An mRNA blueprint for C₄ photosynthesis derived from comparative transcriptomics of closely related C₃ and C₄ species

Andrea Bräutigamᵃ,¹, Kaisa Kajalaᵇ,¹, Julia Wullenweberᵃ, Manuel Sommerᵃ, David Gagneulᵃ, Katrin L. Weberᵃ, Kevin M. Carrᶜ, Udo Gowikᵈ, Janina Maßᵉ, Martin J. Lercherᵉ, Peter Westhoffᵈ, Julian M. Hibberdᵇ,², Andreas P.M. Weberᵃ,²

ᵃInstitute of Plant Biochemistry, Universitätsstrasse 1, Heinrich-Heine-University, 40225 Düsseldorf, Germany
ᵇDepartment of Plant Sciences, Downing Street, University of Cambridge, Cambridge, CB2 3EA, UK
ᶜBioinformatics Core, Research Technology Support Facility, Michigan State University, East Lansing, MI 48824, USA
ᵈInstitute of Plant Molecular and Developmental Biology, Universitätsstrasse 1, Heinrich-Heine-University, 40225 Düsseldorf, Germany
ᵉInstitut für Informatik, Universitätsstrasse 1, Heinrich-Heine-University, 40225 Düsseldorf, Germany

¹These authors have equally contributed to this manuscript
²To whom the correspondence should be addressed

Running title: An mRNA blueprint for C₄ photosynthesis

Page estimate: 13
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors of Plant Physiology) are: Julian M Hibberd (julian.hibberd@plantsci.cam.ac.uk) and Andreas PM Weber (andreas.weber@uni-duesseldorf.de).

Data deposition footnote: The *Cleome* read data have been submitted to the NCBI Short Read Archive. *Cleome spinosa* = SRS002743.1 and *Cleome gynandra* = SRS002744.2

**Keywords:** C₄ photosynthesis, *Cleome, Arabidopsis*, RNA-seq.
Abstract

C4 photosynthesis involves alterations to the biochemistry, cell biology, and development of leaves. Together these modifications increase the efficiency of photosynthesis, and despite the apparent complexity of the pathway, it has evolved at least 45 times independently within the angiosperms. To provide insight into the extent to which gene expression is altered between C3 and C4 leaves, and to identify candidates associated with the C4 pathway we used massively-parallel mRNA sequencing of closely related C3 (Cleome spinosa) and C4 (Cleome gynandra) species. Gene annotation was facilitated by the phylogenetic proximity of Cleome and Arabidopsis. Up to 603 transcripts differ in abundance between these C3 and C4 leaves. This includes 17 transcription factors, putative transport proteins, as well as genes that in A. thaliana are implicated in chloroplast movement and expansion, plasmodesmatal connectivity, and cell-wall modification. These are all characteristics known to alter in a C4 leaf, but which previously had remained undefined at the molecular level. We also document large shifts in overall transcription profiles for selected functional classes. Our approach defines the extent to which transcript abundance in these C3 and C4 leaves differs, provides a blueprint for the NAD-ME C4 pathway operating in a dicotyledon, and furthermore identifies potential regulators. We anticipate that comparative transcriptomics of closely related species will provide deep insight into the evolution of other complex traits.


Introduction

C₄ photosynthesis is a complex biological trait which enables plants to either accumulate biomass at a much faster rate or live in adverse environments compared with ‘ordinary’ plants (Hatch, 1987; Osborne and Freckleton, 2009). These C₄ plants have added a CO₂ concentration mechanism on top of their regular photosynthetic carbon fixation which makes them not only more efficient at assimilating inorganic carbon; they frequently also have higher water and nitrogen use efficiencies (Black, 1973; Oaks, 1994; Osborne and Freckleton, 2009). Beyond the basic biochemistry, our understanding of the C₄ photosynthesis is limited.

The principle of C₄ photosynthesis is deceivingly simple: instead of using Ribulose 1,5-Bisphosphate Carboxylase Oxygenase (RuBisCO) as the primary carbon fixing enzyme, C₄ plants use PEP carboxylase (PEPC). Unlike RuBisCO, PEPC is more specific for inorganic carbon (Hatch, 1987). Since the C₄ cycle is an add-on rather than a replacement for RubisCO and the Calvin-Benson Cycle, the prefixed CO₂ is transported in a bound form, a C₄ acid, hence the name, to the site of RuBisCO. The C₄ cycle generates high concentrations of CO₂ around RubisCO (Hatch, 1987) and this increases the rate of photosynthesis because competition between CO₂ and O₂ at the active site of RuBisCO is reduced (Jordan and Ogren, 1984). In most C₄ plants, concentrating CO₂ around RuBisCO involves the reactions of photosynthesis being partitioned between bundle sheath (BS) and mesophyll (M) cells as well as changes to cell biology and leaf development (Hatch, 1987; Sage, 2004), although in some lineages, C₄ photosynthesis operates within individual cells (Reiskind et al., 1989; Keeley, 1998; Voznesenskaya et al., 2001; Voznesenskaya et al., 2002; Voznesenskaya et al., 2003).

In all known C₄ plants CO₂ enters M cells and is converted into bicarbonate by carbonic anhydrase (CA). Phosphoenolpyruvate carboxylase (PEPC) then combines HCO₃⁻ with phosphoenolpyruvate (PEP) to generate the C₄ oxaloacetic acid (OAA), which is rapidly converted into either aspartate or malate. These C₄ acids then diffuse to the site of RuBisCO.
through abundant plasmodesmata, where C₄ acid decarboxylases release CO₂ (Hatch, 1987). Three distinct C₄ acid decarboxylases known as NADP-dependent malic enzyme (NADP-ME), NAD-dependent malic enzyme (NAD-ME) and phosphoenolpyruvate carboxykinase (PEPCK) have been co-opted into the C₄ pathway, and this has been used to define three biochemical subtypes of C₄ photosynthesis. The three carbon compound released after decarboxylation diffuses back to the M cells, and is converted to PEP catalysed by pyruvate,orthophosphate dikinase (PPDK) (Hatch and Slack, 1968). Because the enzymes involved in the C₄ cycle are found in the cytosol, chloroplasts and mitochondria, a significant amount of transport across organellar membranes is required for the C₄ cycle to operate. However, few genes encoding transporters that allow the increased intracellular flux of metabolites required for C₄ photosynthesis have been identified (Bräutigam et al., 2008a; Majeran and van Wijk, 2009). In addition, we have a very limited understanding of mechanisms controlling the altered cell biology and morphology associated with C₄ leaves. The C₄ cycle likely affects not only the relatively small number of enzymes and transport proteins needed to perform the core reactions but, given the consequences on the ecological performance of the plants, also a range of other processes.

The gaps in our understanding of mechanisms underlying C₄ photosynthesis limits insight into a metabolic pathway that has evolved repeatedly at least 45 times in plants (Sage, 2004), and so is of interest in terms of understanding a remarkable example of convergent evolution. In addition, because C₄ plants are amongst the most productive on the planet and the pathway is associated with increased water and nitrogen use efficiencies (Brown, 1999), it has been suggested that characteristics of C₄ photosynthesis should be placed into C₃ crops (Matsuoka et al., 2001; Mitchell and Sheehy, 2006; Hibberd et al., 2008). A more complete understanding of genes involved in C₄ photosynthesis is fundamental to attempts at placing components of the C₄ pathway into C₃ crops to increase yield.
Recently, a new set of tools has become available to analyze species without sequenced genomes on a genomic scale – next generation sequencing (NGS) technology (summarized in Metzker, 2010). With NGS, the transcriptome of a tissue can be sequenced and quantified at the same time (RNA-seq) (Wang et al., 2009). The 454 FLX genome sequencer provides a quarter million sequence reads of 230 bases in each run from a cDNA template generated from mRNA (http://www.454.com/; Metzker, 2010). The resulting reads can be mapped onto a closely related reference to quantify the number of reads matching a gene locus thus providing a measure of transcript abundance (Flicek and Birney, 2009; Bräutigam and Gowik, 2010). We chose to compare the C4 plant *Cleome gynandra* with the C3 plant *Cleome spinosa* since they are members of the same genus and are closely related to *Arabidopsis thaliana* (Brown et al., 2005; Marshall et al., 2007). Given the close phylogenetic relationship, we can take advantage of the well annotated Arabidopsis genome (Swarbreck et al., 2008) and its known genome history (Bowers et al., 2003; Haberer et al., 2004; Thomas et al., 2006) to identify and quantify the biological functions regulated at the level of transcript abundance in the C4 species compared with the C3 species. Although the experiment will also capture variation in the abundance of transcripts associated with differences between the species that do not relate to C4 photosynthesis, the close proximity of the Cleome species should reduce this effect. We chose to use mature fully differentiated leaves for the analysis since we wanted to minimize the influence of species-specific effects during leaf differentiation but rather focus on transcript profiles when C4 photosynthesis is fully operational. Once this profile is defined, analysis of developmental stages may reveal how the profile is achieved during differentiation.

By comparing the transcriptomes of closely related C3 and C4 species, we will test (i) whether cross species transcriptomic comparisons are feasible, (ii) the degree to which the core C4 cycle enzymes and transport proteins are regulated at the level of transcript abundance, (iii) whether the changes in metabolism associated with C4 photosynthesis are
associated with additional unexpected shifts in transcript profiles in leaves of C₄ compared
with C₃ plants, and (iv) define candidates for additional functions critical to C₄ photosynthesis
based on unbiased observation of the data. By analyzing the complete transcriptome, we
define the maximal extent to which the C₄ pathway alters leaf transcript profiles.

Results

Physiological analysis of C₃ and C₄ leaves confirms C₄ metabolism in C. gynandra

To confirm that the C. spinosa and C. gynandra leaves we used for transcriptomic analysis
were using C₃ and C₄ photosynthesis respectively, we analyzed the steady-state levels of
metabolites associated with the C₄ cycle. For example, large quantities of aspartate, alanine
and pyruvate are produced in M and BS cells of NAD-ME C₄ leaves, and they were 19, 3.9,
and 3.6 times more abundant respectively, in C. gynandra compared with C. spinosa (Table
S1). In contrast, and in agreement with the lower demand for the photorespiration in C₄
leaves, glycerate and glycolate, intermediates of the photorespiratory cycle, were 4.5 and 1.9
times more abundant in C. spinosa (Table S1). We also determined the extractable activities
of PEPC, aspartate aminotransferase (AspAT), NAD dependent malate dehydrogenase (NAD-
MDH), NAD-ME and alanine aminotransferase (AlaAT). Except for NAD-MDH,
significantly higher activities of the enzymes required for the C₄ cycle were measured in C.
gynandra leaf extracts (Figure S1). The metabolite profiling of leaf extracts using GC-EI-
TOF and the enzyme activity assays showed that the plants we used for digital gene
expression analysis had clear differences in their metabolite profiles and enzyme activities,
and these were consistent with functional C₃ and C₄ photosynthesis operating in leaves of C.
spinosa and C. gynandra, respectively.

The leaf transcriptomes for closely related C₃ and C₄ species are qualitatively similar
To obtain sequence tags for digital gene expression (DGE) analysis from *C. spinosa* (C3) and *C. gynandra* (C4), RNA was isolated from mature leaves of each species and prepared for 454 sequencing. One sequencing run on a GS FLX sequencing system was conducted on leaf cDNA isolated from either *C. gynandra* or *C. spinosa*. From *C. spinosa*, we obtained 70,564,592 nucleotides (nts) and from *C. gynandra* 91,851,136 nts of raw sequence that after quality control corresponded to 65,525,139 nts and 85,681,233 nts, respectively (Table 1). The mean read-length of the cleaned sequence reads was 232 nts for *C. gynandra* and 230 nts for *C. spinosa* (Table 1).

To exclude program specific mapping artifacts and to test whether the *C. gynandra* and *C. spinosa* libraries behave robustly similar during mapping, two different programs, BLAST and BLAT, were used to align the reads to Arabidopsis as the reference genome. To define the most suitable mapping parameters an array of parameters for mappings in both the DNA and protein space were tested (Table 2). Neither the *C. gynandra* nor the *C. spinosa* library mapped well to Arabidopsis cDNAs in the DNA space using BLAT or BLAST, although the differences are more dramatic for BLAT (Table 2). In the protein space, however, the proportion of mapped reads increased dramatically. When 75% amino acid sequence identity was required, three quarters of the reads could be mapped with BLAT, resulting in 1.48 and 1.57 average mappings per read, respectively. Even with the most lenient mapping parameters the proportion of mapped reads did not exceed 83% with BLAT and 78.8% with BLAST (Table 2). In all mapping attempts, the *C. gynandra* and the *C. spinosa* read libraries yielded qualitatively similar mapping results, irrespective of mapping program or parameters.

To obtain a stringent yet inclusive mapping, the mapping conducted in protein space at ≥75% identity with BLAT was chosen and this mapping file was parsed by in-house scripts to keep only the read match with the highest number of matching bases. For a more lenient mapping, a BLAST mapping at a cut-off of 1e⁻⁵ was chosen and parsed to keep only the best BLAST Hit for each read. For each AGI the number of matching reads was counted and the hit
count was then transformed to Reads Per Million (RPM) to normalize for the number of reads available for each species. After parsing, the sequenced libraries matched between 50.5% and 55.3% of the genes in the Arabidopsis reference (Table S2).

To assess whether the datasets for the two different species and the two different mappings were qualitatively similar, we tested the coverage of the functional classes. Overall about 50% of all genes were represented in both species with the BLAT (Figure 1A) and the BLAST mapping (Figure 1B). Although the majority of gene classes were represented by more than 50% of genes in each class for both mappings the classes ‘function unknown’, ‘putative lipid transfer protein’, ‘storage protein’ and ‘defence’ were under-represented compared to all genes (Figure 1A and B). Genes present in the organellar genomes were not well represented (Table S3). Genes classified into primary metabolism including ‘photosynthesis’, ‘central carbon’, ‘nitrogen metabolism’, ‘amino acid’ and ‘nucleotide metabolism’ as well as many cellular processes were well-represented categories, and about four-fifths of genes predicted to be involved in the C₄ pathway were detected in both species. Overall the pattern of detection in the different gene classes was similar for both species and independent of the program used for the mapping (Figure 1A and B).

Transcripts of known C₄ genes are more abundant with one exception

Detailed analysis of known C₄ genes showed that all but one gene necessary for the core C₄ cycle of NAD-malic enzyme type plants were massively upregulated in C. gynandra compared with C. spinosa. Transcripts encoding PEP carboxylase were upregulated 78-fold, aspartate aminotransferase were upregulated 343-fold, NAD-malic enzymes upregulated 27-fold and 21-fold, respectively, and alanine aminotransferase was upregulated 29-fold (Table 3). The results for the BLAT and the BLAST mappings were similar with one exception. In the BLAST mapping the reads mapping to PEPC were split onto two genes in the Arabidopsis reference genome whereas they mapped to only one gene in the BLAT mapping (Table 3).
Transcripts encoding mitochondrial malate dehydrogenases were increased only 1.3-fold (Table S3). Not only were genes associated with the C_4 pathway upregulated compared to C_3 but they also had high absolute read counts between 1,800 and 4,806 RPM.

The leaf transcriptomes for closely related C_3 and C_4 species are quantitatively different

Before undertaking detailed analysis of differences in transcript abundance between *C. gynandra* and *C. spinosa* we used quantitative polymerase chain reactions (qPCR) to confirm estimates of transcript abundance identified by RNA-Seq. We chose genes whose transcript abundance differed over four orders of magnitude, and used qPCR to assess their abundance. qPCR was performed on both the cDNA used for RNA-Seq and cDNA generated from RNA isolated from leaves in a separate experiment. This approach provided strong support for the differences in abundance of transcripts between the two species that we determined from RNA-Seq (Figure 2). Overall, this showed that the ratios of transcript abundance obtained by RNA-Seq-based DGE are suitable for calling differentially expressed genes between two related species.

Of the 13,662 transcripts we captured quantitative data for (Table S3), we identified 583 (BLAT) or 603 (BLAST) transcripts whose abundance differed significantly (*p* ≤ 0.01) between *C. spinosa* and *C. gynandra*, with 256/258 (1.2%/1.2%) transcripts being more abundant in *C. gynandra* and 327/345 (1.5%/1.6%) transcripts being more abundant in *C. spinosa* (Figure 3A and B, ‘all’). We tested whether significantly changed transcripts are enriched in functional categories and whether they were more highly expressed in the C_4 or the C_3 species. While the qualitative classification of detected genes showed a very similar pattern between *C. spinosa* and *C. gynandra* (Figure 1A and B), the quantitative analysis revealed massive differences in representation between gene classes in the C_3 and the C_4 species (Figure 3A and B). The transcript profile generated by the BLAT mapping (Figure 3A) is similar to the one generated by the BLAST mapping (Figure 3B), although not all genes called as
The classes containing the highest percentage of changed genes are the photosynthetic classes as well as the C₄ cycle, Calvin-Benson Cycle, and photorespiration (Figure 3A and B). The latter two have lower steady-state mRNA levels in C₄ leaf tissue (Figure 3A and B bottom) while the photosynthetic classes of photosystem I, cyclic electron flow, and cytochrome b₆/f complex as well as the C₄ cycle have higher levels in C₄ leaf tissue (Figure 3A and B top). A number of classes involved in primary metabolism also have lower steady state transcript levels in C₄ tissues: one carbon compound metabolism, other central carbon metabolism, shikimate pathway and amino acid metabolism. Protein synthesis also has lower steady state transcript levels which are limited to cytosolic and plastidic protein synthesis genes (Figure S3). Among the classes with higher steady state transcript levels are starch metabolism, cofactor synthesis, putative lipid transfer proteins, nitrogen metabolism, and beta-1,3 glucan metabolism. The quantitative pattern (Figure 3A and B) is similar to the qualitative pattern (Figure 1A and B) with regard to the influence of the mapping program; the BLAT and BLAST mappings look remarkably similar with the exception of shikimate metabolism.

*Transcripts with similar patterns of abundance compared with bona fide C₄ genes and RuBisCO*

The list of 13,662 transcripts detected in either *C. spinosa* or *C. gynandra* tissues and the list of 603 transcripts which are differentially regulated between both species (Table S3, BLAST mapping) prompted us to determine which transcripts showed changes in abundance similar to the core C₄ genes or RuBisCO subunit encoding genes. Such transcripts display both a large fold-change between the C₃ and the C₄ plants and large absolute read numbers. For example, among the transcripts encoding putative transport proteins, three plastidic transport proteins, the phosphoenolpyruvate phosphate translocator PPT, a putative bile acid: sodium symporter, and a putative proton:sodium antiporter, two mitochondrial dicarboxylate
carriers and one plasmamembrane intrinsic protein were massively upregulated in C₄ C. gynandra (Table 4). No transcripts encoding transport proteins are found to be downregulated to a comparable degree. Among metabolic genes, two cytosolic carbonic anhydrases, one of which (CA4; see Table 4) is likely tethered to the plasma membrane, an adenylate kinase and a pyrophosphatase are upregulated at levels comparable to those of C₄ cycle genes. Many proteins of unknown function showed differential expression, the most striking case being a putative lipid transfer protein, also annotated as an extensin-like protein. Based on annotation and differential expression pattern, several transcripts predicted to encode known C₄ functions that have not yet been assigned to genes, such as CHUP1 and actin for chloroplast positioning or callose degrading enzymes for regulating plasmodesmatal opening were identified (Table 4).

Regulatory genes that are significantly changed

The transcript profiles of these C₃ and the C₄ species identify a number of regulatory proteins that are candidates for maintaining C₄ status. Among transcripts encoding proteins with regulatory functions, 43 were significantly upregulated in either C. gynandra or C. spinosa (Figure 3A and B). These include bona fide transcription factors, protein phosphatases and kinases, and the regulatory proteins of the pyruvate dehydrogenase complex (up in C₄), of PPDK (up in C₄) and of RuBisCO (down in C₄). Only seventeen transcription factors are significantly changed, seven of those have higher steady state mRNA levels compared to the C₃ leaf tissue while ten have lower steady state mRNA levels (Table 5).

In addition to the detailed quantitative and qualitative analysis of read mappings to generate ESTs for both species, contigs were assembled from cleaned reads for each species as described previously (Weber et al., 2007; Bräutigam et al., 2008b) and then annotated by BLASTX vs. the TAIR 9 protein models. 18,992 and 17,655 contigs representing a total
sequence length of 9,062,043 and 7,746,894 nts were obtained for *C. gynandra* and *C. spinosa*, respectively (Table 1).
Discussion

Transcriptomic comparisons of different species with NGS technology are feasible

Read mapping by alignment is a well-established tool to quantify transcript abundance and thus determine mRNA steady state levels (Wall et al., 2009; Metzker, 2010). The concept of mapping to a cross species reference has also been established theoretically (Palmieri and Schlotterer, 2009), although the potential has not been experimentally explored to date (Bräutigam and Gowik, 2010).

To explore cross species mapping, the transcriptome sequencing was carried out using 454 FLX, a long read technology, since theoretical work had established that at least BLAT is capable of mapping reads which contain alterations in comparison to the reference if the reads are at least 100 bases long (Palmieri and Schlotterer, 2009). We also established a reference database, which removes the genome history of Arabidopsis as far as it is known (Bowers et al., 2003; Haberer et al., 2004; Thomas et al., 2006). Tandem duplicated genes and segmentally duplicated genes (remnants of the last whole genome duplications) were removed to prevent genome history from interfering with comparative quantitative mapping (Bräutigam and Gowik, 2010).

Both BLAT and BLAST mappings indicate that using a minimal reference does not diminish read mappings (Table S4) while avoiding mapping problems based on genome history (Bräutigam and Gowik, 2010). The mappings in protein space allowed more successful read mappings because protein sequences diverge slower than nucleotide sequences. Although the proportion of reads mapped varied with changing mapping parameters (Table 2, Table S4), the C. spinosa and C. gynandra libraries yielded similar results indicating that, evolutionary, both species are approximately equally distant to A. thaliana with mapping incurring similar penalties depending on parameters.
Since no read alignment program has emerged as the consensus program for NGS data analysis, two different programs were used for mapping and the output compared in all cases. The output proved robust against changing the mapping program both qualitatively and quantitatively. When we mapped the quarter million reads obtained from each species of *Cleome* to a minimized TAIR 9 release of the *A. thaliana* genome they corresponded to ~11,000 loci. As the minimized TAIR 9 dataset contains 21,972 gene loci, the reads we collected in *C. gynandra* and *C. spinosa* represent approximately 50% of the transcriptome. In *A. thaliana* seedlings approximately 60% of the loci represented in the TAIR 8 release were detectable (Weber et al., 2007), hence we have likely captured a large proportion of the transcripts associated with leaves of *C. spinosa* and *C. gynandra*.

The qualitative representation of gene classes detected reflects that leaf tissues were analyzed. While photosynthetic genes as well as primary metabolism are well represented in all datasets, genes implicated in cell walls, secondary metabolism, and defence responses are under-represented (Figure 1A and B). These classes contain genes that are likely specific to certain tissues, developmental stages, or environmental challenges. For example cell wall genes may be better represented if our sampling had included expanding leaf or stem material (Schmid et al., 2005), and stress-response genes better represented if plants were sampled after exposure to extreme conditions (Kilian et al., 2007).

Likewise, certain pathways of secondary metabolism are likely restricted to defined tissues or developmental stages making it unlikely that we would pick up many of these genes when profiling leaf libraries. Based on the gene detection pattern, the two plant species did not encounter different biotic or abiotic stressors or are not in different stages of growth as very similar genes were detected in both species (Figure 1A and B, Figure 3A and B).

Finally, only a very small proportion of transcripts showed significant differences in abundance between the two different species (Tables S2 and S3) and these changes were enriched in a limited number of functional classes (Figure 3A and B). We conclude that cross
species mapping in protein space is a feasible strategy to compare different species as long as an equidistant reference is available (Bräutigam and Gowik, 2010).

**Transcripts derived from core C_4 cycle genes are more abundant in the C_4 species**

C_4 photosynthesis has evolved convergently in many different lineages of plants (Sage, 2004) and in many cases the alterations to expression of specific genes has been related to transcriptional regulation (summarized in Sheen, 1999). Our genome scale analysis allowed us to compare the steady state transcript levels for all candidate C_4 genes at the same time. For all of the enzymes where a change in total extractable activity could be shown (Figure S1), a higher mRNA level of at least one isoform as judged from the read count is also present (Table 3). The only enzyme showing no changes in transcript level is the mitochondrial NAD-MDH. Possibly, the activity of the mitochondrial NAD-MDH is high enough already in C_3 plants to support a C_4 type metabolic flux. The only transport protein known to date that is involved in the C_4 cycle, the phosphoenolpyruvate phosphate translocator (Fischer et al., 1997; Bräutigam et al., 2008a), is also upregulated 20-fold indicating that this transport protein is regulated at the level of mRNA abundance. Based on similarities in transcript abundance to known C_4 genes, our comparative RNA-seq also identified likely additional components needed for C_4 photosynthesis. When PPDK was characterized it was proposed that adenylate kinase as well as inorganic pyrophosphatase need to be abundant in C_4 chloroplasts (Hatch and Slack, 1968). RNA-seq confirmed this prediction and showed that the upregulation also occurs at the level of transcript abundance. Taken together, we found that almost all transcripts encoding the proteins required for the core C_4 cycle have higher steady state mRNA levels and we propose that, at least in *C. gynandra*, the activity of C_4 cycle enzymes and transport proteins is controlled at least partially at the level of transcript abundance.
Alterations to the abundance of transcripts associated with other metabolic processes

Changes in the abundance of transcripts that are not associated with the core C₄ cycle are also detectable in leaves of *C. gynandra* and *C. spinosa*. The high flux C₄ cycle poses additional demands in terms of ATP and reduction equivalents on the light reaction (Hatch, 1987). Specifically, the recycling of the initial CO₂ acceptor PEP requires additional ATP molecules (Hatch, 1987). In C₄ leaf tissue one third to half of the genes in the photosynthetic gene classes which contribute to ATP production by cyclic electron flow are upregulated compared to C₃ leaf tissue: photosystem I, the cytochrome b6/f complex and the genes mediating cyclic electron flow themselves (Figure 3A and B). It remains an open question as to whether these higher steady state levels are caused by higher ATP demand or whether C₄ photosynthesis requires upregulation of these genes to meet the ATP demand prior to establishing C₄ photosynthesis.

On the other hand, the classes of Calvin-Benson Cycle genes and photorespiratory genes are those with the highest number of genes with significantly lower steady state mRNA levels. It is a well-established fact that most C₄ plants have less RuBisCO protein compared to C₃ plants (Ku et al., 1979) and that flux through the photorespiratory pathway is reduced compared with C₃ species (Chollet and Ogren, 1975; Leegood, 2002). Transcripts encoding the large and small subunits of RuBisCO were reduced from 22,968 and 15,442 RPM to 6,984 and 4,900 RPM in *C. spinosa* and *C. gynandra* respectively. Overall the trend for Calvin-Benson Cycle genes was for them to be downregulated in *C. gynandra* compared with *C. spinosa* (Figure 3A and B). Likewise, a large number of genes encoding photorespiratory proteins, proteins involved in one carbon compound metabolism and the genes involved in ammonia re-assimilation, glutamine synthethase and glutamate synthase, have lower steady state transcriptional levels (Table S3, Figure 3A and B). The reduced flow through the photorespiratory pathway obviously decreases the demand on the expression system to maintain high steady state levels of mRNA for many Calvin-Benson Cycle and
photorespiratory genes. The photosynthetic genes, the Calvin-Benson Cycle and photorespiratory genes (in C₃) and the C₄ cycle genes (in C₄) are those with the highest read counts of the genes with known function (Table S3). Although it is currently not possible to quantify absolute transcript levels since the genomes of neither Cleome species has been sequenced, the high read counts obtained for the genes of central carbon metabolism and photosynthesis indicate that the steady state level of transcripts are high. Since the most altered gene classes are also those that contain the genes with the highest absolute read count, it is not clear whether C₄ photosynthesis lowers or raises the demand on protein synthesis and accessory pathways such as amino acid synthesis. However, both the protein synthesis and the amino acid metabolism classes contain more genes that have lower steady state levels in C₄ leaf tissue (Figure 3). Within the protein synthesis gene class, many transcripts encoding structural components of plastidic and cytosolic ribosomes were reduced (Figure S3). This was not the case for components of mitochondrial ribosomes (Figure S3) indicating that there is not a general effect on translation but that the effect is likely specific to ribosomes involved in translation for the Calvin-Benson Cycle and photorespiration. The protein to fresh-weight ratio is also lower in C₄ leaf tissue compared to C₃ leaf tissue (Figure S2). We propose that plastidic ribosomes are relieved of the high translation load associated with the large subunit of RuBisCO and the cytosolic ribosomes need to translate fewer transcripts associated with central carbon metabolism as well as the small subunit of RuBisCO. The reduced production of proteins in the leaves of C₄ plants is considered important in increasing nitrogen use efficiency because the rate of photosynthesis per unit nitrogen in the leaf is increased (Oaks, 1994). Our data indicate that there is also likely a significant saving in the nitrogen provision in the leaf because fewer ribosomes as well as fewer proteins for central carbon metabolism are required.

The dataset contains two additional gene classes, beta-1,3 glucan metabolism and putative lipid transfer proteins, that showed differences in transcript abundance between C. gynandra
than *C. spinosa* that could be explained within the current framework of knowledge of C4 photosynthesis. The C4 pathway requires efficient exchange of metabolites between M and BS cells via large numbers of plasmodesmata connecting both cell types while the BS cell wall of many C4 plants is suberized to reduce diffusion of CO2 away from RuBisCO (Hatch, 1987). Transcripts encoding three distinct glucan 1,3-beta-glucosidases (Table 4) involved in governing plasmodesmatal conductivity by regulating the turnover of the beta-1,3-glucan callose (Levy et al., 2007) were upregulated in leaves of *C. gynandra* compared with *C. spinosa*. It is therefore possible that these genes are involved in increasing the open probability of plasmodesmata (Roberts and Oparka, 2003) which allows the efficient flux of organic acids between M and BS cells required during C4 photosynthesis (Evert et al., 1977; Botha, 1992; Roberts and Oparka, 2003). A transcript annotated as a putative lipid transfer protein is among those that are most highly up-regulated in *C. gynandra* compared with *C. spinosa*. LTPs are required for the export of lipids to the cell wall during cutin biosynthesis (DeBono et al., 2009). Interestingly, in *A. thaliana* some LTPs are exclusively and abundantly expressed in the root endodermis where suberin biosynthesis is required to establish the Casparian strip.

There are additional changes in the transcript profile that are less easily explained. Among the gene classes containing more genes with significantly higher transcript levels in C4 leaf tissue are starch metabolism, cofactor synthesis and nitrogen metabolism and heat shock/protein folding (in order of decreasing number of significantly different genes). On the other hand it is difficult to conceive why genes involved in metal handling are frequently lower in transcript level in C4 leaf tissues (Figure 3A and B). These changes may be connected to currently unknown phenomena relating to the C4 pathway or be part of differences not relating to C4 photosynthesis between the two species. Overall the global analysis of transcription on the level of functional classes reveals unexpected shifts in
transcript profiles which can be explained based on the current knowledge about the C₄ pathway while a range of, albeit smaller, changes remain enigmatic.

Finally, our global transcriptional analysis of C₄ and C₃ leaf tissues not only allows testing hypotheses about the C₄ pathway on a global scale but also allowed genes with expression patterns similar to those of known C₄ genes to be identified. The phylogenetic proximity of the Cleomaceae to *A. thaliana* allows the identification of the orthologues in *A. thaliana* which will facilitate translational research into the model species (Brown et al., 2005).

**Candidates for additional C₄ related genes**

The identification of transport proteins involved in the C₄ cycle lags behind that of enzymes considering that the C₄ cycle requires the intracellular transport of pyruvate, PEP, aspartate and alanine across different organellar membranes (Bräutigam and Weber, 2010). A wide range of C₄ plants take up pyruvate into chloroplasts from the M in co-transport with sodium (Aoki et al., 1994; Aoki and Kanai, 1997), which might explain the requirement for sodium as a micronutrient in many C₄ species (Brownell and Crossland, 1972). Since the rate of pyruvate transport into C₄ M cell chloroplasts occurs at or exceeds the apparent rate of CO₂ assimilation, sodium-coupled pyruvate import implies a large influx of sodium into these chloroplasts, but the transporter has not yet been identified at the molecular level (Aoki and Kanai, 1997). Our finding that a putative plastidic proton:sodium symporter (NHD1) is 16-fold up-regulated in *C. gynandra* prompts us to hypothesize that it functions in exporting sodium from the chloroplast in order to maintain the sodium gradient required for import of pyruvate. In addition, we found strong upregulation of a putative bile acid:sodium cotransporter in *C. gynandra*. Interestingly, up-regulation of the putative bile acid:sodium cotransporter or of NHD1 was not observed in maize (Bräutigam et al., 2008a), which belongs to a group of C₄ plants that show proton-dependent, not sodium-dependent transport of pyruvate into M cell chloroplasts (Aoki et al., 1994; Aoki and Kanai, 1997). PEP generated
from pyruvate in M cell chloroplasts is exported from these chloroplasts by PPT, thereby providing the substrate for the cytosolic PEPC reaction. Accordingly, transcripts encoding PPT are 20-fold up-regulated in *C. gynandra*, likely reflecting the increased requirement for transport of PEP (Table 3). In contrast to what has been observed for the NADP-ME-type C₄ plant maize by quantitative proteomic analysis (Bräutigam et al., 2008a), we did not detect increased transcript abundance of the putative M chloroplast oxaloacetate/malate exchanger DiT1 (Taniguchi et al., 2002; Renne et al., 2003; Taniguchi et al., 2004) (Table S3). This is consistent with the fact that OAA/malate shuttling across the M cell chloroplast envelope membrane is not required for NAD-ME-type C₄ photosynthesis (Bräutigam and Weber, 2010; Weber and von Caemmerer, 2010). The mitochondrial dicarboxylate carriers are prime suspects for the C₄ acid importer into the mitochondria where decarboxylation takes place (Table 4). The initial uptake of inorganic carbon and its conversion to bicarbonate may be facilitated by the concerted action of a membrane intrinsic protein channelling the gas and a carbonic anhydrase that is predicted to be membrane bound (Table 4).

Chloroplasts in the BS of *C. gynandra* are larger than those in the BS of C₃ species and as in many other C₄ plants are positioned in a strictly centripetal pattern (Marshall et al., 2007; Voznesenskaya et al., 2007). Transcripts derived from the *GC1* (*GIANT CHLOROPLAST1*) gene were more abundant in *C. gynandra* than in *C. spinosa* (Table 4). Although over-expression of *GC1* in *A. thaliana* is reported not to effect chloroplast division (Maple et al., 2004), it is possible that it does so in *C. gynandra*. In addition, we also detected reduced accumulation of transcripts derived from the *CHUP1* and *ACTIN11* genes. In *A. thaliana* the outer chloroplast envelope membrane protein CHUP1 (chloroplast unusual positioning 1) contains an actin-binding motif, and is required for preventing chloroplast aggregation (Oikawa et al., 2003). Differential positioning of chloroplasts in BS and M of C₄ plants finger millet and maize requires the actomyosin system (Kobayashi et al., 2009). Since AtCHUP1 is involved in positioning chloroplasts at the periclinal plasma membrane during the weak-light
acclimation response via a coiled-coil domain and interaction with the cytoskeleton (Oikawa et al., 2003), it is possible that the centripetal positioning of chloroplasts in BS cells is linked to lower expression of the CgCHUP1 and ACTIN11 genes.

Controlling and maintaining a C₄ state in leaf tissue

Our estimate that around 603 transcripts accumulate differentially in leaves of C₃ and C₄ species provides insight into the extent to which gene expression profiles change in C₄ leaves. For example, the fact that 258 transcripts were more abundant in the leaves of the C₄ compared to the species indicates that about 2.8% of the leaf transcriptome differentially accumulates in C₄ leaves (Tables S2 and 6). To compare the complexity of the C₄ pathway with other multigenic traits, we assessed the number of transcripts that are known to be regulated by sugars, cold, diurnal and circadian rhythms as well as attack by pests and pathogens (Table 6). Interestingly, the alterations in transcript abundance of leaves of C. gynandra compared with those of C. spinosa, were greater than those observed in response to cold treatment, and lower than those induced by glucose feeding, those occurring during pathogen attack, and of the response to both diurnal and circadian rhythms. As significant progress has been made in understanding sugar signalling (Rolland et al., 2006), pathogen attack (Wise et al., 2007) and the control of gene expression in response to the diurnal cycle and circadian rhythms (Imaizumi et al., 2007), it should be possible to identify the regulators responsible for these alterations in transcript abundance in a C₄ leaf compared with a C₃ leaf. The changes in transcript abundance that we document in a C₄ leaf compared with a C₃ leaf likely over-represents the changes in transcript abundance actually associated with C₄ photosynthesis on a whole leaf basis, as some differences in gene expression are likely due to the phylogenetic distance between C. gynandra and C. spinosa. A more confident estimate of the extent to which the leaf transcriptome is altered in association with C₄ photosynthesis will be generated when additional con-generic pairs of C₃ and C₄ species are subjected to deep
transcriptome analysis, and shared transcripts are identified. Between M and BS cells, the alterations in gene expression may be greater than those that we have defined for whole leaves. For example, up to 18% of genes are estimated to be differentially expressed between M and BS cells of maize (Sawers et al., 2007). However, it is not clear how different the transcript profiles of M and BS cells are in a dicot C₃ leaf, and until this is defined, it is not possible to infer the extent to which transcript abundance alters in these cell-types in association with C₄ photosynthesis.

As we sampled from mature leaves to capture the differences between C₃ and C₄ leaves at the point of fully differentiated pathways we likely also captured regulatory genes needed to maintain C₄ architecture and metabolism in mature leaves. Of the seventeen transcription factors significantly altered (Table 5), *GOLDEN2-LIKE1* (*GLK1*), has previously been implicated in regulating genes important in C₄ photosynthesis. In maize GOLDEN2 (G2) controls functional differentiation of chloroplasts in BS cells (Langdale and Kidner, 1994), and GOLDEN2-LIKE1 has been implicated in the expression of photosynthesis genes in M cells (Rossini et al., 2001). The fact that *GLK1* transcripts are significantly more abundant in leaves of *C. gynandra* would not necessarily be predicted as previous work indicates it becomes specialised in BS cells of C₄ leaves, but not that its abundance is altered significantly. This implies that the increase in abundance of *GLK1* transcripts may not simply be due to its involvement in C₄ photosynthesis. When over-expression of *GLK1* was induced in *A. thaliana* the abundance of 114 transcripts were altered (Waters et al., 2009). We assessed the extent to which the genes that are controlled by *GLK1* alter in abundance in leaves of *C. gynandra* compared to *C. spinosa*, and found that only 19 genes were shared between the two datasets. This may be due to a number of factors that could include the following; that there are differences in the targets of GLK1 in *A. thaliana* and *C. gynandra*; that a number of other transcriptional regulators are more important than GLK1 in maintaining patterns of photosynthesis gene expression in *C. gynandra*; or that a rapid
induction of GLK1 gene expression has more impact than increasing the steady state level of GLK1. This analysis is also subject to the caveat that in neither case were the amounts of GLK1 protein measured.

In all of our analysis, differences in transcript abundance between the leaves of C. gynandra and C. spinosa may reflect the operation of the C4 and C3 photosynthetic pathways, or alternatively, they may be due to differences in metabolism and cell biology associated with the phylogenetic distance between the two species. However, in many cases, it is striking that our analysis has identified differences in abundance of transcripts derived from genes that have been documented to be involved in processes known to alter in a C4 leaf. Taken together the analysis allows us to significantly extend the number of C4 related genes controlled at the level of transcript abundance and extend the current model for C4 related processes in NAD-ME C4 plants (Figure 4). Analysis of additional pairs of C3 and C4 species will likely facilitate identification of genes specifically involved in the C4 pathway, and exclude genes that are modified for other reasons.

**Methods**

**Plant material and 454 sequencing**

*Cleome spinosa* and *Cleome gynandra* plants for transcript profiling by RNA-Seq were grown in standard potting mix in a glasshouse in August and September of 2007. To obtain sequence tags for DGE analysis from *C. spinosa* and *C. gynandra*, total RNAs were isolated from fully-expanded leaves sampled from 56-day-old plants of each species. mRNA was reverse-transcribed to cDNA after two consecutive rounds of oligo(dT) purification and prepared for 454 sequencing as described previously (Weber et al., 2007).

**Mapping and quantification of the sequence reads**
Evolution did not stop in the lineage to the reference genome of *A. thaliana* after the Cleomaceae branch diverged. Hence there may be genes which were tandem duplicated or retained after the whole genome duplication event of the Brassicaceae that are absent in either of the Cleomaceae (Bräutigam and Gowik, 2010). To avoid mapping problems such as splitting of reads or mapping errors due to differential retention of genes in either Cleomaceae or Arabidopsis, we created a minimal genome for mapping. The remnants of the last whole genome duplication in the lineage of the Brassicaceae (Bowers et al., 2003; Thomas et al., 2006) and the tandem duplicated genes (Haberer et al., 2004) were reduced to one representative for each based on the TAIR9 coding sequence set. In each case, the gene with the lowest AGI code was retained for mapping. For each gene, the supplemental files store whether there are duplicates and which duplicates match the gene (Tables S3 and S5). We recommend recovery of the associated duplicated genes followed by a detailed analysis with phylogenetic trees to define the true orthologue when translating the results of Cleomaceae analyses to Arabidopsis research.

The 454 sequence reads were mapped onto coding sequences of the minimalized TAIR9 genome by BLAT (Kent, 2002) and BLAST (Altschul et al., 1997) with varying parameters and the output was parsed with in-house PERL scripts to retain only the best matching AGI for each sequence read and the best BLAST hit, respectively. Differentially expressed transcripts were identified using the Poisson statistics developed by Audic and Claverie (Audic and Claverie, 1997) followed by a Bonferroni-correction to account for the accumulation of alpha-type errors when conducting multiple pair-wise comparisons (Audic and Claverie, 1997).

**Plant material and qPCR analysis**

Both species were grown in a growth chamber in long day conditions (16 hrs light / 8 hrs dark) under 350 μmol photons m$^{-1}$ s$^{-1}$, at 22°C, and 65% relative humidity prior to samples being taken for qPCR. qPCR was conducted on the same samples used for RNA-Seq, and also
on mature leaves collected at noon grown in the growth cabinet. For qPCR RNA was isolated using TriPure reagent (Roche Applied Science). RNA was treated with DNase I (Promega) and purified with RNeasy Mini Kit (Qiagen). First-strand cDNA was then synthesised with SuperScriptII reverse transcriptase (Invitrogen) using 4 µg of RNA and oligo(dT) primers (Roche Applied Science). qRT-PCR was carried out with 96-well plates using a DNA Engine thermal cycler, Chromo4 real-time detector (BioRad), SYBR Green JumpStart Taq Ready Mix (Sigma) and 15-fold dilution of the cDNA as a template. Initial denaturation was carried out at 94°C for 2 min, then followed by 40 cycles of 94°C for 20 s, 60°C for 30 s, 72°C for 30 s, and 75°C for 5 s. Primers were designed to have melting temperatures of 60±0.5°C and to produce amplicons of 91 to 189 bp. The specificity of the primers and lack of primer dimers in the PCR reaction were verified using agarose gel electrophoresis and melting curve analysis. C_T values were generated for three technical replicates and four independent biological replicates per species. The comparative $2^{-\Delta\Delta CT}$ method was used to quantify relative abundance of transcripts (Livak and Schmittgen, 2001). $ACTIN7$ was chosen as a reference due to the 454 sequencing data showed equal, intermediate levels of $ACTIN7$ transcripts in both species. For the qPCR, standard errors were calculated from $2^{-\Delta\Delta CT}$ values of each combination of biological replicates.

Polar metabolite, chlorophyll, protein and enzyme activity analysis

For metabolite analysis, mature leaves from 56-day-old plants were collected in the middle of the light period and immediately frozen in liquid nitrogen. Three independent biological replicates were used. The tissues were ground in a mortar and a 50 mg fresh weight aliquot was extracted using the procedure described by Lee and Fiehn (Lee and Fiehn, 2008). Ribitol was used as an internal standard for data normalization. For GC-EI-TOF analysis, samples were processed and analyzed according to Lee and Fiehn (Lee and Fiehn, 2008). Enzyme
activities, chlorophyll and protein content were determined according to (Hausler et al., 2001).

Supplemental Material

The following materials are available in the online version of this article.

Supplemental Data Table S1. Relative abundance of predominant metabolites detected by GC-EI-TOF in _C. gynandra_ and in _C. spinosa_ confirms a functional C₄ pathway in the leaves of _C. gynandra_ used for analysis of transcript abundance.

Supplemental Data Table S2. Number of gene loci and number of differentially expressed genes detected with two different mapping programs, BLAT and BLAST.

Supplemental Data Table S3. Excel worksheet providing quantitative information for all reads mapped onto the reference genome from _Arabidopsis thaliana_.

Supplemental Data Table S4. Comparison of mapping parameters with two different programs, BLAST and BLAT, using a minimized and a complete reference genome.

Supplemental Data Table S5. Segmental and tandem duplicates in the _Arabidopsis thaliana_ genome.

Supplemental Figure S1. Quantitation of marker enzyme activities in leaf extracts of _C. spinosa_ and _C. gynandra_. PEPC, PEP carboxylase; NADME, NAD-dependent malic enzyme; PEPCK, PEP carboxykinase; AlaAT, alanine aminotransferase; AspAT, aspartate aminotransferase; NADMDH, NAD-dependent malate dehydrogenase.

Supplemental Figure S2. Protein to fresh weight and protein to chlorophyll ratios in leaves of _C. gynandra_ and _C. spinosa_.

Supplemental Figure S3. The percentage of genes encoding ribosomal components with significantly lower abundance of transcripts in C₄ (blue bars), or significantly higher abundance of transcripts in C₄ (red bars) based on the total number of genes in each annotation class (in parentheses on the y axes).
Acknowledgements

This work was supported by grants of the German Research Council (DFG WE 2231/4-1 to APMW, SFB TR1 to PW and APMW, IRTG 1525/1 to PW and APMW) and the Leverhulme Trust and Isaac Newton Trust (to JMH). We thank Tom Hardcastle for help with R.

Authors’ contributions

JMH and APMW designed the research
AB, KK, JW, SM DG & KLW performed the research
AB, KMC, UG, PW, JMH & APMW analyzed data
AB, JMH & APMW wrote the paper
Literature Cited

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ
(1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs. Nucl Acids Res 25: 3389-3402

Aoki N, Kanai R (1997) Reappraisal of the role of sodium in the light-dependent active
transport of pyruvate into mesophyll chloroplasts of C₄ plants. Plant Cell Physiol 38:
1217-1225

Aoki N, Ohnishi JI, Kanai R (1994) Proton/Pyruvate Cotransport Into Mesophyll
Chloroplasts Of C-4 Plants. Plant Cell Physiol 35: 801-806

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome
Res 7: 986-995

Black CC (1973) Photosynthetic Carbon Fixation In Relation To Net CO₂ Uptake. Annu Rev
Plant Physiol Plant Mol Biol 24: 253-286

Botha CEJ (1992) Plasmodesmatal Distribution, Structure And Frequency In Relation To
Assimilation In C₃ And C₄ Grasses In Southern Africa. Planta 187: 348-358

Bowers JE, Chapman BA, Rong JK, Paterson AH (2003) Unravelling angiosperm genome
evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:
433-438

Bräutigam A, Gowik U (2010) What can next generation sequencing do for you? Next
generation sequencing as a valuable tool in plant research. Plant Biol, in press

Bräutigam A, Weber APM (2010) Transport processes – connecting the reactions of C₄
photosynthesis. In AS Raghavendra, RF Sage (eds.), C₄ Photosynthesis and Related
CO₂ Concentrating Mechanisms, Springer, Berlin, Germany, in press

Bräutigam A, Hofmann-Benning S, Weber APM (2008a) 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

Bräutigam A, Shrestha RP, Whitten D, Wilkerson CG, Carr KM, Froehlich JE, Weber
APM (2008b) Comparison of the use of a species-specific database generated by
pyrosequencing with databases from related species for proteome analysis of pea
chloroplast envelopes. J Biotechnol 136: 44-53

Brown NJ, Parsley K, Hibberd JM (2005) The future of C₄ research - Maize, Flaveria or
Cleome? Trends Plant Sci 10: 215-221
Brown RH (1999) Agronomic implications of C4 photosynthesis. In RF Sage, RK Monson (eds.), C4 Plant Biology, Academic Press, San Diego, USA, pp. 473–507

Brownell PF, Crossland CJ (1972) The Requirement for Sodium as a Micronutrient by Species Having the C4 Dicarboxylic Photosynthetic Pathway. Plant Physiol 49:794-797

Chollet R, Ogren WL (1975) Regulation Of Photorespiration In C3 And C4 Species. Bot Review 41: 137-179

DeBono A, Yeats TH, Rose JKC, Bird D, Jetter R, Kunst L, Samuelsa L (2009) Arabidopsis LTPG Is a Glycosylphosphatidylinositol-Anchored Lipid Transfer Protein Required for Export of Lipids to the Plant Surface. Plant Cell 21: 1230-1238

Evert RF, Eschrich W, Heyser W (1977) Distribution And Structure Of Plasmodesmata In Mesophyll And Bundle-Sheath Cells Of Zea mays L. Planta 136: 77-89

Fischer K, Kammerer B, Gutensohn M, Arbinger B, Weber A, Hausler RE, Flugge UI (1997) A new class of plastidic phosphate translocators: A putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. Plant Cell 9: 453-462

Flicek P, Birney E (2009) Sense from sequence reads: methods for alignment and assembly. Nature Meth 6: S6-S12

Haberer G, Hindemitt T, Meyers BC, Mayer KFX (2004) Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis. Plant Physiol 136: 3009-3022

Hatch MD (1987) C4 Photosynthesis - a Unique Blend of Modified Biochemistry, Anatomy and Ultrastructure. Biochim Biophys Acta 895: 81-106

Hatch MD, Slack CR (1968) A New Enzyme For Interconversion Of Pyruvate And Phosphopyruvate And Its Role In C4 Dicarboxylic Acid Pathway Of Photosynthesis. Biochem J 106: 141-147

Hausler RE, Rademacher T, Li J, Lipka V, Fischer KL, Schubert S, Kreuzaler F, Hirsch HJ (2001) Single and double overexpression of C4-cycle genes had differential effects on the pattern of endogenous enzymes, attenuation of photorespiration and on contents of UV protectants in transgenic potato and tobacco plants. J Exp Bot 52: 1785-1803

Hibberd JM, Sheehy JE, Langdale JA (2008) Using C4 photosynthesis to increase the yield of rice - rationale and feasibility. Curr Opin Plant Biol 11: 228-231

Imaizumi T, Kay SA, Schroeder JI (2007) Circadian rhythms - Daily watch on metabolism. Science 318: 1730-1731
Jordan DB, Ogren WL (1984) The Co2/O2 Specificity Of Ribulose 1,5-Bisphosphate Carboxylase Oxygenase - Dependence On Ribulosebisphosphate Concentration, pH And Temperature. Planta 161: 308-313

Keeley JE (1998) C4 photosynthetic modifications in the evolutionary transition from land to water in aquatic grasses. Oecologia 116: 85-97

Kent WJ (2002) BLAT - The BLAST-like alignment tool. Genome Res 12: 656-664

Kilian J, Whitehead D, Horak J, Wanke D, Weinh S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J 50: 347-363

Kobayashi H, Yamada M, Taniguchi M, Kawasaki M, Sugiyama T, Miyake H (2009) Differential Positioning of C4 Mesophyll and Bundle Sheath Chloroplasts: Recovery of Chloroplast Positioning Requires the Actomyosin System. Plant Cell Physiol 50: 129-140

Ku MSB, Schmitt MR, Edwards GE (1979) Quantitative determination of ribulose bisphosphate carboxylase oxygenase protein in leaves of several C3 and C4 plants. J Exp Bot 114: 89-98

Langdale JA, Kidner CA (1994) Bundle-Sheath Defective, A Mutation That Disrupts Cellular-Differentiation In Maize Leaves. Development 120: 673-681

Lee DY, Fiehn O (2008) High quality metabolomic data for Chlamydomonas reinhardtii. Plant Meth 4: 7

Leegood RC (2002) C4 photosynthesis: principles of CO2 concentration and prospects for its introduction into C3 plants. J Exp Bot 53: 581-590

Levy A, Erlanger M, Rosenthal M, Epel BL (2007) A plasmodesmata-associated beta-1,3-glucanase in Arabidopsis. Plant J 49: 669-682

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods 25: 402-408

Majeran W, van Wijk KJ (2009) Cell-type-specific differentiation of chloroplasts in C4 plants. Trends Plant Sci 14: 100-109

Maple J, Fujiwara MT, Kitahata N, Lawson T, Baker NR, Yoshida S, Moller SG (2004) GIANT CHLOROPLAST 1 is essential for correct plastid division in Arabidopsis. Curr Biol 14: 776-781
Marshall DM, Muhaidat R, Brown NJ, Liu Z, Stanley S, Griffiths H, Sage RF, Hibberd JM (2007) Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C3 to C4 photosynthesis. Plant J 51: 886-896

Matsuoka M, Furbank RT, Fukuyama H, Miyao M (2001) Molecular Engineering of C4 Photosynthesis. Annual Rev Plant Physiol Plant Mol Biol 52: 297-314

Metzker ML (2010) Applications of Next-Generation Sequencing technologies - the next generation. Nature Rev Genet 11: 31-46

Mitchell PL, Sheehy JE (2006) Supercharging rice photosynthesis to increase yield. New Phytol 171: 688-693

Oaks A (1994) Efficiency Of Nitrogen-Utilization In C3 And C4 Cereals. Plant Physiol 106: 407-414

Oikawa K, Kasahara M, Kiyosue T, Kagawa T, Suetsugu N, Takahashi F, Kanegae T, Niwa Y, Kadota A, Wada M (2003) CHLOROPLAST UNUSUAL POSITIONING1 is essential for proper chloroplast positioning. Plant Cell 15: 2805-2815

Osborne CP, Freckleton RP (2009) Ecological selection pressures for C4 photosynthesis in the grasses. Proc Royal Soc B-Biol Sci 276: 1753-1760

Palmieri N, Schlotterer C (2009) Mapping Accuracy of Short Reads from Massively Parallel Sequencing and the Implications for Quantitative Expression Profiling. PLoS ONE 4: 10

Reiskind JB, Berg RH, Salvucci ME, Bowes G (1989) Immunogold Localization Of Primary Carboxylases In Leaves Of Aquatic And A C3-C4 Intermediate Species. Plant Sci 61: 43-52

Renne P, Dressen U, Hebbeker U, Hille D, Flugge UI, Westhoff P, Weber APM (2003) The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2. Plant J 35: 316-331

Roberts AG, Oparka KJ (2003) Plasmodesmata and the control of symplastic transport. Plant Cell Environ 26: 103-124

Rolland F, Baena-Gonzalez E, Sheen J (2006) Sugar sensing and signaling in plants: Conserved and novel mechanisms. Annu Rev Plant Biol 57: 675-709

Rossini L, Cribb L, Martin DJ, Langdale JA (2001) The maize Golden2 gene defines a novel class of transcriptional regulators in plants. Plant Cell 13: 1231-1244

Sage RF (2004) The evolution of C4 photosynthesis. New Phytol 161: 341-370
Sawers RJH, Liu P, Anufrikova K, Hwang JTG, Brutnell TP (2007) A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf. BMC Genomics 8:12

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nature Genet 37: 501-506

Sheen J (1999) C4 gene expression. Annu Rev Plant Physiol Plant Mol Biol 50: 187-217

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucl Acids Res 36: D1009-D1014

Taniguchi M, Taniguchi Y, Kawasaki M, Takeda S, Kato T, Sato S, Tahata S, Miyake H, Sugiyama T (2002) Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in Arabidopsis thaliana. Plant Cell Physiol 43: 706-717

Taniguchi Y, Nagasaki J, Kawasaki M, Miyake H, Sugiyama T, Taniguchi M (2004) Differentiation of dicarboxylate transporters in mesophyll and bundle sheath chloroplasts of maize. Plant Cell Physiol 45: 187-200

Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res 16: 934-946

Voznesenskaya EV, Edwards GE, Kiirats O, Artyusheva EG, Franceschi VR (2003) Development of biochemical specialization and organelle partitioning in the single-cell C4 system in leaves of Borszczowia aralocaspica (Chenopodiaceae). Am J Bot 90: 1669-1680

Voznesenskaya EV, Franceschi VR, Kiirats O, Artyusheva EG, Freitag H, Edwards GE (2002) Proof of C4 photosynthesis without Kranz anatomy in Bienertia cycloptera (Chenopodiaceae). Plant J 31: 649-662

Voznesenskaya EV, Franceschi VR, Kiirats O, Freitag H, Edwards GE (2001) Kranz anatomy is not essential for terrestrial C4 plant photosynthesis. Nature 414: 543-546

Voznesenskaya EV, Koteyeva NK, Chuong SDX, Ivanova AN, Barroca J, Craven LA, Edwards GE (2007) Physiological, anatomical and biochemical characterisation of photosynthetic types in genus Cleome (Cleomaceae). Func Plant Biol 34: 247-267

Wall PK, Leebens-Mack J, Chandler AS, Barakat A, Wolcott E, Liang HY, Landherr L, Tomsho LP, Hu Y, Carlson JE, Ma H, Schuster SC, Soltis DE, Soltis
PS, Altman N, dePamphilis CW (2009) Comparison of next generation sequencing
technologies for transcriptome characterization. BMC Genomics 10

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics.
Nature Rev Genet 10: 57-63

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA (2009) GLK
Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in
Arabidopsis. Plant Cell 21: 1109-1128

Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the
Arabidopsis Transcriptome with Massively Parallel Pyrosequencing. Plant Physiol
144: 32-42

Weber APM, von Caemmerer S (2010) Plastid transport and metabolism of C_3 and C_4
plants - comparative analysis and possible biotechnological exploitation. Curr Opin
Plant Biol, in press

Wise RP, Moscou MJ, Bogdanove AJ, Whitham SA (2007) Transcript profiling in host-
pathogen interactions. Annual Rev Phytopathol 45: 329-369
Figure Legends

Figure 1. The qualitative patterns of transcript abundance between *C. gynandra* and *C. spinosa* are very similar with the same classes under- and over-represented in both libraries. (A) Analysis based on BLAT mapping (B) Analysis based on BLAST mapping

Black bars refer to the C4 *C. gynandra*, open bars to the C3 *C. spinosa*.

Figure 2. Massively parallel sequencing of mRNAs (RNA-Seq) and qPCR generate similar profiles of transcript abundance in *C. gynandra* and *C. spinosa*. Ratios of transcript abundance in *C. gynandra* and *C. spinosa* were calculated, and transcripts selected for this analysis spanned four orders of magnitude. *PPDK* = pyruvate orthophosphate dikinase, *NAD-ME* = NAD dependent Malic Enzyme, *CA* = carbonic anhydrase, *PPCK* = phosphoenolpyruvate carboxylase kinase, *LHCA* = light harvesting complex subunit A, *RbcS1a* = Ribulose Bisphosphate Carboxylase Oxygenase 1a, *RCA* = RuBisCO activase. Solid bars represent data from RNA-Seq, and open bars data from qPCR. The horizontal dashed line represents a ratio of one and indicates no difference in transcript abundance between the two species.

Figure 3. The quantitative patterns of transcript accumulation in *C. gynandra* and *C. spinosa* are distinct. (A) Analysis based on BLAT mapping (B) Analysis based on BLAST mapping

The percentage of genes with significantly higher abundance of transcripts in C4 (red bars), unchanged (open bars, included genes not detected) or significantly lower abundance of transcripts in C4 (blue bars) based on the total number of genes in each annotation class (in parentheses on the y axes).
**Figure 4.** Schematic of components associated with the C₄ cycle in the NAD-ME sub-type based on interpretation of RNA-Seq. Proteins that have been described previously are in grey, and novel proteins are marked in red. Metabolites are in black. PIP1B:CA4 = PIP1B plasma membrane aquaporin:membrane tethered carbonic anhydrase; OAA = oxaloacetic acid; Asp = aspartate; AspAT = aspartate aminotransferase; NAD-ME = NAD-dependent malic enzyme; AlaAT = alanine aminotransferase; Ala = alanine; ACT11-CHUP1 = ACTIN11-CHUP1 complex; NHD1 = putative sodium:proton antiporter; PPDK = pyruvate, orthophosphate dikinase; Pyr = pyruvate; OEP24, chloroplast outer envelope protein 24; PEP = phosphoenolpyruvate; PPT = phosphoenolpyruvate transporter.
Tables

Table 1. Massively parallel signature sequencing allows large-scale assembly of transcripts in both *C. spinosa* and *C. gynandra* after comparison with the TAIR8 *Arabidopsis* database. One GS FLX sequencing run allowed significant generation of sequence for both species and the vast majority of these could be used to assemble contigs, and then matched to *Arabidopsis* genes.

|                  | *C. spinosa* | *C. gynandra* |
|------------------|--------------|---------------|
| Raw reads        | 313,807      | 402,674       |
| Raw nt           | 70,564,592   | 91,851,136    |
| Raw mean length  | 225          | 228           |
| Clean reads      | 284,318      | 368,333       |
| Clean nt         | 65,525,139   | 85,681,233    |
| Clean mean length| 230          | 232           |
| Contigs          | 17,655       | 18,992        |
| Total length (nt)| 7,746,894    | 9,062,043     |
| Total reads      | 245,324      | 319,732       |
| % assembled      | 86.3         | 86.8          |

Table 2. Mapping the sequence reads with different BLAT and BLAST parameters to empirically determine suitable mapping conditions. The percentage of AGIs with at least one mapped read and the average mappings per read are determined prior to parsing the tables to retain only the best match.
| Mapping program | Library      | search space | cut off value* | % reads with at least one hit in the reference | % AGIs with at least one mapped read | Average mappings per read |
|-----------------|--------------|--------------|---------------|-----------------------------------------------|--------------------------------------|--------------------------|
| BLAT            | C. gynandra  | DNA          | 60            | 40.9                                          | 42.0                                 | 1.19                     |
|                 |              |              | 75            | 40.7                                          | 41.7                                 | 1.19                     |
|                 |              |              | 85            | 30.2                                          | 35.8                                 | 1.15                     |
|                 |              |              | 90            | 7.7                                           | 19.5                                 | 1.09                     |
|                 |              | protein      | 25            | 82.6                                          | 70.4                                 | 2.33                     |
|                 |              |              | 50            | 82.6                                          | 70.4                                 | 2.33                     |
|                 |              |              | 75            | 75.4                                          | 62.6                                 | **1.48**                 |
|                 |              |              | 80            | 56.4                                          | 52.2                                 | 1.27                     |
| BLAST           | C. spinosa   | DNA          | 60            | 40.8                                          | 38.9                                 | 1.29                     |
|                 |              |              | 75            | 40.6                                          | 38.5                                 | 1.28                     |
|                 |              |              | 85            | 29.7                                          | 32.3                                 | 1.21                     |
|                 |              |              | 90            | 8.5                                           | 17.1                                 | 1.15                     |
|                 |              | protein      | 25            | 83.0                                          | 67.7                                 | 2.49                     |
|                 |              |              | 50            | 83.0                                          | 67.7                                 | 2.46                     |
|                 |              |              | 75            | **76.0**                                      | **58.9**                             | **1.57**                 |
|                 |              |              | 80            | 57.9                                          | 48.4                                 | 1.32                     |
|                 |              | DNA          | 1e-05:        | 68.9                                          | 56.5                                 | 30.7                     |
|                 |              |              | 1e-10:        | 58.8                                          | 49.1                                 | 27.7                     |
|                 |              |              | 1e-30:        | 29.6                                          | 30.5                                 | 18.9                     |
|                 |              |              | 1e-50:        | 9.9                                           | 15.9                                 | 11.5                     |
|                 |              | protein      | **1e-05:**    | **78.0**                                      | **76.9**                             | **106.6**                |
|                 |              |              | 1e-10:        | 67.8                                          | 71.0                                 | 64.6                     |
|                 |              |              | 1e-30:        | 29.0                                          | 39.5                                 | 22.9                     |
|                 |              |              | 1e-50:        | 0.1                                           | 0.3                                  | 7.5                      |
| BLAST           | C. gynandra  | DNA          | 1e-05:        | 69.7                                          | 53.0                                 | 28.2                     |
|                 |              |              | 1e-10:        | 59.6                                          | 46.3                                 | 25.1                     |
|                 |              |              | 1e-30:        | 29.4                                          | 28.3                                 | 16.2                     |
|                 |              |              | 1e-50:        | 9.8                                           | 14.4                                 | 9.8                      |
|                 | C. spinosa   | protein      | **1e-05:**    | **78.8**                                      | **75.3**                             | **93.7**                 |
|                 |              |              | 1e-10:        | 68.3                                          | 68.7                                 | 56.4                     |
|                 |              |              | 1e-30:        | 29.3                                          | 36.0                                 | 21.2                     |
|                 |              |              | 1e-50:        | 0.1                                           | 0.3                                  | 4.6                      |
**Table 3.** Transcript abundance of C₄ cycle genes that have significantly higher transcript abundance in C₄ leaf tissue

| Enzyme  | Locus   | *C. gynandra* RPM | *C. spinosa* RPM | fold-change | *C. gynandra* RPM | *C. spinosa* RPM | fold-change |
|---------|---------|-------------------|------------------|-------------|-------------------|------------------|-------------|
| AspAT   | AT2G30970 | 4806              | 14               | 343.3       | 4601              | 18               | 257.9       |
| PPDK    | AT4G15530 | 3262              | 14               | 233.0       | 3216              | 13               | 240.3       |
| PEPC    | AT2G42600 | 9702              | 124              | 78.2        | 8321              | 169              | 49.1        |
| AlaAT   | AT1G17290 | 7610              | 267              | 28.5        | 7242              | 259              | 28.0        |
| NAD-ME  | AT4G00570 | 1357              | 51               | 26.6        | 1326              | 49               | 27.0        |
| NAD-ME  | AT2G13560 | 1800              | 87               | 20.7        | 1723              | 85               | 20.3        |
| PEPC-K  | AT1G08650 | 230               | 37               | 6.2         | 226               | 36               | 6.3         |
| NADP-ME*| AT1G79750 | 227               | 60               | 3.8         | 216               | 45               | 4.8         |
| PEPC*   | AT1G53310 | 94                | 248              | 0.4         | 950               | 192              | 5.0         |
| PPDK-RP*| AT4g21210 | 148               | 32               | 4.6         | 198               | 27               | 7.3         |

*denotes changes significant only in BLAST mapping
Table 4. Transcript abundance of selected genes with an expression similar to that of C₄ cycle genes and RuBisCO, respectively. All changes are significant at p ≤ 0.01.

| Locus   | Annotation (TAIR9)                                                                 | C. gynandra (reads per million) | C. spinosa (reads per million) | ratio |
|---------|-----------------------------------------------------------------------------------|---------------------------------|--------------------------------|-------|
| **transport proteins** |                                                                                   |                                 |                                |       |
| AT2G26900 | bile acid:sodium symporter family protein                                         | 4774                            | 55                             | 86.8  |
| AT2G22500 | mitochondrial dicarboxylate carrier                                                | 324                             | 0                              | n/a   |
| AT4G24570 | mitochondrial dicarboxylate carrier                                                | 148                             | 0                              | n/a   |
| AT2G45960 | plasma membrane intrinsic protein subfamily protein                               | 2686                            | 133                            | 20.2  |
| AT5G33320 | phosphoenolpyruvate/phosphate translocator                                        | 1955                            | 97                             | 20.2  |
| AT1G49810 | member of Na+/H+ antiporter family                                                | 1321                            | 83                             | 15.9  |
| **metabolism** |                                                                                   |                                 |                                |       |
| AT3G52720 | alpha carbonic anhydrase 1                                                        | 227                             | 152                            | 1.5   |
| AT1G23730 | beta carbonic anhydrase 4                                                          | 497                             | 87                             | 5.7   |
| AT5G35170 | adenylate kinase family protein                                                    | 1994                            | 235                            | 8.5   |
| AT5G09650 | inorganic pyrophosphatase                                                         | 2664                            | 833                            | 3.2   |
| **proteins of unknown function** |                                                                                  |                                 |                                |       |
| AT1G12090 | extensin-like protein (ELP)                                                        | 6278                            | 147                            | 42.7  |
| **callose degrading enzymes** |                                                                                  |                                 |                                |       |
| AT3G57240 | member of glycosyl hydrolase family 17, likely beta 1,3 glucanase                  | 436                             | 0                              | n/a   |
| AT1G32860 | member of glycosyl hydrolase family 17, likely beta 1,3 glucanase                  | 50                              | 0                              | n/a   |
| AT5G42100 | plasmodesmal associated beta-1,3-glucanase                                         | 173                             | 32                             | 5.4   |
| **cell biology** |                                                                                  |                                 |                                |       |
| AT3G25690 | CHUP1                                                                             | 22                              | 170                            | 0.13  |
| AT3G12110 | actin                                                                              | 122                             | 727                            | 0.2   |
Table 5. Transcription factors which are significantly changed between the leaf tissue samples

| Locus       | transcription factor type | BLAT mapping | BLAST mapping | segmentally duplicated? |
|-------------|---------------------------|--------------|---------------|-------------------------|
|             |                           | C. gynandra RPM | C. spinosa RPM | ratio | C. gynandra RPM | C. spinosa RPM | ratio |
| AT1G25560   | AP2-EREBP                 | 176          | 9             | 19.6     | 219          | 9             | 24.3     | yes |
| AT5G07580   | AP2-EREBP                 | 223          | 51            | 4.4      | 292          | 36            | 8.1      | yes |
| AT1G53910   | AP2-EREBP*                | 32           | 138           | 0.2      | 84           | 268           | 0.3      | yes |
| AT5G10570   | bHLH                      | 0            | 83            | n/a      | 0            | 112           | n/a      | yes |
| AT3G21330   | bHLH*                     | 0            | 74            | n/a      | 0            | 107           | n/a      | |
| AT3G62420   | bZIP                      | 11           | 138           | 0.1      | 10           | 138           | 0.1      | |
| AT2G20570   | G2-like                   | 220          | 0             | n/a      | 292          | 0             | n/a      | |
| AT1G72030   | GNAT                      | 11           | 179           | 0.1      | 10           | 330           | 0.0      | |
| AT2G22430   | HB                        | 515          | 106           | 4.9      | 505          | 116           | 4.4      | yes |
| AT1G10200   | LIM                       | 22           | 230           | 0.1      | 21           | 205           | 0.1      | |
| AT4G30410   | not specified*            | 0            | 32            | n/a      | 0            | 76            | n/a      | |
| AT1G32700   | PLATZ                     | 176          | 9             | 19.6     | 115          | 4             | 28.8     | |
| AT5G02810   | PseudoARR-B               | 0            | 106           | n/a      | 10           | 112           | 0.1      | |
| AT2G36990   | Sigma70-like              | 130          | 0             | n/a      | 143          | 0             | n/a      | |
| AT1G48500   | Tify                      | 11           | 147           | 0.1      | 10           | 174           | 0.1      | yes |
| AT1G17380   | Tify*                     | 18           | 110           | 0.2      | 24           | 161           | 0.1      | yes |
| AT3G02790   | zinc finger               | 374          | 87            | 4.3      | 407          | 112           | 3.6      | yes |

*denotes changes significant only in BLAST mapping

Table 6. Comparison of alterations in transcript abundance in C₄ and C₃ leaves with those induced by cold, sugar-feeding, attack by pests or pathogens, diurnal changes to light or circadian rhythms.

|                            | Estimated change in transcriptome | Percent change (%) | Reference                      |
|---------------------------|-----------------------------------|--------------------|--------------------------------|
| Cold treatment            | 514 (24 000) ATH1                 | 2.1%               | Vogel et al., (2005)           |
| C₄ leaves and C₃ leaves   | 583/603 (13 443/13662) ATH1       | 2.7/2.8%           | this study                     |
| Glucose feeding           | 978 (22 500) ATH1                 | 4.4%               | Price et al., (2001)           |
| *Pseudomonas syringae*    | 2 034 (23 750) ATH1               | 8.6%               | De Vos et al., (2005)          |
| *Mysus persicae*          | 2 181 (23 750) ATH1               | 9.1%               | De Vos et al., (2005)          |
| Diurnal regulation        | 1 115 (11 521) cDNA array         | 11%                | Schaffer et al., (2001)        |
| Circadian regulation      | 2 282 (18 890) ATH1               | 12%                | Dodd et al., (2007)            |
Ratio of transcript abundance
(C. gynandra/C. spinosa)
CO$_2$ → HCO$_3^-$ → OAA → Asp → OAA → Asp

PEPC → PEP → Pyr → Ala → NAD-ME → Pyr + CO$_2$ → OAA → Asp

PIP1B:CA4

PEP → Pyr → Ala

PPDK

NHD1

OEP24

ACT11-CHUP → PCR cycle → CO$_2$