In planta inactivation of the C_4 phospho*enol*pyruvate carboxylase (PEPC) gene of *Flaveria bidentis* and evolutionary analyses of the PEPC protein kinase (PPCK) gene family of *Flaveria*

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Sophia Hilary Aldous
aus Berlin

Düsseldorf, Mai 2014

aus dem Institut für Entwicklungs- und Molekularbiologie der Pflanzen
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Peter Westhoff

Korreferent: Prof. Dr. Georg Groth

Tag der mündlichen Prüfung:

Hiermit versichere ich, dass ich die vorliegende Dissertation eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation habe ich in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, 13. Mai 2014

(Sophia Hilary Aldous)

Contents

VI. Chapters

ABBREVIATIONS

A	Net CO_2 uptake (assimilation)
Ala	Alanine
amiRNA	artificial micro RNA
ATP	Adenine triphosphate
A. thaliana	Arabidopsis
bp	Base pairs
C_I	CO_2 concentration, intercellular
°C	Degree Celsius
CA	Carbonic anhydrase
CDPK	Calcium-dependent protein kinases
CO_2	Carbon dioxide
CRKs	CDPK-related kinases
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
E	Estradiol
E. coli	*Escherichia coli*
F.	*Flaveria*
Ft, Fb, Fp	*Flaveria trinervia*, *Flaveria bidentis, Flaveria pringlei*
Glc-6-P	Glucose-6 phosphate
GFP	Green fluorescent protein
GUS	Glucuronidase
h	Hour (s)
HCO_3^-	Bicarbonate (hydrogen carbonate)
$I_{0.5}$	Inhibition of 50 % enzymatic activity
Ins(1,4,5)P3	Inositol-1,4,5-trisphosphate
K_a	Association constants
K_{cat}	catalytic activity
K_i	Inhibitor constants
K_m	Michaelis constants
K.	*Kalanchoë*
K	Kanamycin
kDa	Kilo Dalton
kb	Kilobases
LB	Media from Luria-Bertani
LD	Light/dark
MES	2 (N-morpholino)-Ethansulfonacid
MDH	Malate dehydrogenase
ME	Malic enzyme
mRNA	Messenger ribonucleic acid
NAD	Nicotinamide adenine dinucleotide

NADP	Nicotinamide adenine dinucleotide phosphate
N.	*Nicotiana*
nt	Nucleotides
O_2	Molecular oxygen
OAA	Oxaloacetate
PCR	Polymerase chain reaction
PEP	Phospho*enol*pyruvate
PEPC	Phospho*enol*pyruvate carboxylase
PEPC-PK	Phospho*enol*pyruvate carboxylase protein kinase
PI-PLC	Phosphatidylinositol-specific phospholipase
PPCK	Phospho*enol*pyruvate carboxylase protein kinase gene
PEPRK	Phospho*enol*pyruvate carboxylase kinase-related kinases
3-PGA	3- Phosphoglycerate
PPDK	Pyruvate, orthophosphate dikinase
PCR	Polymerasenkettenreaktion
P_i	inorganic phosphate
RNA	Ribonucleic acid
RNAse	Ribonuclease
rpm	Rotation per minute
RT	Reverse transcriptase
Rubisco	Ribulose 1,5-biphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-biphosphate
Ser	Serine
SNF	Sucrose non-fermenting-1
SnRK	SNF1-related kinase
T-DNA	Transfer DNA
V_{max}	Maximal velocity (µmol/Min)
WT	wild-type

I. Introduction

1. Evolution of C_4 photosynthesis

1.1. The dual-cell system of the C_4 carbon cycle requires the activity of key enzymes

The photosynthetic carbon pathway is a Ribulose 1,5-biphosphate carboxylase/oxygenase (Rubisco)-based net fixation of atmospheric CO_2 to convert carbon into organic compounds using solar energy. The C_3 photosynthetic pathway represents the ancestral condition operative in a single-cell system, where atmospheric CO_2 is fixed directly by Rubisco of the Calvin-Benson cycle in the mesophyll cells. C_4 photosynthesis evolved by division of labor between two cell types from an ancestor C_3 condition. The C_4 photosynthetic carbon pathway is spatially separated by the compartmentation of mesophyll and bundle-sheath cells of the leaf, first described by Haberlandt (Haberlandt, 1914). In this anatomical arrangement typical for C_4 plants, known as the Kranz anatomy, the mesophyll cells are arranged around large bundle-sheath cells (Hodge et al., 1955; Edwards et al., 2004). C_4 photosynthesis is based on the coordination between these two distinct biochemically specialized and chloroplast-containing cells (Slack and Hatch, 1967). The typical Kranz anatomy of genera like *Atriplex*, *Amaranthus*, *Eleocharis*, *Flaveria*, *Sorghum* and *Zea* is however not present among each C_4 plant, such as the aquatic plant *Hydrilla verticillata* and species of the Chenopodiaceae (Holaday and Bowes, 1980; Freitag and Stichler, 2002).

In the NADP-malic enzyme (ME)-type pathway of C_4 photosynthesis atmospheric CO_2 enters at the periphery of the mesophyll cells, and is converted to bicarbonate (HCO_3^-) and protons by the carbonic anhydrase (CA) (Hatch and Slack, 1970; Hatch, 1987) and [**Figure 1**]. The mesophyll-expressed phospho*enol*pyruvate carboxylase (PEPC) fixes atmospheric CO_2 in form of HCO_3^- in the cytosol to catalyze phospho*enol*pyruvate (PEP) to oxaloacetate (OAA) and phosphoric acid (H_3PO_4) (Uedan and Sugiyama, 1976; Wedding et al., 1988; Chollet et al., 1996). Oxaloacetate is then reduced by the NADP-malate dehydrogenase (NADP-MDH) enzyme to the four-carbon organic acid malate (Edwards and Walker, 1983; Trevanion et al., 1997). In some C_4 species oxaloacetate is converted by transamination by the aspartate aminotransferase (Asp-AT) into aspartate (Edwards and Walker, 1983). Malate, carrying the reducing equivalent and CO_2 diffuses into the inner bundle-sheath-like compartment (Sage et al., 2012). A high concentration of malate in the mesophyll cells creates a gradient along the two cells and thereby a flux of carbon compounds in the C_4 cycle (Häusler et al., 2002). Malate is decarboxylated by the NADP-malic enzyme (NADP-ME) into the three-carbon acid pyruvate and reducing equivalents are released (Ivanov et al., 2005). In the bundle-sheath cells the CO_2 concentration is build up and the net carbon acquisition catalyzed by the carboxylase activity of Rubisco in the Calvin-Benson cycle. Ribulose 1,5-biphosphate (RuBP) is converted to 3-phosphoglycerate (3-PGA), which is further reduced to the precursor of sucrose (triose phosphates). The Calvin-Benson cycle requires energy from NADPH and ATP synthesized by the photosynthetic light-driven reactions. NADPH is provided by 50% from the NADP-ME decarboxylating enzyme (Chapman et al., 1980). Assimilated carbon is exported from the bundle-sheath cells and transformed to sucrose which is translocated through the surrounding vascular

bundles to the sink tissues. To complete the photosynthetic pathway pyruvate (or alanine) diffuses back to the mesophyll cells. Here, the enzyme pyruvate orthophosphate dikinase (PPDK) uses ATP and P_i to rephosphorylate pyruvate to PEP. PEP as the primary carbon acceptor provides new substrate to reinitiate the C_4 cycle. A diffusion barrier of suberin lamella integrated into the bundle-sheath outer cell wall of C_4 grasses was suggested to stabilise the effective reduction of CO_2 effluxes (Dengler and Nelson, 1999).

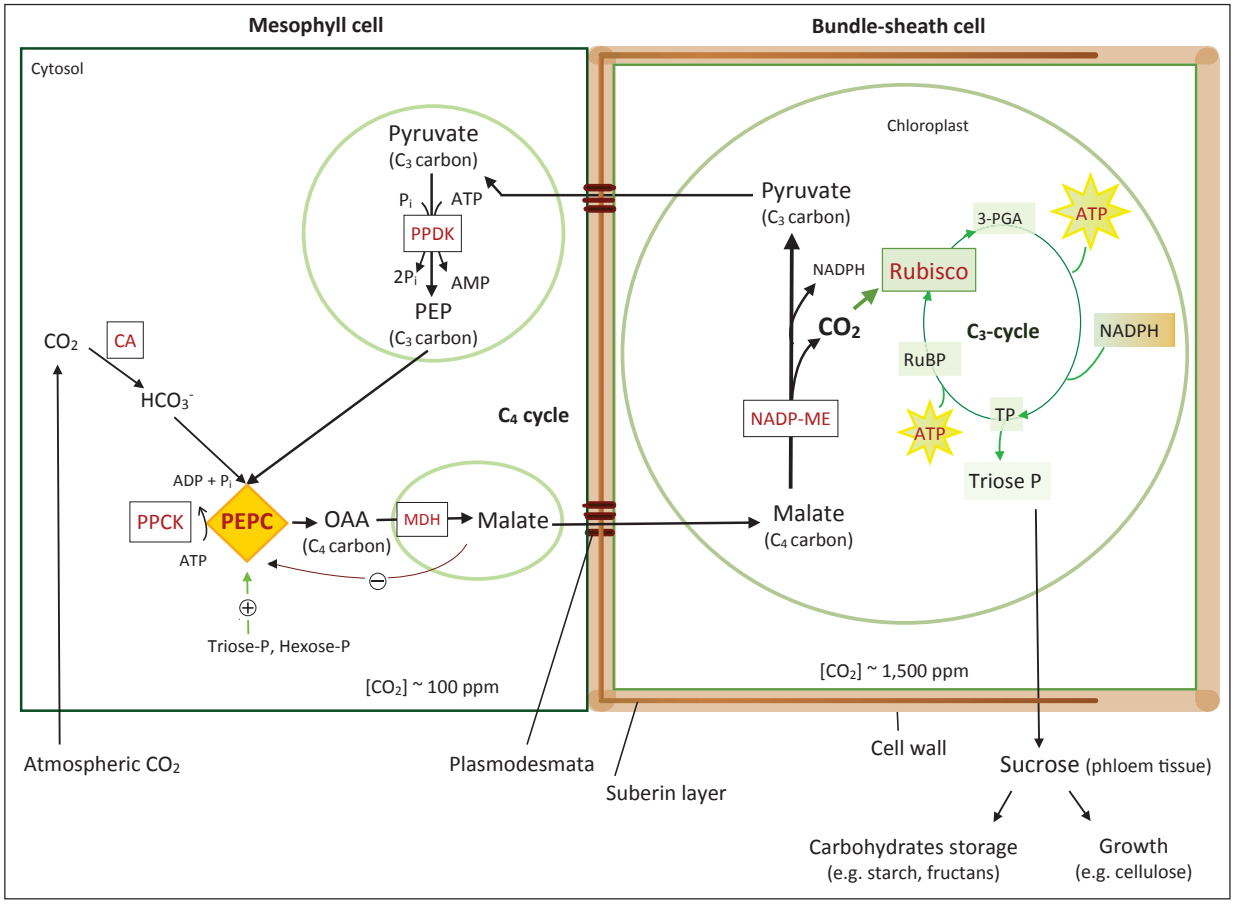


Figure 1: Schematic representation of the main reactions of the photosynthetic carbon cycle in C_4 species of the NADP-ME subtype. Operation of the photosynthetic carbon pathway in C_4 species involves biochemical reactions of key enzymes between mesophyll and bundle-sheath cells. The carbonic anhydrase (CA) located in the mesophyll cells fixes atmospheric CO_2. CA converts CO_2 into HCO_3^-. Phospho*enol*pyruvate carboxylase (PEPC) uses HCO_3^- to catalyze phospho*enol*pyruvate (PEP) to yield oxaloacetate (OAA). The enzyme PEPC is regulated by the phospho*enol*pyruvate carboxylase protein kinase (PPCK) and allosterically regulated by metabolites. OAA is converted to Malate by the NADP-Malate dehydrogenase (NADH-MDH) in the mesophyll cells. Malate diffuses into the bundle-sheath cells, where it becomes decarboxylated during the turn-over into pyruvate by the NADH-malate enzyme (NADH-ME). A high number of plasodesmata allow for the flux of metabolites through the close adjusted cells (van Bel, 1993). The CO_2 is carboxylated by the Ribulose 1,5-biphosphate carboxylase/oxygenase (Rubisco). Additional abbreviations: RuBP, Ribulose 1,5-bisphosphate; 3-PGA, 3-phosphoglyceric acid; P, phosphate. Adapted from Sage et al., 2012.

1.2. The CO_2 concentrating mechanism compensates for photorespiration

Around 33.7 to 25 million years ago during the Oligocene the global partial pressure of the atmospheric carbon dioxide (pCO_2) decreased gradually (Zachos et al., 2001; Pagani et al., 2005; Christin et al., 2008; Beerling and Royer, 2011). The pCO_2 reached a content below 300 parts per million (p.p.m.) by the Pleistocene, around 12 to 3 million years ago (Gerhart and Ward, 2010). In comparison the current atmospheric concentration is around 0,039% or 392 p.p.m. (Sage, 2013). In early times, the gradual decline of atmospheric CO_2 operating in concert with additional ecological selective forces caused plants to adapt to the effects of these climate changes (Monson et al., 1988; Sage et al., 1999; Osborne and Freckleton, 2009; Edwards and Smith, 2010; Christin et al., 2011). This adaptation in turn caused the evolution of natural variations among photosynthetic modes, while contributing to the bio-diversification on earth (Sage et al., 2012).

A unique feature of Rubisco is the ability to bind CO_2 or O_2 as alternative substrate at the catalytic active site, in order to catalyze either the carboxylation or the oxygenation of RbBP. The photosynthetic CO_2 fixation is therefore a competitive reaction to the activity of Rubisco to oxygenate RuBP. Because of the affinity of Rubisco to O_2 being less than to CO_2 at optimal temperatures, the ratio of O_2 to CO_2 has to reach a threshold for oxygenation to occur (Andrews and Lorimer, 1987). Oxygenation of RuBP generates 2-phosphoglycolate (2-PG) and 3-PGA and directs a process common in C_3 plants, namely the photorespiration (Bauwe et al., 2010). Although some of the photosynthetic carbon in 2-PG is recycled in a three organelle complex and 3-PGA is incorporated back to the Calvin-Benson cycle, this pathway while consuming light energy causes the loss of carbon molecules, that reduces the rate and efficiency of CO_2 assimilation (Sharkey, 1988; Ehleringer et al., 1991; Cegelski and Schaefer, 2006).

The photorespiratory pathway is promoted at low CO_2 concentrations, at temperatures above 25°C and under drought stresses. High rates of photorespiration were proposed as the primary condition to drive the C_4 biological innovation (Hatch, 1971; Björkman, 1973; Sage et al., 2012). The challenge was to minimise photorespiration (oxygenation) relative to photosynthesis (carboxylation), by increasing the internal CO_2 concentration above both the atmospheric CO_2 and the internal O_2 levels. Like this, the photosynthetic rates particularly in adaptation to high light/temperatures and salinity were saturated (Sage et al., 2012). C_4 photosynthesis has an increased photosynthetic efficiency with respect to C_3 plants. C_4 photosynthesis is based on a mechanism without either affecting the water-use-efficiency (WUE) by stomata opening, or lowering the leaf temperature by solar radiation avoidance. The WUE is higher in C_4 in than in C_3 plants, due to increased photosynthetic rates per unit leaf area and higher stomatal closure (Ghannoum et al., 2011; Sage et al., 2012). In addition, C_4 plants evolved without promoting the cost of nitrogen investment for the synthesis of the enzymes CA and Rubisco, and thereby not affected in the nitrogen-use-efficiency (NUE) (Farquhar and Sharkey 1982; Sage et al., 1987; Seemann et al., 1987; Evans and von Caemmerer, 1996; Huxman and Monson, 2003).

Concomitantly, the intercellular CO_2 diffusion distances and conductance had to become low in the C_4 cycle (Von Caemmerer and Furbank, 2003; Edwards et al., 2004). C_4 species have overcome the effects of photorespiration by restructuring the leaf anatomy and establishing a CO_2 pump that builds up the CO_2 concentration at the vicinity of Rubisco. This CO_2 concentrating mechanisms increases the carboxylation activity of Rubisco until saturated even at low CO_2 atmospheric conditions (Hatch and Osmond, 1976;

Edwards and Walker, 1983; Ehleringer and Monson, 1993; Sage et al., 2012). Despite the high productivity of C_4 photosynthesis, the expense to concentrate CO_2 is two molecules of ATP per CO_2 molecule absorbed (in the NADP-ME and the NAD-ME type for the PPDK activity) (Osborne and Beerling, 2006). This extra investment is reflected in the reduction of the assimilation of CO_2 in function to a very high intercellular pCO_2 in C_3 with respect to C_4 species (Ehleringer and Börkman, 1977).

1.3. Ecological forces cause the taxonomic expansion of C_4 plants

As a consequence of the pre-adaptation of C_4 species to climate factors, the eco-geographical distribution of C_4 species over several continents was indicated to be a result of high temperatures, drought and salinity (Ehleringer et al., 1997; Ehleringer, 2005). It is suggested that reduced precipitation and grazing are factors causing the change of a C_3 vegetation into grassland (Hattersley, 1983; Edwards and Smith, 2010). In turn the development of grassland reservation often results in aridity, that as such can promote a further expansion of C_4 communities which adaptation rate is faster than for C_3 lineages (Knapp and Medina 1999; Sage, 2001; Osborne and Freckleton, 2009; Edwards and Smith, 2010). There are evidences for the fact that a large taxonomic radiation of C_4 plants occurred in the late Miocene (Cerling et al., 1997) and again in the last glacial maximum (LGM) (Boutton et al., 1994; Street-Perrott et al., 1994).

Globally, C_3 photosynthesis is the most common photosynthetic pathway executed among bushes, herbs, cyanobacteria, algae and the majority of vascular plants. C_3 plants inhabit arctic or temperate regions, whereas C_4 communities are predominantly occupying open ecosystems of semiarid tropics and subtropical grassland biomers of low latitudes, arid and nutrient-poor soils or even hot deserts (Cerling et al., 1993; Ehleringer et al., 1997; Sage, 2004; Edwards et al., 2010; Edwards and Smith, 2010; Sage et al., 1999; Sage et al., 2012).

Despite of the high efficiency of C_4 photosynthesis, nowadays C_4 species (predominantly *Poaceae*) are suggested to contribute to not more than 30 % to the global biomass production on Earth (Lloyd and Farquhar, 1994; Osborne and Beerling, 2006). Both C_3 and C_4 plants are relevant crop plants. For a higher productivity of biomass as in the crop plants (maize, *Sorghum* and sugar cane) a C_4 pathway is desirable. This explains the broader effort to genetically engineer C_3 crops with C_4 traits (Brown, 1999; Matsuoka et al., 2001; Jeanneau et al., 2002a/b; Sage and Sage, 2009; McAllister et al., 2012).

1.4. Evolution of C_4 photosynthesis with a polyphyletic origin

The evolutionary relationship between C_3 and C_4 plants was resolved on comparative analyses regarding different data sets of the morphology, life history, nuclear-/plastid-encoded DNA sequences, the eco-geographical distribution and the determination of the carbon isotope ($\delta^{13}C$) composition (O´Leary, 1981; Kadereit et al., 2003; Mckown and Dengler, 2007; Christin et al., 2011). Using these approaches, an independent origin of phylogenetically quite distant lineages of C_4 angiosperms was estimated that originated several times during the evolution (Christin et al., 2011; Sage et al., 2012). C_4 photosynthesis is a relative common phenomenon in dicots (e.g. Asteraceae, Amaranthaceae and Chenopodiaceae)

(Sage, 2004; Kadereit et al., 2003; Christin et al., 2011; Sage et al., 2011). Monocots with C_4 photosynthesis are (the Poaceae, Cyperaceae and Hydrocharitaceae). In the Chenopodiaceae the first occurrence of C_4 photosynthesis was estimated to 11.5 to 7.9 million years ago, and in the Amaranthaceae 21.6 to 14.5 million years ago (Kadereit et al., 2003). Within many of the families, the C_4 pathway evolved multiple times (Sage, 2004). The *Poaceae* are monocots where C_4 photosynthesis predominates with at least seventeen independent origins over the last 30 million years (Edwards and Smith, 2010; Christin et al., 2011; Gowik and Westhoff, 2011). The C_4 origin within the dicot *Flaveria* genus appeared independently and was estimated to a recent epoch corresponding to the beginning of the Pleistocene (around 2 to 3 million years ago) (Kubien et al., 2008; Christin et al., 2011).

1.5. Photosynthetic evolution in the genus Flaveria

Flaveria belongs to the Asteraceae (Powell, 1978: Mckown et al., 2005; Kapralov et al., 2011). The geographical origin of *Flaveria* and from where it was widely distributed is recognized as the south-central region of Mexico (Mckown et al., 2005). With the radiation of individuals the apparently genetic diversity increased within the genus giving rise to photosynthetic diverse but still closely related species (Powell, 1978; Mckown et al., 2005). Examples for representative *Flaveria* species with C_4 photosynthesis are *F. trinervia* and *F. bidentis*. A *Flaveria* species with C_4-like photosynthesis is *F. brownii* and with C_3-C_4 intermediate photosynthesis are *F. ramosissima* and *F. pubescens, F. anomala* and *F. chloreapholia. F. pringlei* and *F. robusta* are representative species with C_3 photosynthesis. Therefore, *Flaveria* represents the principal model genus in the extensive research of C_4-associated anatomical, biochemical and molecular alterations.

A phylogram of the genus *Flaveria* represents two clades A and B where each hold a different number of phyllary lines. Intermediates evolved at least twice and are represented in both clades (Edwards and Ku, 1987; Kopriva et al., 1996; Mckown et al., 2005). At the base of the phylogram C_3 species represent the ancestral condition, while the C_3-C_4, C_4-like and C_4 species are phylogenetically placed at the more recent evolutionary branch (McKown, 2005).

1.6. C_4 evolution is based on minor genetic alterations of the ancestral condition

C_4 photosynthesis evolved independently within *Flaveria* based on a pre-existing condition in a C_3 ancestry(Apel and Maass, 1981; McKown, 2005; Osborne and Beerling, 2006). This C_3 ancestor is likely to resemble the contemporaneous C_3 *Flaveria* species. The C_4 photosynthetic pathway is an end product of combined C_4-traits (known as the "C_4 syndrome") acquired along a developmental process in an evolutionary transition from C_3 to C_4. Several studies formulate/support the fact, that this evolutionary trajectory in *Flaveria* compasses a step-wise order of photosynthetic transitions until establishing a fully developed C_4 cycle [**Figure 2**]. C_3-C_4 intermediates represent phylogenetically speaking true transitory species that demonstrate that C_4-traits in *Flaveria* were acquired step-wise in a gradual transition (Apel and Maass, 1981; Mckown et al., 2005; Mckown and Dengler, 2007; Heckmann et al., 2013). C_4 traits are accompanied with a gain-of-fitness in each evolutionary step (Westhoff and Gowik, 2004; Sage,

2004; Gowik and Westhoff, 2011), and each gain-of-fitness was comparable with each other along the evolutionary trajectory (Heckmann et al., 2013).

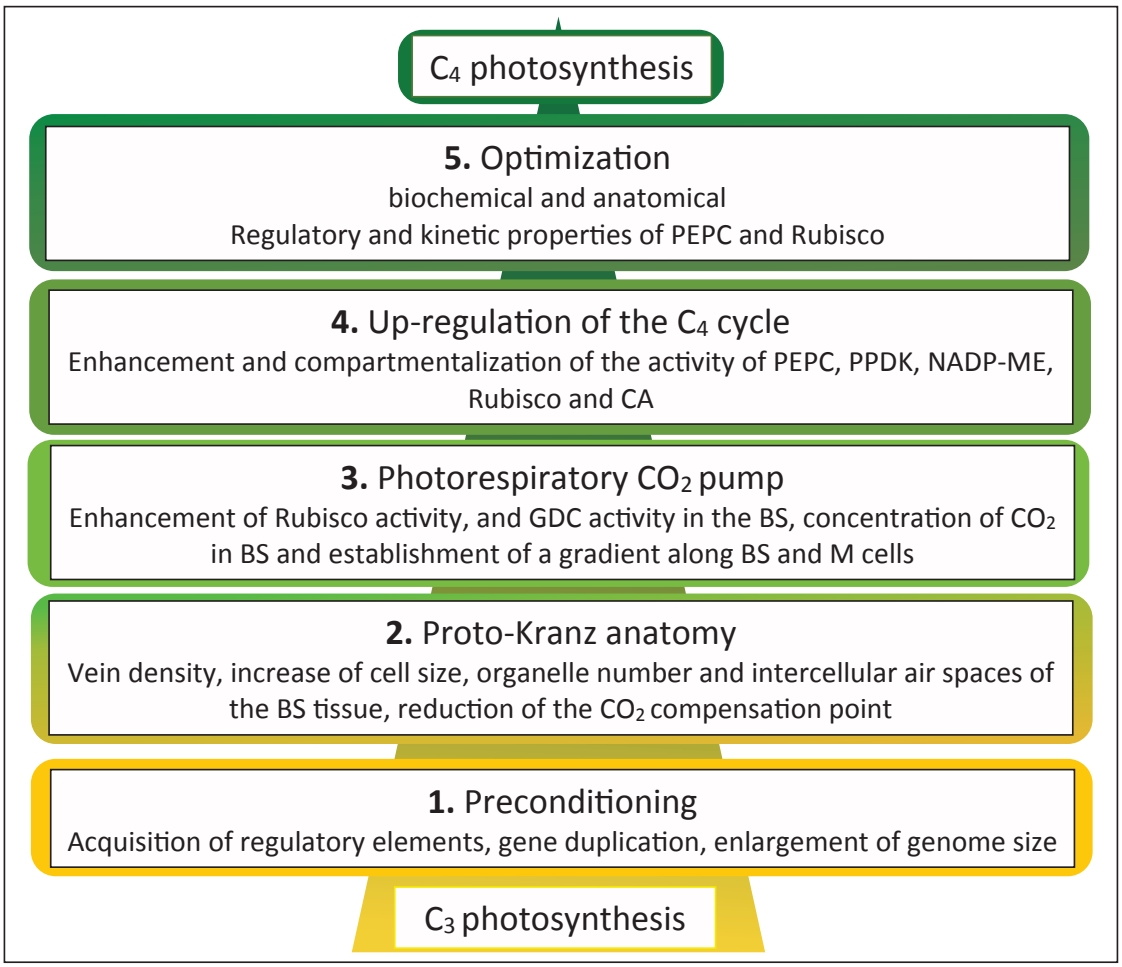


Figure 2: Flowchart of the model of the evolutionary step-wise acquisition of C_4 traits. Indicated are the five critical phases required for the incipient C_4 photosynthesis in *Flaveria*. Abbreviations: BS, Bundle-sheath cells; M, mesophyll cells; GDC, Glycine decarboxylase complex; PEPC, Phospho*enol*pyruvate carboxylase; PPDK, pyruvate, orthophosphate dikinase; NADH-ME, NADH malate enzyme; CA, carbonic anhydrase. Adapted from Sage, 2004; Sage et al., 2012.

The conceptual model of C_4 evolution describes a series of genomic, anatomical and biochemical modifications. Phase 1: Low humidity accelerated the formation of lower vein spacing, required to minimize the distances of metabolite transport and CO_2 conductance between the two cells (Sinha and Kellogg, 1996; Mckown and Dengler, 2007/09). The genome size enlarged and gene duplications occurred from an increased number of chromosomes by autopolyploidization. This precondition allowed for the neo- and sub-functionalization of non-photosynthetic progenitor genes to establish C_4-traits (Rondeau et al., 2005; Christin et al., 2012). As a result most genes of the C_4 cycle are members of small gene families. In terms of molecular alterations, none of the C_4 enzymes is unique to C_4 plants and the isogenes coding for enzymes of the C_4 cycle are already present in C_3 plants (Hattersley, 1987). Therefore, it can be assumed that despite of the complex arrangement of the C_4 cycle, genes evolved through the co-option of pre-existing genes present in the last common ancestor of C_3 and C_4 plants,

without the requirement of major genetic inventions (Sheen, 1999; Dengler and Nelson 1999; Sage, 2004; Aubry et al., 2011; Brown et al., 2011).

Phase 2: Anatomical modifications promoted the generation of a fully differentiated Kranz anatomy from a proto-Kranz, still present in some contemporaneous C_3 species (Muhaidat et al., 2011) as a base for the operative C_4 cycle (Nelson and Dengler, 1992; Mckown and Dengler, 2009). In the undifferentiated Kranz anatomy the bundle-sheath (BS) cells enlarged, the mesophyll cells (M) reduced in size and the organelle number increased (McKown and Dengler, 2007; Muhaidat et al., 2011; Sage et al., 2012). Large BS intercellular spaces required for gas exchange would allow for a higher efficiency of Rubisco to recover photorespired CO_2 (Sage et al., 2012).

Phase 3: The evolution of C_2 photosynthesis. The glycine decarboxylase (GDC) of the photorespiratory cycle decarboxylates glycine to serine. The expression and activity of functional GDC in the BS cells and the absence in the M cells results in a reduced rate of photorespiration (Wiludda et al., 2011).

Phase 4: Establishment of the C_4 metabolic cycle. The non-photosynthetic genes had to be reprogrammed, so that the genes of the C_4 cycle became highly expressed and the encoding enzymes recruited to either the mesophyll or the bundle-sheath cells (Sage, 2004). The C_4-type PEPC from *Flaveria* evolved from a pre-existing ancestral PEPC, similar to the collectively called C_3-type PEPC (Stockhaus et al., 1995/97; Engelmann et al., 2008). This reorganization of the expression pattern of the genes of the C_4 cycle comes along with the fully integration of the C_3 and C_4 cycle, that allowed for the efficient establishment of the CO_2 pump (Sage, 2004). Additional enzymes which had been subjected to C_4 alterations are CA (Ludwig, 2011), PPDK, NADP-MDH, NADP-ME (Sheen, 1999) and Rubisco (Kapralov et al., 2011).

A major factor contributing to the specific expression pattern of genes of the C_4 cycle was the integration of *cis*-acting regulatory elements at the promoter (Nelson and Dengler, 1992; Matsuoka et al., 1993/94; Taylor et al., 1997; Westhoff et al., 1997; Sheen, 1999; Rondeau et al., 2005; Christin et al., 2007; Wang et al., 2009). Promoter elements of PEPC from *F. trinervia* were defined that confer (at least in part) high-level expression (Stockhaus et al., 1995/97) and mesophyll-specificity (Akyildiz et al., 2007).

Along the C_4 evolution the transcript abundances of genes related to various metabolic pathways were altered in *Flaveria* and *Cleome* (Bräutigam et al., 2011; Gowik et al., 2011). The transcript abundances of the core genes of the C_4 cycle (up-regulated in C_4 *Flaveria* with regard to C_3 *Flaveria*) changed early in the C_4 evolution very likely before, and genes related to the Calvin-Benson cycle and to photorespiration (mostly down-regulated in C_4 *Flaveria* with regard to C_3 *Flaveria*) evolved presumably after the establishment of a fully operative C_4 cycle.

Phase 5: Optimisation. One of the final alterations, predicted by the C_4 evolutionary model was towards the maximal efficiency of the C_4 pathway. The adjustment of enzymes of the C_4 cycle to substrate availability and inhibitory metabolites was established by the enhancement of the kinetic efficiencies of the key enzymes PEPC and Rubisco (Ku et al., 1983/91; Engelmann et al., 2003); (see 2.2 for PEPC). In case of Rubisco the activity was not higher in C_4 *Flaveria* with respect to the intermediate species (Ku et al., 1983). However, the carboxylation rate ($k\text{cat}_{CO2}$) and the Michaelis-Menten constant

for CO_2 (K_c) of Rubisco was higher in C_4 than in C_3 plants. The relative specificity ($S_{c/o}$) of Rubisco remained relatively unchanged within the *Flaveria* genus (Wessinger et al., 1989; Kubien et al., 2008).

It appears likely that the kinetic variations of Rubisco between C_3 and C_4 plants was accompanied with changes in quantity of the respective protein amount. The total soluble protein amount of Rubisco per leaf area in C_4 plants decreased relative to the amount in C_3 plants. The nitrogen requirement for enzymes of the C_4 cycle is less in C_4 plants than in C_3 plants and the rate of photosynthesis per unit of N is greater in C_4 than in C_3 plants (high NUE in C_4 plants) (Ku et al., 1979; Sage et al., 1987).

2. The phospho*enol*ypyruvate carboxylase

2.1. Evolution of the PEPC (ppc) gene family

PEPC is ubiquitous among the plant kingdom from green algae to vascular plants (Lepiniec et al., 1993; Toh et al., 1994; Rivoal et al., 2002; O'Leary et al., 2011). Besides of the photosynthetic function of PEPC (C_4-type), plants possess non-photosynthetic PEPC isoforms (C_3-type) specialized to regulate the plants´ physiological and developmental processes (O´Leary *et al*., 2011). The non-photosynthetic PEPC is involved in i.g. coordinating stomata conductance, pH balance and nitrogen assimilation through replenishment of malate in the TCA cycle of the mitochondria (Latzko and Kelly, 1983; Kawamura et al., 1990; Chollet et al., 1996, Cockburn, 1996; O´Leary, 2011).

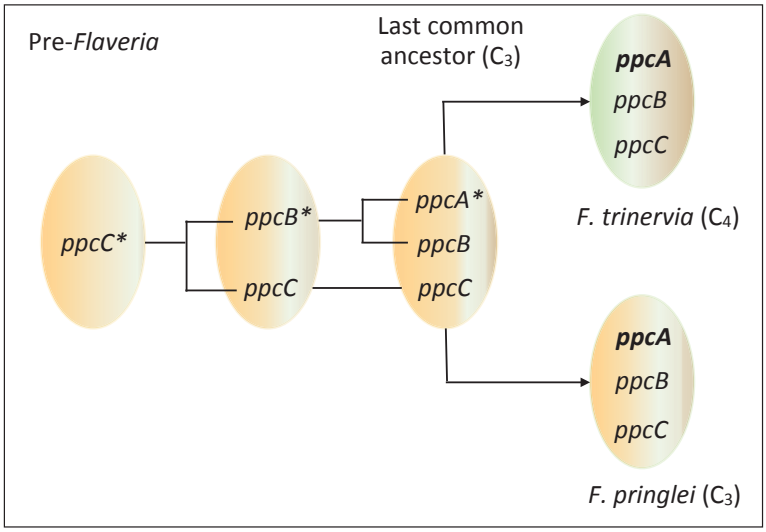


Figure 3: Flowchart of the evolution of PEPC genes (*ppc*) in the genus *Flaveria*. The set of *ppc* gene classes of C_4 *Flaveria trinervia* and C_3 *Flaveria pringlei* were pre-existent in the last common ancestral *Flaveria* plant. The *ppcA* and *ppcB* are sister genes that originated by gene duplication from *ppcB**. The *ppcA** gene represents the gene which evolved from a *ppcB** gene and likely gave origin to the orthologous *ppcA* genes in *F. trinervia* and *F. pringlei*. Adapted from Bläsing et al., 2002; Westhoff & Gowik, 2004.

Based on the diverse functions of PEPC the enzyme is encoded by a *ppc* gene family in *Flaveria* of four different but conserved classes with at least 90 % sequence identity (Poetsch et al., 1991; Herman and Westhoff, 1990) and [**Figure 3**]. The *ppcA*, *ppcB* and *ppcC* genes are present within the C_3, C_3-C_4 intermediate and C_4 *Flaveria* species.

The *ppcA* from *F. trinervia* represents the gene encoding the photosynthetic C_4-type PEPC, whose relative transcript abundance is high and exclusively expressed in the mesophyll cells of *F. trinervia*. In *F. pringlei* the expression of the evolutionary orthologous *ppcA* is not cell-type specific and in the leaves the mRNA is 100 times lower abundant than in *F. trinervia* (Ernst and Westhoff, 1997). The non-photosynthetic C_3-type PEPC isoforms are coded by *ppcB* and *ppcC*, whereby at the level of amino acid sequences the *ppcA* PEPC is more related to *ppcB* PEPC than to *ppcC* PEPC (Bläsing et al., 2002). The *ppcA* PEPC from *F. trinervia* is more similar at the level of amino acid sequences to the *ppcA* PEPC from *F. pringlei* than to the *ppcB* PEPC or *ppcC* PEPC from *F. trinervia*. In the leaves of either *F. trinervia* or *F. pringlei* the abundances of the *ppcB* and *ppcC* mRNAs are lower in regard to that of *ppcA* (Ernst and Westhoff, 1997). In *F. trinervia* this discrepancy is stronger than in *F. pringlei.* The relation between the three *ppc* genes in the leaves is different to that in stems and roots. In stems and roots from *F. trinervia* and *F. pringlei* the relative transcript abundances of *ppcB* and *ppcC* are higher to that of *ppcA* (Ernst and Westhoff, 1997).

2.2. The posttranslational regulation of PEPC

In general, the C-terminal half of the ~110 kDa PEPC polypeptide contains most of the presumed active-site determinants, whereas the N-terminal sub-domain includes motifs which are regulatory in nature (Svensson et al., 1997). PEPC is allosterically inhibited and activated by opposite metabolites (Glucose 6-phosphate, glycine, malate/aspartate and glutamate) and phosphorylated by a protein kinase (Andreo et al., 1987; Echevarría et al., 1994; Chollet et al., 1996; Vidal and Chollet, 1997; Kai et al., 2003). PEPC is dephosphorylated presumably by a phosphatase as experimentally shown in a Crassulacean acid metabolism (CAM) plant (Carter et al., 1990) and in maize (Jiao and Chollet, 1988; Vidal and Chollet, 1997; Dong et al., 2001). Some indications exist for a Ca^{2+}-dependent protein kinase (CDPK) with an unspecified substrate specificity to PEPC from maize and Sorghum (Bakrim et al., 1992; Li and Chollet, 1993; Ogawa et al., 1998).

Despite of the highly conserved amino acid sequences with 94.7 % between the *ppcA* PEPC from *F. trinervia* with the orthologous from *F. pringlei*, the PEPC isoforms from either species differ in their kinetic characteristics (Ting and Osmond, 1973). The C_4-type PEPC has a great affinity for HCO_3^- in several C_4 species (Bauwe, 1986; Bauwe and Chollet, 1986; Dong et al., 1998). The C_4-type PEPC from *Flaveria* has a higher substrate saturation constants for PEP in regard to the C_3-type counterpart enzyme (Svensson et al., 1997; Engelmann et al., 2002; Jacobs et al., 2008). The major C_4-determinant that influence the $K_{0.5}$ (PEP) of PEPC was revealed (Bläsing et al., 2000; Engelmann et al., 2002). At amino acid position 774 in the active site of PEPC, that corresponds to the residue of a conserved alanine in the ancestor condition (as seen in *F. pringlei*), in C_3-C_4 *Flaveria* intermediates and in *F. brownii* (C_4-like), was substituted by a serine (as seen in *F. trinervia*) (Bläsing et al., 2000).

Although malate/aspartate act as compensatory effectors that controls how fast four-carbon acids are produced by PEPC carboxylation, a recovery and a steady-state rate of the activity of PEPC in the presence of high concentrations of malate/aspartate is imperative for a continued flow of carbon in the C_4 cycle. In general a high concentrations of the feedback inhibitor decreases the kinetic efficiency of PEPC from *Flaveria* (Bläsing et al., 2002), *E.coli* (Kai et al., 2003), *Zea* (Wedding et al., 1990; Takahashi-Terada et al., 2005) and *Sorghum* (Duff et al., 1995). The activity of PEPC is also characterized by its inhibition by oxaloacetate (Lowe and Slack, 1971) or glutamate (Gousset-Dupont et al., 2005). The C_4-type PEPC from *F. trinervia* is relatively tolerant towards high malate concentrations when compared to PEPC from *F. pringlei* (Bläsing et al., 2002; Jacobs et al., 2008). Recently, with the approaches of high-resolution X-ray crystal structures of PEPC from *Flaveria* a single site-specific mutation was defined as the molecular determinant for the different responses towards malate/aspartate (Paulus et al., 2013a). This evolutionary alteration at amino acid position 884 in the inhibitory site of PEPC consists of the substitution of an arginine in the C_3 PEPC (as seen in *F. pringlei*) by a glycine in the C_4 PEPC (as seen in *F. trinervia*) (Paulus et al., 2013a/b). Arginine is a conserved residue among the ancestor C_3 PEPC isoforms within *Flaveria* and the *Poaceae*. Whereas, in the C_4-type PEPC isoform of 33 species the substitution of arginine was not limited to glycine (Paulus et al., 2013b).

The regulation of PEPC is adjusted to light/dark transitions, so that atmospheric CO_2 can be fixed primarily during the day in C_4 plants. In several C_4 species the activity of PEPC was light-dependent, but the protein content of PEPC or the transcriptional level of the respective gene light-independent (Jiao and Chollet, 1988; Chastain and Chollet, 1989; Jiao and Chollet, 1990a/b; McNaughton et al., 1991; Rao et al., 2006; Avasti et al., 2011). In C_4 *Flaveria* the activity of PEPC increased during the day (Rajagopalan et al., 1993) but the *ppcA* mRNA abundance from *F. bidentis* was identical between midday and midnight (Furumoto et al., 2007). In contrast, in other C_4 species the diurnally expression of *ppc* accumulated primarily during the day, and was regulated by the circadian clock in maize (Kurotani et al., 1999; Rao et al., 2006; Horst et al., 2009; Christin et al., 2013). The changes of the diurnal *ppc* mRNA abundance might not be translated into changes on the PEPC protein content (Rao et al., 2006; Khan et al., 2010; Avasti et al., 2011). The extent to which light modulates the activity of PEPC is limited in C_3 and C_3-C_4 intermediate species, including *Flaveria* (Chastain and Chollet, 1989; Rajagopalan et al., 1993; Rao et al., 2006).

3. The phospho*enol*pyruvate carboxylase protein kinase

3.1. The influence of regulatory phosphorylation of PEPC

The activity of PEPC is regulated by reversible phosphorylation, a widespread mechanism in plants (Hrabak et al., 2003) catalyzed by a serine/threonine phospho*enol*pyruvate carboxylase protein kinase (PPCK) (Jiao and Chollet, 1988; Chollet et al., 1996; Vidal and Chollet, 1997; Nimmo, 2000; Tsuchida et al., 2001). The phosphorytable residue in the substrate protein domain was identified as an invariant serine residue at the N-terminal of PEPC (Wang et al., 1992; Vidal and Chollet, 1997). Actually, among

most of the C_4, C_3 and CAM species the deduced amino acid sequences XX**S**IDAQ of the phosphorytable domain were shown to be conserved residues (Vidal and Chollet, 1997). PPCK has a high specificity to bind PEPC at this specific serine (Takahashi-Terada et al., 2005). The potential existence of multi-recognition sites outside the phosphorylation-site causing a potential increase of the phosphorylation efficiency, cannot be excluded (Li et al., 1997).

Posttranslational modification of PEPC by phosphorylation causes a shift in the kinetic constants of PEPC. The affinity for PEP or the complex Mg-PEP increased moderately for the phosphorylated PEPC in *Sorghum* (Duff et al., 1995) and maize (Terada et al., 1990; Tovar-Méndez et al., 1998/2000) and [**Figure 4**]. However, more important is the partial desensitization of the phosphorylated PEPC towards malate/aspartate as shown in maize (Jiao and Chollet, 1988/89; McNaughton et al., 1991; Dong et al., 1998; Tovar-Méndez et al., 1998/2000; Tovar-Méndez et al., 1998/2000; Takahashi-Terada et al., 2005), *F. bidentis* (Furumoto et al., 2007), *Sorghum* (Echevarría et al., 1994; Wang et al., 1992), CAM (Nimmo et al., 1986; Webb, 2003), tobacco and wheat (Duff and Chollet, 1995).

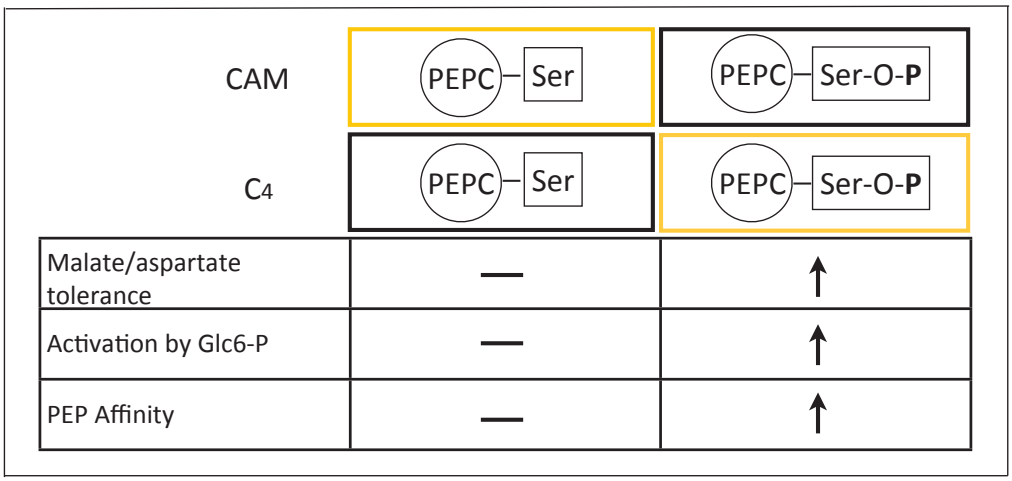


Figure 4: The influence of regulatory phosphorylation of PEPC in CAM and C_4 plants. In CAM the phosphorylated PEPC is most active during the night, whereas in C_4 the phosphorylated PEPC is most active in the light. The PEPC sensitivity to allosteric metabolites and to the substrate PEP changes drastically once PEPC is phosphorylated at the serine residue of the N-terminus by PPCK. Phosphorylation is responsible for the increase in both the Michaelis-Menten constant (K_m) and the inhibitor constants (K_i) for malate. Phosphorylation causes a modest increase of the maximal velocity (V_{max}) and an increase in the sensitivity to the activator of PEPC. The concentration of substrate to reach half of the maximal velocity ($S_{0.5}$) decreases upon phosphorylation. In CAM the catalytic activity of PEPC increases upon phosphorylation (PEP affinity refers to C_4 plants). Night is indicated by the black lined box, day is indicated by the orange lined box. Abbreviations: CAM, Crassulacean acid metabolism; Glc6-P, glucose 6-phosphate.

In addition, the sensitization of the phosphorylated PEPC to glucose-activation was observed in CAM, maize and *Sorghum* (Wang et al., 1992; Arrio-Dupont et al., 1992; Tovar-Méndez et al., 2000; Webb, 2003; Takahashi-Terada et al., 2005). The inhibition of the activity of PEPC by malate was overcome by the activator glucose 6-phosphate (Glc6-P) (Huber and Edwards, 1975; Echevarría et al., 1994). This effect increased once PEPC was phosphorylated and in addition high PEP concentrations abolished the inhibition over the phosphorylated PEPC by malate even more (Echevarría et al., 1994;

Tovar-Méndez et al., 2000). Therefore, sensitization of the activator seams to be required for the ability of the phosphorylated PEPC from maize to become desensitized towards malate (Takahashi-Terada et al., 2005). Although basic arginine residues of PEPC from *Zea mays* involved in the binding of sulfate ions were involved in Glc6-P sensitivation and responsiveness to regulatory phosphorylation (Takahashi-Terada et al., 2005), the exact binding-site for Glc6-P in the PEPC structure is not defined. The electron density of the N-terminal region of the structure of PEPC from *Zea mays* was to low to further reveal the structural explanation behind the serine phosphorylation for the regulation of PEPC (Matsumura et al., 2002), but might be explained by the alteration of the accessibility of metabolites to the inhibitor-site by the charged phosphate group (Kai et al., 1999).

The phosphorylation status and the rates of the effector metabolites are both factors that are not constant throughout the day in many species. Phosphorylation increases in the light in C_4 species and in the dark in CAM plants (Nimmo, 2003) [**Figure 4**].

The light-dependent sensitity of PEPC is linked to its modulation by phosphorylation and to malate, PEP and pH levels in a synergistic manner (Karabourniotis et al., 1985; Budde and Chollet, 1986; Huber and Sugiyama, 1986). The diurnal timing of phosphorylation is an adapted mechanism to the diurnal fluctuations with regard to the rate of the effector metabolites and to light/dark transitions. Thereby, phosphorylation was suggested to cause a rhythmic activity of PEPC in response to diurnal changes (Nimmo et al., 1987b; Kusumi et al., 1994; Vidal and Chollet, 1997).

Phosphorylation of PEPC was shown to be induced by light in maize (Jiao and Chollet, 1988; McNaughton et al., 1991), *Sorghum*, C_4 *H. verticillata* and in the C_3 species wheat (*Triticum aesfivum*) (Duff and Chollet, 1995) and tobacco (Li et al., 1996; Vidal and Chollet, 1997; Rao et al., 2006). The phosphorylation levels of PEPC follow a diurnal and even a seasonal pattern in *Amaranthus* (Avasthi et al., 2011) and in maize (Ueno et al., 2000). In these C_4 plants the amplitude of phosphorylation reached a maximum around midday and gradually diminished until midnight.

Additional roles driven by phophorylation of PEPC are considered as compensatory feedback mechanism of the metabolism of *Arabidopsis*, maize and rice to stress conditions (Ueno et al., 2000; Fontaine et al., 2002; Fukayama et al., 2006; Gregory et al., 2009) and in C_4 species an adjustment to PEPC depletion (Dever et al., 1997). Therefore, phosphorylation is not only light-regulated, but in addition regulated by nutritional factors (Izui et al., 2003).

However, the primary physiological role of PPCK *in planta* is still debatable. In CAM, PPCK was not the major factor controlling the temporal separation of PEPC carboxylation (Taybi et al., 2004). *Arabidopsis* and *F. bidentis* mutants were used to study the role of PPCK for photosynthesis and non-photosynthetic pathways *in planta*. An abolishment of phosphorylation of PEPC in *Arabidopsis* mutants revealed a decrease in the concentration of metabolite intermediates of the tricarboxylic acid cycle (TCA) and an impairment in the cell expansion (Meimoun et al., 2009). An abolishment of phosphorylation in *F. bidentis* mutants did not dimished the net assimilation of CO_2 significantly, and the mutants had no aberrant phenotype in respect to the wild-type (Furumoto et al., 2007).

3.2. Classification of the gene family and protein structure of PPCK

In many C_3, C_4 and CAM species, the PPCK protein is encoded by a *PPCK* gene family (Nimmo, 2003; Chollet et al., 1996). The cDNA sequences and/or genomic DNA sequences of *PPCK* were identified in C_4 *Flaveria bidenis* and *F. trinervia* (Tsuchida et al., 2001; Furumoto et al., 2007), *Zea mays L.* (Shenton et al., 2006; Wang and Chollet, 1993), *Brachypodium dystachon*, *Setaria italica*, *Sorghum bicolor* (Bakrim et al., 1992; Shenton et al., 2006), *Digitaria sanguinalis* (Giglioli-Guivarc'h et al., 1996), rice (*Oryza sativa*) (Fukayama et al., 2006), *Arabidopsis thaliana* (Fontaine et al., 2002), sugar beet (*Beta vulgaris*) (Kloos et al., 2002), soybean (*Glycine max*) (Zhang and Chollet, 1997; Sullivan et al., 2004/05), Solanaceae: *Lycopersicon esculentum* (Marsh et al., 2003) and tobacco (*Nicotiana tabacum L.*) (Li et al., 1996), Fabaceae: legume (*Lotus japonicus*) and bean (*Vicia faba*) (Du et al., 1997), CAM plants: e.g. *Kalanchoë* (Theng et al., 2008), *Opuntia ficus-indica* (Mallona et al., 2011) and *Mesembryanthemum crystallinum*, and in the C_3-CAM plants *Clusia mino*r (Borland and Griffiths, 1997).

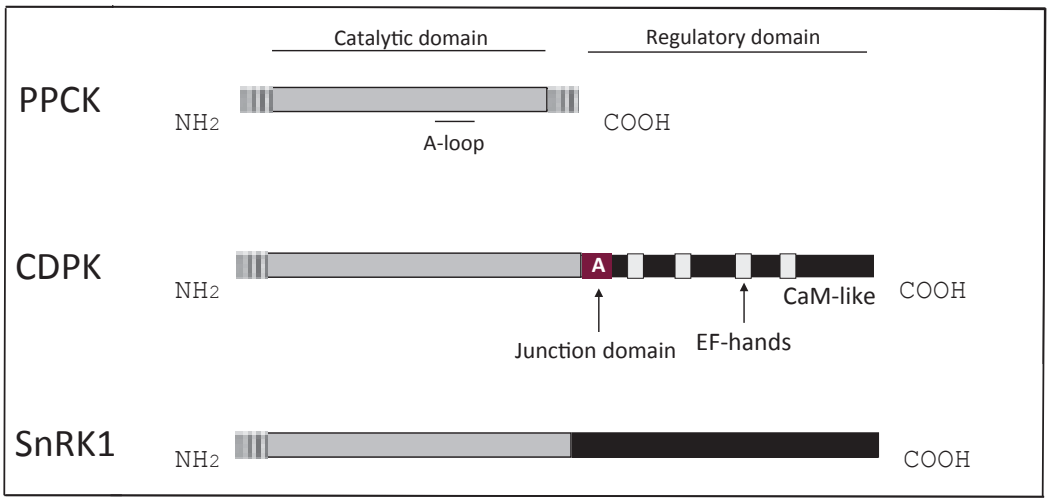


Figure 5: Domain structure of single plant protein kinases of the CDPK-SnRK superfamily. PPCK comprises only the catalytic core domain within which the activation loop is indicated. The amino- and carboxyl-terminal extensions are indicated in dash boxes, which have distinct amino acid lengths and functions between the kinases. The junction or autoinhibitory domain (red box) is located next to the catalytic domain of the CDPK kinases. EF-hands (calcium-binding subdomain motifs) are located within the CaM-like regulatory domain of the CDPK kinases. Abbreviations: PPCK, Phospho*enol*pyruvate carboxylase protein kinase; CDPK, calcium dependent protein kinase; SnRK, sucrose non-fermenting-1 (SNF1)-related kinase; NH2, amino-terminal domain, COOH, C-terminal domain; A-loop, activation-loop; A, autoinhibitory domain; CaM, calmodulin; EF, E- and H-helices. Adapted from Hardie, 1999; Harmon et al., 2001 and Hrabak et al., 2003.

The PPCK protein comprises a catalytic core and ranges in size between 30 to 39 kDa among diverse species (Hanks and Hunter, 1995; Chollet et al., 1996; Tsuchida et al., 2001) and [**Figure 5**]. PPCK belongs to a subfamily of the sucrose non-fermenting-1 (SNF1)-related kinase (SnRK) and is closely related to the Ca^{2+}-dependent protein kinases (CDPKs) (Hartwell et al., 1996; Halford and Hardie, 1998; Hardie, 1999/2000; Harmon et al., 2001; Halford et al., 2003; Hrabak et al., 2003). The PPCK protein shares a putative common evolutionary origin with the CDPKs, CDPK-related kinases (CRKs) and the phospho*enol*pyruvate carboxylase kinase-related kinases (PEPRKs) (Hrabak et al., 2003).

All plant kinases possess a catalytic core formed by two lobes of different sizes (Hanks and Hunter, 1995). One at the N-terminus of β-sheets functional in ATP-binding and the larger C-terminal lobe of mostly α-helices required for the recognition of the substrate and the initiation of the transfer of phosphate groups (Johnson et al., 1996; Dissmeyer and Schnittger, 2011). The binding of ATP and the substrate in a catalytic cleft of the active site between the two lobes of the open form, results in the transfer of γ-phosphate of ATP to the hydroxyl oxygen of the serine (threonine or tyrosine) residue of the substrate in the closed form, followed by the release of ADP and the phosphorylated substrate (Ubersax and Ferrell Jr., 2007). The kinase specificity or recognition is determined by the charge and the hydrophobicity of the active site of the kinase surface residues (Ubersax and Ferrell Jr., 2007).

PPCK lacks the motif of the E and F helixes (EF-hand) to bind calcium. On this account PPCK is excluded from the category of specific plant Ca^{2+}-dependent protein kinases (Hartwell et al., 1996; Hartwell et al., 1999a). The PPCK protein structure does neither integrate an auto- or trans-phosphorylation domain between the carboxyl and the amino terminal domain, nor a regulatory carboxyl terminal domain with alternative extra-regulatory sub-domains (Hartwell et al., 1996; Harmon et al., 2001; Hrabak et al., 2003).

Several kinases are targeted and activated by phosphorylation at residues in the activation loop (A-loop), resulting in a conformation change of the kinase for adaptive substrate recognition and catalysis (Johnson et al., 1996; Huse and Kuriyan, 2002; Marchler-Bauer et al., 2013) and [**Figure 5**]. The PPCK enzyme is assumed to be inaccessible to covalent or non-covalent binding, and to be not regulated directly by second messengers or protein/RNA interactions (Hartwell et al., 1996; Johnson et al., 1996; Hardie, 1999). The assumption of the existence of a phosphate recognition site in the A-loop of PPCK as target protein was rejected. Because a non-phosphorylatable glycine (around amino acid 170) in the active-site of PPCK was suggested to prevent the access to phosphorylation (Taybi et al., 2000). A possible explanation for this specific loss of phosphorylation-dependent kinase activation would be, that the PPCK enzyme does not require a negative charge through the addition of phospho-amino acid residues in order to become activated. On this account PPCK was suggested to be a constitutively active kinase (Hrabak et al., 2003).

3.3. The activity of PPCK is strongly regulated at the transcriptional level

In accordance with the *PPCK* gene family within species a differentiated cell-type and temporal specific expression pattern for each *PPCK* isogene was indicated to play a role for diverse biological functions. In addition, the expression pattern of *PPCK* was suggested to be under evolutionary pressure (Taybi et al., 2004).

The posttranslational control of PEPC by phosphorylation, is in contrast to the regulation of the PPCK enzyme at the transcriptional level (Li et al., 1996; Hartwell et al., 1999a; Hrabak et al., 2003; Rao et al., 2006). In CAM the *PPCK* transcript, translatable mRNA abundances and the activity of PPCK follow a day/night rhythmic pattern in response to diurnal changes with a peak during the night when CO_2 fixation takes place (Carter et al., 1991; Borland and Griffiths, 1997; Borland et al. 1999; Hartwell et al., 1999a; Nimmo, 2000; Mallona et al., 2011). The diurnal transcript abundance of *PPCK* from CAM is

primarily controlled by changes in organic metabolite levels (malate or soluble sugar), a range of other factors (developmental stages, temperature, salt) and secondarily by the circadian clock (Nimmo et al., 1984/87a; Kluge et al. 1988; Kusumi et al., 1994; Li and Chollet, 1994; Carter et al., 1995; Grams et al. 1997; Hartwell et al., 1996; García-Mauriño et al., 2003; Taybi et al., 2000/04; Theng et al., 2008) and [**Figure 6**].

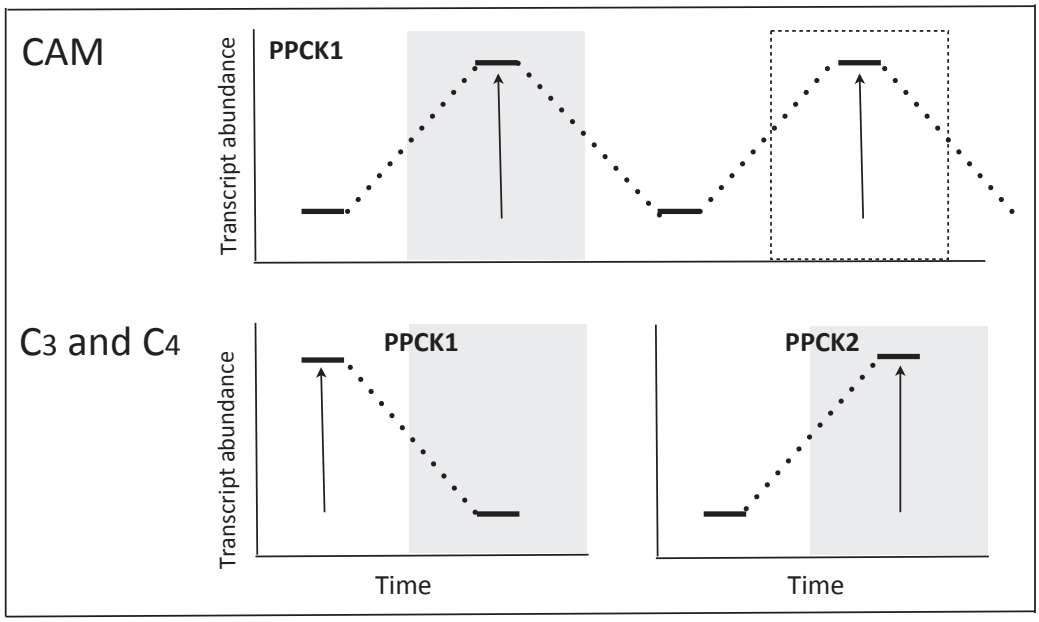


Figure 6: Regulation of PPCK in CAM, C_4 and C_3 plants. In CAM the *PPCK* transcript, translatable mRNA and PPCK activity is day/night regulated and high during the night (grey box). The transcript abundance of *PPCK* from CAM remains rhythmic under constant light and therefore considered as circadian regulated. In C_4 and C_3 species the regulation of the transcript abundance of *PPCK* is controlled by light/dark fluctuations and varies between *PPCK* isogenes within species and among different species. Abbreviations: CAM, Crassulacean acid metabolism.

In C_4 species the expression of *PPCK* varies between isogenes and can be either light or dark-regulated. The integration of circadian factors were suggested to modulate *PPCK* in C_4 plants (Ueno et al., 2000), but in maize no evidence could be shown (Shenton et al., 2006). The increased expression of *PPCK* in response to light indicate for a primary role of the kinase in phosphorylating the photosynthetic PEPC. Whereas, the increased nocturnal-expression of *PPCK* was suggested to be related to non-photosynthetic functions (Aubry et al., 2011). The abundance of the *PPCK* mRNA from *F. trinervia* and *ZmPPCK1* from maize was higher in light-exposed leaf extracts when compared to that of dark-exposed leaf extracts (Tsuchida et al., 2001; Shenton et al., 2006). *ZmPPCK2* was significantly higher during the night in the bundle-sheath cells. The diurnal expression of *PPCK* from the C_4 *Alloteropsis* had a clear peak at midday (Christin et al., 2013). In *Sorghum* the relative transcript abundance of *SbPPCK1* was high during the light and low during the dark, whereas the transcript abundance of the other isogene *SbPPCK2* remained constant in response to light/dark transitions (Shenton et al., 2006).

In C_3 plants the regulation of *PPCK* varies between species and includes light/dark-, circadian-, metabolite- and nitrogen- or phosphate-mediated factors (Vidal et al., 2002). In addition, in soybean nodules the photosynthase supply to the leaves was required for the enhancement of the transcriptional

and translational synthesis of PPCK (Zhang and Chollet, 1997; Sullivan et al., 2004). In a variety of C_3 plants the induction of light/dark signals on either the expression of *PPCK* or the phosphorylation levels was marginal (Leport et al., 1996; Smith et al., 1996; Fukayama et al., 2006; Feria et al., 2008). However, in a total of thirty C_3 plants twelve revealed a light-dependent phosphorylation, whereas in three phosphorylation occurred during the night (Fukayama et al., 2006).

In *F. pringlei* the relative *PPCK* mRNA abundance from either light- or dark-exposed leaf extracts was very low (Tsuchida et al., 2001). The transcript abundance of *PPCK* from C_3 *Alloteropsis* was low relative to the orthologous *PPCK* gene from C_4 *Alloteropsis* and remained constant along the diurnal cycle (Christin et al., 2013). In rice the expression of *OsPPCK3* was high during the night, whereas the one of the *OsPPCK2* isogene was high during the light (Fukayama et al., 2006). *OsPPCK3* was suggested to be controlled by the circadian clock, but primarily controlled through the accumulation of metabolites. A light-dependent expression was defined for both *PPCK* isogenes from *Arabidopsis* (Fontaine et al. 2002), from *L. esculentum* (Marsh et al., 2003) and for the *PPCK* gene from tobacco (Li et al., 1996). In wheat and C_3 *Hydilla verticillata* the phosphorylation of PEPC was light-dependent and increased under nitrogen-treated conditions in wheat (Duff and Chollet, 1995; Rao et al., 2006).

The expression of *GmPPCK4* generates an apparent circadian rhythm with a peak during the day in the leaves of soybean, but not in roots (Sullivan et al., 2004/05). A motif being similar to the *Arabidopsis* ´evening element´ of the promoter of *GmPPCK4* was found to be required for the circadian regulation. The remaining *GmPPCK* genes did not contain a specific motif in the promoter region, elucidating why *GmPPCK2* and *GmPPCK3* were light- but not circadian clock-regulated.

In C_4 species one leaf-expressed *PPCK* cDNA was characterized for *F. trinervia* (Tsuchida et al., 2001) and *F. bidentis* (Furumoto et al., 2007). Although of a high sequence identity between numerous *PPCK* genes from maize, three *ZmPPCK* genes are expressed in the leaf mesophyll cells, whereas one is expressed in the bundle-sheath cells (Shenton et al., 2006). In soybean four *GmPPCK* genes showed different expression patterns along shoots, roots and leave tissues (Sullivan et al., 2004/05). The transcripts of the genes *OsPPCK1* and *OsPPCK3* from rice are highly abundant in the roots, whereas *OsPPCK2* is ubiquitously expressed in each plant organ (Fukayama et al., 2006). The expression of two *PPCK* genes from *Arabidopsis* differed according to their cell-type specificity (*AtPPCK1* in rosette leaves and *AtPPCK2* in flowers) (Fontaine et al. 2002). In *Sorghum* and *Lycopersicon esculentum* two *PPCK* genes were identified within each species, in which the *LePPCK* isogenes share a common cell-type expression (Bakrim et al., 1992; Marsh et al., 2003; Shenton et al., 2006).

PPCK is a low abundant transcript and due to specific elements at the 3´ untranslated region of the *PPCK* mRNA prone to instability (Hartwell et al., 1999b; Tsuchida et al., 2001; Taybi et al., 2004). In addition, PPCK is controlled by synthesis/degradation of the PPCK protein (Hartwell et al., 1999a; Osuna et al., 1999; Echevarría and Vidal, 2003). The PPCK protein is very low abundant (Agetsuma et al., 2005) with a fast turnover rate (Jiao et al. 1991a, Shenton et al., 2006). In *Flaveria* the *in vitro* phosphorylation of PEPC appears to be regulated by thioredoxin-dependent reductions of disulfide bonds (Saze et al., 2001; Tsuchida et al., 2001). Another factor regulating the phosphorylation of PEPC in illuminated *Zea mays L.* and *Kalanchoë fedtschenkoi* leaves is a protein-inhibitor that acts negatively over the protein activity and content of PPCK (Nimmo et al., 2001).

3.4. The light-dependent signal-transduction chain of phosphorylation of PEPC

In C_4 species a signal cascade leading to the phosphorylation of PEPC is induced by light involves intracellular pH changes, Ca^{2+} release from the vacuole and changes in the concentration of inositol-1,4,5-trisphosphate (Ins(1,4,5)P3) (Duff et al., 1996; Coursol et al., 2000; Gousset-Dupont et al., 2005) and [**Figure 7**]. Illumination of protoplasts of *Sorghum* (Pierre et al., 1992) and of intact leaves from maize and *Sorghum* (McNaughton et al., 1991; Jiao et al., 1991b; Jiao and Chollet, 1992; Bakrim et al., 1992/93) was a precondition of phosphorylation. In *Digitaria sanguinalis* ATP and NADPH were required for phosphorylation (Giglioli-Guivarc'h et al., 1996). Cytosolic alkalization is critical for the light-mediated activation of PEPC in C_4 species (Rajagopalan et al., 1993), and also an early signaling factor for the increase of the activity of PPCK and/or the phosphorylation of PEPC in C_4 species (Pierre et al., 1992; Giglioli-Guivarc'h et al., 1996; Vidal and Chollet, 1997; Taybi et al., 2000; Echevarría and Vidal, 2003). The pH levels determine most likely the onset of phosphoinositide-specific phospholipase C (PI-PLC) that functions as a second messenger to induce the formation of $Ins(1,4,5)P_3$ (Giglioli-Guivarc'h et al., 1996; Coursol et al., 2000; Monreal et al., 2010). Ligand-gated Ca^{2+}-channels were required for the activity of PPCK by unknown Ca^{2+}-dependent protein elements (Giglioli-Guivarc'h et al., 1996; Li et al., 1996). *In planta* 3-PGA, a product of the Calvin-Benson cycle, is protonated in the mesophyll cells and contributes to the alkalization (Duff et al., 1996). Due to the fact that the protein synthesis of PPCK was independent from an operative light-activated Calvin-Benson cycle in *Digitaria sanguinalis* and maize (Smith et al., 1998), different compounds or pathways were suggested to act on the signal-transduction chain.

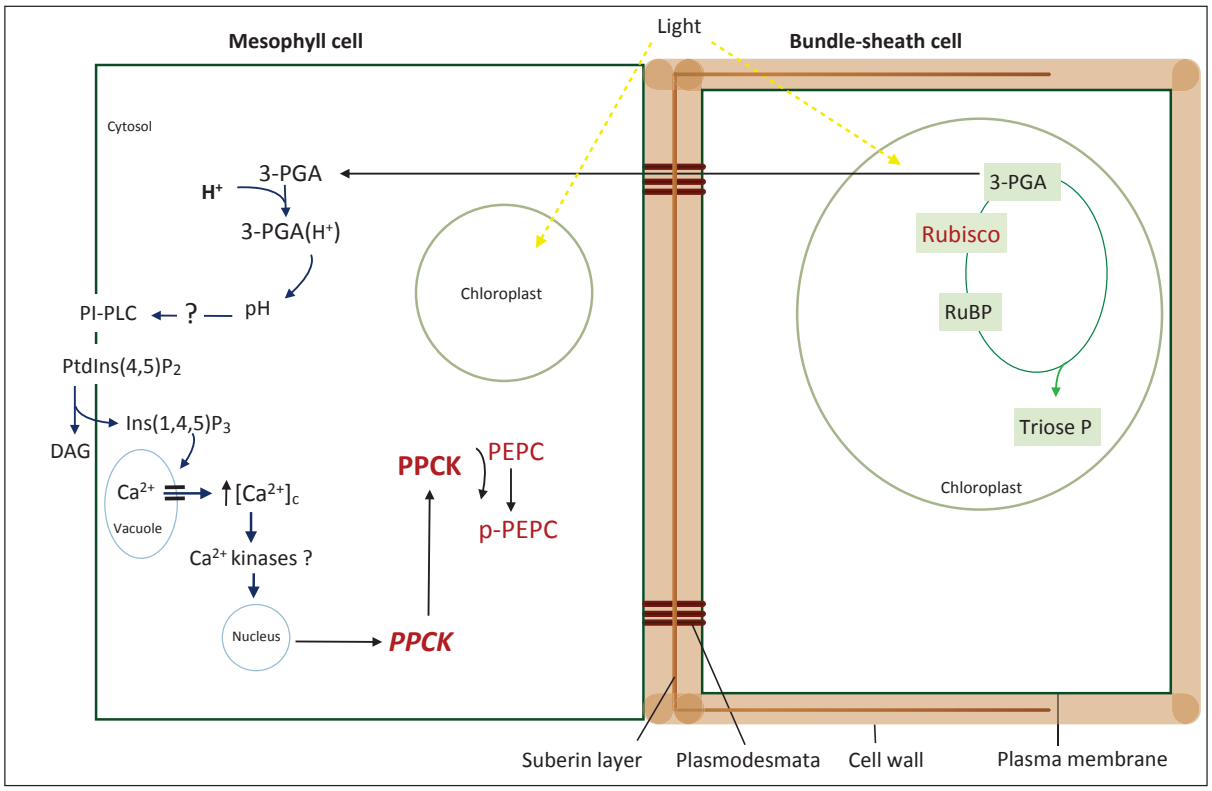


Figure 7: The signal transduction of C_4 PEPC phosphorylation circuitry in illuminated *D. sanguinalis* and *Sorghum*. Cross-talk between the two photosynthetic tissues allows for a protonation in form of 3-phosphoglyceric acid (3-PGA) in the mesophyll cells. The alkalization functions as an early signaling element. D-*myo*-inositol-1,4,5-tiphosphate ($Ins(1,4,5)P_3$) is a product of Phosphatidylinositol-(4,5)-biphosphate (PtdIns (4,5) P_2) and suggested to mobilize Ca^{2+} from vacuole storages. Ca^{2+}/calmodulin-regulated protein kinases are suggested to transmit the light signal to the nuclear-encoded expression of *PPCK*. Additional abbreviations: DAG, Diacylglycerol; PP2A, 2A-type protein phosphatase; PI-PLC, Phosphoinositide-specific phospholipase C; RuBP, Ribulose 1,5-bisphosphate; Triose P, Triose phosphate. Adapted from: Chollet et al., 1996; Giglioli-Guivarc'h et al., 1996; Echevarría and Vidal, 2003.

4. Diurnal regulation in plants

4.1. Light induces biological processes

The light/dark alternation is critical for the triggering of physiological processes, photosynthesis and for anticipating the initiation of developmental stages, such as seed germination and flowering (Pratt, 1982; Koornneef et al., 1998; Mockler et al., 1999; Christensen and Silverthorne, 2001; Halliday et al., 2009). Plants perceive light in form of physical energy that differs in its quality (wavelength), quantity (irradiance), duration (photoperiodism) and direction (phototropism). Light is sensed and converted to chemical energy by red and blue light photoreceptors. Phytochrome, cryptochrome and phototropins are the best characterized photoreceptors to mediate light responses through signaling transduction pathways.

The sensed environmental changes are transduced into changes in expression of genes encoding transporters or enzymes of the assimilatory pathways. Thus, the activation/repression of gene expression

is under the control of the presence/absence of external signals, but in addition to the internal demand and availability of carbon and nitrate in the plant. Changes in diurnal gene expression are required to adjust the activity of physiological and metabolic processes to a close to a 24 hour periodicity. Transcription of genes that are induced and which abundances oscillate significantly in response to diurnal transitions in a 24-h cycle are considered as diurnally regulated genes. Diurnally regulated genes include genes regulated by light or genes that respond to nutrient levels (glucose and nitrate) (Girke et al., 2000; Bläsing et al., 2005; Wang et al., 2000).

Crucial light-regulated genes are the chlorophyll *a/b*-binding proteins of the light-harvesting complex (*Lhcb*) of the photosynthetic light reaction and the small subunit of Rubisco (*rbcS*) (Meyer et al., 1989; Wang and Tobin, 1998). *Cis*-acting-elements of the promoters of these light-regulated genes in *Arabidopsis* are subjected to transcription factors of the phytochrome-mediated signal transduction pathway in order to initiate the rhythmic transcription in a 24-h periodicity (Millar and Kay, 1996; Wang and Tobin, 1998; Christensen and Silverthorne, 2001; Tyagi and Gaur, 2010). Conserved sequence motifs of the light-responsive promoters are collectively called as light regulatory elements (LREs) (Grob and Stüber, 1987; Chattopadhyay et al., 1998).

4.2. The circadian clock controls biological processes

Diurnally regulated genes also include genes which are regulated by the circadian clock. The circadian clock functions to anticipate environmental changes during the day and night. Changes in gene expression and in activity of enzymes of metabolic processes generated by the clock are defined as circadian rhythms. Due to the fact that these persist in the absence of light or temperature stimuli circadian rhythms are considered endogenous and self-sustaining (Giuliano et al., 1988). In case the transcriptional oscillations of entrained conditions persist during a free-running cycle with constant light or darkness, the gene is considered as circadian clock-regulated.

The circadian clock enables plants to predict environmental changes (Millar et al., 1995) and it was suggested to play a major role in a wide range of plants (Filichkin et al., 2011) to even optimize the plants´ fitness and survival (Green et al., 2002; Dodd et al., 2005; Farré and Weise, 2012). It is absolutely clear though that the circadian clock is contributing to the plant growth and developmental processes including flowering (Koornneef et al., 1998; Zhao et al., 2012), photosynthesis (Schaffer et al., 1998; Nusinow et al., 2012), germination (Baskin and Baskin, 1976), plant defense responses (Yakir et al., 2007), CO_2 assimilation (Salomé et al., 2002), starch degradation (Smith and Stitt, 2007), stomata opening (Gorton et al., 1993; Webb, 2003), embryonic stem elongation (Nusinow et al., 2011), photoperiodism (McClung, 2006), operation of CAM (Webb, 2003) and enzymatic activities (Avasthi et al., 2011)

4.3. The circadian oscillator

The cycling of the transcriptome is synchronized through a diurnal and circadian network by clock transcription factors and the *cis*-regulatory elements. Circadian rhythms are functioning in a cell-autonomous system referred to as the oscillator (Thain et al., 2000; Harms et al., 2004; Nagoshi et al., 2004; Salomé et al., 2008). The oscillator keeps a light/dark cycle of close to a 24-h day-length to synchronize the phases of physiological processes to daily and seasonal changes. Over the last decade, the functioning of so-called core clock genes (or clock regulators) was described as the molecular basis of the circadian rhythms (McClung, 2006). Environmental stimuli of the input pathway such as light cause changes in the periodicity of the transcript abundances of the clock core components (Farré et al., 2005; Gould et al., 2006; Harmer, 2009). Environmental stimuli thereby function as critical timers for the onset of the circadian rhythms. The environmental ability to reset the clock depends upon the time of the day when the stimuli (light but also temperature) is given (Millar, 2004; McClung, 2006), and on the photoreceptor-mediated signaling (Koornneef et al., 1998; Somers et al., 1998). The output of the clock also influences the light input pathway[**Figure 8**].

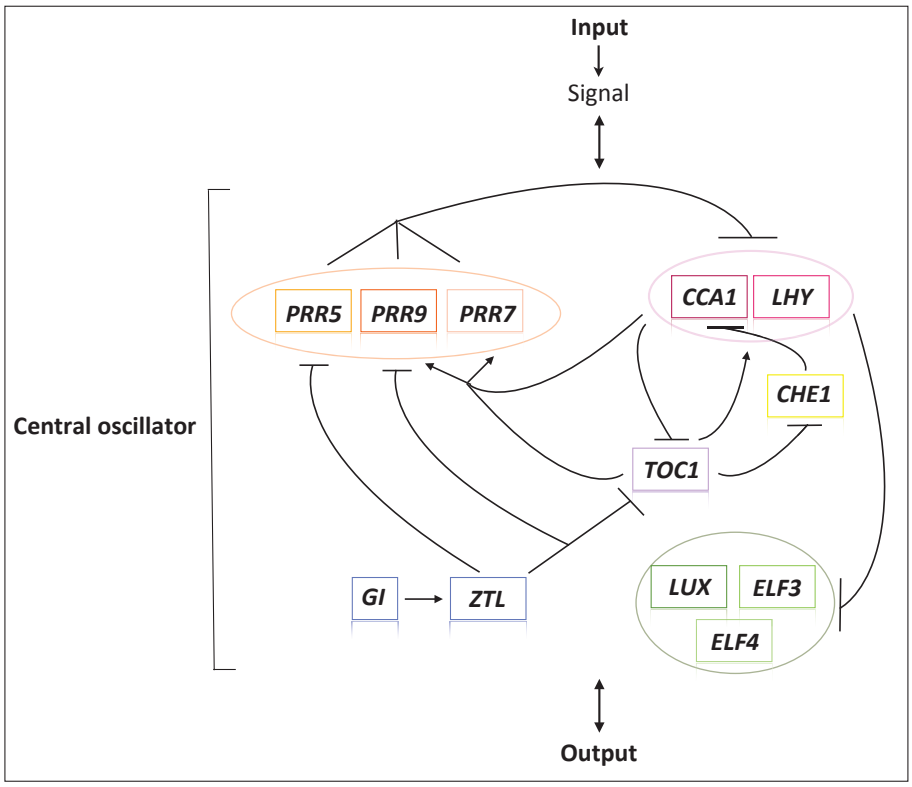


Figure 8: Schematic representation of the feedback-regulated interlocked system of the circadian clock. Operation of the circadian oscillator involves the cooperative interaction between environmental, metabolic and genetic factors with the major clock components of the core clock. Abbreviations: CCA1, CIRCADIAN CLOCK ASSOCIATED 1; LHY, LATE ELONGATED HYPOCOTYL; TOC1, TIMING OF CAB 1; ELF3, EARLY FLOWERING 3; ELF4, EARLY FLOWERING 4; LUX, LUX ARRYTHMO; CHE, CCA1 HIKING EXPEDITION; ZT, ZEITLUPE; GI, GIGANTEA; PRR, PSEUDO-RESPONSE REGULATOR. Adapted from Farré and Weise, 2012.

The light signaling and circadian clock pathways are associated and both connected to photoreceptors-mediated pathways (Fankhauser and Staiger, 2002). Thereby, the stability and translation of the mRNA of light-regulated genes are concomitantly regulated by the biological clock (Schaffer et al., 2001; Bläsing et al., 2005; Franklin and Whitelam, 2007; Guo et al., 2008; Harmer, 2009). Clock regulators can be degraded by light (Harmer, 2009). It was suggested, that the control of light- and circadian clock-regulated genes work through the same essential *cis*-regulatory network conserved across mono- and dicotyledonous species (Dunlap, 1999; Harmer, 2009; Khan et al., 2010; Filichkin et al., 2011).

Core clock genes oscillate in a 24-h interval and are grouped into three major classes of phase-specific *cis*-regulatory modules: the morning, evening and midnight modules (Michael et al., 2008). Core clock genes affect each other positively/negatively by reciprocal regulation in interlocked feedback-loops within the circadian oscillator (Alabadi et al., 2001; Locke et al., 2006) and [**Figure 8**]. The feedback-regulatory system is based on transcription and by anticipating the turnover rate of the encoding core clock proteins (Van Gelder et al., 2003; Harmer, 2009). The core clock genes while binding to clock-regulated genes of the output pathway influence the rhythm of circadian-mediated biological processes (Kiki et al., 2005; Harmer, 2009). The interaction with each other occurs to some extent in a light-dependent manner (Harmer, 2009). The oscillator is required to respond to environmental stimuli and in turn to transmit the information at an appropriate diurnal/seasonal time at the cellular level.

II. Aims

C_4 photosynthesis evolved multiple times and independently out of a C_3 ancestral condition. The evolution of the phospho*enol*pyruvate carboxylase (PEPC) enzyme involved alterations at the molecular level and in the kinetic and regulatory properties of the enzyme, which allowed for the optimization of the C_4 cycle. PEPC is posttranslational modified by the phospho*enol*pyruvate carboxylase protein kinase (PPCK). The aims of this study were to define C_4-related evolutionary changes regarding PEPC, the PPCK family and the global regulation in diurnal gene expression in the genus *Flaveria*.

(1) *Establishment of PEPC-deficient Flaveria bidentis.* The polyphyletic origin of C_4 photosynthesis is restricted to a limited number of defined evolutionary steps. The alterations of the kinetics and regulatory properties of PEPC occurred at a late stage of the C_4 evolutionary trajectory. This ultimate adaptive step is strongly selected and the fixation of the new C_4-trait is short when compared to the early step of the initial establishment of a C_4 metabolic cycle (Sage et al., 2012; Heckmann et al., 2013). Therefore, the manipulation of the C_4-type PEPC in *Flaveria bidentis* (C_4) should reveal substantial effects on C_3-to-C_4 changes *in planta*. The aim was to establish a conditional *knockout* system of C_4 PEPC in *F. bidentis*, and to further complement this mutant by the replacement of a non-photosynthetic C_3 PEPC isoform.

(2) *Evolution of the PPCK family in the genus Flaveria.* Several C_3 and C_4 species are composed by at least two PPCK isoforms. The temporal/spatial expression of *PPCK* is isogene-specific within species and can differ among species. The high-level expression (Gowik et al., 2011) and the diurnal pattern in gene expression (Christin et al., 2013) of *PPCK* in C_4 species indicate for a regulatory C_4 function. The aim was to investigate the evolutionary changes of a putative PPCK family from *Flaveria*. In a comparative analysis of PPCK along C_3 to C_4 species, the question was addressed weather or not the regulation of PPCK is photosynthetic-type dependent. The purpose was to highlight potential adaptive changes in gene expression, sequence order and activity of the kinase in the context of C_4 evolution in the genus *Flaveria*. Therefore, the characterization of the phylogeny of the PPCK family, the diurnal patterns of gene expression and *in vivo* phosphorylation of PEPC and the *in vitro* activity of PPCK recombinant protein was addressed.

(3) *Global regulation of the diurnal leaf transcriptomes from F. pringlei and F. trinervia.* In a previous study, transcript abundances at a single time point of genes from closely related C_3, C_4, and C_3-C_4 intermediate species of *Flaveria* were found to be altered during C_4 evolution (Gowik et al., 2011). The aim was to define alterations in the diurnal transcriptome from *F. pringlei* (C_3) and *F. trinervia* (C_4) and to characterize potential C_4-related evolutionary differences.

III. Thesis

(1) *Establishment of PEPC-deficient Flaveria bidentis plants.* The predictions deduced from previously *in vitro* molecular, biochemical and structural studies considering the evolution of PEPC were extended with an *in vivo* approach. A chemical-inducible *knockout* system was used in this study to trace down the *in vivo* structural/functional relationship of PEPC for C_4 evolution. On this account, artificial miRNA (amiRNA) were directed to the untranslated regions of the *ppcA* PEPC mRNA from *F. bidentis*, and introduced *in planta* in order to trigger the degradation of the targeted endogenous C_4 *ppcA* mRNA in the transgenic *F. bidentis* plants. The attempt to establish a tightly regulated and highly inducible system resulted in transgenic *F. bidentis* plants with no clear decrease in PEPC, improper for analytical analyses at later stages.

(2) *Evolution of the PPCK family in the genus Flaveria.* A pairwise gene family of *PPCKA* and *PPCKB* within photosynthetic diverse *Flaveria* species was identified. Based on the phylogenetic characterization of the PPCK family and the expression of *PPCK*, a trend parallel to the evolution of the PEPC family within the genus *Flaveria* was represented. The expression of *PPCK* and activity of the kinase from *Flaveria* phase-shifted from a C_3 to a C_4 mode along the diurnal cycle. The transcript abundance of *PPCKA* was highest during the day, and *PPCKB* highest during the night in *F. trinervia* (C_4) and *F. pringlei* (C_3). The high-level expression of *PPCKA* from *F. trinervia* is in agreement with C_4-associated amino acid changes. The diurnal *in vivo* phosphorylation pattern of PEPC as estimated by targeted mass spectrometry differed between the C_3 and C_4 *Flaveria* species. The phosphorylation status of the C_4 isoform PEPC (*ppcA*) reached its maximum during the day, whereas in the C_3 *Flaveria* the phosphorylation status was relatively constant throughout the diurnal cycle. C_4-related changes of PPCKA in *Flaveria* indicate, that the evolution of PPCK optimized the C_4 evolutionary process and that PPCKA and PPCKB are prone to execute distinct roles in *Flaveria*.

(3) *Global regulation of the diurnal leaf transcriptomes from F. pringlei and F. trinervia.* A set of diurnally regulated genes was identified in the diurnal transcriptome from *F. pringlei* and *F. trinervia.* The absolute transcript abundances of the total number of genes were mostly different between the C_3 and C_4 species, whereas the global diurnal pattern of gene expression was highly similar between the species. Changes in diurnal gene expression is a mechanism already established in the C_3 *F. pringlei*. Although of a phase-conservation of the diurnal pattern of gene expression between *F. pringlei* and *F. trinervia*, particular genes probably gained or lost the ability to respond to diurnal changes towards C_4 evolution.

IV. Summary

The C_4 photosynthetic carbon cycle is the end product of step-wise acquired combinations of new traits of the C_3 ancestry. One cardinal step, was the evolution of the phospho*enol*pyruvate carboxylase (PEPC) of the C_4 cycle from a non-photosynthetic C_3 PEPC. The C_4 PEPC (*ppcA*) is the entry enzyme for atmospheric CO_2 uptake in the mesophyll cells of the C_4 cycle. The high kinetic and regulatory efficiency of the C_4 PEPC confers the generation of an elevated concentration of CO_2 around of ribulose-1.5-bisphosphate carboxylase/oxygenase (Rubisco) of the Calvin-Benson cycle in the bundle-sheath cells. In this present study of C_4 evolution, the genus *Flaveria* was used for possessing closely related species with C_3-to-C_4 photosynthetic types. The first aim was to extend the *in vitro* analysis of C_3-to-C_4 changes of PEPC from *Flaveria* by an *in planta* approach. Transgenic C_4 *F. bidentis* plants were generated with the intention to decrease the *ppcA* mRNA amount and to inactivate the C_4 PEPC enzyme.

The activity of PEPC is regulated by phosphorylation by the PEPC protein kinase (PPCK). The second aim of this study was the evolutionary characterization of PPCK from *Flaveria*. A gene family of congeneric orthologous pairs, *PPCKA* and *PPCKB*, conserved within each of the C_3-to-C_4 *Flaveria* species was identified. The phylogenetic characterization of PPCK is in agreement with the polyphyletic origin of C_4 photosynthesis and in parallel to the evolution of the PEPC family. While the transcript abundance of *PPCKA* increased, *PPCKB* decreased from the C_3 to the C_4 species. In *F. trinervia* (C_4) and *F. pringlei* (C_3) the diurnal expression of *PPCKA* was phase-shifted towards the day, whereas *PPCKB* had maximal expression during the night. The high-level expression of *PPCKA* during the day matched to the high *in vivo* phosphorylation level of PEPC in *F. trinervia*. The specificity of the recombinant PPCKA protein from C_4 *Flaveria* to the substrate from C_3 and C_4 *Flaveria* species was different, but not abolished. Despite of the high level amino acid sequence similarities between the PPCK isoforms, the amino acid exchanges and the properties of *PPCKA* argue that the encoded protein corresponds to the kinase phosphorylating the C_4 isoform PEPC (*ppcA*) in C_4 *Flaveria*. Whereas, the properties of *PPCKB* are in good agreement to the regulation of the *ppcB* and *ppcC* PEPC isoenzymes in non-photosynthetic tissues, with a higher relevance in C_3 *Flaveria*. It can be concluded, that the PPCK family is likely under evolutionary pressure and linked to the optimization of the PPCKA structure for the operation of the C_4 cycle.

In the *Flaveria* transcriptome neither the global regulation of diurnal gene expression is known, nor to what extent it has changed during C_4 evolution. In a third approach, the diurnal transcriptome from *F. pringlei* and *F. trinervia* was profiled to identify diurnally regulated genes that encode enzymes with predicted function in various biological pathways. The alterations of the transcriptome were associated to C_4 photosynthesis and in response to diurnal changes. Photosynthetic-type differences were ascribed to differences in the absolute transcript abundances, rather than on differences in the diurnal pattern of gene expression. The discrepancy in transcript abundance of some genes varied between *F. pringlei* and *F. trinervia* according to diurnal changes. Although diurnal gene expression is a mechanism already established in the C_3 *F. pringlei* and phase-conserved among diverse taxonomic groups, the diurnal expression of particular genes most probably changed along C_4 evolution.

IV.A Zusammenfassung

Der Kohlenstoff-Kreislauf der C_4-Photosynthese ist das Endprodukt einer schrittweise erlangten Kombination innovativer Eigenschaften aus präexistenten C_3-Vorläuferpflanzen. Ein entscheidender Schritt war die Evolution des Phospho*enol*pyruvat-Carboxylase-Protein (PEPC)-Enzyms des C_4-Kohlenstoff- Kreislaufs ausgehend von einem nicht-photosynthetischen Vorläufer-C_3-PEPC-Protein. Das C_4-PEPC-Enzym (*ppcA*) ist das Schlüsselenzym des photosynthetischen Kreislaufs, welches CO_2 in den Mesophyllzellen fixiert. Die hohe kinetische und regulatorische Effizienz des C_4-PEPC-Enzyms ermöglicht die Anreicherung von CO_2 unmittelbar um die Ribulose 1,5-biphosphat-Carboxylase/ Oxygenase (Rubisco) des Calvin-Benson Zyklus der Bündelscheidenzellen. Zur Untersuchung der photosynthetischen C_4-Evolution wurde für diese Arbeit die Pflanzengattung *Flaveria* ausgewählt, da diese mit nah verwandten C_3-, C_3-C_4 intermediären- und C_4-Pflanzen ausgestattet ist. Erster Bestandteil dieser Arbeit war es die *in vitro* Untersuchungen der C_3-/C_4-Veränderungen von PEPC mit einem *in planta* Versuch mit transgenen Pflanzen von *F. bidentis* (C_4), mit einem verringerten *ppcA* mRNA Gehalt und einer Inaktivierung des C_4-PEPC-Enzyms, hin zu erweitern.

Die Aktivität des PEPC Enzyms wird von der PEPC-Protein-Kinase (PPCK) reguliert. Ein zweiter Bestandteil dieser Arbeit war die evolutionäre Charakterisierung von PPCK in der Gattung *Flaveria*. Es konnte eine Genfamilie nah verwandter orthologer PPCK-Paare, *PPCKA* und *PPCKB*, innerhalb der jeweiligen *Flaveria*-Art, übergreifend von C_3 bis C_4, identifiziert werden. Die phylogenetische Charakterisierung der PPCK Familie stimmt mit dem polyphyletischen Ursprung der Photosynthese als auch mit der Evolution der PEPC-Familie überein. Innerhalb von *F. trinervia* (C_4) war die allgemeine Transkript-Abundanz von *PPCKA* im Vergleich zu *F. pringlei* (C_3) hin gesteigert, während die von *PPCKB* verringert war. Während die Expression von *PPCKA* beider *F. pringlei* und *F. trinervia* Arten bei Tageslicht hochreguliert war, war die Expression von *PPCKB* in der Nacht am höchsten. Das Expressionsmuster von *PPCKA* verhielt sich in Bezug auf die tagesabhängige *in vivo* Phosphorylierung von PEPC in *F. trinervia*, mit einem Anstieg während des Tages, dementsprechend übereinstimmend. Die Spezifität des rekombinanten PPCKA-Proteins der C_4 *Flaveria* Art gegenüber dem Substrat der C_3 und C_4 *Flaveria* Arten war zwar unterschiedlich, aber nicht aufgehoben. Trotz der hohen Ähnlichkeit der Aminosäurensequenz innerhalb der PPCK-Isoenzyme, legen sowohl der Austausch essentieller Aminosäuren als auch die dargestellten Eigenschaften des *PPCKA*-Gens dar, dass es sich hierbei um das photosynthetische Enzym handeln müsste, welches in C_4 *Flaveria* für die Phosphorylierung des C_4-PEPC-(*ppcA*)-Enzyms verantwortlich ist. Im Vergleich dazu sprechen die Eigenschaften des *PPCKB*-Gens eher für eine zuständige Regulation der *ppcB* and *ppcC* PEPC Isoenzyme in nicht-photosynthetischen Geweben, mit ausgeprägter Funktion in C_3 *Flaveria.* Basierend auf den dargestellten Ergebnissen lässt sich abschließend schlussfolgern, dass die PPCK Familie in *Flaveria* unter evolutionärem Selektionsdruck steht und im Zusammenhang mit der Optimierung einer neuen PPCKA Struktur für die Funktion des C_4-Kohlenstoff-Kreislaufs steht.

Im *Flaveria* Transkriptome wurde bislang weder die Veränderung des tagesabhängigen Verlaufes der Genexpression untersucht, noch wie sich diese im Verlauf der C_4 Evolution verändert hat. In dem dritten Abschnitt dieser Arbeit, wurde das tagesabhängige Transkriptom von *F. pringlei* and *F. trinervia* untersucht um tagesabhängig-regulierte Gene zu identifizieren, welche für Proteine kodieren mit Funktion in unterschiedlichen Stoffwechselwegen. Veränderungen des Transkriptoms wurden im Zusammenhang mit der C_4 Photosynthese und als Antwort auf tagesabhängige Veränderungen hin untersucht. Die globalen Unterschiede zwischen den zwei photosynthetisch unterschiedlichen Arten *F. pringlei* and *F. trinervia* beruhten eher auf Unterschiede in der Transkript Abundanz aller erfassten Gene, als auf die globale Regulation des tagesabhängigen Verlaufes der Transkription. Der Unterschied zwischen der Transkript Abundanz einiger Gene veränderte sich im tagesabhängigen Verlauf zwischen *F. pringlei* and *F. trinervia*. Obwohl die Expression tagesabhängig-regulierter Gene ein Mechanismus ist, welcher bereits in der C_3 *F. pringlei* Art zu finden ist, und zwischen entfernten Arten unterschiedlicher taxonomischer Herkunft konserviert ist, veränderte sich höchstwahrscheinlich für bestimmte Gene die Regulation der tagesabhängigen Expression im Verlauf der C_4-Evolution.

V. References

Agetsuma, M., Furumoto, T., Yanagisawa, S. and **Izui, K.** (2005). The Ubiquitin–Proteasome Pathway is Involved in Rapid Degradation of Phospho*enol*pyruvate Carboxylase Kinase for C_4 Photosynthesis. *Plant Cell Physiol*. **46**: 389–398.

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M. and **Westhoff, P.** (2007). Evolution and function of a cis-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. *The Plant Cell* **19**: 3391–3402 .

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P. and **Kay, S.A.** (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* **293:** 880–883.

Andreo, C.S., Gonzalez, D.H. and **Iglesias, A.A.** (1987). Higher plant phospho*enol*pyruvate carboxylase: structure and regulation. *FEBS Lett* **213**: 1–8.

Andrews, T.J. and **Lorimer, G.H**. (1987). Rubisco: structure, mechanisms, and the prospects for improvement. **In**: *The Biochemistry of Plants*, (MD Hatch, NK Boardman eds.), **10**: 131–218. New York: Academic

Apel, P. and **Maass, I.** (1981). Photosynthesis in species of *Flaveria*. CO_2 compensation concentration, O_2 influence on photosynthetic gas exchange and $\delta\ ^{13}C$ values in species of *Flaveria* (Asteraceae). *Biochem. Physiol.* **176**: 396–399.

Arrio-Dupont, M., Bakrim, N., Echevarría, C., Gadal, P., Le Maréchal, P. and **Vidal, J.** (1992). Compared properties of phospho*enol*pyruvate carboxylase from dark- and light-adapted Sorghum leaves: use of a rapid purification technique by immunochromatography. *Plant Science* **81**: 37–46.

Avasti, U., Izui, K. and **Raghavendra, A.S.** (2011). Interplay of light and temperature during the *in planta* modulation of C_4 phospho*enol*pyruvate carboxylase from the leaves of *Amaranthis hypochondriacus L.*: diurnal and seasonal effects manifested at molecular levels. *J. Ecp. Bot.* **62**: 1017–1026.

Aubry, S, Brown, N.J. and **Hibberd, J.M.** (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *J. Exp. Bot.* **62:** 3049–3059.

Bakrim, N., Echevarría, C., Cretin, C., Arrio-Dupont, M., Pierre, J.-N., Vidal, J., Chollet, R. and **Gadal, P.** (1992). Regulatory phosphorylation of *Sorghum* leaf phospho*enol*pyruvate carboxylase, identification of the protein-serine kinase and some elements of the signal transduction cascade. *Eur. J. Biochem*. **204**: 821-830.

Bakrim, N., Prioul, J.L., Deleens, E., Rocher, J-P., Arrio-Dupont, M., Vidal, J., Gadal, P. and **Chollet, R.** (1993). Regulatory phosphorylation of C_4 phospho*enol*pyruvate carboxylase (A cardinal event influencing the photosynthesis rate in *Sorghum* and maize). *Plant Physiol*. **101**: 891–897

Baskin, J.M. and **Baskin, C.C.** (1976). Effect of photoperiod on germination of *Cyperus inflexus* seeds. *Bot Gaz.* **137:** 269–273.

Bauwe H. (1986). An efficient method for the determination of K_m values for HCO_3^- of phospho*enol*pyruvate carboxylase. *Planta* **169**: 356–360.

Bauwe, H. and **Chollet, R.** (1986). Kinetic Properties of Phospho*enol*pyruvate Carboxylase from C_3, C_4 and C_3-C_4 Intermediate Species of *Flaveria* (Asteraceae). *Plant Physiol*. **82**: 695–699

Bauwe, H., Hagemann, M. and **Fernie, A.R.** (2010). Photorespiration: players, partners and origin. *Trends Plant Sci.* **15:** 330–336

Beerling, D.J. and **Royer, D.L.** (2011). Convergent Cenozoic CO_2 history. *Nature Geoscience* **4**: 418–420.

Björkman, O. (1973). Comparative studies on photosynthesis in higher plants. **In**: *Photophysiology* (A. Giese, ed.), **8**: 1–63. New York: Academic Press

Bläsing, O.E., Westhoff, P. and **Svensson, P.** (2000). Evolution of C_4 phospho*enol*pyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. *J. Biol. Chem.* **275**: 27917–27923.

Bläsing, O.E., Ernst, K., Streubel, M., Westhoff, P. and **Svensson, P.** (2002). The non-photosynthetic phospho*enol*pyruvate carboxylases of the C_4 dicot *Flaveria trinervia* – implications for the evolution of C_4 photosynthesis. *Planta* **215**: 448–456.

Bläsing O.E., Gibon, Y., Günther, M., Höhne, M., Morcuende, R., Osuna, D., Thimm, O., Usadel., B., Scheible, W.-R. and **Sitt, M.** (2005). Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *The Plant Cell* **17**: 3257–3281.

Borland, A.M. and **Griffiths, H.** (1997). A comparative study on the regulation of C_3 and C_4 carboxylation processes in the constitutive crassulacean acid metabolism (CAM) plant *Kalanchoë daigremontiana* and the C_3-CAM intermediate *Clusia minor*. *Planta* **201**: 368–378.

Borland, A.M., Hartwell, J., Jenkins, G.I., Wilkins, M.B. and **Nimmo, H.G.** (1999). Metabolite control overrides circadian regulation of phospho*enol*pyruvate carboxylase kinase and CO_2 fixation in Crassulacean acid metabolism. *Plant Physiol*. **121**: 889–896.

Boutton, T. W., Nordt, L. C., and **Archer, S. R.** (1994). Climate, CO_2, and plant abundance. *Nature* **372:** 625–626.

Bräutigam, A., Kajala, K. Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. W., Carr, K. M., Gowik, U., Maß, J., Lercher, M.J., Westhoff, P., hibberd, J.M. and **Wever, A.P.M.** (2011). An

mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiol*. **155**: 142–156.

Brown, HA. (1999). Agronomic implications of C_4 photosynthesis. **In**: C_4 plant biology. (R.F. Sage, R.K Monson, eds.), San Diego: Academic Press, 473–508.

Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K. and **Hibberd, J.M.** (2011). Independent and parallel recruitment of preexisting mechanisms underlying C_4 photosynthesis. *Science* **331:** 1436–1439

Budde, R.J.A. and **Chollet, R**. (1986). *In vitro* phosphorylation of maize leaf phospho*enol*pyruvate carboxylase. *Plant Physiol.* **82**: 1107–1114.

Carter, P.J., Nimmo, H.G., Fewson, C.A. and **Wilkins, M.B.** (1990) *Bryophyllum fedtschenkoi* protein phosphatase type2A can dephosphorylate phospho*enol*pyruvate carboxylase. *FEBS Lett.* **263**: 233–236

Carter, P.J., Nimmo, H.G., Fewson, C.A. and **Wilkins, M.B.** (1991). Circadian rhythms in the activity of a plant protein kinase. *EMBO J.* **10**: 2063–2068.

Carter, P.J., Wilkins, M.B., Nimmo, H.G. and **Fewson, C.A.** (1995). Effects of temperature on the activity of phospho*enol*pyruvate carboxylase and on the control of CO_2 fixation in *Bryophyllum fedtschenkoi*. *Planta* **196**: 375–380.

Cegelski, L. and **Schaefer, J.** (2006). NMR determination of photorespiration in intact leaves using *in vivo* $^{13}CO_2$ labeling. *J. Magn. Reson.* **178:** 1–10.

Cerling, T.E., Wang, Y. and **Quade, J.** (1993). Expansion of C_4 ecosystems as an indicator of globals ecological change in the late Miocene. *Nature* **361**: 345–345.

Cerling, T.E., Harri, J.M., MacFadden, B.J, Leakey, M.G., Quade, J., Eisenmann, V. and **Ehleringer, J.R.** (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature* **389**: 153–157.

Chapman, K.S.R., Berry, J.A. and **Hatch, M.D.** (1980). Photosynthetic metabolism in the bundle sheath cells of the C_4 species *Zea mays*: sources of ATP and NADPH and the contribution of photosystem II. *Arch. Biochem. Biophys*. **202**: 330–341.

Chastain, C.J. and **Chollet, R.** (1989). Interspecific variation in assimilation of $^{14}CO_2$ into C_4 acids by leaves of C_3, C_4 and C_3-C_4 intermediate *Flaveria* species near the CO_2 compensation concentration. *Planta* **179**: 81–88.

Chattopadhyay, S., Ang, L.H., Puente, P., Deng, X.W. and **Wei, N.** (1998). *Arabidopsis* bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. *Plant Cell* **10**: 673–683.

Chollet, R., Vidal, J. and **O´Leary, M.H.** (1996). Phospho*enol*pyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Annu. Rev. Plant Physiol.* **47**: 273–298.

Christin, P.-A., Salamin, N., Savolainen, V., Duvall, M.R. and **Besnard, G.** (2007). C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol*. **17**:1241–1247.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M.R., Hodkinson, T.R., Savolainen, V. and **Salamin, N.** (2008). Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology* **18**: 37–43.

Christin, P.-A., **Osborne, C.P., Sage**, **R.F., Arakaki, M.** and **Edwards, E.J.** (2011). C_4 eudicots are not younger than C_4 monocots. *Journal of Experimental Botany* **62**: 3171–3181.

Christin, P.-A., Wallace, M. J., Clayton, H., Edwards, E. J., Furbank, R.T., Hattersley, W., Sage, R.F., Macfarlane, T. D. and **Ludwig, M.** (2012). Multiple photosynthetic transitions, polyploidy, and laterla gen transfer in the grass subtribe Neurachninae. *Journal of Exp. Bot.* **63**: 6298–6308.

Christin, P.-A., Boxall, S.F., Gregory, R., Edwards, E.J., Hartwell, J. and **Osborne, C.P.** (2013). Parallel recruitment of multiple genes into C_4 photosynthesis. *Genome Biology and Evolution*.

Christensen, S. and **Silverthorne, J.** (2001). Origin of phytochrome-modulated *Lhcb* mRNA expression in seed plants. *Plant Physiol.* **126**: 1609–1618.

Cockburn, W. (1996). Stomatal mechanism as the basis of the evolution of CAM and C_4 photosynthesis. *Plant Cell Environ.* **6**: 275–279.

Coursol, S., Giglioli-Guivarc'h, N., Vidal, J. and **Pierre, J.-N.** (2000). An increase in phosphoinositide-specific phospholipase C_4 activity precedes induction of C_4 phospho*enol*pyruvate carboxylase phosphorylation in illuminated and NH_4Cl^- treated protoplasts from *Digitaria sanguinalis*. *Plant J.* **23**: 497–506.

Dengler, N.G. and **Nelson, T.** (1999). Leaf structure and development in C_4 plants. **In**: *C_4 plant biology* (R. F. Sage and R. K. Monson eds.), 133–172. Academic Press, San Diego, California, USA.

Dever, L.V., Bailey, K.J, Leegood, R.C. and **Lea, R.J.** (1997). Control of Photosynthesis in *Amaranthus edulis* Mutants with Reduced Amounts of PEP Carboxylase. *Australian Journal of Plant Physiology* **24**: 469–476.

Dissmeyer, N. and **Schnittger, A.** (2011). The age of protein kinases. **In**: *Plant Kinases: Methods and Protocols, Methods in Molecular Biology* (N. Dissmeyer and A. Schnittger [Walker, J.M. eds.]), 7–52. Humana Press/Springer Protocols, 1st Edition.

Dodd, A.N., Salathia, N., Hall, A. Kevei, E. Toth, Nagy, F., Hibberd, J.M., Millar, A.J. and Webb, A.A. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. ***Science* 309**: 630–633

Dong, L.Y., Masuda, T., Kawamura, T., Hata, S. and Izui, K. (1998). Cloning, Expression, and Characterization of a Root-Form Phospho*enol*pyruvate Carboxylase from *Zea mays*: Comparison with the C_4-Form Enzyme. *Plant and Cell Physiology* 39: 865–873.

Dong, L.Y., Ermolova, N. and **Chollet, R.** (2001). Partial purification and biochemical characterization of a heteromeric protein phosphatase 2A holoenzyme from maize (*Zea mays* L.) leaves that dephosphorylates C_4 phospho*enol*pyruvate carboxylase. *Planta* **213**: 379–389.

Du, Z., Aghoram, K. and **Outlaw, W.H. Jr.** (1997). *In vivo* phosphorylation of phospho*enol*pyruvate carboxylase in guard cells of *Vicia faba* L. is enhanced by fusicoccin and suppressed by abscisic acid, *Arch. Biochem. Biophys.* **337**: 345–350.

Duff, S.M.G., Andreo, C.S., Pacquit, V., Lepiniec, L. and **Sarath, G.** (1995). Kinetic analysis of the non-phosphorylated, *in vitro* phosphorylated, and phosphorylation-site-mutant (Asp8) forms of intact recombinant C_4 phospho*enol*pyruvate carboxylase from *Sorghum*. *Eur. J. Biochem.* **228**: 92–95.

Duff, S.M.G. and **Chollet, R.** (1995). *In vivo* regulation of wheat-leaf phospho*enol*pyruvate carboxylase by reversible phosphorylation. *Plant Physiol.* **107**: 775–782.

Duff, S.M.G., Giglioli-Guivarc'h, N., Pierre, J.-N., Vidal, J., Condon, S.A. and **Chollet, R.** (1996). *In situ* evidence for the involvement of calcium and bundle sheath-derived photosynthetic metabolites in the C_4 phospho*enol*pyruvate carboxylase kinase signal transduction chain. *Planta* **199**: 467–474.

Dunlap, J.C. (1999). Molecular Bases for Circadian Clocks. *Cell* **96**: 271–290.

Echevarría, C., Pacquit, V., Bakrim, N., Osuna, L. and **Delgado, B.** (1994). The effect of pH on the covalent and metabolic control of C_4 phospho*enol*pyruvate carboxylase from *Sorghum* leaf. *Arch. Biochem. Biophys.* **315**: 425–30.

Echevarría, C. and **Vidal, J.** (2003). The unique phosphoenolpyruvate carboxylase kinase. *Plant Physiol. Biochem*. **41**: 541–547.

Edwards, G.E. and **Walker, D.A.** (1983). C_3, C_4: Mechanisms, and Cellular and Environmental Regulation of Photosynthesis. Textbook on C_3, C_4 Photosynthesis. Blackwell Scientific, Oxford. pp. 542.

Edwards, G.E. and **Ku, M.S.B.** (1987). Biochemistry of C_3-C_4 intermediates. **In**: *The biochemistry of plants* (P. K. Stumpf and E. E. Conn eds.), Vol. 10, Photosynthesis, 275–325. Academic Press, London, UK.

Edwards, G.E., Franceschi, V.R. and **Voznesenskaya, E.V.** (2004). Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. *Annu. Rev. Plant Biol.* **55:** 173–196.

Edwards, E.J., Osborne, C.P., Stromberg, C.A.E., and **Smith, S.A.** (2010). The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science* **328**: 587–591.

Edwards, E.J. and **Smith, S.A.** (2010). Phylogenetic analyses reveal the shady history of C_4 grasses. *Proc. Natl. Acad. Sci. USA* **107**: 2532–2537.

Ehleringer, J.R. and **Björkman, 0.** (1977). Quantum yields for CO_2 uptake in C_3 and C_4 plants: Dependence on temperature, CO_2 and O_2 concentration. *Plant Physiol.* **59**: 86–90.

Ehleringer, J.R., Sage, R.F., Flanagan, L.B. and **Pearcy, R.W.** (1991). Climate change and the evolution of C_4 photosynthesis. *Trends Ecol. Evol.* **6**: 95–99.

Ehleringer, J.R. and **Monson, R.K.** (1993). Evolutionary and ecological aspects of photosynthetic pathway variation. *Annual Review of Ecology and Systematics* **24**: 411–439.

Ehleringer, J.R., Cerling, T.E. and **Helliker, B.R.** (1997). C_4 photosynthesis, atmospheric CO_2 and climate. *Oecologia* **112**: 285–299.

Ehleringer, J.R. (2005). The influence of atmospheric CO_2, temperature, and water on the abundance of C_3/C_4 taxa. **In**: *A History of Atmospheric CO_2 and Its Effects on Plants, Animals, and Ecosystems*, ed. JR Ehleringer, TE Cerling, MD Dearing **10**: 214–232. New York: Spinger

Engelmann, S., Bläsing, OE. Westhoff, P. and **Svensson P.** (2002). Serine 774 and amino acids 296 to 437 comprise the major C_4 determinants of the C_4 phospho*enol*pyruvate carboxylase of *Flaveria trinervia*. *FEBS Lett.* **524**: 11–14.

Engelmann, S., Bläsing, O.E., Gowik, U., Svensson, P. and **Westhoff, P.** (2003). Molecular evolution of C_4 phospho*enol*pyruvate carboxylase in the genus *Flaveria* – a gradual increase from C_3 to C_4 characteristics. *Planta* **217**: 717–725.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., Streubel, M., Cossu, R., Bauwe, H. and **Westhoff, P.** (2008). The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* C_4 and *Arabidopsis* C_3. *Plant Physiol.* **146**: 1773–1785.

Ernst, K. and **Westhoff, P.** (1997). The phospho*enol*pyruvate carboxylase (*ppc*) gene family of *Flaveria trinervia* (C_4) and *F. pringlei* (C_3): molecular characterization and expression analysis of the *ppcB* and *ppcC* genes. *Plant Mol. Biol.* **34**: 427–443.

Evans, J.R. and **von Caemmerer, S.** (1996). Carbon dioxide diffusion inside leaves. *Plant Physiol.* **110**: 339 – 346.

Fankhauser, C. and **Staiger, D.** (2002) Photoreceptors in *Arabidopsis thaliana*: light perception, signal transduction and entrainment of the endogenous clock. *Planta* **216**: 1–16.

Farquhar, G.D. and **Sharkey, T.D.** (1982). Stomatal Conductance and Photosynthesis. *Annual Review* of *Plant. Physiol.* **33**: 317–345.

Farré, E.M., Stacey, L.H., Harmon, F.G, Yanovsky, M.J. and **Kay, S.A.** (2005). Overlapping and Distinct Roles of *PRR7* and *PRR9* in the *Arabidopsis* Circadian Clock. *Current Biology* **15:** 47–54.

Farré E.M. and **Weise S.E.** (2012). The interactions between the circadian clock and primary metabolism. *Current Opinion in Plant Biology* **15**: 293–300.

Feria, A.-B., Alvarez, R., Cochereau, L., Vidal, J., García-Mauriño, S. and **Echevarría, C.** (2008). Regulation of Phospho*enol*pyruvate Carboxylase phosphorylation by metabolites and abscisic acid during the development and germination of barley seeds. *Plant Physiology* **148**: 761–774.

Filichkin, S.A., Breton, G., Priest, H.D., Dharmawardhana, P., Jaiswal, P., Fox, S.E., Michael, T.P., Chory, J., Kay, S.A. and **Mockler, T.C.** (2011). Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PLoS ONE* **6**:e16907.

Fontaine, V., Hartwell, J., Jenkins, G.I. and **Nimmo, H.G.** (2002). *Arabidopsis thaliana* contains two phospho*enol*pyruvate carboxylase kinase genes with different expression patterns. *Plant Cell Environ*. **25**: 115–122.

Franklin, K.A. and **Whitelam, G.C.** (2007). Light-quality regulation of freezing tolerance in *Arabidopsis thaliana*. *Nat. Genet*. **39**:1410–1413.

Freitag, H. and **Stichler, W.** (2002). *Bienertia cycloptera* Bunge ex Boiss. Chenopodiaceae, another C_4 plant without Kranz tissues. *Plant Biology* **4**: 121–134.

Fukayama, H., Tamai, T., Taniguchi, Y., Sullivan, S., Miyao, M. and **Nimmo, H.G.** (2006). Characterization and functional analysis of phospho*enol*pyruvate carboxylase kinase genes in rice. *The Plant Journal* **47**: 258–268.

Furumoto, T., Izui, K., Quinn, V., Furbank, R.T. and **von Caemmerer, S.** (2007). Phosphorylation of phospho*enol*pyruvate carboxylase is not essential for high photosynthetic rates in the C_4 species *Flaveria bidentis*. *Plant Phys.* **144**: 1936–1945.

Ghannoum, O., Evans, J. and **von Caemmerer, S.** (2011), 'Nitrogen and Water Use Efficiency of C_4 Plants', *C4 Photosynthesis and Related CO2 Concentrating Mechanisms*, Springer 9789048194063.

García-Mauriño, S., Monreal, S.A., Alvarez, R., Vidal, J. and **Echevarría, C.** (2003). Characterization of salt stress-enhanced phospho*enol*pyruvate carboxylase kinase activity in leaves of *Sorghum vulgare*: independence from osmotic stress, involvement of ion toxicity and significance of dark phosphorylation. *Planta* **216**: 648–655.

Gerhart, L.M. and **Ward, J.K.** (2010). Plant responses to low [CO_2] of the past. *New Phytol.* **188**: 674–695.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and **Ohlrogge, J.** (2000). Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* **124**: 1570–1581.

Giuliano, G., Hoffmann, N.E., Ko, K., Scolnik, P.A. and **Cashmore, A.R.** (1988). A light-entrained circadian clock controls transcription of several plant genes. *The EMBO Journal.* **7**: 3635–3642.

Gorton, H.L., Williams, W.E. and **Assmann, S.M.** (1993). Circadian Rhythms in Stomatal Responsiveness to Red and Blue Light. *Plant Physiol.* **103**: 399–406.**Gould, P.D., Locke, J.C, Larue, C., Southern, M.M., Davis, S.J., Hanano, S., Moyle, R. Milich, R., Putterill, J.** and **Millar, A.J.** (2006). The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. *Plant Cell* **18:** 1177–1187.

Gregory, A.l., Hurley, B.A., Tran, H.T., Valentine, A.J., She, Y.-M., Knowles, V.L. and **Plaxton, W.C.** (2009). *In vivo* regulatory phosphorylation of the phospho*enol*pyruvate carboxylase *AtPPC1* in phosphate-starved *Arabidopsis thaliana. Biochem. L.* **420**: 57–65

Grams, T.E.E., Borland, A.E., Roberts, A., Griffiths, H., Beck, F. and **Lüttge, U.** (1997). On the mechanism of reinitiation of endogenous crassulacean acid metabolism rhythm by temperature changes. *Plant Physiol.* **113**: 1309–1317.

Green, R.M., Tingay, S., Wang, Z.Y. and **Tobin, E.M.** (2002). Circadian rhythms confer a higher level of fitness to *Arabidopsis* plants. *Plant Physiol.* **129**: 576–584.

Grob, U. and **Stüber, K.** (1987). Discrimination of phytochrome dependent light inducible from non-light inducible plant genes. Prediction of a common light-responsive element (LRE) in phytochrome dependent light inducible plant genes. *Nucl. Acids Res.* **15:** 9957–9973.

Giglioli-Guivarc'h, N., Pierre, J.-N., Brown, S., Chollet, R., Vidal, J. and **Gadal, P.** (1996). The light-dependent transduction pathway controlling the regulatory phosphorylation of C_4 phospho*enol*pyruvate carboxylase in protoplasts from *Digitaria sanguinalis*. *Plant Cell* **8**: 573–586.

Gousset-Dupont, A., Lebouteiller, B., Monreal, J., Echevarría, C., Pierre, J.-N., Hodges, M., and Vidal, J. (2005). Metabolite and post-translational control of phospho*enol*pyruvate carboxylase from leaves and mesophyll cell protoplasts of *Arabidopsis thaliana*. *Plant Science* **169**: 1096–1101.

Gowik, U. and **Westhoff, P.** (2011). The Path from C_3 to C_4 photosynthesis. *Plant Physiol.* **155**: 56–63.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M. and **Westhoff, P.** (2011). Evolution of C_4 Photosynthesis in the genus *Flaveria*: How Many and Which Genes Does It Take to Make C_4? *The Plant Cell* **23**: 2087–2105.

Halford, N.G. and **Hardie, D.G.** (1998). SNF1-related protein kinases: global regulators of carbon metabolism in plants? *Plant Molecular Biology* **37**: 735–748.

Halford, N.G. Hey, S., jhurreea, D., Laurie, S., McKibbin, R.S., Paul, M. and **Zhang, Y.** (2003). Metabolic signalling and carbon partioning: role of Snf1-related (SnRK1) protein kinases. *Journal of Experimental Botany* **54**: 467–475.

Halliday, K.J., Martinez-García, J. F. and **Josse, E.-M.** (2009). Integration of Light and Auxin Signaling. *Cold Spring Harb. Perspect. Biol.* **00**:a001586

Haberlandt, G. (1914). Physiological plant anatomy. London: Macmillan.

Hanks, S.K. and **Hunter, T.** (1995). The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**: 576–596.

Hardie, D.G. (1999). Plant protein serine/threonine kinases: classification and functions. *Annu Rev Plant Physiol Plant Mol Biol* **50**: 97–131.

Hardie, D.G. (2000). Plant protein-serine/threonine kinases: classification into subfamilies and overview of function. *Adv Bot Res* **32:** 1–44.

Harmer, S.L. (2009). The circadian system in higher plants. *Annu. Rev. Plant Biol.* **60**: 357–377.

Harmon, A.C., Gribskov, M., Gubrium, E. and **Harper, J.F.** (2001). The CDPK superfamily of protein kinases. *New Phytologist* **151**: 175–183.

Harms, E., Kivimae, S., Young, M.W. and **Saez, L.** (2004). Posttranscriptional and posttranslational regulation of clock genes. *J. Biol. Rhythms* **19**: 361–373.

Hartwell, J., Smith, L.H., Wilkins, M.B., Jenkins, G.I. and **Nimmo, H.G.** (1996). Higher plant phospho*enol*pyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. *Plant J.* **10**: 1071–1078.

Hartwell, J., Gill, A., Nimmo, G.A., Wilkins, M.B., Jenkins, G.I. and **Nimmo, H.G.** (1999a). Phospho*enol*pyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *Plant J.* **20:** 333–342.

Hartwell, J., Jenkins, G.I., Wilkins, M.B. and **Nimmo, H.G.** (1999b). The light induction of maize phospho*enol*pyruvate carboxylase kinase translatable mRNA requires transcription but not translation. *Plant Cell Environ.* **22:** 883–889.

Hatch, M.D. and **Slack, C.R.** (1970). The C_4 carboxylic acid pathway of photosynthesis. **In**: *Progress in Phytochemistry* (Reinhold L and Liwschitz Y. eds.), pp: 35–106. Wiley-Interscience, New York.

Hatch, M.D. (1971). Mechanism and function of the C_4 pathway of photosynthesis. **In**: *Photosynthesis and photorespiration* (M.D. Hatch et al., eds.), pp. 139–152. New York: Wiley-Interscience.

Hatch, M.D. and **Osmond, C.B.** (1976). Compartmentation and transport in C_4 photosynthesis. **In**: *Transport in Plants* (U. Heber and C.R. Stocking, eds.), **3**: 145–184. Springer-Verlag, Berlin.

Hatch, M.D. (1987). C_4 photosynthesis: A unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta.* **895**: 81–106.

Hattersley, P.G. (1983). The distribution of C_3 and C_4 grasses in Australia in relation to climate. *Oecologia* **57**: *113–128.*

Hattersley, P.W. (1987). **In**: Grass *Systematics and Evolution* (Soderstrom, T.R., Hilu, K.W., Campbell, C.S. and Barkworth, M.E., eds), pp. 49–69, Smithsonian Institution Press.

Häusler, R.E., Hirsch, H.-J., Kreuzaler, F. and **Peterhänsel, C.** (2002). Overexpression of C_4-cycle enzymes in transgenic C_3 plants: a biotechnological approach to improve C_3-photosynthesis. *J. Exp. Bot.* **53**: 591–607.

Heckmann, D., Schulze, S, Denton, A., Gowik, U, Westhoff, P, Weber, A.P. and **Lercher, M.J.** (2013). Predicting C_4 Photosynthesis Evolution: Modular, individually adaptive steps on a Mount Fuji Fitness Landscape. *Cell* **153**: 1579–1588.

Hermans, J. and **Westhoff, P.** (1990). Analysis of expression and evolutionary relationships of phospho*enol*pyruvate carboxylase genes in *Flaveria trinervia* (C_4) and *F. pringlei* (C_3). *Mol. Gen. Genet.* **224**: 459–468.

Hodge, A.J., McLEan, J.D. and **Mercer, FV.** (1995). Ultrastructure of the lamellae and grana in the chloroplasts of *Zea mays L. Journal of Biophysical and Biochemical Cytology* **25:** 605–614.

Holaday, A.S. and **Bowes, G.** (1980). C_4 acid metabolism and dark CO_2 fixation in a submersed aquatic macrophyte (*Hydrilla verticillata*). *Plant Physiology* **65**: 331–335.

Holaday, A.S. and **Chollet, R.** (1983). Photosynthetic/respiratory carbon metabolism in the C_3-C_4 intermediate species, *Moricandia arvensis* and *Panicum milioides*. *Plant Physiol.* **73**: 740–745.

Holaday, A.S., Lee, K.W. and **Chollet, R.** (1984). C_3-C_4 Intermediate species in the genus *Flaveria*: leaf anatomy, ultrastructure, and the effect of O_2 on the CO_2 compensation concentration. *Planta* **160**: 25–32.

Horst, I., Offermann, S., Dreesen, B., Niessen, M. and **Peterhansel, C.** (2009). Core promoter acetylation is not required for high transcription from the phospho*enol*pyruvate carboxylase promoter in maize. *Epigenetics and Chromation* **2**: 17

Hrabak, E.M., Chan, C.W., Gribskov, M., Harper, J.F., Choi, J.H., Halford, N., Kudla, J., Luan, S., Nimmo, H.G., Sussman, M.R., Thomas, M., Walker-Simmons, K., Zhu, J.K. and **Harmon, A.C.** (2003). The *Arabidopsis* CDPK-SnRK superfamily of protein kinases. *Plant Physiol.* **132**: 666–680.

Huber, S.C. and **Edwards, G.E.** (1975). Regulation of oxaloacetate, aspartate and malate formation in mesophyll protoplast extracts of three types of C_4 plants. *Plant Physiol.* **56**: 324–331.

Huber, S.C. and **Sugiyama, T.** (1986). Changes in sensitivity to effectors of maize leaf phospho*enol*pyruvate carboxylase during light/dark transitions. *Plant Physiology* **81**: 674–677.

Huse, M. and **Kuriyan, J.** (2002). The conformational plasticity of protein kinases. *Cell* **109**: 275–282.

Huxman, T.E. and **Monson, R.K.** (2003). Stomatal responses of C_3, C_3-C_4 and C_4 *Flaveria* species to light and intercellular CO_2 concentration: implications for the evolution of stomatal behavior. *Plant Cell Environ*. **26**: 313–322.

Ivanov, B.N., Asada, K., Kramer, D.M. and **Edwards, G.E.** (2005). Characterization of photosynthetic electron transport in bundle sheath cells of maize. I. Ascorbate effectively stimulates cyclic electron flow around PSI. *Planta* **220**: 572–581.

Jacobs, B., Engelmann, S., Westhoff, P. and **Gowik, U.** (2008). Evolution of C_4 phospho*enol*pyruvate carboxylase in *Flaveria*: determinants for high tolerance towards the inhibitor L-malate. *Plant Cell Environ*. **31**: 793–803.

Jeanneau, M., Vidal, J., Gousset-Dupont, A., Lebouteiller, B., Hodges, M., Gerentes, D. and **Perez, P.** (2002a). Manipulating PEPC levels in plants. *J Exp Bot.* **53**: 1837–1845.

Jeanneau, M., Gerentes, D., Foueillassar, X., Zivy, M., Vidal, J., Toppan, A. and **Perez, P.** (2002b). Improvement of drought tolerance in maize: towards the functional validation of the Zm-Asr1 gene and increase of water use efficiency by over-expressing C_4-PEPC. *Biochimie* **84**: 1127–1135.

Jiao, J.-A. and **Chollet, R.** (1988). Light/dark regulation of maize leaf phospho*enol*pyruvate carboxylase by *in vivo* phosphorylation. *Arch Biochem Biophys.* **261**: 409–417.

Jiao, J.-A. and **Chollet, R.** (1989). Regulatory seryl-phosphorylation of C_4 phospho*enol*pyruvate carboxylase by a soluble protein kinase from maize leaves. *Arch Biochem Biophys.* **269**: 526–535.

Jiao, J.-A. and **Chollet, R**. (1990a). Regulatory phosphorylation of serine-15 in maize phospho*enol*pyruvate carboxylase by a C_4-leaf protein-serine kinase. *Arch Biochem Biophys.* **283**: 300–305.**Jiao, J.-A.** and **Chollet, R**. (1990b). Posttranslational Regulation of Phosphoenolpyruvate Carboxylase in C_4 and Crassulacean Acid Metabolism Plants. *Plant Physiol.* **95**: 981–985.

Jiao, J.-A., Echevarría, C., Vidal, J. and **Chollet, R.** (1991a). Protein turnover as a component in the light dark regulation of phospho*enol*pyruvate carboxylase protein serine kinase activity in C_4 plants. *Proc. Natl. Acad. Sci. USA* **88**: 2712–2715.

Jiao, J.-A, Vidal, J., Echevarría, C. and **Chollet, R.** (1991b). *In vivo* regulatory phosphorylation site in C_4 leaf phospho*enol*pyruvate carboxylase from maize and *Sorghum. Plant Physiol.* **96**: 297–301.

Jiao, J.-A., and **Chollet, R.** (1992). Light activation of maize phospho*enol*pruvate carboxylase protein kinase activity is inhibited by mesophyll and bundle sheath-directed photosynthesis inhibitors. *Plant Physiol.* **98**: 152–156.

Johnson, L.N., Noble, M.E.M. and **Owen, D.J.** (1996). Active and inactive protein kinases: structural basis for regulation. *Cell* **85**: 149–158.

Kadereit, G., Borsch, T., Weising, K. and **Freitag, H.** (2003). Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. *Int. J. Plant Sci*. **164**: 959–986.

Kai, Y. Matsumura, H, Inoue, T., Terada, K., Nagara, Y., Yoshinaga, T., Kihara, A., Tsumura, K. and **Izui, K.** (1999). Three-dimensional structure of phospho*enol*pyruvate carboxylase: A proposed mechanism for allosteric inhibition. *Proc. Natl. Acad. Sci. USA* **96**: 823–828.

Kai, Y., Matsumura, H. and **Izui, K.** (2003). Phospho*enol*pyruvate carboxylase: Three-dimensional structure and molecular mechanisms. *Archives of Biochemistry and Biophysics* **414:** 170–179.

Kapralov, M.V., Kubien, D.S., Andersson, I. and **Filatov, D.A.** (2011). Changes in Rubisco kinetics during the evolution of C_4 photosynthesis in *Flaveria* (*Asteraceae*) are associated with positive selection on genes encoding the enzyme. *Mol. Biol. Evol.* **28**: 1491–1503.

Karabourniotis, B.G., Manetas, Y. and **Gavalas, N.A.** (1985). Detecting photoactivation of phospho*enol*pyruvate carboxylase in C_4 plants. An effect of pH. *Plant Physiol.* **77:** 300–302.

Kawamura, T., Shigesada, K., Yanagisawa, S. and **Izui, K.** (1990). Phospho*enol*pyruvate carboxylase prevalent in maize roots: isolation of a cDNA clone and its use for analyses of the gene and gene expression. *J. Biochem.* **107**: 165–168.

Khan, S., Rowe, S.C. and **Harmon, F.G.** (2010). Coordination of the maize transcriptome by a conserved circadian clock. *BMC Plant Biology* **10**: 126

Kloos, D.U., Oltmanns, H., Dock, C., Stahl, D. and **Hehl, R.** (2002). Isolation and molecular analysis of six taproot expressed genes from sugar beet. *J. Exp. Bot.* **53**: 1533–1534

Kluge, M., Maier, P., Brulfert, J., Faist, K. and **Wollny, E.** (1988). Regulation of phospho*enol*pyruvate carboxylase in Crassulacean acid metabolism: *in vitro* phosphorylation of the enzyme. *J. Plant Physiol.* **133**: 252–256.

Koornneef, M., Alonso-Blanco, C., Peeters, A.J.M. and **Soppe, W.** (1998). Genetic control of flowering time in *Arabidopsis*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**: 345–370.

Knapp, A.K. and **Medina, E.** (1999). Success of C_4 photosynthesis in the Field: Lesson from Communities Dominated by C_4 plants. **In**: *C_4 plant biology* (R. F. Sage and R. K. Monson eds.), 251–277. Academic Press, San Diego, California, USA.

Kopriva, S., Chu, C.C., and **Bauwe, H.** (1996). Molecular phylogeny of *Flaveria* as deduced from the analysis of nucleotide sequences encoding the H-protein of the glycine cleavage system. *Plant Cell and Environment* **19**: 1028–1036.

Ku, M.S.B., Schmitt, M.R. and **Edwards, G.E.** (1979). Quantitative determination of RuBP carboxylase-oxygenase protein in leaves of several C_3 and C_4 plants. *J Exp Bot.* **114:** 89–98.

Ku, M.S.B., Monson, R.K., Littlejohn, R.O., Nakamoto, H., Fisher, D.B. and **Edwards, G.E. (**1983). Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria* species: I. Leaf anatomy, photosynthetic responses to O_2 and CO_2, and activities of key enzymes in the C_3 and C_4 pathways. *Plant Physiol.* **71**: 944–948.

Ku, M.S.B., Wu, J.R., Dai, Z.Y., Scott, R.A., Chu, C. and **Edwards, G.E.** (1991). Photosynthetic and photorespiratory characteristics of *Flaveria* species. *Plant Physiol.* **96**: 518–528.

Kubien, D.S., Whitney, S.M., Moore, P.V. and **Jesson, L.K.** (2008). The biochemistry of Rubisco in *Flaveria*. *J. Exp. Bot.* **59**: 1767–1777.

Kurotani, K., Hata, S. and **Izui, K.** (1999). Light-regulated expression of the gene for C_4-form phospho*enol*pyruvate carboxylase in maize: destabilization of mRNA by okadaic acid. *Plant and Cell Physiology* **40**: 423-430.

Kusumi, K., Arata, H., Iwasaki, I. and **Nishimura, M.** (1994). Regulation of PEP-carboxylase by biological clock in a CAM plant. *Plant Cell Physiol*. **35**: 233–242.

Latzko, E. and **Kelly, J.** (1983). The multi-faceted function of phospho*enol*pyruvate carboxylase in C_3 plants. *Physiol. Vég.* **21**: 805–815. **Lepiniec, L., Keryer, E., Philippe, H., Gadal, P.** and **Crétin, C.** (1993). *Sorghum* phospho*enol*pyruvate carboxylase gene family: Structure, function and molecular evolution. *Plant Mol.Biol.* **21**: 487–502.

Leport, L., Kandlbinder, A., Baur, B. and **Kaiser, W.M.** (1996). Diurnal modulation of phospho*enol*pyruvate carboxylation in pea leaves and roots as related to tissue malate concentrations and to the nitrogen source. *Planta* **198**: 495–501.

Li, B., Zhang, X.Q. and **Chollet, R.** (1996). Phospho*enol*pyruvate carboxylase kinase in tobacco leaves is activated by light in a similar but not identical way as in maize. *Plant Physiol*. **111**: 497–505.

Li, B., Pacquit, V., Jiao, J.A., Duff, S.M.G., Maralihalli, G.B. and Sarath, G., Condon, S.A., Vidal, J. and **Chollet, R.** (1997). Structural requirements for phosphorylation of C4-leaf phosphoenolpyruvate carboxylase by its highly regulated protein-serine kinase. A comparative study with synthetic- peptide substrates and native, mutant target proteins. *Aust. J. Plant Physiol.* **24**: 443–449.

Lloyd, J. and **Farquhar, G.D.** (1994). ^{13}C discrimination during CO_2 assimilation by the terrestrial biosphere. *Oecologia* **99**: 201–215.

Locke, J.C., Kozma-Bognar, L., Gould, P.D., Feher, B., Kevei, E., Nagy, F., Turner, M.S., Hall, A. and **Millar, A.J.** (2006). Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. *Mol. Syst. Biol.* **2**: 59

Ludwig, M. (2011). The molecular evolution of β-carbonic anhydrase in *Flaveria*. *J. Exp. Bot.* **62**:3071–3081.

Mallona, I., Egea-Cortines, M. and **Weiss, J.** (2011). Conserved and Divergent Rhythms of Crassulacean Acid Metabolism-Related and Core Clock Gene Expression in the Cactus *Opuntia ficus-indica*. *Plant Physiology* **156**: 1987–1989.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D. and **Bryant, S.H.** (2013). CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* **41:** D348-352

Marsh, J.T., Sullivan, S., Hartwell, J. and **Nimmo, H.G.** (2003). Structure and expression of phospho*enol*pyruvate carboxylase kinase genes in solanaceae. A novel gene exhibits alternative splicing. *Plant Physiol*. **133**: 2021–2028.

Matsumura, H., Xie, Y., Shirakata, S., Inoue, T., Yoshinaga, T., Ueno, Y., Izui, K. and **Kai, Y.** (2002). Crystal structures of C4 form maize and quaternary complex of *E. coli* phospho*enol*pyruvate carboxylases. *Structure* **10**: 1721–1730.

Matsuoka, M. and **Hata, S.** (1987). Comparative studies of phospho*enol*pyruvate carboxylase from C_3 and C_4 plants. *Plant Physiol.* **85**: 947–951.

Matsuoka, M., Tada, Y., Fujimura, T. and **Kano-Murakami, Y.** (1993). Tissue-specific light-regulated expression directed by the promoter of a C_4 gene, maize pyruvate, orthophosphate dikinase, in a C_3 plant, rice. *Proc. Natl. Acad. Sci. USA* **90**: 9586–9590.

Matsuoka, M., Kyozuka,J., Shlmamoto, K., and Kano-Murakami, Y. (1994). The promoters of two carboxylases in a C_4 plant (maize) direct cell-specific, light-regulated expression in a C_3 plant (rice). *Plant J.* **6**: 311–319.

Matsuoka, M., Furbank, R.T., Fukayama, H., and **Miyao, M.** (2001). Molecular engineering of C_4 photosynthesis. *Annu. Rev. Plant Biol*. **52**: 297–314.

McAllister, C.H., Beatty, P.H. and Good, A.G (2012). Engeneering nitrogen use efficiency crop plants: the current status. *Plant Biotechnology Journal* **10**: 1011–1025.

McClung, C.R. (2006). Plant Circadian Rhythms. *The Plant Cell* **18**: 792–803.

Mckown, A.D., Moncalvo, J.-M. and **Dengler, N.G.** (2005). Phylogeny of *Flaveria* (Asteracea) and interference of C_4 photosynthesis evolution. *American Journal of Botany* **32**: 1911–1928.

McKown, A.D. and **Dengler, N.G.** (2007). Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). *Am. J. Bot.* **94**: 382–399.

Mckown, A.D. and **Dengler, N.G.** (2009). Shifts in leaf vein density through accelerated vein formation in C_4 *Flaveria* (Asteraceae). *Annals of Botany* **104**: 1085–1098.

McNaughton, G.A., MacKintosh, C., Fewson, C.A., Wilkins, M.B. and **Nimmo, H.G.** (1991). Illumination increases the phosphorylation state of maize leaf phospho*enol*pyruvate carboxylase by causing an increase in the activity of a protein kinase. *Biochimica et Biophysica Acta* **1093**: 189–195.

Meimoun, P., Gousset-Dupont, A., Lebouteiller, B., Francoise, A-B., Besin, E., Lelarge, C., Mauve, C., Hodges, M. and **Vidal, J.** (2009). The impact of PEPC phosphorylation on growth and developmenta of *Arabidopsis thaliana*: Molecular and phyisological characterization of PEPC kinase mutants. *FEBS Letters* **583**: 1649–1652.**Meyer, H., Thienel, U.** and **Piechulla, B.** (1989). Molecular characterization of the diurnal/circadian expression of the chlorophyll a/b-binding proteins in leaves of tomato and other dicotyledoneous and monocotyledonous plant species. **180**: 5–15

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R., Priest, H.D. Sullivan, C.M., Givan, S.A., Yanovsky, M., Hong, F., Kay, S.A. and **Chory, J.** (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet.* **4**:e14

Millar, A.J., Carré, I.A., Strayer, C.A., Chua, N.-H. and **Kay, S.A.** (1995). Circadian clock mutants in *Arabidopsis* identified by luciferase imaging. *Science* **267**: 1161–1163.

Millar, A.J. and **Kay, S.A.** (1996). Integration of circadian and phototransduction pathways in the network controlling CAB gene expression in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **93**: 15491–15496.

Millar, A.J. (2004). Input signals to the plant circadian clock. *J. Exp. Bot*. **55**: 277–283.

Mockler, T.C., Guo, H., Yang, H., Duong, H. and **Lin, C.** (1999). Antagonistic actions of *Arabidopsis* cryptochrome and phytochrome in the regulation of floral induction. *Development* **126**: 2073–2082

Monreal, J.A., López-Baena F.J., Vidal, J., Echevarría, C. and **García-Mauriño, S.** (2010). Involvement of phospholipase D and phosphatidic acid in the light-dependent up-regulation of *Sorghum* leaf phospho*enol*pyruvate carboxylase-kinase. *Journal of Experimental Botany* 61: 2819–2827.

Monson, R.K., Teeri, J.A., Ku, M.S.B., Gurevitch, J, Mets, L.J. and **Dudley, S.** (1988). Carbon-isotope discrimination by leaves of *Flaveria* species exhibiting different amounts of C_3-cycle and C_4-cycle co-function. *Planta* **174**: 145–51.

Muhaidat, R., Sage, T.L., Frohlich, M.W., Dengler, N.G., and **Sage, R.F.** (2011). Characterization of C_3-C_4 intermediate species in the genus *Heliotropium L.* (Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant Cell Environ.* **34**: 1723–1736.

Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F. and **Schibler, U.** (2004). Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell* **119**: 693–705.

Nelson, T. and **Dengler, N.G.** (1992). Photosynthetic tissue differentiation in C_4 plants. *Int. J. Plant Sci*. **153**: 93–105.

Nimmo, G.A., Nimmo, H.G., Fewson, C.A. and **Wilkins, M.B.** (1984). Diurnal changes in the properties of phospho*enol*pyruvate carboxylase in *Bryophyllum* leaves: a possible covalent modification. *FEBS Lett.* **178**: 199–203.

Nimmo, G.A., Nimmo, H.G., Hamilton, I.D, Fewson, C.A. and **Wilkins, M.B**. (1986). Purification of the phosphorylated night form and dephosphorylated day form of phospho*enol*pyruvate carboxylase from *Bryophyllum fedtschenkoi. Biochern. J.* **239**: 213–220.

Nimmo, G.A., Wilkins, M.B., Fewson, C.A. and **Nimmo, H.G.** (1987a). Persistent circadian rhythms in the phosphorylation state of phospho*enol*pyruvate carboxylase from *Bryophyllum fedtschenkoi* leaves and in its sensitivity to inhibition by malate. *Planta* **170**: 408–415.

Nimmo, G.A., McNaughton, G.A.L., Fewson, C.A., Wilkins, M.B. and **Nimmo, H.G.** (1987b). Changes in the kinetic properties and phosphorylation state of phospho*enol*pyruvate carboxylase in *Zea mays* leaves in response to light and dark. *FEBS Lett.* **213**: 18–22.

Nimmo, H.G. (2000). The regulation of photosynthesis in CAM plants. *Trends in Plant Scienc*e **5**: 75–80.

Nimmo, G.A., Wilkins, W.B. and **Nimmo, H.G.** (2001) Partial purification and characterization of a protein inhibitor of phospho*enol*pyruvate carboxylase kinase. *Planta* **213**: 250–257.

Nimmo, H.G. (2003). Control of the phosphorylation of phospho*enol*pyruvate carboxylase in higher plants. *Arch. Biochem. Biophys*. **414**: 189–196.

Nusinow, D.A, Helfer, A., Hamilton, E.E., King, J.J., Imaizumi, T., Schultz, T.F., Farré, E.M. and **Kay, S.A.** (2011) The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature* **475**: 398–404.

Ogawa, N., Yabuta, N., Ueno, Y. and **Izui, K**. (1998). Characterization of a Maize Ca^{2+}-Dependent Protein Kinase Phosphorylating Phospho*enol*pyruvate Carboxylase. *Plant Cell Physiol.* **39**: 1010–1019

O'Leary, B., Park, J. and **Plaxton, W.C.** (2011). The remarkable diversity of plant PEPC (phospho*enol*pyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. *Biochem. J.* **436**: 15–34.

Osborne, C.P. and **Beerling, D.J.** (2006). Nature's green revolution: the remarkable evolutionary rise of C_4 plants. *Phil. Trans. R. Soc. B*. **361**: 173–194.

Osborne, C.P. and **Freckleton, R.P.** (2009). Ecological selection pressures for C_4 photosynthesis in the grasses. *Proc. R. Soc. B*. **276**: 1753–1760.

Osuna, L., Pierre, J.-N., Gonzáles, M.-C., Alvarez, R., Cejudo, F.J., Echevarría, C. and **J. Vidal** (1999). Evidence for a slow-turnover form of the Ca^{2+}-independent phospho*enol*pyruvate carboxylase kinase in the aleurone-endosperm tissue of germinating barley seeds. *Plant Physiol*. **119**: 511–520.

Pagani, M., Zachos, J.C, Freeman, K.H., Tipple, B.J. and **Bohaty, S.** (2005). Marked Decline in Atmospheric Carbon Dioxide Concentrations during the Paleogene. *Science* **309**: 600–603.

Paulus, J.K, Schlieper, D. and **Groth, G.** (2013a). Greater efficiency of photosynthetic carbon fixation due to single amino acid substitution. *Nature communication* **4**: 1518.

Paulus, J.K, Niehus, C. and **Groth, G.** (2013b). Evolution of C_4 Phospho*enol*pyruvate Carboxylase – enhanced feedback inhibitor tolerance is determined by a single residue. *Mol Plant.*

Pierre, J.-N., Pacquit, V., Vidal, J. and **Gadal, P.** (1992). Regulatory phosphorylation of phospho*enol*pyruvate carboxylase in protoplasts from *Sorghum* mesophyll cells and the role of pH and Ca^{2+} as possible components of the light-transduction pathway. *Eur J Biochem.* **210:** 531–537.

Poetsch, W., Hermans, J. and **Westhoff, P.** (1991). Multiple cDNAs of phospho*enol*pyruvate carboxylase in the C_4 dicot *Flaveria trinervia*. *FEBS Letters* **292**: 133–136.

Powell, A.M. (1978). Systematics of *Flaveria* (Flaveriinae-Asteraceae). *Annals of the Missouri Botanical Garden* **65**: 590–636.

Pratt, L.H. (1982). Phytochrome: the protein moiety. *Annu. Rev. Plant Physiol.* **54**: 275–307.

Rajagopalan, A.V., Devi, M.T. and **Raghavendra, A.S.** (1993). Patterns of phospho*enol*pyruvate carboxylase activity and cytosolic pH during light activation and dark deactivation in C_3 and C_4 plants. *Photosynthetic Research* **38**: 51–60.

Rao, S., Reiskind, J. and **Bowes, G.** (2006). Light regulation of the photosynthetic phospho*enol*pyruvate carboxylase (PEPC) in *Hydrilla verticillata*. *Plant Cell Phys*. **47**: 1206–1216.

Rivoal, J., Turpin, D. and **Plaxton, W.C.** (2002). *In vitro* phosphorylation of phospho*enol*pyruvate carboxylase from the green alga *Selenastrum minutum*. *Plant Cell Physiol.* **43**: 785–792.

Rondeau, P., Rouch, C. and **Besnard, G.** (2005). NADP-malate dehydrogenase gene evolution in Andropogoneae (Poaceae): gene duplication followed by sub-functionalization. *Ann. Bot.* **96**: 1307–1314.

Sage, R.F., Pearcy, R.W. and **Seemann, J.R.** (1987). The nitrogen use efficiency of C_3 and C_4 plants. III. Leaf nitrogen effects on the activity of carboxylating enzymes in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). *Plant Physiol.* **85:** 355–359.

Sage, R.F., Wedin, D.A. and **Li, M. (**1999). The Biography of C_4 photosynthesis: Patterns and Controlling factors. **In**: *C_4 plant biology* (R. F. Sage and R. K. Monson eds.), 313–356. Academic Press, San Diego, California, USA.

Sage, R.F. (2001). Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. *Plant Biology* **3**: 202–213.

Sage, R.F. (2004). The evolution of C_4 photosynthesis. *New Phytologist* **161**: 341–370.

Sage, R.F. and **McKown, A.D.** (2005). Is C_4 photosynthesis less phenotypically plastic than C_3 photosynthesis? *Journal of Exp. Botany* **57**: 303–317.

Sage, T.L. and **Sage, R.F.** (2009) The Functional Anatomy of Rice Leaves: Implications for Refixation of Photorespiratory CO_2 and Efforts to Engineer C_4 Photosynthesis into Rice. *Plant Cell Physiol.* **50**:756–772.

Sage, R.F., Christin, P.-A. and **Edwards, E.J.** (2011). The C_4 plant lineage of plant Earth. *Journal of Experimental Botany* **62**: 3155–3169.

Sage, R.F., Sage, T.L. and **Kocacinar, F.** (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annu Rev. Plant Biol*. **63**: 19–47

Sage, R.F. (2013). Photorespiratory compensation: a driver for biological diversity. *Plant Biology* **15**: 624–638

Salomé, P.A., Michael, T.P., Kearns, E.V., Fett-Neto, A.G., Sharrock, R.A. and **McClung, C.R.** (2002). The *out of phase 1* Mutant Defines a Role for PHYB in Circadian Phase control in *Arabidopsis*. *Plant Physiol*. **129**: 1674–1685.

Salomé, P.A., Xie, Q. and **McClung, C.R.** (2008). Circadian timekeeping during early *Arabidopsis* development. *Plant Physiol*. **147:** 1110–1125.

Saze, H., Ueno, Y., Hisabori, T., Hayashi, H. and **Izui, K.** (2001). Thioredoxin-mediated reductive activation of a protein kinase for the regulatory phosphorylation of C_4-form phospho*enol*pyruvate carboxylase from maize. *Plant Cell Physiol*. **42**: 1295–1302. **Schaffer, R., Ramsay, N., Samach, A., Cordon, S., Putterill, J., Carré, I.A.** and **Coupland, G.** (1998). The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**: 1219–1229.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M. and **Wisman, E.** (2001). Microarray Analysis of Diurnal and Circadian-Regulated Genes in *Arabidopsis*. *The Plant Cell* **13**: 113–123.

Scheible, W.-R., Krapp, A. and **Stitt, M.** (2000). Reciprocal diurnal changes of phosphoenolpyruvate carboxylase expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves. *Plant, Cell and Environment* **23**: 1155–1167.

Seemann, J.R., Sharkey, T.D., Wang, J. and **Osmond, C.B.** (1987). Environmental Effects on Photosynthesis, Nitrogen-Use Efficiency, and Metabolite Pools in Leaves of Sun and Shade Plants. *Plant physiology* **84**: 796–802.

Sharkey, T.D. (1985). Photosynthesis in Intact Leaves of C_3 plants: Physics, Physiological and Rate Limitations. *The Botanical Review*. **51**: 53–105.

Sharkey, T.D. (1988). Estimating the rate of photorespiration in leaves. *Physiol. Plant* **73**: 147–152.

Shenton, M., Fontaine, V., Hartwell, J., Marsh J.T., Jenkins, G.I. and **Nimmo, H.G.** (2006). Distinct patterns of control and expression amongst members of the PEP carboxylase kinase gene family in C_4 plants. *The Plant Journal* **48**: 45–53.

Sheen, J.-Y. and **Bogorad, L.** (1985). Differential expression of the ribulose bisphosphate carboxylase large subunit gene in bundle sheath and mesophyll cells of developing maize leaves is influenced by light. *Plant Physiol*. **79**: 1072–1076.

Sheen, J. (1999). C_4 gene expression. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**: 187–217.

Shu, G., Pontieri, V., Dengler, N.G. and **Mets, L.J.** (1999). Light Induction of Cell Type Differentiation and Cell-Type-Specific Gene Expression in Cotyledons of a C_4 plant, *Flaveria trinervia*. *Plant Physiol.* **121**: 731–741.

Sinha, N.R. and **Kellogg, E.A.** (1996). Parallelism and diversity in multiple origins of C_4 photosynthesis in grasses. *American Journal of Botany* **83**: 1458–1470.

Slack, C.R. and **Hatch, M.D.** (1967). Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochem. J.* **103**: 660–665.

Smith, A.M. and **Stitt, M.** (2007). Coordination of carbon supply and plant growth. *Plant, Cell and Environment* **30**: 1126–1149.

Smith, L.H., Lillo, C., Nimmo, H.G. and **Wilkins, M.B.** (1996). Light regulation of phospho*enol*pyruvate carboxylase in barley mesophyll protoplasts is modulated by protein synthesis and calcium and not necessarily correlated with phospho*enol*pyruvate carboxylase kinase activity. *Planta*. **200**: 174–180.

Smith, L.H., Langdale, J.A. and **Chollet R.** (1998). A Functional Calvin Cycle Is Not Indispensable for the Light Activation of C_4 Phospho*enol*pyruvate Carboxylase Kinase and Its Target Enzyme in the Maize Mutant *bundle sheath defective2-mutable1*. *Plant Phys.* **118**: 191–197.

Somers, D.E., Devlin, P.F. and Kay, S.A. (1998). Phytochrome and Cryptochrome in the Entrainment of the *Arabidopsis* Circadian Clock. *Science* **282**: 1488–1491

Street-Perrott, F.A., Huang, Y., Perrott, R.A., Eglinto, G., Barker, P., Khelifa, L.B., Harkness, D.D. and **Olago, D.O.** (1994). Impact of Lower Atmospheric Carbon Dioxide on Tropical Mountain Ecosystems. *Science* **278**: 1422–1426.

Stock, W.D., Chuba, D.K. and **Verboom, G.A.** (2004). Distribution of South African C_3 and C_4 species of Cyperaceae in relation to climate and phylogeny. *Austral Ecology* **29**: 313–319.

Stockhaus, J., Poetsch, W,, Steinmuller, K. and **Westhoff, P.** (1995). Evolution of the C_4 phospho*enol*pyruvate carboxylase promoter of the C_4 dicot *Flaveria trinervia*: An expression analysis in the C_3 plant tobacco. *MoI. Gen. Genet.* **245:** 286–293.

Stockhaus, J., Schlue, U., Koczor, M., Chitty, J.A., Taylor, W.C. and **Westhoff, P.** (1997) The promoter of the gene encoding the C_4 form of phospho*enol*pyruvate carboxylase directs mesophyll specific expression in transgenic C_4 *Flaveria* spp. *The Plant Cell* **9:** 479–489

Sullivan, S., Jenkins, G.I. and **Nimmo, H.G.** (2004). Roots, cycles and leaves. Expression of the phospho*enol*pyruvate carboxylase kinase gene family in soybean. *Plant Physiol*. **135**: 1–10.

Sullivan, S., Shenton, M. and **Nimmo, H.G.** (2005). Organ specificity and the circadian control of plant gene expression. *Biochem Soc Trans*. **33**: 943–944.

Svensson, P., Bläsing, O.E. and **Westhoff P.** (1997). Evolution of the enzymatic characteristics of C_4, phosphoenolpyruvatecarboxylase. A comparison of the orthologous *ppcA* phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3). *Eur. J. Biochem*. **246**: 452–460.

Takahashi-Terada, A., Kotera, M., Ohshima, K., Furumoto, T., Matsumura, H., Kai, Y. and **Izui, K**. (2005). Maize Phospho*enol*pyruvate carboxylase. *The Journal of biological chemistry* **12:** 11798–11806.

Taybi, T., Patil, S., Chollet, R. and **Cushman, J.C.** (2000). A Minimal Serine/Threonine Protein Kinase Circadianly Regulates Phospho*enol*pyruvate Carboxylase Activity in Crassulacean Acid Metabolism-Induced Leaves of the Common Ice Plant. *Plant Physiology* **123**: 1471–1481.

Taybi, T., Nimmo, H.G. and **Borland, A.M.** (2004). Expression of Phospho*enol*pyruvate Carboxylase and Phospho*enol*pyruvate Carboxylase Kinase Genes. Implications for Genotypic Capacity and Penotypic Capacity and Phenotypic Plasticity in the Expression of Crassulacean Acid Metabolism. *Plant Physiology* **135**: 587–598

Taylor, W.C., Rosche, E., Marshall, J.S., Ali, S., Chastain, C.J. and **Chitty. J.A.** (1997). Diverse mechanisms regulate the expression of genes coding for C_4 enzymes. *Aust. J. Plant Physiol*. **24**: 437–442.

Terada, K., Kai, T., Okuno, S., Fujisawa, H. and **Izui, K.** (1990). Maize leaf phospho*enol*pyruvate carboxylase: phosphorylation of Ser15 with a mammalian cyclic AMP-dependent protein kinase diminishes sensitivity to inhibition by malate. *FEBS Lett.* **259**: 241–244.

Thain, S.C., Hall, A. and **Millar, A.J.** (2000). Functional independence of circadian clocks that regulate plant gene expression. *Current Biology* **10**: 951–956.

Theng, V., Agarie, S. and **Nose, A.** (2008). Regulatory phosphorylation of phospho*enol*pyruvate carboxylase in the leaves of *Kalanchoë pinnata, K. daigremontiana* and *Ananas comosus*. *Journal Biol. Plant.* **52**: 281–29.

Toh, H., Kawamura, T., and **Izui, K.** (1994). Molecular evolution of phospho*enol*pyruvate carboxylase. *Plant Cell Environ.* **17**: 31–43.

Ting, I.P. and **Osmond, C.B.** (1973). Photosynthetic phospho*enol*pyruvate carboxylase. Characteristics of allozymes from leaves of C_3 and C_4 plants. *Plant Physiology* **51**: 439–447.

Tovar-Méndez, A., Rodríguez-Sotres, R., López-Valentín, D. and **Munoz-Clares, R.A.** (1998). Re-examination of the roles of PEP and Mg^{2+} in the reaction catalysed by the phosphorylated and non-phosphorylated forms of phospho*enol*pyruvate carboxylase from leaves in *Zea mays*: effects of the activators glucose-6-phosphate and glycine. *Biochem J.* **332**: 633–642.

Tovar-Méndez, A., Mújica-Jiménez, C. and **Munoz-Clares, R.A.** (2000). Physiological implications of the kinetics of maize leaf phospho*enol*pyruvate carboxylase. *Plant Physiology* **123**: 149–160.

Trevanion, S. J., Furbank, R.T. and **Ashton, A.R.** (1997). NADP-Malate Dehydrogenase in the C_4 Plant *Flaveria bidentis*. Cosense Suppression of Activity in Mesophyll and Bundle-Sheath Cells and Consequences for Photosynthesis. *Plant Physio*. **113**: 1153–1165.

Tsuchida, Y., Furumoto, T., Izumida, A., Hata, S. and **Izui, K.** (2001). Phospho*enol*pyruvate carboxylase kinase involved in C_4 photosynthesis in *Flaveria trinervia*: cDNA cloning and characterization. *FEBS Lett*. **507**: 318–322.

Tyagi, A.K. and **Gaur, T.** (2010) Light Regulation of Nuclear Photosynthetic Genes in Higher Plants. *Critical Reviews in Plant Sciences* **22**: 417–452.

Ubersax, J.A. and **Ferrell Jr., J.E.** (2007). Mechanisms of specificity in protein phosphorylation. *Nature* **8**: 530–541.

Uedan, K. and **Sugiyama, T.** (1976). Purification and Characterization of Phospho*enol*pyruvate Carboxylase from Maize Leaves. *Plant Physiology* **57**: 906–910.

Ueno, Y., Imanari, E., Emura, J., Yoshizawa-Kumagaye, K., Nakajima, K.,Inami, K., Shiba, T., Sakakibara, H., Sugiyama, T. and **Izui, K.** (2000). Immunological analysis of the phosphorylation state of maize C_4-form phospho*enol*pyruvate carboxylase with specific antibodies raised against a synthetic phosphorylated peptide. *Plant J*. **21**: 17–26.

Van Bel A.J.E. (1993). Strategies of phloem loading. Annual Review of Plant Physiology and Plant Molecular Biology **44**: 253–281.

Van Gelder, R.N., Herzog, E.D., Schwartz W.J. and **Taghert, P.H.** (2003). Circadian Rhythms: In the Loop at Last. *Science* **300**: 1534–1535.

Von Caemmerer, S. and **Furbank, R.T.** (2003). The C_4 pathway: an efficient CO_2 pump. *Photosynthesis Research* **77**: 191–207.

Vidal, J. and **Chollet, R**. (1997). Regulatory phosphorylation of C_4 PEP carboxylase. *Trends in Plant Science* **2**: 230–237.

Vidal, J., Bakrim, N. and **Hodges, M.** (2002). The regulation of plant phospho*enol*pyruvate carboxylase by reversible phosphorylation. *Advances in Photosynthesis* **12**: 135–150.

Wang, R., Guegler, K., LaBrie, S.T. and **Crawford, N.M.** (2000). Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell* **12**: 1491–1510.**Wang, Y.H., Duff, S.M., Lepiniec, L., Crétin, C., Sarath, G., Condon, S.A., Vidal, J., Gadal, P.** and **Chollet, R.** (1992). Site-directed mutagenesis of the phosphorylatable serine (Ser8) in C_4 phospho*enol*pyruvate carboxylase from *sorghum*. The effect of negative charge at position 8. *J. Biol. Chem.* **267**: 16759–16762.

Wang, Y.H. and **Chollet, R.** (1993). Partial purification and characterization of phospho*enol*pyruvate carboxylase protein-serine kinase from illuminated maize leaves, *Arch. Biochem. Biophys.* **304**: 496–502.

Wang, X.Y., Gowik, U., Tang, H.B., Bowers, J.E., Westhoff, P. and **Paterson, A.H.** (2009). Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses. *Genome Biol.* **10**: R68.

Wang, Z.Y. and **Tobin, E.M.** (1998). Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**: 1207–1217.

Ward, J.K., Myers, D.A. and **Thomas, R.B.** (2008). Physiological and growth responses of C_3 and C_4 plants to reduced temperature when grown at low CO_2 of the last ice age. *Journal of Integrative Plant Biology* **50**: 1388–1395.

Webb, A.A.R. (2003). The physiology of circadian rhythms in plants. *New Phytologia* **160**: 281–303

Wedding, R.T., Rustin, P., Meyer, C.R. and **Black, M.K.** (1988). Kinetic studies of the form of substrate bound by phospho*enol*pyruvate carboxylase. *Plant Physiol.* **88**: 976–979.

Wedding, R.T., Black, M.K., and **Meyer, C.R.** (1990). Inhibition of phospho*enol*pyruvate carboxylase by malate. *Plant Physiol.* **92**: 456–461.

Wessinger, M.E., Edwards, G.E. and **Ku, M.S.B.** (1989). Quantity and kinetic properties of ribulose 1,5-bisphosphate carboxylase in C_3, C_4, and C_3-C_4 intermediate species of *Flaveria* (Asteraceae). *Plant Cell Physiol.* **30**: 665–671.

Westhoff, P., Svensson, P., Ernst, K., Bläsing, O., Burscheidt, J. and **Stockhaus, J.** (1997). Molecular evolution of C_4 phospho*enol*pyruvate carboxylase in the genus *Flaveria*. *Aust J Plant Physiol.* **24**: 429–436.

Westhoff, P. and **Gowik, U.** (2004). Evolution of C_4 phospho*enol*pyruvate carboxylase. Genes and proteins: a case study with the genus *Flaveria*. *Ann. Bot.* **93:** 13–23.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., Bauwe, H. and **Westhoff, P.** (2012). Regulation of the photorespiratory GLDPA gene in C(4) *Flaveria*: an intricate interplay of transcriptional and posttranscriptional processes. *The Plant Cell* **24**: 137–151.

Yakir, E., Hilman, D., Harir, Y. and **Green, R.M.** (2007). Regulation of output from the plant circadian clock. *FEBS Journal* **274:** 335–345.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. and **Billups, K.** (2001). Trends, rhythms and aberrations in global climate 65 Ma to present. *Science* **292**: 686–693.

Zhang, X. and **Chollet, R.** (1997). Phospho*enol*pyruvate carboxylase protein kinase from soybean root nodules: partial purification, characterization, and up/down-regulation by photosynthate supply from the shoots. *Arch. Biochem. Biophys*. **343**: 260–268.

Zhao, J., Huang, X., Chen W., Du, A., **Zhu, L., Wang, S., Deng X.W,** and **Li, S.** (2012). OsELF3-1, an Ortholog #of *Arabidopsis* EARLY FLOWERING 3, Regulates Rice Circadian Rhythm and Photoperiodic Flowering. *PLoS ONE* **7:** e43705

Chapter 1

In planta inactivation of the C_4 phospho*enol*pyruvate carboxylase (PEPC) gene of *Flaveria bidentis*

INTRODUCTION

Phospho*enol*pyruvate carboxylase (PEPC) is the entry enzyme of the C_4 photosynthetic pathway. According to the origin of the C_4 syndrome, multiple times and independently within clades (Christin et al., 2011; Sage et al., 2012), the *ppcA* gene encoding for the C_4-type PEPC evolved from a non-photosynthetic *ppc* gene (encoding the collectively called C_3-type PEPC) (Westhoff and Gowik, 2004). An early step during C_4 evolution was the recruitment of the C_4-type PEPC to the mesophyll cells that involved changes in expression of *ppcA* with regard to the non-photosynthetic *ppcB* and *ppcC* (Ernst and Westhoff, 1996). The ultimate adaptive step towards C_4 photosynthesis considers changes in PEPC kinetics and regulation, which is strong selected and the fixation short when compared to the early step (Sage et al., 2012; Heckmann et al., 2013).

The aim was to inactivate the C_4-type PEPC and to reveal effects of C_3-to-C_4 changes *in planta* using *Flaveria bidentis* (C_4). Conditional *knockout F. bidentis* were generated with integral artificial micro RNAs (amiRNAs). AmiRNAs are small RNAs that decrease the level of specifically targeted RNAs by base pairing to near-perfect amiRNA-complementary RNA strands (Schwab et al., 2005/06; Ossowski et al., 2008). AmiRNAs were directed against the 5'- and/or 3‘-untranslated regions of the *ppcA* PEPC mRNA from *F. bidentis*, predicted to target the endogenous mRNA *in planta*. As a complementary approach, the aim of a future project was to concomitantly introduce a *ppc* gene with C_3-like characteristics into the transgenic *F. bidentis*, that would not be targeted by the amiRNA.

The partial to complete decrease of the photosynthetic PEPC protein amount would under ambient CO_2 levels cause the impediment of the primary CO_2 fixation *in planta*, and prevent the regeneration of transgenic *F. bidentis* plantlets in tissue culture. To circumvent this problem the expression of amiRNA in transgenic *F. bidentis* was performed with an inducible system.

This chemical inducible system consisted of the Cre/*loxP* DNA excision (CLX) [**Figure 1**]. It combines a β-estradiol-mediated activation of homologous recombination with the expression of the gene of interest (Zuo et al., 2000a/01; Guo et al., 2003). CLX-mediated recombination is induced by the β-estradiol receptor-based transactivator XVE, causing the activation of the Cre-recombinase and the site-specific excision of DNA (CLX-lost) resulting in a rejoined fragment. This brings the genome integrated amiRNA under the control of the upstream located constitutive G10-90 promoter, causing its expression. In principal, the CLX-mediated system allows for the DNA recombination event in the vegetative propagation line of the T_0 generation (Zuo et al., 2000a). It is also applicable to germline cells giving rise to the progeny generations.

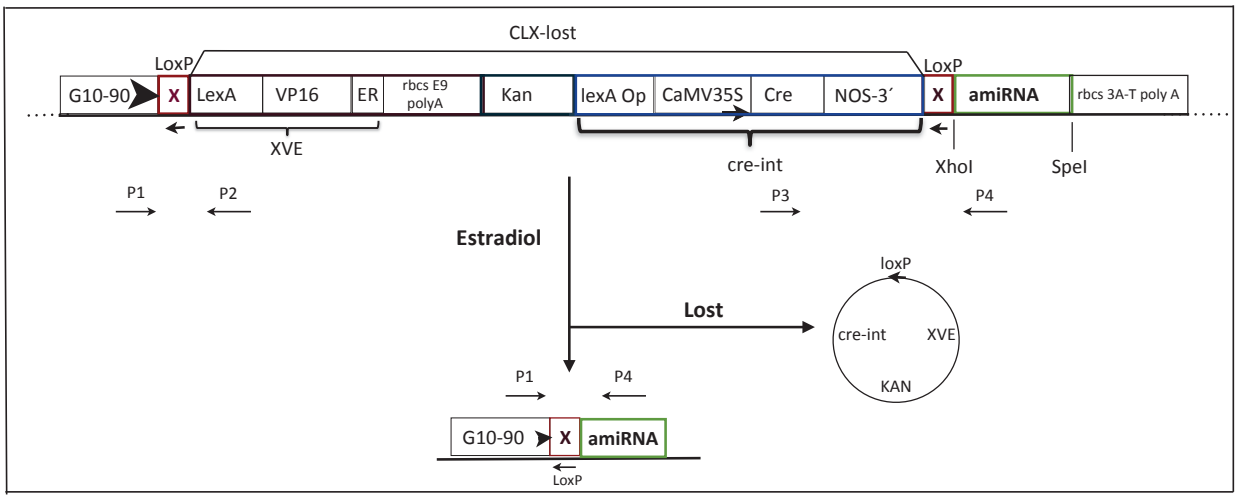


Figure 1: **Diagram of the Cre/*loxP* (CLX) plasmid.** In the original CLX pX6-GFP vector, GFP functions as reporter gene located at the multiple cloning site. GFP was replaced by the amiRNA against the untranslated regions of *ppcA* from *F. bidentis* in the new pX6-amiR-*ppcA* plasmid. The estrogen receptor ligand-binding domain is an integral part of the XVE transcription factor. XVE encodes a transcriptional activator and relies on three transcription units: the DNA-binding domain (DBD) of the bacterial repressor LexA, the acidic transactivating domain of VP16 and the carboxyl region of the human estrogen receptor (ER) (Zuo et al., 2000a). Upon ß-estradiol-induction, the Cre-recombinase recognizes repeated LoxP sites flanked at two boarders of the DNA transcription units of CLX6 (Zuo et al., 2001). The flanked DNA to the loxP sites subunits are removed. DNA recombination results in a rejoined fragment that brings the amiRNA under the control of the upstream located constitutive G10-90 promoter (Ishige et al., 1999). A kanamycin resistance cassette for the plant transformation consists of the nopaline synthase (NOS) gene promoter, the coding sequence of the neomycin transferase II (NPTII) and the *NOS* polyadenylation sequence. The direction of transcription is indicated by arrows. P1 through P4 are the primers used for the molecular characterization of the transgenic plants (Table S1). P2 and P3 are designed to specifically bind to excised and P1 and P4 to flanked non-excised sequences of the pX6-amiR-ppcA plasmid. Adapted from Zuo et al., 2001.

RESULTS

1. Identification of the full-length complementary DNA of ppcA from F. bidentis

The leaf-expressed full-length complementary DNA of the coding region and the 5´-/3´-untranslated regions (UTR) of *ppcA* from *F. bidentis* were defined. The *ppcA* sequences from *F. bidentis* were isolated by the use of Rapid Amplification of 5´ and 3´ Complementary DNA of Polymerase Chain Reactions (5´and 3´ RACE-PCR) (Frohman et al., 1998; Zhu et al., 2001). The 3´-UTR of *ppcA* was identified using gene-specific primers obtained from the transcriptome assembly data set of *F. bidentis* (Gowik et al., 2011). The potentiality of more than a single *ppcA* isogene within *F, bidentis* was analyzed, to ensure that the amiRNA target all possible endogenous *ppcA* copies effectively. Two 5´-UTR sequences of *ppcA* from *F. bidentis*, *ppcA1* (EMBL/ GenBank data libraries AY297087) (Gowik et al., 2004) and a *ppcA2* putative second sequence (Diploma thesis: J. Mallmann) were considered. A single copy of the 5´-/3´-UTR and CDS of *ppcA* from *F. bidentis* were identified [**Figure S1**].

2. *Design of amiRNA precursors*

For the elaboration of the amiRNA constructs the cleavage sites within the RNA of the predicted *ppcA* target from C_4 *F. bidentis* were retrieved. In total four amiRNAs were directed to the 5´- and/or 3´-UTR of *ppcA* from *F. bidentis*. The sequences of the UTR of the *ppc* genes have a higher nucleotide variation relative to the coding sequences, and allows for the amiRNA to specifically target exclusively the C_4 *ppcA* PEPC mRNA. The sequences of 5´- and 3´-UTR of *ppcA* from *F. bidentis* were used as input for the design of amiRNAs according to the parameters of the WMD design tool (http://wmd2.weigelworld.org). The selection of two amiRNAs against each UTR was based on the criteria for the functionality of amiRNAs (Schwab et al., 2006). The position of the amiRNAs in the target sequence (5´- and 3´-UTR *FbppcA*) are illustrated in **Figure 2**. Overlapping PCR reactions were performed for the design of functional sense and antisense amiRNAs ready for cloning [**Figure S2**]. The amiRNA predicted to target either one of the 5´-/3´-UTR of *FbppcA* was cloned as one sense and antisense DNA fragment into the multiple cloning site of the chemical-inducible CLX plasmid pX6 (AF33063), and designated pX6-amiR-*ppcA* (Zuo et al., 2000a/b). The pX6-amiR-ppcA-1 and pX6-amiR-ppcA-2 correspond to the plasmids containing the amiRNA against the 3´UTR (amiR-3´UTR-ppcA_I) and (amiR-3´UTR-ppcA_II), respectively. The pX6-amiR-ppcA-3 and pX6-amiR-ppcA-4 correspond to the plasmids containing the amiRNA against the 5´UTR (amiR-5´UTR-ppcA_I) and (amiR-5´UTR-ppcA_II), respectively.

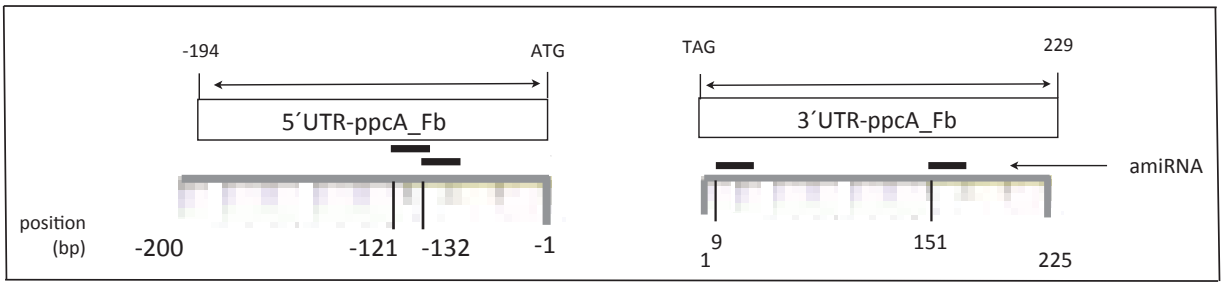


Figure 2: Schematic diagram representing the amiRNAs and predicted target sequence. The 5´UTR- and 3´UTR *FbppcA* isolated by RACE-PCR correspond to 194 bp and 229 bp, respectively (Figure S1). Each sequence was used as input for the WMD design tool. Two amiRNAs (black box) of 21 bp were choosen against the 5´UTR and 3´UTR (Table S1). The amiR-5´UTR-ppcA_I (starts at -121 bp); amiR-5´UTR-ppcA_II (starts at -132 bp); amiR-3´UTR-ppcA_I (starts at 9 bp); amiR-3´UTR-ppcA_II (starts at 151 bp) (Table S2).

3. Efficiency of the CLX-lost excision in Nicotiana bentamiana

We considered to evaluate the efficiency of the original Cre/*loxP*-mediated system in a preliminary test trial. Therefore, the CLX plasmid pX6-GFP (Zuo et al., 2000a/01) containing the gene encoding for the green fluorescence protein (GFP) was transiently transformed into *N. benthamiana* leaves. GFP served as reporter gene, downstream located to the estrogen receptor-based transactivator XVE and the *cre* transcription untis [**Figure 1**], so that upon ß-estradiol-induction GFP would become activated and the gene transiently expressed. The recombination of the CLX system was considered functional revealed by the estradiol-induced expression of GFP.

N. benthamiana leaves were transienly transformed by infiltration with the *Agrobacterium tumefaciens*-mediated technique with the pX6-GFP construct. Prior analysis, transformed *N. benthamiana* leaves were externally treated with a solution containing the inducer ß-estradiol, and subsequently along 24 hours to 3 days of infiltration the leaves were collected and analysed. The pattern of translated GFP accumulation was assessed with leave sections from *N. benthamiana* using an upright fluorescence microscope.

Several ß-estradiol-treated leaf sections from independent transformed *N. benthamiana* plants revealed green fluorescence due to the accumulation of GFP after 3 days of infiltration [**Figure 3A**]. Whereas, the control, non-treated transformed *N. benthamiana* leaves showed no signal of GFP translated protein [**Figure 3C**]. GFP was localized in small aggregates and not uniformly distributed along the leaf sections. Due to a non-uniform GFP accumulation, the frequency and efficiency was considered low, at least in transiently transformed *N. benthamiana*.

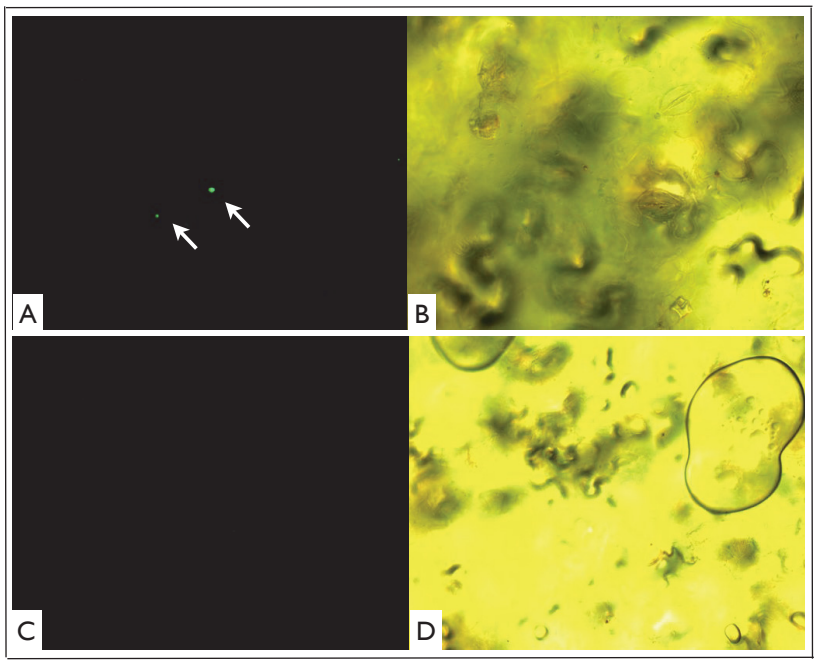


Figure 3: GFP accumulation in estradiol-treated transformed *Nicotiana benthamiana leaves.* *N. benthamiana* leaves were transformed by *A. tumefaciens* infiltration carrying the CLX plasmid pX6-GFP (Zuo et al., 2000a/01). The pX6-GFP contained the GFP-coding sequence as reporter gene. Green fluorescene in (**A**) represents GFP fluorescence in the cells of inducer-treated transformed *N. benthamiana* leaf disc. (**B**) Cells of non-treated transformed *N. benthamiana* leaf disc (**B**) and (**D**) are the negative non-fluorescence images of the corresponding fluorescence images (**A**) and (**C**), respectively.

4. Functionality of the amiRNA with respect to its target in N. benthamiana

An additional preliminary trial considered to test for the functionality of active amiRNA. The binding of amiRNA to the complementary site and the conditional cleavage of the target was functional with the Cre/*loxP*-mediated system in *N. benthamiana* plants, revealed with *gusA* encoding the glucoronidase (GUS) as reporter gene.

N. benthamiana leaves were co-transformated with the target and the amiRNA. Each of the four pX6-amiR-ppcA construct was introduced separately into *N. benthamiana*. The transformants were co-infilrated with the 5´/3´UTR *FbppcA* encoding the target (as part of the ppcA-L-Ft-3UTR plasmid) [**Figure 4**]. GUS was placed under the control of a strong constitutive 35S-promoter in the ppcA-L-Ft-3UTR. In this system it was predicted that the expression of GUS would become prevented upon degradation of the target gene in the inducer-treated transgenic *N. benthamiana* plants. Hence, the absence of GUS would correlate with the functionality of the amiRNA-mediated cleavage of the target in *N. benthamiana*.

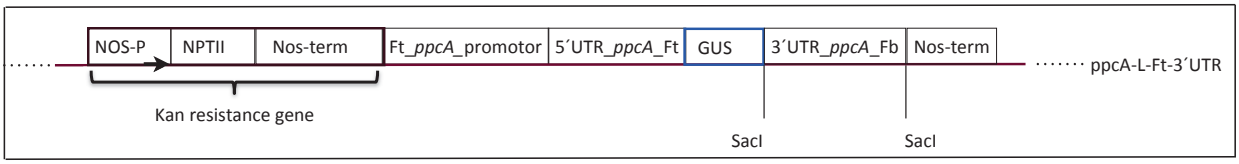


Figure 4: **Schematic representation of the reconstructed ppcA-L-Ft-3UTR plasmid.** The full-length of the 3´flanking sequence of the *ppcA* from *F. bidentis* was placed downstream of the GUS and upstream of the Nos-terminator sequence of the ppcA-L-Ft plasmid (Stockhaus et al., 1994). The ppcA-L-Ft plasmid contains the promotor of from *F. trinervia* (C4), including the 5´UTR from Ft and the 3´UTR *FbppcA*. The *gusA* encoding GUS protein was used as reporter gene, placed between the UTR sequences. The ppcA-L-Ft-3UTR fusion construct was used for the transient transformation *N. benthamiana* to test the functionality of amiRNA.

The transiently transformed *N. benthamiana* leaves were treated with the inducer and the intact leaves collected as described under section 3. The pattern of translated GUS accumulation was then analyzed with an histochemical assay in intact leaves. The co-transformed and inducer-treated *N. benthamiana* leaves had a reduced accumulation of GUS when compared to the controls [**Figure 5**]. The control corresponded to inducer-treated *N. benthamiana* leaves expressing the target, but not the amiRNA. *N. benthamiana* were transformed with different ratios between the amiRNA and target. In the co-transformed inducer-treated *N. benthamiana* leaves with a ratio of 1:1 between ppcA-L-Ft-3UTR and pX6-amiR-*ppcA*, the accumulation of GUS was less, compared to *N.b.* leaves with a ratio of either 1:10 or 1:100 [**Figure 5**]. Indicating, that the overexpression of amiRNA increased the effect on the target cleavage. However, either the recombination leading to the amiRNA expression or the degradation of the target was only partial in transiently transformed *N. benthamiana*.

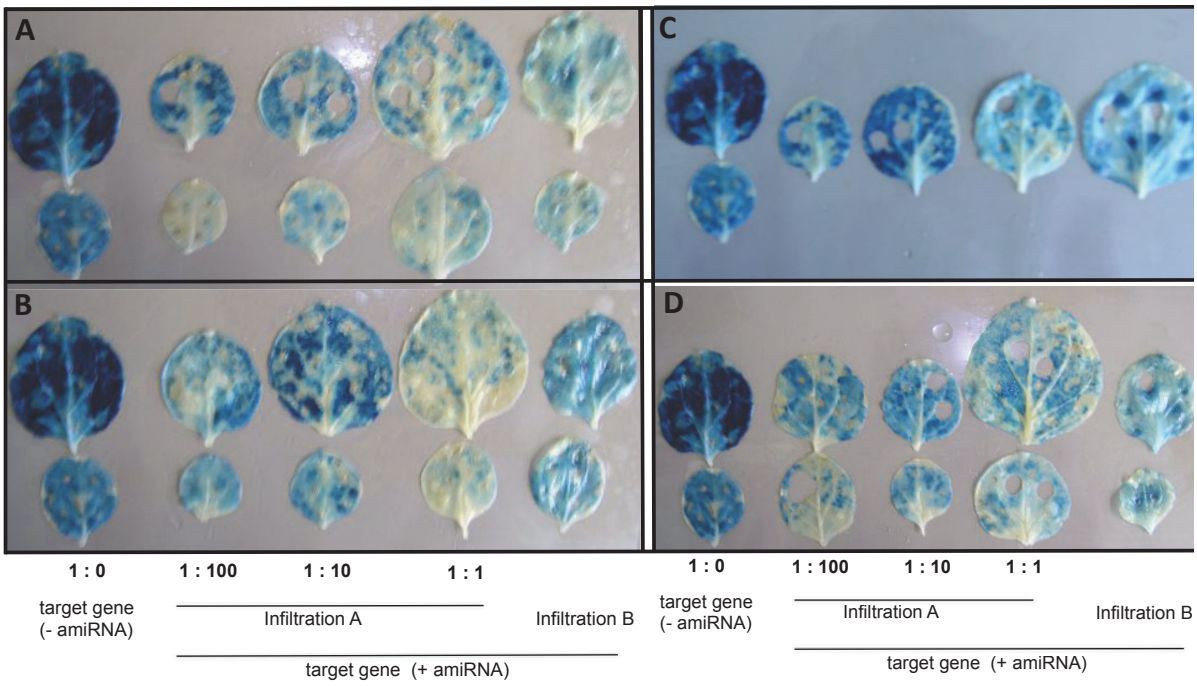


Figure 5: Functionality of the amiRNA with respect to its target in *N. benthamiana.* *N. benthamiana* leaves were co-transformed by *A. tumefaciens* infiltration carrying the ppcA-L-Ft-3UTR and the pX6-amiR-*ppcA* plasmids (Infiltration A and B). ppcA-L-Ft-3UTR (containing the target consistent to the amiRNA-complementary sites) and pX6-amiR-ppcA (containing the amiRNA). Transformed *N. benthamiana* leaves were treated externally with the inducer prior to analysis with a GUS histochemical assay. (**A** - **D**) 1:0 = control, single transformation with the target. Co-Infiltration A: Different ratios between the target:amiRNA (1:100; 1:10; 1:1), co-infiltration simultaneously. Co-Infiltration B: ratio between the target:amiRNA (1:1), co-infiltration successively. (**A**) AmiR-5´UTR-ppcA_I; (**B**) amiR-5´UTR-ppcA_II; (**C**) amiR-3´UTR-ppcA_I; (**D**) amiR-3´UTR-ppcA_II.

5. Generation of amiRNA transgenic F. bidentis

Two pX6-amiR-*ppcA* constructs against the 3´UTR from *ppcA* from *F. bidentis* were introduced independently into *Flaveria bidentis* via tissue-culture-based *Agrobacterium tumefaciens*-mediated transformation (Chitty et al., 1994). 30 lineages of the primary transformants (FbamiR-ppcA-1) were generated [**Table 1**].

30 lineages of the primary transformants were treated with β-estradiol at post-callus stages and as control plants non-treated transformants of the same lineages were considered. A stable inheritance of CLX-recombination (amiRNA expressing plants) from the primary to successive generations, is only possible if the subepidermal cells of the L2 layer (second layer of cells in the primordium of the shoot apical meristem, giving rise to germ cells) are mutated and if recombination occurred in these specific cells (Marcotrigiano and Bernatzky, 1995). Than this specific transgene will pass onto the progeny plants. Therefore, the application of β-estradiol was based on an external treatment at the entire leave surface and the shoot apical meristem of the plant. In addition, varying β-estradiol concentrations (2 to 20 µM) were tested.

Table 1. Characterization of the transgenic *F. bidentis* plants: FbamiR-ppcA-1 of the T_0, T_1 and T_2 generations and FbamiR-ppcA-2 of the T_0 generation. Estradiol (E+ treated; E- non-treated) and Kanamycin (K+ selected; K- unselected). Polymerase chain reaction (PCR); Pulse-amplitude modulation (PAM); Gas-exchange to measure the rate of CO_2 assimilation.

	T_0 FbamiR-ppcA-1	T_1 FbamiR-ppcA-1	T_1 FbamiR-ppcA-1	T_1 FbamiR-ppcA-1	T_2 FbamiR-ppcA-1	T_0 FbamiR-ppcA-2
Lineage no.	1 to 30	8 (2,3,4,5,8,9,16,21)	13 (1,2,4,5,6,8,9,10,11,29 ,20,29,30)	30 (1-30)	6 (2, 5, 9, 29, 30)	19 (1-19)
Plants analyzed	transgenic: lineage no. 1 to 30	160 plants	322 plants	30 plants	180 plants	transgenics no.: 1,4,5,6,9,10,11,13,15,16,17,18
Recombination event	1 plant (no. 5)	0	0	0	0	
Incomplete recombination event leading to genetic chimerical sections	10 plants (no.: 2, 3, 8, 9, 11, 12, 13, 18, 20, 21)	3 plants (no.: 5,9)	19 plants (no.: 2,5,6,9,10, 20,29,30)	8 plants (no.: 6,7,11,15,21,27,28,30)	8 plants (no.: 2, 5, 9)	
		genetic chimeras/ non-recombinants: No. 5: 2/18 (10 %) No. 9: 1/19 (5%)	genetic chimeras: in total 6 % No. 2: 3 No. 5: 4 No. 6: 2 No. 9: 4 No. 10: 2 No. 20: 1 No. 29: 1 No. 30: 2	genetic chimeras: in total 26 % No. 6: 1 No. 7: 1 No. 11: 1 No. 15: 1 No. 21: 1 No. 27: 1 No. 28: 1 No. 30: 1	genetic chimeras/ non-recombinants No. 2: 2/36 (5,5 %) No. 5: 3/36 (8,3%) No. 9: 3/36 (8,3%)	
Treatment and selection	E+/K+ external estradiol treatment on shoot apical meristem and leaves (control: E-/K+ for each lineage; FbWT E+)	E-/K- E+/K+ E-/K+ on medium and followed by external estradiol treatment (control: FbWT E+)	E+/K- on medium and followed by external estradiol treatment (control: E-/K- for each lineage; FbWT E +)	E-/K- (control: FbWT)	E-/K- (control: FbWT)	
Analyses	PCR phenotype Western-blot PAM self-fertilized/ seed propagation	PCR phenotype Western-blot Gas-exchange qRT-PCR PEPC activity self-fertilized/ seed propagation	PCR phenotype self-fertilized/ seed propagation	PCR phenotype self-fertilized/ seed propagation	PCR phenotype self-fertilized/ seed propagation	PCR self-fertilized/ seed propagation

6. CLX-lost excision in T_0 FbamiR-ppcA-1

To evaluate the frequency of recombination in the 30 T_0 FbamiR-ppcA-1 lineages, genomic DNA was probed for PCR analysis. Recombinants are characterized by expressing efficiently the amiRNA transgene uniformly along large leaf sectors. Among the 30 lineages a single plant showed to be a putative amiRNA *knockout* mutant. However, most of the inducer-treated transformants were non-recombinants revealed by DNA fragments corresponding to the complete genome integrated transgene, without excision of CLX-lost. Indicating, that the expression and activity of amiRNA was not functional in these plants. A few of the inducer-treated T_0 FbamiR-ppcA-1 were genetic chimeras [**Table 1**]. Genetic chimeras are characterized by either locally restricted recombination, or an incomplete recombination event, resulting in a combination of recombinant and non-recombinant tissue of the transgene (Zuo et al., 2000a; Guo et al., 2003).

The genome-integrated transgene was amplified at excised and flanked non-excised sequences with the primers P1 to P4 [**Figure 1**; **Table S1**]. P1 and P4 primers were directed within the non-excised sequence and P2 and P3 within excised sequences. Recombination events are characterized by a rejoined PCR fragment (P1/P4) of 493 bp and the absence with P1/P2 and P3/P4 [**Figure 1**]. In order to differentiate non-recombinant and genetic chimeras from true recombinant events the three different primer combinations were used. Non-recombination corresponds to 653 bp (P1-P2) and 982 bp (P3-P4) and no PCR fragment with P1-P4, since a DNA fragment of around 6 kb was above the estimated range of detection.

A single T_0 FbamiR-ppcA-1 (lineage no. 5) had an aberrant phenotype with growth retardation when compared to the non-treated T_0 FbamiR-ppcA-1 no. 5 control plant [**Figure 6A**]. Whereas, the remaining ß-estradiol-treated T_0 FbamiR-ppcA-1 plants had a similar phenotype to the respective control plant.

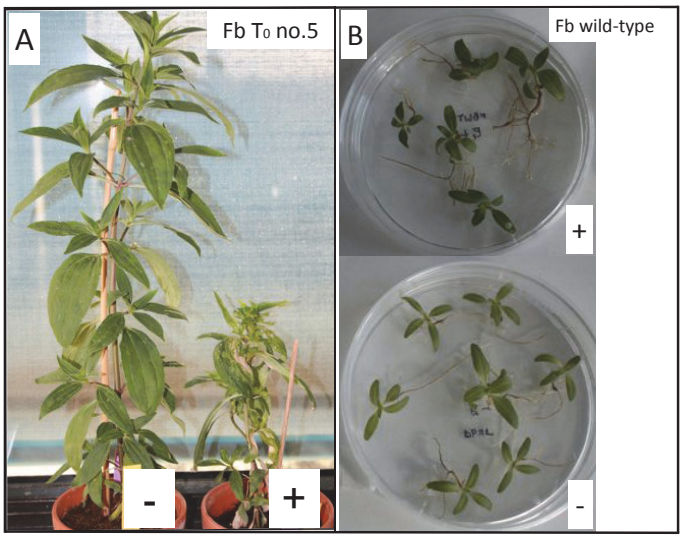


Figure 6. Phenotypic characterization of T_0 FbamiR-ppcA-1 no. 5. (**A**) Phentopic analysis of the ß-estradiol-induced (+) and non-induced (-) of T_0 FbamiR-ppcA-1 no. 5. Both induced and non-induced plants were generated at identical conditions. (**B**) Test of external and through ß-estradiol inductive medium (+) and without ß-estradiol (-) on *F. bidentis* wild-type. The treatment with β-estradiol caused no apparent endogenous elicitation of relative pathogen defensive responses or any other phenotypic anomalies to wild-type *F. bidentis*.

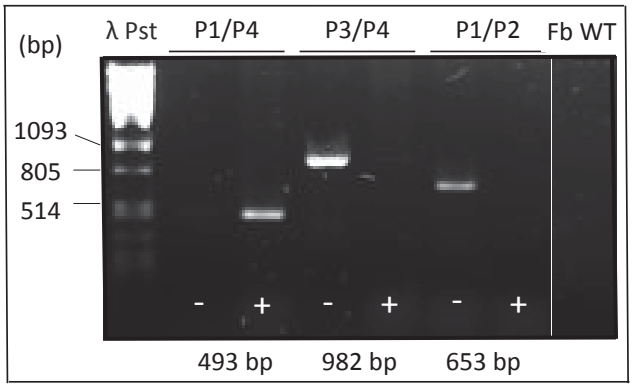


Figure 7. Molecular characterization of T_0 FbamiR-ppcA-1 no. 5. Confirmation of the recombination and CLX-lost excision in T_0 *FbamiR*-ppcA-1 no. 5. The genomic DNA was extracted from ß-estradiol-treated (+) and non-treated (-) leaves and from *F. bidentis* wild-type (Fb WT) and used for PCR analyses. Primers are indicated at the top. Genomic DNA from the ß-estradiol treated T_0 *FbamiR-ppcA-1* no. 5 yielded a DNA fragment with P1-P4 and no product with either P3-P4 and P1-P2. Whereas, the control (-) yielded only a DNA fragment with P3-P4 and P1-P2.

Figure 7 shows the molecular characterization of T_0 FbamiR-ppcA-1 no. 5. In the leaf section analyzed, complete CLX-lost excision and recombination occurred upon β-estradiol-induction. The rejoined DNA fragment of 493 bp was detected, whereas no DNA was amplified with P1-P2 and P3-P4. In the non-treated control T_0 FbamiR-ppcA-1 no. 5, a DNA fragment was amplified with the P1-P2 and P3-P4, but not with the P1-P4.

Figure 8 shows the immunodetection of PEPC in T_0 FbamiR-ppcA-1 no. 5. PEPC was absent in most plant sections. Indicating that a partial reduction of the PEPC content was probably due to the expression of amiRNA to decrease the levels of the *ppcA* mRNA *in planta*. The target effect of amiRNA in the T_0 FbamiR-ppcA-1 no. 5 likely resulted in a reduction of PEPC protein amount. By monitoring the PEPC protein of the remaining treated T_0 FbamiR-ppcA plants the amount did not changed to the control plants.

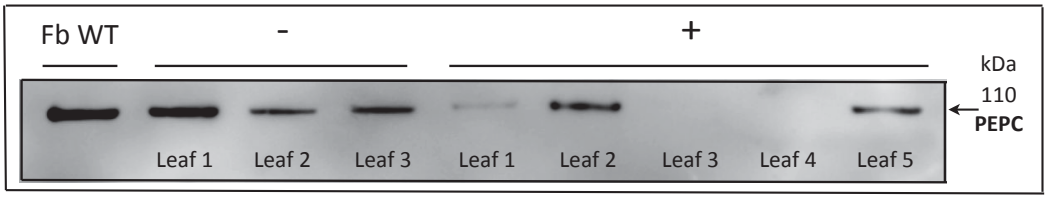


Figure 8. Immunodetection of PEPC in leaf extracts of T_0 FbamiR-ppcA-1 no. 5 separated by SDS-PAGE. PEPC detection in the putative recombinant transgenic *F. bidentis* no. 5 ß-estradiol-induced (E+) in comparison to both ß-estradiol-non-induced transgenic *F. bidentis* no.° 5 (E-) and wild-type *F. bidentis* (Fb WT). Biological replicates from the same plant were loaded on different lanes. The detection of PEPC polypeptides was based on a primary rabbit polyclonal anti-PEPC antiserum, followed by a mouse peroxidase conjugated secondary antibody and visualized with chemiluminescence.

The non photochemical quenching (NPQ) of the T_0 FbamiR-ppcA-1 no. 5 was measured [**Figure 9**]. The NPQ is a process in plants in order to stabilze the energy flow in the light-driven electron transport chain of photosynthesis to stress conditions (high light), which is dissipated as pH-dependent energy. In this study, the NPQ was used as an indicator of an inappropriate regulation of photosynthesis in the mutant as a result of a putative impediment of the PEPC carboxylation efficiency.

The results show, the developmental state of NPQ of T_0 FbamiR-ppcA-1 no. 5 was very low with respect to those of the non-treated and wild-type *F. bidentis*. The induction of the NPQ during illumination was high and occurred fast in the wild-type *F. bidentis* reaching a maximum extent at around 3 minutes. The relaxation of the NPQ occurred after 11 minutes upon darkness, because of the loss of the gradient of the ΔpH across the thylakoid membrane.

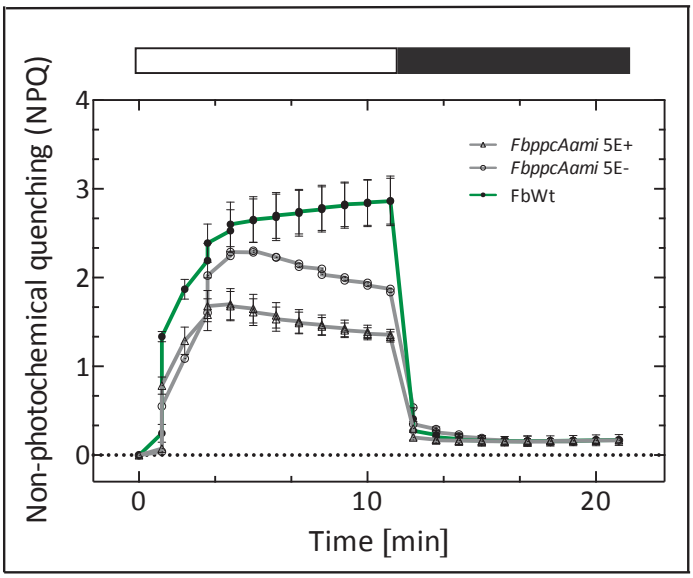


Figure 9. NPQ of chlorophyll fluorescence in T_0 FbamiR-ppcA-1 no. 5. Comparison of the induction of the NPQ in the β-estradiol-induces (E+) transgenic *F. bidentis* no. 5 to those of the β-estradiol-nontreated (E-) non-recombinant *F. bidentis* line 5 and wild-type *F. bidentis* dark adapted plants. The NPQ was measured with actinic light during 11 min and the dark horizontal bar indicates the relaxation of the NPQ in the dark. For each data point three biological replicates were used represented by the means and standard error of the mean (SEM).

7. CLX-lost excision in T_1 FbamiR-ppcA-1

The β-estradiol-treated 30 T_0 FbamiR-ppcA-1 lineages were considered for the regeneration of transgenic plants by self-fertilization and seed germination in the propagation of successive generations. The seeds of the T_1 and T_2 generation were germinated on Murshige-Skoog (MS) medium. DNA excision and recombination in the CLX system in the primary transformants should results in a loss of both the selectable marker (KAN) and the estradiol receptor cassette (XVE). Therefore, to ensure to select only the recombinants of the parent lineage, the T_1 and T_2 transformants were germinated on MS medium without estradiol and kanamycin [**Table 1**]. Some of the T_1 transformants were germinated on medium containing the inducer and/or the selective component, to ensure that non-recombinant cells of the parent lineage would become recombinants in the next generation.

The transgene was successfully passed onto the progeny plants of the T_1 and T_2 generation, as confirmed by PCR analyses. The molecular characterization showed that around 90 % of the inducer-treated transgenic *F. bidentis* within each of the progeny generations revealed to be non-recombinants and around 10 % were genetic chimeras [**Table 1**]. The recombinant transgene of the T_0 FbamiR-ppcA-1 no. 5 was not passed onto the progeny plants in several attempts of the T_1 and T_2 generation. The PEPC protein content was present in the inducer-treated T_1 FbamiR-ppcA-1 identical to the control plants [**Figure 10**]. Therefore, the T_1 FbamiR-ppcA-1 plants could not be considered for downstream analyses.

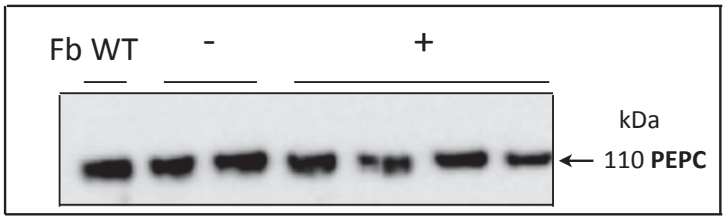


Figure 10. Immunodetection of PEPC in leaf extracts of the T_1 FbamiR-ppcA-1 separated by SDS-PAGE. Transgenic T_1 FbamiR-ppcA-1 were generated on ß-estradiol-inductive and selective medium (+) an on non-induced and non-selected (-) medium. The detection of PEPC polypeptides was based on a primary rabbit polyclonal anti-PEPC antiserum, followed by a mouse peroxidase conjugated secondary antibody and visualized with chemiluminescence.

DISCUSSION

The CLX system was characterized in transgenic *Arabidopsis* and *N. benthamiana*, where it was indicated to be tightly regulated and highly inducible with up to 66 % of efficiency (Zuo et al., 2000/01; Guo et al., 2003). In the transient system preceeding the stable transformation of *F. bidentis*, the efficiency of the CLX system and the functionality of the amiRNA was at least in part functional in transformed *N. benthamiana* leaves. However, the frequency of expressed GFP along the *N. benthamiana* leave sectors was below the predictions (Zuo et al., 2000/01; Guo et al., 2003). Concluding, that the DNA recombination leading to GFP expression occurred only in a limited number of cells.

The CLX system in *F. bidentis* resulted in an extremely low frequency of recombinants in the parent lineages. A single T_0 FbamiR-ppcA-1 no. 5 showed a strong retardation in growth. We have evidences that this plant was a amiRNA *knockout* mutant. The decrease in yield of the maximum NPQ in regard to that of the control plants [**Figure 9**], might have been a consequence of the partial decrease of the PEPC protein content [**Figure 8**]. Indicating, that this plant was affected in the efficiency of photosynthesis in particular in the photosystem II antenna pigment. This might have led to a reduced efficiency of the photosynthetic electron transport chain for the synthesis of both ATP and reducing agents required for the photochemical reaction of photosynthesis. In case the rate or capacity of the Calvin-Benson cycle is limited due to a reduced CO_2 assimilation by PEPC carboxylation, the electron flow in the photosynthetic electron transport chain could as a result become reduced. However, we have no evidence if a low CO_2 assimilation was in fact limiting the C_4 cycle and the ultimate reason of the reduced cell expansion. The transgene with CLX-lost excision and complete recombination of this parent mutant was not passed into the germline cells onto plants of the progenitor cells of descendent T_1 and T_2 generations, and therefore impeded reproducible analyses.

Even with different concentrations and modes of inducer-applications, the CLX system did not allow to be inherited into the germline of the parental cells to the following generations, and to recover recombinants along large leaf sectors in the progeny plants. Recombination was locally restricted in around 10 % of the progeny generations, indicating for the generation of genetic chimerical transgenic plants. In the genetic chimerical *F. bidentis* the regulation of the CLX-mediated recombination was stochastic, rather than strictly and highly regulated. The proportion of chimerical leaf sections was found to be still below the one of non-recombinants. There were no differences displayed at either the PEPC protein level or the rate of CO_2 assimilation between the non-recombinants and the genetic chimeras (data not shown).

The generation of genetic chimeras was also shown in *Arabidopsis* mutants carrying the CLX plasmid (Zuo et al., 2000a) and explained to be a result of the inaccessibility of the chemical inducer to the meristematic cells of the L2 cells. Genetic chimeras can also be a result of the segregation pattern in the T_1 and T_2 progenitors with a positional defect of the T-DNA insertions, which lead to heterozygously lines. The presence of two copies of the same dominant transgene present in the genome enhances the possibility to induce both copies successfully upon estradiol treatment.

The CLX system was used in the attempt to generate conditional *knockout F. bidentis* plants of the phosphoglycolate phosphatase (PGLP) gene (PhD project Myles Levey). The system seamed to be inefficient in a similar way to the one represented in this study. The possibility to enhance the frequency of complete recombinant plants, rather than chimerical plants might be achieved by an increase of the number of generated plants of progeny generations (up to T_5).

Due to the unreproducible generation of recombinants we could not define the C_3-to-C_4 mode changes *in vivo*. Neither we could determine if a depletion of PEPC in *F. bidentis* would affect the process of CO_2 exchange, nor if redundant regulatory mechanisms, such as the increase of phosphorylation of PEPC, would compensate the dysfunction. We could therefore not define if the manipulation of photosynthetic PEPC carboxylation in C_4 *Flaveria* would result in a shift in the CO_2 assimilation response towards a C_3 or a C_3-C_4 intermediate mode.

OUTLOOK

The generation of PEPC-deficient *knockout F. bidentis* mutants could be the base for additional approaches in the future. Therefore, it would be useful to establish a different *knockout* system to the one mentioned in this study. Several attempts exist to introduce an operative C_4 photosynthetic carbon cycle into transgenic C_3 plants in order to increase the photosynthetic capacity of these commercially important crop plants (Jeanneau et al., 2002a/b; Von Caemmerer and Furbank, 2003; Endo et al., 2008). In the opposite direction, it would be interesting to monitor the effect of the introduction of a C_3 *ppc* gene variant in a C_4 *F. bidentis knockout* plant with reduced *ppcA* levels, to deduce C_4 evolutionary changes.

MATERIAL AND METHODS

Plant material, growth conditions and plant transformation

30 lineages (*FbamiR-ppcA-1*) and 19 lineages (*FbamiR-ppcA*-2) of *F. bidentis* of the T_0 generation were generated via the tissue-culture-based *Agrobacterium tumefaciens*-mediated transformation technique (Chitty et al., 1994) using the strain AGL1 (Lazo et al., 1991). T_0 FbamiR-ppcA-1 plants were grown under greenhouse condition (S1) on soil with a photoperiod of 16-h, after generated on medium in the growth chamber. T_1 FbamiR-ppcA-1 were grown in a closed chamber at 120 µmol quanta m^2 s^{-1} at 18 °C (night) and 20 °C (day) during a 12-h photoperiod.

Isolation of RNA, cDNA synthesis and RACE-PCR

Total RNA was extracted from *Flaveria bidentis* and digested with DNAse using the RNA Purification kit (QIAGEN). Synthesis of 5` and 3` RACE-ready complementary DNAs was carried out according to the SMARTer RACE cDNA Amplification Kit User Manual (Clontech, Laboratories). 3´RACE-PCR was performed with the primer pairs Universal Primer A Mix (UPM) (Clontech, Laboratories) and gene specific primers I- to III-3RACE-PCR-ppcA_fw. 5´RACE-PCR was performed with the primer pairs Smart-II-A-5RACE and I- and II-5RACE-PCR-ppcA_rev [**Table S1**]. The 50x Advantage 2 Polymerase Mix (Clontech, Laboratories) was used for RACE-PCR according to the clontech protocol. The complete full length *ppcA* cDNA was amplified with the primer pairs 5´UTR-ppcA and 3´UTR-ppcA and the high-fidelity Phusion polymerase (Biolabs). DNA fragments were cloned with the CloneJET™ PCR Cloning Kit (Fermentas) and several independent clones verified via PCR and sequencing.

Design of amiRNA precursors and plasmid construction

The amiRNAs were directed to the 3´- and 5´-UTR of *ppcA* from *F. bidentis*. Four oligonucleotides for each amiRNA were provided by the WMD 'Oligo' tool (http://wmd2.weigelworld.org) [**Table S2**]**.** To generate the amiRNA precursor, the plasmid from *A. thaliana* (miR319a pBSK) was used for sub-cloning (WMD design tool). In overlapping PCR reactions the DNA fragments were separately amplified in a four-step reaction [**Figure S1**]. Thereby, the amiRNA* (III and IV) replaced the anti-sense endogenous miR319a* of RS3000, and the amiRNA (I and II) the sense endogenous miR319a. PCR reaction (a) was performed with the primer combination OligoA and IV, (b) III and II and (c) I and OligoB [**Table S2**]. Oligo A and B primers aligned to the boarders of the miR319a precursor. In the 4th PCR reaction the entire amiRNA precursor sequence was amplified as a single fragment with A and B using the clean PCR products of a, b and c as template DNA. PCR reactions for the generation of the amiRNA were carried out with the high-fidelity Phusion polymerase (Biolabs), and according to the standard PCR reactions of the protocol: "Cloning of artificial microRNAs" (Rebecca Schwab, MPI for Developmental Biology, Tübingen, Germany, 2006-2007). The amiRNA precursors were cloned with the CloneJET™ PCR Cloning Kit (Fermentas) for sequence confirmation. AmiRNAs were subcloned into the MCS of the pX6 (-GFP) plasmid (AF33063), designated as pX6-amiR-ppcA. In total four constructs were generated, two for each UTR.

Molecular characterization by PCR

For the extraction of genomic DNA frozen plant material was homogenized in 400 µL of DNA extraction buffer (0,2 M Tris/HCl, pH 7,5; 0,25 M NaCl; 0,025 M EDTA; 0,5 % (w/v) SDS). The solution was centrifuged at 13.000 x g for 5 minutes at room temperature and the supernatant transferred to a new tube. DNA precipitation was performed with 300 µL ice cold isopropanol (100 %) added to 300 µL supernatant. The solution was centrifuged at 13.000 x g for 5 minutes at 4 °C. The supernatant was washed with 70 % ice cold ethanol, centrifuged and the DNA pellet air dried. The pellet was resuspended in 50 µl of TE buffer (10 mM Tris/HCl, pH 8,0; 1 mM EDTA) and dissolved for at least 1 hour on ice. The genomic DNA concentration was determined with the nanodrop (Peqlab) and set to 50 ng. The genomic DNA was stored at 4°C. Genome-integrated transgene was confirmed in the *FbppcAami* lines with the primer combinations listed in **Table S2**, P1-Zuo-fw, P2-Zuo-rev, P3-Zuo-fw, P4_amiRNA_rev-I or P4_amiRNA_rev-II. The PCR reaction occurred at 50 cycles with the Phusion high-fidelity polymerase or alternatively with the taq DNA polymerase (5 U/µL) (Biolabs), according to the manufacture protocol.

Plasmid constructions

The recognition sites of BcuI (SpeI) and XhoI in the miR319a plasmid were used for shuttling the amiRNA sequence into the multiple cloning site (MCS). The PCR fragment (d) containing the amiRNA precursor for each 5´ and/or 3´UTR and the pX6-GFP vector (AF330636) was digested with the restriction enzymes BcuI (SpeI) and XhoI. The digested and linearized pX6-GFP vector was replaced with the purified amiRNA precursor using the T4-DNA ligase (Invitrogen). The pX6-amiR-*ppcA* plasmids were confirmed via sequencing and electrophoretically transformed into *Agrobacterium tumefaciens* (AGL1) cells (Hood et al., 1986; Lazo et al., 1991). The bacterial selection was by means of streptomycin, spectinomycin and gentamycin and the plant selection by means of kanamycin.

The recognition sites of SacI in the ppcA-L-Ft (pBi) plasmid (Stockhaus et al., 1994) was used for shuttling the 3´UTR of the *ppcA* from *F. bidentis* into ppcA-L-Ft [**Figure S2**]. The 3´UTR of *FbppcA* plus SacI restriction sites was amplified with the primers 3UTR-ppcAFbi-pbi121-fw and 3UTR-ppcAFbi-pbi121-rev using the 3´UTR_RACE template [**Table S1**]. The DNA product was cloned into the CloneJET^{TM} PCR Cloning Kit (Fermentas) for sequence confirmation**.** PCR reactions were carried out with the Phusion high-fidelity polymerase (Biolabs) for 25 cycles, according to the manufacture protocol. Both DNA fragments 3´UTR and ppcA-L-FT were digested with SacI and dephosphorylated with the Shrimp Alkaline phosphatase (SAP). The linearized ppcA-L-FT plasmid was ligated with the 3 ´UTR using the T4-DNA ligase (Invitrogen). The designated ppcA-L-Ft-3´UTR plasmid was chemically transformed into *E.coli* cells and confirmed via sequencing.

Histochemical analysis of the GUS reporter activity in *N. benthamiana* leaves

Agrobacterium tumefaciens strain (GV3101) (Holster et al., 1980; Koncz and Schell, 1986) was transformed via electroporation with either the pX6-amiR-ppcA or the ppcA-L-Ft-3UTR construct (Mustroph et al., 2013). The incubation time of 2 days at 30 °C on YEB-plates contained kanamycin. An overnight culture was further incubated in a 50 mL volume of YEB medium, containing 1M MES and 0,1 M Acetosyringone. The solution was centrifuged (20 min. at 3500 x g) and the pellet washed with dH_2O. After a second centrifugation step *Agrobacteria* were resuspended in a 5 mL incubation buffer (10 mM $MgCl_2$, 10 mM MES and 0,1 M Acetosyringone) after reaching an optical density of 0,7. The *Agrobacteria* suspension was set to dilution series, consisting of ppcA-L-Ft-3UTR and pX6-amiR-ppcA in a ratio of 1:1; 1:10; 1:100 and 1:0, respectively. Subsequently the solution was incubated at room temperature for 2 hours in the dark. The dilution series were infiltrated into *N. benthaminana* leaves of different plants. The infiltration B consisted of a successive infiltration step of transformed *A. tumefaciens* with ppcA-L-Ft-3UTR and pX6-amiR-*ppcA*. Infiltrated leaves were treated externally with 20 µM estradiol and grown under greenhouse conditions. Leave material was analyzed after 24-h to 3 days of infiltration. GUS as an integral part of the ppcA-L-Ft-3UTR cassette served as a reporter gene. The GUS staining solution (100 mM Na_2HPO_4, pH 7.5, 10 mM EDTA, 50 mM $K_4[Fe(CN)_6]$, 50 mM $K_3[Fe(CN)_6]$, 0,1% (v/v) triton X-100, 2 mM 5-bromo-4-chloro-3-indolyl-ß-D-glucuronid acid) was vaccum infiltrated into entire *N. benthamiana* leaves and incubated for 2 days at 37°C. Fluorescence microscopy was performed with the Axiophot fluorescence microscope (Carl Zeiss AG, Germany). Images of entire leaves were taken with a Canon camera.

Transient expression of GFP in *N. benthamiana*

Agrobacterium tumefaciens strain (GV3101) (pMP90) (Koncz and Schell, 1986) was transformed via electroporation with the pX6-GFP plasmid (Zuo *et al.*, 2000a). The cultivation and infiltration of transformed *Agrobacterium tumefaciens* into *N. benthamiana* was performed as described above, but without dilution series. Infiltrated *N. benthamiana* leaves were treated externally with 20 µM estradiol and analyzed after 24-h to 3 days. Non treated *N. benthamiana* leaves served as a control. The GFP fluorescence was analyzed with *N. benthamiana* leaf discs with a Zeiss (HAL-100) microscope.

Chemical induction

The primary transformants were selected on kanamycin and treated post-callus generation externally with a solution containing 20 µM of β-estradiol and 0.1 % of Tween 20. In addition different concentrations of 2 µM to 20 µM estradiol were tested externally. Estradiol was applied repeatedly along a minimum of two weeks, to ensure that each progenitor cell becomes permanently activated. After self-fertilization the seeds of the T_1 and T_2 generation were sterilized and washed before germination and regenerated on Murashige & Skoog media.

Protein analysis

For the immunodetection of PEPC the total protein content of *F. bidentis* transgenic and wild-type plants was extracted according to (Shen et al., 2007). The protein extraction buffer was composed by the following components: 100 mM Tris/HCl, pH 7.8; 4 M Urea; 5 % SDS; 15 % glycerol; 10 mM MeOH was added in a ratio of 2:1 (v/w) to the dried crude leaf extracts. A mixture of a protease inhibitor mix (Sigma) was added to the extraction buffer prior to homogenization. The homogenized solution was boiled at 100 °C for 4 minutes and centrifuged at 15.000 rpm for 15 minutes at room temperature. The supernatant was frozen at -80°C for further analysis. 30 µg of total protein of the plant crude extracts per sample was determined by standard procedure Bradford (Biorad) (Bradford, 1976) and resolved on a 10 % acrylamide gel (Schägger and von Jagow, 1987). After 2-h blotting of total protein with a standard procedure, nitrocellulose membranes were blocked in TBS (200 mM Tris/HCl (pH 7.6); 1.37 M NaCl; 0.1% Tween-20 and 5% milkpowder) for 1-h. Incubation with the primary antibody (anti-PEPC) from Agrisera, was performed in a dilution of 1:1000 in TBS for 12 hours at 4°C. The secondary antibody (anti-rabbit Agrisera) was added for 1 hour at room temperature in a 1:80000 TBS dilution. Band signals were examined with the reagent of the chemiluminescence assay (SuperSignal West; Thermo Scientific) using the LAS system.

Fluorescence measurements

Plants were dark adapted for 5 hours. NPQ measurements were performed according to (Kramer et al., 2004) with a Dual PAM-100 device (Walz, Effeltrich, Germany). Saturation of actinic light impulses were given at a light intensity of 2000 µmol photons m^{-2} s^{-1}; applied every second during the first 11 s of illumination. Light were switched off for the dark phase of 10 sec to determine the relaxation of NPQ. Biological replicates correspond to different leaves from the same plant.

REFERENCES

Bradford, M.M. (1976) Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principal-Dye Binding. *Anal. Biochem.* **72**: 248–254.

Chitty, J.A., Furbank, R.T., Marshall J.S., Chen Z, and **Taylor, W.C.** (1994). Genetic transformation of the C_4 plant, *Flaveria bidentis. The Plant Journal* **6**: 949–956.

Christin, P.-A., **Osborne, C.P., Sage**, **R.F., Arakaki, M.** and **Edwards, E.J.** (2011). C_4 eudicots are not younger than C_4 monocots. *Journal of Experimental Botany* **62**: 3171–3181.

Ernst, K. and **Westhoff, P.** (1996). The phospho*enol*pyruvate carboxylase (*ppc*) gene family of *Flaveria trinervia* (C_4) and *F. pringlei* (C_3): molecular characterization and expression analysis of the *ppcB* and *ppcC* genes. *Plant Mol Biol.* **34**: 427–443.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85:** 8998–9002.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M. and **Westhoff, P.** (2011). Evolution of C_4 Photosynthesis in the genus *Flaveria*: How Many and Which Genes Does It Take to Make C_4? *The Plant Cell* **23**: 2087–2105.

Guo, H.S., Fei ,J-F., Xie, Q. and **Chua, N.H** (2003). TECHNICAL ADVANCE: A chemical-regulated inducible RNAi system in plants. *The Plant Journal* **34**: 383–392.

Hood, E.E., Helmer, G.L., Fraley, R.T., and Chilton, M.D. (1986). The hypervirulence of *Agrobacterium tumefaciens* A281 is encoded in a region of pTiBo542 outside of T- DNA. *J Bacteriol* **168:** 1291–1301.

Ishige, F., Takaichi, M., Foster, R., Chua, N.-H. and **Oeda, K.** (1999). A G-box motif (GCCACGTGCC) tetramer confers high-level constitutive expression in dicot and monocot plants. *Plant J.* **18**: 443–448.

Jeanneau, M., Vidal, J., Gousset-Dupont, A., Lebouteiller, B., Hodges, M., Gerentes, D. and **Perez, P.** (2002a). Manipulating PEPC levels in plants. *J Exp Bot.* 53: 1837–1845.

Jeanneau, M., Gerentes, D., Foueillassar, X., Zivy, M., Vidal, J., Toppan, A. and **Perez, P.** (2002b). Improvement of drought tolerance in maize: towards the functional validation of the Zm-Asr1 gene and increase of water use efficiency by over-expressing C_4-PEPC. *Biochimie* 84: 1127–1135.

Koncz, C., and **Schell, J.** (1986). The promoter of the TL-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* **204:** 383–396.

Kramer, D.M., Johnson, G., Kiirats, O. and **Edwards, G.E.** (2004). New fluorescence parameters for the determination of QA redox state and excitation energy fluxes**.** *Photosynthesis Research* **79:** 209–218

Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1991). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Biotechnology* **9:** 963-967.

Marcotrigiano, M. and **Bernatzky, R**. (1995). Arrangement of cell layers in the shoot apical meristems of periclinal chimeras influences cell fate. *Plant J.* **7**, 193–202.

Mustroph, A., Stock, J., Hess, N., Aldous, S., Dreilich, A. and **Grimm, B.** (2013). Characterization of the phosphofructokinase gene family in rice and its expression under oxygen deficiency stress. *Front. Plant. Sci.* **125**: 1–16.

Ossowski, S., Schwab, R. & Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J*. **53**: 674–690.

Sage, R.F., Sage, T.L. and **Kocacinar, F.** (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annu Rev. Plant Biol*. **63**: 19–47

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* **8**: 517–527.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N. & Weigel, D. (2006). Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* **18**: 1121–1133.

Shen, Y., Khanna, R., Carle, C.M. and **Quali, P.H.** (2007). Phytochrome Induces Rapid PIF5 Phosphorylation and Degradation in Response to Red-Light Activation. *Plant Physiology* **145**: 1043–1051.

Stockhaus, J., Poetsch, W., Steinmüller, K. and **Westhoff, P.** (1994). Evolution of the C4 phosphoenolpyruvate carboxylase promoter of the C4 dicot *Flaveria trinervia*: an expression analysis in the C3 plant tobacco. *Mol Gen Genet* . **245**: 286–293.

Von Caemmerer, S. and **Furbank, R.T.** (2003). The C_4 pathway: an efficient CO_2 pump. *Photosynthesis Research* **77**: 191–207.

Westhoff, P. and **Gowik., U.** (2004). Evolution of C_4 Phospho*enol*pyruvate carboxylase genes and proteins: a case study with the genus *Flaveria*. *Annals of Botany* **93**: 13–23.

Zhu, Y.Y., Machleder, E. M., Chenchik, A., Li, R. and **Siebert, P. M.** (2001). Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *BioTechniques* **30**: 892–897.

Zuo, J., Niu, Q.W. and **Chua, N.H.** (2000a). Technical advance: an estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.* **24**: 265–273.

Zuo, J., Niu, Q.W., Nishizawa, N. Wu, Y, Kost, B. and **Chua, N.H.** (2000b). KORRIGAN, an Arabidopsis endo-1,4-β-Glucanase, Localizes to the Cell Plate by Polarized Targeting and Is Essential for Cytokinesis. *The Plant Cell* **12**: 1137–1152.

Zuo J., Niu, Q.-W., Moller, S.G. and **Chua, N.H.** (2001). Chemical-regulated, site-specific DNA excision in transgenic plants. *Nature Biotechnology* **19**: 157-161.

SUPPLEMENTAL FILES

Figure S1: DNA sequences of the 5´and 3´ untranslated and the coding sequences of *ppcA* from *F. bidentis*.

3´UTR-ppcA_Fb

>**TAG**GTTAACCGTTATGCAAGTATGTATTTATCTTTTTTTTTCATGCTTTTGGATGTTGTCACATGCCTTGTTTACT
GCATGTGTTCTTGAGTGTTTGTACGCATCATCAATCTATGCATTTCAGAAGTTGTACTGTGTGCATCATTGATGGTG
GTTTATCCAATAAACTAAAGACTTATTGTACACTTGACTACAAGTGATTATATTAATTGTTGTGCTTTCTACG**TAA**

5´UTR-ppcA_Fb

>ACGCGGGGAATCCTGAGTTTGCTCTGTGGATGAGCAACTGTATCGTTGATACTTGATACCTGTAACTCACACACCT
CATATCTCATACTTCATCTATAAATACCCAATTCATTTTGCTCAAAGTCTCAACATTGAGCATACCCAATATTCAGG
GTTGGAGGGGAATTAAGTATTAAGCAAGGGTGTGAGTA**ATG**

Fb_ppcA

>**ATG**GCTAACCGGAATGTGGAGAAATTAGCATCGATCGACGCTCAGTTGAGGCTTTTAGTCCCTGGGAAAGTTTCTG
AGGATGATAAGCTTGTTGAGTATGATGCTTTGCTTTTGGATAAGTTTCTGGATATTCTTCAGGATTTGCATGGGGAA
GATCTCAAGGAAGCGGTTCAACAATGCTATGAGCTATCTGCTGAATATGAAGGAAAACATGACCCGAAGAAGCTGGA
GGAGCTTGGAAGTCTGTTGACAAGTTTAGATCCAGGGGATTCCATTGTCATTGCAAAAGCCTTTTCTCACATGCTTA
ACTTAGCCAATCTGGCTGAAGAAGTTCAGATTGCTTACCGCCGAAGAATCAAACTGAAGAGTGGTGATTTTGCTGAT
GAGGCTAATGCAACAACTGAATCAGATATTGAAGAAACTTTCAAGAGACTTGTGCATAAGCTTAACAAGTCCCCTGA
AGAGGTTTTTGATGCACTCAAGAATCAAACTGTTGAGTTGGTCTTGACTGCTCATCCAACTCAATCCGTCCGCAGAT
CTTTGCTTCAAAAGCATGGAAGGATTCGTAACTGTCTGGCCCAGTTGTATGCCAAAGACATCACTCCTGATGATAAG
CAGGAACTGGATGAGGCTTTGCATAGAGAAATTCAAGCTGCATTCCGTACTGATGAAATCAGGAGGACTCCACCAAC
ACCACAAGATGAAATGAGAGCAGGAATGAGTTACTTCCATGAAACAATCTGGAAGGGTGTTCCAAAATTCTTACGTC
GTGTTGACACTGCCTTAAAGAATATTGGGATTAATGAACGTTTTCCCTATAATGCACCTCTAATTCAATTCTCTTCA
TGGATGGGTGGTGACCGTGATGGCAATCCGAGGGTTACTCCTGAGGTAACAAGGGATGTTTGTTTGCTAGCCAGAAT
GATGACGTCAAACATGTACTTTTCTCAGATAGAGGATCTTATGATTGAGATGTCCATGTGGCGTTGTAATAGTGAAC
TACGTGTTCGAGCAGAAGAACTGTACCGAACATCAAGAAAAGATGTGAAGCACTACATAGAGTTTTGGAAACAGATT
CCAACGAATCAACCTTATCGTGTAATTCTTGGTGATATAAGGGACAAGTTATATAATACACGTGAACGATCTCGTCA
TCTATTGGTCGATGGGAAATCTGACATCCCAGACGAAGCTGTTTATACCAATGTTGAACAGCTCTTGGAACCACTGG
AGCTATGCTACAGATCACTATGTGACTGTGGTGACCATGTGATTGCTGATGGAAGCCTTCTTGATTTTCTAAGACAA
GTGTCGACTTTTGGACTCTCACTTGTAAAACTTGATATAAGGCAAGAATCTGACCGTCACACTGAAGTCCTTGATGC
AATCACTCAACATTTAGGAATTGGGTCCTATCGTGAGTGGTCTGAAGAAAAACGCCAAGAATGGCTTCTAGCTGAAC
TCAGTGGAAAACGTCCTCTTATTGGTCCAGACCTTCCAAAAACTGAGGAAGTTAAGGATTGTTTAGACACGTTTAAA
GTTTTAGCAGAACTCCCATCTGACTGTTTTGGTGCTTACATCATCTCAATGGCCACATCAACTTCTGATGTCCTTGC
TGTTGAGCTTCTCCAGCGCGAATACCATATAAAACATCCGTTACGCGTGGTCCCCTTATTTGAAAAACTTGCTGACC
TGGAGGCGGCCCCTGCGGCCATGACCCGCCTTTTCTCAATGGATTGGTACAGAAACCGGATTGACGGTAAACAAGAA
GTCATGATTGGGTACTCTGATTCAGGAAAAGATGCAGGCCGGTTTTCTGCTGCATGGCAGCTCTACAAAACTCAAGA
ACAGATTATTAAAATTGCAAAAGAGTTTGGAGTCAAACTTGTTATATTTCATGGGCGTGGTGGAACTGTTGGTAGAG
GTGGTGGGCCCACACATCTGGCTCTTCTCTCTCAACCACCAGACACCATTAACGGGTCTTTAAGAGTCACGGTTCAG
GGTGAGGTCATAGAGCAGTCGTTTGGTGAGGAACATTTGTGCTTTAGAACACTTCAGAGATTTTGTGCAGCTACACT
TGAGCATGGGATGAACCCACCAATCTCACCACGGCCCGAGTGGCGTGAACTTATGGACCAGATGGCTGTTGTTGCAA
CCGAGGAGTACCGTTCTGTTGTGTTAAAGGAACCACGTTTTGTGGAGTATTTCCGGCTTGCAACACCTGAACTGGAG
TTCGGGCGTATGAATATTGGAAGTCGCCCATCAAAAAGAAAACCGAGTGGTGGCATTGAATCACTCAGAGCCATTCC
ATGGATCTTTTCATGGACTCAGACCAGGTTCCATCTCCCAGTTTGGCTCGGGTTTGGGGCGGCGTTCAAACACGCCA
TCCAAAAAGACAGCAAGAATCTCCAAATGCTTCAAGAAATGTACAAAACATGGCCTTTCTTTCGGGTCACCATTGAT
TTAGTTGAAATGGTGTTTGCTAAAGGTAACCCAGGCATTGCTGCCCTGAATGACAAGCTCCTTGTTTCTGAAGATCT
AAGGCCCTTTGGAGAATCTTTGAGAGCAAACTATGAAGAAACCAAAAATTTTCTTCTCAAGATTGCTGGACATAAGG
ACCTTCTAGAGGGTGATCCCTACTTGAAACAAGGAATCAGGCTGCGTGATCCGTACATCACAACCTTGAATGTATGC
CAAGCTTATACCCTAAAGAGGATCCGTGACCCGAACTATCATGTGACATTAAGGCCTCATATTTCTAAAGAATATGC
TGCTGAGCCGAGCAAACCAGCTGATGAGCTTATCCACCTGAACCCAACCAGCGAGTACGCACCCGGTTTGGAGGACA
CGCTCATCTTGACCATGAAAGGGATTGCTGCTGGAATGCAGAACACCGGT**TAG**

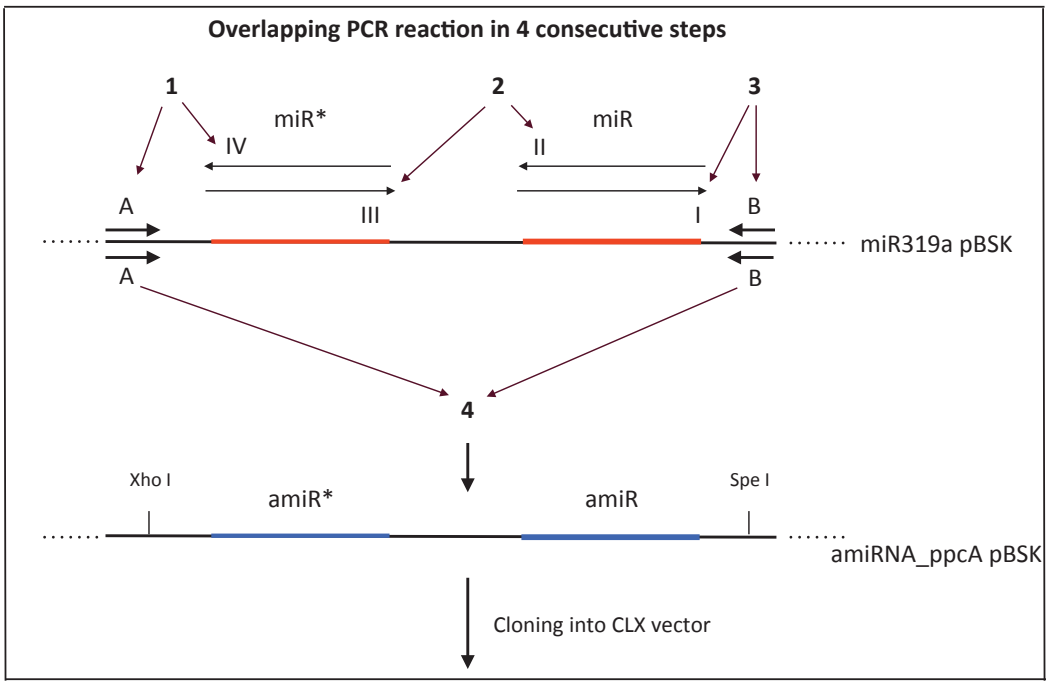


Figure S2: **Diagram of the generation of the amiRNA precursor**. DNA fragments encoding sense (amiRNA) and antisense (amiRNA*) amiRNA of single-stranded 21mer were engineered by overlapping PCR reactions: a combination of primer pairs was used in three separate PCR reactions (1 to 4), corresponding to primer pairs: A-IV; II-III; I-B) and a fusion PCR reaction (A-B). This allowed to amplify the complete amiRNA precursor against the 5´- and 3´-untranslated regions of ppcA from *F. bidentis*. The amiRNA precursor of the 5´- and 3´-UTR of *ppcA* from *F. bidentis* was cloned further into the plasmid pX6-CLX plasmid, and placed behind a constitutive promoter for plant transformation. Abbreviations: amiRNA, artificial micro RNA, rev, reverse direction of the primer; fw, forward direction of the primer.

Table S1: Oligonucleotides used for the isolation of *ppcA* from *F. bidentis* and for cloning.

Method	Primer name	Forward primer	Reverse Primer
ppcA-L-Ft-3UTR	3UTR-ppcAFbi-pbi121-fw	ACCGAGCTCGTTAACCGTTATGCAAGTAT	
	3UTR-ppcAFbi-pbi121-rev		TTCGAGCTCTTAAGTAGAAAGCACAACAA
5´RACE	Smart-II-A-5RACE	AAGCAGTGGTATCAACGCAGAGT	
	I-5RACE-PCR-ppcA_rev		TTACCGCCGAAGAATCAAACTGAAGAG
	II-5RACE-PCR-ppcA_rev		CTGGAGGAGCTTGGAAGT
3´RACE	UPM		ACTCTGCGTTGATACCACTGCTTGCCCTATAGTGAGTCGTATTAG
	I-3RACE-PCR-ppcA_fw	CGCTCATCTTGACCATGAAAGGGATT	
	II-3RACE-PCR-ppcA_fw	GCTCATCTTGACCATGAAGGGG	
	III-3RACE-PCR-ppcA_fw	GCTCATCTTGACCATGAAGGGGATT	
ppcA-CDS	UTR-5RACEppcA	GTATTAAGCAAGGGTGTGAGTA	
	UTR-3RACEppcA		ATACATACTTGCATAACGGTTAAC
molecular analysis	P1-Zuo-fw	CCATCTCCACTGACGTAAGGGAT	
	P2-Zuo-rev		CTCGTCAATTCCAAGGGCATCGGT
	P3-Zuo-fw	CTGGACACAGTGCCCGTGTCGGA	
	P4_amiRNA_rev-I		GGAATACAAAAGAGAGCCGTTATGCAAG
	P4_amiRNA_rev-II		GCATATATGTCACTTAGTGGATCAAGCAT

Table S2: The amiRNA sequences as predicted target for *ppcA* from *F. bidentis.* The position of the amiRNA in the target untranslated sequences correspond to: 3-5amiRNA (nt position -121 bp); 4-5amiRNA (nt position-132 bp); 1-3amiRNA (nt position 151 bp); 2-3amiRNA (nt position 9 bp) (Figure 2).

Gene	Name	Sequence
amiR-5 ´UTR-ppcA	I miR-s-3-5amiRNA	gaTTATGCTCAATGTTGAGACTTtctctcttttgtattcc
	II miR-a-3-5amiRNA	gaAAGTCTCAACATTGAGCATAAtcaaagagaatcaatga
	III miR*s-3-5amiRNA	gaAAATCTCAACATTCAGCATATtcacaggtcgtgatatg
	IV miR*a-3-5amiRNA	gaATATGCTGAATGTTGAGATTTtctacatatatattcct
	I miR-s-4-5amiRNA	gaTTGAATATTGGGTATGCTCAAtctctcttttgtattcc
	II miR-a-4-5amiRNA	gaTTGAGCATACCCAATATTCAAtcaaagagaatcaatga
	III miR*s-4-5amiRNA	gaTTAAGCATACCCATTATTCATtcacaggtcgtgatatg
	IV miR*a-4-5amiRNA	gaATGAATAATGGGTATGCTTAAtctacatatatattcct
amiR- 3 ´UTR-ppcA	I-miR-s-amiRNA-1-3UTR-fw	gaTTAGTTTATTGGATAAACCACctctctttttgtattcca
	II-miR-a-amiRNA-1-3UTR-rev	agGTGGTTTATCCAATAAACTAAtcaaagagaatcaatga
	III-miR*s-amiRNA-1-3UTR-fw	agGTAGTTTATCCAAAAAACTATtcacaggtcgtgatatg
	IV-miR*a-amiRNA-1-3UTR-rev	gaATAGTTTTTTGGATAAACTACctacatatatattccta
	I-miR-s-amiRNA-2-3UTR-fw	gaTAATACATACTTGCATAACGGctctctttttgtattcca
	II-miR-a-amiRNA-2-3UTR-rev	agCCGTTATGCAAGTATGTATTAtcaaagagaatcaatga
	III-miR*s-amiRNA-2-3UTR-fw	agCCATTATGCAAGTTTGTATTTtcacaggtcgtgatatg
	IV-miR*a-amiRNA-2-3UTR-rev	gaAAATACAAACTTGCATAATGGctacatatatattccta
	PCR-amiRNA-OligoA-fw	CTGCAAGGCGATTAAGTTGGGTAAC
	PCR-amiRNA-OligoA-rew	GCGGATAACAATTTCACACAGGAAACAG

AUTHOR CONTRIBUTIONS

S.H.A. performed all experiments.

Monika Streubel and **Maria Koczor** performed the transformation of *F. bidentis*

P.W. and **U.G.** designed the project

Chapter 2

Evolution of the phospho*enol*pyruvate carboxylase protein kinase family in C_3 and C_4 *Flaveria* species

Evolution of the phospho*enol*pyruvate carboxylase protein kinase family in C_3 and C_4 *Flaveria* species

Running title: Evolution of the PPCK family in *Flaveria*

Authors:
Sophia H. Aldous[1], Sean E. Weise[2], Thomas D. Sharkey[2], Daniel M. Waldera-Lupa[3], Kai Stühler[3,4], Julia Mallmann[1], Georg Groth[5], Udo Gowik[1,4], Peter Westhoff [1,4] and Borjana Arsova[1,4*]

[1]Heinrich-Heine-Universität, Institut für Entwicklungs- und Molekularbiologie der Pflanzen, Universitätsstrasse 1, 40225 Düsseldorf, Germany
[2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
[3]Molecular Proteomics Laboratory (MPL), BMFZ, Heinrich-Heine-Universität, Universitätsstrasse 1, 40225 Düsseldorf, Germany
[4]Cluster of Excellence on Plant Sciences (CEPLAS) "From Complex Traits towards Synthetic Modules", Universitätsstrasse 1, 40225 Düsseldorf, Germany
[5]Heinrich-Heine-Universität, Biochemische Pflanzenphysiologie, Universitätsstrasse 1, 40225 Düsseldorf, Germany

*To whom the correspondence should be addressed: Borjana Arsova, Heinrich-Heine-Universität, Institut für Entwicklungs- und Molekularbiologie der Pflanzen, Universitätsstrasse 1, 40225 Düsseldorf, Germany, Tel.: + 49 (0) 211 81 13428. E-mail: Arsova@hhu.de

Research area: Genes, Development and Evolution
Associate Editors Jiming Jiang and Hong Ma

Summary: A gene family of the PEPC kinase from *Flaveria* evolved in parallel to the gene family of PEPC, and this resulted in diurnal regulatory phosphorylation of PEPC that differs between *Flaveria* species with C_3 and C_4 photosynthetic-type.

ABSTRACT

The key enzyme for C_4 photosynthesis, phosphoenolpyruvate carboxylase (PEPC), has evolved from non photosynthetic PEPC found in C_3 ancestors. In all plants PEPC is phosphorylated by phosphoenolpyruvate carboxylase protein kinase (PPCK). However, differences in the phosphorylation pattern exist among plants with these photosynthetic types and it is still not clear if they are due to interspecies differences or depend on photosynthetic type. The genus *Flaveria* contains closely related C_3, C_3-C_4 intermediate and C_4 species, which are evolutionarily young and thus well suited for comparative analysis. To characterize evolutionary differences in PPCK between plants with C_3 and C_4 photosynthesis, transcriptome libraries from nine *Flaveria* species were used and a two member PPCK family (*PPCKA* and *PPCKB*) was identified. Sequence analysis identified a number of C_3 and C_4 specific residues with various occurrence in the intermediates. Quantitative analysis of transcriptome data revealed that *PPCKA* and *PPCKB* exhibit inverse diurnal expression patterns and that C_3 and C_4 *Flaveria* species differ in the expression levels of these genes. *PPCKA* has maximal expression levels during the day, whereas *PPCKB* has maximal expression during night. Phosphorylation patterns of PEPC varied among C_3 and C_4 *Flaveria* species too, with PEPC from the C_4 species being predominantly phosphorylated throughout the day while in the C_3 species phosphorylation level was maintained during the entire 24 hours. Since C_4 *Flaveria* evolved from C_3 ancestors, this work links the evolutionary changes in sequence, PPCK expression and phosphorylation pattern to an evolutionary phase-shift of kinase activity from a C_3 to C_4 mode.

INTRODUCTION

C_4 plants evolved from C_3 plants by developing a spatial separation for the process of carbon fixation in the leaves, and carrying it out in two cell types, mesophyll and bundle-sheath cells. This special leaf anatomy is known as Kranz anatomy and includes enlarged, chloroplast-rich bundle-sheath cells around the closely spaced veins to ensure an intense contact between the mesophyll and bundle-sheath cells (Sage et al., 2012). During C_4 photosynthesis atmospheric CO_2 is initially fixed by phosphoenolpyruvate carboxylase (PEPC) in the mesophyll cells. Here, PEPC catalyzes the β-carboxylation of phosphoenolpyruvate, using HCO_3^- as substrate leading to the formation of the 4-carbon organic acid oxaloacetate. Oxaloacetate is reduced to malate or transaminated to aspartate, and transported into the bundle-sheath cells where CO_2 is released by a decarboxylating enzyme (Edwards and Walker, 1983). By this process, CO_2 is effectively concentrated in the bundle-sheath cells bringing its concentration up to 1500 ppm (Hatch, 1987; Sage et al., 2012). CO_2 is re-fixed by Rubisco (ribulose-1.5-bisphosphate carboxylase/oxygenase), which in C_4 plants is restricted to bundle-sheath cells, and further metabolized by the Calvin-Benson cycle (Sage, 2004). This characteristic allows C_4 plants to survive in more extreme environments where heat and the lack of water cause the closing of stomata and result in low CO_2 concentrations in the intercellular spaces around the mesophyll cells (Ehleringer et al., 1997).

However, the evolution of C_4 photosynthesis involved more than just the development of Kranz anatomy. It included changes in the transcriptional regulation of genes which encode components of the

C_4 pathway, and the adjustment of enzyme properties and their regulation (Sheen, 1999; Sage, 2004; Akyildiz et al., 2007; Engelmann et al., 2008; Aubry et al., 2011; Brown et al., 2011; Ludwig, 2011; Wiludda et al., 2012; Paulus et al., 2013; Paulus et al., 2013). In the process of evolution towards C_4 photosynthesis, PEPC was influenced by a series of single amino acid exchanges which raised the Michaelis-Menten constant for the substrate phosphoenolpyruvate (Bläsing et al., 2000) and lowered the sensitivity to malate and aspartate (Engelmann et al., 2002; Paulus et al., 2013; Paulus et al., 2013). So far there is no information if such exchanges have also occurred in the regulatory protein phosphoenolpyruvate carboxylase protein kinase (PPCK) whose phosphorylation of PEPC changes the kinetic and regulatory properties of the enzyme (Nimmo, 2003). In order to shed light on the questions "What are the differences in PEPC regulation, by phosphorylation, between plants with C_3 and C_4 photosynthesis?" and "How did PEPC phosphorylation in C_4 plants evolve to the state it is today?" we turned to species in the genus *Flaveria*.

Although the first C_4 plants originated about 30 million years ago (Pagani et al., 2005; Tipple and Pagani, 2007), the much younger genus *Flaveria* is the preferred model for studying the evolution of C_4 photosynthesis. Having evolved in the last 3 million years, the species contained within this genus still retain a high level of similarity and allow the study of changes driven predominantly by the evolution of C_4 photosynthesis (Christin et al., 2011; Sage et al., 2012). Most importantly, the genus contains C_4, C_3-C_4 intermediates, and C_3 species e.g. *F. trinervia, F. ramosissima* and *F. pringlei* respectively (Sage, 2004; McKown et al., 2005). This palette of species with various photosynthetic types allows the investigation of the gradual development of C_4-related characteristics by comparative analysis.

Contrary to its photosynthetic function in C_4 plants, in C_3 plants PEPC is non-photosynthetic and involved in diverse functions, for instance in the replenishment of the tricarboxylic acid cycle, in carbon-nitrogen interactions, carbon storage or pH maintenance (O'Leary *et al.*, 2011). This multifaceted role of PEPC in plant metabolism is underlined by a complex regulation system (O'Leary et al., 2011).

In order to carry out all these functions PEPC in all investigated angiosperms is encoded by small gene families (Lepiniec et al., 1994) In *Flaveria* the PEPC gene family is composed of four genes *ppcA* to *ppcD*, with the latter being a bacterial type PEPC and out of the scope of this manuscript (Hermans and Westhoff, 1990, 1992). The *ppcA* gene of C_4 *Flaveria* species encodes the C_4-type photosynthetic PEPC isoform, which is highly expressed in the mesophyll cells of C_4 *Flaveria*, while its evolutionary orthologue in C_3 *Flaveria* species encodes a typical non-photosynthetic C_3 type PEPC (Hermans and Westhoff, 1992; Svensson et al., 2003). *ppcB* and *ppcC* of both C_3 and C_4 *Flaveria* species encode non-photosynthetic PEPC isoforms with ubiquitous expression pattern (Ernst and Westhoff, 1997; Svensson et al., 2003).

As a primary enzyme in C_4 photosynthesis or primary metabolism in C_3 plants, PEPC is controlled both metabolically and post-translationally. Malate, aspartate or oxaloacetate function as negative feedback regulators, while the PEPC activity increases in the presence of triose- and hexose-phosphate (Rajagopalan et al., 1994; Law and Plaxton, 1997; Bläsing et al., 2002; Svensson et al., 2003; Takahashi-Terada et al., 2005; Jacobs et al., 2008).

The C_4 *ppcA* PEPC isoform has been shown to have a lower sensitivity for the allosteric inhibitors mentioned above, a characteristic which is explained by a single amino acid mutation from arginine-884 to glycine (Paulus et al., 2013).

Both phosphorylation and ubiquitination have been indicated as post-translational modifications that affect PEPC activity. Mono-ubiquitination of PEPC has been identified only recently; it was found to be tissue specific and to influence the feedback regulation of PEPC by various metabolites (Uhrig et al., 2008; O'Leary et al., 2011).

Phosphorylation of the photosynthetic PEPC of C_4 and CAM plants has been known for decades (Budde and Chollet, 1986; Nimmo et al., 1986; Jiao and Chollet, 1989). Phosphorylation of PEPC was shown to be induced by several factors e.g. light or availability of nitrogen and phosphorus (Leport et al., 1996). Diurnally regulated phosphorylation is evident for CAM plants, *Arabidopsis*, maize, *Sorghum*, *Digitaria sanguinalis* and wheat (Jiao and Chollet, 1988; Echevarria et al., 1990; McNaughton et al., 1991; Duff and Chollet, 1995; Giglioli-Guivarc'h et al., 1996; Nimmo, 2000; Fontaine et al., 2002). However, these same studies outline some crucial differences among species when grouped by photosynthesis type. In C_4 plants phosphorylation of PEPC is induced by light and decreased in the dark (Ueno et al., 2000). In CAM plants both the expression of the *PPCK* transcript and the phosphorylation level of PEPC peak at night (Nimmo, 2000) while in C_3 plants mixed results were found (Vanquy et al., 1991; Duff and Chollet, 1995; Li et al., 1996; Fukayama et al., 2006; Meimoun et al., 2009).

In context of the changes occurring during C_4 evolution, it is crucial to understand if the above differences result from the different photosynthetic type among the studied species or are due to species individuality. The effects of malate and glucose 6-phosphate on PEPC activity are modulated by phosphorylation, which decreases PEPC's inhibition by malate and renders the enzyme more sensitive to the activator glucose-6-phosphate. Thus phosphorylation seems to broaden the conditions under which PEPC can be active and seems to decrease the K_m, while the V_{max} of PEPC is only modestly affected (Duff et al., 1995; Tovar-Mendez et al., 2000; Takahashi-Terada et al., 2005).

Interestingly the knock-down of PEPC phosphorylation in *F. bidentis* (C_4) by RNAi inhibition of phosphoenolpyruvate carboxylase protein kinase (PPCK) did not affect CO_2 assimilation rate, although the response to malate was observed as in previous studies (Furumoto et al., 2007). The latter indicates that phosphorylation is probably required to adjust PEPC activity in response to various signals or fluctuating conditions, or to potentially mediate the interaction with another protein e.g. 14-3-3 proteins (O'Leary et al., 2011; Grieco et al., 2012). The implicated phosphorylation site is located on the N-terminus of PEPC and has been identified as Serine-6, -8, -11 or -15 among various plants (Jiao and Chollet, 1990; Wang et al., 1992; Lepiniec et al., 1994; Chollet et al., 1996; Vidal and Chollet, 1997; Tripodi et al., 2005).

PEPC phosphorylation is carried out by a kinase found to be ranging in size from 30 to 39 kDa in diverse species (Chollet et al., 1996) named phosphoenolpyruvate carboxylase protein kinase (PPCK). PPCK is most similar to a Ca^{2+}-independent serine/threonine protein kinase. It has been classified as belonging to the Ca^{2+}-sensing calmodulin-like regulated CDPK-SnRK superfamily (Hartwell et al., 1996; Halford and Hardie, 1998; Hartwell et al., 1999; Hrabak et al., 2003). The kinase is composed of a single core-domain, without broader N- and C-terminal extensions. The interaction between kinase catalytic

domain and substrate has been studied for other CDP-kinases whose catalytic domain is composed of two lobes: an N-terminal lobe containing ß-sheets and a C-terminal lobe composed of α-helices. It is thought that the ATP molecule is bound in a cleft between these two lobes while the substrate binds along the cleft (Ubersax and Ferrell, 2007). On the substrate protein, the amino acids flanking the phosphorylation site, probably up to four residues on either side, are thought to contribute to substrate recognition. However residues further away from the phosphorylation-site can interact with portions of the kinase just outside the active site (Ubersax and Ferrell, 2007). Echevarria and Vidal list the consensus domain "acid-base-X-X-S-I-D-A-Q-L-R" as characteristic for phosphorylation by PPCK (Echevarria and Vidal, 2003).

The protein amount of PPCK seems to be regulated at the level of transcription and protein synthesis/turnover in C_4 maize and seeds from C_3 plants (Jiao et al., 1991; Hartwell et al., 1999; Osuna et al., 1999; Echevarria and Vidal, 2003). In addition the activity and/or expression of the PPCK protein are known to be influenced by reduction of di-sulphide bonds (Saze, 2001; Tsuchida et al., 2001; Nimmo, 2003), a proteinaceous inhibitor (Nimmo, 2000) and inhibition by malate (Borland et al., 1999; Uhrig et al., 2008).

PPCK proteins were found to be encoded by a small gene family (*PPCK*), which in most investigated species is composed of at least two isoforms (Fontaine et al., 2002; Nimmo, 2003; Fukayama et al., 2006; Shenton et al., 2006). The genus *Flaveria* is an exception with only one characterized *PPCK* gene, encoding a 31,8 kDa protein identified in *Flaveria trinervia* and *F. bidentis* (Tsuchida et al., 2001; Furumoto et al., 2007). This raises the question whether these species are a true exception in possessing only a single PPCK gene or if the nucleic acid sequences for further isoforms simply have been missed so far. Furthermore the *PPCK* genes are not described in the C_3 and C_3-C_4 intermediate *Flaveria* species. The present study therefore aimed at identifying the genomic representation of *PPCK* genes in the genus *Flaveria* and to investigate the evolutionary characteristics related to C_4 photosynthesis.

RESULTS

PPCK is encoded by two-member gene family in *Flaveria* and both genes were altered in the process of C_4 evolution

To investigate the number of *PPCK* isoforms in Flaveria, previous leaf transcriptome data for the two C_3 species *F. pringlei* and *F. robusta*, the C_3-C_4 intermediate *F. ramosissima* and the two C_4 species *F. trinervia* and *F. bidentis* (Gowik et al., 2011) were complemented by RNA seq data that included four additional C_3-C_4 intermediates, namely *F. chloraefolia*, *F. pubescens*, *F. anomala* and *F. brownii* (Mallmann et al, unpublished). In all nine *Flaveria* species we consistently found two *PPCK* isoforms (Supplemental file 1). A phylogenetic analysis of the PPCK proteins showed that the two isoforms clearly differ from each other and belong to two separate branches of a maximum-likelihood phylogram (Figure 1). We will thus refer to them as PPCKA and PPCKB respectively.

The PPCKA isoform branching reflects the photosynthetic type of the species. At the base of the branch are the sequences from the C_3 plants and the evolutionary time increases towards the sequence form the C_4 species. In the case of PPCKB there is high similarity between the PPCKB isoforms from the species with C_3 and intermediate photosynthesis type. These results are similar to previous phylogenetic analyses of PEPC isoforms in *Flaveria* (Svensson et al., 2003) (Figure 1).

Extending the analysis to a selection of known PPCKs from C_3, CAM and C_4 species, we found that the PPCKs cluster primarily based on their systematic relationships rather than photosynthetic type or PPCK isoform (Supplemental file 2). All PPCK isoforms from the genus *Flaveria* cluster together, just as the PPCKs from e.g *Magnoliophyta* and the rosids. Secondary, within the systematics groups further organization driven by PPCK isoform is observed with *Flaveria, Fabales* (lotus and soja), and the order Brassicales (*Arabidopsis thaliana*- C_3*, Cleome gynandra*- C_4) (Brautigam et al., 2011). Similarly in the family *Poaceae*, here represented by maize (C_4), sorghum (C_4), setaria (C_4) and rice (C_3), all PPCKs are grouped together but within this clade the subdivision is driven by the high similarity of PPCK1 proteins. The data indicate that the PPCK isoforms have evolved independently, after speciation of the highly divergent species (Supplemental file 2).

In order to pin-point any evolutionary changes among the *Flaveria* species that were associated with the transition from C_3 to C_4 photosynthesis in this genus, we compared the amino acid differences in the PPCK sequences. Amino acid exchanges of interest were defined as conserved changes found in both C_3 species (*F. robusta* and *F. pringlei*) on one side and both C_4 species (*F. bidentis* and *F. trinervia*) on the other, while the exchange could be absent or present in the intermediate species. Following these guidelines three types of mutations were observed: (1) exchanges found in all intermediate and C_4 species, (2) exchanges found in some intermediate and C_4 species and (3) exchanges found only in the C_4 species. In the case of PPCKA exchanges of type 1 are E80D, G160A, G165R, G185M and L257F; type 2 are E163D and S270N and type 3 exchanges are A4T, I135L, S147G, Y211H, E217D and E273K (Figure 2, Table 1).

In line with the parallel evolution of the PPCK isoforms seen in the phylogenetic analysis (Supplemental file 2), these amino acid exchanges seem characteristic for the genus *Flaveria* (Table 1, Supplemental file 3). A comparison of sequences from the order *Brassicales* and the family *Poaceae*, both of which have C_3 and C_4 representatives, shows a similar trend of C_3 to C_4 exchanges primarily for PPCK1 (Supplemental file 3). However it must be noted that for *Brassicales* only single PPCK1 representatives from C_3 and C_4 species are available, and in *Poaceae* we have a single C_3 PPCK1 representative from rice (Supplemental file 2). It is thus difficult to conclude if the observed differences are conserved among related species with same type of photosynthesis or if we are seeing species specific characteristics. The presence of the 3 PPCK1 sequences in *Poaceae* originating from C_4 species which all show the same amino acid difference from the rice PPCK1 sequence support the hypothesis that we are observing C_3 to C_4 related changes, like in the genus *Flaveria* (Table 1).

In order to visualize the distribution of these exchanges in *Flaveria* we first summarized the mutations in a schematic model (Figure 2B) and could observe that most of the conserved differences between C_3 and C_4 species are in the C-terminal part of the kinase. Subsequently the tertiary structure from PPCKA from *F. pringlei* (C_3) and *F. trinervia* (C_4) was modelled using Swiss-Model(de Brevern et

al., 2000; Joseph et al., 2010; Benkert et al., 2011), the models were aligned in iPBA (Gelly et al., 2011) and displayed using PyMol version 1.3. The sequence similarity of the templates used for modeling was 0.40 (Figure 2C). The structure of the kinase was composed of an N- and C- terminal lobe, with an unstructured protein region between them. The cluster of amino acid exchanges on position 135, 160, 163, 165 and 185 is found on the unstructured protein region in the three dimensional model, while most other exchanges were located on the C-terminal lobe (figure 2B, C).

For PPCKB only three amino acid exchanges were found among C_3 and C_4 species according to the above stated criteria. S109A is found in some intermediates and the C_4 species (type 2), while T49I and S123G are found in the C4 species only (type 3) (Figure 2A, Table 1). The smaller number of differences in this isoform, which are related to the photosynthetic type of the species, illustrates that higher evolutionary pressure was put in optimization of the structure of PPCKA.

PPCK genes are diurnally expressed in *Flaveria* and their levels vary between C_3 and C_4 species

The differences in the protein sequences among PPCK isoforms, led to the question if changes in regulation of transcript abundance were also present among species with C_3 and C_4 photosynthesis. We therefore investigated *PPCK* transcript levels by a diurnal harvest experiment and profiled the samples by Illumina RNA-Seq. Plants from *F. trinervia* (C_4) and *F. pringlei* (C_3) were grown in a 10/14 hour (light/dark) period and harvesting was performed every 4 hours from the onset of light (Figure 3).

Figure 3 depicts that *PPCKA* and *PPCKB* transcripts exhibit strikingly different patterns of accumulation. Both *PPCKA* and *PPCKB* transcripts, in *F. trinervia* (C_4) as well as in *F. pringlei* (C_3), fluctuate with a phase shift of 12 hours. Both transcripts can be classified as diurnally regulated, since a significant difference in the p-values ($p \leq 0.05$) and at least a 3 fold change can be simultaneously observed between a reference time (4h light) and 3 consecutive time points (Facella et al., 2008). However, *PPCKA* and *PPCKB* transcripts differ in the temporal order of maxima and minima. While *PPCKA* transcript amounts of *F. trinervia* and *F. pringlei* reach their maximum during the day, *PPCKB* transcripts peak in the dark.

The abundance of *PPCKA* transcripts differs between the species with varying photosynthesis type. At the time of expression maximum (4h after onset of light) the amounts of *PPCKA* transcripts are at least 10-fold higher in *F. trinervia* (C_4) than in *F. pringlei* (C_3). As it can be seen in Figure 3B the relative levels in *F. pringlei* seem to drop a little bit faster than in the C_4 species resulting in higher differences in later time points (Figure 3, Supplemental file 4).

This increase in *PPCK* expression was only found in *PPCKA*. In the case of the *PPCKB* transcript, we found that the amounts in *F. pringlei* (C_3) are generally higher than in *F. trinervia* (C_4) and the differences among species are less expressed (Figure 3C, Supplemental file 4). During the night, at the time of expression maximum for *PPCKB,* this transcript is 2 to 6 times higher in the C_3 species. The only time point which showed higher levels of *PPCKB* in the C_4 species was at the end of night at 0/ 24 h (Figure 3D).

Comparison of the isoform transcripts within each species shows that at midday the level of *PPCKA* in *F. trinervia* (C_4) is more than 160 times higher than the level of *PPCKB*, whereas in *F. pringlei* (C_3) the *PPCKA* transcript is only 5 times more abundant than *PPCKB*. During the night, *PPCKB* is twice as abundant as *PPCKA* in *F. trinervia* (C_4) and 44 times more abundant in *F. pringlei* (C_3).

In the context of evolution of C_4 photosynthesis, it is the PPCKA isoform that shows the characteristic increase in transcription seen for many transcripts involved in this process (Gowik et al., 2011; Christin et al., 2013) suggesting that the *PPCKA* isoform is responsible for the day-time phosphorylation of PEPC. On the other hand, in the C_4 species there is a decrease in the amounts of PPCKB transcripts during the night, while this isoform is still relatively highly expressed in the C_3 species. This supports a hypothesis for differential importance of these isoforms in the two compared species, which might be reflected in the levels of PEPC phosphorylation.

The phosphorylation pattern of PEPC from species with C_3 and C_4 photosynthesis is consistent with the characteristic expression patterns of the PPCK isoforms

To confirm the above hypothesis we started two approaches: (i) investigation of the protein levels of PPCKA and (ii) monitoring of the in vivo phosphorylation state of PEPC in *F. trinervia* (C_4) and *F. pringlei* (C_3). In the first approach we generated antibodies against PPCKA isoforms, which showed satisfactory specificity, however due to their very low sensitivity they were not usable in a complex plant extract and could only be applied against higher amounts of recombinant protein (Supplemental file 6). We then attempted to monitor the kinase protein using Selected Reaction Monitoring -SRM (Lange et al., 2008) but the kinase levels were below the detection limit as well.

Consequently, we took the transcript levels of PPCK as a proxy for the protein amount and proceeded to monitor the *in vivo* phosphorylation state of the substrate using the phosphorylated and non-phosphorylated version of the peptide which contains the Serine-11 phosphorylation site LA(pS) IDAQLR and LASIDAQLR respectively (Svensson et al., 1997; O'Leary et al., 2011) (Supplemental file 5). Using the plants harvested over the course of 24 hr, we extracted proteins and spiked known concentrations of the labeled synthetic peptides mentioned above. The signal from the standard was used to determine the amount of the native PEPC peptide, following which the ratio of phosphorylated to non-phosphorylated PEPC was calculated.

The results in Figure 4 show one crucial difference between *F. trinervia* (C_4) and *F. pringlei* (C_3): while the phosphorylation of the PEPC from the species with C_4 photosynthesis peaks during the day, the PEPC from the C_3 species does not, instead it maintains a certain level of phosphorylation throughout the entire day and this level drops only briefly after the end of night.

Comparison of the phosphorylation trend for PEPC from *F. trinervia* (C_4) with the transcript levels of *PPCKA* and *PPCKB* from this species, shows that the peak of phosphorylation closely matches the accumulation pattern of the *PPCKA* transcript, while at night phosphorylation levels are negligible (Figure 4A, Supplemental file S4).

On the other hand, PEPC from *F. pringlei* (C_3) maintains a certain phosphorylation level throughout the entire day fitting the expression profiles of both *PPCKA* and *PPCKB*, the maximum phosphorylation point is during the dark period (Figure 4B, Supplemental file S4).

In the context of C_4 evolution the results show that the *Flaveria* plants with C_4 photosynthesis have most probably abolished the phosphorylation of PEPC in the dark, while this action is still important in the C_3 species where PEPC does not play a photosynthesis-related role but functions in other primary-metabolic pathways.

Functional characterization of recombinant PPCKA proteins shows cross-phosphorylation of PEPC among *Flaveria* species and indicates changes in activity

The first part of this result section outlined the changes that have occurred in the amino acid sequence of PPCKA, which we believe is responsible for the phosphorylation of the photosynthetic PEPC isoform in *F. trinervia*. As can be seen on Figure 2B many of these amino acid exchanges occurred near or at sites which are implicated in binding either ATP or PEPC. Thus, we investigated if these changes are sufficient to abolish substrate recognition between species with C_3 and C_4 photosynthesis.

To this aim we first isolated the *PPCKA* cDNAs from *F. pringlei* (C_3) and *F. trinervia* (C_4) based on existing studies (Tsuchida et al., 2001) and our sequencing data, and purified the recombinant PPCKA proteins to near homogeneity (Supplemental file 6).

In two independent experiments, either recombinant *ppcA* PEPC from *F. pringlei* (C_3) or from *F. trinervia* (C_4) was used as a substrate (obtained from Paulus *et al.,* 2013). Regardless of the origin of PEPC, there was successful phosphorylation by both *F. pringlei* and *F. trinervia* PPCKA recombinant proteins (Figure 5). Consequently the inter-species cross-phosphorylation is not a trait restricted only to the C_4 species.

We find that the signals from the combination of *F. trinervia* (C_4) PPCK and PEPC were much stronger than the signal resulting from the phosphorylation of the *F. trinervia* (C_4) PEPC with the *F. pringlei* (C_3) PPCK (Figure 5B). In comparison, even though both kinases could successfully phosphorylate the PEPC from *F. pringlei* (C_3) the signals were of similar intensity (Figure 5A).

Consequently, although the PPCKs can cross-phosphorylate among species, the stronger signals from the *F. trinervia* (C_4) PPCKA and PEPC suggest that the changes in both proteins are important for the observed difference in phosphorylation.

To identify possible changes near/on the substrate recognition site, we investigated the N-terminus of PEPC in the 9 sequenced *Flaveria* species and the PEPCs from *Brassicales, Amaranthaceae* and *Poaceae* – all of which contain C_3 and C_4 species. The conserved motif from this interspecies comparison was identified as –E-K-X-X-S-I-D-A-Q-L-R- (Figure 6A; Suppelmental File 8) (Bailey and Elkan, 1994; Echevarria and Vidal, 2003).

Focusing on the genus *Flaveria*- where the N-terminal region of the photosynthetic *ppcA* PEPC is well conserved, we found a single L6V exchange which is different between the C_3 and C_4 *Flaveria* species (Figure 6B). The species in the genus are phylogenetically subdivided in two clades- A and B, with the C_4 species belonging to clade A, whereas the furthest evolved species in clade B is the C_4-like *F. brownii* (McKown et al., 2005; Gowik et al., 2011). The exchange L6V was found in the C_4 species *F. trinervia* (C_4) and *F. bidentis* (C_4)- clade A, as well as *F. brownii* (C_4-like) and the C_3-C_4 intermediates *F.pubescens* and *F.anomala*- all belonging to clade B. Thus, valine has been introduced in the further evolved C_3-C_4 intermediates, and the C_4 species in two independent evolutionary paths. Furthermore, the same type of exchange was observed in the C_4 species in *Amaranthaceae*- which represent a totally different origin of C_4 photosynthesis, but was not found in the *Brassicales* or *Poaceae* (Supplemental File 8).

Surprisingly a second amino acid exchange, Q15H, was discovered in the C_4 non-photosynthetic *ppcB* PEPC isoforms in *Flaveria* (Figure 6A, Supplemental File 8).

Using the recently deduced crystal structure of PEPC from plants (Paulus et al., 2013), we used virtual mutagenesis to display the above mentioned C_4- related changes (L6V and Q15H) onto the *F.pringeli* PEPC tetramer (Figure 6C). This *in silico* investigation shows that in both cases, the amino acids lie on the surface of the protein and are exposed to the solvent. Consequently they could be accessible to interact with a PPCK. The findings show that the C_4 related changes, which have been identified to take place in PEPC, also extend to its N-terminus.

DISCUSSION

This study underlines the existence of a two-member PPCK family in the genus *Flaveria*, whose isoforms are expressed in an inverse day/night manner. PPCKA underwent changes characteristic for proteins recruited to C_4 photosynthesis in *F. trinervia* (Gowik et al., 2011; Christin et al., 2013), while PPCKB seems to have a higher importance in the C_3 *Flaveria* species.

Parallel evolution of PEPC and PPCK proteins

Phylogenetic analyses for PPCK (Supplemental file 2) and PEPC (Bläsing et al., 2002; Svensson et al., 2003) show that in both cases the ancestor proteins diverged at the time of speciation, and the respective protein isoforms evolved in parallel within each species. This is in agreement with the polyphyletic evolution of C_4 photosynthesis, where multiple independent origins lead to the full C_4 photosynthesis (Sage et al., 2012).

PPCK isoforms and the *ppcA* PEPC from species with C_4 photosynthesis, are always found on branches marking longest evolutionary distance. This illustrates (i) the ongoing evolutionary process of both kinase and substrate (Wang et al., 2009) and (ii) the presence of a large number of C_4 related amino acid exchanges, as we were able to demonstrate with the case of PPCKA (Figure 2A). PPCKA and *ppcA* PEPC are thus indicated as the kinase-substrate pair that was under strong evolutionary pressure.

In contrast, PPCKB does not show large differences between species with C_3 and C_4 photosynthesis (Figure 1) nor do the *ppcB* and *ppcC* PEPC isoforms from *F. trinervia* (C_4), which is marked by shorter evolutionary distance in previous studies (Bläsing et al., 2002; Svensson et al., 2003).

Changes in the quantity of PPCK transcripts in leaves of C_3 and C_4 *Flaveria* species; potential mechanisms for regulation of transcript stability and transcription rhythms

Suitable expression patterns seem to be one of the prerequisites necessary for the recruitment of genes in the C_4 pathway (Christin et al., 2013). We showed that *PPCKA* was present in the C_3 *Flaveria* species which diverged earlier in evolution and already there it has a day-time maximum in transcript accumulation. This expression pattern is much more suitable for C_4 photosynthesis than the night-time peak in transcription found for *PPCKB*. Similar to the case of the *ppcA* (Ernst and Westhoff, 1997), *PPCKA* also showed the characteristic increase in expression between C_3 and C_4 species. Taken together with the phylogenetic analysis, this change in transcriptions supports a conclusion that *PPCKA* is the isoform that was recruited in the transition from C_3 to C_4 photosynthesis mode.

PPCKB has opposite characteristics from *PPCKA*. In addition to the maximal transcript accumulation at night in both species, a decrease in transcript levels is observed from the C_3 to the C_4 species (Figure 3). Thus, it appears that this isoform has decreased importance in the leaves of the newly evolved C_4 *Flaveria* species, which was also reflected in the decreased phosphorylation levels of PEPC during the night. On the other hand, PPCKB is fully functional in the C_3 species, where night-time phosphorylation is relatively abundant. As a side-finding, the existence of day- and night-time expressed PPCK in the C_3 species, especially if this expression pattern for the PPCK family is confirmed for other C_3 plants, is an explanation for the varying PEPC phosphorylation results observed among various C_3 plants (Fukayama *et al.*, 2006; Meimoun *et al.*, 2009).

PEPC has been found to be under the influence of the circadian clock in a number of plants (O'Leary et al., 2011). The closely related function of *PPCKA* in *Flaveria*, in addition to the similar diurnal expression pattern in C_4 photosynthesis allows for speculation that the same will be the case for this kinase. Indeed extended light experiments in CAM plants show such regulation for PPCK (Taybi et al., 2000), but no evidence for circadian regulation has been found in maize (C_4) (Shenton et al., 2006). In soybean (C_3) *PPCK* has been found under circadian regulation in leaves but not in roots (Sullivan et al., 2005), and in rice (C_3) it appears that there is no circadian regulation (Fukayama et al., 2006). Additional experiments are needed to test whether the accumulation of *PPCKA* and *PPCKB* transcripts in *Flaveria* is controlled by the circadian clock. An analysis of cis- regulatory elements in the *PPCK* promoter regions would also be useful to investigate any regulatory differences between the *Flaveria* species with varying photosynthetic type (Mockler et al., 2007).

The main mechanisms influencing *PPCKA* transcript levels could be (i) increased transcription rate and/or (ii) increased transcript stability. We inspected the 3' UTR region from the *PPCK* isoforms from the available data for the presence of the 10 most significant destabilization or stabilization 6-mers (Tsuchida et al., 2001; Narsai et al., 2007). On average 80% of all identified 6-mers in the PPCKA 3'UTR were destabilization motifs implicated in rapid RNA turnover, while this number was a little bit lower 72% for PPCKB (Supplemental File 1, disregarding the partial UTR sequences). Thus, in both cases we find indications for rapid transcript turnover, that agree with a previously postulated theory (Tsuchida et al., 2001).

The evolutionary changes in PPCKA proteins and their potential functional significance

Eight of the thirteen amino acid exchanges conserved among the PPCKA proteins from C_3 and C_4 plants are concentrated in the region of amino acids 135-217 (Figure 2). Some amino acids in this region are tentatively indicated to participate in ATP and substrate binding based on the NCBI Conserved Domains Database- GenBank: BAB71853.1 and BAF48321.1; cd05123; Marchler-Bauer *et al*, 2013. This is an indication of the strong evolutionary pressure to optimize either the kinetic properties and/or activation mechanism of the kinase.

Some of the potentially important differences are at positions 147, 160-165, and 211. At position 147 there is a serine to glycine exchange. Serine is a neutral amino acid that can be subject to phosphorylation but is also much larger than glycine and this could be a similar mechanism as the Arg 884 to Gly exchange in the PEPC molecule (Paulus et al., 2013). At positions 160, 163 and 165 three amino acid exchanges were identified. The first two are exchanges among amino acids with similar chemical properties, but the last exchange on position 165 replaces glycine, a small non-polar amino acid, with arginine, a larger basic (polar) amino acid, whose side chain is fully protonated at neutral pH, consequently bringing in positive charge at this position (Taniguchi, 2010).

This cluster of point mutations is followed by an arginine to methionine exchange at position 185. In the three-dimensional structure Arg185 is found opposite the region 160-165. Together with a third amino acid exchange at position 135 these amino acid exchanges are predicted to lie in an unstructured protein region, partly annotated as the "activation loop (A-loop)" (Figure 2). Current structural annotations state that phosphorylation of the A-loop, results in transition from disordered to ordered state of the kinase and is part of the kinase activation mechanism (cd05123, Marchler-Bauer *et al*, 2013). As the putative A-loop contains Ser (positions 155, 168), Thr (positions 166, 173) and Tyr (175) which are all succeptible to phosphorylation- the near-by exchanges might influence the interaction with another regulating protein. Towards the C-terminus, at position 211 instead of a tyrosine in C_3 plants, which has been indicated to play part in substrate binding, in the C_4 plants, is a histidine (basic amino acid), which is known to function as an acid/base catalyst (Taniguchi, 2010). However, we also must bear in mind that the crystal structures onto which PPCKA was modeled share only about 40% sequence similarity and that PPCK has not been crystalized from any plant species. The latter will be necessary for proper determination of the influence of these amino acid exchanges in the tertiary structure of PPCK.

If we assume that the evolutionary trajectories are accompanied by gain-of-fitness and that evolution of the PPCK gene family optimized the C_4 evolutionary processes (Sage et al., 2012; Heckmann et al.,

2013), then amino acid exchanges observed in the PPCKA sequences from various species could infer the order of the structure optimization. Starting from a common ancestor in the genus *Flaveria,* species with C_3 photosynthesis diverged earlier, whereas the C_4 species diverged later. If an amino acid exchange is conserved among all intermediates and the C_4 species (type 1), then this point mutation probably occurred before those found only in the C_4 species (type 3; Table 1). Thus, in the case of PPCKA, we believe that the exchanges E80D, G160A, G165R, R185M and L257F (type 1) probably precede the amino acid exchanges A4T, I135L, S147G, Y211H, E217D and E273K (type 3). In the two cases where the C_4 residue is found in some but not all intermediate species- E163D and S270N (type 2) it is difficult to determine what happened. Creating a series of mutated PPCKA proteins will be needed to investigate the importance of each of these amino acid exchanges on the kinase enzymatic activity.

The presented recombinant PPCKA experiments show that these amino acid differences are not enough to abolish substrate recognition among the species in the genus *Flaveria* (Figure 5) which is in agreement with previous findings where PPCK proteins from several plants can phosphorylate exogenous PEPC (Li and Chollet, 1994; Tsuchida et al., 2001; Ermolova et al., 2003). Since a higher signal results from the C_4 form of both kinase and substrate, it is the parallel optimization of these molecules that is responsible for this increased activity (Figure 5B). In this context the so-far ignored amino acid differences in the N-terminus of PEPC, and in the vicinity of the phosphorylated Serine, (Figure 6, Supplemental file 8) gain in importance and will require further study (Echevarria and Vidal, 2003). Simulation of the C_3 to C_4 exchanges on the N-terminal part of PEPC which take into account that PEPC is a tetramer show that despite the macromolecular conformation, these amino acids would still be accessible during a protein- protein interaction (Figure 6C, Paulus et al., 2013). The L6V exchange on the N-terminus of *ppcA* PEPC characteristic for the intermediate and C_4 species, is one of two exchanges in the otherwise conserved N-terminal region in *Flaveria* (Supplemental file 8). The side chains of these two amino acids have similar properties but the smaller side chain from Valine might allow easier uncoupling of the kinase from the substrate- by providing larger separation between the PEPC and PPCK interacting residues. Consequently this might be one of the reasons for observing a stronger signal in the in vitro phosphorylation assay between the C_4 recombinant PEPC and PPCK proteins. *ppcB* PEPC, on the other hand showed a Q15H exchange only in the C_4 species (Figure 6,). This introduces a positive charge close to S11 but the true consequence of this exchange will only be known after further investigation as it will depend on the interacting residue from PPCK. Co-crystallization of both proteins should uncover the amino acid residues involved in the protein-protein interaction and allow the investigation of the observed amino acid exchanges between isoforms from C_3 and C_4 plants.

Although the evolution of the photosynthetic PEPC from C_4 *Flaverias* has been extensively studied and the influence of crucial amino acid exchanges on enzymatic activity has been clarified (Bläsing et al., 2000; Engelmann et al., 2002; Engelmann et al., 2008; Paulus et al., 2013), no such information exists for PPCK at this time. Additionally, as the cumulative findings from previous and these studies reinforce the differences in phosphorylation patterns of PEPC from C_3 and C_4 species, and at a time with ongoing efforts for engineering C_4 traits in C_3 crops, we can only stress the need for detailed knowledge on the regulation of one of the key enzymes of the C_4 pathway.

MATERIALS AND METHODS

Plant material

Plants were initially grown at greenhouse conditions and before harvest for the Illumina and SRM experiment they were acclimated for the last 10 days in an open phytochamber at 120 µmol quanta m^{-2} s^{-1} irradiance with a 10-h photoperiod, at 19 °C night and 22 °C day. Leaf blades were harvested from wild-type 4 to 6 week old plants from *F. trinervia, F. ramosissima* and *F. pringlei.* Leaf samples were collected at 4 hr intervals over a 24 hr period. Plant material was immediately frozen in liquid nitrogen for further storage at –80 °C. Three separate biological replicates for each time point were ground in liquid nitrogen prior to the respective RNA or protein preparations.

Library construction and Illumina sequencing

F. trinervia, and *F. pringlei* were used for transcriptome sequencing. Total RNA was extracted from *Flaveria* leaves using TRIsure Reagent according to the protocol for plant tissue (Bioline) transcriptome sequencing. Digestion with DNAse was performed for 15 min. The RNA was treated with phenol and chloroform and precipitated over night with a sodium acetate/ isopropyl alcohol solution. The RNA was washed with 70 % ethanol and dissolved in H_2O. The RNA concentration was determined with the nano-drop (PEQLAB Biotechnologie GmbH).

Total RNA was set to a final concentration of 1 µg in 50 µL as starting volume. The DNA libraries were generated according to the manufacturer´s TruSeq™ RNA Sample Preparation Kit via the Low-Throughput Protocol (Illumina Proprietary Catalog # RS-930-2001, Part # 15008136 Rev. A, November 2010). RNA aliquots and libraries were validated for qualification and quantitation with the Agilent2100 bioanalyzer (Agilent Technologies). Validated DNA libraries were normalized and pooled to a final concentration of 2 nM. Clusters were generated with the TruSeq SR Cluster Kit v2 according to the Reagent Preparation Guide with the Illumina cBot device. The single read sequencing was performed with the Illumina HiSeq2000.

Transcriptome mapping and analysis

To obtain the abundance of reads for the kinase, clean Illumina reads were aligned against *F. trinervia* cDNA sequences either obtained from public databases (*PEPCK-A* Accession: AB272061) or during the present work (Supplemental file 1). In case of *PEPCK-B*, we used a contig generated by de novo assembly of 454 and Illumina reads (Gowik et al. 2011; Mallman et al., unpublished) (Supplemental file 1). Reads were mapped with Bowtie2 (version 2.0.6.) using the end to end mode with sensitive settings and allowing one mismatch per seed for multiseeding (Langmead and Salzberg, 2012). We counted the number of unambiguous best hits per transcript and normalized these values to reads per million mapped reads and kb transcript length (RPMK).

Means and standard error of the mean (SEM) was calculated from three biological replicates. The data were evaluated with a one-way ANOVA by a multiple comparison using a Tukey-test (multiplicity

adjusted $p \leq 0.05$). In parallel, with a two-tailed paired Student's *t* test we obtained significance for transcripts of two time points, with major differences in abundance ($p \leq 0.05$). Statistical analyses were performed with the program GraphPad PRISM 6 (GraphPad Software, Inc., La Jolla, CA).

Isolation, Cloning and sequencing of *PPCK* for recombinant protein purification

Total RNA was extracted from *F.pringlei* and *F. trinervia* leaves using the RNeasy Plant Mini Purification Kit (QIAGEN), including the RNase-Free DNAse Set (QIAGEN). The RNA concentration was determined with the nano-drop (PEQLAB Biotechnologie GmbH). Synthesis of 5` and 3` RACE-ready first-strand DNAs was carried out according to the SMARTer™ RACE cDNA Amplification Kit (Clontech Laboratories) with 1 µg of total RNA using the SMARTScriber™ Reverse Transcriptase (Clontech Laboratories). 3´- and 5´-RACE-PCR reactions were performed using the Universal Primer A Mix (UPM; Clontech Laboratories) and the Smart-II-A-5´RACE oligonucleotides, respectively (Supplemental file 7). The gene-specific oligonucleotides were selected according to the available *PPCK* sequence from *Flaveria trinervia* (GenBank accession number AB065100). For the non-coding regions of *PPCKA* from *F. pringlei* were identified with the *PPCKA* 5´-end (PEPC-PK_Fp_5´RACE) and the *PPCKA* 3´-end (PEPC-PK_Fp_3´RACE) oligonucleotides. PCR reactions were performed with the Advantage 2 Polymerase Mix according to Clontech Laboratories. The obtained DNA fragments were cloned with the CloneJET™ PCR Cloning Kit (Fermentas). Several independent clones were verified via PCR with the oligonucleotides pJET_fw and pJET_rew (Supplemental file 7)**.** Potential positive clones were verified by DNA sequencing (LGC Genomics GmbH, Berlin, Germany). The cDNA full-length of the *PPCK* gene from *Flaveria* was amplified with the respective oligonucleotides PEPC-PK_CDS_Fp, PEPC-PK_CDS_Fra and PEPC-PK_CDS_Ft listed in Supplemental file 7**.** Additional restriction sites were amplified to *PPCK* with oligonucleotides PEPC-PK-Fp-5´-XhoI and PEPC-PK-Fp-3´-BamHI (*F. pringlei*) and PEPC-PK-Ft-5´-NdeI and PEPC-PK-Ft-3´-BamHI (*F. trinervia*) as listed in Supplemental file 7. All PCR reactions were carried out by means of the Phusion High-Fidelity DNA Polymerase (New England Biolabs). PCR products were sequenced, digested with XhoI/BamHI or NdeI/BamHI and cloned into the corresponding XhoI/BamHI-digested or XhoI/NdeI-digested pETEV15b (+) vector (pET system, Novagene, Madison, USA) together with the coding sequence of an amino-terminal His6-tag and TEV cleavage site. All generated plasmids were verified by DNA sequencing.

Alignments and phylogenetic analysis

Calculations of sequence identity between the PPCK proteins from *Flaveria* were performed with the website http://emboss.bioinformatics.nl. The phylogenetic analysis was performed using MEGA5 (Tamura et al., 2011; Hall, 2013) as listed in the respective Figures. Predicted protein parameters, for *PPCK* which has been isolated with RACE-PCR, were acquired using the ExPASy program (http://web.expasy.org/protparam).

Heterologous expression of recombinant PPCK in *E.coli*

The QIAexpressionist™ protocol was used for the heterologous expression of native protein (QIAGEN). *E.coli* BL21 (DE3) cells (Agilent Technologies) were transformed with *PPCK*-pETEV15b. Culture growth was performed using 2YT media (5 g L^1 NaCl, 10 g L^{-1} yeast extract and 16 gl^{-1} peptone) with ampicillin at 37 °C after reaching an optical density of 0.7. Expression of recombinant proteins was induced by the addition of 500 µM IPTG (isopropyl ß-D-thiogalactopyranoside) and cells continued to grow for 5 hours at 20 °C.

Affinity purification of the recombinant PPCK proteins

Recombinant PPCKA protein was purified by Ni-NTA affinity chromatography according to the manufacturers instruction (Protino, MACHEREY-NAGEL, Germany). In short, intact *E.coli* cells were incubated on ice and subsequently resuspended in lysis buffer (50 mM NaH_2PO_4; 300 mM NaCl; 1mg/mL lysozym, pH 8.0). The cells were disrupted by sonification. Native recombinant proteins were purified with a pre-equilibrated Ni-NTA (nickel- nitrilotriacetic acid) affinity column (Protino, MACHEREY-NAGEL) on the Äktaprime Plus system (GE healthcare). In order to protect protein integrity a gradient from 0 to 500 mM imidazole in 50 mM steps was used. Proteins were concentrated with the Amicon Ultra-10 centrifugal filter units (10 kDa cutoff, Millipore) and desalted and buffered in (50 mM Tris, 5 mM $MgCl_2$, pH 8.0) using PD-10 columns (GE healthcare).

Protein concentrations of purified recombinant proteins were determined with the standard procedure (Biorad) (Bradford, 1976) and examined by Coomassie staining and immunoblot by SDS-PAGE on a 12% polyacrylamide gel (Laemmli, 1970). Signals were quantified using the program Multi Gauge 3.0 (Fuji Photo Film Co., Ltd). The blotted nitrocellulose membrane was incubated with the primary antibody (anti-PPCK) in a dilution of 1:400 for 12-h at 4 °C. Immunoblot analysis was examined using the reagent of the chemiluminescence assay (SuperSignal West; Thermo Scientific).

After investigating the protein purity using SDS-PAGE and Coomasie staining, the intensity of the kinase bands from *F. pringlei* were compared to those of *F. trinervia* and the protein content used for the in vitro phosphorylation assays was adjusted accordingly to equalize the kinase amounts (Supplemental file 6).

Enzymatic activity of PPCK

In vitro phosphorylation activity of recombinant PPCKA was detected by labeling recombinant PEPC with ^{32}P in a 15 µL reaction mixture containing 50 mM Tris/HCl; 5 mM $MgCl_2$, 1 mM EGTA and 1 mM DTT, pH 8.0. Post 20 min incubation on ice, radiolabeled ATP was added as a final step to the reaction at a concentration of (1 µCi/mmol) [g-^{32}P]-ATP/reaction. The reaction run for 15 min at 30 °C and terminated by the addition of 2x SDS-Loading buffer (Laemmli, 1970). Samples were boiled at 100 °C for 1 min and centrifuged at 10,000g for 10 min before loading on a 10% SDS-polyacrylamide gel. Following electrophoresis the radioactive bands were autographically visualized on a FLA3000 Bio-Imaging Analyser (Fuji Photo Film Co. Ltd, Tokyo, Japan).

PPCK antibody synthesis

Peptides conserved among the PPCK from *Flaveria pringlei*, *F. ramosissima* and *F. trinervia* species were considered for the antibody synthesis. Rabbit polyclonal PPCK antibodies were raised against peptides (NH2-) CLQKEPKILHILG (-CONH2), PPCK residues 53-65 and (NH2-) CRLGIAHRDLKPDNV (-CONH2) PPCK residues 127-140 by Agrisera (Sweden). An additional anti-PPCK antibody was synthesized by Agrisera obtained from a mix of purified recombinant PPCK protein from the same *Flaveria* species. Unfortunately, although specific, the antibody sensitivity was determined to above 0.6 µg, which was sufficient for recombinant protein work, but not sensitive enough to detect PPCK in plant extracts.

Selected reaction monitoring (SRM): peptide selection, sample preparation, LC-MS/MS and quantitation

The amino acid sequences of PPCK and PEPC were aligned, and *in silico* tryptic peptides conserved among *F. trinervia* and *F. pringlei* were considered for SRM. Initial testing was performed on the recombinant proteins for *F. trinervia* and *F. pringlei* both in unmodified and phosphorylated state. For the phosphorylated PEPC test, proteins were phosphorylated as above, using 1µg kinase to 10 µg PEPC in a 7.5 µL reaction and non-radioactive ATP.

The peptides that showed good properties in the mass spectrometer were checked against the *Flaveria* transcriptomes (BLASTp) supplemented with the experimentally determined sequences of PEPC and PEPCK, as control. The BLASTp was performed with increasing e-value until the control peptide was found. The peptide covering the PEPC phosphorylation site is conserved among the three isoforms in *F. pringlei* (C_3) and among the two most abundant isoforms in *F.trinervia* (C_4) (Ernst and Westhoff, 1997) (Supplemental file 5). The peptides then underwent a second round of selection using plant extract, and those peptides, which could be successfully monitored in this complex sample were obtained as heavy peptides (^{13}C, ^{15}N) from Thermo Fisher Scientific. None of the peptides from PPCK could be monitored in complex plant extracts using our setup.

Proteins were extracted from *F. trinervia* and *F. pringlei* as already described (Willige et al., 2011) with minor modifications. In short, 300 mg leaf material was treated with 600 µL buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, Protease inhibitor cocktail (Sigma-Aldrich), phosphatase inhibitor cocktail (Serva) and 1 % (w/v) PVPP (polyvinylpolypyrrolidone). Proteins were precipitated overnight at -20 °C, using 5x volume of ice-cold acetone (Sigma-Aldrich). The protein pellet was dissolved in 6 M urea-2 M thiourea (Merck and Sigma-Aldrich, respectively) and 10 µg of protein was digested and desalted as in Arsova et al., 2012. Prior to desalting the synthetic peptides for PEPC (listed in Supplemental file S5) were added to the samples in species-specific amounts. Per 500 ng total protein the amount of synthetic peptides was: *F. trinervia* (non-phosphorylated 10 fmol, phosphopeptide 5 fmol); *F. pringlei* (non-phosphorylated 2.5 fmol, phosphopeptide 1.25 fmol).

The tryptic digests were measured using UHPLC system UltiMate 3000 RSLCnano in combination with a TSQ Vantage triple quadrupole mass spectrometer (both, Thermo Fisher Scientific, Bremen, Germany). 500 ng digested plant protein was loaded at a rate of 20 µL/min and separated on a 20 minute gradient (5 %-30 % B, where solution A is 0.1% formic acid in water; solution B- 80% acetonitrile, 10% trifluoroethanol, 0.1% formic acid in water), using a column of 75 µm internal diameter, 15 cm length, C18, 2 µm particle size (Acclaim PepMap RSLC, Thermo Fisher Scientific, Bremen, Germany). Samples were sprayed into the mass spectrometer by electro spray ionization. The resolution of Q1 and Q3 was set to 0.7 u FWHM. Cycle time was 1 s (dwell time: > 50 ms for each transition). Peptides were fragmented using Argon gas for collision-induced dissociation at collision energies specific for each peptide (Supplemental file S5). Instrument scan mode was iSRM, and 3 primary and 2 secondary fragment ions per peptide were collected; allowed time was 0.1 min, threshold intensity 300 counts, and dynamic exclusion was activated after 3 fragmentation events for 0.3 minutes.

Data was analyzed using PinPoint 1.3 (Thermo Fisher Scientific), signal to noise ratio was set < 10. In general 3 transitions per peptide for each light and heavy form were used for peak area summation (except FP peptides LASIDAQLR (Phos): 2 transitions per peptide). The phosphorylation of PEPC on S11 was monitored using LASIDAQLR and LA(pS)IDAQLR. The absolute concentration was calculated as $Light_{(area)}/Heavy_{(area)}$ * concentration of specific standard peptide, and the ratios of these two concentrations were taken as the ratio of phosphorylated to non-phosphorylated PEPC.

ACCESSION CODES

DNA Sequences of *PPCK* from C_4 *Flaveria* are available at the EMBL/GenBank data libraries with the accession numbers AB065100 (*Flaveria trinervia* PPCK) and AB272061 (*Flaveria bidentis* PPCK). The *Arabidospis thaliana* genes used as references for the transcriptome mapping are annotated as At1g08650 (*PPCK*1) and At3g04530 (*PPCK*2). Accession numbers for the phylogram were taken from deposited EMBL-GenBank data libraries according to Marsh et al. 2003, Nimmo, 2003, (Sullivan et al., 2004; Fukayama et al., 2006; Shenton et al., 2006).

REFERENCES

Akyildiz M, Gowik U, Engelmann S, Koczor M, Streubel M, Westhoff P (2007). Evolution and function of a cis-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. *Plant Cell* **19:** 3391-3402.

Arnold K, Bordoli L, Kopp J, Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22:** 195-201.

Arsova B, Zauber H, Schulze WX (2012). Precision, proteome coverage, and dynamic range of Arabidopsis proteome profiling using (15)N metabolic labeling and label-free approaches. *Mol Cell Proteomics* **11:** 619-628.

Aubry S, Brown NJ, Hibberd JM (2011). The role of proteins in C(3) plants prior to their recruitment into the C(4) pathway. *J Exp Bot* **62:** 3049-3059.

Bailey TL, Elkan C (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2:** 28-36.

Benkert P, Biasini M, Schwede T (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics **27:** 343-350.

Bläsing OE, Ernst K, Streubel M, Westhoff P, Svensson P (2002). The non-photosynthetic phosphoenolpyruvate carboxylases of the C_4 dicot *Flaveria trinervia* - implications for the evolution of C_4 photosynthesis. *Planta* **215:** 448-456.

Bläsing OE, Westhoff P, Svensson P (2000). Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C4-specific characteristics. *J Biol Chem* **275:** 27917-27923.

Borland AM, Hartwell J, Jenkins GI, Wilkins MB, Nimmo HG (1999). Metabolite control overrides circadian regulation of phosphoenolpyruvate carboxylase kinase and CO_2 fixation in Crassulacean acid metabolism. *Plant Physiol* **121:** 889-896.

Bradford MM (1976). Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding. *Anal Biochem* **72:** 248-254.

Brautigam A, Mullick T, Schliesky S, Weber AP (2011). Critical assessment of assembly strategies for non-model species mRNA-Seq data and application of next-generation sequencing to the comparison of C (3) and C(4) species. *J Exp Bot* **62:** 3093-3102.

Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM (2011). Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C_4 Photosynthesis. *Science* **331:** 1436-1439.

Budde RJA, Chollet R (1986). *In vitro* Phosphorylation of Maize Leaf Phosphoenolpyruvate Carboxylase. *Plant Physiol* **82:** 1107-1114.

Chollet R, Vidal J, O'Leary MH (1996). PHOSPHOENOLPYRUVATE CARBOXYLASE: A Ubiquitous, Highly Regulated Enzyme in Plants. *Annu Rev Plant Physiol Plant Mol Biol* **47:** 273-298.

Christin PA, Boxall SF, Gregory R, Edwards EJ, Hartwell J, Osborne CP (2013). Parallel recruitment of multiple genes into C_4 photosynthesis. *Genome Biol Evol*

Christin PA, Osborne CP, Sage RF, Arakaki M, Edwards EJ (2011). C_4 eudicots are not younger than C_4 monocots. *J Exp Bot* **62:** 3171-3181.

De Brevern AG, Etchebest C, Hazout S (2000). Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins* **41:** 271-287.

Duff S, Chollet R (1995). *In Vivo* Regulation of Wheat-Leaf Phosphoenolpyruvate Carboxylase by Reversible Phosphorylation. *Plant Physiol* **107:** 775-782.

Duff SMG, Andreo CS, Pacquit V, Lepiniec L, Sarath G, Condon SA, Vidal J, Gadal P, Chollet R (1995). Kinetic-Analysis of the Nonphosphorylated, in-Vitro Phosphorylated, and Phosphorylation-Site-Mutant (Asp8) Forms of Intact Recombinant C_4 Phosphoenolpyruvate Carboxylase from *Sorghum*. *Eur J Biochem* **228:** 92-95.

Echevarria C, Vidal J (2003.) The unique phosphoenolpyruvate carboxylase kinase. *Plant Physiol Bioch* **41:** 541-547.

Echevarria C, Vidal J, Jiao JA, Chollet R (1990). Reversible light activation of the phosphoenolpyruvate carboxylase protein-serine kinase in maize leaves. *FEBS Lett* **275:** 25-28.

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:** 113.

Edwards GE, and Walker, D.A. (1983). C3, C4 mechanisms and cellular and environmental regulation of photosynthesis. Blackwell Scientific, Oxford

Ehleringer JR, Cerling TE, Helliker BR (1997). C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia* **112:** 285-299

Engelmann S, Bläsing OE, Westhoff P, Svensson P (2002) Serine 774 and amino acids 296 to 437 comprise the major C_4 determinants of the C_4 phosphoenolpyruvate carboxylase of *Flaveria trinervia*. *FEBS Lett* **524:** 11-14.

Engelmann S, Zogel C, Koczor M, Schlue U, Streubel M, Westhoff P (2008). Evolution of the C_4 phosphoenolpyruvate carboxylase promoter of the C_4 species *Flaveria trinervia*: the role of the proximal promoter region. *BMC Plant Biol* **8:** 4.

Ermolova NV, Ann Cushman M, Taybi T, Condon SA, Cushman JC, Chollet R (2003). Expression, purification, and initial characterization of a recombinant form of plant PEP-carboxylase kinase from CAM-induced *Mesembryanthemum crystallinum* with enhanced solubility in *Escherichia coli*. *Protein Expr Purif* **29:** 123-131.

Ernst K, Westhoff P (1997). The phosphoenolpyruvate carboxylase (*ppc*) gene family of *Flaveria trinervia* (C_4) and *F. pringlei* (C_3): Molecular characterization and expression analysis of the *ppcB* and *ppcC* genes. *Plant Mol Biol* **34:** 427-443

Facella P, Lopez L, Carbone F, Galbraith DW, Giuliano G, Perrotta G (2008). Diurnal and circadian rhythms in the tomato transcriptome and their modulation by cryptochrome photoreceptors. *PLoS One* **3:** e2798.

Fontaine V, Hartwell J, Jenkins GI, Nimmo HG (2002). *Arabidopsis thaliana* contains two phosphoenolpyruvate carboxylase kinase genes with different expression patterns. Plant Cell Environ **25:** 115-122.

Fukayama H, Tamai T, Taniguchi Y, Sullivan S, Miyao M, Nimmo HG (2006). Characterization and functional analysis of phosphoenolpyruvate carboxylase kinase genes in rice. *Plant J* **47:** 258-268.

Furumoto T, Izui K, Quinn V, Furbank RT, von Caemmerer S (2007). Phosphorylation of phosphoenolpyruvate carboxylase is not essential for high photosynthetic rates in the C_4 species *Flaveria bidentis*. *Plant Physiol* **144:** 1936-1945.

Gelly JC, Joseph AP, Srinivasan N, de Brevern AG (2011). iPBA: a tool for protein structure comparison using sequence alignment strategies. *Nucleic Acids Res* **39:** W18-23.

Giglioli-Guivarc'h N, Pierre JN, Brown S, Chollet R, Vidal J, Gadal P (1996). The Light-Dependent Transduction Pathway Controlling the Regulatory Phosphorylation of C_4 Phosphoenolpyruvate Carboxylase in Protoplasts from *Digitaria sanguinalis*. Plant Cell **8:** 573-586.

Gowik U, Brautigam A, Weber KL, Weber AP, Westhoff P (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C_4? *Plant Cell* **23:** 2087-2105.

Grieco M, Tikkanen M, Paakkarinen V, Kangasjarvi S, Aro EM (2012). Steady-state phosphorylation of light-harvesting complex II proteins preserves photosystem I under fluctuating white light. *Plant Physiol* **160:** 1896-1910.

Halford NG, Hardie DG (1998). SNF1-related protein kinases: global regulators of carbon metabolism in plants? *Plant Mol Biol* **37:** 735-748.

Hall BG (2013). Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol* **30:** 1229-1235.

Hartwell J, Gill A, Nimmo GA, Wilkins MB, Jenkins GL, Nimmo HG (1999). Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *Plant J* **20:** 333-342.

Hartwell J, Smith LH, Wilkins MB, Jenkins GI, Nimmo HG (1996). Higher plant phosphoenolpyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. *Plant J* **10:** 1071-1078.

Hatch MD (1987). C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim Biophys Acta* **895:** 81–106.

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber Andreas PM, Lercher Martin J (2013). Predicting C_4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell* **153:** 1579-1588.

Hermans J, Westhoff P (1990). Analysis of expression and evolutionary relationships of phosphoenolpyruvate carboxylase genes in *Flaveria trinervia* (C4) and *F. pringlei* (C3). *Mol Gen Genet* **224:** 459-468.

Hermans J, Westhoff P (1992). Homologous Genes for the C_4 Isoform of Phosphoenolpyruvate Carboxylase in a C_3 and a C_4 *Flaveria* Species. *Mol Gen Genet* **234:** 275-284.

Hrabak EM, Chan CW, Gribskov M, Harper JF, Choi JH, Halford N, Kudla J, Luan S, Nimmo HG, Sussman MR, Thomas M, Walker-Simmons K, Zhu JK, Harmon AC (2003). The *Arabidopsis* CDPK-SnRK superfamily of protein kinases. *Plant Physiol* **132:** 666-680.

Jacobs B, Engelmann S, Westhoff P, Gowik U (2008). Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*: determinants for high tolerance towards the inhibitor L-malate. *Plant Cell Environ* **31:** 793-803.

Jiao J, Chollet R (1990). Regulatory Phosphorylation of Serine-15 in Maize Phosphoenolpyruvate Carboxylase by a C_4 Leaf Protein-Serine Kinase. *Arch Biochem Biophys* **283:** 300-305.

Jiao J, Echevarria C, Vidal J, Chollet R (1991). Protein turnover as a component in the light/dark regulation of phosphoenolpyruvate carboxylase protein-serine kinase activity in C_4 plants. *Proc Natl Acad Sci USA* **88:** 2712-2715.

Jiao JA, Chollet R (1988). Light Dark Regulation of Maize Leaf Phosphoenolpyruvate Carboxylase by Invivo Phosphorylation. *Arch Biochem Biophys* **261:** 409-417.

Jiao JA, Chollet R (1989). Regulatory seryl-phosphorylation of C_4 phosphoenolpyruvate carboxylase by a soluble protein kinase from maize leaves. Arch Biochem Biophys **269:** 526-535.

Jones DT, Taylor WR, Thornton JM (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8:** 275-282.

Joseph AP, Agarwal G, Mahajan S, Gelly JC, Swapna LS, Offmann B, Cadet F, Bornot A, Tyagi M, Valadie H, Schneider B, Etchebest C, Srinivasan N, De Brevern AG (2010). A short survey on protein blocks. *Biophys Rev* **2:** 137-147.

Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Re*s **37:** D387-392.

Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680-685.

Lange V, Picotti P, Domon B, Aebersold R (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **4:** 222.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods **9:** 357-359

Law RD, Plaxton WC (1997). Regulatory phosphorylation of banana fruit phosphoenolpyruvate carboxylase by a copurifying phosphoenolpyruvate carboxylase-kinase. Eur J Biochem **247:** 642-651

Lepiniec L, Vidal J, Chollet R, Gadal P, Cretin C (1994). Phosphoenolpyruvate Carboxylase - Structure, Regulation and Evolution. *Plant Science* **99:** 111-124.

Leport L, Kandlbinder A, Baur B, Kaiser WM (1996). Diurnal modulation of phosphoenolpyruvate carboxylation in pea leaves and roots as related to tissue malate concentrations and to the nitrogen source. *Planta* **198:** 495-501.

Li B, Chollet R (1994). Salt Induction and the Partial Purification/Characterization of Phosphoenolpyruvate Carboxylase Protein-Serine Kinase from an Inducible Crassulacean-Acid-Metabolism (CAM) Plant, *Mesembryanthemum crystallinum L. Arch Biochem Biophys* **314:** 247-254.

Li B, Zhang XQ, Chollet R (1996). Phosphoenolpyruvate carboxylase kinase in tobacco leaves is activated by light in a similar but not identical way as in maize. *Plant Physiol* **111:** 497-505.

Ludwig M (2011). The molecular evolution of beta-carbonic anhydrase in *Flaveria*. *J Exp Bot* **62:** 3071-3081

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH (2013). CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* **41:** D348-352.

McKown AD, Moncalvo JM, Dengler NG (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. *Am J Bot* **92:** 1911-1928

McNaughton GA, MacKintosh C, Fewson CA, Wilkins MB, Nimmo HG (1991). Illumination increases the phosphorylation state of maize leaf phosphoenolpyruvate carboxylase by causing an increase in the activity of a protein kinase. *Biochim Biophys Acta* **1093:** 189-195.

Meimoun P, Gousset-Dupont A, Lebouteiller B, Ambard-Bretteville F, Besin E, Lelarge C, Mauve C, Hodges M, Vidal J (2009). The impact of PEPC phosphorylation on growth and development of *Arabidopsis thaliana*: molecular and physiological characterization of PEPC kinase mutants. *FEBS Lett* **583:** 1649-1652.

Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay SA, Chory J (2007). The diurnal project: Diurnal and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb Symp Quant Biol* **72:** 353-363.

Narsai R, Howell KA, Millar AH, O'Toole N, Small I, Whelan J (2007). Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell* **19:** 3418-3436.

Nimmo GA, Nimmo HG, Hamilton ID, Fewson CA, Wilkins MB (1986). Purification of the Phosphorylated Night Form and Dephosphorylated Day Form of Phosphoenolpyruvate Carboxylase from *Bryophyllum fedtschenkoi. Biochem J* **239:** 213-220.

Nimmo HG (2000). The regulation of phosphoenolpyruvate carboxylase in CAM plants. **Trends Plant Sci 5:** 75-80.

Nimmo HG (2003). Control of the phosphorylation of phosphoenolpyruvate carboxylase in higher plants. *Arch Biochem Biophys* **414:** 189-196.

O'Leary B, Park J, Plaxton WC (2011). The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. *Biochem J* **436:** 15-34

Osuna L, Pierre JN, Gonzalez MC, Alvarez R, Cejudo FJ, Echevarria C, Vidal J (1999). Evidence for a slow-turnover form of the Ca2+-independent phosphoenolpyruvate carboxylase kinase in the aleurone-endosperm tissue of germinating barley seeds. *Plant Physiol* **119:** 511-520.

Pagani M, Zachos JC, Freeman KH, Tipple B, Bohaty S (2005). Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309:** 600-603.

Paulus JK, Niehus C, Groth G (2013). Evolution of C_4 Phosphoenolpyruvate Carboxylase: Enhanced Feedback Inhibitor Tolerance Is Determined by a Single Residue. *Mol Plant.*

Paulus JK, Schlieper D, Groth G (2013). Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat Commun* **4:** 1518

Rajagopalan AV, Devi MT, Raghavendra AS (1994). Molecular-Biology of C_4 Phosphoenolpyruvate Carboxylase - Structure, Regulation and Genetic-Engineering. *Photosynth Res* **39:** 115-135

Sage RF (2004). The evolution of C_4 photosynthesis. *New Phytologist* **161:** 341-370.

Sage RF, Sage TL, Kocacinar F (2012). Photorespiration and the evolution of C4 photosynthesis. *Annu Rev Plant Biol* **63:** 19-47.

Saze H, Ueno, Y., Hisabori, T., Hayashi, H., Izui, K. (2001). Thioredoxin-Mediated Reductive Activation of a Protein Kinase for the Regulatory Phosphorylation of C_4 form Phosphoenolpyruvate Carboxylase from Maize. *Plant Cell Physiol* **42:** 1295–1302.

Schwede T, Kopp J, Guex N, Peitsch MC (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31:** 3381-3385.

Sheen J (1999) C-4 gene expression. Annu Rev Plant Physiol Plant Mol Biol **50:** 187-217

Shenton M, Fontaine V, Hartwell J, Marsh JT, Jenkins GI, Nimmo HG (2006). Distinct patterns of control and expression amongst members of the PEP carboxylase kinase gene family in C_4 plants. *Plant J* **48:** 45-53.

Sullivan S, Jenkins GI, Nimmo HG (2004). Roots, cycles and leaves. Expression of the phosphoenolpyruvate carboxylase kinase gene family in soybean. *Plant Physiol* **135:** 2078-2087.

Sullivan S, Shenton M, Nimmo HG (2005). Organ specificity in the circadian control of plant gene expression. *Biochem Soc Trans* **33:** 943-944

Svensson P, Bläsing OE, Westhoff P (1997). Evolution of the enzymatic characteristics of C_4 phosphoenolpyruvate carboxylase - A comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of *Flaveria trinervia* (C_4) and *Flaveria pringlei* (C_3). *Eur J Biochem* **246:** 452-460.

Svensson P, Bläsing OE, Westhoff P (2003). Evolution of C_4 phosphoenolpyruvate carboxylase. *Arch Biochem Biophys* **414:** 180-188.

Takahashi-Terada A, Kotera M, Ohshima K, Furumoto T, Matsumura H, Kai Y, Izui K (2005). Maize phosphoenolpyruvate carboxylase. Mutations at the putative binding site for glucose 6-phosphate caused desensitization and abolished responsiveness to regulatory phosphorylation. *J Biol Chem* **280:** 11798-11806.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731-2739.

Taniguchi N (2010). Amino Acids and Proteins. *In* J Baynes, and Dominiczak, M., ed, Medical Biochemistry, Ed 3. Elsevier

Taybi T, Patil S, Chollet R, Cushman JC (2000). A minimal serine/threonine protein kinase circadianly regulates phosphoenolpyruvate carboxylase activity in crassulacean acid metabolism-induced leaves of the common ice plant. *Plant Physiol* **123:** 1471-1482.

Tipple BJ, Pagani M (2007). The Early Origins of Terrestrial C_4 Photosynthesis. *Annu Rev Earth Planet Sci* **35:** 435-461.

Tovar-Mendez A, Mujica-Jimenez C, Munoz-Clares RA (2000). Physiological implications of the kinetics of maize leaf phosphoenolpyruvate carboxylase. *Plant Physiol* **123:** 149-160.

Tripodi KE, Turner WL, Gennidakis S, Plaxton WC (2005). In vivo regulatory phosphorylation of novel phosphoenolpyruvate carboxylase isoforms in endosperm of developing castor oil seeds. *Plant Physiol* **139:** 969-978.

Tsuchida Y, Furumoto T, Izumida A, Hata S, Izui K (2001) Phosphoeno/pyruvate carboxylase kinase involved in C_4 photosynthesis in *Flaveria trinervia*: cDNA cloning and characterization. *Febs Lett* **507:** 318-322

Ubersax JA, Ferrell JE, Jr. (2007). Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* **8:** 530-541.

Ueno Y, Imanari E, Emura J, Yoshizawa-Kumagaye K, Nakajima K, Inami K, Shiba T, Sakakibara H, Sugiyama T, Izui K (2000). Immunological analysis of the phosphorylation state of maize C_4 form phosphoenolpyruvate carboxylase with specific antibodies raised against a synthetic phosphorylated peptide. *Plant J* **21:** 17-26.

Uhrig RG, She YM, Leach CA, Plaxton WC (2008). Regulatory monoubiquitination of phosphoenolpyruvate carboxylase in germinating castor oil seeds. *J Biol Chem* **283:** 29650-29657.

Vanquy L, Foyer C, Champigny ML (1991). Effect of Light and No3- on Wheat Leaf Phosphoenolpyruvate Carboxylase Activity - Evidence for Covalent Modulation of the C_3 Enzyme. *Plant Physiol* **97:** 1476-1482.

Vidal J, Chollet R (1997). Regulatory phosphorylation of C_4 PEP carboxylase. *Trends Plant Sci* **2:** 230-237.

Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH (2009). Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses. *Genome Biol* **10:** R68.

Wang YH, Duff SMG, Lepiniec L, Cretin C, Sarath G, Condon SA, Vidal J, Gadal P, Chollet R (1992) Site-Directed Mutagenesis of the Phosphorylatable Serine (Ser8) in C_4 Phosphoenolpyruvate Carboxylase from Sorghum - the Effect of Negative Charge at Position-8. *J Biol Chem* **267:** 16759-16762.

Willige BC, Isono E, Richter R, Zourelidou M, Schwechheimer C (2011). Gibberellin Regulates PIN-FORMED Abundance and Is Required for Auxin Transport-Dependent Growth and Development in Arabidopsis thaliana. *Plant Cell* **23:** 2184-2195.

Wiludda C, Schulze S, Gowik U, Engelmann S, Koczor M, Streubel M, Bauwe H, Westhoff P (2012). Regulation of the photorespiratory GLDPA gene in C_4 *Flaveria*: an intricate interplay of transcriptional and posttranscriptional processes. *Plant Cell* **24:** 137-151.

LEGENDS

FIGURE LEGENDS

Figure 1: Phylogenetic analysis of the PPCK family in *Flaveria*. Two PPCK isoforms A and B are found in each *Flaveria* species which cluster together to form individual groups. In both situations the isoforms from the C_4 species show largest evolutionary distance from the C_3 species. Species: *F. trinervia* (Ft), *F. bidentis* (Fb), *F. brownii* (Fbr); *F. ramosissima* (Fra), *F. anomala* (Fa), *F. pubescencs* (Fpu), *F. chloraefolia* (Fc), *F. pringlei* (Fp), *F. robusta* (Fro). FaPPCKB is not used in the phylogeny due to partial sequence coverage. The photosynthesis type for each species is indicated on the phylogenetic tree. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011; Hall, 2013). Sequences were aligned using Muscle (Edgar, 2004) with default settings for 1000 iterations. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-1177.4396) is shown. Reliability of the evolutionary relationship was estimated using bootstrap variances (1,000 replicates), indicated in percentage at the branches.

Figure 2: Amino acid exchanges between *Flaveria* species with C_3 and C_4 photosynthesis. A: Multiple sequence alignment of PPCK A and B isoforms from nine *Flaveria* species, using MUSCLE (Edgar, 2004a/b); B: Schematic model of PPCKA from *F. pringlei* with marked positions of amino acid exchange among C_3 and C_4 species for both isoforms; C: Three dimensional model of PPCKA from *F. trinervia* (blue) is aligned onto the model from *F. pringlei* (red) and the positions of the amino-acid exchanges are marked in yellow and numbered according to the alignment shown in Figure 2A.The protein kinase catalytic core domain starts at amino acid position 15 at the N-terminus and ending at position 268 at the C-terminus (according to BAB71853 and BAF4832). Amino acids involved in ATP-binding are indicated in red, amino acids involved in the substrate-binding are marked in light blue. Amino acids in green are involved in ATP- and substrate-binding; Dark blue boxes refer to the activation (A)-loop. Highlighted background or stars indicates amino acid residues which differ between species with C_3 and C_4 photosynthesis. PPCK from C_4 *Flaveria* were taken from the accession numbers BAB71853 and BAF4832. The remaining *Flaveria* PPCK sequences were cloned and obtained from RNAseq. Sequences used for the alignment and modelling are listed in **Supplemental file 1.** C_4 species: *F. trinervia* (Ft) and *F. bidentis* (Fb); C4-like: *F. brownii* (Fbr); C_3-C_4 species: *F. ramosissima* (Fra); *F. anomala* (Fa); *F. pubescens* (Fpu); *F. chloraefolia* (Fch); C_3 species: *F. pringlei* (Fp) and *F. robusta* (Fro). The 3D models are calculated using Swissmodel (Schwede et al., 2003; Arnold et al., 2006; Kiefer et al., 2009) using the sequences for FtPPCKA and FpPPCKA listed in Supplemental file 1. PPCKA from *F. pringlei* and *F. trinervia* is modeled onto a Calmodulin-domain protein kinase 1 from *Toxoplasma gondii*, with sequence similarity of 0.40, (sequence identity of 38.49%) and sequence similarity 0.40 (sequence identity of 39.62%) respectively. The models with highest QMEAN value -0.85 and -0.78 for FpPPCKA and FtPPCKA were aligned in iPBA (de Brevern et al., 2000; Joseph et al., 2010; Benkert et al., 2011) and the resulting file was displayed in in PyMol v 1.3 http://www.pymol.org/.

Figure 3: Transcript abundances of the *PPCK* genes from *Flaveria* along a light/dark cycle. Each graph represents the circadian transcript expression of *PPCKA* (**A**) and normalized *PPCKA* (**B**), *PPCKB* (**C**) and normalized *PPCKB* (**D**) genes from *F. pringlei* (C_3) and *F. trinervia* (C_4) plants grown under light (white background) and dark (gray background) conditions, also indicated as the white (light) and black (dark) horizontal bar. The transcript abundances of *PPCK* genes are presented in read per kilobase per million mapped reads (RPMK) as the average of three biological replicates measured by RNAseq. Normalized RPMK values for each time point, were expressed as the percentage from the sum of RPMK values for the respective time series for each transcript **B** and **D**. *F. trinervia* (C_4)- blue line; *F. pringlei* (C_3)- red line, *PPCKA*- full circle, *PPCKB*-open circle. Lights were turned on at 0h and off at 10 h diurnal time. The standard error of the mean (SEM) was calculated for each time course of three biological replicates. The significant differences between each time point were determined with the RPMK values using a one-way ANOVA (Tukey-test) and additionally by the Student's t-test, $p \leq 0.05$ (Supplemental file 4).

Figure 4: Circadian variation of PEPC phosphorylation levels in comparison to PPCK transcripts. The ratio of PEPC phosphorylation calculated as ratio of the concentrations of the phosphorylated and non-phosphorylated version of the peptide LASIDAQR monitored by SRM, and expressed in percent from the total sum of ratios for each species. The diurnal variation of PPCKA and PPCKB transcripts for each species is expressed in percent RPMK as in Figure 3. Three biological replicates were averaged for each time point. *F. trinervia* (A), *F. pringlei* (B), PPCKA- grey line and full circle, PPCKB- grey line and open circle, ratio of phosphorylated PEPC- black line and filled squares. The horizontal bar above the graph and the graph background indicate the light conditions light- white, dark – black/grey. The original results for each measurement are listed in Supplemental file 4.

Figure 5: *In vitro* phosphorylation of *ppcA* PEPC protein from *Flaveria* species. Recombinant *ppcA* PEPC proteins from species with C_3 and C_4 photosynthesis were phosphorylated in vitro using radioactive ATP and recombinant PPC kinase from 2 *Flaveria* species. PPCKs are not specific to the PEPC from the respective species and can cross-phosphorylate in vitro. The phosphorylation of the Fp *ppcA* PEPC (1 µg) was tested, with equal amounts of purified recombinant kinase based on based on supplemental file 6: FpPPCKA (25µg total protein) and Ft PPCKA (1µg total protein) (**A**); Indications of higher activity by the PPCK from C_4 species only when the FtPEPC (3µg) is used as a substrate, FpPPCK (25µg totyl protein), Ft (1µg total protein) **(B).** Each lane represents an independent PPCK expression and purification. Fp- *F. pringlei* (C_3), Ft- *F. trinervia* (C_4) Controls: C1 – temperature deactivated PPCK; C2 - no PEPC; C3- no FpPPCK; C4- no FtPPCK.

Figure 6: Analysis of the phosphorylation motif on the N terminus of PEPC: The first 18-21 amino acids (based on the sequence alignment in supplemental file 8) were analyzed for conserved motifs using the online tool 'MEME' http://meme.nbcr.net/meme/ (Bailey and Elkan, 1994). A- The consensus motif E-K-X-X-S-I-D-A-Q-L-R was found in 50 of the 56 submitted sequences from *Flaveria*, *Brassicales*, *Amarantheaceae* and *Poaceae*. Several C_3 to C_4 amino acid exchanges can be observed directly in front of the phosphorylated Serine (Supplemental file 8). B- The consensus sequence for *Flaveria* shows the exchanges L6V (found in C_4 *ppcA* PEPC) and Q15H (found in C_4 *ppcB* PEPCs), which differ between C_3 and C_4 species. C- Virtual mutagenesis is applied onto the crystal structure of the PEPC tetramer from *F.pringlei* which starts at position L6. The applied mutations L6V and Q15H lie on the surface of the molecule and are accessible to the solute. Shades of grey indicate the individual monomers as part of the PEPC tetramer, Valine- cyan; Serine 11- yellow; Histidine- green.

TABLE LEGENDS

Table 1: Amino acid exchanges in PPCK isoforms occurring among C_3 and C_4 *Flaveria* species compared to C_3 and C_4 species from *Brassicales* and *Poaceae*. Amino acids are listed using standard single amino acid code and with positions marked according to the *F. pringlei* PPCKA sequence. An alignment of all sequences is presented in Supplemental File 3. Light grey- amino acid exchanges found in some intermediate species and species with C_4, medium grey –amino acid exchanges found in all intermediate species and species with C_4, dark grey- amino acid exchanges found only in species with C_4 photosynthesis. Red - C_3 species. Blue - C_4 species. White - C_3-C_4 intermediates and C_4-like *F. brownii* (*Fbr*). "-" -Sequence information is missing or residue is not present. Species abbreviations: *Brassicales- Arabidopsis thaliana (At), Cleome gynandra* (*Cg*); *Poaceae - Zea mays* (*Zm*)*, Oryza sativa* (*Os*), *Setaria italica* (*Si*), *Sorghum bicolor* (*Sb*). The accession numbers for the sequences are listed in supplemental file 2, except Cg. The sequence for Cg is included at the end of supplemental file 1.

SUPPLEMENTAL LEGENDS

S1: Supplemental file 1- *PPCK* and *ppc* sequences from *Flaveria*. *PPCK* and *ppc* sequences from *Cleome gynandra*

S2: Supplemental file 2- Phylogenetic analysis of PPCK isoforms from various species. The analysis shows that the PPCKs have evolved post speciation of the vascular plants, and evolved in parallel within each closely related group. Within the genus Flaveria, the PPCKA and PPCKB are clearly different from each other, and the segregation in branches is driven by the PPCK type rather than the type of photosynthesis. However, in the case of PPCKA there is further separation of the species with C_3 photosynthesis, and a group formed by the C_3-C_4 intermediates and species with C_4 photosynthesis. A similar grouping of PPCK isoforms is also seen in all included systematics groups where we have a C_4 representative. The isoforms of the genus *Flaveria* are indicated by the two blue backgrounds. Pink and purple indicate the PPCKs from the order *Brassicales*, yellow and orange – family *Poaceae.* Evolutionary analyses were conducted in MEGA5 (Tamura *et al*., 2011; Hall, 2013). Sequences were aligned using Muscle (Edgar, 2004) with default settings for 1000 iterations. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al, 1992). The tree with the highest log likelihood (-5116.5439) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 49 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 173 positions in the final dataset. Species abbreviations: *Arabidopsis thaliana (At), Cleome gynandra (Cg), Zea mays (Zm),Brachypodium distachyon (Bd), Beta vulgaris (Bv), Clusia minor (Cm); Glycine max (Gm), Kalanchoe daigremontiana (Kd), Kalanchoe fedtschenkoi (Kf), Kalanchoe pinnata (Kp), Lotus japonicus (Lj), Lycopersicon esculentum (Le), Mesembryanthemum crystallinum (Mc), Oryza sativa (Os), Setaria italica (Si), Solanum tuberosum (St), Sorghum bicolor (Sb),* The accession numbers for the sequences used for this are: At1g08650.1 (*PPCK*1), At3g04530.1 (*PPCK*2); NP_001105808 (*ZmPPCK1*), NP_001105772 (*ZmPPCK2*), NP_001105773 (*ZmPPCK3*), NP_001105774 (*ZmPPCK4*), ABD39238.1 (*SbPPCK1*), AAK81871.1 (*SbPPCK2*), Sb06G022690 (SbPPCK3), AC069158 (*OsPPCK1*), BAE80213.1 (*OsPPCK2*), BAE80227.1 (*OsPPCK3*), AAN12511.1 (*GmPPCK1*), AAN12513.1 (*GmPPCK2*), AAQ82624.1 (*GmPPCK3*), AAS75449.1 (*GmPPCK4*), CAC43293.1 (*BePPCK*) , AAF06970.1 (*KfPPCK*), ABO65214 (*KdPPCK*), AAF19403.1 (*LePPCK1*), AAO32075.1 (*LePPCK2*), AAF05112.1 (*McPPCK*), BAC20363 (*LjPPCK*), ABO65213 (*KpPPCK*), AAR31830 (*CmPPCK1*), AAR31831 (*CmPPCK2*), AAO61489 (*StPPCK*), XP_003575394 (*BdPPCK*), XP_003580142 (*BdPPCK1-like*), XP_003572921 (*BdPPCK2-like*), XP_004976235 (*SiPPCK3*), XP_004954289 (*SiPPCK1*), XP_004953151 (SiPPCK2). The sequence for *Cleome gynandra (Cg)* is at the end of supplemental file 1.

S3: Supplemental file 3- Alignments of PPCK proteins from various C_3 and C_4 species The amino acid residues which changed in the evolution of C_4 *Flaveria* species are specific for this genus. Full length protein sequences from the PPCK protein family from the genus *Flaveria,* genus *Poaceae* and order *Brassicales* were aligned using Muscle (Edgar, 2004), as part of MEGA5 (Tamura *et al*., 2011; Hall, 2013) using default settings for 1000 iterations. The C_4 specific amino acid exchanges, characteristic for the genus *Flaveria* (Table 1) are highlighted in grey: Light grey- amino acid exchanges found in some intermediate species and species with C_4, medium grey –amino acid exchanges found in all intermediate species and species with C_4, dark grey- amino acid exchanges found only in species with C_4 photosynthesis isoforms. Species abbreviations: *Brassicales- Arabidopsis thaliana (At),* Cleome gynandra (Cg); *Poaceae - Zea mays (Zm), Oryza sativa (Os), Setaria italica (Si), Sorghum bicolor (Sb)*. The accession numbers for the sequences used for this are listed in supplemental file 2, except Cg. The sequence for Cg is included at the end of supplemental file 1.

S4: Supplemental file 4- RNA_Seq and SRM data

S5: Supplemental file 5- Multiple alignment of N-terminal PEPC isoforms, showing the specificity of LASIDAQLR SRM peptide (A); Characteristics of the LASIDAQLR and LA(pS)IDAQLR SRM peptides (B). Amino acid sequences of PEPC isoforms were analyzed with MUSCLE (Edgar, 2004a/b). Residues marked with bold and grey background indicate the peptide sequence used for SRM. Ser11 residues at the N-terminal in bold red indicate the phosphorylation site. *F. trinervia* (Ft); *F. pringlei* (Fp). In the table "X" marks that the peptide transition has been successfully monitored in three biological replicates in the time series for the respective species.

S6: Supplemental file 6- Recombinant PPCKA expression and anti-PPCK antibodies. Representative SDS-PAGE of purified recombinant PPCK protein (A) and western-blot with the custom anti-PPCK antibody (B). The adjustment of kinase amount for the *in-vitro* phosphorylation assay was performed by quantifying the kinase band intensity of the Coomasie-stained gel, and protein amount for FpPPCKA was adjusted to match the strongest signal of FtPPCKA.

S7: Supplemental file 7- List of oligonucleotides

S8: Supplemental file 8- Supplemental File 8: N-terminal alignment of PEPC isoforms from *Flaveria*, Amarantheaceae, Brassicales, and Poaceae. Approximately 50 amino acids from the N-terminus of PEPC have been taken for this analysis and the sequences alligned using Muscle Muscle (Edgar, 2004) within the software MEGA5 (Tamura *et al*., 2011; Hall, 2013). The sequences are sub-grouped within their Phylogenetic clades and within each systematic group a C_3 PEPC sequence is taken as reference. Amino acid that are close to the phosphorylated Serine (marked in red, with yellow background) and that vary either between PEPC isoforms form same systematics group, or among C_3 and C_4 species from same systematics group have been marked in grey. Species abbreviations- *Amaranthaceae: Alternanthera sessilis* (Al.s)*, Alternanthera tenella* (Al.t)*, Alternanthera pungens* (Al.p)*, Amaranthus hypochondriacus* (Am.h)*, Bienertia sinuspersici* (Bs); *Brassicales: Arabidopsis thaliana (At), Cleome gynandra* (Cg)*, Poaceae: Zea mays* (Zm), *Brachypodium distachyon* (Bd), *Oryza sativa* (Os), *Setaria italica* (Si)*, Sorghum bicolor* (Sb), *Chloris_gayana* (Cga), *Saccharum offcinarum* (So), *Saccharum spontaneum* (Ss). For most species accession codes are included within the alignment names, whereas PEPC sequences for *Flaveria* and *Cleome* are listed in supplemental file 1. The phosphorylated Serine is in red and highlighted in yellow, the consensus motif from the N-terminal alignment is highlighted in blue, (*)- a single cell C_4 species.

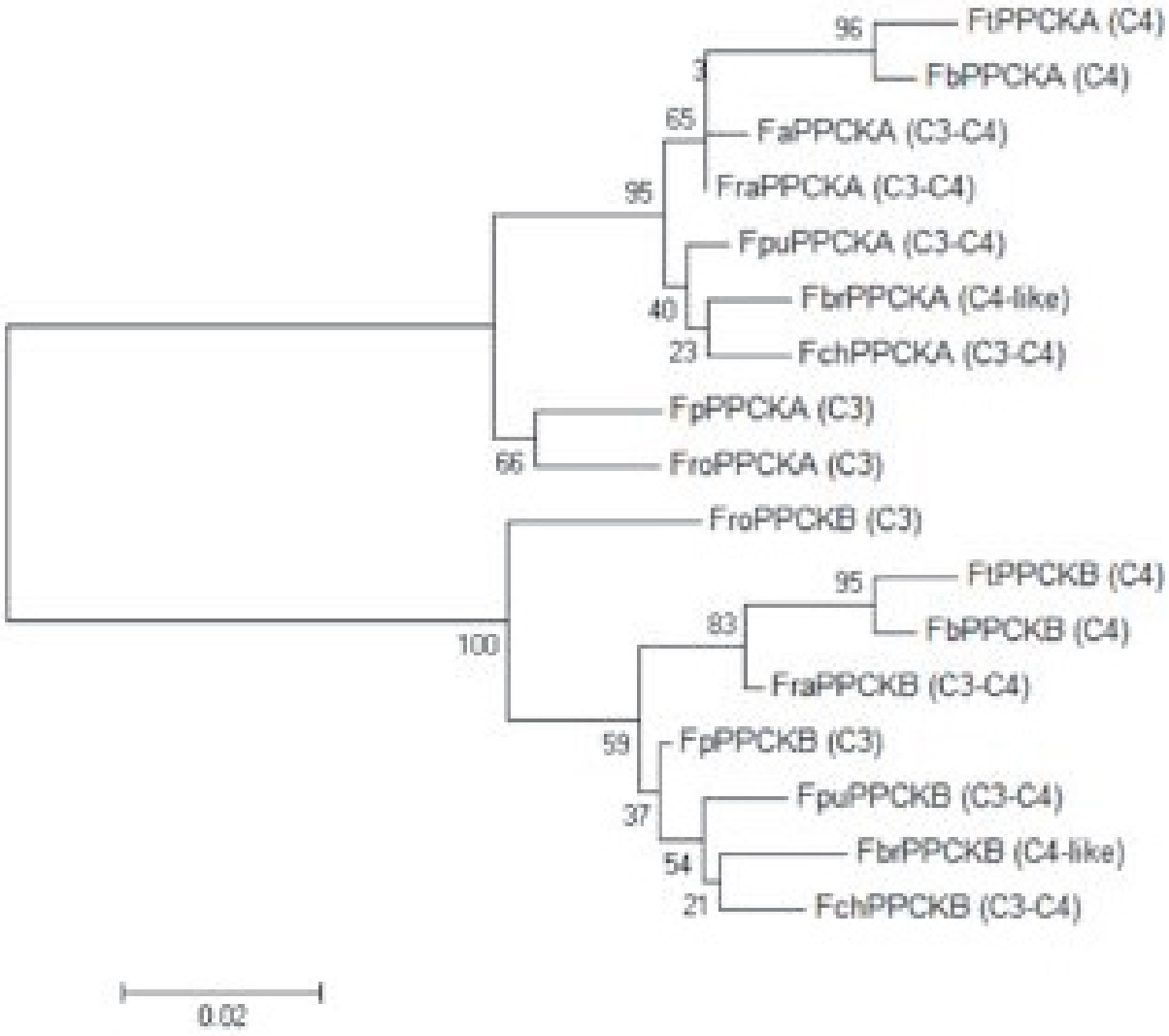


Figure 1: Phylogenetic analysis of the PPCK family in *Flaveria*. Two PPCK isoforms A and B are found in each *Flaveria* species which cluster together to form individual groups. In both situations the isoforms from the C_4 species show largest evolutionary distance from the C_3 species. Species: *F. trinervia* (Ft), *F. bidentis* (Fb), *F. brownii* (Fbr); *F. ramosissima* (Fra), *F. anomala* (Fa), *F. pubescencs* (Fpu), *F. chloraefolia* (Fc), *F. pringlei* (Fp), *F. robusta* (Fro). FaPPCKB is not used in the phylogeny due to partial sequence coverage. The photosynthesis type for each species is indicated on the phylogenetic tree. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011; Hall, 2013). Sequences were aligned using Muscle (Edgar, 2004) with default settings for 1000 iterations. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-1177.4396) is shown. Reliability of the evolutionary relationship was estimated using bootstrap variances (1,000 replicates), indicated in percentage at the branches.

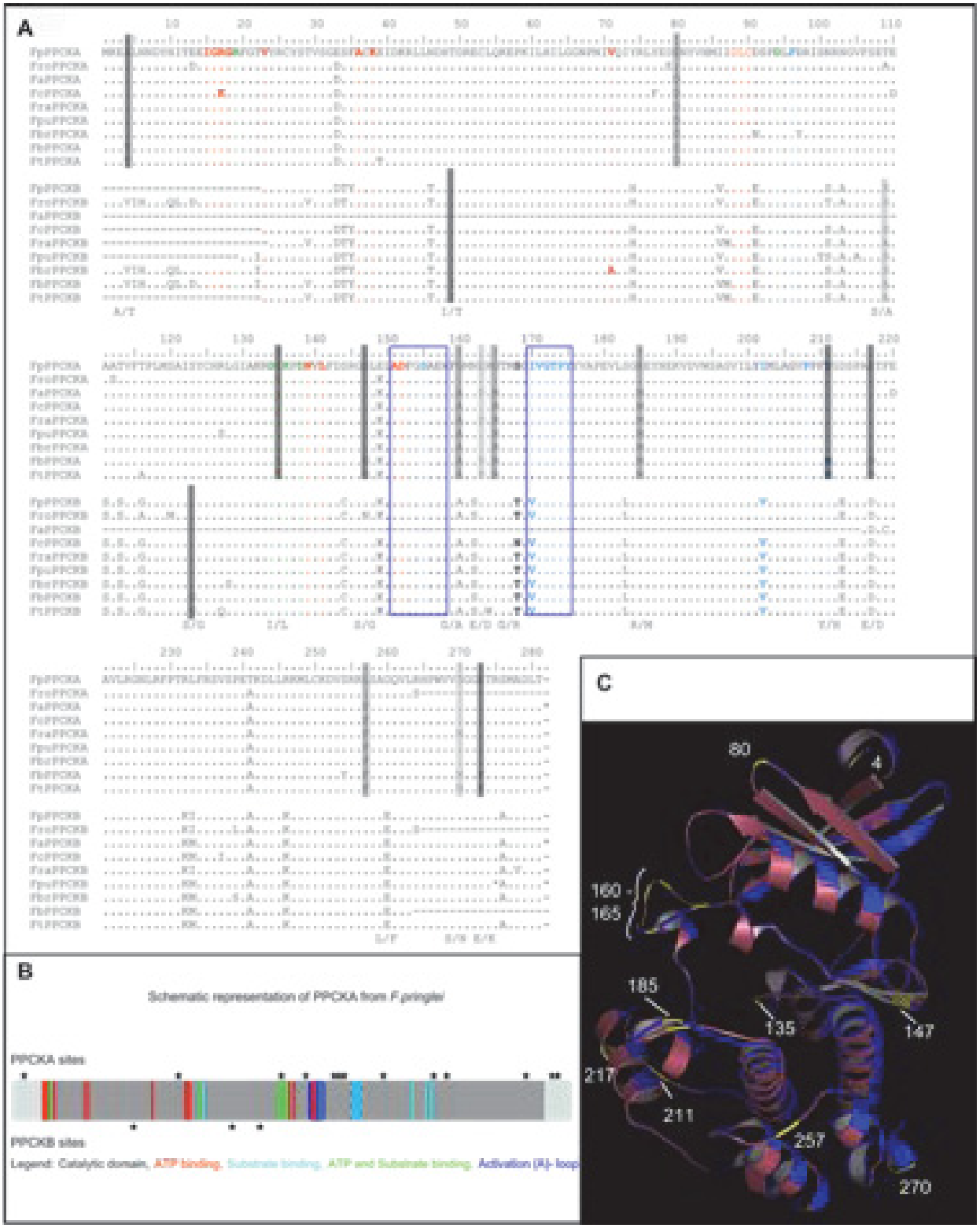

A
B
Schematic representation of PPCKA from P. pringlei
PPCKA sites:
PPCKB sites:
C
80
4
160 -
165
185
135
147
217
211
257
270

Figure 2: Amino acid exchanges between *Flaveria* species with C_3 and C_4 photosynthesis. A: Multiple sequence alignment of PPCK A and B isoforms from nine *Flaveria* species, using MUSCLE (Edgar, 2004a/b); B: Schematic model of PPCKA from *F. pringlei* with marked positions of amino acid exchange among C_3 and C_4 species for both isoforms; C: Three dimensional model of PPCKA from *F. trinervia* (blue) is aligned onto the model from *F. pringlei* (red) and the positions of the amino-acid exchanges are marked in yellow and numbered according to the alignment shown in Figure 2A.The protein kinase catalytic core domain starts at amino acid position 15 at the N-terminus and ending at position 268 at the C-terminus (according to BAB71853 and BAF4832). Amino acids involved in ATP-binding are indicated in red, amino acids involved in the substrate-binding are marked in light blue. Amino acids in green are involved in ATP- and substrate-binding; Dark blue boxes refer to the activation (A)-loop. Highlighted background or stars indicates amino acid residues which differ between species with C_3 and C_4 photosynthesis. PPCK from C_4 *Flaveria* were taken from the accession numbers BAB71853 and BAF4832. The remaining *Flaveria* PPCK sequences were cloned and obtained from RNAseq. Sequences used for the alignment and modelling are listed in **Supplemental file 1.** C_4 species: *F. trinervia* (Ft) and *F. bidentis* (Fb); C4-like: *F. brownii* (Fbr); C_3-C_4 species: *F. ramosissima* (Fra); *F. anomala* (Fa); *F. pubescens* (Fpu); *F. chloraefolia* (Fch); C_3 species: *F. pringlei* (Fp) and *F. robusta* (Fro). The 3D models are calculated using Swissmodel (Schwede et al., 2003; Arnold et al., 2006; Kiefer et al., 2009) using the sequences for FtPPCKA and FpPPCKA listed in Supplemental file 1. PPCKA from *F. pringlei* and *F. trinervia* is modeled onto a Calmodulin-domain protein kinase 1 from *Toxoplasma gondii*, with sequence similarity of 0.40, (sequence identity of 38.49%) and sequence similarity 0.40 (sequence identity of 39.62%) respectively. The models with highest QMEAN value -0.85 and -0.78 for FpPPCKA and FtPPCKA were aligned in iPBA (de Brevern et al., 2000; Joseph et al., 2010; Benkert et al., 2011) and the resulting file was displayed in in PyMol v 1.3 http://www.pymol.org/.

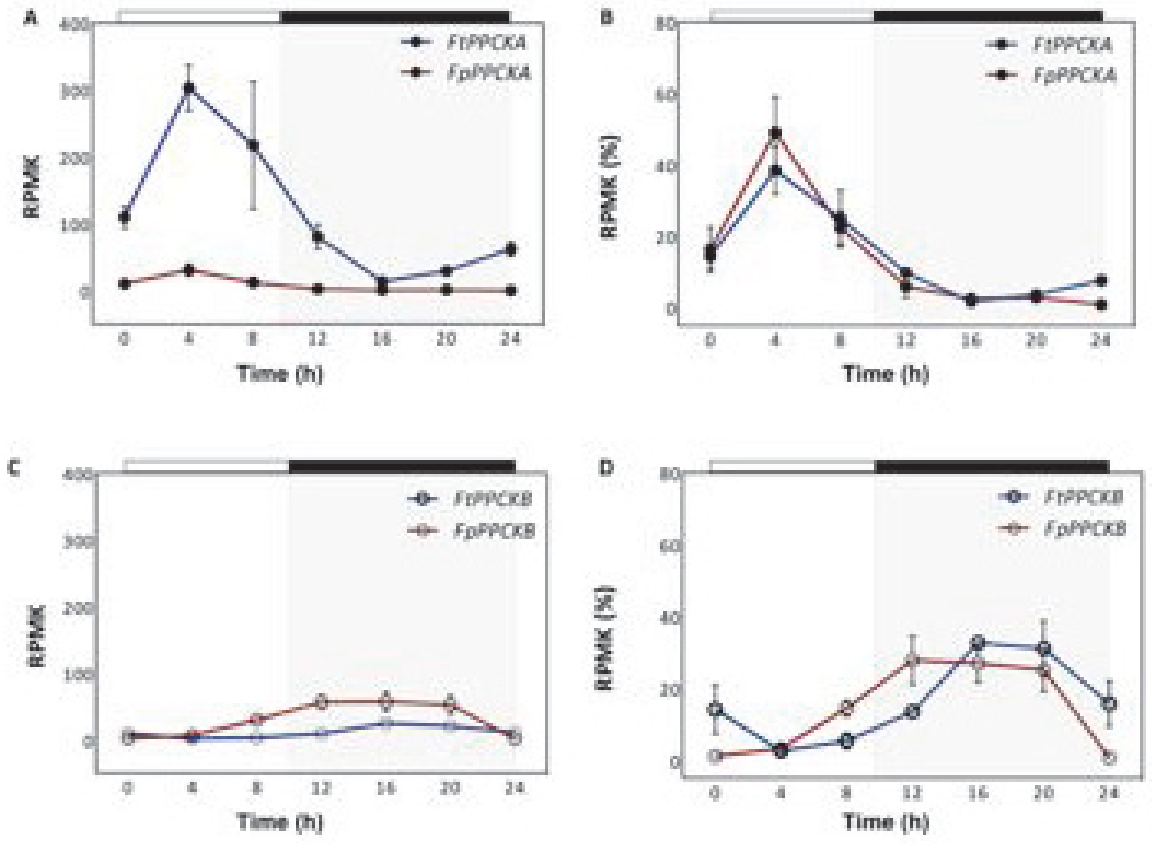


Figure 3: Transcript abundances of the *PPCK* genes from *Flaveria* along a light/dark cycle. Each graph represents the circadian transcript expression of *PPCKA* (**A**) and normalized *PPCKA* (**B**), *PPCKB* (**C**) and normalized *PPCKB* (**D**) genes from *F. pringlei* (C_3) and *F. trinervia* (C_4) plants grown under light (white background) and dark (gray background) conditions, also indicated as the white (light) and black (dark) horizontal bar. The transcript abundances of *PPCK* genes are presented in read per kilobase per million mapped reads (RPMK) as the average of three biological replicates measured by RNAseq. Normalized RPMK values for each time point, were expressed as the percentage from the sum of RPMK values for the respective time series for each transcript **B** and **D**. *F. trinervia* (C_4)- blue line; *F. pringlei* (C_3)- red line, *PPCKA*- full circle, *PPCKB*-open circle. Lights were turned on at 0h and off at 10 h diurnal time. The standard error of the mean (SEM) was calculated for each time course of three biological replicates. The significant differences between each time point were determined with the RPMK values using a one-way ANOVA (Tukey-test) and additionally by the Student's t-test, $p \leq 0.05$ (Supplemental file 4).

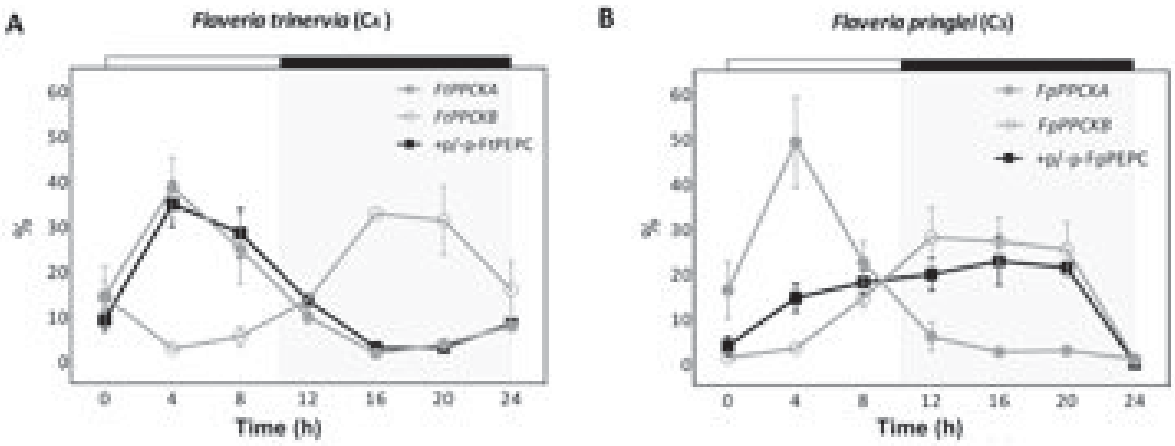


Figure 4: Circadian variation of PEPC phosphorylation levels in comparison to PPCK transcripts. The ratio of PEPC phosphorylation calculated as ratio of the concentrations of the phosphorylated and non-phosphorylated version of the peptide LASIDAQR monitored by SRM, and expressed in percent from the total sum of ratios for each species. The diurnal variation of PPCKA and PPCKB transcripts for each species is expressed in percent RPMK as in Figure 3. Three biological replicates were averaged for each time point. *F. trinervia* (A), *F. pringlei* (B), *PPCKA*- grey line and full circle, *PPCKB*- grey line and open circle, ratio of phosphorylated PEPC- black line and filled squares. The horizontal bar above the graph and the graph background indicate the light conditions light- white, dark – black/grey. The original results for each measurement are listed in Supplemental file 4.

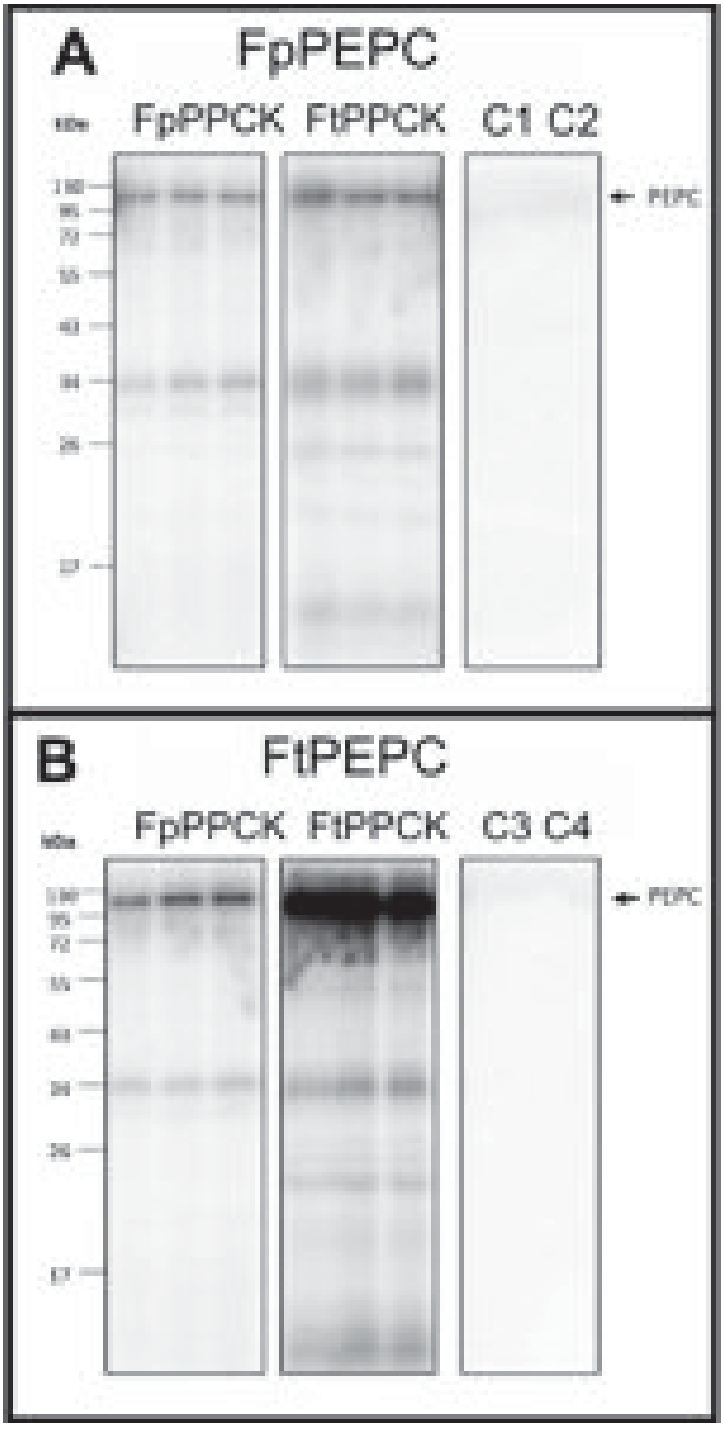


Figure 5: *In vitro* phosphorylation of *ppcA* PEPC protein from *Flaveria* species. Recombinant *ppcA* PEPC proteins from species with C_3 and C_4 photosynthesis were phosphorylated in vitro using radioactive ATP and recombinant PPC kinase from 2 *Flaveria* species. PPCKs are not specific to the PEPC from the respective species and can cross-phosphorylate in vitro. The phosphorylation of the Fp *ppcA* PEPC (1 µg) was tested, with equal amounts of purified recombinant kinase based on based on supplemental file 6: FpPPCKA (25µg total protein) and Ft PPCKA (1µg total protein) (**A**); Indications of higher activity by the PPCK from C_4 species only when the FtPEPC (3µg) is used as a substrate, FpPPCK (25µg totyl protein), Ft (1µg total protein) **(B).** Each lane represents an independent PPCK expression and purification. Fp- *F. pringlei* (C_3), Ft- *F. trinervia* (C_4) Controls: C1 – temperature deactivated PPCK; C2 - no PEPC; C3- no FpPPCK; C4- no FtPPCK.

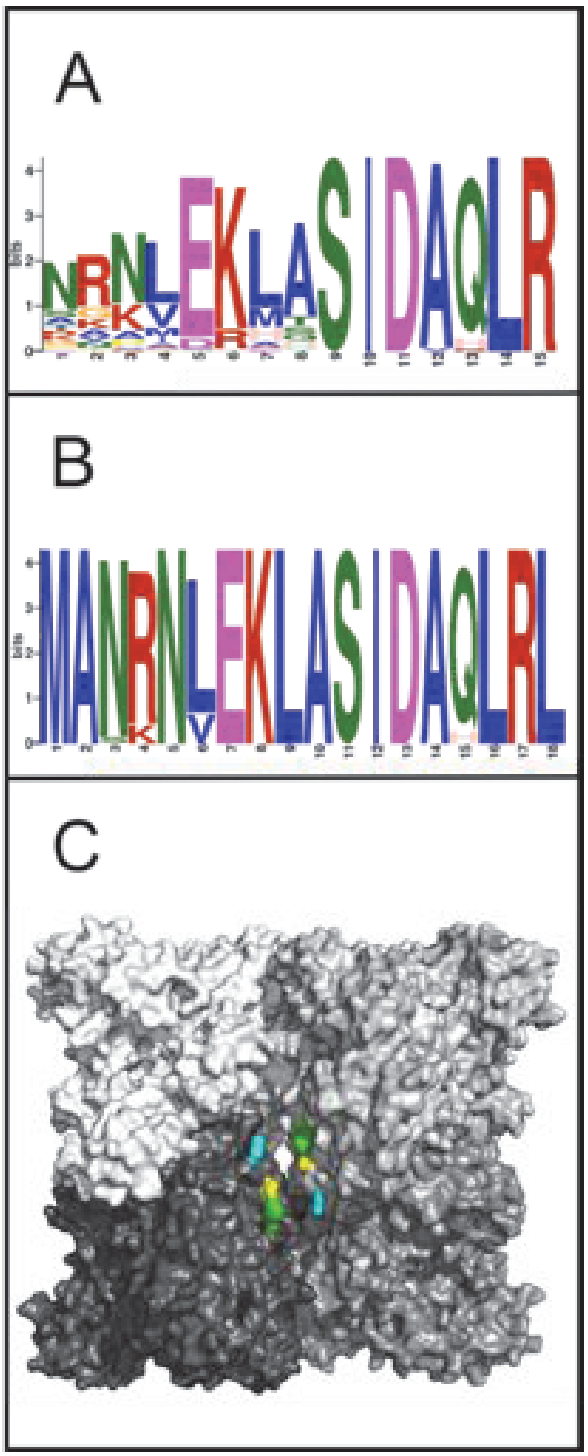
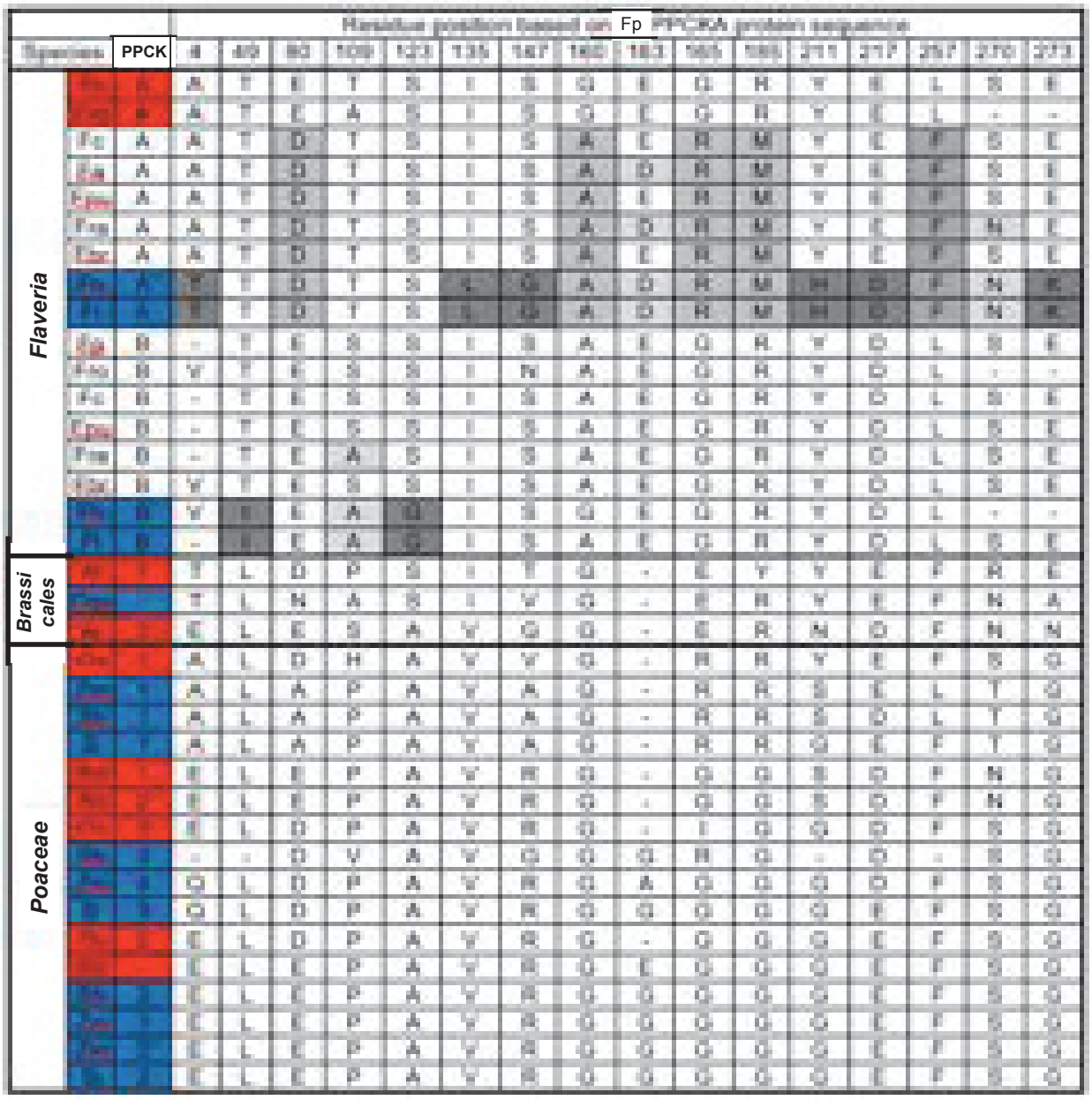


Figure 6: Analysis of the phosphorylation motif on the N terminus of PEPC: The first 18-21 amino acids (based on the sequence alignment in supplemental file 8) were analyzed for conserved motifs using the online tool 'MEME' http://meme.nbcr.net/meme/ (Bailey and Elkan, 1994). A- The consensus motif E-K-X-X-S-I-D-A-Q-L-R was found in 50 of the 56 submitted sequences from *Flaveria*, *Brassicales*, *Amarantheaceae* and *Poaceae*. Several C_3 to C_4 amino acid exchanges can be observed directly in front of the phosphorylated Serine (Supplemental file 8). B- The consensus sequence for *Flaveria* shows the exchanges L6V (found in C_4 *ppcA* PEPC) and Q15H (found in C_4 *ppcB* PEPCs), which differ between C_3 and C_4 species. C- Virtual mutagenesis is applied onto the crystal structure of the PEPC tetramer from *F.pringlei* which starts at position L6. The applied mutations L6V and Q15H lie on the surface of the molecule and are accessible to the solute. Shades of grey indicate the individual monomers as part of the PEPC tetramer, Valine- cyan; Serine 11- yellow; Histidine- green.

Table 1: Amino acid exchanges in PPCK isoforms occurring among C_3 and C_4 *Flaveria* species compared to C_3 and C_4 species from *Brassicales* and *Poaceae*. Amino acids are listed using standard single amino acid code and with positions marked according to the *F. pringlei* PPCKA sequence. An alignment of all sequences is presented in Supplemental File 3. Light grey- amino acid exchanges found in some intermediate species and species with C_4, medium grey –amino acid exchanges found in all intermediate species and species with C_4, dark grey- amino acid exchanges found only in species with C_4 photosynthesis. Red - C_3 species. Blue - C_4 species. White - C_3-C_4 intermediates and C_4-like *F. brownii* (*Fbr*). "-" -Sequence information is missing or residue is not present. Species abbreviations: *Brassicales- Arabidopsis thaliana (At), Cleome gynandra* (*Cg*); *Poaceae - Zea mays* (*Zm*)*, Oryza sativa* (*Os*), *Setaria italica* (*Si*), *Sorghum bicolor* (*Sb*). The accession numbers for the sequences are listed in supplemental file 2, except Cg. The sequence for Cg is included at the end of supplemental file 1.



S1: Supplemental file 1- *PPCK* and *ppc* sequences from *Flaveria*. *PPCK* and *ppc* sequences from *Cleome gynandra*

>FpPPCKA
ATGAAAGAAGCTCTGAACAACGACTATCACATAACTGAAGAGATCGGAAGAGGCCGTTTCGGCACCGTCTACCGCTGCTACTCCACCGTCTCCGGC
GAATCCTTCGCCTGTAAATCGATCGACAAACGCCTCCTCAACGACCACACCGACCGCGAGTGCCTGCAAAAAGAGCCTAAAATCCTTCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTATCGTCTGTACGAAGATGAGAATTACGTTCATATGATTATCGATCTCTGCGACTCGCCGGATCTCTTC
GATCGGATCTCGAACCGCAACGGAGTGTTCTCGGAGACTGAAGCCGCGACGGTGTTCACACCGTTAATGTCAGCGATTAGTTATTGTCATCGGTTA
GGTATCGCGCATCGGGATATTAAGCCGGATAATGTGTTGTTCGATTCGAGAGGGAGTTTGGAGCTAGCGGATTTTGGATCTGCGGAGTGGTTTGGG
ATGAATGAGCGGGGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTCGCGCCGGAGGTGTTGTCCGGTAGAGAGTATAATGAGAAGGTGGAT
GTGTGGAGCGCCGGTGTGATTTTGTACATAATGCTCGCCGGAGTTCCTCCGTTTTACGGTGACTCGCCGGCGGAGACTTTTGAAGCGGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCGGAAACCAAGGATCTGTTGAGGAAGATGCTCTGCAAAGATGTTTCTAGAAGA
TTGTCCGCCGATCAAGTCCTCAGGCATCCATGGGTTGTAAGTGGAGGTGAAACTCGATCTATGGCGGATCTAACCTAA

>FroPPCKA
ATGAAAGAAGCTCTGAACAACGATTATCACATCACCGATGAGATCGGCAGAGGCCGTTTCGGCACTGTATACCGGTGCTACTCCACCGTCTCCGGC
GACTCCTTCGCCTGCAAATCAATCGACAAACGCCTCCTCAACGACCACACCGACCGCGAGTGCCTGCAAAAAGAGCCTAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTATCGCCTGTACGAAGAAGAGAACTACGTTCACATGATCATCGATCTCTGCGACTCGCCGGATCTCTTC
GACAGGATCTCGAACCGCAACGGAGTGTTCTCGGAGGCCGAAGCCTCGACGGTCTTCACACCGTTAATGTCAGCGATTAGTTATTGTCATCGGTTA
GGTATCGCGCATCGAGATATTAAGCCGGATAACGTGTTGTTCGATTCGCGAGGGAGTTTGAAGCTTGCCGATTTTGGATCTGCGGAGTGGTTCGGG
ATGAATGAGCGGGGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTGGCGCCGGAGGTGTTGTCCGGTAGGGAGTATAACGAGAAGGTGGAT
GTGTGGAGCGCCGGTGTGATTTTGTACATAATGCTCGCCGGAGTTCCTCCGTTCTACGGTGACTCACCGGCGGAGACTTTTGAGGCGGTTTTGCGC
GGGAATTTGCGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCGGAAGCCAAGGATCTGTTGAGGAAGATGCTCTGCAAAGATGTTTCTAGAAGA
TTGTCCGCCGATCAAGTCCTCAGGCATTCGTGGGTTGTAAGTGGAGGTGAAACTCGATCTATGGCGAATCTAACCTAA

>FraPPCKA
ATGAAAGAAGCTCTGAACAACGACTATCACATCACTGAAGAGATCGGCAGAGGCCGTTTCGGCACCGTCTACCGCTGCTACTCCACCGTCTCCGGC
GACTCCTTCGCCTGCAAATCAATCGACAAACGCCTCCTCAACGACCACACCGACCGCGAATGCCTGCAAAAAGAGCCCAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTACCGTCTCTACGAAGACGACAACTACGTTCACATGATCATCGACCTCTGCGACTCACCGGATCTCTTC
GACCGGATCTCGAACCGCAACGGCGTGTTCTCGGAGACCGAAGCCGCGACGGTGTTCACACCGTTAATGTCAGCGATTAGCTACTGTCATCGGTTA
GGTATCGCGCATCGTGATATCAAGCCTGATAACGTGTTGTTCGATTCGCGAGGGAGTTTGAAGCTTGCGGATTTCGGATCTGCGGAGTGGTTCGCG
ATGAATGATCGGAGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTGGCGCCGGAGGTGTTGTCCGGTATGGAGTATAACGAGAAGGTGGAT
GTGTGGAGCGCCGGTGTGATTCTGTACATAATGCTCGCCGGAGTTCCTCCGTTTTACGGTGACTCGCCGGCGGAGACTTTTGAAGCTGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCGGAAGCTAAGGATCTGTTGAGGAAGATGCTCTGTAAAGACGTTTCTAGAAGA
TTCTCCGCAGATCAAGTCCTCAGGCATCCATGGGTTGTAAATGGGGGAGAAACTCGATCAATGGCGGATCTAACCTAA

>FcPPCKA
ATGAAAGAAGCTCTGAACAACGACTATCACATCACTGAAGAGATCGGCAAAGGCCGTTTCGGCACCGTCTACCGCTGCTACTCCACCGTCTCCGGC
GACTCCTTCGCCTGCAAATCAATCGACAAACGCCTCCTCAACGACCACACCGACCGAGAGTGCCTGCAAAAAGAGCCCAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTACCGTCTCTTCGAAGACGACAACTACGTCCACATGATCATCGACCTCTGCGACTCGCCGGATCTATTC
GACCGGATCTCGAACCGCAACGGCGTGTTCTCGGAGACCGACGCCGCGACGGTGTTCACACCGTTAATGTCAGCGATTAGCTACTGTCATCGGTTA
GGTATCGCGCATCGTGATATCAAGCCTGATAACGTGTTGTTCGATTCGAGAGGGAGTTTGAAGCTTGCGGATTTCGGATCTGCGGAGTGGTTCGCG
ATGAATGAGCGGAGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTCGCCCCGGAGGTGTTGTCAGGTATGGAGTATAACGAGAAGGTGGAT
GTGTGGAGCGCCGGAGTGATTCTGTACATAATGCTCGCCGGAGTTCCTCCGTTTTACGGTGATTCGCCGGCGGAGACTTTTGAAGCGGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTTCGGTCAGTTTCGCCTGAAGCTAAGGATCTGTTGAGGAAGATGCTCTGCAAAGATGTTTCTAGAAGA
TTCTCCGCAGATCAAGTCCTCAGGCATCCATGGGTTGTAAGTGGAGGAGAAACTCGATCAATGGCGGATCTAACCTAA

>FpuPPCKA
ATGAAAGAAGCTCTGAACAACGACTATCACATCACTGAAGAGATCGGCAGAGGCCGTTTCGGCACCGTCTACCGCTGCTACTCCACCGTCTCCGGC
GACTCGTTCGCCTGCAAATCAATCGACAAACGCCTCCTCAACGACCACACCGACCGCGAGTGCCTGCAAAAAGAGCCCAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTACCGTCTCTACGAAGACGACAACTACGTCCACATGATCATCGACCTCTGCGACTCGCCGGATCTCTTC
GACCGGATCTCGAACCGCAACGGCGTGTTCTCGGAGACCGAAGCCGCGACGGTGTTCACACCGTTAATGTCAGCGATTAGCTACTGTCATAGCTTA
GGTATCGCGCATCGTGATATCAAGCCTGATAACGTGTTGTTCGATTCGAGAGGGAGTTTGAAGCTTGCGGATTTCGGATCTGCGGAGTGGTTCGCG
ATGAATGAGCGGAGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTCGCGCCGGAGGTGTTGTCCGGTATGGAGTATAACGAGAAGGTGGAT
GTGTGGAGCGCCGGTGTGATTCTGTACATAATGCTCGCCGGAGTTCCTCCGTTTTACGGCGACTCGCCGGCGGAGACTTTTGAAGCGGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCTGAAGCTAAGGATCTGTTGAGGAAGATGCTCTGCAAAGATGTTTCTAGAAGA
TTCTCCGCAGATCAAGTCCTCAGGCATCCATGGGTTGTAAGTGGAGGAGAAACTCGATCAATGGCGGATCTAACCTAA

>FaPPCKA
ATGAAGGAAGCTCTGAACAACGACTATCACATCACTGAAGAGATCGGCAGAGGCCGTTTCGGCACCGTCTACCGGTGCTACTCCACCGTCTCCGGC
GACTCCTTCGCCTGCAAATCAATCGACAAACGCCTCCTCAACGACCACACCGACCGTGAGTGCCTGCAAAAAGAGCCCAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTACCGTCTCTACGAAGACGACAACTACGTCCACATGATCATCGACCTCTGCGACTCGCCGGATCTCTTC
GACCGGATCTCGAACCGCAACGGCGTGTTCTCGGAGACCGAAGCCGCGACGGTGTTCACACCGTTAATGTCAGCGATTAGCTACTGTCATCGGTTA
GGTATCGCGCATCGTGATATCAAGCCTGATAACGTGTTGTTCGATTCGCGAGGGAGTTTGAAGCTTGCGGATTTCGGATCTGCGGAGTGGTTCGCG
ATGAATGATCGGAGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTCGCGCCGGAGGTGTTGTCCGGTATGGAGTATAACGAGAAGGTGGAT
GTGTGGAGCGCCGGTGTGATTCTGTACATAATGCTCGCCGGAGTTCCTCCGTTTTACGGTGACTCGCCGGCGGAGACTTTCGATGCGGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCGGAAGCTAAGGATCTGTTGAGGAAGATGCTCTGTAAAGACGTTTCTAGAAGA
TTCTCCGCAGATCAAGTCCTCAGGCATCCATGGGTTGTAAGTGGAGGAGAAACTCGATCAATGGCGGATCTAACCTAA

>FbrPPCKA
ATGAAAGAAGCTCTGAACAACGACTATCACATCACTGAAGAGATCGGCAGAGGCCGTTTCGGCACCGTCTACCGCTGCTACTCCACCGTCTCCGGC
GACTCCTTCGCCTGCAAATCAATCGACAAACGCCTCCTCAACGACCACACCGACCGCGAGTGCCTGCAAAAAGAGCCCAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTACCGTCTCTACGAAGACGACAACTACGTCCACATGATCATCGACCTCTGCAACTCGCCGGATCTCTTC
TACCGGATCTCGAACCGCAACGGCGTGTTCTCGGAGACCGAAGCCGCGACGGTGTTCACACCGTTAATGTCAGCGATTAGCTACTGTCATCGGTTA
GGTATCGCGCATCGTGATATCAAGCCTGATAACGTGTTGTTCGATTCGAGAGGGAGTTTGAAGCTTGCGGATTTCGGATCTGCGGAGTGGTTCGCG
ATGAATGAGCGGAGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTCGCGCCGGAGGTGTTGTCCGGTATGGAGTACAACGAGAAGGTGGAT
GTGTGGAGCGCCGGTGTGATTCTGTACATAATGCTCGCCGGAGTTCCTCCGTTTTACGGTGACTCGCCGGCGGAGACTTTTGAAGCGGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCTGAAGCTAAGGATCTGTTGAGGAAGATGCTCTGCAAAGATGTTTCTAGAAGA
TTCTCCGCAGATCAAGTCCTCAGGCATCCATGGGTTGTAAGTGGAGGAGAAACTCGATCAATGGCGGATCTAACCTAA

>FtPPCKA
ATGAAGGAAACTCTGAACAACGACTATCACATCACTGAAGAGATCGGCAGAGGCCGTTTCGGCACCGTCTACCGCTGCTACTCCACCGTCTCCGGC
GACTCCTTCGCCTGCAAAACAATCGACAAACGCCTCCTCAACGACCACACCGACCGCGAATGCCTGCAAAAAGAGCCCAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTACCGTCTCTACGAAGACGACAACTACGTCCACATGATCATCGACCTCTGCGACTCGCCGGATCTCTTC
GACCGGATCTCGAACCGCAACGGCGTGTTCTCCGAGACCGAAGCCGCGACGGTGTTCGCACCGTTAATGTCAGCGATTAGCTACTGTCACCGGTTA
GGTATCGCGCATCGTGATCTCAAGCCTGATAACGTGCTGTTCGATTCGCGAGGGGGTTTGAAGCTTGCGGATTTCGGATCTGCGGAGTGGTTCGCG
ATGAATGATCGGAGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTCGCGCCGGAGGTGTTGTCCGGTATGGAGTATAACGAGAAGGTTGAT
GTGTGGAGCGCCGGTGTGATTCTGTACATAATGCTCGCCGGAGTTCCTCCGTTTCACGGTGACTCGCCGGCGGACACTTTTGAAGCTGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCGGAAGCTAAGGATCTGTTGAGGAAGATGCTCTGTAAAGATGTTTCTAGAAGA
TTCTCCGCAGATCAAGTCCTCAGGCATCCATGGGTTGTAAACGGGGGAAAAACTCGATCAATGGCGGATCTGACCTAA

>FbPPCKA
ATGAAGGAAACTCTGAACAACGACTATCACATCACTGAAGAGATCGGCAGAGGCCGTTTCGGCACCGTCTACCGCTGCTACTCCACCGTCTCCGGC
GACTCCTTCGCCTGCAAATCAATCGACAAACGCCTCCTCAACGACCACACCGACCGCGAATGCCTGCAAAAAGAGCCCAAAATCCTCCACATCCTC
GGCGGAAACCCTAACATCGTCCAGATCTACCGTCTCTACGAAGACGACAACTACGTACACATGATCATCGACCTCTGCGACTCGCCGGATCTCTTC
GACCGGATCTCGAACCGCAACGGCGTGTTCTCCGAGACCGAAGCCGCGACGGTGTTCACACCGTTAATGTCAGCGATTAGCTACTGTCACCGGTTA
GGTATCGCGCATCGTGATCTCAAGCCTGATAACGTGTTGTTCGATTCGCGAGGGGGTTTGAAGCTCGCGGATTTCGGATCTGCGGAGTGGTTCGCG
ATGAATGATCGGAGGACGATGAGTGGAATCGTCGGAACGCCGTACTACGTCGCGCCGGAGGTGTTGTCCGGTATGGAGTATAACGAGAAGGTTGAT
GTGTGGAGCGCCGGTGTGATTCTGTACATAATGCTCGCCGGAGTTCCTCCGTTTCACGGTGACTCGCCGGCGGACACTTTTGAAGCTGTTTTGCGC
GGGAATTTACGGTTTCCGACGAGGCTTTTCCGGTCAGTTTCGCCGGAAGCTAAGGATCTGTTGAGGAAGATGCTCTGTAAAGATGTTTATAGAAGA
TTCTCCGCAGATCAAGTCCTCAGGCATCCATGGGTTGTAAATGGGGGAAAAACTCGATCAATGGCGGATCTAACCTAA

>FpPPCKB
ATGAAGGAAGTTATTCACAACGACTATCAACTTACCGACGAAATCGGCAGAGGTCGTTTCGGAACCGTCTACCGCTGTTACTCCACCGTCTCCGGC
GACACTTACGCCTGTAAATCAATCGACAAACGCCTCCTAACGGATCACACCGACCGTGAGTGCCTTCAAAAAGAGCCCAAAATCCTTCATATCCTC
GGCGGAAACCCTAACATCGTTCAGATCCATCGGTTGTACGAAGATGAGAACTATGTTCATATGGTTATTGACCTCTGTGAGTCGCCGGACCTTTTC
GATCGGATTTCGAGTCGTGCCGGAGTGTTTTCGGAGTCAGAATCGGCGTCGGTTTTTGGTCCGTTGATGTCGGCGATCAGTTATTGTCATCGGTTA
GGTATTGCGCATCGGGATATTAAGCCGGATAATGTGTTGTTTGATTGTAGGGGGAGTTTGAAGCTGGCGGATTTCGGATCGGCGGAGTGGTTTGCG
ATGAGTGAGAGGGGGACGATGACCGGAGTTGTTGGGACACCGTATTACGTGGCGCCGGAGGTGTTGTTAGGGCGGGAGTATAATGAGAAGGTGGAT
GTGTGGAGTGCTGGTGTGATTTTGTACGTAATGTTGGCCGGAGTTCCTCCGTTTTACGGTGAGTCGCCGGCGGATACTTTTGAGGCGGTTTTGAGG
GGGAATTTACGGTTCCCGACGAAGATTTTCAGGTCGGTTTCGCCGGAAGCGAAGGATCTGTTGAAGAAGATGCTTTGTAAAGATGTATCTAGAAGA
TTATCTGCAGAACAAGTTCTGAGGCATCCGTGGGTTGTAAGTGGAGGTGAAACTCGAGCAATGGCGGATCTAACCTAA

>FroPPCKB
ATGAAGGAAGTTATTCACAACGACTATCAACTTACCGACGAAATCGGCAGAGGTCGTTTTGGAACCGTCTACCGCTGTTACTCCGTCGTCTCCGGC
GACACTTTCGCCTGTAAATCAATCGACAAACGCCTTCTCACTGATCACACTGACCGTGAGTGCCTACAAAAAGAGCCTAAAATTCTTCATATACTT
GGCGGAAACCCTAACATCGTTCAGATCCATCGGTTGTACGAAGATGAGAACTACGTTCATATGGTTATTGACCTCTGTGAGTCACCGGACCTTTTC
GATCGGATTTCGACGCGTGCCGGAGTGTTTTCGGAGTCCGAATCGGCGTCGGTTTTCGCTCCGTTGATGATGGCGATCAGTTATTGTCATCGGTTA
GGTATAGCGCATCGGGATATTAAGCCGGATAATGTGTTGTTTGATTGCAGGGGGAATTTGAAGCTGGCGGATTTCGGATCGGCGGAGTGGTTTGCG
ATGAGTGAGAGGGGGACGATGACCGGAGTTGTAGGTACGCCGTACTATGTGGCGCCGGAGGTTTTGTTAGGGAGGGAGTATAATGAGAAGGTTGAT
GTGTGGAGTGCTGGTGTGATTTTGTACATAATGTTGGCCGGAGTTCCTCCGTTTTACGGCGAGTCGCCGGCGGATACTTTTGAGGCGGTTTTGCGA
GGGAATTTACGGTTCCCGACGAAGATTTTCAGGTCGGTTTCGCCGGAAGCGAAGGATCTGTTGAAGAAGATGCTGTGTAAAGATGTATCAAGACGA
TTATCCGCAGAACAAGTTCTGAGGCATCCGTGGGTTGTAAGTGGAGGTGAAACTCGATCTATGGCGGATCTAACCTAA

>FraPPCKB
CTACCGCTGTTACTCCGTCGTCTCCGGCGACACTTACGCCTGTAAATCAATCGACAAACGCCTCCTAACAGATCACACCGACCGTGAGTGCCTTCA
AAAAGAGCCCAAAATCCTTCATATTCTCGGTGGAAACCCTAACATCGTTCAGATCCATCGGTTGTACGAAGATGAGAACTATGTTCATATGGTTAT
GGACCTCTGTGAGTCGCCGGACCTTTTTGATCGGATTTCGAGCCGTGCCGGAGTGTTTTCGGAGGCTGAATCGGCGTCGGTTTTTGGTCCGTTGAT
GTCGGCGATTAGTTATTGTCATCGGTTAGGTATTGCGCATCGGGATATTAAGCCGGATAACGTGTTGTTTGATTGTAGGGGGAGTTTGAAGCTGGC
GGATTTCGGATCGGCGGAGTGGTTTGCGATGAGTGAGAGGGGGACGATGACCGGAGTTGTTGGGACACCGTATTACGTGGCGCCGGAGGTGTTGTT
AGGGCGGGAGTATAACGAAAAGGTGGATGTGTGGAGTGCTGGTGTGATTTTGTACGTAATGTTGGCCGGAGTTCCTCCGTTTTACGGCGAGTCTCC
GGCGGATACTTTTGAGGCGGTTTTGAGGGGGAATTTACGGTTCCCGACGAAGATTTTCAGGTCGGTTTCGCCGGAAGCGAAGGATCTGTTGAAGAA
GATGCTTTGTAAAGATGTTTCTAGACGATTATCTGCAGAACAAGTTCTGAGGCATCCATGGGTTGTAAGTGGAGGTGAAACTCGAGCAATGGTGGA
TCTAACCTAG

>FcPPCKB
TCGTCTACCGCTGTTACTCCACCGTCTCCGGCGACACTTACGCCTGTAAATCAATCGACAAACGCCTCCTAACGGATCACACCGACCGTGAGTGCC
TACAAAAAGAGCCCAAAATTCTTCATATTCTCGGTGGAAACCCTAACATCGTTCAGATCCATCGGTTGTACGAAGATGAGAACTATGTTCATATGG
TTATTGACCTCTGTGAGTCGCCGGACCTTTTCGATCGGATTTCGAGTCGCGCCGGAGTGTTTTCGGAGTCGGAATCGGCGTCGGTTTTCGGTCCGT
TGATGTCGGCGATCAGTTATTGTCATCGGTTAGGTATTGCGCATCGGGATATTAAGCCGGATAACGTGTTGTTTGATTGTAGGGGGAGTTTGAAGC
TGGCGGATTTCGGATCGGCGGAGTGGTTTGCGATGAGTGAGAGGGGGACGATGAACGGAGTTGTAGGGACACCGTATTATGTGGCGCCGGAGGTGC
TGTTAGGGCGGGAGTATAATGAGAAGGTGGATGTGTGGAGTGCTGGTGTGATTTTGTACGTAATGTTGGCCGGAGTTCCTCCGTTTTACGGCGAGT
CTCCGGCGGATACTTTTGAGGCGGTTTTGAGGGGGAATTTACGGTTCCCGACGAAGATGTTCAGGTCGATTTCGCCGGAAGCGAAGGATCTGTTGA
AGAAGATGCTTTGTAAAGATGTATCTAGACGATTATCTGCAGAACAAGTTCTGAGGCATCCGTGGGTTGTAAGTGGAGGTGAAACTCGAGCAATGG
CGGATCTAACCTA

>FpuPPCKB
GTTTCGGCATCGTCTACCGCTGTTACTCCACCGTCTCCGGCGACACTTACGCCTGTAAATCAATCGACAAACGCCTCCTAACGGATCACACCGACC
GTGAGTGCCTACAAAAAGAGCCCAAAATTCTTCATATTCTCGGTGGAAACCCTAACATCGTTCAGATCCATCGGTTGTACGAAGATGAGAACTATG
TCCATATGGTTATTGACCTCTGTGAGTCGCCGGACCTTTTCGATCGGATTACGAGTCGCGCCGGAGCGTTTTCGGAGTCGGAATCGGCGTCGGTTT
TCGGTCCGTTGATGTCGGCGATCAGTTATTGTCATCGGTTAGGTATTGCGCATCGGGATATTAAGCCGGATAACGTGTTGTTTGATTGTAGGGGGA
GTTTGAAGCTGGCGGATTTCGGATCGGCGGAGTGGTTTGCGATGAGTGAGAGGGGGACGATGACCGGAGTTGTAGGGACACCGTATTATGTGGCGC
CGGAGGTGCTGTTAGGGCGGGAGTATAATGAGAAGGTGGATGTGTGGAGTGCTGGTGTGATTTTGTACGTAATGTTGGCCGGAGTTCCTCCGTTTT
ACGGCGAGTCTCCGGCGGATACTTTTGAGGCGGTTTTGAGGGGGAATTTACGGTTCCCGACGAAGATGTTCAGGTCGGTTTCGCCGGAAGCGAAGG
ATCTGTTGAAGAAGATGCTTTGTAAAGATGTATCTAGACGATTATCTGCAGAACAAGTTCTGAGGCATCCGTGGGTTGTAAGTGGAGGTGAAACTT
GAGCAATGGCGGATCTAACCTA

>FaPPCKB
GGCGGATACTTGTGAGGCGGTTTTGAGGGGGAATTTACGGTTCCCGACGAAGATGTTCAGGTCGGTTTCGCCGGAAGCGAAGGATCTGTTGAAGAA
GATGCTTTGTAAAGATGTATCTAGACGATTATCTGCAGAACAAGTTCTGAGGCATCCGTGGGTTGTAAGTGGAGGTGAAACTCGAGCAATGGCGGA
TCTAACGTAG

>FbrPPCKB
ATGAAGGAAGTTATTCACAACGACTATCAACTCACCGAAGAAATCGGCAGAGGTCGTTTCGGCATCGTCTACCGCTGTTACTCCACCGTCTCCGGC
GACACTTACGCCTGTAAATCAATCGACAAACGCCTCCTGACGGATCACACCGACCGTGAGTGCCTACAAAAAGAGCCCAAAATTCTTCATATTCTC
GGTGGAAACCCTAACATCGCTCAGATCCATCGGTTGTACGAAGATGAGAACTATGTTCATATGGTTATTGACCTCTGTGAGTCGCCGGACCTTTTC
GATCGGATTTCGAGTCGCGCCGGAGTGTTTTCGGAGTCGGAATCGGCGTCGGTTTTCGGTCCGTTGATGTCGGCGATCAGTTATTGTCATCGGTCA
GGTATTGCGCATCGGGATATTAAGCCGGATAACGTGTTGTTTGATTGTAGGGGGAGTTTGAAGCTGGCGGATTTCGGATCGGCGGAGTGGTTTGCG
ATGAGTGAGAGGGGGACGATGACCGGAGTTGTAGGGACACCGTATTATGTGGCGCCGGAGGTGCTGTTAGGGCGGGAGTATAATGAGAAGGTGGAT
GTGTGGAGTGCTGGTGTGATTTTGTACGTAATGTTGGCCGGAGTTCCTCCGTTTTACGGCGAGTCTCCGGCGGATACTTTTGAGGCGGTTTTGAGG
GGGAATTTACGGTTCCCGACGAAGATGTTCAGGTCGGTTTCGTCGGAAGCGAAGGATCTGTTGAAGAAGATGCTTTGTAAAGATGTGTCTAGACGA
TTATCTGCAGAACAAGTTCTGAGGCATCCGTGGGTTGTAAGTGGAGGTGAAACTCGAGCAATGGCGGATCTAACCTAG

>FtPPCKB
CTACCGCTGTTACTCCGTCGTCTCCGGCGACACTTACGCCTGTAAATCAATCGACAAACGCCTCCTAACGGATCACATCGACCGTGAGTGCCTTCA
AAAAGAGCCCAAAATCCTTCATATCCTCGGTGGAAACCCTAACATCGTTCAGATCCATCGGTTGTACGAAGACGAGAACTATGTTCATATGGTTAT
GGACCTCTGTGAGTCGCCGGACCTTTTTGACCGGATTTCGAGCCGTGCCGGAGTGTTTTCGGAGGCTGAATCGGCGTCGGTTTTTGGTCCGTTGAT
GTCGGCGATCGGTTATTGTCATCAGTTAGGTATTGCGCATCGGGATATTAAGCCGGATAATGTGTTGTTTGATTGTAGGGGGAGTTTGAAGCTGGC
GGATTTCGGATCGGCGGAGTGGTTTGCGATGAGTGAGTGGGGGACGATGACCGGAGTTGTCGGGACACCGTATTACGTGGCGCCGGAGGTGCTGTT
AGGGCGTGAGTATAACGAGAAGGTGGATGTGTGGAGTGCTGGTGTGATTTTGTACGTAATGTTGGCCGGAGTTCCTCCGTTTTACGGTGAGTCTCC
GGCGGATACTTTTGAGGCGGTTTTGAGGGGGAATTTACGGTTCCCGACGAAGATGTTCAGGTCGGTTTCGCCGGAAGCGAAGGATCTGTTGAAGAA
GATGCTCTGTAAAGATGTATCTAGAAGATTATCCGCAGAACAAGTTCTGAG

>FbPPCKB
ATGAAGGAAGTTATTCACAACGACTATCAACTCACCGACGAAATCGGCAGAGGTCGTTTCGGCATCGTCTACCGCTGTTACTCCGTCGTCTCCGGC
GACACTTACGCCTGTAAATCAATCGACAAACGCCTCCTAACAGATCACATCGACCGTGAGTGCCTTCAAAAAGAGCCCAAAATCCTTCATATCCTC
GGCGGAAACCCTAACATCGTTCAGATCCATCGGTTGTATGAAGATGAGAACTATGTTCATATGGTTATGGACCTCTGTGAGTCGCCGGACCTTTTT
GATCGGATTTCGAGCCGTGCCGGAGTGTTTTCGGAGGCTGAATCGGCGTCGGTTTTTGGTCCGTTGATGTCGGCGATCGGTTATTGTCATCGGTTA
GGTATTGCGCATCGGGATATTAAGCCGGATAATGTGTTGTTTGATTGTAGGGGGAGTTTGAAGCTGGCGGATTTCGGATCGGCGGAGTGGTTTGGG
ATGAGTGAGCGGGGGACGATGACCGGAGTTGTTGGGACACCGTATTACGTGGCGCCGGAGGTGCTGTTAGGACGTGAGTATAACGAGAAGGTGGAT
GTGTGGAGTGCTGGTGTGATTTTGTACGTAATGTTGGCCGGAGTTCCTCCGTTTTACGGTGAGTCTCCGGCGGATACTTTTGAGGCGGTTTTGAGG
GGGAATTTACGGTTCCCGACGAAGATGTTCAGGTCGGTTTCTCCGGAAGCGAAGGATCTGTTGAAGAAGATGCTCTGTAAAGATGTATCTAGAAGA
TTATCTGCAGAACAAGTTCTGAGGCATCCGTGGGTTGTAAGTGGGGGTGAAACTCGAGCAATGGCGGATCTAACCTAG

3‘ UTR regions

>3´UTR-FpPPCKA
TCCGAGATGAAATCAAGCTTTTTCAAACTCTTGGGTATGAAATCAACACTGGTTTTTTGATCTGTTGATTTGAATCCATGATGAATCAACAGAAAG
AAAGAGGGAGAGCGAAAGTTCAATGAATCTTTTGAAGTTGAGGAAGTCTGCATGTACTTTTTTTGTTTGTTTTTGTAATGAATCAAGATGAATGTT
CATCTTCTCCAACTAGTATTACATGGAGAATGAATGAATGAATAATCATCTCTGAGTAGGAGATTTTGTAAATATAGTGAAGAGAGCTTAGGGAGC
TACGATGAATAAAGTGATGGTGTATATGTTAATGGCGGAAAGATGTCATCATCCTCTGTTTTTCTCTCTCTGGTTTTGTTGCCTTTGAATCTGAAT
AATCTGAGTTTTG

>3´UTR-FroPPCKA (partial)
TTTTTCAAACTCATTGGGTATGAAATCAACACTGGTTTTTTGATCTGTTGATTTGAATACATGATGAATCAACAGAAAGAAAGAGAGAGAGCGAAA
GTTCAATGAATCTTCTGAAGTTGAGGAAGTCTGCATGTAGTTGTCTTATGTTTTTGTAATGAATCAAGATGAATGTTCATCTTCTCCAACTAGTAT
TATATGGAGGATGAATGAATGAAT

>3´UTR-FaPPCKA
TTCAAGAAGAAATCAAGGTTTTTAACTCATTGGGTACGAAATCAACACTGGTTTTTTGATCTGTTGATTTGAATACATGATGAATCAACAGAAAGA
AAGAGAGAGAGCGAAAGTTCAATGAATCTTTTGAAGTTGAGGAAGTCTGCATGTAG

>3´UTR-FcPPCKA
TTCGAGAAGAAATCAAGGTTTTTCAAAGTCATTGGGTACGAAATTAACACTGGTTTTTGATCTGTTGATTTGAATACATGATGAATCAACAGAAAG
AAAGAGAGAGAGCGAAAGTTCAATGAATCTTTTGAAGTTGAGGAAGTCTGCATGTAGTTTTTTGTTTTTGTTTTTGTTTTTGTAATGAATCAAGAT
GAATGTTCATCTTCTCCAACTAGTATTATATGGAGGATGAATGAATGAATAATCATCTCTGAGTAGGAGATTTTGTAAATATATGAAGAGAGTTTA
CTCCGATGAATAAAGTGATGGTGTATATGTTAATGGCATAAAGATGTCATCATCCTCTGTTTTTCTTCATCTTGTTTTTTCTGTTTTTATTGCCTA
TGAATCTGAATAATAATCGAGTTATAAATTTGAGTGTTTTTTTTTTTCCTATGAGGTGTAAACTGTAAAGTTCATGTTGCAATTGTTTTCCCATCT
GAATAATAATTGAG

>3´UTR-FraPPCKA
TTCGAGAAGAAATCAAGGTGTTTTTAAACTCATTGGGTACGAGAAATCAATCAACACTGGTTTTTTGATCTGTTGATTTGAATACATGATGAATCA
ACAGAAAGAAAGAGAGCGAAAGTTCAATGAATCTTTTGAAGTTGAGGAAGTCTGCATGTAGTTTTTTTGTTTGTGTAATGAATGAAGATGAATGTT
GAATGTTCATCTTCTCCAACTAGTATTATATGGAGGATGAATGAATGAATAATCATCTCTGAGTAGGAGATTTTGTAAATATATGAAGAGAGTTTA
CTGCGATGAATAAAGTGATTGTGTATATGTTAATGGCA

>3´UTR-FpuPPCKA
TTCGAGAAGAAATCAAGGTTTTTCAAACTCATTGGGTACGAAATCAACACTGGTTTTTGATCTGTTGATTTGAATACATGATGAATCAACAGAAAG
AAAGAGAGAGAGCGAAAGTTCAGTGAATCTTTTGAAGTTGAGGAAGTCTGCATGTAGTTTTTTGTTTTTCTGTTTTTGTAATGAATCAAGATGAAT

GTTCATCTTCTCCAACTAGAATTATATGGAGGATGAATGAATGAATAATCATCTCTGAGTAGGAGATTTTGTAAATATATGAAGAGAGTTTACTCC
GATGAATAAAGTGATGGTGTATATGTTAATGGCATAAAGATGTCATCATCCTCTGTTTTTCTTCATCTTGTTTTTTCTGTTTTTATTGCTTATGAA
TCTGAATAATAATCGAGTTATAAATTTGAG

>3´UTR-FbrPPCKA
TTCGAGAAGAAATCAAGGTTTTTCAAACTCATTGAGTACGAAATCAACACTGGTTTTTGATCTGTTGATTTGAATACATGATGAATCAACAGAAAG
AAAGAGAGAGAGCGAAAGTTCAATGAATCTTTTGAAGTTGAGGAAGTCTGCATGTAGTTTTTTGTTTTTGTTTTTGTTTTTATAATGAATCAAGAT
GAATGTTCATCTTCTCCAACTAGTATTATATGGAGGGTGAATGAATGAATAATCATCTCTGAGTAGGAGATTTTGTAAATATATGAAGAGAGTTTA
CTCCGATGAATAAAGTGATGGTGTATATGTTAATGGCATAAAGATGTCATCATCCTCTGTTTTTCTTCATCTTGTTTTTTCTGTTTTTATTGCCTA
TGAATCTGAATAATAATCTAGTTATAAATTTGAGTGTTTTTTTTTCCTATGAGATGTAAAGTGTAAAGTTCCTGTTGCAATTGTTTTCCCATCTGA
ATAATAATTGCGTTATAAACGAG

>3´UTR-FbPPCKA
GTCGAGAAGAAATCAAGGTTTTTCAATACGAAATCAAGTAATCAACACTGGTTTTTTGATCTGTTGATTTGAATACATGATGAATCAACAGAAAGA
GAGAGAGAGCGAAAGTTCAGTGAATCTTTTGAAGTTGAGCAAGTCTGCATGTAGTATTTTTGTGTTTTGTTTTTTGTTTGTGTAATGAATGAAGAT
GAATGTTCATCTTCTCCGACTAGTATTATATGTGGAGGATGAATGAGTGAATAATCATCTCTGAGCAATAGATTTTGTAAATATATGAAGAGAGAT
TACTGTTTAGGGACCTACGATGAATAAAGTGATGGTGTATATGTT

>3´UTR-FtPPCKA
GTCGAGAAGAAATCAAGGTTTTTCAAACTCATTGGGTACGAAATCAAGTAATCAACACTGGTTTTTTGATCTGTTGATTTGAATACATGATGAATC
AACAGAAAGAGAGAGAGAGCGAAAGTTCAATGAATCTTTTGAAGTTGAGCAAGTCTGCATGTAGTATTTTTGTGTTTTGTTTTTAGTTTTTTGTTT
GTGTAATGAATGAAGATGAATGTTTCGTCTTCTCCGACTAGTATTATCTGGAGGATGAACGAGTGAATAATCATCTCTGAGTAAAAATAGATTTTG
TAAATATATGAAGAGAGTTTACTGTTTTAAAGATGTCATCATCCTCTGTTTCTCATGTTGTTTTTTTTTTTCTGTTTTTACTGCCTATGAATCTGA
ATAATAATCGGGAGTTATA

>3´UTR-FpPPCKB (partial)
AGGAAAGGTTTCTTCATGGGTATGAAATCAACACTGGTTTTTTGATCTGTTGATTTGAATCCATGATGAATCAACAGAAAGAAAGAGGGAGAGCGA
AAGTTCAATGAATCT

>3´UTR-FroPPCKB
TAGAGGAAAGGTTATTTCAGGGGTATGAAATCAACACTGGTTTTTTTCATCGGTTGATTTTGATCGTTGATGAATCAACAGAAAGAAAGAGAGAGA
GCAAAAGTTCAATGAATCTTCTGAAGTTGAGGAATTCTGCACTTGTTTTTTGGTTTTGTAATCGGATCGAGATGATGCTGATTCTTCTTATTCTCT
TGCTAGTATTATTTGGGGAATGAATGAATGTATGTATGAATGAATCGCATCTAAGTAGGAGTACCTTGTACTATTTGGTAAAAG

>3´UTR-FcPPCKB
AGGAAAGGTTTCTTCATGGGTATGAAATCAACACTGGTTTTTTCATCTGTTGATTTGAATGGTTGAGAATCAACAGAAAGAAAGAGAGAGAGCGAA
AGTTCAATGAATCTTCTGAAGTTGAGGAGTTCTGCACTTGTTTTTTGGGTTTTTTTAATCGAGATGATGATGATTATGATTCTTCTTCTCTTACTA
GTATTGTTTGGAGAATGAATGAATGTATGAATGAATCGCATCAAGTAGAAGATCTTGTACATTTGGTAAAGAGAAAGTATGGAGTTTACGATGTAT
AAAGCGTTGGTGTAAATGTTCATCCTCTGTTTTCCTTCATCTGGTTTTTTTCTTTACTCTGAGTTTGAATAGCAGAGTGTCAAAGTTATCTGATCAG
AAATTTGCATCAAAATATGTTTGCTTTAAAAATTCTTAAGGAAATATCATTAAGAGTGAATTGCCGTTTTCTCCCCTAG

>3´UTR-FraPPCKB (partial)
AGGAAAGGTTTCTTCATGGGTATGAAATCAACACTGGTTTTTTCAT

>3´UTR-FpuPPCKB
AGGAAAGGTTTCTTCATGGGTATGAAATCAACACTGGTTTTTTCATCTGTTGATTTGAATGGTTGATGAATCAACAGAAAGAAAGAGAGAGAGCGA
AAGTTCAATGAATCTTCTGAAGTTGAGGAGTTCTGCACTTTTTTTTTTTGGGTTTTTTTAATCGAGATGATGATGATTATGATTCTTCTCTTACTAG
TATAGTTTGGAGAATGAATGAATGTATGAATGAATCGCATCAAGTAGAAGATCTTGTACATTTGGTAAAGAGAAAGTATGGAGTTTACGATGTATA
AAGCGTTGGTGTAAATGTTCATCCTCTGTTTTCTTCATCTGGTTTTTTCTTTACTCTGAGTTTGAATAGCAGAGTGTCAAGGTTATCTGATCAGAA
ATTTGCAT

>3´UTR-FbrPPCKB
AGGAAAGGTTTCTTCATGGGTATGAAATCAACACTGGTTTTTTCATCTGTTGATTTGAATGGTTGAGAATCAACAGAAGAAAGAGAGAGAGCGAAA
GTTCAATGAATCTTCTGAAGTTGAGGAGTTCTGCACTTGTTTTTTGGGTTTTTTTAATCGAGATGATGATGATTATGATTCTTCTTCTCTTACTAG
TATTGTTTGGAGAATGAATGAATGTATGAATGAATCGCATCTAAGTAGAAGATCTTGTACATTTGGTAAAGATAAAGTATATGGAGTTTACGATGT
ATAAAGCGTTGGTGTAAATGTTCATACTCTGTTTCCTTCATCTGGTTTTTTCTTTACTCTGAGTTTGAATAGCAGAGTGTCAAAGTTATCTGATCA
GAAATTTG

>3´UTR-FbPPCKB
AGGAAAGGTTTCTTCATGGGTATGAAATCAACACTGGTTTTTTCATCTG

>3´UTR-FtPPCKB
AGGAAAGGTTTCTTCATGGGTATGAAATCAACACTGGTTTTTTCATCTGTTGATTTGAATGGTTGATGAATCAACAGAAAAGAGAGTGAGAGCGAA
AGTTCAATGAATCTTCTAAAGTTGAGGAGTTCTGCACTTGTTTTTTGGTTTTTGTAATCGAGATGATGATGATGATTATGATTCTTCTTCTCTTAC
TAGTATTGTTTGGAGAATGAATGAATGTATGAATGAATTGCATCAACTAGAAGATCTTGTATAAAGAGAAAGTATGGAGTTTACGATGTATAAAAG
CTTTGGTGTAAATGTTCATCCTCTGTTTTCCTTCATCTGGTTTTTTTTTTATTTCTCTGAGTTTGAATAGCAGAGTGTGAAAGTTATCTGAT

Flaveria PEPC protein sequences

>Ft_ppcA_C4
MANRNVEKLASIDAQLRLLVPGKVSEDDKLVEYDALLLDKFLDILQDLHGEDLKEAVQQCYELSAEYEGKHDPKKLEELGSLLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKSGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVELVLTAHPTQSIRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERFPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMTSNMYFSQIEDLMIEMSMWRCNSELRVRAEELYRTSRKDVKHYIEFWKQIPPNQPYRVILGDVRDKLYNTRERSR
HLLVDGKSDIPDEAVYTNVEQLLEPLELCYRSLCDCGDHVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTEVLDAITQHLGIGSYREWSEEKR
QEWLLAELSGKRPLIGPDLPKTEEVKDCLDTFNVLAELPSDCFGAYIISMATSTSDVLAVELLQREYHIKHPLRVVPLFEKLADLEAAPAAMTRLF
SMDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKTQEQIVKIAKEFGVKLVIFHGRGGTVGRGGGPTHLALLSQPPDTINGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSVVFKEPRFVEYFRLATPELEFGRMNIGSRPSKRKPSGGIESLRA
IPWIFSWTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGNPGIAALNDKLLVSEDLRPFGESLRANYEETKNY
LLKIAGHKDLLEGDPYLKQGIRLRDPYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fb_ppcA_C4
MANRNVEKLASIDAQLRLLVPGKVSEDDKLVEYDALLLDKFLDILQDLHGEDLKEAVQQCYELSAEYEGKHDPKKLEELGSLLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKSGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVELVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERFPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMTSNMYFSQIEDLMIEMSMWRCNSELRVRAEELYRTSRKDVKHYIEFWKQIPTNQPYRVILGDIRDKLYNTLERSR
HLLVDGKSDIPDEAVYTNVEQLLEPLELCYRSLCDSGDHVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTEVLDAITQHLRIGSYREWSEEKR
QEWLLAELSGKRPLIGPDLPKTEEVKDVLDTFKVVAELPSDCFGAYIISMATSTSDVLAVELLQREYHIKHPLRVVPLFEKLADLEAAPASMTRLF
SMDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKTQEQIIKIAKEFGVKLVIFHGRGGTVGRGGGPTHLALLSQPPDTINGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSGVLKEPRFVEYFRLATPELEFGRMNIGSRPSKRKPSGGIESLRA
IPWIFSWTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGNPGIAALNDKLLVSEDLRPFGESLRANYEETKNF
LLKIAGHKDLLEGDPYLKQGIRLRDPYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fbr_ppcA_C4-like
MANKNVEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASKMYFSQIEELMFEMSMWRCNSEVRVRAEELYRTARREVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTIERSR
HLLAHGMSDIPDEAVYTNVEQFLEPLELCYRSLCDSGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLKIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEVKDVLDTFNVIAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKHPLRVVPLFEKLADLEAAPASMARLF
SIDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEEIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLALLSQPPDTIHGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSTVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGTAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRIRLRDAYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fa_ppcA_C3-C4
MANKNVEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVISK
AFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKKLVHKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASNMYFSQIEDLMFEMSMWRCNSEGRVRAEELYRTARKDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTLERSR
HLLAHGMSDIPEEAVYTNVEQFLEPLELCYRSLCDSGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLKIGSYREWSEEKR
QEWLLAELSGKRPLLGSDLPKTEEVKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKHPLRVVPLFEKLADLESAPATMARLF
SIDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEDIIKVAKDFGVKLVIFHGRGGTVGRGGGPTHLAILSQPPDTIHGALRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSSVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRIRLRDAYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fpu_ppcA_C3-C4
MANKNVEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASNMYFSQIEDLMFEMSMWRCNSEVRVRAEELYRTARKDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTIERSR
HLLAHGMSDIPDEAVYTNVEQFLEPLELCYRSLCDSGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLKIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEVKDVLDTFNVIAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKHPLRVVPLFEKLADLEAAPASMARLF
SIDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEEIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLALLSQPPDTIHGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSTVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGTAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRIRLRDPYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fc_ppcA_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASNMYFSQIEDLMFEMSMWRCNSELRVRAEELYRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERSR
HLLAHGMSDIPDEAVYTNVEQFLEPLELCYRSLCDCGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEVKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKHPLRVVPLFEKLADLEAAPAAMARLF
SIDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEEIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGTAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRIRLRDSYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fra_ppcA_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASNMYFSQIEDLMFEMSMWRCNSELRVRAEELYRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERSR
HLLAHGISDIPDEAVYTNVEQFLEPLELCYRSLCDCGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTDEVKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKHPLRVVPLFEKLADLEGAPAAMARLF
SIDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEDIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRIRLRDSYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fro_ppcA_C3
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKRLVHKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASNMYFSQIEDLMFEMSMWRCNSELRVRAEELYRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERSR
HLLAHGISDIPEEAVYTNVEQFLEPLELCYRSLCDCGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEVKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKNPLRVVPLFEKLADLEAAPAAMARLF
SIDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEEIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGNPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRIRLRDSYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Fp_ppcA_C3
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKEAVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKRGDFADEANATTESDIEETFKKLVLKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASNMYFSQIEDLMFEMSMWRCNSELRVRAEELYRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERSR
HLLAHGISDIPEEAVYTNVEQFLEPLELCYRSLCDCGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEVKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKHPLRVVPLFEKLADLEAAPAAMARLF
SIDWYRNRIDGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEEIIKVAKEFGVKLVIFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEEYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKHAIKKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHRDLLEGDPYLKQRIRLRDSYITTLNVCQAYTLKRIRDPNYHVTLRPHISKEYAAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAA
GMQNTG*

>Ft_ppcB_C4
MANRNLEKLASIDAHLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSLLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEANATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMTSNMYFSQIEDLMIEMSMWRCIDELRVRAEELYRTARKDVKHYIEFWKQIPPNQPYRVILGDVRDKLYNTRERAR
HLLAHDISDIPEEAVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSTSDVLAVELLQREYHIKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLALLSQPPDTINGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEFGRMNIGSRPSKRKPSGGIESLRA
IPWIFSWTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGNPGIAALNDKLLVSEDLRPFGESLRANYEETKNY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fb_ppcB_C4
MANRNLEKLASIDAHLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMTSNMYFSQIEDLMFEMSMWRCTDELRVRAEELFRTARRDVKHYIEFWKQIPTNQPYRVILGDIRDKLYNTRERAR
HLLAHDISDIPEEAVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEVKDCLDTFKVLAELPSDCFGAYIISMATSTSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTINGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFSWTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGNPGIAALNDKLLVSEDLRPFGESLRANYEETKNF
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fbr_ppcB_C4-like
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEANATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELGEALHREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCTDELRVRAEQLFRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERAR
HLLAHDISDIPEETVYTDVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTIQGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKQAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fa_ppcB_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCTDELRVRAEELFRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYSTRERAR
HLLAHDISDIPEEAVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLDIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTIQGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKQAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fra_ppcB_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMASNMYFSQIEDLMFEMSMWRCTDELRVRAEELFRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERAR
HLLAHDISDIPEEAVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTIQGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPANELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fpu_ppcB_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKHDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCTDELRVRAEQLFRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERAR
HLLAHDISDIPEEAVYTDVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGFSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTIQGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKQAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fc_ppcB_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCTDELRVRAEQLFRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERAR
HLLAHDISDIPEETVYTDVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTIQGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKQAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fro_ppcB_C3
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCTDELRVRAEELFRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERAR
HLLAHDISEIPEEAVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTIQGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKQAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPADELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Ft_ppcC_C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGSLLTSLDAGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEASATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEESVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGPDLPKTEEIADVLDTFHVLAELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSMDWYRNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGVKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVYATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

>Fp_ppcB_C3
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEDLKETVQECYELSAEYEGKRDPKKLEELGSVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDMEETLKKLVYKLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTQPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCTDELRVRAEELFRTARRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERAR
HLLAHDISDIPEEAVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVKLDIRQESDRHTDVLDAITQHLEIGSYREWSEEKR
QEWLLAELSGKRPLFGSDLPKTEEIKDVLDTFNVLAELPSDCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAMARLF
SIDWYKNRINGKQEVMIGYSDSGKDAGRFSAAWQLYKAQEELKNVAKEFGVKLVMFHGRGGTVGRGGGPTHLAILSQPPDTIQGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRELMDQMAVVATEQYRSIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFAWTQTRFHLPVWLGFGAAFKHAIQKDSKNLQMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWPFGESLRANYEETKDY
LLKIAGHKDLLEGDPYLKQRLKLRDSYITTLNVCQAYTLKRTRDPNYHVTLRPHISKEYAEPSKPANELIHLNPTSEYAPGLEDTLILTMKGIAAG
MQNTG*

>Fb_ppcC_C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGSLLTSLDAGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDLADEASATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEESVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGPDLPKTEEIADVLDTFHVLAELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGVKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVYATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

>Fbr_ppcC_C4-like
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGNVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRDCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEESVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGPDLPKTEEIADVLDTFHVLAELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGIKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVYATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSQDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

>Fa_ppcC_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGNVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRDCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEESVYTDVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGPDLPKTADIAGVLDTFHVLAELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGVKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVYATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSQDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

>Fc_ppcC_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDVLLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGNVLTRLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRDCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEESVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGPDLPKTVDIAGVFGYIPCVGELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGIKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVYATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSQDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

>Fpu_ppcC_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDVLLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGNVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTKFDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRDCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEESVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGLDLPKTVDIAGVLDTFHVLAELPADCFSAYIISMATSPSDVLAVELLQRECHVKQPLPVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGVKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVYATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSQDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRISDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

>Fra_ppcC_C3-C4
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGNVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRNCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEESVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGPDLPKTEEIADVLDTFHVLAELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGVKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEIAVHATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

>Fro_ppcC_C3
MASRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGNVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGDFADEAHATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRDCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEELVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAVTQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGTDLPKTEEIADVLDTFHVLAELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGVKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVHATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSQDLWSFGESLRANYEETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYHVTFRPHISKEYSEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

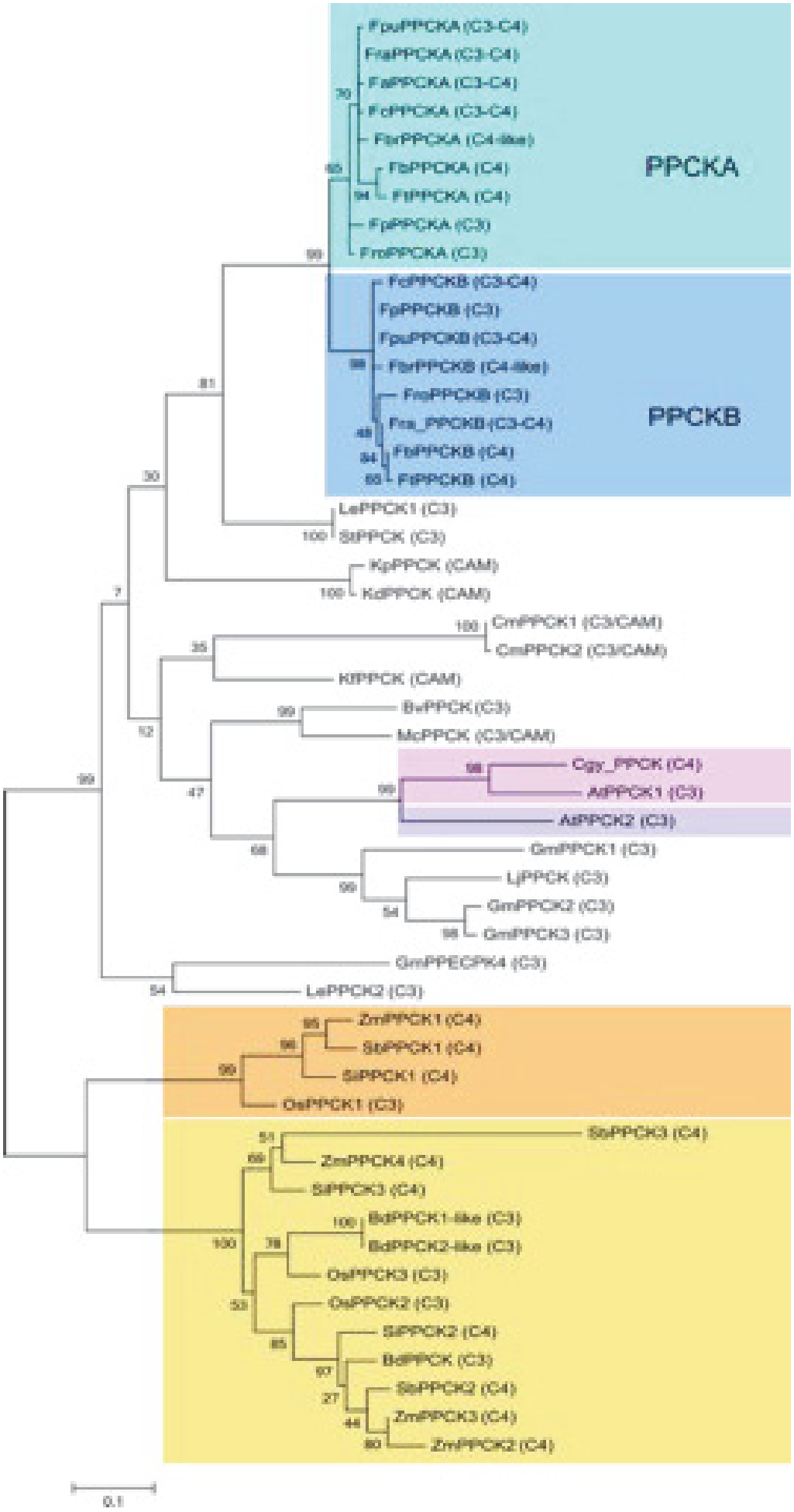
>Fp_ppcC_C3
MANRNLEKLASIDAQLRLLVPGKVSEDDKLIEYDALLLDKFLDILQDLHGEGLKETVQECYELSAEYEGKHDPKKLEELGNVLTSLDPGDSIVIAK
AFSHMLNLANLAEEVQIAYRRRIKLKKGNFADEAHATTESDIEETFKKLVHKLKKSPEEVFDALKNQTVDLVFTAHPTQSVRRSLLQKHGRIRDCL
AQLYAKDITPDDKQELDEALHREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANMYFSQIEDLMFEMSMWRCSDELRVRAEELHRSSSKRDVKHYIEFWKQVPPTEPYRVILGDVRDKLYNTRERA
RHLLAHDVSDIPEELVYTNVEQFLEPLELCYRSLCACGDRVIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITQHLEIGSYREWSEEK
RQEWLLSELSGKRPLFGPDLPKTEEIADVLDTFHVLAELPADCFGAYIISMATSPSDVLAVELLQRECHVKQPLRVVPLFEKLADLEAAPAAVARL
FSIEWYKNRIDGKQEVMIGYSDSGKDAGRLSAAWQLYKAQEELINVAKKFGVKLTMFHGRGGTVGRGGGPTHLAILSQPPETIHGSLRVTVQGEVI
EQSFGEEHLCFRTLQRFCAATLEHGMNPPISPRPEWRALMDEMAVHATEQYREIVFKEPRFVEYFRLATPELEYGRMNIGSRPSKRKPSGGIESLR
AIPWIFAWTQTRFHLPVWLGFGAAFKYAIEKDIKNLHMLQEMYKTWPFFRVTIDLVEMVFAKGDPGIAALNDKLLVSEDLWSFGESLRANYQETKN
LLLKIAGHKDLLEGDPYLRQRLRLRDSYITTLNVCQAYTLKRIRDPNYRVTCRPHISKEYTEPSSKPADEYIKLNPKSEYAPGLEDTLILTMKGIA
AGMQNTG*

Cleome gynandra PPCK protein

>Cgy_contig_1223 (C4)
MSYTQTLHNSSNKYQFCEEIGRGRFGTVTRVYAPSTGDFFACKTIDKASLADDALDRACIDAEPKIMALLSPHPNIVKIHDLIETNDSLSIVMELI
DPAVTIYDRLVSSGPLSEAQTAAYAKQILSAVSHSHRHGVVHRDIKPENILVDLRKDVVKICDFGSGIWLGEGETTVGVVGTPYYVAPEVLTGRPY
AEKVDVWSVGVVLYTMLAGTPPFYGETAEEIFDAVIRGNLRFPPKIFRGVSSAAKDLLRKMICRDVSRRFSAEQSLRHPWIVNAAATEVEESFI

***Cleome gynandra* PEPC protein**

>CgPEPC (C4)
MAAKNLEKMASIDAQLRLLAPGKVSEDDKLVEYDALLLDRFLDILQDLHGEDVRGFVQECYEVAAEYDGKGNTKKLEELGKMLTSLDPGDSIVVTK
AFSNMLSLTNLAEEVQIAHRRRIDKVKKGDFADEASAATESDIEETLKKLLQLKKSPQEVFDALKNQTVDLVLTAHPTQSVRRSLLQKFGRIRNCL
SQLYAKDITPDDKQELDEALQREIQAAFRTDEIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGIDERIPYNAPLIQFSSWMGGDR
DGNPRVTPEVTRDVCLLARMMAANLYFSQIGDLMFELSMWRCNEELRSRAEQLQSSSKKDAKHYIEFWKHIPPNEPYRVVLGDVRDKLYNTRERAR
QLLTDGVSDIPEDAVFTNIDQFLEPLELCYRSLCDCGDRSIADGSLLDFLRQVSTFGLALARLDIRQESERHTDVLDAITKHLDIGSYREWSEEKR
QEWLLSELRGKRPLIGPDLPKTEEIADVLGAFHVISELPPDSFGPYIISMATSTSDVLAVELLQRECGIKHPLRVVPLFEKLADLENAPASVARLF
SIDWYRDRIKGKQEVMIGYSDSGKDAGRLSAAWQLYKTQEQLVKVAKEFGVKLTMFHGRGGTVGRGGGPTHLALLSQPPDTIHGSLRVTVQGEVIE
QSFGEEHLCFRTLQRFTAATLEHGMHPPISPKPEWRALMDEMTVIATQEYRSMVLKEPRFVEYFRSATPELEYGRMNIGSRPSKRKPSGGIESLRA
IPWIFSWTQTRFHLPVWLGIGSAFRHVIQKDAKNLDMLREMYNQWPFFRVTIDLVEMVFAKGDPEIAAMYDKLLVSEELQPFGQQLRTNYEETKSL
LLQVAGHKDILEGDPYLKQMLRLREPYITILNVCQAYTLKQIRDPSFHVKVRPHLSKEYMESSKPAAELVKLNPKSEYAPGLEDTLILTMKGIAAG
MQNTG


FpuPPCKA (C3-C4)
FnaPPCKA (C3-C4)
FalPPCKA (C3-C4)
FcPPCKA (C3-C4)
FsvPPCKA (C4-like)
FtPPCKA (C4)
FtlPPCKA (C4)
FpPPCKA (C3)
FnPPCKA (C3)
PPCKA
FcPPCKB (C3-C4)
FpPPCKB (C3)
FpuPPCKB (C3-C4)
FsvPPCKB (C4-like)
FnPPCKB (C3)
Fna_PPCKB (C3-C4)
FtPPCKB (C4)
FtlPPCKB (C4)
PPCKB
LePPCK1 (C3)
StPPCK (C3)
KpPPCK (CAM)
KdPPCK (CAM)
CmPPCK1 (C3/CAM)
CmPPCK2 (C3/CAM)
KlPPCK (CAM)
BvPPCK (C3)
McPPCK (C3/CAM)
Cgy_PPCK (C4)
AtPPCK1 (C3)
AtPPCK2 (C3)
GmPPCK1 (C3)
LjPPCK (C3)
GmPPCK2 (C3)
GmPPCK3 (C3)
GmPPEPCK4 (C3)
LePPCK2 (C3)
ZmPPCK1 (C4)
SbPPCK1 (C4)
SiPPCK1 (C4)
OsPPCK1 (C3)
SbPPCK3 (C4)
ZmPPCK4 (C4)
SiPPCK3 (C4)
BdPPCK1-like (C3)
BdPPCK2-like (C3)
OsPPCK3 (C3)
OsPPCK2 (C3)
SiPPCK2 (C4)
BdPPCK (C3)
SbPPCK2 (C4)
ZmPPCK3 (C4)
ZmPPCK2 (C4)
0.1

S2: Supplemental file 2- Phylogenetic analysis of PPCK isoforms from various species. The analysis shows that the PPCKs have evolved post speciation of the vascular plants, and evolved in parallel within each closely related group. Within the genus Flaveria, the PPCKA and PPCKB are clearly different from each other, and the segregation in branches is driven by the PPCK type rather than the type of photosynthesis. However, in the case of PPCKA there is further separation of the species with C_3 photosynthesis, and a group formed by the C_3-C_4 intermediates and species with C_4 photosynthesis. A similar grouping of PPCK isoforms is also seen in all included systematics groups where we have a C_4 representative. The isoforms of the genus *Flaveria* are indicated by the two blue backgrounds. Pink and purple indicate the PPCKs from the order *Brassicales*, yellow and orange – family *Poaceae.* Evolutionary analyses were conducted in MEGA5 (Tamura *et al*., 2011; Hall, 2013). Sequences were aligned using Muscle (Edgar, 2004) with default settings for 1000 iterations. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al, 1992). The tree with the highest log likelihood (-5116.5439) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 49 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 173 positions in the final dataset. Species abbreviations: *Arabidopsis thaliana (At), Cleome gynandra (Cg), Zea mays (Zm),Brachypodium distachyon (Bd), Beta vulgaris (Bv), Clusia minor (Cm); Glycine max (Gm), Kalanchoe daigremontiana (Kd), Kalanchoe fedtschenkoi (Kf), Kalanchoe pinnata (Kp), Lotus japonicus (Lj), Lycopersicon esculentum (Le), Mesembryanthemum crystallinum (Mc), Oryza sativa (Os), Setaria italica (Si), Solanum tuberosum (St), Sorghum bicolor (Sb),* The accession numbers for the sequences used for this are: At1g08650.1 (*PPCK*1), At3g04530.1 (*PPCK*2); NP_001105808 (*ZmPPCK1*), NP_001105772 (*ZmPPCK2*), NP_001105773 (*ZmPPCK3*), NP_001105774 (*ZmPPCK4*), ABD39238.1 (*SbPPCK1*), AAK81871.1 (*SbPPCK2*), Sb06G022690 (SbPPCK3), AC069158 (*OsPPCK1*), BAE80213.1 (*OsPPCK2*), BAE80227.1 (*OsPPCK3*), AAN12511.1 (*GmPPCK1*), AAN12513.1 (*GmPPCK2*), AAQ82624.1 (*GmPPCK3*), AAS75449.1 (*GmPPCK4*), CAC43293.1 (*BePPCK*) , AAF06970.1 (*KfPPCK*), ABO65214 (*KdPPCK*), AAF19403.1 (*LePPCK1*), AAO32075.1 (*LePPCK2*), AAF05112.1 (*McPPCK*), BAC20363 (*LjPPCK*), ABO65213 (*KpPPCK*), AAR31830 (*CmPPCK1*), AAR31831 (*CmPPCK2*), AAO61489 (*StPPCK*), XP_003575394 (*BdPPCK*), XP_003580142 (*BdPPCK1-like*), XP_003572921 (*BdPPCK2-like*), XP_004976235 (*SiPPCK3*), XP_004954289 (*SiPPCK1*), XP_004953151 (SiPPCK2). The sequence for *Cleome gynandra (Cg)* is at the end of supplemental file 1.

```
                            10        20        30        40        50        60        70        80        90       100       110       120       130       140       150       160       170       180
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
FpPPCKA(C3)         ---------------MKEALNND----YHITEEIGRGRFGTVYRCYSTVSGESFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEDENYVHMIIDLCDSP-DLFDRISNRNGV-FSETEAATVFTPLMSAISYCHRLGIAHRDIKPDNVLFDSR---
FroPPCKA(C3)        ---------------MKEALNND----YHITDEIGRGRFGTVYRCYSTVSGDSFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEEENYVHMIIDLCDSP-DLFDRISNRNGV-FSEAEASTVFTPLMSAISYCHRLGIAHRDIKPDNVLFDSR---
FcPPCKA(C3-C4)      ---------------MKEALNND----YHITEEIGKGRFGTVYRCYSTVSGDSFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLFEDDNYVHMIIDLCDSP-DLFDRISNRNGV-FSETDAATVFTPLMSAISYCHRLGIAHRDIKPDNVLFDSR---
FaPPCKA(C3-C4)      ---------------MKEALNND----YHITEEIGRGRFGTVYRCYSTVSGDSFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEDDNYVHMIIDLCDSP-DLFDRISNRNGV-FSETEAATVFTPLMSAISYCHRLGIAHRDIKPDNVLFDSR---
FpuPPCKA(C3-C4)     ---------------MKEALNND----YHITEEIGRGRFGTVYRCYSTVSGDSFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEDDNYVHMIIDLCDSP-DLFDRISNRNGV-FSETEAATVFTPLMSAISYCHSLGIAHRDIKPDNVLFDSR---
FraPPCKA(C3-C4)     ---------------MKEALNND----YHITEEIGRGRFGTVYRCYSTVSGDSFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEDDNYVHMIIDLCDSP-DLFDRISNRNGV-FSETEAATVFTPLMSAISYCHRLGIAHRDIKPDNVLFDSR---
FbrPPCKA(C4-like)   ---------------MKEALNND----YHITEEIGRGRFGTVYRCYSTVSGDSFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEDDNYVHMIIDLCNSP-DLFYRISNRNGV-FSETEAATVFTPLMSAISYCHRLGIAHRDIKPDNVLFDSR---
FbPPCKA(C4)         ---------------MKETLNND----YHITEEIGRGRFGTVYRCYSTVSGDSFACKSIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEDDNYVHMIIDLCDSP-DLFDRISNRNGV-FSETEAATVFTPLMSAISYCHRLGIAHRDLKPDNVLFDSR---
FtPPCKA(C4)         ---------------MKETLNND----YHITEEIGRGRFGTVYRCYSTVSGDSFACKTIDKRLL-----NDH-TDRECLQKEPKILHILG-----GNPNIVQIYRLYEDDNYVHMIIDLCDSP-DLFDRISNRNGV-FSETEAATVFAPLMSAISYCHRLGIAHRDLKPDNVLFDSR---
FpPPCKB(C3)         -----------------------------------------VYRCYSTVSGDTYACKSIDKRLL-----TDH-TDRECLQKEPKILHILG-----GNPNIVQIHRLYEDENYVHMVIDLCESP-DLFDRISSRAGV-FSESESASVFGPLMSAISYCHRLGIAHRDIKPDNVLFDCR---
FroPPCKB(C3)        ---------------MKEVIHND----YQLTDEIGRGRFGTVYRCYSVVSGDTFACKSIDKRLL-----TDH-TDRECLQKEPKILHILG-----GNPNIVQIHRLYEDENYVHMVIDLCESP-DLFDRISTRAGV-FSESESASVFAPLMMAISYCHRLGIAHRDIKPDNVLFDCR---
FcPPCKB(C3-C4)      -----------------------------------------VYRCYSTVSGDTYACKSIDKRLL-----TDH-TDRECLQKEPKILHILG-----GNPNIVQIHRLYEDENYVHMVIDLCESP-DLFDRISSRAGV-FSESESASVFGPLMSAISYCHRLGIAHRDIKPDNVLFDCR---
FpuPPCKB(C3-C4)     --------------------------------------FGIVYRCYSTVSGDTYACKSIDKRLL-----TDH-TDRECLQKEPKILHILG-----GNPNIVQIHRLYEDENYVHMVIDLCESP-DLFDRITSRAGA-FSESESASVFGPLMSAISYCHRLGIAHRDIKPDNVLFDCR---
Fra_PPCKB(C3-C4)    ------------------------------------------YRCYSVVSGDTYACKSIDKRLL-----TDH-TDRECLQKEPKILHILG-----GNPNIVQIHRLYEDENYVHMVMDLCESP-DLFDRISSRAGV-FSEAESASVFGPLMSAISYCHRLGIAHRDIKPDNVLFDCR---
FbrPPCKB(C4-like)   ---------------MKEVIHND----YQLTEEIGRGRFGIVYRCYSTVSGDTYACKSIDKRLL-----TDH-TDRECLQKEPKILHILG-----GNPNIAQIHRLYEDENYVHMVIDLCESP-DLFDRISSRAGV-FSESESASVFGPLMSAISYCHRSGIAHRDIKPDNVLFDCR---
FbPPCKB(C4)         ---------------MKEVIHND----YQLTDEIGRGRFGIVYRCYSVVSGDTYACKSIDKRLL-----TDH-IDRECLQKEPKILHILG-----GNPNIVQIHRLYEDENYVHMVMDLCESP-DLFDRISSRAGV-FSEAESASVFGPLMSAIGYCHRLGIAHRDIKPDNVLFDCR---
FtPPCKB(C4)         -----------------------------------------VYRCYSVVSGDTYACKSIDKRLL-----TDH-IDRECLQKEPKILHILG-----GNPNIVQIHRLYEDENYVHMVMDLCESP-DLFDRISSRAGV-FSEAESASVFGPLMSAIGYCHQLGIAHRDIKPDNVLFDCR---

AtPPCK1(C3)         -------------MTCSQTLGNNNTNKYQICEEIGRGRFGTVSRVYAPATGDFFACKTIDKASL-----SDD-LDRACLDNEPKLMALLS-----YHPNIVQIHDLIDTDSTLSIFMELVHPSVSIYDRLVS-SGT-FFEPQTASFAKQILQALSHCHRYGVVHRDIKPENILVDLRN--
Cgy_PPCK            --------KKKKVMSYTQTLHNS-SNKYQFCEEIGRGRFGTVTRVYAPSTGDFFACKTIDKASL-----ADDALDRACIDAEPKIMALLS-----PHPNIVKIHDLIETNDSLSIVMELIDPAVTIYDRLVSSGP--LSEAQTAAYAKQILSAVSHSHRHGVVHRDIKPENILVDLRK--
AtPPCK2(C3)         -------------MTREFELENN----YQLCDEIGRGRFGTITRCFSPATKEFYACKTIDKRVL-----IDA-LDRECIETEPRIMAMLP-----PHPNIIRIFDLYETEDSLAIVMELVDPPMTIYDRLISAGGR-LSESESASYAKQILSALAHCHRCDVVHRDVKPDNVLVDLVS--
OsPPCK1(C3)         -----------MSSGGELALRQE----YVIGDEIGRGRFGTVRRCYAVATGEAFAVKSTPKAPLREAEAADA-LDLALAEQEPKVHLVASAPGPGASPHVVALHAAFEDDAAVHLVVDLCAGG-DLLSLVSSRGR--LPEHEAADLVAQLASALASCHRRGVAHRDVKPDNLLFDGG---
ZmPPCK1(C4)         -----------MGGAAEEALRRE----YVVGEEIGRGRFGTVRRCYAAATGEPFALKTTPNAALR----ADP-LDLALAEQEPKVHHLASA----ASRHVVALHAAFEDADAVHLVLDLCAGG-DLFALVSARGA--LPEPEAAELMAQLADALAGCHRRGVAHRDVKPDNLLFDAAA--
SbPPCK1(C4)         -----------MSGAAEEALLRE----YVIGEEVGRGRFGTVRRCYAASTGEPFALKTTPKAALRELE-ADP-LDLALAEQEPKVHLLASA----ASRHVVALHAAFEDADAVHLVLDLCAGG-DLFALVSARGP--LPEPEAAGLVAQLADALAGCHRRGVAHRDVKPDNLLFDAS---
SiPPCK1(C4)         ------------MSGGEEALRRE----YVIGDEIGRGRFGTVRRCYAAATGAPFALKSTPKEPLRELE-ADP-LDLALAEQEPKVHLLVSAP-PPASRHVVALHAAFEDADAVHLVLDLCEGG-DLFSLVSARGP--LPEPEAADLAAQLADALAGCHRRGVAHRDVKPDNLFFDAG---
OsPPCK3(C3)         ---------------MSEELKRD----FEIGEEIGRGRFGVVHRCASRSTGEAYAVKSVDRSRL-----SDD-LDRSLAALEPKLARLAAA----GNPGVVQVHAVYEDDDWTHTVMDLCTGP-DLLDWVRLRCGQPVPEPDAAAVVAQIAEALALCHRRGVAHRDVKPDNVLLDAT---
BdPPCK1-like(C3)    ---------------MSEELKRD----YEIGEEIGRGRFGVVHRCASRSTGDLYAVKSVDRSRL-----SDD-LDRGLAELEPKLAQLAAA----GNPGVVQVHAVYEDESWTHMVMDLCEGP-DLLDWVRLRQGAPVPEPEAAAVALQIAEALAICHRRGVAHRDVKPDNVLLDSGY--
BdPPCK2-like(C3)    ---------------MSEELKRD----YEIGEEIGRGRFGVVHRCASRSTGDLYAVKSVDRSRL-----SDD-LDRGLAELEPKLAQLAAA----GNPGVVQVHAVYEDESWTHMVMDLCEGP-DLLDWVRLRQGAPVPEPEAAAVALQIAEALAICHRRGVAHRDVKPDNVLLDSGY--
OsPPCK2(C3)         ---------------MSGELKRD----YEIGEEIGRGRFGVVHRCTSRATGEAFAVKSVDRSQL-----ADD-LDRELAELEPKLAQLAAA----GNPGVVQVHAVYEDDAWTHMVMDLCSGP-DLLDWIRLRRGAPVPEPVAAAVVAQLAEALAHCHRRGVAHRDVKPDNILLDVV---
BdPPCK(C3)          ---------------MSGELNRD----YEICEEIGRGRFGVVHRCTSRSTGEAYAVKSVDRSSL-----ADG-LDRELAELEPKLAQLAGQ----GNPGVVQVHAVYEDEAWTHVVMDLCSGP-DLLDWVGLRRGAPVPEPVAADIVAQVAQALALCHRRGVAHRDVKPDNILLDVA---
ZmPPCK4(C4)         ---------------MTEQLNRE----YDIGDEIGRGRFGVVRRCVSRATGDAFAVKSVDRSLL-----GDD-LDRALAELEPKLARLAGGE-GDGNPGVLQVRAVYEGDAWTHTVMDLCTGP-DLLDWIRLRRGAPVPEPDAARVVAQLAEALAGCHRRGVVHRDVKPDNVLLDVDPA-
SbPPCK3(C4)         ------------------MSKRD----YG-----GRGR-GVVHRCVSRATG---AVKSVDRS-------ADD--DRAAKARAG------G-----GDNGVV-VRAVY-DDAWTHTVMD-CTGD-D---WVR-RRGA----VDAARVVA----AAAGCHRRGVVHRDVK-DNV--DDA---
SiPPCK3(C4)         MELSRNQRAAKRNGSMSEQLKRD----YEIGEEIGRGRFGVVHRCVSRATGEPFAVKSVDRSAL-----ADD-LDRALAELEPKLAQLAGA--GGANPGVVQVRAVYEDDAWTHTVMDLCTGP-DLLDWVRLRRGAPVPEPDAARVVAQLAEALAGCHRAGVVHRDVKPDNVLLDLD---
SbPPCK2(C4)         ---------------MSAELKRD----YEIGAEIGRGRFGVVHRCTSRATGEAFAVKSVDRSQL-----ADD-LDRELAELEPKLAQLAGA----GNPGVVQTHAVYEDETWTHTVMDLCAGP-DLLEWVRLRRGAPVPEPLAAAIVAQVAQALALCHRRGVAHRDVKPDNILIDTAP--
ZmPPCK3(C4)         ---------------MSAELKRD----YEIGAEIGRGRFGVVHRCTSRATGEAFAVKSVDRSRL-----GDD-LDRELAELEPKLAQVAGA----GNPGVVQVHAVYEDEAWTHVVMDLCAGP-DLLEWVRIRRGAPVPEPVAAAIVAQVAQALALCHRRGVAHRDVKPDNILIDAAAAA
ZmPPCK2(C4)         ---------------MSAELKRD----YEIGEEIGRGRFGVVHRCTSRATGEAFAVKSVDRSRL-----GDD-LDRELAQLEPKLAQLAGA----GNPGVVQVHAAYEDEAWTHVVINLCPGP-DLLEWVGRRRGAPVPEPVAAAIVAQVAQALALCHRRGVAHRDVKPENILIDAAA--
SiPPCK2(C4)         ---------------MSAELKRD----YEIGEELGRGRFGVVHRCTSRATGEAFAVKSVDRSQL-----ADD-LDRELAQVEPKLAQLAGA----GNPGVVQVHAVYEDEAWTHTVMDLCTGP-DLLGWMGLRRGAPVPEPVAAAIVAQLAQALALCHRRGVAHRDVKPDNILIDAA---
```

```
                        190       200       210       220       230       240       250       260       270       280       290       300       310       320       330       340       350
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
FpPPCKA(C3)         ----------------G---------SLELADFGSAEWFGMNE-----RGTMSGIVGTPYYVAPEVLSGREYNEKVDVWSAGVILYIMLA-GVPPFYGDSPAETFEAVLRGNLRFPTRLFRSVSPETKDLLRKMLCKDVSRRLSADQVLRHPWVVSGG-ETRSMADLT-
FroPPCKA(C3)        ----------------G---------SLKLADFGSAEWFGMNE-----RGTMSGIVGTPYYVAPEVLSGREYNEKVDVWSAGVILYIMLA-GVPPFYGDSPAETFEAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVSRRLSADQVLS------------------
FcPPCKA(C3-C4)      ----------------G---------SLKLADFGSAEWFAMNE-----RRTMSGIVGTPYYVAPEVLSGMEYNEKVDVWSAGVILYIMLA-GVPPFYGDSPAETFEAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVSRRFSADQVLRHPWVVSGG-ETRSMADLT-
FaPPCKA(C3-C4)      ----------------G---------SLKLADFGSAEWFAMND-----RRTMSGIVGTPYYVAPEVLSGMEYNEKVDVWSAGVILYIMLA-GVPPFYGDSPAETFDAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVSRRFSADQVLRHPWVVSGG-ETRSMADLT-
FpuPPCKA(C3-C4)     ----------------G---------SLKLADFGSAEWFAMNE-----RRTMSGIVGTPYYVAPEVLSGMEYNEKVDVWSAGVILYIMLA-GVPPFYGDSPAETFEAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVSRRFSADQVLRHPWVVSGG-ETRSMADLT-
FraPPCKA(C3-C4)     ----------------G---------SLKLADFGSAEWFAMND-----RRTMSGIVGTPYYVAPEVLSGMEYNEKVDVWSAGVILYIMLA-GVPPFYGDSPAETFEAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVSRRFSADQVLRHPWVVNGG-ETRSMADLT-
FbrPPCKA(C4-like)   ----------------G---------SLKLADFGSAEWFAMNE-----RRTMSGIVGTPYYVAPEVLSGMEYNEKVDVWSAGVILYIMLA-GVPPFYGDSPAETFEAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVSRRFSADQVLRHPWVVSGG-ETRSMADLT-
FbPPCKA(C4)         ----------------G---------GLKLADFGSAEWFAMND-----RRTMSGIVGTPYYVAPEVLSGMEYNEKVDVWSAGVILYIMLA-GVPPFHGDSPADTFEAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVYRRFSADQVLRHPWVVNGG-KTRSMADLT-
FtPPCKA(C4)         ----------------G---------GLKLADFGSAEWFAMND-----RRTMSGIVGTPYYVAPEVLSGMEYNEKVDVWSAGVILYIMLA-GVPPFHGDSPADTFEAVLRGNLRFPTRLFRSVSPEAKDLLRKMLCKDVSRRFSADQVLRHPWVVNGG-KTRSMADLT-
FpPPCKB(C3)         ----------------G---------SLKLADFGSAEWFAMSE-----RGTMTGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYVMLA-GVPPFYGESPADTFEAVLRGNLRFPTKIFRSVSPEAKDLLKKMLCKDVSRRLSAEQVLRHPWVVSGG-ETRAMADLT-
FroPPCKB(C3)        ----------------G---------NLKLADFGSAEWFAMSE-----RGTMTGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYIMLA-GVPPFYGESPADTFEAVLRGNLRFPTKIFRSVSLEAKDLLKKMLCKDVSRRLSAEQVLS------------------
FcPPCKB(C3-C4)      ----------------G---------SLKLADFGSAEWFAMSE-----RGTMNGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYVMLA-GVPPFYGESPADTFEAVLRGNLRFPTKMFRSISPEAKDLLKKMLCKDVSRRLSAEQVLRHPWVVSGG-ETRAMADLT-
FpuPPCKB(C3-C4)     ----------------G---------SLKLADFGSAEWFAMSE-----RGTMTGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYVMLA-GVPPFYGESPADTFEAVLRGNLRFPTKMFRSVSPEAKDLLKKMLCKDVSRRLSAEQVLRHPWVVSGG-ET*AMADLT-
Fra_PPCKB(C3-C4)    ----------------G---------SLKLADFGSAEWFAMSE-----RGTMTGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYVMLA-GVPPFYGESPADTFEAVLRGNLRFPTKIFRSVSPEAKDLLKKMLCKDVSRRLSAEQVLRHPWVVSGG-ETRAMVDLT-
FbrPPCKB(C4-like)   ----------------G---------SLKLADFGSAEWFAMSE-----RGTMTGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYVMLA-GVPPFYGESPADTFEAVLRGNLRFPTKMFRSVSSEAKDLLKKMLCKDVSRRLSAEQVLRHPWVVSGG-ETRAMADLT-
FbPPCKB(C4)         ----------------G---------SLKLADFGSAEWFGMSE-----RGTMTGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYVMLA-GVPPFYGESPADTFEAVLRGNLRFPTKMFRSVSPEAKDLLKKMLCKDVSRRLSAEQVL-------------------
FtPPCKB(C4)         ----------------G---------SLKLADFGSAEWFAMSE-----WGTMTGVVGTPYYVAPEVLLGREYNEKVDVWSAGVILYVMLA-GVPPFYGESPADTFEAVLRGNLRFPTKMFRSVSPEAKDLLKKMLCKDVSRRLSAEQVLRHPWVVSGG-ETRAMADLT-

AtPPCK1(C3)         ----------------D---------TVKICDFGSGIWLGEGE-------TTEGVVGTPYYVAPEVLMGYSYGEKVDLWSAGVVLYTMLA-GTPPFYGETAEEIFEAVLRGNLRFPTKIFRGVSSMAKDFLRKLICKDASRRFSAEQALRHPWIQRAG---ETEERFI-
Cgy_PPCK            ----------------D---------VVKICDFGSGIWLGEGE-------TTVGVVGTPYYVAPEVLTGRPYAEKVDVWSVGVVLYTMLA-GTPPFYGETAEEIFDAVIRGNLRFPPKIFRGVSSAAKDLLRKMICRDVSRRFSAEQSLRHPWIVNAA-ATEVEESFI-
AtPPCK2(C3)         ----------------G---------GVKLCDFGSAVWLG-GE-------TAEGVVGTPYYVAPEVVMGRKYDEKVDIWSAGVVIYTMLA-GEPPFNGETAEDIFESILRGNLRFPPKKFGSVSSEAKDLLRKMICRDVSRRFSAEDALRHSWMMNVG-NLQSN-----
OsPPCK1(C3)         ----------------G---------VLKLGDFGSAGWFGDG-------RPMTGLVGTPYYVAPEVVAGREYGEKVDVWSAGVVLYMMLS-GTLPFYGATAAEVFQCVLRGNLRFPPRAFASVSPEAKDLMRRMLCKDVSRRFSADQVLRHPWIVSRG-GAAVMG----
ZmPPCK1(C4)         ---------------GG---------ALRLGDFGSAAWFGDG-------RPMTGLAGTPYYVAPEVVAGREYTEKVDVWSAGVVLYMMLS-GTVPFSGATAGEIFQAVLRGNLRFPPRAFAAVSPEAKDLLRRMLCRDVWRRLSAEQVLRHPWIVTRG-GSVPGGRLV-
SbPPCK1(C4)         ----------------G---------ALRLGDFGSADWFGDG-------RPLTGLVGTPYYVAPEVVAGREYTEKVDVWSAGVVLYMMLS-GTVPFSGATAGDVFQAVLRGNLRFPPRAFPAVSPEAKDLMRRMLCKDVWRRLSAEQVLSHPWIVTRG-GSVAVA----
SiPPCK1(C4)         ----------------G---------ALRLGDFGSAGWFGDG-------RPMTGLVGTPYYVAPEVVAGREYTEKVDVWSAGVVLYVMLS-GTVPFGGATAGEIFEAVLRGNLRFPPRAFAGVSPEAKDLMRRMLCKDVSRRFSAEQVLRHPWIVTRG-GSAA------
OsPPCK3(C3)         ----------------GDG-----PPRVRLADFGSAAWVGDG-------ISAEGLVGTPHYVAPEVVAGGEYGEKADVWSAGVVMYVLLTGGALPFGGETASDVFAAVLRGNLRFPPRLFSGVSPAAKDLMRRMMCRDVYRRFSAEQVLRHPWIVSGG-GARDV-QPT-
BdPPCK1-like(C3)    -------------E--G---GSPAPLRARLADFGSAAWVGGG-------GSAQGLVGTPHYVAPEVVAGGEYGEKADVWSAGVVLYVLLSGGALPFSGETASDVFAAVLRGGPRFSPRLFAGVSPLAKDLMRRMMCRDVSRRFSAEQVLGHPWILNGG-GAREAVQPT-
BdPPCK2-like(C3)    -------------EG-G----SPAPLRARLADFGSAAWVGGG-------GSAQGLVGTPHYVAPEVVAGGEYGEKADVWSAGVVLYVLLSGGALPFSGETASDVFAAVLRGGPRFSPRLFAGVSPLAKDLMRRMMCRDVSRRFSAEQVLGHPWILNGG-GAREAVQPT-
OsPPCK2(C3)         ----------------DDGIDGGGTPRARLADFGSAAWVGESG------GSAEGLVGTPHYVAPEVVAGGEYGEKADVWSAGVVLYVLLSGGALPFGGETAAEVLASVLRGSVRFPPRLFAGVSPAAKDLMRRMMCRDTWRRFSAEQVLGHPWIVSGG-GARAMEQPT-
BdPPCK(C3)          ---------------EDG---EGEGSPRARLADFGSAAWVGGAE-----GGRAEGLVGTPHYVAPEVVSGGEYGEKADVWSAGVVMYVLLSGGALPFGGETAKEVLSAVMRGSVRFPPRLFSGVSPAAKDLMRRMICRDEWRRFSAEQVLRHPWILSGG-GSRAMEQPT-
ZmPPCK4(C4)         ----------------G-----EAPPRVRLADFGSAAWLGEGA-----RGRAEGLVGTPHDVAPEVVAGGEYGEKVDVWSAGVVMYALLSGGALPFGGETAADVFAAVLRGSLRFPPALFAGVSPAAKDLMRRMMCRDVSRRFSAEQVLRHPWIQSGG-GGRELVRPT-
SbPPCK3(C4)         ----------------G---------TRVRAD-GSAAWGGGGG-----GRRRGAGVGT-HYVA--VVAGG-YG-KVDVWSAGVVMYA---------SGGAGGTADVAAVRGSRA------AGVS-AAKD-MRRMMCRDVSRR----SAVRH-W--SGG-GVR-VVRTT-
SiPPCK3(C4)         ----------------DDPAGEGAPPRVRLADFGSAAWLGGGGGEGARRGSVEGLVGTPHYVAPEVVAGGEYGEKVDVWSAGVVAYVLLSGGALPFGGETPAEVFAAVLRGSLRFPPALFAGVSPAAKDLMRRMMCRDASRRFSAEQVLRHPWIQSGG-GAREVVQPT-
SbPPCK2(C4)         ---AGSDDGEDEEEDSGEAEEAETAPRARLADFGSAAWVGAGG-----LGRAEGLVGTPHYVAPEVVGGGEYGAKADVWSAGVVMYALLSGGALPFGGETAAEVLAAVLRGSVRFPPRLFSGVSPAAKDLLRRMICRDEWRRFTAEQVLAHPWIVSGG-GARAIERPT-
ZmPPCK3(C4)         AGTDGDGDGEDEEEEGGDG-EAEAAPRARLADFGSAAWVGAGG-----LGRAEGLVGTPHYVAPEVVGGGEYGAKADVWSAGVVMYALLSGGALPFGGESAAEVLAAVLRGSARFPPRLFGGVSPAAKDLMRRMICRDEWRRFTAEQVLAHPWIVSGG-GARGMERPT-
ZmPPCK2(C4)         ------GSDDDEED------EAEAAPRARLADFGSAAWVGAGG-----LGRAEGLVGTPHYVAPEVVAGGEYGAKADVWSAGVVMYALLSGGAFPFGGDSAAEVLAAVLRGSARFPPRLFGGVSPAAKDLMRRMICRDEWRRFTAEQVLAHPWIVSGG-GARAMERPT-
SiPPCK2(C4)         ------EEDDEEDGAEGEG-EGEAAPRARLADFGSAAWVGVGG-----LGRAEGLVGTPHYVAPEVVSGGEYGAKADVWSAGVVMYTLLSGGALPFGGESGAEVLAAVMRGSVRFPPRLFGGVSPAAKDLMRRMICRDEWRRFTAEQVLAHPWIVSGG-GARAVGQ---
```

S3: Supplemental file 3- Alignments of PPCK proteins from various C_3 and C_4 species. The amino acid residues which changed in the evolution of C_4 *Flaveria* species are specific for this genus. Full length protein sequences from the PPCK protein family from the genus *Flaveria,* genus *Poaceae* and order *Brassicales* were aligned using Muscle (Edgar, 2004), as part of MEGA5 (Tamura *et al*., 2011; Hall, 2013) using default settings for 1000 iterations. The C_4 specific amino acid exchanges, characteristic for the genus *Flaveria* (Table 1) are highlighted in grey: Light grey- amino acid exchanges found in some intermediate species and species with C_4, medium grey –amino acid exchanges found in all intermediate species and species with C_4, dark grey- amino acid exchanges found only in species with C_4 photosynthesis isoforms. Species abbreviations: *Brassicales- Arabidopsis thaliana (At),* Cleome gynandra (Cg); *Poaceae - Zea mays (Zm), Oryza sativa (Os), Setaria italica (Si), Sorghum bicolor (Sb)*. The accession numbers for the sequences used for this are listed in supplemental file 2, except Cg. The sequence for Cg is included at the end of supplemental file 1.

S4: Supplemental file 4- Table with RNA_Seq and SRM data. The significant differences between each time point were determined using a one-way ANOVA (Tukey-test) and the student's t-test. NS (no asterisk) indicates $p > 0.05$; * indicates $p \leq 0.05$; ** indicates $p \leq 0.01$; *** indicates $p \leq 0.001$; **** indicates $p \leq 0.0001$. Red asterisk refers to multiplicity adjusted p-values from the ANOVA test, and black asterisk refers to t-test p-values.

Raw Data

RnaSeq Data	Description	Species	Unit	Diurnal time (h) = 0			Diurnal time (h) = 4h			Diurnal time (h) = 8h		
				1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
PPCKA	phosphoenolpyruvate carboxylase prot	*F. trinervia*	RPKM	77.9545344	121.05079	130.87247	300.157965	363.465908	246.348874	403.129938	87.2318105	162.797999
PPCKA	phosphoenolpyruvate carboxylase prot	*F. pringlei*	RPKM	4.3791614	22.0267644	5.30691908	11.78493	36.04012	46.26201	11.3009085	14.3402857	11.5778941
PPCKB	phosphoenolpyruvate carboxylase prot	*F. trinervia*	RPKM	80.5940471	4.36972869	18.3175362	1.47584609	2.3393259	1.8691655	2.01146331	5.67472153	2.87520402
PPCKB	phosphoenolpyruvate carboxylase prot	*F. pringlei*	RPKM	2.47930651	5.05094371	2.09046438	5.67588752	9.96715033	4.90347916	21.8099862	36.7134587	31.3876374
SRM data												
Ratio of phosphorylated PEPC	ratio of phosphorylated phosphoenolpy	*F. trinervia*	ratio	0.04699097	0.06885275	0.04207216	0.17646747	0.21252519	0.22875307	0.23312852	0.08464869	0.22225135
Ratio of phosphorylated PEPC	ratio of phosphorylated phosphoenolpy	*F. pringlei*	ratio	0.05020625	0.07441439	0	0.21960672	0.10449767	0.00957231	0.1704318	0.15712486	0.17414845

Comparison of Time Points within an experiment

RnaSeq	Species	Unit	ANOVA	0h/ 4h	0h/ 8h	0h/ 12h	0h/ 16h	0h/ 20h	0h/ 24h	4h/ 8h	4h/ 12h	4h/ 16h
PPCKA	*F. trinervia*	RPKM	0.0027	0.38*	0.5	1	8	4*	2	1	4*/*	22**
PPCKA	*F. pringlei*	RPKM	0.0027	0.34*	0.65	4	8	7	21.5	2.5	11**	24**
PPCKB	*F. trinervia*	RPKM	0.0018	5	3	1	0.4	0.5	0.86	0.5	0.18*	0.07**
PPCKB	*F. pringlei*	RPKM	0.0003	0.34	0.08*	0.04**/*	0.04**	0.05*	2	0.23*	0.122*/*	0.12*
SRM												
Ratio of phosphorylated PEPC	*F. trinervia*	ratio	0.0003	0.26**/*	0.3**	0.7	5	5	1	1	2.5 */*	12**/**
Ratio of phosphorylated PEPC	*F. pringlei*	ratio	0.002	0.3	0.25*	0.2	0.17 *	0.2 */**		0.86	0.71*	0.63

Raw Data

RnaSeq Data	Description	Species	Unit	Diurnal time (h) = 12h			Diurnal time (h) = 16h			Diurnal time (h) = 20h			Diurnal time (h) = 24h		
				4-1	4-2	4-3	5-1	5-2	5-3	6-1	6-2	6-3	7-1	7-2	7-3
PPCKA	phosphoenolpyruvate carboxylase prot	*F. trinervia*	RPKM	110.385986	47.9763401	83.4081675	12.4734675		15.4555003	31.5182441	19.2277815	37.8389785	79.2637181	65.3021769	42.5343611
PPCKA	phosphoenolpyruvate carboxylase prot	*F. pringlei*	RPKM	4.36035183	1.97945645	2.10495036	1.59911909	1.07918733	1.20443978	1.70762791	1.27658094	1.43450592	0.51208983	0.37079035	0.58980885
PPCKB	phosphoenolpyruvate carboxylase prot	*F. trinervia*	RPKM	12.0912582	8.88506845	7.75899962	27.7826119		22.8734924	16.9470183	26.8682725	19.3281669	21.1298686	10.015175	3.03842366
PPCKB	phosphoenolpyruvate carboxylase prot	*F. pringlei*	RPKM	43.7086871	49.2209604	76.9338016	59.859206	80.0275167	32.5369343	80.7416835	81.327458	56.7435657	1.60318604	1.76727697	1.61614525
SRM data															
Ratio of phosphorylated PEPC	ratio of phosphorylated phosphoenolpy	*F. trinervia*	ratio	0.11158563	0.05820457	0.07993601	0.02205101	0.01669743	0.0151631	0.02105638	0.01490288	0.01884701	0.06891973	0.03019275	0.05634333
Ratio of phosphorylated PEPC	ratio of phosphorylated phosphoenolpy	*F. pringlei*	ratio	0.18650847	0.15876279	0.18669309	0.20252756	0.35740625	0.11736487	0.24266642	0.21955933	0.15281716	0	0	0

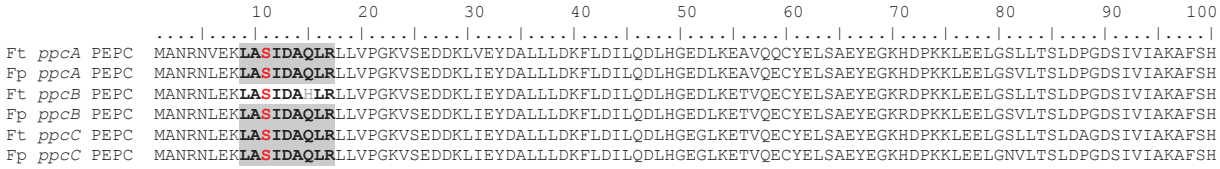
Comparison of Time Points within an experiment

RnaSeq	Species	Unit	ANOVA	4h/ 20h	4h/ 24h	8h/ 12h	8h/ 16h	8h/ 20h	8h/ 24h	12h/ 16h	12h/ 20h	12h/ 24h	16h/ 20h	16h/ 24h	20h/ 24h
PPCKA	*F. trinervia*	RPKM	0.0027	10**/*	5*/*	3	16	7	3.5	6	3	1	0.5	0.2	0.5
PPCKA	*F. pringlei*	RPKM	0.0027	21**	84**	4*	10**	8**	25**	2	2	6	0.88	3*	3
PPCKB	*F. trinervia*	RPKM	0.0018	0.09**/*	0.16	0.37	0.14**	0.17*/*	0.31	0.34*	0.5	0.84	1	2	2
PPCKB	*F. pringlei*	RPKM	0.0003	0.13*	4	0.53	0.52	0.56	19*	1	1	35**/*	1	36**	33*
SRM															
Ratio of phosphorylated PEPC	*F. trinervia*	ratio	0.0003	11.5 **/**	4**/*	2	10.5 **	10**	3.5 *	5*	5*	2*	0.99	0.34	0.35
Ratio of phosphorylated PEPC	*F. pringlei*	ratio	0.002	0.7		0.9	0.7	0.8		0.8	0.9		1		

Comparison of time points between species or PPCK isoforms

Comparison	Parameter	Unit	0h (Ft)/ 0h (Fp)	4h (Ft)/ 4h (Fp)	8h (Ft) /8h (Fp)	12h (Ft)/ 12h (Fp)	16h (Ft)/ 16h (Fp)	20h (Ft)/ 20h (Fp)
F. tinervia with *F. pringlei*	PPCKA	RPKM	10*	10****/*	18**	29*	11	20*
			0h (Fp)/ 0h (Ft)	**4h (Fp)/ 4h (Ft)**	**8h (Fp) /8h (Ft)**	**12h (Fp)/ 12h (Ft)**	**16h (Fp)/ 16h (Ft)**	**20h (Fp)/ 20h (Ft)**
F. pringlei with *F. tinervia*	PPCKB	RPKM	0.3	4	8*	6**	2	3
			4h (PPCKA)/ 4h (PPCKB)	**4h (PPCKA)/16h (PPCKB)**	**16h (PPCKA)/ 16h (PPCKB)**	**16h (PPCKA)/ 8h (PPCKB)**		
PPCKA/PPCKB	*F. tinervia*	RPKM	166****/*	13****	0.6	4		
			4h (PPCKB)/ 4h (PPCKA)	**4h (PPCKB)/ 16h (PPCKA)**	**16h (PPCKB)/ 4h (PPCKA)**	**16h (PPCKB)/ 16h (PPCKA)**		
PPCKB/PPCKA	*F. pringlei*	RPKM	0.22	5.4	2	44***		
			0h (Fp)/ 0h (Ft)	**4h (Fp)/ 4h (Ft)**	**12h (Fp) / 12h (Ft)**	**16h (Fp)/ 16h (Ft)**	**20h (Fp)/ 20h (Ft)**	
F. pringlei with *F. tinervia*	Ratio of phosphorylated PEPC	ratio (SRM)	0.8	0.7	2*	13**/***	11**/*	

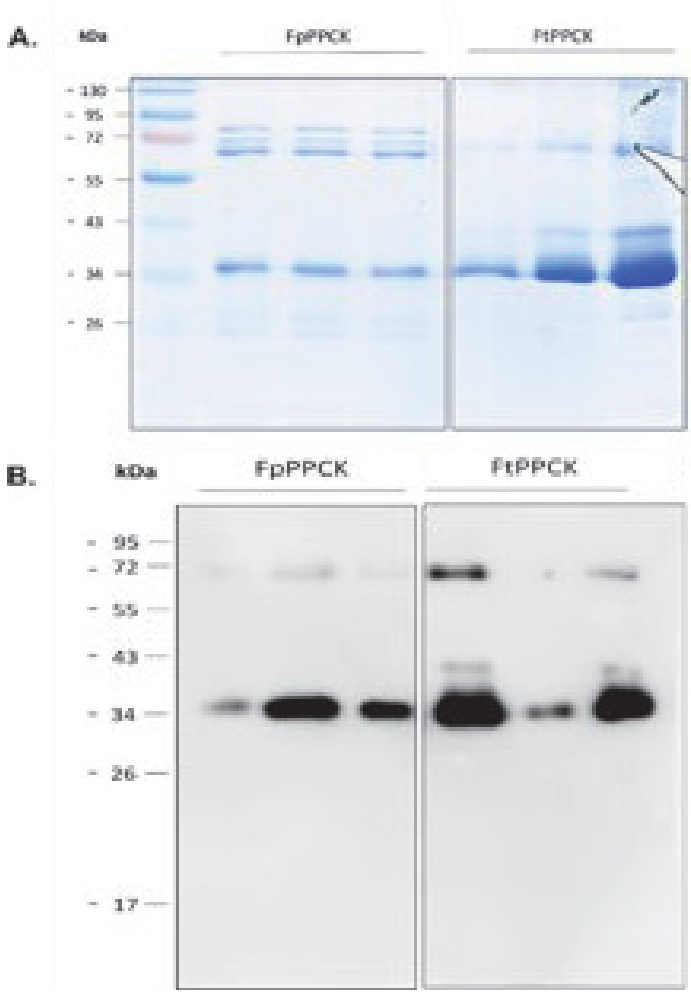
A



B

Peptide	CID (eV)	Retention time start-end (min)	Precursor m/z	Fragments (m/z)	Ion Type	Monitored transitions	
						F. tr	*F. pr*
LASIDAQLR	17.5904	20-25	493.785	487.2982	y4	x	x
				602.3251	y5	x	
				802.4412	y7	x	x
LASIDAQLR[HeavyR]	17.7153	20-25	498.789	497.3064	y4	x	
				612.3334	y5	x	x
				812.4495	y7	x	x
LASIDAQLR Phos	18.5888	25.6-27	533.768	416.261	y3	x	
				487.2982	y4	x	x
				602.3251	y5	x	x
LASIDAQLR[HeavyR] Phos	18.7137	25.6-27	538.772	426.2693	y3	x	
				497.3064	y4	x	x
				612.3334	y5	x	x

S5: Supplemental file 5- Multiple alignment of N-terminal PEPC isoforms, showing the specificity of LASIDAQLR SRM peptide (A); Characteristics of the LASIDAQLR and LA(pS)IDAQLR SRM peptides (B). Amino acid sequences of PEPC isoforms were analyzed with MUSCLE (Edgar, 2004a/b). Residues marked with bold and grey background indicate the peptide sequence used for SRM. Ser11 residues at the N-terminal in bold red indicate the phosphorylation site. *F. trinervia* (Ft); *F. pringlei* (Fp). In the table "X" marks that the peptide transition has been successfully monitored in three biological replicates in the time series for the respective species.



S6: Supplemental file 6- Recombinant PPCKA expression and anti-PPCK antibodies. Representative SDS-PAGE of purified recombinant PPCK protein (A) and western-blot with the custom anti-PPCK antibody (B). The adjustment of kinase amount for the *in-vitro* phosphorylation assay was performed by quantifying the kinase band intensity of the Coomasie-stained gel, and protein amount for FpPPCKA was adjusted to match the strongest signal of FtPPCKA.

S7: Supplemental file 7- List of oligonucleotides

List of oligonucleotides used for RACE-PCR

Name	Forward Primer	Reverse Primer	Method
Smart-II- A-5RACE	AAGCAGTGGTATCAACGCAGAGT		5´RACE
PEPC-PK-Fp_5´RACE		CGAAGAGATCCGGCGAGTCGC	
UPM		ACTCTGCGTTGATACCACTGCTTGCCCTATAGTGAGTCGTATTAG	3´RACE
PEPC-PK_Fp_ 3´RACE	GATCAAGTCCTCAGGCATCC		

List of oligonucleotides used for the amplification of PPCK coding regions from *F. trinervia* and *F. pringlei*

Name	Forward Primer	Reverse Primer	Method
PEPC-PK-CDS_Ft	CGGAAGTCCACCTACAATCG	CTACATGCAGACTTGCTCAA	PPCK coding sequence (incl. UTR regions)
PEPC-PK-CDS_Fp	CCAACACCACCAACAACCACC	ATCAAAAAACCAGTGTTGATTGAT	
PEPC-PK-Ft-5´-NdeI	TATCATATGAAGGAAACTCTGAACAAC		PPCK coding sequence incl. restrictions site
PEPC-PK-Ft-3´-BamHI		TATGGATCCTTAGGTCAGATCCGCCATTGATCG	
PEPC-PK-Fp-5´-XhoI	TATCTCGAGATGAAAGAAGCTCTGAACAAC		
PEPC-PK-Fp-3´-BamHI		TATGGATCCTTAGGTTAGATCCGCCATAGATCG	
pJET	ATCAACTGCTTTAACACTTGTGC	CGGTTCCTGATGAGGTGGTTAG	Sequencing
pET15b	TGGCAGCAGCCAACTCAGCTTCCTT	GCAGCAGCCATCATCATCATCATC	

Flaveria

```
Fp_ppcA_C3                    ----MANRNL EKLASIDAQL RLLVPGKVSE DDKLIEYDAL LLDKFLDILQ DLHG [54]
Fro_ppcA_C3                   ----..... .......... .......... .......... .......... .... [54]
Fc_ppcA_C3-C4                 ----..... .......... .......... .......... .......... .... [54]
Fpu_ppcA_C3-C4                ----...K.V .......... .......... .......... .......... .... [54]
Fa_ppcA_C3-C4                 ----...K.V .......... .......... .......... .......... .... [54]
Fra_ppcA_C3-C4                ----..... .......... .......... .......... .......... .... [54]
Fbr_ppcA_C4-like              ----...K.V .......... .......... .......... .......... .... [54]
Fb_ppcA_C4                    ----.....V .......... .......... ....V..... .......... .... [54]
Ft_ppcA_C4                    ----.....V .......... .......... ....V..... .......... .... [54]

Fp_ppcB_C3                    ----..... .......... .......... .......... .......... .... [54]
Fro_ppcB_C3                   ----..... .......... .......... .......... .......... .... [54]
Fc_ppcB_C3-C4                 ----..... .......... .......... .......... .......... .... [54]
Fpu_ppcB_C3-C4                ----..... .......... .......... .......... .......... .... [54]
Fa_ppcB_C3-C4                 ----..... .......... .......... .......... .......... .... [54]
Fra_ppcB_C3-C4                ----..... .......... .......... .......... .......... .... [54]
Fbr_ppcB_C4-like              ----..... .......... .......... .......... .......... .... [54]
Fb_ppcB_C4                    ----..... ........H. .......... .......... .......... .... [54]
Ft_ppcB_C4                    ----..... ........H. .......... .......... .......... .... [54]

Fp_ppcC_C3                    ----..... .......... .......... .......... .......... .... [54]
Fro_ppcC_C3                   ----..S... .......... .......... .......... .......... .... [54]
Fc_ppcC_C3-C4                 ----..... .......... .......... ........V. .......... .... [54]
Fpu_ppcC_C3-C4                ----..... .......... .......... ........V. .......... .... [54]
Fa_ppcC_C3-C4                 ----..... .......... .......... .......... .......... .... [54]
Fra_ppcC_C3-C4                ----..... .......... .......... .......... .......... .... [54]
Fbr_ppcC_C4-like              ----..... .......... .......... .......... .......... .... [54]
Fb_ppcC_C4                    ----..... .......... .......... .......... .......... .... [54]
Ft_ppcC_C4                    ----..... .......... .......... .......... .......... .... [54]
```

Brassicales

```
AtPEPC1_NP_175738_(C3)        ----MANRKL EKMASIDVHL RQLVPGKVSE DDKLVEYDAL LLDRFLDILQ DLHG [54]
AtPEPC3_NP_188112_(C3)        ----..G.NI .......AQ. .....A.... .......... .......... .... [54]
AtPEPC2_NP_850373_(C3)        ----..A.N. .......AQ. .L.A...... ....I..... .......... .... [54]
CgPEPC_(C4)                   ----..AKN. .......AQ. .L.A...... .......... .......... .... [54]
```

Amaranthaceae

```
Al.sPEPC_AAY28731_(C3)        ----MATVKL EKLTSIDAQL RLLAPRKVSE DDKLVEYDAL LLDRFLDILD SLHG [54]
Al.tPEPC_AAY28730_(C3-C4)     ----..... ...A...... .......... .......... .......... .... [54]
Al.pPEPC_AAY28729_(C4)        ----.....V ..MA...... .....G.... .......... ......ES.. .... [54]
Am.hPEPC_Q43299_(C4)          ----..SG.V ..MA...... .....K.... .......... .........E .... [54]
BsPEPC_ABG20459_(C4)*         ----..SG.. ..MA...... .A...G.... .......... .......S-Q A... [54]
```

Poaceae

PEPC isoforms with C_3 characteristics

```
OsPEPC1_NP_001046417_(C3)     MAAAAGKAAM ERHQSIDAQL RLLAPGKVSE DDKLVEYDAL LVDRFLDILQ DLHG [54]
BdPEPC1-like_XP_003571949_(C3) ...P.....I .......... .......... .......... .......... .... [54]
SiPEPC2_NP_001267742_(C4)     ---MT...P. .......... .......... .......... .......... .... [54]
SbPEPC1_P29195_(C4)           --------MP .......... .......... .......... .......... .... [54]
ZmPEPC_NP_001105503_(C4)      --------MP .......... .......... .......... .......... .... [54]
```

PEPC isoforms with C_4 characteristics

```
SiPEPC1-like_NP_001267758_(C4) ----MASKPV .K.H...... .......... .......... .I......F. .... [54]
CgaPEPC_AAG42288_(C4)         ----MATK.- .K.H...... .......... .......... .I........ N... [54]
ZmPEPC1_NP_001105418_(C4)     ..STKAPGPG .K.H...... .Q.V...... ....I..... ......N... .... [54]
SbPEPC3_P15804_(C4)           ----MAS--- ...H...... .A........ E--.IQ.... .......... .... [54]
SoPEPC1_CAC08829_(C4)         ----MAS--- ...H...... .A........ E--.IQ.... .......... .... [54]
SsPEPC1_CAC85930_(C4)         -------M.S ...H...... .A...D.... E--.IQ.... .A........ .... [54]
```

PEPC isoforms with C_3 characteristics

```
OsPEPC1_NP_001044301_(C3)     ----MARN.A DKGT...... .I...K.L.. .......... .L........ .... [54]
BdPEPC-like_XP_003569851_(C3) ----MARNV. DKAT...... .....N.L.. .......... .LE....... ..Q. [54]
OsPEPC_NP_001061646_(C3)      ----MAGK-V .KMA...... .M...A.L.. .......... .L........ .... [54]
BdPEPC2-like_XP_003577980_(C3) -M.SSAPGKI ..LS...... .M.V.A.... ....I..... .L.....V.. G... [54]
SiPEPC-likeX1_XP_004970018_(C4) -MRGMARN.V DKAT...... .....Q.L.. .......... .L........ .... [54]
SbPEPC2_P29194_(C4)_ppc2      ----------. ..LS...... .M.V...... ....I..... .L........ .... [54]
SiPEPC2-like_XP_004956559_(C4) --M..LG.KV ..LS...... .M.V...... ....I..... .L........ .... [54]
ZmPEPC2_NP_001105438_(C4)     --M..LGPK. ..LS...... .M.V...... ....I..... .L........ .... [54]
OsPEPC-likeB1_NP_001042358_(C3) --M.VAEGRR QHEY....R. .Q...E...A ..R..D.ET. ..A......E .... [54]
```

S8: Supplemental file 8- Supplemental File 8: N-terminal alignment of PEPC isoforms from Flaveria, Amarantheaceae, Brassicales, and Poaceae. Approximately 50 amino acids from the N-terminus of PEPC have been taken for this analysis and the sequences alligned using Muscle Muscle (Edgar, 2004) within the software MEGA5 (Tamura *et al*., 2011; Hall, 2013). The sequences are sub-grouped within their Phylogenetic clades and within each systematic group a C_3 PEPC sequence is taken as reference. Amino acid that are close to the phosphorylated Serine (marked in red, with yellow background) and that vary either between PEPC isoforms form same systematics group, or among C_3 and C_4 species from same systematics group have been marked in grey. Species abbreviations- *Amaranthaceae: Alternanthera sessilis* (Al.s)*, Alternanthera tenella* (Al.t)*, Alternanthera pungens* (Al.p)*, Amaranthus hypochondriacus* (Am.h)*, Bienertia sinuspersici* (Bs); *Brassicales: Arabidopsis thaliana (At), Cleome gynandra* (Cg)*, Poaceae: Zea mays* (Zm), *Brachypodium distachyon* (Bd), *Oryza sativa* (Os), *Setaria italica* (Si)*, Sorghum bicolor* (Sb), *Chloris_gayana* (Cga), *Saccharum offcinarum* (So), *Saccharum spontaneum* (Ss). For most species accession codes are included within the alignment names, whereas PEPC sequences for *Flaveria* and *Cleome* are listed in supplemental file 1. The phosphorylated Serine is in red and highlighted in yellow, the consensus motif from the N-terminal alignment is highlighted in blue, (*)- a single cell C_4 species.

AUTHOR CONTRIBUTIONS

S.H.A., **S.E.W.** and **T.D.S.** performed the circadian harvest experiments

S.H.A., **B.A.**, **D.W.-L.**, **K.S.** performed the selected reaction monitoring experiment

U.G., **S.H.A.** and **J.M.** assembled the transcriptomes

G.G. performed the virtual mutagenesis

S.H.A. performed all other experiments

P.W. and **U.G.** designed the project with contribution from **B.A.**

B.A. wrote the manuscript with contributions from **S.H.A.**, **U.G.** and **P.W**.

All authors read and approved the manuscript.

ACKNOWLEDGEMENTS

PEPC recombinant proteins were kindly provided by Judith K. Paulus of the group of Prof. Dr. Georg Groth (Paulus et al., 2013). We thank Bianca Jacobs for providing the *PPCKA* complementary DNA from *F. trinervia.* The single read sequencing was performed with the Illumina HiSeq2000, by the BMFZ (Biologisch-Medizinisches Forschungszentrum), MZL, Universitätsstr.1, 40225 Düsseldorf, Heinrich-Heine-Universität.

FOOTNOTES

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) through the International Research Training Group IRTG 1525 (SA, BA and PW) and the Excellence Cluster EXC 1028 (PW). Additional funding for work in East Lansing came from Michigan AgBioResearch.

The manuscript of Chapter 2 was submitted for publication to *Plant Physiology*

Chapter 3

Global regulation of the diurnal leaf transcriptomes from *F. pringlei* and *F. trinervia*

ABSTRACT

Higher plants sense environmental changes which are transduced into changes in expression of diurnally regulated genes and physiological and metabolic processes. In the genus *Flaveria* with photosynthetic diverse C_3 to C_4 species, the understanding of the diurnal regulation of gene expression is in the initial stages. Photosynthetic genes of the C_4 cycle of *Flaveria* evolved towards C_4 photosynthesis through co-option from non-photosynthetic genes of C_3 ancestor *Flaveria*, that resulted in high-level and cell-type specific expression. Thus, this study reveals to what extent the diurnal regulation of gene expression was altered during C_4 evolution in *Flaveria*. The diurnal transcriptomes from closely related *Flaveria pringlei* (C_3) and *Flaveria trinervia* (C_4) were used to profile the gene expression along a 24-hour day/night cycle. The alterations of the transcriptome were associated to C_4 photosynthesis and in response to diurnal changes. Diurnally regulated genes of encoding enzymes with predicted function in various biological pathways were identified in *F. pringlei* and *F. trinervia*, and compared between the species. Photosynthetic-type differences between *F. pringlei* and *F. trinervia* were ascribed to differences in the absolute transcript abundances, rather than on differences in the diurnal pattern of gene expression. However, particular genes gained or lost the ability to respond to diurnal changes towards C_4 evolution.

INTRODUCTION

C_4 photosynthesis derived out of a C_3 plant by - in almost all C_4 plants - the development of a two-cell type structure (Kranz anatomy). The C_4 cycle evolved in a step-wise order by the assembly of various C_4 traits. In the C_4 cycle the phospho*enol*pyruvate carboxylase (PEPC) is the entry enzyme for atmospheric CO_2 uptake in the mesophyll cells, and is spatially separated from the activity of ribulose-1.5-bisphosphate carboxylase/oxygenase (Rubisco) of the Calvin-Benson cycle in the bundle-sheath cells. PEPC generates an elevated concentration of CO_2 around Rubisco that improves the photosynthetic efficiency, that evolved in adaptation to low atmospheric CO_2 concentrations and high temperatures. This CO_2 concentrating mechanism reduces the oxygenation rate of Rubisco and thereby the energy loss through photorespiration (Peterhänsel and Maurino, 2011; Sage et al., 2012). The rate of photoassimilate efflux required for the biomass production is higher and faster in the light in C_4 plants (e.g maize, Sorghum, *Flaveria, Panicum*) than in C_3 plants (e.g. wheat, oat, barley, *Flaveria, Panicum*) (Leonardos and Grodzinski, 2000).

Molecular changes of pre-existing non-photosynthetic genes in C_3 ancestor plants resulted in high-level and cell-type specific expression of genes that execute photosynthetic functions in the C_4 cycle (Aubry et al., 2011; Christin et al., 2013; Gowik et al., 2011; Sheen and Bogorad, 1985; Sheen, 1999; Shu et al., 1999). Many genes of the C_4 cycle are therefore members of a small gene family within C_3 and C_4 species.

With the application of the technology of RNA sequencing it is possible to extent the understanding of the molecular and biochemical basis of the C_4 cycle. A comparative study of the transcriptome from C_3 to C_4 species revealed global changes in the absolute transcript abundances of genes as a

predisposition for the evolution of C_4 photosynthesis (Bräutigam et al., 2011; Gowik and Westhoff, 2011; Gowik et al., 2011).

Photosynthesis is a light-dependent process adjusted to the day/night cycle in either C_3 or C_4 plants. Therefore, it is reasonable to study, in the context of C_4 evolution, the regulation of the transcriptome from C_3 or C_4 species by diurnal changes. The leaf transcriptome from *F. pringlei* (C_3) and *F. trinervia* (C_4) was profiled along an entrained 24-hour day/night cycle using RNA sequencing. The *Flaveria* genus was used for this purpose, for possessing different photosynthetic modes with a continuous gradation from C_3 to C_4 photosynthesis (Mckown et al., 2005).

During the day sucrose is synthesized from triose phosphate (TP) to be immediately exported for growth. Some TP is converted in the leaves into storage carbon compounds (mostly starch) (Leonardos and Grodzinski, 2000; Leonardos et al., 2006). At night, starch is degraded back into TP used for sucrose synthesis, that becomes remobilized in the phloem and used in the non-photosynthetic tissue (Sharkey, 1985). Therefore, the sucrose content is maintained throughout the day (Geiger and Servates, 1994; Bläsing et al., 2005), whereas starch accumulates in a transitory manner (Weise et al., 2011; Smith et al., 2004). Although of similarities in starch and sucrose metabolism, the photo-assimilated carbon transfer varies between C_3 and C_4 plants. In C_4 plants the starch metabolism is preferentially restricted to the bundle-sheath cells and sucrose synthesis to the mesophyll cells (Leonardos and Grodzinski, 2000; Weise et al., 2011).

Sucrose, starch and light function as crucial diurnal signals that regulate the plants´ growth and development (Smith et al., 2004; Bläsing et al., 2005; Smith and Stitt, 2007). Photosynthesis and the sucrose/starch allocation are processes controlled by the expression of genes either induced or repressed by these diurnal signals (Krapp et al., 1993; Furbank and Taylor, 1995; Wang et al., 2000; Schaffer et al., 2001; Bläsing et al., 2005; Smith and Stitt, 2007). Genes oscillating significantly in a diurnal cycle are considered as diurnally regulated and include those that are controlled by either light (Thum et al., 2004; Tyagi and Gaur, 2010), sugars (Price et al., 2004; Bläsing et al., 2005), nutrients (Scheible et al., 2004), water deficit (Kawaguchi et al., 2004) or/and by the circadian clock (Harmer, 2009). Sugar levels *per se* could potentially regulate gene expression directly (Bläsing et al., 2005).

The circadian clock enables the plant to anticipate changes of the environment. The circadian clock is composed by clock regulators that are interacting with signaling pathways and influencing each other by transcriptional activation/repression in feedback-loops (Harmer, 2009; Farré and Weise, 2012). These are oscillating in response to day/night changes, and thereby induce rhythmic transcription of circadian-clock regulated genes involved in physiological and metabolic processes. Light resets the clock through the control over the rhythmic amplitude of the clock regulators. Clock regulators can be degraded by light, or interacting with each other in a light-dependent manner (Harmer, 2009). The stability and translation of the mRNA of some clock-regulated genes are concomitantly controlled by light signals (Harmer, 2009).

The diurnal/circadian system is a widespread mechanism and was suggested to be conserved among mono and dicots (Young and Kay, 2001; Murakami et al., 2007; Khan et al., 2010; Filichkin et al., 2011).

Although the entire set of genes required for the regulation of diurnal biological processes is unknown, a great overview of the diurnally/circadian regulated transcriptome is characterized in *Arabidopsis* (Schaffer et al., 2001; Harmer et al., 2000), poplar, rice (Mockler et al., 2007; Murakami et al., 2007; Filichkin et al., 2011; Xu et al., 2011), maize (Khan et al., 2010) and soybean (Hudson, 2010). In *Arabidopsis* an estimation of around 30 % of the total number of expressed genes are diurnally regulated (Hudson, 2000; Bläsing et al., 2005) and 6 % are indicated to be circadian regulated (Harmer et al., 2000). In soybean 1.8 % (Hudson, 2000) and in maize 10 % (Khan et al., 2010) are circadian clock-regulated genes.

This study reveals the C_3/C_4 differences and similarities in terms of diurnal patterns of gene expression and absolute transcript abundances in *Flaveria*. In addition, the diurnal transcriptome from *Flaveria* was compared to that of other taxonomic groups (Bläsing et al., 2005; Khan et al., 2010; Filichkin et al., 2011). The discrepancy in the absolute transcript abundances of genes between *F. trinervia* and *F. pringlei* changes in response to diurnal changes. The diurnal pattern of expression of genes with function in various pathways was highly similar in *F. trinervia* and *F. pringlei*. However, particular genes indicate for the impact of diurnal regulation of gene expression as adaptive response to new functions for C_4 photosynthesis.

RESULTS

The diurnal transcriptome from *F. pringlei* and *F. trinervia* was used to define the global changes in diurnal pattern of gene expression and absolute transcript abundances within each of these species and to reveal C_4-related alterations. *F. pringlei* and *F. trinervia* leaves exposed to a 10-h (light)/14-h (dark) cycle were harvested at different time points every 4 hours from the onset of light. The samples corresponded to seven harvested time points along the diurnal cycle for each species. These were analyzed by transcriptome profiling using RNA sequencing by the Solexa/Illumina technology. RNA sequencing resulted in clean reads that were mapped and aligned against the *Arabidopsis thaliana* transcriptome using the BLAST-like alignment tool (BLAT) (Kent, 2002). The best hit retained from the alignment of the reads were then mapped against the transcriptome of *F. robusta* with Bowtie2 (version 2.0.6.) (Langmead and Salzberg, 2012). This program aligned the sequenced reads as a measure of the transcript abundances into reads per million mapped reads and kilobase transcript length (RPMK). The RPMK counts from *F. pringlei* and *F. trinervia* were annotated against the reference genome of *Arabidopsis*, so that each gene from *Flaveria* was assigned to the respective biological functional class defined by BINs (Thimm et al., 2004; Usadel et al., 2005). The same RNA-seq data set from *F. pringlei* and *F. trinervia* was visualized through various tools. Each tool considered to condense and simplify the multidimensional RNA-seq data to a core set.

1. Photosynthetic-type separation relate to differences in terms of absolute transcript abundances

The average of the normalized RPMK expression values from three biological replicates for each gene at each time point was $\log_2$ transformed and processed with the TIGR MultiExperiment Viewer (MeV) program for principle component analysis (PCA) and the generation of clusters (Saeed et al., 2003).

The diurnal transcriptome data set of both species was reduced to a limited number of principle components by PCA (Raychaudhuri et al., 2000). The principle component represents individual genes and their significant contribution to the variability in the data set. The results from the PCA analysis are shown on the PCA plot, that summarised the principle components into two groups located opposite to each other [**Figure 1**]. The two groups are separated according to photosynthetic-type differences. Indicating, that the variability within the RNA data set is more related to differences between *F. pringlei* and *F. trinervia* than between identical time points.

A second approach of the generation of clusters was used to reveal whether the photosynthetic-type differences relate to differences in terms of either transcript abundances or to the diurnal pattern of gene expression. Clustering of the diurnal transcriptome from *F. pringlei* and *F. trinervia* partitioned the total number of genes in the RNA-seq data set. Each cluster represents a subset of genes grouped together according to similarities in diurnal pattern of gene expression and transcript abundances. With KMC-clustering the number of clusters and iterations to run were predetermined (Ben-Dor et al., 1999).

In total 10 KMC clusters were generated from the diurnal transcriptome from *F. pringlei* and *F. trinervia* [**Figure 2**]. The separation by photosynthetic-type differences relate to differences in terms of absolute transcript abundances between the species. Cluster 1 to 5 represent genes with differences in the absolute transcript abundances between *F. pringlei* and *F. trinervia*. In summary, more than the half of the total number of detected genes varied quantitatively in terms of absolute transcript abundances between the species. Within this group of genes the expression was up-regulated in *F. pringlei* with regard to *F. trinervia* (around 37 % of the total number of detected genes). Cluster 1 and 2 show genes with transcript abundances up-regulated in *F. pringlei* relative to *F. trinervia.* Cluster 3 to 5 correspond to genes up-regulated in *F. trinervia* with regard to *F. pringlei* (around 26 % of the total number of detected genes). Whereas, the remaining 37 % of the total number of genes had similar absolute transcript abundances between the species.

2. Photosynthetic-type similarities in terms of diurnal pattern of gene expression

KMC-clustering further shows that by 78 % of the total number of detected genes the diurnal pattern of gene expression was similar between *F. pringlei* and *F. trinervia* [**Figure 2**]. The relation of the diurnal transcriptome from *F. pringlei* to that of *F. trinervia* was further revealed by means of the pearson correlation coefficient (*r*). The pearson correlation considers the diurnal trend or pattern of gene expression, regardless of the overall absolute transcript abundance. The pearson correlation coefficient was estimated by a pair-wise comparison between two variables (RPMK values for each annotated gene along the diurnal cycle from *F. pringlei* to those of *F. trinervia*).

The diurnal pattern of gene expression was highly similar between *F. pringlei* and *F. trinervia*. A positive correlation in diurnal patterns of gene expression was estimated by 76.5 % of the total number of annotated genes [**Figure 3**]. Within this group 28.5 % of all genes were strongly correlated with *r* greater than 0.75. In a minor group of 5 % the diurnal pattern of gene expression from *F. pringlei* was opposite to that of *F. trinervia*. Finally, by 18.5 % of all genes the diurnal pattern of gene expression was unrelated between the species.

3. Functional classes of genes that exhibit similar diurnal patterns of expression between F. pringlei and F. trinervia

In order to reveal C_4-related changes in terms of diurnal pattern of gene expression or in transcript abundance associated to biological pathways, the genes from the diurnal transcriptome from *F. pringlei* and *F. trinervia* were assigned to functional classes. The functional classes were identified and displayed using the PageMan software (Usadel et al., 2006) and [**Figure 4**;**5**]. This program compares the changes of a single gene to that of the total number of genes, and over-represents those that surpass a fold-change in expression (a high absolute p-value is converted to a z-value and represented as blue) (Usadel et al., 2006).

The correlation coefficient values (from figure 3) for each annotated gene were used as input in PageMan. The functional classes were thereby associated to C_3/C_4-differences or -similarities in terms of diurnal pattern of gene expression [**Figure 4**]. Note that the probability that the functional class becomes over-represented is high in a group with a low number of genes (such as the inverse correlated group) (see figure 2).

Genes with differential diurnal pattern of expression between the species were over-represented in the functional class of photorespiration, the Calvin-Benson cycle, cell wall metabolism, synthesis and elongation of fatty acids, isoleucine, glucosinolates, flavonoids, biotic stress, C1-metabolism, protein glycolysation, the development in late embryogenesis, major carbon metabolism, degradation of lipids, glutamate and tryptophan, phytohormone metabolism (ethylene and brassinosteroids), tetrapyrrole synthesis, nitrilases, RNA regulation of lateral organ boundaries and receptor kinases. Genes with a similar diurnal pattern between the species were assigned to myb-transcription factors and the transport plasma membrane intrinsic protein (PIP). Genes with no correlation in diurnal pattern between the species were over-represented were assigned to transcriptional machinery, DNA synthesis, glycolysis, gluconeogenesis, synthesis and degradation of leucine, jasmonate, nucleotide metabolism, β-1,3 glucan hydrolase, protein degradation and mitochondrial protein targeting.

4. Functional classes of genes associated to clusters

Further, the diurnal transcriptome from *F. pringlei* and *F. trinervia* was clustered by using the hierarchical clustering module, in order to identify functional classes with C_4-related diurnal pattern of gene expression. 200 clusters were generated by a non-predetermined number [**Figure 5A**]. Each cluster was linked to the functional class identified and shown by PageMan [**Figure 5B**]. 6 clusters were

selected to demonstrate the photosynthetic-type differences in terms of diurnal pattern of gene expression and transcript abundances. Many of the remaining clusters represent genes with no clear contrast either within species or between the species.

Several functional categories were associated to more than one cluster. Examples are genes of the Calvin-Benson cycle associated to cluster 1, 3 and 5. Genes involved in abiotic stresses were found within cluster 4, 5 and 6. Genes involved in nitrate metabolism were found within cluster 1 and 6. Indicating, that genes of the same functional categories or metabolic pathway can be differently expressed in response to diurnal changes

Cluster 1 and 2 correspond to genes with absolute transcript abundances up-regulated in *F. pringlei* relative to *F. trinervia*. Genes with a greater transcript abundance in *F. pringlei* relative to *F. trinervia,* but with a similar diurnal pattern of gene expression were related to photorespiration, fermentation, oxidative pentose phosphate (OPP), mitochondrial electron chain, brassinosteroids, jasmonate, glycine hydromethyltransferase, oxygenase, posttranslational protein modification (PTM) and transport (cluster 1). Genes with a greater transcript abundance in *F. pringlei* relative to that in *F. trinervia* and different diurnal pattern of gene expression between the species corresponded to the glucose-6-phosphate isomerase and salicylic acid (cluster 2).

Cluster 3 and 4 correspond to genes up-regulated in *F. trinervia* relative to *F. pringlei.* Genes with a greater transcript abundance in *F. trinervia* relative to the one in *F. pringlei,* but with similar diurnal pattern of gene expression between the species corresponded to the synthesis of serine, glycine and cysteine, co-factors and vitamines, glucosidases, galactosidases and mannosidases, invertase, signaling (phosphoinositide) (cluster 3). Genes with a greater transcript abundance in *F. trinervia* relative to the one in *F. pringlei* and different diurnal pattern of gene expression between the species, corresponded to the PEPC kinase of glycolysis (cluster 4).

Genes up-regulated during the day in both species and with similar absolute transcript abundance related to the state transition of the light reaction of photosynthesis, steroids, synthesis of starch, glutamate and histidine, choline kinase, auxin metabolism, biotic stresses, phosphotransfer, β-1,3 glucan hydrolases and phosphate transport shared a similar diurnal pattern of gene expression and comparable transcript abundances between the species (cluster 5 and 6).

5. The discrepancy in transcript abundance between the species varies according to diurnal changes

The alterations in transcript abundances were compared between the diurnal transcriptome from C_3 and C_4 *Flaveria* in association to responses in gene expression to diurnal changes. The ratio of the transcript abundance of each gene from *F. pringlei* to that of *F. trinervia* was evaluated at the most contrasting time points. The ratio was assigned to functional categories and displayed by the MapMan diagram of metabolic pathways (Thimm et al., 2004; Usadel et al., 2005) [**Figure 6**;**7**]. Blue indicates for a higher expression of *F. trinervia* in respect to *F. pringlei*, whereas red indicates for a higher expression of *F. pringlei* in respect to *F. trinervia* (Usadel et al., 2005). Genes related to the electron transport, the Calvin-Benson cycle and photorespiration [**Figure 6**] and genes involved in other biological pathways were visualized in this manner [**Figure 7**].

The discrepancy in transcript abundance of some genes between C_3 and C_4 *Flaveria* varied during the diurnal cycle. Transcript abundances of genes of the light reactions were in general higher in *F. trinervia* than in *F. pringlei* at dawn and noon [**Figure 6**;**7**]. The light harvesting complex (*lhc5*) of photosystem I and *lhcb6*, *PSB-like 1*, PSB-like protein 2 (*PPL2*), NDH-dependent cyclic electron flow (*NDF2* and *NDF4*), *PSBY* and the ferredoxin-related gene of PSII were significant more abundant in *F. trinervia* relative to *F. pringlei* during the day [**Figure 11**]. A dependency between the photoperiod and the expression of transcripts of the light reaction in *F. trinervia* might be more accentuated than in *F. pringlei.*

The transcript abundances of most of the genes assigned to the photorespiratory cycle were reduced in *F. trinervia* with respect to *F. pringlei* along the 24-h cycle [**Figure 6**;**7**;**10**]. The transcript abundances of some genes assigned to the Calvin-Benson were reduced in *F. trinervia* and higher in *F. pringlei*. This trend can be ascribed especially to the gene expression during the dark (16-h) [**Figure 6**;**7**].

The transcript abundances of most of the genes involved in nitrate reduction and amino acid synthesis were more abundant in *F. pringlei* than in *F. trinervia* along the diurnal cycle. Transcript abundances of genes involved in starch and sucrose degradation, cell wall extension, sulfate reduction and the synthesis of phenylpropanoids were higher in *F. pringlei* than in *F. trinervia* at midnight [**Figure 7**]. Whereas, the expression of genes involved in synthesis of starch and sucrose, sulfate reduction and the synthesis of phenylpropanoids was higher in *F. trinervia* than in *F. pringlei* at dawn and noon [**Figure 7**].

Diurnal pattern of expression of single genes

In order to reveal alterations in the diurnal pattern of gene expression associated to C_4-related changes, diurnally regulated genes were identified and the diurnal pattern characterized in *F. pringlei* and *F. trinervia.* Diurnally regulated genes from publicly available data from *Arabidopsis*, tomato, maize, rice, soybean, *Alloteropsis* and poplar were used as a proxy and the gene counterparts selected from the diurnal transcriptome from *F. pringlei* and *F. trinervia* (Murakami et al., 2007; Hudson, 2010; Khan et al., 2010; Taygi and Gaur, 2010; Filichkin et al., 2011; Christin et al., 2013). A transcript was considered as significant diurnally regulated according to a ratio of at least 3-fold difference between the transcript abundance at the maxima to that of the minima, determined by ANOVA [**Table 1 to 10**]. The diurnal patterns of gene expression from *Flaveria* are represented as the average of the normalized RPMK expression values from three biological replicates at each time point [**Figure 8-17**]. The diurnal pattern of gene expression of diurnally regulated genes from *Flaveria* were compared to that of the counterpart genes from other species of publicly available literature and further at (http://diurnal.cgrb.oregonstate.edu/) (Mockler et al., 2007). The alterations in the diurnal pattern of gene expression were compared among C_3 and C_4 species of different taxonomic groups.

6. Diurnal expression of genes of the C_4 cycle

To study the evolution of C_4 photosynthesis it is important to reveal the diurnal expression of genes of the C_4 cycle. Transcripts of genes of the NADP-ME type C_4 pathway were selected which transcript abundances (Bräutigam et al., 2011; Gowik et al., 2011) or diurnal regulation (Christin et al., 2013) were altered during C_4 evolution. Genes encoding enzymes of the C_4 cycle in *F. trinervia* were analyzed and compared to the likely orthologs in *F. pringlei* with potential non-photosynthetic functions.

The transcript abundance of most of the genes of the C_4 cycle were significantly higher in *F. trinervia* than in *F. pringlei* at several time points [**Figure 8**]. This observation is consistent to a previous study with defined transcript abundances from C_3 to C_4 *Flaveria* at noon (Gowik et al., 2011) and C_3 and C_4 *Alloteropsis* along a 24-h diurnal cycle (Christin et al., 2013). The absolute transcript abundances of the genes encoding the chloroplastic malate dehydrogenase (MDH), pyruvate orthophosphate dikinase (PPDK), carbonic anhydrase (CA) and the NADP-malate enzyme (NADP-ME4) were expressed with highest absolute transcript abundances.

Genes encoding the CA, PEPC and PPDK enzymes are circadian-regulated in maize (Horst et al., 2009; Khan et al., 2010) and diurnally regulated in C_4 *Alloteropsis* (Christin et al., 2013). Although not considered as diurnally regulated in *Flaveria*, the diurnal phase of CA and PEPC from *F. trinervia* and *F. pringlei* were comparable to that of *Alloteropsis* (Christin et al., 2013). The Alanine aminotransferase (*Ala-AT*) encodes the enzyme which converts pyruvate to alanine in the NADP-ME C_4-type pathway. *Ala-AT* was classified in *F. trinervia* as diurnally regulated by a 3-fold increase at noon with respect to midnight [**Figure 8D**] and [**Table 1**]. NADP-ME converts the oxidative decarboxylation of malate into pyruvate in the C_4 cycle. Two *NADP-ME* isogenes are know from *Flaveria* from which one was suggested to be photosynthetic and the low-expressed the non-photosynthetic isoform (Marshall et al., 1996; Drincovich et al., 2000; Lai et al., 2002; Maier et al., 2011; Alvaréz et al., 2013). *NADP-ME1* was diurnally regulated and co-expressed between the species. *NADP-ME1* was significantly low abundant at dawn and peaked by the evening [**Figure 8G**]. The second isogene, *NADP-ME4* from *F. trinervia* was significantly higher than *NADP-ME1* from *F. trinervia* and *NADP-ME4* from *F. pringlei*. *NADP-ME4* may therefore encode the photosynthetic isoform of the C_4 cycle in *F. trinervia*.

7. Diurnal expression of genes of the Calvin-Benson cycle

Light-activated enzymes of the Calvin-Benson cycle are essential for the operation of photosynthesis. The transcript abundances of genes encoding enzymes of the Calvin-Benson cycle including the fructose-1,6-biphosphatase and phosphoribulokinase (PRK) were significantly more abundant in *F. pringlei* than in *F. trinervia* during the dark [**Figure 9**]. The CP12 domain containing protein 2 (CP12-2) peptide is regulated by light in *Arabidopsis* in the formation of a complex with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and PRK. CP12 was significantly more abundant in *F. trinervia* than in *F. pringlei* [**Figure 9A**]. The expression of the gene encoding the rubisco activase, required for the light activation of Rubisco (Portis et al., 1986), was by a 6-fold increase significantly higher in *F. trinervia* than in *F. pringlei* at dawn [**Figure 9B**].

Many genes of the Calvin-Benson cycle are circadian-regulated in soybean, maize and *Arabidopsis* (Hudson, 2010; Khan et al., 2010). The expression of *PRK* from wheat is induced by light (Raines et al., 1989) and circadian-regulated in soybean (Hudson, 2010). The amount of the mRNA of CP12 was reduced in the darkness in *Arabidopsis* (Marri et al., 2004). In *Flaveria* no significant changes in diurnal expression of genes of the Calvin-Benson cycle were observed. With exception to the expression of fructose-1,6-biphosphatase (*FBPase*) from *F. pringlei,* which peaked significantly at midday, but remained constant in *F. trinervia* [**Figure 9D**].

8. Diurnal expression of genes of the photorespiratory carbon cycle

Photorespiration is a light-dependent process linked to the photosynthetic electron transport. Therefore, genes assigned to photorespiration were selected from the diurnal transcriptome of *F. pringlei* and *F. trinervia*. The transcript abundances of the genes encoding enzymes of the photorespiratory pathway including the serine hydroxymethyl transferase (SHM), the 2-phosphoglycolate phosphatase 1 (PGLP1) and the D-isomer specific 2-hydroxyacid dehydrogenase were significantly higher in *F. pringlei* than in *F. trinervia* at near to all diurnal time points.

Among the mitochondrial genes of the photorespiratory pathway, the *SHM1* and the glycine decarboxylase P-protein (*GLDP*) are encoding enzymes that convert glycine to serine. In *Arabidopsis* *SHM1* increased significantly during the light and was circadian-regulated (McClung et al., 2000). *SHMT3* from *F. pringlei* peaked significantly during the day, but remained constant in *F. trinervia* [**Figure 10A**] and [**Table 3**]. The *GLDP2* identified in the transcriptome from *F. pringlei* and *F. trinervia* was not diurnally regulated in *Flaveria* [**Figure 10F**].

PGLP is an essential enzyme of the first reaction of the photorespiratory pathway that converts 2-phosphoglycolate to glycolate in the chloroplast (Schwarte and Bauwe, 2007). The PGLP enzyme was light-induced in corn (Baldy et al., 1989). Two *PGLP* isogenes were identified in the diurnal transcriptome from *Flaveria*. *PGLP1* from *F. pringlei* had a significant higher abundance in comparison to *PGLP2* from either *F. pringlei* or *F. trinervia* and to *PGLP1* from *F. trinervia*. *PGLP1* may therefore be more relevant for the photorespiratory pathway in *F. pringlei* than in *F. trinervia*. *PGLP1* from *Flaveria* was not diurnally regulated [**Figure 10D**]. Whereas, *PGLP2* was co-expressed between *F. pringlei* and *F. trinervia* with a significant increase by the middle of the dark period [**Figure 10E**].

9. Diurnal expression of genes of the photosynthetic electron transport chain

Based on the information that light, the circadian clock and other diurnal signals regulate the mRNA amount of genes of the light reactions in *Arabidopsis* these were considered for analyses in *Flaveria*. The *lhcb* gene encodes a apoprotein of the light-harvesting chlorophyll *a*/*b* binding complex (LHC) that binds carotenoids and chlorophyll *a* for the absorption and distribution of the light energy to the light reaction centers (photosystem (PS) I and II) (Liu et al., 2004). Several members of LHC proteins are encoded by a gene family, as found in *Flaveria* and other species, and are homologous in structure and associated to either PSI or PSII (Sheen, 1999). The *lhcb* genes of PSII are known to be light- and clock-regulated with

a peak in expression during the morning in a wide range of C_3 and C_4 species (Piechulla and Gruissem, 1987; Giuliano et al., 1988; Meyer et a., 1989; Ernst et al., 1990; Kaldis et al., 2003; Kellmann et al., 1993; Oberschmidt et al., 1995; Christensen and Silverthorne, 2001; Murakami et al., 2007; Khan et al., 2010). The three *lhcb* genes of PSII identified in *Flaveria* shifted significantly towards the day [**Figure 11A;C;D**] and [**Table 4**]. The *lhcb* transcripts were highly co-expressed between the *Flaveria* species and resembled the diurnal pattern to that of other species.

Many of the remaining genes of the light reaction were diurnally regulated in *Flaveria*, although not necessarily co-expressed between the species. The NDH-dependent cyclic electron flow (*NDF*) gene encodes the subunits of the plastoquinone. *NDF2* and *NDF4* from *F. pringlei* were diurnally regulated with an increase during the night [**Figure 11F**;**G**]. Several *Psb* genes from maize and *Arabidopsis* are known to be light-regulated (Taygi and Gaur, 2010). However, in *F. pringlei* the transcript abundances of *Psb-like 2* (*PPL2*), the photosystem II subunit T (*PSBTN*), the *PsbQ-like 1* were oscillating significantly in the diurnal cycle with a high abundance at midnight. By contrast, *PsbY* was diurnally regulated in *F. trinervia* with a peak in expression at noon [**Figure 11K**].

The function of the encoding protein kinase state transition 7 (*STN7*) from *Arabidopsis* is important for photosynthesis through the phosphorylation of the LHCII protein and for the adaptation to light. *STN7* was circadian clock-regulated in *Arabidopsis* (Samol et al., 2012). In *Flaveria STN7* is diurnally regulated and co-expressed with a peak in expression at noon and a decrease towards the night [**Figure 11L**]. The gene encoding the ATP synthase was co-expressed between the species with low levels during the morning [**Figure 11O**]. The expression of the genes encoding for the 2Fe-2S ferredoxin-like superfamily protein and the PsbP family protein reached their maxima during the day in *F. trinervia* [**Figure 11M**].

The thioredoxin f and m (*TRXf* and *TRXm*) transcripts from *Arabidopsis* are circadian clock regulated and the encoding proteins can mediate the activation of the enzymes of the Calvin-Benson cycle (Barajas-López et al., 2011). *TRXf1* was however not diurnally regulated in *Flaveria* [**Figure 11Q**]. Another gene involved in the cyclic electron transport chain, the PROTON GRADIENT REGULATION 5 (*PGR5*), is highly expressed in the bundle-sheath chloroplasts of C_4 *Flaveria* (Munekage et al., 2007/10). The function of *PGR5* is to promote the ATP demand for C_4 photosynthesis. The PGR5 protein was four-times higher in *F. trinervia* and *F. bidentis* compared to that in *F. pringlei* (Munekage et al., 2007/10). In this present study, the *PGR5* transcript abundance was similar between the C_3 and C_4 species, however only diurnally regulated in *F. trinervia* maximal expressed at dawn [**Figure 11R**].

10. Diurnal expression of circadian clock genes

Components of the circadian oscillator from *Arabidopsis* were considered as a proxy for the analyses in *Flaveria* (Schaffer et al., 2001; Bläsing et al., 2005; Smith and Stitt, 2007; Tyagi and Gaur, 2010). A set of circadian regulators was identified from the diurnal transcriptome of *Flaveria,* which were either expressed during the morning or the evening.

The absolute transcript abundances of the circadian regulators were mostly similar between *F. pringlei* and *F. trinervia* throughout the diurnal cycle. Exceptions are the pseudo-response regulator 3 and 5 (*PRR3* and *PRR5*), PHYTOCLOCK 1 (*LUX/PCL1*) and GIGANTEA (*GI*), which were significantly higher in *F. pringlei* than in *F. trinervia* during the day [**Figure 12C**;**D**;**F**;**N**]. Whereas, the transcript abundances of *PRR7* was significantly higher in *F. trinervia* than in *F. pringlei* during the day [**Figure 12E**]. A phase-shift in gene expression between *F. pringlei* and *F. trinervia* was also observed for the light-insensitive period1 (*LIP1*), TIMING OF CAB1 (*TOC1*) and the EARLY FLOWERING 3 (*ELF3*), which were significantly higher in *F. trinervia* than in *F. pringlei* during the night [**Figure 12K**;**I**;**O**]. The CIRCADIAN CLOCK ASSOCIATED (*CCA1*), the LATE ELONGATED HYPOCOTYL (*LHY1*) and the CCR-like (*CCL*) from *F. pringlei* and *F. trinervia* exhibited high absolute transcript abundances.

The three *PRR* genes from *Flaveria* were diurnally regulated [**Figure 12C-E**] and [**Table 5**]. *PRR7* from *Flaveria* and *Arabidopsis* are morning-expressed genes (Farré et al., 2005; Khan et al., 2010) and [**Figure 12E**]. In *Flaveria* the expression of *PRR3*, *PRR5, LUX/PCL1*, *CCL* and *GI* was highest at the day-to-night transition similar to that from *Arabidopsis,* rice, maize and tomato (Yakir et al., 2006; Murakami et al., 2007; Para et al., 2007; Facella et al., 2008; Khan et al., 2010). In *Flaveria* the expression of ZEITLUPE (*ZTL*) was not oscillating in the diurnal cycle, consistent to *ZTL* from rice (Murakami et al., 2007) and [**Figure 12G**].

The morning gene complex *CCA1/LHY* from *Arabidopsis* is anti-phase regulated to the evening-expressed *TOC1* (Farré et al., 2005). The transcript abundances of *CCA1*/*LHY* are high at dawn, while repressing the expression of *TOC1* (Schaffer et al., 1998; Alabadi et al*.,* 2001; Murakami et al., 2007). The diurnal patterns of *CCA1*, *LHY* and TOC1 from *F. pringlei* and *F. trinervia* are resembling those from *Arabidopsis* [**Figure 12H-J**].

The phytochrome interacting factor 3 (PIF3) can directly bind at the G-box motif of the promotor of *CCA1* and *LHY* to induce the expression of these genes (Martinez-Garcia et al., 2000). *PIF3* was degraded upon light induction in *Arabidopsis* (Bauer et al., 2004). The expression of *PIF3* from *F. pringlei* decreased with the onset of light [**Figure 12L**].

ELF3-ELF4-LUX is a complex that regulates the late evening/night timing of the clock in *Arabidopsis* (Hicks et al., 2001; Nusinow et al., 2011). In *Arabidopsis* ELF4-ELF3-LUX are repressed by *CCA1/LHY* (Hazen et al., 2005; Pokhilko et al., 2012). *ELF3* from *F. trinervia* was highly expressed during the night [**Figure 12O**] and [**Table 5**]. As found in rice, *ELF4* was not diurnally regulated in *Flaveria* (Murakami et al., 2007) and [**Figure 12P**].

11. Diurnal expression of light-regulated genes

Genes encoding proteins of photoreceptor signaling pathways and genes known to be light-responsive in *Arabidopsis* were considered for analyses in *Flaveria*. The suppressor of phyA-105 (*SPA3*) suppresses photomorphogenesis and regulates the elongation growth in *Arabidopsis*. The expression of *SPA3* was circadian regulated in tomato (Facella et al., 2008) and diurnally regulated in *Arabidopsis* (Ishikawa et al., 2006). *SPA3* decreased gradually upon dawn in *F. trinervia* [**Figure 13A**]. The phototropic-

responsive (*NPH3*) encodes a signal transducer element in response to light, and functions in the regulation of phototropism in *Arabidopsis* (Motchoulski and Liscum, 1999). *NPH3* was repressed after dawn in *F. pringlei* [**Figure 13B**].

Phototropin (*PHOT*) encodes a blue light photoreceptor kinase. *PHOT1* is considered as circadian regulated in maize and *Arabidopsis* (Khan et al., 2010). Two *PHOT* isogenes were identified in *Flaveria* from which *PHOT1* peaked in the late evening in *F. pringlei* and *F. trinervia* [**Figure 13C**]. The gene which encodes the light-harvesting-like protein (LIL3) functions in the light harvesting process of the light reactions (Tanaka et al., 2010). *LIL3* was significantly up-regulated during the day in *F. trinervia* with respect to the night, but not diurnally regulated in *F. pringlei* [**Figure 13E**].

12. Diurnal expression of genes related to transport

In plants, transport can be a process adjusted to light or the nutrient availability (Schaffer et al., 2001; Bläsing et al., 2005; Facella et al., 2008). Several genes encoding transporter proteins from *Arabidopsis* that transfer ions, carbon or nitrate function in a diurnal-manner. The transcript abundances of the sulfate transporter (*SULTR4*) gene from *F. pringlei* and *F. trinervia* peaked at dawn and decreased gradually until the start of the night [**Figure 14A**]. The gene encoding the anion exchange transporter (HCO_3^- transporter) from *F. trinervia* was significantly oscillating with a peak at the late evening [**Figure 14B**].

The phosphate ion transporter proteins which are encoded by a phosphate transporter (*PHT*) gene family are located in the chloroplast inner envelope membrane in *Arabidopsis*. Phosphate transport is required for starch biosynthesis potentially generated by PHT transporters, such as *PHT2;1* and *PHT4;4* (Guo et al., 2008). In *Arabidopsis PHT2;1* and *PHT4;4* are light-induced genes, whereas *PHT4;2* transcript is expressed in the dark and *PHT4;1* circadian-regulated by CCA1 (Schulz et al., 1993; Guo et al., 2008).

In *F. trinervia PHT4;4* peaked at dusk and the co-expression of *PHT2;1* from *F. pringlei* and *F. trinervia* was high during the day [**Figure 14C**;**D**]. Indicating, that *PHT2;1* might be light-regulated in *Flaveria*. The transcript abundance of the phosphate transporter *PHT4;4* was significantly higher in *F. trinervia* than in *F. pringlei*. Another *PHT4;2* isogene was not diurnally regulated in *Flaveria* (data not shown). The phosphate transporter traffic facilitator (*PHF1*) is involved in phosphate transport. *PHF1* from *F. trinervia* peaked during the night and was anti-phase between *F. pringlei* and *F. trinervia* [**Figure 14E**].

The potassium transporter (*AKT*) from *Arabidopsis* encodes a photosynthate- and light-dependent transporter (Deeken et al., 2001). Among the members of the *AKT* gene family, the transcript abundances of *AKT2/3* from *Flaveria* peaked at dawn and decreased with a trough gradually until the start of the night [**Figure 14F**]. The expression of the plasma membrane intrinsic protein 1C (*PIP1C*) that function e.g. in water transport, was diurnally regulated in *F. pringlei* and more abundant in *F. pringlei* than in *F. trinervia* at dawn [**Figure 14G**].

The nitrate, starch and sucrose metabolisms are diurnally regulated processes, which involve the transcriptional regulation of transporters (Choiu and Bush, 1998; Coruzzi and Bush, 2001). Therefore, genes encoding transporters known to be carbon- or nitrate-responsive were analyzed in *Flaveria*.

The nitrate transporter NRT is involved in responses to nitrate and auxin-mediated signaling. *NRT1-1* from tomato and *NRT2.1* from tobacco are diurnally regulated genes (Stitt and Krapp, 1998; Facella et al, 2008). *NRT1.5* was diurnally regulated and co-expressed between *F. trinervia* and *F. pringlei* with maximum abundance at the start of the light period and a strong decline towards the night [**Figure 14H**]. The diurnal expression of *NRT2.1* from tobacco was consistent to that of *NRT1.5* from *Flaveria*.

Among the members of the glucose-responsive sucrose transporter *STP* gene family (Price et al., 2004), the *STP13* from *F. trinervia* and *F. pringlei* peaked significantly at midnight [**Figure 14I;J**] and [**Table 7**]. The glucose transporter GLT1 located in the chloroplast envelope, contributes to the export of glucose (Majeran et al., 2010). The transcript levels of *GLT1* changed in a 24-h cycle in *Arabidopsis* (Smith et al., 2004; Stitt et al., 2010). However, in *Flaveria* the expression of *GLT1* remained constant in the diurnal cycle [**Figure 14K**]. The dicarboxylate transporter DiT1 has antiporter activity at the chloroplast in order to shuttle the oxaloacetate/malate intermediates (Taniguchi et al., 2002). Among the members of the *DiT* gene family, the *DiT2.1* was identified in the transcriptome data set from *F. pringlei* and *F. trinervia*. *DiT2.1* from *F. trinervia* was significantly high during the photoperiod and more abundant than *DiT2.1* from *F. pringlei* [**Figure 14L**].

13. Diurnal expression of genes encoding transcription factors

Among the genes coding for proteins required for the control of transcription, some were classified as responsive to diurnal stimuli in *Arabidopsis*. The Golden2-like (GLK) transcription factor is known to influence the expression of photosynthetic genes required for chlorophyll biosynthesis, light harvesting and the electron transport in *Arabidopsis* (Waters et al., 2009). *GLK1* was diurnally co-expressed between *F. pringlei* and *F. trinervia* with a peak in expression at dawn [**Figure 15B**] and [**Table 8**]. The ETHYLENE-INSENSITIVE 3 (EIN3) transcription factor from *Arabidopsis* responds to far red and red light for ethylene-induced responses. EIN3 is associated to the PIF3-dependent growth-promoting pathway to activate the growth during the light (Zhong et al., 2012). *EIN3* was not diurnally regulated in *Flaveria* [**Figure 15C**]. The expression of the homeobox 1 (*HB1*) transcription factor encodes a homeodomain protein from *Arabidopsis* involved in the responses to blue light, nitrate and salt levels (Wang et al., 2000; Seki et al., 2002). *HB-1* remained constant in the diurnal cycle of both *Flaveria* species [**Figure 15D**].

Members of a gene family encoding the sigma factor transcriptional regulators (SIG) respond to light stimuli and are involved in the photosystem stoichiometry adjustment. The expression of *SIG1* is light-induced in maize (Lahiri et al., 1999), but was not diurnally regulated in *Flaveria* (data not shown). However, *SIG2* and *SIG6* from *F. pringlei* and *SIG5* from *F. trinervia* and *F. pringlei* were oscillating significantly in the diurnal cycle. The expression of *SIG2* and *SIG6* increased upon dawn, whereas the diurnal pattern from *SIG5* had the lowest amplitude by the start of the night [**Figure 15E-15H**].

14. Diurnal expression of genes related to the hormone metabolism

Circadian clock-associated genes are known to regulate the phytohormone biosynthesis and responses in *Arabidopsis* (Covington and Harmer, 2007; Harmer, 2009). Therefore, the diurnal pattern of gene expression of hormone-related genes known to be diurnally regulated in *Arabidopsis*, were defined in *F. pringlei* and *F. trinervia.* The constitutive photomorphogenic dwarf (*CPD*) and the brassinosteroid-6-oxidase (*BR6OX2* or *CYP85A2*) are both light- and circadian clock-responsive genes and part of the brassinosteriod pathway in *Arabidopsis* (Bancos et al., 2006). *CPD* and *CYP85A2* were not diurnally regulated in *Flaveria* (data not shown). The 1-aminocyclopropane-1-carboxylate synthase 6 (*ACS6*) and S-adenosylmethionine synthase 4 (*MTO3*) are involved in ethylene biosynthesis and circadian-controlled transcripts in maize and *Arabidopsis*(Khan et al., 2010). *ACS6* and *MTO3* were not diurnally regulated in *Flaveria* (data not shown). The abscisic acid 8´-hydroxylase 1 (*CYP707A1*) encodes a protein member of the cytochrome p450 family, that is involved in the regulation of the abscisic acid metabolism in respond to red or far red light (Bolwell et al., 1994; Okamoto et al., 2006). The peak in expression of *CYP707A1* from *F. pringlei* was significant at noon, but remained constant in *F. trinervia* [**Figure 16A**] and [**Table 9**].

Gibberellin biosynthesis is involved in the control of the flowering time in *Arabidopsis* (Blázquez et al., 2002). The circadian-regulated genes ent-copalyl diphosphate synthase (*GA1*) and ent-kaurenoic acid oxidase 2 (*KAO2*) from maize are required for gibberellin biosynthesis (Khan et al., 2010). *GA1* and *KAO2* were not diurnally regulated in *F. pringlei* and *F. trinervia*. The GIBBERELLIC ACID INSENSITIVE (GAI) is degraded in response to gibberellin in *Arabidopsis*. The transcript abundances of *GAI* from *F. pringlei* and *F. trinervia* peaked at dawn and decreased with a trough until the start of the night [**Figure 16C**]. Indicating, that the levels of GA may change in *Flaveria* throughout the day, and act over *GAI*. The expression of *GASA6* encoding the gibberellin-regulated family protein, respond to GA and sucrose stimuli in *Arabidopsis*. *GASA6* from *F. pringlei* increased during the night with respect to the expression at noon [**Figure 16D**].

Some auxin-induced genes from *Arabidopsis* are concomitantly circadian clock regulated (Jouve et al., 1999; Covington and Harmer, 2007). The auxin-responsive PIN-formed 7 (*PIN7*) required for auxin efflux was circadian regulated in *Arabidopsis*. In *Flaveria PIN7* increased at midnight [**Figure 16E**]. The auxin F-box protein 5 (*AFB5*) encodes a putative auxin receptor of the auxin signal transduction pathway from *Arabidopsis* and is expressed at night (Covington and Harmer, 2007; Tan et al., 2007). In *F. pringlei AFB5* increased significantly at noon, and remained constant in *F. trinervia* [**Figure 16F**]. Among several members of the auxin response factors (*ARFs*) and of the INDOLE-3-ACETIC ACID INDUCIBLE (IAA) gene family identified in the transcriptome from *Flaveria* none was significantly diurnally regulated (data not shown).

15. Diurnal expression of genes related to carbon and nitrogen metabolism

Genes encoding proteins predicted to be involved in starch, sucrose and nitrogen biosynthesis and degradation are controlled by various diurnal signals including the circadian clock. In *Arabidopsis,* genes involved in the starch metabolism are diurnally controlled at the transcriptional and posttranscriptional level, whereas the amount of starch metabolic enzymes remains constant in the diurnal cycle (Smith et al., 2004). The nitrogen assimilation is light-dependent and the expression of nitrate- and nitrogen-responsive genes was suggested to be circadian-regulated at the transcriptional level (Stitt and Krapp, 1999; Farré and Weise, 2012).

Glucan water dikinase (*GWD1* and *GWD3*) are regulating the amount of starch by degradation required for growth and flowering (Weise et al., 2011). *GWD1* and alpha-amylase-like 3 (*AMY3*) are involved in circadian rhythms in *Arabidopsis*. *GWD1, GWD3, AMY3*, glucan phosphorylase 2 (*PHS2*) and the disproportionating enzyme 2 (*DPE2*) were low expressed during the first hours of the day, increased gradually until reaching a peak by the day/night transition and decreased during the night [**Figure 17A-D;F**]. Indicating, that the genes related to the degradation of starch were diurnally co-expressed between *F. pringlei* and *F. trinervia* and resembled those from *Arabidopsis* (Smith et al., 2004; Lu et al., 2005; Smith and Stitt, 2007). Although of the similar trend in expression between the species, *AMY3* was significantly more abundant in *F. trinervia* than in *F. pringlei* at 16-h [**Figure 17C**]. Whereas, the transcript abundances of the α-amylase 1 (*BAM1*) decreased with the onset of light [**Figure 17E**]. *GWD1* from rice was expressed consistently to that from *Flaveria* (Hirose et al., 2013). The maltose transporter 1 (MEX1) exports most of the carbon derived from the degradation of starch from the chloroplast (Weise et al., 2004). The expression of *MEX1* increased in *F. pringlei* during the night [**Figure 17G**].

Among several genes encoding proteins of the starch biosynthetic pathway, the starch branching enzyme 2.2 (*SBE2.2*) from *F. pringlei* and *Arabidopsis* remained low during the photoperiod, but increased during the night [**Figure 17H**] and (Smith et al., 2004). Similar to the diurnal pattern of the starch synthase 2 (*SS2*) from *Arabidopsis* and maize, *SS2* from *F. pringlei* and *F. trinervia* peaked by the start of the day (Khan et al., 2010) and [**Figure 17I**] and [**Table 10**]. Among the genes involved in the synthesis of sucrose, the transcript abundances of the gene fructose-1,6-biphosphatase from *F. pringlei* was diurnally regulated with high expression during the night [**Figure 17K**].

Further, genes encoding enzymes involved in cell wall modifications, which were found to be diurnally regulated in other species were analysed in *Flaveria*. The gene encoding the enzyme required for the degradation of the cell wall (pectine lyase) from *F. pringlei* increased at midnight with regard to the expression at noon [**Figure 17L**]. Among the members of the gene family encoding the protein cellulose synthase (CESA) required in the synthesis of cellulose, *CESA5* from *F. pringlei* and *F. trinervia* peaked significantly by the end of the day [**Figure 17O**]. The diurnal peak was consistent to that of *CESA6* from *Arabidopsis* and maize (Khan et al., 2010).

The nitrate reductase (*NIA*) encodes the enzyme (NR) that reduced nitrate to nitrite. Nitrate assimilation is stimulated in the light. *NIA* from leaves of tobacco was diurnally regulated with a decline by the middle of the photoperiod and the rhythmic transcription translated into the levels of protein (Scheible et al., 1997). The diurnally regulated *NIA2* was co-expressed between *F. pringlei* and *F.*

trinervia with a gradual decrease after dawn [**Figure 17P**]. In *Flaveria* the diurnal pattern of gene expression of *NIA2* correlates to the diurnal expression of the nitrate transporter *NRT1.5* and in addition to the assimilation of nitrate (decreased levels of nitrate upon light). The glutamate synthase 1 (*Glu1*) gene encodes the ferredoxin-dependent glutamate synthase (Fd-GOGAT), which responses to nitrate and accumulates in the light in *Arabidopsis* and barley to reassimilate photorespired ammonium (Stitt and Krapp, 1998; Feraud et al., 2005). *Glu1* was significantly more abundant in *F. pringlei* than in *F. trinervia* along the diurnal cycle, but not diurnally regulated [**Figure 17Q**]. The glutamine synthetase 1;4 (*GLN1;4*) encodes the glutamine synthetase (GS1) protein and is a sugar- and nitrate-responsive gene in *Arabidopsis*. *GLN1;4* from *F. pringlei* peaked significantly at noon and remained low in *F. trinervia* [**Figure 17R**].

16. Diurnal expression of genes related to stress, cell organization, calcium signaling and the secondary metabolism

The transcript abundances of the gene encoding the heat shock factor (HSF4) peaked significantly at 4-h in *F. trinervia* [**Figure 18A**]. This is consistent to the diurnally controlled expression of *HSF* in non-stressed plants of tomato (Facella et al., 2008), rice and poplar (Filichkin et al., 2011; Xu et al., 2011). The gene encoding the protein kinase superfamily family was circadian regulated in soybean (Hudson, 2010). The diurnal expression of the protein kinase superfamily protein peaked significantly during the night in *F. trinervia* [**Figure 18B**].

The calcium dependent protein kinase (CDPK) is required for calcium signalling and involved in abscisic acid signalling (Hrabak et al., 2003; Zhu et al., 2007). Among the numerous *CPK* isogenes identified in *Flaveria* merely the *CPK11* was considered as diurnally regulated in *F. trinervia* and *F. pringlei* with an increased expression during the night [**Figure 18E**]. This is in agreement with the circadian-regulated expression of *CPK11* from soybean (Hudson, 2010).

The expression of tubulin beta-1-chain (*TUB1*) encodes a protein involved in the cell organization and the transcript of some *TUB* isogenes from *Arabidopsis* were down-regulated by light (Leu et al., 1995). However, in *F. pringlei TUB1* peaked significantly at midday and was significantly higher to that of *F. trinervia* [**Figure 18F**]. *TUB1* from *F. trinervia* was not diurnally regulated. The diurnal patterns of expression of the genes *TUB2, TUB7*, alpha-tubulin (*TUA2* and *TUA4*) from *F. pringlei* and *F. trinervia* were highly resembling those of *TUB1* (data not shown).

The gene encoding the geranylgeranyl reductase (GGR) is involved in the chlorophyll/cysteine biosynthesis and in the regulation of photosynthesis. *GGR* was significantly co-expressed between *F. pringlei* and *F. trinervia* and up-regulated at noon with respect to the night [**Figure 18G**]. In *Arabidopsis* and maize the genes encoding enzymes of the secondary metabolism, cinnamyl coa reductase 1 (*CCR1*) and cinnamyl alcohol dehydrogenase (*CAD*) were circadian regulated (Khan et al., 2010). In *Flaveria* the expression of *CCR1* required for the synthesis of lignin was high at dawn [**Figure 18H**]. Among various *CAD* isogenes the *CAD5* from *F. pringlei* peaked significantly at dawn [**Figure 18I**].

DISCUSSION

Photosynthetic-type differences relate to variations in transcript abundances

The aim of this study was to analyse the diurnal transcriptome in the genus *Flaveria* in order to characterize potential C_4-related evolutionary changes. The major variability in the leaf diurnal transcriptomes from *F. pringlei* and *F. trinervia* profiled by Illumina RNAseq was related to photosynthetic-type differences [**Figure 1**]. The photosynthetic-type separation is likely the result of a high frequency of genes with different transcript abundances between *F. pringlei* and *F. trinervia* [**Figure 2**]. The differences in transcript abundances between the species are ascribed to photosynthetic and non-photosynthetic genes. The evolutionary pressure was high for photorespiratory genes, that resulted in a downregulation of the transcript abundances in *F. trinervia* with regard to *F. pringlei* [**Figure 6**;**7**;**10**]. The photosynthetic-type expression of *SHMT3*, *SHM2* and *PGLP1* may reflect the minimised photorespiration rates in C_4 *Flaveria* [**Figure 10**]. Genes recruited to the C_4 cycle were mostly upregulated in *F. trinervia* with regard to *F. pringlei* [**Figure 8**]. The high-level expression of genes of the C_4 cycle in *F. trinervia* represent an adaptive response to photosynthetic functions that evolved during C_4 evolution (Bräutigam et al., 2011; Gowik et al., 2011; Christin et al., 2013). Most of the genes involved in the N-metabolism were downregulated in the C_4 *Flaveria* relative to the C_3 *Flaveria* along the diurnal cycle [**Figure 7**;**17**]. This may correlate to the higher rate of photosynthesis per unit of N, that is greater in C_4 than in C_3 plants and results in a lower allocation of N to enzymes involved in the Calvin-Benson cycle and the photorespiratory cycle (Ku et al., 1979; Sage et al., 1987).

C_4-related differences in diurnal gene expression in *Flaveria*

This study describes that C_4 evolution includes alterations of transcript abundances in response to diurnal changes. In *Alloteropsis* the diurnal cycling of some genes of the C_4 cycle were potentially acquired during C_4 evolution (Christin et al., 2013). The genes of the C_4 cycle were not diurnally regulated in *F. trinervia* and *F. pringlei*. An exception was *Ala-At* which potentially gained the ability to respond to diurnal changes along the transition from C_3 to C_4 photosynthesis [**Figure 8D**]. In *Pinus sylvestris L.* alanine that is synthesised by *Ala-At* is required for protein synthesis and as amino donor for the amino acid metabolism (Otter et al., 1992). *Ala-At1* from *Pinus sylvestris L.* accumlated in a light-dependent manner. *Ala-At* from *F. trinervia* peaked at noon, which may indicate for an optimization of this gene for the C_4 cycle. Experiments are required to underline this conclusion. Light is required for the activity of some enzymes of the C_4 cycle and for the recruitment of gene expression to either the mesophyll or bundle-sheath cells (Sheen, 1999; Shu et al., 1999). Concluding, that the diurnal regulation of the C_4 cycle from *Flaveria* is regulated at the translational and transcriptional level. However, the expression of genes of the C_4 cycle from *Flaveria* and of the orthologues in C_3 *Flaveria* did not change remarkable in response to diurnal changes

FBPase of the Calvin-Benson cycle and *SHMT3* of the photorespiratory pathway were diurnally regulated only in *F. pringlei*, and therefore potentially lost the ability to respond to diurnal changes in *F.*

trinervia due to C_4-associated changes [**Figure 9D**;**10A**]. The mRNA of *SHM1* from *Arabidopsis* accumulated during the light, but is not directly regulated by illumination (McClung et al., 2000). A similar regulation in diurnal expression of *SHMT3* may be operative in *F. pringlei*. Although the photorespiratory pathway and the Calvin-Benson cycle are light-dependent processes in *F. pringlei* and *F. trinervia*, the diurnal regulation of *SHMT3* and *FBPase* might have a higher relevance in C_3 *Flaveria* than in C_4 *Flaveria.*

Many genes of the light reactions and the Calvin-Benson cycle were more abundant in *F. trinervia* than in *F. pringlei* during the day, but more abundant in *F. pringlei* than in *F. trinervia* during the night [**Figure 6**;**7**]. This may reflect on one site the synthesis of ATP and NADPH driven by the photosynthetic electron transport, which is required to match to the demand of the carbon assimilation in the Calvin-Benson cycle. And on the other, the diurnal demand of energy for photosynthesis, which is slightly higher in C_4 plants than in C_3 plants during the day (Kramer and Evans, 2011). In *F. trinervia* the light-dependent electron transport correlated to the high expression of the genes encoding for the 2Fe-2S ferredoxin-like superfamily protein and the PsbP family protein during the day [**Figure 11M**]. However, the expression of the same genes were not diurnally regulated in *F. pringlei*. The high PGR5 protein amount in C_4 *Flaveria* was found to promote the synthesis of ATP in the cyclic electron transport chain (Munekage et al., 2010). *PGR5* from *F. trinervia* was strongly expressed during the morning, which may indicate for an additional regulation of *PGR5* for the C_4 cycle in *Flaveria* [**Figure 11R**].

The trend in transcript abundances in response to diurnal changes, with a higher abundance in *F. pringlei* than in *F. trinervia* during the night, was similarly observed for genes of various pathways (cell wall extension, sulfate metabolism and phenylpropanoid synthesis) [**Figure 7**]. Ultimately, this may reflect a different regulation in gene expression of the night metabolism of *F. pringlei* to that of *F. trinervia* at different physiological levels.

In *Arabidopsis PHT4;4* is required for transport related to starch biosynthesis (Guo et al., 2008). In *F. trinervia* the diurnal regulation of *PHT4;4* may be required in the photosynthetic regulation in terms of partitioning fixed carbon. *PHF1* from *F. pringlei* was expressed in the morning and in anti-phase to that of *PHF1* from *F. trinervia* [**Figure 14E**]. The differentiated expression in terms of diurnal pattern of gene expression and absolute transcript abundances among the various genes of the encoding phosphate transporter isoforms, indicate for distinct physiological functions within *Flaveria* species. The high expression of *DIT2.1* from *F. trinervia* during the day, may reveal that the oxaloacetate/malate transport through the chloroplast envelope is a diurnally regulated process in C_4 *Flaveria* [**Figure 14L**].

In addition to the reduced expression of genes of the nitrate metabolism in C_4 *Flaveria* relative to C_3 *Flaveria* along the diurnal cycle, *GLN1;4* was only diurnally regulated in *F. pringlei* [**Figure 17R**]. Indicating, that in *F. pringlei* the expression of the *GLN1;4* for the assimilation of ammonia is adjusted to diurnal changes.

Global regulation of diurnal gene expression is conserved between *F. pringlei* and *F. trinervia*

This study emphasizes that the global regulation of the diurnal pattern of gene expression is not necessarily associated to C_4 evolutionary changes. The pearson correlation and cluster analyses was used to compare photosynthetic-type similarities in the global diurnal regulation of gene expression. The results withdrew a high correlation between *F. pringlei* and *F. trinervia* in terms of diurnal pattern of gene expression [**Figure 2**;**3**]. The phase-conservation of genes related to various pathways suggests for a common diurnal regulation of gene expression in *Flaveria*. This diurnal regulation already established in the C_3 photosynthetic pathway , indicates for a potential adaptation to diurnal changes in either *F. pringlei* or *F. trinervia.* In addition, the phase-conservation in *Flaveria* is in good agreement with the high degree of conserved diurnal/circadian profiles and phases between various C_3 and C_4 species (Khan et al., 2010; Filichkin et al., 2011).

Conserved diurnal pattern of photosynthetic genes among *Flaveria* and higher plants

In the C_3 plant *Arabidopsis NADP-ME* was suggested to be involved in the stabilization of the content of sugar, amino acids and glucosamides (Brown et al., 2010). In *F. pringlei* and *F. trinervia* the expression of *NADP-ME1* oscillated significantly in the diurnal cycle [**Figure 8G**]. Experimental approaches are required to elucidate the reason of the induction of the expression of *NADP-ME1* in *Flaveria* during the day for putative non-photosynthetic processes. *NADP-ME4* potentially represents the photosynthetic isogene in *F. trinervia*. The loss of the diurnal expression of *NADP-ME4* may reflect that the C_4 cycle is not operative through the diurnal regulation of this gene. A similar relationship was found between the two *PGLP* isogenes. The low-expressed *PGLP2* peaked during the night in *F. trinervia* and *F. pringlei* [**Figure 10E**]. The expression of *PGLP2* may therefore be important for the regulation of photorespiration in response to diurnal changes in both species.

The activation state of enzymes of the Calvin-Benson cycle increases during the light and deactivation occurs during the night (Buchanan, 1980/1991; Geiger and Servaites, 1994). The diurnal regulation of the enzymes of the Calvin-Benson cycle is not reflected at the transcriptional level in both *Flaveria* species. Indicating, that the adjustment of the Calvin-Benson cycle to diurnal conditions in *Flaveria* occurs mainly at the enzyme level.

The prominent light-/circadian-dependent expression of *lhcb* in C_3 and C_4 plants, was found to be regulated in response to diurnal changes in phylogenetically distant *Flaveria* species. The expression of *lhcb* may be induced by light signals, indistinguishable between the *F. pringlei* and *F. trinervia* species. The identification of potentially conserved *CCA1* binding sequence in the promoter of the chlorophyll *a/b* binding protein genes in *Flaveria*, would permit to elucidate the regulation of *lhcb* by the circadian clock and the interaction with phytochrome. Equally, the interacting partner of *lhcb,* the encoding protein *STN7* from *Arabidopsis* was diurnally regulated in *Flaveria* similar to the diurnal pattern of expression of *lhcb* [**Figure 11L**]. Therefore, a protein-protein interaction between *STN7* and *lhcb* may exist in both C_3 and C_4 *Flaveria* and as suggested for *Arabidopsis*, may have similar functions for the light reactions in *Flaveria* (Kramer and Evans, 2011).

Conserved diurnal pattern of clock components among *Flaveria* and higher plants

Light-/circadian-regulated genes and circadian clock regulators characterised in *Arabidopsis*, tomato, rice and maize (Murakami et al., 2007; Facella et al., 2008; Pokhilko et al., 2012) were identified in *Flaveria*. The peak in expression of the circadian components was shifted towards either the dusk or the dawn in *Flaveria*, according to the anticipation of changes caused by the day/night transitions [**Figure 12**]. The time-phased expression between the major clock regulators was mostly indistinguishable between *F. pringlei* and *F. trinervia*, and consistent to that of rice, poplar (Filichkin et al., 2011), maize and *Arabidopsis* (Michael et al., 2008; Khan et al., 2010). Therefore, the diurnal regulation of the clock components might be highly conserved among these species to that of *F. pringlei* and *F. trinervia*.

The circadian oscillator from *Arabidopsis* is composed by feedback-loops of the circadian clock regulators which are activating/repressing each other in response to diurnal changes. The *PRR* genes are components of the feedback-loop and have the function to regulate the transmission of light and the signals of temperature to the oscillator in *Arabidopsis* (Farré et al. 2005; Gould et al., 2006). The diurnal patterns of the three *PRR* isogenes from *Flaveria* were similar to those of *Arabidopsis,* rice, maize and tomato (Murakami et al., 2007; Para et al., 2007; Facella et al., 2008; Khan et al., 2010).

PRR5, *PRR7* and *PRR9* from *Arabidopsis* are repressing the transcription of *CCA1/LHY,* whereas *PRR7* and *PRR9* are in turn positively regulated by *CCA1/LHY* (Farré et al. 2005; Farré and Weise, 2012). *CCA1* directs both the photosynthetic light-dependent process and the nitrogen assimilation in *Arabidopsis* (Wang and Tobin, 1998; Gutierrez et al., 2008). CCA1 and LHY from *Arabidopsis* are closely related within the MYB-repeat region of these transcription factors. *CCA1*/*LHY* are induced by *TOC1* and in turn repressing *TOC,* and thereby oscillating anti-phase to each other in *Arabidopsis*, tomato and rice (Makino et al., 2002; Farré et al., 2005; Kim et al., 2007; Murakami et al., 2007; Facella et al., 2008; Pokhilko et al., 2012). This trend was observed in *Flaveria* with *TOC1* being transcribed at the evening and *CCA1*/*LHY* at dawn. This anti-phase expression could be interpreted as the inhibitory effect of (1) *CCA1*/*LHY* over *TOC1* and (2) *PRR5*/*PRR7* over *CCA1/LHY* in the feedback-loop in a potential oscillator in *Flaveria*.

LUX/ELF3/ELF4 is an evening-expressed complex of the circadian oscillator repressed by *CCA1/LHY* in *Arabidopsis* (Hazen et al., 2005; Murakami et al., 2007; Khan et al., 2010; Farré and Weise, 2012). The expression of *LUX/PCL1* from *F. pringlei* and *F. trinervia* increased gradually upon dawn, consistent to the diurnal pattern of *Arabidopsis*. Indicating, that the feedback regulation between the *LUX* and *CCA1/LHY* may be consistent between *Flaveria* and *Arabidopsis*. However, a photosynthetic-type specific expressions was found for *ELF3* between *F. pringlei* and *F. trinervia* and for *ELF4* between *Flaveria* and *Arabidopsis.*

GI is part of the photoperiodic flowering pathway of *Arabidopsis* and stabilizes the transcription of the clock gene *ZTL*, which in turn inhibits *TOC1* (Kim et al., 2007). The diurnal pattern of expression of *GI* from *Flaveria* was similar among tomato, maize, *Arabidopsis* with a peak at the day/night transition (Facella et al., 2008; Khan et al., 2010). Potentially, *GI* has an indirect effect over *TOC1* in *Flaveria* which remains to be demonstrated. However, *ZTL* from *Flaveria* was not rhythmic in the diurnal cycle, consistent to *ZTL* from rice (Murakami et al., 2007).

PHOT1 is required for photosynthetic-associated responses in *Arabidopsis* such as stomatal opening, chloroplast accumulation and phototropism (Sullivan et al., 2008) and circadian-regulated in *Arabidopsis* and maize (Khan et al., 2010). These physiological responses may be also be diurnally regulated by *PHOT1* in *Flaveria*.

Conserved diurnal pattern of genes of the carbon and nitrogen metabolism among *Flaveria* and higher plants

The partitioning of sucrose and starch synthesis to either the mesophyll or the bundle-sheath cells is a consequence of C_4 evolution (Leonardos, 2000; Weise et al., 2011). The changes in transcript abundances of major genes related to starch/sucrose synthesis/degradation between *F. pringlei* and *F. trinervia* were not significant, but highly co-expressed at similar diurnal time points [**Figure 17**]. This study emphasises, that the regulation of the starch/sucrose metabolism in *Flaveria* is controlled by diurnal transcription of the respective genes. In *Flaveria* numerous genes involved in starch degradation peaked by the start of the night, which is consistent to the diurnal pattern of those from *Arabidopsis* (Smith and Stitt, 2007). Whereas, *SS2* for synthesis of starch peaked around noon and was low during midnight in *Flaveria* [**Figure 16F**], consistent to the peak in phase of those from *Arabidopsis* and maize (Khan et al., 2000). The activity of the sucrose-phosphate synthase (SPS) protein required for the synthesis of sucrose is higher in C_4 species than in C_3 species (Leonardos et al., 2000). The transcript abundances of *SPS* were similar between *F. pringlei* and *F. trinervia*, and not diurnally regulated (data not shown). Therefore, the expression of *SPS* may not contribute significantly to the high and fast rate of exported photoassimilates in C_4 *Flaveria*.

Enzymes involved in the nitrate assimilation have a high activity during the morning. Nitrate levels in turn affect the diurnal changes in gene expression of the carbon assimilatory pathways for starch and sucrose synthesis. The transcript levels of the nitrate reductase gene from *Arabidopsis* correlated to the levels of nitrate and were inverse correlated to the levels of glutamine (Scheible et al., 2004). *NIA2* was significantly oscillating in *F. pringlei* and *F. trinervia* with high levels at dawn consistent to that from *Arabidopsis* (Scheible et al., 1997/2004) and [**Figure 16M**]. It remains an open question weather *NIA2* is a nitrate-responsive gene in *Flaveria.*

OUTLOOK

The gene expression in *F. pringlei* and *F. trinervia* under entrained diurnal day/night conditions needs to be extended further by a free-running (light) experiments to analyze the circadian regulation of genes of the transcriptome from the genus *Flaveria*. To ascertain the conclusions statistical-based analyses are required to identify the percentage of genes of the transcriptome from *F. pringlei* and *F. trinervia* that is governed by the diurnal/circadian system, including the percentage of diurnally regulated photosynthetic-related genes. Further, in order to avoid species-specific differences it is important to perform the diurnal and circadian experiments with various C_3 and C_4 *Flaveria* species.

MATERIALS AND METHODS

Plant material and RNA extraction

Plants were grown at greenhouse conditions and transferred for acclimation for 10 days in an open phytochamber at 120 µmol quanta m^{-2} s^{-1} irradiance at 19°C (night) and 22°C (day). Leaf blades were harvested from wild-type 4 to 6 week old *F. trinervia* and *F. pringlei* from the top of the plant. Leaf samples were collected every 4 hours over a 10-h (light)/14-h (dark) diurnal cycle. Plant material was immediately frozen in liquid nitrogen for further storage at –80 °C. Three separate biological replicates for each time point were ground in liquid nitrogen prior to RNA extraction. Total RNA was extracted from *Flaveria* leaves using TRIsure Reagent according to the protocol for plant tissue (Bioline) transcriptome sequencing. Digestion with DNAse was performed for 15 min. The RNA was treated with phenol and chloroform and precipitated over night with a sodium acetate/ isopropyl alcohol solution. The RNA was washed with 70 % ethanol and dissolved in H_2O. The RNA concentration was determined with the nano-drop (PEQLAB Biotechnologie GmbH). RNA aliquots was validated for qualification and quantitation with the Agilent2100 bioanalyzer (Agilent Technologies) and with agarose gel electrophoresis.

Library construction and Illumina sequencing

Total RNA was set to a final concentration of 1 µg in 50 µL as starting volume. The DNA libraries were generated according to the manufacturer´s TruSeq™ RNA Sample Preparation Kit via the Low-Throughput Protocol (Illumina Proprietary Catalog # RS-930-2001, Part # 15008136 Rev. A, November 2010). DNA libraries were validated for qualification and quantitation with the Agilent2100 bioanalyzer (Agilent Technologies). Validated DNA libraries were pooled to a final concentration of 2 nM. Clusters were generated with the TruSeq SR Cluster Kit v2 according to the Reagent Preparation Guide with the Illumina cBot device. The single read sequencing was performed with the Illumina HiSeq2000.

Transcriptome mapping and analysis

To count the number of unambiguous best hits per transcript the clean Illumina reads from *F. trinervia* and *F. pringlei* were aligned against *F. robusta* to avoid species-related errors. Reads were mapped with Bowtie2 (version 2.0.6.) using the end to end mode with sensitive settings and allowing one mismatch per seed for multiseeding (Langmead and Salzberg, 2012). The transcript best hits values were normalized to reads per million mapped reads and kb transcript length (RPMK).

Statistical analysis

The Pearson´s correlation coefficient was calculated by a pair-wise *F. pringlei* and *F. trinervia* comparison. For each species the average of normalized RPMK from three biological replicates of the RNA-seq data set was used. In figure 8- 18, the means and standard error of the mean (SEM) was calculated from three biological replicates of normalized RPMK.

The data were evaluated with a one-way ANOVA test by a multiple comparison using a Tukey-test (multiplicity adjusted $p \leq 0.05$). The significance and the ratio for transcripts of two time points with major differences in abundance was calculated. Statistical analysis was performed with the program GraphPad PRISM 6 (GraphPad Software, Inc., La Jolla, CA).

Cluster generation and data processing

Normalized read counts (of the average of three biological replicates for each detected gene from *F. pringlei* and *F. trinervia*) were $\log_2$ transformed and used as input for clustering, principal component analysis (PCA) and MapMan analysis. Clustering and the principal component analysis (PCA) was performed with TIGR MultiExperiment Viewer (MeV) http://www.tigr.org/software/tm4/mev.html (Saeed et al., 2003). KMC (K-means clustering) modules was used with user-specified parameters considered as input the mean and the number of 10 clusters after 50 iterations to run. The HCL module was used to generate 200 clusters. For both modules the Euclidean parameter was considered to estimate the distances of closest neighbors. The transcript abundances of expressed genes were analyzed for their biological function according to the annotation of BINs from *Arabidopsis* with the software MapMan and PageMan (Thimm et al., 2004; Usadel et al., 2005). The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction are included statistical parameters of the MapMan and PageMan programs and were used to reveal the significance of the BIN assignment to a particular category.

REFERENCES

Aubry, S, Brown, N.J. and **Hibberd, J.M.** (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *J. Exp. Bot.* **62:** 3049–3059.

Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Mas, P. and **Kay, S.A.** (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* **293:** 880–883.

Alvarez, C.E., Saigo, M., Margarit, E., Andreo, C.S. and **Drincovich, M.F.** (2013). Kinetics and functional diversity among the five members of the NADP-malic enzyme family from *Zea mays*, a C_4 species. *Photosynth Res.* **115**: 65–80.

Bancos, S., Szatmari, A.M., Castle, J., Kozma-Bognar, L., Shibata, K., et al. (2006). Diurnal regulation of the brassinosteroid-biosynthetic *CPD* gene in *Arabidopsis*. *Plant Physiol.* **141**: 299–309.

Baldy, P., Jacquot, J.P., Lavergne, D. and **Champigny, M.L**. (1989). Corn phosphoglycolate phosphatase: modulation of activity by pyridine-nucleotides and adenylate energy-charge. *Photosynth Res* **22**: 147–155.

Barajas-López, J.d.D., Serrato, A.J., Cazalis, R., Meyer, Y., Chueca, A., Reichheld, J.P. and **Sahrawy, M.** (2011). Circadian regulation of chloroplastic f and m thioredoxins through control of the CCA1 transcription factor. *Journal of Exp. Bot.* **62**: 2039–2051.

Bauer, D., Viczián, A. Kircher, S., Nobis, T., Nitschke, R., Kunkel, T., Panigrahi, KCS., Ádam, E., Fejes, E., Schäfer, E. and **Nagby, F.** (2004). Constitutive Photomorphogenesis 1 and Multiple Photoreceptors Control Degradation of Phytochrome Interacting Factor 3, a Transcription Factor Required for Light Signaling in *Arabidopsis*. *Plant Cell* **16**: 1433–1445.

Bläsing, O.E., Westhoff, P. and **Svensson, P.** (2000). Evolution of C_4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. *J Biol Chem* **275:** 27917-27923.

Bläsing O.E., Gibon, Y., Günther, M., Höhne, M., Morcuende, R., Osuna, D., Thimm, O., Usadel., B., Scheible, W.-R. and **Sitt, M.** (2005). Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *The Plant Cell* **17**: 3257–3281.

Blázquez, M.A., Trénor, M. and **Weigel, D.** (2002). Independent Control of Gibberellin Biosynthesis and Flowering Time by the Circadian Clock in *Arabidopsis*. *Plant Physiology.* **130**: 1770–1775.

Bolwell, G.P., Bozak, K. and **Zimmerlin, A.** (1994). Plant cytochrome p450. *Phytochemistry* **37**: 1491–506.

Bräutigam, A. and **Gowik, U**. (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol.* **12**: 831–841.

Bräutigam, A., Kajala, K. Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. W., Carr, K. M., Gowik, U., Maß, J., Lercher, M.J., Westhoff, P., hibberd, J.M. and **Wever, A.P.M.** (2011). An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiol*. **155**: 142–156.

Brown, N.J., Palmer, B.G., Stanley, S., Hajaji, H., Janacek, S.H., Astley, H.M., Parsley, K., Kajala, K., Qick, W.P., Trenkamp, S., Fernie, A.R., Maurion, V.G. and **Hibberd, J.M.** (2010). C_4 acids decarboxylases required for C_4 photosynthesis are active in the mid-vein of the C_3 species *Arabidopsis thaliana*, and are important in sugar and amino acid metabolism. *The Plant Journal* **61**: 122–133.

Buchanan, B.B. (1980). Role of light in the regulation of chloroplast enzymes. *Annual Review of Plant Physiology* **31**: 341-374.

Buchanan, B.B. (1991). Regulation of CO_2 assimilation in oxygenic photosynthesis: the ferredoxin thioredoxin system. Perspective on its discovery, present status, and future development. *Arch. Biochem. Biophys.* **288**: 1-9.

Choiu, T.J. and **Bush, D.R.** (1998). Sucrose is a signal molecules in assimilate partitioning. *Proc. Natl. Acad. Sci. USA*. **95**: 4784–4788.

Christensen, S. and **Silverthorne, J.** (2001). Origins of Phytochrome-Modulated *Lhcb* mRNA Expression in Seed Plants. *Plant Physiology* **126**: 1609–1618.

Christin, P.-A., Boxall, S.F., gregory, R., Edwards, E.J., Hartwell, J. and **Osborne, C.P.** (2013). Parallel recruitment of multiple genes into C_4 photosynthesis. *Genome Biology and Evolution*.

Coruzzi, G. and **Bush, D.R.** (2001). Nitrogen and carbon nutrient and metabolite signaling in plants. *Plant Physiol.* **125**: 61–64.

Covington, M.F. and **Harmer, S.L**. (2007). The circadian clock regulates auxin signaling and responses in *Arabidopsis*. *PLoS Bio*l: 5:e222.

Deeken, R., Sanders, C, Ache, P. and **Hedrich, R.** (2001). Developmental and light-dependent regulation of a phloem-localised K+ channel of *Arabidopsis thaliana*. *The Plant Journal.* 23: 285–290.

Drincovich, M.F., Casati, P. and **Andreo, C.S.** (2000). NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. *FEBS Letters* **490**: 1–6.

Ehleringer, J.R., Cerling, T.E. and **Helliker, B.R.** (1997). C_4 photosynthesis, atmospheric CO_2 and climate. *Oecologia* **112**: 285–299.

Engelmann, S., Bläsing, O.E., Westhoff, P. and **Svensson, P.** (2002). Serine 774 and amino acids 296 to 437 comprise the major C_4 determinants of the C_4 phosphoenolpyruvate carboxylase of *Flaveria trinervia*. *FEBS Lett* **524:** 11–14.

Ernst, D., Apfelbock, A., Bergmann, A., and Weyrauch, C. (1990). Rhythmic regulation of the light-harvesting chlorophyll a/b protein and the small subunit of ribulose-1,5-bisphosphate carboxylase mRNA in rye seedlings. Photochem. *Photobiol.* **52:** 29–33.

Facella, P., Lopez, L., Carbone, F., Galbraith, D.W., Giuliano, G., and **Perrotta, G.** (2008). Diurnal and circadian rhythms in the tomato transcriptome and their modulation by cryptochrome photoreceptors. PLoS One **3,** e2798.

Farré, E.M., Stacey, L.H., Harmon, F.G, Yanovsky, M.J. and **Kay, S.A.** (2005). Overlapping and Distinct Roles of *PRR7* and *PRR9* in the *Arabidopsis* Circadian Clock. *Current Biology* **15:** 47–54.

Farré E.M. and **Weise S.E.** (2012). The interactions between the circadian clock and primary metabolism. *Current Opinion in Plant Biology* **15**: 293–300.

Feraud, M., Masclaux-Daubresse, C., Ferrario-Méry, S., Pageau, K., Lelandais, :, Ziegler, C., Leboeuf, E., Jouglet, T., Viret, L., Spampinato, A., Paganelli, V., Hammouda, M.B. and **Suzuki, A**. (2005). Expression of a ferredoxin-dependent glutamate synthase gene in mesophyll and vascular cells and fucntions of the enzyme in ammonium assimilation in *Nicotiana tabacum* (L.). **4**: 667–677.

Filichkin, S.A., Breton, G., Priest, H.D., Dharmawardhana, P., Jaiswal, P., Fox, S.E., Michael, T.P., Chory, J., Kay, S.A. and **Mockler, T.C.** (2011). Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PLoS ONE* **6**:e16907.

Fluhr, R., Kuhlemeier, C., Nagy, F. and **Chua, N.H.** (1986). Organ-specific and light-induced expression of plant genes. *Science* **232**: 1106–1112.

Furbank, R.T. and T**aylor, W.C.** (1995). Regulation of Photosynthesis in C_3 and C_4 Plants: A Molecular Approach. *The Plant Cell* **7**: 797–807

Geiger, D.R. and **Servaites, J.C.** (1994). Diurnal regulation of photosynthetic carbon metabolism in C_3 plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **45**: 235–256.

Giuliano G., Hoffman, N.E., Ko, K., Scolnik, P.A. and **Cashmore, A.R.** (1988). A light-entrained circadian clock controls transcription of several plant genes. *EMBO J* **7**: 3635–3642.

Gould, P.D., Locke, J.C, Larue, C., Southern, M.M., Davis, S.J., Hanano, S., Moyle, R. Milich, R., Putterill, J. and **Millar, A.J.** (2006). The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. *Plant Cell* **18:** 1177–1187.

Gowik, U. and **Westhoff, P.** (2011). The Path from C_3 to C_4 photosynthesis. *Plant Physiol.* **155**: 56–63.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M. and **Westhoff, P.** (2011). Evolution of C_4 Photosynthesis in the genus *Flaveria*: How Many and Which Genes Does It Take to Make C_4? *The Plant Cell* **23**: 2087–2105.

Guo, B., Irigoyen, S., Fowler, T.B and **Versaw, W.K.** (2008). Differential expression and phylogenetic analysis suggest specialization of plastid-localized members of the PHT4 phosphate transporter family for photosynthetic and heterotrophic tissues. *Plant Signaling & Behaviour* **3**: 784–790.

Gutierrez, R.A., Stokes, T.L., Thum, K., Xu, X. and **Obertello, M.** (2008). Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. *Proc. Natl. Acad. Sci. USA* **105**: 4939–4944.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H-S., Han, B, Zhu, T., Wang, X., Kreps, J.A. and **Kay, S.A.** (2000). Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**: 2110–2113.

Harmer, S.L. (2009). The circadian system in higher plants. *Annu. Rev. Plant Biol.* **60**: 357–377.

Hazen, S.P., Schultz, T.F., Pruneda-Paz, J.L., Borevitz, J.O., Ecker, J.R and **Kay, S.A.** (2005). LUX ARRHYTHMO encodes a Myb domain protein esential for circadian rhythms. *PNAS*. **102**: 10387–10392.

Heckmann, D., Schulze, S, Denton, A., Gowik, U, Westhoff, P, Weber, A.P. and **Lercher, M.J.** (2013). Predicting C_4 Photosynthesis Evolution: Modular, individually adaptive steps on a Mount Fuji Fitness Landscape. *Cell* **153**: 1579–1588.

Hicks, K.A., Albertson, T.M. and **Wagner, D.Ry** (2001). EARLY FLOWERING 3 encodes a novel protein that regulates circadian clock function and flowering in *Arabidopsis*. *The Plant Cell* **13**: 1281–1292

Hirose, T., Aoki, N., Harada, Y., Okamura, M., Hashida, Y., Ohsugi, R., Miyao, A., Hirochika, H. and **Terao, T**. (2013). Disruption of a rice gene for α-glucan water dikinase, OsGWD1, leads to hyperaccumulation of starch in leaves but exhibitslimited effects on growth. *Front. Plant Sci.* 10.3389

Horst, I., **Offermann, S., Dreesen, B., Niessen, M.** and **Peterhansel, C**. (2009). Core promotor acetylation is not required for high transcription from the phospho*enol*pyruvate carboxylase promotor in maize. *Epigenetics and Chromatin* **2**: 17.

Hudson, K.A. (2010). The circadian clock-controlled transcriptome of developing soybean seeds. *The Plant Genome* **3**: 3–13.

Ishikawa, M., Kiba, T. and Chua, N.H. (2006). The *Arabidopsis SPA1* gene is required for circadian clock function and photoperiodic flowering. *Plant J.* **46**: 736–746.

Jouve, L., Gaspar, T., Kevers, C., Greppin, H. and **Degli Agosti, R**. (1999). Involvement of indole-3-acetic acid in the circadian growth of the first internode of *Arabidopsis*. *Planta* **209**:136–142.

Kaldis, A.D., Kousidis, P., Kesanopoulos, K. and **Prombona, A.** (2003). Light and circadian regulation in the expression of *LHY* and *Lhcb* genes in *Phaseolus vulgaris*. *Plant Mol. Biol.* **52**: 981–997.

Kawaguchi, R., Girke, T., Bray, E.A. and **Bailey-Serres, J.** (2004). Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant J.* **38**: 823–839.

Kellmann, J.-W., Merforth, N., Wiese, M., Pichersky, E. and **Piechulla, B.** (1993). Concerted circadian oscillations in transcript levels of nineteen *Lha/b* (*cab*) genes in *Lycopersicon esculentum* (tomato). *Mol. Gen. Genet.* **237**: 439–448.

Kent, W.J. (2002). BLAT-The BLAST-like alignment tool. Genome Res. **12**: 656–664.

Khan, S., Rowe, S.C. and **Harmon, F.G.** (2010). Coordination of the maize transcriptome by a conserved circadian clock. *BMC Plant Biology* **10**: 126.

Kim, W.Y., Fujiwara, S., Suh, S.S., Kim, J., Kim, Y., Han, L., David, K., Putterill, J., Nam, H.G. and **Somers, D.E.** (2007). ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nature* **449**: 356–360.

Krapp, A., Hofmann, B., Schäfer, C. and **Stitt, M.** (1993). Regulation of the expression of *rbcS* and other photosynthetic genes by carbohydrates: a mechanism for the ´sink regulation´ of photosynthesis. *The Plant Journal* **3**: 817–828.

Kramer, D.M. and **Evans, J.R.** (2011). The Importance of Energy Balance in Improving Photosynthetic Productivity. *Plant Physiology* **155**: 70–78.

Ku, M.S.B., Schmitt, M.R. and **Edwards, G.E.** (1979). Quantitative determination of RuBP carboxylase-oxygenase protein in leaves of several C_3 and C_4 plants. *J Exp Bot.* **114:** 89–98.

Kuhlemeier, C., Green, P. J. and **Chua, N.-H.** (1987). Regulation of gene expression in higher plants. *Annu. Rev. Plant Physiol.* **38:** 221–257.

Lahiri, S.D., Yao, J., McCumbers, C. and **Allison, L.A.** (1999). Tissue-specific and light-dependent expression within a family of nuclear-encoded sigma-like factors from *Zea mays*. *Mol Cell Biol. Res. Commun.* **1**: 14–20.

Lai, L.B., Tausta, L. and **Nelson, T.M.** (2002). Differential Regulation of Transcripts Encoding Cytosolic NADP-Malic Enzyme in C_3 and C_4 *Flaveria* species. *Plant Physiology* **128**: 140–149.

Langmead, B. and **Salzberg, S.L**. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357–359.

Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X. and **Chang, W.** (2004). Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* **428**: 287–292.

Leonardos, E.D. and **Grodzinski, B.** (2000). Photosynthesis, immediate export and carbon partitioning in source leaves of C_3, C_3-C_4 intermediate and C_4 *Panicum* and *Flaveria* species at ambient and elevated CO_2 levels. *Plant, Cell and Environment* 23, 839–851.

Leonardos. E.D., Micallef, B.J., Carreno Micallef, M. and **Grodzinski, B.** (2006). Diel patterns of leaf C export and of main shoot growth for *Flaveria linearis* with altered leaf sucrose-starch partitioning. *Journal of Experimental Botany* **57**: 801–814.

Leu, W.-M., Cao, X.-L., Wilson, T.J., Snusstad, P. and **Chua, N.-H.** (1995). Phytochrome A and Phytochrome B Mediate the Hypocotyl-Specific Downregulation of *TUB1* by light in *Arabidopsis*. *The Plant Cell*. **7**: 2187–2196.

Lu, Y., Gehan, J.P. and **Sharkey, T.D.** (2005). Daylength and circadian effects on starch degradation and maltose metabolism. *Plant Physiology* **138**: 2280–2291.

Majeran, W., Friso, G., Ponnala, L., et al. (2010). Structural and metabolic transitions of C_4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. *The Plant Cell* **22**: 3509–3542.

Makino, S., Matsushika, A., Kojima, M., Yamashino, T. and **Mizuno, T.** (2002). The APRR1/TOC1 quintet implicated in circadian rhythms of *Arabidopsis thaliana*: I. Characterization with APRR1-overexpressing plants. *Plant Cell Physiol.* 43: 58–69.

Marri, L., Sparla, F., Pupillo, P. and Trost, P. (2004). Co-ordinated gene expression of photosynthetic glyceraldehyde-3-phosphate dehydrogenase, phosphoribulokinase, and CP12 in *Arabidopsis thaliana*. *Journal of Exp. Bot.* **56**: 73–80.

Marshall, J.S., Stubbs, J.D. and **Taylor, W.C.** (1996). Two genes encode highly similar chloroplastic NADP-malic enzyme enzymes in *Flaveria*. Implications for the evolution of C_4 photosynthesis. *Plant Physiology* **111**: 1251–1261

Martinez-Garcia, JF, Huq, E. and **Quail, PH.** (2000). Direct targeting of light signals to a promotor element-bound transcription factor. *Science* **288**: 859–863.

Maier, A., Zell, M.B. and **Maurino, V.G.** (2011). Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C_4 and C_3 photosynthesis. *J. Exp. Bot.* **62**: 3061–3069.

McClung, C.R., Hsu, M., Painter, J.E., Gagne, J.M., Karlsberg, S.D. and **Salome, P.A.** (2000). Integrated temporal regulation of the photorespiratory pathway. Circadian regulation of two *Arabidopsis* genes encoding serine hydroxymethyltransferase. *Plant Physiol* **123**: 381–392.

Mckown, A.D., Moncalvo, J.-M. and **Dengler, N.G.** (2005). Phylogeny of *Flaveria* (Asteracea) and interference of C_4 photosynthesis evolution. *American Journal of Botany* **32**: 1911–1928.

Meyer, H., Theinel, U. and **Piechulla, B.** (1989). Molecular characterization of the diurnal/circadian expression of the chlorophyll *a*/*b*-binding proteins in leaves of tomato and other dicotyledonous and monocotyledonous plant species. *Planta* **180**: 5–15.

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., trout, J.D., Hazen, S.P., Shen, R., Priest, H.D., Sullivan, C.M., Givan, S.A., Yanovsky, M., Hong, F., Kay, S.A. and **Chory, J.** (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLOS genetics.* **4**: e14

Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., McEntee, C., Kay, S.A. and **Chory, J**. (2007). The diurnal project: diurnal and circadian expression profiling, model-based pattern matching, and promotor analysis. *Cold Spring Harbor Laboratory Press*. **72**: 353–363.

Motchoulski, A. and **Liscum, E.** (1999). *Arabidopsis* NPH3: A NPH1 photoreceptor-interacting protein essential for phototropism. *Science*. **29**: 961–964.

Munekage, Y.N., Eymery, F., Rumeau, D., Yokota, A., Genty, B. and **Peltier, G.** (2007). Localization of PGR5 and NDH in leaf tissues of *Flaveria* C_3 or C_4 species. *Plant Cell Physiol* **48**: S95

Munekage, Y.N., Eymery, F., Rumeau, D., Cuiné, S., Oguri, M., Nakamura, N., Yokota, A., Genty, B. and **Peltier, G.** (2010). Elevated expression of PGR5 and NDH-H in bundle sheath chloroplasts in C_4 *Flaveria* species.

Murakami, M., Tago, Y., Yamashino, T. and **Mizuno, T.** (2007). Comparative overviews of clock-associated genes of *Arabidopsis thaliana* and *Oryza sativa*. *Plant Cell Physiol* **48**: 110–121.

Nusinow, D.A, Helfer, A., Hamilton, E.E., King, J.J., Imaizumi, T., Schultz, T.F., Farré, E.M. and **Kay, S.A.** (2011). The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature* **475**: 398–404.

Oberschmidt, O., Hücking, C. and **Piechulla, B.** (1995). Diurnal *Lhc* gene expression is present in many but not all species of plant kingdom. *Plant Mol. Biol.* **27**: 147–153.

Okamoto, M., Kuwahara, A., Seo, M., Kushiro, M., Asami, T., Hirai, N., Kamiya, Y., Koshiba, T., and **Nambara, E.** (2006). CYP707A1 and CYP707A2, Which Encode Abscisic Acid 8′-Hydroxylases, Are Indispensable for Proper Control of Seed Dormancy and Germination in *Arabidopsis*. *Plant Physiology* **141**: 97–107

Otter, T., Penther, J.M. and **Mohr, H.** (1992). Control of the appeareance of alanine aminotransferase in the Scots pine (*Pinus sylvestris L.*) seedling. *Planta* **188**: 376–383.

Paulus, J.K., Schlieper, D. and **Groth, G.** (2013). Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat Commun* **4:** 1518.

Peterhänsel, C. and **Maurino, V. G.** (2011). Photorespiration redesigned. *Plant Physiol.***155**: 49.

Pokhilko, A., Fernández, A.P., Edwards, K.D, Southern, M.M., Halliday, K.J. and **Millar, A.J.** (2012). The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback-loops. *Molecular Systems Biology* **574:** 1–13.

Portis, Ar.R., Salvucci, JR.M.E. and **Ogren, W.L.** (1986). Activation of Ribulosebiphospahte carboxylase/ oxygenase at physiological CO_2 and ribulosebiphosphate concentrations by Rubisco activase. *Plant Physiol.* **82**: 967–971

Price, J., Laxmi, A., St Martin, S.K. and **Jang, J.C.** (2004). Global transcription profiling reveals multiple sugar signal transduction mechanisms in *Arabidopsis*. *Plant Cell* **16**: 2128–2150.

Raines, C.A., Longstaff, M., Lloyd, J.C. and **Dyer, T.A**. (1989). Complete coding sequence of phosphoribulokinase: developmental and light- dependent expression of the mRNA. *Molecular and General Genetics* **220**: 43–48.

Rausch, C. and **Bucher, M**. (2002). Molecular mechanisms of phosphate transport in plants. *Planta* **216**: 23–37.

Raychaudhuri, S., Stuart, J.M. and **Altman, R.B.** (2000). Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pacific Symposium on Biocomputing*, Honolulu, Hawaii, 452–463.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T. Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and **Quackenush, J.** (2003). TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* **34**: 374–378.

Sage, R.F., Pearcy, R.W. and **Seemann, J.R.** (1987). The nitrogen use efficiency of C_3 and C_4 plants. III. Leaf nitrogen effects on the activity of carboxylating enzymes in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). *Plant Physiol.* **85:** 355–359.

Sage, R.F. and **McKown, A.D.** (2005). Is C_4 photosynthesis less phenotypically plastic than C_3 photosynthesis? *Journal of Exp. Botany* **57**: 303–317.

Sage, R.F., Sage, T.L. and **Kocacinar, F.** (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annu. Rev. Plant Biol.* **63**: 19–47.

Samol, I., Shapiguzov, A., Ingelsson, B., Fucile, G., Crèvecoeur, M., Vener, A.V., Rochaix, J.-D. and **Goldschmidt-Clermont, M.** (2012). Identification of a photosystem II phosphatase involved in light acclimation in *Arabidopsis*. *Plant Cell* **24**: 2596–2609.

Schaffer, R., Ramsay, N., Samach, A., Cordon, S., Putterill, J., Carré, I.A. and **Coupland, G.** (1998). The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**: 1219–1229.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M. and Wisman, E. (2001). Microarray Analysis of Diurnal and Circadian-Regulated Genes in *Arabidopsis*. *The Plant Cell* **13**: 113–123.

Scheible, W.-R., Gonzales-Fontes, A., Morcuende, R., Lauerer, M., Geiger, M., Glaab, J., Schulze, E.D. and **Stitt, M.** (1997). Tobacco mutants with a decreased number of functional nia-genes compensate by modifying the diurnal regulation transcription, post-translational modification and turnover of nitrate reductase. *Planta* **203**: 305–319.

Scheible, W.-R., Morcuende, R., Czechowski, T., Osuna, D., Fritz, C., Palacios-Rojas, N., Thimm, O., Udvardi, M.K. and **Stitt, M.** (2004). Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes and signaling infrastructure by nitrogen. *Plant Physiol.* **136**: 2483–2499.

Schwarte, S. and **Bauwe, H.** (2007). Identification of the Photorespiratory 2- Phosphoglycolate Phosphatase, PGLP1, in *Arabidopsis*. *Plant Physiology* **144**: 1580–1586.

Schulz, B., Frommer, W.B., Flügge, U.I., Hummel, S., Fischer, K. and **Willmitzer, L**. (1993). Expression of the triose phosphate translocator gene from potato is light dependent and restricted to green tissues. *Mol. Gen. Genet.* **238**: 357–361.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M. Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and **Shinozaki, K.** (2002). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J*. **31**: 279–292.

Sharkey, T.D. (1985). Photosynthesis in Intact Leaves of C_3 plants: Physics, Physiological and Rate Limitations. *The Botanical Review*. **51**: 53–105.

Sheen, J. (1999). C_4 gene expression. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**: 187–217.

Shu, G., Pontieri, V., Dengler, N.G. and **Mets, L.J**. (1999). Light induction of cell type differentitaion and cell-type-specific gene expression in cotyledons of a C_4 plant *Flaveria trinervia*. *Plant Phys.* **121**: 731–741.

Smith, S.M., Fulton, D.C., Chia, T., Thorneycroft, D., Chapple, A., Dunstan, H., Hylton, C., Zeeman, S.C. and **Smith, A.M.** (2004). Diurnal Changes in the Transcriptome Encoding Enzymes of Starch Metabolism Provide Evidence for Both Transcriptional and Posttranscriptional Regulation of Starch Metabolism in *Arabidopsis* Leaves. *Plant Physiology.* **136**: 2687–2699.

Smith, A.M. and **Stitt, M.** (2007). Coordination of carbon supply and plant growth. *Plant, Cell and Environment* **30**: 1126–1149.

Stitt, M. and **Krapp, A.** (1999). The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. *Plant, Cell and Environment* **22**: 583–621.

Stitt, M., Lunn, J. and **Usadel, B.** (2010). *Arabidopsis* and primary photosynthetic metabolism – more than the icing on the cake. **61**: 1067–1091.

Sullivan, S., Thomson, C.E., Lamont, D.J., Jones, M.A. and **Christie, J.M.** (2008). *In vivo* phosphorylation site mapping and functional characterizaion of *Arabidopsis* phototropin 1. *Mol Plant.* **1**: 178–194.

Tan, X., Calderon-Villalobos, L.I., Sharon, M., Zheng, C., Robinson, C.V., et al. (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**: 640–645

Tanaka, R., Rothbart, M., Oka, S., Takabayashi, A., Takahashi, K., Shibata, M., Myouga, F., Motohashi, R., Shinozaki, K., Grimm, B. and **Tanaka, A.** (2010). LIL3, a light-harvesting-like protein, plays an essential role in chlorophyll and tocopherol biosynthesis. *Proc. Natl. Acad. Sci USA*. **107**: 16721–16725.

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

Thimm, O, Bläsing, O. Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.J. and **Stitt, M**. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J*. **37**: 914–939.

Thum, K.E., Shin, M.J., Palenchar, P.M., Kouranov, A. and **Coruzzi, G.M**. (2004). Genome-wide investigation of light and carbon signaling interactions in *Arabidopsis*. *Genome Biol.* **5**: R10.

Tyagi, A.K. and **Gaur, T.** (2010) Light Regulation of Nuclear Photosynthetic Genes in Higher Plants. *Critical Reviews in Plant Sciences* **22**: 417–452.

Usadel, B., Nagel, A., Thimm, O., Redestig, H., Bläsing, O.E., Palacios-Rojas, N., Selbig, J., Hannemann, J., Conceição Piques, M., Steinhauser, Scheible, W.-R., Gibon, Y., Morcuende, R., Weicht, D., Meyer, S. and **Mark Stitt** (2005). Extension of the Visualization Tool MapMan to Allow Statistical Analysis of Arrays, Display of Corresponding Genes, and Comparison with Known Responses. *Plant Phys.* **138**: 1195–1204.

Usadel, B., Nagel, A. Steinhauser, D., Gibon, Y., Bläsing, O.E., Redestig, H., Screenivasulu, N., Krall, L., Hannah, M.A., Poree, F., Fernie, A.R. and **Stitt, M.** (2006). PageMan: An interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC Bioinformatics* **7**: 535.

Wang, R., Guegler, K., LaBrie, S.T. and **Crawford, N.M.** (2000). Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell* **12**: 1491–1510.

Wang, Z.Y., Kenigsbuch, D., Sun, L., Harel, E., Ong, M.S. and **Tobin, E.M.** (1997). A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis Lhcb* gene. *Plant Cell* **9**: 491–507.

Wang, Z. Y. and **Tobin, E. M.** (1998). Constitutive expression of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**: 1207–1217.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J. and **Langdale, J.A.** (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. *Plant Cell* **21**: 1109–1128.

Wedel, N., **Soll, J.** and **Paap, B.K**. (1997). CP12 provides a new mode of light regulation of Calvin cycle activity in higher plants. *Proc. Natl. Acad. Sci. USA* **94**: 10479–10484.

Weise, S.E., Weber, A. and **Sharkey, T.D.** (2004). Maltose is the major form of carbon exported from the chloroplast at night. *Planta* **218**: 474–482.

Weise, S.E., van Weijk, K. and **Sharkey, T.D.** (2011). The role of transitory starch in C_3, CAM, and C_4 metabolism and opportunities for engineering leaf starch accumulation. *Journal of Experimental Botany* **62**: 3109–3118.

Yakir, E., Hilman, D., Harir, Y. and **Green, R.M.** (2006). Regulation of output from the plant circadian clock. *FEBS Journal.* **274**: 335–345.

Young, M.W. and **Kay, S.A**. (2001). Times zones: a comparative genetics of circadian clocks. *Nature* **2**: 702–715.

Xu, W., Yang, R.., Li, M., Xing, Z., Yang, W., Chen, G., Guo, H., Gong, X., Du, Z., Zhang, Z., Hu, X., Wang, D., Qian, Q., Wang, T., Su, Z. and **Xue, Y.** (2011). Transcriptome Phase Distribution Analysis Reveals Diurnal Regulated Biological Processes and Key Pathways in Rice Flag Leaves and Seedling Leaves. *PLoS ONE* **6**: e17613.

Zhong, S., Shi, H., Xue, C., Wang, L., Xi, Y., Li, J., Quail, P.H., Deng, X.W. and **Guo, H.** (2012). A molecular framework of light-controlled phytochrome action in *Arabidopsis. Curr. Biol.* **22**: 1530–1535.

Zhu, Sh., Yu, X.C., Wang, X.J., Zhao, R., Li, Y., Fan, R.C., Shang, Y., Du, S.Y., Wang, X.F., Wu, F.Q., Xu, Y.H., Zhang, X.Y. and **Zhang, D.P.** (2007). Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in *Arabidopsis*. *Plant Cell* **10**: 30193–3036.

LEGENDS

FIGURE LEGENDS

Figure 1. The relation in the global diurnal gene expression between *F. pringlei* and *F. trinervia*. Principal component analysis (PCA) evaluated the samples that contribute most to the variability within the entire RNAseq data set. The RNAseq data set corresponded to the combination of the diurnal transcriptome from *F. pringlei* (Fp) and *F. trinervia* (Ft). For each gene and each time point the average of the log_2 transformed normalized RPMK accounts of three biological replicates was used. PCA analysis was based on TIGR.

Figure 2. Comparison in the changes in gene expression in response to diurnal changes between *F. pringlei* and *F. trinervia*. 10 clusters were generated with the KMC modules showing the respective transcript profile of all detected genes from the transcriptome of *F. pringlei* and *F. trinervia*. The normalized RPMK accounts from each time point and each gene for the two species were log_2 transformed and combined to be processed with the TIGR MultiExperiment Viewer (MeV) program. Each cluster represents a set of genes that exhibit similar diurnal expression pattern. These were compared between *F. pringlei* (C_3) and *F. trinervia* (C_4). The horizontal bar shows the day (open) and night (closed) period length (10-h day/14-h night). Lights were turned on at 8 a.m. and off at 6 p.m. Measurements started at dawn at time point 0-h. Cluster 1 and 2: 16424 genes upregulated in *F. pringlei* relative to *F. trinervia*. Cluster 3 to 5: 11736 genes are upregulated in *F. trinervia* relative to *F. pringlei.*

Figure 3. Correlation in diurnal pattern of gene expression between *F. pringlei* and *F. trinervia*. The Pearson´s correlation analysis was calculated for all detected genes from the transcriptome of *F. pringlei* and *F. trinervia*, in a species pair-wise comparison. The Pearson correlation shows the trend or pattern in diurnal expression irrespectively to the overall absolute transcript abundance. The histogram represents the total number of transcripts assigned to a particular Pearson´s correlation coefficient (r). The Pearson´s correlation coefficient corresponds to the degree of correlation in diurnal gene expression between the species. $0.75 \leq r < 1$ (strong to perfect correlation), $0.25 \leq r < 0.75$ (correlation), $0 \leq r < 0.25$ (weak correlation), $-0.25 \leq r \leq 0$ (weak inverse correlation), $-0.75 \leq r < -0.25$ (inverse correlation), $-1 \leq r \leq -0.75$ (strong to perfect inverse correlation).

Figure 4. Functional classes of genes from *F. pringlei* and *F. trinervia* linked to diurnal gene expression. Identification of over-represented functional classes in association to a particular correlation coefficient. Overrepresented (blue) and underrepresented functional gene classes (red). Genes that surpass a fold-change in expression assigned to a particular functional class in comparison to the expression of all genes in the RNA seq data set, is considered as over-represented. In the inverse correlated group a high number of functional classes are over-represented, because the group has a low number of genes. The normalized RPMK accounts from each time point and each gene for both species were combined and processed with PageMan. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction are included statistical parameters of the PageMan programs and were used to reveal the significance of the BIN assignment to a particular category. Genes that surpass a fold-change in expression assigned to a particular functional class in comparison to the expression of all genes in the RNAseq data set, are considered as over-represented (blue). The p-values are set to z-values (Usadel et al., 2006). PS, photosynthesis, TCA, tricarboxylic acid cycle, CHO, carbohydrate; OPP, oxidative pentose phosphate.

Figure 5. Functional classes of genes from *F. pringlei* and *F. trinervia* linked to clusters. (**A**) 6 clusters within 200 clusters were selected that represent the diurnal expression pattern from *F. pringlei* compared to that of *F. trinervia*. The normalized RPMK accounts from each time point and each gene for the two species were log_2 transformed and combined to be processed with the TIGR MultiExperiment Viewer (MeV) program. Clusters were generated with Euclidean distances using the HCL module. White box (light period); Black box (dark period). (**B**) Identification of functional classes overrepresented in a particular cluster. The normalized RPMK accounts from each time point and each gene for both species were combined and processed with the PageMan program. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction are included

statistical parameters of the PageMan program and were used to reveal the significance of the BIN assignment to a particular category. Genes that surpass a fold-change in expression assigned to a particular functional class in comparison to the expression of all genes in the RNAseq data set, are considered as over-represented (blue). The p-values are set to z-values (Usadel et al., 2006). PS, photosynthesis, TCA, tricarboxylic acid cycle, CHO, carbohydrate; OPP, oxidative pentose phosphate.

Figure 6. Ratio of gene expression in *F. trinervia* to *F. pringlei*. The ratio of the transcript abundance at dawn (0-h), noon (4-h) and midnight (16-h) from *F. trinervia* to that of *F. pringlei*. The average of the normalized RPMK for each gene from three biological replicates between *F. pringlei* and *F. trinervia* was $\log_2$ transformed. The ratio was used as input in MapMan. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction were used for correction. Blue indicates for a higher expression of *F. trinervia* in respect to *F. pringlei*, whereas red indicates for a higher expression of *F. pringlei* in respect to *F. trinervia.* White squares indicate no significant differences between the species in gene expression. (1) PS. Light reaction. PSII. LHC-II; (2) PS. Light reaction. PSII. polypeptide subunit; (3) PS. Light reaction. ATP synthase; (4) PS. Light reaction. cytochrome b6/f; (5) PS. Light reaction.PSI-LHC-I; (6) PS. Light reaction. PSI polypeptide subunit; (7) PS. Light reaction. other electron carrier. ox/red. ferredoxin reductase; (8) PS. Light reaction. other electron carrier. ox/red. ferredoxin. (9) PS. calvin cycle. FBPase; (10) PS. calvin cycle. transketolase; (11) PS. calvin cycle. transketolase; (12) PS. calvin cycle. Rib5P isomerase; (13) PS. calvin cycle. PRK; (14) PS. calvin cycle. Rubisco large (left) small (right) subunits; (15) PS. calvin cycle. rubisco interacting; (16) PS. calvin cycle. GAP; (17) PS. photorespiration. hydroxypyruvate reductase; (18) PS. photorespiration. phosphoglycolate phosphatase; (19) PS. photorespiration. glycolate oxydase; (20) glycine cleavage T subunit; (21) PS. photorespiration. serine hydroxymethyl transferase.

Figure 7. Ratio of gene expression in *F. trinervia* to *F. pringlei*. The ratio of the transcript abundance at dawn (0-h), noon (4-h) and midnight (16-h) from *F. trinervia* to that of *F. pringlei*. The average of the normalized RPMK for each gene from three biological replicates between *F. pringlei* and *F. trinervia* was $\log_2$ transformed. The ratio was used as input in MapMan. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction were used for correction. Blue indicates for a higher expression of *F. trinervia* in respect to *F. pringlei*, whereas red indicates for a higher expression of *F. pringlei* in respect to *F. trinervia.* White squares indicate no significant differences between the species in gene expression.

Figure 8. Changes of transcript abundances of genes of the C_4 cycle during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). The reads per million mapped reads and kilobase transcript length (RPMK) is a measure of the transcript abundances. The average of RPMK values correspond to three biological replicates for each gene and species. The horizontal bar shows the day (open) and night (closed) period length (10-h day/14-h night). Lights were turned on at 8 a.m. and off at 6 p.m. Measurements started at dawn at time point 0-h. The harvest of leave samples continued repeatedly every 4-h along the 24-h cycle. The standard error of the mean (SEM) was calculated for each time course of three biological replicates. The significant differences between each time point were determined with the RPMK values using a one-way ANOVA (Tukey-test) $p \leq 0.05$ summarized in Table 1. (**A**) NADP-MDH (**B**) PPDK (**C**) CA1 (**D**) Ala-AT (**E**) Asp-AT (**F**) PPDK-RP (**G**) NADP-ME1 (**H**) NADP-ME4 (**I**) NAD-ME1 (**J**) NAD-ME2 (**K**) PCK (**L**) NAD-MDH.

Figure 9. Changes of transcript abundances of genes of the Calvin-Benson cycle during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 2. (**A**) CP12 (**B**) RUBISCO activase (**C**) PRK (**D**) fructose-1,6-biphosphatase family protein (**E**) D-ribulose-5-phosphate-3-epimerase.

Figure 10. Changes of transcript abundances of photorespiratory genes during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 3. (**A**) SHMT3 (**B**) SHM2 (**C**) Oxidoreductase family protein (**D**) PGLP1 (**E**) PGLP2 (**F**) GLDP2 (**G**) HPR.

Figure 11. Changes of transcript abundances of genes involved in the photosynthetic light reactions during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 4. (**A**) Lhcb2.2 (**B**) Lhcb6 (**C**) Lhcb4.2 (**D**) Lhb (cab) (**E**) LHC5 (**F**) NDF2 (**G**) NDF4 (**H**) PsbQ-like 1 (**I**) PPL2 (**J**) PSBTN (**K**) PSBY (**L**) STN7 (**M**) ferredoxin-related (**N**) thylakoid lumenal (**O**) ATP synthase (**P**) FTR1 (**Q**) TRXf1 (**R**) PGR5.

Figure 12. Changes of transcript abundances of circadian clock genes and clock-associated genes during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 5. (**A**) CCL (**B**) XCT (**C**) PRR3 (**D**) PRR5 (**E**) PRR7 (**F**) LUX/PCL1 (**G**) ZTL (**H**) CCA1 (**I**) TOC1 (**J**) LHY (**K**) LIP1 (**L**) PIF3 (**M**) TIC (**N**) GI (**O**) ELF3 (**P**) ELF4.

Figure 13. Changes of transcript abundances of light-regulated genes during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 6. (**A**) SPA3 (**B**) NPH3 (**C**) PHOT1 (**D**) PHOT2 (**E**) LIL3.

Figure 14. Changes of transcript abundances of genes involved in transport during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 7. (**A**) SULTR4 (**B**) HCO_3^- transporter family (**C**) PHT2;1 (**D**) PHT4;4 (**E**) PHF1; (**F**) AKT (**G**) PIP1C (**H**) NRT1.5 (**I**) STP13 (**J**) STP1 (**K**) GLT1 (**L**) DIT2.1 (**M**) BASS2

Figure 15. Changes of transcript abundances of genes encoding transcription factors during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 8. (**A**) Acyl-CoA N-acyltransferase (**B**) GLK1 (**C**) EIN3 (**D**) HB1 (**E**) SIG1 (**F**) SIG2 (**G**) SIG5 (**H**) SIG6.

Figure 16. Changes of transcript abundances of genes involved in hormone metabolism during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 9. (**A**) CYP707A1 (**B**) CYP88A3 (**C**) GAI (**D**) GASA6 (**E**) PIN7 (**F**) AFB5 (**G**) IAA9

Figure 17. Changes of transcript abundances of genes involved in carbon and nitrogen metabolism during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 10. (**A**) GWD1 (**B**) GWD3 (**C**) AMY3 (**D**) PHS2 (**E**) BAM1 (**F**) DPE2 (**G**) MEX1 (**H**) SBE2.2 (**I**) SS2 (**J**) SPP1 (**K**) FBPase (**L**) Pectin lyase (**M**) PGM1 (**N**) CESA3 (**O**) CESA5 (**P**) NIA2 (**Q**) GLU1 (**R**) GLN1;4 (**S**) ALDH11A3.

Figure 18: Changes of transcript abundances of genes involved in stress, protein post-translational modification, calcium signalling and in the secondary metabolism during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 11. (**A**) HSF4 (**B**) Protein kinase superfamily protein (**C**) DNAse-I-like (**D**) CPK1 (**E**) CPK11 (**F**) TUB1 (**G**) GGR (**H**) CCR1 (**I**) CAD5

FIGURES

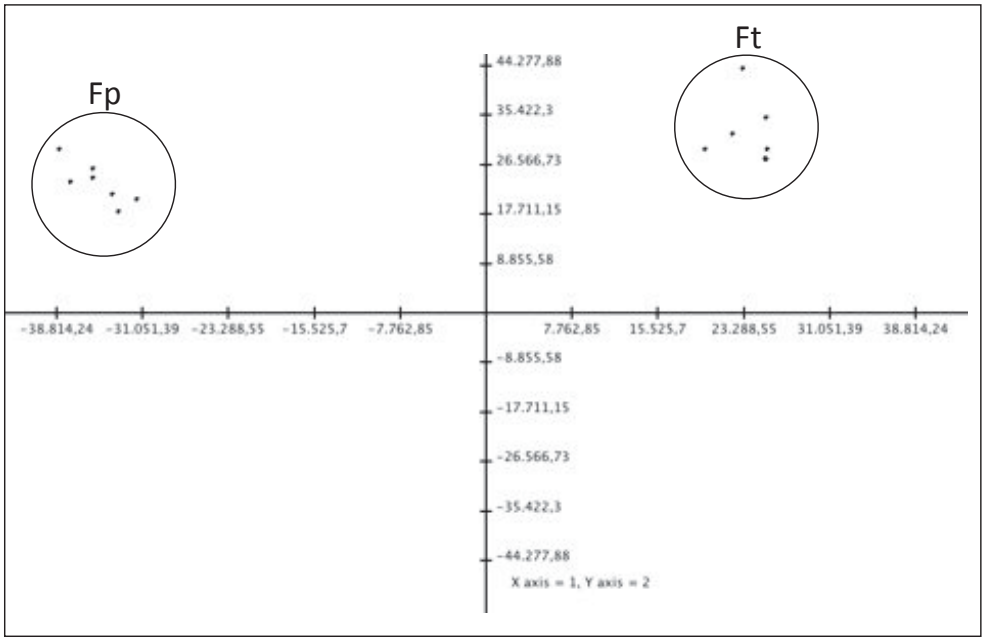


Figure 1. The relation in the global diurnal gene expression between *F. pringlei* and *F. trinervia*. Principal component analysis (PCA) evaluated the samples that contribute most to the variability within the entire RNAseq data set. The RNAseq data set corresponded to the combination of the diurnal transcriptome from *F. pringlei* (Fp) and *F. trinervia* (Ft). For each gene and each time point the average of the $\log_2$ transformed normalized RPMK accounts of three biological replicates was used. PCA analysis was based on TIGR.

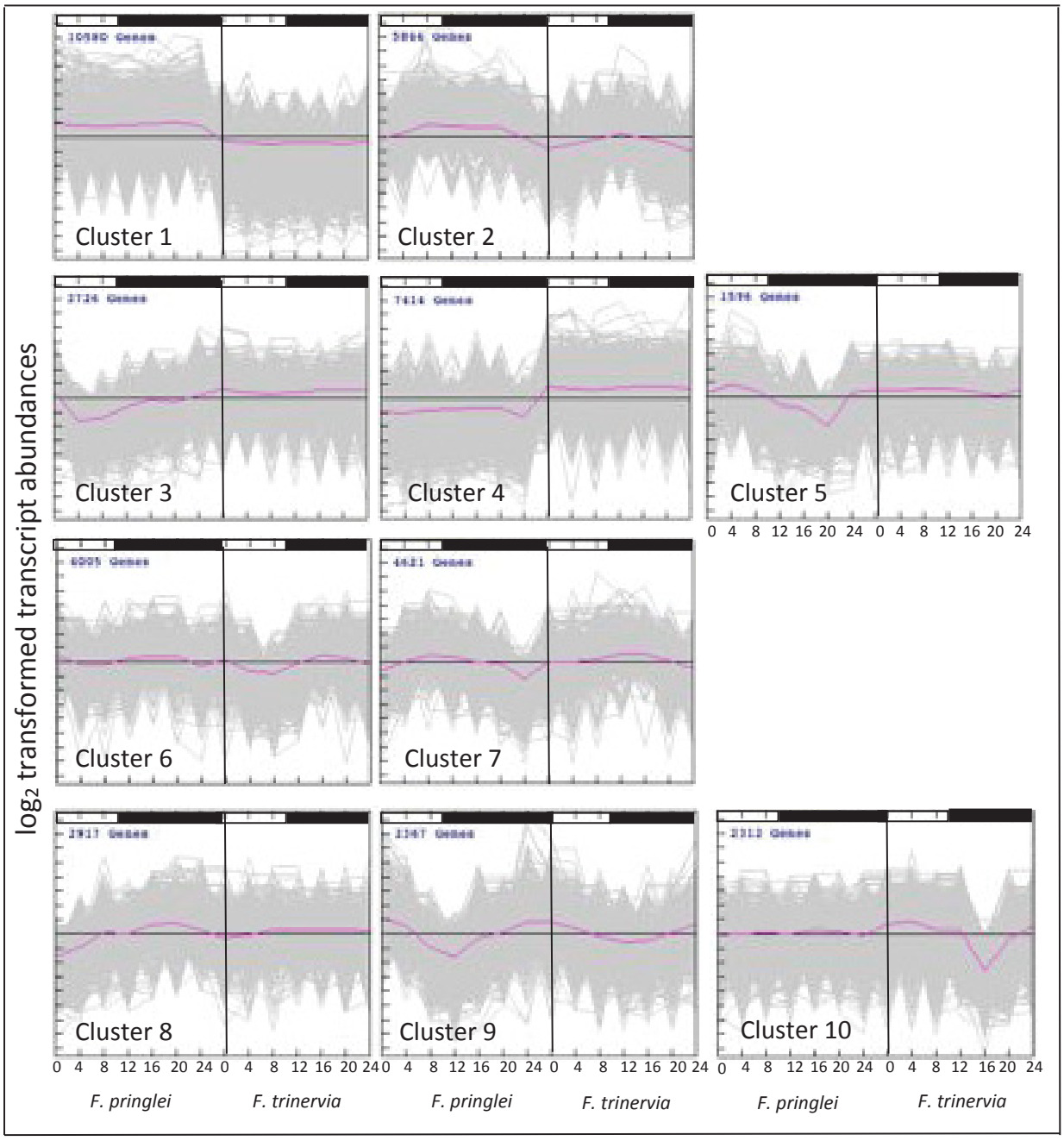


Figure 2. Comparison in the changes in gene expression in response to diurnal changes between *F. pringlei* and *F. trinervia*. 10 clusters were generated with the KMC modules showing the respective transcript profile of all detected genes from the transcriptome of *F. pringlei* and *F. trinervia*. The normalized RPMK accounts from each time point and each gene for the two species were log_2 transformed and combined to be processed with the TIGR MultiExperiment Viewer (MeV) program. The horizontal bar shows the day (open) and night (closed) period length (10-h day/14-h night). Lights were turned on at 8 a.m. and off at 6 p.m. Measurements started at dawn at time point 0-h. Each cluster represents a set of genes that exhibit similar diurnal expression pattern. These were compared between *F. pringlei* (C_3) and *F. trinervia* (C_4). White box (light period); Black box (dark period). Cluster 1 and 2: 16424 genes upregulated in *F. pringlei* relative to *F. trinervia*. Cluster 3 to 5: 11736 genes are upregulated in *F. trinervia* relative to *F. pringlei.*

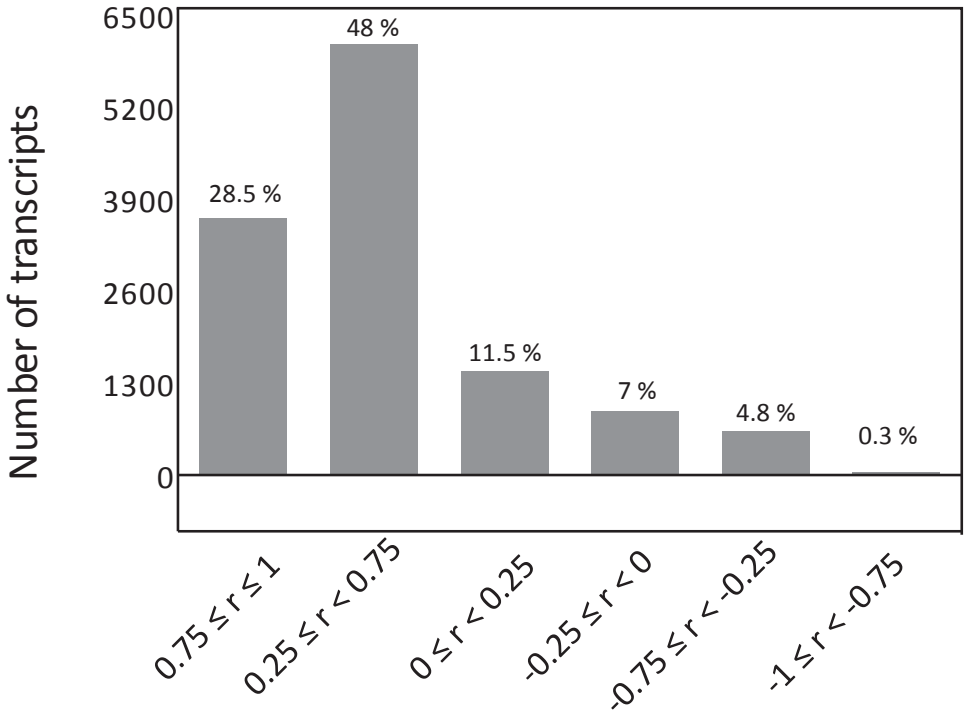
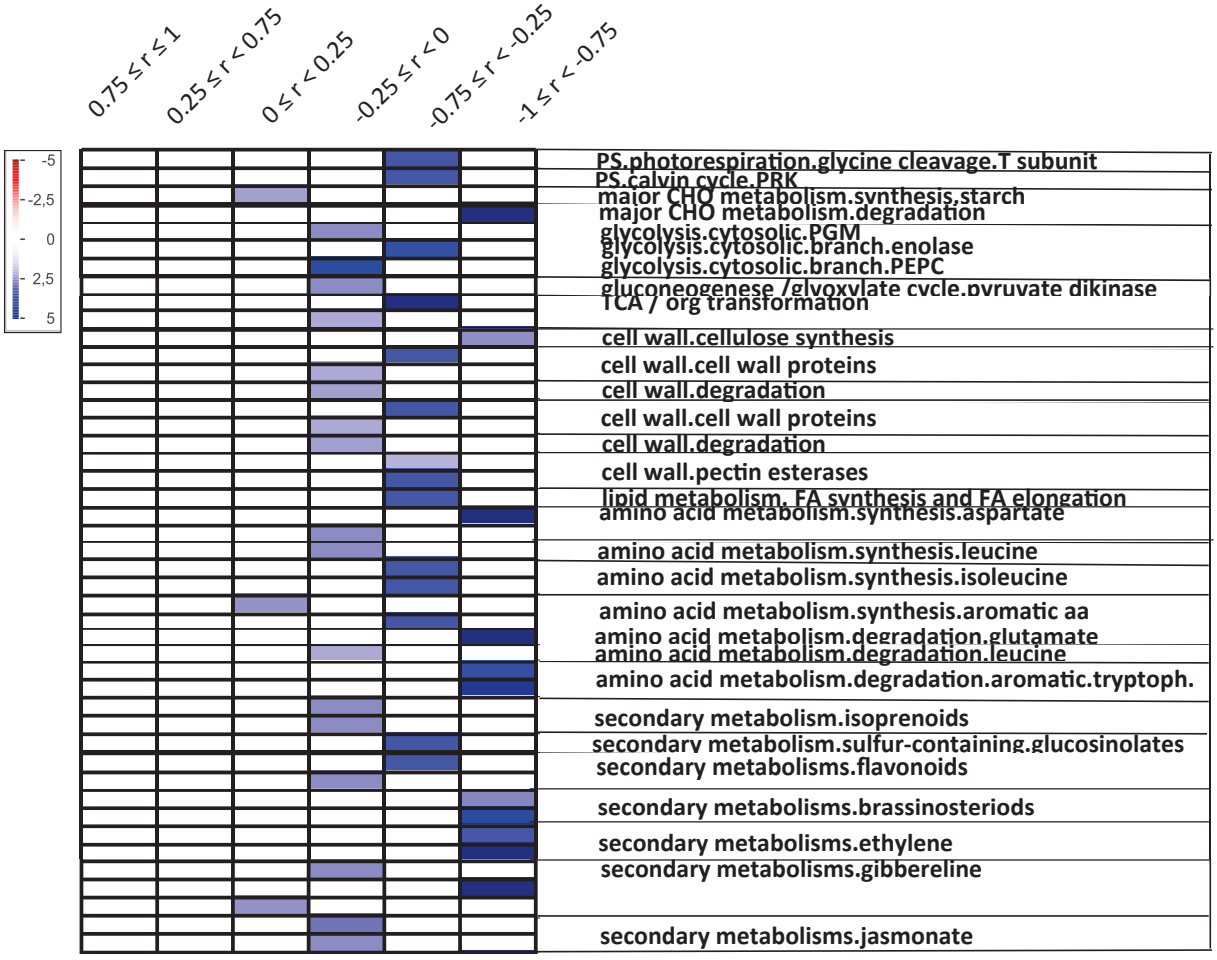


Figure 3. Correlation in diurnal pattern of gene expression between *F. pringlei* and *F. trinervia*. The Pearson´s correlation analysis was calculated for all detected genes from the transcriptome of *F. pringlei* and *F. trinervia*, in a species pair-wise comparison. The Pearson correlation shows the trend or pattern in diurnal expression irrespectively to the overall absolute transcript abundance. The histogram represents the total number of transcripts assigned to a particular Pearson´s correlation coefficient (r). The Pearson´s correlation coefficient corresponds to the degree of correlation in diurnal gene expression between the species. $0.75 \leq r < 1$ (strong to perfect correlation), $0.25 \leq r < 0.75$ (correlation), $0 \leq r < 0.25$ (weak correlation), $-0.25 \leq r \leq 0$ (weak inverse correlation), $-0.75 \leq r < -0.25$ (inverse correlation), $-1 \leq r \leq -0.75$ (strong to perfect inverse correlation).


0.75 ≤ r ≤ 1
0.25 ≤ r < 0.75
0 ≤ r < 0.25
-0.25 ≤ r < 0
-0.75 ≤ r < -0.25
-1 ≤ r < -0.75
-5
-2,5
0
2,5
5
PS.photorespiration.glycine cleavage.T subunit
PS.calvin cycle.PRK
major CHO metabolism.synthesis.starch
major CHO metabolism.degradation
glycolysis.cytosolic.PGM
glycolysis.cytosolic.branch.enolase
glycolysis.cytosolic.branch.PEPC
gluconeogenese /glyoxylate cycle.pyruvate dikinase
TCA / org transformation
cell wall.cellulose synthesis
cell wall.cell wall proteins
cell wall.degradation
cell wall.cell wall proteins
cell wall.degradation
cell wall.pectin esterases
lipid metabolism. FA synthesis and FA elongation
amino acid metabolism.synthesis.aspartate
amino acid metabolism.synthesis.leucine
amino acid metabolism.synthesis.isoleucine
amino acid metabolism.synthesis.aromatic aa
amino acid metabolism.degradation.glutamate
amino acid metabolism.degradation.leucine
amino acid metabolism.degradation.aromatic.tryptoph.
secondary metabolism.isoprenoids
secondary metabolism.sulfur-containing.glucosinolates
secondary metabolisms.flavonoids
secondary metabolisms.brassinosteriods
secondary metabolisms.ethylene
secondary metabolisms.gibbereline
secondary metabolisms.jasmonate

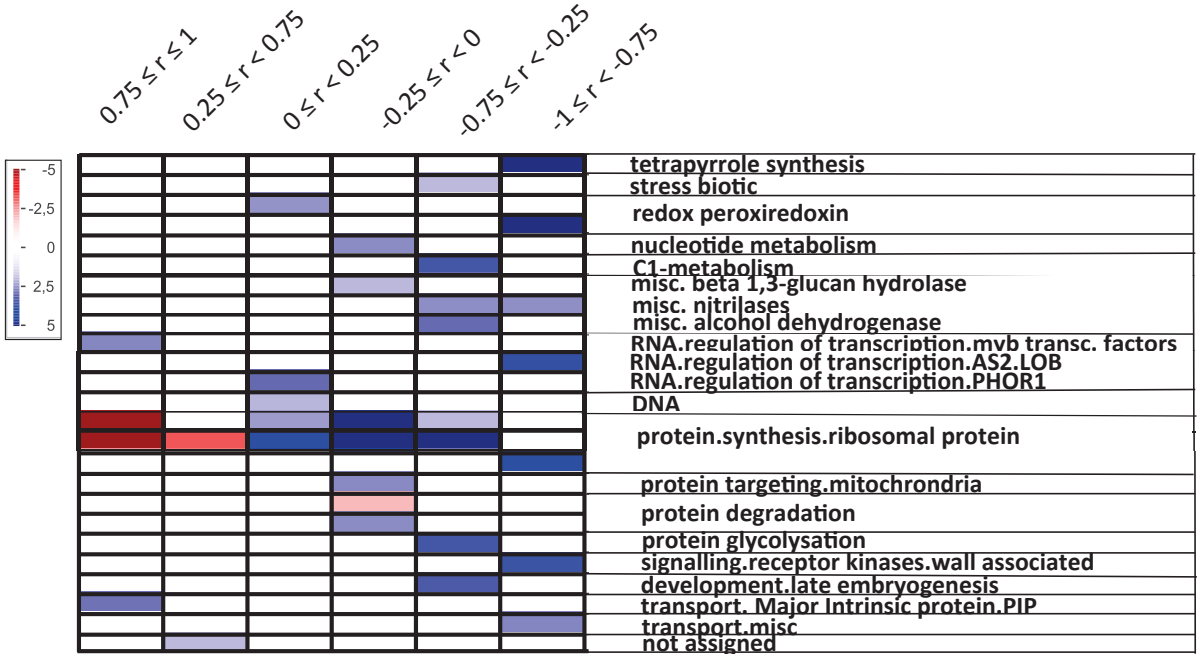
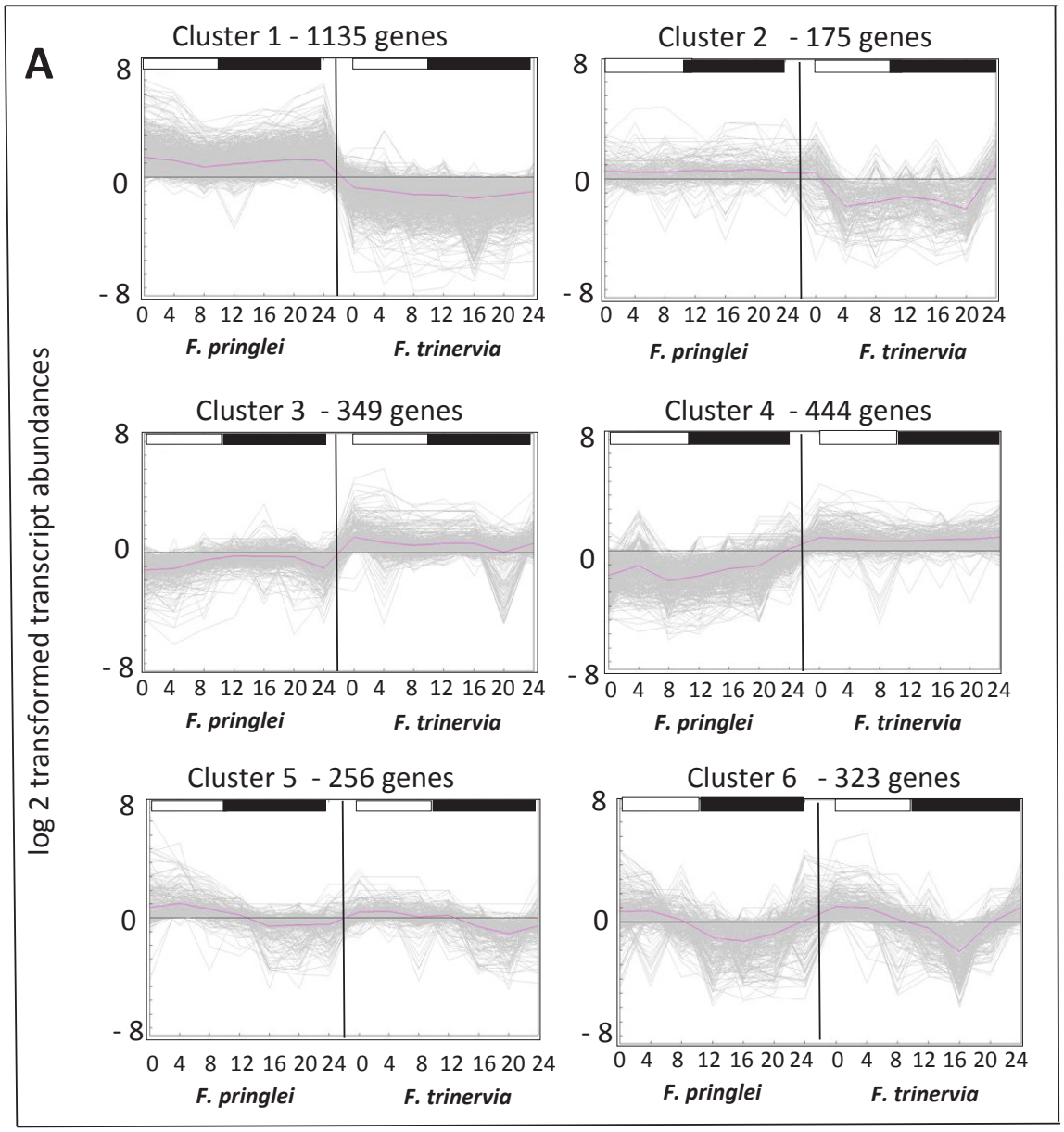
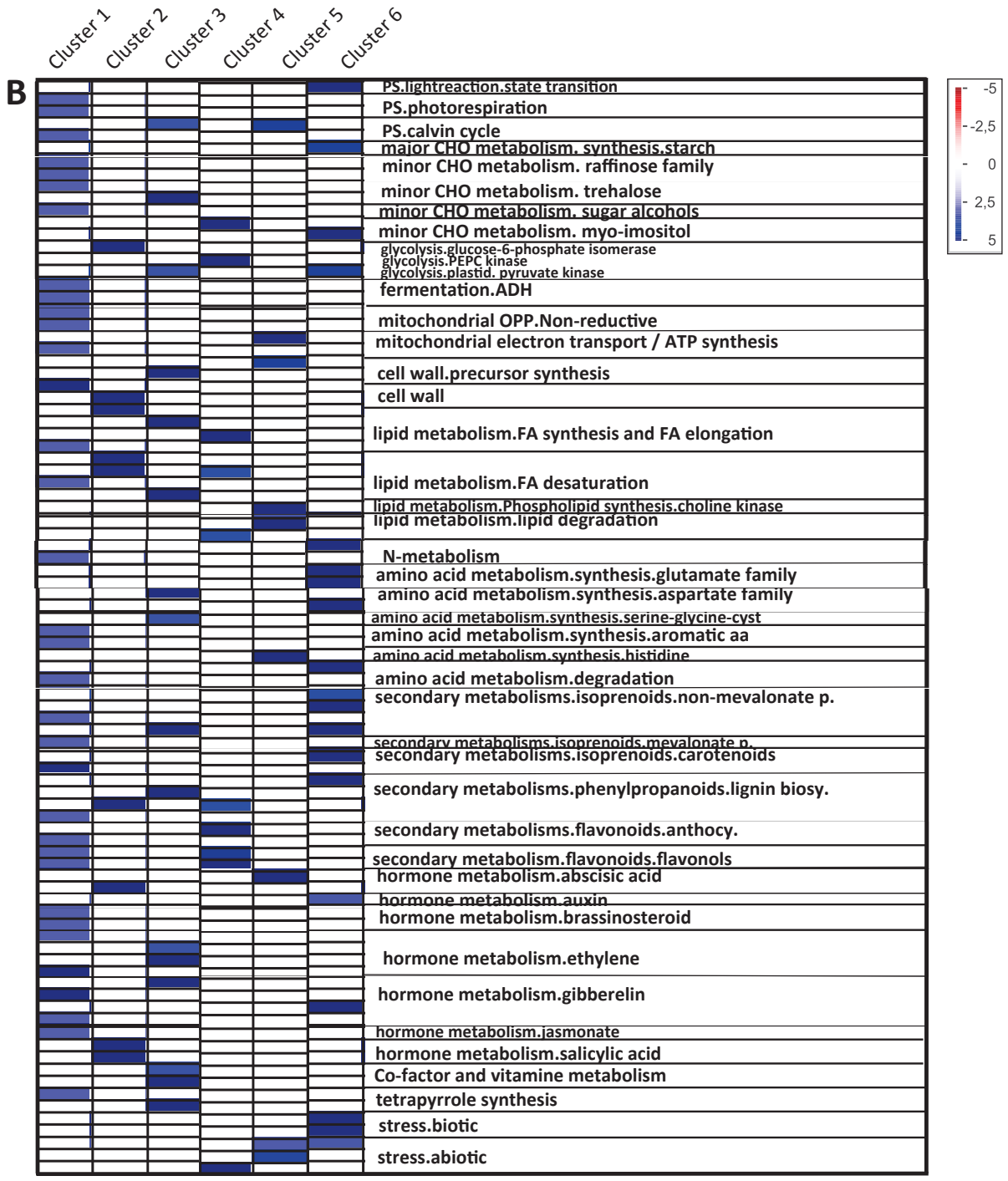

0.75 ≤ r ≤ 1
0.25 ≤ r < 0.75
0 ≤ r < 0.25
-0.25 ≤ r < 0
-0.75 ≤ r < -0.25
-1 ≤ r < -0.75
-5
-2,5
0
2,5
5
tetrapyrrole synthesis
stress biotic
redox peroxiredoxin
nucleotide metabolism
C1-metabolism
misc. beta 1,3-glucan hydrolase
misc. nitrilases
misc. alcohol dehydrogenase
RNA.regulation of transcription.myb transc. factors
RNA.regulation of transcription.AS2.LOB
RNA.regulation of transcription.PHOR1
DNA
protein.synthesis.ribosomal protein
protein targeting.mitochrondria
protein degradation
protein glycolysation
signalling.receptor kinases.wall associated
development.late embryogenesis
transport. Major Intrinsic protein.PIP
transport.misc
not assigned

Figure 4. Functional classes of genes from *F. pringlei* and *F. trinervia* linked to diurnal gene expression. Identification of over-represented functional classes in association to a particular correlation coefficient. Overrepresented (blue) and underrepresented functional gene classes (red). Genes that surpass a fold-change in expression assigned to a particular functional class in comparison to the expression of all genes in the RNA seq data set, is considered as over-represented. In the inverse correlated group a high number of functional classes are over-represented, because the group has a low number of genes. The normalized RPMK accounts from each time point and each gene for both species were combined and processed with PageMan. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction are included statistical parameters of the PageMan programs and were used to reveal the significance of the BIN assignment to a particular category. Genes that surpass a fold-change in expression assigned to a particular functional class in comparison to the expression of all genes in the RNAseq data set, are considered as over-represented (blue). The p-values are set to z-values (Usadel et al., 2006). PS, photosynthesis, TCA, tricarboxylic acid cycle, CHO, carbohydrate; OPP, oxidative pentose phosphate.




B
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
PS.lightreaction.state transition
PS.photorespiration
PS.calvin cycle
major CHO metabolism. synthesis.starch
minor CHO metabolism. raffinose family
minor CHO metabolism. trehalose
minor CHO metabolism. sugar alcohols
minor CHO metabolism. myo-imositol
glycolysis.glucose-6-phosphate isomerase
glycolysis.PEPC kinase
glycolysis.plastid. pyruvate kinase
fermentation.ADH
mitochondrial OPP.Non-reductive
mitochondrial electron transport / ATP synthesis
cell wall.precursor synthesis
cell wall
lipid metabolism.FA synthesis and FA elongation
lipid metabolism.FA desaturation
lipid metabolism.Phospholipid synthesis.choline kinase
lipid metabolism.lipid degradation
N-metabolism
amino acid metabolism.synthesis.glutamate family
amino acid metabolism.synthesis.aspartate family
amino acid metabolism.synthesis.serine-glycine-cyst
amino acid metabolism.synthesis.aromatic aa
amino acid metabolism.synthesis.histidine
amino acid metabolism.degradation
secondary metabolisms.isoprenoids.non-mevalonate p.
secondary metabolisms.isoprenoids.mevalonate p.
secondary metabolisms.isoprenoids.carotenoids
secondary metabolisms.phenylpropanoids.lignin biosy.
secondary metabolisms.flavonoids.anthocy.
secondary metabolism.flavonoids.flavonols
hormone metabolism.abscisic acid
hormone metabolism.auxin
hormone metabolism.brassinosteroid
hormone metabolism.ethylene
hormone metabolism.gibberelin
hormone metabolism.jasmonate
hormone metabolism.salicylic acid
Co-factor and vitamine metabolism
tetrapyrrole synthesis
stress.biotic
stress.abiotic
-5
-2,5
0
2,5
5

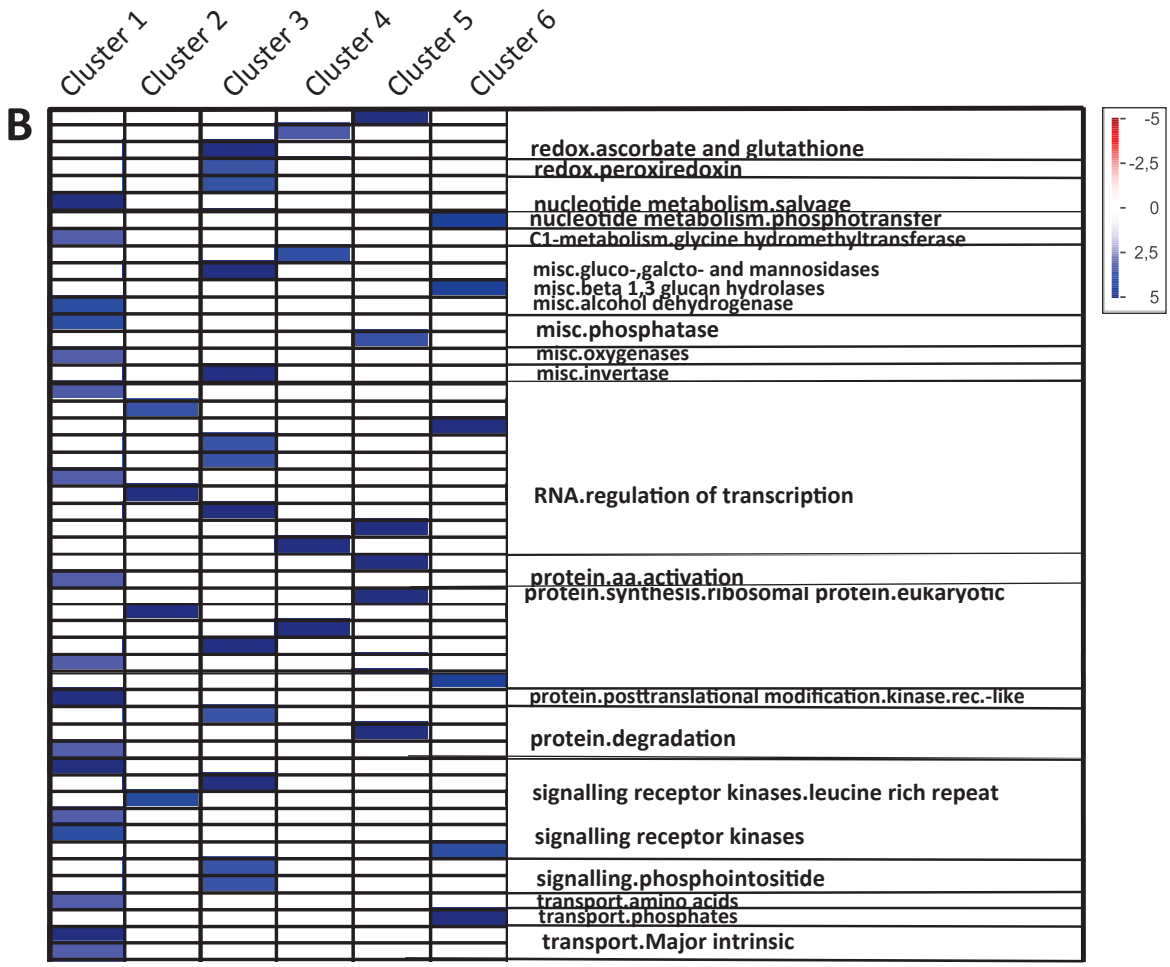


Figure 5. Functional classes of genes from *F. pringlei* and *F. trinervia* linked to clusters. (**A**) 6 clusters within 200 clusters were selected that represent the diurnal expression pattern from *F. pringlei* compared to that of *F. trinervia*. The normalized RPMK accounts from each time point and each gene for the two species were log_2 transformed and combined to be processed with the TIGR MultiExperiment Viewer (MeV) program. Clusters were generated with Euclidean distances using the HCL module. White box (light period); Black box (dark period). (**B**) Identification of functional classes overrepresented in a particular cluster. The normalized RPMK accounts from each time point and each gene for both species were combined and processed with the PageMan program. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction are included statistical parameters of the PageMan program and were used to reveal the significance of the BIN assignment to a particular category. Genes that surpass a fold-change in expression assigned to a particular functional class in comparison to the expression of all genes in the RNAseq data set, are considered as over-represented (blue). The p-values are set to z-values (Usadel et al., 2006). PS, photosynthesis, TCA, tricarboxylic acid cycle, CHO, carbohydrate; OPP, oxidative pentose phosphate.

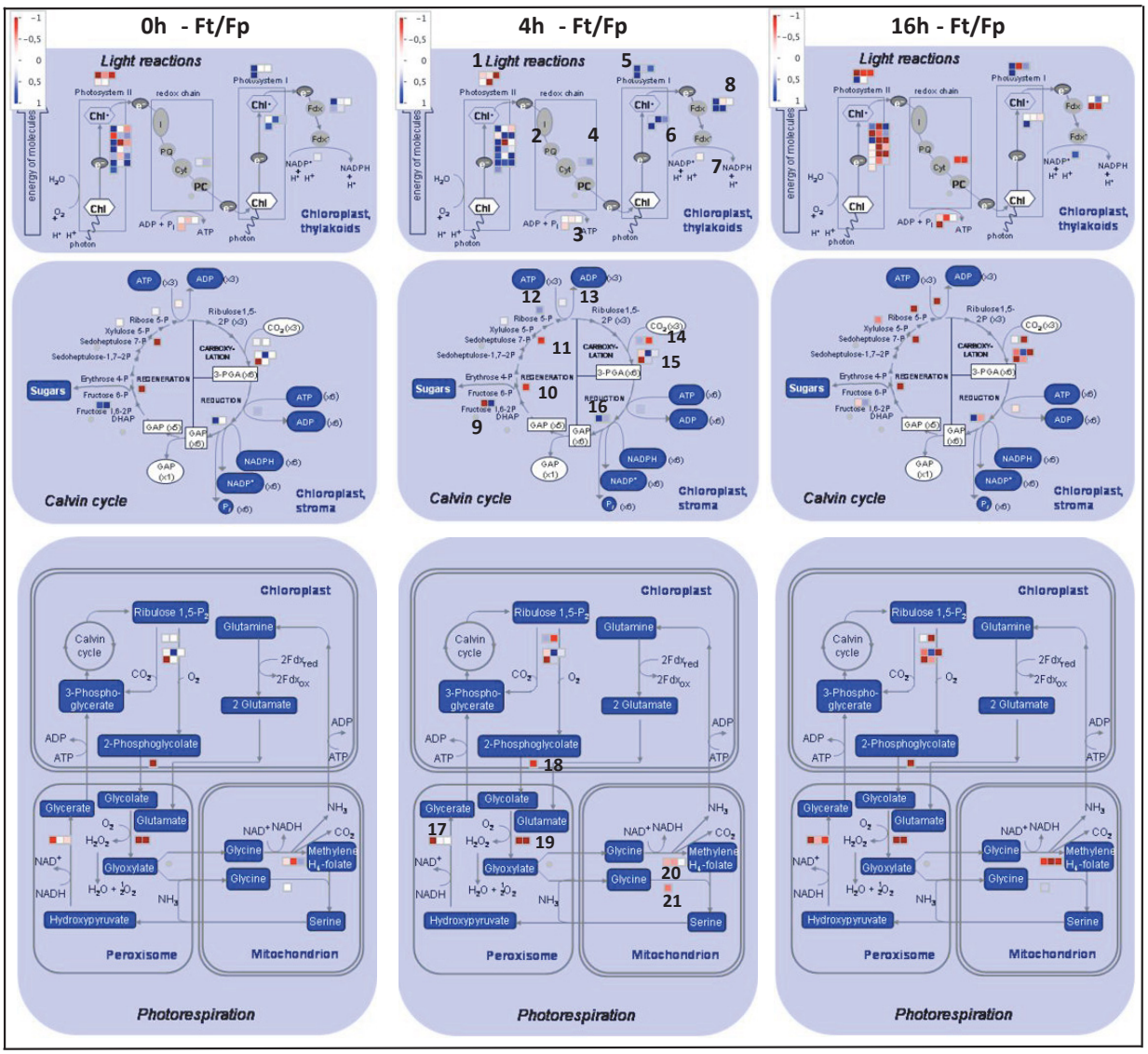


Figure 6. Ratio of gene expression in *F. trinervia* to *F. pringlei*. The ratio of the transcript abundance at dawn (0-h), noon (4-h) and midnight (16-h) from *F. trinervia* to that of *F. pringlei*. The average of the normalized RPMK for each gene from three biological replicates between *F. pringlei* and *F. trinervia* was $\log_2$ transformed. The ratio was used as input in MapMan. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction were used for correction. Blue indicates for a higher expression of *F. trinervia* in respect to *F. pringlei*, whereas red indicates for a higher expression of *F. pringlei* in respect to *F. trinervia.* White squares indicate no significant differences between the species in gene expression. (1) PS. Light reaction. PSII. LHC-II; (2) PS. Light reaction. PSII. polypeptide subunit; (3) PS. Light reaction. ATP synthase; (4) PS. Light reaction. cytochrome b6/f; (5) PS. Light reaction.PSI-LHC-I; (6) PS. Light reaction. PSI polypeptide subunit; (7) PS. Light reaction. other electron carrier. ox/red. ferredoxin reductase; (8) PS. Light reaction. other electron carrier. ox/red. ferredoxin. (9) PS. calvin cycle. FBPase; (10) PS. calvin cycle. transketolase; (11) PS. calvin cycle. transketolase; (12) PS. calvin cycle. Rib5P isomerase; (13) PS. calvin cycle. PRK; (14) PS. calvin cycle. Rubisco large (left) small (right) subunits; (15) PS. calvin cycle. rubisco interacting; (16) PS. calvin cycle. GAP; (17) PS. photorespiration. hydroxypyruvate reductase; (18) PS. photorespiration. phosphoglycolate phosphatase; (19) PS. photorespiration. glycolate oxydase; (20) glycine cleavage T subunit; (21) PS. photorespiration. serine hydroxymethyl transferase.

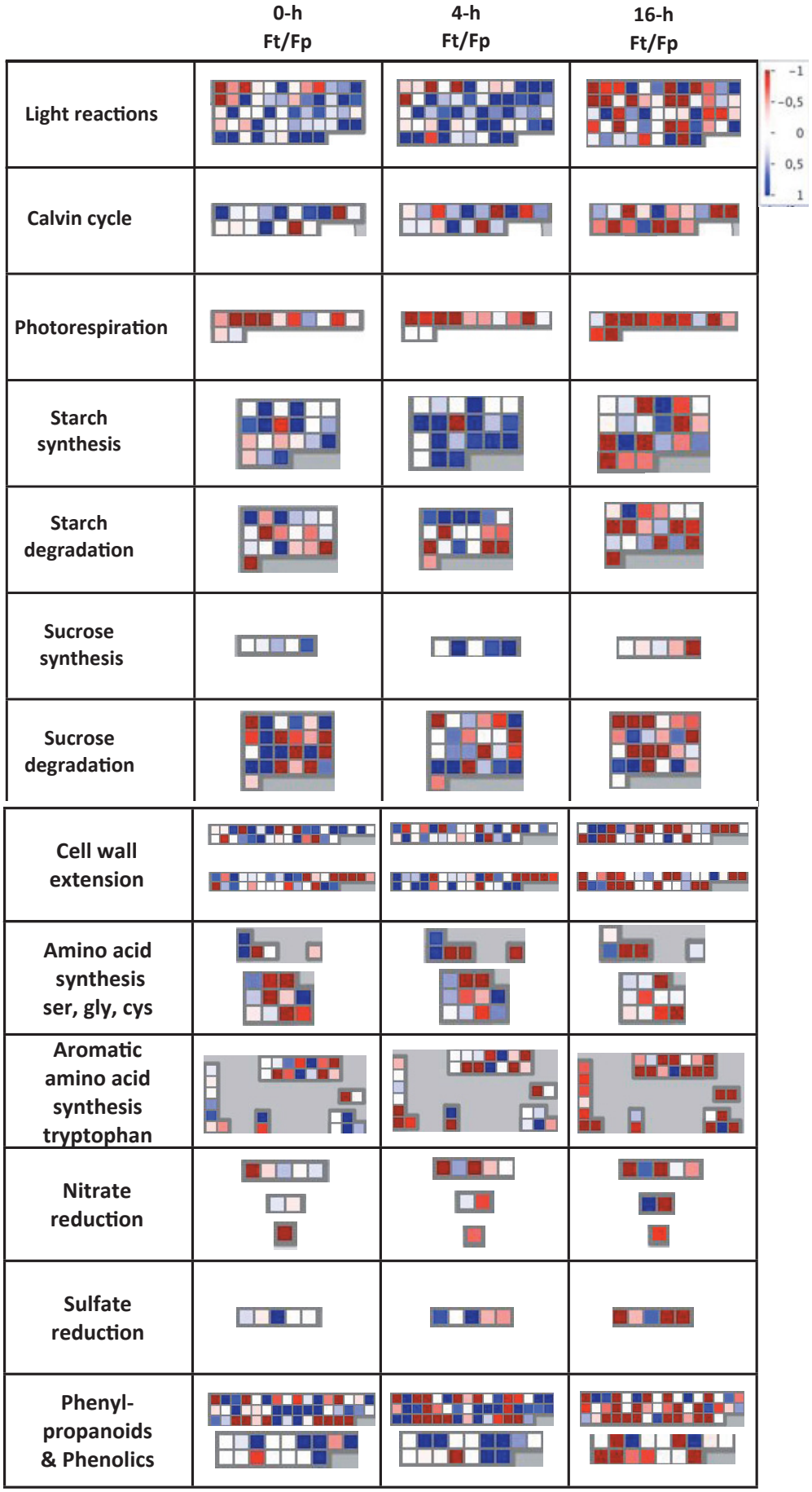

0-h
Ft/Fp
4-h
Ft/Fp
16-h
Ft/Fp
Light reactions
Calvin cycle
Photorespiration
Starch synthesis
Starch degradation
Sucrose synthesis
Sucrose degradation
Cell wall extension
Amino acid synthesis ser, gly, cys
Aromatic amino acid synthesis tryptophan
Nitrate reduction
Sulfate reduction
Phenyl-propanoids & Phenolics
-1
-0,5
0
0,5
1

Figure 7. Ratio of gene expression in *F. trinervia* to *F. pringlei*. The ratio of the transcript abundance at dawn (0-h), noon (4-h) and midnight (16-h) from *F. trinervia* to that of *F. pringlei*. The average of the normalized RPMK for each gene from three biological replicates between *F. pringlei* and *F. trinervia* was $\log_2$ transformed. The ratio was used as input in MapMan. The Wilcoxon-Mann-Whitney test followed by the Bonferroni correction were used for correction. Blue indicates for a higher expression of *F. trinervia* in respect to *F. pringlei*, whereas red indicates for a higher expression of *F. pringlei* in respect to *F. trinervia.* White squares indicate no significant differences between the species in gene expression.

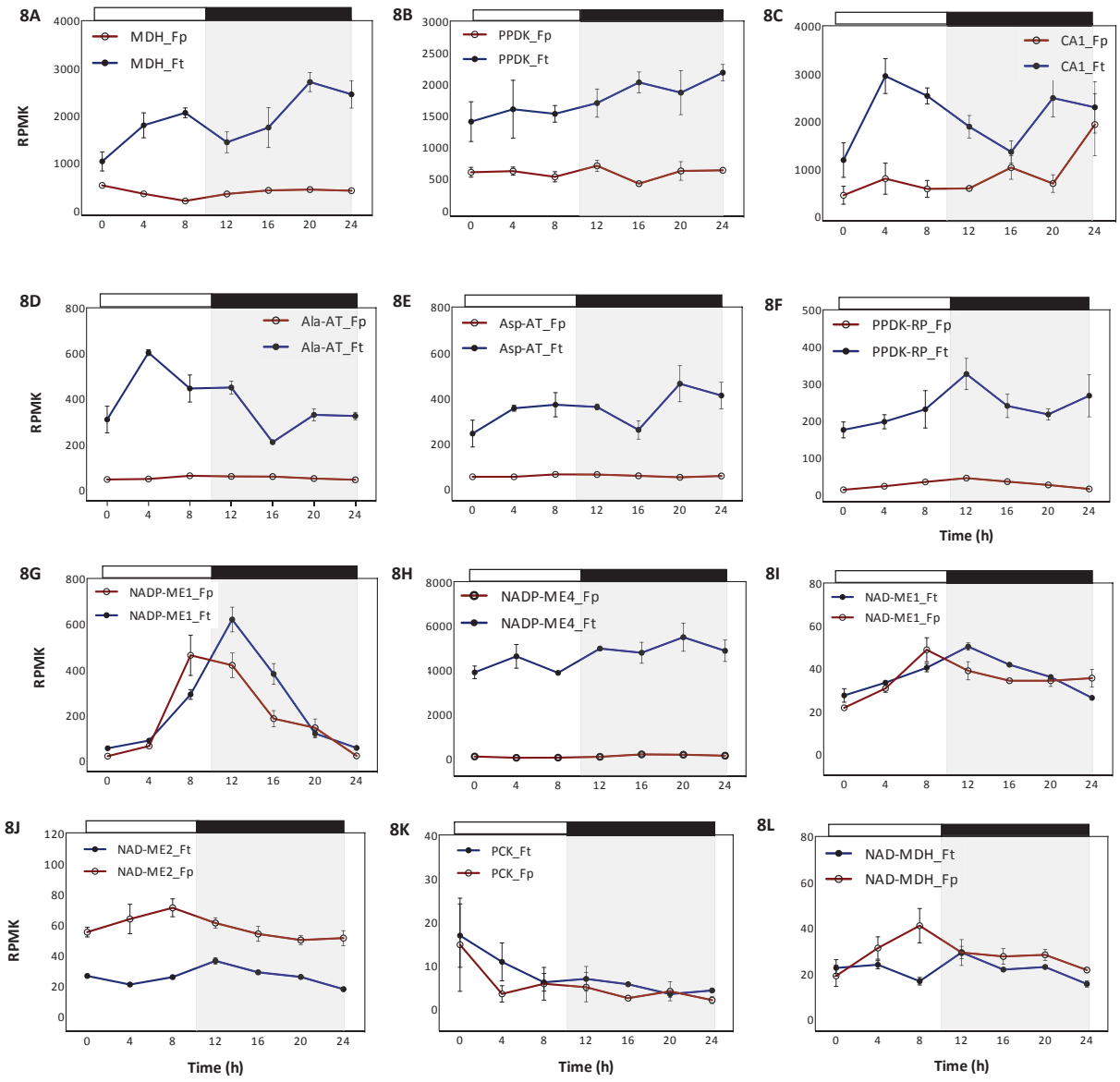


Figure 8. Changes of transcript abundances of genes of the C_4 cycle during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). The reads per million mapped reads and kilobase transcript length (RPMK) is a measure of the transcript abundances. The average of RPMK values correspond to three biological replicates for each gene and species. The horizontal bar shows the day (open) and night (closed) period length (10-h day/14-h night). Lights were turned on at 8 a.m. and off at 6 p.m. Measurements started at dawn at time point 0-h. The harvest of leave samples continued repeatedly every 4-h along the 24-h cycle. The standard error of the mean (SEM) was calculated for each time course of three biological replicates. The significant differences between each time point were determined with the RPMK values using a one-way ANOVA (Tukey-test) $p \leq 0.05$ summarized in Table 1. (**A**) NADP-MDH (**B**) PPDK (**C**) CA1 (**D**) Ala-AT (**E**) Asp-AT (**F**) PPDK-RP (**G**) NADP-ME1 (**H**) NADP-ME4 (**I**) NAD-ME1 (**J**) NAD-ME2 (**K**) PCK (**L**) NAD-MDH.

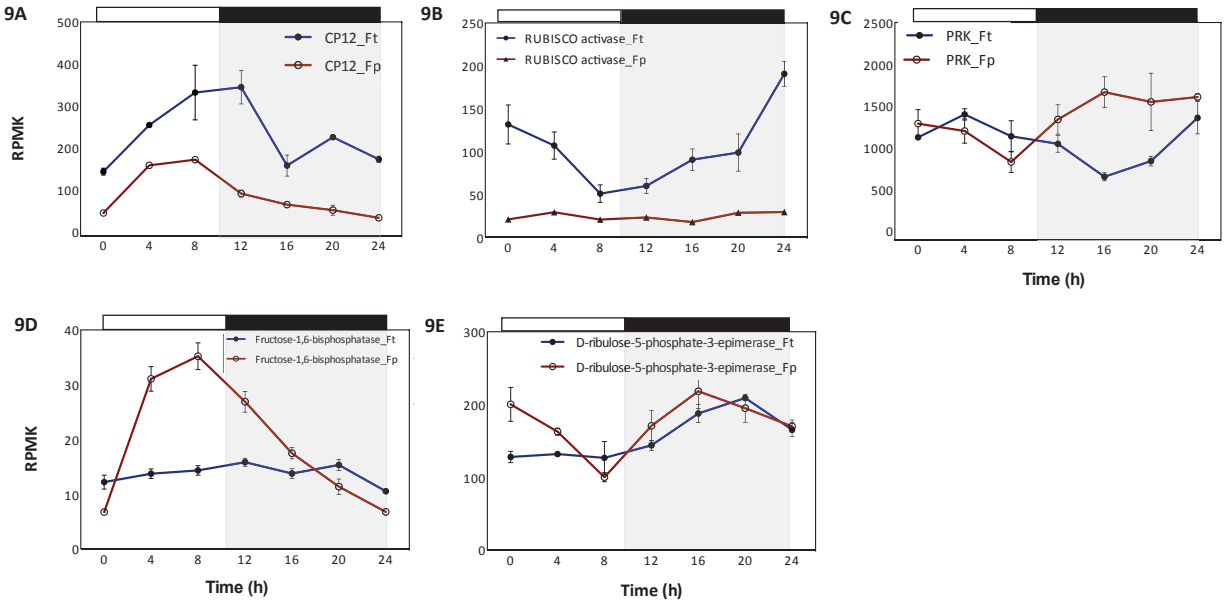


Figure 9. Changes of transcript abundances of genes of the Calvin-Benson cycle during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 2. (**A**) CP12 (**B**) RUBISCO activase (**C**) PRK (**D**) fructose-1,6-biphosphatase family protein (**E**) D-ribulose-5-phosphate-3-epimerase.

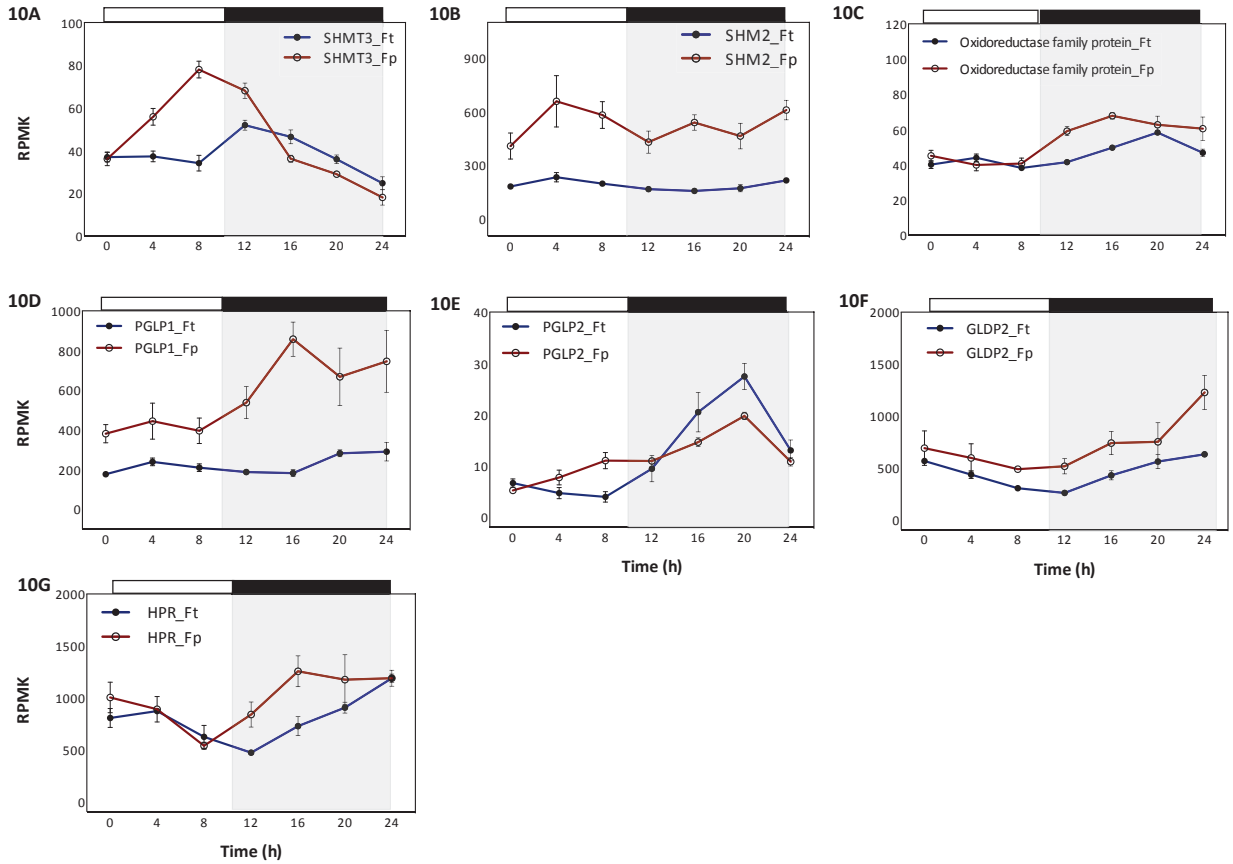


Figure 10. Changes of transcript abundances of photorespiratory genes during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 3. (**A**) SHMT3 (**B**) SHM2 (**C**) Oxidoreductase family protein (**D**) PGLP1 (**E**) PGLP2 (**F**) GLDP2 (**G**) HPR.

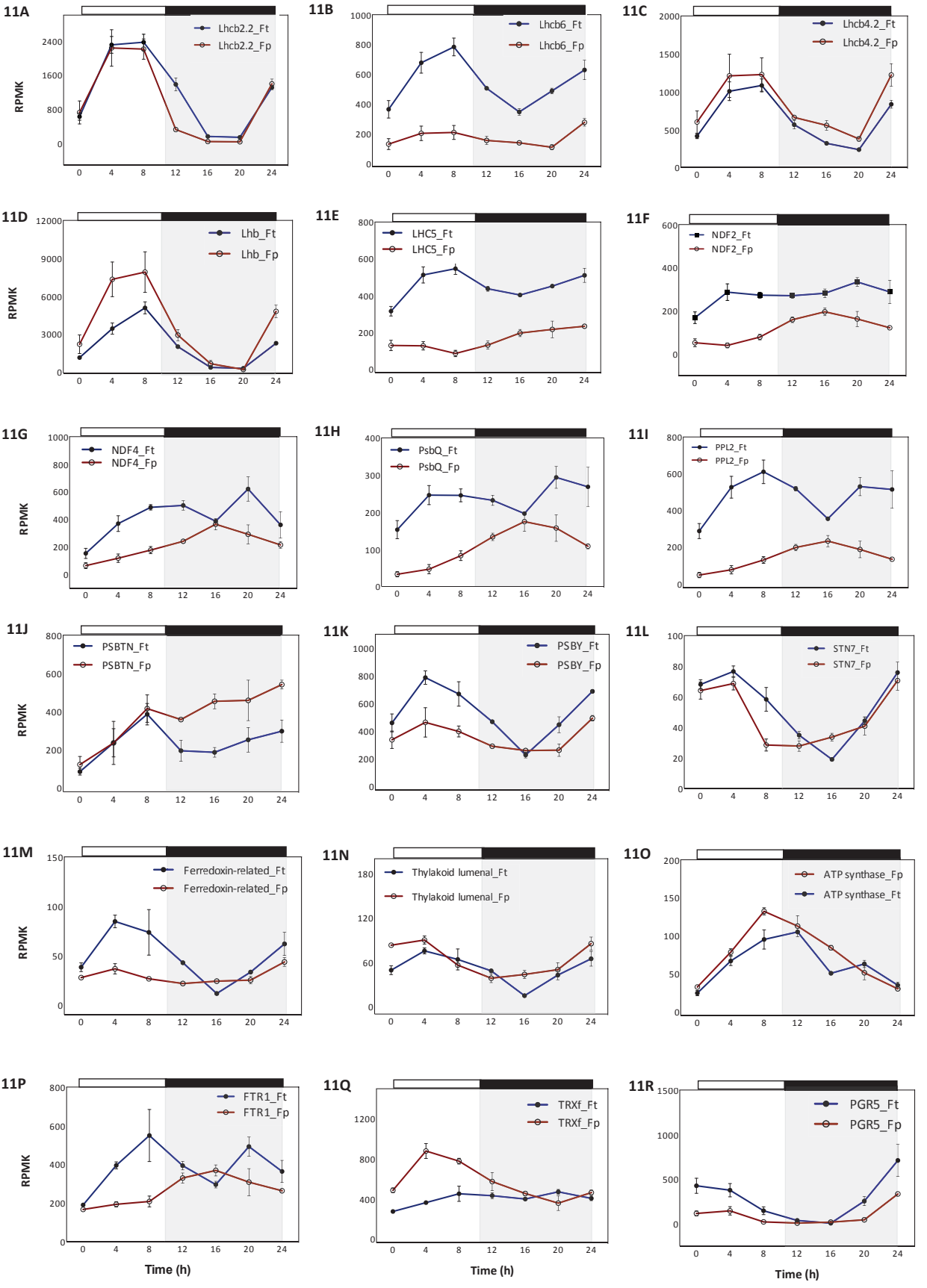
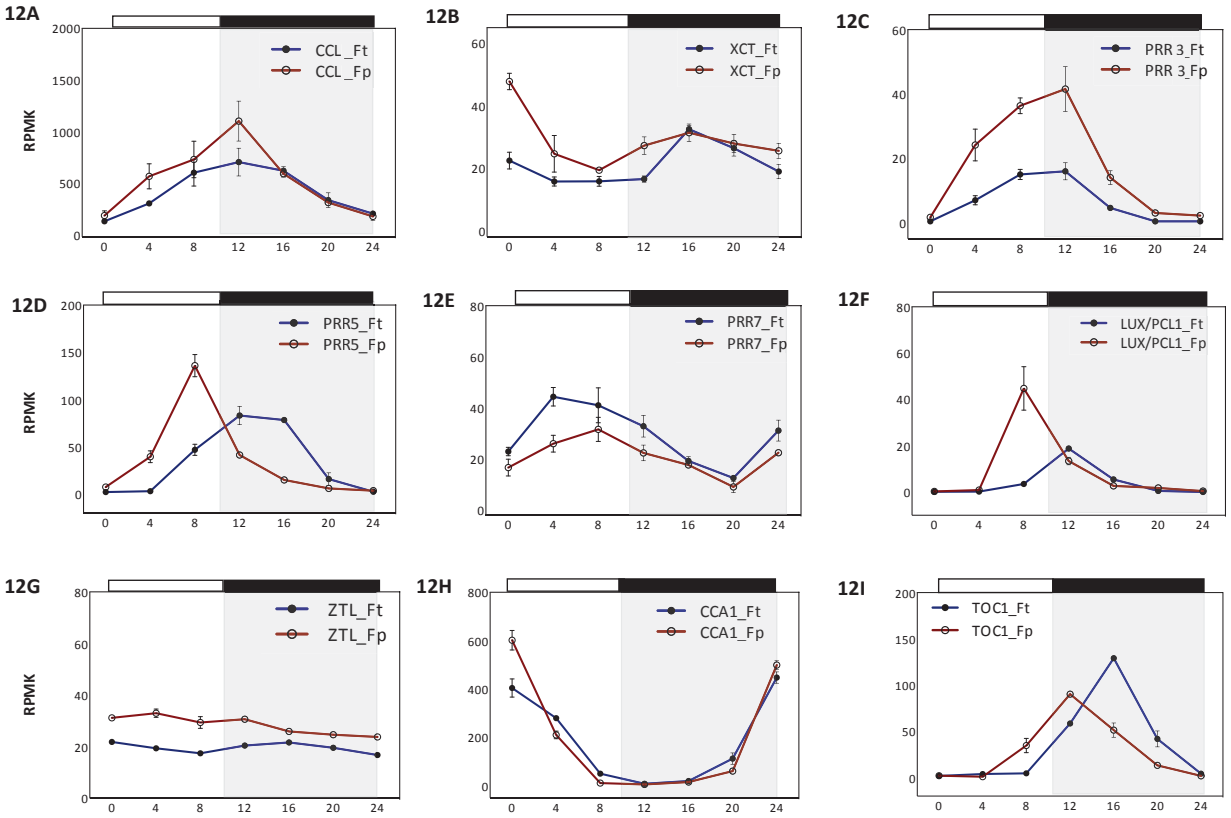

11A
Lhcb2.2_Ft
Lhcb2.2_Fp
11B
Lhcb6_Ft
Lhcb6_Fp
11C
Lhcb4.2_Ft
Lhcb4.2_Fp
11D
Lhb_Ft
Lhb_Fp
11E
LHC5_Ft
LHC5_Fp
11F
NDF2_Ft
NDF2_Fp
11G
NDF4_Ft
NDF4_Fp
11H
PsbQ_Ft
PsbQ_Fp
11I
PPL2_Ft
PPL2_Fp
11J
PSBTN_Ft
PSBTN_Fp
11K
PSBY_Ft
PSBY_Fp
11L
STN7_Ft
STN7_Fp
11M
Ferredoxin-related_Ft
Ferredoxin-related_Fp
11N
Thylakoid lumenal_Ft
Thylakoid lumenal_Fp
11O
ATP synthase_Fp
ATP synthase_Ft
11P
FTR1_Ft
FTR1_Fp
11Q
TRXf_Ft
TRXf_Fp
11R
PGR5_Ft
PGR5_Fp
RPMK
Time (h)

Figure 11. Changes of transcript abundances of genes involved in the photosynthetic light reactions during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 4. (**A**) Lhcb2.2 (**B**) Lhcb6 (**C**) Lhcb4.2 (**D**) Lhb (cab) (**E**) LHC5 (**F**) NDF2 (**G**) NDF4 (**H**) PsbQ-like 1 (**I**) PPL2 (**J**) PSBTN (**K**) PSBY (**L**) STN7 (**M**) ferredoxin-related (**N**) thylakoid lumenal (**O**) ATP synthase (**P**) FTR1 (**Q**) TRXf1 (**R**) PGR5.



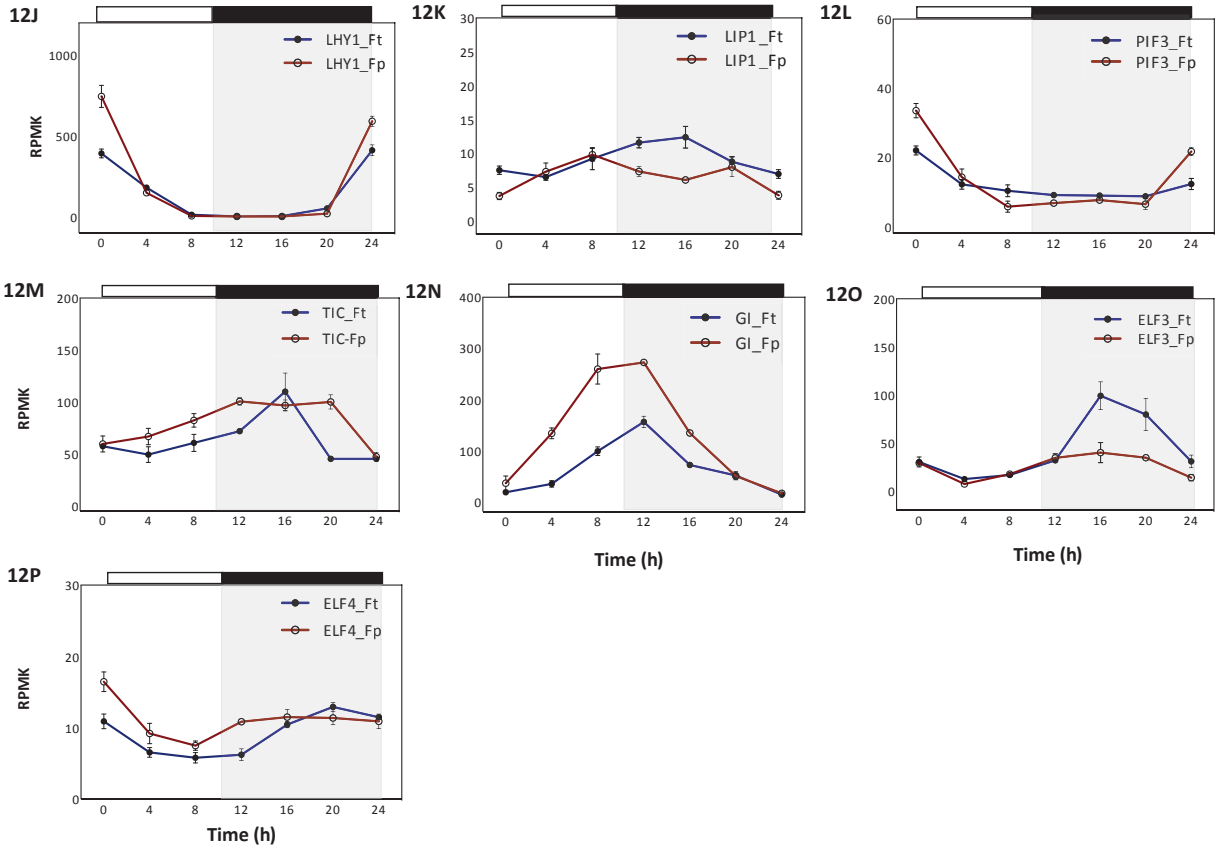


Figure 12. Changes of transcript abundances of circadian clock genes and clock-associated genes during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 5. (**A**) CCL (**B**) XCT (**C**) PRR3 (**D**) PRR5 (**E**) PRR7 (**F**) LUX/PCL1 (**G**) ZTL (**H**) CCA1 (**I**) TOC1 (**J**) LHY (**K**) LIP1 (**L**) PIF3 (**M**) TIC (**N**) GI (**O**) ELF3 (**P**) ELF4.

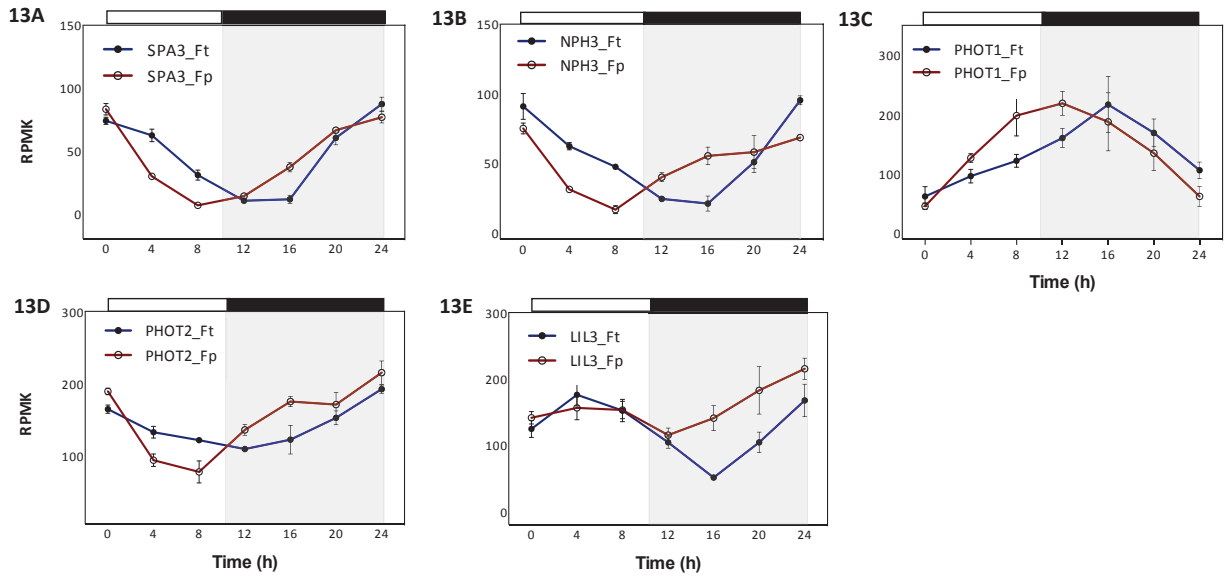


Figure 13. Changes of transcript abundances of light-regulated genes during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 6. (**A**) SPA3 (**B**) NPH3 (**C**) PHOT1 (**D**) PHOT2 (**E**) LIL3.

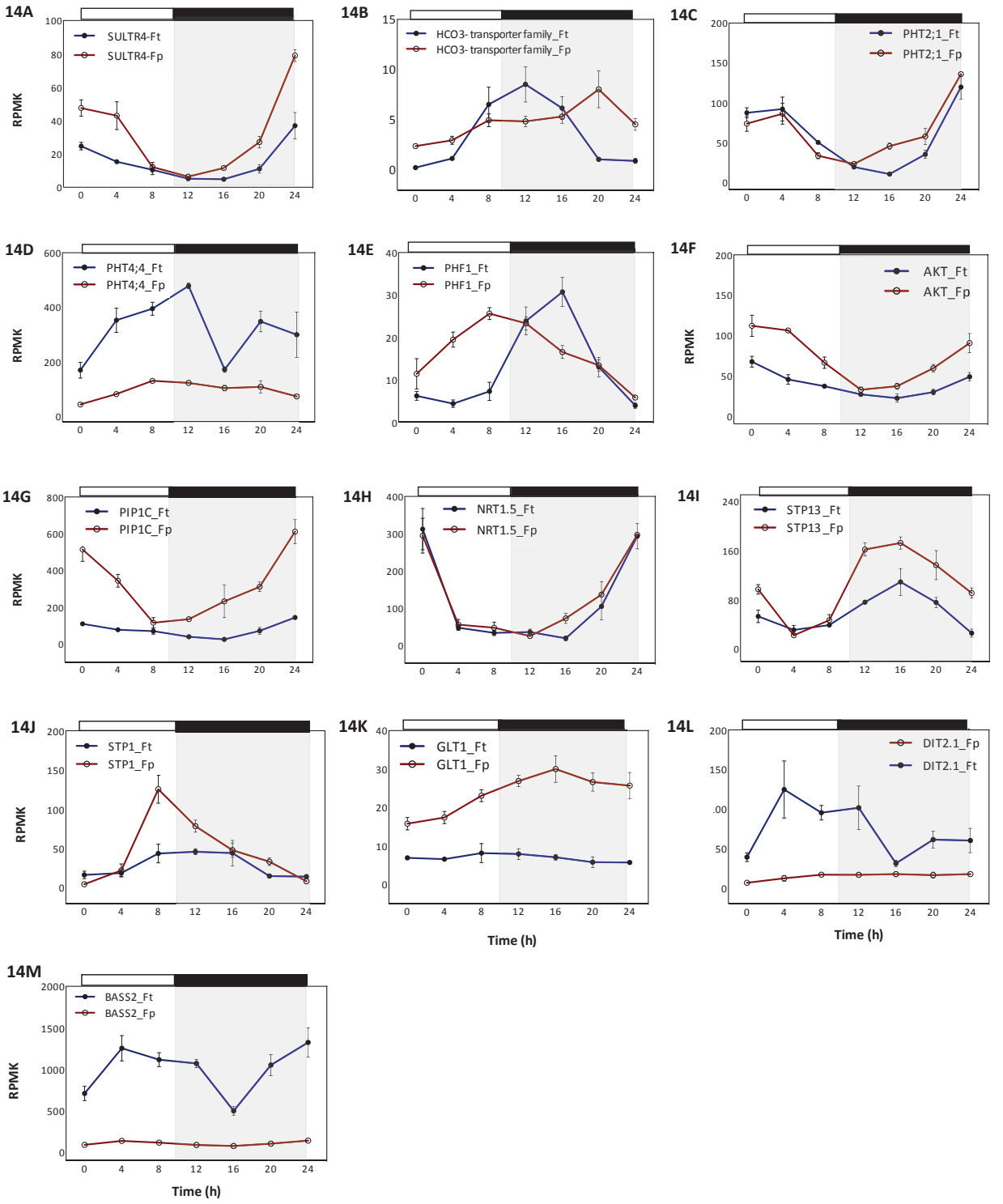


Figure 14. Changes of transcript abundances of genes involved in transport during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 7. (**A**) SULTR4 (**B**) HCO_3^- transporter family (**C**) PHT2;1 (**D**) PHT4;4 (**E**) PHF1; (**F**) AKT (**G**) PIP1C (**H**) NRT1.5 (**I**) STP13 (**J**) STP1 (**K**) GLT1 (**L**) DIT2.1 (**M**) BASS2

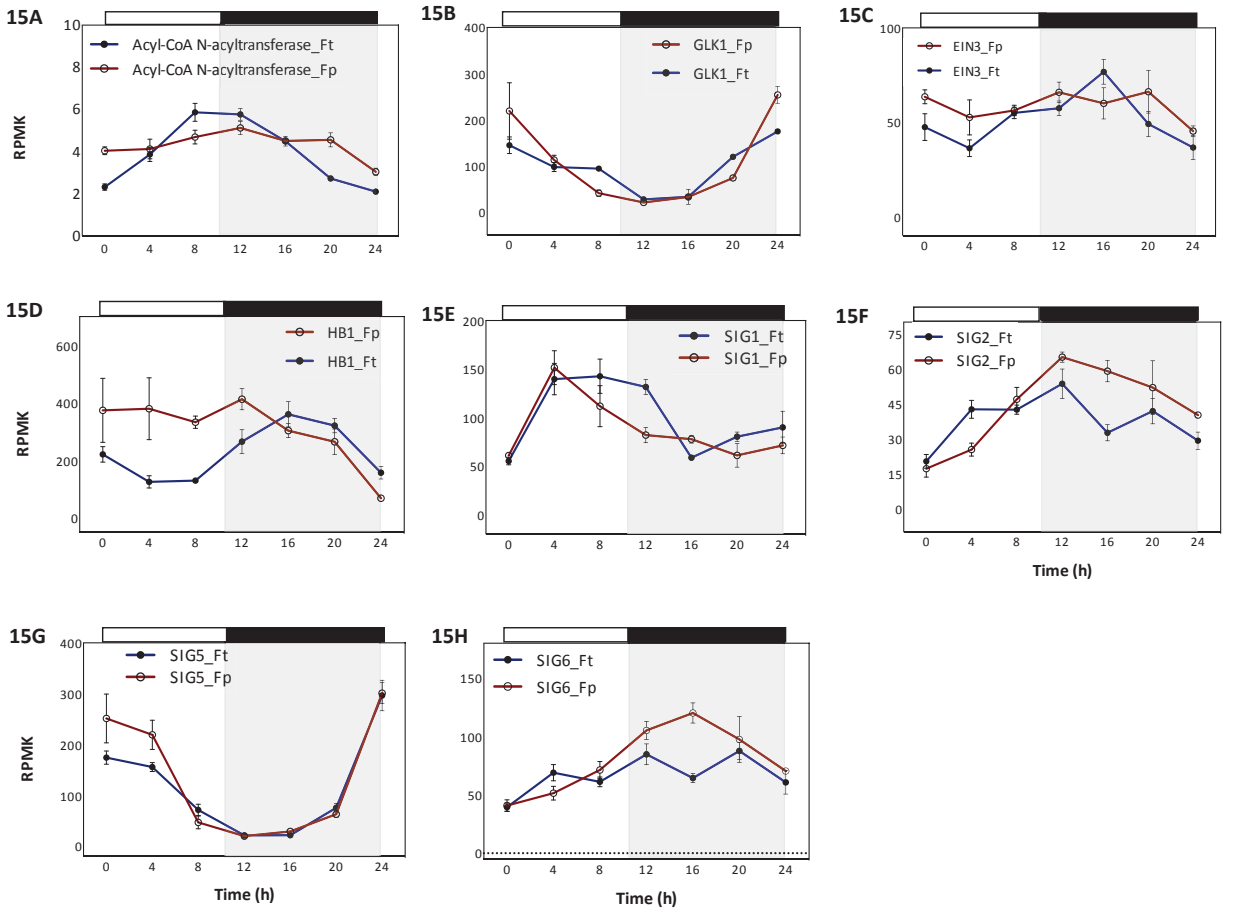


Figure 15. Changes of transcript abundances of genes encoding transcription factors during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 8. (**A**) Acyl-CoA N-acyltransferase (**B**) GLK1 (**C**) EIN3 (**D**) HB1 (**E**) SIG1 (**F**) SIG2 (**G**) SIG5 (**H**) SIG6.

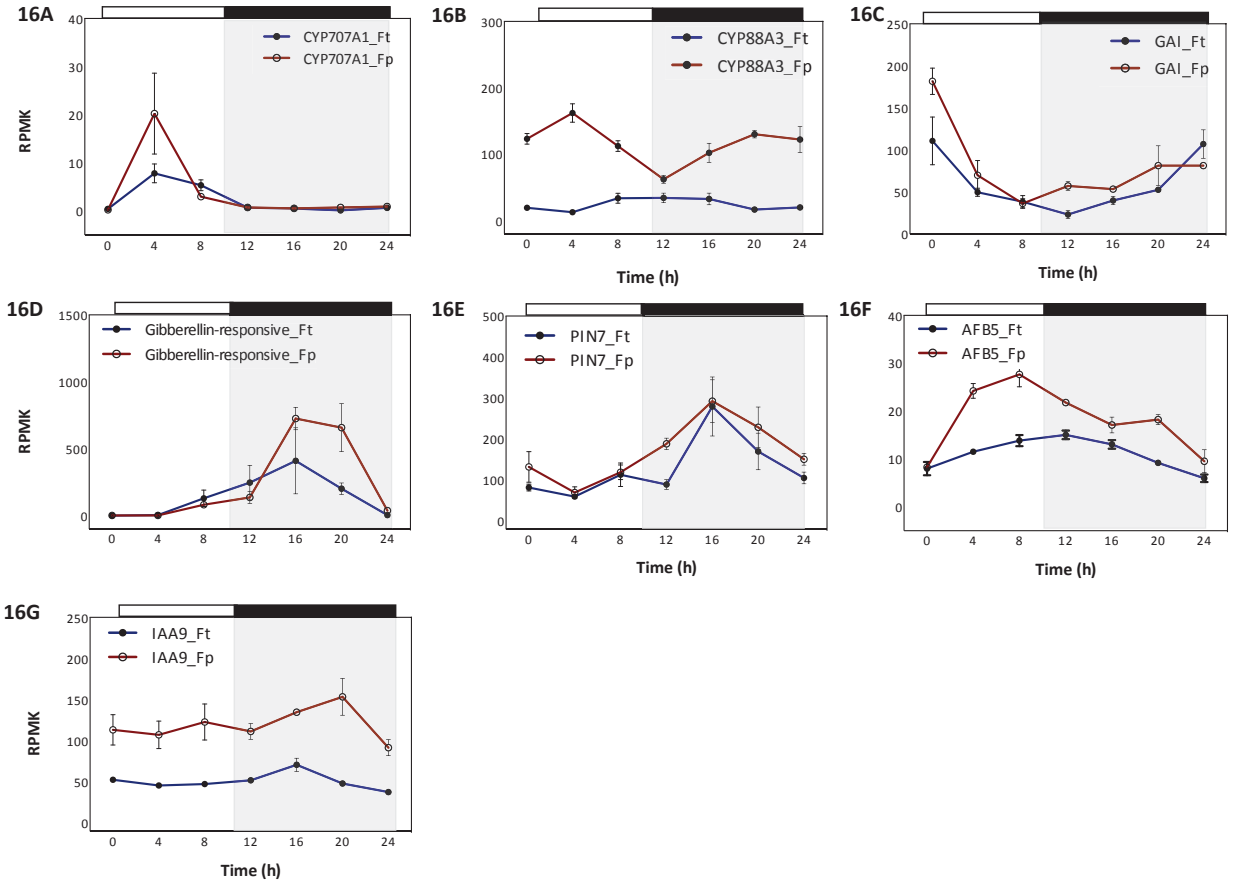
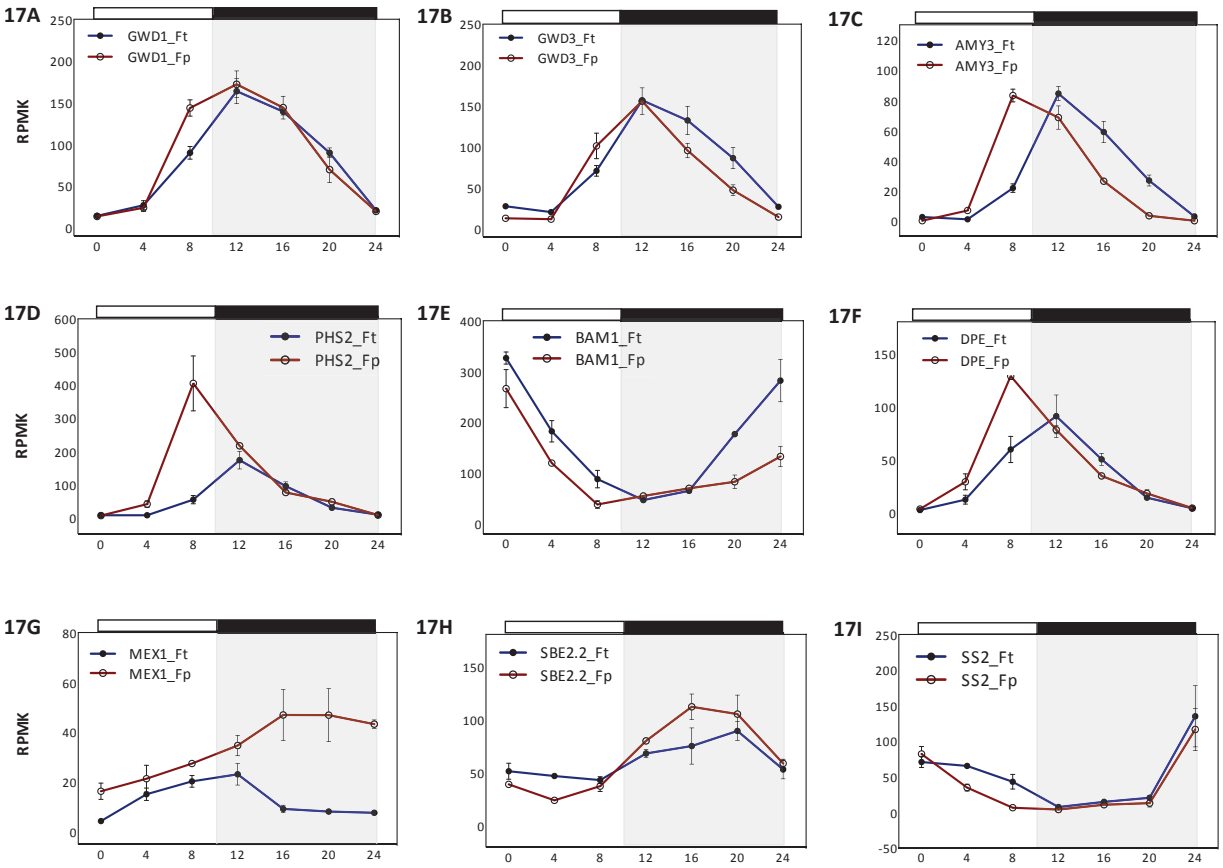


Figure 16. Changes of transcript abundances of genes involved in hormone metabolism during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 9. (**A**) CYP707A1 (**B**) CYP88A3 (**C**) GAI (**D**) GASA6 (**E**) PIN7 (**F**) AFB5 (**G**) IAA9


17A
GWD1_Ft
GWD1_Fp
RPMK
17B
GWD3_Ft
GWD3_Fp
17C
AMY3_Ft
AMY3_Fp
17D
PHS2_Ft
PHS2_Fp
RPMK
17E
BAM1_Ft
BAM1_Fp
17F
DPE_Ft
DPE_Fp
17G
MEX1_Ft
MEX1_Fp
RPMK
17H
SBE2.2_Ft
SBE2.2_Fp
17I
SS2_Ft
SS2_Fp

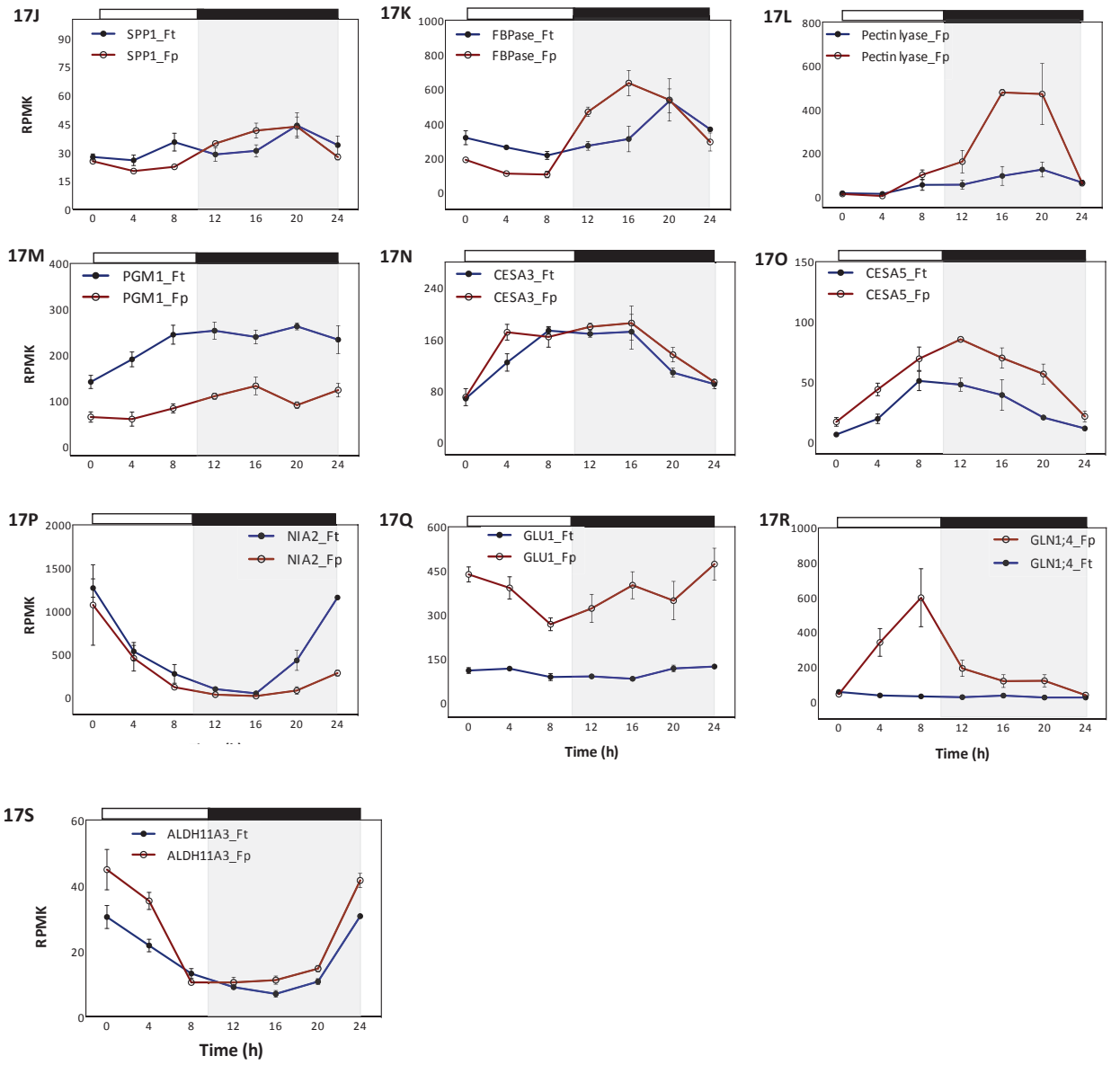


Figure 17. Changes of transcript abundances of genes involved in carbon and nitrogen metabolism during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 10. (**A**) GWD1 (**B**) GWD3 (**C**) AMY3 (**D**) PHS2 (**E**) BAM1 (**F**) DPE2 (**G**) MEX1 (**H**) SBE2.2 (**I**) SS2 (**J**) SPP1 (**K**) FBPase (**L**) Pectin lyase (**M**) PGM1 (**N**) CESA3 (**O**) CESA5 (**P**) NIA2 (**Q**) GLU1 (**R**) GLN1;4 (**S**) ALDH11A3.

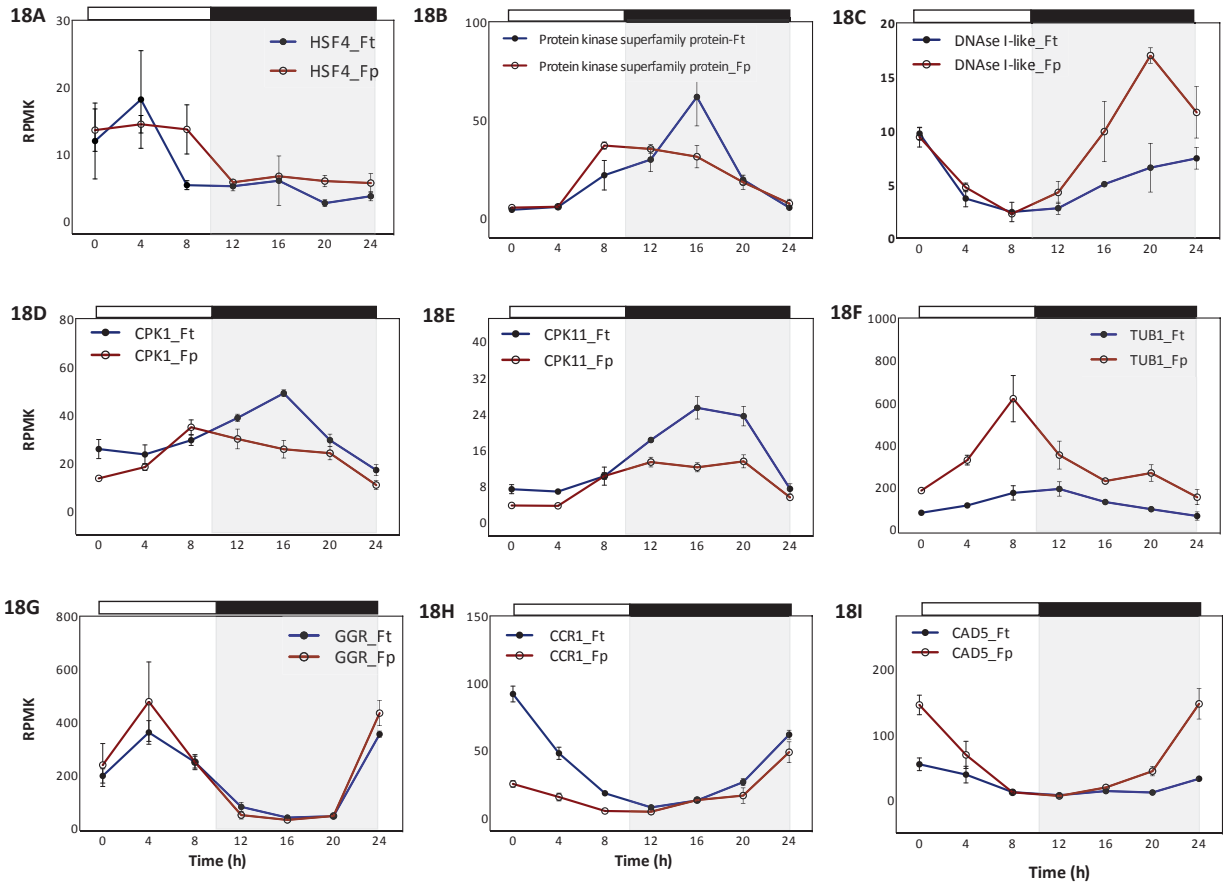


Figure 18: Changes of transcript abundances of genes involved in stress, protein post-translational modification, calcium signalling and in the secondary metabolism during the diurnal cycle in *F. pringlei* and *F. trinervia*. Each graph represents the changes in RPMK to diurnal changes of a gene from *F. trinervia* (blue) and *F. pringlei* (red). Growth conditions, evaluation of samples and statistical analysis were performed as described in figure 8 and summarized in Table 11. (**A**) HSF4 (**B**) Protein kinase superfamily protein (**C**) DNAse-I-like (**D**) CPK1 (**E**) CPK11 (**F**) TUB1 (**G**) GGR (**H**) CCR1 (**I**) CAD5

TABLES

Table 1: Expression of genes of the C_4 cycle
The significant differences between each time point were determined using a one-way ANOVA (tukey-test). N.S. (no asterisk, non significant) indicates $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$. *F. pringle*i (Fp) ; *F.trinervia* (Ft). Ratios above 3-fold are marked in red.

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Malate dehydrogenase	**MDH**	Fp	TCA /org transformation	4h to 20h: 1-fold n.s.	AT5G58330.1	8 A
Malate dehydrogenase	**MDH**	Ft	TCA /org transformation	4h to 20h: 1.5-fold**	AT5G58330.1	8 A
Pyruvate orthophosphate dikinase	**PPDK**	Fp	gluconeogenesis	4h to 16h: 1.2-fold n.s.	AT4G15530.5	8 B
Pyruvate orthophosphate dikinase	**PPDK**	Ft	gluconeogenesis	4h to 16h: 1.5-fold n.s.	AT4G15530.5	8 B
Carbonic anhydrase 1	**CA1**	Fp	TCA /org transformation	4h to 16h: 1-fold n.s.	AT3G01500.2	8 C
Carbonic anhydrase 1	**CA1**	Ft	TCA /org transformation	4h to 16h: 2-fold n.s.	AT3G01500.2	8 C
Alanine aminotransferase	**Ala-AT**	Fp	amino acid metabolism	4h to 16h: 1-fold n.s	AT1G17290.1	8 D
Alanine aminotransferase	**Ala-AT**	Ft	amino acid metabolism	4h to 16h: **3-fold****	AT1G17290.1	8 D
Aspartate aminotransferase	**Asp-AT**	Fp	amino acid metabolism	4h to 16h: 1-fold n.s.	AT4G31990.3	8 E
Aspartate aminotransferase	**Asp-AT**	Ft	amino acid metabolism	4h to 16h: 1-fold n.s.	AT4G31990.3	8 E
Pyruvate orthophosphate dikinase regulatory protein	**PPDK-RP**	Fp	not assigned.unknown	4h to 16h: 2-fold n.s	AT4G21210.1	8 F
Pyruvate orthophosphate dikinase regulatory protein	**PPDK-RP**	Ft	not assigned.unknown	4h to 16h: 1-fold n.s	AT4G21210.1	8 F
NADP malic enzyme 1	**NADP-ME1**	Fp	TCA /org transformation	0h to 8h: **24-fold****	AT2G19900.1	8 G
NADP malic enzyme 1	**NADP-ME1**	Ft	TCA /org transformation	0h to 12h: **11.5-fold****	AT2G19900.1	8 G
NADP malic enzyme 4	**NADP-ME4**	Fp	TCA /org transformation	0h to 20h: 1.8-fold n.s.	AT1G79750.1	8 H
NADP malic enzyme 4	**NADP-ME4**	Ft	TCA /org transformation	8h to 20h: 1.4-fold*	AT1G79750.1	8 H
NAD-dependent malic enzyme 1	**NAD-ME1**	Fp	TCA / org transformation	0h to 8h: 2-fold ****	AT2G13560.1	8 I
NAD-dependent malic enzyme 1	**NAD-ME1**	Ft	TCA / org transformation	0h to 12h: 1.8-fold ***	AT2G13560.1	8 I
NAD-dependent malic enzyme 2	**NAD-ME2**	Fp	TCA / org transformation	8h to 20h: 1.4-fold*	AT4G00570.1	8 J
NAD-dependent malic enzyme 2	**NAD-ME2**	Ft	TCA / org transformation	4h to 12h: 1.7-fold n.s.	AT4G00570.1	8 J
Phosphoenolpyruvate carboxykinase 1	**PCK1**	Fp	gluconeogenesis	4h to 16h: 1.4-fold n.s.	AT4G37870.1	8 K
Phosphoenolpyruvate carboxykinase 1	**PCK1**	Ft	gluconeogenesis	4h to 16h: 1.9-fold n.s.	AT4G37870.1	8 K
NAD-malate dehydrogenase	**NAD-MDH**	Fp	TCA / org transformation	0h to 8h: 2.1-fold *	AT3G47520.1	8 L
NAD-malate dehydrogenase	**NAD-MDH**	Ft	TCA / org transformation	0h to 8h: 1.3-fold n.s.	AT3G47520.1	8 L

Table 2: Expression of genes involved in the Calvin-Benson cycle

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
CP12 domain-containing protein 2	**CP12**	Fp	PS.calvin cycle	0h to 8h: 2-fold*	AT3G62410.1	9 A
CP12 domain-containing protein 2	**CP12**	Ft	PS.calvin cycle	0h to 8h: 2-fold***	AT3G62410.1	9 A
RUBISCO activase		Fp	PS.calvin cycle.rubisco interacting	0h to 12h: 2-fold n.s.	AT2G39730.1	9 B
RUBISCO activase		Ft	PS.calvin cycle.rubisco interacting	0h to 12h: 2-fold**	AT2G39730.1	9 B
phosphoribulokinase	**PRK**	Fp	PS.calvin cycle	8h to 16h: 2-fold*	AT1G32060.1	9 C
phosphoribulokinase	**PRK**	Ft	PS.calvin cycle	4h to 16h: 2-fold n.s.	AT1G32060.1	9 C
Inositol monophophatase /Fructose-1,6-bisphosphatase	**FBPase**	Fp	PS.carbohydrate metobolism.calvin cycle	0h to 8h: **5-fold****	AT5G64380.1	9 D
Inositol monophophatase /Fructose-1,6-bisphosphatase	**FBPase**	Ft	PS. carbohydrate metobolism.calvin cycle	0h to 8h: 1-fold n.s.	AT5G64380.1	9 D
D-ribulose-5-phosphate-3-epimerase	**RPE**	Fp	PS.calvin cycle	8h to 16h: 2-fold***	AT5G61410.1	9 E
D-ribulose-5-phosphate-3-epimerase	**RPE**	Ft	PS.calvin cycle	0h to 20h: 2-fold*	AT5G61410.1	9 E

Table 3: Expression of genes involved in the photorespiratory pathway

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Serine hydroxymethyl transferase 3	**SHMT3**	Fp	PS.photorespiration	12h to 24h: **4-fold******	AT4G32520.1	10 A
Serine hydroxymethyl transferase 3	**SHMT3**	Ft	PS.photorespiration	12h to 24h: 2-fold****	AT4G32520.1	10 A
serine hydroxymethyltransferase 2	**SHM2**	Fp	glycine hydroxymethyltransferase	0h to 4h: 1.6-fold n.s.	AT5G26780.3	10 B
serine hydroxymethyltransferase 2	**SHM2**	Ft	glycine hydroxymethyltransferase	0h to 4h: 1.2-fold n.s.	AT5G26780.3	10 B
D-isomer specific 2-hydroxyacid dehydrogenase family protein	**oxido-reductase family**	Fp	PS.photorespiration	4h to 16h: 2-fold****	AT1G79870.1	10 C
D-isomer specific 2-hydroxyacid dehydrogenase family protein	**oxido-reductase family**	Ft	PS.photorespiration	8h to 20h: 2-fold**	AT1G79870.1	10 C
2-phosphoglycolate phosphatase 1	**PGLP1**	Fp	PS.photorespiration	4h to 16h: 2-fold*	AT5G36700.1	10 D
2-phosphoglycolate phosphatase 1	**PGLP1**	Ft	PS.photorespiration	4h to 16h: 1-fold n.s.	AT5G36700.1	10 D
2-phosphoglycolate phosphatase 2	**PGLP2**	Fp	PS.photorespiration	0h to 20h: **4-fold ******	AT5G47760.1	10 E
2-phosphoglycolate phosphatase 2	**PGLP2**	Ft	PS.photorespiration	0h to 20h: **4-fold ******	AT5G47760.1	10 E
Glycine decarboxylase P-protein 2	**GLDP2**	Fp	PS.photorespiration	0h to 12h: 1-fold n.s.	AT2G26080.1	10 F
Glycine decarboxylase P-protein 2	**GLDP2**	Ft	PS.photorespiration	0h to 12h: 2-fold n.s.	AT2G26080.1	10 F
Hydroxypyruvate reductase	**HPR**	Fp	PS.photorespiration	8h to 16h: 1-fold **	AT1G68010.2	10 G
Hydroxypyruvate reductase	**HPR**	Ft	PS.photorespiration	12h to 24h: 2.5-fold **	AT1G68010.2	10 G

Table 4: Expression of genes involved in the light reaction

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Photosystem II light harvesting complex gene 2.2	**LHCB2.2**	Fp	PS.Lightreaction.photosystem II	4h to 20h: **50-fold******	AT2G05070.1	11 A
Photosystem II light harvesting complex gene 2.2	**LHCB2.2**	Ft	PS.Lightreaction.photosystem II	4h to 20h: **16-fold******	AT2G05070.1	11 A
Photosystem II light harvesting complex gene 6	**LHCB6**	Fp	PS.Lightreaction.photosystem II	8h to 16h: 1.7-fold n.s	AT1G19150.1	11 B
Photosystem II light harvesting complex gene 6	**LHCB6**	Ft	PS.Lightreaction.photosystem II	8h to 16h: 2.3-fold****	AT1G19150.1	11 B
Photosystem II light harvesting complex gene 4.2	**LHCB4.2**	Fp	PS.Lightreaction.photosystem II	8h to 20h: **3-fold ****	AT3G08940.2	11 C
Photosystem II light harvesting complex gene 4.2	**LHCB4.2**	Ft	PS.Lightreaction.photosystem II	8h to 20h: **4.6-fold****	AT3G08940.2	11 C
Photosystem II light harvesting complex gene B1B2	**Lhb (cab)**	Fp	PS.Lightreaction.photosystem II	8h to 16h: **12-fold******	AT2G34420.1	11 D
Photosystem II light harvesting complex gene B1B2	**Lhb (cab)**	Ft	PS.Lightreaction.photosystem II	8h to 16h: **14-fold****	AT2G34420.1	11 D
Photosystem I light harvesting complex gene 5	**LHC5**	Fp	PS.Lightreaction.photosystem I.LHC-I	4h to 16h: 2-fold n.s.	AT1G45474	11 E
Photosystem I light harvesting complex gene 5	**LHC5**	Ft	PS.Lightreaction.photosystem I.LHC-I	4h to 16h: 1-fold n.s.	AT1G45474	11 E
NDH-dependent cyclic electron flow	**NDF2**	Fp	PS.Lightreaction	4h to 16h: **5-fold****	AT1G64770.3	11 F
NDH-dependent cyclic electron flow	**NDF2**	Ft	PS.Lightreaction	0h to 20h: 2-fold**	AT1G64770.3	11 F
NDH-dependent cyclic electron flow 1	**NDF4**	Fp	PS.lightreaction.other electron carrier	4h to 16h: **3-fold****	AT3G16250.1	11 G
NDH-dependent cyclic electron flow 1	**NDF4**	Ft	PS.lightreaction.other electron carrier	0h to 16h: **3-fold***	AT3G16250.1	11 G
Encodes a subunit of the NAD(P)H complex located in the chloroplast thylakoid lumen	**PsbQ-like 1**	Fp	PS.Lightreaction.photosystem II	0h to 16h: **5-fold***	AT3G01440.1	11 H
Encodes a subunit of the NAD(P)H complex located in the chloroplast thylakoid lumen	**PsbQ-like 1**	Ft	PS.Lightreaction.photosystem II	0h to 20h: 2-fold*	AT3G01440.1	11 H
PSB-like protein 2	**PPL2**	Fp	PS.Lightreaction.PS II.polypeptide subunit	0h to 16h: **5-fold***	AT2G39470	11 I
PSB-like protein 2	**PPL2**	Ft	PS.Lightreaction.PS II.polypeptide subunit	0h to 4h: 2-fold**	AT2G39470	11 I
Photosystem II subunit T	**PSBTN**	Fp	PS.Lightreaction.PS II.polypeptide subunit	0h to 16h: **4-fold***	AT3G21055	11 J
Photosystem II subunit T	**PSBTN**	Ft	PS.Lightreaction.PS II.polypeptide subunit	0h to 16h: 2-fold n.s	AT3G21055	11 J
Photosystem II BY	**PSBY**	Fp	PS.Lightreaction.PS II.polypeptide subunit	4h to 16h: 2-fold n.s	AT1G67740	11 K
Photosystem II BY	**PSBY**	Ft	PS.Lightreaction.PS II.polypeptide subunit	4h to 16h: **3,5-fold****	AT1G67740	11 K
STATE TRANSITION7 protein kinase	**STN7**	Fp	PS.lightreaction.state transition	4h to 12h: **3-fold*****	AT1G68830.1	11 L
STATE TRANSITION7 protein kinase	**STN7**	Ft	PS.lightreaction.state transition	4h to 16h: **4-fold******	AT1G68830.1	11 L
2Fe-2S ferredoxin-like superfamily protein	**ferredoxin-related**	Fp	PS.lightreaction	4h to 16h: 1.5-fold n.s.	AT4G32590.3	11 M
2Fe-2S ferredoxin-like superfamily protein	**ferredoxin-related**	Ft	PS.lightreaction	4h to 16h: **7-fold*****	AT4G32590.3	11 M
PsbP family protein	**thylakoid lumenal**	Fp	PS.lightreaction. PSII	4h to 16h: 2-fold**	AT1G77090.1	11 N
PsbP family protein	**thylakoid lumenal**	Ft	PS.lightreaction. PSII	4h to 16h: **5-fold*****	AT1G77090.1	11 N

Table 5: Expression of circadian clock components or circadian-associated genes in *Arabidopsis*.

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
CCR-like	**CCL**	Fp	signalling.light	0h to 12h: **6-fold******	AT3G26740.1	12 A
CCR-like	**CCL**	Ft	signalling.light	0h to 12h: **5-fold***	AT3G26740.1	12 A
XAP5 CIRCADIAN TIMEKEEPER	**XCT**	Fp	signaling.light	0h to 20h: 2-fold**	AT2G21150.1	12 B
XAP5 CIRCADIAN TIMEKEEPER	**XCT**	Ft	signaling.light	4h to 16h: **4-fold***	AT2G21150.1	12 B
ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 3	**PRR3**	Fp	RNA.regulation of transcription	4h to 20h: **8-fold*****	AT5G60100.2	12 C
ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 3	**PRR3**	Ft	RNA.regulation of transcription	8h to 20h: **33-fold***	AT5G60100.2	12 C
ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 5	**PRR5**	Fp	RNA.regulation of transcription	8h to 16h: **9-fold******	AT5G24470.1	12 D
ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 5	**PRR5**	Ft	RNA.regulation of transcription	0h to 12h: **40-fold******	AT5G24470.1	12 D
ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 7	**PRR7**	Fp	RNA.regulation of transcription	8h to 20h: **3-fold*****	AT5G02810.1	12 E
ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 7	**PRR7**	Ft	RNA.regulation of transcription	8h to 20h: **3.5-fold****	AT5G02810.1	12 E
PHYTOCLOCK 1	**LUX/PCL1**	Fp	RNA.regulation of transcription	8h to 16h: **17-fold******	AT3G46640.3	12 F
PHYTOCLOCK 1	**LUX/PCL1**	Ft	RNA.regulation of transcription	12h to 20h: **34-fold****	AT3G46640.3	12 F
ZEITLUPE	**ZTL**	Fp	RNA.regulation of transcription	4h to 16h: 1-fold**	AT5G57360.2	12 G
ZEITLUPE	**ZTL**	Ft	RNA.regulation of transcription	4h to 16h: 1-fold n.s.	AT5G57360.2	12 G
Circadian clock associated 1	**CCA1**	Fp	RNA.regulation of transcription	0h to 16h: **39-fold******	AT2G46830.1	12 H
Circadian clock associated 1	**CCA1**	Ft	RNA.regulation of transcription	0h to 16h: **20-fold******	AT2G46830.1	12 H
TIMING OF CAB EXPRESSION 1	**TOC1**	Fp	RNA.regulation of transcription	4h to 16h: **51-fold******	AT5G61380.1	12 I
TIMING OF CAB EXPRESSION 1	**TOC1**	Ft	RNA.regulation of transcription	4h to 16h: **33-fold******	AT5G61380.1	12 I
LATE ELONGATED HYPOCOTYL	**LHY1**	Fp	RNA.regulation of transcription	0h to 16h: **265-fold******	AT1G01060.1	12 J
LATE ELONGATED HYPOCOTYL	**LHY1**	Ft	RNA.regulation of transcription	0h to 16h: **78-fold******	AT1G01060.1	12 J
Light Insensitive Period1	**LIP1**	Fp	signaling.G-proteins	0h to 8h: **3-fold****	AT5G64813.1	12 K
Light Insensitive Period1	**LIP1**	Ft	signaling.G-proteins	4h to 16h: 2-fold*	AT5G64813.1	12 K
Phytochrome interacting factor 3	**PIF3**	Fp	signaling.light	0h to 16h: **4-fold******	AT1G09530.1	12 L
Phytochrome interacting factor 3	**PIF3**	Ft	signaling.light	0h to 16h: 2-fold***	AT1G09530.1	12 L
TIME FOR COFFEE	**TIC**	Fp	signaling.light	0h to 16h: 2-fold*	AT3G22380.2	12 M
TIME FOR COFFEE	**TIC**	Ft	signaling.light	4h to 16h: 2-fold***	AT3G22380.2	12 M
GIGANTEA	**GI**	Fp	development	4h to 20h: **3-fold*****	AT1G22770.1	12 N
GIGANTEA	**GI**	Ft	development	0h to 12h: **9-fold******	AT1G22770.1	12 N
EARLY FLOWERING 3	**ELF3**	Fp	RNA.regulation of transcription	4h to 16h: 5-fold n.s.	AT2G25930.1	12 O
EARLY FLOWERING 3	**ELF3**	Ft	RNA.regulation of transcription	4h to 16h: **8-fold******	AT2G25930.1	12 O
EARLY FLOWERING 4	**ELF4**	Fp	signaling.light	0h to 12h: 1.5-fold*	AT2G40080.1	12 P
EARLY FLOWERING 4	**ELF4**	Ft	signaling.light	0h to 12h: 1.7-fold*	AT2G40080.1	12 P

Table 6: Expression of transcripts of genes involved in light signaling pathways.

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Suppressor of phyA-105 protein family	**SPA3**	Fp	signaling.light	4h to 20h: 2-fold****	AT3G15354.1	13 A
Suppressor of phyA-105 protein family	**SPA3**	Ft	signaling.light	4h to 16h: **5-fold****	AT3G15354.1	13 A
Phototropic-responsive NPH3 family protein	**NPH3**	Fp	signaling.light	0h to 8h: **5-fold****	AT5G67385.1	13 B
Phototropic-responsive NPH3 family protein	**NPH3**	Ft	signaling.light	4h to 16h: 2-fold**	AT5G67385.1	13 B
Phototropin 1	**PHOT1**	Fp	protein.postranslational modification	0h to 12h: **5-fold****	AT3G45780.1	13 C
Phototropin 1	**PHOT1**	Ft	protein.postranslational modification	0h to 16h: **3.5-fold***	AT3G45780.1	13 C
Phototropin 2	**PHOT2**	Fp	protein.postranslational modification	4h to 20h: 2-fold***	AT5G58140.2	13 D
Phototropin 2	**PHOT2**	Ft	protein.postranslational modification	4h to 24h: 1-fold*	AT5G58140.2	13 D
Light-harvesting-like protein	**LIL3**	Fp	RNA. regulation	12h to 24h: 2-fold**	AT5G47110.1	13 E
Light-harvesting-like protein	**LIL3**	Ft	RNA.regulation	4h to 16h: **4-fold****	AT5G47110.1	13 E

Table 7: Expression of transcripts of genes involved in transport

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
sulfate transporter	**SULTR4**	Fp	transport.sulphate	16h to 24h: **7-fold******	AT5G13550.1	14 A
sulfate transporter	**SULTR4**	Ft	transport.sulphate	16h to 24h: **8-fold****	AT5G13550.1	14 A
HCO_3^--transporter		Fp	transport	0h to 20h: **3-fold***	AT4G32510.1	14 B
HCO_3^--transporter		Ft	transport	0h to 12h: **35-fold******	AT4G32510.1	14 B
phosphate transporter 2;1	**PHT2;1**	Fp	transport.phosphate	4h to 12h: **4-fold****	AT3G26570.1	14 C
phosphate transporter 2;1	**PHT2;1**	Ft	transport.phosphate	24h to 16h: **12-fold****	AT3G26570.1	14 C
phosphate transporter 4;4	**PHT4;4**	Fp	transport.sugars	0h to 12h: 3-fold n.s.	AT4G00370.1	14 D
phosphate transporter 4;4	**PHT4;4**	Ft	transport.sugars	0h to 12h: **3-fold******	AT4G00370.1	14 D
Phosphate transporter traffic facilitator1	**PHF1**	Fp	development	12h to 4h: **4-fold******	AT3G52190.1	14 E
Phosphate transporter traffic facilitator1	**PHF1**	Ft	development	4h to 16h: **7-fold******	AT3G52190.1	14 E
Potassium transport 2/3	**AKT**	Fp	transport.potassium	4h to 20h: **4-fold****	AT4G22200.1	14 F
Potassium transport 2/3	**AKT**	Ft	transport.potassium	0h to 16h: **3-fold****	AT4G22200.1	14 F
Plasma membrane intrinsic protein 1C	**PIP1C**	Fp	transport.	0h to 8h: **4.5-fold ******	AT1G01620.1	14 G
Plasma membrane intrinsic protein 1C	**PIP1C**	Ft	transport.	0h to 12h: 4.5-fold n.s.	AT1G01620.1	14 G
Nitrate transporter 1.5	**NRT1**	Fp	transport.peptides	0h to 12h: **12-fold******	AT1G32450.1	14 H
Nitrate transporter 1.5	**NRT1**	Ft	transport.peptides	0h to 12h: **9-fold******	AT1G32450.1	14 H
Hexose transporter	**STP13**	Fp	transport.sugars	4h to 16h: **8-fold******	AT5G26340.1	14 I
Hexose transporter	**STP13**	Ft	transport.sugars	4h to 16h: **3.5-fold****	AT5G26340.1	14 I
Sugar transporter 1	**STP1**	Fp	transport.sugars	4h to 16h: 2-fold n.s	AT1G11260.1	14 J
Sugar transporter 1	**STP1**	Ft	transport.sugars	4h to 16h: 2-fold n.s	AT1G11260.1	14 J
Plastidic glucose transporter	**GLT1**	Fp	major CHO metabolism.degradation. starch.transporter	0h to 16h: 2-fold **	AT5G16150.1	14 K
Plastidic glucose transporter	**GLT1**	Ft	major CHO metabolism.degradation. starch.transporter	0h to 16h: 1-fold n.s	AT5G16150.1	14 K
Dicarboxylate transport 2.1	**Dit2.1**	Fp	transport.metabolite transporters	4h to 16h: 1.5-fold n.s.	AT5G64290.1	14 L
Dicarboxylate transport 2.1	**Dit2.1**	Ft	transport.metabolite transporters	4h to 16h: **4-fold ***	AT5G64290.1	14 L
Sodium Bile acid symporter	**BASS2**	Fp	transport	4h to 16h: 1.8-fold n.s.	AT2G26900.1	14 M
Sodium Bile acid symporter	**BASS2**	Ft	transport	4h to 16h: 2.5-fold ****	AT2G26900.1	14 M

Table 8: Expression of transcripts of genes involved in RNA regulation. Transcription factors

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Acyl-CoA N-acyltransferase		Fp	RNA.regulation of transcription.PHD finger transcription factor	8h to 24h: 2-fold*	AT1G05380.1	15 A
Acyl-CoA N-acyltransferase		Ft	RNA.regulation of transcription.PHD finger transcription factor	8h to 20h: 2-fold****	AT1G05380.1	15 A
Golden2-like 1	**GLK1**	Fp	RNA.regulation of transcription	0h to 16h: **7-fold ******	AT2G20570.2	15 B
Golden2-like 1	**GLK1**	Ft	RNA.regulation of transcription	0h to 16h: **4-fold ***	AT2G20570.2	15 B
ETHYLENE-INSENSITIVE 3	**EIN3**	Fp	RNA.regulation of transcription	0h to 16h: 1-fold n.s.	AT3G20770.1	15 C
ETHYLENE-INSENSITIVE 3	**EIN3**	Ft	RNA.regulation of transcription	4h to 16h: 2-fold *	AT3G20770.1	15 C
Homeobox 1	**HB-1**	Fp	RNA.regulation of transcription	12h to 24h: **6-fold****	AT3G01470.1	15 D
Homeobox 1	**HB-1**	Ft	RNA.regulation of transcription	4h to 16h: 3-fold n.s.	AT3G01470.1	15 D
Sigma factor 1	**SIG1**	Fp	RNA.regulation of transcription	4h to 16h: 2-fold *	AT1G64860.1	15 E
Sigma factor 1	**SIG1**	Ft	RNA.regulation of transcription	4h to 16h: 2.3-fold *	AT1G64860.1	15 E
Sigma factor 2	**SIG2**	Fp	RNA.regulation of transcription	0h to 12h: **4-fold ******	AT1G08540.1	15 F
Sigma factor 2	**SIG2**	Ft	RNA.regulation of transcription	0h to 12h: 2.6-fold **	AT1G08540.1	15 F
Sigma factor 5	**SIG5**	Fp	RNA.regulation of transcription	0h to 16h: **9-fold ******	AT5G24120.1	15 G
Sigma factor 5	**SIG5**	Ft	RNA.regulation of transcription	0h to 16h: **8-fold ****	AT5G24120.1	15 G
Sigma factor 6	**SIG6**	Fp	RNA.regulation of transcription	0h to 16h: **3-fold ******	AT2G36990.1	15 H
Sigma factor 6	**SIG6**	Ft	RNA.regulation of transcription	0h to 12h: 2-fold *	AT2G36990.1	15 H

Table 9: Expression of transcripts of genes involved in hormone metabolism.

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Abscisic acid 8´-hydroxylase 1	**CYP707A1**	Fp	hormone metabolism.abscisic acid	4h to 16h: **35-fold****	AT4G19230.2	16 A
Abscisic acid 8´-hydroxylase 1	**CYP707A1**	Ft	hormone metabolism.abscisic acid	4h to 16h: 15-fold n.s.	AT4G19230.2	16 A
Ent-kaurenoic acid oxidase 1	**CYP88A3**	Fp	hormone metabolism.gibberelin	4h to 12h: 2.6-fold****	AT1G05160.1	16 B
Ent-kaurenoic acid oxidase 1	**CYP88A3**	Ft	hormone metabolism.gibberelin	4h to 16h: 2.6-fold n.s.	AT1G05160.1	16 B
GRAS family transcription factor: GIBBERELLIC ACID INSENSITIVE	**GAI**	Fp	hormone metabolism.gibberelin	0h to 12h: **3-fold****	AT1G14920.1	16 C
GRAS family transcription factor: GIBBERELLIC ACID INSENSITIVE	**GAI**	Ft	hormone metabolism.gibberelin	0h to 12h: **5-fold****	AT1G14920.1	16 C
Gibberellin-regulated family protein	**GASA6**	Fp	hormone metabolism.gibberelin	4h to 16h: **500-fold****	AT1G74670.1	16 D
Gibberellin-regulated family protein	**GASA6**	Ft	hormone metabolism.gibberelin	4h to 16h: 82-fold n.s.	AT1G74670.1	16 D
PIN-FORMED 7	**PIN7**	Fp	hormone metabolism.auxin	4h to 16h: **4-fold****	AT1G23080.1	16 E
PIN-FORMED 7	**PIN7**	Ft	hormone metabolism.auxin	4h to 16h: **4.6-fold****	AT1G23080.1	16 E
Auxin F-box protein 5	**AFB5**	Fp	hormone metabolism.auxin	0h to 8h: **3.3-fold****	AT5G49980.1	16 F
Auxin F-box protein 5	**AFB5**	Ft	hormone metabolism.auxin	12h to 24h:2.5-fold**	AT5G49980.1	16 F
INDOLE-3-ACETIC ACID INDUCIBLE 9	**IAA9**	Fp	hormone metabolism.auxin	4h to 16h: 1.2-fold n.s	AT5G65670.1	16 G
INDOLE-3-ACETIC ACID INDUCIBLE 9	**IAA9**	Ft	hormone metabolism.auxin	4h to 16h: 1.5-fold n.s	AT5G65670.1	16 G

Table 10: Expression of transcripts of genes involved in carbon and nitrogen metabolism.

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Glucan water dikinase 1	**GDW1**	Fp	major CHO metabolisms.degradation.starch	12h to 24h: **9-fold******	AT1G10760.1	17 A
Glucan water dikinase 1	**GDW1**	Ft	major CHO metabolisms.degradation.starch	12h to 24h: **8-fold******	AT1G10760.1	17 A
Glucan water dikinase 3	**GDW3**	Fp	major CHO metabolisms.degradation.starch	12h to 24h: **11-fold******	AT5G26570.1	17 B
Glucan water dikinase 3	**GDW3**	Ft	major CHO metabolisms.degradation.starch	12h to 24h: **6-fold******	AT5G26570.1	17 B
Alpha-amylase-like 3	**AMY3**	Fp	major CHO metabolisms.degradation.starch	12h to 24h: **150-fold******	AT1G69830.1	17 C
Alpha-amylase-like 3	**AMY3**	Ft	major CHO metabolisms.degradation.starch	12h to 24h: **26-fold******	AT1G69830.1	17 C
α-glucan phopshorylase 2	**PHS2**	Fp	major CHO metabolisms.degradation.starch	0h to 8h: **66-fold******	AT3G46970.1	17 D
α-glucan phopshorylase 2	**PHS2**	Ft	major CHO metabolisms.degradation.starch	0h to 12h: **24-fold****	AT3G46970.1	17 D
β-amylase 1	**BAM1**	Fp	major CHO metabolisms.degradation.starch	0h to 16h: **4-fold******	AT3G23920.1	17 E
β-amylase 1	**BAM1**	Ft	major CHO metabolisms.degradation.starch	0h to 16h: **5-fold******	AT3G23920.1	17 E
Disproportionating enzyme 2	**DPE2**	Fp	major CHO metabolism.degradation.starch	0h to 8h: **37-fold******	AT2G40840.1	17 F
Disproportionating enzyme 2	**DPE2**	Ft	major CHO metabolism.degradation.starch	0h to 12h: **36-fold******	AT2G40840.1	17 F
Maltose transporter 1	**MEX1**	Fp	major CHO metabolism.degradation.starch	0h to 16h: **3-fold****	AT5G17520.1	17 G
Maltose transporter 1	**MEX1**	Ft	major CHO metabolism.degradation.starch	0h to 12h: 5-fold n.s.	AT5G17520.1	17 G
Starch branching enzyme 2.2	**SBE2.2**	Fp	major CHO metabolism.synthesis.starch	4h to 16h: **5-fold******	AT5G03650.1	17 H

Table 11: Expression of transcripts of genes involved in other biological pathways.

Name	Symbol	Species	Mapman_bin	ANOVA	AGI ID	Fig.
Heat shock factor 4	**HSF4**	Fp	stress.abiotic.heat	4h to 20h: 2.4-fold n.s.	AT4G36990.1	18 A
Heat shock factor 4	**HSF4**	Ft	stress.abiotic.heat	4h to 20h: **6-fold***	AT4G36990.1	18 A
Protein kinase superfamily protein		Fp	protein.postranslational modification	0h to 16h: 2-fold**	AT3G59110.1	18 B
Protein kinase superfamily protein		Ft	protein.postranslational modification	0h to 16h: **6-fold***	AT3G59110.1	18 B
DNAse I-like superfamily protein		Fp	DNA.synthesis/chromatin structure	8h to 20h: **7-fold****	AT2G48030.1	18 C
DNAse I-like superfamily protein		Ft	DNA.synthesis/chromatin structure	0h to 8h: **4-fold***	AT2G48030.1	18 C
calcium dependent protein kinase 1	**CPK1**	Fp	signaling.calcium	0h to 12h: 2-fold*	AT5G04870.1	18 D
calcium dependent protein kinase 1	**CPK1**	Ft	signaling.calcium	4h to 16h: 2-fold****	AT5G04870.1	18 D
calcium dependent protein kinase 2	**CPK11**	Fp	signaling.calcium	4h to 16h: **3.4-fold****	AT1G35670.1	18 E
calcium dependent protein kinase 2	**CPK11**	Ft	signaling.calcium	4h to 16h: **3.7-fold****	AT1G35670.1	18 E
Tubulin beta-1 chain	**TUB1**	Fp	cell.organisation	0h to 8h: **3-fold****	AT1G75780.1	18 F
Tubulin beta-1 chain	**TUB1**	Ft	cell.organisation	0h to 8h: 2-fold n.s.	AT1G75780.1	18 F
Geranylgeranyl reductase	**GGR**	Fp	secondary metabolism.isoprenoids. non-mevalonate pathway	4h to 16h: **15-fold****	AT1G74470.1	18 G
Geranylgeranyl reductase	**GGR**	Ft	secondary metabolism.isoprenoids. non-mevalonate pathway	4h to 16h: **9-fold***	AT1G74470.1	18 G
Cinnamoyl coa reductase 1	**CCR1**	Fp	secondary metabolism. phenylpropanoids.lignin biosynthesis	0h to 12h: **6-fold****	AT1G15950.1	18 H
Cinnamoyl coa reductase 1	**CCR1**	Ft	secondary metabolism. phenylpropanoids.lignin biosynthesis	0h to 12h: **12-fold****	AT1G15950.1	18 H
Cinnamyl alcohol dehydrogenase 5	**CAD5**	Fp	secondary metabolism. phenylpropanoids.lignin biosynthesis	0h to 16h: **7.5-fold****	AT4G34230.1	18 I
Cinnamyl alcohol dehydrogenase 5	**CAD5**	Ft	secondary metabolism. phenylpropanoids.lignin biosynthesis	0h to 16h: 4-fold n.s.	AT4G34230.1	18 I

AUTHOR CONTRIBUTIONS

U.G. and **S.H.A** assembled the transcriptomes

P.W. and **U.G.** designed the project

ACKNOWLEDGEMENTS

With sincerity I would like to acknowledge my **committee members**:

Prof. Dr. Peter Westhoff to which I owe my deepest gratitude. Thanks for the guidance, advices, confidence and independency which I received. Thanks to impart your severity, enthusiasm and profound knowledge in research. Thanks for being the classical "Dr. Vater".

Prof. Dr. Georg Groth to have reviewed the manuscript. Thanks for the critical feedback on my topic and for the brainstorming.

Prof. PhD. Thomas D. Sharkey thanks for the discussions, experimental feedback, advices, cooperation and hospitality.

In addition, I wish to thank several persons that made this work possible:

Prof. Dr. Andreas Weber to have given me the opportunity to be part of this privileged interdisciplinary research program. I always admired your dynamic and broad-minded character; qualities you implemented into this program.

Prof. PhD. Barbara B. Sears to have provided us with "survival equipments" and thanks for the support.

Prof. Dr. Tina Romeis for being a long-lasting considerable inspiration.

The Westhoff lab:

Thanks **to** the team for charitableness, nice conversations, celebrations and consideration.
Christian Wiludda for your well preserved qualities, all the support and funny moments.
Chris for nice conversations.
Florian for the fun and the delivery of mountains of chocolate. Thanks for your commitment as my "FedEx Courier".
Sandra thanks to have provided me with "Energienahrung" and for your work in handicrafts.
Maria and **Monika** and the friendly **gardeners** to have taken care of the *Flaveria* plants. Maria for distracting conversations.
Heidi Schwan for the arrangements and caring nature.
Udo for the useful suggestions, skills, support and the nice supervision.
Borjana for your scientific and team skills. Thanks for the irreplaceable support, patience and critical feedback.
Julia for your time and conversations.
The Leveys for sharing the break of dawn.
Susanne and **Myles** to have taken hands-on during the accident (s).

The **Sharkey lab**:

James for your scientific inspiration, adorable skills and your simple way of explaining.
Aparajita for your sentimental and empathetic nature, and for your admiration.
Chris for the nice time at MSU and in Germany. Thanks for your support.
Sean & family thanks for the great time. To have introduced us to American culture, habits and humor. Thanks for the effort in entertaining Merlin and for the reassurance. Thanks for your time in editing the manuscript, and yes, for all the good-humored communications.

The **iGRAD *plant* fellows** and **faculties**,
to have brought the german culture and humor closer to me. Thanks for all the fun and vitality.

Sigrun and **Prof. PhD. Susanne Hoffmann-Benning** for the effort it took to organize the stay at MSU. Sigrun for your support and disposability, while knowing what it means to be a mother.

Judy, **Vivi, Tobi, René, Sarah R., Christopher, Sarah V., Dennis, Chrissy, Jamie, Fabi, Lena, Katie, Sebastian and Tom** to have made our time at MSU an unforgettable and substantial one!

Wei for your frankness and your lucid mind.

Freddy, **Dominik B.** and **Sarah R.** for the great and supportive time you have given to Merlin.

Dominik B. to have stayed up late during the circadian experiment.

René for your humored character and your hidden qualities.

Judy for your heartiness. For the passionated exchange of experimental ideas and materials. Thanks for your cooperative support and for your esprit. I appreciated your smooth, but still intense way of being a scientist - keep it up!

Viviana my sweet sister heart. For being a special person with a clear perspective of life and a genuine spirit. That you were there for me in "the bad or good days" always with your recreational humor and composure. Thanks for the sharing - of precious moments and life lessons. Thanks for touching souls.

The "**Strümpell-Dorf**", the "**Pony-Hof**" and the "**Spartan Village** " community for the fun, support and all the sharing.

Dominik Jozic for the company, the empathetic regard, "das weiche Fallen".

Friederike & family Schoeller/Frey for the everlasting friendship and faith in me. For your irreplaceable charitableness, supportive reception and distractions from daily routine.

Thanks to all the members of my German, British, Polish and Portuguese **family** for their Love. That although from the distance, would always remind me of the persons I belong to. Thanks for having me in mind. Thanks for all the admiration I receive and that I have for you. For the care and support for Merlin!

Merlin for going hand-in-hand with me to reach the goal. For making me into the person I am today. For being my constancy, my real concern and my deepest meaning.

I would like to thank everyone I mentioned, and additional friends that shared with me this time, for their human kindness enabling unforgettable moments.

Financial support was provided by the German Research Foundation (DFG) /IRGT1525 iGRAD-Plant