partially hydrolyzed wood.

It is also of interest to investigate if the laccases produced by *C. curvignathus* were unique. In this study, the gene sequences of *C. curvignathus* laccases were analyzed for their phylogenetic distances with 377 laccases transcripts from 73 genus of living organisms covering plant, bacteria, fungi, and insect. The result is shown in Figure 3, 4 and 5. From the result, we can conclude that the laccase genes expressed by *C. curvignathus* were considerably diverse. For laccases that were expressed in all three cDNA libraries, which regarded as core laccases, are considerably distantly placed in the phylogenetic tree. The core laccase, CC_007 is closely related to laccases produced by a subterranean fungus-growing termite, *Macrotermes barneyi* (Figure 4). Whereas another core laccase, CC_001 was phylogenetically more closely related to laccase produced by the Pacific oyster, *Crassostrea gigas*.

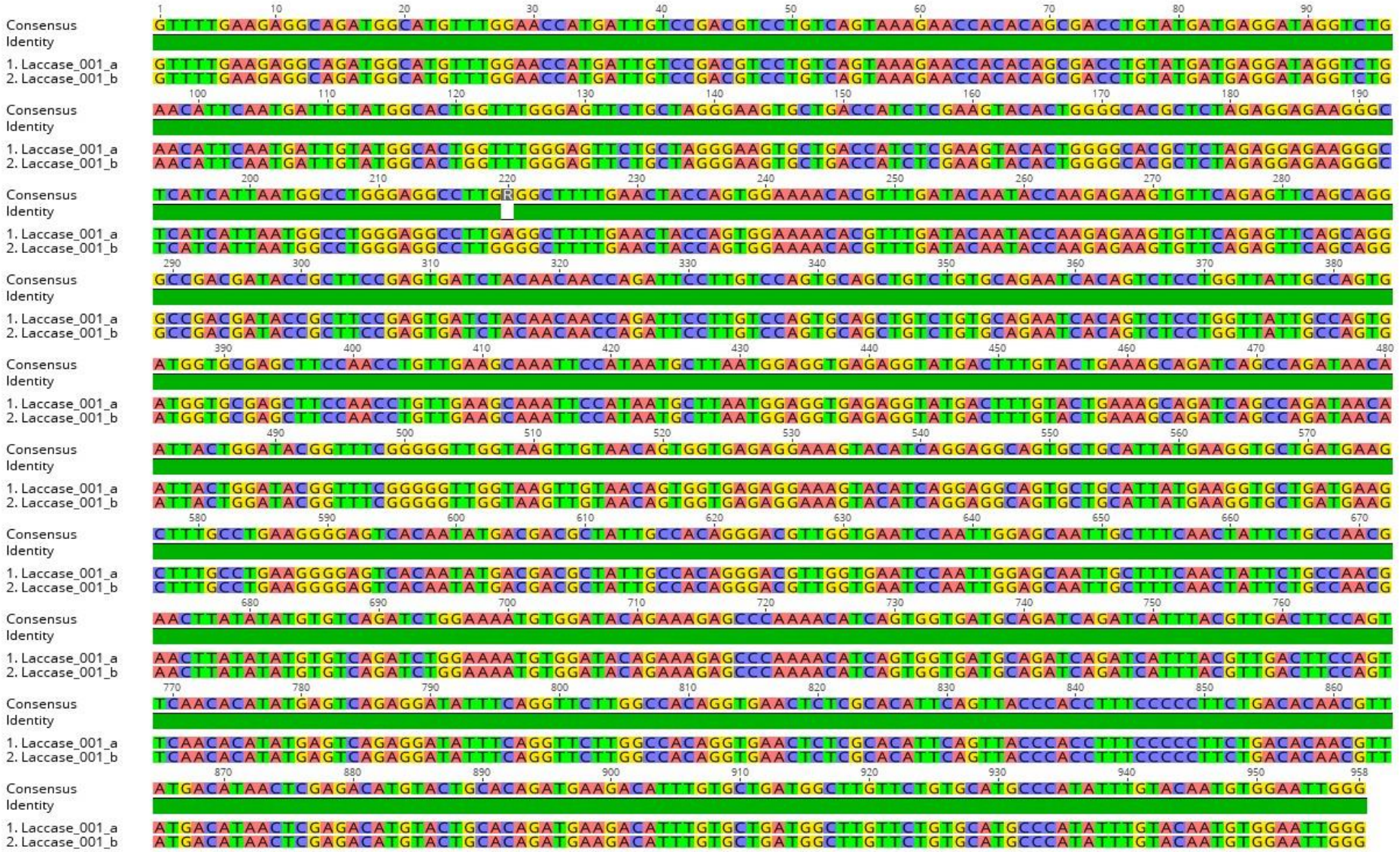

Figure 1. Multiple sequences alignment with ClustalW for CC-Laccase_001 nucletotide isoforms. The box shows the SNP site.

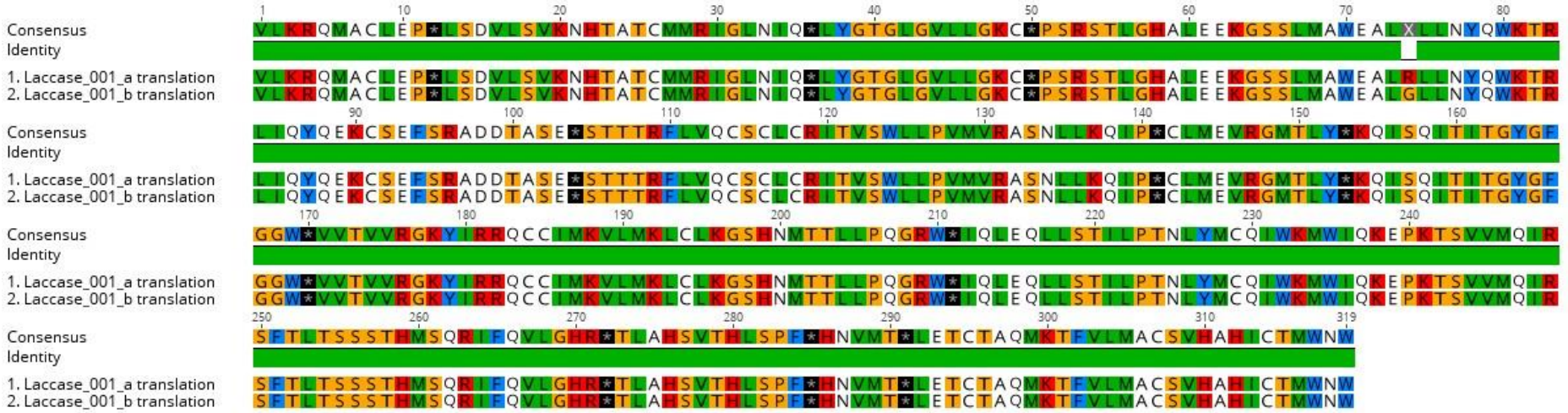

Figure 2. Multiple sequences alignment with ClustalW for CC-Laccase_001 amino acid isoforms. The box shows the changes of amino acid.

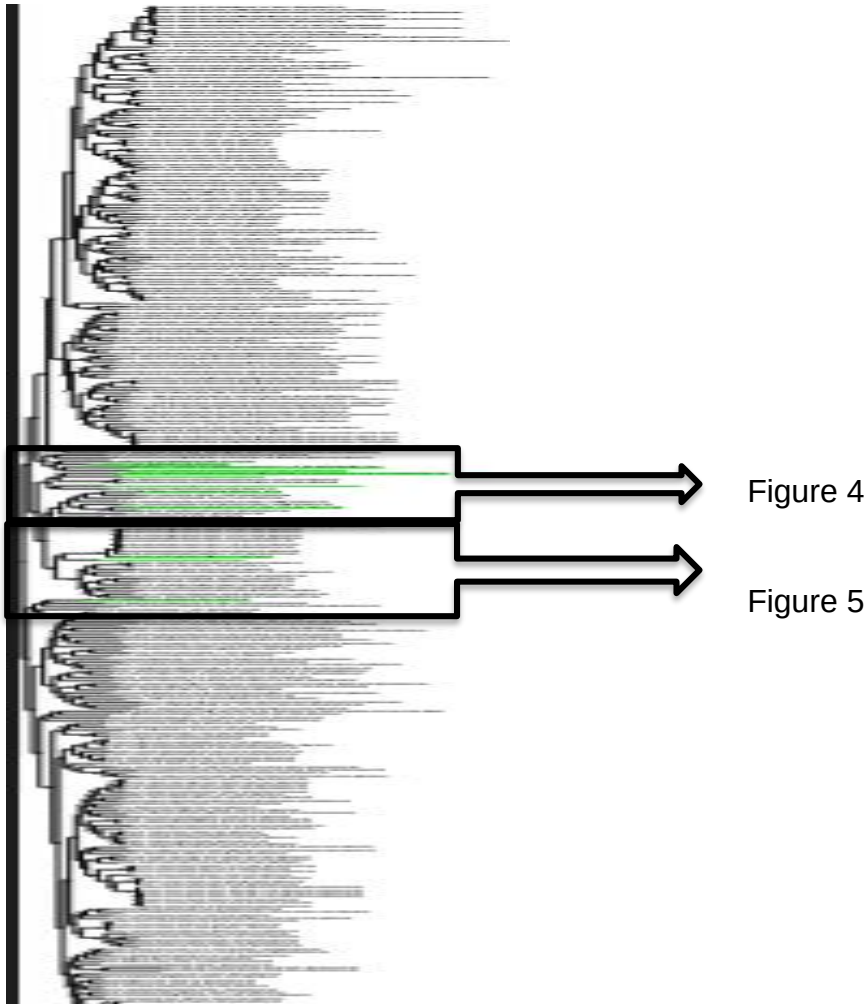

Figure 3. Phylogenetic tree for laccases gene generated by neighbour joining

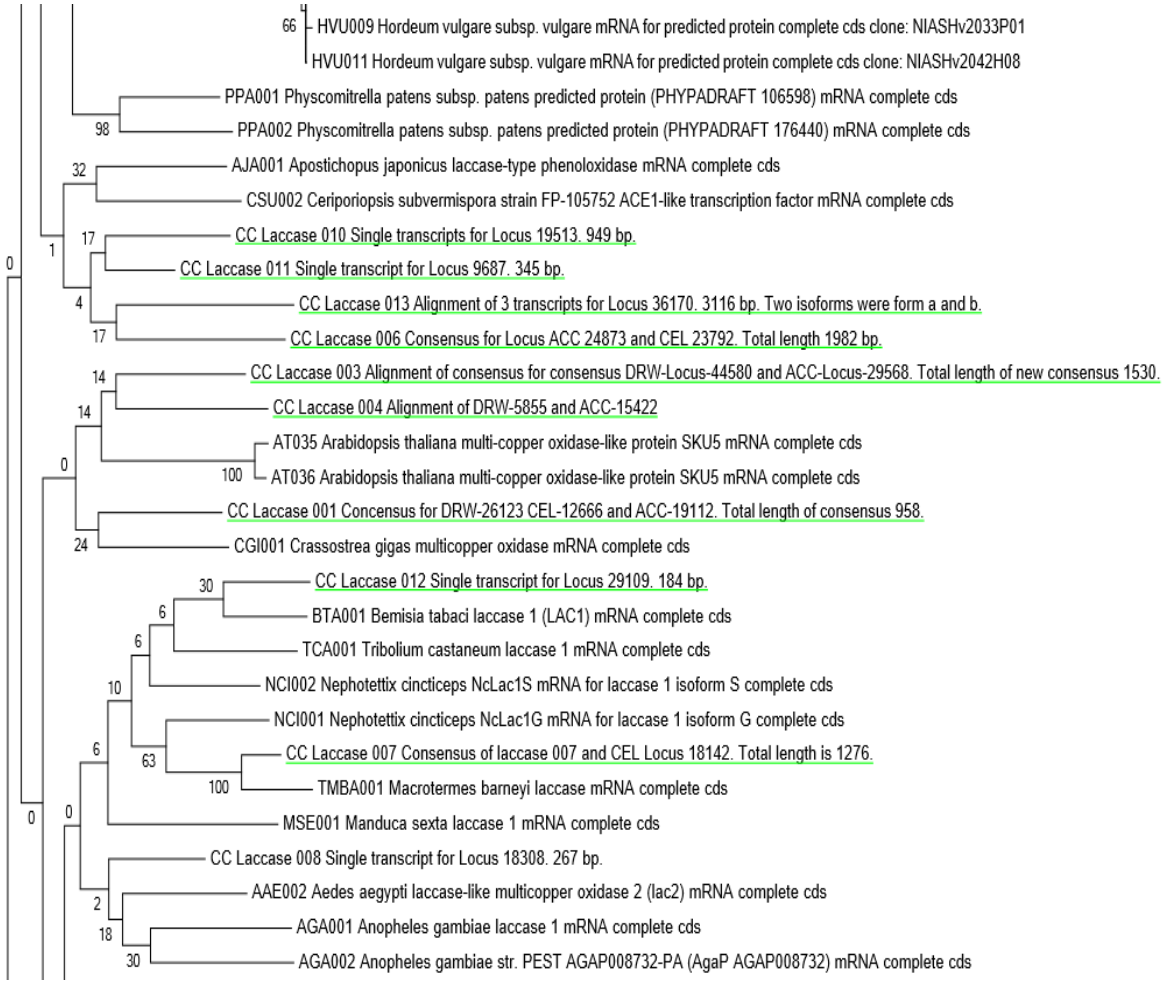

Figure 4. Part 2 of Phylogenetic tree for laccases gene generated by Neighbor joining

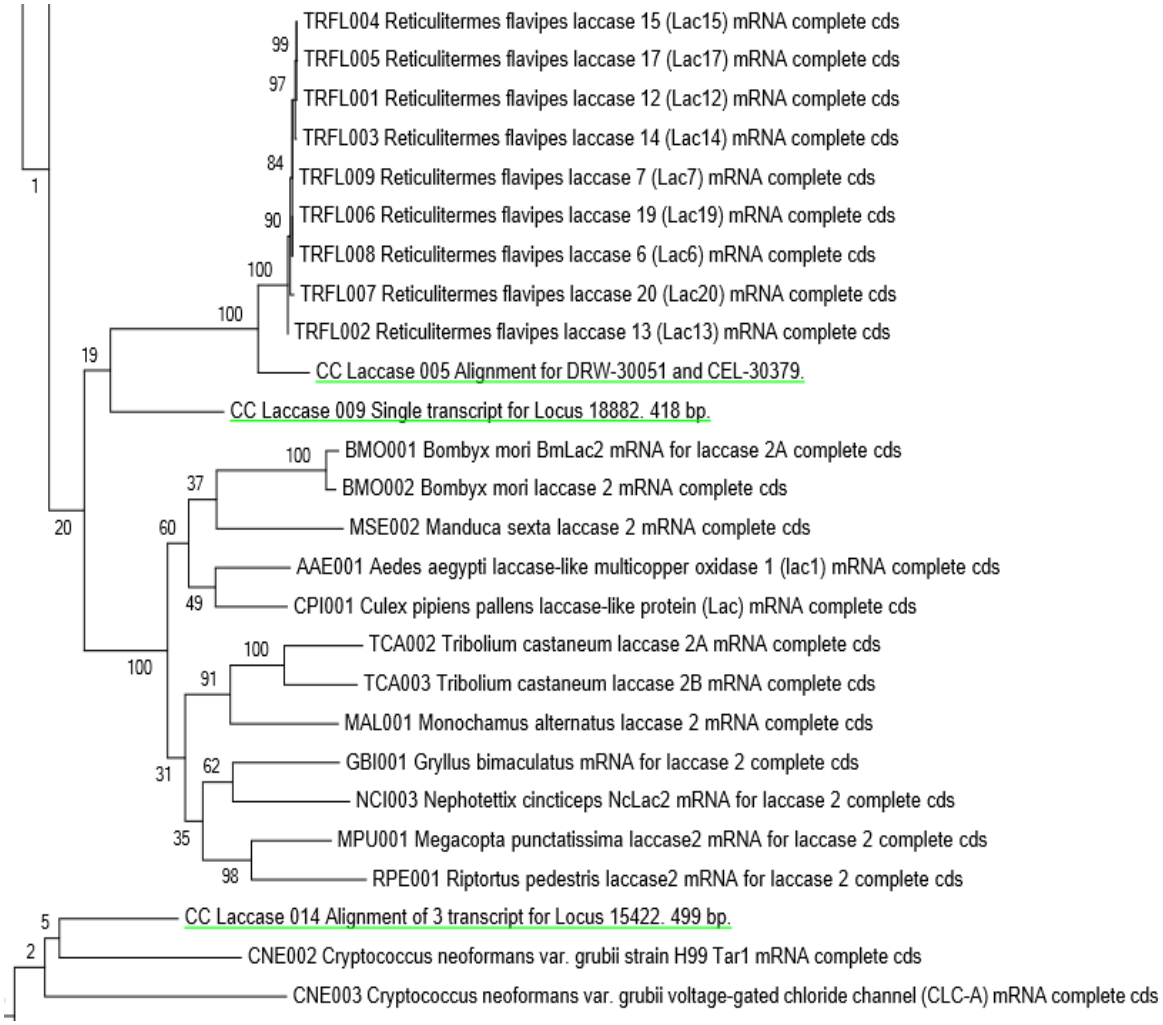

Figure 5. Part 2 of Phylogenetic tree for laccases gene generated by Neighbour joining

Nonetheless, the findings of this study also observed that the laccases of *C. curvignathus* are considerably conserved among termite species. Other than CC_007, CC_005 was another *C. curvignathus* laccase contig that is closely related to laccases from another termite species. CC_005 is closely related to laccase produced by termite of the same family Rhinotermitidae, *Reticulitermes flavipes* (Figure 3); In addition, based on Figure 4, it can be observed that CC_006, CC_010, CC_011 and CC_013 formed a monophyletic clade. Whereas CC_003 and CC_004 formed another clade. This indicates that the laccase genes are considerably conserved among termite species. This agrees with Geng et al. (2018) who characterized laccase from a wood-feeding termite, *Coptotermes formosanus* and noted that the characteristics were very similar to known termite laccases. The laccases produced by *C. curvignathus* are distinct from the fungal or bacterial laccases. The unique laccase repertoire of *C. curvignathus* may assist the termites to efficiently accomplish their degradation of biomass in hours instead of weeks as in fungal systems (Ke et al. 2012).

Referring to Figure 4, most of the *C. curvignathus* laccases are more closely related to insect laccase than either of plants, fungal, or bacterial. CC_008 had high similarity in their sequences to laccase produced by *Aedes aegypti*; and CC_012 was closely related to laccase produced by silverleaf whitefly (*Bemisia tabaci*), which is one of several species of whitefly that are currently important agricultural pests (Yang et al. 2017).

On the other hand, one of the *C. curvignathus* laccases contig is more related to bacteria laccases, CC_014 which is closely related to laccase produced by *Cryptococcus neoformans*, an encapsulated yeast that can live in both plants and animals (Figure 5). *Cryptococcus neoformans* was occasionally found in the wood infested by termites (Randhawa et al. 2003). It is not known if occasional ingestion of the yeast may facilitate the horizontal gene transfer between termite and the yeast. None of the laccases expressed by *C. curvignathus* are closely related to plant laccase. The wood-based diet preference of *C. curvignathus* does not promote any form of gene transfer in the case of laccase genes.

CONCLUSION

This study showed that termite, *C. curvignathus* expressed a diverse range of laccases that were phylogenetically similar to other termites' laccases and distinctly related to fungal or bacterial laccases. This assists the termites ability to achieve highly efficient biomass conversion.

REFERENCES

Assavanig, A., Amornkitticharoen, B., Ekpaisal, N., Meevootisom, V. & Flegel, T.W. 1992. Isolation, characterization and function of laccase from *Trichoderma*. *Applied Microbiology and Biotechnology* 38: 198–202.

Benfield, G., Bocks, S.M., Bromley, K. & Brown, B.R. 1964. Studies in fungal and plant laccases. *Phytochemistry* 3: 79 – 88.

Cox, M.P., Peterson, D.A. & Biggs, P.J. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. 2010. *BMC Bioinformatics* 11: 485.

Diamantidis, G., Effosse, A., Potier, P. & Bally, R. 2000. Purification and characterization of the first bacterial laccase in rhizospheric bacteria, *Azospirillium lipoferum*. *Soil Biology and Biochemistry* 32: 919–927.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martínez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., Aerts, A., Benoit, I., Boyd, A., Carlson, A., Copeland, A., Coutinho, P.M., de Vries, R.P., Ferreira, P., Findley, K., Foster, B., Gaskell, J., Glotzer, D., Górecki, P., Heitman, J., Hesse, C., Hori, C., Igarashi, K., Jurgens, J.A., Kallen, N., Kersten, P., Kohler, A., Kües, U., Kumar, T.K., Kuo, A., LaButti, K., Larrondo, L.F., Lindquist, E., Ling, A., Lombard, V., Lucas, S., Lundell, T., Martin, R., McLaughlin, D.J., Morgenstern, I., Morin, E., Murat, C., Nagy, L.G., Nolan, M., Ohm, R.A., Patyshakuliyeva, A., Rokas, A., Ruiz-Dueñas, F.J., Sabat, G., Salamov, A., Samejima, M., Schmutz, J., Slot, J.C., St John, F., Stenlid, J., Sun, H., Sun, S., Syed, K., Tsang, A., Wiebenga, A., Young, D., Pisabarro, A., Eastwood, D.C., Martin, F., Cullen, D., Grigoriev, I.V. & Hibbett, D.S. 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336: 1715–1719.

Geng, A., Wu, J., Xie, R.R., Li, X., Chang, F.X. & Sun, J.Z. 2018. Characterization of a laccase from a wood-feeding termite, *Coptotermes formosanus*. *Insect Science* 25: 251-258.

Götz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talŏn, M., Dopazo, J. & Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36: 3420–3435.

Ke, J., Laskar, D.D., Gao, D.F. & Chen, S.L. 2012. Advanced biorefinery in lower termite-effect of combined pretreatment during the chewing process. *Biotechnology for Biofuels* 5: 11.

Levine, W.G. 1965. Laccase: A Review. In Peisach, J. (ed.). *The Biochemistry of Copper*, pp. 371–385. New York. Academic Press Inc.

Madhavi, V. & Lele, S.S. 2009. Laccase: Properties and applications. *Bioresources* 4: 1694–1717.

Messerschmidt, A. 1997. *Multi-copper Oxidases*. Singapore: World Scientific Pub Co Inc.

Morozova, O.V., Shumakovich, G.P., Shleev, S.V. & Yaropolov, Y.A. 2007. Laccase-mediator systems and their applications: a review. *Applied Biochemistry and Microbiology* 43: 523–535.

Octavio, L.C., Irma, P.P.M.C., Ricardo, B.R.J. & Franscisco, V.O. 2006. Laccases. In Guevara-Gonzalez, R.G. & Torres-Pacheco, I. (ed.). *Advances in Agricultural and Food Biotechnology. India*, pp. 323-340. Kerala, India: Departamenta de Ingenieria Bioquimica, Instituto Technologico de Celaya.

Randhawa, H.S., Kowshik, T. & Khan, Z.U. 2003. Decayed wood of *Syzygium cumin*i and *Ficus religiose* living trees in Delhi/New Delhi metropolitan area as natural habitat of *Cryptococcus neoformans*. *Medical Mycology* 41: 199–209.

Schultz, M.H., Zerbino, D.R., Vingron, M. & Birney, E. 2012. Oases: Robust *de novo* RNA-seq assembly across the dynamic range of expression levels. B*ioinformatics* 28: 1086–1092.

Sigoillot, F.D., Lyman, S., Huckins, J.F., Adamson, B., Chung, E., Quattrochi, B. & King, R.W. 2012. A Bioinformatics method identifies prominent off-target transcripts in RNAi screens. *Nature Methods* 9: 363-366.

Solomon, E.I., Sundaram, U.M. & Machonkin, T.E. 1996. Multicopper oxidases and oxygenases. *Chemical Reviews* 96 : 2563–2605.

Solomon, E.I., Chen, P., Metz, M., Lee, S.K. & Palmer, A.E., 2001. Oxygen binding, activation, and reduction to water by copper proteins. *Angewandte Chemie International Edition English* 40: 4570–4590.

Sun, J.Z. & Scharf, M.E. 2010. Exploring and integrating cellulolytic systems of insects to advance biofuel technology. *Insect Science* 17: 163–165.

Tamura, K., Peterson, D., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: Molecular evolutionary genetic analysis using maximum likelihood, evolutionary distance, and maximum parsimony. *Molecular Biology and Evolution* 28: 2731–2739.

Thapa, R.S. 1982. *Termites of Sabah (East Malaysia).* Sabah: Sabah Forest Record.

Tho, Y.P. 1992. *Termites of Peninsular Malaysia. Malayan Forest Records No. 36*. Kuala Lumpur: Forest Research Institute Malaysia.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weigh matrix choice. *Nucleic Acids Research* 22: 4673–4680.

Thurston, C.F. 1994. The structure and functal of fungal laccases. *Microbiology* 140: 19–26.

Xu, F. 1996. Oxidation of phenols, anilines, and benzenethiols by fungal laccases: correlation between activity and redox potentials as well as halide inhibition. *Biochemistry* 35: 7608–7614.

Yang, C.H, Guo, J.Y. & Wan, F.H. 2017. Secretory laccase1 in *Bemisia tabici* MED is involved in whitefly-plant interaction. *Scientific Reports* 7: 3623.

Zerbino, D.R. & Birney, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 18: 821 – 829.

Table 6. ORFs and protein translations for contig CC-Laccase_001

Sequence	ORF number	Start	End	Length (nt)	Found at strand	Start codon	Start of sequence	Size (aa)
CC-Laccase_001	1	166	312	147	positive	CTG	LGHALEEKGSSLMAWEAL**X**LLNYQWKTRLIQYQEKCSEFSRADDTASE*	49
CC-Laccase_001	2	640	816	177	positive	TTG	LEQLLSTILPTNLYMCQIWKMWIQKEPKTSVVMQIRSFTLTSSSTHMSQRIFQVLGHR*	59
CC-Laccase_001	3	3	956	954	positive	TTT	FEEADGMFGTMIVRRPVSKEPHSDLYDEDRSEHSMIVWHWFGSSAREVLTISKYTGARSRGEGLIINGLGGL**X**AFELPVENTFDTIPREVFRVQQGRRYRFRVIYNNQIPCPVQLSVQNHSLLVIASDGASFQPVEANSIMLNGGERYDFVLKADQPDNNYWIRFRGLVSCNSGERKVHQEAVLHYEGADEALPEGESQYDDAIATGTLVNPIGAIAFNYSANELIYVSDLENVDTERAQNISGDADQIIYVDFQFNTYESEDISGSWPQVNSRTFSYPPFPLLTQRYDITRDMYCTDEDICADGLFCACPYLYNVEL	318
CC-Laccase_001	4	863	958	96	negative	CCC	QFHIVQIWACTEQAISTNVFICAVHVSSYVI	31
CC-Laccase_001	5	557	658	102	negative	TTG	LKAIAPIGFTNVPVAIASSYCDSPSGKASSAPS*	34
CC-Laccase_001	6	2	277	276	negative	CTG	LNTSLGIVSNVFSTGSSKA**X**RPPRPLMMSPSPLERAPVYFEMVSTSLAELPNQCHTIIECSDLSSSYRSLCGSLLTGRRTIMVPNMPSASS	91
CC-Laccase_001	7	829	957	129	negative	CCA	QFHIVQIWACTEQAISTNVFICAVHVSSYVITLCQKGERWVT	42
CC-Laccase_001	8	427	549	123	negative	CTG	LPPDVLSSHHCYNLPTPETVSSNCYLADLLSVQSHTSHLH*	41
CC-Laccase_001	9	795	956	162	negative	CAA	QFHIVQIWACTEQAISTNVFICAVHVSSYVITLCQKGERWVTECARVHLWPRT*	54
CC-Laccase_001	10	231	557	327	negative	ATG	MQHCLLMYFPLTTVTTYQPPKPYPVIVIWLICFQYKVIPLTSIKHYGICFNRLEARTITGNNQETVILHRQLHWTRNLVVVDHSEAVSSALLNSEHFSWYCIKRVFHW*	109

CC-Laccase_001	11	3	101	99	negative	ATG	MFRPILIIQVAVWFFTDRTSDNHGSKHAICLFK	33

Note: The bold 'X' in the contig and isoform indicate the unknown acid amino

Table 7. Nucleotide and amino acid sequences for CC-Laccase_001 isoforms

Contig	Isoform	Sequence
CC-Laccase_001 (nucleotide)	a	GTTTTGAAGAGGCAGATGGCATGTTTGGAACCATGATTGTCCGACGTCCTGTCAGTAAAGAACCACACA GCGACCTGTATGATGAGGATAGGTCTGAACATTCAATGATTGTATGGCACTGGTTTGGGAGTTCTGCTA GGGAAGTGCTGACCATCTCGAAGTACACTGGGGCACGCTCTAGAGGAGAAGGGCTCATCATTAATGGCC TGGGAGGCCTTGaGGCTTTTGAACTACCAGTGGAAAACACGTTTGATACAATACCAAGAGAAGTGTTCA GAGTTCAGCAGGGCCGACGATACCGCTTCCGAGTGATCTACAACAACCAGATTCCTTGTCCAGTGCAGC TGTCTGTGCAGAATCACAGTCTCCTGGTTATTGCCAGTGATGGTGCGAGCTTCCAACCTGTTGAAGCAAA TTCCATAATGCTTAATGGAGGTGAGAGGTATGACTTTGTACTGAAAGCAGATCAGCCAGATAACAATTA CTGGATACGGTTTCGGGGGTTGGTAAGTTGTAACAGTGGTGAGAGGAAAGTACATCAGGAGGCAGTGCT GCATTATGAAGGTGCTGATGAAGCTTTGCCTGAAGGGGAGTCACAATATGACGACGCTATTGCCACAGG GACGTTGGTGAATCCAATTGGAGCAATTGCTTTCAACTATTCTGCCAACGAACTTATATATGTGTCAGAT CTGGAAAATGTGGATACAGAAAGAGCCCAAAACATCAGTGGTGATGCAGATCAGATCATTTACGTTGAC TTCCAGTTCAACACATATGAGTCAGAGGATATTTCAGGTTCTTGGCCACAGGTGAACTCTCGCACATTCA GTTACCCACCTTTCCCCCTTCTGACACAACGTTATGACATAACTCGAGACATGTACTGCACAGATGAAGA CATTTGTGCTGATGGCTTGTTCTGTGCATGCCCATATTTGTACAATGTGGAATTGGG
CC-Laccase_001 (nucleotide)	b	GTTTTGAAGAGGCAGATGGCATGTTTGGAACCATGATTGTCCGACGTCCTGTCAGTAAAGAACCACACA GCGACCTGTATGATGAGGATAGGTCTGAACATTCAATGATTGTATGGCACTGGTTTGGGAGTTCTGCTA GGGAAGTGCTGACCATCTCGAAGTACACTGGGGCACGCTCTAGAGGAGAAGGGCTCATCATTAATGGCC TGGGAGGCCTTGgGGCTTTTGAACTACCAGTGGAAAACACGTTTGATACAATACCAAGAGAAGTGTTCA GAGTTCAGCAGGGCCGACGATACCGCTTCCGAGTGATCTACAACAACCAGATTCCTTGTCCAGTGCAGC TGTCTGTGCAGAATCACAGTCTCCTGGTTATTGCCAGTGATGGTGCGAGCTTCCAACCTGTTGAAGCAAA TTCCATAATGCTTAATGGAGGTGAGAGGTATGACTTTGTACTGAAAGCAGATCAGCCAGATAACAATTA CTGGATACGGTTTCGGGGGTTGGTAAGTTGTAACAGTGGTGAGAGGAAAGTACATCAGGAGGCAGTGCT GCATTATGAAGGTGCTGATGAAGCTTTGCCTGAAGGGGAGTCACAATATGACGACGCTATTGCCACAGG GACGTTGGTGAATCCAATTGGAGCAATTGCTTTCAACTATTCTGCCAACGAACTTATATATGTGTCAGAT

		CTGGAAAATGTGGATACAGAAAGAGCCCAAAACATCAGTGGTGATGCAGATCAGATCATTTACGTTGAC TTCCAGTTCAACACATATGAGTCAGAGGATATTTCAGGTTCTTGGCCACAGGTGAACTCTCGCACATTCA GTTACCCACCTTTCCCCCTTCTGACACAACGTTATGACATAACTCGAGACATGTACTGCACAGATGAAGA CATTTGTGCTGATGGCTTGTTCTGTGCATGCCCATATTTGTACAATGTGGAATTGGG
CC-Laccase_001 (amino acid)	a	VLKRQMACLEP*LSDVLSVKNHTATCMMRIGLNIQ*LYGTGLGVLLGKC*PSRSTLGHALEEKGSSLMAWE ALRLLNYQWKTRLIQYQEKCSEFSRADDTASE*STTTRFLVQCSCLCRITVSWLLPVMVRASNLLKQIP*CLM EVRGMTLY*KQISQITITGYGFGGW*VVTVVRGKYIRRQCCIMKVLMKLCLKGSHNMTTLLPQGRW*IQLEQ LLSTILPTNLYMCQIWKMWIQKEPKTSVVMQIRSFTLTSSSTHMSQRIFQVLGHR*TLAHSVTHLSPF*HNVM T*LETCTAQMKTFVLMACSVHAHICTMWNW
CC-Laccase_001 (amino acid)	b	VLKRQMACLEP*LSDVLSVKNHTATCMMRIGLNIQ*LYGTGLGVLLGKC*PSRSTLGHALEEKGSSLMAWE ALGLLNYQWKTRLIQYQEKCSEFSRADDTASE*STTTRFLVQCSCLCRITVSWLLPVMVRASNLLKQIP*CLM EVRGMTLY*KQISQITITGYGFGGW*VVTVVRGKYIRRQCCIMKVLMKLCLKGSHNMTTLLPQGRW*IQLEQ LLSTILPTNLYMCQIWKMWIQKEPKTSVVMQIRSFTLTSSSTHMSQRIFQVLGHR*TLAHSVTHLSPF*HNVM T*LETCTAQMKTFVLMACSVHAHICTMWNW