

***DCT4* - a new member of the dicarboxylate transporter family in C₄ grasses**

Sarit Weissmann¹, Pu Huang¹⁺, Koki Furuyama², Madeline A. Wiechert¹, Mitsutaka Taniguchi², James C. Schnable^{3*}, Thomas P. Brutnell⁴, and Todd C. Mockler^{1*}

¹ Donald Danforth Plant Science Center, St. Louis, MO, 63132, USA.

² Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi, Japan.

³ Computational Sciences Initiative, Center for Plant Science Innovation, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA.

⁴ Chinese Academy of Agricultural Sciences, Biotechnology Research Institute, Beijing, China.

⁺ Current Address: BASF Corporation, 26 Davis Dr., Durham, NC 27709, USA.

* Corresponding author

ABSTRACT

Malate transport shuttles atmospheric carbon into the Calvin-Benson cycle during NADP-ME C₄ photosynthesis. Previous characterizations of several plant dicarboxylate transporters (DCT) showed that they efficiently exchange malate across membranes. Here we identify and characterize a previously unknown member of the *DCT* family, *DCT4*, in *Sorghum bicolor*. We show that SbDCT4 exchanges malate across membranes and its expression pattern is consistent with a role in malate transport during C₄ photosynthesis. *SbDCT4* is not syntenic to the characterized photosynthetic gene *ZmDCT2*, and an ortholog is not detectable in the maize reference genome. We found that the expression patterns of *DCT* family genes in the leaves of *Z. mays*, and *S. bicolor* varied by cell type. Our results suggest that sub-functionalization of members of the *DCT* family for the transport of malate into the bundle sheath (BS) plastids occurred during the process of independent recurrent evolution of C₄ photosynthesis in grasses of the PACMAD clade. This study confirms the value of using both syntenic information and gene expression profiles to assign orthology in evolutionarily related genomes.

INTRODUCTION

Three subtypes of C_4 photosynthesis are generally recognized as defined by the primary decarboxylase in the bundle sheath (BS) cells: chloroplastic NADP-dependent malic enzyme (NADP-ME); mitochondrial NAD-dependent malic enzyme (NAD-ME); and cytosolic phosphoenolpyruvate carboxykinase (PEPCK) (Hatch and Slack, 1966; Hatch, 1971; Rathnam and Edwards, 1977). Different plant species may contain various combinations of these three subtypes (Hatch, 1971; Chapman and Hatch, 1979; Furbank, 2011; Pick *et al.*, 2011; Wang *et al.*, 2014b). The movement and exchange of malate across membranes, by dicarboxylate transporters (DCTs/DiTs), plays a significant role during photosynthesis in NADP-ME and NAD-ME C_4 species (Ding *et al.*, 2015). In C_3 plants, DCTs are crucial to nitrate assimilation, such as the GS/GOGAT cycle and photorespiration (Linka and Weber, 2010; Kinoshita *et al.*, 2011). Taniguchi *et al.* characterized several plant DCTs that efficiently exchange malate across membranes (Taniguchi *et al.*, 2002; Taniguchi *et al.*, 2004). The differential expression of C_4 photosynthesis genes in mesophyll (M) and BS cells (John *et al.*, 2014; Tausta *et al.*, 2014; Wang *et al.*, 2014a) suggests that different malate transporters may be needed to move malate out of the chloroplasts of M cells and into the chloroplasts of BS cells. In *Zea mays*, an NADP-ME C_4 grass, dicarboxylate transporter-2 (*ZmDCT2*, GRMZM2G086258) moves malate into the chloroplast of BS cells during C_4 photosynthesis (Ding *et al.*, 2015). *ZmDCT2* plays a critical role during C_4 photosynthesis in *Z. mays*, and its absence severely impairs plant growth and development (Ding *et al.*, 2015). The role of DCTs in C_4 photosynthesis in other species, however, remains unknown.

Z. mays is the best characterized and functionally annotated C_4 grass species. As such, it is a useful reference for identification of photosynthesis-related genes in poorly characterized C_4 grasses and for resolving orthology (John *et al.*, 2014; Ding *et al.*, 2015; Huang *et al.*, 2017). Microsynteny, the comparison of collinearity among related species, is a reliable approach to determine orthology and predict the function of a gene (Bennetzen and Freeling, 1997; Chen *et al.*, 1997; Tikhonov *et al.*, 1999; Bennetzen, 2000; Kumar *et al.*, 2009; Jin *et al.*, 2016). Davidson *et al.* (Davidson *et al.*, 2012) showed that syntenic orthologs are likely to have conserved functions and expression patterns across lineages. Here we identify a new member of the *DCT* family, *DCT4*, which is not syntenic to the photosynthetic gene *ZmDCT2* and is not detected in the maize reference genome. We demonstrate that *Sorghum bicolor* DCT4 (*SbDCT4*) efficiently exchanges malate across membranes, consistent with a malate transport role in C_4 photosynthesis. We characterize the diverse expression patterns of *DCT* genes in leaves of multiple

grass species. We also propose that sub-functionalization of *DCTs* in grasses of the PACMAD clade (Sanchez-Ken and Clark, 2010) occurred during independent recurrent evolution of C_4 photosynthesis.

RESULTS

Identification of *DCT4* in *Sorghum bicolor*

To learn more about C_4 -related dicarboxylate transporters in species evolutionarily related to maize, we identified the syntenic ortholog of *ZmDCT2* in *S. bicolor*. Two genes, Sobic.007G226700 and Sobic.007G226800, are present at the predicted syntenic orthologous position on chromosome 7. We refer to them as *SbDCT2.1* and *SbDCT2.2*, respectively (Figure 1). *ZmDCT2* is abundantly expressed (Table 1), and its expression is enriched in BS cells of maize leaves (Figure 2) (Li *et al.*, 2010; Tausta *et al.*, 2014; Ding *et al.*, 2015). In contrast, the expression profiles of *SbDCT2.1* and *SbDCT2.2* in *S. bicolor* leaves are low (Table 1). *SbDCT2.1* expression is slightly enriched in the M cells whereas *SbDCT2.2* is enriched in BS cells (Figure 2). We also analyzed the transcript levels of two other *S. bicolor* dicarboxylate transporters, *SbDCT1* (Sobic.002G233700) and *SbOMT1* (Sobic.008G112300). These genes are the orthologs of the *Z. mays* genes *ZmDCT1* (GRMZM2G040933) and *Zm-oxoglutarate/malate transporter 1* (*ZmOMT1*; GRMZM2G383088), respectively. We found that *SbDCT1* expression, similar to that of *ZmDCT1*, is relatively low (Table 1), and only slightly differentially expressed in M cells relative to BS cells (Figure 2). The expression of both *ZmOMT1* and *SbOMT1* is relatively high (Table 1), and both are slightly enriched in M cells (Figure 2).

In C_4 species, the expression of many photosynthetic genes is enriched in either BS and M cells (Li *et al.*, 2010; John *et al.*, 2014; Tausta *et al.*, 2014; Weissmann *et al.*, 2015; Rao *et al.*, 2016). In NADP-ME species, two transporters, one within the BS cells and another in M cells, move malate in and out of the chloroplast during C_4 photosynthesis (Brautigam *et al.*, 2008; Weissmann and Brutnell, 2012; John *et al.*, 2014; Tausta *et al.*, 2014; Wang *et al.*, 2014a). However, in sorghum leaves, we found only one highly expressed dicarboxylate transporter, *SbOMT1*, that showed slightly enriched expression in M cells. Therefore we screened the sorghum genome for additional homologs of known maize *DCTs*. We identified the gene Sobic.004G035500 that showed homology to *ZmDCT1* and *ZmDCT2* but was not syntenic to either gene (Figure 1). *DCT3* is the name of the second transcript of *ZmDCT2* (Taniguchi *et al.*, 2004), so we named this new gene *SbDCT4*. No syntenic ortholog of *SbDCT4* is present in the reference genomes of *Z. mays* or *Oryza sativa*. The absence of syntenic conservation between *S. bicolor* and *Z. mays* and the lack of direct orthologs in *Z. mays* or C_3 species prevented identification of *DCT4* in a

previous bioinformatic screen for C_4 photosynthesis genes (Huang *et al.*, 2017). The expression of *SbDCT4* is moderately abundant (Table 1) and strongly enriched in the BS cells of *S. bicolor* leaves (Figure 2).

SbDCT4 is an efficient malate transporter

To verify the ability of SbDCT4 to transport malate, we cloned coding sequences from the three sorghum *DCT* genes, *SbDCT1*, *SbDCT2*, and *SbDCT4*. We measured the malate transport activities of the recombinant proteins expressed in yeast. SbDCT4 was an efficient malate transporter (Table 2). The K_m of SbDCT4 was similar to that of SbDCT1, and the affinity for malate was highest in SbDCT2 among the three SbDCTs (Table 2), consistent with the relative malate transport activities reported for maize DCT1 and DCT2 (Taniguchi *et al.*, 2004).

Phylogenetic distribution of *DCT* genes in grasses

To understand the relationship of *SbDCT4* to other grass *DCT* genes, we searched the genomes of the grass species *Setaria italica*, *Urochloa fusca*, *Brachypodium distachyon*, and *Dichanthelium oligosanthes*. In *S. italica*, an NADP-ME C_4 species, we identified a dicarboxylate transporter, Seita.9G375100, that showed no syntenic orthologous relationship with dicarboxylate transporter genes in other available grass genomes. Phylogenetic analysis showed that this gene clustered with *SbDCT4* but not with *SbDCT1* and *SbDCT2* (Figure 3). We designated this gene *SiDCT4*. We did not detect orthologs, syntenic or otherwise, in *U. fusca*, a PEPCK C_4 species, or in the two C_3 species. To expand the search for *DCT4* in other grasses currently lacking genome assemblies, we examined leaf-derived transcript assemblies for *Aristida congesta*, *Eriachne aristidea*, *Chasmanthium laxum*, *Danthoniopsis dinteri*, *Antheophora pubensis*, *Echinochloa esculenta*, *Paspalum vaginatum*, and *Arundinella hirta* (Huang *et al.*, in preparation). We then used the predicted coding sequences of the *DCT* genes from available genomes and from the *de novo* leaf transcriptome assemblies to generate a phylogenetic tree of the *DCT* family. The resulting phylogeny shows that *DCT4* transcripts form a distinct subclade from the *DCT1* clade (Figure 3). The absence of *DCT4* transcript expression does not rule out the existence of the gene in the genome. We also used polymerase chain reaction (PCR) to survey for *DCT* genes in the genomes of grass species for which whole-genome assemblies were not available. We designed conserved primers (non-degenerate or minimally degenerate) to small regions unique to each of the three *DCT* genes using PrimaClade (Gadberry *et al.*, 2005). We detected *DCT1* and *DCT2* in the genomes of all species tested (Table 1,

Supplemental Figure 1). *DCT4*, however, was detected only in the genomes of NADP-ME C₄ species of the PACMAD clade, excluding *Z. mays* (Table 1, Supplemental Figure 2).

Expression of malate transporter genes in NADP-ME C₄ grasses

C₃ species and *U. fusca*, a PEPCK C₄ species, express both *DCT1* and *DCT2* at low levels in leaves (Table 1). C₄ NADP-ME species of the PACMAD clade generally express one *DCT* gene in leaves at a high level and also express one or two other *DCT* genes at low levels (Table 1). We did not find an apparent lineage-specific pattern for the expression of the predominant *DCT* gene in the NADP-ME species we analyzed. This finding is consistent with random evolutionary processes underlying the sub-functionalization of members of the *DCT* family. Interestingly, *Z. mays* is the only species we examined in which *DCT2* is the predominantly expressed *DCT* gene (Table 1). We also examined the expression of the non-*DCT* malate transporter *OMT1* gene in the leaves of grasses (Table 1). Interestingly, we found that while *OMT1* expression was generally abundant, there was no consistent pattern of relative expression between the *DCT* and *OMT* genes within the NADP-ME C₄ species (Table 1).

DISCUSSION

Evolution of the *DCT* gene family in grasses

We identified *DCT4* as a new member of the *DCT* gene family in the grasses (Figure 1). Our analysis suggests that *DCT4* is present in some C₄ NADP-ME PACMAD grasses. *DCT1* and *DCT2* appear to have originated from a duplication of a single *DCT* gene after the monocot-eudicot split (Taniguchi *et al.*, 2004) and *DCT4* arose from a duplication of *DCT1* at the root of the PACMAD grasses (Figure 3). The expression of *DCT* genes in the grasses that we analyzed exhibited no clear lineage-specific patterns (Table 1). Therefore, we propose that different members of the *DCT* family subfunctionalized for photosynthetic malate transport in the BS cells of C₄ grasses of the PACMAD clade.

This work also confirms the importance of including syntenic and expression data in assigning orthology across species, and of developing multiple models for C₄ photosynthesis in the grasses. *SiDCT4* was previously mis-annotated as the ortholog of *DCT2* (John *et al.*, 2014), likely because of the lower expression level of *SiDCT2* (Seita.9G375100) in leaf tissue. The use of different malate transporters, for example, *DCT4* in *S. bicolor* and *S. italica*, or *DCT2* in *Z. mays*, suggests that multiple evolutionary paths resulted in the development of an active C₄ NADP-ME photosynthetic cycle. It is interesting to note

common origins of C_4 photosynthesis are often defined based on the predominant decarboxylase utilized, thus maize and sorghum are considered to have evolved from a common C_4 ancestor. This analysis suggests that rather than being static, biochemical adaptations continued after the divergence of maize and sorghum lineages. Thus, optimizations of C_4 activities may be continuous as breeding pressures or climate change alters ecological niches of individual species.

Various C_4 subtype combinations have different transport requirements

The variation of expression levels among the different malate transporters within each NADP-ME species (Table 1) suggests different transport requirements during C_4 photosynthesis. This supposition is in agreement with the view that the three subtypes of C_4 photosynthesis are mixed rather than exclusive (Hatch, 1971; Chapman and Hatch, 1979; Furbank, 2011; Pick *et al.*, 2011; Wang *et al.*, 2014b). For example, *Z. mays* utilizes both the NADP-ME (75%) and PEPCK (25%) pathways to fix carbon (Chapman and Hatch, 1979; Wingler *et al.*, 1999; Weissmann *et al.*, 2015), and has similar expression levels of *DCT2* and *OMT1* and low expression of *DCT1*. *S. bicolor* moves carbon through both malate and aspartate, although no PEPCK activity was detected in its leaves (Chapman and Hatch, 1979). *S. bicolor* has similar expression levels for *DCT4* and *DCT1* and high expression of *OMT1* (Table 1). Other grass species may have dicarboxylate transporter expression ratios that correspond to their unique combination of C_4 subsystems. For example, *OMT1* is highly expressed in *U. fusca*, ~3-7 fold higher than *DCT2* or *DCT1*, respectively. *OMT1* transports dicarboxylates, excluding those containing an amino group (Taniguchi *et al.*, 2002; Taniguchi *et al.*, 2004). Thus, in PEPCK C_4 plants, *OMT1* may move oxaloacetate into the mesophyll chloroplast, and 2-oxoglutarate out, to support the high production of aspartate needed to maintain the photosynthetic cycle (Rathnam and Edwards, 1977). Interestingly, both *ZmOMT1* and *SbOMT1* are only slightly differentially expressed in the M cells (Figure 2). As the loss of *DCT2* in *Z. mays* prevents movement of malate into the BS chloroplast (Weissmann *et al.*, 2015), *OMT1* cannot be moving malate into the BS chloroplast alongside *DCT2*. But *OMT1* may also have a role in organic acid metabolism in both cell types, such as shuttling reducing equivalents in organelles other than the chloroplast (Pleite *et al.*, 2005).

CONCLUSIONS

Our results show that the newly identified member of the *DCT* gene family, *SbDCT4*, is an efficient malate transporter. Based on the expression patterns of malate transporters among the grasses, we suggest that different members of the *DCT* family may have evolved multiple roles in C_4 photosynthesis.

Further studies will be needed to verify the subcellular localization of these proteins and to define their specific metabolic functions. Characterizing the various combinations of C₄ photosynthetic subsystems in grasses will facilitate the exploitation of *DCT* genes, through breeding or engineering, to improve the performance of crop plants and increase yield.

MATERIALS AND METHODS

Identification of *DCT4* genes in *Sorghum* and other grasses

We used QUOTA-ALIGN (Tang *et al.*, 2011) to identify syntenic orthologous regions in grass species with sequenced genomes, following the protocol described in (Zhang *et al.*, 2017). To find homologous genes at non-syntenic locations, we used two complementary approaches. For species with sequenced genomes, we used LASTZ (Harris, 2007) to align the coding sequence of the primary transcript annotated in Phytozome (<https://phytozome.jgi.doe.gov>) to the genome assembly. For species without assembled genomes, we used LASTZ to align the coding sequence of the primary transcript from Phytozome to transcript assemblies generated by Trinity (Grabherr *et al.*, 2011).

Measurements of malate transport

We cloned each of the three *SbDCT* cDNAs between the promoter and terminator of yeast *GAL2* in the pTV3e vector (Nishizawa *et al.*, 1995). We transformed the plasmids into yeast LBY416 cells and selected transformants on tryptophan-deficient agar plates. We prepared a crude membrane fraction from the selected yeast transformants. We used a freeze-thaw technique to reconstitute liposomes for the measurement of the uptake of [¹⁴C]malate (Taniguchi *et al.*, 2002).

Phylogenetic analysis of *DCT* homologs

DCT coding sequences for *Z. mays*, *S. bicolor*, *S. italica*, *B. distachyon*, *O. sativa*, *D. oligosanthos*, and *U. fusca* were from Phytozome (<https://phytozome.jgi.doe.gov>). We used BLASTN (Altschul *et al.*, 1990) to search *de novo* assembled leaf transcriptomes (Huang, manuscript in preparation) from the C₄ grass species *A. congesta*, *E. aristidea*, *C. laxum*, *D. dinteri*, *A. pubensis*, *E. esculenta*, *P. vaginatum*, and *A. hirta* with the *DCT* sequences from maize, *Setaria*, and *Sorghum* as queries. We used ProGraphMSA to generate a codon-based sequence alignment (Szalkowski, 2012). We used MEGA6 (Tamura *et al.*, 2013), with default parameters and the branch support values based on 1,000 bootstraps, to generate the phylogenetic reconstruction with the maximum likelihood method and based on the nucleotides in the third position of codons (Simmons *et al.*, 2006).

Analysis of gene expression for decarboxylase transporters in grasses

For species with published leaf transcriptome profiles (Ouyang *et al.*, 2007; Li *et al.*, 2010; Zhang *et al.*, 2012; Schnable, 2014; Wang *et al.*, 2014a; Studer *et al.*, 2016), gene expression levels were calculated and normalized, for each species, as Transcripts Per Million (TPM). For the other species, the normalized TPM values were based on *de novo* transcriptome assemblies (Huang *et al.*, manuscript in preparation). The values in Table 1 only allow for intraspecies comparisons among the decarboxylase transporters.

Identification of *DCT4* in species without sequenced genomes

We aligned the coding sequences from each of the *DCT* genes from *Z. mays*, *S. bicolor*, *S. italica*, *U. fusca*, *B. distachyon*, *O. sativa*, *D. oligosanthos*, *A. congesta*, *E. aristidea*, *C. laxum*, *D. dinteri*, *A. pubensis*, *E. esculenta*, *P. vaginatum*, and *A. hirta* using PAL2NAL (Suyama *et al.*, 2006). The resulting multiple sequence alignment enabled the design of non-degenerate or minimally degenerate PCR primers (Table 3) using PrimaClade (Gadberry *et al.*, 2005).

Jacob D. Washburn and J. Chris Pires (University of Missouri, Columbia) kindly provided genomic DNA from *A. congesta*, *E. aristidea*, *D. dinteri*, *A. pubensis*, *E. esculenta*, and *A. hirta* (Washburn *et al.*, 2015). We used a CTAB-based method to extract genomic DNA from *C. laxum*, *P. vaginatum*, *Z. mays*, *S. bicolor*, *S. italica*, and *B. distachyon* (Weissmann *et al.*, 2015). *Z. mays* and *B. distachyon* were the negative controls for *DCT4* and the positive controls for *DCT1* and *DCT2*. *S. bicolor* and *S. italica* were the positive controls for *DCT1*, *DCT2*, and *DCT4*.

We conducted amplification of *DCT* genes by PCR using a 25- μ l reaction mix and an ABI 2720 Thermal cycler. The reaction mixture included 2.5 μ l of 10X Buffer, 2.5 μ l of 10 μ M solutions of forward and reverse primers, 2 μ l of 2.5 mM dNTP stock, 14 μ l of nuclease-free water, 0.5 μ l of Choice Taq enzyme, and 1 μ l of 100 ng/ μ l DNA. We performed PCR reactions as described in Table 3 with 5 μ l of loading dye added to each reaction. Aliquots of 13 μ l were loaded on 3% agarose gels (Invitrogen UltraPure Agarose 1000, 1X TAE buffer, Invitrogen SYBR Safe Gel Stain) and electrophoresed for 30 minutes at 100 volts. We based size estimates on 100bp and 50 bp DNA markers (GoldBio).

Table 1. Genomic presence or absence and whole leaf TPM values for dicarboxylate transporter genes in grass leaves¹.

Species	Presence in genomic DNA			RNA-seq TPM values			Photosynthesis	
	DCT1	DCT2	DCT4	DCT1	DCT4	DCT2	OMT1	
<i>Brachypodium distachyon</i>	+	+	-	33	N/A	119	18	C ₃
<i>Oryza sativa</i>	+	+	-	16	N/A	11	101	C ₃
<i>Aristida congesta</i>	+	+	+	393	0	18	871	C ₄ NADP-ME
<i>Eriachne aristidea</i>	+	+	+	1348	0	89	874	C ₄ NADP-ME
<i>Chasmanthium laxum</i>	+	+	-	8	N/A	11	18	C ₃
<i>Danthoniopsis dinteri</i>	+	+	+	0	1555	123	305	C ₄ NADP-ME
<i>Antheophora pubensis</i>	+	+	+	1611	0	0	514	C ₄ NADP-ME
<i>Echinochloa esculenta</i>	+	+	+	1588	0	0	683	C ₄ NADP-ME
<i>Urochloa fusca</i>	+	+	-	50	N/A	86	344	C ₄ PEPCK
<i>Setaria italica</i>	+	+	+	29	658	41	328	C ₄ NADP-ME
<i>Dichanthelium oligosanthes</i>	+	+	-	11	N/A	13	13	C ₃
<i>Paspalum vaginatum</i>	+	+	+	591	666	0	354	C ₄ NADP-ME
<i>Arundinella hirta</i>	+	+	+	0	481	23	398	C ₄ NADP-ME
<i>Sorghum bicolor</i>	+	+	+	126	229	0 ³	483	C ₄ NADP-ME
		+				9 ⁴		C ₄ NADP-ME ²
<i>Zea mays</i>	+	+	-	5	N/A	166	153	C ₄ NADP-ME + PEPCK

¹ Note that interspecies comparison is not possible, because expression levels were normalized within each species.

² Despite no PEPCK activity detected in *Sorghum*, carbon moves into the BS through aspartate (Chapman and Hatch, 1979).

³ *SbDCT2.1*

⁴ *SbDCT2.2*

Table 2. K_m of malate for recombinant DCT proteins¹ demonstrates the ability of SbDCT4 to transport malate efficiently.

	K_m (mM)		
	DCT1	DCT2	DCT4
<i>S. bicolor</i>	1.24 ± 0.14	0.71 ± 0.10	1.13 ± 0.10
<i>Z. mays</i> ²	1.1 ± 0.1	0.85 ± 0.44	N/A

¹ The values are the means of three independent experiments ± SE.

² Kinetic values from a previous report (Taniguchi *et al.*, 2004).

Table 3. Primers and PCR conditions for the amplification of grass *DCT* genes.

Primer Pair	Forward 5'-3' ¹	Reverse 5'-3' ¹	Cycling Conditions
DCT1	CACCAACGAGGTCATCTGG	AGTAGGTGGCGATDCGGTC	94°C 3 min, [94°C 45 sec, 58°C 30 sec, 72°C 1 min] x 30, 72°C 10 min, 4°C ∞.
DCT2	CVTGGATGTCRAATTGTGTTG	TGGCTTGCAAABADATAGTGAA	94°C 3 min, [94°C 45 sec, 58°C-52°C (-0.5°C/cycle) 30 sec, 72°C 1 min] x 14, [94°C 45 sec, 52°C 30 sec, 72°C 1 min] x 16, 72°C 10 min, 4°C ∞.
DCT4	CTTYGTCAAGTGGCTCGG	GACTTGATGATSGGCAGGA	94°C 3 min, [94°C 45 sec, 60°C 30 sec, 72°C 1 min] x 30, 72°C 10.0 min, 4°C ∞.

¹ B = C+G+T, D = A+G+T, R = A+G, S = C+G, V = A+C+G, Y = C+T.

FIGURE LEGENDS

Figure 1. CoGe (<https://genomeevolution.org/coge/>) genome viewer screenshots depicting the conservation and genomic contexts of *DCT* genes in *Z. mays* and *S. bicolor*. Colored lines between panels show conserved genes. *SbDCT4* shows high sequence conservation with other *DCT* genes, but is not a syntenic ortholog of *ZmDCT2*, as shown by the lack of conservation in neighboring genes.

Figure 2. Differential expression of malate transporters in *Z. mays*, and *S. bicolor* leaves between the BS and M cells. The genome of *Z. mays* has one copy of *DCT2* and does not contain *DCT4*, and *ZmDCT2* is highly enriched in BS cells. *Sorghum bicolor* has two copies of *DCT2* (*DCT2.1*, and *DCT2.2*) in the syntenic genomic location that are the result of gene duplication. *S. bicolor* also expresses *DCT4*, which is highly enriched in BS cells. Both species express *OMT1* and *DCT1*, which are only slightly enriched in the M cells. Red bars represent enrichment in the BS cells. Blue bars represent enrichment in the M cells. The white numbers inside the bars represent the significance (p-value) of the log₂(FoldChange).

Figure 3. A phylogenetic tree of the *DCT* family in the grasses showing that *DCT4* is a subclade of *DCT1*. The *DCT1*, *DCT2*, and *DCT4* gene lineages are black, blue, and red, respectively. The length of the branches represents the evolutionary distance between ancestor to descendent nodes. The numbers represent the confidence level of the specific branch.

Supplemental Figure 1. Gel image showing that *DCT1* and *DCT2* genes are present in all grass species tested.

Supplemental Figure 2. Gel image showing the presence or absence of *DCT4* genes from species lacking genome assemblies. Negative controls were *Z. mays* and *B. distachyon*, and positive controls were *S. bicolor* and *S. italica*.

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.
- Bennetzen, J.L.** (2000). Comparative sequence analysis of plant nuclear genomes:m microcolinearity and its many exceptions. *Plant Cell* **12**, 1021-1029.
- Bennetzen, J.L., and Freeling, M.** (1997). The unified grass genome: synergy in synteny. *Genome research* **7**, 301-306.
- Brautigam, A., Hoffmann-Benning, S., and Weber, A.P.** (2008). Comparative proteomics of chloroplast envelopes from C3 and C4 plants reveals specific adaptations of the plastid envelope to C4 photosynthesis and candidate proteins required for maintaining C4 metabolite fluxes. *Plant Physiol* **148**, 568-579.
- Chapman, K.S., and Hatch, M.D.** (1979). Aspartate stimulation of malate decarboxylation in Zea mays bundle sheath cells: possible role in regulation of C4 photosynthesis. *Biochemical and biophysical research communications* **86**, 1274-1280.
- Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.S., Zhang, H., Wing, R.A., and Bennetzen, J.L.** (1997). Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3431-3435.
- Davidson, R.M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S.H., Jiang, N., and Robin Buell, C.** (2012). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *The Plant journal : for cell and molecular biology* **71**, 492-502.
- Ding, Z., Weissmann, S., Wang, M., Du, B., Huang, L., Wang, L., Tu, X., Zhong, S., Myers, C., Brutnell, T.P., Sun, Q., and Li, P.** (2015). Identification of Photosynthesis-Associated C4 Candidate Genes through Comparative Leaf Gradient Transcriptome in Multiple Lineages of C3 and C4 Species. *PLoS one* **10**, e0140629.
- Furbank, R.T.** (2011). Evolution of the C(4) photosynthetic mechanism: are there really three C(4) acid decarboxylation types? *Journal of experimental botany* **62**, 3103-3108.
- Gadberry, M.D., Malcomber, S.T., Doust, A.N., and Kellogg, E.A.** (2005). Primafade--a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* **21**, 1263-1264.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A.** (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644-652.
- Harris, R.S.** (2007). Improved Pairwise Alignment of Genomic DNA. In *Computer Science and Engineering* (Philadelphia: pennsylvania state university).
- Hatch, M.D.** (1971). The C 4 -pathway of photosynthesis. Evidence for an intermediate pool of carbon dioxide and the identity of the donor C 4 -dicarboxylic acid. *The Biochemical journal* **125**, 425-432.
- Hatch, M.D., and Slack, C.R.** (1966). Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *The Biochemical journal* **101**, 103-111.
- Huang, P., Studer, A.J., Schnable, J.C., Kellogg, E.A., and Brutnell, T.P.** (2017). Cross species selection scans identify components of C4 photosynthesis in the grasses. *Journal of experimental botany* **68**, 127-135.
- Jin, J., Kong, J., Qiu, J., Zhu, H., Peng, Y., and Jiang, H.** (2016). High level of microsynteny and purifying selection affect the evolution of WRKY family in Gramineae. *Development genes and evolution* **226**, 15-25.

- John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S., and Hibberd, J.M.** (2014). Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. *Plant Physiology* **165**, 62-75.
- Kinoshita, H., Nagasaki, J., Yoshikawa, N., Yamamoto, A., Takito, S., Kawasaki, M., Sugiyama, T., Miyake, H., Weber, A.P.M., and Taniguchi, M.** (2011). The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism. *Plant Journal* **65**, 15-26.
- Kumar, S., Mohan, A., Balyan, H.S., and Gupta, P.K.** (2009). Orthology between genomes of Brachypodium, wheat and rice. *BMC research notes* **2**, 93.
- Li, P.H., Ponnala, L., Gandotra, N., Wang, L., Si, Y.Q., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers, C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., and Brutnell, T.P.** (2010). The developmental dynamics of the maize leaf transcriptome. *Nature Genetics* **42**, 1060-U1051.
- Linka, N., and Weber, A.P.M.** (2010). Intracellular Metabolite Transporters in Plants. *Molecular Plant* **3**, 21-53.
- Nishizawa, K., Shimoda, E., and Kasahara, M.** (1995). Substrate recognition domain of the Gal2 galactose transporter in yeast *Saccharomyces cerevisiae* as revealed by chimeric galactose-glucose transporters. *The Journal of biological chemistry* **270**, 2423-2426.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., and Buell, C.R.** (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic acids research* **35**, D883-887.
- Pick, T.R., Brautigam, A., Schluter, U., Denton, A.K., Colmsee, C., Scholz, U., Fahnenstich, H., Pieruschka, R., Rascher, U., Sonnewald, U., and Weber, A.P.M.** (2011). Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C-4 Model and Provides Candidates for Regulation. *Plant Cell* **23**, 4208-4220.
- Pleite, R., Pike, M.J., Garcés, R., Martínez-Force, E., and Rawsthorne, S.** (2005). The sources of carbon and reducing power for fatty acid synthesis in the heterotrophic plastids of developing sunflower (*Helianthus annuus* L.) embryos. *Journal of experimental botany* **56**, 1297-1303.
- Rao, X., Lu, N., Li, G., Nakashima, J., Tang, Y., and Dixon, R.A.** (2016). Comparative cell-specific transcriptomics reveals differentiation of C4 photosynthesis pathways in switchgrass and other C4 lineages. *Journal of experimental botany* **67**, 1649-1662.
- Rathnam, C.K.M., and Edwards, G.E.** (1977). C4 acid decarboxylation and CO₂ donation to photosynthesis in bundle sheath strands and chloroplasts from species representing 3 groups of C4 plants. *Arch. Biochem. Biophys.* **182**, 1-13.
- Sanchez-Ken, J.G., and Clark, L.G.** (2010). Phylogeny and a new tribal classification of the Panicoideae s.l. (Poaceae) based on plastid and nuclear sequence data and structural data. *American journal of botany* **97**, 1732-1748.
- Schnable, J.C.** (2014). *Urochloa fusca* LBJWC-52 Annotated Standard Draft (Joint Genome Institute).
- Simmons, M.P., Zhang, L.B., Webb, C.T., and Reeves, A.** (2006). How can third codon positions outperform first and second codon positions in phylogenetic inference? An empirical example from the seed plants. *Systematic biology* **55**, 245-258.
- Studer, A.J., Schnable, J.C., Weissmann, S., Kolbe, A.R., McKain, M.R., Shao, Y., Cousins, A.B., Kellogg, E.A., and Brutnell, T.P.** (2016). The draft genome of the C3 panicoid grass species *Dichanthelium oligosanthos*. *Genome Biol* **17**, 223.
- Suyama, M., Torrents, D., and Bork, P.** (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research* **34**, W609-612.
- Szalkowski, A.M.** (2012). Fast and robust multiple sequence alignment with phylogeny-aware gap placement. *BMC bioinformatics* **13**, 129.

- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013).** MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729.
- Tang, H., Lyons, E., Pedersen, B., Schnable, J.C., Paterson, A.H., and Freeling, M. (2011).** Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC bioinformatics* **12**, 102.
- Taniguchi, M., Taniguchi, Y., Kawasaki, M., Takeda, S., Kato, T., Sato, S., Tahata, S., Miyake, H., and Sugiyama, T. (2002).** Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in *Arabidopsis thaliana*. *Plant and Cell Physiology* **43**, 706-717.
- Taniguchi, Y., Nagasaki, J., Kawasaki, M., Miyake, H., Sugiyama, T., and Taniguchi, M. (2004).** Differentiation of dicarboxylate transporters in mesophyll and bundle sheath chloroplasts of maize. *Plant Cell Physiol* **45**, 187-200.
- Tausta, S.L., Li, P., Si, Y., Gandotra, N., Liu, P., Sun, Q., Brutnell, T.P., and Nelson, T. (2014).** Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. *Journal of experimental botany* **65**, 3543-3555.
- Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L., and Avramova, Z. (1999).** Colinearity and its exceptions in orthologous adh regions of maize and sorghum. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 7409-7414.
- Wang, L., Czedik-Eysenberg, A., Mertz, R.A., Si, Y., Tohge, T., Nunes-Nesi, A., Arrivault, S., Dedow, L.K., Bryant, D.W., Zhou, W., Xu, J., Weissmann, S., Studer, A., Li, P., Zhang, C., LaRue, T., Shao, Y., Ding, Z., Sun, Q., Patel, R.V., Turgeon, R., Zhu, X., Provart, N.J., Mockler, T.C., Fernie, A.R., Stitt, M., Liu, P., and Brutnell, T.P. (2014a).** Comparative analyses of C(4) and C(3) photosynthesis in developing leaves of maize and rice. *Nature biotechnology* **32**, 1158-1165.
- Wang, Y., Brautigam, A., Weber, A.P.M., and Zhu, X.G. (2014b).** Three distinct biochemical subtypes of C-4 photosynthesis? A modelling analysis. *Journal of experimental botany* **65**, 3567-3578.
- Washburn, J.D., Schnable, J.C., Davidse, G., and Pires, J.C. (2015).** Phylogeny and photosynthesis of the grass tribe Paniceae. *American journal of botany* **102**, 1493-1505.
- Weissmann, S., and Brutnell, T.P. (2012).** Engineering C4 photosynthetic regulatory networks. *Curr Opin Biotechnol* **23**, 298-304.
- Weissmann, S., Ma, F., Furuyama, K., Gierse, J., Berg, H., Shao, Y., Taniguchi, M., Allen, D.K., and Brutnell, T.P. (2015).** Interactions of C4 Subtype Metabolic Activities and Transport in Maize Are Revealed through the Characterization of DCT2 Mutants. *Plant Cell* **28**, 466-484.
- Wingler, A., Walker, R.P., Chen, Z.H., and Leegood, R.C. (1999).** Phosphoenolpyruvate carboxykinase is involved in the decarboxylation of aspartate in the bundle sheath of maize. *Plant Physiol* **120**, 539-546.
- Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z., Wang, W., Tao, Y., Bian, C., Han, C., Xia, Q., Peng, X., Cao, R., Yang, X., Zhan, D., Hu, J., Zhang, Y., Li, H., Li, H., Li, N., Wang, J., Wang, C., Wang, R., Guo, T., Cai, Y., Liu, C., Xiang, H., Shi, Q., Huang, P., Chen, Q., Li, Y., Wang, J., Zhao, Z., and Wang, J. (2012).** Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature biotechnology* **30**, 549-554.
- Zhang, Y., Ngu, D.W., Carvalho, D., Liang, Z., Qiu, Y., Roston, R.L., and Schnable, J.C. (2017).** Differentially Regulated Orthologs in Sorghum and the Subgenomes of Maize. *Plant Cell* **29**, 1938-1951.

Figure 1

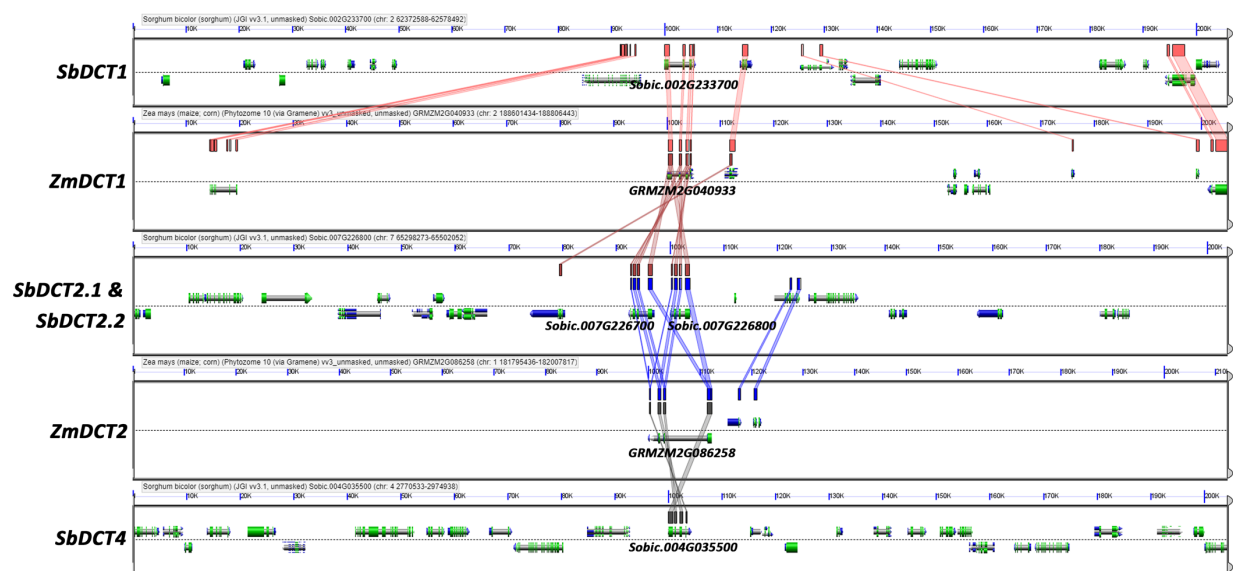


Figure 2

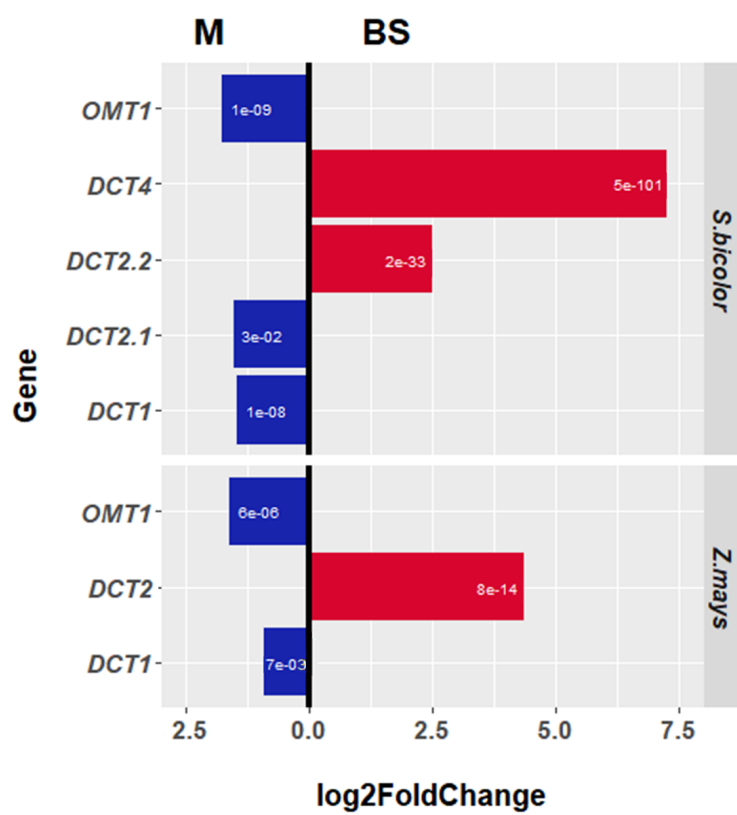
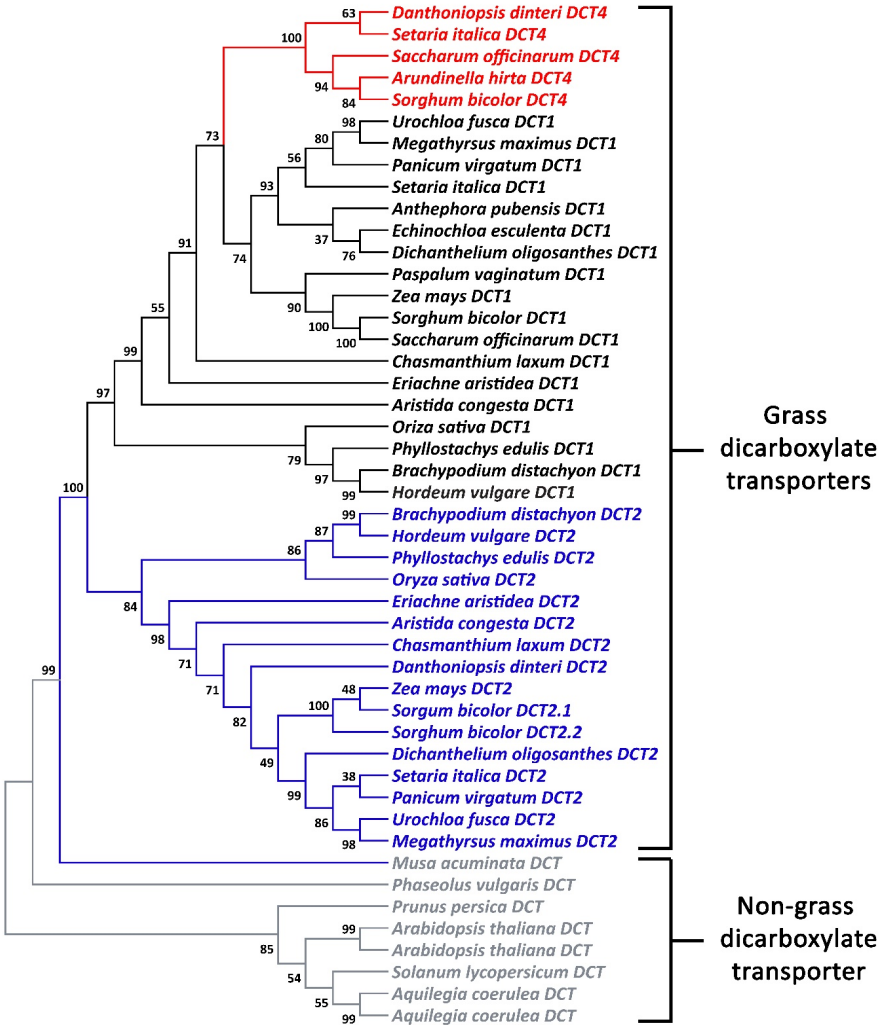
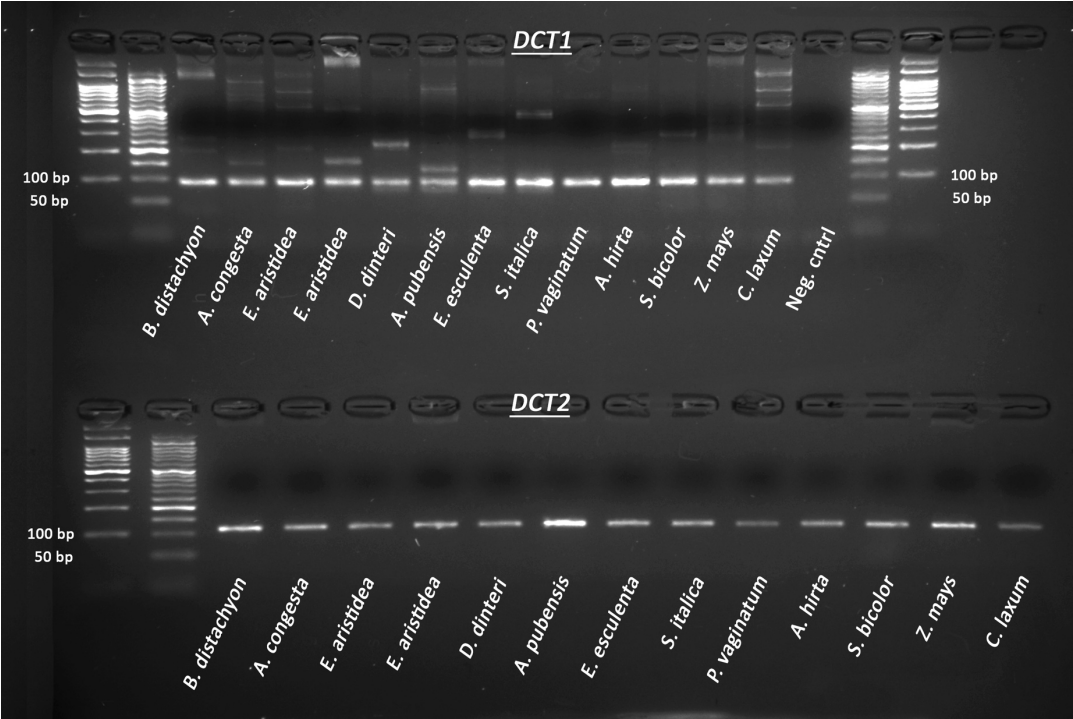


Figure 3



Supplemental Figure 1



Supplemental Figure 2

