

## Research Article

# Isolation of Alcohol Dehydrogenase cDNA and Basal Regulatory Region from *Metroxylon sagu*

**Ching Ching Wee and Hairul Azman Roslan**

*Genetic Engineering Laboratory, Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Kota Samarahan, 94300 Sarawak, Malaysia*

Correspondence should be addressed to Hairul Azman Roslan, hairulroslan@hotmail.com

Received 19 June 2012; Accepted 10 July 2012

Academic Editors: A. Montecucco and T. Tozaki

Copyright © 2012 C. C. Wee and H. A. Roslan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Alcohol dehydrogenase (Adh) is a versatile enzyme involved in many biochemical pathways in plants such as in germination and stress tolerance. Sago palm is plant with much importance to the state of Sarawak as one of the most important crops that bring revenue with the advantage of being able to withstand various biotic and abiotic stresses such as heat, pathogens, and water logging. Here we report the isolation of sago palm *Adh* cDNA and its putative promoter region via the use of rapid amplification of cDNA ends (RACE) and genomic walking. The isolated cDNA was characterized and determined to be 1464 bp long encoding for 380 amino acids. BLAST analysis showed that the *Adh* is similar to the *Adh1* group with 91% and 85% homology with *Elaeis guineensis* and *Washingtonia robusta*, respectively. The putative basal *msAdh1* regulatory region was further determined to contain promoter signals of TATA and AGGA boxes and predicted amino acids analyses showed several *Adh*-specific motifs such as the two zinc-binding domains that bind to the adenosine ribose of the coenzyme and binding to alcohol substrate. A phylogenetic tree was also constructed using the predicted amino acid showed clear separation of *Adh* from bacteria and clustered within the plant *Adh* group.

## 1. Introduction

Alcohol dehydrogenase (Adh) is an enzyme involved in various biological activities such as in the germination and abiotic stresses in plants [1–3]. Previous studies have shown that there are between two or three *Adh* loci in flowering plants with exception in *Arabidopsis* [4, 5]. Previous Adh protein work on sago palm, a flood-tolerant plant, by Roslan et al. [6] detected the presence of Adh in the leaf and roots. A higher Adh enzyme expression was observed in sago palm young shoots compared to the other part of *Metroxylon sagu* [6]. The finding was consistent with those of Padmanabhan and Sahi [7] that reported a greater increase in Adh activity in the leaves than the roots of sunflower that was treated with high phosphorus. In contrast, in flood-intolerant plants such as *Arabidopsis* and pea, increased Adh activity was determined in the roots than in the shoots under anaerobic condition [8, 9]. A higher expression level in different tissue and developmental stage may be

because the cells are dividing and exposed to many stresses [10].

The discovery of Adh protein expression in young leaf prompted the work to isolate the *Adh* gene from sago palm. The isolation of the regulatory region was also conducted to further understand the regulation of *Adh* in sago palm. *Adh* gene have been isolated from several techniques from a number of plants such as in *Arabidopsis thaliana*, barley, maize [4, 5, 11], and including *Washingtonia robusta*, a member of same Arecaceae family with *Metroxylon sagu* [12, 13].

In this study, we report the isolation of full length *Adh* cDNA and the regulatory sequences from sago palm leaf using RACE and genomic walking methods. Full length cDNA was isolated using the RACE technique that is faster and less laborious compared to the screening of cDNA library by using gene-specific probe [14]. The *Adh* regulatory sequences of *M. sagu* was also isolated using the genomic walking.

## 2. Material and Methods

**2.1. Plant Material.** Young leaves of sago palm (*Metroxylon sago*) were obtained from Universiti Malaysia Sarawak (UNIMAS) plant house. The stems of the leaves were discarded. The samples were sterilised with 70% ethanol before being cut into small pieces. The samples were stored at  $-80^{\circ}\text{C}$  to preserve the RNA integrity.

**2.2. Isolation of RNA and DNA.** RNA isolation was carried out according to the method described by Wee and Roslan [15] while total genomic DNA was extracted from young leaves by using modified Doyle and Doyle method [16].

**2.2.1. Isolation of RNA and 1st Strand cDNA Synthesis.** Approximately 5 g of leaves was ground to fine powder in liquid nitrogen. The powdered tissue was transferred into a pre-warmed ( $60^{\circ}\text{C}$ ) 10 mL extraction buffer [2% cetyl trimethylammonium bromide (CTAB), 2% polyvinylpyrrolidone (PVP 40), 100 mM Tris HCl (pH 8.0), 25 mM ethylenediaminetetraacetic acid (EDTA), 2.0 M NaCl, and 2%  $\beta$ -mercaptoethanol (added just before use)]. The mixture was incubated for 30 minutes and proceeded with addition of an equal volume of chloroform : isoamylalcohol [24 : 1 (v/v)]. The mixture was then vortexed for 1 minute and centrifuged at 8000 rpm for 30 min at  $4^{\circ}\text{C}$ . The upper aqueous phase was transferred to a 1.5 mL microcentrifuge tube and reextracted with equal volume of chloroform : isoamylalcohol. Next, one-third volume of 8 M lithium chloride solution was added to the upper aqueous phase in a new tube to a final concentration of 2 M. The tube was then incubated overnight at  $-20^{\circ}\text{C}$  and centrifuged at 13,000 rpm for 30 min at  $4^{\circ}\text{C}$ . The supernatant was discarded and the pellet was washed with 500  $\mu\text{L}$  of 70% ethanol followed by 500  $\mu\text{L}$  of 80% ethanol. The pellet was air-dried and resuspended in 35  $\mu\text{L}$  DEPC-treated water. Finally, RNA was kept at  $-80^{\circ}\text{C}$  for long-term storage.

First strand cDNA synthesis was carried out according to RevertAid First strand cDNA synthesis kit protocol (Fermentas). Approximately 1  $\mu\text{g}$  of total RNA isolated was treated with DNase I (Promega), to remove traces of genomic DNA. A 0.5  $\mu\text{g}$  of adaptor(dt)<sub>17</sub> (5'-GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT-3') primer was added to the purified RNA and made up to a final volume of 12.5  $\mu\text{L}$  using nuclease-free water. The mixture was incubated at  $70^{\circ}\text{C}$  for 5 min and chilled on ice. Then, 4  $\mu\text{L}$  of 5X RT-buffer (Fermentas), 2  $\mu\text{L}$  dNTPs (10 mM), and 0.5  $\mu\text{L}$  of 40 U/ $\mu\text{L}$  RNase Inhibitor (Fermentas) were added and incubated at  $37^{\circ}\text{C}$  for 5 min. Lastly, 1  $\mu\text{L}$  of 200 U/ $\mu\text{L}$  RevertAid M-MuLV RT (Fermentas) was added and incubated for 1 hour at  $42^{\circ}\text{C}$ . The reaction was then stopped by heating at  $70^{\circ}\text{C}$  for 10 minutes.

**2.2.2. Isolation of DNA and DNA Purification.** Approximately 0.1 g of sample was ground in liquid nitrogen and mixed with prewarmed ( $65^{\circ}\text{C}$ ) 1 mL of CTAB extraction buffer (0.1 M Tris-HCl pH 8, 0.02 M EDTA pH 8, 1.4 M NaCl, 2% (w/v) CTAB powder, 1% (w/v) PVP, and 0.2%  $\beta$ -mercaptoethanol

and incubated at  $65^{\circ}\text{C}$  for 1 hour. After incubation, the mixture was left for 5 min to cool to room temperature and mixed gently with 400  $\mu\text{L}$  of chloroform : isoamylalcohol (24 : 1). The tube was centrifuged at 13,000 rpm for 5 min at  $4^{\circ}\text{C}$ . The aqueous fraction was reextracted once with chloroform : isoamylalcohol (24 : 1). One volume of ice-cold isopropanol was added and DNA was precipitated overnight at  $-20^{\circ}\text{C}$ . The next day, DNA was collected by centrifugation at 13,000 rpm for 15 min at  $4^{\circ}\text{C}$ . The pellet was washed with 1 mL of wash buffer (76% ethanol and 0.01 M ammonium acetate), air-dried and resuspended in 40  $\mu\text{L}$  TE buffer (10 mM Tris HCl pH 8.0 and 1 mM EDTA, pH 8.0).

Approximately 50  $\mu\text{L}$  of the isolated genomic DNA ( $\approx 15 \mu\text{g}$ ) was mixed with 0.1  $\mu\text{L}$  RNase A (10 mg/mL) and incubated at  $37^{\circ}\text{C}$  for 1 hour. Then, equal volume of P : C : I was added and the mixture was centrifuged at 13,000 rpm for 15 min at  $4^{\circ}\text{C}$ . The aqueous phase was transferred to a new tube. Next, 2/3 volume of isopropanol and 0.1 volume of 3 M NaOAc (pH 5.2) was added to the solution and left to precipitate overnight at  $-20^{\circ}\text{C}$ . After centrifugation at 13,000 rpm for 15 min at  $4^{\circ}\text{C}$ , the pellet was washed with 1 mL wash buffer, air-dried, and dissolved in 50  $\mu\text{L}$  of TE buffer.

**2.3. RACE Primer Design.** Initially *Adh* screening was done via RT-PCR using primers designed from *Washingtonia robusta* described by Morton et al. [12]: morADH-f (5'-GGGTGCTGTAGGCCTTGC-3') and morADH-r (5'-GAT-ATCTGCATTTGAATGCG-3'). Subsequently, sago palm *Adh*-specific primers were designed (msadh-f: 5'-CTAGAG-CTTCAGGGGCATCA-3'; msadh-r: 5'-TCAAACCTCTTG-GGGTTCAC-3') and used to isolate the cDNA via RACE. For the 5'RACE, another primer was designed from *Adh* sequences of *Elaeis guineensis* (GenBank Accession No. ACF06607) (5'-ATGGCAAGCACTGTTGGTCA-3'), denoted as egadh-f.

**2.4. RACE Method.** RACE was performed as described by Frohman et al. [19] with minor modifications. The 3' RACE was conducted in a final volume of 25  $\mu\text{L}$  containing 3  $\mu\text{L}$  cDNA template, 1X PCR buffer (Fermentas), 1.5 mM  $\text{MgCl}_2$  (Fermentas), 0.2 mM dNTPs (Fermentas), 0.4  $\mu\text{M}$  msadh-f (1st BASE), 0.4  $\mu\text{M}$  adaptor(dt)<sub>17</sub> primer (1st BASE), 1.25 U *Taq* polymerase (Fermentas), and 15.25  $\mu\text{L}$  nuclease-free water. PCR program parameters were:  $94^{\circ}\text{C}$  for 2 min, 35 cycles of  $94^{\circ}\text{C}$  for 45 sec,  $65^{\circ}\text{C}$  for 45 sec,  $72^{\circ}\text{C}$  for 1 min, and  $72^{\circ}\text{C}$  for 5 min. A 5'RACE was performed using egadh-f primer (5'ATGGCAAGCACTGTTGGTCA-3') and PCR amplified with msadh-r. It was conducted in a final volume of 25  $\mu\text{L}$  containing 3  $\mu\text{L}$  cDNA template, 1X PCR buffer (Fermentas), 1.5 mM  $\text{MgCl}_2$  (Fermentas), 0.2 mM dNTPs (Fermentas), 0.4  $\mu\text{M}$  RACE2 (1st BASE), 0.4  $\mu\text{M}$  GSP3 (1st BASE), and 1.25 U *Taq* polymerase (Fermentas). PCR conditions were  $94^{\circ}\text{C}$  for 2 min, 35 cycles of  $94^{\circ}\text{C}$  for 30 sec,  $60^{\circ}\text{C}$  for 30 sec,  $72^{\circ}\text{C}$  for 45 sec, and  $72^{\circ}\text{C}$  for 5 min.

**2.5. Full-Length *Adh* Amplification.** Full length *Adh* cDNA amplification was carried out by mixing 2.5  $\mu\text{L}$  10X HotStart buffer (Fermentas), 1.5  $\mu\text{L}$  25 mM  $\text{MgCl}_2$  (Fermentas), 0.5  $\mu\text{L}$

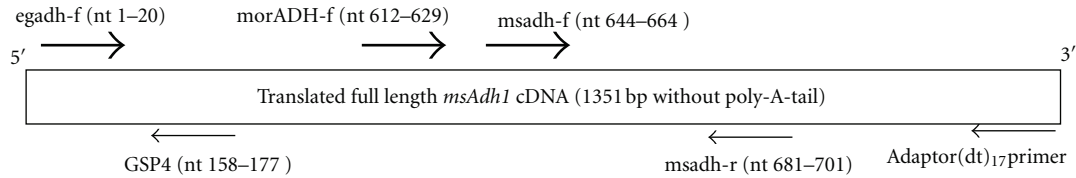


FIGURE 1: The orientation and position of primers used to isolate full length Adh cDNA.

TABLE 1: Nucleotide sequence of primers used in genomic walking. Overhanging primers (OHP) with *KpnI*, *PstI*, and *SacI* overhang sequence indicated in italic bold adaptor primer (AP), nested primer (NP), and *Adh* gene specific primer (GSP4).

Primer name	Sequence
OHP <i>KpnI</i>	5'-GAATTCGAGCTCGCCCGGGATCCTCTAGAGTAC-3'
OHP <i>PstI</i>	5'-GAATTCGAGCTCGCCCGGGATCCTCTAGATGCA-3'
OHP <i>SacI</i>	5'-GAATTCGAGCTCGCCCGGGATCCTCTAGAAGCT-3'
AP	5'-GAATTCGAGCTCGCCCGGGAT-3'
NP	5'-GCTCGCCCGGGATCCTCTAGA-3'
GSP4	5'-GCCCTTAGCTTCCCAGAAGT-3'

10 mM dNTPs (Fermentas), 1.0  $\mu$ L each for primer 10  $\mu$ M egadh-f (1st BASE) and 10  $\mu$ M adaptor(dt)<sub>17</sub> (1st BASE), respectively, 1.25 U HotStart *Taq* DNA polymerase (Fermentas), 5.0  $\mu$ L first strand cDNA, and 13.25  $\mu$ L deionized water. The mixture was run under this condition: 95°C 4 min for 1 cycle, 95°C 45 sec, 60°C 1 min, 72°C 1.5 min for 35 cycles, and final extension for 5 min at 72°C. Figure 1 shows the RACE strategy employed to isolate the full length Adh cDNA.

**2.6. Genomic Walking.** Genomic walking was carried out using a modified DNA walking method described by Ashoub and Abdalla [20]. Approximately 1.65  $\mu$ g of DNA was digested with 10 U of *KpnI*, *PstI*, and *SacI* individually in a final volume of 50  $\mu$ L at 37°C overnight. The *KpnI* and *SacI* enzymes were heat-inactivated at 65°C while *PstI* was heat-inactivated at 80°C for 20 minutes. Ten microlitres of the digested DNA was mixed with 3 U T4 DNA ligase (Fermentas), 1X ligation buffer (Fermentas), and 10 pmol of corresponding overhanging primers (Table 1) in a final volume of 20  $\mu$ L. The reaction was incubated at 16°C for 2 hours and then 4°C overnight.

The first round PCR was carried out in 25  $\mu$ L of reaction mixture containing 1X PCR buffer, 2 mM MgCl<sub>2</sub>, 200  $\mu$ M of each dNTPs, 10 pmol of each primer AP and GSP4, 1.25 U *Taq* polymerase (Fermentas), and 1  $\mu$ L of adaptor-ligated genomic DNA. The GSP4 primer was designed from the sequence derived from RACE and located 153 bp downstream from the *msAdh1* start codon. Thermal cycling condition was set as pre-denaturing at 94°C for 3 min, denaturing at 94°C for 45 sec, annealing at 54–58°C for 45 sec, extension at 72°C for 1 min, repeat for 34 cycles, and a final extension at 72°C for 5 min. For nested PCR, 1  $\mu$ L of PCR product from the first round of PCR was used as the template and mixed with 1 X PCR buffer, 2 mM MgCl<sub>2</sub>, 0.2 mM dNTP, 0.4  $\mu$ M NP, and 0.4  $\mu$ M GSP4. The reaction mix was made up to 25  $\mu$ L with nuclease-free water. The PCR program was the same as indicated above. PCR products

were analyzed on a 1.5% agarose gel and fragments detected were purified using GF-1 Gel DNA Recovery kit (Vivantis). Subsequently, the purified PCR products were cloned into pGEM-T easy vector and positive clones were sequenced.

**2.7. DNA Sequencing and Data Analysis.** All PCR products were cloned into pGEM-T vector and sequenced. Several softwares were used to analyse the sequences such as the GenScan software [18] that was used to predict the protein coding region and intron region. The promoter region and transcription start site were predicted using the program Promoter Prediction [17]. Meanwhile, the amino acid sequences were predicted using an ORF detection program available from National Center for Biotechnology Information (NCBI). Alignment of protein sequences between ADH protein of sago palm with Adh from animal, bacteria, and other plants species derived from the GenBank database was done using European Bioinformatics Institute's (EBI) ClustalW multiple alignment software. The phylogenetic tree was produced to analyse sequence divergence using the Lasergene MegAlign program (DNASTAR Sequence Analysis Package, version 7.1.0) with the Clustal alignment algorithm.

### 3. Results and Discussion

**3.1. RACE.** RNA isolated from leaf samples were analysed on 1.0% agarose gel-stained with ethidium bromide. Two distinct RNA fragments, the 28S and 18S rRNA, were clearly observed (data not shown) which showed no apparent RNA degradation. The total RNA extract were then treated with RQ1 RNase-Free DNase (Promega) to remove trace amount of genomics DNA from the total RNA. The mRNA was converted to cDNA using adaptor(dt)<sub>17</sub>primer and subsequently used in the RACE. Initially several PCRs were conducted using primers designed from *W. robusta*. These PCR produced several *Adh*-specific amplicons (data not shown) which were isolated and sequenced. From these

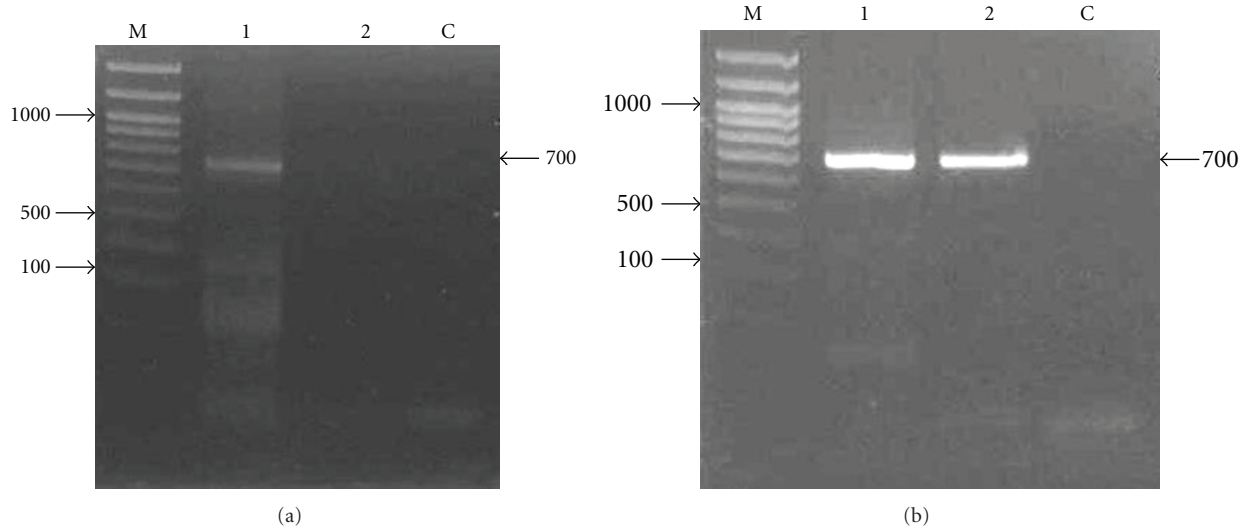


FIGURE 2: Agarose gel electrophoresis of RACE PCR products. (a) Gel electrophoresis picture of 3'RACE PCR product of young leaves on a 1.5% agarose gel. Lane 1 is the 3'RACE product and lane 2 is the PCR product with no RT template. (b) Gel electrophoresis picture of PCR product using primer egadh-f and msadh-r on 1.5% agarose gel. Lanes 1 and 2 are the 5'RACE-PCR. Lane C is the control and lane M contained the Forever 100 bp DNA ladder (Seegene). Visible fragment showing the 700 bp fragment.

results, the primers used in RACE were designed and used in the 3'- and 5'RACE. The 3'RACE was conducted using msadh-f and adaptor(dt)<sub>17</sub> and at the annealing temperature of 65°C. A fragment of approximately 700 bp was detected, isolated, and sequenced (Figure 2(a)).

Meanwhile, the forward primer used for the 5'RACE (egadh-f) was designed based on the *Adh* sequence of oil palm that have showed a high-sequence similarity with the 3'RACE sequence (data not shown). Thus, the 5'RACE was performed using egadh-f primer along with sago palm *Adh*-specific primer. The amplification produced a fragment of approximately 700 bp fragment (Figure 2(b)) that was then isolated and sequenced. Subsequently the assembly of the 5'- and 3'RACE contigs showed overlapping region (58 bp) to produce a complete *Adh* cDNA. A complete *Adh* cDNA was then amplified using the primers egadh-f and adaptor(dt)<sub>17</sub>. Approximately 1.3 kb fragment was obtained and sequenced that confirm the nucleotide sequence from the two assembled contigs. Sequence search via BLAST against the nucleotide database in Genebank NCBI showed a high degree of similarity with *Adh1* from different plant species.

**3.2. Genomic Walking.** The genomic walking was performed with two aims; first was to verify the translation start site sequence of sago palm *msAdh1* cDNA since the egadh-f primer sequence used in 5'RACE was designed from oil palm. The second aim was to isolate the regulatory sequences of *msAdh1*. The nested PCR of genomic walking produced two fragments with the sizes of 400 bp (YKpn400) and 500 bp (YKpn500) when restriction enzyme of *KpnI* was used. For the *PstI*-restricted genomic DNA, the nested PCR produced three fragments; two clear fragments at approximately 700 bp (YPst700) and 300 bp (YPst300) and

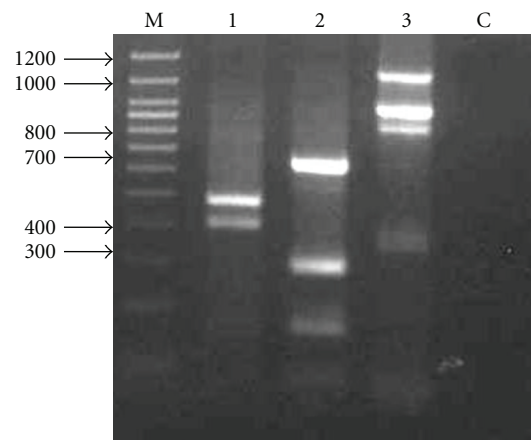


FIGURE 3: Agarose gel electrophoresis of genomic walking nested PCR products of young leaves. The restriction enzymes used for DNA digestion were indicated above the gel photo. Lanes 1, 2, and 3 contained the PCR products using *KpnI*, *PstI*, and *SacI*, respectively. Lane M contained the 1 Kb ladder (Fermentas) and C is negative control.

one faint fragment at approximately 200 bp. Lastly, genomic walking using *SacI*-restricted genomic samples produced three clear fragments at around 1.2 kb (YSac1200), 1 kb (YSac1000), and 800 bp (YSac800) and one faint fragment at around 350 bp (YSac300) (Figure 3).

The fragments produced from nested PCR were excised and sequenced. Sequencing results obtained were analysed using BLAST against NCBI genebank nonredundant nucleotide collection database. BLASTX search showed that the three sequences (YPst300, YSac1200, and YKpn400) matched with *Adh* gene. The 1.2 kb fragment (YSac1200) was selected for further analysis of the protein coding sequences



Transcription start		
1	TCTCCAACCTTGTGCACAAGGGGATTAGAAGGCATTCCACTGTGATTTCTATAGGGTCATT	60
	Translation start	
61	CGAAGAAGATCGATAGAAGCAGAGGAGTATTGCAAGGTCTAGTATTGGAGAAGATGTTGA	120
	M L S	
121	GCACAGCTGGTCGAGTGATCAAATGCAGAGCGGGCTCATGGGAGGCCGGGAAGCCGC	180
	T A G R V I K C R A A V S W E A G K P L	
181	TGGTGTGGAGGAGGTCGAGGTTGCCGCCCGCAGGCGATGGAGGTTCCGATGAAGATCC	240
	V M E E V E V A P P Q A M E V R M K I L	
241	TTTATACTTCCCTCTGCCACACTGATGTCTACTTCTGGGAAGCTAAGGGCCAGACTCCTG	300
	Y T S <b>L C</b> H T D V Y F W E A K G Q T P V	
301	TCTTTCCTCGGATCTTGGCCATGAAGCTGGAGGATTATAGAGAGTGTGGGGAGGGTG	360
	F P R I F G <b>H</b> E A G G I I E S V G E G V	
361	TGACTGAACTTGACCAGGAGACCATGTCTCCCTATATTCCTGGAGAATGCAAAGAGT	420
	T E L A P G D H V L P I <b>F T G E C K E C</b>	
421	GTGCTCACTGTAAGTCCGAGGAGAGCAACATGTGTGATCTCCCTCAGGATAAACACGGATC	480
	A H <b>C K S E H S N M C D L</b> L R I N T D R	
481	GGGAGTGATGATCAATGATGGGAAATCGAGGTTCACTATCAATGGAAAGCCATTACC	540
	G V M I N D G K S R F T I N G K P I Y H	
541	ATTTCCTAGGAACATCCACTTTCAGCGAGTACACCGTTGTCCATGTTGGCTGTGTTGCCA	600
	F L G T S T F S E Y T V V H V G C V A K	
601	AGATCAACCCCTTGGCTCCCTTGATAAAGTTGTGTTCTTAGCTGTGGCATTTCACAG	660
	I N P L A P L D K V C V L S C G I S T G	
661	GATTTGGTGGCAGTGTAAATGTTGCAAAACCACAAAGGGATCGACGGTGGCTGTTTTTG	720
	F G A T V N V A K P P K G S T V A V F G	
721	GCTTGGGAGCTGTAGGCCCTTGCTGCTGCAGAAGGTGCTAGAGCTTCAGGGGCATCAAGAA	780
	L G A V G L A A A E G A R A S G A S R I	
781	TCATTTGGTGTGATGTGAACCCCAAGAGGTTTGAGGAAGCAATGAAGTTCGGTTCACGG	840
	I G V <b>D</b> V N P K R F E E A M K F G C T E	
841	AGTTTGTGAATCCAATGGACCATGACAAGCCAGTCCAAGAGGTGATTGCTGAGATGACAA	900
	F V N P M D H D K P V Q E V I A E M T N	
901	ATGGTGGAGTTGATCGAAGCGTTGAATGCACCTGGCAACATAAATGCCATGATATCTGCAT	960
	G G V D R S V E C T G N I N A M I S A F	
961	TCGAATGTGTTGATGATGGGCTGGGGTGTGCTGTACTGGTTGGGGTGCCTCACAAGAAG	1020
	E C V H D G W G V A V L V G V P H K E A	
1021	CTGAGTTCAAAACCCACCCCTATGAACCTCCTTAACGAAAGAACTCTTAAAGGAACCTTCT	1080
	E F K T H P M N F L N E R T L K G T F F	
1081	TTGGGAATAAACCACCGCTCTGACATTCCTGCAAGTGTGTTGAGAAGTACATGAACAGG	1140
	G N Y K P R S D I P A V V E K Y M N K E	
1141	AGCTAGAATTGGAGAAGTTTATCACCCACAGTGTGCCTTTCTCTGAGATCAACAAGACCT	1200
	L E L I T H S V P F S E I N K T F D Y M	
1221	TTGACTACATGCTCAAGGGGAGAGCCTTAGGTGCATCATTACATGGATGGTTAGAGCA	1260
	E K F L K G E S L R C I I H M D G *	
1261	GGCTTACACCTGTCTGAGGCCACTCTGAAGAGAAAGAACTCATTACAAGTTAAAGCATG	1320
1321	GCCTTTGTCATGTCCATCTTAGCCTTCCCTTGCCTTATGAATAAAAAAAGGCTGCTGAG	1380
1381	TTTACTGTGAGATAGATGGGTCCAGTCCCTTATCCTTCTGCTAGTTGAAAAATAAATG	1440
1441	<u>TGTTTAAATAATCCCTTCTCTGTT</u>	1464

FIGURE 5: Nucleotide *msAdh1* sequence of sago palm leaf cDNA and deduced amino acid sequence. Underlined regions represent putative polyadenylation sites (AATAAA) and terminal transcription factor (TGTGTTTA). The conserved amino acids of Adh are indicated in bold.

sequence similarity of was found between the 5' end in both genomic DNA and cDNA of *msAdh1*.

Promoter region and transcription start site were also predicted from the isolated sequence using the program Promoter Prediction [17]. From the predicted transcription start site, a putative TATA box with the sequence CTATAAAAA was found at the positions -31 bp and -23 bp from transcription start site. The sequence and location closely corresponds to the plant TATA box consensus sequence (C/G)TATA(T/A)(A1-3)(C/T)A [25, 26]. Furthermore, Breathnach and Chambon [21] and Lin et al. [25] suggested that the distance between putative TATA box and transcription start site in most genes are between 25-30 bp or  $32 \pm 7$  bp, respectively. The CA dinucleotide that is a typical initiation site for eukaryotes class II genes usually found at region -1 to +1 [23, 27] was also found in this

study, however, the CA-dinucleotide was located -6 from transcription start site. In addition, putative CAT and AGGA box [26, 28] that may play a role in promoter efficiency were also found near the TATA box at position -98 bp to -95 bp and -162 bp to -159 bp, respectively, (Figure 4). All these elements found in the promoter region are necessary for accurate initiation of basal transcription.

**3.3. Analysis of the Predicted Protein Sequence of Sago Palm *Adh1*.** The *msAdh1* cDNA sequence was determined to be 1464 bp in length with 1143 bp open reading frame encoding the Adh protein, 113 bp of 5' untranslated region, and a 208 bp 3' untranslated region not including the poly(A) tail (Figure 5). Further *in silico* sequence analysis showed that the nucleotide encodes for 380 amino acids. Based on the cDNA

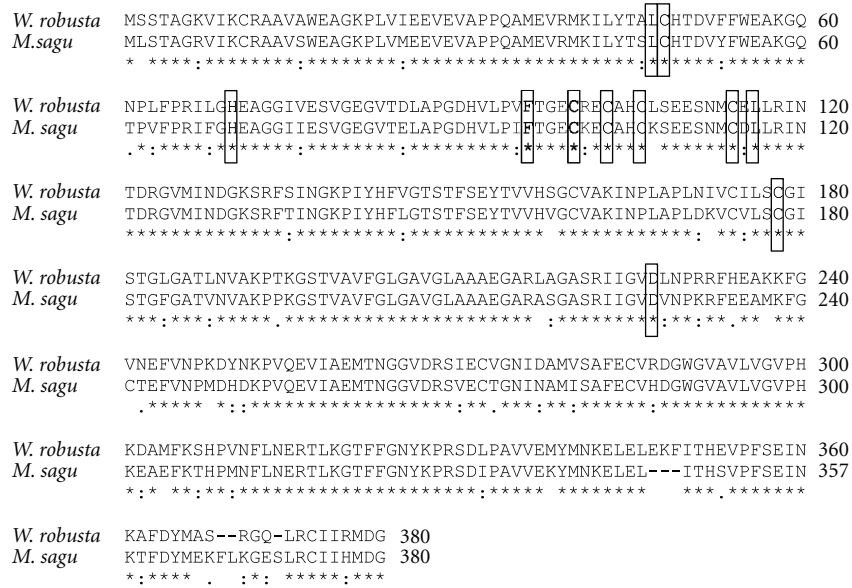


FIGURE 6: Alignment of alcohol dehydrogenase amino acid sequence between *W. robusta* and *M. sagu*. The boxes showed the conserved regions (Cys-48/His-70/Cys-178, Cys-100/Cys-103/Cys-106/Cys-114, Asp-227, Phe-93, Leu-47, and Leu-116) in the amino acid sequences. Dash is included in the sequences to maximize the homology.

TABLE 2: Comparison between msAdh1 with other species Adh deduced amino acid sequences.

Species	Protein	Length of deduced amino acid	Amino acid identity (%)	Accession number
<i>Elaeis guineensis</i>	Adh	380	91	ACF06607
<i>Washingtonia robusta</i>	AdhB	380	85	AAB39597
<i>Zea mays</i>	Adh1	379	87	CAA27681
<i>Oryza sativa</i>	Adh1	379	87	Q75ZX4
<i>Arabidopsis thaliana</i>	Adh	379	82	AAL90991
<i>Drosophila melanogaster</i>	Adh	379	49	AAB02520
<i>Mus musculus</i>	Adh1	375	48	NP_031435
<i>Homo sapiens</i>	Adh1	375	49	AAC41757
<i>Equus caballus</i>	Adh	374	51	1MG0_A
<i>Saccharomyces cerevisiae</i>	Adh1	347	18	2HCY_A
<i>Bacillus cereus</i>	Adh1	345	15	ZP_04323237

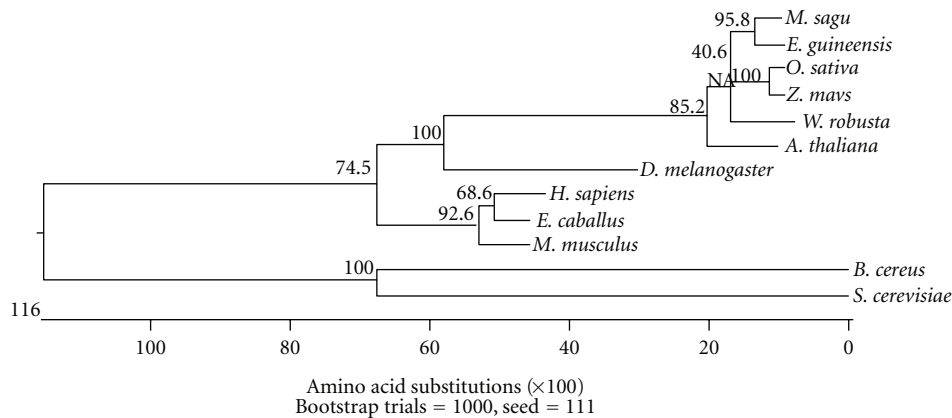


FIGURE 7: Phylogenetic tree of msAdh1 of sago palm and other species Adh. Tree was constructed by using clustal method of Lasergene Megalign (DNASTAR, Inc., Madison, WI) based on amino acid similarities. Numbers represent bootstrap values. Scale bar indicates levels of sequence divergence.

sequence data, the predicted molecular mass of msAdh1 was 41.8 kDa. At the 3' terminus of the msAdh1 cDNA sequence (Figure 5), two polyadenylation signals (AATAAA) conserved in plants [29] were located at positions 1218 bp and 1319 bp, respectively. The sequence TGTGTTTA that is homologous to terminal transcription factor consensus [30] was also found between positions 1326 to 1333.

The msAdh1 sequence was further analysed for motifs and factors specific for Adh regulation. Conserved amino acids of two zinc-binding domains found in oil palm and *W. robusta* located at Cys-48/His-70/Cys-178 and Cys-100/Cys-103/Cys-106/Cys-114 [12], respectively, were also present and conserved in sago palm msAdh1 (Figure 6). Other motifs include the Asp-227 that binds to the adenosine ribose of the coenzyme and the amino acids Phe-93, Leu-57, and Leu-116 that bind to alcohol substrate [31], were also found with exception that Leu-57 was found at location 47 for both oil palm and *W. robusta* (Figure 6).

**3.4. Similarity Analysis of Sago Palm msAdh1 with Other Species.** Sequence homology search of deduced msAdh1 amino acid sequence using NCBI BLASTX showed 91% identity with the oil palm (*E. guineensis*) deduced Adh amino acid sequence. Comparisons were also conducted with amino acid sequences of other selected organisms (Table 2). The msAdh1 of sago palm showed highest identity (91%) with oil palm Adh. This was followed by 87% identity to rice and maize Adh1, 85% identity to *W. robusta* AdhB, and 82% identity to Adh from *Arabidopsis*.

The percentage of divergence was calculated by comparing sequence pairs in relation to the phylogeny reconstructed by MEGALIGN. The percent divergences of *M. sagu* from *E. guineensis* and *W. robusta* were determined to be 8.7% and 15.8%, respectively. In contrast, both microorganisms (yeast and *Bacillus cereus*) had higher percentages of divergence from sago palm. From the phylogenetic tree (Figure 7), animals, plants, and microorganism are classified into three distinct groups. In plants, monocots and dicots Adhs form two distinct groups. The Adh1 amino acid sequence between animals and plants is closely related compared to microorganisms. As expected, the sago palm msAdh1 is similar to other plant forms and in particular, closer to the monocot compared to dicots. Furthermore, the result suggested that sago palm msAdh1 is more likely related to oil palm than to rice and maize, and shares a common Adh ancestor.

## 4. Conclusion

Adh cDNA from sago palm have been successfully isolated using a combination of RACE and genomic walking method. Analysis of nucleotide sequence and predicted amino acid indicated that the sago palm cDNA is Adh1. The full length of msAdh1 cDNA was determined to be 1464 bp containing the 5' and 3' untranslated regions and a deduced amino acid of 380. The regulatory region of basal msAdh1 was isolated and found to contain the promoter sequences and conserved

motifs corresponding to Adh regulation such as the two zinc-binding domains, binding domains to adenosine ribose of the coenzyme, and alcohol substrate.

## References

- [1] B. R. Benz, J. M. Rhode, and M. B. Cruzan, "Aerenchyma development and elevated alcohol dehydrogenase activity as alternative responses to hypoxic soils in the *Piriqueta caroliniana* complex," *American Journal of Botany*, vol. 94, no. 4, pp. 542–550, 2007.
- [2] T. Fukao, R. A. Kennedy, Y. Yamasue, and M. E. Rumpho, "Genetic and biochemical analysis of anaerobically-induced enzymes during seed germination of *Echinochloa crus-galli* varieties tolerant and intolerant of anoxia," *Journal of Experimental Botany*, vol. 54, no. 386, pp. 1421–1429, 2003.
- [3] P. K. Macnicol and J. V. Jacobsen, "Regulation of alcohol dehydrogenase gene expression in barley aleurone by gibberellin and abscisic acid," *Physiologia Plantarum*, vol. 111, no. 4, pp. 533–539, 2001.
- [4] C. Chang and E. M. Meyerowitz, "Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 5, pp. 1408–1412, 1986.
- [5] E. S. Dennis, W. L. Gerlach, A. J. Pryor et al., "Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize," *Nucleic Acids Research*, vol. 12, no. 9, pp. 3983–4000, 1984.
- [6] H. A. Roslan, Y. Sundaraj, and A. A. S. A. Husaini, "Multiple forms of alcohol dehydrogenase (*Adh*) genes in sago palm: a preliminary study," *BANWA*, vol. 5, no. 1, 2008.
- [7] P. Padmanabhan and S. V. Sahi, "Suppression subtractive hybridization reveals differential gene expression in sunflower grown in high P," *Plant Physiology and Biochemistry*, vol. 49, no. 6, pp. 584–591, 2011.
- [8] H. J. Chung and R. J. Ferl, "Arabidopsis alcohol dehydrogenase expression in both shoots and roots is conditioned by root growth environment," *Plant Physiology*, vol. 121, no. 2, pp. 429–436, 1999.
- [9] H. Kato-Noguchi, "Osmotic stress increases alcohol dehydrogenase activity in maize seedlings," *Biologia Plantarum*, vol. 43, no. 4, pp. 621–623, 2000.
- [10] J. Bailey-Serres and R. Chang, "Sensing and signalling in response to oxygen deprivation in plants and other organisms," *Annals of Botany*, vol. 96, no. 4, pp. 507–518, 2005.
- [11] M. Trick, E. S. Dennis, K. J. R. Edwards, and W. J. Peacock, "Molecular analysis of the alcohol dehydrogenase gene family of barley," *Plant Molecular Biology*, vol. 11, no. 2, pp. 147–160, 1988.
- [12] B. R. Morton, B. S. Gaut, and M. T. Clegg, "Evolution of alcohol dehydrogenase genes in the palm and grass families," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 21, pp. 11735–11739, 1996.
- [13] R. Ellen, "Processing *Metroxylon sagu* Rottboell (Arecaceae) as a technological complex: a case study from south central Seram, Indonesia," *Economic Botany*, vol. 58, no. 4, pp. 601–625, 2004.
- [14] K. H. Wan, C. Yu, R. A. George et al., "High-throughput plasmid cDNA library screening," *Nature Protocols*, vol. 1, no. 2, pp. 624–632, 2006.
- [15] C. C. Wee and H. A. Roslan, "Expressed sequence tags (ESTs) from young leaves of *Metroxylon sagu*," *3Biotech*. In press.
- [16] J. J. Doyle and J. L. Doyle, "Isolation of plant DNA from fresh tissue," *Focus*, vol. 12, pp. 13–15, 1990.



- [17] S. E. Celniker, D. A. Wheeler, B. Kronmiller et al., "Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence," *Genome biology*, vol. 3, no. 12 article RESEARCH 0079, 2002.
- [18] C. Burge and S. Karlin, "Prediction of complete gene structures in human genomic DNA," *Journal of Molecular Biology*, vol. 268, no. 1, pp. 78–94, 1997.
- [19] M. A. Frohman, M. K. Dush, and G. R. Martin, "Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 23, pp. 8998–9002, 1988.
- [20] A. Ashoub and K. S. Abdalla, "A primer-based approach to genome walking," *Plant Molecular Biology Reporter*, vol. 24, no. 2, pp. 237–243, 2006.
- [21] R. Breathnach and P. Chambon, "Organization and expression of eucaryotic split genes coding for proteins," *Annual Review of Biochemistry*, vol. 50, pp. 349–383, 1981.
- [22] M. Burset, I. A. Seledtsov, and V. V. Solovyev, "SpliceDB: database of canonical and non-canonical mammalian splice sites," *Nucleic Acids Research*, vol. 29, no. 1, pp. 255–259, 2001.
- [23] B. A. Hanley and M. A. Schuler, "Plant intron sequences: evidence for distinct groups of introns," *Nucleic Acids Research*, vol. 16, no. 14, pp. 7159–7176, 1988.
- [24] J. W. Brown, C. G. Simpson, G. Thow et al., "Splicing signals and factors in plant intron removal," *Biochemical Society Transactions*, vol. 30, no. 2, pp. 146–149, 2002.
- [25] J. Lin, X. Zhou, Y. Pang et al., "Cloning and characterization of an agglutinin gene from *Arisaema lobatum*," *Bioscience Reports*, vol. 25, no. 5-6, pp. 345–362, 2005.
- [26] J. Messing, D. Geraghty, G. Herdecker, H. Nien-Tai, J. Kridl, and I. Rubenstein, "Plant gene structure," in *Genetic Engineering in Plants*, T. Kosuge, O. Merridith, and A. Hollaender, Eds., pp. 211–227, Plenum Press, New York, NY, USA, 1983.
- [27] H. K. Bayele, "Trypanosoma brucei: a putative RNA polymerase II promoter," *Experimental Parasitology*, vol. 123, no. 4, pp. 313–318, 2009.
- [28] W. Qin and V. K. Walker, "Tenebrio molitor antifreeze protein gene identification and regulation," *Gene*, vol. 367, no. 1-2, pp. 142–149, 2006.
- [29] G. Ji, J. Zheng, Y. Shen et al., "Predictive modeling of plant messenger RNA polyadenylation sites," *BMC Bioinformatics*, vol. 8, article no. 43, 2007.
- [30] P. S. Manchado, C. Verriès, and C. Tesnière, "Molecular characterization and structural analysis of one alcohol dehydrogenase gene (GV-Adh 1) expressed during ripening of grapevine (*Vitis vinifera* L.) berry," *Plant Science*, vol. 125, no. 2, pp. 177–187, 1997.
- [31] H. Jörnvall, B. Persson, and J. Jeffery, "Characteristics of alcohol/polyol dehydrogenases. The zinc-containing long-chain alcohol dehydrogenases," *European Journal of Biochemistry*, vol. 167, no. 2, pp. 195–201, 1987.