J Mol Model (2010) 16:919–928
DOI 10.1007/s00894-009-0576-0

ORIGINAL PAPER

Evaluation of the impact of functional diversification on Poaceae, Brassicaceae, Fabaceae, and Pinaceae alcohol dehydrogenase enzymes

Claudia E. Thompson · Cláudia L. Fernandes · Osmar Norberto de Souza · Loreta B. de Freitas · Francisco M. Salzano

Received: 15 June 2009 / Accepted: 26 July 2009 / Published online: 16 October 2009

Abstract The plant alcohol dehydrogenases (ADHs) have been intensively studied in the last years in terms of phylogeny and they have been widely used as a molecular marker. However, almost no information about their three-dimensional structure is available. Several studies point to functional diversification of the ADH, with evidence of its importance, in different organisms, in the ethanol, norepinephrine, dopamine, serotonin, and bile acid metabolism. Computational results demonstrated that in plants these enzymes are submitted to a functional diversification process, which is reinforced by experimental studies indicating distinct enzymatic functions as well as recruitment of specific genes in different tissues. The main objective of this article is to establish a correlation between the functional diversification occurring in the plant alcohol dehydrogenase family and the three-dimensional structures predicted for 17 ADH belonging to Poaceae, Brassicaceae, Fabaceae, and Pinaceae botanical families. Volume, molecular weight and surface areas are not markedly different among them. Important electrostatic and pI differences were observed with the residues responsible for some of them identified, corroborating the function diversification hypothesis. These data furnish important background information for future specific structure-function and evolutionary investigations.

Keywords ADH · Alcohol dehydrogenase · Functional diversification · Molecular evolution · Molecular modeling · Protein structure

Electronic supplementary material The online version of this article (doi:10.1007/s00894-009-0576-0) contains supplementary material, which is available to authorized users.

C. E. Thompson (✉) · L. B. de Freitas · F. M. Salzano
Departamento de Genética, Instituto de Biociências,
Universidade Federal do Rio Grande do Sul,
Caixa Postal 15053, 91501-970,
Porto Alegre, RS, Brazil
e-mail: claudia.thompson@ufrgs.br

C. L. Fernandes · O. Norberto de Souza
Laboratório de Bioinformática, Modelagem e Simulação de Biossistemas, Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul,
Av. Ipiranga, 6681, Prédio 32 - Sala 608,
90619-900 Porto Alegre, RS, Brazil

Introduction

The alcohol dehydrogenase (ADH) proteins belong to the medium-chain dehydrogenases/reductases (MDR) superfamily which has almost 1000 members spread in all types of organisms. MDR-ADH have been described in bacteria, archaea, yeast, plants and animals, and have been additionally implicated in ethanol oxidation, norepinephrine, dopamine, serotonin and bile acid metabolism [1], as well as in the in vitro and in vivo oxidation of retinol [2, 3].

Alcohol dehydrogenases (ADHs) are dimeric enzymes of the glycolytic pathway which encode two types of enzymes, one characterized by short protein chains (~250 residues), represented by *Drosophila* ADHs, which do not require zinc as a cofactor; and another characterized by long protein chains (~370 residues), represented by ADHs from organisms as diverse as mammals, plants and yeasts, which require zinc as a cofactor, and are called class P in dicot and monocot plants. The highest specificity of the ADHs among the latter is for ethanol, aldehyde, and acetaldehyde substrates, but they can also utilize other primary alcohols as well [4].

The catalysis, NAD interactions, evolution, and conformational changes of ADHs have been investigated [5–7], using three-dimensional structures from the horse liver and focusing on the analysis of the differences among the

enzymes of distinct species. Other studies have considered plant *Adh* evolution [8–16]. Plant *Adh* transcription has been demonstrated to increase by environmental stresses such as low oxygen levels, dehydration, low temperatures, and in response to the ABA phytohormone [17]. The activation of the fermentation pathway compensates the decrease of the tricarboxylic acid cycle function and of oxidative phosphorilation, regenerating NAD^+ and producing energy. Phylogenetic studies indicated two or three isozymes, sometimes more than three, in all flowering dicot and monocot plant species, except in *Arabidopsis*, where a single *Adh* locus is found. Differences in *Adh1* allele's specific activity were detected in maize, while different patterns of tissue-specific expression were observed in the *Adh1* and *Adh2* loci [9, 14]. However, no consideration was given to the relationship between structure and evolution, since there was no three-dimensional model of the plant alcohol dehydrogenases available. Gaut et al. [14] assumed that the horse and plant ADH structures were similar and mapped some amino acid replacements of plant onto the horse secondary structure. Actually, there is a powerful method to model protein three-dimensional (3D) structures, which makes it easier to locate the amino acid residues important to the functional diversification of enzymes and predict substrate preferences. This method (comparative protein structure modeling) estimates the 3D structure of a given protein sequence based on its alignment to one or more templates [18].

Experimental studies have shown the ADH involvement in additional metabolic pathways in plants, indicating putative distinct enzymatic functions during tobacco's pollen tube growth [19] and seed storage [20–23], in potato's pollinated pistils [24] and in *Petunia*'s seed detoxification [4].

We recently proposed the first plant ADH three-dimensional model using *Arabis blepharophylla* data [25], obtaining evidence for variation in the subunit-subunit interacting segment, active site and the loop around the second zinc atom. The present work provides 16 other 3D structures, which are considered together with the first described especially in relation to their electrostatic and pI properties. The amino residues theoretically important to the functional divergence among the Poaceae, Brassicaceae, Fabaceae, and Pinaceae modeled ADHs were indicated, as well as those between ADH subtypes, and their position in the 3D structure evaluated to contribute to the elucidation of their functional divergence and molecular evolution.

Materials and methods

Source of the data and sequence alignment

A total of 16 alcohol dehydrogenase sequences were retrieved from the National Center of Biotechnology Information (NCBI) and added to a previous one reported by Thompson et al. [25]. They are listed in Table 1. As indicated there, the 12 species from which they were isolated could be classified in four botanical families. Representatives from ADH1, ADH2, and ADH3 proteins were considered. The ClustalW program [26] was used to perform the alignments, which were inspected and manually changed when necessary using GeneDoc 2.6 (Multiple Sequence Alignment Editor & Shading Utility) [27].

Modeling

Three-dimensional structures for the 17 ADH enzymes were built using the *Equus caballus* liver form (PDB code 1N8K) as a template, obtained through Blastp [28, 29]. Its structure has been solved to a 1.13 Å resolution [30]. The ClustalW program [31] was employed to perform the amino

Table 1 Alcohol dehydrogenase sequences considered, their NCBI accession numbers, and the species from which they were obtained[a]

Botanical family	Abbreviation	NCBI accession number	Species
Brassicaceae	1BRAOLE	BAA34686	*Brassica oleraceae*
	2ARABLE	AAF23531	*Arabis blepharophylla*
	1ARAGRI	AAF23538	*Arabidopsis griffithiana (Arabidopsis pumila* var. *griffithiana)*
	1ARAPAR	AAF23548	*Arabis parishii*
	2LEAST[b]	AAC79416	*Leavenworthia stylosa*
	3LEAST	AAC79418	*Leavenworthia stylosa*
Poaceae	1HORVUL	AAK49116	*Hordeum vulgare subsp. vulgare*
	2HORVUL	P10847	*Hordeum vulgare*
	3HORVUL	CAA31231	*Hordeum vulgare subsp. vulgare*
	1ORYSAT	BAC87776	*Oryza sativa subsp. indica*
	2ORYSAT	BAE00044	*Oryza sativa subsp. indica*
	1ZEAMAY	Q5GA23	*Zea mays*
	2ZEAMAY	P04707	*Zea mays*
Fabaceae	1LOTCOR	CAG30579	*Lotus corniculatus*
	1TRIREP	CAA32934	*Trifolium repens*
	1PISSAT	P12886	*Pisum sativum*
Pinaceae	1PINBAN	AAC49539	*Pinus banksiana*

[a] The number before the sequence identification indicates the ADH subtype (ADH1, ADH2, ADH3)

[b] Only a partial sequence of *Leavenworthia stylosa* ADH1 sequence was described, preventing its modeling

acid sequence alignments, using the BLOSUM62 matrix [32] for scoring. The penalties for gap opening and gap extension were 10.0 and 0.2, respectively. The GeneDoc 2.6 program [27] was used to plot the percent identity of the sequences and manually adjust the alignment. The plot is created by sorting the data to be plotted into ascending order. For each data point the fraction of data points which have the same or a smaller value is computed. The data is then compressed to eliminate multiple points with the same value. The highest value is retained during the compression.

The protein models were obtained through MODELLER 8v2 [33], which implements an approach to comparative protein structure modeling by satisfaction of spatial restraints. The best model was selected using PROCHECK [34] and VERIFY-3D [35]. The PROCHECK program calculates the stereochemical parameters of the main and side-chains, the residues in the most favored regions, bond lengths, and the angle's standard deviation. VERIFY-3D evaluates the compatibility of a 3D model with the amino acid sequence considered using a 3D profile. Each residue position in the 3D model is characterized by its location and environment (alpha, beta, loop, polar, nonpolar, etc.), and it is represented by 20 numbers in the profile. These numbers are called 3D_1D scores. The residue environments are defined by three parameters: the residue area that is buried, the fraction of side-chain area that is covered by polar atoms (O and N), and the local secondary structure. If the model is correct, the sum of the 3D profile scores is high, preferentially above zero. The protein signatures were obtained using a database of protein domains known as PROSITE [36]. The protein volumes and surface areas were calculated according to the Richards' Rolling Probe Method [37, 38], using the 3 V program (Voss Volume Voxelator) [39], with a 1.5 Å probe radius and a high grid resolution (0.5 Å). The theoretical isoelectric point and molecular weight were obtained using the ExPASy Tools available at http://ca.expasy.org/tools/pi_tool.html [40]. Koch et al. obtained a value (5.81) not significantly different from our results (5.65) for the *Arabis blepharophylla* ADH isoelectric point.

The Swiss PDB Viewer [41] was used to calculate the root mean square deviation (RMSD) between the template and the model and also to compute the electrostatic potential using the Coulomb method, as well as to draw all the figures and to generate the molecular surface. The nicotinamide-adenine-dinucleotide (acidic form) and two zinc atoms present in the PDB 1N8K code were located in the modeled three-dimensional structures using its fitting tool. The theoretical models are available upon request.

Functional divergence analysis

The amino acid residues responsible for the functional divergence of the plant ADHs were predicted based on site-specific profiles in combination with suitable cut-off values derived from the posterior probability of each comparison, using Gu's [42] methodology, as in our previous analysis [25]. It is known that functional changes are highly correlated to variations in the evolutionary rates occurring during a certain period of time. Therefore, the identification of the residues submitted to this process in our material were evaluated by finding sites with very different patterns (*e.g.*, very few changes in one cluster but many in the others).

The site-specific profile to identify responsible amino acid sites uses a Qk to be the posterior probability that site k is in state S1 ($0 \leq Qk \leq 1$). A large Qk indicates a high possibility that the functional constraint (or the evolutionary rate) of a site is different between two clusters. We used three cut-off values, equal to or above respectively 0.80, 0.85, and 0.90.

Results

Sequence alignment and modeling

The results obtained with the multiple alignments are presented in Fig. 1S, and they show high similarity among the sequences. The degree of identity between the sequence of the selected template and the models was around 48%. In general, the number of gaps in the template's primary sequence is very low (Fig. 1S), so it does not significantly affect the comparative molecular modeling. The inserted region of the alignment (Fig. 1S – positions 75 up to 83), which do not have an equivalent segment in the template, was modeled in the context of the whole molecule, using its primary sequence alone. The percent identity of the sequences is presented in Fig. 2S. All target proteins have the signature of the zinc alcohol dehydrogenase family, which has a consensual pattern corresponding to G-H-E-X (2)-G-X(5)-[GA]-X(2)-[IVSAC]. Ten models were initially created, and they were considered using PROCHECK and VERIFY-3D, as well as the root mean square deviations (RMSD).

The stereochemical parameters used to verify the quality of the models are listed in Tables 1S, 2S, and 3S (Electronic Supplementary Information). The mean of percentage of amino acid residues in most favored regions according to the Ramachandran plot shows variation from 91.67% (Poaceae) to 93.2% (Fabaceae), which is not significant since all results above 90% are considered of good quality. No model value was lower than 90.8%, confirming the excellent quality of the initial models.

It is important to observe that the G factor measures how "normal" is a given stereochemical property, considering the torsion angles and the bond lengths in the main chain. Therefore, when applied to a specific residue, a low G factor

indicates that the property corresponds to a low probability of conformation. A G factor value smaller than -1.0 could indicate geometry problems. In this work, all G factor results are near -0.1 (Tables 1S, 2S, and 3S, Supplementary Information). Observing the VERIFY_3D (Figs. 3S, 4S, and 5S, Supplementary Information) results, we can see that the sum of 3D profile scores is high in all cases. The region near the 301 amino acid residue, however, shows the smaller 3D_1D average scores for all botanical families, which means that this is most likely the area with the higher number of structural problems. In a general way, the graphics show a similar pattern. Taken together, these data suggest that the models were stereochemically valid.

Structural information, electrostatic and pI differences

Information concerning number of residues, molecular weight, surface area, and volume is shown in columns 3-6 of Table 2. *Brassica oleraceae* (1BRAOLE) has a reduced number of amino acids (350), conditioning also lower values for the molecular weight and volume. The opposite occurs in 1ZEAMAY which presents the highest number of residues (388). No clear differences in relation to these variables were observed in the ADHs of different botanical families.

The molecular surface of this protein is electrostatically polarized (Figs. 1, 2, and 3). The Brassicaceae have the most acid ADH proteins when compared to the other families (Table 2), with *Brassica oleraceae* (1BRAOLE) having the most negative pI (5.47) value, followed by *Arabis blepharophylla* (2ARABLE; 5.65), *Arabis griffthiana* (1ARAGRI; 5.69), and *Arabis parishii* (1ARAPAR; 5.88). The regions of the active site, the second zinc atom, and of the subunit-subunit interacting segment (middle portion, upper and lower right region of the figures, respectively) show the greatest differences (Fig. 1). These proteins have a pI value significantly different from those of the *Leavenworthia* proteins (2LEAST and 3LEAST), which show pI values equal to 6.37 and 6.40, respectively.

Considering now the Poaceae group (Table 2 and Fig. 2), it is seen that the ADH1 forms 1HORVUL and 1ORYSAT (pI 6.28 and 6.20; nos. 1 and 4 in the Figures) are more basic than the ADH2 forms of the same species (respectively 5.52 and 6.04; nos. 2 and 5 in the Figure), the same occurring in *Zea mays* (6.43 and 5.72, nos. 6 and 7 in the Figure). The most significant differences in electrostatic potential is in the region near the second zinc atom and in the subunit-subunit interacting segment (upper and lower right, Fig. 2), a smaller contrast being observed in the active site region.

The pI values for the Fabaceae are not much different (Table 2). However, there is a different concentration of negative charges between the models, the subunit-subunit segment of *Lotus corniculatus* (Fig. 3.1) showing a clear

Table 2 Theoretical values obtained for the ADH models and the template from *Equus caballus*[a]

			Theoretical values			
Botanical families	Abbreviation	Number of residues	Molecular weight (D)	Surface area (Å^2)	Volume (Å^3)	Isoelectric point
Brassicaceae	1BRAOLE	350	38001.58	14006.5	47976.37	5.47
	2ARABLE	379	40994.03	13889.1	51463.00	5.65
	1ARAGRI	379	41308.23	13923.9	51977.75	5.69
	1ARAPAR	379	41165.19	13729.7	52032.37	5.88
	2LEAST	379	41454.79	13613.9	52275.25	6.37
	3LEAST	380	41255.52	14081.2	52307.50	6.40
Poaceae	1HORVUL	379	40903.29	13873.0	51820.50	6.28
	2HORVUL	373	40511.62	13683.4	51113.75	5.52
	3HORVUL	379	41011.48	14243.1	52027.37	6.08
	1ORYSAT	379	40984.30	13994.8	52085.75	6.20
	2ORYSAT	379	41176.75	14089.5	52134.00	6.04
	1ZEAMAY	388	41975.50	14529.8	53186.50	6.43
	2ZEAMAY	379	41054.43	14467.6	52977.87	5.72
Fabaceae	1LOTCOR	380	41096.13	14156.2	51981.75	5.92
	1TRIREP	380	41172.33	14336.3	52204.00	6.08
	1PISSAT	380	41155.37	14198.7	52050.25	6.09
Pinaceae	1PINBAN	375	40465.59	13794.6	51078.00	5.91
Template	1N8K	374	39806.29	13187.4	51493.12	8.31

[a] The number before the sequence identification indicates the ADH subtype (ADH1, ADH2, ADH3)

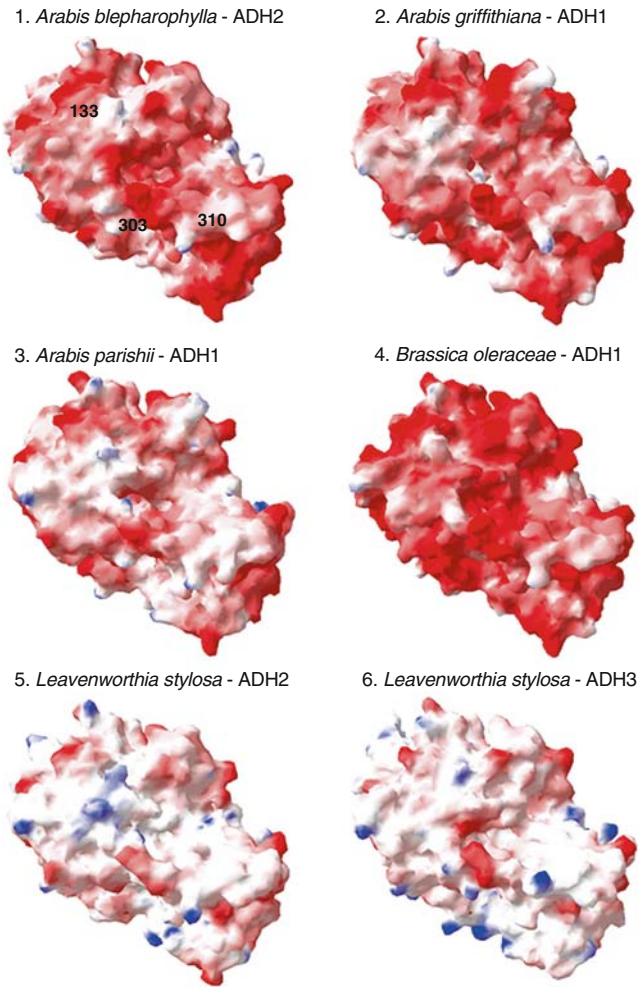

Fig. 1 View of the surface topology of the Brassicaceae ADH models with the electrostatic potential represented as red (most negative), white (neutral), and blue (most positive). Numbers in black refer to the sites identified as showing functional divergence ($Qk \geq 0.90$) among botanical families. Those numbered 315 and 337 are placed on the other side of the figure and cannot be displayed. Since the molecules are shown in the same position, only the first was labeled

difference from the other two (Figs. 3.2 and 3.3). The *Pinus banksiana* model have a pI value of 5.91 (Table 2), and the protein has the negative charges concentrated in the active site region (Fig. 3.4).

Functional divergence analysis

Sites showing *Qk* values above 0.8 and therefore suggestive of being associated with functional divergences are listed in Table 3 for the comparisons involving different botanical families (60 sequences considered); while in Table 4 the comparisons are between the ADH1 and ADH2 forms. Data related to ADH3 could not be used for this analysis because the number of sequences available was less than those needed for statistical comparisons [43].

Concentrating our attention to the comparisons which yielded $Qk \geq 0.9$ only, we see that in the Brassicaceae vs. Fabaceae contrast, three residues which occur in loops (133, 303, 310) (Table 4S) show different amino acid

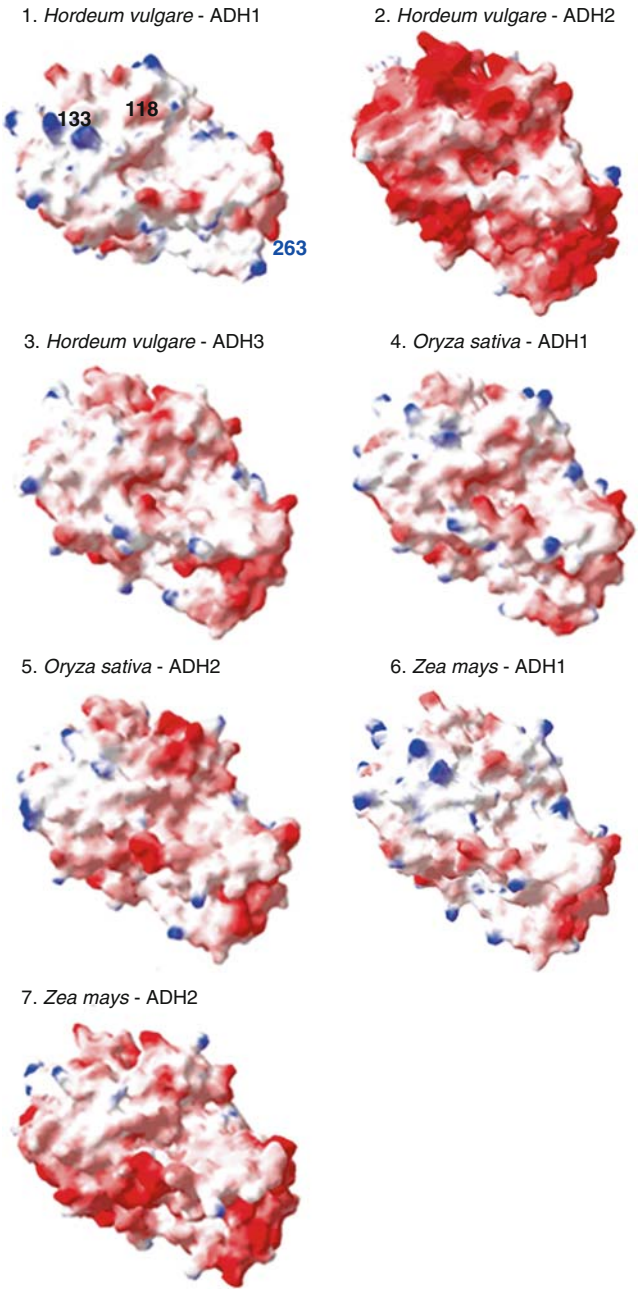

Fig. 2 View of the surface topology of the Poaceae ADH models with the electrostatic potential represented as red (most negative), white (neutral), and blue (most positive). Numbers in black refer to the sites identified as showing functional divergence ($Qk \geq 0.90$) among botanical families. That numbered 236 is placed on the other side of the figure and cannot be displayed. The number in blue refers to the site showing functional divergence ($Qk \geq 0.90$) between ADH forms. Sites nos. 234 and 329 are placed on the other side of the figure. Since the molecules are shown in the same position, only the first was labeled

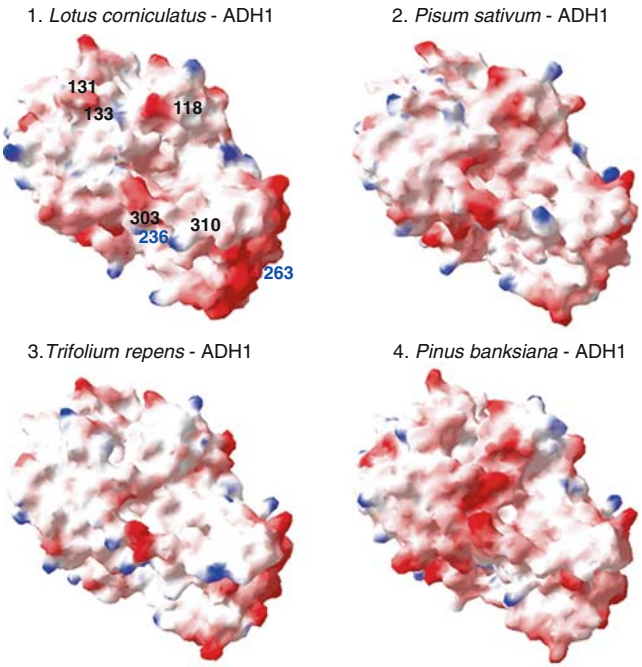

Fig. 3 View of the surface topology of the Fabaceae and Pinaceae ADH models with the electrostatic potential represented as red (most negative), white (neutral), and blue (most positive). Numbers in black refer to the sites identified as showing functional divergence ($Qk \geq 0.90$) among botanical families. Those numbered 315 and 337 are on the other side of the figure. The number in blue refers to the site showing functional divergence ($Qk \geq 0.90$) between ADH forms. Site no. 329 is on the other side of the figure. Since the molecules are shown in the same position, only the first was labeled

conservation, the same being true for two others (315, 337) that are located in helices (Table 4S). Positions 133 (Gly), 303 (Ser), and 337 (Gly) are conserved within Brassicaceae, but highly divergent in Fabaceae (133: Asn, Gly, Ser; 303: Asn, Ser, Lys; 337: Asn, Leu, Ser, Gly) (Table 3). The three Fabaceae modeled show variability in position 303 only (Table 4S). In the Poaceae vs. Fabaceae comparative analysis, two amino acid residues of loop regions (118, 133) and one in the helix secondary structure (236) (Table 4S) should be considered; while in the Fabaceae vs. Pinaceae comparison the amino acids to be distinguished are 131 and 133 (loop) and 337 (helix) (Tables 3 and 4S). The Poaceae ADHs present conservation in residues 118 (Asp) and 133 (Gly), and the six Fabaceae in residue 236 (Phe) (Table 3). Considering the three Fabaceae ADH modeled, position 118 is variable in the Poaceae vs. Fabaceae, and position 131 in the Fabaceae vs. Pinaceae comparisons (Table 4S).

As presented in Tables 4 and 5S, both in the functional divergence analysis and in the models, amino acids that show different rates of change between Poaceae's ADH1 and ADH2 are nos. 234 and 263 (in helices) and 329 (loop), ADH2 being conserved for all of them. In the Poaceae vs. Fabaceae comparison the Poaceae ADH1s

Table 3 Amino acid residues changes associated with the functional divergence among the botanical families[a]

Comparison[b]	Amino acid residue position	Amino acid residue	
Brassicaceae (31) vs. Poaceae (16)		in Brassicaceae	in Poaceae
	236	F	F, Y, H
Brassicaceae (31) vs. Pinaceae (7)		in Brassicaceae	in Pinaceae
	271	R	Y, C
	310	T, S	T
	315	F, L	F
	317	N	N, C, T, S
Brassicaceae (31) vs. Fabaceae (6)		in Brassicaceae	in Fabaceae
	45	F	*Y, F*
	49	C, *S, W*	C
	57	E	E, D
	64	L, *W, R*	L
	82	V, I, A	V
	90	Q, A, K	K
	112	E, *V, G*	E
	125	E, D	D
	127	G, V, R	G
	128	G, V	V
	130	I	I, L
	133*	G	N, G, S
	135	S	S, T
	139	I	I, K
	178	I	I, V
	187	L	*F, L*
	188	G, E, R	G
	190	T, V, I, P	T
	194	A, V	A
	213	A, G	A
	219	R, K	R
	221	A, S	S
	224	S, G	S
	237	D, E	E
	241	K, E	K
	295	V	V, *L, T*
	303*	S	*N, S, K*
	310*	T, S	T
	311	H	H, *A, N*
	315*	F, L	F
	337*	G	N, *L, S, G*
	338	V, I, L	V
	344	N	*N, K, R, S*
Poaceae (16) vs. Fabaceae (6)		in Poaceae	in Fabaceae
	118*	D	D, E, N
	133*	G	N, G, S
	236*	F, Y, H	F

Table 3 (continued)

Comparison[b]	Amino acid residue position	Amino acid residue	
	279	I, V, A	I
Fabaceae (6) vs. Pinaceae (7)		in Fabaceae	in Pinaceae
	131*	*S, H, N*	S
	133*	N, G, S	G
	209	A	G, A, *T, S*
	271	R	Y, C
	337*	N, *L, S, G*	G
Poaceae (16) vs. Pinaceae (7)		in Poaceae	in Pinaceae
	161	V	V, A, S
	209	A	G, A, *T, S*
	271	R	Y, C
	313	M	V, L, I

[a] Only sequences which yielded $Qk \geq 0.80$ are listed; amino acid residues with Q(k)≥0.85 are in bold face, and those with $Qk \geq 0.90$ are distinguished by an asterisk (*). The amino acid residues are displayed by decreasing order of frequency. Residues in italics are those with the same frequency. Those in italics and underlined have smaller frequencies than the residues placed before them

[b] Numbers in parentheses indicate the number of sequences used in this analysis (data supplied on request)

exhibit differences in residues 263, located in helix and 329, in a loop.

A ribbon representation of one model of each botanical family showing sites identified as functional divergent ($Qk \geq 0.85$) is presented in Figs. 4 and 5. The subunit-subunit interaction segment seems to be the region with the highest number of functionally important residues in Brassicaceae and Fabaceae (Figs. 4.1 and 5.1, respectively). In Fabaceae the amino acids forming helices and loops around the second zinc atom region are variable (Fig. 5.1). Amino acid changes near the same region distinguish the Poaceae ADH forms, as well as substitutions in the dimer interaction zone (Fig. 4.2). The same regions are fundamental for the diversification of Pinaceae ADHs (Fig. 5.2). There are also some differences among all ADHs near the coenzyme region (in green).

Discussion

Alcohol dehydrogenase is an essential enzyme in the anaerobic metabolism, and it has been widely used as a molecular marker in plants due to its convenient size (2-3 kb in length with a ~1000 nucleotide coding sequence, 10 exons, 9 introns) and low copy number. The enzyme is important primarily for responses to hypoxic conditions, when its expression is highly induced. Moreover, it has an

Table 4 Amino acid residues changes associated with the functional divergence between ADH1 and ADH2[a]

Comparison[b]	Amino acid residue position	Amino acid residue	
Poaceae ADH1 (9) vs. Poaceae ADH2 (6)		Poaceae ADH1	Poaceae ADH2
	25	V, S, T	S
	41	V	V, *D, I*
	45	F, Y	Y
	62	T, I	T
	64	V, M	V
	79	V, I	V
	109	C, S	C
	112	A, P	E
	170	A, E, Q	E
	178	V	I, L
	183	I	F, I
	185	T, S	T
	190	T, S	T
	200	S	Q, *M, S*
	204	I, V	I
	221	A	S, A
	229	I, V	V
	233	A, P	P
	234*	N, S, V	A
	236	F	F, *H, Y*
	240	R, K	K
	259	Q, E	E
	263*	E, D	E
	285	A	C, A
	329*	Y, F	Y
	337	N	N, G
Poaceae ADH1 (9) vs. Fabaceae ADH1 (6)		Poaceae ADH1	Fabaceae ADH1
	64	V, M	L
	263*	E, D	E
	329*	Y, F	Y
Poaceae ADH2 (6) vs. Fabaceae ADH1 (6)		Poaceae ADH2	Fabaceae ADH1
	41	V, *D, I*	L
	118	D	D, E, N
	133	G	N, G, S
	221	S, A	S
	236	F, *Y, H*	F
	238	Q	*L, E,* <u>*G, Q*</u>
	279	I	I, *V, A*
	285	C, A	A

[a] Only sequences which yielded $Qk \geq 0.80$ are listed; amino acid residues with Q(k)≥0.85 are in bold face, and those with $Qk \geq 0.90$ are distinguished by an asterisk (*). The amino acid residues are displayed by decreasing order of frequency. Residues in italics are those with the same frequency. Those in italics and underlined have smaller frequencies than the residues placed before them

[b] Numbers in parentheses indicate the number of sequences used in this analysis (data supplied on request)

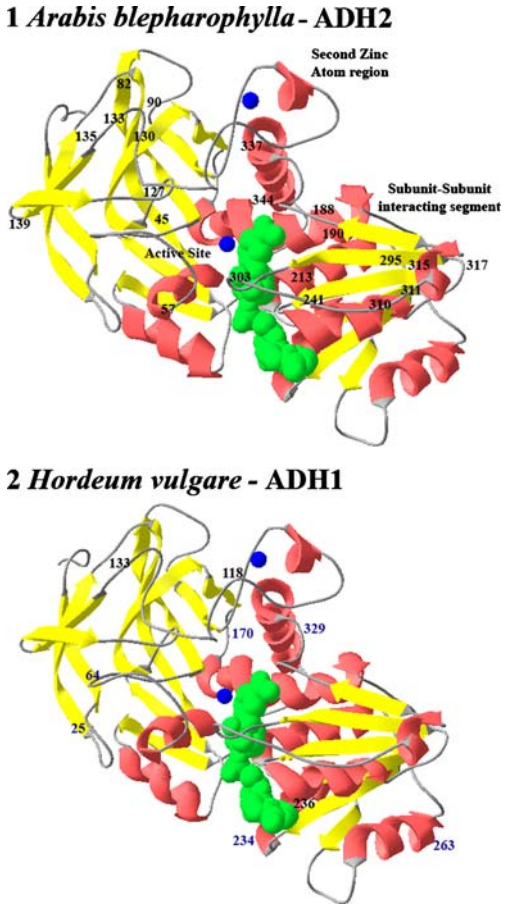

Fig. 4 Ribbon representation of the ADHs three-dimensional structures in the same orientation shown in Figs. 3 and 4: (1) 2ARABLE and (2) 1HORVUL. Numbers in black refer to the sites identified as showing functional divergence ($Qk \geq 0.85$) among botanical families. Numbers in blue identified sites showing functional divergence ($Qk \geq 0.85$) between ADH forms. Zinc atoms are displayed in blue, and the nicotinamide-adenine-dinucleotide (acidic form) is shown in green

important role in fruit ripening, seedling and pollen development [44]. Despite the large number of phylogenetic investigations performed, no extensive work correlating its sequence and structure in plants exists.

Studies in *Zea mays* have revealed that different alloenzyme types of *Adh1* exhibit different specific activity, and distinct pattern of organ-specific gene expression [45, 46]. An exchange of Tyr for Asp at residue 52, located in a helix structure in the *Adh1-C* allele, alters enzymatic properties by reducing the specific activity. Additionally, amino acid replacements changing the secondary structure were also reported [9].

In humans, ADH is a cytosolic enzyme able to metabolize ethanol and a wide variety of substrates, including aliphatic alcohols, hydroxysteroids and lipid peroxidation products. Its catalytic properties are variable. The *Adh2* gene may be present as *Adh2*1*, *Adh2*2*, and *Adh2*3* encoding for β1, β2, and β3 subunits, respectively, which differ by a single nucleotide change. The enzyme containing the β1 subunit has high affinity and low capacity for ethanol, whereas the β2 and β3 forms show lower affinity and higher capacity. Additionally, the human tissues show measurable different *Adh* gene expressions [47].

The proteins modeled in this work are composed by two domains and have a similar fold. The nucleotide binding domain is formed by a structural motif known as Rossmann fold [48], consisting of parallel beta strands linked by alpha helices (Figs. 4 and 5, region of nicotinamide binding at lower right). The catalytic region containing residues involved in substrate binding has a zinc atom located deep in the cleft formed between the two domains. There are divergent amino acid residues localized in three important regions (the loop around the zinc atom, an important cofactor for the enzyme's function; the subunit-subunit interacting segment, responsible for the dimer formation; and the active site) which are probably submitted to functional diversifica-

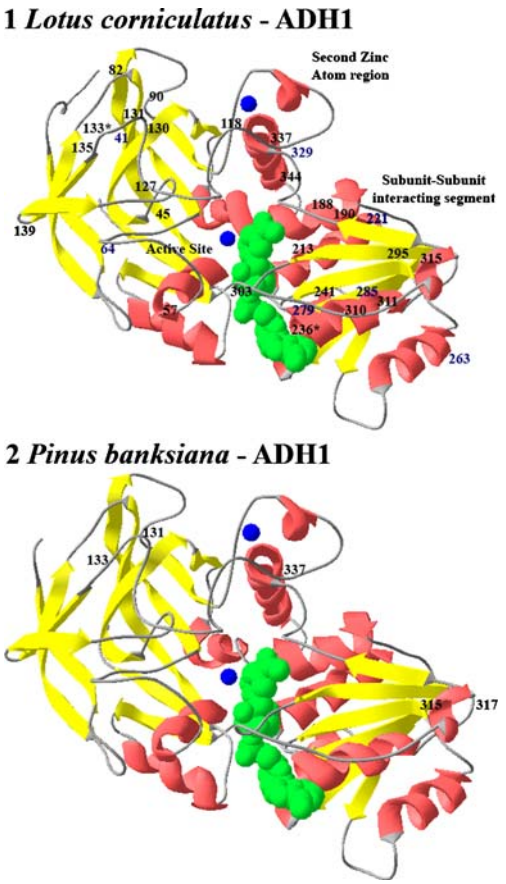

Fig. 5 Ribbon representation of the ADHs three-dimensional structures in the same orientation shown in Fig. 5: (1) 1LOTCOR and (2) 1PINBAN. Sites showing functional divergence ($Qk \geq 0.85$) among botanical families are in black. Those showing functional divergence ($Qk \geq 0.85$) between ADH forms are in blue. Residues 133 and 236 are distinguished by an asterisk (*) in 1LOTCOR, since they are important both to the divergence among botanical families and between ADH forms, as can be seen in Tables 3, 4, 4S, and 5S. Zinc atoms are displayed in blue, and the nicotinamide-adenine-dinucleotide (acidic form) is shown in green

tion. Zinc seems to be important for the catalysis and geometry stabilization of the active site. These two processes could be achieved by moderating the electrostatic potential near the substrate or by zinc acting as ligand during the enzyme's catalysis [49]. Thus the residues indicated as functionally divergent near the zinc atom region possibly have an impact on ADH function. Some residues located near the zinc atom region, such as 109 and 112, which were not previously discussed since they have $0.80 \leq Qk \leq 0.85$, may also be candidates for future investigations. The same can be said of 313 that is related to the subunit-subunit interaction, and residues nos. 49, 62 and 178, present near the active site. The first helix, located in residues 49 up to 55 using 1HORVUL as reference, can accommodate large movements associated with the loop near the active site [49]; consequently, amino acid no. 64 (loop) has high probability to contribute to these movements.

Clearly the modeled proteins show electrostatic potential differences in the molecular surface. Comparing proteins of the same species, ADH1 seems to be more basic than the ADH2 enzymes. *Arabis blepharophylla* ADH1, which was not model, has a theoretical pI equal to 5.74, greater than the 5.65 from the modeled ADH2, corroborating the pattern observed between the ADH forms.

Electrostatic interactions have an important role in the structure and function of biological molecules. Association of proteins in solution and in membranes, enzyme-substrate complexation, chemical reactions in enzyme active sites, charge transfer, are all drastically affected by the strength and distribution of the electrostatic field around regions in biological molecules. The protein-protein interactions are affected by several surfaces properties, such as cavities, hydrophobic residues, specific interaction residue pockets, and electrostatics. This latter has a high potential for functional protein classification [50], since it plays an important role in the specificity of protein-ligand or protein-protein interactions. Due to its attractive or repulsive forces, certain protein-protein interactions could be more or less favorable [50]. The electrostatic and pI differences described here most certainly lead to dissimilar functional efficiency, a subject that is now open for further investigation. Note that the number of plant proteomic papers is still quite reduced as compared to those of other organisms (only about 3% according to Jorrín et al. [51]).

It is well-known that variation in a specific DNA region not necessarily correlates with the evolutionary pattern of the organism as a whole. Our results summarized in Tables 3 and 4 add new information on this point. In a previous study [25], based on 1155 sites from 176 sequences, we found a close relationship between the Brassicaceae and Fabaceae families. But it is between them that we find the largest number of site differences (a total of 33 with $Qk \geq 0.80$; 20 with $Qk \geq 0.85$; and five with $Qk \geq 0.90$). The other between-family comparisons show much fewer differences, despite the fact that they are placed far away in the phylogenetic tree [25].

The dissimilarities between the Poaceae ADH1 and ADH2 [26 sites with $Qk \geq 0.80$; six with $Qk \geq 0.85$; three with $Qk \geq 0.90$; Table 4] point to the functional differences which exist between these two forms. Our models clearly differentiate them structurally in *H. vulgare*, *O. sativa*, and *Z. mays* (Fig. 2). On the other hand, the ADH1 from the Poaceae and Fabaceae show three sites with clear functional differences. All these findings point to the subtle quantitative changes that occur at the molecular level as a result of the evolutionary process.

Acknowledgments We would like to thank Dr. Ana Tereza R. Vasconcelos, Dr. Élgion L. S. Loreto, Dr. Laurent E. Dardenne, and Dr. Paulo M. Bisch for their helpful comments. This work was supported by the Institutos do Milênio and Programa de Apoio a Núcleos de Excelência, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul, and Pró-Reitoria de Pesquisa da Universidade Federal do Rio Grande do Sul.

References

1. Höög JO, Hedberg JJ, Stromberg P, Svesson S (2001) Mammalian alcohol dehydrogenase – functional and structural implications. J Biomed Sci 8:71–76
2. Boleda MD, Saubi N, Farrés J, Parés X (1993) Physiological substrates for rat alcohol dehydrogenase classes: aldehydes of lipid peroxidation, omega-hydroxyfatty acids, and retinoids. Arch Biochem Biophys 307:85–90
3. Martras S, Alvarez R, Martinez SE, Torres D, Gallego O, Duester G, Farrés J, de Lera AR, Parés X (2004) The specificity of alcohol dehydrogenase with *cis*-retinoids. Activity with 11-*cis*-retinol and localization in retina. Eur J Biochem 271:1660–1670
4. Garabagi F, Duns G, Strommer J (2005) Selective recruitment of *Adh* genes for distinct enzymatic functions in *Petunia* hybrid. Plant Mol Biol 58:283–294
5. Eklund H, Bränden CI (1979) Structural differences between apo- and holoenzyme of horse liver alcohol dehydrogenae. J Biol Chem 254:3458–3461
6. Danielsson O, Atrian S, Luque T, Hjelmqvist L, Gonzalez-Duarte R, Jörnvall H (1994) Fundamental molecular differences between alcohol dehydrogenase classes. Proc Natl Acad Sci USA 91:4980–4984
7. Persson B, Bergman T, Keung WM, Waldenström U, Holmquist B, Vallee BL, Jörnvall H (1994) Structural and functional divergence of class II alcohol dehydrogenase – cloning and characterisation of rabbit liver isoforms of the enzyme. Eur J Biochem 216:49–56
8. Chang C, Meyerowitz EM (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 83:1408–1412
9. Gaut BS, Clegg MT (1993) Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc Natl Acad Sci USA 90:5095–5099
10. 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

11. 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
12. Miyashita NT, Kawabe A, Innan H, Terauchi R (1998) Intra- and interspecific DNA variation and codon bias of the alcohol dehydrogenase (*Adh*) locus in *Arabis* and *Arabidopsis* species. Mol Biol Evol 15:1420–1429
13. Charleswort D, Liu FL, Zhang L (1998) The evolution of the alcohol dehydrogenase gene family by loss of introns in plants of the genus *Leavenworthia* (Brassicaceae). Mol Biol Evol 15:552–559
14. Gaut BS, Peek AS, Morton BR, Clegg MT (1999) Patterns of genetic diversification within the *Adh* gene family in the grasses (Poaceae). Mol Biol Evol 16:1086–1097
15. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidospis*, *Arabis* and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498
16. Lin J-Z, Brown AHD, Clegg MT (2001) Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). Proc Natl Acad Sci USA 98:531–536
17. Dolferus R, Jacobs M, Peacock WJ, Dennis ES (1994) Differential interactions of promoter elements in stress responses of the *Arabidopsis Adh* gene. Plant Physiol 105:1075–1087
18. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325
19. Bucher M, Brander KA, Sbicego S, Mandel T, Kuhlemeier C (1995) Aerobic fermentation in tobacco pollen. Plant Mol Biol 28:739–750
20. Zhang M, Maeda Y, Furihata Y, Nakamaru Y, Esashi Y (1994) A mechanism of seed deterioration in relation to the volatile compounds evolved by dry seeds themselves. Seed Sci Res 4:49–56
21. Zhang M, Yajima H, Umezawa Y, Nakagawa Y, Esashi Y (1995) GC-MS identification of volatile compounds evolved by dry seeds in relation to storage conditions. Seed Sci Technol 23:59–68
22. Zhang M, Nakamaru Y, Tsuda S, Nagashima T, Esashi Y (1995) Enzymatic conversion of volatile metabolites in dry seeds during storage. Plant Cell Physiol 36:157–164
23. Zhang M, Nagata S, Miyazawa K, Kikuchi H, Esashi Y (1997) A competitive enzyme-linked immunosorbent assay to quantify acetaldehyde-protein adducts that accumulate in dry seed during aging. Plant Physiol 113:397–402
24. Van Eldik GJ, Ruiter RK, Van Herpen MMA, Schrauwen JAM, Wullems GJ (1997) Induced ADH gene expression and enzyme activity in pollinated pistils of *Solanum tuberosum*. Sex Plant Reprod 10:107–109
25. Thompson CE, Salzano FM, Souza ON, Freitas LB (2007) Sequence and structural aspects of the functional diversification of plant alcohol dehydrogenases. Gene 396:108–115
26. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. Trends Biochem Sci 23:403–405
27. Nicholas KB, Nicholas HB Jr (1997) GeneDoc: a tool for editing and annotating multiple sequence alignment. Distributed by the authors (www.psc.edu/biomed/genedoc)
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
29. Altschul SF, Madden TL, Schaffer AA, Zhang JZ, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402
30. Rubach JK, Plapp BV (2003) Amino acid residues in the nicotinamide binding site contribute to catalysis by horse liver alcohol dehydrogenase. Biochemistry 42:2907–2915
31. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalty and weight matrix choice. Nucleic Acids Res 22:4673–4680
32. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89:10915–10919
33. Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 243:779–815
34. Laskowski RA, McArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
35. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85
36. Gattiker A, Gasteiger E, Bairoch A (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. Appl Bioinformat 1:107–108
37. Richards FM (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. J Mol Biol 82:1–14
38. Richards FM (1977) Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng 6:151–176
39. Voss NR (2006) The geometry of the ribosomal polypeptide exit tunnel. J Mol Biol 360:893–906
40. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy Server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, Totowa, pp 571–607
41. Guex N, Peitsch MC (1997) SWISS-MODEL and Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2733
42. Gu X (2001) Mathematical modeling for functional divergence after gene duplication. J Comp Biol 3:221–234
43. Gu X, Vander Velden K (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics 18:500–501
44. 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
45. Schwartz D, Laughner WJ (1969) A molecular basis for heterosis. Science 166:626–627
46. Freeling M, Bennet DC (1985) Maize *Adh 1*. Annu Rev Genet 19:297–323
47. Gemma S, Vichi S, Testai E (2006) Individual susceptibility and alcohol effects: biochemical and genetic aspects. Ann Ist Super Sanità 42:8–16
48. Lesk AM (1995) NAD-binding domains of dehydrogenases. Curr Opin Struct Biol 5:775–783
49. Baker PJ, Britton KL, Fisher M, Esclapez J, Pire C, Bonete MJ, Ferrer J, Rice DW (2008) Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. PNAS 106:779–784
50. Valeyev NV, Downing AK, Sondek J, Deane C (2008) Electrostatic and functional analysis of the seven-bladed WD β-Propellers. Evol Bioinform 4:203–216
51. Jorrín JV, Maldonado AM, Castillejo MA (2007) Plant proteome analysis: a 2006 update. Proteomics 7:2947–2962