Theor Appl Genet (2002) 105:404–412
DOI 10.1007/s00122-001-0851-7

G. Besnard · B. Offmann · C. Robert · C. Rouch
F. Cadet

Assessment of the C_4 phosphoenolpyruvate carboxylase gene diversity in grasses (Poaceae)

Received: 27 July 2001 / Accepted: 10 September 2001 / Published online: 19 June 2002

Abstract C_4 phosphoenolpyruvate carboxylase (PEPC) is a key enzyme in the C_4 photosynthetic pathway. To analyze the diversity of the corresponding gene in grasses, we designed PCR primers to specifically amplify C_4 PEPC cDNA fragments. Using RT-PCR, we generated partial PEPC cDNA sequences in several grasses displaying a C_4 photosynthetic pathway. All these sequences displayed a high homology (78–99%) with known grass C_4 PEPCs. PCR amplification did not occur in two grasses that display the C_3 photosynthetic pathway, and therefore we assumed that all generated sequences corresponded to C_4 PEPC transcripts. Based on one large cDNA segment, phylogenetic reconstruction enabled us to assess the relationships between 22 grass species belonging to the subfamilies Panicoideae, Arundinoideae and Chloridoideae. The phylogenetic relationships between species deduced from C_4 PEPC sequences were similar to those deduced from other molecular data. The sequence evolution of the C_4 PEPC isoform was faster than in the other PEPC isoforms. Finally, the utility of the C_4 PEPC gene phylogeny to study the evolution of C_4 photosynthesis in grasses is discussed.

Keywords C_4 photosynthesis · cDNA sequence · Poaceae · Phosphoenolpyruvate carboxylase · Phylogeny

Introduction

In the grass family, there are more than 10,000 species among which are numerous cultivated C_4 plants that commonly grow in tropical areas (e.g. maize, sorghum, sugarcane, mil, panicum). All C_4 grasses are classified in the "PACC clade" (Davis and Soreng 1993) which includes the subfamilies Panicoideae, Arundinoideae, Chloridoideae and Centothecoideae. C_4 photosynthesis occurs in all Chloridoideae (except *Eragrostis walteri* Pilg.), in a few Arundinoideae [i.e. the genera *Aristida* L. (tribe Aristideae) or *Centropodia* Reichenb. (tribe Arundineae Dumort.)] and in most Panicoideae (Watson and Dallwitz 1992). Grasses display very diversified photosynthetic pathways, C_3, C_4 PCK (Phosphoenolpyruvate Carboxylase Kinase), C_4 NADP-ME (Malic Enzyme) or C_4 NAD-ME, due to the differential or absence of C_4 gene expression (Watson and Dallwitz 1992; Ku et al. 1996; Sinha and Kellogg 1996). Consequently, this family seems to be well adapted for an exhaustive evolutionary study of the C_4 pathway.

The C_4 pathway involves several enzymes which catalyze the fixation and transport of carbon dioxide. This results in a cyclical reaction, called the C_4 cycle or the Hatch and Slack cycle, which enables one to concentrate CO_2 in the bundle-sheath cells where photosynthesis occurs (Hatch and Slack 1966). Thus, Rubisco oxygenase activity, which decreases the photosynthetic yield, is reduced in plants displaying C_4 photosynthesis (Ku et al. 1996). This feature is especially advantageous for tropical plants since Rubisco oxygenase activity is enhanced with increasing temperature, in contrast to Rubisco carboxylase activity.

The first enzyme, which is involved in the C_4 cycle, is the C_4 phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31). The C_4 PEPC gene has a potential biotechnological interest (Ku et al. 1999) and should be worthwhile to assess the diversity of this nuclear gene. PEPC enzymes are encoded by an oligogenic family (Lepiniec et al. 1994). C_4 and non-C_4 PEPC sequences are now available in different plant families. Several isoforms have been characterised in maize, sorghum and sugarcane. These isoforms are involved in several functions such as the initial fixation of atmospheric CO_2 (= C_4 PEPC) and anaplerotic functions associated with nitrogen assimilation or amino-acid biosynthesis (for a review see Lepiniec et al. 1994). The evolution of such

Communicated by H.F. Linskens

G. Besnard (✉) · B. Offmann · C. Robert · C. Rouch · F. Cadet
Laboratoire de Biochimie et Génétique moléculaire,
Faculté des Sciences Université de la Réunion,
15 Avenue René Cassin, 97715 Saint Denis Messag Cedex 9,
La Réunion, France
e-mail: gbesnar@hotmail.com
Tel.: +33-262-93-86-45, Fax: +33-262-93-82-37

nuclear oligogenic families has involved gene duplication and ectopic expression of new isoforms (Clegg et al. 1997; Kellogg 2000). In fact, Lepiniec et al. (1994) have postulated that C_4 PEPC isoforms were derived from non-C_4 PEPC isoforms. Interestingly, the C_4 and non-C_4 isoforms of grasses have been shown to be well-differentiated, whereas it appears that, in dicotyledonous plants, C_4 PEPCs are very related to non-C_4 PEPCs. This would suggest that grass C_4 PEPCs have diverged from non-C_4 PEPCs over a long time, while divergence in dicots would be more recent (Lepiniec et al. 1994). C_4 plants would have appeared independently several times during evolution, as attested by phylogenetic reconstruction based on the different PEPC gene sequences (Lepiniec et al. 1994; Gehrig et al. 2001). Furthermore, multiple independent appearances of the grass C_4 photosynthetic pathway are supposed to have occurred (Sinha and Kellogg 1996) and, hence, C_4 PEPC isoforms would have appeared several times. However, only a limited number of grass C_4 PEPC sequences were then available.

Other molecular grass phylogenies have been reconstructed using different DNA sequences (Hamby and Zimmer 1988; Cummings et al. 1994; Nadot et al. 1994; Clark et al. 1995; Duvall and Morton 1996; Mathews and Sharrock 1996; Mason-Gamer et al. 1998; Soreng and Davis 1998; Gaut et al. 1999; Hilu et al. 1999; Hsiao et al. 1999; Mathews et al. 2000; Zhang 2000). The available molecular information is very useful for inferring the evolutionary mode of phenotypic traits inside this family (i.e. convergence, reticulated evolution, or divergence from a common ancestor). The comparison of morphological and molecular classifications has led to new insights about Poaceae evolution and tend to support grass C_4-pathway multiple appearances (e.g. Soreng and Davis 1998; Hsiao et al. 1999; Kellogg 2000). Nevertheless, such studies have rarely been based on genes that are directly involved in the analyzed phenotypes and could be spurious to simulate the evolution of an adaptive trait from the evolution of other *a priori* independent characters. To obtain a better understanding of the C_4-pathway evolution in grasses, phylogenetic relationships between C_4 grass species based on C_4 PEPC sequences should be assessed.

In the present study, we checked the usefulness of PCR primers to generate grass C_4 PEPC cDNA fragments. This enabled us to define the strategy to generate a long C_4 PEPC segment in a large sample of grasses displaying a C_4 photosynthetic pathway. From this cDNA segment, a phylogenetic approach was performed to estimate the evolutionary relationships between 22 C_4 grass species.

Materials and methods

Plant material

Wild, weed or cultivated species belonging to the family Poaceae were collected on Reunion Island (Table 1). A voucher sample for each of these plants was deposited at the herbarium of La Réunion University. We worked only on green leaves in which C_4 PEPC has been reported to be highly expressed in comparison to the other PEPC isoforms (Lepiniec et al. 1994; Dong et al. 1998; Besnard et al. 2001).

Molecular analysis

RNA extractions were carried out using the protocol employed by Atanassova et al. (1995). Reverse transcription was performed on total RNA as previously described by Cesari et al. (1999), using oligo-dT primers and the Reverse Transcriptase Kit (Gibco-BRL). To define PEPC specific primers, we compared C_4 and non-C_4 PEPC cDNA sequences, which have been isolated in sorghum, maize and sugarcane [EMBL accession numbers: for maize, X61489 (Kawamura et al. 1992), AB012228 (Dong et al. 1998), X03613 (Hudspeth and Grula 1989) and X15238 (Izui et al. 1986); for sorghum, X55664 (Crétin et al. 1991), X59925 (Lepiniec et al. 1991) and X17379 (Crétin et al. 1991); and for sugarcane, M86661 (Albert et al. 1992)]. Sequences were aligned using Clustal W software (Thompson et al. 1994). The cDNA length was approximately 3,300 bp for each isoform. C_4 specific primers were defined in the C_4 PEPC conserved sequences (see Table 2; Fig. 1). Thus, the primers were defined to specifically amplify C_4 PEPC cDNA fragments in grasses related to maize and sorghum. The PCR amplifications were performed using a thermocycler (PTC100-v7, MJ Research). The PCR reaction mixture contained 1 µl of cDNA template, 0.25 mM of dNTP, 10 mM of KCl, 10 mM of Tris-HCl (pH 8.3), 1.5 mM of $MgCl_2$, 20 pmol of each oligonucleotide primer, 1.25 µl (5% vol) of DMSO, and 1.2 units of Red Hot DNA polymerase (Advanced Biotechnologies) in a total volume of 25 µl. The samples were incubated for 4 min at 94 °C, followed by 36 cycles consisting of 50 s at 94 °C, 1 min at a temperature between 55 °C and 65 °C according to the primer combination employed, and 2 min at 72 °C (see Table 2). The last cycle was followed by a 6-min extension at 72 °C. PCR amplification products were separated in 2% agarose gels and were revealed under UV light after staining with ethidium bromide. We checked the PCR-amplification specificity of the five primer pairs which generated overlapping cDNA fragments on the entire translated C_4 PEPC mRNA plus a 3′ untranslated section of about 200 bp (see Fig. 1). Using the primer combinations number 2 and number 5, we generated and directly sequenced a fragment for some studied accessions. In most cases, we cloned a RT-PCR fragment before sequencing. We used the pGEM-T vector (PROMEGA) and *Escherichia coli* strain DH5α according to the provider's recommendations. Double-stranded DNA sequencing reactions were performed by the ESGS Society (Evry, France). Sequences were deposited in the EMBL databank.

Phylogenetic analyses

Known C_4 PEPC cDNAs from *Sorghum bicolor* (X17379), *Zea mays* (X03613, X15238) and *Chloris gayana* (AF268091: O.E. Blaesing et al., unpublished data) plus the cDNA segments generated using primer pair combination number 2 were aligned using Clustal W software (Thompson et al. 1994). C_4 PEPC sequences were compared to the root PEPC from maize (accession number AB012228) which was chosen as an outgroup since it was the non-C_4 PEPC grass sequence closest to the generated C_4 PEPC sequences (data not shown). Maximum-parsimony analysis was conducted using MEGA version 2 (Kumar et al. 2001). All characters were equally weighted and the sequence input order was jumbled five times. Gaps were treated as missing data. A heuristics search was used to find the most-parsimonious trees. Bootstrap values were computed using 100 replicates to evaluate support of the branches. A strict consensus tree was generated from the equally most-parsimonious trees. As an alternative to the parsimony analysis, the maximum-likelihood method (Felsenstein 1981) implemented in PHYLIP software was used. We considered a transition/transversion ratio of 2 and used empirical base frequencies.

Table 1 List, codes and EMBL accession numbers of the studied species. We followed the classification proposed by Clayton and Renvoize (1986). "(C_3)" indicated that the considered species displayed a C_3 photosynthetic pathway. Other species displayed a C_4 photosynthetic pathway. *Means that the fragment was previously cloned in pGEM-T vector before sequencing

Subfamily	Tribe	Species	Voucher reference	Code	EMBL accession number
Panicoideae Link.	Andropogoneae Dumort.	*Saccharum officinarum* L. (clone "*Big Tana Ray*")	–	A	AJ293346[c]
		Saccharum spontaneum L. (clone "*SES 14*")	–	B	AJ318338*[c]
		Sorghum verticilliflorum (Steud.) Stapf	G. Besnard no. 431	C	AJ318575*[a] AJ293347[b]
		Sorghum bicolor (L.) Moench	–	–	X17379 (Crétin et al. 1991)
		Vetiveria zizanioides (L.) Nash	G. Besnard no. 422	D	AJ318573*[a] AJ318583[b]
		Hemarthria altissima (Poir.) Stapf & C.E. Hubb.	G. Besnard no. 435	E	–
		Hyparrhenia rufa (Ness) Stapf	G. Besnard no. 434	F	AJ318580[a]
		Ischaemum koleostachys (Steud.) Hack.	G. Besnard no. 437	G	AJ318574*[a]
		Eulalia aurea (Borg) Kunth.	G. Besnard no. 436	H	AJ318576*[a] AJ318582[b]
		Themeda quadrivalvis (L.) Kuntze	G. Besnard no. 440	I	AJ318581[a]
		Heteropogon contortus (L.) P. Beauv. ex Roem. & Schult.	G. Besnard no. 441	J	AJ318577*[a]
		Pogonatherum paniceum (Lam.) Hack.	G. Besnard no. 444	K	AJ318578*[a]
		Zea mays L.	–	L	X03613[d], X15238[e]
		Coix lacryma-jobi L.	G. Besnard no. 429	M	AJ318579[a] AJ293348[b]
	Paniceae R. Br.	*Panicum maximum* Jacq.	G. Besnard no. 423	N	AJ318586*[a]
		Melinis repens (Willd.) Zizka	G. Besnard no. 426	O	AJ318585*[a]
		Paspalum urvillei Steud.	G. Besnard no. 421	P	–
		Paspalum paniculatum L.	G. Besnard no. 439	Q	AJ318587[a]
		Paspalidium geminatum (Forssk.) Stapf	G. Besnard no. 442	R	AJ318584*[a]
		Oplismenus compositus (L.) P. Beauv. – (C_3)	G. Besnard no. 446	S	–
Chloridoideae Burmeist.	Cynodonteae Dumort.	*Cynodon dactylon* (L.) Pers.	G. Besnard no. 428	T	AJ279036[a]
		Chloris barbata Sw.	G. Besnard no. 456	U	AJ318589*[a]
		Chloris gayana Kunth	–	–	AF268091[f]
	Eragrostideae Stapf	*Eleusine indica* (L.) Gaertn.	G. Besnard no. 427	V	AJ318591[a]
		Eragrostis tenuifolia (A. Rich.) Hochst. ex Steud.	G. Besnard no. 443	W	AJ318590*[a]
Arundinoideae Burmeist.	Aristideae C.E. Hubb.	*Aristida mauritiana* Kunth.	G. Besnard no. 449	X	AJ318588[a]
Pooideae Benth.	Aveneae Dumort.	*Holcus lanatus* L. – (C_3)	G. Besnard no. 432	Y	–

[a] Fragment generated using primer combination no. 2
[b] Fragment generated using primer combination no. 5
[c] Complete translated sequence obtained using the primers "0F" and "3210F"
[d] From Hudspeth and Grula (1989)
[e] From Izui et al. (1986)
[f] From O.E. Blaesing et al., unpublished data

Substitution rates

We computed the rates of nonsynonymous and synonymous substitutions between pairs of sequences from sorghum and maize independently for C_4, root and housekeeping PEPC isoforms. For these computations, we used the PEPC fragment number 2 and did not consider the codons in the segment displaying indels. We calculated these estimations using the method of Nei and Gojobori (1986). The heterogeneity in the mutation rates between each PEPC isoform was thus estimated.

Results and discussion

Specificity of primers

In Table 2, we reported the PCR amplification specificity of the primers. Each primer combination did not lead to amplification in all the species. For instance, the primer combinations numbers 4 and 5 (see Table 2; Fig. 1) led to amplification in all species belonging to the tribe Andropogoneae but not in the other species (except in the genus

Table 2 Codes, names, sequences, annealing temperatures (T) and specificity of the different primer pairs

Primer pair code	Primer name	5′–3′ sequence	Fragment size[a]	T (°C)	Species (code correspondence in Table 1)																								
					A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	0F	CCG MGG MSC KCC ATG GCG TC	781	55	+[b]	+	+	+	–[c]	–	–	+	+	+	–	+	+	–	–	–	–	–	–	–	–	–	–	+	–
	800R	GGG CTG TAT CCA CAC GGC GC																											
2	500F	GTT CGA GGC GCT CAA GAA CCA G	1,237	55	+	+	+	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+	+	–
	1700R	CTG CAG GTC GGC CAG CCT CTC																											
3	1400F	CCG ACG TGA TCG ACG CCA TCA C	1,207	60	+	+	+	+	+	+	+	–	+	+	+	+	+	+	+	–	–	+	–	–	+	+	+	+	–
	2550R	ACG TAT TTG TCC CTG AGC TGC TTC																											
4	2000F	CGC CGG ACA CCA TCA ACG GG	831	65	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–	–	–	–	–	–	–	–	–	–
	2780R	CGG CGG GCT TGT TCT CGT CG																											
5	2550F	GAA GCA GCT CAG GGA CAA ATA CGT	546	55	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–	–	+	–	–	–	–	–	–	–

Table 2 (continued)

Primer pair code	Primer name	5′–3′ sequence	Fragment size[a]	T (°C)	Species (code correspondence in Table 1) A B C D E F G H I J K L M N O P Q R S T U V W X Y
5	3210R	ATG CCA AGA TTT TCC ACT TGG AC			

[a] PCR fragment size observed in *S. officinarum*
[b] "+" means that fragment amplification occurred
[c] "–" means that fragment amplification did not occur

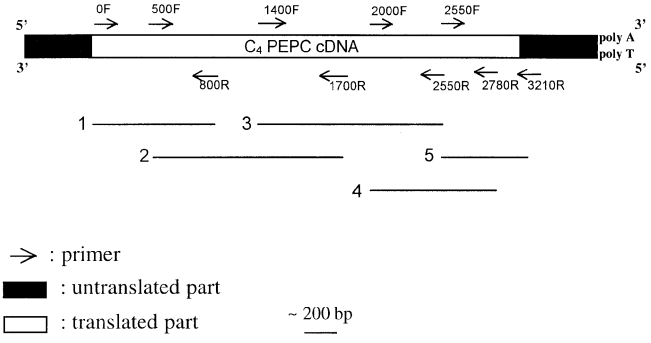

Fig. 1 PCR primer position on the C_4 PEPC cDNA sequence. The expected PCR fragments (coded 1 to 5) are indicated under the cDNA drawing

Paspalidium). In contrast, the primer combination number 2 led to amplification in most of the studied genera. Generally, the PCR amplification success was correlated with the close relationships of the analyzed species with maize or sorghum, in which the primers were defined. Indeed, the five primer pairs were successfully used only for the Andropogoneae species (Table 2: columns A, B, C, D, I, J, L, M). Finally, all primer pairs failed to amplify fragments in the genera *Oplismenus* and *Holcus* (Table 2: columns S and Y), which display a C_3 photosynthetic pathway. Thus, we can assume that all amplified fragments corresponded to partial C_4 PEPC cDNAs. Moreover, their expression in green leaves and their high homology with the grass C_4 PEPCase isoform (see below) also supported this fact (Besnard et al. 2002)

C_4 PEPC sequence analysis

We chose the primer combinations numbers 2 and 5 to generate C_4 PEPC cDNA sequences from some taxa. The choice of the first primer pair (number 2) was selected by the fact that it can generate fragments in C_4 grass accessions belonging to all studied subfamilies (Table 2). In contrast, the second primer pair (number 5) enabled us to estimate the level of variation in the 3′ untranslated part of the cDNAs in a few Andropogoneae. Sequencing revealed a mixing of alleles in *Saccharum spontaneaum*, *Heteropogon*, *Melinis* and *Eragrostis* (data not shown) and thus PCR fragment cloning was sometimes required to generate exploitable sequences. We obtained 18 sequences for fragment number 2 and four sequences with fragment number 5 (Table 1). In addition, we generated complete translated cDNA for *Saccharum* species using primers "0F" and "3210R" (Fig. 1).

All the sequences displayed a high level of homology with known grass C_4 PEPCs (between 78% and 99%). Homology with grass non-C_4 PEPCs ranged between 70 and 75%. A high level of recombination in the 3′ untranslated part was found between the sequences of fragment number 5 (data not shown). In this cDNA segment, a very low homology level with non-C_4 PEPC isoforms was also found (not computable because of the very high recombination frequency). Consequently, the 3′ cDNA part is not appropriate for phylogenetic reconstruction. On the other hand, a few recombinations were found in fragment number 2 between the non-C_4 and C_4 PEPCs. Only four indels (Fig. 2) occurred in a small region (located between nucleotides 999 and 1,041 of the sugarcane PEPC accession AJ293346). Interestingly, these indels appeared to be specific to the subfamily Chloridoideae (indel 1) and to the tribe Andropogoneae (indels 2 and 3; Fig. 2). *Aristida* (Arundinoideae) also displayed a specific indel (indel 4; Fig. 2). Moreover, the four indels appeared as parsimony informative in the phylogenetic analyses (see below; Fig. 3). Hilu and Alice (1999) and Zhang (2000) have already reported such observations using indels in the chloroplast *matK* and *rpl16* genes.

Phylogenetic structure of grasses using C_4 PEPC sequences

For phylogenetic reconstruction, we considered cDNA fragment number 2 using a segment of 1,127 nucleotides (after alignment). This segment is located between nucleotides 516 and 1,639 in the sugarcane sequence AJ293346. Using maximum-parsimony analysis, we obtained two equally parsimonious trees that differ with respect to the placement of *Heteropogon* as either sister to

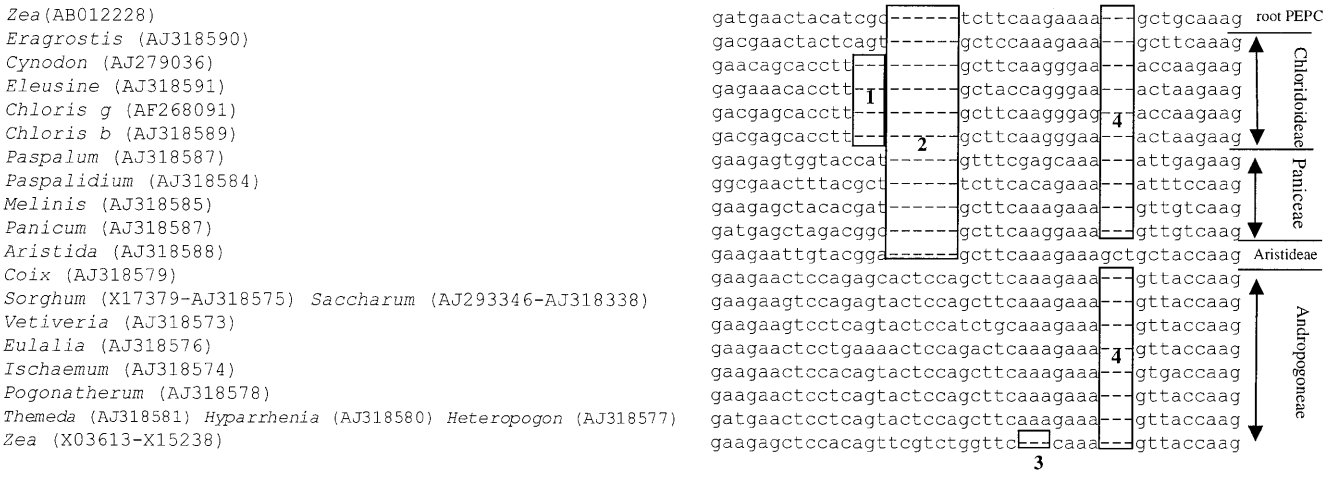

Fig. 2 Alignment of a nucleotidic sequence that flank a cDNA PEPC region displaying indels. This segment (inside fragment number 2) is located between the nucleotides 999 and 1,041 in the sugarcane sequence (accession AJ293346). The informative indels are boxed and numbered 1 to 4

Table 3 Non-synonymous and synonymous mutation rates computed from the segment number 2 by comparison of pair of sequences from sorghum and maize. These parameters were independently computed for each PEPC isoform using the method of Nei and Gojobori (1986)

Pair of sequences (maize/sorghum)	Non-synonymous mutation rate (d_N)	Synonymous mutation rate (d_S)
C_4 PEPC isoform: X15238/X17379	0.063	0.177
C_4 PEPC isoform: X03613/AJ318575	0.067	0.167
C_4 PEPC isoform: X03613/X17379	0.067	0.167
C_4 PEPC isoform: X15238/AJ318575	0.063	0.177
Root PEPC isoform: AB012228/X55664	0.005	0.110
Housekeeping PEPC isoform: X61489/X59925	0.011	0.130

Themeda or *Hyparrhenia*. Minimal-length trees have 1,678 steps and a consistency index (CI) of 0.47 (excluding autapomorphic characters) and a retention index (RI) of 0.61. Bootstrap analysis revealed also that relationships between *Vetiveria*, *Coix* and *Eulalia* were not clearly assigned. The consensus tree is given in Fig. 3A. The maximum-likelihood method led to a very similar phylogram (Fig. 3B). Phenetic reconstruction, based on Jukes and Cantor or Kimura distances, and the Neighbor-joining algorithm also gave similar structures (data not shown). Phylogenetic analyses enabled us to recognize the subfamily Chloridoideae and the tribe Andropogoneae, as in previous molecular studies based on chloroplast or nuclear gene sequences (Mathews and Sharrock 1996; Soreng and Davis 1998; Hilu et al. 1999; Hsiao et al. 1999; Mathews et al. 2000). In addition, the genus *Aristida* displayed a basal position in the PACC clade, and the genus *Eragrostis* displayed a basal position in comparison to the genera *Cynodon* or *Chloris* in the subfamily Chloridoideae. Similar observations were already made using the gene sequences of ITS (Hsiao et al. 1999), phytochrome B (Mathews et al. 2000) or *matK* (Hilu et al. 1999), and cpDNA restriction sites (Soreng and Davis 1998). This shows that chloroplast and nuclear DNA sequences give very similar results for the classification of the species analyzed in our study. Nevertheless, the tribe Paniceae displayed an intermediary position between the Chloridoideae and the Andropogoneae. Indeed, the genus *Paspalum* was more related to the tribe Andropogoneae, whereas the genera *Melinis*, *Panicum* and *Paspalidium* were more related to the subfamily Chloridoideae. ITS analysis has led to a similar conclusion (Hsiao et al. 1999). To explain such an observation, we can hypothesize that the Chloridoideae and the Panicoideae have been derived from an arundinoid-like common ancestor (Hsiao et al. 1999) and that the tribe Paniceae has diverged early and maintained a higher genetic diversity.

Grass PEPC evolution rate

Variations in the mutation rates between each PEPC isoform were detected for non-synonymous mutations (Table 3). Indeed, sorghum and maize C_4 PEPC isoforms displayed higher non-synonymous mutation rates (d_N=0.063 to 0.067) than sorghum and maize non-C_4 PEPC isoforms (d_N=0.005 and 0.011). This tends to confirm the faster evolution of C_4 PEPC compared to non-C_4 PEPC isoforms (Besnard et al. 2002). Such heterogeneous gene evolution has already been shown for other multigene families (Zhang et al. 2001). In our case, we can assume that selection pressure should be relatively weak on the C_4 PEPC isoform because the C_4 system is an adaptive trait not essential for plant subsistence. Alternatively, it can also be suggested that, since the

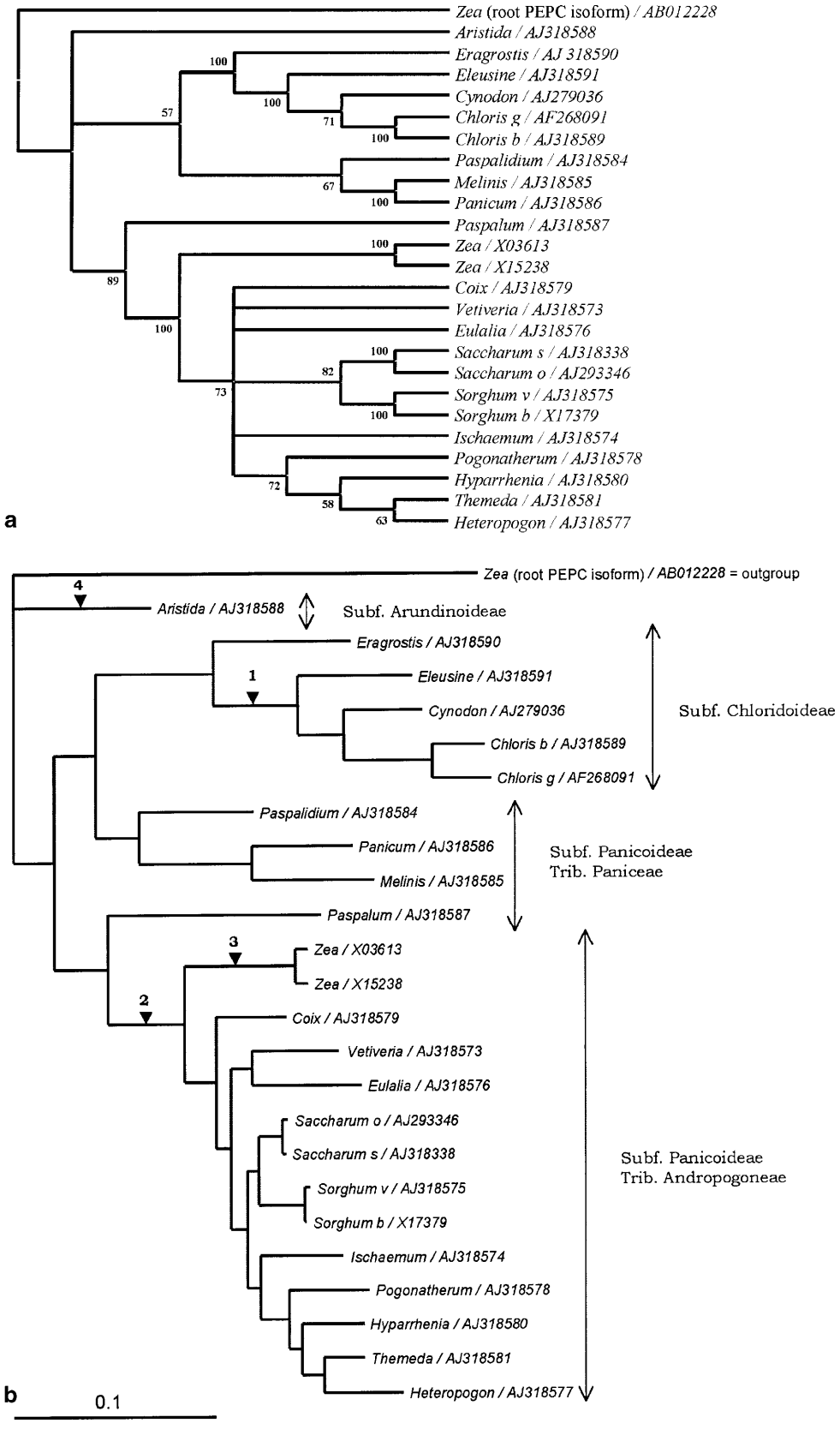

Fig. 3 Phylogenetic trees based on grass C_4 PEPC partial cDNA sequences (1,127 nucleotides, 428 informative characters): **a** Strict consensus of two equally parsimonious trees (1,678 steps: CI = 0.47, RI = 0.61; excluding autapamorphic characters). Bootstrap values were obtained from 100 replicates and were indicated on each corresponding node when superior to 55%; **b** Phylogenetic tree revealed using the maximum-likelihood method. Probable position of the insertion-deletion events is indicated on the branches by the arrows. Indel numbers are defined in Fig. 2. The maize root PEPC isoform (accession AB012228) was used as an outgroup in each analysis

C_4 system has recently appeared in comparison to anaplerotic functions in which other PEPCs are involved (Lepiniec et al. 1994), a high selective divergence could have consequently occurred to improve the efficacy of the C_4 photosynthetic pathway.

Conclusion

In our work, we showed that primer pair combination number 2 allowed us to generate a C_4 PEPC cDNA segment in most C_4 grasses. Based on this fragment, we also demonstrated that the phylogenetic relationships between the studied subfamilies or tribes were similar to those deduced from other molecular data. Of course, we must keep in mind that this isoform, characteristic of C_4 species, is limited to PACC grass systematics. Nevertheless, for the phylogeny of the PEPC multigenic family, it will be possible to study C_4 photosynthesis evolution in grasses and check the existence of convergent evolution (Sinha and Kellogg 1996). This will give some insights into the pertinence of the C_4 photosynthetic trait for grass classification (Watson and Dallwitz 1992) and this issue will be discussed in another forthcoming paper (Besnard et al. 2002; Besnard and Offmann, in preparation).

Acknowledgements We thank Dr. J. Bosser, Dr. C. Lavergne, Dr. D. Strasberg, Pr. J. Figier and Pr. T. Pailler for their contribution to the prospects and identification of all grass samples. We particularly thank Dr. G. Pinçon for her technical support and comments.

References

Albert HA, Martin T, Sun SSM (1992) Structure and expression of a sugarcane gene encoding a housekeeping phospoenolpyruvate carboxylase. Plant Mol Biol 20:663–671

Atanassova R, Favet N, Martz F, Chabbert B, Tollier MT, Monties B, Fritig B, Legrand M (1995) Altered lignin composition in transgenic tobacco expressing *O*-methyltransferase sequences in sense and antisense orientation. Plant J 8:465–477

Besnard G, Offmann B, Robert C, Baret P, Rouch C, Cadet F (2002) Phosphoenolpyruvate carboxylase cDNA phylogeny to investigate the C_4 photosynthetic pathway evolution in grasses. Proc XIIth Int Congress of Photosynthesis, Brisbane, Australia, August 2001 (in press)

Besnard G, Robert C, Rondeau P, Baret P, Rouch C, Cadet F (2001) Phosphoenolpyruvate carboxylase gene expression in the sugarcane. Proc 2nd Symp Sci Res Indian Ocean, Reduit, Mauritius, April 2001, J Nature (in press)

Cesari M, Caillens H, Cadet F, Pabion M (1999) In vivo analysis of *HLA-DQ* gene expression in heterozygous cell lines. Immunogenetics 50:309–318

Clark LG, Zhang W, Wendel JF (1995) A phylogeny of the grass family (Poaceae) based on *ndh*F sequence data. Syst Bot 20:436–460

Clayton WD, Renvoize SA (1986) Genera Graminum: grasses of the world. Kew Bull 13:1–389

Clegg MT, Cummings MP, Durbin ML (1997) The evolution of plant nuclear genes. Proc Natl Acad Sci USA 94:7791–7798

Crétin C, Santi S, Keryer E, Lepiniec L, Tagu D, Vidal J, Gadal P (1991) The phosphoenolpyruvate carboxylase gene family in sorghum: promoter structure, amino acid sequences and expression of genes. Gene 99:87–94

Cummings MP, King LM, Kellogg EA (1994) Slipped-strand mispairing in a plastid gene: *rpo*C2 in grasses (Poaceae). Mol Biol Evol 11:1–8

Davis JI, Soreng RJ (1993) Phylogenetic structure in the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. Am J Bot 80:1444–1454

Dong L, Masuda T, Kawamura T, Hata S, Izui K (1998) Cloning, expression and characterization of a root-form phosphoenolpyruvate carboxylase from *Zea mays*: comparison with the C_4-form enzyme. Plant Cell Physiol 39:865–873

Duvall MR, Morton BR (1996) Molecular phylogeny of Poaceae: an expanded analysis of *rbc*L sequence data. Mol Phylogenet Evol 5:352–358

Felsenstein J (1981) Evolution trees from DNA sequences: a maximum-likelihood approach. J Mol Evol 17:368–376

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

Gehrig HH, Heute V, Kluge M (2001) New partial sequences of phosphoenolpyruvate carboxylase as molecular phylogenetic markers. Mol Phylogenet Evol 20:262–274

Hamby RK, Zimmer EA (1988) Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). Plant Syst Evol 160:29–37

Hatch MD, Slack CR (1966) Photosynthesis by sugar cane leaves. A new carboxylation reaction and the pathway of sugar formation. Biochem J 101:103–111

Hilu KW, Alice LA (1999) Evolutionary implications of *matK* indels in Poaceae. Am J Bot 86:1735–1741

Hilu KW, Alice LA, Liang H (1999) Phylogeny of Poaceae inferred from *matK* sequences. Ann Mo Bot Gard 86:835–851

Hsiao C, Jacobs SWL, Chatterton NJ, Asay KH (1999) A molecular phylogeny of the grass family (Poaceae) based on the sequences of nuclear ribosomal DNA (ITS). Austral Syst Bot 11:667–688

Hudspeth RL, Grula JW (1989) Structure and expression of the maize gene encoding the phosphoenolpyruvte carboxylse isozyme involved in C_4 photosynthesis. Plant Mol Biol 12:579–589

Izui K, Ishijima S, Yamaguchi Y, Katagiri F, Murata T, Shigesada K, Sugiyama T, Katsuki H (1986) Cloning and sequence analysis of cDNA encoding active phosphoenolpyruvate carboxylase of the C_4-pathway from maize. Nucleic Acids Res 14:1615–1628

Kawamura T, Shigesada K, Toh H, Okumura S, Yanagisawa S, Izui K (1992) Molecular evolution of phosphoenol pyruvate carboxylase for C_4 photosynthesis in maize: comparison of its cDNA sequence with a newly isolated cDNA encoding an isozyme involved in the anaplerotic function. J Biochem 112:147–154

Kellogg EA (2000) The grasses: a case study in macroevolution. Annu Rev Ecol Syst 31:217–238

Ku MSB, Kano-Murakami Y, Matsuoka M (1996) Evolution and expression of C_4 photosynthesis genes. Plant Physiol 111:949–957

Ku MSB, Agarie S, Nomura M, Fukayama H, Tsuchida H, Ono K, Hirose S, Toki S, Miyao M, Matsuoka M (1999) High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants. Nat Biotechnol 17:76–80

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Biometrics 17:1244–1245

Lepiniec L, Santi S, Keryer E, Amiet V, Vidal J, Gadal P, Crétin C (1991) Complete nucleotide sequence of one member of the *Sorghum* phosphoenolpyruvate carboxylase gene family. Plant Mol Biol 17:1077–1079

Lepiniec L, Vidal J, Chollet R, Gadal P, Crétin C (1994) Phosphoenolpyruvate carboxylase: structure, regulation and evolution. Plant Sci 99:111–124

Mason-Gamer RJ, Weil CF, Kellogg EA (1998) Granule-bound starch synthase: structure, function and phylogenetic utility. Mol Biol Evol 15:1658–1673

Mathews S, Sharrock RA (1996) The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot Angiosperms. Mol Biol Evol 13:1141–1150

Mathews S, Tsai RC, Kellogg EA (2000) Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. Am J Bot 87:96–107

Nadot S, Bajon R, Lejeune B (1994) The chloroplast gene *rps*4 as a tool for the study of Poaceae phylogeny. Plant Syst Evol 191:27–38

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Sinha NR, Kellogg EA (1996) Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. Am J Bot 83:1458–1470

Soreng RJ, Davis JI (1998) Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction site character sets. Bot Rev 64:1–85

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalities and weight matrix choice. Nucleic Acids Res 22:4673–4680

Watson L, Dallwitz MJ (1992) Grass genera of the world: descriptions, illustrations, identification, and information retrieval. http://biodiversity.uno.edu/delta/

Zhang L, Pond SK, Gaut BS (2001) A survey of the molecular evolutionary dynamics of twenty-five multigene families from four grass taxa. J Mol Evol 52:144–156

Zhang W (2000) Phylogeny of the grass family (Poaceae) from *rpl16* intron sequence data. Mol Phylogenet Evol 15:135–146