

## NADP-Malate Dehydrogenase Gene Evolution in Andropogoneae (Poaceae): Gene Duplication Followed by Sub-functionalization

P. RONDEAU<sup>1</sup>, C. ROUCH<sup>1</sup> and G. BESNARD<sup>2,\*</sup>

<sup>1</sup>Université de la Réunion, LBG, 15 Avenue R. Cassin, 97715 St-Denis Messag Cedex 9, La Réunion, France and

<sup>2</sup>Department of Ecology and Evolution, Biology Building, University of Lausanne, 1015 Lausanne, Switzerland

Received: 22 November 2004 Returned for revision: 18 May 2005 Accepted: 15 September 2005 Published electronically: 21 October 2005

• **Background and Aims** Plastid NADP-dependent malate dehydrogenase (MDH) catalyses the conversion of oxaloacetate to malate. In C<sub>4</sub> plants, it is involved in photosynthetic carbon assimilation. In Poaceae, one NADP-MDH gene has been identified in rice (C<sub>3</sub>; Ehrhartoideae) and maize (C<sub>4</sub>; Panicoideae), whereas two tandemly repeated genes have been identified in *Sorghum* (C<sub>4</sub>; Panicoideae). In the present study, the molecular evolution of the NADP-MDH multigene family was investigated in order to analyse how the C<sub>4</sub> isoform has evolved over a broader range of panicoid grasses.

• **Methods** Polymerase chain reaction (PCR)-based cloning was used to isolate cDNAs encoding NADP-MDHs from 15 species of Panicoideae. A gene phylogeny was reconstructed based on cDNA sequences using distance and maximum parsimony methods. Episodic selection along some branches of the phylogenetic tree was tested by analysing non-synonymous and synonymous rate ratios. Transcription of NADP-MDH genes was compared in green leaves of five accessions of *Saccharum*, *Sorghum* and *Vetiveria* using a semi-quantitative PCR approach.

• **Key Results** Phylogenetic analyses of these data support the existence of two NADP-MDH gene lineages (NMDH-I and NMDH-II) in several Andropogoneae (i.e. *Saccharum*, *Sorghum* and *Vetiveria*). Episodic positive selection was shown along the basal branch of the NMDH-II clade. Three amino acid modifications allow the two gene lineages to be distinguished, suggesting a positive selection at these sites. In green leaves, we showed that the transcript accumulation was higher for NMDH-I than for NMDH-II.

• **Conclusions** It is hypothesized that the maintenance of both NADP-MDH genes in some Andropogoneae is due to a partition of the original functions across both copies. NMDH-I probably corresponds to the C<sub>4</sub> isoform as previously suggested. Nevertheless, some C<sub>4</sub> species (e.g. maize) only have one gene which should be selected for its high expression level in leaves. This study confirms that gene duplicates have been recruited for C<sub>4</sub> photosynthesis but are not required in every case.

**Key words:** C<sub>4</sub> photosynthesis, gene duplication, multigene family, NADP-MDH, Poaceae, positive selection, sub-functionalization.

### INTRODUCTION

C<sub>4</sub> photosynthesis provides a selective advantage for plants growing in tropical areas (Sage, 2004). The C<sub>4</sub> pathway serves as a pump, which provides an elevation of the CO<sub>2</sub> concentration at the site of Rubisco, leading to limited photorespiration and high water use efficiency in C<sub>4</sub> species (von Caemmerer and Furbank, 2003). This pathway has appeared several times during plant evolution from the Oligocene and Miocene, at a period when atmospheric CO<sub>2</sub> levels were gradually declining (Edwards *et al.*, 2001; Sage, 2004). C<sub>4</sub> genes are of great interest since an aim of biotechnology is to introduce the C<sub>4</sub> traits to cultivated C<sub>3</sub> plants, possibly improving their performance under warm conditions (Matsuoka *et al.*, 2001). Phylogenetic and biochemical studies of C<sub>4</sub> genes have also provided insights into the molecular evolution of the C<sub>4</sub> pathway (e.g. Sheen, 1999; Gehrig *et al.*, 2001; Besnard *et al.*, 2003; Monson, 2003; Westhoff and Gowik, 2004). In Poaceae, multiple appearances of C<sub>4</sub> photosynthesis during the PACCAD (Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Aristidoideae, Danthonioideae) clade diversification have probably occurred (Kellogg, 2000). Using a phylogenetic approach on the grass phosphoenolpyruvate

carboxylase multigene family, it has been proposed that such multiple appearances should be associated with multiple inductions of a pre-existing system (Besnard *et al.*, 2003). Study of other multigene families including a C<sub>4</sub> isoform (NADP-malate dehydrogenase, pyruvate orthophosphate dikinase) is required to understand thoroughly how the C<sub>4</sub> pathway evolved during grass diversification.

In plants, plastid NADP-dependent malate dehydrogenases (NADP-MDHs; EC 1.1.1.82) are essential enzymes involved in photosynthesis (Scheibe, 1987; Sheen, 1999). These enzymes are implicated in two main functions. First, in all plants, NADP-MDH helps control export of the reducing equivalents (malate) from the chloroplast to the cytosol (Scheibe, 1987; Ocheretina *et al.*, 2000). Secondly, NADP-MDH is also involved in photosynthetic carbon assimilation of both C<sub>4</sub> NADP-malic enzyme type and C<sub>4</sub> phosphoenolpyruvate carboxykinase (PCK) enzyme type plants because it converts oxaloacetate to malate, which is then transferred from the mesophyll cells to the bundle sheath cells (von Caemmerer and Furbank, 2003). NADP-MDH is a redox-regulated enzyme (Miginiac-Maslow *et al.*, 2000). Redox regulation appeared relatively late in evolution (probably in green algae; Ocheretina *et al.*, 2000). NADP-MDH is a highly conserved enzyme in plants, and is apparently encoded by a small multigene family

\* For correspondence. E-mail gbesnar@hotmail.com

TABLE 1. Species origin, accession number and reference of NADP-MDH sequences used for analyses

Species	Accession number	Reference
Poaceae		
Sub-family Panicoideae		
Tribe Andropogoneae		
<i>Dichanthium aristatum</i> (Poir.) C. Hubb.	AJ512368	This study
<i>Heteropogon contortus</i> (L.) P. Beauv.	AJ621582	This study
<i>Hypparrhenia rufa</i> (Nees) Stapf	AJ621580	This study
<i>Ischaemum koleostachys</i> (Steud.) Hack.	AJ512369	This study
<i>Pogonatherum paniceum</i> (Lam.) Hack.	AJ512373	This study
<i>Saccharum officinarum</i> L.	AJ344432, AJ565857, AJ833960	This study
<i>Saccharum spontaneum</i> L.	AJ416567, AJ565856	This study
<i>Sorghum arundinaceum</i> (Desv.) Stapf	AJ512374	This study
<i>Sorghum bicolor</i> (L.) Moench	X53453, S55884, M31965	Luchetta <i>et al.</i> (1991)
<i>Themeda quadrivalvis</i> (L.) Kuntze	AJ512376	This study
<i>Vetiveria zizanioides</i> (L.) Nash	AJ344433, AJ512375	This study
<i>Zea mays</i> L.	X16084	Metzler <i>et al.</i> (1989)
Tribe Paniceae		
<i>Melinis repens</i> (Willd.) Zizka	AJ621581	This study
<i>Oplismenus compositus</i> (L.) P. Beauv.	AJ512370	This study
<i>Paspalidium geminatum</i> (Forssk.) Stapf	AJ833959	This study
<i>Paspalum paniculatum</i> L.	AJ512372	This study
<i>Urochloa maxima</i> (Jacq.) R. Webster	AJ512371	This study
Sub-family Erhartoideae		
<i>Oryza sativa</i> L.	AK105935	Kikuchi <i>et al.</i> (2003)
Asteraceae		
<i>Flaveria trinervia</i> (Spreng.) C. Mohr	U22533	McGonigle and Nelson (1995)

(Luchetta *et al.*, 1991). Phylogenetic trees based on this gene family should be appropriate for increasing our understanding of the evolution of NADP-MDH isoforms and particularly to study the appearance(s) of the C<sub>4</sub> NADP-MDH isoform during plant evolution.

In maize and sorghum, two grasses belonging to tribe Andropogoneae of subfamily Panicoideae (GPWG, 2001), NADP-MDH genes have been characterized (Metzler *et al.*, 1989; Luchetta *et al.*, 1991). Only one NADP-MDH nuclear gene was identified in maize (Metzler *et al.*, 1989) and in the complete genome sequence of rice (Kikuchi *et al.*, 2003). Conversely, in sorghum, the existence of two tandemly repeated genes encoding NADP-MDH was demonstrated (Luchetta *et al.*, 1991). The regulation of the two genes was differential, with one being light induced (C<sub>4</sub> gene; NMDH-I) and the other (NMDH-II) being constitutively expressed at a low level. The function of the latter was not clearly elucidated but it could be involved in the export of reducing equivalents. Based on these data, Luchetta *et al.* (1991) postulated that the grass C<sub>4</sub> NADP-MDH gene could have been derived from a non-C<sub>4</sub> NADP-MDH. In contrast, McGonigle and Nelson (1995) proposed that only changes in the regulation of a pre-existing gene were involved in the C<sub>4</sub> NADP-MDH appearance in the genus *Flaveria* (Asteraceae), without gene duplication during the evolution from C<sub>3</sub> to C<sub>4</sub> plants. Only a few DNA sequences of NADP-MDH genes are currently available, and consequently the evolution of the NADP-MDH multigene family is still not well understood (Ocheretina *et al.*, 2000).

In this study, PCR-based cloning was used to isolate cDNAs encoding NADP-MDHs from 15 species belonging to Panicoideae. From green leaves, cDNAs related to the two sorghum isoforms were isolated. A phylogenetic

approach was used to provide insights into the evolution of the NADP-MDH multigene family in Panicoideae. Episodic selection along some branches of the phylogenetic tree was tested. In addition, an analysis of transcription specificities of NADP-MDH genes was conducted for three genera of Andropogoneae.

## MATERIALS AND METHODS

### Molecular analysis

In order to isolate cDNA sequences encoding NADP-MDHs in grasses, 15 species belonging to Panicoideae were analysed (Table 1). Twelve are C<sub>4</sub> NADP-malic enzyme type plants (representatives of Andropogoneae plus *Paspalum paniculatum* and *Paspalidium geminatum*), *Urochloa maxima* and *Melinis repens* are C<sub>4</sub> PCK enzyme type plants, and *Oplismenus compositus* is a C<sub>3</sub> plant. For each accession, green leaves were sampled at noon. Total RNA was extracted and then reverse transcription reactions were performed as previously described (Besnard *et al.*, 2002).

From the available NADP-MDH sequences of sorghum [EMBL accessions X53453 (NMDH-I) and S55884 (NMDH-II); Luchetta *et al.*, 1991] and maize (EMBL accession X16084; Metzler *et al.*, 1989), specific primer pairs were designed to amplify cDNA segments of NADP-MDH using reverse transcription-polymerase chain reaction (RT-PCR). Primers were defined in conserved regions of available grass NADP-MDH genes. Two forward primers were defined: 5' GAGTGCTTCGGGGTCTTCTGC 3' (A) and 5' ATGGGCCTGTCAACAGYTTACTCC 5' (B). The same reverse primer was used in all reactions: 5' TACATTTCTCCTGGTAGCATTGTATC 3' (R). Long cDNA segments

(of approx. 1100 bp with primers AR and 1300 bp with primers BR) were PCR generated. The PCR conditions described in Besnard *et al.* (2002) were used with an annealing temperature of 50 °C for each primer pair. RT-PCR fragments were cloned before sequencing. The pGEM<sup>®</sup>-T vector (Promega, Madison, WI, USA) and *Escherichia coli* strain DH5 $\alpha$  were used according to the supplier's recommendations. For each species of *Andropogoneae*, distinct sequences were looked for among 15 clones using their restriction profile. The insert in each clone was amplified by PCR using SP6 and T7 primers (Promega). PCR products were restricted with *Hpa*II, separated on a 3% agarose gel and stained with ethidium bromide. Each clone displaying a different profile was then sequenced. Double-stranded DNA sequencing reactions were performed by the ESGS Society (Euro Sequence Gene Service, Evry, France). Sequences have been deposited in the EMBL databank.

#### Phylogenetic analysis

NADP-MDH cDNA sequences from *Sorghum*, *Zea* and *Oryza* from GenBank were added to the data (Table 1). For *Saccharum* and *Sorghum*, both cultivated forms (*Saccharum officinarum* and *Sorghum bicolor*) and close wild relatives [*Saccharum spontaneum* and *Sorghum arundinaceum* (= *S. verticilliflorum*)] were studied (Dillon *et al.*, 2001; Hodkinson *et al.*, 2002). Only coding sequences (cds) were used for analysis. In addition, the phylogenetic trees were rooted with an outgroup sequence. Since no complete NADP-MDH sequence from another monocot was available in GenBank, a cDNA sequence from *Flaveria* (U22533; Asteraceae) was used. A sample of 25 sequences (Table 1) displaying the major part of the cds was considered. Sequence alignments were performed using Clustal W (Thompson *et al.*, 1994). Phylogenetic analyses were made using a large internal cDNA segment of 1050 bp in which indels are very infrequent (only one 3 bp deletion was detected in *S. spontaneum* sequences). In the sorghum sequence X53453, the 1050 bp segment is located between nucleotides 220 and 1269. Phylogenetic analyses were conducted using MEGA version 2 (Kumar *et al.*, 2004). Mean frequencies of each nucleotide were similar (T = 25%, A = 29%, G = 26%, C = 20%), but transitions were twice as frequent as transversions. Consequently, to consider this unequal probability of the substitution types, the Kimura 2-parameter (Kimura, 1980) was used to compute distances between each pair of sequences. Phylogenetic trees were then constructed using the neighbour joining algorithm (NJ; Saitou and Nei, 1987). Bootstrap values were computed using 1000 replicates to evaluate support for the groupings. In addition, a maximum parsimony (MP) analysis was conducted. All characters were equally weighted. Gaps were treated as missing data. A heuristic search was used to find the most-parsimonious trees. The close-neighbour-interchange algorithm was used with a search level of 3 (Kumar *et al.*, 2004). Searches included 100 replications of random addition sequences. All the best trees were retained. A strict consensus tree was generated from the equally most-parsimonious trees. Bootstrap values were computed using 1000 replicates with

MEGA (Kumar *et al.*, 2004) to evaluate support for the branches.

#### Compared evolution of NADP-MDH genes in *Andropogoneae*

Within *Andropogoneae*, two NADP-MDH genes were identified (NMDH-I and NMDH-II; see below). Using all available sequences, amino acid sites providing a distinction between these two genes were investigated. A segment of 350 amino acids was used. In the *Sorghum* NMDH-I protein (429 amino acids; X53453), this segment is situated between amino acids 70 and 419. In addition, in order to determine possible selective divergence between the two NADP-MDH genes, the rates of non-synonymous and synonymous substitutions between pairs of sequences from *S. officinarum*, *S. bicolor* and *Vetiveria zizanioides* were compared independently for NMDH-I and NMDH-II. The method of Nei and Gojibori (1986) was applied.

Within *Andropogoneae*, the occurrence of episodic positive selection following gene duplication was also tested on the 1050 bp segment. Eleven sequence accessions from *Saccharum*, *Sorghum*, *Vetiveria*, *Zea* and *Oryza* were considered (M31965, S55884, X16084, X53453, AJ344432, AJ344433, AJ416567, AJ512374, AJ512375, AJ565857 and AK105935). The AJ565856 accession was not used in this analysis since it probably encodes a non-functional enzyme (see below). Furthermore, in the genus *Saccharum*, only one NMDH-I sequence (AJ565857) was used because accessions AJ565857 and AJ833960 only differ by one non-synonymous substitution. The structure of the phylogenetic tree was defined based on an MP analysis as previously described. The  $d_N/d_S$  ratios ( $\omega$ ) were estimated for all branches of the tree by using a maximum likelihood (ML) analysis as described by Yang (1998). We used the CODEML program implemented in the PAML package (Yang, 1997) using the 'free ratio' model assuming an independent  $d_N/d_S$  ratio for each branch of the phylogenetic tree (Yang, 1998). The F3X4 codon model was used; it takes into account differences in transversion *vs.* transition and codon usages between analysed sequences.

#### NADP-MDH gene transcription analysis in *Andropogoneae*

To quantify differential NADP-MDH RNA accumulation between different plant materials (i.e. green leaf, etiolated leaf, root and stem) in *S. officinarum*, we used a Northern dot-blot membrane (see Besnard *et al.*, 2003). For each tissue sample of cultivar R570, 10 mg of total RNA was directly located onto an N+-Hybond nylon membrane (Amersham, Bucks, UK). An NADP-MDH segment of 357 bp (accession AJ320266) was hybridized as a probe using the procedure described in Besnard *et al.* (2003). In *Saccharum*, this probe displays great similarity (at least 94%) with both genes NMDH-I and NMDH-II. Consequently, this probe cannot be considered as specific to one NADP-MDH gene. The membrane was also independently hybridized with a ribosomal 18S probe. Ribosomal 18S gene expression was considered to be constitutive and

proportional to the amount of deposited RNA. The light induction of the NADP-MDH transcription was estimated using the relative signal intensity of 18S and NADP-MDH genes on the Northern dot-blot for both green and etiolated leaves.

A semi-quantitative PCR approach was developed to compare transcript accumulation of NMDH-I and NMDH-II in green leaves of *Saccharum*, *Sorghum* and *Vetiveria*. One accession each for *S. arundinaceum* and *V. zizanioides* and three for *S. officinarum* accessions [i.e. SES14 (wild accession), Big Tana Rayé and R570 (cultivated accessions)] were characterized. Several restriction sites allowing distinction between NMDH-I and NMDH-II were identified. One primer pair was defined to amplify simultaneously an internal cds region of 253 bp of both NADP-MDH genes containing a diagnostic *Hpa*II restriction site: For-5' GCCCTTCTTATTGGTGCTAAG 3' and Rev-5' CTAGCTGGCACTTTGCTCTAT 3'. PCR mixtures, containing 0.25 mM dNTP, 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl<sub>2</sub> and 20 pmol of each oligonucleotide primer in a total volume of 50 µL, were added to 0.3 mg of cDNA and 1.2 U of *Taq* DNA polymerase (Advanced Biotechnologies, Inc., Columbia, MD, USA) and amplified in a thermocycler (PTC100-v7, MJ Research). Samples were incubated for 4 min at 94 °C, followed by 36 cycles consisting of 45 s at 94 °C, 45 s at 50 °C and 45 s at 72 °C. This PCR program was followed by steps of decreasing temperature from 94 to 50 °C on the basis of 3 min per degree, to avoid heteroduplex formation. A 10 µL aliquot of PCR product was then restricted with *Hpa*II (Sigma-Aldrich Corp., Missouri, USA) following the supplier's recommendations. Restriction fragments were separated on a 3% agarose gel and stained with ethidium bromide. The gels were photographed using a Kodak DC290 digital camera. Photo treatment for a relative quantification of NMDH-I or NMDH-II mRNA of each gene was carried out using the Kodak 1D Image analyser software (version 3.6). A correction was made between the ethidium bromide staining intensity and the size of the fragments [253 bp for the unrestricted fragment (NMDH-I) and 220 bp for the restricted fragment (NMDH-II)]. To test the reproducibility of the method, three independent assays were performed.

## RESULTS

### Data sequences and phylogenetic relationships between NADP-MDH

Nineteen 19 cDNA sequences encoding NADP-MDHs were generated in the species analysed (Table 1). One sequence from *S. spontaneum* (AJ565856) probably corresponds to a pseudo-gene due to the presence of a termination codon in the middle of the sequence. The phylogenetic trees based on the sequence analysis of the 1050 bp segment are presented in Fig. 1. The different methods used to reconstruct phylogenies led to similar results. *Oryza*, as expected with this level of sampling, is sister to Panicoideae. In a non-monophyletic tribe Paniceae, *Paspalum* is separated from *Paspalidium*, *Melinis*, *Urochloa* and *Oplismenus*, although with only weak bootstrap support. *Paspalum* is sister to

Andropogoneae in the MP tree, whereas it is sister to the rest of the Panicoideae in the NJ tree. Lastly, phylogenetic reconstruction shows that the clade including all NADP-MDHs from Andropogoneae is well supported (>98%; Fig. 1). Nevertheless, in this clade, two groups of sequences can be recognized (named NMDH-I and NMDH-II; Fig. 1B). In each of these two sub-clades, there is at least one sequence from *Saccharum*, *Sorghum* and *Vetiveria*. The sequence of *Zea* is not clearly assignable to either of these two sub-clades.

### Evolution of NADP-MDH genes in Andropogoneae

The comparison of mutation rates between the same species for each NADP-MDH gene revealed that the ratio  $\omega$  ( $d_N/d_S$ ) is 1.3–2.0 times higher in NMDH-II than in NMDH-I (Table 2). This suggests that the evolution of NMDH-II was slightly faster than that of NMDH-I. Using an ML analysis to estimate the  $d_N/d_S$  ratios along phylogenetic branches, a value >1.0 was obtained at the base of the NMDH-II clade ( $\omega = 1.60$ ; Fig. 2). This probably indicates an episodic positive selection on the NMDH-II gene. Within the 350 amino acid sequence, three sites allow a distinction between NMDH-I and NMDH-II. First, at position 329 (in the sorghum sequence X53453), NMDH-I proteins have an alanine whereas all NMDH-II proteins have a valine. The other sequences from Paniceae, *Zea* and *Oryza* also have a valine. (This modification is a conservative mutation that should not induce a major change in the enzyme characteristics.) Secondly, at position 292 (in X53453), a lysine is found in proteins belonging to NMDH-I in Paniceae, *Zea* and *Oryza*, whereas a glutamine or a glutamate is found in NMDH-II proteins. Thirdly, at position 414, an asparagine is found in all proteins belonging to NMDH-I in Paniceae, *Zea* and *Oryza*, whereas an aspartate is found in NMDH-II proteins.

### Comparative NADP-MDH gene transcription in Andropogoneae

Using a Northern dot-blot approach, higher NADP-MDH gene transcription was revealed in green leaves of sugarcane cultivar R570 (Fig. 3A) than in roots, stem and etiolated leaves. This confirms that the transcription of NADP-MDH genes is light induced in *Saccharum*. It was estimated that the transcript accumulation was about seven times higher in green leaves than in etiolated leaves. Using the semi-quantitative PCR method, it was revealed that both NADP-MDH genes are transcribed in green leaves of *Saccharum*, *Sorghum* and *Vetiveria* (Fig. 3B), and it was shown that the amount of cDNA was usually higher for NMDH-I than for NMDH-II. However, similar amounts of cDNA were observed for both genes in cultivar R570 (54 and 46%; Fig. 3B).

## DISCUSSION

The phylogenetic trees based on NADP-MDH sequences (Fig. 1) allow a distinction to be made between a paraphyletic Paniceae and a monophyletic Andropogoneae in subfamily Panicoideae. The separation of *Paspalum* from other Paniceae analysed (*Paspalidium*, *Melinis*, *Urochloa* and

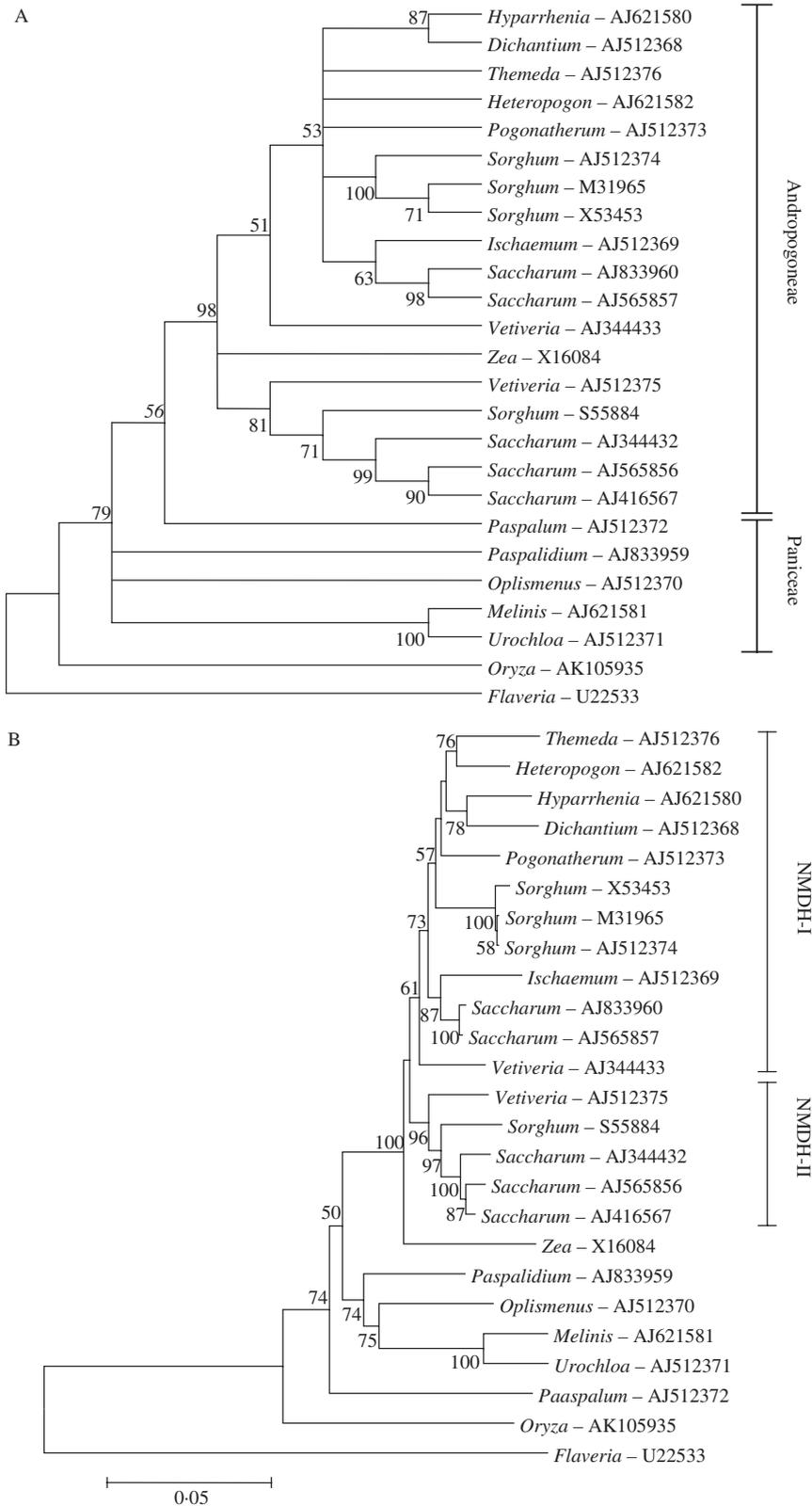


FIG. 1. Phylogenetic relationships between cDNAs encoding NADP-malate dehydrogenase in grasses. The analysis is based on a 1050 bp segment (423 variable sites of which 228 are potentially parsimony informative). (A) An MP tree based on a maximum parsimony analysis (904 steps, CI = 0.64, RI = 0.57). The consensus tree is given (from the three most-parsimonious trees). Species belonging to either Andropogoneae or Paniceae are indicated. (B) An NJ tree based on Kimura 2-parameter distance and the neighbour-joining algorithm. The two gene lineages NMDH-I and NMDH-II are specified. Bootstrap values >50% are indicated on each tree.

TABLE 2. Non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) mutation rates computed by comparison of pairs of sequences from *Saccharum officinarum*, *Sorghum bicolor* and *Vetiveria zizanioides*

NMDH gene	Pair of sequences		$d_N$	$d_S$	$\omega^*$
NMDH-I	<i>Saccharum</i> (AJ565857)	<i>Sorghum</i> (X53453)	0.014	0.098	0.143
NMDH-II	<i>Saccharum</i> (AJ344432)	<i>Sorghum</i> (S55884)	0.020	0.105	0.190
NMDH-I	<i>Saccharum</i> (AJ565857)	<i>Vetiveria</i> (AJ344433)	0.010	0.095	0.105
NMDH-II	<i>Saccharum</i> (AJ344432)	<i>Vetiveria</i> (AJ512375)	0.017	0.081	0.210
NMDH-I	<i>Sorghum</i> (X53453)	<i>Vetiveria</i> (AJ344433)	0.015	0.143	0.105
NMDH-II	<i>Sorghum</i> (S55884)	<i>Vetiveria</i> (AJ512375)	0.015	0.101	0.149

The parameters were computed independently for each NADP-MDH gene (NMDH-I and NMDH-II) using the method of Nei and Gojobori (1986). These computations were based on the 1050 bp segment.

\*  $\omega$  is the ratio  $d_N/d_S$ .

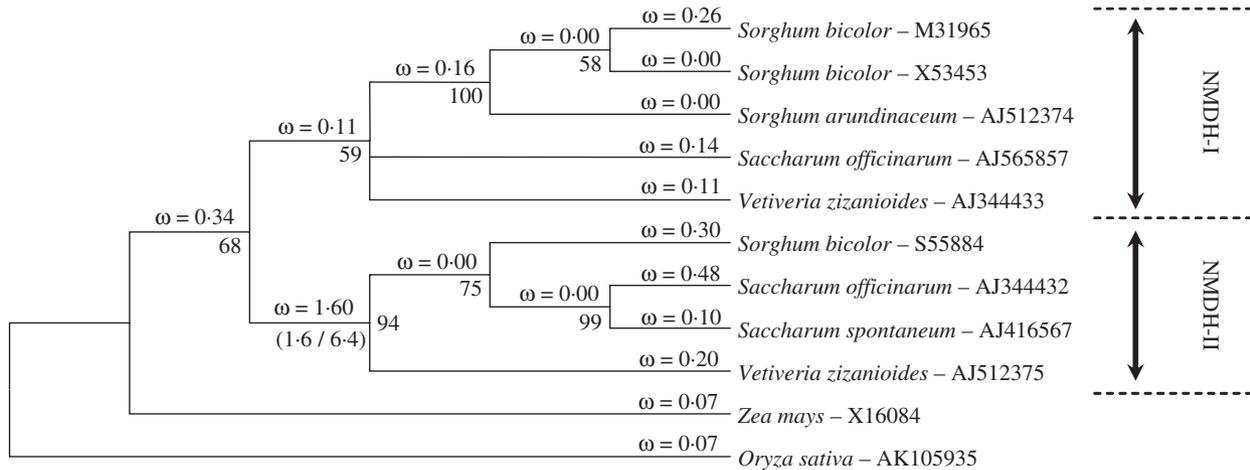


FIG. 2. An MP tree based on a subset of 11 NADP-MDH cDNAs representative of the gene lineages NMDH-I and NMDH-II in the tribe Andropogoneae (genera *Saccharum*, *Sorghum* and *Vetiveria*). Only one most-parsimonious tree was obtained (289 steps, CI = 0.85, RI = 0.71). Bootstrap values >50 % are indicated. The tree topology was used to estimate  $d_N/d_S$  ( $\omega$ ) values along each branch based on a maximum likelihood analysis and a free ratio model. The numbers in parentheses are the maximum likelihood estimates of the numbers of synonymous and non-synonymous substitutions, respectively, along the basal branch of NMDH-II.

*Oplismenus*; Fig. 1B) was expected since paraphyly in this tribe has been demonstrated using other gene sequences (e.g. Giussani *et al.*, 2001; Besnard *et al.*, 2002; Duvall *et al.*, 2003). In Andropogoneae, the existence of two distinct genes encoding NADP-MDH (named NMDH-I and NMDH-II) is supported by our study, and this confirms the observation in sorghum (Luchetta *et al.*, 1991). These two genes are closely related, but the sequence from *Zea* falls in an unresolved position relative to the clades representing NMDH-I and NMDH-II (Fig. 1B). This suggests that the gene duplication event leading to the appearance of these two gene lineages is specific to some Andropogoneae (at least in species deriving from the common ancestor of *Saccharum*, *Sorghum* and *Vetiveria*). The duplicate gene probably exists in some of the other genera analysed (*Dichanthium*, *Heteropogon*, *Hyparrhenia*, *Ischaemum* and *Pogonatherum*), but this was not demonstrated in our analysis (probably due to a lower transcription level of NMDH-II than of NMDH-I).

Theoretical models predict that maintenance of duplicate genes could be associated with selective pressures via neo-functionalization, in which one copy acquires a new

function, or by sub-functionalization, in which the original function is partitioned across both copies (Monson, 2003; Walsh, 2003). Gene duplication and subsequent modifications are considered to be essential for the multiple C<sub>4</sub> pathway appearance. Many of the duplicated genes destined for C<sub>4</sub> roles should have been shaped by selection to produce novel C<sub>3</sub> roles, prior to being recruited for C<sub>4</sub> photosynthesis (Monson, 2003). However, the C<sub>4</sub> NADP-MDH isoform selection could be a relatively simple event compared with the selection of other C<sub>4</sub> genes (e.g. C<sub>4</sub> PEPC genes) since it may not always have involved gene duplication (McGonigle and Nelson, 1995; Monson, 2003). As proposed by McGonigle and Nelson (1995), NADP-MDH enzymes involved in C<sub>4</sub> photosynthesis have been selected for their high expression level in the mesophyll cells (as is probably the case in *Zea* in which only one NADP-MDH gene has been identified). Nevertheless, selective pressures favouring new mutations during the evolution from C<sub>3</sub> to C<sub>4</sub> NADP-MDH are likely to have occurred. Gene duplication should thus be an opportunity to partition the original functions of NADP-MDHs across different copies. In *S. bicolor*, NMDH-I was reported to have a high transcription level in

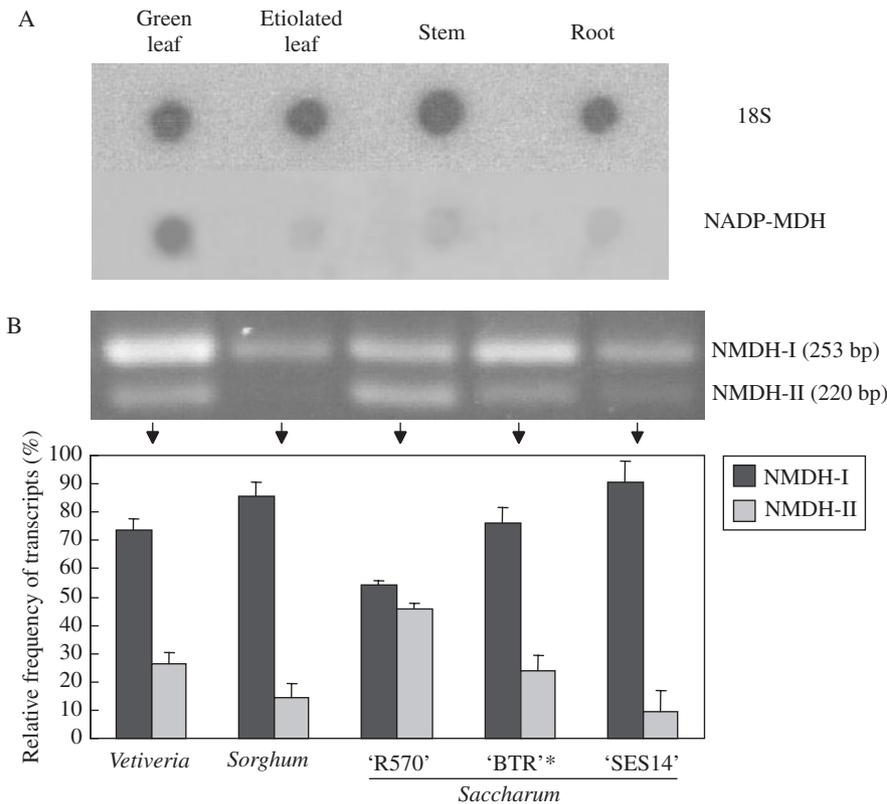


FIG. 3. NADP-malate dehydrogenase gene transcription analyses. (A) Gene transcript accumulation according to different plant materials in sugarcane cultivar R570 revealed using Northern dot-blot analysis. A nylon membrane with RNA from four types of plant material (green leaf, etiolated leaf, stem and root) was hybridized successively with two probes: an 18S rDNA probe (30 min exposure) and an NADP-MDH probe (7 d exposure). (B) Relative frequency of NMDH-I and NMDH-II transcripts in green leaves of five accessions belonging to *Saccharum*, *Sorghum* and *Vetiveria* assessed using the PCR quantification method. The photograph illustrates an example of restricted DNA fragments of NMDH-I (253 bp) and NMDH-II (220 bp) separated on a 3 % agarose gel. Histograms give the relative transcript abundance of NADP-MDH genes in each plant based on three repeats (data are shown as the means  $\pm$  s.d.). \*BTR = Big Tana Rayé.

green leaves and is likely to be involved in  $C_4$  photosynthesis, whereas NMDH-II, which displays a lower transcription level, could be involved in the reducing equivalent export (Luchetta *et al.*, 1991). Our results (Fig. 3) confirm that the transcript accumulation in green leaves is usually stronger for NMDH-I than for NMDH-II, sustaining the hypothesis that these two genes were maintained in some *Andropogoneae* due to sub-functionalization. Furthermore, we expect that each isoform displays different kinetic properties since positive selection has probably occurred during the period of time following the duplication event (on the basal branch of the NMDH-II clade; Fig. 2). In this way, two amino acid sites (Lys292 and Asn414 in sorghum NMDH-I) were detected which could induce significant differences in kinetic properties between NMDH-I and NMDH-II. Different electric properties of the amino acids at these variable sites should result in conformational changes between NMDH-I and NMDH-II. To assess the importance of such sites, additional studies are needed. For example, experiments could be done using site-directed mutagenesis (Schepens *et al.*, 2000) or by construction of reciprocal enzyme chimeras (Bläsing *et al.*, 2000), offering new perspectives for biochemical research on the NADP-MDH enzymes (Miginiac-Maslow *et al.*, 2000).

#### ACKNOWLEDGEMENTS

We would like to thank Trevor Hodkinson, Gerald Edwards and Mike Fay for their constructive suggestions. We are also grateful for helpful discussions with J. P. Jacquot, N. Corradi, C. Parisod and B. Offmann.

#### LITERATURE CITED

- Besnard G, Offmann B, Robert C, Rouch C, Cadet F. 2002. Assessment of the  $C_4$  phosphoenolpyruvate carboxylase gene diversity in grasses (Poaceae). *Theoretical and Applied Genetics* **105**: 404–412.
- Besnard G, Pinçon G, D'Hont A, Hoarau JY, Cadet F, Offmann B. 2003. Characterisation of the phosphoenolpyruvate carboxylase gene family in sugarcane (*Saccharum* spp.). *Theoretical and Applied Genetics* **107**: 470–478.
- Bläsing OE, Westhoff P, Svensson P. 2000. Evolution of  $C_4$  phosphoenolpyruvate carboxylase in *Flaveria*: a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for  $C_4$ -specific characteristics. *Journal of Biological Chemistry* **275**: 27917–27923.
- von Caemmerer S, Furbank RT. 2003. The  $C_4$  pathway: an efficient  $CO_2$  pump. *Photosynthesis Research* **77**: 191–207.
- Dillon SL, Lawrence PK, Henry R. 2001. The use of ribosomal ITS to determine phylogenetic relationships within *Sorghum*. *Plant Systematics and Evolution* **230**: 97–110.
- Duvall MR, Saar DE, Grayburn WS, Holbrook GP. 2003. Complex transitions between  $C_3$  and  $C_4$  photosynthesis during the evolution of Paniceae: a phylogenetic case study emphasizing the position of

- Steinchisma hians* (Poaceae), a C<sub>3</sub>–C<sub>4</sub> intermediate. *International Journal of Plant Sciences* **164**: 949–658.
- Edwards GE, Furbank RT, Hatch MD, Osmond CB. 2001.** What does it take to be C<sub>4</sub>? Lessons from the evolution of C<sub>4</sub> photosynthesis. *Plant Physiology* **125**: 46–49.
- Gehrig H, Heute V, Kluge M. 2001.** New partial sequences of phosphoenolpyruvate carboxylase as molecular phylogenetic markers. *Molecular Phylogenetics and Evolution* **20**: 262–274.
- Giussani LM, Cota-Sanchez JH, Zuloaga FO, Kellogg EA. 2001.** A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C<sub>4</sub> photosynthesis. *American Journal of Botany* **88**: 1993–2012.
- GPWG. 2001.** Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* **88**: 373–457.
- Hodkinson TR, Chase MW, Lledó D, Salamin N, Renvoize SA. 2002.** Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. *Journal of Plant Research* **115**: 381–392.
- Kellogg EA. 2000.** The grasses: a case study in macroevolution. *Annual Review of Ecology and Systematics* **31**: 217–238.
- Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, et al. 2003.** Collection, mapping and annotation of over 28,000 cDNA clones from *japonica* rice. *Science* **301**: 376–379.
- Kimura M. 1980.** A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111–120.
- Kumar S, Tamura K, Nei M. 2004.** MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**: 150–163.
- Luchetta P, Crépin C, Gadal P. 1991.** Organization and expression of the two homologous genes encoding the NADP-malate dehydrogenase in *Sorghum vulgare* leaves. *Molecular and General Genetics* **228**: 473–481.
- McGonigle B, Nelson T. 1995.** C<sub>4</sub> isoform of NADP-malate dehydrogenase: cDNA cloning and expression in leaves of C<sub>4</sub>, C<sub>3</sub>, and C<sub>3</sub>–C<sub>4</sub> intermediate species of *Flaveria*. *Plant Physiology* **108**: 1119–1126.
- Matsuoka M, Furbank RT, Fukayama H, Miyao M. 2001.** Molecular engineering of C<sub>4</sub> photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**: 297–314.
- Metzler MC, Rothermel BA, Nelson T. 1989.** Maize NADP-malate dehydrogenase: cDNA cloning, sequence, and mRNA characterization. *Plant Molecular Biology* **12**: 713–722.
- Miginic-Maslow M, Johansson K, Ruelland E, Issakidis-Bourguet E, Schepens I, Goyer A, et al. 2000.** Light-activation of NADP-malate dehydrogenase: a highly controlled process for an optimized function. *Physiologia Plantarum* **110**: 322–329.
- Monson RK. 2003.** Gene duplication, neofunctionalization, and the evolution of C<sub>4</sub> photosynthesis. *International Journal of Plant Sciences* **164**: S43–S54.
- Nei M, Gojobori T. 1986.** Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**: 418–426.
- Ocheretina O, Haferkamp I, Tellioglu H, Scheibe R. 2000.** Light-modulated NADP-malate dehydrogenases from mossfern and green algae: insights into evolution of the enzyme's regulation. *Gene* **258**: 147–154.
- Sage RF. 2004.** The evolution of C<sub>4</sub> photosynthesis. *New Phytologist* **161**: 341–370.
- Saitou N, Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.
- Scheibe R. 1987.** NADP-malate dehydrogenase in C<sub>3</sub> plants: regulation and role of light-activated enzyme. *Physiologia Plantarum* **71**: 393–400.
- Schepens I, Ruelland E, Miginic-Maslow M, Le Maréchal P, Decottignies P. 2000.** The role of active site arginines of sorghum NADP-malate dehydrogenase in thioredoxin-dependent activation and activity. *Journal of Biological Chemistry* **275**: 35792–35798.
- Sheen J. 1999.** C<sub>4</sub> gene expression. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**: 187–271.
- Thompson JD, Higgins DG, Gibson TJ. 1994.** Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673–4680.
- Walsh B. 2003.** Population-genetics models of the fates of duplicate genes. *Genetica* **118**: 279–294.
- Westhoff P, Gowik U. 2004.** Evolution of C<sub>4</sub> phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus *Flaveria*. *Annals of Botany* **94**: 13–23.
- Yang Z. 1997.** PAML, a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**: 555–556.
- Yang Z. 1998.** Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* **15**: 568–573.