

Full Length Research Paper

***In silico* characterization and evolution studies of alcohol dehydrogenase gene from *Phoenix dactylifera* L.cv Deglet Nour**

Imen Rezik^{1*}, Amine Elleuch¹, Nouredine Drira¹ and Foued Cheour²

¹Laboratory of plant biotechnology, Faculty of sciences of Sfax, University of Sfax, Tunisia.

²High Institute of Applied Biology of Medenine, Tunisia.

Received 21 March, 2015; Accepted 15 May, 2015

The aim of our study was to isolate the alcohol dehydrogenase (ADH) mRNA from *Phoenix dactylifera*, and examine the molecular evolutionary history of this nuclear gene with others ADH genes from palms and other plants species. The *DnADH* gene has been isolated *in silico* by BLAST2GO from a cDNA library of date palm cv *Deglet Nour*. The prediction of candidate's mRNA and protein for *ADH* gene from *khalas* were performed *in silico* from whole genome shotgun sequence (ACYX02009373.1) using FGENESH prediction program. Nucleotide polymorphism using DnaSPv5 was examined in four palm ADH mRNA sequences across the entire 1.098 kb length of ADH mRNA. A primary conclusion of the present study is that nucleotide diversity for ADH between palm species is very low. In order to assess selective pressure, we calculated the ratio of non-synonymous to synonymous substitutions. We conclude that ADH palms genes appear to be under very different selective constraints. Phylogenetic analyses using PHYLIP and Notung 2.8 programs suggest that ADH genes of some plants species resulted from relatively ancient duplication events. In this study, we present for the first time a molecular characterization of ADH protein of *P. dactylifera* L cv Deglet nour and a phylogeny analysis between plants ADH.

Keys word: Alcohol dehydrogenase, palms species, evolution, duplication.

INTRODUCTION

Sequencing of date palm genome and cDNA or expressed sequence tags (EST) using Next generation sequencing provides a rapid method for gene discovery and can be used to identify transcripts associated with specific biological processes (Al-Mssallem et al., 2013).

The alcohol dehydrogenase (ADH) genes encode a glycolytic enzyme and have been characterized at the molecular level in a wide range of flowering plants (Clegg et al., 1997; Miyashita et al., 2001), California fan Palm (*Washingtonia robusta*) (Morton et al., 1996) and Oil palm

*Corresponding author. E-mail: imenbmc@yahoo.fr. Tel: 0021674676616. Fax: 0021674274437.

Author(s) agree that this article remains permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Abbreviations: EST, Expressed sequence tags; ADH, alcohol dehydrogenase; CDD, conserved domains database; ML, maximum-likelihood; SNP, single nucleotide polymorphisms.

(*Elaeis guineensis*) (Low et al., 2008; Singh et al., 2013). The ADH enzyme is essential for anaerobic metabolism (Dolferus et al., 1994; Dolferus et al., 1997), found in plants and prokaryotes, its activity is required for lignin biosynthesis and apparently also for defense-related functions (Umezawa, 2010). The ADH genes in *Arabidopsis thaliana* (Innan et al., 1996), *Arabidopsis gemmifera* (Miyashita et al., 2013) and *Leavenworthia* (Charlesworth et al., 1998) in Brassicaceae, cottons (Small et al., 2000), and grasses (Gaut et al., 1996; Gaut et al., 1999) have been subjected to molecular evolutionary studies. However, the broader evolutionary histories of the ADH genes in the angiosperms remain unclear since few studies have investigated the evolution of the ADH genes in a wide range of angiosperms (Strommer, 2011).

Recently, Small and Wendel (Small et al., 2000) suggested that some ADH gene duplications may have predated the origin of each of the flowering plant families. However, the details of the gene duplications and deletions experienced by the ADH genes of most groups of the angiosperms remain unclear. Additional studies are needed to understand the evolutionary history of the ADH genes in various plant groups. A number of studies have been conducted on the evolutionary dynamics of plant gene families, including the gene families coding for the R and MADS-box regulatory proteins (Purugganan et al., 1995), the small heat-shock proteins (Waters et al., 1995), chalcone synthase (Durbin et al., 1995), and the chlorophyll *a/b* binding proteins (Demmin et al., 1989). Most of these gene families consist of numerous loci and have a great deal of variation in copy number between species. The evolutionary picture emerging for these gene families is one of dynamic fluctuations of copy number through multiple gene duplication/deletion events.

The glycolytic proteins in plants are coded by small multigene families, which provide an interesting contrast to the high copy number gene families studied to date. Isozyme surveys covering an array of dicot and monocot species have revealed that most glycolytic enzymes have two forms in all species (Gottlieb et al., 1982), probably reflecting a small, and stable, number of loci. The apparent stability of these gene families raises important questions regarding evolutionary dynamics. One issue is whether any given gene family emerged once by duplication and then differentiated, as suggested by Gottlieb (Gottlieb et al., 1982). An alternative view posits a continuous, albeit slow, flux of gene duplication and loss that leads to an approximate dynamic equilibrium in copy number. The narrow range of gene family size for glycolytic enzymes suggests that additional constraints may act to determine copy number for this important class of genes.

An analysis of animal and plant ADH genes indicated that the grass ADH1 and ADH2 genes diverged following the divergence of monocots and dicots (Yokoyama et al., 1993). This result provides evidence that the gene family

did not emerge from a single duplication event early in angiosperm evolution. Additionally, the isolation of a recent duplication product in barley (Trick et al., 1988), as well as the duplication of ADH I in other species, suggests that the gene family undergoes some copy number fluctuation. Here, we report the isolation of ADH mRNA from *Phoenix dactylifera* L cv Deglet Nour and the prediction of ADH from *P. dactylifera* L cv Khalas. In order to study ADH gene evolution between recalcitrant vs. orthodox palm species, we compared ADH mRNAs and proteins of Khalas, Deglet Nour, *E. guineensis* and *W. robusta* varieties. We also investigated the molecular evolutionary history of the palms species ADH genes with the others species to gain further understanding of the evolutionary dynamics of nuclear gene families.

MATERIALS AND METHODS

cDNA library normalization and *in silico* isolation of DnADH

Fresh leaf tissue from *P. dactylifera* L cv Deglet Nour were processed and flash frozen in liquid nitrogen. Tissues were immediately sent to Bio S&T Inc. (Montreal, QC, Canada) where RNA extraction, cDNA synthesis and normalization were performed. Briefly, RNA was extracted using a modified TRIzol method (Invitrogen, USA). cDNA synthesis was carried out using 18 µg total RNA by a modified SMART™ Cdna synthesis method and then were normalized by a modified normalization method where full-length cDNA was synthesized with two sets of primers for driver and tester cDNA. Single-stranded cDNA was used for hybridization instead of double-stranded cDNA. Excess amounts of sense-stranded cDNA was hybridized with antisense-stranded cDNA. After hybridization, duplex DNA was removed by hydroxyapatite chromatography. Normalized tester cDNA was re-amplified and purified with tester specific primer L4N by PCR, while driver cDNA was unable to amplify using L4N primer. Size fractionation of re-amplified cDNA was done in a 1% agarose gel. Greater than 0.5 kb cDNA fragments were purified by electroelution and after determining the concentrations, purified cDNAs were precipitated and stored in 80% EtOH at -80°C. The normalized cDNA library was prepared for sequencing and approximately 8 µg of purified cDNA was sheared into small fragments via Covaris E210 Acoustic Focusing Instrument and sequenced in three-fourths 454 plate run on a 454 GS-FLX Titanium platform (Roche). To identify DnADH cDNA, assembled contigs were analysed using Blast2GO2.8 bioinformatic (Conesa and Götz, 2008) to provide Gene Ontology, BLAST and domain/Interpro annotation. Evaluation of DnMRE11 predicted protein was done based on the identification of domains in the NCBI Conserved Domains Database (CDD), phytozome of June 2013 (<http://www.phytozome.net/>) and most recent version of HMMER (HMMERV3.0; Eddy, 2009).

mRNA and protein prediction of KhADH

Candidate's mRNA and protein for ADH gene from *P. dactylifera* L cv *khalas* (KhADH) were identified *in silico* from whole genome shotgun sequence (accession number: ACYX02009373.1) using FGENESH prediction program. Our approach to the annotation was based on applying basic gene prediction tools and using the BLASTN and BLASTP programs to improve the accuracy of gene prediction. Sequence motifs related to function were identified through PFAM (Henikoff et al., 1999), and PROSITE (Altschul et al., 1998).

Table 1. List of accession numbers of mRNA and protein for palms species used in this study.

Nom latin	Abbreviation	mRNA accession number	Protein accession number
<i>Phoenix dactyfera</i> v Deglet Nour	DnADH	KF961040	AHH32692
<i>Phoenix dactyfera</i> v khalas	KhADH		
<i>Elaeis guineensis</i>	EgADH	EU284998.1	ACF06607.1
<i>Washingtonia robusta</i>	WrADH	U65972.1	AAB39597.1

Sequences analysis

The sequences of the ADH genes used in this study were obtained from the GenBank/EMBL/DDBJ database (Table 1). Alignments of four ADH proteins sequences (Table 1) for palms species were performed with ClustalX (Thompson et al., 1997) and ESPript (Robert and Gouet, 2014). The number of segregating sites and levels of nucleotide diversity π (π), the average number of nucleotide differences per site between two sequences (Nei et al., 1987) and θ , an estimate of $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per nucleotide (Watterson et al., 1975), were computed in DnaSP (version 5.10.00). Tajima's D test and Fu and Li's D test were also performed in DnaSP for testing selections deviating from neutrality (Librado and Rozas, 2009). The divergence distance of all mRNA date palm sequence was estimated by using the Kimura two-parameter model (Kimura, 1980) employed by PHYLIP (Felsenstein, 2000) with a transition/transversion ratio of 2.0. Estimation of dN and dS values in order to get information about functional constraints on palm ADH sequences, we also estimated the number of synonymous substitutions per synonymous site (dS), and the number of non-synonymous substitutions per non-synonymous site (dN), using PAML yn00 program with default parameters (Yang, 1979) using Yang and Nielsen (2000) method. A distance matrix based on the aligned amino acid sequences was constructed by using the jones taylor thornston method of the PROTDIST program on PHYLIP.

The phylogenetic relationships between the 38 ADH proteins (Table 2) for different species were analyzed using the maximum-likelihood (ML) method. For the ML analyses, we used the PROTML program of PHYLIP version 3.6 (Felsenstein, 2000). We employed the JTT model of amino acid substitution. All indels were counted as missing. We performed ten random sequence addition searches using the J option and global branch swapping using the G option to isolate the ML tree with the best log-likelihood. In addition, we performed boot-strap analysis with 100 replications. To infer the evolutionary events affecting the ADH genes, an analysis using Notung2.8 (Chen et al., 2000) was performed. The ML tree with the highest log-likelihood was used for the gene tree. Both gene duplications and losses were considered to reconcile the gene tree with the species tree. Evidence of recombination was sought by the program RDP4 (Version 4.16) (Martin et al., 2010).

RESULTS AND DISCUSSION

Sequence analysis of isolated DnADH and predicted KhADH proteins

DnADH (ADH gene from *P. dactyfera Deglet Nour*) and KhADH (ADH gene from *P. dactyfera* L cv Khalas) encode proteins of 380 residues, with the predicted molecular weight of 41.14 and 41.18 kDa, and isoelectric points of 6.16 and 6.59, respectively. The alignment of DnADH

with sequences of different palms ADH proteins shows the presence of a large number of conserved domains (Figure 1), that are typical of this sub-family (Chase, 1999). The identity at the amino acid level between DnADH and other palm species ADHs sub-family is very high and ranges between 78 and 91%. The genetic distance between the four proteins is very low (Table 3). Many very well conserved amino acids that have been implicated in the fixation of zinc are present in DnADH: Cys, His and Cys at the 48, 70 and 178 positions (Figure 1) and four Cys at positions 100, 103, 106, 114 (Figure 1) (Eklund et al., 1976; Yokoyama and Harry, 1993). The Asp in position 237, corresponding to DnADH sequence has been described as implicated in the preference of NAD as cofactor in the dehydrogenase reaction (Eklund et al., 1976; Fan et al., 1991).

Divergence of the palms ADH loci

Pairwise distances based on the Kimura two-parameter model for mRNA sequences of four palm ADH loci are given in Table 4. Two points are apparent from this table. One is that the two KhADH and EgADH mRNA are the most similar. The second is that the ADH gene from palms is moderately diverged and must represent duplication event.

Sequence diversity of ADH gene between palm species

Nucleotide polymorphism was examined in four palm ADH mRNA sequences across the entire length of ADH palms mRNA. The examination yielded 48 single nucleotide polymorphisms (SNP) and 2 insertions or deletions (Indels) in this region (1.098 kb). Nucleotide diversity p of the entire mRNA sequence was 0.15523 and θ was 0.10594. Several statistical tests were used to test the hypothesis that ADH sequences have been evolving in accordance with expectations under neutral theory. Several statistical tests were used to test the hypothesis that ADH sequences have been evolving in accordance with expectations under neutral theory. The tests of Tajima (-0.75403) and Fu and Li (-0.34314) compare different estimates of θ ($4N_e\mu$) and p ; they made assumptions that the four ADH sequences have a

Table 2. List of accession numbers and list of taxa used in this study.

Family	Accession number
Eukaryota	
<i>Aspergillus niger</i>	AnADH : XP001395505.2
Core eudicotyledons	
<i>Dianthus caryophyllus</i>	DcADH : <u>AAP96921.1</u>
Rosids	
<i>Vitis vinifera</i>	VvADH1 : <u>NP001268079.1</u> , VvADH2 : NP001268083.1
Sapindales	
<i>Dimocarpus longan</i>	DIADH : ABF61806.1
<i>Mangifera indica</i>	MiADH1 : ADB43613.1, MiADH2 : ADB43614.1
Brassicaceae	
<i>Arabis hirsuta</i>	AhADH : AAF23543.1
<i>Leavenworthia stylosa</i>	LsADH3 : AAC79418.1
<i>Brassica rapa</i>	BrADH1 : AEC13713.1
Fabids	
<i>Alnus glutinosa</i>	AgADH : CAJ21172.1
<i>Cucumis melo</i>	CmADH1 : ABC02081.1
Papilionoideae	
<i>Lotus corniculatus</i>	LcADH1 : AAO72531.1
<i>Trifolium repens</i>	TrADH1 : P13603.1
<i>Glycine max</i>	GmADHL1 : NP001242142.1
Lamiales	
<i>Salvia miltiorrhiza</i>	SmADH : ACZ48689.1
Solanaceae	
<i>Solanum tuberosum</i>	StADH1 : AAA33806.1
<i>Nicotiana tabacum</i>	NtADH : CAA57446.1
<i>Lactuca sativa</i>	LsADH : AFP72378
Arecaceae	
Coryphoideae	
<i>Phoenix dactylifera</i>	DnADH : AHH32692
<i>Washingtonia robusta</i>	WrADH : AAB39597.1
<i>Elaeis guineensis</i>	EgADH : ACF06607.1
Poaceae	
<i>Oryza sativa</i>	OsADH1 : CAA34363.1, OsADH2 : Q4R1E8.1
Aegilops	
<i>Aegilops tauschii</i>	AtADH3 : ABL74268.1
	AtADH2 : ABL74260.1
<i>Aegilops speltoides</i>	AsADH1 : ABL74271
Triticum	
<i>Triticum turgidum</i>	TtADH1 : ABL74262
<i>Triticum aestivum</i>	TaADH1 : ABL74258, TaADH2 : ABL74254, TaADH3 : ABL74253
<i>Hordeum vulgare</i>	HvADH1 : AAK49116
Panicoideae	
<i>Cenchrus americanus</i>	CaADH : CAC37632
Andropogoneae	
<i>Miscanthus sinensis</i>	MsADH1 : ADI24330
	ZmADH1 : NP001105409
<i>Zea mays</i>	ZmADH2 : NP001105410
<i>Coix lacryma-jobi</i>	CljADH1 : ABE68381

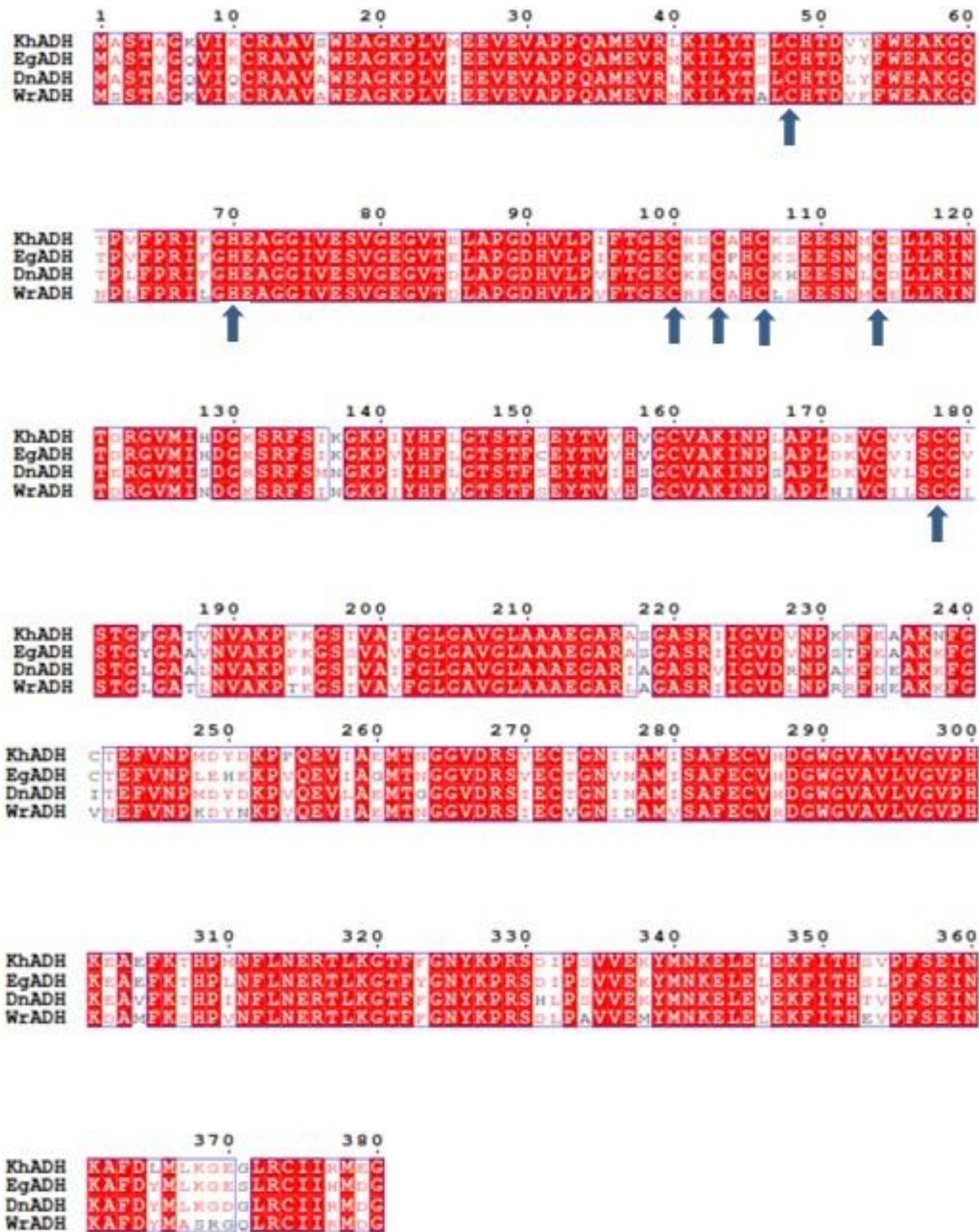


Table 3. Homology percent of amino acid sequence between full length palm ADH protein and protein distance using Jones Taylor Thornton method of PHYLIP program.

Amino acid sequence	KhADH %	EgADH	DnADH	WrADH
KhADH	100	0.090943	0.139001	0.170949
EgADH	91	100%	0.162583	0.197522
DnADH	86	84%	100%	0.168466
WrADH	84	82%	83%	100%

Table 4. Divergences based on Kimura's two-parameter model, between palms ADH mRNA.

	DnADH	KhADH	EgADH	WrADH
DnADH	0.000000			
KhADH	0.202561	0.000000		
EgADH	0.203524	0.086238	0.000000	
WrADH	0.225311	0.221938	0.232504	0.000000

Table 5. Ratio of non-synonymous to synonymous substitutions rate among mRNA palms ADH

Seq	S	N	dN±SE	dS±SE	dN/dS
KhADH-DnADH	285.7	812.3	0.0851±0.0107	0.7850±0.1109	0.202561
EgADH-DnADH	288.2	809.8	0.0969±0.0115	0.6501±0.0823	0.203524
EgADH- KhADH	284.3	813.7	0.0417±0.0073	0.2344±0.0339	0.086238
WrADH-DnADH	288.9	809.1	0.0910±0.0111	0.9349±0.1430	0.225311
WrADH-KhADH	286.2	811.8	0.0914±0.0111	0.8472±0.1205	0.226587
WrADH-EgADH	289.4	808.6	0.1052±0.0121	0.8086±0.1095	0.232504

neutral evolution (Simonsen et al., 1995; Wayne and Simonsen, 1998). None of these tests returned significant *P* values. This is not surprising, given the small number of variable positions and the relatively low statistical power of these tests (Wayne and Simonsen 1998). A primary conclusion of the present study is that nucleotide diversity for ADH between palm species is very low. Estimates reported here are lower than previously reported values not only for plant ADH sequences (Cummings and Clegg, 1998; Liu et al., 1998), but for other plant nuclear genes such as *C1* in maize, (Hanson et al., 1996); *ChiA* in *Arabidopsis* (Kawabe et al., 1997); *ChsA* in *Ipomoea* (Huttley et al., 1997) and *Pgi* in *Dioscorea* (Terachi and Miyashita, 1997). Tests for conversion among the four mRNA palms sequences using RDP v 4.16 with Maxchi program, detected two recombination's: one between DnADH and WrADH (ADH from *Washingtonia robusta*) (KhADH is the recombinant) with a *P* value of $2.97 \cdot 10^{-2}$ and length of 801 nt; one between DnADH and WrADH with EgADH (ADH from *E. guineensis*) is the recombinant (*P* value of $1.45 \cdot 10^{-2}$ and length of 793 nt).

Selection pressure

In order to assess selective pressure, we calculated the

ratio of non-synonymous to synonymous substitutions (dN/dS) among mRNA palms ADH. Estimates of dN and dS for the entire coding region between the four palms mRNA are given in Table 5. Comparisons within plant ADH genes show dN/dS < 0.3 (Table 5). For all genes, dS exceeded dN in both comparisons, as would be expected for genes under purifying selection (Nei, 1987). All comparisons with *P* values ≤ 0.001 remain significant after correcting for multiple tests. The same results were found for the others ADH evolution studies (Yokoyama et al., 1990). By comparing dN/dS ratios of ADH palms genes, we found that the ratio EgADH –KhADH has the lowest value. Therefore, we conclude that ADH genes appear to be under very different selective constraints. This result shows that dN/dS ratios are lower for duplicated genes than for unique genes (Davis et al., 2004; Jordan et al., 2004).

Phylogenetic analyses

The palm family emerged -80 million years ago and as such it represents one of the lineages that radiated early in monocot evolution (Wilson et al., 1990; Duvall et al., 1993). The comparative analysis of these four palms monocot families presents an ideal opportunity to

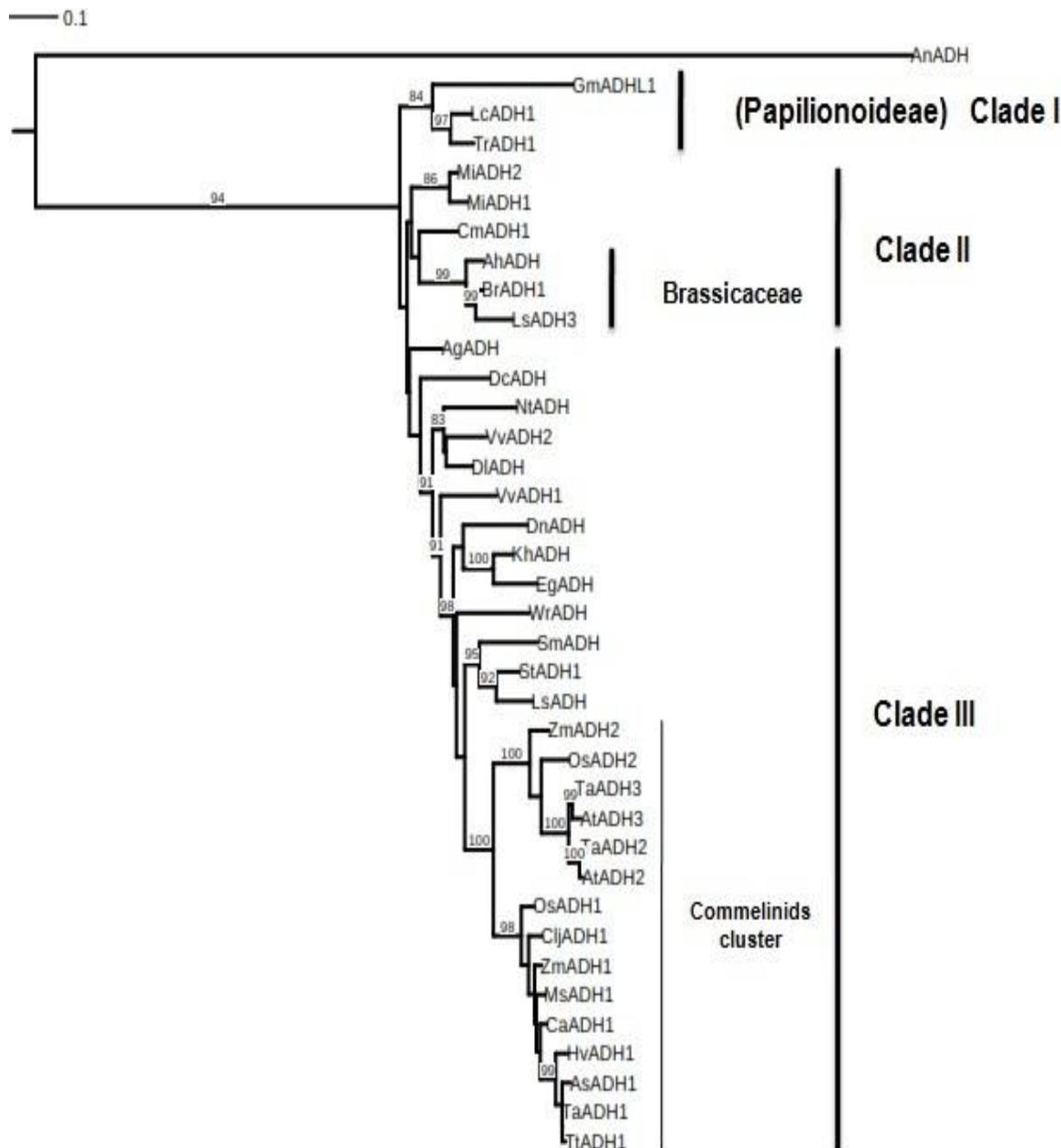


Figure 2. The phylogenetic tree based on ADH gene sequences obtained by the maximum-likelihood method. The log-likelihood of the best ML tree is -3981.05. The numbers below the branches are the bootstrap values of 50% or more support. The ADH genes from plants roughly fall into two clades that we denoted as Clade I and Clade II.

investigate the dynamics of angiosperm gene family evolution, and in particular, to expand our understanding of the evolution of the ADH gene family. A previous analysis of ADH has provided evidence against ADH genes in grasses emerging from a single duplication event early in the evolutionary history of the angiosperms (Yokoyama and Harry, 1993). We conducted phylogenetic analyses of the ADH genes using sequence from *Aspergillus niger* as outgroups. To determine the phylogenetic position of the palms ADH genes isolated and predicted in this study, we subjected their sequences

to ML analysis by employing a data set including the previously published ADH gene family sequences from various phylogenetic groups (Clegg et al., 1997; Small et al., 2000). Our resulting ADH gene tree roughly consisted of three monophyletic groups that we denoted "Clade I", "Clade II" and Clade III (Figure 2). Clade I contains only ADH genes from Papilionoideae species, while Clade II contains ADH genes from rosids species which contains the Brassicaceae species. The palms ADH proteins appeared with eudicots species in Clade III and not with the Commelinids cluster (monocots) (Figure 2).

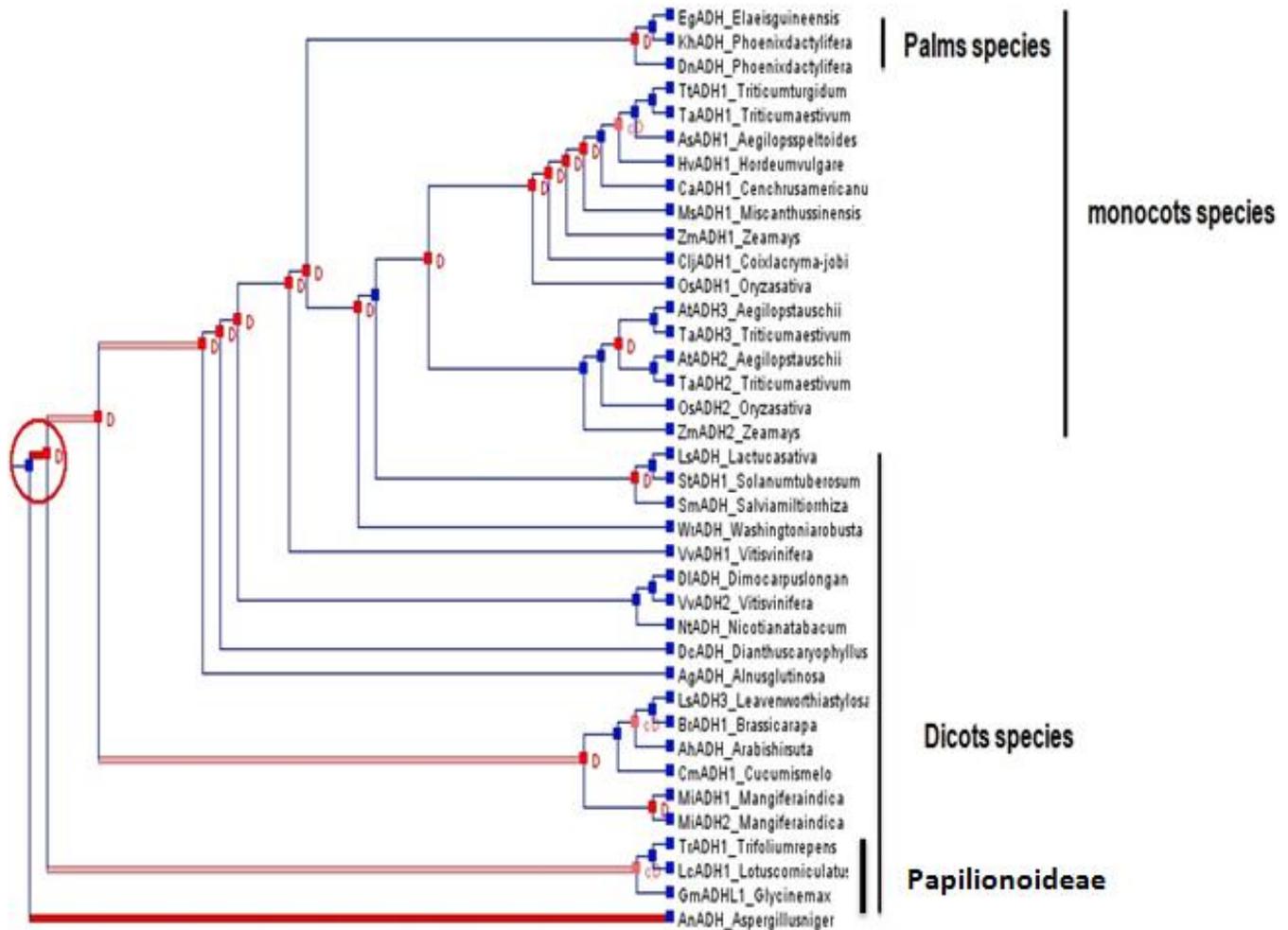


Figure 3. Reconciled tree for the ADH plants family. The ML tree of ADH proteins (figure 2) was reconciled using Notung 2.8 with a species tree compiled from a phylogeny of model organisms. The reconciled tree involves 18 gene duplications (D) and 3 gene co-duplication (cD). The solid red and pink boxes indicate gene duplications that were inferred on the basis of mismatches between the gene tree and the species tree.

Notung2.8 analysis using the ADH gene sequences suggested that the first ADH gene divergence event shown in Figure 3 by a circle separate monocot and dicots species from Papilionoideae species. The palms ADH proteins diverged within the monocots species (Figure 3) except WrADH and before the others monocots species (Figure 3) (Strommer, 2011). Our analysis revealed that in palms, the divergence of WrADH genes occurred after the others palms species diverged. This study revealed the complicated evolution of the ADH gene family that occurred during the course of plant diversification.

In our study, the phylogenetic tree resulting from Notung 2.8 analysis showed that some ADH genes in flowering plants evolved in complex manner that included several duplication events (Figure 3). Duplication events in ADH genes have also been detected in other plant groups at various evolutionary levels. For example, we revealed

duplication events in *A. thaliana* ADH copies (AtADH2 and AtADH3) and *Triticum astivum* ADH copies (TaADH2 and TaADH3) (Figure 3). Sang et al. (1997) showed that diploid species of *Paeonia* (Paeoniaceae) had two or three ADH sequences and that repeated duplication or deletion events occurred after the diversification of this genus. Small and Wendel analyzed ADH genes in *Gossypium* (Malvaceae) in great detail and found that these ADH sequences had experienced duplication events both before and after the divergence in *Gossypium*. Duplicated genes arise frequently in eukaryotic genomes through local events that generate tandem duplications, large-scale events that duplicate chromosomal regions or entire chromosomes, or genome-wide events that result in complete genome duplication (Dujon et al., 2004). Indeed, the existence of multigene families is evidence of the repeated gene duplication that has occurred over the history of life. One

of the examples of the comprehensive analysis of gene duplication events in plants is the study of the MADS-box gene family. This gene family, which plays a central role in the morphogenesis of plant reproductive organs such as ovules and flowers, had experienced duplication events before the origin of angiosperms (Theissen et al., 2000). Moreover, some specific functions were gained through duplication events that took place after the diversification of flowering plants (Theissen et al., 2000). Thus, gene duplication has long been recognized as an important mechanism for the creation of new gene functions (Wagner, 1998; Wagner, 2001). It is likely that each of the ADH genes in the palms that were identified in the present study would have been subjected to different selective pressures over a long period. To determine whether this resulted in new functions, functional analysis of the palms ADH genes in each clade will have to be performed in the future.

Conclusion

The *Adh* genes in the date palm that were identified and analysed in the present study would have been subjected to different selective pressures over a long period. This is the first report revealing that palms species have a ADH genes loci belonging to the same clade. Phylogenetic analyses suggest that these genes resulted from relatively ancient divergence and duplication events.

Conflict of interests

The authors did not declare any conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by grants from the Tunisian minister of Higher Education and Scientific Research.

REFERENCES

- Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yu X, Liu J, Pan L, Zhang T, Yin Y, Xin C, Wu H, Zhang G, Ba Abdullah MM, Huang D, Fang Y, Alnakhli YO, Jia S, Yin A, Alhuzimi EM, Alsaihati BA, Al-Owayyed SA, Zhao D, Zhang S, Al-Otaibi NA, Sun G, Majrashi MA, Li F (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nat Commun.* 4:2274.
- Altschul S, Gish W, Miller W, Myersand E, Lipman D (1997). Basic logical alignment search tool. *Journal of Molecular Biology.* 215:403-410.
- Altschul SF, Koonin EV (1998). Iterated profile searches with PSI-BLAST: a tool for discovery in protein databases. *Trends Biochem. Sc.* 23:444-447.
- Charlesworth D, Liu F, Zhang L (1998). The evolution of the alcohol dehydrogenase gene family in plants of the genus *Leavenworthia* (Brassicaceae): loss of introns, and an intronless gene. *Mol Biol Evol.* 15:552-559.
- Chase T (1999) Alcohol dehydrogenases: identification and names for gene families. *Plant Mol Biol Rep* 17:333-350.
- Chen K, D Durand, M Farach-Colton (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* 7:429-447.
- Clegg MT, Cummings MP, Durbin ML (1997). The evolution of plant nuclear genes. *Proc Natl Acad Sci USA.* 94:7791-7798.
- Conesa A, Götz S (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics.* 76:1-13.
- Cummings MP, Clegg MT (1998) Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. *Proc. Natl. Acad. Sci. USA.* 95:5637-5642.
- Davis JC, Petrov DA (2004). preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol.* 2:318-326.
- Demmin DS, Stockinger EJ, Chang YC, Walling LL (1989). Phylogenetic relationships between the chlorophyll *a/b* binding proteins (*CAB*) multigene family: An intra- and interspecies study. *J. Mol. Evol.* 29:266-279.
- Dolferus R, Bruxelles GD, Dennis ES, Peacock WJ (1994). Regulation of the *Arabidopsis* ADH gene by anaerobic and other environmental stress. *Ann Bot.* 74:301-308.
- Dolferus R, Osterman JC, Peacock WJ, Dennis ES (1997). Cloning of the *Arabidopsis* and rice formaldehyde dehydrogenase genes: implications for the origin of plant ADH enzymes. *Genetics.* 146:1131-1141.
- Dujon B, Sherman D, Fischer G, Durrrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anhouard V, et al (2004). Genome evolution in yeasts. *Nature.* 430:35-44.
- Durbin ML, Learn GH, Huttley GA, Clegg MT (1995). Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proc Natl Acad Sci USA.* 92: 3338-3342.
- Duvall MR, Learn GH, Eguiarte LE, Clegg MT (1993). Phylogenetic analysis of *rbcL* sequences identifies *Acorus calamus* as the primal extant monocotyledon. *Proc. Natl. Acad. Sci. USA.* 90: 4641-4644.
- Eddy SR (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23:205-211.
- Eklund H, Nordstrom B, Zeppezauer E, Soderlund G, Ohlsson I, Boiwe T, Soderberg BO, Tapia O, Branden CI, Akesson (1976). A Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* 102:27-59.
- Fan F, Lorenzen JA, Plapp BV (1991). An aspartate residue in yeast alcohol dehydrogenase I determines the specificity for coenzyme. *Biochemistry.* 30:6397-6401.
- Felsenstein J (2000). PHYLIP: Phylogeny Inference Package, version 3.6 (alpha). University of Washington, Seattle, USA.
- Gaut BS, Morton BR, McCaig BC, Clegg MT (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene ADH parallel rate differences at the plastid gene *rbcL*. *Proc Natl Acad Sci USA.* 93:10274-10279.
- Gaut BS, Peek AS, Morton BR, Clegg MT (1999). Patterns of generic diversification within the ADH gene family in the grasses (Poaceae). *Mol Biol Evol.* 16:1086-1097.
- Gottlieb LD (1982). Conservation and duplication of isozymes in plants. *Science.* 216:373-380.
- Hanson MA, Gaut BS, Stec AO, Fuerstenber SI, Goodman MM, Coe EH, Doebley JF (1996). Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics.* 143:1395-1407.
- Henikoff J, Henikoff S, Pietrovovski S (1999). New features of the Blocks Database servers. *Nucleic Acids Res.* 27:226-228.
- Huttley GA, Durbin ML, Glover DE, Clegg MT (1997). Nucleotide polymorphism in the chalcone synthase-A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. *Mol. Ecol.* 6:549-558.
- Innan H, Tajima F, Terauchi R, Miyashita NT (1996). Intragenic recombination in the ADH locus of the wild plant *Arabidopsis thaliana*. *Genetics.* 143:1761-1770.
- Jordan IK, Wolf YI, Koonin EV (2004). Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol. Biol.* 4, 22.
- Kawabe A, Innan H, Terauchi R, Miyashita NT (1997). Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. *Mol. Biol. Evol.* 14:1303-1315.
- Kimura M (1980). A simple method for estimating evolutionary rates of

- base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111-120.
- Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 25:1451-1452.
- Liu F, Zhang G L, Charlesworth D (1998). Genetic diversity in *Leavenworthia* populations with different inbreeding levels. *Proc. R. Soc. Lond. B Biol. Sci.* 265:293-301.
- Low ETL, Halimah A, Soo-Heong B, Elyana MS, Chi-Yee AT, Leslie CL O, Suan CC, Abdul RR, Kiew LW, Rajinder S (2008). Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: Identifying genes associated with callogenesis and embryogenesis. *BMC Plant Biology.* 8:62-70.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 26:2462-2463.
- Miyashita NT (2001). DNA variation in the 5' upstream region of the ADH locus of the wild plants *Arabidopsis thaliana* and *Arabidopsis gemmifera*. *Mol Biol Evol.* 18:164-171.
- Morton BR, Gaut BS, Clegg MT (1996) Evolution of alcohol dehydrogenase genes in the palm and grass families. *Proc. Natl. Acad. Sci. USA.* 93:11735-11739.
- Nei M (1987). *Molecular Evolutionary Genetics.* Columbia University Press, New York.
- Nicholas KB, Nicholas HBJ, Deerfield DW (1997). GeneDoc: Analysis and visualization of genetic variation. *Embnet News* 4.
- Perry DJ, Furnier GR (1996). *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. *Proc Natl Acad Sci USA.* 93:13020-13023.
- Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky M (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics.* 140:345-356.
- Robert X and Gouet P (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* 42:320-324.
- Sang T, Donoghue MJ, Zhang D (1997). Evolution of alcohol dehydrogenase genes in paeonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species. *Mol Biol Evol.* 14:994-1007.
- Singh R, Abdullah M, Low ETL, Abdul Manaf MA, Rosil R, Rajanaidu N, Ooi CL, Ooi SE, Chan KL, Ab-Halim MA, Azizi N, Jayanthi N, Bacher B, Lakey N; Smith SW, He D, Hogan M, Budiman MA, Lee EK, Desalle R, Kudrana D, Goicoechea JL, Wing RA, Wilson RK, Fulton RS, Ordaway JM, Martienssen RA, Sambanthamur R (2013). Oil palm genome sequences reveals divergence of interfertile species in old and new worlds. *Nature.* 500(7462): 335-339.
- Small RL, Wendel JF (2000). Copy number lability and evolutionary dynamics of the ADH gene family in diploid and tetraploid cotton (*Gossypium*). *Genetics.* 155:1913-1926.
- Strommer J (2011). The plant ADH gene family. *The Plant Journal.* 66:128-142.
- Terauchi RT, Terachi NT, Miyashi TA (1997). DNA polymorphism at the *Pgi* locus of a wild yam, *Dioscorea tokoro*. *Genetics.* 147:1899-1914.
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H (2000). A short history Of MADS-box genes in plants. *Plant Mol Biol.* 42:115-149.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24:4876-4882.
- Trick M, Dennis ES, Edwards KJR, Peacock WJ (1988). Molecular analysis of the alcohol dehydrogenase gene family of barley. *Plant Mol. Biol.* 11:147-160.
- Umezawa T (2010). The cinnamate/monolignol pathway. *Phytochem. Rev.* 9:1-17.
- Wagner A (1998). The fate of duplicated genes: loss or diversification. *BioEssays.* 20:785-788.
- Wagner A (2001). Birth and death of duplicated genes in completely sequenced eukaryotes. *Trends Genet.* 17:237-239.
- Waters E (1995). The molecular evolution of the small heat-shock proteins in plants. *Genetics.* 141:785-795.
- Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul Biol.* 7:256-276.
- Wayne ML, Simonsen KL (1998) Statistical tests of neutrality in the age of weak selection. *TREE* 13:236-240.
- Wilson MA, Gaut B, Clegg MT (1990). Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Molec Biol Evol.* 7:303-314.
- Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 17: 32-43.
- Yang, Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13:555-556.
- Yokoyama S, Harry DE (1993). Molecular phylogeny and evolutionary rates of alcohol dehydrogenases in vertebrates and plants. *Mol. Biol. Evol.* 10:1215-1226.
- Yokoyama S, Yokoyama R, Kinlaw CS, Harry DE (1990). Molecular evolution of the zinc-containing long-chain alcohol dehydrogenase genes. *Mol. Biol. Evol.* 7:143-154.