

EVOLUCIÓN MOLECULAR DE LOS GENES *BADH*MOLECULAR EVOLUTION OF *BADH* GENES

Juan Porfirio Legaria Solano

<sup>1</sup> Departamento de Fitotecnia, Universidad Autónoma Chapingo. Km. 38.5 Carr. México-Texcoco. C.P. 56230 Chapingo, Edo. de México. Tel: 01 (595) 952-1642. Correo electrónico: jlegaria@taurus1.chapingo.mx

## RESUMEN

Se hizo un análisis comparativo entre 18 secuencias de aminoácidos de enzimas Betaina Aldehído Deshidrogenasas (BADH), que participan en la síntesis del osmoprotector glicina betaina, en algunas especies vegetales pertenecientes a familias económicamente importantes, como Chenopodiaceae, Amaranthaceae, Brassicaceae, Acanthaceae y Poaceae. El objetivo fue estudiar la evolución molecular de la vía de síntesis de la glicina betaina y de los genes que codifican para las enzimas que en ella participan, a fin de definir estrategias que permitan transferir la vía de síntesis a especies de interés agronómico que carecen de la misma, y así conferirles tolerancia a sequía, salinidad y frío. La comparación de las secuencias y el análisis filogenético se llevó a cabo con los programas de cómputo ClustalX y ClustalW. Se encontró que las proteínas BADH de especies dicotiledóneas están filogenéticamente más relacionadas entre sí que con aquellas de monocotiledóneas. Las BADHs de amaranto (*Amaranthus hypochondriacus* L.) mostraron entre sí sólo 12 aminoácidos sustituidos (98 % de identidad), lo que sugiere que la evolución de dichas enzimas constituye un evento relativamente reciente. La estructura y la secuencia codificadora de los genes BADH de amaranto (dicotiledónea) y de arroz (*Oryza sativa* L.) (monocotiledónea) se ha conservado a través del proceso evolutivo; sin embargo, en el gen del arroz algunos de los intrones son más pequeños. Las evidencias sugieren que las enzimas BADH de plantas han divergido en el sitio donde realizan su actividad, en los cloroplastos y los peroxisomas

**Palabras clave:** Evolución, genes, betaina aldehído deshidrogenasa (BADH), osmoprotección.

## SUMMARY

A comparative analysis among 18 Betaine Aldehyde Dehydrogenase (BADH) amino acid sequences was done. BADH enzymes participate in the synthesis of glycinebetaine osmolyte in plant species of economic importance, such as in Chenopodiaceae, Amaranthaceae, Brassicaceae, Acanthaceae and Poaceae. The molecular evolution of glycinebetaine synthesis pathway and the involved genes, in relation to regulatory sequences and intracellular localization, were analyzed as a step toward molecular genetic manipulation of the betaine pathway in plants. The amino acid sequence analyses and phylogenetic tree construction were based on sequence alignment obtained with the ClustalX and ClustalW software. The deduced amino acid sequences show that BADH enzymes of dicotyledonous species are closely related. Dicotyledonous and monocotyledonous BADH enzymes are phylogenetically distant. The homology between two BADHs from amaranto (*A. hypochondriacus*) was high (98 %), with only 12 amino acid substitutions. This suggests that the evolution of amaranto BADHs constituted relatively recent events. The BADH

gene structure and protein coding sequence were conserved during evolution in amaranth (dicotyledonous) and rice (*Oryza sativa* L.) (monocotyledonous); nevertheless, the rice gene contains more small size introns than the amaranth BADH gene. The plant BADH enzymes had significantly different subcellular localization, chloroplasts and peroxisomes

**Index words:** Evolution, genes, betaine aldehyde dehydrogenase (BADH), osmoprotection.

## INTRODUCCIÓN

Dobzhansky (1970) estableció que nada tiene sentido en biología, excepto bajo el prisma de la evolución. La biología molecular ha desarrollado los métodos necesarios para descifrar la información contenida en las cadenas de ADN y de las proteínas, es decir, la secuencia lineal de los componentes de estas moléculas. Con ello se ha abierto el camino más certero y eficaz para reconstruir la historia evolutiva de los organismos. Los conceptos subyacentes al proceso de reconstrucción histórica son los mismos en la evolución molecular que en la anatomía comparada, la paleontología y otras disciplinas semejantes. El proceso fundamental consiste en interpretar el grado de diferencia en términos de distancias filogenéticas. Pero los estudios moleculares tienen una ventaja notable sobre la anatomía comparada y las otras disciplinas clásicas, de que la información es cuantificable; el número de elementos diferentes puede ser exactamente determinado al comparar las cadenas de ADN o de proteína entre dos especies, simplemente alineándolas y contando las diferencias (Ayala, 1987).

Ambas moléculas informativas (ADN y proteínas) proporcionan información no sólo sobre la topología de la evolución (las relaciones entre especies), sino también sobre el número de cambios que han ocurrido en un linaje particular (Lewin, 1999).

Los organismos ancestrales que vivieron hace tres mil millones de años o más tenían pocos genes. A partir de tan modestas cantidades de información genética han evolucionado miríadas de clases de organismos con estructuras

genéticas (genomas) que tienen centenares de millones y millares de millones de nucleótidos de longitud. El aumento en el número de genes desde los organismos más simples a los más complejos se hace evidente en los organismos actuales, por la diferencia que existe entre los procariotas y los eucariotas (Graur y Li, 2000).

A partir del desarrollo de las técnicas de ADN recombinante y de los métodos de secuenciación automática y bioinformática, se han abordado de manera sistemática los proyectos de secuenciación de genomas de diferentes especies procariotas y eucariotas que permiten determinar con gran precisión el tamaño de su genoma y el número de genes. En la actualidad pueden contabilizarse 56 genomas totalmente secuenciados y publicados (10 de arqueobacterias, 41 de bacterias y 5 de eucariotas: *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *A. thaliana* y *H. sapiens*). La bacteria *E. Coli* cepa K12 tiene un genoma de 4639 kb y aproximadamente 1000 genes, mientras que los eucariotes como *S. cerevisiae* presenta un genoma de 12 Mb con alrededor de 5800 genes; el genoma de *C. elegans* tiene 97 Mb y un número estimado de 19099 genes; *D. melanogaster* tiene 180 Mb y 13601 genes; *A. thaliana*, 125 Mb y alrededor de 25498 genes; en tanto que *H. sapiens* tiene un genoma de 3000 Mb y entre 31780 y 39114 genes (Monfort *et al.*, 2002).

Los nuevos genes y otras secuencias definidas de ADN pueden aparecer a través de una gama de procesos como son la recombinación, duplicación, delección, sustitución, inserción o inversión de regiones específicas de ADN. Estos procesos están integrados uno dentro del siguiente. El proceso de elongación genera genes sencillos por repetición una tras otra de unidades constituidas por sólo unos pocos nucleótidos (a menudo entre nueve y doce), seguido por una divergencia gradual de las unidades repetidas originalmente idénticas. Los genes originados independientemente por elongación se combinan entre ellos para dar lugar a genes mayores con funciones complejas. El total de estos genes e incluso familias de genes puede duplicarse (Lewin, 1999).

Zuckerlandl y Pauling (1962, 1965) fueron los primeros en proponer que el ritmo de sustituciones de aminoácidos y nucleótidos en la evolución de un organismo podría ser constante en el tiempo. En tal caso las macromoléculas de información podrían constituir un reloj molecular de la evolución. Esta notable hipótesis ha revolucionado la reconstrucción de la historia evolutiva y la datación de los sucesos evolutivos. El razonamiento de los dos autores es que muchas de las sustituciones en la secuencia de aminoácidos pueden tener consecuencias funcionales relativamente pequeñas, de manera que el número de cambios observados en la secuencia de aminoácidos de una proteína pue-

da ser aproximadamente proporcional al tiempo evolutivo transcurrido.

Esta propuesta fue establecida junto con otras que serían tema dominante en los estudios de evolución molecular durante las dos décadas siguientes: 1) Que el ritmo de sustitución en la secuencia de aminoácidos de una proteína dada sería directamente proporcional al número de posiciones que pueden experimentar sustituciones sin causar una alteración radical de la función; 2) Que el ritmo de la evolución morfológica se debería en gran parte a los cambios en la regulación génica y no se reflejaría en el ritmo de evolución de la mayoría de las cadenas polipeptídicas, y 3) Como las sustituciones de aminoácidos funcionalmente importantes son bastante raras, la frecuencia de las sustituciones evolutivamente significativas durante los periodos de evolución morfológica rápida pueden no diferir en grado notorio de la frecuencia que se da durante los periodos de evolución lenta (Bryson y Vogel, 1965).

La teoría neutralista de la evolución de las proteínas propone que los ritmos de sustitución de nucleótidos en el ADN -y por tanto, de reemplazo de aminoácidos en las proteínas- pueden ser aproximadamente constantes porque la mayoría de esos cambios son relativamente neutros (Kimura, 1987).

Los osmoprotectores (también llamados solutos compatibles) ocurren en muchos organismos, desde las bacterias hasta las plantas superiores y los animales. Son compuestos altamente solubles que no transportan carga neta a pH fisiológico y no son tóxicos a concentraciones altas. Sirven para incrementar la presión osmótica en el citoplasma y pueden también estabilizar las proteínas y las membranas de las células cuando los niveles de sales o la temperatura del ambiente son desfavorables. Los osmoprotectores juegan, por tanto, un papel muy importante en la adaptación de las células a varias condiciones ambientales adversas.

La manipulación metabólica de la acumulación de osmoprotectores ha atraído el interés de los ingenieros genéticos como una forma para mejorar la resistencia al estrés de los cultivos. Uno de los objetivos fundamentales es la ingeniería de la vía de síntesis de la glicina betaína, un potente osmoprotector que ocurre ampliamente entre las plantas, que incluyen algunas especies pertenecientes a familias económicamente importantes como Chenopodiaceae, Amaranthaceae, Brassicaceae, Acanthaceae, Malvaceae, Compositae y Poaceae.

Las plantas sintetizan glicina betaína vía la oxidación de la colina, en dos pasos. El primer paso es catalizado por una colina monooxigenasa (CMO), una enzima dependiente de ferredoxina que produce la forma hidratada de la

betaína-aldehído. El segundo paso es mediado por la enzima betaina aldehído deshidrogenasa (BADH), dependiente del cofactor  $\text{NAD}^+$  (Brouquisse *et al.*, 1989; Burnet *et al.*, 1995). En *E. coli* el primer paso es catalizado por la colina deshidrogenasa (CDH) y la conversión a glicina betaína es realizada tanto por la CDH como por una BADH (Landfald y Strom, 1986; Lamark *et al.*, 1991).

Existen reportes del aislamiento de genes que codifican para la BADH en *E. coli* (Boyd *et al.*, 1991), arroz (*Oryza sativa* L.) (Nakamura *et al.*, 1997), espinaca (*Spinacea oleraceae* L.) (Shu *et al.*, Institute of Genetics, Academia Sínica, Plant Biotechnology Laboratory, Beijing, China) y amaranto (Legaria *et al.*, 1998) y los ADN complementarios (ADNc) de espinaca (Weretilnyk y Hanson, 1990), remolacha (*Beta vulgaris* L.) (McCue y Hanson, 1992), cebada (*Hordeum vulgare* L.) (Ishitani *et al.*, 1995; Nakamura y Takabe, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya Japan), sorgo (*Sorghum bicolor* L. Moench.) (Wood *et al.*, 1996), *Atriplex hortensis* (Xiao *et al.*, Institute of Genetics, Academia Sínica, Plant Biotechnology Laboratory, Beijing China), *Atriplex prostrata* (Wang y Showalter, Ohio University, Athens USA), *Atriplex centralasiática* (Yin *et al.*, Institute of Genetics, Academia Sínica, Plant Biotechnology Laboratory, Beijing China), *Avicennia marina* (Hibino *et al.*, 2001), *Arabidopsis thaliana* (Theologis *et al.*, Plant Gene Expression Center/USDA-U.C Berkeley, California USA; Town *et al.*, The Institute of Genomic Research, Medical Center, Rockville USA), amaranto (*Amaranthus hypochondriacus* L.) (Legaria *et al.*, 1998), trigo (Li *et al.*, College of Life Sciences, Shandong Agricultural University, Shandong China) *Suaeda liaotungensis* (Li *et al.*, Bioengineering Department, Dalian University of Technology, Liaoning China) y el ADNc parcial de arroz (Nakamura *et al.*, 1997). En dichas plantas la actividad de la enzima y la transcripción son inducidos por estrés hídrico, salinidad o frío, coincidente con un incremento en los niveles de glicina betaína.

Durante la evolución o durante la colonización de nuevos nichos ecológicos las plantas y otros organismos han estado sujetos a la presión de selección debida al estrés en un ambiente cambiante. Para contrarrestar los efectos del estrés osmótico (escasez de agua, salinidad de los suelos o bajas temperaturas) algunas especies vegetales han desarrollado estrategias adaptativas, como la acumulación de solutos compatibles.

El objetivo del presente trabajo fue estudiar la evolución molecular de la vía de síntesis de la glicina betaína y de los genes que codifican las enzimas que en ella participan, lo que permitiría definir estrategias para transferir los genes a especies de interés agronómico que carecen del

osmoprotector y así conferirles tolerancia a sequía, salinidad y frío.

## MATERIALES Y MÉTODOS

Se hizo una búsqueda de secuencias reportadas para la proteína BADH en la literatura, y en las bases de datos del Genebank (<http://www.ncbi.nlm.nih.gov/GenBank>) y del Swiss-prot/trEMBL (<http://www.ebi.ac.uk/swissprot/>). De ella se seleccionaron 18 secuencias porque incluyeron a todas las especies reportadas, y por su localización intracelular. En el Cuadro 1 se indican el organismo, el número de acceso y los autores.

Para determinar el número de aminoácidos sustituidos y el porcentaje de identidad entre las diferentes secuencias BADH, se realizaron alineamientos con el programa FASTA (<http://www.ebi.ac.uk>) y para construir el dendrograma correspondiente se hizo un alineamiento múltiple de las secuencias de las diferentes especies mediante los programas de cómputo ClustalX y ClustalW (<http://www.ebi.ac.uk>), que integran métodos de alineamiento progresivo múltiple de secuencias. Ambos programas se basan en el cálculo de distancias (porcentaje de divergencia) entre todos los pares de secuencias de un alineamiento múltiple y luego aplican el método NJ (Neighbor-Joining, por unión al vecino más cercano) para construir una matriz de distancias. Finalmente, los dos programas pueden realizar un análisis para estimar la robustez del árbol generado (Bootstrapping) (Mount, 2001). La diferencia entre los dos métodos consiste en que ClustalW asigna un peso (o valor) determinado a cada alineamiento de pares de secuencias, mientras que ClustalX no lo hace.

## RESULTADOS Y DISCUSIÓN

La secuencia deducida de aminoácidos de la proteína AHY4 de *Amaranthus hypochondriacus* L. mostró 12 sustituciones de aminoácidos con respecto a AHY17, 297 sustituciones en relación con la BADH de *E. coli* y 101, 110, 91, 87, 120, 82, 82, 82, 119, 104, 150, 196, 154, 137 y 136 sustituciones de aminoácidos con proteínas BADH de *Arabidopsis thaliana*, *Atriplex hortensis*, *Atriplex prostrata*, *Atriplex centralasiática*, espinaca, remolacha, *Suaeda liaotungensis*, *Avicennia marina*, arroz, sorgo, cebada, y trigo, respectivamente (Cuadro 2).

La identidad de tales secuencias BADH indica una divergencia relativa entre plantas dicotiledóneas y monocotiledóneas, y probablemente un origen evolutivo común de las enzimas de plantas y de bacterias. Wolfe *et al.* (1989) sugieren que las monocotiledóneas y dicotiledóneas divergieron desde hace 200 millones de años, aproximadamente. Las BADH se han reportado en bacterias y plantas

Cuadro 1. Organismo, número de acceso a la base de datos y autores de las secuencias de proteínas utilizadas para los alineamientos.

Organismo	Número de acceso	Autor
<i>E. coli</i> (ECOLI)	sp P17445, M77739	Boyd <i>et al.</i> (1991)
Arroz (ORYSA)	sp O24174, AB001348	Nakamura <i>et al.</i> (1997)
<i>A. hortensis</i> (ATRHO)	sp P42757, X69770	Xiao <i>et al.</i> ; Institute of Genetics, Academia Sínica, Beiging, China
<i>A. prostrata</i> (ATPRP)	tr Q8RX99, AY082068	Wang y Showalter; Ohio University, Athens, USA
<i>A. centralasiatica</i> (ATRCE)	tr Q8L811	Yin <i>et al.</i> ; Institute of Genetics, Beiging, China
<i>A. marina</i> (AVMAR1)	tr Q94LA0	Hibino <i>et al.</i> (2001)
(AVMAR2)	tr Q9FRY2	Hibino <i>et al.</i> (2001)
<i>A. thaliana</i> (ARATH1)	sp Q9S795	Theologis <i>et al.</i> Plant Gene Expression Ceter/USDA, Berkeley California, USA
(ARATH2)	tr Q9STS1	Town <i>et al.</i> ; The Institute of Genomic Research, Medical Center, Rockville, USA
Amaranto (AHY4)	tr 022477, AF000132	Legaria <i>et al.</i> (1998)
(AHY17)	sp 004895, AF017150	Legaria <i>et al.</i> (1998)
Cebada (HORVU1)	tr Q94 IC1, AB063178	Nakamura y Takabe; Graduate School Of Bioagricultural Sciences, Nagoya, University, Japan
(HORVU2)	sp Q40024, D26448	Ishitani <i>et al.</i> (1995)
Espinaca (SPIOL)	sp P17202, M31480	Weretilnyk Hanson (1990)
Remolacha (BETVU)	sp P28237, X58463	McCue y Hanson (1992)
Sorgo (SORBI)	tr Q43829, U12196	Wood <i>et al.</i> (1996)
<i>S. liaotungensis</i> (SUALI)	tr Q8W5A1	Li <i>et al.</i> ; Bioengineering Department, Dalian University of Technology, China
Trigo (TRIAE)	tr Q8LGQ9	Li <i>et al.</i> ; College of Life Sciences, Shadong Agricultural University, China

pero no en animales y levaduras; en estos últimos organismos se han encontrado proteínas relacionadas conocidas como Aldehído Deshidrogenasas (ALDHs). Las ALDHs representan un grupo de enzimas capaces de oxidar aldehídos tanto aromáticos como alifáticos, involucrados en diversos procesos fisiológicos. Son altamente inespecíficas para un sustrato, y se distinguen tres prototipos: ALDHs citosólicas dependientes de  $\text{NAD}^+$ , formas mitocondriales codificadas por genes nucleares, y enzimas inducibles dependientes de  $\text{NADP}^+$  (Boyd *et al.*, 1991).

En un análisis comparativo entre secuencias BADHs y ALDHs de diferentes organismos Boyd *et al.* (1991) y Habenicht *et al.* (1994), sugirieron que ambas proteínas tienen un origen ancestral común debido a que muestran similitudes tanto estructurales como funcionales, cuyo origen se debe a duplicaciones en un gene ocurridas antes o poco después de la separación de las eubacterias y los eu-

cariotes. Los mismos autores proponen que la proteína ALDH ancestral era una enzima no especializada que utilizaba múltiples aldehídos como sustrato. Este tipo de enzimas ha sobrevivido en organismos animales y levaduras, pero en las plantas y las bacterias han evolucionado ALDHs un tanto sustrato específicas, como las BADH.

Probablemente la especificidad de sustrato de las BADH se originó dos veces durante la evolución, pues dio como resultado las BADH de bacterias y las BADH de las plantas superiores (Habenicht, 1994). El concepto de evolución de la especificidad de sustrato de las BADHs no está muy claro, pues recientemente Vojtechová *et al.* (1997) y Trossat *et al.* (1997) demostraron que la BADH de amaranto y de remolacha tienen alta afinidad por el sustrato betaína aldehído así como por aldehídos con carga positiva, y muestran actividad de aldehído deshidrogenasas,

Cuadro 2. Número de aminoácidos sustituidos entre las secuencias BADH de diferentes especies. *Arabidopsis thaliana* (ARATH1 y ARATH2); *Atriplex hortensis* (ATRHO); *Atriplex prostrata* (ATRPR); *Atriplex centralasiática* (ATRCE); *espinaca* (SPIOL); *remolacha* (BETVU); *Suaeda liaotungensis* (SUALI); *amaranto* (AHY4 y AHY17); *Avicennia marina* (AVMAR1 y AVMAR2); *arroz* (ORYSA); *sorgo* (SORBI); *cebada* (HORVU1 y HORVU2); *trigo* (TRIAE); *E. Coli* (ECOLI).

	ARATH1	ARATH2	ATRHO	ATRPR	ATRCE	SPIOL	BETVU	SUALI	AHY4	AHY17	AVMAR1	AVMAR2	ORYSA	SORBI	HORVU2	HORVU1	TRIAE
ARATH2	105																
ATRHO	129	147															
ATRPR	123	140	16														
ATRCE	121	139	23	11													
SPIOL	119	139	53	45	49												
BETVU	120	142	60	54	44	49											
SUALI	120	141	60	54	50	57	56										
AHY4	101	110	91	87	120	82	82	82									
AHY17	102	116	93	85	85	82	86	82	12								
AVMAR1	122	138	138	130	130	128	138	125	119	118							
AVMAR2	113	113	136	129	131	129	139	119	104	110	80						
ORYSA	133	109	154	148	152	145	155	156	150	146	156	147					
SORBI	182	178	203	189	194	198	204	201	196	190	197	196	143				
HORVU2	150	149	155	151	157	152	159	166	154	148	174	160	89	148			
HORVU1	127	141	148	143	145	146	144	145	137	142	153	150	140	192	156		
TRIAE	128	142	150	145	147	145	146	146	136	144	154	150	136	191	152	7	
ECOLI	324	327	315	318	309	300	305	322	297	297	320	319	304	348	325	317	316

mientras que en *Avicennia marina* las enzimas son específicas para el sustrato betaina aldehído (Hibino *et al.*, 2001). El análisis filogenético de la superfamilia de las ALDHs reveló un complejo patrón evolutivo, con numerosas ramas grandes que incluyen genes de eubacterias y eucariotas, o ambos, que codifican enzimas específicas o no específicas para un sustrato aldehído particular. Dicha topología sugiere una emergencia concomitante de múltiples especificidades de sustrato de ALDHs no especializadas, durante una fase evolutiva temprana de intensa diversificación metabólica (Habenicht, 1994).

A juzgar por la relativa conservación de las secuencias, diríase que la BADH es una proteína que evoluciona rápidamente. Todas las especies estudiadas en este trabajo comparten el mismo residuo en sólo 106 de las 489 posiciones comunes a todas las secuencias. Este grado de homología (22%) es característico de proteínas que evolucionan muy rápidamente (Ayala, 1994).

En la Figura 1 se muestra un dendograma de relaciones entre proteínas BADH de especies pertenecientes a las familias Chenopodiaceae, Amaranthaceae, Brassicaceae, Acanthaceae y Poaceae. Las secuencias formaron cuatro grupos bien definidos. Un primer grupo lo constituyeron las BADHs correspondientes a especies de la familia Poaceae (monocotiledónea): cebada (HORVU2), sorgo (SORBI) y arroz (ORYSA), en cuyo extremo carboxilo-terminal portan el tripéptido SKL (Figura 1, Cuadro 3), una señal que dirige preproteínas a peroxisomas. El segundo grupo, muy relacionado con el primero, quedó integrado por secuencias BADH también de especies de la familia Poaceae

como trigo (TRIAE) y cebada (HORVU1), que carecen de la señal SKL y son presumiblemente cloroplásticas (Figura 1, Cuadro 3). El tercer grupo lo conformaron secuencias correspondientes a *Arabidopsis thaliana*, una dicotiledónea de la familia Brassicaceae. Dentro de este grupo la proteína ARATH2 lleva la señal SKL y ARATH1 es un precursor cloroplástico. El cuarto grupo lo integraron las especies dicotiledóneas pertenecientes a las familias Chenopodiaceae (ATRHO, ATRPR, ATRCE, SPIOL, BETVU), Amaranthaceae (SUALI, AHY4, AHY17) y Acanthaceae (AVMAR1, AVMAR2). La mayoría de las BADH de este grupo son presumiblemente cloroplásticas, salvo AVMAR2 que porta la señal SKL (Figura 1, Cuadro 3).

El dendograma establece una clara divergencia entre las BADH de especies monocotiledóneas y las BADH de especies dicotiledóneas. También indica que la divergencia es mayor entre especies, géneros y familias que dentro de géneros y dentro de especies. Dentro de las especies monocotiledóneas las BADHs que portan la señal SKL (presumiblemente con actividad en los peroxisomas), agruparon a mayores distancias que las correspondientes cloroplásticas, lo que sugiere una mayor diversificación en las primeras. Además, las proteínas BADH de especies monocotiledóneas, en general, agruparon a mayores distancias en relación con las distancias que agruparon las BADH de especies o de familias de dicotiledóneas, lo que indica una mayor tasa de cambio, divergencia o evolución en las primeras.

El número de aminoácidos sustituidos entre las dos BADHs de amaranto (12 aminoácidos), entre las tres

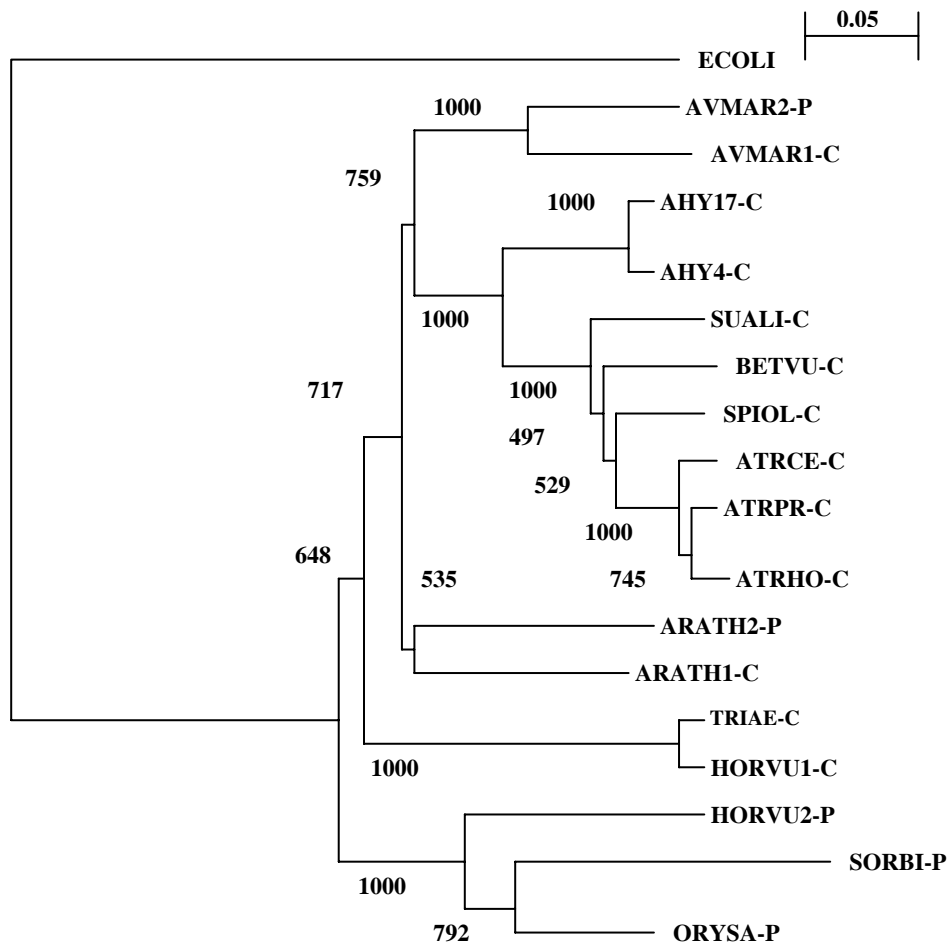


Figura 1. Dendrograma de relaciones entre proteínas BADH. La comparación se realizó con base en los alineamientos de secuencias de aminoácidos obtenidos mediante el programa de cómputo ClustalX y ClustalW (<http://www.ebi.ac.uk>). Las secuencias de aminoácidos que se incluyeron en la comparación son de las monocotiledóneas (HORVU1 y HORVU2, cebada; ORYSA, arroz; SORBI, sorgo y TRIAE, trigo), de las dicotiledóneas (ARATH1 y ARATH2, Arabidopsis thaliana; ATRHO, Atriplex hortensis; ATRPR, Atriplex prostrata; ATRCE, Atriplex centralasiática; SPIOL, espinaca; BETVU, remolacha; SUALI, Suaeda liaotungensis; AHY4 y AHY17, amaranto y AVMAR1 y AVMAR2, Avicennia marina) y de la bacteria (ECOLI, E. coli). La longitud de la barra de la escala de distancias genéticas corresponde a 0.05 de divergencia entre las secuencias (el número de posiciones no apareadas en un alineamiento dividido por el número total de posiciones apareadas). Los números en las ramas del árbol indican el número de veces en que la topología de una rama particular se repite durante el análisis de robustez del árbol. (C) precursor cloroplástico, (P) con la señal SKL en el extremo carboxilo-terminal.

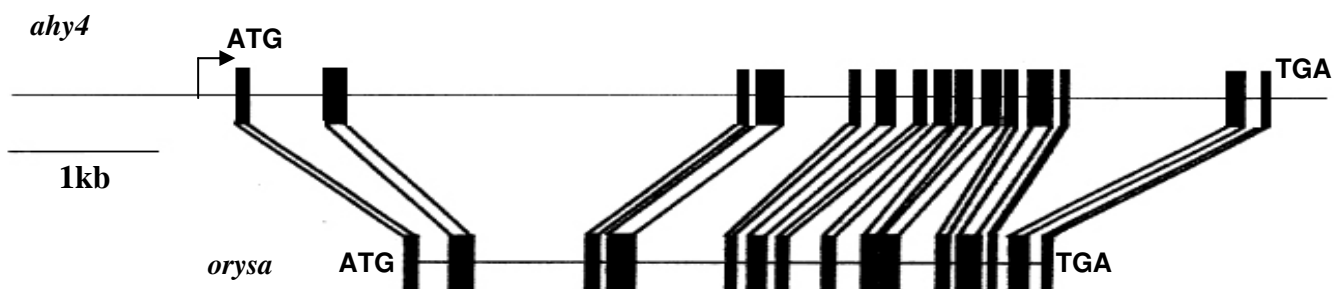


Figura 2. Comparación de la estructura de los genes que codifican para la BADH de amaranto (ahy4) y arroz (oryza). Las barras negras verticales denotan los exones y la línea horizontal los intrones y secuencias de los extremos 5' y 3'.

BADHs de *Atriplex* (11-23 aminoácidos) y entre las BADHs TRIAE y HORVU1 (7 aminoácidos) de trigo y cebada, respectivamente (todas precursoras cloroplásticas), fue muy bajo en comparación a las de otras especies, lo que hace suponer que la evolución de dichas enzimas constituye un evento relativamente reciente. Finalmente, las secuencias de aminoácidos de las BADHs de *Amaranthus hypochondriacus* están más relacionadas con BADHs de las familias Amaranthaceae, Chenopodiaceae y Acanthaceae que con la Poaceae, como se muestra en el dendrograma (Figura 1), lo que está de acuerdo con las relaciones filogenéticas.

Nakamura *et al.* (1997) reportaron la clonación y secuencia del gene que codifica para la BADH de arroz. El gene contiene 14 intrones y la secuencia de nucleótidos de la región codificadora mostró ser altamente similar a la BADH de otras plantas. En el presente trabajo, al comparar las secuencias genómicas entre amaranto (*ahy4*) y arroz (*orysa*) se detectó que la posición de los intrones está conservada en ambas especies, aunque su tamaño y secuencia es diferente (Figura 2); esto es explicable dada la falta de presión de selección sobre los intrones, como ya se ha observado en numerosos casos (Graur y Li, 2000).

La longitud del gene *ahy4* de amaranto es de 8998 pb mientras que la longitud del gene *orysa* de arroz es de 4428 pb, que representa sólo 49.21 % de la secuencia del primero. Dicha diferencia se debe fundamentalmente a que algunos de los intrones del gene de arroz son más cortos que los de amaranto. En amaranto la longitud de los intrones varía de 75 a 2723 pb, y comprenden 5624 pb de la secuencia total, mientras que en el gene *orysa* la longitud de los intrones comprende sólo 2900 pb del total de su secuencia. Al comparar la secuencia codificadora de *ahy4* con la secuencia del gen *oryza* se detectó que la estructura de los dos genes es semejante; ambos presentan 15 exones de igual tamaño con los sitios de ruptura, que corresponden exactamente en las mismas posiciones, si bien los 14 intrones de *orysa* son más cortos (Figura 2). Ello sugiere que los genes BADH de plantas dicotiledóneas y monocotiledóneas descienden de un gene ancestral común pero que divergieron en algún punto del proceso evolutivo hace aproximadamente 200 millones de años (Wolfe *et al.*, 1989).

El decapeptido VTLELGKSP y los residuos circundantes (Figura 3) presentes en AHY4 y AHY17, están altamente conservados entre las aldehído deshidrogenasas y las BADHs, y se consideran parte fundamental del sitio activo de la enzima y sitios de unión al cofactor NAD<sup>+</sup> (Weretilnyk y Hanson, 1990; Boyd *et al.*, 1991; McCue y Hanson, 1992; Wood *et al.*, 1996). Su presencia en AHY4 y AHY17 sugiere que se trata de enzimas funcionales.

Tanto el residuo glutámico del decapeptido como la cisteína localizada a 34 residuos, han sido implicados en la catálisis de las aldehído deshidrogenasas de mamíferos (Pietruszko, 1989). Se ha propuesto que las secuencias alrededor del residuo cisteína confieren especificidad de sustrato. En enzimas aldehído deshidrogenasas se ha propuesto la existencia del residuo cisteína en el sitio activo y se considera esencial para su actividad, ya que el grupo SH es el responsable de la oxidación del aldehído (Tu-G-C y Weiner, 1988; Kitson *et al.*, 1991) para formar un tiohemiacetal intermediario (Jacobi, 1963). Valenzuela-Soto (Comunicación Personal, 1994) consideró de interés determinar si este aminoácido está involucrado en la actividad catalítica de la BADH de amaranto; encontró que el ácido 5-5'-ditio-bis-2-nitrobenzoico (DTNB), un reactivo específico para grupos sulfhidrilos, fue capaz de inactivar a la enzima. Esta inactivación se caracterizó y se encontró que la BADH de hojas de amaranto tiene residuos cisteína esenciales para su actividad.

Vojtechová *et al.* (1997) y Trossat *et al.* (1997) demostraron que la BADH de amaranto y de remolacha muestran alta afinidad por el sustrato betaína aldehído, así como por aldehídos con carga positiva, al mostrar actividad de aldehído deshidrogenasas. Dado que en amaranto se detectaron tres secuencias relacionadas con la BADH, porque codifican al menos dos de ellas para un polipéptido de peso molecular semejante (Legaria *et al.*, 1998), es posible que alguna de las isoenzimas muestre especificidad por el sustrato betaína aldehído y que alguna de las isoenzimas restantes tenga además afinidad por los aldehídos. La estructura primaria de la BADH permitiría identificar residuos blanco para modificación por mutagénesis sitio-dirigida al probar esta idea.

La biosíntesis de la glicina betaína ocurre en los cloroplastos de la familia Chenopodiaceae (Hanson *et al.*, 1985). La BADH está codificada por un gene nuclear (Weretilnyk y Hanson, 1988) y la enzima se localiza en el cloroplasto (Weigel *et al.*, 1986). Rathinasabapathi *et al.* (1994) también demostraron que plantas transgénicas de tabaco que expresan los ADNc de espinaca y remolacha, producen una BADH cloroplástica. Las secuencias deducidas de las BADHs de espinaca y remolacha contienen una región de aminoácidos **QLFIDGE** en los residuos 9 a 15. La secuenciación directa de la proteína madura extraída de espinaca mostró que este péptido corresponde a su extremo amino-terminal (Weretilnyk y Hanson, 1990). Esto indica que la BADH cloroplástica tiene un péptido de tránsito atípico de 8 aminoácidos, porque los péptidos de tránsito para dirigir las proteínas al cloroplasto son de 30 a 50 aminoácidos (Keegstra *et al.*, 1989; Berry-Lowe y Schmidt, 1991).

<b>ARATH1</b>	PSLAAGCTAILKPSELASVTCLELADICREVGLPPGVNLNLTGFGSEAGAPLASHPGVDKIAF	<b>236</b>
<b>ARATH2</b>	PSLAAGCTAILKPSELASLTCLADICREVGLPPGVNLNLTGLGTEAGAPLASHPHVDKIVF	<b>236</b>
<b>ATRHO</b>	PALAAGCTTVLKPSELASVTCLFEGFVCNEVGLPPGVNLNLTGLGPDAGAPIVSHPDIDKVAF	<b>238</b>
<b>ATRPR</b>	PALAAGCTAVLKPSELASVTCLFEGFVCNEVGLPPGVNLNLTGLGPDAGAPIVSHPDIDKIAF	<b>236</b>
<b>ATRCE</b>	PALAAGCTAVLKPSELASVTCLFEGFVCNEVGLPPGVNLNLTGLGPDAGAPIVSHPDVDKIAF	<b>236</b>
<b>SPIOL</b>	PALAAGCTAVLKPSELASVTCLFEGFVCNEVGLPPGVNLNLTGLGPDAGAPLVSHPDVDKIAF	<b>233</b>
<b>BETVU</b>	PALAAGCTAVLKPSELASITCLFEGFVCNEVGLPPGVNLNLTGLGPDAGAPLAHPDVKVAF	<b>236</b>
<b>SUALI</b>	PALAAGCTAVLKPSELASVTCLFEGFVCNEVGLPPGVNLNLTGLGPDAGAPLVSHPDVDKVAF	<b>236</b>
<b>AHY4</b>	PALAAGCSAVLKPSELASVTCLAEVCREVGLPPGVNLNLTGLGPEAGGPLACHPDVDKVAF	<b>236</b>
<b>AHY17</b>	PALAAGCSAVLKPSELASVTCLAEVCREVGLPPGVNLNLTGLGPEAGGPLACHPDVDKVAF	<b>236</b>
<b>AVMAR1</b>	PALASGCTAILKPSELASVTCLAEVCMVGLPPGVNLNLTGLGPEAGAPLVTHPHVAKISF	<b>232</b>
<b>AVMAR2</b>	PALAAGCTAILKPSELASVTCLAEVQVCKEVLPPGVNLNLTGLGPEAGAPLASHPHVDKITF	<b>235</b>
<b>ORYSA</b>	PALAAGCTAVLKPSELASLTCLGGAICAEIGLPPGVNLNLTGLGTEAGAPLASHPHVDKIAF	<b>238</b>
<b>SORBI</b>	PALAAGCTAVLKPSELASVSCLELGAICMEIGLPPGVNLNLTGLGLKLVLPHYPHIPCGIRLL	<b>229</b>
<b>HORVU2</b>	PALAAGCTAVLKPSELASLTCLGGAICEEIGLPSGVNLNLTGLGPDAGAPIASHPHVDKIAF	<b>237</b>
<b>HORVU1</b>	PALAAGCTAVLKPSELASVTCLGVDVCKEVLPSGVNLNLTGLGNEAGAPLSSHPDVKVAF	<b>237</b>
<b>TRIAE</b>	PALAAGCTAVLKPSELASVTCLGVDVCKEIGLPSGVNLNLTGLGHEAGAPLSSHPDVKVAF	<b>237</b>
<b>ECOLI</b>	PALAAGNAMI FKPSEVTPLTALKLAETISEAGLPDGVFNVLPGVGAETGQYLTEHPGIAKVSF	<b>226</b>
<b>ARATH1</b>	TGSFATGSKVMT-AAAQLVKPVSMELGGKSPDIVFDD-VLDKAAEWALFGCFWNTNGQICSATS	<b>298</b>
<b>ARATH2</b>	TGSTTTGSSIMT-SAAQLVKPVSELEGGKSPDIVFDD-VDIDKAVEWTFMFGCFWNTNGQICSATS	<b>298</b>
<b>ATRHO</b>	TGSSATGSKIMA-SAAQLVKPVTTLEGGKSPVIMFED-IDIETAVEWTLFGVFWNTNGQICSATS	<b>300</b>
<b>ATRPR</b>	TGSSATGSKIMA-SAAQLVKPVTTLEGGKSPVIMFED-IDIETAVEWTLFGVFWNTNGQICSATS	<b>298</b>
<b>ATRCE</b>	TGSSATGSKIMA-SAAQLVKPVTTLEGGKSPVIMFED-IETVVAVEWTLFGVFWNTNGQICSATS	<b>298</b>
<b>SPIOL</b>	TGSSATGSKVMA-SAAQLVKPVTTLEGGKSPVIMFED-VDIDKVVWETIFGCFWNTNGQICSATS	<b>295</b>
<b>BETVU</b>	TGSSATGSKVMA-SAAQLVKPVTTLEGGKSPVIMFED-VDIDQVVEWTFMFGCFWNTNGQICSATS	<b>298</b>
<b>SUALI</b>	TGSSATGSKVMG-SAAQLVKPVTTLEGGKSPDIVFDDVVDLDVAAEWTFMFGCFWNTNGQICSATS	<b>299</b>
<b>AHY4</b>	TGSTATGSKVMS-SAAQLVKPVTTLEGGKSPDIVFED-VLDKAAEWTFMFGCFWNTNGQICSATS	<b>298</b>
<b>AHY17</b>	TGSTATGSKVMS-SAAQLVKPVTTLEGGKSPDIVFED-VLDKAAEWTFMFGCFWNTNGQICSATS	<b>298</b>
<b>AVMAR1</b>	TGSDTTGVKIMT-AAAQLVKPVTTLEGGKSPVIMFED-VLDTAAEWTLFGCFWNTNGQICSATS	<b>294</b>
<b>AVMAR2</b>	TGSGATGSKIMT-AAAQLVKPVTTLEGGKSPVIMFED-VLDTAAEWTLFGCFWNTNGQICSATS	<b>297</b>
<b>ORYSA</b>	TGSTETGKRIMI-TASQMVKPVSELEGGKSPDIVFDDV-DIDKAVEWAMFGCFANAGQVCSATS	<b>300</b>
<b>SORBI</b>	LGSTETGKRIMI-TASQMVKPVSELEGGKSPDIVFDDIRIDKAVEWTFMFGILPNAGQVCSAAS	<b>292</b>
<b>HORVU2</b>	TGSTATGKTIMT-AAAQMVKPVSELEGGKSPVITFDDVADIDKAVEWPMGLGCFNNGQVCSATS	<b>300</b>
<b>HORVU1</b>	TGSYATGQKIMV-AAAPTVPVTTLEGGKSPVIMFDDV-DIDKAVEWTLFGCFWNTNGQICSATS	<b>299</b>
<b>TRIAE</b>	TGSYATGQKIMV-AAAPTVPVTTLEGGKSPVIMFDDV-DIDKAVEWTLFGCFWNTNGQICSATS	<b>299</b>
<b>ECOLI</b>	TGGVASGKKVMANSAASSLKEVTMELGGKSPDIVFDD-ADLDLAADIAMMANFSSGQVCNTGT	<b>289</b>

Figura 3. Posible sitio activo de la enzima BADH de diferentes especies. El decapeptido subrayado se considera el sitio activo de la enzima, se encuentra muy conservado entre las BADHs de plantas (ARATH1 y ARATH2, *Arabidopsis thaliana*; ATRHO, *Atriplex hortensis*; ATRPR, *Atriplex prostrata*; ATRCE, *Atriplex centralasiática*; SPIOL, espinaca; BETVU, remolacha; SUALI, *Suaeda liaotungensis*; AHY4 y AHY17, amaranto; AVMAR1 y AVMAR2, *Avicennia marina*; ORYSA, arroz; SORBI, sorgo; HORVU1 y HORVU2, cebada; TRIAE, trigo) y de microorganismos (ECOLI, *E. coli*). Los aminoácidos circundantes subrayados posiblemente están involucrados en la unión con el NADH.

La secuencia de aminoácidos deducida del gene *ahy4* y del ADNc *ahy17*, así como las secuencias BADH de remolacha, *Atriplex hortensis*, *Atriplex prostrata*, *Atriplex centralasiática*, *Suaeda liaotungensis*, trigo y secuencias de *Avicennia marina*, *Arabidopsis thaliana* y cebada mostraron la presencia de la secuencia conservada **QLFIDGE** (Cuadro 3), lo que sugiere una localización celular semejante a la de la BADH de espinaca.

Se ha reportado que la proteína ORYSA de arroz se localiza en los peroxisomas de la célula, dirigida probablemente por la secuencia carboxi-terminal **SKL** altamente conservada (Nakamura *et al.*, 1997). Este tripéptido está presente en algunas proteínas BADH reportadas tanto de monocotiledóneas (ORYSA, SORBI, HORVU2) como de dicotiledóneas (ARATH2, AVMAR2) (Cuadro 3).



Cuadro 3. Comparación de posibles péptidos de tránsito para dirigir proteínas BADH al cloroplasto, y de secuencias C-terminal de proteínas BADH de diferentes géneros. Las secuencias deducidas de aminoácidos del extremo N-terminal son comparadas con base en el sitio de procesamiento (RQ) y secuencia N-terminal de la proteína madura (subrayado) de la BADH de espinaca (Weretilnyk y Hanson, 1990). En las secuencias C-terminal se resalta el tripéptido SKL, una señal que dirige preproteínas a peroxisomas. En las referencias, tr y sp significan que las secuencias se obtuvieron de las bases de datos TrEMBL y SWISS-PROT, respectivamente.

Género	Secuencia N-terminal	Secuencia C-terminal	Referencia
<i>Amaranthus</i>	MAIRVPS <b>RQ</b> LFI DGEW	KSP	Legaria et al. (1998)
	MAIRVPS <b>RQ</b> LFI DGEW	KSP	Legaria et al. (1998)
<i>Beta</i>	MSMPIPS <b>RQ</b> LFI DGEW	KSP	McCue y Hanson, 1992
<i>Spinacia</i>	MAFP <b>IPARQ</b> LFI DGEW	KSP	Weretilnyk y Hanson, 1990
<i>Atriplex</i>	MAFP <b>IPARQ</b> LFI DGEW	KSP	Genebank x69770
	MAFP <b>IPVRQ</b> LFI DGEW	KSP	tr   Q8RX99
	MAFPMP <b>VRQ</b> LFI DGEW	KSP	tr   Q8L8T1
<i>Avicennia</i>	MP <b>TRQ</b> LFI DGEW	KSP	tr   Q94LA0
	MAIRIPS <b>RQ</b> LFI DGEW	KSP <b>SKL</b>	tr   Q9FRY2
<i>Arabidopsis</i>	MAIPMP <b>TRQ</b> LLFI DGEW	KSPN	tr   Q9S795
	MAFP <b>IPVRQ</b> LFI DGEW	KPP <b>SKL</b>	tr   Q9STS1
<i>Suaeda</i>	MSIPIPS <b>RQ</b> LFI DGEW	KSP	tr   Q8W5A1
<i>Triticum</i>	MVAPAAIP <b>QRQ</b> LFI DGDW	KAPAN	tr   Q8LGQ9
<i>Hordeum</i>	MVAPAKIP <b>QRQ</b> LFI DGED	KAPAN	tr   Q94IC1
	MAAPPAIPRRGLFI GGGW	QRP <b>SKL</b>	Ishitani et al. (1995)
<i>Sorghum</i>	MAAADVPRPSFI GGDW	QPP <b>SKL</b>	Wood et al. (1996)
<i>Oryza</i>	MAAPSAIPRRGLFI GGGW	RPP <b>SKL</b>	Nakamura et al. (1997)

La información existente sugiere una diferenciación en la localización intracelular de las enzimas BADH en las plantas, aunque en todos los casos las enzimas son codificadas por genes nucleares. Dado que en los peroxisomas o en las mitocondrias se han detectado niveles mas altos de NAD<sup>+</sup> (el cofactor requerido por la BADH para su actividad enzimática) que en los cloroplastos, es posible que durante el proceso de evolución algunas proteínas BADH puedan haber cambiado el lugar de su actividad para desarrollar una catálisis mas eficiente (Arakawa et al., 1990).

En algunas especies sólo se han reportado secuencias posiblemente cloroplásticas o secuencias que poseen la señal SKL, por lo que cabe la posibilidad de que aquéllas que poseen actividad enzimática en ambos compartimentos sean ya intermedias o terminales en el proceso evolutivo (Hibino et al., 2001). En una misma especie, además, pueden existir formas alélicas que presenten actividad en un tejido específico, como lo demuestra el trabajo realizado por Onofre et al. (2003).

## CONCLUSIONES

Los resultados indican que las proteínas BADH de amaranto están más relacionadas filogenéticamente con aquéllas de plantas dicotiledóneas que con las monocotiledóneas. La identidad de las BADHs dentro de la especie *Amaranthus hypochondriacus* fue muy alta (98 %, con sólo 12 aminoácidos sustituidos), lo que sugiere que la evolución de dichas enzimas constituye un evento relativamente reciente. La estructura y la secuencia codificadora de los genes BADH de amaranto y de arroz se ha conservado a lo largo del proceso evolutivo; sin embargo, en el gen de arroz algunos intrones son más pequeños y la secuencia de los mismos es no homóloga a la del gene de amaranto. En las plantas superiores las proteínas BADH han divergido en relación al sitio en el que realizan su actividad dado que algunas se localizan en los cloroplastos y otras en los peroxisomas.

## BIBLIOGRAFÍA

- Arakawa K, M Katayama, T Takabe (1990) Levels of betaine and betaine aldehyde dehydrogenase in the green leaves and etiolated leaves and roots of barley. *Plant Cell Physiol.* 31: 797-803.
- Ayala F J (1987) La Naturaleza Inacabada. Biblioteca Científica Salvat. Barcelona, España. 270 p.

- Ayala F J (1994) El reloj molecular de la evolución. Ciencia y Desarrollo 119: 43-55.
- Berry-Lowe S L, G W Schmidt (1991) Chloroplast protein transport. In: Cell Culture and Somatic Cell Genetics of Plants. Vol 7 A : The Molecular Biology of Plastids. L Bogorad, I K Vasil (eds). Academic Press. New York. pp: 63-80.
- Boyd L A, L Adam, L E Pelcher, A McHughen, R HirjiR, G Selvaraj (1991) Characterization of an *Escherichia coli* gene encoding betaine aldehyde dehydrogenase (BADH): structural similarity to mammalian ALDHs and a plant BADH. Gene 103: 45-52.
- Brouquisse R, P Weigel, D Rhodes, C F Yocum, A D Hanson (1989) Evidence for a ferredoxin-dependent choline monooxygenase from spinach chloroplast stroma. Plant Physiol. 90: 322-329.
- Burnet M, P J Lafontaine, A D Hanson (1995) Assay, purification, and partial characterization of choline monooxygenase from spinach. Plant Physiol. 108: 581-588.
- Bryson V, H J Vogel (1965) Evolving Genes and Proteins. Academic Press. New York. pp: 97-116.
- Dobzhansky T (1970) Genética del Proceso Evolutivo. Ed. Textos Extemporáneos. México. 463p.
- Graur D, W-H Li (2000) Fundamentals of Molecular Evolution. Second Edition. Sinauer Associates Inc. USA. 481p.
- Habenicht A, U Hellman, R Cerff (1994) Non-phosphorylating GAPDH of higher plants is a member of the aldehyde dehydrogenase superfamily with no sequence homology to phosphorylating GAPDH. J. Mol. Biol. 237: 165-171.
- Hanson A D, A M May, R Grumet, J Bode, G C Jamieson, D Rhodes (1985) Betaine synthesis in chenopods : localization in chloroplasts. Proc. Natl. Acad. Sci. USA 82: 3678-3682.
- Hibino T, Y-L Meng, Y Kawamitsu, U Naoko, N Matsuda, Y Tanaka, H Ishikawa, S Baba, T Takabe, K Wada, T Ishii, T Takabe (2001) Molecular cloning and functional characterization of two kinds of betaine-aldehyde dehydrogenase in betaine-accumulating mangrove *Avicennia marina* (Forsk.) Vierh. Plant Mol. Biol. 45: 353-363.
- Ishitani M, T Nakamura, S Youn-Han, T Takabe (1995) Expression of the betaine aldehyde dehydrogenase gene in barley in response to osmotic stress and abscisic acid. Plant Mol. Biol. 27: 307-315.
- Jacobi W B (1963) Aldehyde dehydrogenase. In: The Enzymes. Boyer, Lardy and Myrbere (eds.). Academic Press. New York. Vol 7. pp: 203-221.
- Keegstra K, L J Olsen, S M Theg (1989) Chloroplastic precursors and their transport across the envelope membranes. Annu. Rev. Plant Physiol. Plant Mol. Biol. 40: 471-501.
- Kimura M (1987) Teoría neutralista de la evolución. En: Genética Molecular. Libros de Investigación y Ciencia. 232-241 pp.
- Kitson T M, J P Hill, G G Midwinter (1991) Identification of a catalytically essential residue in sheep liver cytoplasmic aldehyde dehydrogenase. Biochem. J. 275: 207-210.
- Lamark T, I Kaasen, M W Eshoo, P Falkenberg, J McDougall, A R Strom (1991) DNA sequence and analysis of the bet genes encoding the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. Mol. Microbiol. 5: 1049-1064.
- Landfald B, A R Strom (1986) Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. J. Bacteriol. 165: 849-855.
- Legaria J, R Rajsbaum, R A Muñoz-Clares, N Villegas-Sepúlveda, J Simpson, G Iturriaga (1998) Molecular characterization of two genes encoding betaine aldehyde dehydrogenase from amaranth. Expression in leaves under short-term exposure to osmotic stress or abscisic acid. Gene 218: 69-76.
- Lewin R (1999) Patterns in Evolution. The New Molecular View. Scientific American Library. New York. pp: 19-119.
- Li Y, J Cui, B Li, Q Zuo (2001) College of Life Sciences, Shandong Agricultural University, Shandong, China. European Molecular Biology Laboratory (EMBL) entry, accession no. [Q8LGQ9](#).
- Li Q L, X R Gao, X D Yuan, W L Li, Q Fan, L J An (2001) Bioengineering Department, Dalian University of Technology, Liaoning, China. European Molecular Biology Laboratory (EMBL) entry, accession no. [Q8W5A1](#).
- McCue K F, A D Hanson (1992) Salt-inducible betaine aldehyde dehydrogenase from sugar beet: cDNA cloning and expression. Plant Mol. Biol. 18: 1-11.
- Montfort A, J García-Mas, P Puigdoménech, P Arús (2002) Estructura del genoma. En: Genómica y Mejora Vegetal, F.Nuez, J.M. Carrillo y R. Lozano (eds.). Sociedad Española de Ciencias Hortícolas. Sociedad Española de Genética. Mundi-Prensa Libros, SA. Consejería de Agricultura y Pesca de la Junta de Andalucía. Pp: 67-103.
- Mount D W (2001) Bioinformatics. Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press. New York. pp: 139-280.
- Nakamura T, S Yokota, Y Muramoto, K Tsutsui, Y Oguri, K Fukui, T Takabe (1997) Expression of a betaine aldehyde dehydrogenase gene in rice, a glycinebetaine nonaccumulator, and possible localization of its protein in peroxisomes. Plant J. 11: 1115-1120.
- Nakamura T, T Takabe (2001) Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan. European Molecular Biology Laboratory (EMBL) entry, accession no. [Q94IC1](#).
- Onofre LJ, JP Legaria, NV Avonce, G Iturriaga (2003) La expresión del promotor del gen *Ahybadh4* en plantas transgénicas de *Arabidopsis thaliana* es específica en la raíz. Rev. Fitotec. Mex. 26 (3): 133-140.
- Pietruszko R (1989) Aldehyde dehydrogenase (EC 1.2.1.3). In: RR Watson (ed.), Biochemistry and Physiology of Substance Abuse. CRC Press. Boca Ratón, FL. Pp: 89-127.
- Rathinasabapathi B, K F McCue, D A Gage, AD Hanson (1994) Metabolic engineering of glycine betaine synthesis : plant betaine aldehyde dehydrogenase lacking typical transit peptides are targeted to tobacco chloroplasts where they confer betaine aldehyde resistance. Planta 193: 155-162.
- Shu W, W Ai, S Chen (1997) Institute of Genetics, Academia Sínica, Plant Biotechnology Laboratory, Beijing, China. National Center for Biotechnology Information (NCBI) sequence viewer, accession no. [U69142](#).
- Theologis A, J R Ecker, C J Palm, N A Federspiel (2002) Plant Gene Expression Center/USDA-U.C Berkeley, California, USA. European Bioinformatics Institute. Swiss-Prot entry, accession no. [Q9S795](#).
- Town C D, B J Hass, R Maiti, L I Hannick (2003) The Institute of Genomic Research, Medical Center, Rockville, USA. European Molecular Biology Laboratory (EMBL) entry, accession no. [Q9STS1](#).
- Trossat C, B Rathinasabapathi, A Hanson (1997) Transgenically expressed betaine aldehyde dehydrogenase efficiently catalyzes oxidation of dimethylsulfoniopropionaldehyde and w-aminoaldehydes. Plant Physiol. 113: 1457-1461.
- Tu-GC, H Weiner (1988) Identification of the cysteine residue in the active site of horse liver mitochondrial aldehyde dehydrogenase. J. Biol. Chem. 263: 1212-1217.
- Vojtechová M, A D Hanson, R A Muñoz-Clares (1997) Betaine aldehyde dehydrogenase from amaranth leaves efficiently catalyzes the NAD-dependen oxidation of dimethylsulfopropionaldehydes to dimethylsulfopropionate. Arch. Biochem. Biophys. 237: 81-88.
- Wang L-W, A M Showalter (2002) Ohio University, Athens, USA. European Molecular Biology Laboratory (EMBL) entry, accession no. [Q8RX99](#).
- Weigel P, E A Weretilnyk, A D Hanson (1986) Betaine aldehyde oxidation by spinach chloroplasts. Plant Physiol. 82: 753-759.
- Weretilnyk E A, A D Hanson (1988) Betaine aldehyde dehydrogenase polymorphism in spinach: genetic and biochemical characterization. Biochem. Genet. 26: 143-151.

- Weretilnyk E A, A D Hanson (1990)** Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. *Proc. Natl. Acad. Sci. USA* 87: 2745-2749.
- Wolfe K H, P M Sharp, W-H Li (1989)** Rates of synonymous substitution in plant nuclear genes. *J. Mol. Evol.* 208-211.
- Wood A J, H Saneoka, D Rhodes, R J Joly, P B Goldsbrough (1996)** Betaine aldehyde dehydrogenase in sorghum. Molecular cloning and expression of two related genes. *Plant Physiol.* 110: 1301-1308.
- Xiao G, G Zhang, F Liu, S Chen (1992)** Institute of Genetics, Academia Sinica, Plant Biotechnology Laboratory, Beijing, China. National Center for Biotechnology Information (NCBI) Sequence Viewer, accession no. [X69770](#).
- Yin X, Y Zhao, H Zhang (2002)** Institute of Genetics, Academia Sinica, Plant Biotechnology Laboratory, Beijing, China. European Molecular Biology Laboratory (EMBL) entry, accession no. [Q8L8I1](#).
- Zuckerkandl E, L Pauling (1962)** Molecular disease, evolution and genic heterogeneity. *In*: M. Kash and B. Pullman (eds.) *Horizons in Biochemistry*. Academic Press. New York. pp: 189-225.
- Zuckerkandl E, L Pauling (1965)** Evolutionary divergence and convergence in proteins. *In*: V. Bryson and H.J. Vogel (eds.). *Evolving Genes and Proteins*. Academic Press. New York. pp: 97-166.