Theor Appl Genet (2003) 107:470–478
DOI 10.1007/s00122-003-1268-2

G. Besnard · G. Pinçon · A. D'Hont · J.-Y. Hoarau · F. Cadet · B. Offmann

Characterisation of the phosphoenolpyruvate carboxylase gene family in sugarcane (*Saccharum* spp.)

Received: 17 October 2002 / Accepted: 4 December 2002 / Published online: 21 May 2003
© Springer-Verlag 2003

Abstract Phosphoenolpyruvate carboxylases (PEPCs) are encoded by a small multigenic family. In order to characterise this gene family in sugarcane, seven DNA fragments displaying a high homology with grass PEPC genes were isolated using polymerase chain reaction-based cloning. A phylogenetic study revealed the existence of four main PEPC gene lineages in grasses and particularly in sugarcane. Moreover, this analysis suggests that grass C_4 PEPC has likely derived from a root pre-existing isoform in an ancestral species. Using the Northern-dot-blot method, we studied the expression of the four PEPC gene classes in sugarcane cv. R570. We confirmed that transcript accumulation of the C_4 PEPC gene (ppc-C_4) mainly occurs in the green leaves and is light-induced. We also showed that another member of this gene family (ppc-aR) is more highly transcribed in the roots. The constitutive expression for a previously characterised gene (ppc-aL2) was confirmed. Lastly, the transcript accumulation of the fourth PEPC gene class (ppc-aL1) was not revealed. Length polymorphism in non-coding regions for three PEPC gene lineages enabled us to develop sequence-tagged site PEPC markers in sugarcane. We analysed the segregation of PEPC fragments in self-pollinated progenies of cv. R570 and found co-segregating fragments for two PEPC gene lineages. This supports the hypothesis that diversification of the PEPC genes involved duplications, probably in tandem.

Keywords C_4 photosynthesis · Evolution · Gene family · Mapping · PEPC · STS · Sugarcane

Communicated by H.F. Linskens

G. Besnard (✉) · G. Pinçon · F. Cadet · B. Offmann
Université de la Réunion, LBGM, 15 Avenue R. Cassin,
97715 St-Denis Messag Cedex 9, La Réunion, France
e-mail: gbesnar@hotmail.com
Fax: +33-383-903277

A. D'Hont
CIRAD, TA 40/03, Avenue Agropolis,
34398 Montpellier Cedex 5, France

J.-Y. Hoarau
CERF, B.P. 315, 97494 Ste-Clotilde, La Réunion, France

Present address:
G. Besnard, UMR INRA/UHP 1136 Iam, Université Nancy I,
Bd des Aiguillettes, B.P. 239, 54506, Vandœuvre Les Nancy,
France

Introduction

Sugarcane (*Saccharum* spp.) is an important crop in tropical areas. It represents the first source of sugar in the world. This crop displays a NADP-ME C_4 photosynthetic pathway, similar to that of maize and sorghum (subfamily Panicoideae, tribe Andropogoneae; Clayton and Renvoize 1986), but our knowledge about sugarcane C_4 photosynthesis enzymes is still limited, particularly on the encoding genes and their expression regulation. Nevertheless, the characterisation of C_4 enzymes, especially the C_4 phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31), has been studied in a few plants, notably in two related grass species, sorghum and maize (Ku et al. 1996; Sheen 1999). It appears that light regulates transcript accumulation of C_4 genes but that enzyme post-transductional modifications are also involved (Vidal and Chollet 1997; Kai et al. 1999; Bläsing et al. 2000). Our question is whether such a model can be verified in sugarcane.

A small multigenic family encodes the PEPC enzymes, which are involved in different functions (Lepiniec et al. 1994; Gehrig et al. 2001). PEPC enzymes enable atmospheric CO_2 fixation in the C_4 photosynthetic pathway (Hatch 1999) and are also involved in anaplerotic functions (Lepiniec et al. 1994; Gehrig et al. 2001). Each PEPC isoform is then specifically expressed (Lepiniec et al. 1994; Dong et al. 1998). Thus, gene transcript accumulation either occurs mainly in an organ (C_4 PEPC in the green leaf, root PEPC in the root) or is constitutive (housekeeping PEPC). Some recent studies have demonstrated that the C_4 PEPC isoform displays a potential biotechnological interest (Ku et al. 1999; Matsuoka et al.

2001). The PEPC gene family has also been specifically studied for evolutionary purposes, in order to understand the origins of C_4 and CAM photosynthesis (Gehrig et al. 2001; Besnard et al. 2002b; Bläsing et al. 2002). In grasses, isoforms of PEPC have been characterised in different species, such as wheat (Gonzalez et al. 1998; 2002), sugarcane (Albert et al. 1992), maize (Izui et al. 1986; Kawamura et al. 1992; Dong et al. 1998) and sorghum (Crétin et al. 1991; Lepiniec et al. 1991). Three functional isoforms (namely C_4, root and housekeeping PEPCs) were identified in this plant family (for a review, see Gehrig et al. 2001).

The evolution of the PEPC multigene family is not well-elucidated. It is assumed that gene duplication from pre-existing genes, a few amino acid changes and the acquisition of a new gene transcription control have led to the appearance of new isoforms such as C_4 PEPC (Sheen 1999; Bläsing et al. 2002). In addition, the genomic organisation of PEPC genes is still unknown, and new information could bring some important insights on the PEPC gene family evolution. In sugarcane, a non-photosynthetic PEPC isoform is known (M86661; Albert et al. 1992). It has been supposed that this enzyme controls an anaplerotic function since it is constitutively expressed. It also displays a particular phylogenetic position in comparison with other grass PEPCs (Besnard et al. 2002a). Thus, we chose to analyse the PEPC gene family in sugarcane for two reasons: the system provides some opportunities for such a study (gene mapping, sequencing) and because this plant displays a specific PEPC member (M86661) which is not present in sorghum and maize.

We studied the diversity and the organisation of the PEPC gene family in sugarcane using cDNA or genomic DNA fragments encoding PEPCs to deduce phylogenetic relationships between four PEPC gene lineages. We compared transcript accumulation levels of each one in different organs using Northern blot analysis and confirmed the light-induced transcript accumulation of the C_4 PEPCs in the leaf. We showed that one PEPC gene was mainly transcribed in the roots and another was constitutively transcribed. We also developed STS (sequence-tagged site) genetic markers of three PEPC gene lineages that were located on a sugarcane genetic map. Based on these results, the mode of evolution of the PEPC gene family is discussed.

Materials and methods

Vegetal material and RNA extractions

Sugarcane is a polyploid species deriving from interspecific hybrids between *S. officinarum* L. and a few accessions of *S. spontaneum* L. (D'Hont et al. 1996; Irvine 1999). We used sugarcane cultivar R570 and two individuals belonging to *Saccharum officinarum* (accession Big Tana Rayé) and *S. spontaneum* (accession SES14). We analysed gene transcript accumulation in leaves, stem, roots, and etiolated leaves of cv. R570. Etiolated leaves were taken from plants that had been germinated and grown in total darkness for 3 weeks. RNA was extracted from each organ studied using the protocol described by Atanassova et al. (1995).

Design of primers for PCR cloning

To define PEPC specific primers, we compared complete C_4, root and housekeeping PEPC cDNAs from grasses (Table 1). Sequences were aligned using CLUSTAL W software (Thompson et al. 1994).

Table 1 Species origin, EMBL accession number, reference and function of the complete or partial PEPCs used in this study

	Species	Accession[a] no.	Reference	Function (Gehrig et al. 2001)
Monocotyledonous (Poaceae)	*Zea mays*	X61489 [c]	Kawamura et al. 1992	Housekeeping isoform (ppc-aL)
		AB012228[c]	Dong et al. 1998	Root isoform (ppc-aR)
		X03613[p]	Izui et al. 1986	C_4 isoform (ppc-C_4)
		X15642[c]	Matsuoka and Minami 1989	C_4 isoform (ppc-C_4)
		X15238[c]	Hudspeth and Grula 1989	C_4 isoform (ppc-C_4)
		X15239[c]	Hudspeth and Grula 1989	C_4 isoform (ppc-C_4)
	Sorghum bicolor	X55664[c]	Crétin et al. 1991	Root isoform (ppc-aR)
		X59925[c]	Lepiniec et al. 1991	Housekeeping isoform (ppc-aL)
		X17379[c]	Crétin et al. 1991	C_4 isoform (ppc-C_4)
	Sorghum verticilliflorum	AJ293347[p]	Besnard et al. 2002b	C_4 isoform (ppc-C_4)
	Saccharum spp.	M86661[c],	Albert et al. 1992	Housekeeping isoform (ppc-aL)
	Triticum aestivum	AJ007705[c]	Gonzalez et al. 2002	Housekeeping isoform (ppc-aL)
		Y15897[p]	Gonzalez et al. 1998	?[b]
	Oryza sativa	AF271995[c]	N. Yamamoto, unpublished	Root isoform?
	Chloris gayana	AF268091[c]	Bläsing et al. 2002	C_4 isoform (ppc-C_4)
	Setaria italica	AF495586[c]	Z. Ding et al., unpublished	C_4 isoform (ppc-C_4)
	Coix lacryma-jobi	AJ293348[p]	Besnard et al. 2002b	C_4 isoform (ppc-C_4)
	Vetiveria zizanioides	AJ318583[p]	Besnard et al. 2002b	C_4 isoform (ppc-C_4)
	Eulalia aurea	AJ318582[p]	Besnard et al. 2002b	C_4 isoform (ppc-C_4)
Dicotyledonous	*Flaveria trinervia*	X61304[c]	Poetsch et al. 1991	C_4 isoform (ppc-C_4)
		X64143[c]	Hermans and Westhoff 1992	?
Gymnosperm	*Picea abies*	X79090[c]	Relle and Wild 1996	?

[a] c, Complete coding sequence; p, partial coding sequence (3′ part of cDNA)
[b] ?, the function of the considered sequence has not been accurately defined

Table 2 Code, name, sequence and annealing temperature (T) of the designed primers used for PCR-based cloning and to generate specific probe fragments

Code	Primer name	5′ → 3′ sequence	T (°C)
1	PEPC/C_4-0F	CCGMGGMSCKCCATGGCGTC	55
2	PEPC/C_4-3200R	ATGCCAAGATTTTCCACTTGGAC	55
3	PEPC/1-F	GGTTTTCGCCAAGGGAGAC	52
4	PEPC/1-R	TGCATCARGCTTCAACAGCC	52
5	PEPC/2-F	GTAACCCTGGCATTGCTGCAC	52
6	PEPC/2-R	ACGGTGTAGAGAGACGGCCG	52
7	PEPC/3-3000F	CCGGCTGAAAGAACAGGATCC	52
8	PEPC/3-3250R	CACCGTAACAAGATGCCCTGC	52
9	PEPC/C_4-Fprobe	CACCGGCTAGGCCGCTTCCC	55
10	PEPC/1-Fprobe	TGATCCTGACCATGAAGGGC	52
11	PEPC/2-Fprobe	GACCATGAAGGGCATAGCAG	52
12	rDNA18SF	GTAGGATAGGGGCCTACCATG	55
13	rDNA18SR	TACCAGACTCTAACGCCCGG	55

Firstly, primers specific for grass C_4 PEPC genes were defined in conserved segments in the 5′ and 3′ parts flanking the translated sequence of sorghum and maize C_4 PEPC cDNAs (Table 2, primers 1 and 2). Then, two primer pairs were defined to specifically amplified non-photosynthetic PEPC genes analogue to those of grasses – primers 3 and 4 to amplify PEPC fragments homologous to the EMBL accessions X55664/AB012228/AF271995, and primers 5 and 6 to amplify PEPC fragments homologous to the EMBL accessions X59925/X61489/AJ007705 (Table 2). Lastly, specific primers (Table 2, primers 7 and 8) were defined to generate the untranslated 3′ cDNA fragment from a housekeeping sugarcane PEPC (EMBL accession M86661), which has been previously characterised by Albert et al. (1992).

Reverse transcription, polymerase chain reaction (PCR) and sequencing

Reverse transcription reaction was performed with the Reverse Transcriptase kit (Gibco-BRL, Gaithersburg, Md.) in 20 μl using 3 μg of total RNA. PCR-based cloning of the PEPC fragments was performed either from total cDNA or from genomic DNA (gDNA). We used the PCR conditions previously defined by Besnard et al. (2002b). Using the primer combinations 1–2 and 7–8, we generated cDNA PEPC fragments for both *S. officinarum* and *S. spontaneun*, whereas primer pairs 3-4 and 5-6 were used to amplify gDNA PEPC fragments of cv. R570. Due to a mixing of allelic sequences in the PCR products, we cloned a few fragments using the pGEM®-T vector according to the manufacturer's recommendations (Promega, Madison, Wis.). Double-stranded DNA sequencing reactions were performed by the ESGS society (Evry, France).

Sequence analysis

Phylogenetic relationships between DNA sequences encoding grass PEPCs were analysed. PEPC cDNAs from *Picea abies* (Gymnosperm) and *Flaveria trinervia* (Dicotyledonous) were arbitrary used as outgroups (Table 1). Phylogenetic analysis was conducted using MEGA version 2 (Kumar et al. 2001). The distance between each pair of sequences was computed according to Jukes and Cantor (1969). Phylogenetic trees were then constructed using the Neighbor joining algorithm (Saitou and Nei 1987). Bootstrap values were computed using 1,000 replicates to evaluate the support of branches. For phylogenetic reconstruction, we considered either complete cDNA sequences or only the 330 coding-nucleotides in the 3′ terminal part.

Northern procedure

To quantify differential PEPC RNA accumulation between different organs (i.e. leaf, etiolated leaf, root and stem), we prepared a Northern dot-blot membrane. For each organ of cv. R570, 10 μg of total RNA was directly located onto N^+-Hybond nylon membrane (Amersham, UK). Probes were generated by PCR using genomic DNA from cv. R570. The PEPC probes used were chosen in the 3′ untranslated region of the genes because this part was shown to be very variable between the different PEPC genes. Indeed, each probe has led to a different restriction fragment length polymorphism (RFLP) pattern in sugarcane (G. Besnard, unpublished data), suggesting that each probe is specific for a PEPC gene lineage as already has been shown in maize or sorghum (Kawamura et al. 1992; Dong et al. 1998). A specific C_4 PEPC fragment was generated using the primer pair 9-2, specific non-C_4 PEPC fragments were generated using primer pairs 7–8, 4–10 and 6–11 and, lastly, an 18S rDNA fragment was generated using the primer pair 12–13. Probes were labelled by α-[^{32}P]-dCTP using the protocol described by Besnard et al. (2000). Northern dot-blot membranes were successively hybridised with the five probes. Ribosomal 18S gene expression was considered as to be constitutive and proportional to the amount of deposited RNA.

Development of STS markers and PEPC gene mapping

From the sugarcane PEPC sequences obtained in this study (see below), we developed genetic STS markers. A primer pair was designed to specifically amplify the last intron, the last exon plus a part of the 3′ untranslated region for each PEPC gene class (Table 3). This enabled us to reveal length polymorphism (see below). PCR amplifications were performed using a thermocycler (I-cycler, BioRad, Hercules, Calif.). For each primer pair checked, one primer was labelled in 5′ with the fluorochrome IRD800 (MWG-Biotech). The PCR reaction mixture contained 60 ng of total DNA template, 2.5 μl (10% vol) DMSO, 1 × reaction buffer, 0.2 mM dNTP, 0.2 μmol of each oligonucleotide primer and 0.75 U DNA polymerase (Invitrogen, Carlsbad, Calif.) in a total volume of 25 μl. The 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 or 55 °C, according to the primer combination used (Table 3), and 90 s at 72 °C. The last cycle was followed by a 6-min extension at 72 °C. Electrophoresis of the sequence-tagged site (STS) markers was carried out on an automated sequencer DNA analyser GeneReadIR 4200 (LiCor, Lincoln, Neb.) according to the manufacturer's recommendations. We used 6% acrylamide denaturing gels and TBE 1 × migration buffer. We characterised 112 self-pollinated progenies derived from R570 (Hoarau et al. 2001). Simplex (3:1) alleles were scored and positioned on the cv. R570 amplified fragment length polymorphism (AFLP) genetic map (Hoarau et al. 2001). The linkage relationships of PEPC simplex markers in coupling phase were determined using the MAPMAKER software (Lander et al. 1987). Two-point analyses were performed at a LOD score of 5. Multipoint analysis allowed us to order the markers inside a linkage group.

The usefulness of the primer pairs was also checked on three other species belonging to the PACCAD clade (GPWG 2001): *Sorghum bicolor* (L.) Moench (Andropogoneae, C_4), *Digitaria*

Table 3 Locus code, corresponding EMBL accessions, sequence and annealing temperature (T) of the designed primers used to develop STS PEPC genetic markers. The number (*n*) and size of bands revealed in cv. R570 are given for each primer pair

Locus code	EMBL accessions	Forward primer[a]	Reverse primer	T	*n*	Fragment size (bp)
SO-ppc-C_4	AJ318338–AJ293346	CCCAGACGAGGTTCCACCTC	ATAAAGAGGCCAAAGCCGAGG	50	3	831, 836, 841
SO-ppc-aR	AJ318339 to AJ318341	GTGGCTGAGGATTTGCAATCC	AAGCTCCCCTTGGCATCTAC	55	6	Series A. 663, 666, 678 Series B. 704, 707, 719
SO-ppc-aL1	AJ318342–AJ318343	CAGAGGAACTGCAGCCATTG	GAGACGGCCGAAGATGCAGC	55	6	Series A. 515, 517, 560 Series B. 522, 524, 567
SO-ppc-aL2	M86661	GGCAAGCGTCTTAGAGCCAAC	CACCGTAACAAGATGCCCTGC	55	1	889

[a] Forward primers were labelled with the fluorochrome IRD800

sanguinalis Scop. (Paniceae, C_4) and *Phragmites australis* (Cav.) Trin. ex Steud. (Arundinoideae, C_3).

Results

In this work, we isolated and sequenced seven PEPC fragments in sugarcane. We analysed the sequence characteristics of each one and then we investigated and compared the transcript accumulation level of four PEPC gene classes in different organs. Lastly, we developped specific genetic markers in order to localise alleles of PEPCs in the cv. R570 genetic map.

Sugarcane C_4 PEPCs sequences

Firstly, using C_4 primers (pair no. 1–2), two partial cDNA sequences of C_4 PEPC analogues were obtained, each corresponding to one of the clones SES14 (EMBL accession AJ318338; 3,015 bp) and Big Tana Rayé (EMBL accession AJ293346; 3,067 bp). In sugarcane, the size of the corresponding complete mRNAs, as revealed by Northern, was estimated to be about 3,300 bp (data not shown). Both cDNA sequences exhibit an open reading frame of 2,883 bp and a 3′ untranslated part. Twenty polymorphisms (19 site mutations and one indel) were detected between the two obtained sugarcane sequences. The indel and one site mutation were detected in the 3′ untranslated region. Nucleotide homology with sorghum and maize C_4 PEPC cDNAs is high (96% and 89.7%, respectively). The proteins resulting from these isolated cDNAs contain 961 amino acids (108.5 kDa) and were named *ppc-C4-So1* and *ppc-C4-Ss1* for *S. officinarum* and *S. spontaneum*, respectively. The two proteins can be distinguished by eight amino acids (e.g. nos. 19, 33, 192, 197, 264, 390, 856 and 907). They display a high identity with C_4 PEPC from maize (90.3%), and sorghum (96.3%). The three conserved amino acid motifs of the PEPCs supposed to be involved in the carboxylation activity and the seven cysteine residues characteristic of all plant PEPCs are present (Lepiniec et al. 1994): VFTAHPT (no. 163–169), QQVMVGYSDSGKDAG (no. 586–600), FHGRGGTVGRGGGP (no. 629–642), plus cysteine residues nos. 55, 325, 410, 415, 417, 544 and 678. The motif SIDAQLR (no. 8–14), which is involved in the regulation of activity by light or darkness in C_4 and CAM PEPCs (Lepiniec et al. 1994), is also present. In addition, the serine in the carboxyl-terminal part of the enzyme, which is a major determinant of C_4-specific characteristics (Bläsing et al. 2000), was found in position 771. In *Flaveria*, the region between amino acids 296 and 437 was described as being essential for activation by glucose 6-phosphate (Bläsing et al. 2000). The homologous region in *Saccharum* C_4 PEPC is situated between amino acids 291 and 434. By carrying out a comparison with other available sequences of C_4 PEPCs in grass (Besnard et al. 2002b), *Flaveria* (Poetsch et al. 1991) or *Amaranthus* (Rydzik and Berry 1996) species, we looked

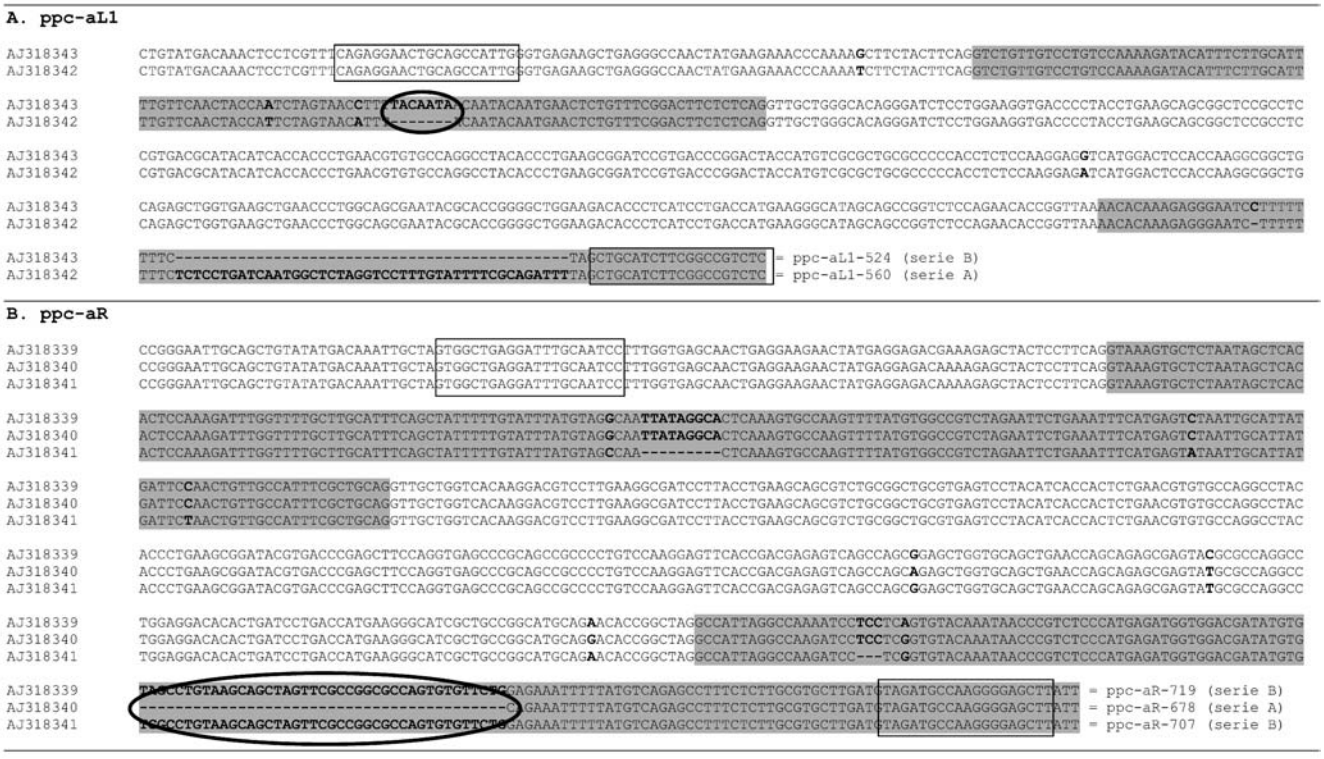

Fig. 1 Alignment of sugarcane nucleotide sequences homologous to sorghum housekeeping (ppc-aL1) and root (ppc-aR) PEPC gene segments. Polymorphisms are indicated in bold characters. Non-coding regions are shaded in grey. Boxes correspond to the location of primers used to generate STS markers. Circled indels correspond to polymorphisms supposed to enable the distinction between co-segregating PEPC fragments

for amino acids discriminating the C_4 PEPC isoforms from the other PEPC isoforms. This was unsuccessful; we did not find any amino acid leading to a strict discrimination (data not shown), suggesting that the activation by glucose 6-phosphate is dependant on several amino acid residues.

Non-photosynthetic PEPC analogues in sugarcane

Using primer pair 3–4, we cloned and sequenced three fragments homologous to a gene part of sorghum and maize root PEPCs: EMBL accession AJ318339, 755 bp; EMBL accession AJ318340, 714 bp; EMBL accession AJ318341, 743 bp. Similarly, using primer pair 5–6, we cloned and sequenced two fragments homologous to a gene part of sorghum and maize housekeeping PEPCs: EMBL accession AJ318342, 573 bp; EMBL accession AJ318343, 537 bp. These five fragments display the terminal part of the translated gene sequence plus one intron and the beginning of the untranslated 3′ part. These two untranslated parts display indel polymorphism, thereby explaining the difference in size of the homologous fragments (Fig. 1). Conversely, we amplified the 3′ untranslated part of the sugarcane housekeeping PEPC gene using primers 7–8 (M86661; Albert et al. 1992) in R570, *S. officinarum* (Big Tana Rayé) and *S. spontaneum* (SES14). Direct sequencing of this fragment (EMBL accession AJ279035) did not reveal any polymorphism between these individuals.

Protein sequences of 133 and 129 amino acids (C-terminal fragment) were deduced from coding regions in accessions AJ318339 to AJ318343. These protein fragments display a very high homology with PEPCs from sorghum and maize: (1) 97–100% identity between the protein fragments deduced from the accessions AJ318339-AJ318340-AJ318341 and root PEPCs (AB012228-X55664); (2) 97–99% identity between the protein fragments deduced from accessions AJ318342-AJ318343 and housekeeping PEPCs (X59925-X61489).

Phylogenetic relationships between PEPC genes in grasses

In the phylogenetic trees, we recognised four classes of PEPC genes in grasses (Fig. 2). As expected, sugarcane C_4 PEPC cDNAs appear to be tightly related to all grass cDNAs encoding C_4 PEPCs (ppc-C_4, Fig. 2). In the same way, we recognised one clade including cDNAs encoding root PEPCs (ppc-aR, Fig. 2). Furthermore, we showed that grass housekeeping PEPC cDNAs are classified into two classes (ppc-aL1 and ppc-aL2, Fig. 2B); for example, a sugarcane housekeeping sequence (ppc-aL2, M86661) is separated from all other grass PEPC sequences. Thus, we obtained evidence that sugarcane displays at least four

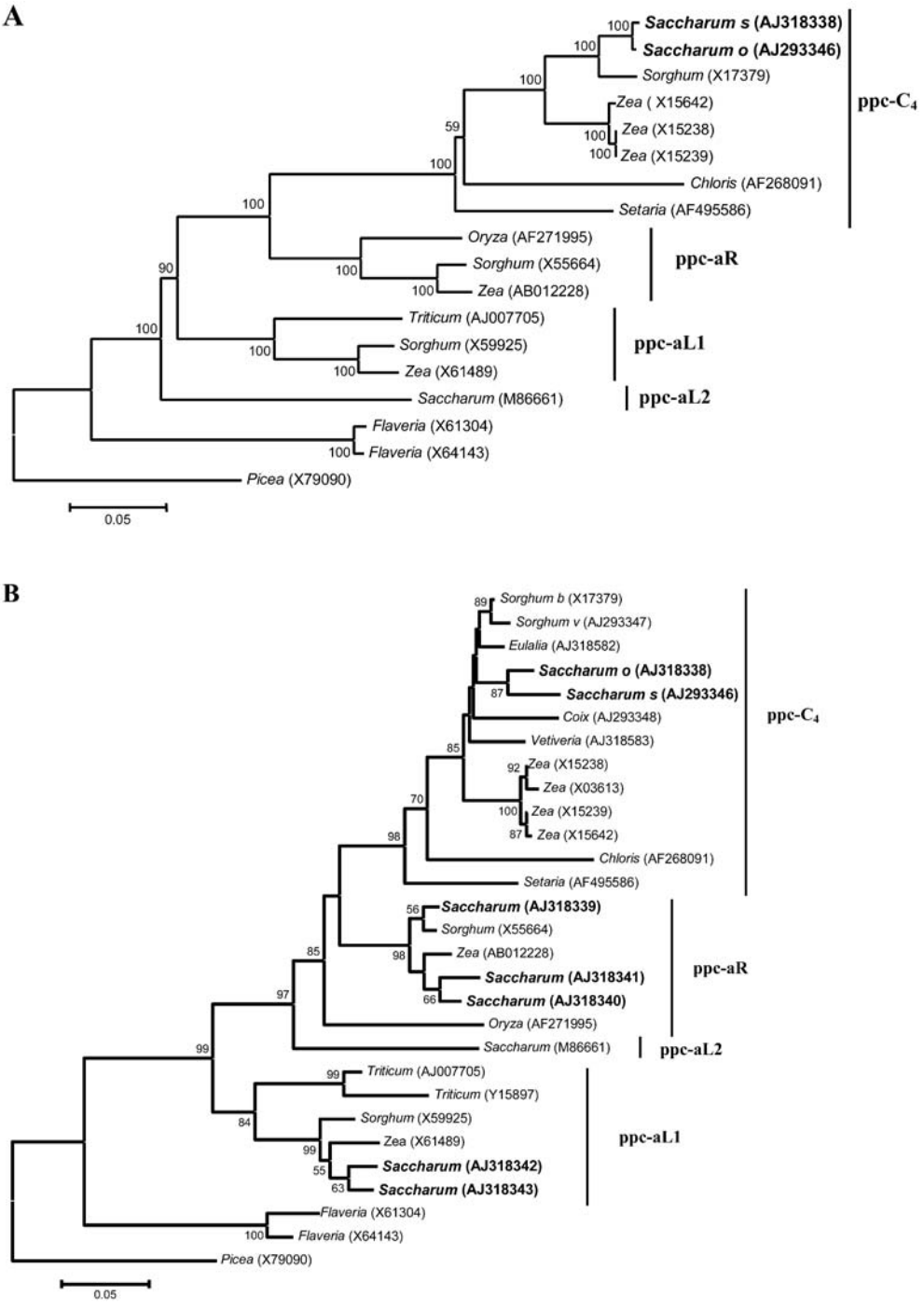

Fig. 2A, B Evolutionary relationships between the grass PEPC sequences toward phylogenetic trees based on Jukes and Cantor distances (1969) and constructed using the Neighbor joining algorithm. **A** Tree based on complete PEPC cDNA sequences (2,940 nucleotides: 1,350 informative sites/1,585 variable sites). **B** Tree based on the 3′ coding cDNA segment (330 nucleotides: 152 informative sites/189 variable sites). PEPC cDNA sequence from *Picea abies* was used to root the two phylogenetic trees. The sequences isolated in this study are in bold characters. The bootstrap values are indicated on each corresponding node when greater than to 50%. Four gene lineages encoding grass PEPCs are indicated. *Saccharum S = S. spontaneum*; *Saccharum o = S. officinarum*

PEPC gene lineages. Lastly, it is noticeable that C_4 and root PEPC clusters are sister clades. This was supported by a strong bootstrap value (100%, Fig. 2A). A very similar tree topology was obtained using Maximum-parsimony analysis (data not shown).

Organ expression specificity

The approach based on Northern dot-blot revealed that the ppc-C_4 gene was mainly transcribed in green leaves (Fig. 3, lane E). Moreover, transcript accumulation of this gene was low in the etiolated leaf. We also confirmed that ppc-aL2 housekeeping gene transcript accumulation was weak but constitutive and quite homogeneous between the studied organs (Albert et al. 1992). Indeed, its transcript accumulation (Fig. 3, lane B) was proportionally similar

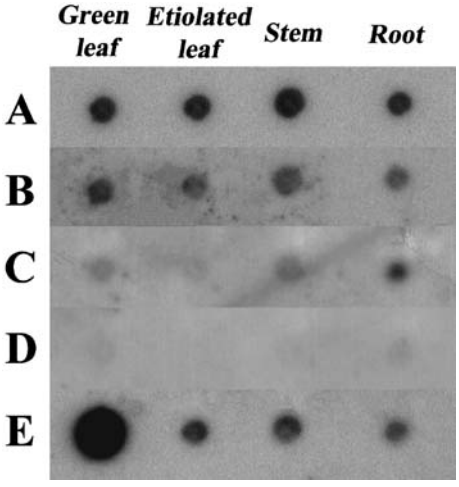

Fig. 3A–E Comparative gene transcript accumulation of four PEPC gene lineages according to different organs in sugarcane cv. R570 revealed using Northern dot-blot. Nylon membrane with RNA from four distinct organs (green leaf, etiolated leaf, stem and root) was successively hybridised with five probes: **A** 18S rDNA probe (30-min exposure), **B** specific probe of the sugarcane ppc-aL2 housekeeping gene (10-day of exposure), **C** specific probe of the ppc-aR gene (7-day exposure), **D** specific probe of the ppc-aL1 gene (10-day exposure), **E** specific probe of the ppc-C_4 gene (5-day exposure)

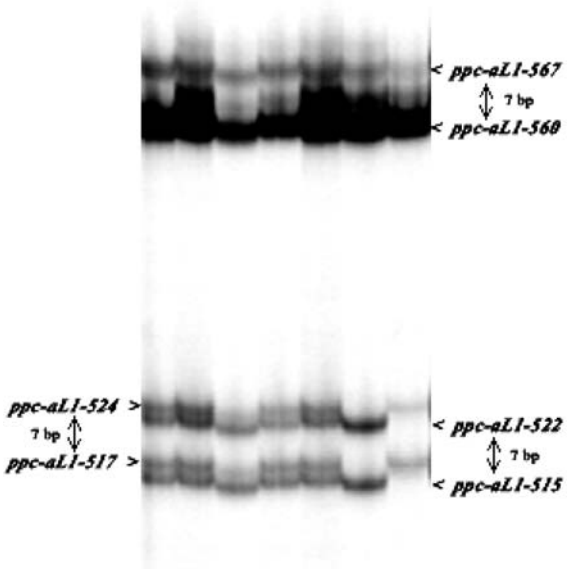

Fig. 4 Segregation of ppc-aL1 fragments in cv. R570 progenies. Size of each band is indicated. Fragments were separated on a 6% acrylamide gel

to that of the 18S rDNA gene (Fig. 3, lane A). Transcipts of a third sugarcane PEPC gene lineage (related to accessions AJ318339, AJ318340 and AJ318341; ppc-aR), which is homologous to sorghum and maize roots PEPCs, accumulated to a higher level in the root than in the other organs studied (Fig. 3, lane C). Lastly, transcript accumulation of the fourth sugarcane gene lineage (related to the accessions AJ318342 and AJ318343; ppc-aL1) was not detected under our experimental conditions (Fig. 3, lane D).

Genomic localisation of each PEPC gene

Using PCR, we looked for length variations in non-coding regions in order to analyse allele segregation of different PEPC gene lineages. In cv. R570, the number of fragments revealed is reported in the Table 3. We detected one to six fragments depending on the primer pair used. No length variation was found for the ppc-aL2 gene. Conversely, we found three fragments for the C_4 PEPC gene and six fragments for the ppc-aL1 and ppc-aR gene lineages. For the latter two gene lineages, two series of three fragments were revealed (named A and B; Table 3). Thus, for ppc-aL1, we can define pairs of fragments which differ by 7 bp (Fig. 4). In the same way, for ppc-aR, we can define pairs of fragments which differ by 41 bp. Interestingly, an indel of 7 bp was revealed in the intron between accessions AJ318342 and AJ318343 (ppc-aL1), and an indel of 41 bp was revealed in the 3′ untranslated part between accessions AJ318339/AJ318341 and AJ318340 (ppc-aR) (Fig. 1). We can assume that these indels explain the length variation between the pairs of fragments revealed. In the cv. R570 progenies, we noticed the segregation of seven simplex fragments (allele 841 for the C_4 PEPC, fragments 663 and 704 for the root PEPC and fragments 515, 517, 522 and 524 for ppc-aL1). Three pairs of fragments co-segregated: ppc-aR-663 (A) and -704 (B), ppc-aL1-515 (A) and -522 (B) and ppc-aL1-517 (A) and -524 (B). Fragments of two PEPC gene lineages were located on the cv. R570 genetic map (Hoarau et al. 2001): ppc-aL1-515/522 was positioned on co-segregation group 28 (X1), ppc-aL1-517/524 on co-segregation group 26 (X2) and ppc-aR-663/704 on co-segregation group 53. The ppc-C_4 841 marker was unlinked.

Usefulness of STS primers in other grasses

Checking STS primer pairs on *Sorghum, Digitaria* and *Phragmites*, PCR amplification occurred for locus ppc-aR (data not shown). Conversely, the ppc-aL2 locus was not amplified in these genera. Lastly, ppc-C_4 and ppc-aL1 loci were revealed in *Sorghum* but not in *Digitaria* and *Phragmites*.

Discussion

Relationships between PEPC isoforms in sugarcane

Phylogenetic trees of grass PEPC cDNAs (Fig. 2) support the existence of at least four PEPC gene lineages in sugarcane: one encodes C_4 PEPC (ppc-C_4), another root PEPC (ppc-aR) and, lastly, two correspond to housekeep-

ing PEPCs (ppc-aL1 and ppc-aL2). The existence of four PEPC gene classes in grasses has only been shown in recent studies (Gehrig et al. 2001; Bläsing et al. 2002), and we report here, for the first time, their presence in the same species. Furthermore, root PEPC clade appears to be a sister to C_4 PEPC clade. The high bootstrap value for such an observation [also revealed by Bläsing et al. (2002) who analysed complete amino acid sequences of PEPC from different plant families] supports the hypothesis that the grass C_4 PEPC gene could have derived from a root pre-existing PEPC gene. This suggests that gene duplication of a root PEPC gene followed by changes in gene regulatory controls have led to the appearance of the grass C_4 PEPC gene (Sheen 1999). This event should have occurred in a C_4 grass ancestor (maybe in a pre-grass monocot) since grass C_4 PEPCs (representative of different C_4 grass taxa) are gathered in only one clade. An ancient genome duplication (Levy and Feldman 2002) could explain such a gene duplication and also the independent evolution of the ppc-C_4 and ppc-aR gene lineages.

Organ expression of each PEPC gene lineage

As expected, based on sequence homologies, C_4 and root PEPC transcripts are more highly accumulated in green leaves and roots, respectively. In addition, a strong light activation of the C_4 PEPC transcript accumulation was evidenced. Consequently, we can conclude that sugarcane C_4 and root PEPCs are similar to those described in sorghum and maize (Izui et al. 1986; Crétin et al. 1991; Dong et al. 1998). Nevertheless, for the housekeeping isoforms, we propose another hypothesis. Indeed, the presence of sugarcane PEPCs analogous to housekeeping PEPCs from sorghum and maize (ppc-aL1; Lepiniec et al. 1991; Kawamura et al. 1992) was observed but not their transcript accumulation. Nevertheless, transcript accumulation may be very low or it occurs in another organ not studied here. We can also assume that the ppc-aL2 housekeeping isoform complements the ppc-aL1 isoform in sugarcane. Indeed, PEPC genes homologous to the ppc-aL2 have not yet been detected in sorghum and maize.

STS genetic markers and genomic organisation of PEPC genes

STS markers for the sugarcane PEPC genes were developed as already proposed for a *Picea* PEPC gene (Ipsen and Ziegenhagen 2001). Using our primers, we showed that STS PEPC markers can also be successfully used in related grass genera. In sugarcane, non-coding regions display length variations, with the exception of the ppc-aL2 gene. Using these length variations, genetic diversity of each PEPC gene could be assessed in sugarcane. These polymorphisms also enable us to locate PEPC genes on genetic maps. Thus, PEPC gene family organisation in sugarcane genome can be revealed. In our study, the number of segregating alleles was insufficient to determine if PEPC genes (belonging to distinct lineages) are located on homologous chromosomes. Nevertheless, pairs of co-segregating fragments were revealed for the ppc-aL1 and ppc-aR gene lineages. This result strongly suggests that duplicated genetic elements exist for these two gene lineages. Furthermore, these duplicated elements seem to have undergone a complex evolution. Gene conversion may explain why co-segregating fragments display specific polymorphism enabling them to be distinguished from other co-segregating fragments. Thus, an indel of 2 bp enables the ppc-aL1-515/ppc-aL1-522 markers to be distinguished from the ppc-aL1-517/ppc-aL1-524 markers (Fig. 4). Nevertheless, indels of 7 bp or 41 bp, characteristic of the ppc-aL1 and ppc-aR gene lineages, respectively, enable co-segregating repeats to be identified. Consequently, these latter mutations should not have undergone a concerted evolution. An example of incomplete homogenisation by gene conversion has already been reported for duplicated *amylase* genes in *Drosophila* (Inomata and Yamazaki 2002). Selective forces could explain the maintenance of differences between each repeat of a duplicate (Inomata and Yamazaki 2002). To check this hypothesis, comparison of a greater number of sequences is required in order to detect linkage disequilibrium between polymorphic sites located either in introns or exons. PEPC genes duplicated in tandem could be assigned to distinct functions; in *Flaveria* genus, three very related PEPC enzymes have been described with very distinct kinetic properties (Bläsing et al. 2002).

In conclusion, our results suggest that both polyploidisation events and gene duplications in tandem should have been involved in the diversification of the PEPC gene family. For a better understanding of the mode of evolution of this gene family, it should be of great interest to compare complete PEPC gene sequences and their genomic organisation in different grass species belonging to different subfamilies (i.e. *Oryza, Triticum, Sorghum, Saccharum*). This could be achieved by comparying BAC clones displaying PEPC genes (or pseudo-genes; e.g. clone BAC OSJNBa0016I09). The evolution of regulatory elements between each gene should be also studied in order to gain a better understanding of the mechanisms leading to new PEPC isoform appearances.

Acknowledgements We thank Dr. A. Bervillé, Dr. M. Morel, Ph. Rondeau, Dr. S. Santoni, and H. Tellismart for their helpful comments and/or technical assistance. We also particularly thank Dr. C. Rouch and Dr. C. Robert for their encouragement.

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, Khadari B, Villemur P, Bervillé A (2000) Cytoplasmic male sterility in the olive (*Olea europaea* L.). Theor Appl Genet 100:1018–1024
Besnard G, Offmann B, Robert C, Baret P, Rouch C, Cadet F (2002a) Phosphoenolpyruvate carboxylase cDNA phylogeny to investigate the C_4 photosynthetic pathway evolution in grasses. In: Proc 12th Int Congr Photosynthesis. Brisbane, Australia, S 17-007. http://www.publish.csiro.au/PS2001
Besnard G, Offmann B, Robert C, Rouch C, Cadet F (2002b) Assessment of the C_4 phosphoenolpyruvate carboxylase gene diversity in grasses. Theor Appl Genet 105:404–412
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. J Biol Chem 275:27,917–27,923
Bläsing OE, Ernst K, Streubel M, Westhoff P, Svensson P (2002) The non-photosynthetic PEP carboxylases of the C_4 dicot *Flaveria trinervia* – implications for the evolution of C_4 photosynthesis. Planta 215:448–456
Clayton WD, Renvoize SA (1986) Genera Graminum: grasses of the world. Kew Bull 13:1–389
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
D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glazmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol Gen Genet 250:405–413
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
Gehrig H, Heute V, Kluge M (2001) New partial sequences of phosphoenolpyruvate carboxylase as molecular phylogenetic markers. Mol Phylogenet Evol 20:262–274
Gonzalez MC, Osuna L, Echevarria C, Vidal J, Cejudo FJ (1998) Expression and localization of phosphoenolpyruvate carboxylase in developing and germinating wheat grains. Plant Physiol 116:1249–1258
Gonzales MC, Echevarria C, Vidal J, Cejudo FJ (2002) Isolation and characterization of a wheat phosphoenolpyruvate carboxylase gene. Modelling of the encoded protein. Plant Sci 162:233–238
GPWG (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). Ann Mo Bot Gard 88:373–457
Hatch MD (1999) C_4 photosynthesis: a historical overview. In: Sage RF, Monson RK (eds) C_4 plant biology. Academic Press, New York, pp 17–46
Hermans J, Westhoff P (1992) Homologous genes for the C_4 isoform of phosphoenolpyruvate carboxylase in C_3 and C_4*Flaveria* species. Mol Gen Genet 243:275–284
Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. Theor Appl Genet 103:84–97
Hudspeth RL, Grula JW (1989) Structure and expression of the maize gene encoding the phosphoenolpyruvte carboxylase isozyme involved in C_4 photosynthesis. Plant Mol Biol 12:579–589
Ipsen A, Ziegenhagen B (2001) New insights into allelic diversity of a phosphoenolpyruvate carboxylase in the conifer *Picea abies* (L.) Karst. Planta 214:265–273
Inomata N, Yamazaki T (2002) Nucleotide variation of the duplicated *amylase* genes in *Drosophila kikkawai*. Mol Biol Evol 19:678–688
Irvine JE (1999) *Saccharum* species as horticultural classes. Theor Appl Genet 98:186–194
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
Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, Academic Press, New York, pp 21–132
Kai Y, Matsumura H, Inoue T, Terada K, Nagara Y, Yoshinaga T, Kihara A, Tsumura K, Izui K (1999) Three-dimensional structure of phosphoenolpyruvate carboxylase: a proposed mechanism for allosteric inhibition. Proc Natl Acad Sci USA 96:823–828
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
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. Bioinformatics 17:1244–1245
Lander E, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage map of experimental and natural populations. Genomics 1:174–181
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
Levy AA, Feldman M (2002) The impact of polyploidy on grass genome evolution. Plant Physiol 130:1587–1593
Matsuoka M, Minami E (1989) Complete structure of the gene for phosphoenolpyruvate carboxylase from maize. Eur J Biochem 181:593–598
Matsuoka M, Furbank RT, Fukayama H, Miyao M (2001) Molecular engineering of C_4 photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 52:297–314
Poetsch W, Hermans J, Westhoff P (1991) Multiple cDNAs of phosphoenolpyruvate carboxylase in the C_4 dicot *Flaveria trinervia*. FEBS Lett 292:133–136
Rao SK, Magnin NC, Reiskind JB, Bowes G (2002) Photosynthetic and other phosphoenolpyruvate carboxylase isoforms in the single-cell, facultative C_4 system of *Hydrilla verticillata*. Plant Physiol 130:876–886
Relle M, Wild A (1996) Molecular characterization of a phosphoenolpyruvate carboxylase in the gymnosperm *Picea abies* (Norway spruce). Plant Mol Biol 32:923–936
Rydzik E, Berry J (1996) The C_4 photosynthetic phosphoenolpyruvate carboxylase from grain amaranth. Plant Physiol 110:713–713
Saitou N, Nei M (1987) The Neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Sheen J (1999) C_4 gene expression. Annu Rev Plant Physiol Plant Mol Biol 50:187–217
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
Vidal J, Chollet R (1997) Regulatory phosphorylation of C_4 PEP carboxylase. Trends Plant Sci 6:230–237