Reference-Aided Full-length Transcript Assembly, cDNA Cloning, and Molecular Characterization of *Coronatine-insensitive 1b (COI1b)* Gene in Coconut (*Cocos nucifera* L.)

Frenzee Kroeizha L. Pammit^{1*}, Anand Noel C. Manohar¹, Darlon V. Lantican¹, Roanne R. Gardoce¹, Hayde F. Galvez^{1,2}

¹Genetics Laboratory, Institute of Plant Breeding, College of Agriculture and Food Science, University of the Philippines Los Baños

²Institute of Crop Science, College of Agriculture and Food Science, University of the Philippines Los Baños

*Corresponding authors: flpammit@up.edu.ph

Abstract

In the Philippines, 26% of the total agricultural land is devoted to coconut production making coconut one of the most valuable industrial crop in the country. However, the country's multimillion-dollar coconut industry is threatened by the outbreak of coconut-scale insect (CSI) and other re-emerging insect pests promoting national research institutes to work jointly on developing new tolerant coconut varieties. Here, we report the cloning and characterization of *coronatine-insensitive 1 (COI1)* gene, one of the candidate insect defense genes, using 'Catigan Green Dwarf' (CATD) genome sequence assembly as reference. Two (2) splicing variants were identified and annotated – CnCOI1b-1 and CnCOI1b-2. The full-length cDNA of CnCOI1b-1 was 7,919 bp with an ORF of 1,176 bp encoding for a deduced protein of 391 amino acids while CnCOI1b-2 has 2,360 bp full-length cDNA with an ORF of 1,743 bp encoding a deduced protein of 580 amino acids. The 3D structural model for the two (2) isoforms were generated through homology modelling. Functional analysis revealed that both isoforms are involved in various physiological and developmental plant processes including defense response of plants to insects and pathogens. Phylogenetic analysis confirms high degree of COI1 protein conservation during evolution, especially among monocot species.

Keywords: coconut, coronatine-insensitive 1 gene, full-length transcript, functional annotation, genomics

Key Message

This paper reports the molecular cloning and characterization of *coronatine-insensitive 1 (COI1)* gene in coconut using reference-guided transcript assembly approach. As a well-known insect defense-response gene in other

crops, the results of this study are expected to assist in the development of new resistant coconut varieties as one of the strategies to address threats in coconut production.

Acknowledgements

This research was conducted through funds provided by Department of Science and Technology- Philippine Council of Agriculture, Aquatic and Natural Resources Research and Development (DOST- PCAARRD) through the Coconut Project 8 headed by Dr. Hayde F. Galvez. We gratefully acknowledge the Institute of Plant Breeding for allowing us to use the facilities for the conduct of this study. Also, the authors extend their sincerest gratitude to Dr. Susan R. Strickler and Dr. Lukas A. Mueller of the Boyce Thompson Institute for Plant Research (BTI), Ithaca, New York, USA for sharing their expertise in generating the high-quality reference genome assembly. Also, the authors would like to thank Ms. Cynthia R. Gulay for her technical assistance during the conduct of the laboratory experiments.

Introduction

More than 90 countries produce over 60M metric tons of coconut annually (FAOSTAT 2014). Currently, Philippines is the second largest producer of coconut around the world but remained as the top global producer of coconut oil (PSA 2017). However, the outbreak of coconut-scale insect (CSI) in 2014 which affected more than 2.6 million nut-bearing coconut trees threatened the coconut production in the country (PCA 2017). Various insect pest control methods were utilized in order to address the outbreak including physical, chemical, and biological control approaches.

Recombinant DNA technologies allow the effective cloning and characterization of candidate insect defense genes. However, most of the genes involved for insect defense in coconut have not been fully sequenced and characterized. One of the candidate insect defense response genes in coconut is *coronatine-insensitive 1* (COI1) gene which is reported to be involved in the jasmonate (JA) signaling pathway (Howe 2008). Jasmonate is known as an immunity hormone that promotes plant growth- and defense-related processes, such as, wounding and necrophitic pathogens (Zhai et al. 2017). In the JA signaling pathway, COI1 protein forms a complex with SCF (Skp/Cullin/F-box) to form an E3 ubiquitin ligase known as SCFCOI1 complex (Xu et al. 2002). This protein complex is necessary for the ubiquitination and subsequent degradation of JAZ proteins, transcription repressors in JA signaling pathway,

causing the de-repression of transcription factors and activation of JA response genes. To date, the only identified jasmonate receptor is the COI1 protein (Yan et al. 2013). In *Arabidopsis*, *COI1* mutants were found to be insensitive to JA resulting in defects in pest and pathogen resistance (Feys et al. 1994; Mewis et al. 2005). COI1 plays a vital role in wound- and JA-signaling since it was found to be required for the activation and repression of genes associated with wounding and JA signaling (Devoto et al. 2004). Similarly, COI1 in rice was reported to be indispensable in signaling component of rice defense response to chewing insects (Ye et al. 2012).

Several studies on molecular characterization of COI1 from dicots have been done but not much on monocots. In *Arabidopsis*, COI1 encodes 66 kDa protein (592 amino acid) containing N-terminal F-box motif and 16 LRRs which adopts the horseshoe-like shape (Xie et al. 1998; Katsir et al. 2008; Yan et al. 2013). There were three (3) closely related *COI1* homologs in rice (Lee et al. 2013) while only a single copy was identified in rubber tree (Peng et al. 2009). COI1 protein in *Aquilaris sinensis* is composed of 29.95% α -helix, 12.25% β -sheets, and 58.50% random coil (Liao et al. 2015).

With the advent of next-generation sequencing (NGS) and bioinformatics tools, the first whole genome sequence assembly of ‘Catigan Green Dwarf’ (CATD) variety of coconut has been generated using PacBio SMRT sequencing platform and 50X Illumina Miseq, and improved through Dovetail Chicago[®] sequencing (Lantican et al. 2019). With the aid of the genome assembly and information from transcriptome data, full-length complementary deoxyribonucleic acid (cDNA) can be obtained. Furthermore, defense and host-response genes such as *COI1* can be characterized at the genome level with the aid of the genome assembly.

The main objective of this study is to clone and elucidate the molecular characteristics of *COI1* gene in coconut. Specifically, this study aims to (1) generate the full-length cDNA sequence of *COI1b* from transcriptome data in reference to ‘Catigan Green Dwarf’ (CATD) genome sequence assembly, (2) clone partial *COI1b* cDNA sequence to validate the predicted full-length cDNA sequence of *COI1b*, and (3) to characterize *COI1b* gene and obtain phylogenetic relationships of coconut COI1b against known COI1 protein sequences.

Materials and Methods

Generation of *COI1b* full-length cDNA sequences

The full-length assembly of *COIIb* transcript was generated using genome-guided transcript assembly approach. The gene sequence of *COII* in *Elaeis guineensis* (NW_011551140.1) was downloaded from the database and used as input for BLASTn (Altschul et al. 1997) search for gene homolog in the existing build of coconut ‘Catigan Green Dwarf’ genome (Lantican et al. 2019). The *COII*-harboring genomic scaffold of coconut was extracted using a local perl script (https://github.com/solgenomics/sgn-biotools/blob/master/bin/fasta_extract.pl) and utilized as reference in the genome-guided assembly of *COIIb* full-length transcript. Sequence Read Archive (SRA) files from ‘Chowghat Green Dwarf’ (CGD; SRR1173229), ‘Taro Green Dwarf’ (TGD; SRR1273070), ‘Aromatic Green Dwarf’ (AGD; SRR1273070), and ‘West Coast Tall’ (WCT; SRR1137438) coconut varieties were retrieved from NCBI and preprocessed using Trimmomatic v0.36 (Bolger et al. 2014; SLIDING WINDOW: 5:30 LEADING: 5 TRAILING:5 MINLEN:85) to remove low-quality base calls. The preprocessed RNA-seq reads were aligned to the *COIIb*-containing coconut genomic scaffold using TopHat2 (Kim et al. 2013) aligner with default parameters. The resulting binary map alignment (BAM) files from each data set were used as input to the genome-guided Trinity software module (Grabherr et al. 2011) to assemble the full length *COIIb* transcript expressed across the four (4) coconut varieties.

Plant Material

Healthy and coconut scale insect (CSI)-infested coconut leaf samples were collected from CATD variety of coconut at the Institute of Plant Breeding, University of the Philippines Los Baños. Sterile distilled water and RNase AWAY™ (Invitrogen Corporation, USA) was used to wash sample surface and remove possible RNA-degrading contaminants.

Extraction of total RNA

Approximately 250 mg of leaf tissue was used for total RNA extraction using PureLink™ Plant RNA reagent following the manufacturer’s protocol (Invitrogen Corporation, USA). A total of 20 µL diethyl pyrocarbonate (DEPC) water was used to resuspend each sample. In order to ensure the purity of the samples, the isolated total RNA was treated with DNase (Invitrogen Corporation, USA) followed by 1% agarose gel electrophoresis in 1x TBE buffer (90mM tris-borate, 2mM EDTA) to check the RNA integrity. The quantity of the RNA was further determined using

Qubit 2.0 Fluorometer (Life Technologies Corporation, USA). The first-strand cDNA was synthesized using Superscript III First-Strand Synthesis System (Invitrogen Corporation, USA) following the kit's protocol.

PCR Amplification

The mRNA sequence of *COII* among different palm species was mined in the NCBI public repository (<https://ncbi.nlm.nih.gov>) and *Elaeis guineensis* coronatine-insensitive 1 homolog b (XM_010909322.2) cDNA sequence was retrieved. With the retrieved cDNA sequence, the gene-specific primers (GSPs) were designed using Primer3Plus software (Untergasser et al. 2012). The designed GSPs were subjected to *in silico* PCR using the CATD genome sequence assembly to determine complementation and specificity prior to primer synthesis. The primer pair exhibiting *in silico* amplification product with specificity to the target gene of interest was sent for outsourced primer synthesis (Invitrogen Corporation, USA).

The thermal cycler profile to amplify the mid-transcript of the target gene is as follows: 94°C for 2 min; 35 cycles of 94°C for 15 s, 62°C for 30 s, and 68°C for 1 min. The PCR reaction cocktail has a total reaction volume of 25 µL composed of 10X HiFi Buffer (1X), MgSO₄ (2mM), dNTPs (0.2 mM), 0.2µM each of forward (5'-GCATTGGAAGAGTTTGGTGGGGGCTCA-3') and reverse primer (5'-TCCAGCTTCTGCAGGCTTGGGCAAC-3'), Platinum® Taq DNA Polymerase HiFi (1U), and cDNA (30ng). The PCR products were subjected to 1% agarose gel electrophoresis in 1x TBE buffer (90mM tris-borate, 2mM EDTA) for 40 min at 100V to validate fragment size, and specificity and quality of the amplification product. Once validated, amplified cDNA fragments were excised from the agarose gel and further purified using the PureLink® Quick Gel Extraction Kit (Life Technologies Corporation, USA). The gel extracted cDNA fragments were eluted with DEPC water to a final volume of 35 µL.

Gene Cloning

The mid-transcript PCR product was ligated to pGEM®-T Easy vector (Promega Corporation, USA), following the manufacturer's instructions and transformed into *E. coli* JM109 competent cells. The transformants were screened through blue-white screening and T7/SP6 colony PCR. The positive transformants were grown overnight in Lauria-Bertani (LB) broth containing ampicillin (100µg/mL). The plasmids were isolated using the PureLink® Quick

Plasmid Miniprep (Life Technologies Corporation, USA) following manufacturer's instructions. Then, 3 µL of the purified plasmid DNA recovered was subjected to plasmid PCR using T7/SP6 sequencing primers to a total volume of 10 µL in order to detect the integrity of the recovered DNA plasmids. High quality plasmid extracts were sent for double-pass outsourced capillary sequencing (AIT Biotech, Singapore).

Raw sequence files (.abl and .seq file) that were generated from the capillary sequencing were pre-processed based on the sequence alignment of the designed primer pairs using the Vector NTI software (Invitrogen, USA). The trimmed sequences were then loaded to the CAP3 Sequence Assembly Program (Huang and Madan 1999) in order to generate the contig assembly from the double-pass paired reads.

Characterization of reference-aided full-length COI1b cDNA sequences and deduced amino acid sequences

Multiple sequence alignment (MSA) was performed on the assembled COI1b full-length assembled variants from CGD, TGD, AGD, and WCT and the corresponding CATD gene sequence of COI1b gene in CATD using Clustal Omega (Sievers et al. 2011). The structural gene annotations were manually done using SnapGene software (GSL Biotech; www.snapgene.com). Similarly, the cloned mid-transcript sequence was also aligned to the consensus sequences to further validate the COI1b gene identity. The cDNA sequences of CnCOI1 were also compared with sequences found in the database using BLASTn (Altschul et al. 1997).

The deduced amino acid sequences of coconut COI1b were derived by translating the mRNA sequences at different reading frame using the ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) and conserved domains were identified using NCBI Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). The composition of the amino acid sequences, molecular weight, and isoelectric point (pI) were predicted and analyzed using ProtParam (<https://web.expasy.org/protparam/>). The 3D molecular models were constructed through homology modelling using I-TASSER (Zhang 2008; Roy et al. 2010; Yang et al. 2015) and functional analysis using COFACTOR and COACH (<https://zhanglab.ccmb.med.umich.edu>). The coconut COI1b amino acid sequences were also used for BLASTn search (Altschul et al. 1997) and retrieval of homologous COI1 proteins from other species in the NCBI database (<https://ncbi.nlm.nih.gov>). Multiple sequence alignment using Clustal Omega (Sievers et al. 2011) was also performed followed by phylogenetic tree construction using FastTree (Price et al. 2010) to establish protein divergence and relationships among the COI1 amino acid sequences.

Results

Generation and validation of full-length COI1b cDNA

Multiple sequence alignment of the *COI1b* full-length assembled transcripts showed two (2) splicing variant detected in CGD and TGD while only one (1) splicing variant for WCT and AGD. CGD assembled transcript 1 and TGD assembled transcript 2 have 7590 bp and 7919 bp transcript length, respectively while all the remaining transcripts are around 2000bp in length. All the varieties have 5'-3' orientation except for WCT. Manual annotation from the transcript assembly showed that there are two (2) *COI1b* splicing variants in coconut (**Fig 1**).

The cloned mid-transcript sequence (752 bp; **Appendix 1**) of coconut *COI1b* gene was aligned to the two (2) *CnCOI1b* splicing variant in coconut generated using genome-guided transcript assembly approach (**Fig 1**). It covers approximately 10% of *CnCOI1b-1* cDNA (7,919 bp) and 32% of *CnCOI1b-2* cDNA (2,360 bp). The cloned mid-transcript sequence also showed 96% similarity with *E. guineensis COI1b* as revealed by Blastn results. No sequence variation (**Appendix 2**) on the cloned mid-transcript was detected for healthy and insect-infested leaf samples.

Fig 1 Visualization of the consensus cDNA sequences of CnCOI1b-1 and CnCOI1b-2 in reference transcript assembly. The arrow indicates the direction of transcription with the gene units indicated: 5'UTR (gray), ORF (blue), and 3'UTR (black).

Characterization of COI1b cDNA sequences

The full-length cDNA of the CnCOI1b-1 was 7,919 bp, containing a 1,176 bp ORF, with a 5' UTR of 6,477 bp upstream the start codon and a 3'UTR of 266 bp downstream the stop codon (**Table 1**). On the other hand, CnCOI1b-2 has 2,360 bp full-length cDNA, containing a 1,743 bp ORF, with a 5' UTR of 351 bp upstream the start codon and a 3'UTR of 266 bp downstream the stop codon. Both *COI1b* transcript variants in coconut have more than 8000 bp gene length. However, CnCOI1b-1 has significantly longer full-length cDNA, but shorter ORF compared to CnCOI1b-2. This observation is likely caused by the non-splicing of the intron 2 in the transcript variant CnCOI1b-1 relative to the sequence of the CnCOI1b-2. Upon further investigation within the sequence of the intron 2, several stop codons were found in the nucleotide stretch of the expressed sequence. Thus, it is classified as the 5'UTR of the

longest ORF predicted by ORF Finder. Moreover, CnCOIb-2 and EgCOIb have a very long 5561 bp intron which is absent in CnCOIb-1, AtCOI, AsCOI, and HbCOI. Regardless, the cDNA of coconut COIb sequences share 80.63% identical nucleotides (**Fig 2**) and more than 95% identical to the COIb of cDNA sequence of *E. guineensis*.

Table 1 Comparison of cDNA and deduced amino acid sequences among different plant species.

Fig 2 Pairwise nucleotide sequence alignment of CnCOIb-1 and CnCOIb-2. The 5'UTR and 3'UTR regions are labeled. Start site (green), stop site (red), conserved regions (blue asterisk).

In terms of cDNA sequence characteristic, *E. guineensis* COI (XM_010909322.2) showed same number of exons (4) and introns (3), and same length of ORF region (1,743 bp) with CnCOIb-2. Of all the coconut and representative dicot cDNA sequences of COI, CnCOIb-1 had the longest gene length while AtCOI had the shortest. In terms of full-length cDNA sequence, CnCOIb-1 was the longest and the shortest is HbCOI. Although CnCOIb-1 had the longest gene and full-length cDNA, it had the shortest ORF. COI from *Aquilaria sinensis* had the longest ORF compared to CnCOIb, *E. guineensis*, and other dicot species. Furthermore, known COI sequences of representative dicot species (**Table 1**) displayed more similarity to CnCOIb-2 compared to CnCOIb-1.

Structural and functional analysis of coconut COIb protein

The information provided by structural and functional analysis of coconut COI can provide the essential and novel details about its biological and cellular function. Though COI protein has been extensively characterized in other crops (Xie et al. 1998; Xu et al. 2002; Peng et al. 2009; Liao et al. 2015), its specific function in coconut is yet to be described. The translated amino acid sequences from the two (2) coconut isoforms, CnCOIb-1 and CnCOIb-2, span 391 and 580 amino acid residues, respectively (**Table 1**). The molecular weight was predicted to be higher in CnCOIb-2 (65.5 kDa) as compared to CnCOIb-1 (44.0 kDa). Both coconut COIb isoforms are rich in leucine residues (>14%). CnCOIb-1 protein contains 31.7% helix, 11.8% strand, and 56.5% coiled. On the other

hand, CnCOI1b-2 protein contains 35.5% helix, 11.6% strand, and 52.9% coiled. The two isoforms have AMN1 domain and leucine-rich repeat domain (LRR) – 8 LRRs for CnCOI1b-1 and 13 LRRs for CnCOI1b-2 (**Fig 3**).

Fig 3 The deduced amino acid sequence of CnCOI1b-1 and CnCOI1b-2. The characteristic motifs are shown as follows: 8 LRRs for CnCOI1b-1 and 13 LRRs for CnCOI1b-2 (highlighted), and AMN1 conserved domain (black box).

Compared to known COI1 protein sequences from other species, CnCOI1b-1 has the lowest molecular weight (44.0 kDa) and the least number of LRR domains (8) (**Table 1**). In terms of secondary protein structure, no significant difference can be observed across COI1 proteins. Only CnCOI1b-1 showed different tertiary structure, having a horse-shoe like shape. All the compared protein sequences showed AMN1 as the top domain hit.

Threading templates used by I-TASSER are based on Z-score or the difference between the raw and average scores in the unit of standard deviation. Based on the result, the highest scoring threading template for both isoforms was the structure of COI1-ASK1 in complex with coronatine and incomplete JAZ1 degran (3ogkB) detected 5X out of 10, from *Arabidopsis thaliana*, classified as protein binding, and expressed in *Spodoptera frugiperda*. The final model for CnCOI1b-1 and CnCOI1b-2 (**Fig 4**) were chosen based on the C-score among the model predictions. The highest structural analog for both isoforms was the Transport Inhibitor Response 1 or TIR1 (2p1oB and 2p1nB), a signaling protein from *A. thaliana* and expressed in *S. frugiperda*.

Fig 4 Homology model of CnCOI1b-1 (left) and CnCOI1b-2 (right).

All the predicted protein functions for CnCOI1b-1 and CnCOI1b-2 were also summarized in **Table 2**. Same ligand, OGK (Coronatine), was detected for both CnCOI1b-1 and CnCOI1b-2. The top enzyme for CnCOI1b-1 was found to be Beta-N-acetylhexosaminidase (EC no. 3.2.1.52), which catalyze the hydrolysis of terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides, and with active site residue of E123 (Glutamic acid). For CnCOI1b-2, the enzyme membrane alanyl aminopeptidase, which catalyze the removal of N-terminal amino acid from a peptide, amide or arylamide, was found to be present but with no active site residue. The consensus

prediction of Gene ontology (GO) terms for CnCOI1b-1 and CnCOI1b-2 were as follows: protein binding, auxin binding, ubiquitin-protein ligase activity, inositol hexakisphosphate binding, ribonuclease inhibitor activity, stomatal movement, jasmonic acid mediated signaling pathway, response to far red light, SCF-dependent proteasomal ubiquitin-dependent protein catabolic process, SCF ubiquitin ligase complex, and Angiogenin-PRI complex. Unique GO terms for each isoform were presented in **Table 2**.

Table 2. Predicted protein function of CnCOI1 isoforms based using COFACTOR and COACH.

Homologous alignment and phylogenetic analysis of coconut COI1b

Evolutionary analysis of the coconut COI1b isoforms and their orthologs revealed that these proteins form two (2) subfamilies, one composed of monocots and the other from dicots. The two (2) coconut COI1b isoforms are most related to *E. guineensis* COI1b, which indicate that both have almost similar structure and likely share some same gene function. Two (2) subclades can be observed in the monocot group, one group composed of Arecaceae and the other composed of Musaceae, Bromeliaceae, and Poaceae. COI1 protein in Arecaceae family is more related to Poaceae compared to Musaceae and Bromeliaceae based on the phylogenetic tree.

Fig 3 Phylogenetic relationships of the 2 isoforms of COI1b in coconut against known COI1 protein sequences of selected monocot and dicot species. The protein sequences used and their GenBank Accession numbers were: *M. acuminata* (XP_009416210); *A. comosus* (XP_020101802); *O. sativa* (XP_015639870); *Z. mays* (NP_001150429); *P. dactylifera* (XP_008775055); *E. guineensis* (XP_010907624); *A. thaliana* (ABR45948); *N. tabacum* (XP_016479272); *S. lycopersicum* (NP_001234464); *G. max* (NP_001238590); and *H. brasiliensis* (XP_021664949).

Discussion

Previously, *COI1* genes have been cloned and characterized from *Arabidopsis*, *Aquilaria*, rubber, and a few other plant species (Xie et al. 1998; Xu et al. 2002; Peng et al. 2009; Liao et al. 2015), but not from *Cocos nucifera*.

The present study is the first report on the cloning and characterization of the gene encoding COI1b protein in coconut. COI1, the first identified F-box protein, is one of the components of the SCF complex that mediates the activation of JA response genes (Xie et al. 1998). Mutations in *COI1* in other plant species resulted to JA insensitivity to defense response and physiological processes (Xie et al. 1998; Peng et al. 2009; Liao et al. 2015) which suggest that COI1 is an essential conserved component of JA signaling pathway in plant secondary metabolism.

The multiple sequence alignment of the assembled *COI1b* full-length assembled transcripts against the coconut CATD gene scaffold showed a different orientation for WCT. Possible explanation to this observations are RNA-seq library preparation method (stranded and non-stranded), the treatment (infested or diseased) of the raw read sequence source, and varietal differences. Furthermore, the coconut *COI1b* was found to have two (2) splicing variant – low MW (CnCOI1b-1) and high MW (CnCOI1b-2). Both splicing variants share high similarity with the coronatine-insensitive protein homolog 1b in *E. guineensis*. Long full-length cDNA sequence in *CnCOI1b-1* and short number of amino acid residue suggest that the two mRNAs might be generated by an alternative splicing from a single gene. The occurrence of this kind of post-transcriptional modification has also been reported in human proteasomal modulator subunit, p27 (*PSMD9*) wherein the longer cDNA sequence encodes a shorter polypeptide sequence (Watanabe et al. 1998). Based on transcript data, both splicing variants can be expressed in coconut. Moreover, the differential gene expression response of these coconut COI1b isoforms against insect damage is still subject for further research as previously demonstrated in *Arabidopsis* (Rehrig et al. 2014) and rice (Ye et al. 2012).

Based on striking similarity between COI1b and TIR1, a good structural model (C-score = 0.05, 1.08) for CnCOI1b-1 and CnCOI1b-2 were presented. The observed resemblance between the structures of COI1b and TIR1 provided an indirect evidence to explain the similarity of the signaling pathways for JA and auxin as previously described (Katsir et al. 2008). The conserved LRR domains in CnCOI1b-1 assemble into a solenoid fold which resembles horseshoe shape while CnCOI1b-2 has a spiral structure. The characteristics difference of the LRR domains among the two (2) isoforms might imply that the integrity of the LRR domains is necessary to the structural framework of COI1. Ja-Ile, ligand for the COI-JAZ protein interaction, was reported to bind at the LRR domain of COI1 (Xie et al. 1998). The LRR domain of TIR1, highest structural analog for both isoforms, was known to contain inositol hexakisphosphate co-factor and recognizes auxin (Tan et al. 20017).

The identified ligand OGK (Coroatine) of the CnCOI1 was known to have functions associated to plant fertility and defense response, regulation of genes induced by wounding, regulation of jasmonate, among many others (Xie et al 1998, Reymond et al 2000; Feng et al 2003). Some GO terms are unique for each isoform while some are in consensus with one another. Both isoforms almost have the same molecular function except that CnCOI1b-1 has GTPase activator activity which binds to and increases the activity of a GTPase, an enzyme that catalyzes the hydrolysis of GTP. CnCOI1b-1 have biological process associated to auxin mediated signaling pathway while response to insect was only detected in CnCOI1b-2. Both have SCF ubiquitin ligase complex and Angiogenin-PRI complex in their cellular component.

Multiple alignment analysis showed that CnCOI1b had more than 80% sequence identity with COI1 protein of monocots and more than 65% in dicots, which suggest that COI1 is highly conserved specially in monocots. Phylogenetic analysis confirms the high degree of COI1 conservation during the evolution, which reflects the selective pressure imposed by the vital functions of COI1 in plants.

Conclusion

The *coronatine-insensitive 1* gene (*COI1*) was cloned from coconut using 'Catigan Green Dwarf' (CATD) genome sequence assembly as reference in this paper. Two (2) splicing variants were identified - CnCOI1b-1 (low MW) and CnCOI1b-2 (high MW). Although both isoforms revealed different 3D structural models, various physiological and developmental plant processes including defense response to insect and pathogens were conserved. However, differential gene expression response of these coconut COI1b isoforms against insect damage is still a subject of further research. This work may provide the foundation of future research on understanding the role of COI1 in coconut. Results of this study are also expected to assist in the development of new resistant coconut varieties as one of the strategies to address threats in coconut production.

References

304 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-
305 BLAST: a new generation of protein database search programs. *Nucleic acids research* 25(17):3389-402.
306 <https://doi.org/10.1093/nar/25.17.3389>

307 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*
308 30(15):2114-20. <https://doi.org/10.1093/bioinformatics/btu170>

309 Devoto A, Ellis C, Magusin A, Chang HS, Chilcott C, Zhu T, Turner JG (2005) Expression profiling reveals COI1 to
310 be a key regulator of genes involved in wound-and methyl jasmonate-induced secondary metabolism,
311 defence, and hormone interactions. *Plant Molecular Biology* 58(4):497-513. [https://doi.org/10.1007/s11103-](https://doi.org/10.1007/s11103-005-7306-5)
312 005-7306-5

313 [FAOSTAT] Food and Agriculture Organization of the United Nations: Statistical Division (2014) Coconuts,
314 Production/Crops. Retrieved from <http://www.fao.org/> on 31 August 2017.

315 Feng S, Ma L, Wang X, Xie D, Dinesh-Kumar SP, Wei N, Deng XW (2003) The COP9 signalosome interacts
316 physically with SCFCOI1 and modulates jasmonate responses. *The Plant Cell* 15(5):1083-94.
317 <https://doi.org/10.1105/tpc.010207>

318 Feys BJ, Benedetti CE, Penfold CN, Turner JG (1994) Arabidopsis mutants selected for resistance to the phytotoxin
319 coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. *The Plant*
320 *Cell* 6 (5):751-759. <https://doi.org/10.1105/tpc.6.5.751>

321 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q,
322 Chen Z (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data.
323 *Nature biotechnology* 29(7):644. 10.1038/nbt.1883

324 Howe GA (2008) New weapons and a rapid response against insect attack. *Plant Physiology* 146 (3):832-
325 838.<https://doi.org/10.1104/pp.107.115683>

326 Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome research* 9 (9):868-877.
327 10.1101/gr.9.9.868

328 Katsir L, Schilmiller AL, Staswick PE, He SY, Howe GA (2008) COI1 is a critical component of a receptor for
329 jasmonate and the bacterial virulence factor coronatine. *Proceedings of the National Academy of Sciences*
330 105(19):7100-7105. <https://doi.org/10.1073/pnas.0802332105>

331 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of
332 transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* 14(4):R36.
333 <https://doi.org/10.1186/gb-2013-14-4-r36>

334 Lantican DV, Strickler SR, Canama AO, Gardoce RR, Mueller LA, Galvez HF (2019) De Novo Genome Sequence
335 Assembly of Dwarf Coconut (*Cocos nucifera* L.'Catigan Green Dwarf') Provides Insights into Genomic
336 Variation Between Coconut Types and Related Palm Species. *G3: Genes, Genomes, Genetics* 9 (8):2377-
337 2393. <https://doi.org/10.1534/g3.119.400215>

338 Lee HY, Seo JS, Cho JH, Jung H, Kim JK, Lee JS, Rhee S, Do Choi Y (2013) *Oryza sativa* COI Homologues Restore
339 Jasmonate Signal Transduction in *Arabidopsis* *coil-1* Mutants. *PLoS ONE* 8(1): e52802.
340 <https://doi.org/10.1371/journal.pone.0052802>

341 Liao Y, Wei J, Xu Y, Zhang Z (2015) Cloning, expression and characterization of COI1 gene (AsCOI1) from *Aquilaria*
342 *sinensis* (Lour.) Gilg. *Acta Pharmaceutica Sinica B* 5(5):473-81. <https://doi.org/10.1016/j.apsb.2015.05.009>

343 Mewis I, Appel HM, Hom A, Raina R, Schultz JC (2005) Major signaling pathways modulate *Arabidopsis*
344 glucosinolate accumulation and response to both phloem-feeding and chewing insects. *Plant Physiology* 138
345 (2):1149-1162. <https://doi.org/10.1104/pp.104.053389>

346 [PCA] Philippine Coconut Authority (2017) No more'cocolisap' outbreak. Retrieved from [http://pca.gov.ph/10-](http://pca.gov.ph/10-news/49-no-more-cocolisap-outbreak)
347 [news/49-no-more-cocolisap-outbreak](http://pca.gov.ph/10-news/49-no-more-cocolisap-outbreak) on 30 August 2019.

348 Peng SQ, Xu J, Li HL, Tian WM (2009) Cloning and molecular characterization of HbCOI1 from *Hevea brasiliensis*.
349 *Bioscience, biotechnology, and biochemistry* 73(3):665-70. <https://doi.org/10.1271/bbb.80721>

350 Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments.
351 *PloS one* 5(3):e9490. <https://doi.org/10.1371/journal.pone.0009490>

352 [PSA] PHILIPPINE STATISTICS AUTHORITY (2017) Major Non - Food and Industrial Crops Quarterly Bulletin,
353 October-December 2017. Retrieved from <https://psa.gov.ph/non-food/coconut> 10 May 2018.

354 Rehrig, E. M., Appel, H. M., Jones, A. D., & Schultz, J. C. (2014). Roles for jasmonate-and ethylene-induced
355 transcription factors in the ability of Arabidopsis to respond differentially to damage caused by two insect
356 herbivores. *Frontiers in plant science*, 5, 407.

357 Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical
358 wounding and insect feeding in Arabidopsis. *The Plant Cell* 12(5):707-19.
359 <https://doi.org/10.1105/tpc.12.5.707>

360 Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function
361 prediction. *Nature protocols* 5(4):725. <https://doi.org/10.1038/nprot.2010.5>

362 Sievers F, Wilm A., Dineen D, Gibson TJ, Karplus K, Li W, Thompson JD (2011). Fast, scalable generation of high-
363 quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1).

364 Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N (2007) Mechanism of auxin
365 perception by the TIR1 ubiquitin ligase. *Nature* 446(7136):640. <https://doi.org/10.1038/nature05731>

366 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities
367 and interfaces. *Nucleic acids research* 40 (15):e115-e115. <https://doi.org/10.1093/nar/gks596>

368 Watanabe TK, Saito A, Suzuki M, Fujiwara T, Takahashi EI, Slaughte, CA, DeMartino GN, Hendil KB, Chung CH,
369 Tanahashi N, Tanaka, K (1998) cDNA cloning and characterization of a human proteasomal modulator
370 subunit, p27 (PSMD9). *Genomics*, 50(2) 241-250. <https://doi.org/10.1006/geno.1998.5301>

371 Xie DX, Feys BF, James S, Nieto-Rostro M, Turner JG (1998) COI1: an Arabidopsis gene required for jasmonate-
372 regulated defense and fertility. *Science* 28(5366):1091-4. [10.1126/science.280.5366.1091](https://doi.org/10.1126/science.280.5366.1091)

373 Xu L, Liu F, Lechner E, Genschik P, Crosby WL, Ma H, Peng W, Huang D, Xie D (2002) The SCFCOI1 ubiquitin-
374 ligase complexes are required for jasmonate response in Arabidopsis. *The Plant Cell* 14 (8):1919-1935.
375 <https://doi.org/10.1105/tpc.003368>

376 Yan Y, Borrego E, Kolomiets MV (2013) Jasmonate biosynthesis, perception and function in plant development and
377 stress responses. In: Baez RV (ed) *Lipid metabolism*, pp. 393-442. InTech, Croatia. 10.5772/52675

378 Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function
379 prediction. *Nature methods* 12(1):7. <https://doi.org/10.1038/nmeth.3213>

380 Ye M, Luo SM, Xie JF, Li YF, Xu T, Liu Y, Song YY, Zhu-Salzman K, Zeng RS (2012) Silencing COI1 in rice
381 increases susceptibility to chewing insects and impairs inducible defense. *PLoS one* 7(4):e36214.
382 <https://doi.org/10.1371/journal.pone.0036214>

383 Zhai Q, Yan C, LI L, Xie D, Li C (2017) Jasmonates. In: Li J, Li C, AND Smith SM (eds) *Hormone Metabolism and*
384 *Signaling in Plants*, pp. 243-272. Elsevier, United Kingdom.

385 Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* 9(1):40.
386 <https://doi.org/10.1186/1471-2105-9-40>

Table 1. Comparison of cDNA and deduced amino acid sequences among different plant species.

	CnCOI1b-1	CnCOI1b-2	¹ EgCOI1b	² AtCOI	³ AsCOI1	⁴ HbCOI1
Gene length (bp)	8204	8204	7081	3151	4566	4003
Full-length cDNA (bp)	7919	2360	2408	2291	2330	2181
No. of exons	3	4	4	11	3	3
No. of introns	2	3	3	10	2	2
ORF (bp)	1176	1743	1743	1779	1839	1794
5'UTR (bp)	6477	351	384	262	191	50
3'UTR (bp)	266	266	281	250	300	337
No. of amino acid residue	391	580	580	592	612	597
Molecular weight (kDa)	44.0	65.5	65.5	67.7	68.9	68.0
Isoelectric point (pI)	5.24	6.04	5.79	6.06	6.56	6.66
% Leucine	14.8	14.5	14.5	13.3	14.7	14.9
Secondary structure of protein	31.7% helix	35.5% helix	35.9% helix	36.5% helix	35.6% helix	36.5% helix
	11.8% strand	11.6% strand	11.4% strand	10.6% strand	9.8% strand	9.2% strand
	56.5% coiled	52.9% coiled	52.7% coil	52.9% coil	54.6% coil	54.3% coil
Tertiary structure of protein	Horseshoe-like	spiral	spiral	spiral	spiral	spiral
No. of LRRs	8	13	13	14	15	13

Conserved domain	AMN1 super family	AMN1 super family	AMN1 super family	AMN1 super family FBOX	AMN1 superfamily	AMN1 super family
------------------	----------------------	----------------------	----------------------	------------------------------	------------------	----------------------

| ¹ *Elaeis guineensis* GenBank Accession No: XM_010909322.2 | | | ² *Arabidopsis thaliana* GenBank Accession No: AF036340.1 | | |
| ³ *Aquilaria sinensis* GenBank Accession No: KM189194.1 | | | ⁴ *Hevea brasiliensis* GenBank Accession No: EU136026.1 | | |

Table 2. Predicted protein function of CnCOI1b isoforms based using COFACTOR and COACH.

	CnCOI1b-1	CnCOI1b-2
Ligand Binding Site	OGK (Coronatine)	OGK (Coronatine)
Active Site	Beta-N-acetylhexosaminidase mutant D539A complexed with Di-N-acetyl-Beta-D-glucosamine (chitobiase)	Crystal Structure of Aminopeptidase N complexed with bestatin
EC Number	Beta-N-acetylhexosaminidase	Membrane alanyl aminopeptidase
Molecular Function	Protein binding Auxin binding Ubiquitin-protein ligase activity Inositol hexakisphosphate binding Ribonuclease inhibitor activity ¹ GTPase activator activity	Protein binding Auxin binding Ubiquitin-protein ligase activity Inositol hexakisphosphate binding Ribonuclease inhibitor activity
Biological process	¹ Auxin mediated signaling pathway ¹ Protein ubiquitination ¹ Cellular response to phosphate starvation ¹ Ethylene mediated signaling pathway ¹ Response to molecule of bacterial origin ¹ Lateral root formation Stomatal movement Jasmonic acid mediated signaling pathway Response to far red light SCF-dependent proteasomal ubiquitin-dependent protein catabolic process	² Response to insect ² Shade avoidance ² Regulation of flower development ² Jasmonic acid and ethylene-dependent systemic resistance ² Negative regulation of defense response ² Defense response to bacterium Stomatal movement Jasmonic acid mediated signaling pathway Response to far red light SCF-dependent proteasomal ubiquitin-dependent protein catabolic process
Cellular component	SCF ubiquitin ligase complex Angiogenin-PRI complex ¹ Cytosol ¹ Nuclear membrane ¹ Perinuclear region of cytoplasm	SCF ubiquitin ligase complex Angiogenin-PRI complex ² Cytoplasm ² Nucleus

Legend: 1= unique for CnCOI1b-1 2=unique for CnCOI1b-2

Fig 1 Visualization of the consensus cDNA sequences of CnCOI1b-1 and CnCOI1b-2 in reference transcript assembly. The arrow indicates the direction of transcription with the gene units indicated: 5'UTR (gray), ORF (blue), and 3'UTR (black).

		5'UTR CnCOI1b-2	
CnCOI1b-1	TTCCATCAAATTTCTTAACCGTGCAAATATCTTATCGGACCAATGTCTCTGTCACTGG	5925	
CnCOI1b-2	TCGGATCGAAGCT---GGAGCGGCGGAGCTTGGCTCGGGCGATGAGCTGCGGGATATCG	369	
	* * * * *		
CnCOI1b-1	AATTCATGCTTGCTTTGTTGATGGCCACTTGGTGCTTCATGTAGAATTCCTTAAAAA	5985	
CnCOI1b-2	GATGTGGCGCTCGATCTGGTCATGGG-----GTACATC---GACGACCCCGCGACCGG	420	
	* * * * *		
CnCOI1b-1	AATGTTATGTCTGTTTAAATTGGTGTTCGGTGCTTCATTTGTTTATGCTTCTGGATG	6045	
CnCOI1b-2	GA-----GGCGTCTCGCTGGTGTGCAAGAAGTGGTACC-ACATCGACGCCATCACCCG	473	
	* * * * *		
CnCOI1b-1	CAATGAATAGAATATCTGTATAAGC-----TAGAATCAAGATGCAATT	6088	
CnCOI1b-2	CAAGCACATCACCATCGCCATCTGCTACTCCACCAGCCCCGCGCGCTCCGGCGGCGCTT	533	
	* * * * *		
CnCOI1b-1	GTGACAAACCAATGACCATCTCTGATTATGATGTTTGTAGTTATCTCTTTGATGTTTAC	6148	
CnCOI1b-2	CCCCCGCTCGAGTCCCTCAAGCTCAAGGGCAAGCCCCGGCGGCCATGTTTAACTCAT	593	
	* * * * *		
CnCOI1b-1	AGTCGTAAACTATGGCTTTCCCAAAGAACACCTTTGGGCATGTTGCATTGGTCGATATG	6208	
CnCOI1b-2	ACCCGAGGACTGGGGCGGGT---ACGCCGCCCTTGGGTCC---GCGAGATCG---C	641	
	* * * * *		
CnCOI1b-1	GTGGTGATTTAATCACTGCATGATTTTTTTAGTTTTTATGCTTATTTGTATA---GACT	6265	
CnCOI1b-2	CGAGGCCCTTCAACTGCCCAAGG-TCCCTCACTTCCGCCGGATGATCGTGACAGACGCCG	700	
	* * * * *		
CnCOI1b-1	ACTGGTGCTTTTGTGTCCCT-----AGAAGATGGTTTTGTGTAACAATGATAGTTTCA	6319	
CnCOI1b-2	ACATCGCCCTCCTGGTCAAGGAGAGGGGCCACATGCTTCAGACGC-TCAAGCTGGATAAG	759	
	* * * * *		
CnCOI1b-1	TGCTGGTTGCTCTATGA-CTGGTTTCAGTTTCCAAATGTCTTATCTACAGATGCCTTAA	6378	
CnCOI1b-2	TGCTCGGGGTTTCCACCGATGGCCTTTTGCTGCTTGGCCGTTCTTGAGATGCCTTAA	819	
	* * * * *		
CnCOI1b-1	ACCTTATTTTGAAGAAAGCTCAATCATTGAAAATGATGGTGCATGGCTCCATGAACCT	6438	
CnCOI1b-2	ACCTTATTTTGAAGAAAGCTCAATCATTGAAAATGATGGTGCATGGCTCCATGAACCT	879	
	* * * * *		
		5'UTR CnCOI1b-1	
CnCOI1b-1	GCTGTAAACAATTCTGTGCTCGAGACTTTGAATTTCTATATGACAGAGCTCAGGATATCG	6498	

57	CnCOI1b-2	GCTGTTAACAATTCTGTGCTCGAGACTTTGAATTTCTATATGACAGAGCTCAGGATATCG	939
58		*****	
59			
60	CnCOI1b-1	TCACAAGACCTCGAGCTACTTGCTAGGAAGTGCAGGTCTTTGATTTCCTTGAAGATCAGT	6558
61	CnCOI1b-2	TCACAAGACCTCGAGCTACTTGCTAGGAAGTGCAGGTCTTTGATTTCCTTGAAGATCAGT	999
62		*****	
63			
64	CnCOI1b-1	GAGTGTGATGTCTCTGATCTAGTTGGCTTTTTTCGCCTAGCAACAGCATTGGAAGAGTTT	6618
65	CnCOI1b-2	GAGTGTGATGTCTCTGATCTAGTTGGCTTTTTTCGCCTAGCAACAGCATTGGAAGAGTTT	1059
66		*****	
67			
68	CnCOI1b-1	GGTGGGGGCTCATTTAATGACCAAGCAGGAGAAACCAAAATATGAGATGGTTCACTTT	6678
69	CnCOI1b-2	GGTGGGGGCTCATTTAATGACCAAGCAGGAGAAACCAAAATATGAGATGGTTCACTTT	1119
70		*****	
71			
72	CnCOI1b-1	CCTCCCAGGCTATGCTGTTTGGGTCTCATTTATATGGAGACAAATGAAATGCAAAATAATA	6738
73	CnCOI1b-2	CCTCCCAGGCTATGCTGTTTGGGTCTCATTTATATGGAGACAAATGAAATGCAAAATAATA	1179
74		*****	
75			
76	CnCOI1b-1	TTTCCATTGCTGCTTCACTCAAGAAGCTAGACTTGCAATATACTTTTCTCAGCACTGAA	6798
77	CnCOI1b-2	TTTCCATTGCTGCTTCACTCAAGAAGCTAGACTTGCAATATACTTTTCTCAGCACTGAA	1239
78		*****	
79			
80	CnCOI1b-1	GATCATTGTCAGCTCATTCACGATGTCCCAATCTAGAAATCTTGAGGTGAGGAATGTC	6858
81	CnCOI1b-2	GATCATTGTCAGCTCATTCACGATGTCCCAATCTAGAAATCTTGAGGTGAGGAATGTC	1299
82		*****	
83			
84	CnCOI1b-1	ATAGGAGACAGAGGTTTAGAAGTTGTTGCCAGACATGTAAGAACTACGAAGGCTTAGA	6918
85	CnCOI1b-2	ATAGGAGACAGAGGTTTAGAAGTTGTTGCCAGACATGTAAGAACTACGAAGGCTTAGA	1359
86		*****	
87			
88	CnCOI1b-1	ATAGAGCGAGGGGATGATGAACAAGGCCTGGAGGATGAACAGGGTAGGGTTTCCCAGGTG	6978
89	CnCOI1b-2	ATAGAGCGAGGGGATGATGAACAAGGCCTGGAGGATGAACAGGGTAGGGTTTCCCAGGTG	1419
90		*****	
91			
92	CnCOI1b-1	GGGTTGTCAACTTTGGCCAGGGTTGCCCTGAACTGGAATACCTGGCAGTATATGTCTCT	7038
93	CnCOI1b-2	GGGTTGTCAACTTTGGCCAGGGTTGCCCTGAACTGGAATACCTGGCAGTATATGTCTCT	1479
94		*****	
95			
96	CnCOI1b-1	GATATCAGAAACGCAGCCCTGGAATCTGTTGGCACTTATAGCAAAAACCTCTGTGATTTT	7098
97	CnCOI1b-2	GATATCAGAAACGCAGCCCTGGAATCTGTTGGCACTTATAGCAAAAACCTCTGTGATTTT	1539
98		*****	
99			
100	CnCOI1b-1	CGGCTTGTTACTTGTACAGAGAAGAAAGGATAGCAGATTTACCCCTTGATAATGGTGTT	7158
101	CnCOI1b-2	CGGCTTGTTACTTGTACAGAGAAGAAAGGATAGCAGATTTACCCCTTGATAATGGTGTT	1599
102		*****	
103			
104	CnCOI1b-1	CGTGCTCTGTTGAGAGGATGCACCAAACCTAGAAGGTTGCTCTTTATTTGAGACCTGGA	7218
105	CnCOI1b-2	CGTGCTCTGTTGAGAGGATGCACCAAACCTAGAAGGTTGCTCTTTATTTGAGACCTGGA	1659
106		*****	
107			
108	CnCOI1b-1	GGTCTTTTCTAGATGTGGGTCTTGGTTTTCATTGGTGAGTGCAGTAGGAATATCCGGTGGATG	7278
109	CnCOI1b-2	GGTCTTTTCTAGATGTGGGTCTTGGTTTTCATTGGTGAGTGCAGTAGGAATATCCGGTGGATG	1719
110		*****	
111			
112	CnCOI1b-1	CTATTGGGTAATGTAGGGGAATCTGATGTTGGACTTCTGCGATTTGCCAGAGGTTGCCCA	7338
113	CnCOI1b-2	CTATTGGGTAATGTAGGGGAATCTGATGTTGGACTTCTGCGATTTGCCAGAGGTTGCCCA	1779
114		*****	
115			
116	CnCOI1b-1	AGCCTGCAGAAGCTGGAGCTGAGAAGCTGTTGCTTTACTGAGCGTGCTTTGGCCCTTGCA	7398
117	CnCOI1b-2	AGCCTGCAGAAGCTGGAGCTGAGAAGCTGTTGCTTTACTGAGCGTGCTTTGGCCCTTGCA	1839
118		*****	
119			
120	CnCOI1b-1	GCAACACAGCTGCCTTCATTGAGATACTTATGGGTGCAGGGTTACAAGGCATCACCAACT	7458
121	CnCOI1b-2	GCAACACAGCTGCCTTCATTGAGATACTTATGGGTGCAGGGTTACAAGGCATCACCAACT	1899
122		*****	
123			

124	CnCOI1b-1	GGCAGGGACCTTATGGCCATGGCCCGCCATTCTGGAACATCGAATTCATCCCTCCTAGA	7518
125	CnCOI1b-2	GGCAGGGACCTTATGGCCATGGCCCGCCATTCTGGAACATCGAATTCATCCCTCCTAGA	1959
126		*****	
127			
128	CnCOI1b-1	CAACAGATTGTTTATCATGATCTTGATGATAGGCCAGGTGCAATAGAGAGTCAGGCTCAA	7578
129	CnCOI1b-2	CAACAGATTGTTTATCATGATCTTGATGATAGGCCAGGTGCAATAGAGAGTCAGGCTCAA	2019
130		*****	
131			
132	CnCOI1b-1	ATACTTGCACTACTCCCTTGCTGGAAAGAGGACCGATTGTCCAGAATCTGTGGTTCCT	7638
133	CnCOI1b-2	ATACTTGCACTACTCCCTTGCTGGAAAGAGGACCGATTGTCCAGAATCTGTGGTTCCT	2079
134		*****	
135			
136	CnCOI1b-1	TTGTATCCAGAA TGA GTGTGTACATATGTAGGCCTCTTGTATTTGGGACCTTTTCCCT	7698
137	CnCOI1b-2	TTGTATCCAGAA TGA GTGTGTACATATGTAGGCCTCTTGTATTTGGGACCTTTTCCCT	2139
138		*****	
139			
140	CnCOI1b-1	TTTCTTTTGGTTGCAGTCCCTTTTCTACTCTTAACTGCCAAAAATATCTGGGCGGTGAACG	7758
141	CnCOI1b-2	TTTCTTTTGGTTGCAGTCCCTTTTCTACTCTTAACTGCCAAAAATATCTGGGCGGTGAACG	2199
142		*****	
143			
144	CnCOI1b-1	TCAATATATCTCGTTTGTAAACATCTTAAAGATGCGAGCCGTGAGCTTTGTAATAAGCTGT	7818
145	CnCOI1b-2	TCAATATATCTCGTTTGTAAACATCTTAAAGATGCGAGCCGTGAGCTTTGTAATAAGCTGT	2259
146		*****	
147			
148	CnCOI1b-1	GATATGCTTTCCGTAATTCAAAAGAACTTTAGCGGCAGCCATGTGGTCTTCTTAGAGCTC	7878
149	CnCOI1b-2	GATATGCTTTCCGTAATTCAAAAGAACTTTAGCGGCAGCCATGTGGTCTTCTTAGAGCTC	2319
150		*****	
151			
152	CnCOI1b-1	CAAGGTATGGGCTCTGAACCTTTTAGATTTAAATTAGTTCT	7919
153	CnCOI1b-2	CAAGGTATGGGCTCTGAACCTTTTAGATTTAAATTAGTTCT	2360
154		*****	

Fig 2 Pairwise nucleotide sequence alignment of CnCOI1b-1 and CnCOI1b-2. The 5'UTR and 3'UTR regions are labeled. Start site (green), stop site (red), conserved regions (blue asterisk). The pairwise alignment of 1-5865 bp of CnCOI1b-1 and 1-313 bp of the CnCOI1b-2 were not shown in the figure.

Fig 3 The deduced amino acid sequence of CnCOI1b-1 and CnCOI1b-2. The characteristic motifs are shown as follows: 8 LRRs for CnCOI1b-1 and 13 LRRs for CnCOI1b-2 (highlighted), and AMN1 conserved domain (black box).

Fig 4 Homology model of CnCOI1b-1 (left) and CnCOI1b-2 (right).

Fig 5 Phylogenetic relationships of the 2 isoforms of COI1b in coconut against known COI1 protein sequences of selected monocot and dicot species. The protein sequences used and their GenBank Accession numbers were: *M. acuminata* (XP_009416210); *A. comosus* (XP_020101802); *O. sativa* (XP_015639870); *Z. mays* (NP_001150429); *P. dactylifera* (XP_008775055); *E. guineensis* (XP_010907624); *A. thaliana* (ABR45948); *N. tabacum* (XP_016479272); *S. lycopersicum* (NP_001234464); *G. max* (NP_001238590); and *H. brasiliensis* (XP_021664949).

APPENDIX 1. Gel electrophoretogram of insert PCR product: cDNA from healthy CATD leaf samples (Lanes 2-3), CSI-infested CATD leaf samples (lanes 4-7), and genomic DNA of CATD coconut variety (lanes 8-9). A 1kb+ ladder was used.

APPENDIX 2. Pairwise alignment of cloned mid-transcript sequences from healthy and CSI-infested CATD leaf samples showed 100% alignment or no sequence variation.

Healthy	GCATTGGAAGAGTTTGGTGGGGGCTCATTTAATGACCAAGCAGGAGAAACCAACAAATAT	60
Infested	GCATTGGAAGAGTTTGGTGGGGGCTCATTTAATGACCAAGCAGGAGAAACCAACAAATAT	60

Healthy	GAGATGGTTCACTTTCCTCCCAGGCTATGCTGTTTGGGTCTCATTTATATGGAGACAAAT	120
Infested	GAGATGGTTCACTTTCCTCCCAGGCTATGCTGTTTGGGTCTCATTTATATGGAGACAAAT	120

Healthy	GAAATGCAAATAATATTTCCATTTGCTGCTTCACTCAAGAAGCTAGACTTGCAATATACT	180
Infested	GAAATGCAAATAATATTTCCATTTGCTGCTTCACTCAAGAAGCTAGACTTGCAATATACT	180

Healthy	TTTCTCAGCACTGAAGATCATTGTGCTAGCTCATTCAACGATGTCCCAATCTAGAAATTCTT	240
Infested	TTTCTCAGCACTGAAGATCATTGTGCTAGCTCATTCAACGATGTCCCAATCTAGAAATTCTT	240

Healthy	GAGGTGAGGAATGTCATAGGAGACAGAGGTTTAGAAGTTGTTGCCAGACATGTAAGAAA	300
Infested	GAGGTGAGGAATGTCATAGGAGACAGAGGTTTAGAAGTTGTTGCCAGACATGTAAGAAA	300

Healthy	CTACGAAGGCTTAGAATAGAGCGAGGGGATGATGAACAAGGCCTGGAGGATGAACAGGGT	360
Infested	CTACGAAGGCTTAGAATAGAGCGAGGGGATGATGAACAAGGCCTGGAGGATGAACAGGGT	360

276				
277	Healthy	AGGGTTTCCCAGGTGGGGTTGTCAACTTTGGCCCAGGGTTGCCCTGAACTGGAATACCTG	420	
278	Infested	AGGGTTTCCCAGGTGGGGTTGTCAACTTTGGCCCAGGGTTGCCCTGAACTGGAATACCTG	420	
279		*****		
280				
281	Healthy	GCAGTATATGTCTCTGATATCACAAACGCAGCCCTGGAATCTGTTGGCACTTATAGCAAA	480	
282	Infested	GCAGTATATGTCTCTGATATCACAAACGCAGCCCTGGAATCTGTTGGCACTTATAGCAAA	480	
283		*****		
284				
285	Healthy	AACCTCTGTGATTTTCGGCTTGTATTACTTGACAGAGAAGAAAGGATAGCAGATTTACCC	540	
286	Infested	AACCTCTGTGATTTTCGGCTTGTATTACTTGACAGAGAAGAAAGGATAGCAGATTTACCC	540	
287		*****		
288				
289	Healthy	CTTGATAATGGTGTTTCGTGCTCTGTTGAGAGGATGCACCAAACCTAGAAGGTTTGCTCTT	600	
290	Infested	CTTGATAATGGTGTTTCGTGCTCTGTTGAGAGGATGCACCAAACCTAGAAGGTTTGCTCTT	600	
291		*****		
292				
293	Healthy	TATTTGAGACCTGGAGGTCCTTCAGATGTGGGTCTTGGTTTCATTGGTGAGTGCAGTAGG	660	
294	Infested	TATTTGAGACCTGGAGGTCCTTCAGATGTGGGTCTTGGTTTCATTGGTGAGTGCAGTAGG	660	
295		*****		
296				
297	Healthy	AATATCCGGTGGATGCTATTGGGTAATGTAGGGGAATCTGATGTTGGACTTCTGCGATTT	720	
298	Infested	AATATCCGGTGGATGCTATTGGGTAATGTAGGGGAATCTGATGTTGGACTTCTGCGATTT	720	
299		*****		
300				
301	Healthy	GCCAGAGGTTGCCCAAGCCTGCAGAAGCTGGA	752	
302	Infested	GCCAGAGGTTGCCCAAGCCTGCAGAAGCTGGA	752	
303		*****		
304				
305				
306				
307				
308				