Philippine Journal of Crop Science (PJCS) 2022, 47 December (3):69-77

RESEARCH NOTE

Isolation and Sequence Analysis of Putative *CnSHELL* Gene Region Provides Insights on Phylogeny and Origin of Coconut Tall Cultivars

Charles Anthon E. Cadorna[1], Ramon A. Rivera[2], and Jessica D. Rey[1*]

[1]Institute of Biology, College of Science, University of the Philippines, Diliman, Quezon City 1101, Philippines; [2]Philippine Coconut Authority Zamboanga Research Center, San Ramon, Zamboanga City, Zamboanga del Sur 7000, Philippines. [*]Corresponding author, jdrey@up.edu.ph

***Cocos nucifera* (coconut) is an economically important crop in the Philippines. Coconut is the only species in the genus *Cocos* of the family Arecaceae and is widely cultivated for its extensive use in agriculture and industry. The *SHELL* gene has already been characterized in oil palm which is responsible for its coconut-like shell and various fruit forms. However, this gene has not yet been isolated and characterized in coconut. In this study, the *CnSHELL* gene region was successfully amplified across 22 coconut cultivars originating from different regions and amplicon size ranges at about 450-550 bp. Based on BLAST analysis, this gene is homologous with *Elaeis guineesis* shell-specific gene. The isolated gene can also be mapped on the whole genome sequences of coconut deposited in NCBI. Sequence analysis of the *CnSHELL* gene revealed low level of genetic diversity (Hd = 0.039394, = 0.00131) which indicates that the amplified region is highly conserved. Nevertheless, phylogenetic analysis using the *CnSHELL* gene region showed two groups of coconut, which is consistent with previous reports on separate domestication origins of coconut. Indeed, this is the first report of *SHELL* gene region isolated from 22 *Cocos nucifera* tall cultivars.**

Keywords: *CnSHELL* gene, *Cocos nucifera*, origin, phylogeny

INTRODUCTION

Cocos nucifera (coconut) is a tropical crop and widely cultivated in more than 98 countries due to its extensive use in agroindustry (Arulandoo et al. 2017). Worldwide, the recorded annual production of coconut is approximately 61.5 million tonnes (FAOSTAT, 2020). The Philippines is a major exporter of coconut products such as coconut oil with an annual production of 14.4 million tonnes (FAOSTAT 2020). According to the Philippine Coconut Authority (PCA), the coconut industry provides livelihood to one-third of the country's population (PCA 2019). However, the production of coconut in the Philippines has started to decline in 2010 and in 2013 due to the infestation of coconut scale insect or "cocolisap" (*Aspidiotus rigidus*) and the occurrence of major typhoons which destroyed large number of trees (FFTC-AP, 2019). With the recent COVID-19 pandemic brought about by SARS-CoV-2, virgin coconut oil (VCO) received quite an attention because of its potential antiviral properties. Moreover, the Department of Science and Technology (DOST) even conducted a clinical trial for its possible use as an adjunctive therapy for hospitalized COVID-19 patients (Alejandria and Dalmacio 2021).

Coconut has a diploid genome with a chromosome number of 32 (2n=32) as revealed by karyotype analysis of various coconut varieties (Pereira et al. 2017; Sisunandar and Adkins, 2007; Perera et al. 2008). Based on flow cytometric analysis, coconut has a mean genome size of about 2.72-2.88 Gbp per haploid set (Gunn et al. 2015; Freitas Neto et al. 2016). The closest relatives of coconut among the economically important palms are *Elaeis* sp. (oil palm) and *Phoenix dactylifera* (date palm). According to Xiao et al. (2017), *Cocos nucifera* and *Elaeis guineensis* appear to diverge at about 46.0 Mya (25.4-83.3 Mya). However, Wang et al. (2021) reported that the divergence of *C. nucifera* and *E. guineensis* happened at about 17-19 Mya in which their shared whole-genome duplication (WGD) occurred at about 47-53 Mya. Phylogenetic analysis using single copy WRKY genes showed sister relationship between *Cocos* and *Attalea*, where they diverged from each other at about 23.9-44.4 Mya (Meerow et al. 2015). Furthermore, the cultivation of coconut is reported to have two independent origins based on analysis of microsatellite markers – one in the Indo-Atlantic basin and the other in the Pacific (Gunn et al. 2011).

Generally, there are two varieties of coconut based on morphological features and mode of pollination – the "Tall" and the "Dwarf" cultivars. The "Dwarf" coconut varieties are usually self-pollinated while the "Tall" varieties are cross-pollinated (Menon & Pandalai 1958). Microsatellite data revealed that "Dwarf" coconuts originated from a typical domestication process which appeared in Southeast Asia (Perera et al. 2016). It was also reported that the divergence time between that tall and dwarf coconut varieties was dated at 2-8 Mya (Wang et al. 2021).

Several genomic resources for coconut have already been reported. The genomes of several cultivars of coconut – Hainan Tall (HAT), Catigan Green Dwarf (CATD), and Chowghat Green Dwarf (CGD) have been published, which provides insights on salinity tolerance, genomic variation between coconut types and other palms, and disease resistance (Xiao et al. 2017; Lantican et al. 2019; Muliyar et al. 2020). A high-quality reference genome of a tall and a dwarf coconut cultivar was also produced and comparison of the two genomes revealed two large deletions in the dwarf coconut genome compared to the tall coconut genome, which might be associated with the divergence of germination time and rates as well as cellulose content (Wang et al. 2021).

The *SHELL* gene is already identified in oil palm, which is responsible for the coconut-like shell and the occurrence of the different fruit forms. This gene is a type II MADS-box transcription factor and homologous with *SEEDSTICK*, which is responsible for ovule and seed development in *Arabidopsis*. Likewise, this gene is an ortholog of *OsMADS13* in rice (Singh et al. 2013). In angiosperms, the function and expression of the *SHELL* is conserved (Favaro et al. 2003). However, this gene has not yet been isolated and characterized in *C. nucifera*.

In this study, the putative *CnSHELL* gene region has been amplified in 22 coconut tall cultivars originating from different regions. This study was conducted to analyze the putative *CnSHELL* gene region and investigate the phylogenetic relationships of different tall coconut cultivars. By analyzing the *CnSHELL* gene, the degree of relatedness was assessed and the genetic relationships of the different tall coconut cultivars across regions was elucidated. This is the first study to report the isolation and analysis of *CnSHELL* gene region in coconut.

MATERIALS AND METHODS

Plant material and collection of leaf samples
Eight (8) coconut tall varieties both from local and foreign sources were used in this study (Table 1). Disease-free, young leaf samples were obtained from the germplasm collection of the Philippine Coconut Authority – Zamboanga Research Center (PCA-ZRC). The leaf samples were then transported to the Plant Molecular Phylogenetics Laboratory (PMPL) at the Institute of Biology, University of the Philippines, Diliman, Quezon City, Philippines and stored at -80C until use.

Table 1. Coconut tall cultivars used in this study and collected from PCA-ZRC

Sample Code	Variety	Primary Region
TAGT	Tagnanan Tall	SE Asia
PYT	Polynesia Tall	Polynesia
RIT	Rennell Island Tall	Melanesia
BAOT	Bago-Oshiro Tall	SE Asia
LAGT	Laguna Tall	SE Asia
BAYT	Baybay Tall	SE Asia
WAT	West African Tall	Africa
TTPT	Tutupaen Tall	SE Asia

Genomic DNA extraction
Total genomic DNA (gDNA) was isolated from young leaf samples using DNeasy® Plant Mini Kit (QIAGEN, Germany) following manufacturer's instructions. The total genomic DNA samples were then assessed for quality and quantity using NanoDrop™ Lite spectrophotometer (Thermo Fisher Scientific, USA).

PCR amplification and Sequencing
Amplification of the putative *CnSHELL* gene region was accomplished using primer pairs of 5'-TTGCTTT TAATTTTGCTTGAATACC-3' (EgSHP-Forward) and 5'-TTTGGATCAGGGATAAAAGGGAAGC-3' (EgSHP-Reverse) (Singh et al. 2013; Babu et al. 2017). PCR reactions were conducted in 50 L volume containing 25 L 2x MyTaq HS Red Mix (Meridian Bioscience Inc., USA), 0.5 L of 20 pmol each of forward and reverse primer, 23 L of nuclease-free water (Vivantis Technologies Sdn Bhd, Malaysia) and 1 L of DNA template. The optimized PCR cycle consists of an initial denaturation at 95ºC for 3 min, followed by 35 cycles of denaturation at 95ºC for 30 sec, annealing at 55.4ºC for 30 sec, extension at 72ºC for 10 sec, and one cycle of final extension at 72ºC for 3 min. The amplified PCR products were electrophoresed in a 1.5% agarose gel stained with GelRed® (Biotium, USA) in 1X TAE buffer at 80V for 55 min. After electrophoresis, the gel was imaged using Gel Doc EZ system (Bio-Rad Laboratories Inc., USA). The PCR products were purified and sequenced at Macrogen Inc. (Seoul, South Korea).

Molecular analysis
DNA sequences were processed using AB1 trace files in SnapGene version 6.1 (GSL Biotech; www.snapgene.com). For each sample, forward and reverse sequences were trimmed, assembled and consensus sequences were generated using CAP3 (Huang and Madan 1999) in SnapGene v6.1 (GSL Biotech; www.snapgene.com). Multiple sequence alignment was performed by using ClustalW (Thompson et al. 1994) algorithm in MEGA 11 (Tamura et al. 2021). Alignments were manually edited and Gblocks 0.91b (Castresana, 2000) was used to remove poorly aligned positions and divergent regions which allows smaller final blocks.

BLAST (Altschul et al. 1990) analysis was performed on individual sequences to find homologous sequences deposited at the NCBI GenBank. Also, the consensus sequence derived from multiple sequence alignment was used as query for BLASTN against the reference genomes of *Cocos nucifera* deposited at NCBI. A physical map was then constructed using the R package "chromoMap" (Anand, 2022) in RStudio version 4.1.2 (https://www.r-project.org/, https://www.rstudio.com) to localize the sequence on the reference genomes.

The optimal model of DNA substitution was determined using Akaike Information Criterion (AIC) method (Posada and Buckley, 2004) in the jModelTest ver. 2.1.10 v20160303 (Darriba et al. 2012) on XSEDE on the CIPRES Science Gateway v3.3 (http://www.phylo.org/, Miller et al. 2010). DNA sequences were also evaluated for evidence of saturation using PAUP* version 4.0a169 (Swoffold, 2003) to compute for the corrected and uncorrected distances. Xia test implemented in DAMBE5 (Xia, 2013) was conducted to check for substitution oversaturation based on the concept of entropy information theory (Xia et al. 2003).

To estimate genetic diversity, the number of haplotypes (*H*), haplotype diversity (*Hd*), and nucleotide diversity () was calculated by using DnaSP version 6.12.03 (Rozas et al. 2017).

The maximum likelihood analysis was conducted using RAxML-NG version 1.0.1 (Kozlov et al. 2019) on

Table 2. Nucleotide composition of putative *CnSHELL* gene region of tall coconut cultivars

Cultivar	Actual Sequence Length (bp)	Nucleotide (%)			
		T(U)	C	A	G
Cocos nucifera cv. Tagnanan Tall "TAGT3517"	508	32.68	19.69	26.97	20.67
Cocos nucifera cv. Tagnanan Tall "TAGT5531"	522	34.1	19.73	25.86	20.31
Cocos nucifera cv. Tagnanan Tall "TAGT0112"	504	33.93	19.44	25.79	20.83
Cocos nucifera cv. Polynesian Tall "PYT0313"	500	33.8	19.4	26	20.8
Cocos nucifera cv. Polynesian Tall "PYT0118"	505	34.46	19.41	25.35	20.79
Cocos nucifera cv. Polynesia Tall "PYT0416"	506	32.61	19.76	26.68	20.95
Cocos nucifera cv. Rennell Island Tall "RIT1215"	507	34.52	19.33	25.44	20.71
Cocos nucifera cv. Rennell Island Tall "RIT1812"	526	33.65	19.58	26.05	20.72
Cocos nucifera cv. Rennell Island Tall "RIT2518"	485	32.78	19.79	26.19	21.34
Cocos nucifera cv. Bago-Oshiro Tall "BAOT1512"	485	32.78	19.79	26.19	21.24
Cocos nucifera cv. Bago-Oshiro Tall "BAOT0305"	503	34	19.48	25.65	20.87
Cocos nucifera cv. Bago-Oshiro Tall "BAOT0410"	510	34.31	19.41	25.69	20.59
Cocos nucifera cv. Laguna Tall "LAGT1014"	526	33.46	19.77	26.24	20.53
Cocos nucifera cv. Laguna Tall "LAGT0413"	525	33.9	19.43	26.29	20.38
Cocos nucifera cv. Laguna Tall "LAGT0511"	486	32.92	19.75	26.13	21.19
Cocos nucifera cv. Baybay Tall "BAYT0813"	492	33.13	19.72	26.22	20.93
Cocos nucifera cv. Baybay Tall "BAYT0404"	486	32.92	19.75	26.13	21.19
Cocos nucifera cv. Baybay Tall "BAYT1008"	481	32.43	19.75	26.61	21.21
Cocos nucifera cv. West African Tall "WAT0509"	486	32.92	19.54	26.34	21.19
Cocos nucifera cv. West African Tall "WAT0219"	528	33.14	19.13	27.27	20.45
Cocos nucifera cv. West African Tall "WAT0310"	527	33.78	19.35	26.38	20.49
Cocos nucifera cv. Tutupaen Tall "TTPT04"	505	34.26	19.21	25.74	20.79
Average	504.68	33.49	19.55	26.15	20.81

the CIPRES Science Gateway v.3.3 (http://www.phylo.org/, Miller et al. 2010) with an F81 substitution model, as determined in jModelTest, and 1,000 bootstrap iterations. The tree was then visualized and edited in FigTree version 1.4.4 (Rambaut 2014). DNA sequence from a closely related taxon *Elaeis guineensis* was used as an outgroup and was retrieved from GenBank with accession number KX465099.

RESULTS AND DISCUSSION

Amplification and Analysis of Putative *CnSHELL* Gene Region

PCR amplification was successfully performed on the 22 tall coconut cultivars using primers EgSHP-Forward and EgSHP-Reverse, which resulted in 450-550 bp DNA fragments (Figure 1). Bidirectional sequencing of the amplicons generated 481-528 bp DNA sequences (Table 2). Based on BLASTn analysis, the sequences are homologous to *Elaeis guineensis* shell-specific gene (87.4-87.7%)(Supplementary Table 2). Moreover, BLASTn result of the putative *CnSHELL* gene region against NCBI's *Cocos nucifera* whole genome sequence (WGS) database showed significant hits on *Cocos nucifera* Hainan Tall coconut COCNU_scaffold004872 (99.0%), *Cocos nucifera* cultivar Catigan Green Dwarf Scaff_1 (99.0%), and *Cocos nucifera* Chowghat Green Dwarf scaffold1993 whole genome shotgun sequences (99.0%)(Supplementary Table 3). The putative *CnSHELL* gene region is located in the following scaffolds: 521,945 to 522,432 nt position in *Cocos nucifera* cultivar Hainan Tall coconut COCNU_scaffold004872; 5,181,909 to 5,182,396 nt position in *Cocos nucifera* cultivar Catigan Green Dwarf Scaff_1; and 49,838 to 50,325 nt position in *Cocos nucifera* cultivar Chowghat Green Dwarf scaffold1993 (Figure 3).

The nucleotide composition of putative *CnSHELL* gene region in the tall coconut cultivars was dominated by T(U) and A nucleotide bases rather than C and G (Table 2). The average composition value of each base was 33.49% T(U), 26.15% A, 20.81% G, and 19.55% C (Table 2). Using the Akaike Information Criterion (AIC), the identified optimal DNA substitution model in jModelTest is F81 for the putative *CnSHELL* gene partial sequence (Supplementary Table 4). Also, the putative *CnSHELL* gene in the different tall cultivars varied only in 2 out of 470 sites. The uncorrected vs corrected distance plot (Figure 2) showed a direct relationship between the corrected distances based on the optimal F81 model and the uncorrected distances, which signifies adequate correction of the model and an indication of a highly conserve dataset without saturation. Furthermore, Xia test yielded an Iss<Iss.c, in which Iss values are significantly lower ($P<0.0000$) than Iss.c in the putative *CnSHELL* gene region (Supplementary Table 5). This indicate that there was no substantial saturation in the dataset.

Genetic Diversity Analysis

Analysis of genetic diversity revealed three (3) *CnSHELL* haplotypes (Table 3) with two (2) segregating sites (Table 4). Philippines had 2 haplotypes while Polynesia, Melanesia, and Africa had only one haplotype. Haplotype diversity was higher in the Philippines at 0.28205 compared to Polynesia, Melanesia, and Africa at 0.0000 for all three regions.

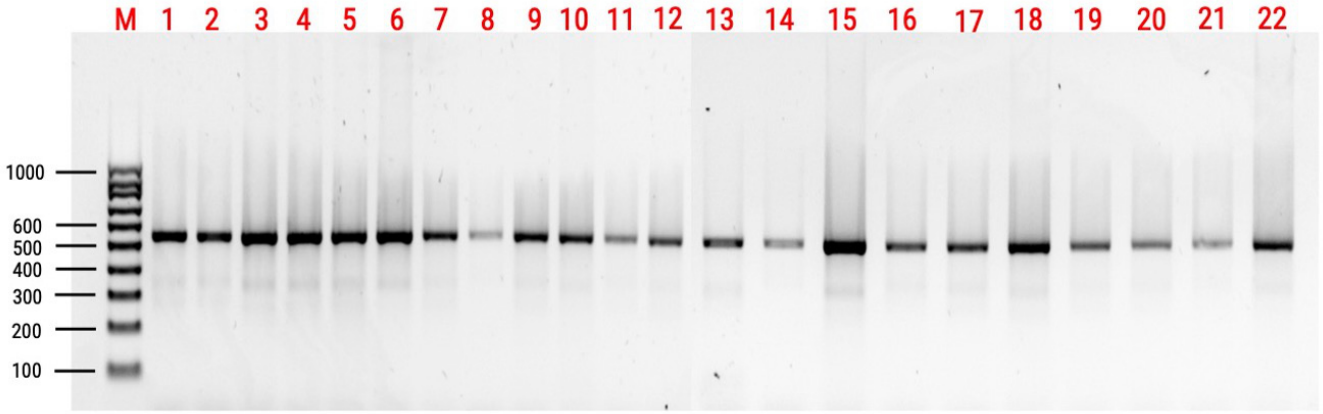

Figure 1. Agarose gel electrophoresis image showing PCR amplicon of putative *CnSHELL* gene region of *Cocos nucifera*. Lane 1 - TAGT3517, Lane 2 - TAGT5531, Lane 3 - TAGT0112, Lane 4 - PYT0313, Lane 5 - PYT0118, Lane 6 - PYT0416, Lane 7 - RIT1215, Lane 8 - RIT1812, Lane 9 - RIT2518, Lane 10 - BAOT1512, Lane 11 - BAOT0305, Lane 12 - BAOT0410, Lane 13 - LAGT1014, Lane 14 - LAGT0413, Lane 15 - LAGT0511, Lane 16 - BAYT0813, Lane 17 - BAYT0404, Lane 18 - BAYT1008, Lane 19 - WAT0509, Lane 20 - WAT0219, Lane 21 - WAT0310, Lane 22 - TTPT04, M - EZ load 100 bp molecular rules. TAGT: *Cocos nucifera* cv. "Tagnanan Tall", PYT: *Cocos nucifera* cv. "Polynesia Tall", RIT: *Cocos nucifera* cv. "Rennell Island Tall", BAOT: *Cocos nucifera* cv. "Bago-Oshiro Tall", LAGT: *Cocos nucifera* cv. "Laguna Tall", BAYT: *Cocos nucifera* cv. "Baybay Tall", WAT: *Cocos nucifera* cv. "West African Tall", TTPT: *Cocos nucifera* cv. "Tutupaen Tall"

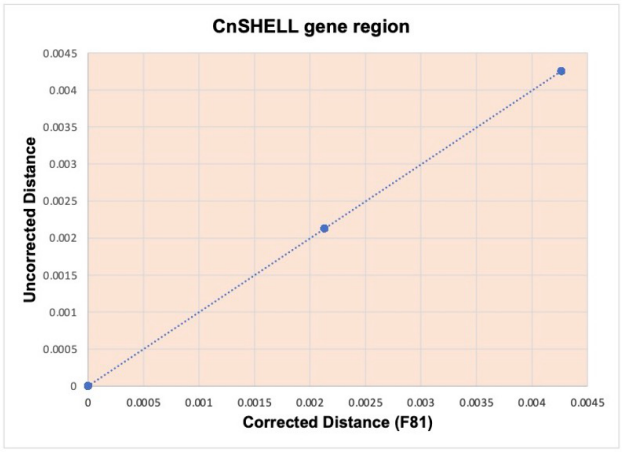

Figure 2. Plot of pairwise uncorrected distance versus corrected distance (F81) for the putative *CnSHELL* gene region of *Cocos nucifera* tall cultivars. The distances were computed based only on variable sites.

Table 3. Diversity indices based on the putative *CnSHELL* gene region of coconut tall cultivars.

Population	Number of Haplotypes (H)	Haplotype Diversity (Hd)	Nucleotide Diversity (p)
Philippines	2	0.28205	0.0006
Polynesia	1	0	0
Melanesia	1	0	0
Africa	1	0	0
Total Data Estimates	3	0.39394	0.00131

Also, nucleotide diversity was higher in the Philippines at 0.00060 compared to the other three regions – Polynesia, Melanesia, and Africa at 0.0000. The average of haplotype and nucleotide diversities in nine (9) tall coconut cultiars was 0.39394 and 0.00131, respectively. The average nucleotide diversity

Table 4. Variable nucleotide sites of the *CnSHELL* gene sequences in the coconut tall cultivars.

Population	Haplotype	CnSHELL (nt position) 37	70
TAGT, PYT, RIT, BAOT, LAGT, BAYT, TTPT	Hap1	T	C
BAYT	Hap2	*/A	*
WAT	Hap3	A	T

obtained from this study was lower compared to those reported by previous studies using *rbc*L (= 0.51) and *mat*K (= 0.0258) (Mursyidin & Ahyar, 2022; Mursyidin et al. 2022). This could indicate that the putative *CnSHELL* gene has lower genetic variation compared to the chloroplastic genes *rbc*L and *mat*K. Low level of nucleotide diversity has also been reported in *Rhizophora spp*. and *Vitex rotundifolia* which are coastal plants (Lo et al. 2014; Sun et al. 2019) like coconut. Also, long distance dispersal of sea-drifted fruits such as in the case of coconut can result in the uniform distribution of DNA haplotypes in spatial scale (Miryeganeh, 2013). The coconut fruit can float in water due to the internal air cavity and the fibrous external husk protects the internal seed which enable it to be dispersed via sea drift. Thus, this ability of the coconut to be dispersed over long distance by means of sea drift may be responsible for its low genetic diversity. In addition, the genetic diversity of various plant species depends on their breeding system, life form, dispersal mechanism, geographic variation and range (Nybom, 2004; Huang et al. 2016). Furthermore, the primers used in this study targets the exon 1 region of *E. guineensis SHELL* gene, which could have also amplified an exon region of coconut *SHELL* gene. Exons are usually more conserved at non-synonymous sites (Small et al. 2004).

Maximum Likelihood (ML) Phylogenetic Analysis

The ML phylogenetic tree generated from the putative *CnSHELL* gene region of 22 coconut tall cultivars

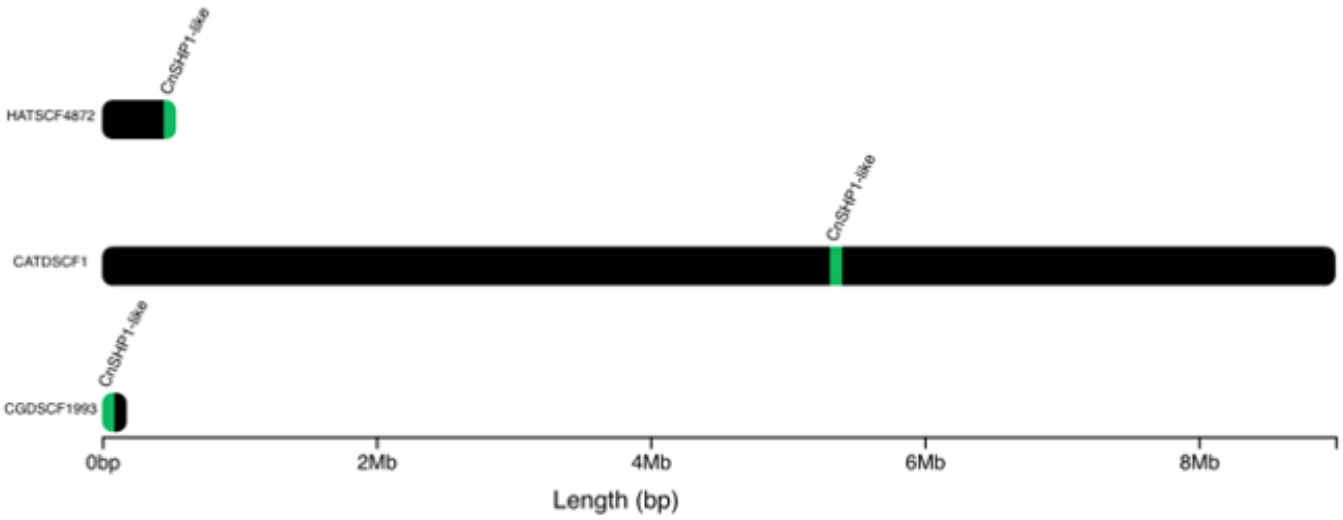

Figure 3. Physical map of *Cocos nucifera* SHP1-like gene sequence in the genomic scaffold assembly of *Cocos nucifera* Hainan Tall cultivar (HATSCF4872), *Cocos nucifera* Green Dwarf cultivar (CATDSCF1), and *Cocos nucifera* Chowghat Green Dwarf cultivar (CGDSCF1993). The map was visualized through the R package chromoMap.

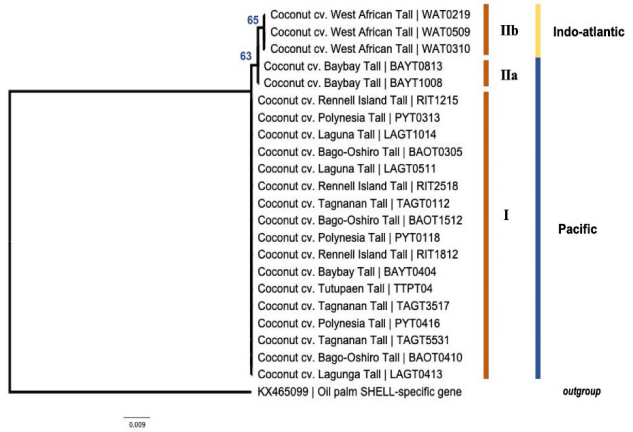

Figure 4. Maximum likelihood phylogenetic tree of *Cocos nucifera* (coconut) tall cultivars using the putative *CnSHELL* gene region (470 nucleotides). The tree was rooted on the taxon *Elaeis guineensis* (oil palm). The values on nodes represent bootstrap branch support with 1,000 replicates. Bootstrap support <50% is not shown. The bar marker indicates average number of substitutions per site (or 9 nucleotide substitutions per 1000 nucleotides).

using the optimal F81 model of DNA substitution showed only one major cluster, nonetheless, groupings can be observed (Figure 4). The first group includes coconut tall cultivars coming from the South Pacific and Southeast Asia while the second group includes cultivars introduced from the Southeast Asia and Africa (BS = 63%). However, sequence diversity may be insufficient to be used for tree construction which implies that the putative *CnSHELL* region is too conserved to provide topology to the tree. Clustering analysis using microsatellite markers detected the formation of two main groups in which the first group was formed by coconuts from Africa and South America while the second group included coconuts introduced from the South Pacific and Southeast Asia (Loiola et al. 2016). Moreover, Geethanjali et al. (2018) conducted SSR marker analysis on several dwarf and tall coconut cultivars wherein they also detected two major clusters in which the first group (C1) was comprised of coconuts from the Indo-atlantic region, Carribean Islands, and the South Asian countries while the second group (C2) included dwarfs and tall coconut cultivars from the Indian Ocean, Oceania, and South Pacific. The groupings generated by the ML algorithm represents groups from the Pacific and the Indo-Atlantic which have also been noted by Gunn et al. (2011). The groupings were likely due to separate domestication events in the Pacific and Indian Ocean basins. Island Southeast Asia was also proposed to be one of the geographic origins of coconut cultivation (Gunn et al. 2011). Interestingly, group 2 possess subgroups which indicates that absence of allele sharing in these coconut cultivars and those introduced from Southeast Asia and South Pacific. These introductions clearly shows that it did not involve admixed populations. Furthermore, the genetic closeness between Baybay Tall (BAYT) and West African Tall (WAT) cultivars confirms the common origin of these tall coconut varieties. It has been reported that historical accounts show that tall coconut was introduced from the Cape Verde Islands (Loiola et al. 2016).

CONCLUSION

The putative *CnSHELL* gene region was successfully amplified through PCR across 22 tall coconut cultivars originating from different regions. The size of amplicons ranged from 450-550 bp and it is homologous with *E. guineensis* shell-specific gene. The *CnSHELL* gene partial sequence can also be mapped to whole genome sequences deposited in NCBI (Hainan Tall, Catigan Green Dwarf, and Chowghat Green Dwarf). The average haplotype and nucleotide diversities of *CnSHELL* gene region is quite less (Hd = 0.39394, = 0.00131) which indicates low genetic variation. This suggests that the amplified segment of the *CnSHELL* is a highly conserve region.

Supplementary Table 1. Quality and quantity of extracted total genomic DNA from *Cocos nucifera* tall cultivars

Sample Code	Concentration (ng/uL)	Purity (A260/A280)
TAGT3517	46.8	1.84
TAGT5531	39.3	1.84
TAGT0112	118	1.84
PYT0313	146.8	1.83
PYT0118	93.5	1.84
PYT0416	173.8	1.85
RIT1215	54.6	1.84
RIT1812	5.5	1.77
RIT2518	47.4	1.84
BAOT1512	133.1	1.83
BAOT0305	42.4	1.85
BAOT0410	61.7	1.85
LAGT1014	8.6	1.84
LAGT0413	4.8	1.74
LAGT0511	637.8	1.85
BAYT0813	28.9	1.86
BAYT0404	71	1.84
BAYT1008	79.7	1.84
WAT0509	25.8	1.85
WAT0219	29.5	1.82
WAT0310	13.4	1.74
TTPT04	649.5	1.82

TAGT: *Cocos nucifera* cv. "Tagnanan Tall"; PYT: *Cocos nucifera* cv. "Polynesia Tall"; RIT: *Cocos nucifera* cv. "Rennell Island Tall"; BAOT: *Cocos nucifera* cv. "Bago-Oshiro Tall"; LAGT: *Cocos nucifera* cv. "Laguna Tall"; BAYT: *Cocos nucifera* cv. "Baybay Tall"; WAT: *Cocos nucifera* cv. "West African Tall"; TTPT: *Cocos nucifera* cv. "Tutupaen Tall"

Thus, the results of this study shows that the isolated *CnSHELL* gene region is too conserved to infer relationships among the tall coconut cultivars.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Susan Rivera and Mr. Ernesto E. Emmanuel of the Philippine Coconut Authority – Zamboanga Research Center (PCA) for providing assistance during the collection of leaf samples in the germplasm. We would also like to thank the Plant Molecular Phylogenetics Laboratory (PMPL) for the provision of necessary equipment and materials needed for the conduct of the study.

LITERATURE CITED

Alejadria L, Dalmacio LM. 2021. Virgin Coconut Oil as Adjunctive Therapy for Hospitalized COVID-19 Patients: Identifier NCT0489637. Available online: https://clinicaltrials.gov/ct2/show/NCT0489637 (accessed on 1 July 2022).

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3): 403-10.

Supplementary Table 2. BLAST analysis result based on putative *CnSHELL* gene region of coconut tall cultivars

Sample	BLAST result				
	Description	Query Coverage	E value	Percent Identity	Accession No.
TAGT3517	Elaeis guineensis shell-specific protein 6 gene, partial cds	81%	3.00E-134	87.65%	KX465099
TAGT5531	Elaeis guineensis shell-specific protein 6 gene, partial cds	84%	3.00E-134	87.65%	KX465099
TAGT0112	Elaeis guineensis shell-specific protein 6 gene, partial cds	83%	3.00E-134	87.65%	KX465099
PYT0313	Elaeis guineensis shell-specific protein 6 gene, partial cds	81%	3.00E-134	87.65%	KX465099
PYT0118	Elaeis guineensis shell-specific protein 6 gene, partial cds	84%	3.00E-134	87.65%	KX465099
PYT0416	Elaeis guineensis shell-specific protein 6 gene, partial cds	83%	3.00E-134	87.65%	KX465099
RIT1215	Elaeis guineensis shell-specific protein 6 gene, partial cds	83%	3.00E-134	87.65%	KX465099
RIT1812	Elaeis guineensis shell-specific protein 6 gene, partial cds	84%	3.00E-134	87.65%	KX465099
RIT2518	Elaeis guineensis shell-specific protein 6 gene, partial cds	82%	3.00E-134	87.65%	KX465099
BAOT1512	Elaeis guineensis shell-specific protein 6 gene, partial cds	81%	3.00E-134	87.65%	KX465099
BAOT0305	Elaeis guineensis shell-specific protein 6 gene, partial cds	83%	3.00E-134	87.65%	KX465099
BAOT0410	Elaeis guineensis shell-specific protein 6 gene, partial cds	84%	3.00E-134	87.65%	KX465099
LAGT1014	Elaeis guineensis shell-specific protein 6 gene, partial cds	84%	3.00E-134	87.65%	KX465099
LAGT0413	Elaeis guineensis shell-specific protein 6 gene, partial cds	84%	3.00E-134	87.65%	KX465099
LAGT0511	Elaeis guineensis shell-specific protein 6 gene, partial cds	81%	3.00E-134	87.65%	KX465099
BAYT0813	Elaeis guineensis shell-specific protein 3 gene, partial cds	81%	4.00E-133	87.63%	KX465097
BAYT0404	Elaeis guineensis shell-specific protein 6 gene, partial cds	81%	3.00E-134	87.65%	KX465099
BAYT1008	Elaeis guineensis shell-specific protein 3 gene, partial cds	81%	4.00E-133	87.63%	KX465097
WAT0509	Elaeis guineensis shell-specific protein 3 gene, partial cds	80%	5.00E-132	87.37%	KX465097
WAT0219	Elaeis guineensis shell-specific protein 3 gene, partial cds	83%	5.00E-132	87.37%	KX465097
WAT0310	Elaeis guineensis shell-specific protein 3 gene, partial cds	83%	5.00E-132	87.37%	KX465097
TTPT04	Elaeis guineensis shell-specific protein 6 gene, partial cds	81%	3.00E-134	87.65%	KX465099

Supplementary Table 3. BLASTn result of the consensus *Cocos nucifera* SHP1-like partial gene sequence against NCBI's *Cocos nucifera* WGS database

Query	BLASTn Result								
	Description	% Identity	Length	Max Score	Total Score	Query Cover	E-value	Acc. Len	Accession
Consensus CnSHELL gene partial sequence	Cocos nucifera cultivar Hainan Tall coconut COCNU_scaffold004872, whole genome shotgun sequence	99.02	530	893	958	96%	0	543696	VOII01004016
	Cocos nucifera cultivar Catigan Green Dwarf Scaff_1, whole genome shotgun sequence	99.02	530	893	958	96%	0	8779653	QRFJ01000002
	Cocos nucifera Chowghat Green Dwarf scaffold1993, whole genome sequence	99.02	530	893	958	96%	0	150385	PDMH01001993

Supplementary Table 4. Output of JModelTest showing the relevant parameters for the Felsenstein substitution model based on Akaike Information Criterion (AIC) on the putative *CnSHELL* gene region of *Cocos nucifera* tall cultivars. The frequency of each of the bases (Freq) and values for the rate matrix (R) are provided.

Parameter	Value
Model Selected	F81
Freq [A]	0.2648
Freq [C]	0.1906
Freq [G]	0.2161
Freq [T]	0.3285
R [AC]	0.3088
R [AG]	0.5597
R [AT]	0.3386
R [CG]	0.3199
R [CT]	0.5597
R [GT]	0.5597

Supplementary Table 5. Assessment of substitution saturation of the putative *CnSHELL* gene region with DAMBE

DF	Iss	Iss.c Sym	P-value	If Iss is significantly smaller that Iss.c Sym
469	0.0015	0.7075	0	Yes

Anand L, Rodriguez Lopez CM. 2022. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. BMC Bioinform. 23(33).

Arulandoo X, Sritharan K, Subramaniam M. 2017. The coconut palm: Encyclopedia of Applied Plant Sciences. 2nd ed. Academic Press. 426-430.

Babu BK, Mathur RK, Kumar PN, Ramajayam D, Ravichandran G, Venu MV, Babu SS. 2017. Development, identification, and validation of CAPS marker for SHELL trait which governs dura, pisifera and tenera fruit forms in oil palm (*Elaeis guineensis* Jacq.). PLoS One. 12(2): e0171933.

Castresana J. 2000. Selection of conserved blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol. 17(4): 540-552.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 9(8): 772.

FAOSTAT. 2020. Commodities by country. Available online: http://www.fao.rg/faostat/en/#rankings/commodities_by_country (accessed on 1 July 2022).

FFTC-AP. 2019. The Philippine Coconut Industry: Status, Policies and Strategic Directions for Development. Available online: https://ap.fftc.org.tw/article/1382 (accessed on 1 July 2022).

Favaro R, Pinyopic A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, Kater MM, Colombo L. 2003. MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. Plant Cell. 15(11): 2603-2611.

Freitas Neto M, Pereira TN, Geronimo IG, Azevedo AO, Ramos SR, Pereira MG. 2016. Coconut genome size determined by flow cytometry: Tall versus Dwarf types. Genet Mol Res. 15(1).

Geethanjali S, Rukmani A, Rajakumar D, Kadirvel P, Viswanathan P. 2018. Genetic diversity, population structure and association analysis in coconut (Cocos nucifera L.) germplasm using SSR markers. Plant Genet Resour. 16(2): 156-168.

Gunn BF, Baudouin L, Beulé T, Ilbert P, Duperray C, Crisp M, Issali A, Konan JL, Rival A. 2015. Ploidy and domestication are associated with genome size variation in Palms. Am J Bot. 102(10): 1625-33.

Gunn BF, Baudouin L, Olsen KM. 2011. Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. PLoS One. 6(6): e21143.

Huang W, Zhao X, Zhao X, Li Y, Lian J. 2016. Effects of environmental factors on genetic diversity of Caragana microphylla in Horqin Sandy Land, northeast China. Ecol Evol. 6(22):8256-8266.

Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9(9): 868-77.

Kozlov AM, Darriba D, Tomáš F, Benoit M, Stamatakis A. 2019. RAxML-NG: a fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 35(21): 4453-4455.

Lantican DV, Strickler SR, Canamo AO, Gardoce RR, Mueller LA, Galvez HF. 2019. *De Novo* Genome Sequence Assembly of Dwarf Coconut (*Cocos nucifera* L. 'Catigan Green Dwarf') Provides Insights into Genomic Variation Between Coconut Types and Related Palm Species. G3 (Bethesda). 9(8): 2377-2393.

Lo EY, Duke NC, Sun M. 2014. Phylogeographic pattern of Rhizophora (Rhizophoraceae) reveals the importance of both vicariance and long-distance oceanic dispersal to modern mangrove distribution. BMC Evol Biol. 14(83).

Loiola CM, Azevedo AON, Diniz, LEC, Aragão WM, Azevedo CDdO, Santos PHAD, Ramos HCC, Pereira MG, Ramos SRR. 2016. Genetic Relationships among Tall Coconut Palm (*Cocos nucifera* L.) Accessions of the International Coconut Genebank for Latin America and the Carribean (ICG-LAC), Evaluated Using Microsatellite Markers (SSRs). PLoS One 11(3): e0151309.

Meerow AW, Noblick L, Salas-Leiva DE, Sanchez V, Francisco-Ortega J, Jestrow B, Nakamura K. 2015. Phylogeny and historical biogeography of the cocosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. Cladistics. 31(5): 509-534.

Menon KPV, Pandalai KM. 1958. The coconut palm: a monograph. Ernakulam India: Indian Central Coconut Committee. 384 p.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop (GCE). 8 p.

Miryeganeh M. 2013. Phylogeography of a pantropical plant with sea-drifted seeds, Ipomoea pes-caprae (L.) Roth (Convolvulaceae). [D. Phil Thesis]. 86 p. (Available at Chiba University Library).

Muliyar RK, Chowdappa P, Behera SK, Kasaragod S, Gangaraj KP, Kotimoole CN, Nekrakalaya B, Mohanty V, Sampgod RB, Banerjee G, Das AJ, Niral V, Karun A, Mahato AK, Gaikwad K, Singh NK, Prasad TSK. 2020. Assembly and Annotation of the Nuclear and Organellar Genomes of a Dwarf Coconut (Chowghat Green Dwarf) Possessing Enhanced Disease Resistance. OMICS. 24(12): 726-742.

Mursyidin DH, Ahyar MZ. 2022. Genetic diversity of coconut germplasm native to South Kalimantan, Indonesia: a molecular study. Pesq Agropec Trop. 52: e71016.

Mursyidin DH, Nazari YA, Ahyar GMZ, Makruf MI. 2022. Molecular identity of native coconut (Cocos nucifera) germplasm from South Kalimantan, Indonesia. Aust J Crop Sci. 16(03):424-430.

Nybom H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol. 13(5):1143-55.

PCA. 2019. Coconut: Major Export Crop of Filipino Farmers. Available online: https://pca.gov.ph/index.php/10-news/234-coconut-major-export-crop-of-filipino-farmers (accessed on 1 July 2022.

Pereira TNS, Neto MF, de Souza MM, da Costa Geronimo IG, de Melo CAF, Pereira MG. 2017. Cytological Characterization of Brazilian Green Dwarf Coconut (*Cocos nucifera* L.) via Meiosis and Conventional and Differential Karyotyping. Cytologia. 82(2): 167-174.

Perera L, Baudouin L, Mackay I. 2016. SSR markers indicate a common origin of self-pollinating dwarf coconut in South-East Asia under domestication. Sci Hortic. 211: 255-262.

Perera PIP, Wickremasinghe IP, Fernando WMU. 2008. Morphological, cytogenetic, and genotypic differences between spicata and ordinary tall coconut (*Cocos nucifera* L.). Nat Sci Found Sri Lanka. 36(1): 103-108.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol. 53(5): 793-808.

Rambaut A. 2018. FigTree-v1.4.4. Available online: https://www. http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 1 July 2022).

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Garcia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol. 34(12): 3299-3302.

Singh R, Low ET, Ooi LC, Ong-Abdullah M, Ting NC, Nagappan J, Nookiah R, Amiruddin MD, Rosli R, Manaf MA, Chan KL, Halim MA,

Azizi N, Lakey N, Smith SW, Budiman MA, Hogan M, Bacher B, Van Brunt A, Wang C, Ordway JM, Sambanthamurthi R, Martienssen RA. 2013. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature. 500(7462): 340-4.

Sisunandar S, Adkins SW. 2007. Karyotype of some Indonesian coconut (*Cocos nucifera* L.) cultivars. Chrom Res. 15:36-37.

Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. Aust Syst Bot. 17: 145-170.

Sun Y, Yang H, Zhang Q, Qin L, Li P, Lee J, Chen S, Rahman K, Kang T, Jia M. 2019. Genetic diversity and its conservation implications of *Vitex rotundifolia* (Lamiaceae) populations in East Asia. PeerJ 7: e6194.

Swoffold DL. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 38(7): 3022-3027.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22): 4673-80.

Wang S, Xiao Y, Zhou ZW et al. 2021. High-quality reference genome sequences of two coconut cultivars provide insights into evolution of monocot chromosomes and differentiation of fiber content and plant height. Genome Biol. 22(304).

Xia X. 2013. DAMBE5: A Comprehensive Software Package for Data Analysis in Molecular Biology and Evolution. Mol Biol Evol. 30(7): 1720-1728.

Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003. An index of substitution saturation and its application. Mol Phylogenet Evol. 26(1): 1-7.

Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, Xu J, Li Q, Guo A, Zhou L, Li J, Wu Y, Ma Z, Armero A, Issali AE, Liu N, Peng M, Yang Y. 2017. The genome draft of coconut (Cocos nucifera). Gigascience. 6(11): 1-11.