Evolution and diversity of photosynthetic metabolism in C_3, C_3-C_4 intermediate and C_4 plants

---

Gian Luca Borghi

Cumulative Thesis

For the obtainment of the academic title
"Doctor rerum naturalium"
(Dr. rer. nat.)
in scientific discipline "Molecular Plant Physiology"

Disputed on 18/08/2021

submitted to:
Faculty of Mathematics and Natural Sciences
Institute for Biochemistry and Biology
University of Potsdam

Thesis title: Evolution and diversity of photosynthetic metabolism in C3, C3-C4 intermediate and C4 plants
Author / Doctoral student: Gian Luca Borghi
Main supervisor: Prof. Dr. Mark Stitt
Affiliations: MPI of Molecular Plant Physiology – University of Potsdam

Thesis reviewers: Prof. Dr. Mark Stitt – Prof. Dr. Andrea Bräutigam – Prof. Dr. Thomas Sharkey

Date of disputation: 18/08/2021
Doctoral disputation committee members: Prof. Dr. Michael Lenhard (chair) – Prof. Dr. Markus Grebe – Prof. Dr. Alisdair Fernie

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-52220
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-522200

Acknowledgements

(ENGLISH)

During these many years of doctoral studies I have amassed a long list of people that I would like to thank, and to whom I owe, in large or small part, the success of this enterprise.

Heartfelt thanks go to my supervisors Mark Stitt, John Lunn and Stéphanie Arrivault, whom have taught and guided me for these years. In particular Mark, for his final push which led me to write and complete this thesis. Outside the circle of supervisors, I also thank very much Martha Ludwig, Alisdair Fernie, Joachim Kopka and Ina Talke for their valuable advice, both inside and outside the PhD Advisory Committee.

I cannot fail to thank the numerous colleagues of the Stitt research group; among the many, a special mention goes, for friendship and help, to Manu, Jana, Thiago, Ginie, Rúben, Sandi, Chiara, Luiza, Armin and Hiro.

Life inside the research group is not everything, therefore I want to thank other people whom, inside or outside the MPI, have made these years in Potsdam more beautiful. A special mention goes to Vicky, who managed to make me appreciate running and kept me in good spirits. Another one goes to the whole bunch of aficionados whom, every month, came to the Board Game Nights that I organized, before COVID stroke.

At last, but not least, I want to thank all my familiar friendly baggage that I unfortunately left in Italy, and to which, I hope, I will be reunited soon. I am immensely grateful to my parents, who have supported me from afar and have never deprived me of their unconditional affection. I am also very grateful to the rest of the family which, in each and every moments of reunion, made me feel as if I had never left home. Finally, like the icing on the cake, I want to thank the best company of friends I could wish for, “la Ciiurma”, whom have always kept me close, despite the distance.

Ringraziamenti

(ITALIANO)
Durante questi numerosi anni di dottorato ho accumulato una lunga lista di persone che vorrei ringraziare, e a cui devo, in grande o piccola parte, il successo di questa impresa.
I ringraziamenti più sentiti vanno ai miei supervisori Mark Stitt, John Lunn e Stéphanie Arrivault, i quali mi hanno insegnato e guidato per questi anni. In particolare Mark, per la sua spinta finale che mi ha indotto a scrivere ed ultimare questa tesi. Fuori dalla cerchia dei supervisori, ringrazio molto anche Martha Ludwig, Alisdair Fernie, Joachim Kopka ed Ina Talke per i preziosi consigli, fuori e dentro il "PhD Advisory Committee".
Non posso esimermi dal ringraziare i numerosi colleghi del gruppo di ricerca Stitt; tra i tanti, una menzione speciale va, per amicizia ed aiuto, a Manu, Jana, Thiago, Ginie, Rúben, Sandi, Chiara, Luiza, Armin ed Hiro.
La vita dentro il gruppo di ricerca non è tutto, per questo ci tengo a ringraziare altre persone che, fuori o dentro il MPI, hanno reso questi anni a Potsdam più belli: un grazie particolare va a Vicky, che è riuscita a farmi apprezzare la corsa e tenermi su di morale, e poi a tutta combriccola di affezionati che ogni mese si trovava alle Serate Giochi da Tavolo che organizzavo, prima che il COVID colpisse.
Alla fine, ma non meno importanti, voglio ringraziare tutto il mio bagaglio di famiglia e amici che purtroppo ho lasciato in Italia, e che spero presto di raggiungere. Sono immensamente grato ai miei genitori, che mi hanno supportato da lontano e non mi hanno mai fatto mancare il loro affetto incondizionato. Sono molto grato anche al resto della famiglia che, nei momenti di ricongiungimento, mi ha sempre fatto sentire come se non me ne fossi mai andato da casa. Per finire, come la ciliegina sulla torta, voglio ringraziare la più bella compagnia di amici che possa desiderare, la Ciurma, che in questi anni mi ha sempre tenuto vicino, nonostante la distanza.

It is hereby declared that this doctoral thesis and the related scientific research was not submitted to any other university or similar institution, and it has been carried out independently and exclusively with the means specified.

Signature of the Doctoral student Signature of the Principal Supervisor

Abstract

In C_3 plants, CO_2 diffuses into the leaf and is assimilated by the Calvin-Benson cycle in the mesophyll cells. It leaves Rubisco open to its side reaction with O_2, resulting in a wasteful cycle known as photorespiration. A sharp fall in atmospheric CO_2 levels about 30 million years ago have further increased the side reaction with O_2. The pressure to reduce photorespiration led, in over 60 plant genera, to the evolution of a CO_2-concentrating mechanism called C_4 photosynthesis; in this mode, CO_2 is initially incorporated into 4-carbon organic acids, which diffuse to the bundle sheath and are decarboxylated to provide CO_2 to Rubisco. Some genera, like *Flaveria*, contain several species that represent different steps in this complex evolutionary process. However, the majority of terrestrial plant species did not evolve a CO_2-concentrating mechanism and perform C_3 photosynthesis.

This thesis compares photosynthetic metabolism in several species with C_3, C_4 and intermediate modes of photosynthesis. Metabolite profiling and stable isotope labelling were performed to detect inter-specific differences changes in metabolite profile and, hence, how a pathway operates. The results obtained were subjected to integrative data analyses like hierarchical clustering and principal component analysis, and were deepened by correlation analyses to uncover specific metabolic features and reaction steps that were conserved or differed between species.

The main findings are that Calvin-Benson cycle metabolite profiles differ between C_3 and C_4 species and between different C_3 species, including a very different response to rising irradiance in Arabidopsis and rice. These findings confirm Calvin-Benson cycle operation diverged between C_3 and C_4 species and, most unexpectedly, even between different C_3 species. Moreover, primary metabolic profiles supported the current C_4 evolutionary model in the genus *Flaveria* and also provided new insights and opened up new questions. Metabolite profiles also point toward a progressive adjustment of the Calvin-Benson cycle during the evolution of C_4 photosynthesis. Overall, this thesis point out the importance of a metabolite-centric approach to uncover underlying differences in species apparently sharing the same photosynthetic routes and as a valid method to investigate evolutionary transition between C_3 and C_4 photosynthesis.

Table of contents

List of abbreviations

This list contains all the abbreviations included in the General Introduction, General Discussion and Conclusions of this thesis, including those present in figures. Abbreviation included in the three manuscripts are not included, having their own abbreviation referral.

The abbreviations are listed below in alphabetical order.

1,3-PGA = 1,3-bisphosphoglycerate
2OG = 2-oxoglutarate
2PG = 2-phosphoglycolate
3PGA = 3-phosphoglycerate
ADP = adenosine diphosphate
Ala = alanine
AlaAT = alanine aminotransferase
AMP = adenosine monophosphate
Asn = asparagine
Asp = aspartate
AspAT = aspartate aminotransferase
ATP = adenosine triphosphate
BS = bundle sheath
CA = carbonic anhydrase
CAM = Crassulacean acid metabolism
CBC = Calvin-Benson cycle
CCM = CO_2-concentrating mechanism
CO_2 = carbon dioxide
DHAP = dihydroxyacetone phosphate
DNA = Deoxyribonucleic acid
E4P = erythrose 4-phosphate
F6P = fructose 6-phosphate
FAO = Food and Agriculture Organization of the United Nations
FBA = fructose-1,6-bisphosphate aldolase
FBP = fructose 1,6-bisphosphate
FBPase = fructose-1,6-bisphosphatase
GAP = glyceraldehyde 3-phosphate
GAPDH = glyceraldehyde-3-phosphate dehydrogenase
GDC = glycine decarboxylase
GGT = glutamate-glyoxylate aminotransferase
GLDC = glycine decarboxylase complex
Gln = glutamine
Glu = glutamate
GLYK = glycerate 3-kinase
GOGAT = glutamine oxoglutarate aminotransferase
GOX = glycolate oxidase
HPR1 = hydroxypyruvate reductase (peroxisomal)
HPR2 = hydroxypyruvate reductase (cytosolic)
LC-MS/MS = liquid chromatography coupled with tandem mass spectrometry

NAD = nicotinamide adenine dinucleotide
NADP = nicotinamide adenine dinucleotide phosphate
NADP-MDH = NADP-dependent malate dehydrogenase
NADP-ME = NADP-dependent malic enzyme
NUE = nitrogen use efficiency
O_2 = molecular oxygen
OAA = oxaloacetate
PEP = phosphoenolpyruvate
PEPC = phospho*enol*pyruvate carboxylase
PEPCK = phospho*enol*pyruvate carboxykinase
PGK = phosphoglycerate kinase
PGP = 2-phosphoglycolate phosphatase
PPDK = pyruvate, phosphate dikinase
ppm = parts per million
PRK = phosphoribulokinase
Pyr = pyruvate
R5P = ribose 5-phosphate
RPE = ribulose-5-phosphate 3-epimerase
RPI = ribose-5-phosphateisomerase
Ru5P = ribulose 5-phosphate
Rubisco = ribulose-bisphosphate carboxylase-oxygenase
RuBP = ribulose 1,5-bisphosphate
S7P = sedoheptulose 7-phosphate
SBP = sedoheptulose 1,7-bisphosphate
SBPase = sedoheptulose-1,7-bisphosphatase
Ser = serine
SGT = serine-glyoxylate aminotransferase
Triose-P = triose phosphates
TK = transketolase
TPI = triose phosphate isomerase
WUE = water use efficiency
Xu5P = xylulose 5-phosphate

List of figures

General Introduction

The work presented in my thesis was performed against the background of current attempts to improve photosynthetic yield, and in particular by engineering C_4 photosynthesis into rice. More specifically, I used analyses of metabolic intermediates to gain insights into the operation of the Calvin-Benson cycle (CBC) in C_3 species, including rice, and C_4 species, carrying out a detailed analysis of metabolism in nine species in the *Flaveria* genus to provide new insights into the evolution of C_4 photosynthesis. In the Introduction, I will first review background knowledge about C_3 photosynthesis focusing on the decrease in photosynthetic efficiency due to photorespiration, and strategies that have evolved to suppress photorespiration, especially C_4 photosynthesis. I will then discuss current ideas about the evolution of C_4 photosynthesis, and what steps would be needed to engineer C_4 photosynthesis into a C_3 species like rice. I will then define the specific aims of the work presented in this thesis.

Photosynthetic pathways in plants and algae

Photosynthesis and photorespiration

Oxygenic photosynthesis evolved in an ancestor of modern cyanobacteria around 2.5 billion years ago in a world extremely rich in carbon dioxide (CO_2) and devoid of molecular oxygen (O_2) (Rasmussen et al., 2008). From that moment on, all photosynthetic organisms relied on the carboxylation reaction of the enzyme ribulose-bisphosphate carboxylase-oxygenase (Rubisco; EC 4.1.1.39) to fix CO_2 and produce triose-phosphates (Triose-P) to fuel growth. Rubisco is one of the 11 enzymes forming the CBC which, in a series of steps, regenerates the Rubisco substrate, ribulose-1,5-bisphosphate (RuBP), from various sugar phosphates (Fig. 1)..

The prolonged activity of photosynthetic organisms led to the decrease of atmospheric CO_2 concentrations over the eons, while at the same time O_2, which is a byproduct of the light-dependent photosynthesis, increased (Hagemann et al., 2016). Rubisco enzymatic activity is not specific for CO_2, as it has a competing oxygenation reaction in which O_2 is used instead of CO_2.

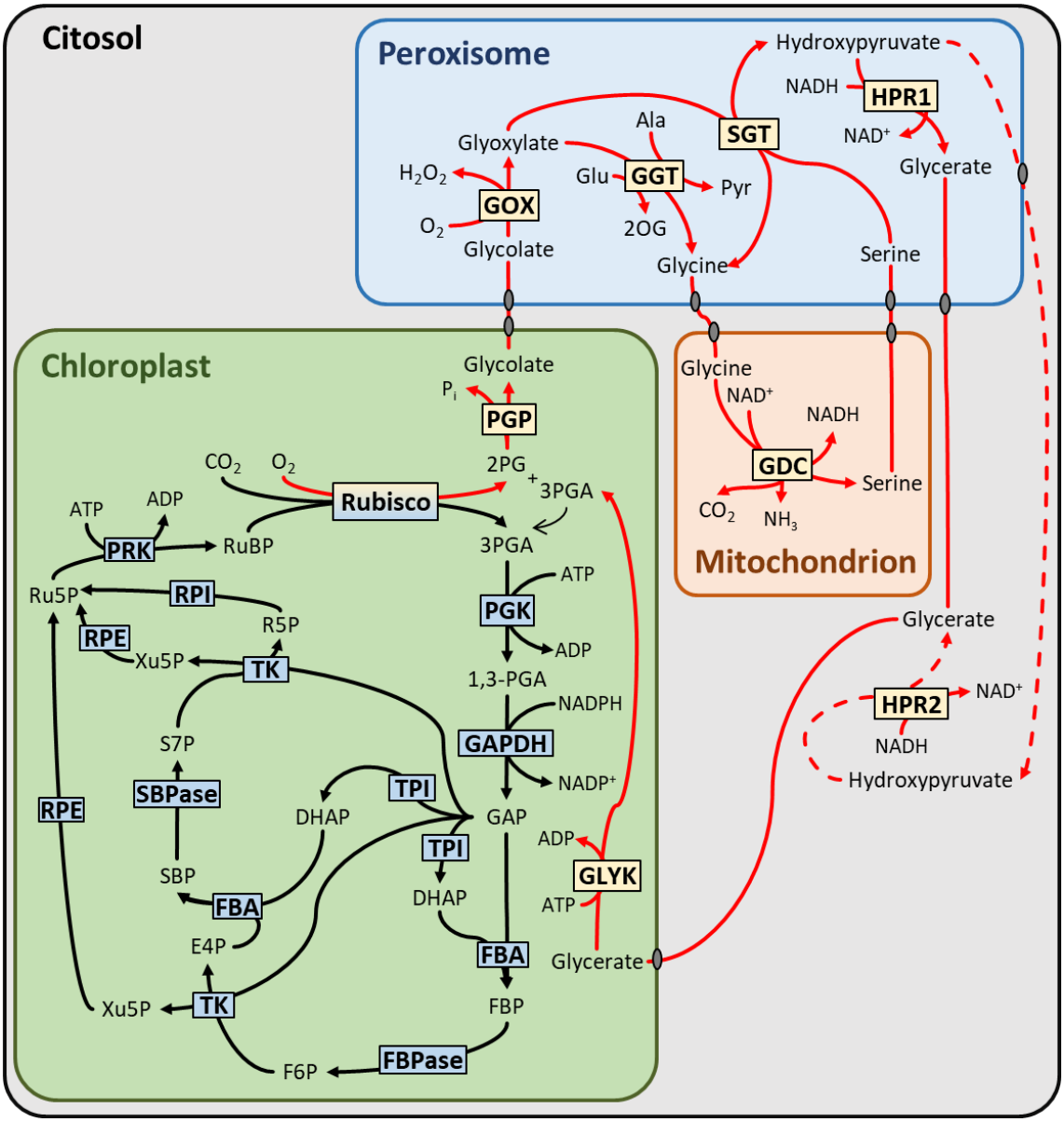


Fig. 1 – **Calvin-Benson cycle and photorespiration** – Simplified schematics of both Calvin Benson cycle (CBC, black arrows) and photorespiration (red arrows) with key enzymes, highlighted in azure (CBC) and yellow (photorespiration) boxes. Transmembrane transporters are indicated by small dark-grey oval shapes. CBC enzyme abbreviations are, in order: Rubisco, ribulose 1,5-bisphosphate carboxylase/oxygenase; PGK, phosphoglycerate kinase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; TPI, triose phosphate isomerase; FBA, fructose 1,6-bisphosphate aldolase; FBPase, fructose-1,6-bisphosphatase; TK, transketolase; SBPase, sedoheptulose-1,7-bisphosphatase; RPE, ribulose-5-phosphate 3-epimerase; RPI, ribose-5-phosphateisomerase; PRK, phosphoribulokinase. Photorespiration enzyme abbreviations are, in order: PGP, 2-phosphoglycolate phosphatase; GOX, glycolate oxidase; GGT, glutamate:glyoxylate aminotransferase; GDC, glycine decarboxylase; SGT, serine:glyoxylate aminotransferase; HPR1, hydroxypyruvate reductase (peroxisomal); HPR2, hydroxypyruvate reductase (cytosolic); GLYK, glycerate 3-kinase. Metabolite abbreviations, alphabetically ordered, are: 1,3-PGA, 1,3-bisphosphoglycerate; 2OG, 2-oxoglutarate; 2PG, 2-phosphoglycolate; 3PGA, 3-phosphoglycerate; Ala, alanine; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; GAP, glyceraldehyde 3-phosphate; Glu, glutamate; Pyr, pyruvate; R5P, ribose 5-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose-1,5-bisphosphate; S7P, sedoheptulose 7-phosphate; SBP, sedoheptulose 1,7-bisphosphate; Xu5P, xylulose 5-phosphate.

The rate of oxygenation increases relative to the rate of carboxylation as the CO_2 concentration decreases and the O_2 concentration increases. In current atmospheric conditions, about every fourth Rubisco reaction is an oxygenation (Fernie et al., 2013). Oxygenation leads to formation of 2-phosphoglycolate (2PG) (Bowes et al., 1971), which acts as an enzymatic inhibitor (Anderson, 1971; Kelly and Latzko, 1976; Norman and Colman, 1991). Furthermore, given the high rate of 2PG formation, this compound must be scavenged to prevent its accumulation and sequestration of large amounts of carbon and phosphorous, which would be lethal. The pressure to reduce the consequences of Rubisco oxygenation led photosynthetic organisms to evolve several different ways of coping with this unwanted reaction. The photorespiration pathway (Fig. 1), the first countermeasure evolved, appeared when cyanobacteria encountered rising atmospheric O_2 levels caused by their own photosynthetic activity, and was subsequently optimized in photosynthetic eukaryotes by compartmentation in different organelles (Bauwe et al., 2012). This metabolic pathway is now employed by all photosynthetic organisms to avoid the build-up of 2PG; the photorespiratory pathway converts 2PG back to 3-phosphoglycerate (3PGA) through a series of reactions that collectively salvage 75% of the carbon in the 2PG molecule whilst releasing 25% as CO_2 (Bauwe et al., 2010). Together with CO_2, NH_3 is also released in the glycine decarboxylase reaction of photorespiration.

Photorespiration, despite being necessary for plants to live at current atmospheric CO_2 and O_2 levels (Eisenhut et al., 2008; Zelitch et al., 2009; Eisenhut et al., 2017), is nevertheless a wasteful cycle that consumes energy. The following shortly explains most of the drawbacks. First, it decreases light use efficiency and decreases the rate of photosynthesis. Running photorespiration requires energy in the form of ATP and reducing power, and the pathway also does not recapture all of the carbon in 2PG. Second, photorespiration releases fixed nitrogen as NH_3, which in turn has to be re-assimilated mainly by the enzyme glutamine synthase (GS; EC 6.3.1.2) in the glutamine oxoglutarate aminotransferase (GOGAT) pathway, again at the cost of ATP (Wallsgrove et al., 1983). Third, photorespiration causes an indirect lowering of water use efficiency (WUE) for the following reason. Due to the competitive reaction of Rubisco with O_2, a higher CO_2 concentration has to be maintained within the leaf to achieve a given net rate of CO_2 fixation, which in turn requires that stomata open more leading to greater evaporative water loss. Fourth, nitrogen use efficiency (NUE) is decreased due to Rubisco's intrinsic properties. Selective pressure on Rubisco to favor higher CO_2 selectivity over O_2 appears

to be inextricably linked to a decrease in V_{max} (Morell et al., 1992; Tcherkez et al., 2006). Moreover, the k_{cat} of Rubisco in terrestrial plants is about 3-4 mol CO_2 mol^{-1} sites s^{-1} (Sage, 2002), and this slow rate of catalysis is compensated by allocating up to 50% of the total leaf protein to Rubisco (Feller et al., 2007). In addition, the amount of Rubisco, and thus allocated nitrogen, to achieve a given net rate of CO_2 fixation is increased even further because, at any one time, about 25% of Rubisco molecules are carrying out oxygenation in current atmospheric conditions (Fernie et al., 2013).

Different CO_2-concentration mechanisms

The second way of coping with the oxygenation reaction of Rubisco was to evolve CO_2-concentrating mechanisms (CCMs). These increase the local concentration of CO_2 around Rubisco, thus enhancing carboxylation over oxygenation rates and partly suppressing photorespiration. Broadly speaking, two types of CCM evolved, by temporally separating CO_2 uptake and release, or by spatially isolating Rubisco.

The only temporal CCM known so far is Crassulacean acid metabolism (CAM), which separates CO_2 fixation and release along the whole day. In CAM photosynthesis, carbon is assimilated into malate and other organic acids during the night period (when stomata are kept open), then these acids are mobilized and decarboxylated to release CO_2 within the chloroplast to be captured by Rubisco during the light period (when stomata are close) (detailed in Ranson and Thomas, 1960; Osmond, 1978).

However, most CCMs evolved by photosynthetic organisms require spatial separation, coupled with efficient ways of concentrating CO_2 (Badger et al., 1998). These CCMs are:

a) Active HCO_3^- pumping coupled with **carboxysomes**. Carboxysomes are intracellular proteic structures, found in cyanobacteria, which contain both carbonic anhydrase (CA; EC 4.2.1.1) and Rubisco. HCO_3^- is concentrated within the cell by active transport, then it moves into the carboxysome where it is converted to CO_2 by CA to generate a high concentration of CO_2 around Rubisco (Shively et al., 1973; Price et al., 1992; Yeates et al., 2008);

b) Active HCO_3^- pumping coupled with **pyrenoids**. Pyrenoids are starch-covered chloroplastic micro-compartments, found in many eukaryotic algae and all hornworts, which contain both CA and Rubisco. HCO_3^- is concentrated within the cell and the chloroplast by active transport, and moves into the pyrenoid where it is converted to CO_2

by CA, again generating a high concentration of CO_2 around Rubisco (Lacoste-Royal and Gibbs, 1987; Ramazanov et al., 1994).

c) **C_4 photosynthesis,** carried out by about 4% of all plant species, some of which are important crops like corn (*Zea mays*), sugarcane (*Saccharum officinarum*) and sorghum (*Sorghum bicolor).* C_4 photosynthesis represents a complex biochemical and anatomical restructuring of photosynthetic organs (or rarely conducted within single cells; Voznesenskaya et al., 2001), which evolved within certain angiosperm genera. In this CCM, Rubisco is confined in cells located deep inside the leaf, and it is separated from the initial reaction in which CO_2 (actually HCO_3^-) is fixed into 4-carbon organic acids, the latter occurring in cells that are in direct contact with the leaf air spaces. The organic acids move into the cells where Rubisco is located and are decarboxylated to generate a high CO_2 concentration around Rubisco. Thus, C_4 photosynthesis employs a biochemical shuttle to link the two cell types (Hatch and Slack, 1966; von Caemmerer and Furbank, 2003; a more detailed description of C_4 photosynthesis is provided in the next section).

d) **C_2 photosynthesis** is employed by C_3-C_4 intermediate species, which are much less common than other photosynthetic types and distributed in a limited number of angiosperm genera, the best known of which are *Flaveria* (yellowtops), *Moricandia* and *Panicum* (panic grasses). C_2 photosynthesis employs photorespiration to enact a CCM by spatially separating the glycine decarboxylation step of this pathway in cells deeper within the leaf. The released CO_2 can be more effectively recaptured because it has to move out past Rubisco, which at least in early forms of this variant is mainly located in cells that are in contact with the leaf air spaces. Like in C_4 photosynthesis, metabolites are thought to move between different cell types (Monson et al., 1984; Rawsthorne, 1992; see below for details). The general consensus within the scientific community put this CCM as an early step in the evolution of C_4 photosynthesis, hence the association to names like 'C_3-C_4' or 'intermediate' (Monson and Moore, 1989; Sage et al., 2012; Bräutigam and Gowik, 2016; see below for more details).

C_3 versus C_4 photosynthesis

C_3 plants fix CO_2 almost exclusively via the CBC in the abundant chloroplasts that are located inside leaf mesophyll cells. These cells are in direct contact with air spaces within the leaf, which in turn are connected with the outer air through the stomatal pores. This arrangement allows rapid gaseous exchange within the leaf. Nevertheless, several factors

like CO_2 diffusion rates through stomatal pores and intercellular spaces, biological diffusional barriers (cell walls and membranes), solubilization of CO_2 and Rubisco carboxylation activity itself force Rubisco to operate at lower CO_2 concentrations than those in the atmosphere. The rate of CO_2 entry will depend on the magnitude of the concentration gradient and on how wide open the stomata are (i.e. stomatal conductance), and how much hindrance the mesophyll cells exert on gaseous exchanges (i.e. mesophyll conductance). As already mentioned, the competitive reaction of O_2 with Rubisco means that the internal CO_2 concentration needs to be kept relatively high, which restricts the amplitude of the CO_2 concentration gradient. Hence, in C_3 plants, rapid CO_2 entry requires high stomatal conductance, which will result in rapid water loss (Farquhar and Sharkey, 1982).

In contrast to C_3, in C_4 plants Rubisco and the CBC are located in bundle sheath (BS) cells which surround the vascular tissue. C_4 plants initially fix carbon in the mesophyll cells using phospho*enol*pyruvate carboxylase (PEPC; EC 4.1.1.31). PEPC's substrate is HCO_3^-, for which it has a high affinity, and at the same time it has no side reaction with O_2. This allows PEPC to operate at lower CO_2 concentrations in the leaf than those required by C_3 plants. The product is the 4-carbon organic acid oxaloacetate (OAA), which is converted to malate (thereafter termed C_4 acids) by NADP-malate dehydrogenase (NADP-MDH; EC 1.1.1.82) or to aspartate by aspartate aminotransferase (AspAT; EC 2.6.1.1). These diffuse into the BS where they are decarboxylated to release CO_2, creating a CO_2-enriched environment in which Rubisco carboxylation is enhanced and oxygenation is decreased (Hatch and Slack, 1966; Hatch and Slack, 1970). First evidence for this pathway was collected by observing a rapid fixation of radiolabeled $^{14}CO_2$ into mainly malate and aspartate (instead of 3PGA) using illuminated sugarcane leaves (Kortschak et al., 1965). Just one year later the final proof of C_4 acid decarboxylation and CO_2 re-fixation came also from sugarcane leaves following the same labelling approach, with crucial evidence being delivered by pulse-chase experiments, showing that label moved from C_4 acids into 3PGA and other CBC intermediates (Hatch and Slack, 1966). During the following decades many plants were identified as C_4 by detecting their $\delta^{13}C$ signatures, a parameter derived from the $^{13}C/^{12}C$ isotopic ratio of plant material versus a carbonate standard. C_3 plants, due to Rubisco's selectivity over lighter carbon isotopes, have lower ^{13}C content than the standard and more negative $\delta^{13}C$ values. On the other hand PEPC's isotopic discrimination is lower and C_4 plants have higher ^{13}C content and less negative $\delta^{13}C$ values (Bender, 1971).

After the characterization of several C_4 plants among different genera, three variants of C_4 pathways were identified based on the major decarboxylating enzyme (Fig. 2): NADP-dependent malic enzyme (NADP-ME; EC 1.1.1.40), NAD-dependent malic enzyme (NAD-ME; EC 1.1.1.39) or phospho*enol*pyruvate carboxykinase (PEPCK; EC 4.1.1.49). However, this classification may be overly simplistic as there is evidence for parallel operation of multiple decarboxylation pathways, with the predominant pathway being partly dependent on environmental conditions (Furbank, 2011; Weissmann et al., 2016; Arrivault et al., 2017).

C_4 photosynthesis depends not only on this "biochemical CO_2 pump", but also on specialized anatomical features, the best known of which is Kranz anatomy (from German "Kranz"=wreath; Haberlandt, 1918). These anatomical changes include reduced interveinal space compared to C_3 plants, and enlarged BS cells containing more organelles arranged in a particular way: mitochondria and chloroplast are tightly packed together close to the vasculature, while a conspicuous vacuole resides toward the mesophyll. The latter is thought to reduce CO_2 back leakage and O_2 infiltration (Lundgren et al., 2014). Other morphological traits usually associated with Kranz anatomy are a high density of plasmodesmata connecting mesophyll and BS cells favouring metabolite diffusion, and a highly suberized BS cell wall having similar function of obstructing gas diffusion like the above mentioned vacuole (Weiner et al., 1988; Botha, 1992; Sage and Monson, 1999). Moreover, C_4 plants have a reduced mesophyll area and mesophyll organelle content compared to C_3 plants, including the loss of Rubisco and most of the CBC enzymes from the mesophyll chloroplasts (Sage and Monson, 1999; Weber and von Caemmerer, 2010; Wang et al., 2011).

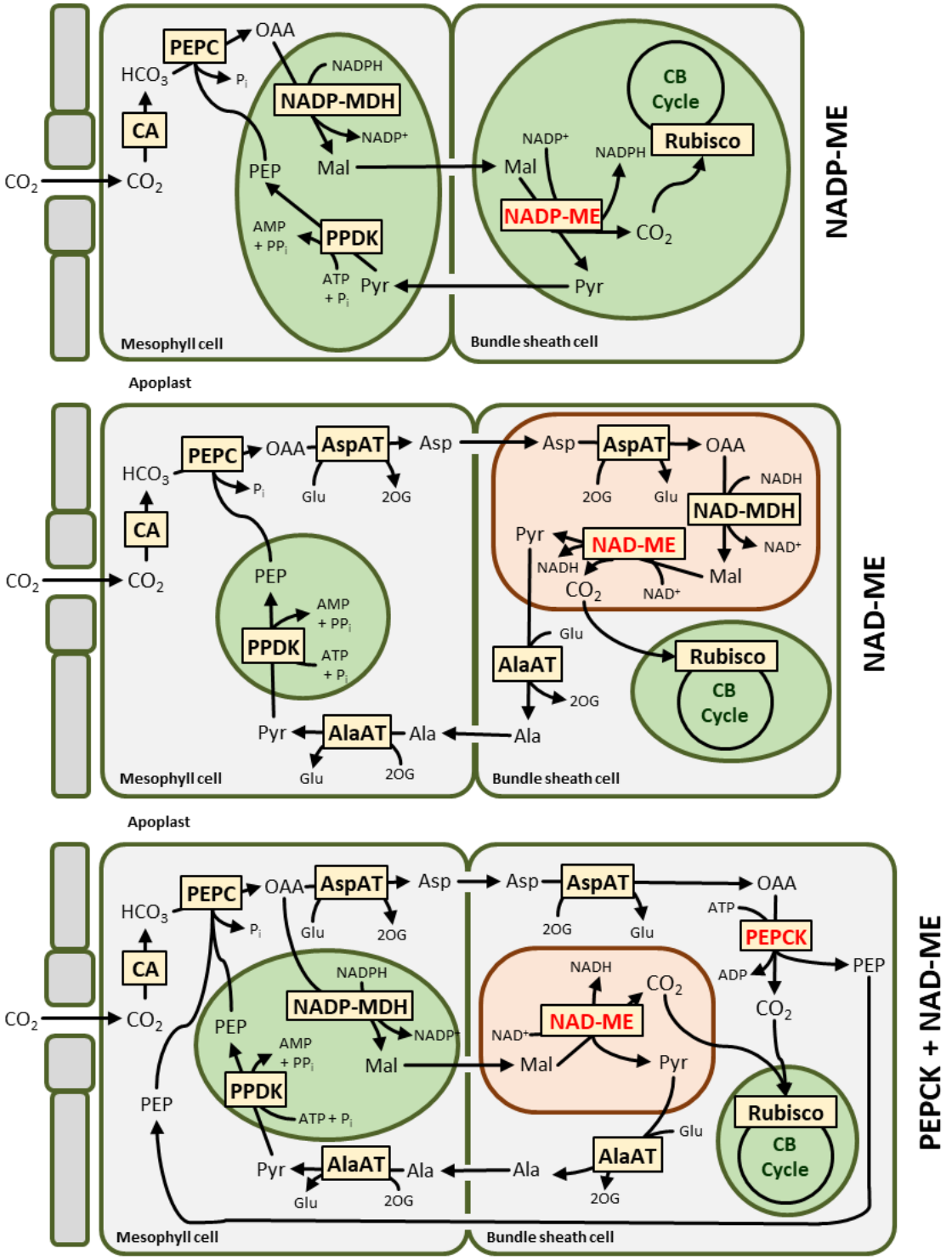


Fig. 2 – **Different C_4 metabolic pathways** – Simplified schematics of the three main C_4 pathways, showing the most important enzymes in yellow boxes. Decarboxylation enzymes are highlighted in red. Chloroplasts are represented by green oval shapes, while mitochondria by orange rectangles. Enzyme abbreviations are, in clockwise order within the pathway: CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; AspAT, aspartate aminotransferase; NADP-MDH, NADP-dependent malate dehydrogenase; NAD-MDH, NAD-dependent malate dehydrogenase; NADP-ME, NADP-dependent malic enzyme; NAD-ME, NAD-dependent malic enzyme; PEPCK, phosphoenolpyruvate carboxykinase; RuBisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; AlaAT, alanine aminotransferase; PPDK, pyruvate phosphate dikinase; Metabolite abbreviations, alphabetically ordered, are: 2OG, 2-oxoglutarate; Ala, alanine; Asp, aspartate; Glu, glutamate; Mal, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; Pyr, pyruvate.

Pros and cons of C_4 photosynthesis

Running C_4 photosynthesis gives several advantages to those species employing it. These advantages stem from the ability of C_4 photosynthesis to operate efficiently even at lower internal CO_2 concentrations (low CO_2 compensation point), and from the reduced rate of the Rubisco oxygenation reaction and photorespiration:

a) better WUE due to the ability of C_4 plants to keep higher carboxylation rates at lower internal CO_2 concentration (c_i), therefore limiting the need to open stomata during day (Ghannoum, 2009; Ghannoum et al., 2011);

b) better NUE due to much lower loss of fixed nitrogen from photorespiration and decreased Rubisco investment in photosynthetic organs (Brown, 1978; Oaks, 1994; Ghannoum et al., 2011), as well as relaxation of selection for a high selectivity for CO_2 over O_2 that has allowed an increase in Rubisco k_{cat} (Sage, 2002);

c) increased temperature optimums for photosynthesis, since photorespiration rates rise with temperature (Sage, 2002; Sage and Kubien, 2007);

d) improved quantum yield, because the number of photons that need to be absorbed to fix each CO_2 molecule is decreased due to the low rate of photorespiration (Ehleringer and Björkman, 1977; Ehleringer and Pearcy, 1983; Zhu et al., 2008).

On the other hand, running a C_4 cycle requires more energy than running only the CBC: the equivalent of two extra ATP molecules are needed during the regeneration of phospho*enol*pyruvate (PEP) by pyruvate, phosphate dikinase (PPDK; EC 2.7.9.1) and one extra ATP is required during the decarboxylation step carried out by PEPCK.

Considering the above-mentioned pros and cons, there is a tradeoff between the advantages of running C_4 photosynthesis versus the extra energy it requires. Under environmental conditions in which C_3 plants would experience high rates of photorespiration, like in areas with seasonal high temperatures and drought, C_4 photosynthesis represents an advantage. In less challenging environments the benefits of running C_4 are marginal or absent (Long, 1983; Zhu et al., 2008). This explains why C_4 plants currently dominate in environments like the tropical and subtropical grasslands, while at the same time being less common in the temperate zones and almost absent within the sub-polar regions (Pearcy and Ehleringer, 1984; Collatz et al., 1998; Sage, 2017).

In summary, C_4 photosynthesis allows higher rates of photosynthesis, WUE and NUE. It is not a coincidence that C_4 species, while representing only about 4% of all land plant

species, are currently responsible for roughly 25% of the total land primary productivity (Still et al., 2003; Sage et al., 2012). All the above-mentioned advantages of C_4 photosynthesis are traits desirable for improving crops, especially in the light of global climate change.

The evolution of C_4 photosynthesis

Convergent evolution of a complex trait

C_4 photosynthesis is a composite trait, whose evolution required complex metabolic and anatomical modifications that are thought to have evolved in several steps starting from ancestral C_3 plants in response to selective pressure from environmental changes. C_4 photosynthesis evolved in the last 20-30 million years. About 30 million years ago there was a transition in the Earth's climate from hot and wet conditions with a relatively high atmospheric CO_2 concentration (over 1000 ppm) to cooler and drier conditions and low CO_2 concentration (less than 300 ppm) (Zachos et al., 2008). The fall in atmospheric CO_2 concentrations is thought to have been a key factor driving the evolution of C_4 photosynthesis, by favouring the survival and reproduction of plants that had acquired adaptations that made their photosynthesis more efficient (Sage, 2003). Phylogenetic studies indicated that C_4 photosynthesis evolved independently in over 60 plant lineages (Sage, 2017); this is a classic example of convergent evolution toward a complex trait. We call C_4 photosynthesis a complex trait because it involves changes in leaf anatomy, cell specification, chloroplast development and changes in the expression level and/or spatial patterning of probably hundreds of genes (Gowik et al., 2011; Gowik and Westhoff, 2011; Schlüter and Weber, 2020). It is generally accepted that it must have evolved in a sequence of steps, in which each step brought an advantage and at the same time predisposed plants for the next step towards the full C_4 syndrome (Schlüter and Weber, 2016).

A recent study compared this evolutionary convergence as a "stroll over Mount Fuji" by creating a mathematical evolutionary fitness model based on several biochemical parameters. The model predicted, if C_4 enabling conditions are in place, a single peak of increasing fitness toward a complete C_4 status, which mimics the "greedy pathway" of little biochemical tweaks that most likely happened during C_3 to C_4 evolution. This model was successfully tested on several C_3, C_4 and intermediate species to confirm this evolutionary trajectory (Heckmann et al., 2013; Heckmann, 2016).

Despite some controversial claims made over the years (P'Yankov et al., 1997; Ibrahim et al., 2009), solid proofs of C_4 to C_3 reversion are still to be found, and some previous reversion claims have been disproved (Ingram et al., 2011).

Evolutionary steps toward C_4 within the model genus *Flaveria* (Asteraceae)

An important model system to investigate C_3 to C_4 evolution is provided by the genus *Flaveria* (Asteraceae), which emerged relatively recently (2-3 million years ago – Sage, 2017) and have its center of origin in the Mesoamerican region (Powell, 1978). Due to its late emergence, it contains many intermediate forms of photosynthesis, starting from C_3 to a complete C_4 photosynthesis (NADP-ME subtype).

Other plant genera studied for C_4 evolution and containing several C_3-C_4 intermediates are *Blepharis* (Fisher et al., 2015), *Cleome* (Marshall et al., 2007; Voznesenskaya et al., 2007), *Heliotropium* (Muhaidat et al., 2011), *Neurachne* (Khoshravesh et al., 2019) and *Panicum* (Sternberg et al., 1986). However, the *Flaveria* genus has been studied more deeply and for a longer time within the scientific community working on C_4 photosynthesis evolution (Brown et al., 2005).

Studies on *Flaveria* species have led to the following picture, in which C_4 evolution is seen as a stepwise process with plants progressively acquiring C_4-like features in the following sequence (Fig. 3):

1) **Preconditioning**: Some C_3 species were predisposed to start down the path to C_4 evolution by acquiring leaf anatomical adaptations like increased vein density/reduced vein spacing. This is thought to favor better leaf hydraulics during drought (Sage et al., 2013a). Another preconditioning factor is thought to be the early stochastic duplications of future pivotal genes (e.g. glycine decarboxylase subunit P [GLDP] – Schulze et al., 2013).

2) **Proto-Kranz**: selection for drought-resistance and suppression of photorespiration favored the emergence of a primordial Kranz anatomy characterized by enlarged BS cells that were enriched in mitochondria (Sage et al., 2013a). It is proposed that in this way part of the photorespiratory glycine decarboxylation would occur deep within the leaf, increasing the probability that the CO_2 released by this reaction would be re-fixed before escaping into the atmosphere (Sage et al., 2014). However, this stage still lacks a CCM and it is still considered to be C_3.

3) **C_3-C_4 Type I**: in this stage, the appearance of a complete Kranz anatomy is associated with C_2 photosynthesis, in which photorespiratory glycine decarboxylation occurs only in the BS cell mitochondria. These are arranged in a centripetal manner within the BS cells, often with chloroplasts clustered nearby (Sage et al., 2013a). This arrangement not only maximizes the path length for released photorespiratory CO_2 to escape from the leaf (and so increase the chance of re-fixation) but also concentrates the released CO_2 around Rubisco, enhancing the carboxylation reaction (Keerberg et al., 2014). Presumably, C_2 photosynthesis emerged from selection on Proto-Kranz plants to increase the efficiency of CO_2 recapture by decreasing glycine decarboxylation in the mesophyll. In *Flaveria*, the BS cell specific expression of glycine decarboxylase arose by differential changes in expression of *GLDP* genes (encoding catalytic subunits of the glycine decarboxylase complex [GLDC, EC 1.4.4.2]), with the mesophyll expressed genes becoming non-functional pseudogenes while the gene encoding the BS isoform was retained (Schulze et al., 2013).

The C_2 cycle increased carboxylation over oxygenation, putting a positive evolutionary pressure over BS and mesophyll cell for further anatomical and biochemical differentiation to occur. Some C_4 features may have arisen in C_2 plants as a way to balance carbon and nitrogen between BS and mesophyll compartments (Mallmann et al., 2014). Crucially, rapid movement of glycine into the BS was required downstream of the oxygenation reactions catalyzed by that part of the Rubisco that was still located in the mesophyll. Intercellular movement of other metabolites was also needed to balance the intercellular movement of amino groups and carbon (for details see the Introduction to the third manuscript in this thesis). Briefly, for every two molecules of 2PG formed by Rubisco in the mesophyll, the stoichiometry of the photorespiration pathway requires that two molecules of glycine move from the mesophyll into the BS, where they are converted to one serine and one NH_3. The serine must then move back to the mesophyll as it is or, alternatively, as another metabolite like glycerate or 3PGA, in order to complete the cycle, while the NH_3 must be re-fixed via the GS-GOGAT cycle. If nitrogen re-fixation occurs in the BS cells then an organic acid needs to move from the mesophyll to function as a nitrogen acceptor and the resulting amino acid must move back to the mesophyll. The simplest form of exchange between mesophyll and BS cells would be movement of 2OG and glutamate. However, more complex shuttles could operate, involving other organic acids, amino acids and transfer of amino groups to and from 2OG and glutamate via aminotransferase reactions in BS cells. Alternatively, the NH_3 could move back to the

mesophyll and be re-fixed by the GS-GOGAT cycle there. Irrespective of details, this intercellular shuttling of organic and amino acids is thought to have set the scene for the next step in the evolution of C_4 photosynthesis (Mallmann et al., 2014; Bräutigam and Gowik, 2016).

4) **C_3-C_4 Type II**: this step represents a clear shift from C_2 to C_4 photosynthesis. In this stage expression of Rubisco started to shift from the mesophyll to the BS, and PEPC started to operate at increasingly higher rates in the mesophyll cells, allowing the CCM to shift from C_2 to a partial C_4 cycle. How Type II species gradually suppressed photorespiration (and the related C_2 cycle) and moved toward a complete C_4 cycle is not yet clear. As mentioned, it was proposed that this may have been aided by the occurrence of intercellular shuttles of organic acids and amino acids that had been set up in the C_3-C_4 Type I plants (Mallmann et al., 2014). It is reasonable to assume that selection pressure

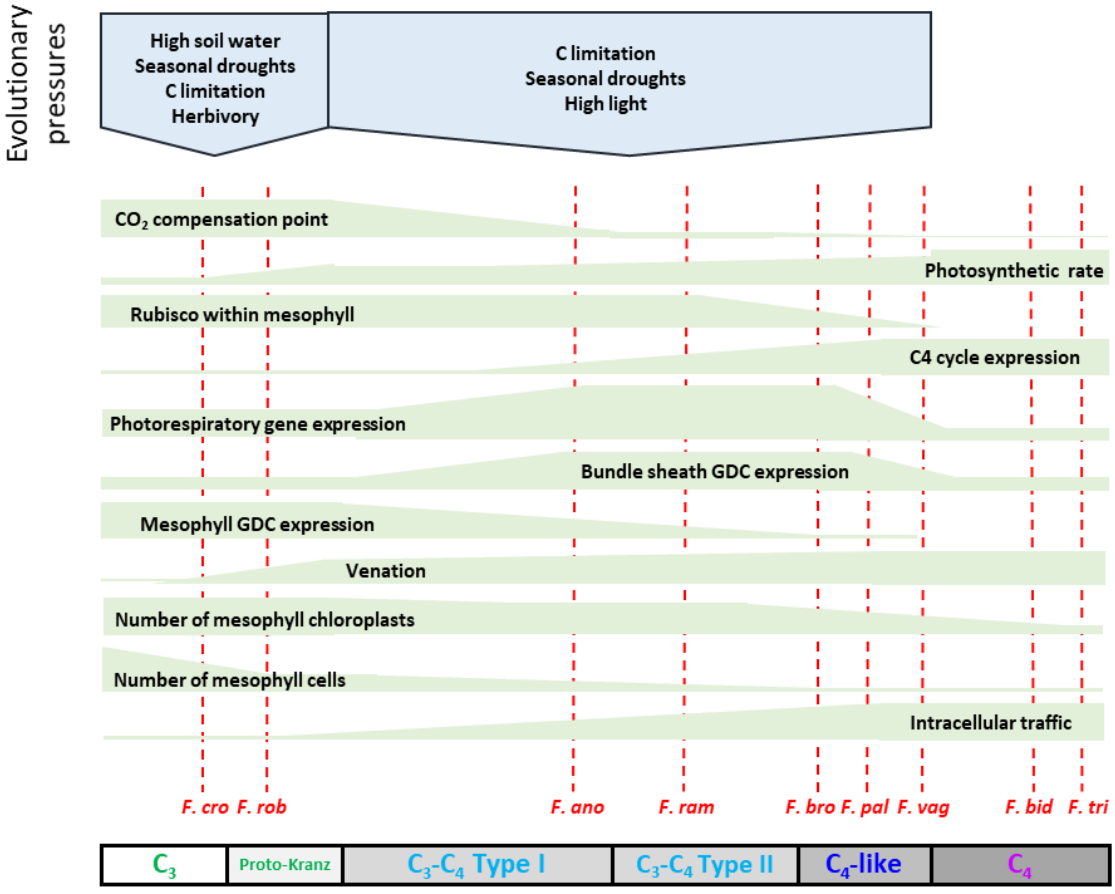


Fig. 3 – **From C_3 to C_4 photosynthesis in the genus *Flaveria*** – A representation of major evolutionary pressures and main trait transitions during the evolution from C3 to C4 photosynthesis, encompassing all the recognized intermediate states. Dashed red lines represents the position regarding each trait for each of the nine *Flaveria* species included in the third paper of this thesis. Species abbreviations are, alphabetically: *F. ano*, *F. anomala*; *F. bid*, *F. bidentis*; *F. bro*, *F. brownii*; *F. cro*, *F. cronquistii*; *F. pal*, *F. palmeri*; *F. ram*, *F. ramosissima*; *F. rob*, *F. robusta*; *F. tri*, *F. trinervia*; *F. vag*, *F. vaginata*. This figure represents a modified version of the one appearing in Bräutigam & Gowik (2016).

over time may have led to Rubisco expression shifting from mesophyll to BS cells, where the higher CO_2 concentration would decrease the wasteful reaction with O_2. Once this had started, the scene might be set to increase PEPC activity in the mesophyll without this leading to a drawdown of CO_2 in the vicinity of Rubisco and a resulting rise in the wasteful reaction with O_2.

5) **C_4–like**: in this state a complete C_4 cycle is operating and represents the main carbon assimilation strategy. There is, however, still a residual activity of primary carboxylation reaction through Rubisco which accounts for 5-15% of total carbon fixation; this is due to incomplete segregation of Rubisco into BS cells or an incomplete compartmentation of C_4 related enzymes (Moore et al., 1989).

6) **C_4**: the final step to achieve a complete C_4 status is fine-tuning of the cell-specific compartmentation of enzymes, and changes in enzyme kinetics (e.g. K_m and V_{max}) and regulatory properties. For example, Rubisco enzymes from C_4 plants generally have a lower affinity for CO_2 than the C_3 enzyme, but a higher V_{max} (Sage, 2002; Kubien et al., 2008; Kapralov et al., 2011). C_4 PEPC enzymes have lower sensitivity to malate inhibition (Gowik et al., 2011). It is probable that activities of several transporters were increased on the road to complete C_4-ness (Weber and von Caemmerer, 2010; Schuler et al., 2016). Furthermore, changes in the regulation of enzymes of end-product synthesis like the cytosolic fructose-1,6-bisphosphatase (FBPase; EC 3.1.3.11) were needed to allow the maintenance of large pools of metabolites that are needed for C_4 photosynthesis (Stitt and Heldt, 1985b; Stitt and Heldt, 1985a). In the case of the NADP-ME subtypes, the chloroplast dimorphism advanced even further with partial or near complete loss of photosystem II expression in the BS chloroplasts (Laetsch and Price, 1969; Woo et al., 1970; Höfer et al., 1992), most likely to favor the decarboxylation reaction of NADP-ME by keeping NADPH/$NADP^+$ ratio low within the chloroplast (Blätke and Bräutigam, 2019).

The phylogeny of the genus *Flaveria* has been extensively covered in many publications based on different morphological and molecular evidence (Powell, 1978; Kopriva et al., 1996; McKown et al., 2005; Lyu et al., 2015). According to these studies, all C_3 Flaveria species lie at the base of the phylogenetic tree. The lineage in which the first steps towards C_4 photosynthesis occurred subsequently split into two clades. In Clade A, all extant species, including the only C_3-C_4 Type II intermediate in the clade (*Flaveria ramosissima*), have a partial or fully functional C_4 cycle, and true C_4 *Flaveria* species are found only within this clade. Therefore, the extant species in Clade A presumably share

a common ancestor that had started to move along the path towards a functional C_4 cycle earlier. On the other hand, Clade B contains mostly C_3-C_4 type I and II intermediate species, with only one extant species (*Flaveria brownii*) having evolved further to achieve a C_4-like status; therefore in Clade B there are no true C_4 species (McKown et al., 2005).

Improving rice yields in the new millennium

Rice in the post-Green Revolution era

Rice is the third most important food crop for mankind in terms of worldwide annual biomass production, after sugarcane and maize, and makes up the majority of daily calories intake for over half the world population (FAOSTAT, 2018). Rice production worldwide had a significant boost in the second half of the 20th century during the so called "Green Revolution". This was achieved in part by allocation of more land to rice cultivation and increased use of fertilizers and pesticides, but a major contributing factor was breeding to create new rice cultivars with improved yield potential. Initial advances were linked to selection for a semi-dwarfing habitus that increased the number of spikelets and productive tillers per plant, increasing the overall harvest index, followed by improvements in pathogen and insect resistance. However, it is suspected that the improvements based on selection for these traits may have reached a plateau (Cassman, 1999), especially in irrigated rice cultivation in Asia (Cassman, 1994).

More than half of undernourished people in the world live in Asia, based on recent reports from the Food and Agriculture Organization of the United Nations (2017; www.fao.org), and it was estimated that, by 2050, the Asian population will have increased by a further 1.5 billion people. Combined with the projected increase in population for whom rice is a staple crop, and negative effects on food production (Ray et al., 2012) due to climate change, loss of agricultural land through urbanization and soil degradation, and increased demand for biofuels, Mitchell and Sheehy (2006) estimated that an increase of around 50% in rice production would be needed to meet human demand by 2050.

Improving photosynthesis with a focus on rice

Within our current knowledge, the most promising and possibly the only way to achieve the necessary boost in yield to sustain population growth and global changes is to improve the efficiency and rate of rice photosynthesis (Long et al., 2006; Mitchell and Sheehy,

2006; Sage et al., 2017). Rubisco was the first logical target for improvement since its carboxylation reaction in C_3 plants is quite slow, partially due to evolutionary selection to dampen the oxygenation reaction (Morell et al., 1992; Tcherkez et al., 2006). Research on Rubisco improvement has focused on speeding up carboxylation activity and increasing carboxylation over oxygenation rates (Evans, 2013; Parry et al., 2013). Rubisco in rice is no exception (Makino et al., 1985), and different approaches were attempted in the last two decades, albeit with not very promising results so far (Suzuki et al., 2009; Suzuki et al., 2012; Suzuki et al., 2019).

Another approach to improve photosynthetic yield is to minimize the metabolic burden of photorespiration; this can be achieved by bypassing existing inefficient steps in the pathway, particularly the glycine decarboxylation step which release fixed CO_2 and NH_3, or by introducing novel and energetically cheaper pathways to recycle 2PG (Trudeau et al., 2018). In fact both approaches led to promising biomass improvements in both laboratory and field experiments using two C_3 plants: *Arabidopsis thaliana* and tobacco (Simkin et al., 2017; López-Calcagno et al., 2019; South et al., 2019). However, these approaches remain untested in rice (or other cereals) so far, and do not promise large gains in yield.

On the other hand, introducing a CCM, and in particular C_4 photosynthesis, is predicted by modelling to bring an ample increase in photosynthetic capacity, and therefore is considered to have the greatest potential for crop improvement (Hibberd et al., 2008; Covshoff and Hibberd, 2012). That said, the introduction of C_4 photosynthesis is a major challenge because it simultaneously requires: a) major changes in leaf anatomy, with the veins becoming more closely spaced; b) major changes in cell specification, with BS cells becoming larger and acquiring large numbers of functional chloroplasts; c) major changes in the expression levels and pattern of large numbers of enzymes and metabolite transporters; d) fine tuning and coordination of these far-reaching changes.

The international C_4 Rice Project

Before the Project

The higher productivity of C_4 species compared to their C_3 relatives, especially in tropical and sub-tropical environments, prompted the idea of harnessing this pathway to improve crops in the late 1960s, shortly after the pathway was discovered. It was estimated that

converting a C_3 crop to C_4 could increase the productivity by up to 50% (Kiniry et al., 1989; Hibberd et al., 2008), making this idea even more interesting.

The first approach involved hybridization of C_3 plants with close C_4 relatives to generate intermediate species. However, this approach was seriously hampered because it was limited to a few specific genera and was not applicable to important crops, and because of general infertility of the hybrids due to chromosomal anomalies (Brown and Bouton, 1993).

The second approach made use of transgenic techniques, as they emerged during the 1980s, to introduce C_4 specific genes into several C_3 crops in order to confer C_4 traits (Matsuoka et al., 2001; Häusler et al., 2002; Miyao et al., 2011). *Agrobacterium tumefaciens*-mediated transformation opened up the transgenic approach in many species; however most monocot cereals, including rice, proved to be recalcitrant to transformation until the development of biolistic approaches (Li et al., 1993). Only in the early 1990s was *Agrobacterium*-mediated transformation of rice achieved, by optimizing the binary vectors (promoters, selectable marker genes), *Agrobacterium* strains, and choice of source material (genotype and tissue type) used for transformation (Hiei et al., 1994). These advances in rice transformation facilitated attempts, primarily in Japan, to introduce a single celled C_4 pathway into rice, comparable to that in *Hydrilla verticillata*, by over-expression of major C_4 pathway enzymes. Despite limited progress towards the ultimate goal, some valuable lessons were learned for achieving high-level transgene expression in rice by using genomic rather than cDNA sequences, along with a clearer recognition of the need to modify leaf anatomy as well as biochemistry, and to improve the targeting of C_4 enzymes to appropriate compartments (Miyao et al., 2011).

Despite these difficulties, several developments sustained or even increased the interest in introducing C_4 photosynthesis into rice. These included: a) phylogenetic analyses showing that C_4 photosynthesis had arisen independently in multiple plant lineages in a relatively short period of geological time (Sage, 2017); b) the recognition that many elements of the C_4 pathway already exist in photosynthetically active stem tissues in some C_3 plants (Hibberd and Quick, 2002); c) breakthroughs in our understanding of the molecular mechanisms involved in the development of Kranz anatomy (Nelson and Langdale, 1992). The first two advances indicated that the transition from C_3 to C_4 might not be as difficult to achieve as previously thought, and the third opened up possibilities to engineer C_4 leaf anatomical traits in parallel with C_4 biochemistry. All this accumulated knowledge created an optimistic momentum that culminated in 2006 with the foundation

of the C_4 Rice Project at the International Rice Research Institute (IRRI). The structured and funded C_4 Rice Project, as we see it today (www.c4rice.com), is a multinational effort started in 2008 and backed by financial support from the Bill & Melinda Gates Foundation.

General goals and structure of the international C_4 Rice Project

The ultimate goal of the international C_4 Rice Project is to obtain rice varieties employing a complete or impartial C_4 photosynthesis to improve yield and radiation, water- and nitrogen-use efficiencies (Zhu et al., 2008; Ghannoum et al., 2011).

Consistent with the humanitarian principles of the Bill & Melinda Gates Foundation, improved rice varieties generated by the project are to be made available at an affordable price to smallholder farmers, in order to alleviate hunger and poverty in developing countries.

Within the C_4 Rice Project sequential phases were planned, each of them of approximately four years (www.c4rice.com):

- **Phase I – 2008-2011**: The first phase of the project revolved around deepening the knowledge regarding Kranz specification in C_4 species and developing tools for better transformation of rice.
- **Phase II – 2012-2015**: The second phase had the major objective of obtaining a C_4 rice prototype from which to work on in the following phase; however the complexity of this task was clearly underestimated and the phase ended with the creation of a large number of transgenic rice lines that still needed to be tested for C_4-ness.
- **Phase III – 2016-2019**: In the third phase the focus was shifted from creating a prototype to analyzing the many lines generated so far, to search for evidence of incorporation of assimilated carbon into C_4 acids and subsequent decarboxylation and refixation. An additional component of this phase was metabolic modelling, using *in silico* approaches to identify and overcome potential bottlenecks in the introduced C_4 pathway. New cutting edge technologies were also introduced to aid introduction of multiple genes.
- **Phase IV – 2020 – onward**: The advances achieved in Phase III were sufficient for securing funding for a next phase; all the efforts are now in developing a

functional C_4 prototype by building on the most promising lines obtained in Phase III.

Engineering C_4 photosynthesis in rice

Rice, as a member of the Bambusoidae subfamily of grasses, has no close C_4 relatives. Thus, introgression of C_4 photosynthesis by conventional crossing is not feasible, and we lack an ideal blueprint for a transgenic approach (Mitchell and Sheehy, 2006).

At the beginning of the C_4 Rice Project in 2008, the tools for transforming rice were more advanced than those employed during the first attempts of C_4 rice creation, but were still limited to one insertion per transformation event, therefore relying on multiple crossing cycles and genotyping to stack all the desired homozygous loci within one line. When the C_4 Rice Project started, IRRI made a very extensive effort in planning, transforming and crossing high yield lowland rice (*Oryza sativa ssp. indica*) cultivar IR64 to generate the first multigene lines; these lines were analyzed during Phase II and, more extensively, during Phase III. However, this initial approach had limitations, which contributed to delays in the development of transgenic lines and highlighted the need for new tools within the consortium. These included: a) the lengthy multiple crossings and rounds of selection required to stack multiple transgenes together; b) the requirement for extensive genotyping and confirmation of protein expression in each generation, to overcome the common problem of transgene silencing; and c) the use of cv. IR64, which is late flowering (i.e. has a long generation time) and grows poorly under laboratory conditions, most likely due to sub-optimal light levels and spectral quality available in most controlled environment facilities.

Recent molecular biology tools in the form of Golden Gate cloning allowed multiple transgenes to be inserted in the same vector and within plants in a single transformation event (Engler et al., 2014; Andreou and Nakayama, 2018). The employment of this toolbox within the C_4 Rice Project considerably speeded up the process of generating new lines. Moreover, the gradual transition from IR64 cultivar to *O. sativa ssp. japonica* cultivar Kitaake, which is smaller, fast growing and early flowering, making it more suitable for manipulation and indoor growth (Li et al., 2017), further boosted the creation of transgenic lines. It should be noted that cv. Kitaake is a useful model for development of a C_4 prototype, but the C_4 pathway and anatomical traits will eventually need to be introduced into *O. sativa ssp. indica* cultivars to allow the creation of high yield C_4 rice varieties for agricultural use.

Introducing the C_4 cycle

Theoretically, a minimal NADP-ME C_4 cycle could be engineered into rice by introducing just five C_4 enzymes (tentatively from maize or *Setaria sp.*) with appropriate cell specificity and subcellular compartmentation. Cytosolic CA and PEPC enzymes are required in rice mesophyll to convert CO_2 into HCO_3^- and then fix it into OAA. Still in the mesophyll, the chloroplastic enzyme NADP-MDH is needed to convert OAA into malate, using NADPH produced by the photosynthetic electron transport chain in the mesophyll chloroplasts. By diffusion, malate should then reach the BS cells where it is going to be decarboxylated by chloroplastic NADP-ME enzyme to form pyruvate (Pyr). Pyr should then diffuse back to the mesophyll cell and be transformed back into PEP by PPDK inside the mesophyll cell chloroplast. This minimal cycle would contribute to establishing higher CO_2 partial pressure around Rubisco in BS cells. Modelling indicates that it could even be beneficial in plants that still retained a C_3 leaf anatomy (higher BS/mesophyll ratio) (Ermakova et al., 2019).

This minimalist approach however, could be restricted by further factors. The first is in the form of metabolite transporters; many C_4 reactions are carried out within chloroplasts, presenting an obvious diffusional barrier in the form of their multiple membranes. For polar metabolites like OAA, malate and Pyr, it would be therefore necessary to have specific transporters to allow rapid movement in and out of chloroplasts for a properly functioning C_4 cycle (Schuler et al., 2016). C_3 plants like rice have relatively few plasmodesmata between mesophyll and BS cells, compared to the high density of plasmodesmata between these cell types in C_4 plants. Thus, a second factor is the potential limitation of diffusional movement of malate and Pyr between cells (Danila et al., 2016). A third factor is that achieving the full potential of introducing C_4 photosynthesis will depend on modifying rice leaf anatomy (Sage and Sage, 2009).

Introducing Kranz anatomy

The introduction of Kranz anatomical traits into rice posed a new challenge to the community, but was seen as a necessary step because previous attempts to introduce C_4 photosynthesis as a single-cell process had been unsuccessful (see above). As already mentioned, there is a general consensus for most studied C_4 genera that the first step in evolution of Kranz anatomy was a reduced vein spacing and increased photosynthetic activity of the BS layer (Sage et al., 2012). However, the molecular mechanisms behind this shift are for the most part unknown. Denser venation can be attributed to increased

and prolonged procambial initiation within the leaf, while activated BS activation can be associated with plastid proliferation (Fouracre et al., 2014). Given its importance as a crop species and longstanding use as a model for plant genetics, maize has been the principal species used for investigating the development of Kranz anatomy and as potential source of genes for introducing Kranz formation into rice. In fact, research in maize has uncovered the most promising candidates for Kranz induction; the transcriptional factors SCARECROW and SHORT-ROOT, which are involved in minor leaf vein patterning (Li et al., 2010; Slewinski et al., 2012; Tausta et al., 2014), and the GOLDEN-LIKE family of transcriptional factors, which are involved in chloroplast proliferation in both mesophyll and BS cells (Rossini et al., 2001; Fitter et al., 2002; Waters et al., 2008; Wang et al., 2013; Wang et al., 2017).

The importance of flux analysis and modelling

Whilst molecular analyses can verify if the introduced genes are expressed or not, they cannot establish that the encoded proteins are active and, in particular, that flux has been increased at individual enzymes or enzymatic pathways. The operation of a complete C_4 cycle can be detected by gas exchange (e.g. decreased CO_2 compensation point) and $\delta^{13}C$ discrimination, even if the C_4 cycle is operating at only a low rate. However, these approaches would not detect incomplete operation of a C_4 cycle, i.e. increased activities and fluxes at one or several, but not all, of the enzymes and transporters in the cycle. They might not necessarily detect the operation of a complete C_4 cycle, if it were not well coordinated with changes in the location of Rubisco and the CBC (i.e. if the C_4 cycle did not lead to increased CO_2 levels around a substantial part of leaf Rubisco, and without lowering CO_2 levels in the vicinity of the remainder of the Rubisco). Incomplete or uncoordinated operation are, however, quite likely outcomes in the first steps of engineering this complex pathway into rice. These considerations highlighted the need to employ methods that could detect increased *in vivo* flux at the level of individual enzymes or subsets of enzymes from the C_4 pathway.

When our group became involved in the C_4 Rice project, in the latter stage of Phase II, the main task was to use high performance liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) (like those in Arrivault et al., 2009; Arrivault et al., 2015) to measure metabolite levels in some rice transgenic lines, and determine their ^{13}C labeling patterns after supplying $^{13}CO_2$, and thereby assess whether any photosynthetic CO_2 fixation was occurring via a C_4-like pathway. We became full partners in the C_4 Rice

project in Phase III, contributing to the biochemical analysis of more advanced transgenic rice lines. In parallel, these methods were applied to analyze labelling patterns and fluxes in the C_4 species maize (Arrivault et al., 2017). Labelling patterns in rice plants with a partial C_4 pathway, or even a complete but minor C_4 pathway, will be more complicated than those observed in C_4 species, due to a substantial proportion of CO_2 still being fixed via the original C_3 pathway. For this reason these labelling patterns would require careful interpretations (Arrivault et al., 2017).

Another goal of our experiments was to provide detailed data about metabolite levels and enzyme activities data in wild-type rice to be used for parameterization of a rice leaf metabolic model (Zhu et al., 2005; Zhu et al., 2007; Zhu et al., 2013) developed by Prof. Xin-Guang Zhu and colleagues at the Shanghai Institute of Plant Physiology and Ecology (Chinese Academy of Sciences, Shanghai). The metabolic model will be used to deepen the information that can be won from ^{13}C-labeling data from the transgenic rice lines.

More generally, metabolic network modelling is an *in-silico* approach to incorporate diverse data (genetic, transcriptomic, enzymatic and metabolic) to create a mathematical model of a specific organism in order to understand it better, but most importantly foresee what could be its physiological responses in case of external changes or internal manipulations. The diffusion of metabolic network modelling followed hand in hand with the diffusion and refinement of the so called "omics" techniques, which allowed the accumulation of massive amounts of biological data.

Aim of this thesis

In this cumulative thesis, I present two published papers and one ready for submission manuscript draft. These papers come from work done within the C_4 Rice Project or from experiments that were carried out to provide background knowledge regarding evolution of C_4 photosynthesis.

The first paper involves a comparison of CBC metabolite levels between C_3 and C_4 species. Using LC-MS/MS I was able to obtain quantitative information about the levels of almost all CBC metabolites in a range of C_3 and C_4 species. This information provides a top-down snapshot of the balance between the different steps in the pathway. The first aim of this paper was to provide more information on whether the CBC operates differently between C_3 and C_4 species. In C_4 species, the CBC had many millions of years to evolve while operating in a high-CO_2 environment within BS cells. After the

introduction of a C_4 cycle into rice, we anticipate that it will be necessary to fine tune the CBC and this study was carried out to provide more information about changes that have occurred in three different C_4 lineages. The second aim of this paper was to find out if the CBC and related metabolism vary between different C_4 species. This might be expected because they have evolved separately and differ in details of the pathways. To provide a good reference, my study included several C_3 species. Unexpectedly, I also found large species-to-species differences in CBC metabolite levels between C_3 species. Analysis of these differing patterns in C_3 species grew to be a third and major focus of the paper.

This unexpected finding of diversity in CBC operation between different C_3 species triggered experiments that are presented in the second paper of this thesis. In the first paper, metabolite levels were compared at ambient CO_2 and a single limiting irradiance. In the second paper, I carried out a detailed comparison of metabolite levels in *Arabidopsis thaliana* and rice at different light intensities and in low CO_2. This paper allowed me to publish the detailed analysis of rice metabolites levels, independently of their use for modelling within the C_4 Rice Project. This paper confirmed that there are major differences in how the CBC operates between the two species and provided first insights into what is responsible for these differences. The findings in these two papers are not only of academic interest, but also imply that the best strategies for improving photosynthesis may vary from crop to crop, with relevance not only for the C_4 Rice Project, but also for ongoing research in other projects to improve photosynthetic performance by other routes e.g., the B&MGF 'RIPE' project (https://ripe.illinois.edu/).

The third, and at this time unpublished manuscript, represents a large part of the experimental work carried out during my PhD project. In this work I compared metabolite levels across nine species belonging to the *Flaveria* genus to learn how primary metabolism, or more specifically metabolite levels, changed during the course of evolution from C_3 to C_3-C_4 intermediate, C_4-like and finally to C_4 photosynthesis. The main aim was to test predictions made by current models for the operation of photosynthesis in intermediate species. In particular, the idea that there are intercellular shuttles of metabolites (see above) is based mainly on considerations of pathway stoichiometry and of gene expression and protein localization patterns. The operation of such shuttles, which are presumably driven by diffusion, will require large pools of the involved metabolites. These predictions have, up to now, been poorly experimentally tested. Furthermore, measurements of metabolite content could also reveal which metabolites likely move between cell types. Measurement of metabolites could also shed

light on changes that might have occurred during the evolution of C_4-like photosynthesis into C_4 photosynthesis. More generally, the hope is to provide important background knowledge to inform strategies for engineering C_4 photosynthesis into crops.

Manuscripts overview and candidate contributions

This section contains the overview of the three manuscript included in this doctoral thesis, and my personal contribution for each paper.

Manuscript 1 - Metabolite profiles reveal interspecific variation in operation of the Calvin–Benson cycle in both C_4 and C_3 plants

Authors: Stéphanie Arrivault, Thiago Alexandre Moraes, Toshihiro Obata, David B. Medeiros, Alisdair R. Fernie, Alix Boulouis, Martha Ludwig, John E. Lunn, **Gian Luca Borghi**, Armin Schlereth, Manuela Günther, Mark Stitt

Status: Published on *Journal of Experimental Botany*, Volume 70, Issue 6, 1 March 2019, Pages 1843–1858, https://doi.org/10.1093/jxb/erz051

Candidate contributions: I contributed in growing the plants belonging to the two C_4 *Flaveria* species present in this study (*F. bidentis* and *F. trinervia*). I also harvested plant material from these plants, processed and extracted these samples and analyzed them on LC-MS/MS, providing CBC and 2PG data for these species. Moreover, I contributed by providing feedback during figure creation process and critical review of the whole manuscript.

Manuscript 2 - Relationship between irradiance and levels of Calvin–Benson cycle and other intermediates in the model eudicot Arabidopsis and the model monocot rice

Authors: **Gian Luca Borghi**, Thiago Alexandre Moraes, Manuela Günther, Regina Feil, Virginie Mengin, John E. Lunn, Mark Stitt, Stéphanie Arrivault

Status: Published on *Journal of Experimental Botany*, Volume 70, Issue 20, 15 October 2019, Pages 5809–5825, https://doi.org/10.1093/jxb/erz346

Candidate contributions: I contributed to this study by planning the irradiance experiments in rice and growing all the rice plants in controlled conditions. I also harvested all plant material from rice plants at different irradiances and CO_2 concentrations. Samples generated from this harvest were partially processed, extracted, and analyzed using LC-MS/MS by me, obtaining part of the metabolite data used for this manuscript. I also performed most of the gas exchange measurements on rice plants. I participated less in the creation of the figures, in which I only participated directly in Fig.

6 and relative statistical analyses. Overall, I contributed by providing feedback during figure creation process and critical review of the whole manuscript.

Manuscript 3 – Metabolic characterization of the transition from C_3 to C_4 photosynthesis in the genus *Flaveria*

Authors: **Gian Luca Borghi**, Stéphanie Arrivault, Manuela Günther, David B. Medeiros, Emilia Dell'Aversana, Giovanna M. Fusco, Petronia Carillo, Martha Ludwig, Alisdair R. Fernie, John E. Lunn, Mark Stitt

Status: Ready for submission in 2021

Candidate contributions: I contributed to this study by planning the normal and labelling experiments for all the *Flaveria* species included in this study. I took care of growth and propagation of the plants, and carried out both harvests (normal and $^{13}CO_2$ labelling one) by directly harvesting or directly supervising the harvest of all the plant material sampled in this study. I also processed samples generated from these harvest and carried out most of the sample extractions and downstream analyses using LC-MS/MS. I created the core datasets of this study, carried out statistical analyses and created the final figures and supplementary figures-tables of this paper. At the end, I wrote the whole manuscript, with fundamental feedback from my Principal Supervisor.

Manuscript one

Metabolite profiles reveal interspecific variation in operation of the Calvin–Benson cycle in both C_4 and C_3 plants

METABOLITE PROFILES REVEAL INTERSPECIFIC VARIATION IN OPERATION OF THE CALVIN–BENSON CYCLE IN BOTH C_4 AND C_3 PLANTS

Stéphanie Arrivault[1], Thiago Alexandre Moraes[1], Toshihiro Obata[1,*], David B. Medeiros[1], Alisdair R. Fernie[1], Alix Boulouis[1,†], Martha Ludwig[2], John. E. Lunn[1], Gian Luca Borghi[1], Armin Schlereth[1], Manuela Guenther[1], and Mark Stitt[1,‡]

[1] Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, D-14476 Potsdam-Golm, Germany

[2] School of Molecular Sciences, The University of Western Australia, 35 Stirling Hwy, Crawley WA 6009, Australia

[*] Present address: Department of Biochemistry, Center for Plant Science Innovation, University of Nebraska-Lincoln, 1901 Vine Str, Lincoln, NE 68588, USA.

[†] Present address: Institut de Biologie Physico-Chimique, CNRS - Sorbonne Université, Paris, France.

[‡] Correspondence: mstitt@mpimp-golm.mpg.de

Received 2 October 2018; Editorial decision 22 January 2019; Accepted 29 January 2019

Editor: Christine Raines, University of Essex, UK

ABSTRACT

Low atmospheric CO_2 in recent geological time led to the evolution of carbon-concentrating mechanisms (CCMs) such as C_4 photosynthesis in >65 terrestrial plant lineages. We know little about the impact of low CO_2 on the Calvin–Benson cycle (CBC) in C_3 species that did not evolve CCMs, representing >90% of terrestrial plant species. Metabolite profiling provides a top-down strategy to investigate the operational balance in a pathway. We profiled CBC intermediates in a panel of C_4 (*Zea mays*, *Setaria viridis*, *Flaveria bidentis*, and *F. trinervia*) and C_3 species (*Oryza sativa*, *Triticium aestivum*, *Arabidopsis thaliana*, *Nicotiana tabacum*, and *Manihot esculenta*). Principal component analysis revealed differences between C_4 and C_3 species that were driven by many metabolites, including lower ribulose 1,5-bisphosphate in C_4 species. Strikingly, there was also considerable variation between C_3 species. This was partly due to different chlorophyll and protein contents, but mainly to differences in relative levels of metabolites. Correlation analysis indicated that one contributory factor was the balance between fructose-1,6-bisphosphatase, sedoheptulose-1,7-bisphosphatase, phosphoribulokinase, and Rubisco. Our results point to the CBC having experienced different evolutionary trajectories in C_3 species since the ancestors of modern plant lineages diverged. They underline the need to understand CBC operation in a wide range of species.

INTRODUCTION

The Calvin–Benson cycle (CBC) evolved ~2 billion years ago (Rasmussen *et al.*, 2008), is the most abundant biochemical pathway on Earth in terms of nitrogen investment (Ellis, 1979; Raven, 2013), and plays a dominant role in the global carbon (C) and O_2 cycles. The CBC can be divided into three partial processes; fixation of CO_2 (ribulose-1,5-bisphosphate carboxylase-oxygenase) RuBisCO into a 3-C compound, 3-phosphoglycerate (3PGA), reduction of 3PGA to triose phosphate (triose-P) using ATP and NADPH from the light reactions, and a series of reactions that use triose-P to regenerate ribulose 1,5-bisphosphate (RuBP) (von Caemmerer and Farquhar, 1981; Heldt, 2005; Stitt *et al.*, 2010; Adam, 2017). The net gain in C exits the CBC and is converted into end-products. Despite its evolutionary age, the pathway's structure is essentially unchanged from cyanobacteria to angiosperms.

This conservation of the CBC pathway structure is remarkable. The CBC evolved in a world in which CO_2 concentrations were very high and O_2 concentrations were very low. Over geological time, there has been a dramatic rise in atmospheric O_2 and decline in atmospheric CO_2. This uncovered a side reaction with O_2, which competes with CO_2 as a substrate for RuBisCO, leading to the formation of 2-phosphoglycolate (2PG) (Lorimer and Andrews, 1973; Lorimer, 1981; Tcherkez *et al.*, 2006). 2PG is recycled via an energetically wasteful process termed photorespiration that results in the loss of 0.5 CO_2 per scavenged molecule of 2PG (Somerville, 2001; Heldt, 2005). In the current atmosphere with 0.04% CO_2 and 21% O_2, in C_3 plants about every fourth reaction is with O_2 instead of CO_2, leading to a 20–30% decrease in the net rate of photosynthesis (Osmond, 1981; Sharkey, 1988; Long *et al.*, 2006; Betti *et al.*, 2016). This side reaction decreases nitrogen use efficiency, because higher amounts of protein must be invested in the photosynthetic apparatus. This includes an especially large investment in RuBisCO, which has a relatively low catalytic rate and represents up to half of leaf protein (Ellis, 1979; Betti *et al.*, 2016). It negatively impacts water use efficiency because a higher internal CO_2 concentration is required to support a given net rate of photosynthesis, which in turn requires higher stomatal conductance and higher evaporative water loss (Ort *et al.*, 2015; Betti *et al.*, 2016).

Cyanobacteria and eukaryotic algae possess C-concentrating mechanisms (CCMs) that accumulate CO_2 in RuBisCO-containing microstructures, the carboxysome in cyanobacteria and the pyrenoid in eukaryotic algae (Badger *et al.*, 1998; Giordano *et al.*,

2005; Kerfeld and Melnicki, 2016; Raven *et al.*, 2017). These microstructures were lost in plant lineages that colonized the land. A second type of CCM evolved in terrestrial plants in the last 30 million years (Sage *et al.*, 2012; Raven *et al.*, 2017), coinciding with the decline of CO_2 from ~1000 ppm to <300 ppm during the Oligocene (Christin *et al.*, 2008; Zachos *et al.*, 2008; Edwards *et al.*, 2010). These CCMs are in essence biochemical CO_2 pumps, in which bicarbonate is fixed into 4-C acids that are subsequently decarboxylated to generate a high internal CO_2 concentration. In C_4 plants, bicarbonate is typically captured by phosph*enol*pyruvate (PEP) carboxylase in mesophyll cells, and 4-C acids diffuse to bundle sheath cells, which are located internally within the leaf and contain RuBisCO and the rest of the CBC (Hatch, 2002; von Caemmerer and Furbank, 2003; Sage *et al.*, 2012; Sage, 2017). There is substantial diversity in the pathway of C_4 photosynthesis; for example, which 4-C and 3-C metabolites are involved in the CCM, how the 4-C acid is decarboxylated, and to what extent PSII activity is lost in the bundle sheath chloroplasts. C_4 photosynthesis evolved independently >65 times in separate lineages among the angiosperms, and C_4 species currently represent ~3% of terrestrial plant species and account for 23% of total terrestrial C gain (Still *et al.*, 2003; Sage *et al.*, 2011; Sage, 2017). An analogous biochemical CO_2 pump evolved in plants with Crassulacean acid metabolism (CAM); bicarbonate is assimilated in the dark into 4-C acids, which are decarboxylated in the light to provide CO_2 for the CBC (Shameer *et al.*, 2018). CAM evolved in at least 35 independent lineages and is found in ~6% of current terrestrial plant species (Silvera *et al.*, 2010). Parallel evolution of C_4 and CAM in many lineages underlines the strong selective pressure exerted by low CO_2 in the recent geological past.

CCMs are complex traits. For example, C_4 photosynthesis requires major changes in leaf development and anatomy, gene expression patterns, and the location, levels, and properties of hundreds of enzymes and transporters (Sage *et al.*, 2012; Heckmann *et al.*, 2013; Sage, 2017). It is likely that its evolution involved successive steps, including the development of denser venation, modification of the size and functionality of bundle sheath cells, and stepwise specialization of metabolism in the bundle sheath and mesophyll cells (McKown and Dengler, 2007; Kocacinar *et al.*, 2008; Nelson, 2011; Sage *et al.*, 2013; Mallmann *et al.*, 2014). This multistep evolutionary trajectory may explain why CCMs evolved in only a relatively small fraction of terrestrial plant lineages (Heckmann, 2016).

Low CO_2 will have exerted massive selective pressure on the CBC in species that did not evolve a CCM, representing ~90% of existing terrestrial plant species (Silvera *et al.*, 2010; Sage, 2017). Pressure will also have been exerted by other environmental factors such as water availability, temperature, and nutrient availability (Raven *et al.*, 2017). Indeed, terrestrial C_3 plants exhibit substantial variation in photosynthetic rate, with large differences between annuals and perennials, and considerable differences within these groups (Evans, 1989; Wullschleger, 1993). This includes variation in photosynthetic rate between phylogenetically related species (Galmés *et al.*, 2014*b*) and within species (Driever *et al.*, 2014). Factors contributing to variation in photosynthetic rate include differences in the rate of electron transport and carboxylation (Wullschleger, 1993), leaf nitrogen content and photosynthetic nitrogen use efficiency (Field and Mooney, 1986; Evans, 1989; Hikosaka, 2010), and differing investment strategies in short-lived (deciduous) and long-lived (evergreen) leaves (Wright *et al.*, 2004; Donovan *et al.*, 2011).

We know relatively little about whether there is interspecific variation in the CBC in C_3 plants (Lawson *et al.*, 2012). It is well established that RuBisCO kinetics have evolved over a long geological time scale, with selectivity for CO_2 rising and catalytic rate declining between cyanobacteria and higher plants (Jordan and Ogren, 1981; Badger *et al.*, 1998; Tcherkez *et al.*, 2006; Savir *et al.*, 2010; Sharwood *et al.*, 2016a, b). Intriguingly, there is also variance over shorter evolutionary time scales. RuBisCO kinetics vary between quite closely related C_3 species (Yeoh *et al.*, 1980; Galmés *et al.*, 2014*a*; Prins *et al.*, 2016). In perennial oak, ecological adaptations have been linked to specific amino acid polymorphisms in RuBisCO (Hermida-Carrera *et al.*, 2017). RuBisCO is inhibited by RuBP and low molecular weight inhibitors that derive from catalytic infidelities of RuBisCO or, like 2-carboxyarabinitol 1-phosphate, are synthesized by other enzymes (Yeoh *et al.*, 1980; Parry *et al.*, 2008). There is surprising diversity in the levels and dynamics of these low molecular weight inhibitors in different C_3 species (Servaites *et al.*, 1986; Moore *et al.*, 1993; Charlet *et al.*, 1997; Parry *et al.*, 2008) and, incidentally, different C_4 species (Carmo-Silva *et al.*, 2010). CP12 is a small regulatory protein that interacts with NADP-glyceraldehyde-3-phosphate dehydrogenase (NADP-GAPDH) and phosphoribulokinase (PRK) (Gontero and Maberly, 2012; López-Calcagno *et al.*, 2014). The action of CP12 varies between C_3 species (Howard *et al.*, 2011; López-Calcagno *et al.*, 2014), again pointing to interspecies variation in CBC regulation.

Some of the strongest evidence that the CBC can adapt to selection or relaxation of selection in a relatively short evolutionary time comes from studies of C_4 species. Compared with C_3 species, C_4 species contain forms of RuBisCO with a lower affinity for CO_2 and faster catalytic turnover (Yeoh *et al.*, 1980; Sage and Seemann, 1993; Kapralov *et al.*, 2011; Galmés *et al*., 2014*b*; Sharwood *et al.*, 2016a, b), allowing a substantial decrease in RuBisCO abundance (Long, 1999; Ghannoum *et al.*, 2005; Sharwood *et al.*, 2016*a*, *b*, *c*). Such changes are found even within the tribe Paniceae in which C_4 photosynthesis evolved recently (Sharwood *et al.*, 2016*a*).

The operation of a pathway depends on many factors, including the abundance of the participating enzymes, their kinetic properties, and the action of regulatory mechanisms on individual enzymes and sets of enzymes. It is laborious to characterize variation in all these potential factors. Analyses of steady-state metabolite levels provide a top-down strategy to search for variation in pathway operation. This is because changes in enzyme abundance, properties, or regulation will all lead to changes in the relative levels of the metabolic intermediates in a pathway.

Information about CBC intermediate levels in different C_3 species is rather sparse. Most previous studies in C_3 plants focused on RuBP (e.g. Sage and Seemann, 1993) or a handful of metabolites such as 3PGA, triose-P, and fructose 1,6-bisphosphate (FBP), and were restricted to single species (see Stitt *et al.*, 2010 for references). A similar picture holds for C_4 plants (Stitt and Heldt, 1985; Usuda, 1987; Leegood and von Caemmerer, 1988, 1989). The reason was partly conceptual, reflecting the idea that photosynthesis is usually limited by the light reactions or RuBisCO (Farquhar *et al.*, 1980). Subsequent work has highlighted that photosynthesis can also be limited by reactions in the remainder of the CBC (see Stitt *et al.*, 2010 for a review), especially sedoheptulose-1,7-bisphosphatase (SBPase) (Raines *et al.*, 2000; Lefebvre *et al.*, 2005; Zhu *et al.*, 2007; Ding *et al.*, 2016; Driever *et al.*, 2017; Simkin *et al.*, 2017). There were also technical reasons; until ~10 years ago it was impossible to quantify many CBC intermediates routinely. This is now possible using HPLC-MS/MS (Cruz *et al.*, 2008; Arrivault *et al.*, 2009; Hasunuma *et al.*, 2010; Ma *et al.*, 2014).

In this study, we have profiled CBC intermediates in four C_4 species and five C_3 species, representing diverse plant lineages including eudicots and monocots. We used these data to address two questions. The first is whether CBC intermediates display different profiles in C_3 and C_4 species, as would be expected if the presence of a CCM allows a different mode of CBC operation. This question provides a check that expected differences in CBC

operation can be detected as changes in CBC metabolite profiles. In particular, we might expect that the lower abundance of RuBisCO (see above) results in lower levels of RuBP. Furthermore, C_4 species with dimorphic chloroplasts might have enhanced levels of 3PGA and triose-P to support an intercellular shuttle that transfers energy from the mesophyll to the bundle sheath cells. The second and major question is whether there are interspecific differences between C_3 species. This would have important implications for the evolution of the CBC and the need for a better understanding of the pathway in a broader range of C_3 species, including many of our major crops.

MATERIALS AND METHODS

Chemicals

Carbon dioxide ($^{13}CO_2$, isotopic purity 99 atom%) was from Campro Scientific GmbH (Berlin, Germany; www.campro.eu), N_2, O_2, and unlabelled CO_2 from Air Liquide (Germany; https://industrie.airliquide.de/), and chemicals were obtained from Sigma-Aldrich (Darmstadt, Germany; www.sigmaaldrich.com), Roche Applied Science (Mannheim, Germany; lifescience.roche.com), or Merck (www.merckmillipore.com).

Plant growth and harvest

Nine species (of which eight were phylogenetically diverse; Supplementary Fig. S1 at *JXB* online) were grown as described in Supplementary Table S1. Material was harvested by cutting leaves and quenching them immediately in a bath of liquid N_2 under growth irradiance, avoiding shading.

Metabolite analyses

Plant material was ground to a fine powder by hand in a mortar pre-cooled with liquid N_2 or in a cryo-robot (Stitt *et al.*, 2007) and stored at –80 °C. Metabolites were extracted and quantified by LC-MS/MS (Arrivault *et al.*, 2009). All samples were spiked with stable isotope-labelled internal standards for correction of ion suppression and other matrix effects (Arrivault *et al.*, 2015). 3PGA gives a broad, poorly defined peak in LC-MS/MS and was therefore quantified enzymatically (Merlo *et al.*, 1993).

Chlorophyll and protein

Chl *a* and *b* were extracted and quantified as in Gibon *et al.* (2002). Protein was extracted from 20 mg FW ground plant material in 750 µl of buffer [0.1 M Tris–HCl, pH 8, 0.2 M NaCl, 5 mM EDTA, 2% (w/v) SDS, 0.2% (v/v) β-mercaptoethanol, and protease inhibitor cocktail (P9599, Sigma, Germany)]. The suspension was mixed well, incubated (30 min, room temperature), re-mixed, centrifuged (10 min, 1500 *g*, 4 °C), and the supernatant collected. Supernatants were pooled from two or (*Oryza sativa* and *Manihot esculenta*) three successive extractions. Protein was quantified colorimetrically with bicinchoninic acid (BCA Protein Assay-Reducing Agent Compatible, Thermo Fisher Scientific, Germany; www.thermofisher.com) with BSA as standard.

Gas exchange

CO_2 assimilation was measured using the fourth fully expanded *Zea mays* leaf or 5-week-old *Arabidopsis thaliana* rosettes using an open-flow infrared gas exchange analyser system (LI-COR Inc., Lincoln, NE, USA; www.licor.com) equipped with an integrated fluorescence chamber head (LI-6400-40, 2 cm^2 leaf chamber for *Z. mays*; LI-6400-17 whole-plant Arabidopsis chamber for *A. thaliana*; LI-COR Inc.). CO_2 was kept at 400 µmol mol^{-1}, leaf temperature at 29 °C for *Z. mays* and at 20 °C for *A. thaliana*, and relative humidity at 65%.

$^{13}CO_2$ labelling with *M. esculenta*

The fifth or sixth fully expanded leaf from the top of a 9-week-old plant was labelled (Supplementary Fig. S2A), starting 2 h into the light period. The leaf was placed in the labelling chamber (Supplementary Fig. S2B, C; see Arrivault *et al.*, 2017). Gases were supplied from individual bottles and controlled by gas-flow controllers (Brooks instruments; www.brooksinstrument.com). The labelling chamber was initially supplied with 79% N_2, 21% O_2, and 420 ppm $^{12}CO_2$. After 1 min, $^{12}CO_2$ was replaced by $^{13}CO_2$. Samples were collected after 10, 20, 40, or 60 s, or 2, 5, 10, 30, or 60 min, in random order. Gas flow was 10 l min^{-1} for pulses of up to 1 min, and 5 l min^{-1} for longer pulses. Unlabelled samples (t=0) were collected after 1 min in unlabelled gas mixture. The chamber was maintained at growth cabinet temperature (28 °C) by circulating water from a water bath. Gases were passed through a humidifier in the water bath after mixing and before entering the measuring chamber. Light intensity at the leaf surface was kept as in

the growth cabinet (250 µmol m^{-2} s^{-1}) by supplying additional light (FL-460 Lighting Unit, Walz, Effeltrich, Germany). Material was quenched by dropping a copper rod, pre-cooled in liquid N_2, down a hollow tube incorporated in the chamber lid, thereby freeze-clamping a 1.9 cm diameter (~40 mg FW) leaf disc (Supplementary Fig. S2C, D). $^{13}CO_2$-labelled samples were analysed by LC-MS/MS and GC-MS, and isotopomer distribution (%) and enrichment (%) were calculated as in Arrivault *et al.* (2017).

Statistical analyses

Statistical analysis was performed in R Studio Version 0.99.896 (www.rstudio.com) with R version 3.3.0 (https://cran.r-project.org/) using either Student's *t*-test (R default package stats) or an ANOVA (Sums of Squares Type II) followed by the Tukey's Honest Significant Differences (HSD) post-test (R package agricolae). Details are provided in the figure legends.

RESULTS

Metabolite levels at growth irradiance

We profiled CBC metabolites in four C_4 species from the NADP-malic enzyme subtype including two monocots (*Zea mays* and *Setaria viridis*) and two eudicots (*Flaveria bidentis* and *F. trinervia*), and five C_3 species including two monocots (*Oryza sativa*, *Triticium aestivum*) and three eudicots (*Arabidopsis thaliana*, *Nicotiana tabacum*, and *Manihot esculenta*). Each species was grown with non-saturating irradiance (range of 60–133% of that required for half-maximal rates of photosynthesis) and appropriate temperature for rapid, healthy growth, and harvested under growth irradiance at least 2 h after the beginning of the light period (for details, see Supplementary Table S1). CBC intermediates and 2PG levels were determined by LC-MS/MS, using isotope-labelled internal standards to obtain reliable quantification, or enzymatically (3PGA). The signals for ribulose-5-phosphate (Ru5P) and xylulose-5-phosphate (Xu5P) overlapped, so they were combined ('Ru5P+Xu5P'). Otherwise, we were able to quantify all CBC intermediates except 1,3-bisphosphoglycerate, glyceraldehyde 3-phosphate, and erythrose 4-phosphate. Metabolites were initially normalized on FW.

CBC metabolite levels varied greatly between species (Fig. 1; Supplementary Dataset S1). This involved differences in the absolute and the relative levels of metabolites. Some

of the observed changes were expected, for example the low levels of 2PG in C_4 compared with C_3 species, reflecting the lower rate of photorespiration in the C_4 plants (note, 2PG amounts are multiplied by 10 for better visualization in Fig. 1). RuBP levels were lower in C_4 compared with C_3 species, probably reflecting lower abundance of RuBisCO in C_4 plants. However, other interspecies differences were unexpected, in particular the rather diverse profiles in the five C_3 species. Features that varied between the C_3 species included the absolute levels of individual metabolites such as 3PGA, triose-P, Ru5P+Xu5P, the level of RuBP compared with metabolites involved in RuBP regeneration, and the relative levels of metabolite pairs, for example FBP and fructose 6-phosphate (F6P) or sedoheptulose 1,7-bisphosphate (SBP) and sedoheptulose 7-phosphate (S7P).

Metabolite levels in *Z. mays* and *A. thaliana* at different irradiances

One potential complication of a cross-species comparison is that each species has a different light saturation response, making it difficult to standardize growth and harvest conditions across species. We grew and harvested all species at moderate and limiting irradiance, using lower irradiance for species whose photosynthesis saturates at lower light intensities (Supplementary Table S1). In addition, for *Z. mays* and *A. thaliana*, we asked whether short-term changes in irradiance lead to major changes in the metabolite profile, using an additional lower irradiance for *Z. mays* (Fig. 2A, covering the range from 40% to 133% of that required for half-maximal rates of photosynthesis), and a lower and a higher near-saturating irradiance for *A. thaliana* (Fig. 2B, covering the range from 67% to 233% of that required for half-maximal rates of photosynthesis). The metabolite profiles were not greatly altered for either species (Fig. 2C), except that higher irradiance tended to lead to a general increase in metabolite levels. Metabolite levels in a given species were strongly correlated irrespective of irradiance (r>0.98), whereas metabolite levels were poorly correlated between species (Fig. 2D).

Participation of pools in photosynthesis

Our approach assumes that the investigated metabolites are predominantly involved in the CBC. If they are also involved in another pathway, the total content will not provide reliable information about the size of the CBC pool. Published ^{13}C labelling kinetics validate this assumption for *N. tabacum*, *A. thaliana*, and *Z. mays* (Hasunuma *et al.*,

2010; Szecowka *et al.*, 2013; Arrivault *et al.*, 2017); after pulsing with $^{13}CO_2$, all of the CBC metabolites showed a rapid rise in ^{13}C enrichment to reach a final value of ≥80%. One exception was SBP in maize, where ^{13}C enrichment plateaued at ~14%. We performed analogous $^{13}CO_2$ labelling experiments for *M. esculenta* which, like *Z. mays*, is a subtropical species adapted to high-light conditions. We also chose *M. esculenta* because it has been suggested to be a C_4 or C_3–C_4 intermediate species (Cock *et al.*, 1987; El-Sharkawy and Cock, 1987). A subsequent study showed that *M. esculenta* performs C_3 photosynthesis (Edwards *et al.*, 1990; see also De Souza *et al.*, 2017; De Souza and Long, 2018). Time-resolved $^{13}CO_2$ labelling would provide a further test that *M. esculenta* is a C_3 species

In *M. esculenta*, CBC intermediates rose rapidly to high (>75%) ^{13}C enrichment (Supplementary Fig. S3A; Supplementary Dataset S2) except for SBP where enrichment plateaued at ~40% and about half of the SBP remained in the unlabelled form after 60 min (Supplementary Fig. S3B). Otherwise, the labelling time series in *M. esculenta* resembled published time series for the C_3 plants *A. thaliana* (Szecowka *et al.*, 2013) and *N. tabacum* (Hasunuma *et al.*, 2010). In particular, labelling of 4-C acids was very slow (Supplementary Fig. S3C).

Chlorophyll and protein

Leaf composition varies between species (see the Introduction). This could contribute to interspecific differences in absolute metabolite levels; in particular, differences in leaf composition could lead to systematically higher or lower levels of all metabolites. We therefore determined total chlorophyll and protein contents in the leaf material used for metabolite analyses. Total chlorophyll content (Fig. 3A) was similar on a FW basis in all species except for *O. sativa* and *M. esculenta*, which had considerably higher values. Protein content on a FW basis (Fig. 3B) was similar in all species except for lower values in *N. tabacum*, and higher values in *O. sativa* and, especially, *M. esculenta.* These results partly explain why CBC metabolite levels on a FW basis tended to be low in *N. tabacum* and high in *O. sativa* and *M. esculenta* (Fig. 1).

Principal component analysis

We performed principal component (PC) analyses to provide an integrated overview of the CBC metabolite profiles in the nine species. PC analysis gives information about

which samples (here, different species) are closely related or separated, and which variables (here, metabolites) contribute to this relationship. The analysis was performed with *z*-scored data (i.e. normalizing the individual values of a given variable on the mean value for that variable) to ensure that each metabolite adopted an equally important role in the analysis, independent of its absolute abundance. Each individual sample was included separately in the analysis to provide an overview of the quality of within-species replication. We included the low light maize and the low and high light Arabidopsis samples to further test the impact of prevailing irradiance. In the analyses shown in Fig. 4, we omitted 2PG to focus solely on the CBC and exclude effects due to lower photorespiration in C_4 plants. We also omitted SBP because of the labelling data (Supplementary Fig. S3; Arrivault *et al.*, 2017) indicating that in some species part of the SBP pool is not involved in the CBC. For comparison, analyses including 2PG and SBP are provided in Supplementary Figs S4–S7.

As previously mentioned, some cross-species variation in metabolite levels may be driven by changes in leaf composition. We therefore performed PC analyses on data sets in which the metabolites were normalized on FW (Fig. 4A; Supplementary Fig. S4), total chlorophyll content (Fig. 4B; Supplementary Fig. S5), or protein content (Fig. 4C; Supplementary Fig. S6). We also performed PC analysis on a dimensionless data set in which, for a given species, the amount of C in a given metabolite was divided by the total amount of C in all CBC intermediates plus 2PG (Fig. 4D; Supplementary Fig. S7). In total, we performed 16 PC analyses with different metabolite data sets and normalizations. In interpreting the plots, we focused on features that were seen in all or the vast majority of these analyses.

In analyses with the FW-, chlorophyll-, and protein-normalized data sets and the dimensionless data set, PC1 accounted for 44–46, 33–35, 34–36, and 39–40%, respectively, of the total variance, while PC2 accounted for 17–20, 20–23, 21–22, and 19–22%, respectively (Fig. 4A–C; Supplementary Figs S4–S6). In all cases, replicates for a given species grouped together, showing that within-species variance was smaller than interspecies differences. This included samples harvested at low and ambient light intensities for *Z. mays* and for *A. thaliana*. The *A. thaliana* samples collected at high light grouped separately from the other *A. thaliana* samples, but well removed from the other species in PC analyses with the FW-, chlorophyll-, and protein-normalized data sets. In PC analyses with the dimensionless data set, *A. thaliana* samples from all three light intensities grouped together (Fig. 4D; Supplementary Fig. S7), showing that increasing

light intensity led mainly to a general increase in metabolite levels rather than to changes in their relative levels.

Inspection of the species distribution in the PC plots leads to three main conclusions. First, the PC analyses almost always separated C_4 species from C_3 species; this holds irrespective of how the metabolite data are normalized, and whether 2PG and SBP were excluded (Fig. 4) or included (Supplementary Figs S4–S7). *Manihot esculenta* showed a slight overlap with the *Flaveria* spp. in the analyses using metabolites minus SBP and 2PG, when the data set was normalized on protein (Fig. 4C), but was fully separated from all of the C_4 species in the 15 other PC analyses. *A. thaliana* in low light showed a slight overlap with *Z. mays* or *S. viridis* in two (all metabolites normalized on FW, metabolites minus SBP normalized on FW; Supplementary Fig. S4) of the 16 data permutations. Secondly, within the C_4 species, *Z. mays* and *S. viridis* separated from each other and from the *Flaveria* spp. in most of the PC analyses, while the two *Flaveria* spp. always overlapped with each other. Thirdly, the five C_3 species were almost always clearly separated from each other. In the analyses based on FW-normalized data, *O. sativa* and *M. esculenta* separated strongly from other C_3 species in PC1 (Fig. 4A; Supplementary Fig. S4). This was less marked in the PC analysis based on chlorophyll- or protein-normalized data (Fig. 4B, C; Supplementary Figs S5, S6), indicating that the strong separation in the analysis with FW-normalized data is partly driven by secondary effects due to leaf composition. Similarly, *N. tabacum* was less strongly separated from the other four C_3 species in the PC analysis with protein-normalized data than with FW- or chlorophyll-normalized data. Despite these small shifts in the relationships, the five C_3 species still separated from each other in the PC analyses with the chlorophyll- and protein-normalized data sets, as well as with the dimensionless data set (Fig. 4D; Supplementary Fig. S7).

The metabolite loadings (Fig. 4; Supplementary Figs S4–S7) reveal that the separation of C_4 from C_3 species was driven not only by lower levels of RuBP and (when included) 2PG, but also by other CBC metabolites. 3PGA and triose-P contributed to the separation of the C_4 species *Z. mays* and *S. viridis* from *F. trinervia* and *F. bidentis* (see the Discussion). Almost every metabolite contributed to the separation between the five C_3 species, with large contributions from RuBP, FBP, F6P, S7P, ribose 5-phosphate (R5P), triose-P, and 3PGA.

We repeated the PC analysis on a data set including only C_3 species and with metabolites normalized on total chlorophyll content or protein content, and with a dimensionless data

set (Fig. 5; Supplementary Figs S8–S10). Replicate samples from a given species grouped closely together. *A. thaliana*, *N. tabacum*, and *M. esculenta* were clearly separated from *T. aestivum* and *O. sativa*, which were only weakly separated. The high irradiance *A. thaliana* samples grouped separately from the low and medium light *A. thaliana* samples, but in the same tangent, and were clearly separated from the other four C_3 species. Metabolite loadings revealed strong contributions from 3PGA, triose-P, RuBP, FBP, F6P, and S7P to the separation.

Coefficient of variance

We calculated the coefficient of variance (CV) to determine which metabolites showed the greatest interspecies variance for all nine species together, for the four C_4 species, and for the five C_3 species (Fig. 6). To avoid influence due to leaf composition, this analysis was performed on the dimensionless data set. Across all species (Fig. 6A), the highest CV was for 2PG, followed by FBP, RuBP, triose-P, SBP, and R5P. When only C_4 species are considered (Fig. 6B), the highest CV was for SBP, followed by 2PG, RuBP, R5P, FBP, and triose-P. When only C_3 species are considered (Fig. 6C), the highest CV was for FBP, followed by Ru5P+Xu5P, 2PG, R5P, and 3PGA.

Correlation analysis

When metabolite profiles are compared across different genotypes, they typically generate a correlation network (Meyer *et al.*, 2007; Sulpice *et al.*, 2009, 2013; Zhang *et al.*, 2015; Wu *et al.*, 2016). This reflects features of the underlying metabolic pathways that are maintained across genotypes and generate conserved relationships between metabolites. Our data set allowed us to apply this approach to interspecies variation in the CBC. We performed pairwise PC analysis and clustering on CBC metabolites (Fig. 7) using the dimensionless data set to avoid bias from changes in leaf composition. We also searched for relationships between 2PG and the CBC metabolites. Metabolite pairs that are linked by irreversible reactions (Bassham and Krause, 1969; Mettler *et al.*, 2014) are indicated by black boxes in the figure. Correlation analysis and clustering were performed for all nine species (Fig. 7A), for the four C_4 species (Fig. 7B), and for the five C_3 species (Fig. 7C). To aid visual comparison across the species sets, correlation coefficients are also shown in Supplementary Fig. S11 with the metabolites in a fixed order corresponding to CBC topology.

The CBC correlation network for all nine species (Fig. 7A) contained six positive correlations (e.g. F6P versus S7P; all pairwise comparisons between RuBP, FBP, and SBP), many non-significant relationships [e.g. FBP versus F6P; SBP versus S7P; RuBP versus R5P and Ru5P+Xu5P (here collectively called pentose-P)], and 13 negative correlations (e.g. 3PGA or triose-P versus most other CBC metabolites). In some cases, the correlations were driven by differences between C_4 and C_3 species; for example, the positive correlation between 2PG and RuBP is driven by the lower levels of both metabolites in C_4 compared with C_3 species (see Fig. 1). However, in many cases, the correlations were also seen within the subset of C_4 and within the subset of C_3 species (see Supplementary Fig. S11 and below).

The correlation network for CBC metabolites in C_4 species (Fig. 7B) contained nine positive (e.g. FBP versus SBP; FBP versus RuBP; and triose-P versus FBP, SBP, and RuBP) and 13 negative (e.g. 3PGA versus triose-P, FBP, and SBP; triose-P versus S7P and pentose-P; RuBP versus F6P, S7P, and pentose-P; and SBP versus S7P) relationships. There was no significant relationship between FBP and F6P. The correlation network for CBC metabolites in C_3 species (Fig. 7C) contained six positive (e.g. all pairwise comparisons between RuBP, FBP, and SBP; and F6P versus S7P) and 11 negative (e.g. 3PGA versus triose-P, FBP, SBP, and RuBP; RuBP versus F6P and S7P; and SBP versus S7P) relationships. There was no significant relationship between FBP and F6P, or between RuBP and pentose-P. 2PG correlated positively with S7P and negatively with RuBP in C_4 and C_3 species, respectively, and positively with 3PGA and negatively with triose-P and FBP in C_3 species.

The correlation networks can be interpreted by relating them to CBC topology (Fig. 7D; see also Supplementary Fig. S11). Figure 7D focuses on correlations seen within the subset of C_4 species and within the subset of C_3 species. Triose-Ps are used to synthesize FBP and SBP in reversible reactions catalysed by aldolase. This may explain the positive correlations between triose-P and FBP or SBP (except for SBP in C_3 plants). FBPase and SBPase catalyse irreversible reactions. The non-significant or negative correlations between FBP and F6P and between SBP and S7P point to interspecies variance in the regulation of FBPase and SBPase. This may also explain the absence of a positive correlation between triose-P and pentose-P that otherwise might have been expected because pentose-Ps are formed from triose-P and F6P or S7P in reversible reactions catalysed by transketolase (TK). The negative relationship between pentose-P and RuBP points to interspecies variation in the regulation of PRK. Further, the positive correlations

of FBP and SBP with RuBP (see Supplementary Fig. S11) indicate that FBPase and SBPase activity vary reciprocally to PRK activity and/or co-ordinately with binding or use of RuBP by RuBisCO.

DISCUSSION

The CBC is an ancient pathway that has been under selective pressure due to the long-term increase of the O_2:CO_2 ratio in the atmosphere and particularly over the last 30 million years due to falling CO_2 concentrations, which led to independent evolution of a CCM in >100 terrestrial plant lineages. However, the vast majority of terrestrial species did not evolve a CCM, probably because they were unable to follow the multistep evolutionary trajectory that was required to acquire this complex trait (Sage *et al.*, 2012; Christin and Osborne, 2013; Heckmann *et al.*, 2013). Present-day C_3 plants nevertheless will have been subject to similar selective pressures to those that drove the evolution of C_4 or CAM photosynthesis. Indeed, in the absence of a CCM, the selective pressures on the CBC may have been even greater than in plants that did evolve a CCM. In addition to low CO_2, it is likely that environmental factors such as irradiance, temperature, and nutrient and water availability exerted more or less selective pressure, depending on the local environment, and leading to different evolutionary trajectories in different populations. While it is well documented that there is large variation in photosynthetic rate between terrestrial species (Evans, 1989; Wullschleger, 1993; Lawson *et al.*, 2012), previous studies of the underlying causes have focused on leaf morphology and composition (Field and Mooney, 1986; Evans, 1989; Hikosaka, 2010; Poorter *et al.*, 2015; Díaz *et al.*, 2016), stomatal conductance (Lawson *et al.*, 2012), and the kinetic characteristics of RuBisCO (Yeoh *et al.*, 1980; Jordan and Ogren, 1981; Badger *et al.*, 1998; Tcherkez *et al.*, 2006; Galmés *et al.*, 2014*b*; Prins *et al.*, 2016; Sharwood *et al.*, 2016a, b). Little is known about whether the CBC operates in a highly conserved manner or in different modes in different C_3 species.

We have used metabolite profiling as an unbiased strategy to search for interspecific variance in CBC operation. The underlying assumption is that changes in the balance between different enzymatic steps will lead to changes in the relative levels of pathway intermediates. This approach is top down, in the sense that it does not make assumptions about whether the observed variance is due to changes in gene expression and protein

abundance, enzyme kinetics, or regulatory networks that act on the enzymes. We applied it to search for differences in CBC operation between C_4 and C_3 plants, and within C_3 species. As our aim was to compare CBC operation across species, we focused exclusively on the metabolites that are involved in the CBC plus 2PG, the immediate product of the RuBisCO oxygenation reaction. We excluded metabolites involved further downstream in photorespiration and metabolites involved in the CO_2-concentrating shuttle in C_4 plants, which have non-photosynthetic functions in C_3 plants.

Our interspecies comparison required important control experiments and cross-checks during data analysis. First, plant species differ in their photosynthetic rate and its dependence on light, temperature, and the availability of water, nutrients, and CO_2 (see the Introduction). We grew and harvested plants in a light regime that was limiting for that species, rather than using identical conditions for all species. In these conditions, RuBP regeneration is likely to be limiting, and effects of light stress are avoided. Importantly, we showed for one C_4 species (*Z. mays*) and one C_3 species (*A. thaliana*) that although increased harvest irradiance led to higher levels of metabolites, it did not strongly alter their relative levels (Fig. 2). Secondly, it is important that the CBC pool accounts for most or all of the total content of a given metabolite. Analysis of published data for two C_3 (*N. tabacum* and *A. thaliana*), one C_4 (*Z. mays*) species (Hasunuma *et al.*, 2010; Szecowka *et al.*, 2013; Arrivault *et al.*, 2017), and a new data set for the C_3 species *M. esculenta* (Supplementary Fig. S3) showed that CBC intermediates exhibit a rapid rise in ^{13}C enrichment to a high level after supplying $^{13}CO_2$. This provides evidence that most of the total pool is indeed involved in the CBC. This conclusion is supported by published subcellular fractionation studies, in which most CBC intermediates are exclusively or largely confined to the plastid (Gerhardt *et al.*, 1987; Szecowka *et al.*, 2013). The only exception was SBP, which was only partially labelled in *Z. mays* and *M. esculenta*. We do not know whether there is a separate pool of SBP that is not involved in CO_2 fixation, or if these plant species contain an unknown metabolite with an identical chromatographic behaviour, mass, and fragmentation pattern to SBP. In our interpretation of the metabolite profiles, we took care that our conclusions did not depend on inclusion of SBP. A third set of controls addressed the issue that leaf composition varies between species, with the result that absolute values for metabolite content will depend on the unit in which they are given. We analysed metabolite data normalized on FW, chlorophyll, or protein content, and also used a dimensionless data set in which metabolite levels were expressed relative to each other. Our interpretation

focused on results that were independent of how the data were normalized. Importantly, inclusion of the dimensionless data set eliminated secondary correlations due to differences in leaf composition, and placed the emphasis on relative rather than absolute levels of metabolites. It minimizes contributions from differing light regimes, which had less effect on relative than on absolute metabolite levels (see above).

We included four C_4 species in our panel to test if CBC profiles could distinguish between species in which it is known that the CBC operates in a different context from that of C_3 plants. The CBC operates at a much higher intercellular CO_2 concentration in C_4 than in C_3 plants, and RuBisCO has a higher affinity for CO_2, and an increased catalytic rate in C_4 compared with C_3 species (see the Introduction). PC analysis confirmed that CBC metabolite profiles allow C_4 and C_3 species to be distinguished (Fig. 4; Supplementary Figs S4–S7). As expected, C_4 species had lower 2PG and RuBP than C_3 species (Fig. 1). However, the separation in the PC analysis was also seen when 2PG was excluded, and was driven by several other CBC intermediates, pointing to broader changes in CBC operation between C_4 and C_3 species.

The four C_4 species belong to the NADP-malic enzyme C_4 subtype. Interestingly, PC analysis separated Z. *mays* and *S. viridis* from the two *Flaveria* spp. Whilst this might reflect a difference between monocots and eudicots, the PC vectors indicated that this separation reflected higher levels of 3PGA and, in particular, triose-P in *Z. mays* and *S. viridis* (Fig. 4; Supplementary Figs S4–S7; see also Fig. 1). Most NADP-malic enzyme C_4 subtypes, including *Z. mays*, have dimorphous chloroplasts with little or no PSII activity in the bundle sheath cells (Munekage, 2016). They operate an intercellular shuttle in which 3PGA moves from the bundle sheath to the mesophyll cells and is reduced to triose-P, which returns to the bundle sheath. Intercellular movement is thought to occur by diffusion (Hatch and Osmond, 1976), driven by concentration gradients that require the build-up of large pools of 3PGA and triose-P in the bundle sheath and mesophyll cells, respectively (Leegood, 1985; Stitt and Heldt, 1985; Arrivault *et al.*, 2017). *Flaveria bidentis* and *F. trinervia* can have PSII activity in the bundle sheath chloroplasts, although to a varying extent depending on conditions (Laetsch and Price, 1969; Höfer *et al.*, 1992; Meister *et al.*, 1996; Nakamura *et al.*, 2013). Their separation in the PC analysis from *Z. mays* and *S. viridis* might reflect decreased reliance on this intercellular shuttle.

Our panel included five C_3 species, two monocots (*O. sativa* and *T. aestivum*) and three eudicots (*A. thaliana*, *N. tabacum*, and *M. esculenta*), with the individual species representing different phylogenetic lineages (Supplementary Fig. S1) and originating in

differing climatic zones. The three C_3 eudicot species represent two of the major lineages within the eudicots, namely the asterids (*N. tabacum*) and rosids (*A. thaliana* and *M. esculenta*), that contain 41% and 24% of all angiosperms, respectively. There was considerable interspecies variation in CBC metabolite profiles. This was evident from visual inspection of the metabolite levels (Fig. 1) and was confirmed by PC (Figs 4, 5; Supplementary Figs S4–S10) and variance (Fig. 6) analyses.

When metabolites were expressed on a FW basis, some of the variation was due to differences in leaf composition, with a strong trend to higher absolute levels in *O. sativa* and *M. esculenta*, reflecting their high chlorophyll and protein content. The high protein content in *O. sativa* may be linked to changes in leaf anatomy that enhance mesophyll transfer conductance, including small deeply lobed cells and densely arranged chloroplasts and stromules at the cell surface (Sage and Sage, 2009; Busch *et al.*, 2013). This high mesophyll transfer conductance may prevent internal CO_2 from being drawn down by the high CBC activity that results from the high protein and metabolite content per unit FW in *O. sativa*. The high protein content in *M. esculenta* resembles the findings of previous reports (Awoyinka *et al.*, 1995; Nassar and Marques, 2006), and could explain the high rates of photosynthesis in this species.

However, the five C_3 species still showed differing CBC metabolite profiles when metabolites were expressed on a chlorophyll or protein basis, or when the analyses were performed with a dimensionless data set. Variation was driven by many metabolites including RuBP, 3PGA, triose-P, FBP, F6P, S7P, and Ru5P+Xu5P. This variation points to different operating modes of the CBC in different C_3 species. There were also differences in 2PG content; this might be related to the rate of RuBisCO oxygenation or removal of 2PG by 2-phosphoglycolate phosphatase.

Cross-species correlation analysis (Fig. 7; Supplementary Fig. S11) revealed that in both C_4 and C_3 species, the interspecies variance often included parallel changes of FBP, SBP, and RuBP, and unrelated or even reciprocal changes of these metabolites to F6P, S7P, and pentose-P. This is consistent with interspecies variation in the balance between FBPase, SBPase, PRK, and RuBisCO activity. It could reflect differences in the abundance or the regulation of these enzymes, both within C_3 species and within C_4 species, and between C_3 and C_4 species. Little is known about the expression, characteristics, and regulation of CBC enzymes in different species, with (see the Introduction) the exception of RuBisCO.

Our results do not reveal when and under what circumstances the variation in CBC function in C_3 species appeared. It is tempting to link it with the selection pressure that led to the appearance of C_4 and CAM photosynthesis, but it is likely to have started even earlier. Further, as pointed out by Zhu *et al.* (2007), it is possible that different C_3 species are following different trajectories during the increase in CO_2 levels in very recent evolutionary time. Our results also indicate that there is no strong connection between phylogeny and the diversity in CBC metabolite profiles in C_3 species. In the PC analyses (Figs 4, 5; Supplementary Figs S4–S10), the two monocot species are often closely related, but the three eudicot species are highly diverse, and a given eudicot is often more closely related to the monocot species than to the other eudicot species. Unlike changes in genome sequence, complex emergent phenotypes may not accrue in a linear manner, and phylogenetically distinct species may undergo convergent evolution whilst phylogenetically related species may undergo divergent evolution, depending on the selective pressure they experience. Better understanding of the relationship between diversity in CBC profile, phylogeny, and evolution will require studies both with more phylogenetically diverse species and with more dense sampling in short evolutionary space.

In conclusion, marked differences in CBC metabolite profiles between five C_3 species, including the major crop plants *O. sativum*, *T. aestivum*, and *M. esculenta*, and the important model plants *A. thaliana* and *N. tabacum*, reveal interspecies variation in the operating mode of the CBC in C_3 plants. This probably reflects independent evolution of CBC regulation in different plant lineages, in analogy to the independent evolution of a CCM in different plant lineages. These findings, together with emerging evidence for interspecies variation in the properties of specific CBC enzymes (see the Introduction) and the growing realization that efficient photosynthesis requires integrated operation of the CBC (Stitt *et al.*, 2010; Raines, 2011; Simkin *et al.*, 2017), highlight the need for a mechanistic understanding of CBC regulation in a wider range of species. This will be an important step towards improving C_3 photosynthesis and crop productivity.

SUPPLEMENTARY DATA

Supplementary data are available at *JXB* online.

- Fig. S1. Phylogenetic distribution based on APGIII of the tested plant species.
- Fig. S2. Experimental set-up for $^{13}CO_2$ labelling of *M. esculenta.*
- Fig. S3. ^{13}C enrichment (%) of CBC metabolites, relative abundance (%) of SBP isotopomers, and ^{13}C enrichment (%) of malate, aspartate, pyruvate, and alanine in *M. esculenta.*
- Fig. S4. PC analyses of all species using metabolite data normalized on FW (supplementary analyses to Fig. 4A).
- Fig. S5. PC analyses on all species using metabolite data normalized on total chlorophyll content (supplementary analyses to Fig. 4B).
- Fig. S6. PC analyses on all species using metabolite data normalized on protein content (supplementary analyses to Fig. 4C).
- Fig. S7. PC analyses on all species using a dimensionless data set (supplementary analyses to Fig. 4D).
- Fig. S8. PC analyses on C_3 species only, using metabolite data normalized on total chlorophyll content (supplementary analyses to Fig. 5A).
- Fig. S9. PC analyses on C_3 species only, using metabolite data normalized on protein content (supplementary analyses to Fig. 5B).
- Fig. S10. PC analyses on C_3 species only, using a dimensionless data set (supplementary analyses to Fig. 5C).
- Fig. S11. Correlation between levels of CBC metabolites, with metabolites shown in a fixed order reflecting the reaction sequence in the CBC.
- Table S1. Growth conditions and photosynthetic rates.
- Dataset S1. Metabolite levels, total chlorophyll, and protein contents in different species.
- Dataset S2. Labelling kinetics of CBC and other intermediates after exposing *M. esculenta* to $^{13}CO_2$ (supplementary data to Supplementary Fig. S3).

ABBREVIATIONS

- **C** = carbon
- **CAM** = Crassulacean acid metabolism
- **CBC** = Calvin–Benson cycle
- **CCM** = carbon-concentrating mechanism
- **CV** = coefficient of variance
- **DHAP** = dihydroxyacetone phosphate
- **E4P** = erythrose 4-phosphate
- **FBP** = fructose 1,6-bisphosphate
- **FBPase** = fructose-1,6-bisphosphatase
- **F6P** = fructose 6-phosphate
- **GAP** = glyceraldehyde 3-phosphate
- **NADP-GAPDH** = NADP-glyceraldehyde-3-phosphate dehydrogenase
- **PC** = principal component
- **PEP** = phospho*enol*pyruvate
- **2PG** = 2-phosphoglycolate
- **3PGA** = 3-phosphoglycerate
- **PRK** = phosphoribulokinase
- **R5P** = ribose 5-phosphate
- **RuBP** = ribulose 1,5-bisphosphate
- **Ru5P** = ribulose 5-phosphate
- **SBP** = sedoheptulose 1,7-bisphosphate
- **SBPase** = sedoheptulose-1,7-bisphosphatase
- **S7P** = sedoheptulose 7-phosphate
- **TK** = transketolase
- **triose-P** = triose phosphate
- **Xu5P** = xylulose 5-phosphate

ACKNOWLEDGEMENTS

This research was supported by the Max Planck Society (TAM, ARF, AB, JEL, AS, MG, and MS), the Bill and Melinda Gates Foundation (CASS to SA and TO; C_4 Rice to GLB), the German Ministry of Education and Research (FullThrottle, grant 031B0205A to DBM), CNPq (to TAM), and the Australian Research Council (to ML, JEL, and MS). We thank Christin Abel, Ina Krahnert, and Dr. Mark Aurel Schöttler for help with plant growth.

MANUSCRIPT ONE – MAIN FIGURES

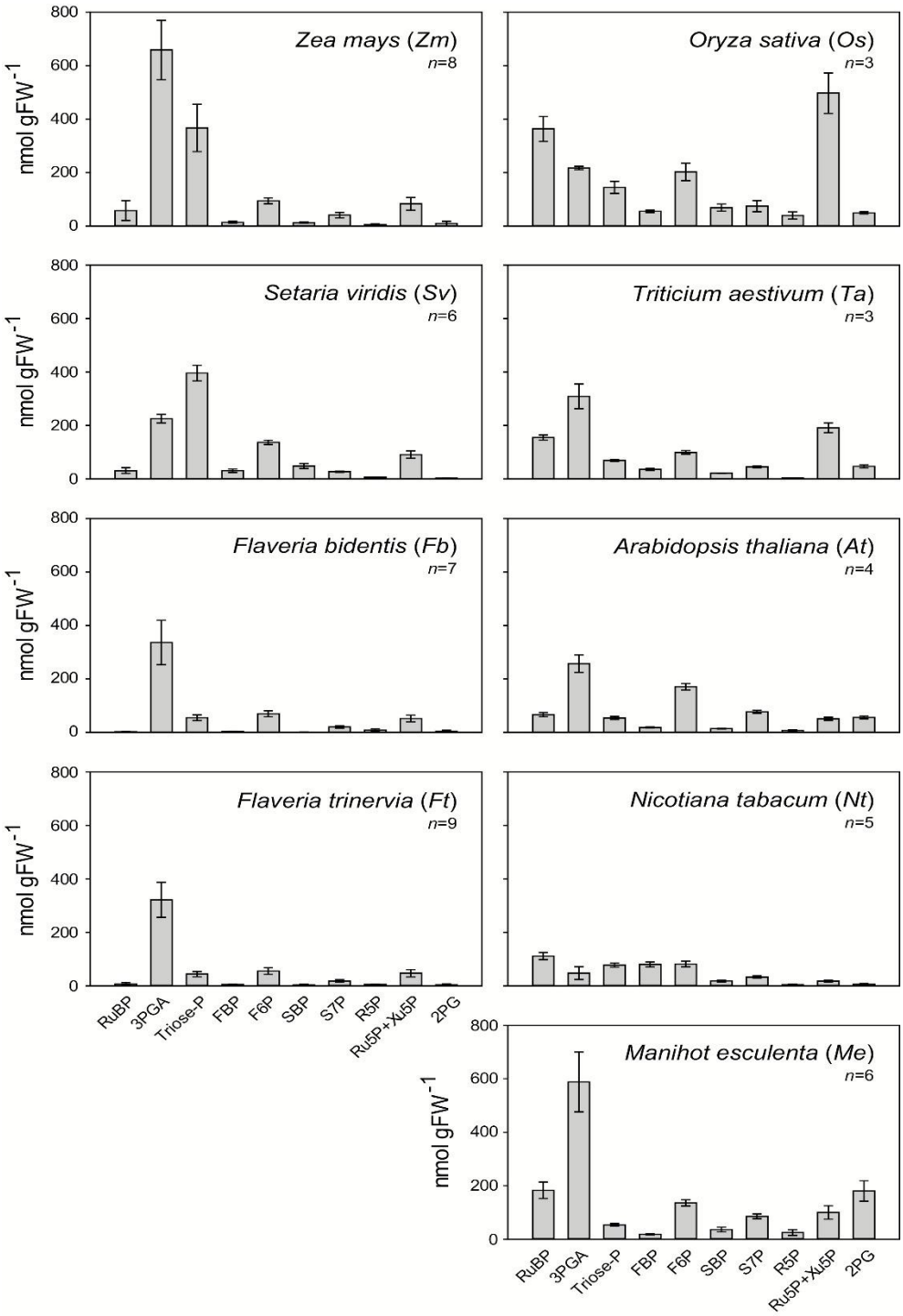
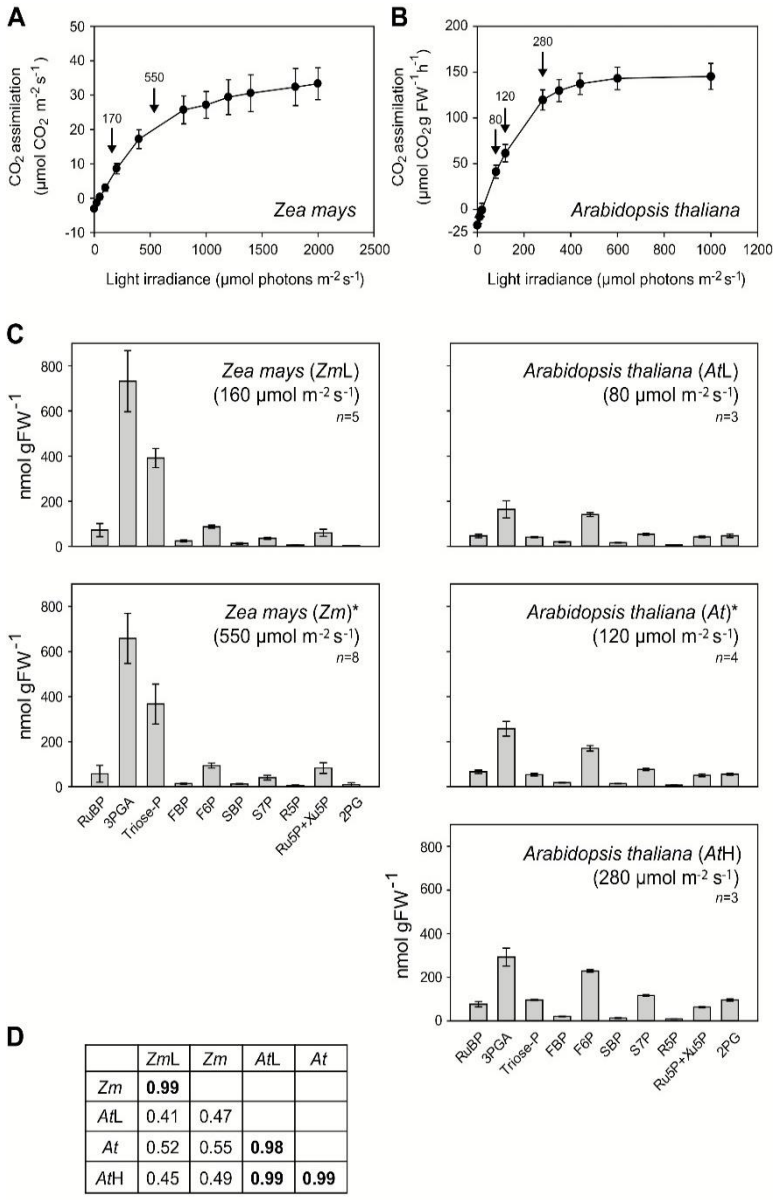


Fig. 1. CBC metabolite and 2PG profiles in different species. Growth and harvest conditions can be found in Supplementary Table S1. Note that 2PG amounts are multiplied by 10 for better visibility. The results are shown as mean (nmol g FW^{-1}) ±SD. The original data are provided in Supplementary Dataset S1.



	ZmL	Zm	AtL	At
Zm	**0.99**			
AtL	0.41	0.47		
At	0.52	0.55	**0.98**	
AtH	0.45	0.49	**0.99**	**0.99**

Fig. 2. CO_2 assimilation rate in *Z. mays* and *A. thaliana*, and CBC metabolite profiles in *Z. mays* and *A. thaliana* at different short-term irradiances. *Zea mays* and *A. thaliana* were grown at 550 µmol m^{-2} s^{-1} and 120 µmol m^{-2} s^{-1} irradiance, respectively. CO_2 assimilation rate in (A) Z. mays (n=10) and (B) A. thaliana (n=9). The results are shown as mean (µmol CO_2 m^{-2} s^{-1} and µmol CO_2 g FW^{-1} h^{-1}, for *Z. mays* and *A. thaliana*, respectively) ±SD. Arrows indicate the irradiances at which leaves were sampled for metabolite analysis. (C) Zea mays was harvested at growth irradiance (Zm, medium irradiance) or after being subjected for 4 h to 160 µmol m^{-2} s^{-1} (ZmL, low irradiance) from the beginning of the light period. *Arabidopsis thaliana* was harvested at growth irradiance (At, medium irradiance) or subjected for 15 min to 80 µmol m^{-2} s^{-1} or 280 µmol m^{-2} s^{-1} (AtL, low and AtH, high irradiances, respectively). Quenching of metabolism and harvest of leaf tissue were performed at least 4 h after the beginning of the light period. 2PG amounts are multiplied by 10 for better visibility. Asterisks indicate graphs already presented in Fig. 1. The results are shown as mean (nmol g FW^{-1}) ±SD. (D) Correlation analysis. The metabolite data shown in (B) and (C) were used to perform Pearson's correlation analysis between data sets from the same species at different irradiances, and correlations between different species. Before performing the correlation analysis, each data set was normalized by calculating the amount of carbon in a given metabolite, and dividing it by the total carbon in all metabolites in that data set. This was done to avoid secondary correlation due to any interspecies differences in leaf composition. The results are given as r and the higher correlations are indicated in bold. All correlations were positive. The original data are presented in Supplementary Dataset S1.

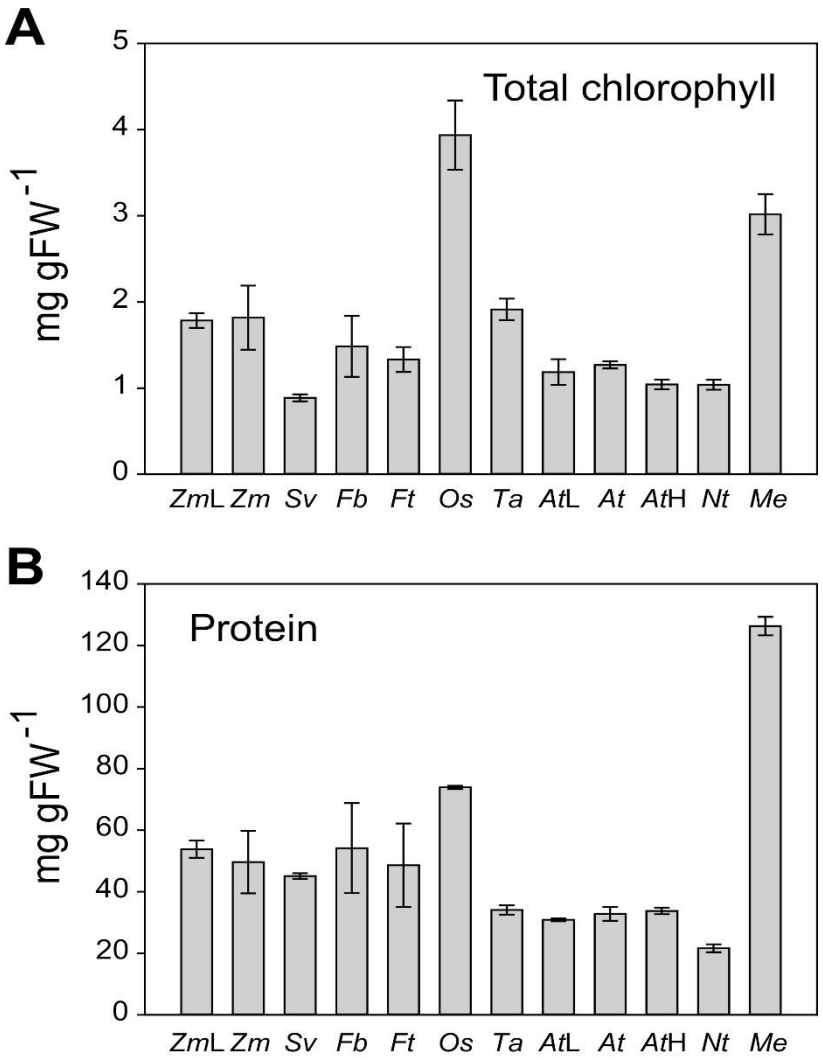


Fig. 3. Total chlorophyll (A) and protein (B) content in different species. Measurements were performed in *Z. mays* at low and medium irradiance (ZmL and Zm, respectively), *S. viridis* (Sv), *F. bidentis* (Fb), *F. trinervia* (Ft), *O. sativa* (Os), *T. aestivum* (Ta), and *A. thaliana* at low, medium, and high irradiance (AtL, At, and AtH, respectively), *N. tabacum* (Nt) and *M. esculenta* (Me). Growth and harvest conditions can be found in Supplementary Table S1. The results are shown as mean (mg g FW^{-1}) ±SD. The original data are presented in Supplementary Dataset S1.

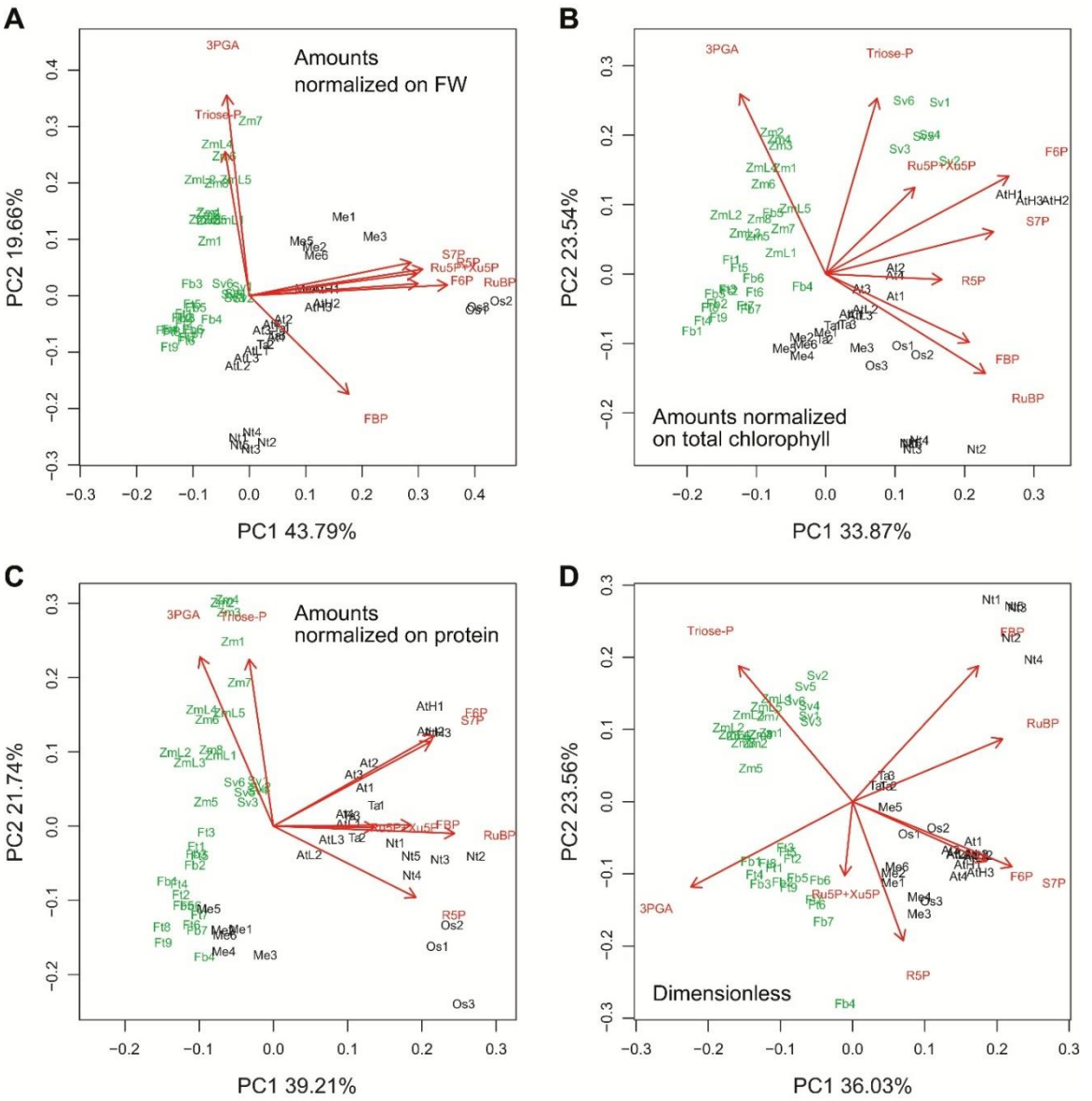


Fig. 4. Principal component analyses of the CBC metabolite profiles in all tested species. The analyses were performed on the metabolite data set excluding 2PG and SBP (2PG was omitted to avoid systematic bias between C_3 and C_4 species due the differing rates of RuBP oxygenation, and SBP was omitted because in some species part of the pool may not be involved in the CBC; see text for details). Metabolite amounts were normalized on (A) FW, (B) total chlorophyll content, or (C) protein content, or (D) were transformed into a dimensionless data set. For dimensionless data set determination, in a given sample, the level of each metabolite was first transformed to C equivalent values by multiplying the amount (nmol g FW^{-1}) by the number of C atoms in the metabolite. The C equivalent amounts of all CBC intermediates plus 2PG were then summed. In the last step, the C equivalent value of a given metabolite was divided by the summed C equivalent values. The transformed values and calculation steps are provided in Supplementary Dataset 1. This transformation generates a dimensionless data set (provided in Supplementary Dataset S1) in which each metabolite receives a value equal to its fractional contribution to all the C in CBC metabolites plus 2PG. As this data set is dimensionless, there is no systematic bias due to differences in leaf composition. The distribution of C_4 species (green) and C_3 species (black) is shown on PC1 and PC2 (*Z. mays*, Zm and ZmL; *S. viridis*, Sv; *F. bidentis*, Fb; *F. trinervia*, Ft; *O. sativa*, Os; *T. aestivum*, Ta; *A. thaliana*, AtL, At, and AtH; *N. tabacum*, Nt; *M. esculenta*, Me). The loadings of CBC intermediates in PC1 and PC2 are shown in red. Principal component analyses with the full metabolite data set and with all metabolites except either 2PG or SBP are shown in Supplementary Fig. S4 (amounts normalized on FW), Supplementary Fig. S5 (amounts normalized on total chlorophyll content), Supplementary Fig. S6 (amounts normalized on protein content), and Supplementary Fig. S7 (dimensionless). The original data are presented in Supplementary Dataset S1.

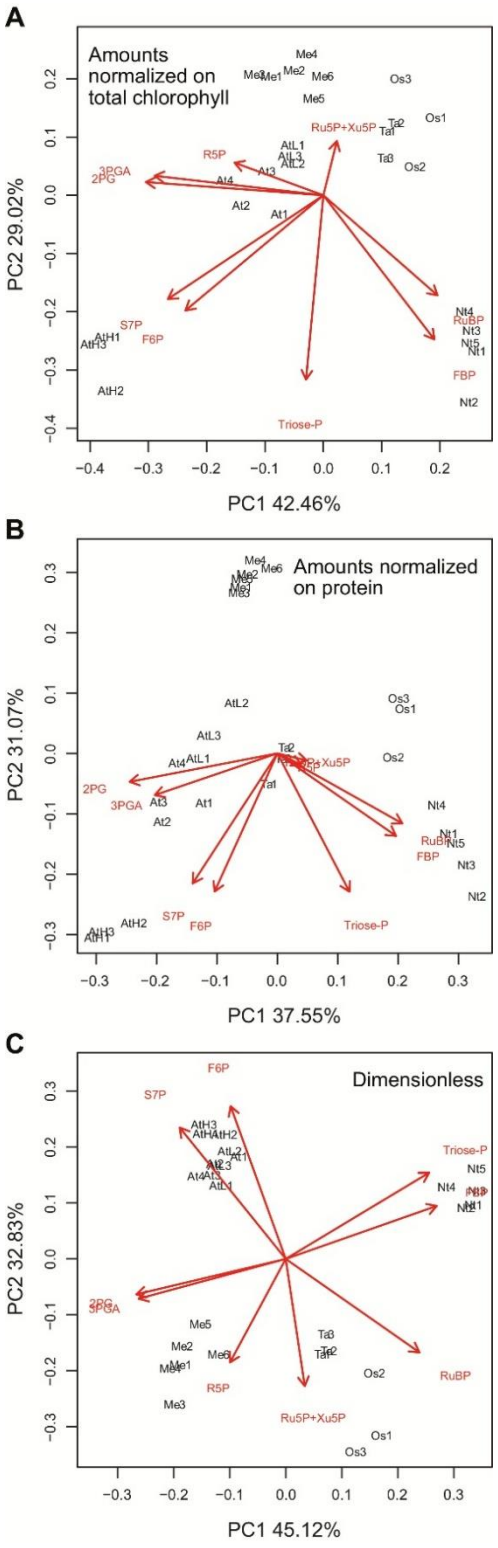


Fig. 5. Principal component analysis of the CBC metabolite contents in C_3 species only. The analyses were performed on the metabolite data set excluding SBP. Metabolite data were normalized on (A) total chlorophyll content and (B) total protein, and (C) using a dimensionless data set (see legend of Fig. 4). The distribution of C3 species is shown on PC1 and PC2 (*O. sativa*, Os; *T. aestivum*, Ta; *A. thaliana*, AtL, At, and AtH; *N. tabacum*, Nt; *M. esculenta*, Me). The loadings of CBC intermediates in PC1 and PC2 are shown in red. PC analyses with the full metabolite data set and with all metabolites except either 2PG or SBP are shown in Supplementary Fig. S8 (amounts normalized on total chlorophyll content), Supplementary Fig. S9 (amounts normalized on total protein), and Supplementary Fig. S10 (dimensionless). The original data are presented in Supplementary Dataset S1.

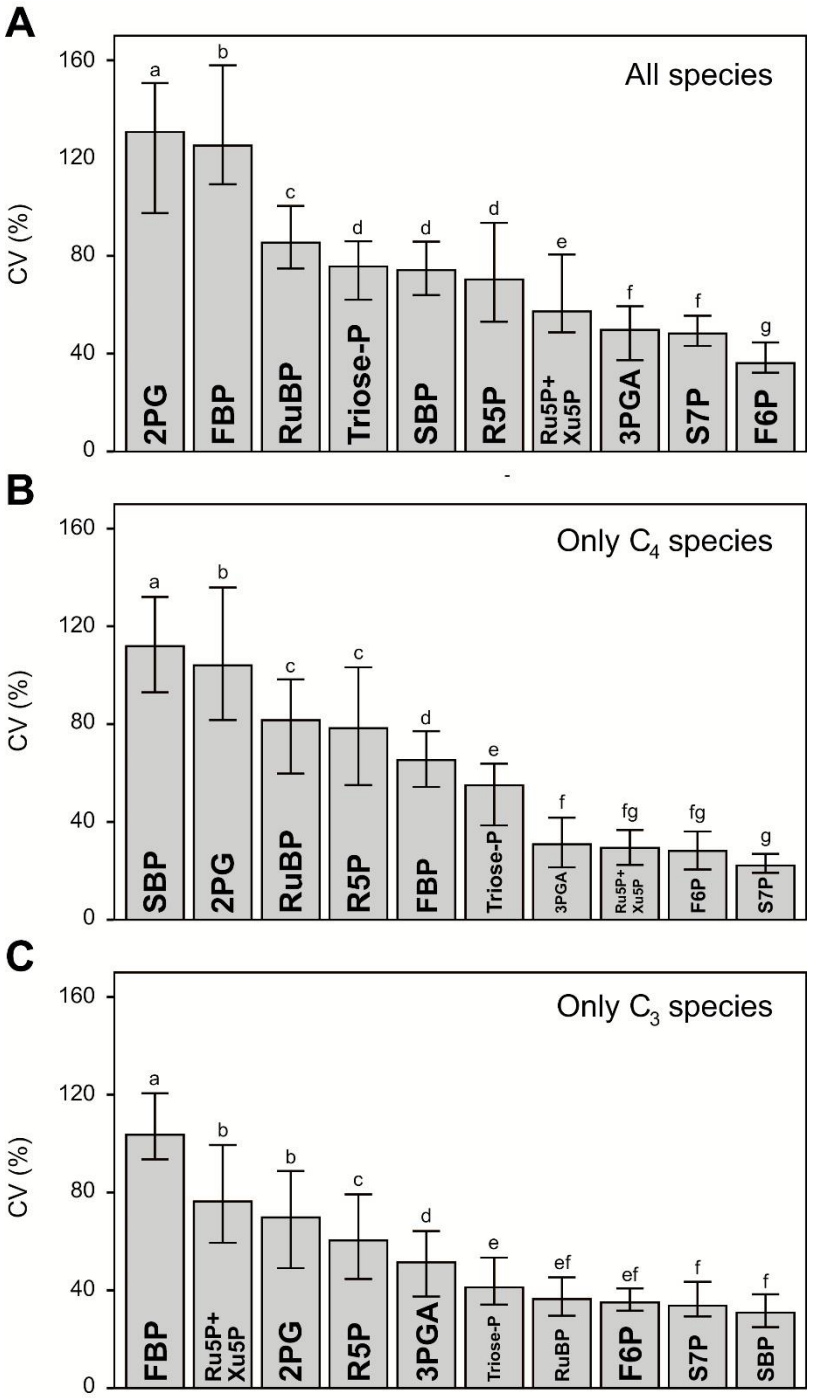


Fig. 6. Coefficient of variance (CV) of CBC metabolites between species. The CV (SD/mean×100) is a standardized quantity describing the dispersion of a population distribution (Simpson and Roe, 1939). The analysis was performed on a dimensionless data set that was generated as described for Fig. 4D. The transformed data (presented in Supplementary Dataset S1) were used to calculate the bootstrapped CV for each metabolite (30 bootstrap iterations). The 95% confidence interval was estimated using the basic bootstrap method. Statistically significant differences between metabolites are indicated by letters (ANOVA on the bootstrap results followed by the Tukey's HSD post-test). (A) All species, (B) only C_4 species, and (C) only C_3 species.

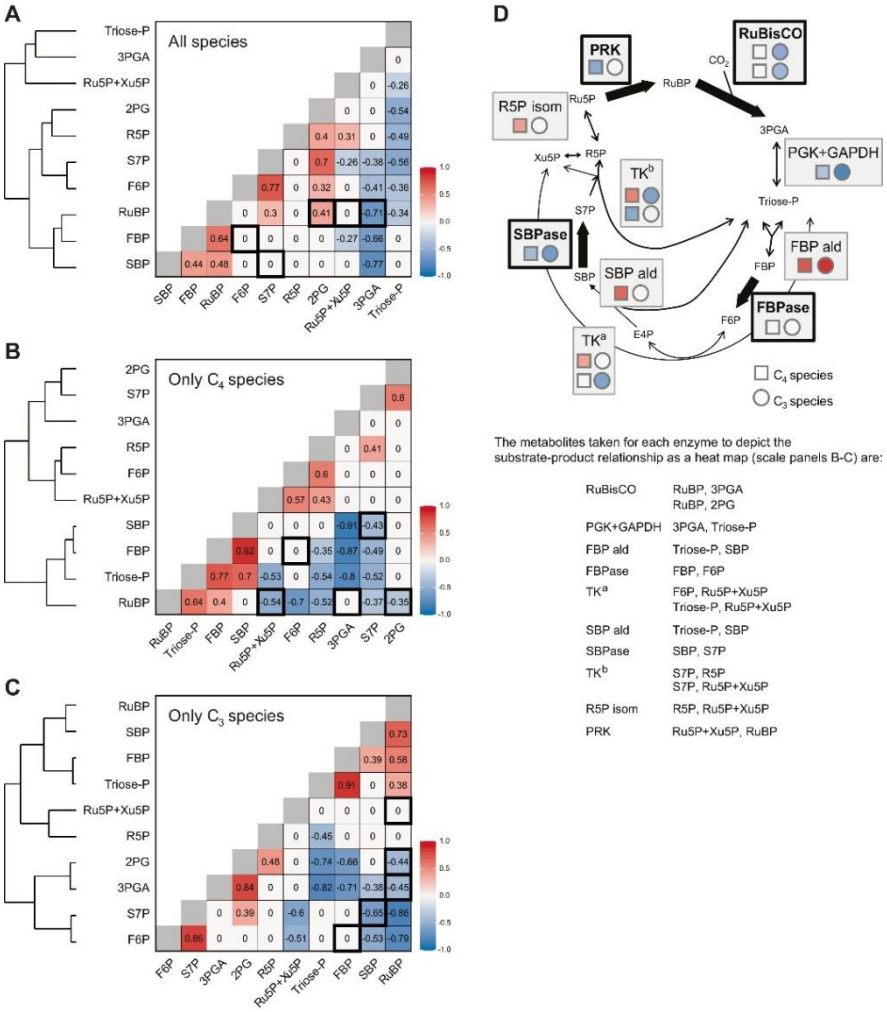


Fig. 7. Correlation between levels of CBC metabolites. The correlations were performed with all individual samples from a dimensionless data set, generated as described for Fig. 4D. The transformed data were used to calculate the Pearson's correlation matrix on every pair of metabolites. Correlation values are given in the figure panels and indicated by a heat map. The adjacent dendrograms show clusters defined using the complete linkage method (Sørensen, 1948). Non-significant correlations ($P \geq 0.05$; two-sided Student's t-test) are set as zero. Metabolite pairs that are linked by irreversible reactions are indicated by a black box. (A) All species, (B) only C_4 species, and (C) only C_3 species. An alternative display is provided in Supplementary Fig. S11, with the same fixed order of metabolites in each panel, corresponding to the reaction sequence in the CBC. The same heat map scale is used for (A–C). (D) Schematic representation of interspecies variance in the ratio of substrate abundance:product abundance for different CBC enzymes. Enzymes that catalyse irreversible reactions are highlighted in bold. For each enzyme reaction, the substrate and product that were compared are indicated in the list below the display. This display is schematic because some metabolites were not measured (erythrose 4-phosphate, E4P; and glyceraldehyde 3-phosphate, GAP) or were not separated (Ru5P and Xu5P). For reactions using GAP, it is assumed that GAP and dihydroxyacetone phosphate (DHAP) are in equilibrium. For transketolase (TK), two reactions were separated (termed TKa and TKb). For TKa, the reactant E4P was missing, and only the relationships between triose-P and F6P and Ru5P+Xu5P are shown. For Tkb, the plot focuses on the relationship between S7P and R5P or Ru5P+Xu5P. The display shows the alternative pairs of metabolites compared, with the upper and lower symbols in the display corresponding to the upper and lower pair in the list. A similar display mode is used for the carboxylation and oxygenation reactions of Rubisco. The correlation coefficients are taken from (B) and (C), using the same heat map scale. Results are shown separated for correlations between the four C_4 species (squares) and the five C_3 species (circles). The analysis is not shown for the combined C_4 plus C_3 species set because, in this case, some relationships are driven by differences between C_4 and C_3 species. Additional abbreviations: fructose 1,6-bisphosphate aldolase (FBP ald), phosphoglycerate kinase (PGK), ribose 5-phosphate isomerase (R5P isom), sedoheptulose 1,7-bisphosphate aldolase (SBP ald).

REFERENCES

Adam NR. 2017. C_3 carbon reduction cycle: eLS. Chichester: John Wiley & Sons, Ltd.

Arrivault S, Guenther M, Fry SC, Fuenfgeld MM, Veyel D, Mettler-Altmann T, Stitt M, Lunn JE. 2015. Synthesis and use of stable-isotope labeled internal standards for quantification of phosphorylated metabolites by LC-MS/MS. Analytical Chemistry 87, 6896–6904.

Arrivault S, Guenther M, Ivakov A, Feil R, Vosloh D, van Dongen JT, Sulpice R, Stitt M. 2009. Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. The Plant Journal 59, 824–839.

Arrivault S, Obata T, Szecówka M, Mengin V, Guenther M, Hoehne M, Fernie AR, Stitt M. 2017. Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation. Journal of Experimental Botany 68, 283–298.

Awoyinka AF, Abegunde VO, Adewusi SR. 1995. Nutrient content of young cassava leaves and assessment of their acceptance as a green vegetable in Nigeria. Plant Foods for Human Nutrition 47, 21–28.

Badger MR, Andrews TJ, Whitney SM, Ludwig M, Yellowlees DC, Leggat W, Price GD. 1998. The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. Canadian Journal of Botany 76, 1052–1071.

Bassham JA, Krause GH. 1969. Free energy changes and metabolic regulation in steady-state photosynthetic carbon reduction. Biochimica et Biophysica Acta 189, 207–221.

Betti M, Bauwe H, Busch FA, *et al*. 2016. Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement. Journal of Experimental Botany 67, 2977–2988.

Busch FA, Sage TL, Cousins AB, Sage RF. 2013. C_3 plants enhance rates of photosynthesis by reassimilating photorespired and respired CO_2. Plant, Cell & Environment 36, 200–212.

Carmo-Silva AE, Keys AJ, Andralojc PJ, Powers SJ, Arrabaça MC, Parry MA. 2010. Rubisco activities, properties, and regulation in three different C_4 grasses under drought. Journal of Experimental Botany 61, 2355–2366.

Charlet T, Moore BD, Seemann JR. 1997. Carboxyarabinitol 1-phosphate phosphatase from leaves of *Phaseolus vulgaris* and other species. Plant & Cell Physiology 38, 511–517.

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N. 2008. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Current Biology 18, 37–43.

Christin PA, Osborne CP. 2013. The recurrent assembly of C_4 photosynthesis, an evolutionary tale. Photosynthesis Research 117, 163–175.

Cock JH, Riaño NM, El-Sharkawy MA, Yamel LF, Bastidas G. 1987. C_3–C_4 intermediate photosynthetic characteristics of cassava (*Manihot esculenta* Crantz): II. Initial products of $^{14}CO_2$ fixation. Photosynthesis Research 12, 237–241.

Cruz JA, Emery C, Wüst M, Kramer DM, Lange BM. 2008. Metabolite profiling of Calvin cycle intermediates by HPLC-MS using mixed-mode stationary phases. The Plant Journal 55, 1047–1060.

De Souza AP, Long SP. 2018. Toward improving photosynthesis in cassava: characterizing photosynthetic limitations in four current African cultivars. Food and Energy Security 7, e00130.

De Souza AP, Massenburg LN, Jaiswal D, Cheng S, Shekar R, Long SP. 2017. Rooting for cassava: insights into photosynthesis and associated physiology as a route to improve yield potential. New Phytologist 213, 50–65.

Díaz S, Kattge J, Cornelissen JH, *et al*. 2016. The global spectrum of plant form and function. Nature 529, 167–171.

Ding F, Wang M, Zhang S, Ai X. 2016. Changes in SBPase activity influence photosynthetic capacity, growth, and tolerance to chilling stress in transgenic tomato plants. Scientific Reports 6, 32741.

Donovan LA, Maherali H, Caruso CM, Huber H, de Kroon H. 2011. The evolution of the worldwide leaf economics spectrum. Trends in Ecology & Evolution 26, 88–95.

Driever SM, Lawson T, Andralojc PJ, Raines CA, Parry MA. 2014. Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. Journal of Experimental Botany 65, 4959–4973.

Driever SM, Simkin AJ, Alotaibi S, *et al*. 2017. Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. Philosophical Transactions of the Royal Society B: Biological Sciences 372, 1730.

Edwards EJ, Osborne CP, Strömberg CA, *et al*. 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. Science 328, 587–591.

Edwards GE, Sheta E, Moore BD, Dai Z, Franceschi VR, Cheng SH, Lin C-H, Ku MSB. 1990. Photosynthetic characteristics of cassava (*Manihot esculenta* Crantz), a C_3 species with chlorenchymatous bundle sheath cells. Plant & Cell Physiology 31, 1199–1206.

El-Sharkawy MA, Cock JH. 1987. C3–C4 intermediate photosynthetic characteristics of cassava (*Manihot esculenta* Crantz): I. Gas exchange. Photosynthesis Research 12, 219 235.

Ellis RJ. 1979. The most abundant protein in the world. Trends in Biochemical Sciences 4, 241–244.

Evans JR. 1989. Photosynthesis and nitrogen relationships in leaves of C_3 plants. Oecologia 78, 9–19.

Farquhar GD, von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta 149, 78–90.

Field C, Mooney HA. 1986. The photosynthesis–nitrogen relationship in wild plants. In: Givnish T, ed. On the economy of plant form and function. Cambridge: Cambridge University Press, 25–55.

Galmés J, Andralojc PJ, Kapralov MV, Flexas J, Keys AJ, Molins A, Parry MA, Conesa MÀ. 2014*a*. Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C_3 genus Limonium (Plumbaginaceae). New Phytologist 203, 989–999.

Galmés J, Kapralov MV, Andralojc PJ, Conesa MÀ, Keys AJ, Parry MA, Flexas J. 2014*b*. Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. Plant, Cell & Environment 37, 1989–2001.

Gerhardt R, Stitt M, Heldt HW. 1987. Subcellular metabolite levels in spinach leaves: regulation of sucrose synthesis during diurnal alterations in photosynthetic partitioning. Plant Physiology 83, 399–407.

Ghannoum O, Evans JR, Chow WS, Andrews TJ, Conroy JP, von Caemmerer S. 2005. Faster Rubisco is the key to superior nitrogenase efficiency in NADP-malic enzyme relative to NAD-malic enzyme C_4 grasses. Plant Physiology 137, 638–650.

Gibon Y, Vigeolas H, Tiessen A, Geigenberger P, Stitt M. 2002. Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system. The Plant Journal 30, 221–235.

Giordano M, Beardall J, Raven JA. 2005. CO_2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. Annual Review of Plant Biology 56, 99–131.

Gontero B, Maberly SC. 2012. An intrinsically disordered protein, CP12: jack of all trades and master of the Calvin cycle. Biochemical Society Transactions 40, 995–999.

Hasunuma T, Harada K, Miyazawa S, Kondo A, Fukusaki E, Miyake C. 2010. Metabolic turnover analysis by a combination of *in vivo* 13C-labelling from $^{13}CO_2$ and metabolic profiling with CE-MS/MS reveals rate-limiting steps of the C_3 photosynthetic pathway in *Nicotiana tabacum* leaves. Journal of Experimental Botany 61, 1041–1051.

Hatch MD. 2002. C_4 photosynthesis: discovery and resolution. Photosynthesis Research 73, 251–256.

Hatch MD, Osmond CB. 1976. Compartmentation and transport in C_4 photosynthesis. In: Stocking CR, Heber U, eds. Encyclopedia of Plant Physiology, Vol. 3. Berlin: Springer-Verlag, 144–184.

Heckmann D. 2016. C_4 photosynthesis evolution: the conditional Mt Fuji. Current Opinion in Plant Biology 31, 149–154.

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber AP, Lercher MJ. 2013. Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell 153, 1579–1588.

Heldt HW. 2005. Plant biochemistry. Oxford: Elsevier Academic Press.

Hermida-Carrera C, Fares MA, Fernández Á, *et al*. 2017. Positively selected amino acid replacements within the RuBisCO enzyme of oak trees are associated with ecological adaptations. PLoS One 12, e0183970.

Hikosaka K. 2010. Mechanisms underlying interspecific variation in photosynthetic capacity across wild plant species. Plant Biotechnology 27, 223–229.

Höfer MU, Santore UJ, Westhoff P. 1992. Differential accumulation of the 10-, 16- and 23-kDa peripheral components of the water-splitting complex of photosystem II in mesophyll and bundle-sheath chloroplasts of the dicotyledonous C_4 plant *Flaveria trinervia* (Spreng.) C. Mohr. Planta 186, 304–312.

Howard TP, Lloyd JC, Raines CA. 2011. Inter-species variation in the oligomeric states of the higher plant Calvin cycle enzymes glyceraldehyde-3-phosphate dehydrogenase and phosphoribulokinase. Journal of Experimental Botany 62, 3799–3805.

Jordan DB, Ogren WL. 1981. Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. Nature 291, 513.

Kapralov MV, Kubien DS, Andersson I, Filatov DA. 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. Molecular Biology and Evolution 28, 1491–1503.

Kerfeld CA, Melnicki MR. 2016. Assembly, function and evolution of cyanobacterial carboxysomes. Current Opinion in Plant Biology 31, 66–75.

Kocacinar F, McKown AD, Sage TL, Sage RF. 2008. Photosynthetic pathway influences xylem structure and function in *Flaveria* (Asteraceae). Plant, Cell & Environment 31, 1363–1376.

Laetsch WM, Price I. 1969. Development of the dimorphic chloroplasts of sugar cane. American Journal of Botany 56, 77–87.

Lawson T, Kramer DM, Raines CA. 2012. Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. Current Opinion in Biotechnology 23, 215–220.

Leegood RC. 1985. The intercellular compartmentation of metabolites in leaves of *Zea mays* L. Planta 164, 163–171.

Leegood RC, von Caemmerer S. 1988. The relationship between contents of photosynthetic metabolites and the rate of photosynthetic carbon assimilation in leaves of *Amaranthus edulis* L. Planta 174, 253–262.

Leegood RC, von Caemmerer S. 1989. Some relationships between contents of photosynthetic intermediates and the rate of photosynthetic carbon assimilation in leaves of *Zea mays* L. Planta 178, 258–266.

Lefebvre S, Lawson T, Zakhleniuk OV, Lloyd JC, Raines CA, Fryer M. 2005. Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. Plant Physiology 138, 451–460.

Long SP. 1999. Environmental responses. In: Sage RF, Monson R, eds. C_4 plant biology. San Diego: Academic Press, 215–249.

Long SP, Ainsworth EA, Leakey AD, Nösberger J, Ort DR. 2006. Food for thought: lower-than-expected crop yield stimulation with rising CO_2 concentrations. Science 312, 1918–1921.

López-Calcagno PE, Howard TP, Raines CA. 2014. The CP12 protein family: a thioredoxin-mediated metabolic switch? Frontiers in Plant Science 5, 9.

Lorimer G. 1981. The carboxylation and oxygenation of ribulose 1,5-bisphosphate: the primary events in photosynthesis and photorespiration. Annual Review of Plant Physiology 32, 349–382.

Lorimer GH, Andrews TJ. 1973. Plant photorespiration—an inevitable consequence of the existence of atmospheric oxygen. Nature 243, 359.

Ma F, Jazmin LJ, Young JD, Allen DK. 2014. Isotopically nonstationary ^{13}C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. Proceedings of the National Academy of Sciences, USA 111, 16967–16972.

Mallmann J, Heckmann D, Brautigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. eLife 16, 02478.

McKown AD, Dengler NG. 2007. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). American Journal of Botany 94, 382–399.

Meister M, Agostino A, Hatch MD. 1996. The roles of malate and aspartate in C_4 photosynthetic metabolism of *Flaveria bidentis* (L.). Planta 199, 262–269.

Merlo L, Geigenberger P, Hajirezaei M, Stitt M. 1993. Changes of carbohydrates, metabolites and enzyme activities in potato tubers during development, and within a single tuber along a stolon–apex gradient. Plant Physiology 142, 392–402.

Mettler T, Mühlhaus T, Hemme D, *et al*. 2014. Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism *Chlamydomonas reinhardtii*. The Plant Cell 26, 2310–2350.

Meyer RC, Steinfath M, Lisec J, *et al*. 2007. The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences, USA 104, 4759–4764.

Moore BD, Isidoro E, Seemann JR. 1993. Distribution of 2-carboxyarabinitol among plants. Phytochemistry 34, 703–707.

Munekage NY. 2016. Light harvesting and chloroplast electron transport in NADP-malic enzyme type C4 plants. Current Opinion in Plant Biology 31, 9–15.

Nakamura N, Iwano M, Havaux M, Yokota A, Munekage YN. 2013. Promotion of cyclic electron transport around photosystem I during the evolution of NADP-malic enzyme-type C_4 photosynthesis in the genus *Flaveria*. New Phytologist 199, 832–842.

Nassar NMA, Marques AO. 2006. Cassava leaves as a source of protein. Journal of Food, Agriculture and Environment 4, 187–188.

Nelson T. 2011. The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C_4 leaves. Journal of Experimental Botany 62, 3039–3048.

Ort DR, Merchant SS, Alric J, *et al*. 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proceedings of the National Academy of Sciences, USA 112, 8529–8536.

Osmond CB. 1981. Photorespiration and photoinhibition: some implications for the energetics of photosynthesis. Biochimica et Biophysica Acta 639, 77–98.

Parry MA, Keys AJ, Madgwick PJ, Carmo-Silva AE, Andralojc PJ. 2008. Rubisco regulation: a role for inhibitors. Journal of Experimental Botany 59, 1569–1580.

Poorter H, Jagodzinski AM, Ruiz-Peinado R, Kuyah S, Luo Y, Oleksyn J, Usoltsev VA, Buckley TN, Reich PB, Sack L. 2015. How does biomass distribution change with size and differ among species? An analysis for 1200 plant species from five continents. New Phytologist 208, 736–749.

Prins A, Orr DJ, Andralojc PJ, Reynolds MP, Carmo-Silva E, Parry MA. 2016. Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. Journal of Experimental Botany 67, 1827–1838.

Raines CA. 2011. Increasing photosynthetic carbon assimilation in C_3 plants to improve crop yield: current and future strategies. Plant Physiology 155, 36–42.

Raines CA, Harrison EP, Ölçer H, Lloyd JC. 2000. Investigating the role of the thiol-regulated enzyme sedoheptulose-1,7-bisphosphatase in the control of photosynthesis. Physiologia Plantarum 110, 303–308.

Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR. 2008. Reassessing the first appearance of eukaryotes and cyanobacteria. Nature 455, 1101–1104.

Raven JA. 2013. Rubisco: still the most abundant protein of Earth? New Phytologist 198, 1–3.

Raven JA, Beardall J, Sánchez-Baracaldo P. 2017. The possible evolution and future of CO_2-concentrating mechanisms. Journal of Experimental Botany 68, 3701–3716.

Sage RF. 2017. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. Journal of Experimental Botany 68, 4039–4056.

Sage RF, Christin PA, Edwards EJ. 2011. The C_4 plant lineages of planet Earth. Journal of Experimental Botany 62, 3155–3169.

Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the evolution of C_4 photosynthesis. Annual Review of Plant Biology 63, 19–47.

Sage RF, Seemann JR. 1993. Regulation of ribulose-1,5-bisphosphate carboxylase/oxygenase activity in response to reduced light intensity in C_4 plants. Plant Physiology 102, 21–28.

Sage TL, Busch FA, Johnson DC, Friesen PC, Stinson CR, Stata M, Sultmanis S, Rahman BA, Rawsthorne S, Sage RF. 2013. Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. Plant Physiology 163, 1266–1276.

Sage TL, Sage RF. 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 Physiology 50, 756–772.

Savir Y, Noor E, Milo R, Tlusty T. 2010. Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. Proceedings of the National Academy of Sciences, USA 107, 3475–3480.

Servaites JC, Parry MA, Gutteridge S, Keys AJ. 1986. Species variation in the predawn inhibition of ribulose-1,5-bisphosphate carboxylase/oxygenase. Plant Physiology 82, 1161–1163.

Shameer S, Baghalian K, Cheung CYM, Ratcliffe RG, Sweetlove LJ. 2018. Computational analysis of the productivity potential of CAM. Nature Plants 4, 165–171.

Sharkey TD. 1988. Estimating the rate of photorespiration in leaves. Physiologia Plantarum 73, 147–152.

Sharwood RE, Ghannoum O, Kapralov MV, Gunn LH, Whitney SM. 2016*a*. Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C_3 photosynthesis. Nature Plants 2, 16186.

Sharwood RE, Ghannoum O, Whitney SM. 2016*b*. Prospects for improving CO_2 fixation in C_3-crops through understanding C_4-Rubisco biogenesis and catalytic diversity. Current Opinion in Plant Biology 31, 135–142.

Sharwood RE, Sonawane BV, Ghannoum O, Whitney SM. 2016*c*. Improved analysis of C_4 and C_3 photosynthesis via refined *in vitro* assays of their carbon fixation biochemistry. Journal of Experimental Botany 67, 3137–3148.

Silvera K, Neubig KM, Whitten WM, Williams NH, Winter K, Cushman JC. 2010. Evolution along the crassulacean acid metabolism continuum. Functional Plant Biology 37, 995–1010.

Simkin AJ, Lopez-Calcagno PE, Davey PA, Headland LR, Lawson T, Timm S, Bauwe H, Raines CA. 2017. Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphophate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO_2 assimilation, vegetative biomass and seed yield in Arabidopsis. Plant Biotechnology Journal 15, 805–816.

Simpson G, Roe A. 1939. Quantitative zoology, numerical concepts and methods in the study of recent and fossil animals. New York and London: McGraw-Hill Book Co.

Somerville CR. 2001. An early Arabidopsis demonstration. Resolving a few issues concerning photorespiration. Plant Physiology 125, 20–24.

Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biologiske Skrifter 5, 1–34.

Still CJ, Berry JA, Collatz GJ, DeFries RS. 2003. Global distribution of C_3 and C_4 vegetation: carbon cycle implications. Global Biogeochemical Cycles 17, 6-1–6-14.

Stitt M, Heldt HW. 1985. Generation and maintenance of concentration gradients between the mesophyll and bundle sheath in maize leaves. Biochimica et Biophysica Acta 808, 400–414.

Stitt M, Lunn J, Usadel B. 2010. Arabidopsis and primary photosynthetic metabolism—more than the icing on the cake. The Plant Journal 61, 1067–1091.

Stitt M, Sulpice R, Gibon Y, Whitwell A, Skilbeck R, Parker S, Ellison R. 2007. Cryogenic Grinder System. Vol. German Patent No. 08146.0025U1. MPG/SFX Link Resolver.

Sulpice R, Nikoloski Z, Tschoep H, *et al*. 2013. Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions. Plant Physiology 162, 347–363.

Sulpice R, Pyl E-T, Ishihara H, *et al*. 2009. Starch as a major integrator in the regulation of plant growth. Proceedings of the National Academy of Sciences, USA 106, 10348–10353.

Szecowka M, Heise R, Tohge T, *et al*. 2013. Metabolic fluxes in an illuminated Arabidopsis rosette. The Plant Cell 25, 694–714.

Tcherkez GG, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proceedings of the National Academy of Sciences, USA 103, 7246–7251.

Usuda H. 1987. Changes in levels of intermediates of the C_4 cycle and reductive pentose phosphate pathway under various light intensities in maize leaves. Plant Physiology 84, 549–554.

von Caemmerer S, Farquhar GD. 1981. Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. Planta 153, 376–387.

von Caemmerer S, Furbank RT. 2003. The C_4 pathway: an efficient CO_2 pump. Photosynthesis Research 77, 191–207.

Wright IJ, Reich PB, Westoby M, *et al*. 2004. The worldwide leaf economics spectrum. Nature 428, 821–827.

Wu S, Alseekh S, Cuadros-Inostroza Á, Fusari CM, Mutwil M, Kooke R, Keurentjes JB, Fernie AR, Willmitzer L, Brotman Y. 2016. Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*. PLoS Genetics 12, e1006363.

Wullschleger SD. 1993. Biochemical limitations to carbon assimilation in C_3 plants—a retrospective analysis of the A/C_i curves from 109 species. Journal of Experimental Botany 44, 907–920.

Yeoh HH, Badger MR, Watson L. 1980. Variations in k_m (CO_2) of ribulose-1,5-bisphosphate carboxylase among grasses. Plant Physiology 66, 1110–1112.

Zachos JC, Dickens GR, Zeebe RE. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature 451, 279–283.

Zhang N, Gibon Y, Wallace JG, *et al*. 2015. Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiology 168, 575–583

Zhu XG, de Sturler E, Long SP. 2007. Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. Plant Physiology 145, 513–526.

Manuscript two

Relationship between irradiance and levels of Calvin–Benson cycle and other intermediates in the model eudicot Arabidopsis and the model monocot rice

RELATIONSHIP BETWEEN IRRADIANCE AND LEVELS OF CALVIN–BENSON CYCLE AND OTHER INTERMEDIATES IN THE MODEL EUDICOT ARABIDOPSIS AND THE MODEL MONOCOT RICE

Gian Luca Borghi, Thiago Alexandre Moraes, Manuela Günther, Regina Feil, Virginie Mengin, John E. Lunn, Mark Stitt and Stéphanie Arrivault*

Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany

* Correspondence: arrivault@mpimp-golm.mpg.de

Received 16 April 2019; Editorial decision 15 July 2019; Accepted 22 July 2019

Editor: Christine Raines, University of Essex, UK

ABSTRACT

Metabolite profiles provide a top-down overview of the balance between the reactions in a pathway. We compared Calvin–Benson cycle (CBC) intermediate profiles in different conditions in Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*) to learn which features of CBC regulation differ and which are shared between these model eudicot and monocot C_3 species. Principal component analysis revealed that CBC intermediate profiles follow different trajectories in Arabidopsis and rice as irradiance increases. The balance between subprocesses or reactions differed, with 3-phosphoglycerate reduction being favoured in Arabidopsis and ribulose 1,5-bisphosphate regeneration in rice, and sedoheptulose-1,7-bisphosphatase being favoured in Arabidopsis compared with fructose-1,6-bisphosphatase in rice. Photosynthesis rates rose in parallel with ribulose 1,5-bisphosphate levels in Arabidopsis, but not in rice. Nevertheless, some responses were shared between Arabidopsis and rice. Fructose 1,6-bisphosphate and sedoheptulose-1,7-bisphosphate were high or peaked at very low irradiance in both species. Incomplete activation of fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase may prevent wasteful futile cycles in low irradiance. End-product synthesis is inhibited and high levels of CBC intermediates are maintained in low light or in low CO_2 in both species. This may improve photosynthetic efficiency in fluctuating irradiance, and facilitate rapid CBC flux to support photorespiration and energy dissipation in low CO_2.

INTRODUCTION

Engineering photosynthesis is a promising route to improve crop yield (Long *et al.*, 2015; Ort *et al.*, 2015). In addition to the light reactions (Kromdijk *et al.*, 2016) and photorespiration (Betti *et al.*, 2016; South *et al.*, 2019), there is increasing interest in optimizing the Calvin–Benson cycle (CBC) (Zhu *et al.*, 2007; Driever *et al.*, 2017; Simkin *et al.*, 2017). Comprehensive information about metabolite levels will further these efforts by providing insights into the operation and regulation of the CBC, and by supporting the development and validation of metabolic models. The need for such information is emphasized by mounting evidence for interspecies variance in C_3 photosynthesis (Parry *et al.*, 2008; Sage and Sage, 2009; Lawson *et al.*, 2012; Galmés *et al.*, 2014; Prins *et al.*, 2016; Arrivault *et al.*, 2019).

The CBC consists of three subprocesses; fixation of CO_2 by Rubisco to form two molecules of 3-phosphoglycerate (3PGA), reduction of 3PGA to triose phosphate (triose-P) using NADPH and ATP from the light reactions, and regeneration of ribulose 1,5-bisphosphate (RuBP) from triose-P (Heldt *et al.*, 2005). Rubisco and three enzymes involved in RuBP regeneration [fructose-1,6-bisphosphatase (FBPase), sedoheptulose-1,7-bisphosphatase (SBPase), and phosphoribulokinase (PRK)] catalyse irreversible reactions, whilst the other reactions are near equilibrium (Bassham and Krause, 1969; Stitt *et al.*, 1980; Dietz and Heber, 1984; Mettler *et al.*, 2014).

The CBC is regulated for two main reasons. First, to balance flux at the various enzymatic reactions and maintain metabolite levels in a range that allows optimal CBC operation (Lawson *et al.*, 2012). Regulation must be rapid because environmental conditions, especially irradiance, can change rapidly, CBC fluxes are very high and metabolite pools are small, with half-times of <1 s (Stitt *et al.*, 1980; Arrivault *et al.*, 2009; Mettler *et al.*, 2014). Secondly, unless they are inhibited by the order of 1000-fold, several CBC enzymes would catalyse wasteful futile cycles in the dark (Heldt *et al.*, 2005).

Pioneering experimental and theoretical analyses emphasized that photosynthesis is restricted by the energy supply in light-limiting conditions and by Rubisco in light-saturating conditions (Farquhar *et al.*, 1980; von Caemmerer and Farquhar, 1981). In some conditions, such as low temperature, photosynthesis can be restricted by the rate of Pi recycling during end-product synthesis (Sharkey, 1985a, b; McClain and Sharkey, 2019). Subsequent work showed that photosynthesis can also be restricted by enzymes that convert triose-P to RuBP (Stitt and Sonnewald, 1995; Stitt, 1999; Stitt *et al.*, 2010)

including SBPase (Raines *et al.*, 2000; Lefebvre *et al.*, 2005; Zhu *et al.*, 2007; Driever *et al.*, 2017) and aldolase (Haake et al., 1998, 1999; Simkin *et al.*, 2017). The distribution of control in the CBC depends on pre-history and current conditions (Stitt and Schulze, 1994; Stitt *et al.*, 2010), implying that optimal operation will require coordinated regulation at multiple sites.

A hierarchy of regulation mechanisms operating in different time scales and often affecting multiple enzymes regulates the CBC. Mass action rapidly propagates changes in metabolite levels through sequences of reactions that are close to thermodynamic equilibrium. For example, when a change in the supply of ATP or NADPH from the light reactions alters the 3PGA:triose-P ratio (Dietz and Heber, 1984), reversible reactions catalysed by triose phosphate isomerase, aldolase, transketolase, and phosphoriboseisomerase propagate the change in triose-P to fructose 1,6-bisphosphate (FBP), erythrose 4-phosphate (E4P), sedoheptulose 1,7-bisphosphate (SBP), sedoheptulose 7-phosphate (S7P), ribose 5-phosphate (R5P), xylulose 5-phosphate (Xu5P), and ribulose 5-phosphate (Ru5P). Product or allosteric regulation also mediate rapid responses (Stitt *et al.*, 2010); for example, feedback inhibition of FBPase by fructose 6-phosphate (F6P) (Gardemann *et al.*, 1986) and SBPase by S7P (Schimkat *et al.*, 1990), and inhibition of PRK by RuBP, ADP, and 3PGA (Gardemann *et al.*, 1983). The latter serves to prevent excess ATP consumption at the irreversible PRK reaction. Another fairly rapid response involves light-driven increases in stromal pH and Mg^{2+}, which activate many CBC enzymes (Heldt *et al.*, 2005). Slightly slower responses are facilitated by post-translational regulation, in particular thioredoxin-mediated redox regulation (Buchanan and Balmer, 2005). Thioredoxins activate many CBC enzymes including NADP-glyceraldehyde-3-phosphate dehydrogenase (NADP-GAPDH), FBPase, SBPase, and PRK, as well as Rubisco activase (Laing *et al.*, 1981; Scheibe, 1991; Buchanan and Balmer, 2005) and CP12, a small protein that interacts with and modulates NADP-GAPDH and PRK (Gontero and Maberly, 2012; Lopez-Calcagno et al., 2014). Rubisco is inhibited by binding of RuBP to the non-carbamylated form (Sharwood *et al.*, 2016), by low molecular weight inhibitors that derive from catalytic infidelities of Rubisco, and by inhibitors such as 2-carboxy-D-arabitinol 1-phosphate that are synthesized by dedicated pathways (Yeoh *et al.*, 1980; Moore *et al.*, 1993; Parry *et al.*, 2008). Release of these inhibitors and re-establishment of the active carbamylated form of Rubisco require dedicated enzymes such as Rubisco activase (Portis and Parry, 2007; Parry *et al.*, 2008; Portis *et al.*, 2008).

The CBC is autocatalytic, and pathway stoichiometry requires that, at steady state, five-sixths of the triose-P is retained in the CBC to regenerate RuBP. The remaining one-sixth is available for end-product synthesis. If triose-Ps are withdrawn too quickly, there will be a shortfall in RuBP, and if triose-Ps are withdrawn too slowly there will be a shortfall in inorganic phosphate (Pi) for ATP synthesis (Stitt, 1990; Stitt et al., 1987, 2010). CBC operation therefore depends on tight regulation of end-product synthesis. Cytosolic FBPase catalyses the first dedicated step in the conversion of triose-P to sucrose. This enzyme has strongly cooperative kinetics and is inhibited by the signal metabolite fructose 2,6-bisphosphate, whose level is regulated by 3PGA and Pi (Stitt, 1990). This regulatory network facilitates a threshold response, in which use of triose-P is inhibited until a threshold concentration is reached, and is strongly activated by further small increases in triose-P (Herzog *et al.*, 1984; Stitt, 1990; Stitt *et al.*, 1987). Starch synthesis is regulated via allosteric activation of ADP glucose pyrophosphorylase by a rising 3PGA/Pi ratio (Ballicora *et al.*, 2004).

In addition to carboxylation of RuBP, Rubisco catalyses a side reaction with O_2 (Lorimer, 1981) leading to formation of 2-phosphoglycolate (2PG) (Osmond, 1981; Sharkey, 1988; Betti *et al.*, 2016; Hagemann *et al.*, 2016). 2PG is salvaged via the photorespiratory pathway, in which two 2PG molecules are recycled to one 3PGA molecule with concomitant loss of CO_2. In current atmospheric conditions, about every fourth reaction of Rubisco leads to formation of 2PG, and this rises even further when stomata close during stress. Regulation is needed to maintain CBC flux and allow energy dissipation via photorespiration (Long *et al.*, 2006; Arrivault *et al.*, 2009; Eisenhut *et al.*, 2017). For example, non-productive consumption of RuBP in the Rubisco oxygenase reaction means that more than five-sixths of the triose-P must be retained in the CBC, and less than one-sixth is available for end-product synthesis. It is also becoming clear that photorespiratory intermediates influence CBC enzymes (Schimkat *et al.*, 1990; Flügel *et al.*, 2017)

It is not easy to predict how this hierarchy of regulation mechanisms interacts *in vivo*. Many CBC enzymes have highly cooperative kinetics and are influenced by multiple interacting effectors. For example, redox regulation often affects substrate affinity (see, for example, Gardemann *et al.*, 1986), and substrate levels affect the midpoint redox potential of enzymes and hence the extent to which they are activated by thioredoxin (Ashton and Hatch, 1983; Scheibe, 1991; Ashton *et al.*, 2000). These interactions allow >1000-fold changes in activity between light and darkness (Laing *et al.*, 1981; Gardemann *et al.*, 1983; Heldt *et al.*, 2005). It can be envisaged that they facilitate

coordinated regulation at multiple sites in the light, with fine modulation of enzymes allowing further fine tuning of flux in the CBC (Scheibe, 1991; Stitt, 1987; Fridlyand and Scheibe, 1999; Stitt *et al.*, 2010). However, better understanding requires information about fluxes and metabolite levels in different conditions.

Full understanding of a metabolic network requires integration of knowledge into predictive models. Many CBC models are available with varying levels of complexity (see Arnold and Nikoloski, 2011, 2012). They range from simple but powerful abstractions based on Rubisco kinetics and pathway topology (Farquhar *et al.*, 1980; von Caemmerer and Farquhar, 1981) to complex models utilizing detailed information about enzyme kinetic properties (Pettersson and Ryde-Pettersson, 1988; Fridlyand *et al.*, 1999; Poolman, 2000; Laisk *et al.*, 2006; Zhu *et al.*, 2007). These models are only approximations due to our incomplete knowledge of the regulatory properties of enzymes, complex interactions between different mechanisms (see above), and subcellular compartmentation (Stitt *et al.*, 1980; Gerhardt *et al.*, 1987; Heldt *et al.*, 2005). Computationally tractable models are likely to always require simplification, and their testing and validation require detailed information about metabolite levels in different conditions.

Information about the levels of CBC intermediates is surprisingly sparse. There were several studies of CBC intermediates levels under different conditions in the 1980s (Stitt et al., 1982a, 1983, 1984a, b; Badger *et al.*, 1984; Dietz and Heber, 1984; Seemann and Sharkey, 1986, 1987; von Caemmerer and Edmondson, 1986; Sharkey and Seemann, 1989; Servaites *et al.*, 1989; summarized in Supplementary Table S1 at *JXB* online; for details, see Discussion). However, for technical and partly conceptual reasons, these early studies focused on RuBP and a handful of other intermediates (3PGA, triose-P, FBP, and F6P). In the last 10 years, new analytical platforms have been developed that combine chromatographic separation with tandem MS (LC-MS/MS), allowing near-comprehensive analysis of CBC intermediates (Cruz *et al.*, 2008; Arrivault *et al.*, 2009; Hasunuma *et al.* 2010; Ma *et al.*, 2014). Arnold and Nikoloski (2012) found poor agreement between the metabolite levels predicted by various CBC models and those measured in Arabidopsis under ambient CO_2 and limiting irradiance (Arrivault *et al.*, 2009). This underlines the need for a compendium of metabolite data sets. In particular, these new methods have not been systematically used to profile metabolites in different conditions, except for Arabidopsis at different CO_2 concentrations (Arrivault *et al.*, 2009)

and the model alga *Chlamydomonas reinhardtii* at different light intensities (Mettler *et al.*, 2014).

It is also important to learn whether CBC operation is conserved across species. Since it originated ~2 billion years ago, the CBC has been subject to continued and powerful selection pressure because of changing atmospheric CO_2 and O_2 levels, as well as variation in temperature and the water and nutrient supply (Raven *et al.*, 2017; Sage, 2017). There is considerable diversity in the rate of photosynthesis in C_3 species (Evans, 1989; Wullschleger, 1993), and mounting evidence for diversity in Rubisco kinetics (Galmés *et al.*, 2014; Prins *et al.*, 2016), how Rubisco is regulated (Parry *et al.*, 2008), leaf anatomy, chloroplast location and density (Sage and Sage, 2009; Busch *et al.*, 2013), and stomatal responses (Lawson *et al.*, 2012) in C_3 species. Metabolite profiling provides an unbiased strategy to search for interspecies variance; changes in the balance between different enzymatic steps will lead to changes in the relative levels of pathway intermediates, irrespective of whether the underlying cause is changes in gene expression and protein abundance, enzyme kinetics, or regulatory networks. Arrivault *et al.* (2019) recently reported variation in CBC intermediate profiles between five C_3 species, including Arabidopsis and rice. However, a major challenge of interspecies studies is to establish whether the observed differences are due to species differences, environmental conditions, or complex genotype×environment interactions.

In this study, we have analysed the levels of CBC and related intermediates in Arabidopsis and rice at several light intensities, and in rice at two CO_2 concentrations. In combination with previously published data for the response to CO_2 in Arabidopsis (Arrivault *et al.*, 2009), these data allow us to compare the response of the CBC to irradiance and CO_2 in a model eudicot and a model monocot C_3 species. The results confirm that there is interspecies variance in CBC operation. They also uncover conserved responses that may minimize energy loss in low irradiance and allow CBC flux to be maintained when photorespiration is rapid.

MATERIALS AND METHODS

Chemicals

N_2, O_2, and CO_2 were obtained from Air Liquide (Germany; https://industrie.airliquide.de), and chemicals from Sigma-Aldrich (Darmstadt, Germany; https://www.sigmaaldrich.com), Roche Applied Science (Mannheim, Germany; https://lifescience.roche.com), or Merck (www.merckmillipore.com).

Plant growth and harvest

Arabidopsis (*Arabidopsis thaliana* Col-0) was grown with an irradiance of 120 µmol m^{-2} s^{-1} until 35 d after sowing as in Arrivault *et al.* (2019). On the day of harvest, plants were illuminated at growth irradiance for 4–5 h and then were subjected to darkness, or illuminated at 20, 80, 280, or 440 µmol m^{-2} s^{-1} for 15 min, or left at 120 µmol m^{-2} s^{-1}, before harvesting by cutting rosettes and quenching them instantaneously in a bath of liquid N_2 under the prevailing irradiance. Three rosettes were pooled per sample.

Rice (*Oryza sativa* ssp. *indica* cv. IR64) was grown at an irradiance of 350 µmol m^{-2} s^{-1} until 45–47 d after sowing as in Arrivault *et al.* (2019). Before harvest, plants were illuminated at growth irradiance for 4–8 h, adapted for 10 min under a mobile LED light array with dimmable LED lights (RHENAC Greentec AG, Hennef, Germany, www.rhenac-greentec.de) that were tuned to mimic the light spectrum within the growth chamber at 350 µmol m^{-2} s^{-1}, and then illuminated at 50, 150, 250, 350, 450, 550, 650, or 730 µmol m^{-2} s^{-1} for 20 min. The first fully expanded leaf blades (two or three per sample) were harvested without shading by freeze clamping between two aluminium blocks that had been pre-cooled in liquid N_2. Leaves were also harvested in darkness. In experiments where CO_2 concentration was changed, 60-day-old rice plants were placed under the LED light array (350 µmol m^{-2} s^{-1}) for 10 min and the first fully expanded leaf blades (two to three per sample) were then enclosed in a transparent gas-tight Plexiglas chamber (volume 68 ml), provided with 78% N_2, 21% O_2, and 500 ppm (ambient level) or 200 ppm CO_2 (5 l min^{-1}) for 20 min, and quenched without shading by opening a small inlet and outlet and pouring liquid N_2 into the chamber.

Metabolite analyses

Plant material was ground to a fine powder in a mortar pre-cooled with liquid N_2 or in a cryo-robot (Stitt *et al.*, 2007) and stored at –80 °C. Metabolites were extracted and quantified by LC-MS/MS using a reverse phase (Arrivault *et al.*, 2009; all CBC metabolites, ADP-glucose (ADPG), UDP glucose (UDPG), glucose 1-phosphate (G1P), glucose 6-phosphate (G6P)], or an anion exchange (Lunn *et al.*, 2006; sucrose 6'-phosphate (Suc6P)] column. Samples were spiked with stable isotope-labelled standards to correct for ion suppression and other matrix effects (Arrivault *et al.*, 2015). 3PGA, phospho*enol*pyruvate (PEP), and pyruvate were measured enzymatically in fresh trichloroacetic acid extracts (Jelitto *et al.*, 1992; Merlo *et al.*, 1993). Chl *a* and *b* were extracted and quantified as in Gibon *et al.*, (2002) and protein as in Arrivault *et al.* (2019).

Gas exchange

CO_2 assimilation was measured using whole rosettes of 5-week-old Arabidopsis or the mid-section of the first fully expanded leaves of 55- to 58-day-old rice plants using an open-flow infrared gas exchange analyser system (LI-6400XT, LI-COR Inc., Lincoln, NE, USA; www.licor.com) equipped with an integrated fluorescence chamber head (LI-6400-17 whole-plant chamber for Arabidopsis; LI-6400-40, 2 cm^2 leaf chamber for rice). CO_2 was 400 µmol mol^{-1}, leaf temperature 20 °C for Arabidopsis and 30 °C for rice, and relative humidity 65–75%.

Statistical analyses

Statistical analysis was performed in R Studio Version 1.1.463 (www.rstudio.com) with R version 3.5.1 (https://cran.r-project.org/) (details in figure legends and Supplementary Dataset S1).

RESULTS

Light saturation response

Arabidopsis and rice were grown under limiting irradiance (120 µmol m^{-2} s^{-1} and 350 µmol m^{-2} s^{-1}, respectively). Photosynthesis saturated at ~500 µmol m^{-2} s^{-1} irradiance with a light compensation point of 20±7 µmol m^{-2} s^{-1} in Arabidopsis, and ~2000 µmol m^{-2} s^{-1} with a light compensation point of 87±20 µmol m^{-2} s^{-1} in rice (Supplementary

Fig. S1). The maximum rate of photosynthesis on a fresh weight basis was ~8-fold higher in rice than in Arabidopsis (1250 µmol CO_2 g^{-1} FW h^{-1} and 145 µmol CO_2 g^{-1} FW h^{-1}, respectively). This partly reflects the higher protein and chlorophyll content in rice (2.95±0.32 mg chlorophyll g^{-1} FW, 71.6±12.2 mg protein g^{-1} FW) compared with Arabidopsis (1.14±0.12 mg chlorophyll g^{-1} FW, 32.2±2.6 mg protein g^{-1} FW). A relatively high light compensation point, light saturation response, and maximum rate of photosynthesis were previously reported for rice (Baker *et al.*, 1990; Dingkuhn *et al.*, 1990; Murchie *et al.*, 1999; Sage and Sage, 2009).

Harvest and metabolite analyses

Arabidopsis was pre-illuminated at growth irradiance and then transferred for 15 min to a new irradiance (20, 80, 120, 280, and 440 µmol m^{-2} s^{-1}) before harvesting whole rosettes. The chosen irradiances correspond, respectively, to the compensation point and ~67, 100, 230, and 360% of the irradiance required to half-saturate photosynthesis (Supplementary Fig. S1A). Rice was pre-illuminated at growth irradiance and then transferred for 20 min to a new irradiance (50, 150, 250, 350, 450, 550, 650, and 730, µmol m^{-2} s^{-1}) before harvesting the first fully expanded leaf. The chosen irradiances correspond to the compensation point and ~30, 60, 80, 100, 120, 150, and 170% of the irradiance that half-saturates photosynthesis (Supplementary Fig. S1B). The highest irradiance used for Arabidopsis and rice was ~90% and 75%, respectively, of that required to saturate photosynthesis. Metabolism was quenched instantaneously in Arabidopsis by submerging the rosettes in liquid N_2 and in rice by clamping the leaf between two aluminium blocks pre-cooled in liquid N_2. Quenching was performed at the prevailing irradiance. Samples were also collected in the dark.

Most intermediates of the CBC were quantified, including 3PGA, dihydroxyacetone phosphate (DHAP, also referred to as triose-P; glyceraldehyde 3-phosphate levels were below the detection limit and are probably 10–20 times lower than DHAP, due to the equilibrium constant of triose phosphate isomerase), FBP, F6P, SBP, S7P, R5P, and RuBP, as well as the photorespiratory intermediate 2PG. Ru5P and Xu5P were not resolved by LC-MS/MS and so were quantified together (termed ‘Ru5P+Xu5P’). Several intermediates of end-product synthesis were quantified including the starch synthesis intermediate ADPG, the sucrose synthesis intermediates G6P, G1P, UDPG, and Suc6P, as well as PEP and pyruvate. We assigned F6P to the CBC and G6P and G1P to sucrose

synthesis because, due to plastid phosphoglucoisomerase being removed from equilibrium in the light, F6P is mainly located in the plastid and G6P and G1P are mainly in the cytosol (Dietz and Heber, 1984; Gerhardt and Heldt, 1984; Gerhardt *et al.*, 1987; Szecowka *et al.*, 2013). Metabolite levels normalized on fresh weight and chlorophyll are provided in Supplementary Dataset S1. In addition, in a given sample, the amount of C within a given metabolite was divided by the total amount of C in all CBC intermediates plus 2PG. This transformation (termed 'dimensionless' or 'normalized on total C') removes effects due to a general change in metabolite levels and emphasizes changes in relative metabolite levels.

Principal component analysis of CBC intermediates

We first analysed the response of CBC intermediates and 2PG, starting with principal component (PC) analysis. PC analysis is a dimension-reduction technique that gives information about which samples (here, different species or irradiance) are closely related or separated, and which variables (here, metabolites) contribute to this relationship. The analysis was performed with *z*-scored data. The plots show the individual samples, and the mean value and 95% confidence limits for each species–irradiance treatment.

PC analysis on chlorophyll-normalized data (Fig. 1A) revealed a progressive shift of the metabolite profile in both species as irradiance increased. This was captured mainly in PC1, which represented 57% of total variance. Separation in PC1 was driven by almost all of the CBC intermediates, indicating that it is driven by a general increase in metabolite levels (see below for more data). Strikingly, Arabidopsis and rice showed different trajectories. This was captured in PC2, which accounted for 18% of total variance. Separation in PC2 was driven by FBP, F6P, and DHAP (higher in Arabidopsis; Fig. 2A–C), and SBP, Ru5P+Xu5P, 3PGA, and RuBP (higher in rice; Fig. 2A, B, D).

When PC analysis was performed on the dimensionless data sets (Fig. 1B), the response to increasing irradiance was weakened, especially for Arabidopsis. This normalization removes the effect of any general increase in metabolite levels (see above) allowing changes in relative levels to be seen more clearly. Arabidopsis and rice again showed different trajectories. This was captured in PC1, which represented 35% of the variance.

The analysis with the dimensionless data sets also highlighted that the lowest irradiance (20 µmol m^{-2} s^{-1} and 50 µmol m^{-2} s^{-1} in Arabidopsis and rice, respectively; in both cases, close to the light compensation point) separated strongly from other irradiances. This

separation was mainly in PC2, which represented 23% of total variance, and was driven by high SBP and FBP (Fig. 2B). The low irradiance samples were also outliers when PC analysis was performed on the Arabidopsis samples only (Supplementary Fig. S3) or the rice samples only (Supplementary Fig. S4). The response was seen for both the chlorophyll-normalized and dimensionless data sets, and was driven by high SBP and FBP.

Responses of individual CBC intermediates

Responses of individual metabolites are shown in Fig. 2, using a shared *y*-axis scale to allow direct comparison of qualitative trends and absolute levels in the two species. To visualize the balance between different reactions, selected metabolite ratios were calculated (Fig. 3). The combined data set was analysed by one-way ANOVA to detect interspecies differences and differences within a species as irradiance increased (see Supplementary Dataset S1 for details). Responses to increasing irradiance in the individual species are highlighted in Supplementary Figs S2A and S5, which provide displays of metabolite levels and ratios using scales that optimally visualize the response in a given species and the results of one-way ANOVA performed separately for each species.

Visual inspection reveals some conserved responses between Arabidopsis and rice. First, many metabolites showed a progressive and significant increase as irradiance is increased, including 3PGA (except at the highest irradiance), DHAP, F6P, S7P, and 2PG in both species, and R5P, Ru5P+Xu5P, and RuBP across all light intensities in Arabidopsis and in the lower irradiance range in rice (Fig. 2; see also Supplementary Fig. S2A). Secondly, SBP and FBP rose to high levels at the lowest irradiance and then declined (Fig. 2). This response is especially marked in rice but is also clear for SBP in Arabidopsis. In both species the FBP/F6P and SBP/S7P ratios peaked at the lowest irradiance and then declined (Fig. 3). The peak of the SBP/S7P ratio at low irradiance was significant in both species in ANOVA performed on the combined data sets and on the data sets for individual species. The peak of the FBP/F6P ratio was significant in rice in ANOVA performed on the combined data sets and the rice data set, and in Arabidopsis in ANOVA performed on the Arabidopsis data set (Fig. 3; Supplementary Fig. S5). This response points to a selective restriction of flux at FBPase and SBPase in both species at irradiance around the respective light compensation point.

There were also marked differences between the two species. For example, the absolute levels of 3PGA and SBP were consistently and significantly higher in rice, whilst FBP and F6P were consistently and significantly higher in Arabidopsis (Fig. 2). Inspection of metabolite ratios (Fig. 3) reveals that the 3PGA/DHAP ratio was consistently higher and the DHAP/RuBP ratio lower in rice than in Arabidopsis, indicating that the balance between the light reactions and RuBP regeneration is shifted to favour RuBP regeneration in rice. Except at the lowest irradiance, the FBP/F6P ratio was consistently lower and the SBP/S7P ratio consistently higher in rice than in Arabidopsis, with the interspecies differences often being significant (Fig. 3; see also Supplementary Fig. S5). This points to a shift in the balance between FBPase and SBPase activity, to favour the former in rice and probably the latter in Arabidopsis. Further, whereas in Arabidopsis, R5P, Ru5P+Xu5P, and RuBP rose progressively with irradiance, in rice they reached high values at 150 µmol m^{-2} s^{-1} and then did not increase much further (Fig. 2; Supplementary Fig. S2). This response indicates that rising rates of photosynthesis are driven by higher RuBP levels in Arabidopsis and by factors that stimulate RuBP utilization in rice.

Correlations between CBC intermediates

Correlation analyses were performed to identify sets of metabolites that show a similar response to rising irradiance, in either Arabidopsis alone (Fig. 4A, B), rice alone (Fig. 4C, D), or the combined Arabidopsis and rice data set (Fig. 4E, F). The analyses were performed with chlorophyll-normalized (Fig. 4A, C, E) and dimensionless (Fig. 6B, D, F) data sets. The analyses in Fig. 4 were performed with data sets that included the dark treatment; parallel analyses using data sets from which the dark treatment was excluded are provided in Supplementary Fig. S6. In analyses with chlorophyll-normalized data, most CBC intermediates correlated with each other in Arabidopsis, in rice, and in the combined data sets. The only exceptions are FBP (especially in rice) and SBP (especially in Arabidopsis). Correlations were often weaker for rice, probably reflecting the plateauing of R5P, Ru5P+Xu5P, and RuBP at relatively low irradiance. Many correlations were weakened or disappeared in the dimensionless data sets, as expected because this normalization removes correlations that are driven by the general increase of metabolite levels as irradiance increases. Positive correlations were retained in Arabidopsis for DHAP versus F6P, FBP versus F6P, SBP versus FBP, SBP versus F6P, S7P versus R5P, and DHAP versus Ru5P+Xu5P, in rice for DHAP versus S7P, R5P versus Ru5P, Ru5P versus RuBP, 2PG versus DHAP, and 2PG versus S7P, and in the

combined data sets for DHAP versus FBP, DHAP versus F6P, DHAP versus S7P, FBP versus F6P, FBP versus SBP, F6P versus S7P, R5P versus Ru5P+Xu5P, Ru5P+Xu5P versus RuBP, 2PG versus DHAP, and 2PG versus S7P. Several negative correlations appeared especially between 3PGA and other CBC intermediates.

Overall, these analyses reveal that although the absolute levels of metabolites vary between Arabidopsis and rice, the qualitative relationships between many CBC intermediates are rather conserved between these two species. The differing absolute levels may reflect differences in the balance between different reactions, whilst correlations common to both species may reflect conserved pathway structure.

Metabolites in pathways for end-product synthesis

The response of metabolites in end-product synthesis pathways is shown in Fig. 5 and Supplementary Fig. S2B. PC analyses with the entire metabolite data set recapitulated many features of the analyses with CBC intermediates (Supplementary Fig. S7). In particular, the lowest irradiance was an outlier in PC analyses with Arabidopsis (Supplementary Fig. S7A, B) and rice (Supplementary Fig. S7C, D), and Arabidopsis and rice showed different trajectories in PC analyses with the combined data sets (Supplementary Fig. S7E, F).

Inspection of the responses of individual metabolites to rising irradiance revealed some features that are shared in Arabidopsis and rice. In particular, there was a progressive and significant increase in dedicated intermediates for starch synthesis (ADPG; Fig. 5A) and sucrose synthesis (Suc6P; Fig. 5C) except for the highest irradiance in Arabidopsis. However, there were also species-dependent differences. G6P, UDPG (Fig. 5B), and G1P (Supplementary Fig. S2) were significantly lower and Suc6P (Fig. 5C) was significantly higher in rice than in Arabidopsis. Whereas PEP peaked at low irradiance and then declined and pyruvate rose progressively with irradiance in rice, PEP rose and pyruvate declined at high irradiance in Arabidopsis (Fig. 5D). These results point to cytosolic fluxes being regulated differently in Arabidopsis and rice.

Correlation analyses (Supplementary Fig. S8) revealed many correlations between metabolites of end-product synthesis, and between these metabolites and CBC intermediates in Arabidopsis. As in the analysis with CBC intermediates, many of the correlations were weakened or lost in the dimensionless data set, indicating that they are driven by a general increase in metabolite levels at higher irradiance. Correlations were weaker in rice, recapitulating the picture for CBC intermediates alone (see above, Fig. 4).

Response of metabolites in rice to low CO_2

CBC intermediates remain high in Arabidopsis in low CO_2 (Arrivault *et al.*, 2009). To test if this is a general response, CBC and other intermediates were compared in rice leaves after 20 min at 500 ppm or 200 ppm CO_2 (Fig. 6; original data provided in Supplementary Dataset S2). Decreasing the CO_2 concentration to 200 ppm led to a small increase in 3PGA (significant; $P<0.05$), no significant change in RuBP (Fig. 6A), and small (<30%) non-significant decreases of metabolites involved in RuBP regeneration (Fig. 6A). There was a 2-fold ($P<0.01$) increase in 2PG (Fig. 6B), consistent with an increase in the rate of photorespiration. There were significant decreases in the levels of metabolites involved in end-product synthesis, including a 3-fold decrease ($P<0.01$) in ADPG (Fig. 6C), a smaller non-significant decreases in G6P, G1P, and Suc6P (Fig. 6D), and a significant ($P<0.01$) decrease of PEP (Fig. 6E).

DISCUSSION

Differences in CBC operation between Arabidopsis and rice

Arrivault *et al.* (2019) reported that profiles of CBC intermediates vary between five C_3 species, and proposed that there is interspecies variance in how the CBC operates. One potential issue with such studies is that as each species shows a different response of photosynthesis to irradiance, cross-species differences in metabolite profiles might be due to the conditions in which plants were harvested. Based on analyses at two irradiances in the C_4 species maize and three irradiances in the C_3 species Arabidopsis, Arrivault *et al.* (2019) concluded that irradiance did not have a strong impact on CBC intermediate profiles.

The current study compares CBC intermediate profiles over a larger irradiance range in two C_3 species, the model eudicot Arabidopsis and the model monocot rice. Due to the differing light response of photosynthesis in Arabidopsis and rice (Supplementary Fig. S1), samples from a given irradiance cannot be directly compared but are better related to the light saturation response of the species. Material was harvested in the dark, close to the light compensation point, and at several irradiances up to ~90% and ~75% of that required to saturate photosynthesis in Arabidopsis and rice, respectively. PC analysis revealed that metabolite levels in Arabidopsis and rice were rather similar in darkness and

very low light, but diverged as irradiance was increased (Fig. 1). This differing trajectory shows that Arabidopsis and rice use partly differing strategies to increase CBC flux.

Loading of metabolites in the PC analysis (Fig. 1) and inspection of the responses of individual metabolites (Figs 2, 3) revealed that several factors drive the separation of Arabidopsis and rice. First, there was a lower 3PGA/triose-P ratio and higher triose-P/RuBP ratio in Arabidopsis than in rice, indicating that the balance between the light reactions and RuBP regeneration is shifted in favour of the light reactions in Arabidopsis and RuBP regeneration in rice. Secondly, except at very low irradiance, there was a consistent trend to a higher FBP/F6P ratio and lower SBP/S7P ratio in Arabidopsis compared with rice, pointing to a shift in the balance between FBPase and SBPase to favour FBPase in rice and SBPase in Arabidopsis. Thirdly, CBC intermediates were more strongly correlated with each other in Arabidopsis than in rice (Fig. 4), indicating that increasing CBC flux is driven by a general increase in metabolite levels in Arabidopsis, whereas regulation of individual enzymes plays a larger role in rice.

One specific example where metabolites were more strongly correlated in Arabidopsis than rice concerns the later steps in RuBP regeneration. Whereas in Arabidopsis the levels of R5P, Ru5P+Xu5P, and RuBP rose progressively with irradiance, in rice R5P, Ru5P+Xu5P, and RuBP rose to high levels at relatively low irradiance and then plateaued. An earlier study with rice also reported that RuBP levels plateaued at lower irradiances than those needed to saturate photosynthesis (Makino *et al.*, 1985). Rising irradiance leads to higher rates of photosynthesis and, hence, of RuBP regeneration and utilization. Our results indicate that this increase in RuBP utilization is largely driven by increased RuBP in Arabidopsis, whilst further factors contribute in rice. Earlier studies also pointed to interspecies variation in the response of RuBP to rising irradiance; for example, RuBP levels rose progressively in bean leaves (Badger *et al.*, 1984) and *Chlamydomonas* (Mettler *et al.*, 2014), whereas RuBP plateaued at irradiances at which photosynthesis was still increasing in wheat (Perchorowicz *et al.*, 1981), spinach (Dietz *et al.*, 1984), and radish (von Caemmerer and Edmondson, 1986) leaves. The latter response resembles that in rice, and points to flux in these species being increased by removing Rubisco inhibitors and/or increasing Rubisco activation.

Chlorophyll and protein content per unit leaf mass are higher in rice than in Arabidopsis. Rice has an unusual leaf anatomy with small lobed mesophyll cells, a high density of chloroplasts with proliferous stromules at the cell surface, and mitochondria located in the inner part of the cell (Sage and Sage, 2009). This increases mesophyll conductance

and promotes recapture of photorespired CO_2 (Sage and Sage, 2009; Busch *et al.*, 2013). Rice also differs from Arabidopsis in having lobe-shaped guard cells and subsidiary cells. These are typical for grasses and allow more efficient CO_2 exchange with the atmosphere (Hetherington and Woodward, 2003; Franks and Farquhar, 2007; Raissig et al., 2016, 2017). These anatomical and ultrastructural features may explain how rice can function efficiently with a higher leaf chlorophyll and protein content, which would otherwise tend to draw down internal CO_2, increase RuBP oxygenation relative to carboxylation, and decrease photosynthetic efficiency. 2PG levels and the RuBP/2PG ratio in rice resembled those in Arabidopsis over much of the light saturation response (Supplementary Figs S3, S5), which is consistent with photorespiration not being higher in rice. Further studies are required to learn to what extent anatomical and ultrastructural adaptations facilitate changes in CBC operation in rice.

There were also differences between Arabidopsis and rice in the responses of metabolites in end-product synthesis (Fig. 5; Supplementary Figs S7, S8). In Arabidopsis, most metabolites in end-product synthesis rose progressively, correlating with CBC intermediates. In rice the response was more diverse; for example, G6P and G1P did not increase with rising irradiance. Further, G6P and UDPG levels were lower and Suc6P levels were higher in rice than in Arabidopsis. These observations point to differences in how sucrose synthesis is regulated, for example a larger role for sucrose-phosphate synthase in rice. There are also implications for CBC operation. The G6P pool typically contains a large part of the total P in photosynthetic metabolites [see Fig. 2 and Szecowka *et al.* (2013) for Arabidopsis, Dietz and Heber (1984), Stitt et al. (1984a, b), and Gerhardt *et al.* (1987) for spinach, Stitt *et al.* (1983) for wheat, and Badger *et al.* (1984) for bean]. The small G6P pool in rice may allow larger cytosolic pools of triose-P and 3PGA, which may affect triose-P export from the chloroplast and energy shuttling via the triose-P/3PGA shuttle.

Shared features in the response to rising irradiance

There are also shared features in the response of Arabidopsis and rice to increasing irradiance. First, both species show a general increase in the levels of many metabolites as irradiance increases, including 3PGA, triose-P, F6P, and S7P over most of the irradiance range, and R5P, Ru5P+Xu5P, and RuBP in the lower part of the irradiance response (Fig. 2; Supplementary Fig. S2B). Secondly, from the light compensation point on, increasing irradiance led to a slight decline in the 3PGA/triose-P ratio and a slight rise

in the triose-P/RuBP ratio. However, in both species, this shift was small compared with the massive rise in CBC flux. Whilst rising irradiance drives 3PGA reduction by supplying more ATP and NADPH, regulation of the CBC enzymes will be needed to speed up conversion of triose-P to RuBP.

Thirdly, in both species, triose-P rose sharply between darkness and the light compensation point, where triose-P already reached a level that was about one-fifth (Arabidopsis) or one-third (rice) of that in high irradiance (Fig. 2). The CBC is an autocatalytic pathway and if CBC intermediates fell to very low levels in low light, time would be needed to build metabolite pools up again after moving back to high light. Maintenance of relatively high levels of CBC intermediates in low light will allow a rapid increase in CBC flux and photosynthesis after sudden increases in light intensity due to changing cloud cover and, especially, canopy movement (Gibbs *et al.*, 2018; Burgess *et al.*, 2019). A rapid response of photosynthetic rate to a rise in light intensity is thought to be an important factor for photosynthetic efficiency in the dense canopies used in modern agriculture (Pearcy *et al.*, 1990; Taylor and Long, 2017; Burgess *et al.*, 2019). The regulatory network that maintains these relatively high levels of metabolites in low light will be discussed in the next subsection.

A fourth shared feature is that FBP and especially SBP rose to a high level or even peaked in very low irradiance (Fig. 2), and that the FBP/F6P and SBP/S7P ratios peaked in low irradiance and then declined (Fig. 3). These results point to FBPase and SBPase restricting CBC flux at irradiances around the light compensation point, presumably due to incomplete redox activation of these enzymes. As mentioned in the Introduction, wasteful futile cycles will occur if enzymes such as FBPase and SBPase are active in the dark when respiratory pathways are operating. Futile cycles would also be especially deleterious in low irradiance, when they would substantially decrease the efficiency with which light energy is used to drive CO_2 fixation, resulting in a lower quantum yield and higher light compensation point.

After relaxing from the low irradiance peak, the FBP/F6P and SBP/S7P ratios remained rather constant (Arabidopsis) or continued to fall (rice) across a wide range of irradiance, even though the rate of photosynthesis rose many fold. This implies that CBC fluxes may be promoted by rising redox activation of FBPase and SBPase over a quite broad irradiance range, partly explaining how CBC flux is speeded up as the light intensity rises. Redox activation of FBPase and SBPase increases strongly between darkness and high light (Buchanan and Balmer, 2005; Heldt *et al.*, 2005). There is, however, little

information on the irradiance range over which this occurs. This is partly because it is technically challenging to extract and assay these enzymes without altering their redox state (Laing *et al.*, 1981; Wirtz *et al.*, 1982). New MS-based methods to quantify redox states of cysteine in proteins (Lennicke *et al.*, 2016; Zhang *et al.*, 2016) will allow testing of these predictions, and may contribute important new insights into CBC regulation.

Our findings are supported by earlier, albeit more fragmentary, studies of CBC intermediate levels in other species (see Supplementary Table S1). For example, studies in wheat protoplasts (Stitt *et al.*, 1983) and leaves of bean (Badger *et al.*, 1984) and spinach (Dietz *et al.*, 1984; Stitt *et al.*, 1984a) reported a general increase in metabolite levels with rising irradiance, relatively small changes in the 3PGA/triose-P and triose-P/RuBP ratios with rising irradiance, a relatively high level of triose-P at low irradiance, and a high level or peak of FBP at low irradiance. A more recent study in *C. reinhardtii* reported marked peaks of FBP and SBP and high FBP/F6P and SBP/S7P ratios in low irradiance (Mettler *et al.*, 2014).

Shared features in the response to low CO_2

Comparison of Fig. 6 with published metabolite profiles for Arabidopsis (Arrivault *et al.*, 2009) reveals that rice and Arabidopsis respond in a similar manner to low CO_2, with only a slight decrease in RuBP and metabolites involved in its regeneration. A similar picture emerges from earlier more fragmentary analyses of the response to low CO_2 in wheat protoplasts (Stitt *et al.*, 1983) and bean and spinach leaves (Badger *et al.*, 1984; Dietz and Heber, 1984), and to treatments such as salinity, water stress, or abscisic acid application that lead to stomatal closure and low intracellular [CO_2] (c_i) (Seemann and Sharkey, 1986, 1987; Sharkey and Seemann, 1989) (Supplementary Table S1). The relatively high levels of CBC intermediates in low CO_2 and under stress conditions that lead to low internal CO_2 imply that end-product synthesis has been inhibited, and this idea is supported by the decline of dedicated metabolites for sucrose and starch synthesis. Maintenance of high levels of CBC intermediates allows continued flux in the CBC to generate RuBP that is oxygenated to 2PG and metabolized via photorespiration. This provides a way in which to dissipate energy in these stressful conditions (see Arrivault *et al.*, 2009; Foyer *et al.*, 2009; Voss *et al.*, 2013).

As outlined in the Introduction, a regulatory network around the cytosolic FBPase involving fructose 2,6-bisphosphate restricts removal of triose-P from the CBC until a threshold concentration of triose-P is reached, and facilitates a large increase in flux to

sucrose when this threshold is exceeded (Herzog *et al.*, 1984; Stitt et al., 1987, 2010; Stitt, 1990). The conserved response of CBC intermediate levels to low CO_2 indicates that this represents a general principle across many species. The observation that triose-P and other CBC intermediates are maintained at relatively high levels at the light compensation point in many species (see the previous section) also points to widespread operation of this regulatory mechanism. In addition to sucrose synthesis, flux to starch needs to be restricted, especially in species where starch is a major product of photosynthesis. Allosteric regulation of ADP glucose pyrophosphorylase by the 3PGA/Pi ratio (Ballicora *et al.*, 2004) will restrict starch synthesis in low irradiance, when most phosphorylated CBC intermediates decline and stromal Pi is likely to be high. Allosteric regulation may be less effective in low CO_2, when CBC intermediates show a smaller decrease than in low light. In these conditions, plastid glucan phosphorylase may contribute by remobilizing starch; this will decrease the net rate of starch synthesis and recycle carbon to maintain CBC metabolite levels (Weise *et al.*, 2006). Whilst starch degradation is under circadian regulation (Stitt and Zeeman, 2012), water stress induces a novel starch degradation pathway involving *α-AMYLASE3* (Zanella *et al.*, 2016). This may bypass regulation by the circadian clock and provide glucans that can be degraded by plastid phosphorylase. Interestingly, recent pioneering studies in which the CBC was introduced into *Escherichia coli* showed that selective evolution was required to allow CO_2-dependent growth of *E. coli*, and that several of the underlying mutations involved a decrease in the activity of enzymes that withdraw intermediates from the CBC (Antonovsky *et al.*, 2017; Herz *et al.*, 2017). This finding underlines that CBC operation depends on regulation of the rate at which intermediates exit this autocatalytic cycle.

In conclusion, the CBC exhibits both contrasting and shared features between the model eudicot Arabidopsis and the model monocot rice. Contrasting features include a shift in the balance between the light reactions and RuBP regeneration, a shift in the balance between FBPase and SBPase, and differences in the regulation of RuBP utilization, which depends strongly on RuBP levels in Arabidopsis and factors that promote Rubisco activity in rice. These differences may be due to interspecies variation in protein abundance, post-translational regulation, or the regulatory properties of enzymes. From the viewpoint of crop improvement, these findings imply that the best targets to increase CBC flux may differ between species. However, our results also reveal shared features between Arabidopsis and rice, and inspection of the fragmentary information for other species indicates that these features may be more widely conserved. One is a restriction on flux

at FBPase and SBPase at low irradiance, indicating incomplete post-translational activation of these enzymes. Another is maintenance of high levels of CBC intermediates at the light or CO_2 compensation point, providing evidence for effective inhibition of end-product synthesis when net C fixation is low. These shared features may be important for efficient CBC function in suboptimal or fluctuating conditions. For example, restricting FBPase and SBPase activity in low irradiance may suppress wasteful futile cycles and improve photosynthetic efficiency in low light. Maintenance of metabolites at relatively high levels in low light or low CO_2 will keep the CBC 'poised to go', allowing a rapid response when irradiance suddenly increases and supporting rapid RuBP regeneration, photorespiration, and energy dissipation when stomata are closed.

SUPPLEMENTARY DATA

Supplementary data are available at *JXB* online.

- **Fig. S1.** Light saturation response of photosynthesis in Arabidopsis and rice.
- **Fig. S2.** Metabolite levels displayed on scales to optimally display the irradiance response for each species, and with ANOVA performed on the Arabidopsis and rice data sets separately.
- **Fig. S3.** Principal component (PC) analysis of CBC intermediates and 2PG on the Arabidopsis data set.
- **Fig. S4.** Principal component (PC) analysis of CBC intermediates and 2PG on the rice data set.
- **Fig. S5.** Metabolite ratios displayed on scales to optimally display the irradiance response for each species, and with one-way ANOVA performed on the Arabidopsis and rice data sets separately.
- **Fig. S6.** Correlation analysis of CBC intermediates and 2PG on the Arabidopsis and rice data sets (excluding dark samples).
- **Fig. S7.** Principal component (PC) analysis of all metabolites on Arabidopsis and rice data sets.
- **Fig. S8.** Correlation analysis for all metabolites in the Arabidopsis and rice data sets.
- **Table S1.** Analysis of published responses of metabolites to changing irradiance or CO_2 concentration.
- **Dataset S1.** Metabolite levels and metabolite ratios in Arabidopsis and rice at different irradiances, and statistical analyses of the combined Arabidopsis and rice data sets by one-way ANOVA.
- **Dataset S2.** Metabolite levels and metabolite ratios in rice at different CO_2 concentrations.

ABBREVIATIONS

- **ADPG** = ADP-glucose
- **CBC** = Calvin–Benson cycle
- **DHAP** = dihydroxyacetone phosphate
- **E4P** = erythrose 4-phosphate
- **FBP** = fructose 1,6-bisphosphate
- **FBPase** = fructose-1,6-bisphosphatase
- **F6P** = fructose 6-phosphate
- **GAP** = glyceraldehyde 3-phosphate
- **G1P** = glucose 1-phosphate
- **G6P** = glucose 6-phosphate
- **NADP-GAPDH** = NADP-glyceraldehyde-3-phosphate dehydrogenase
- **PC** = principal component
- **PEP** = phospho*enol*pyruvate
- **2PG** = 2-phosphoglycolate
- **3PGA** = 3-phosphoglycerate
- **P_i** = inorganic phosphate
- **PRK** = phosphoribulokinase
- **R5P** = ribose 5-phosphate
- **Ru5P** = ribulose 5-phosphate
- **RuBP** = ribulose 1,5-bisphosphate
- **SBP** = sedoheptulose 1,7-bisphosphate
- **SBPase** = sedoheptulose-1,7-bisphosphatase
- **S7P** = sedoheptulose 7-phosphate
- **Suc6P** = sucrose 6^F-phosphate
- **TK** = transketolase
- **triose-P** = triose phosphate (DHAP was measured but is referred to as triose-P in the text)
- **UDPG** = UDP-glucose
- **Xu5P** = xylulose 5-phosphate

ACKNOWLEDGEMENTS

This research was supported by the Max Planck Society (GLB, TAM, MG, RF, VG, JEL, and MS), CNPg (TAM), and the Bill and Melinda Gates Foundation (C_4 Rice, GLB; CASS, SA). We are grateful to Amin Omidbakhshfard for help with harvest of rice plants and to Martha Ludwig for providing the customized Plexiglas chambers. This article is dedicated to the memory of Hans W. Heldt, who carried out pioneering work on regulation of the CBC.

MANUSCRIPT TWO – MAIN FIGURES

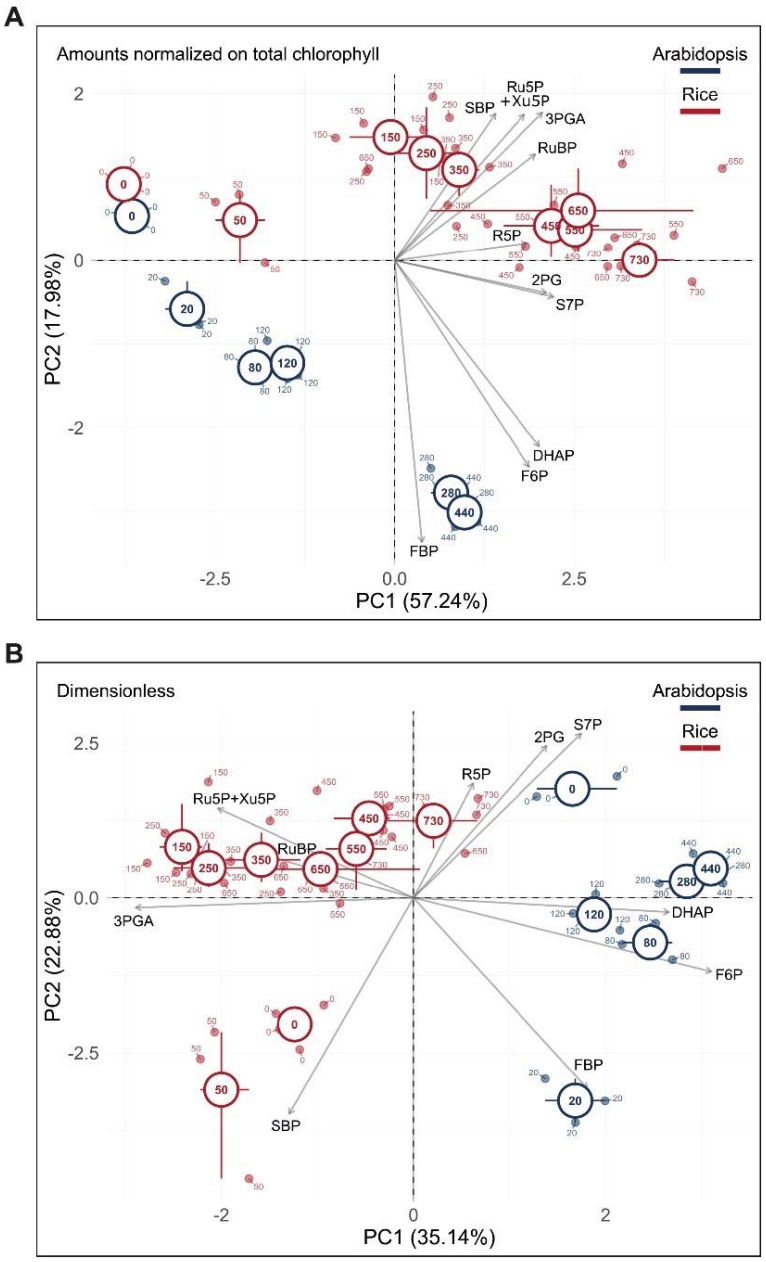


Fig. 1. Principal component (PC) analysis of the response of Arabidopsis and rice to rising irradiance. The plots show PC1 and PC2 for analyses performed on (A) chlorophyll-normalized data sets and (B) dimensionless data sets in which the amount of C in a given metabolite is normed on the total C in CBC intermediates plus 2PG in that sample (see legend to Supplementary Dataset S1 and Arrivault et al., 2019). Samples were collected in darkness and at 20, 80, 120, 280, and 440 µmol m−2 s−1 irradiance for Arabidopsis (shown in blue) and in darkness and at 50, 150, 250, 350, 450, 550, 650, and 730 µmol m−2 s−1 for rice (shown in red). Positions of individual samples at each irradiance are indicated by the corresponding number in small font. The average of these samples in the space defined by PC1 and PC2 is indicated by a circle (irradiance in large font) and the 95% confidence limits are depicted by bars. Metabolite loadings are depicted as grey arrows and the respective metabolite is shown in black font. The unusual asymmetric distribution of loadings in PC1 in the PC analysis with chlorophyll-normalized data indicates that a general increase in the levels of most metabolites contributes to the separation in PC1. PC analyses were also performed separately with the Arabidopsis and rice data sets (see Supplementary Figs S3 and S4, respectively).

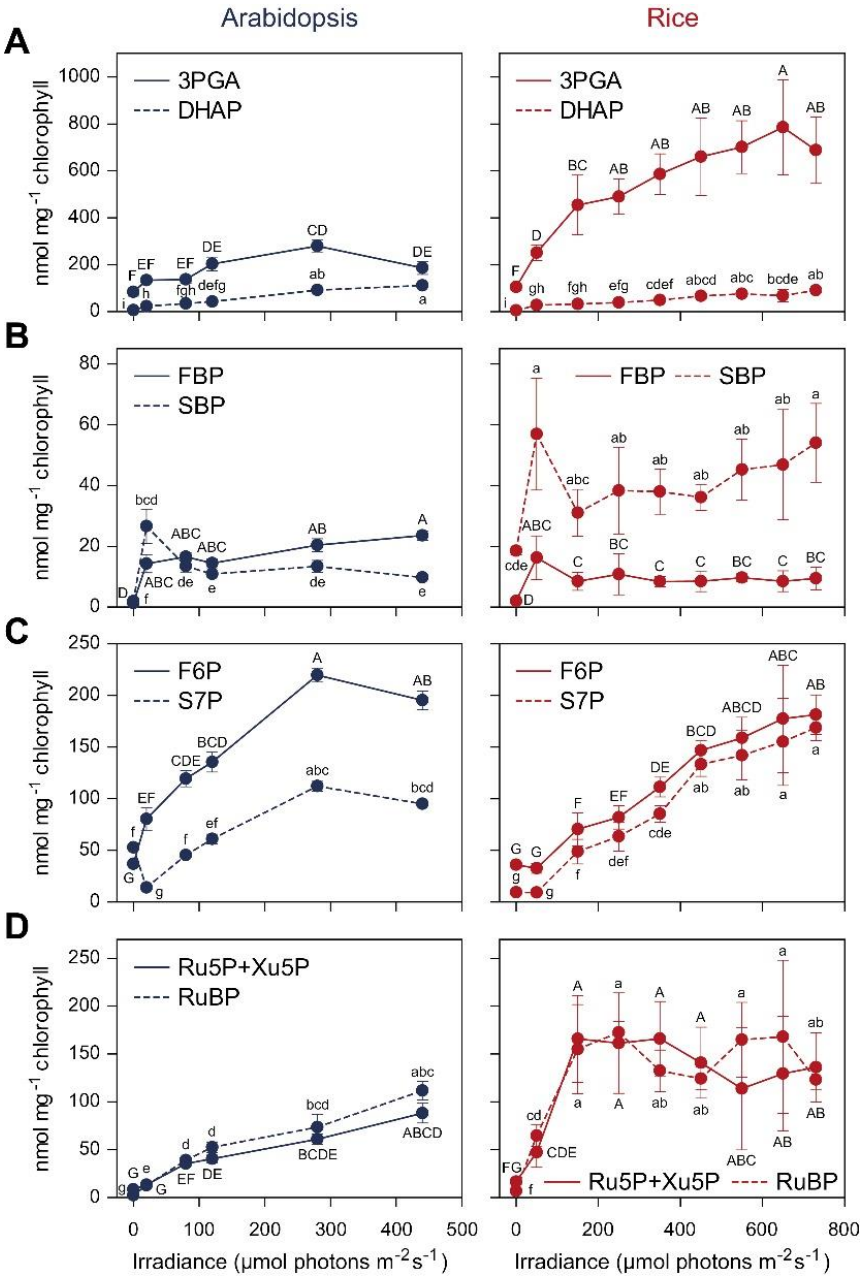


Fig. 2. Response of selected CBC intermediates to rising irradiance in Arabidopsis and rice. (A) 3PGA (line), DHAP (dashed), (B) FBP (line), SBP (dashed), (C) F6P (line), S7P (dashed), (D) Xu5P+Ru5P (line), RuBP (dashed). The response of Arabidopsis is shown in the left-hand side (blue symbols) and of rice in the right-hand (red symbols) of each panel. In a given panel, the same y-axis scale is used for the two metabolites shown in the panel and for both species. Metabolites are normalized on total chlorophyll content. The results are shown as mean ±SD (n=4 in almost all cases; see Supplementary Dataset S1). One-way ANOVA with false discovery rate (FDR) was performed for each metabolite on the entire data set after log transformation, treating each species–irradiance combination as a separate treatment. This tests whether responses of a given species to changes in irradiance are significant, and whether metabolite levels differ between species whereby the latter comparison can be made independently of absolute irradiance. This was followed by a Tukey's HSD post-hoc test. Treatments that are not significantly different share a letter. As two different metabolites are shown in each panel, the Tukey's HSD results are shown with different cases for the two metabolites (e.g. 'a' and 'A'). Letters are assigned such that 'a' or 'A' denotes the treatment group with the highest level. The original data are provided in Supplementary Dataset S1. Plots of all metabolites plus R5P and 2PG with scales selected to optimally view the response in a given species are provided in Supplementary Fig. S2A.

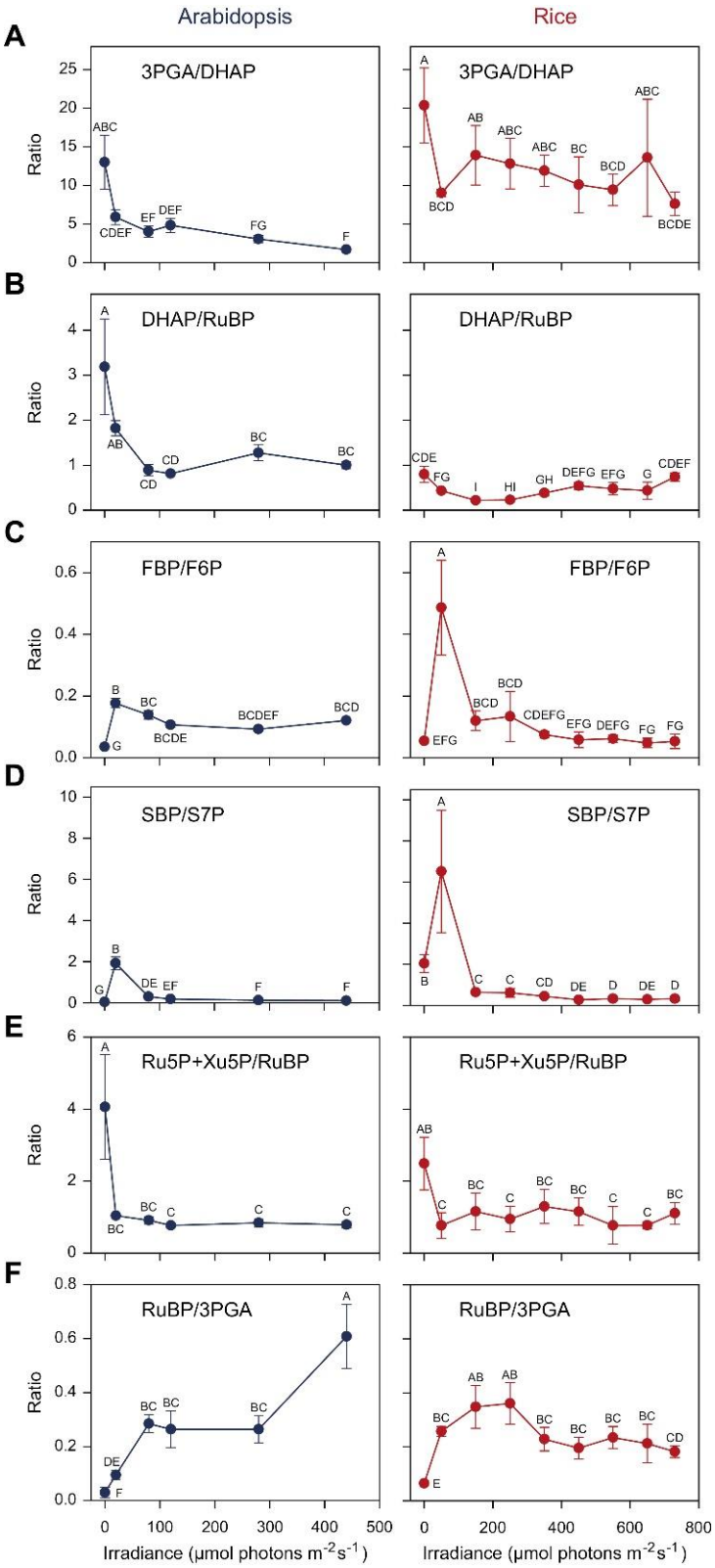


Fig. 3. Metabolite ratios. (A) 3PGA/DHAP, (B) DHAP/RuBP, (C) FBP/F6P, (D) SBP/S7P, (E) Ru5P+Xu5P/RuBP, and (F) RuBP/3PGA. The response of Arabidopsis is shown in the right-hand side (blue symbols) and of rice in the left-hand (red symbols) of each panel. In a given panel, the same y-axis scale is used for both species. The results are shown as mean ±SD (n=4 in almost all cases; see Supplementary Dataset S1). One-way ANOVA with FDR was performed on log-transformed data as in Fig. 2. This was followed by a Tukey's HSD post-hoc test. Treatments that are not significantly different share a letter. Plots of these and further ratios with scales selected to optimally view the response in a given species are provided in Supplementary Fig. S5.

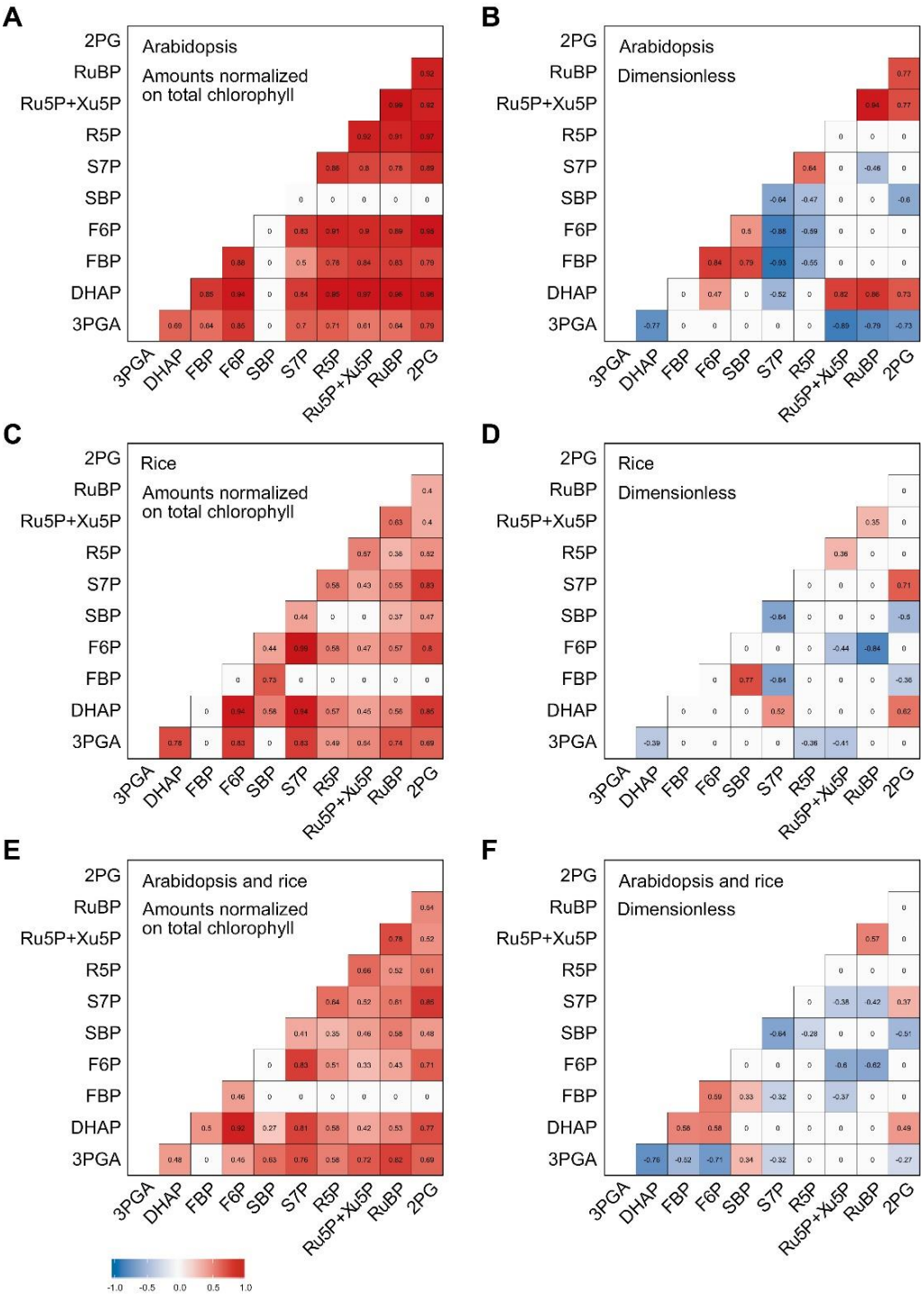


Fig. 4. Correlations between CBC intermediates and 2PG. (A, B) Arabidopsis, (C, D) rice, and (E, F) Arabidopsis and rice combined, using (A, C, E) chlorophyll-normalized data sets or (B, D, F) dimensionless data sets (see legend of Fig. 1) to exclude correlations that are driven by a general increase in metabolite levels. Spearman's r coefficient is denoted in the display and is visualized by a heatmap (red denotes a positive correlation and blue a negative correlation; see panel for scale which is the same for all panels). Non-significant correlations (P>0.05) are not coloured. The original data are provided in Supplementary Dataset S1 and plotted in Fig. 2 and Supplementary Fig. S2A. This analysis used the samples from the dark and all irradiances. A similar analysis excluding dark samples is provided in Supplementary Fig. S6.

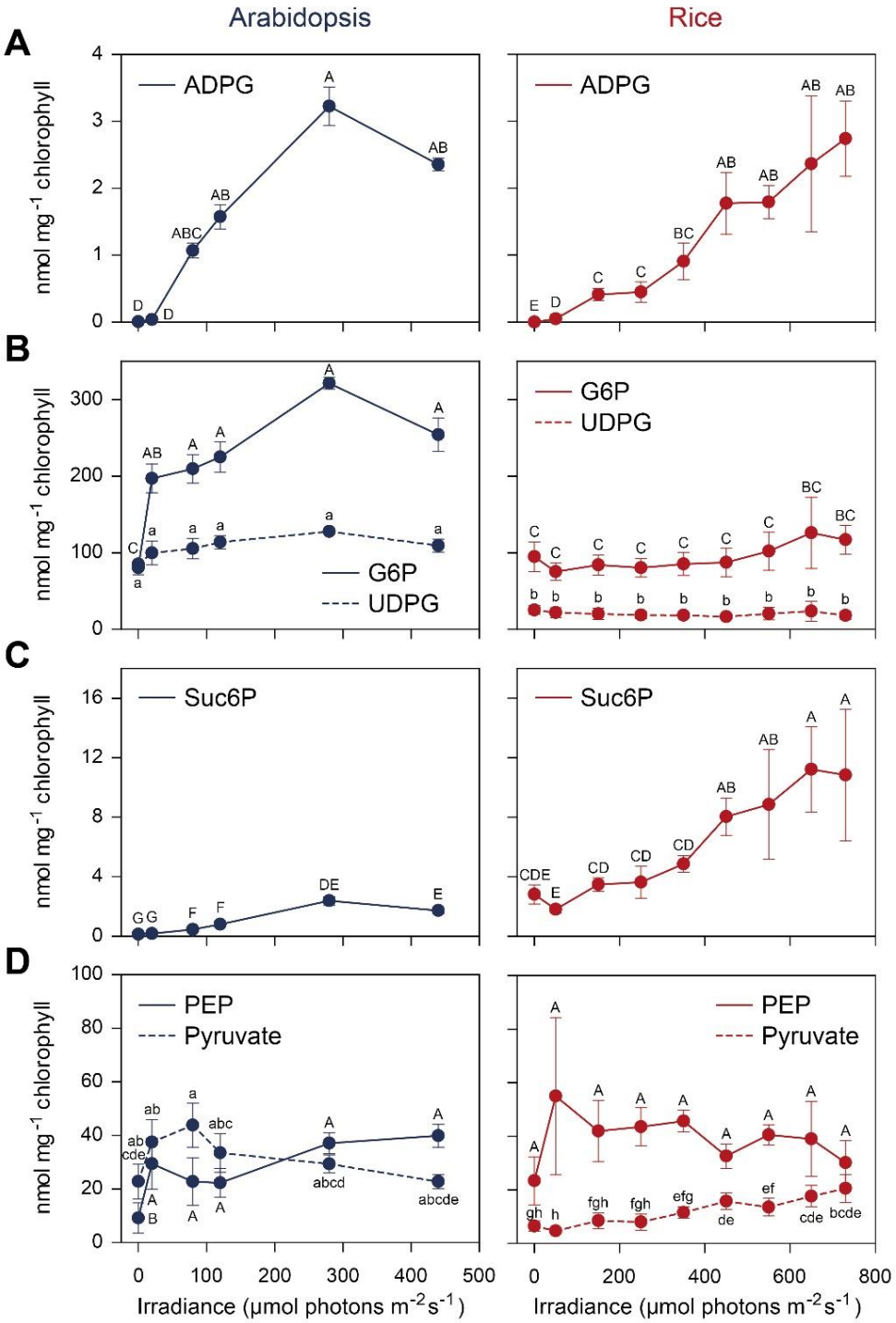


Fig. 5. Response of selected metabolites in pathways of end-product synthesis to rising irradiance in Arabidopsis and rice. (A) ADPG (starch synthesis), (B) G6P and UDPG, (C) Suc6P (sucrose synthesis), and (D) PEP and pyruvate (organic acid synthesis). Metabolites were normalized on total chlorophyll content. The response of Arabidopsis is shown on the left-hand side (blue symbols) and of rice on the right-hand (red symbols) of each panel. In a given panel, the same y-axis scale is used for both of the metabolites shown in the panel and for both species. Metabolite levels are normalized on total chlorophyll content. The results are shown as mean ±SD (n=4 in almost all cases, see Supplementary Dataset S1). One-way ANOVA with FDR was performed on log-transformed data as in Fig. 2. This was followed by a Tukey's HSD post-hoc test. Treatments that are not significantly different share a letter. Plots of all metabolites plus G1P with scales selected to optimally view the response in a given species are provided in Supplementary Fig. S2B.

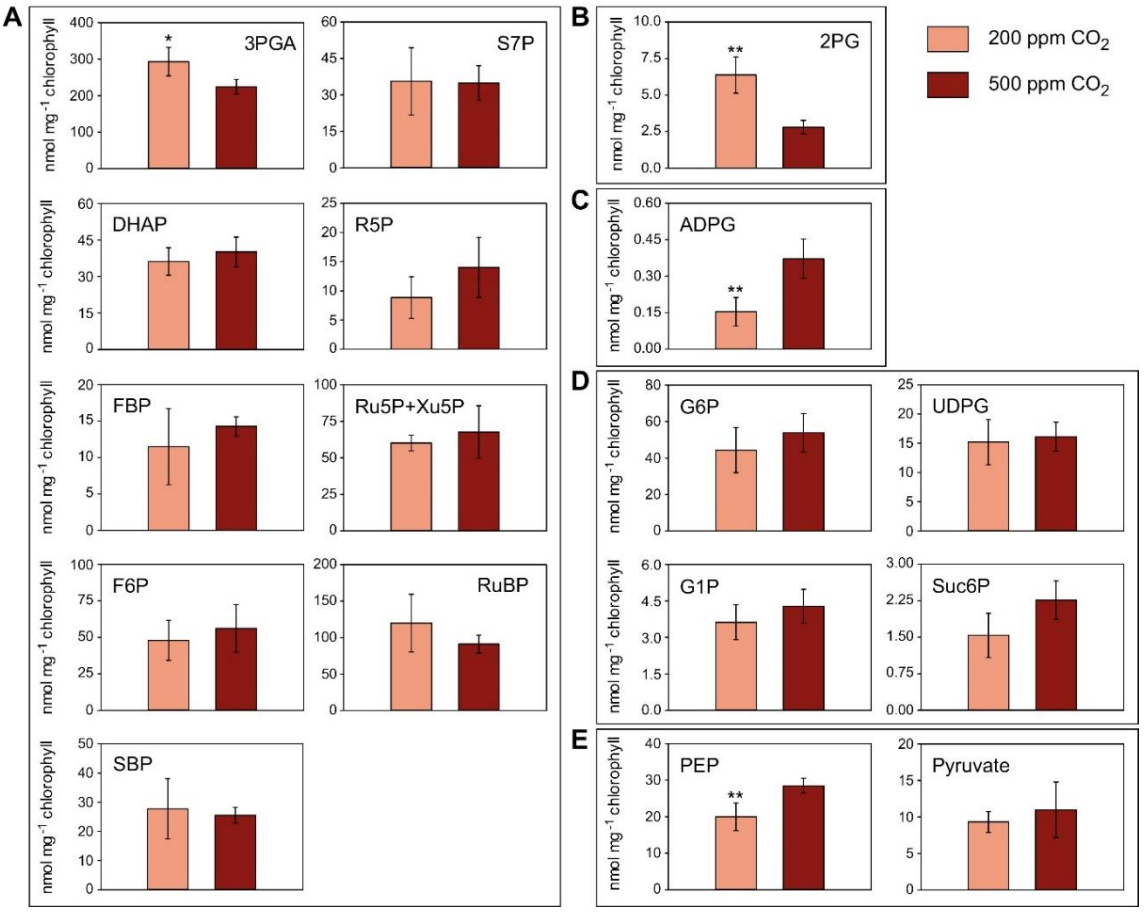


Fig. 6. Response of CBC and other intermediates in rice to a decrease in CO_2 concentration. (A) CBC intermediates (3PGA, DHAP, FBP, F6P, SBP, S7P, RuP, Ru5P+Xu5P, and RuBP), (B) photorespiration (2PG), (C) starch synthesis (ADPG), (D) sucrose synthesis (G6P, UDPG, and Suc6P), and (E) organic acid synthesis (PEP and pyruvate). Rice was supplied with 500 ppm or 200 ppm CO_2 for 20 min at 350 µmol m^{-2} s^{-1} irradiance prior to harvest. Metabolite levels are normalized on total chlorophyll content. The results are shown as mean ±SD (n=4- see Supplementary Dataset S2). Significance was tested by a two-tailed Student's t-test (*P<0.05; **P<0.01).

REFERENCES

Antonovsky N, Gleizer S, Milo R. 2017. Engineering carbon fixation in *E. coli*: from heterologous Rubisco expression to the Calvin–Benson– Bassham cycle. Current Opinion in Biotechnology 47, 83–91.

Arnold A, Nikoloski Z. 2011. A quantitative comparison of Calvin–Benson cycle models. Trends in Plant Science 16, 676–683.

Arnold A, Nikoloski Z. 2012. In search for an accurate model of the photosynthetic carbon metabolism. Mathematics and Computers in Simulation 96, 171–194.

Arrivault S, Alexandre Moraes T, Obata T, *et al.* 2019. Metabolite profiles reveal interspecific variation in operation of the Calvin–Benson cycle in both C_4 and C_3 plants. Journal of Experimental Botany 70, 1843–1858.

Arrivault S, Guenther M, Fry SC, Fuenfgeld MM, Veyel D, Mettler- Altmann T, Stitt M, Lunn JE. 2015. Synthesis and use of stable-isotope labeled internal standards for quantification of phosphorylated metabolites by LC-MS/MS. Analytical Chemistry 87, 6896–6904.

Arrivault S, Guenther M, Ivakov A, Feil R, Vosloh D, van Dongen JT, Sulpice R, Stitt M. 2009. Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. The Plant Journal 59, 826–839.

Ashton AR, Hatch MD. 1983. Regulation of C_4 photosynthesis: physical and kinetic properties of active (dithiol) and inactive (disulfide) NADP-malate dehydrogenase from *Zea mays*. Archives of Biochemistry and Biophysics 227, 406–415.

Ashton AR, Trevanion SJ, Carr PD, Verger D, Ollis DL. 2000. Structural basis for the light regulation of chloroplast NADP malate dehydrogenase. Physiologia Plantarum 110, 314–321.

Badger MR, Sharkey TD, von Caemmerer S. 1984. The relationship between steady-state gas exchange of bean leaves and the levels of carbon reduction-cycle intermediates. Planta 160, 305–313.

Baker JT, Allen LH, Boote KJ, Jones P, Jones JW. 1990. Rice photosynthesis and evapotranspiration in subambient, ambient, and superambient carbon dioxide concentrations. Agronomy Journal 82, 834–840.

Ballicora MA, Iglesias AA, Preiss J. 2004. ADP-glucose pyrophosphorylase: a regulatory enzyme for plant starch synthesis. Photosynthesis Research 79, 1–24.

Bassham JA, Krause GH. 1969. Free energy changes and metabolic regulation in steady-state photosynthetic carbon reduction. Biochimica et Biophysica Acta 189, 207–221.

Betti M, Bauwe H, Busch FA, *et al.* 2016. Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement. Journal of Experimental Botany 67, 2977–2988.

Buchanan BB, Balmer Y. 2005. Redox regulation: a broadening horizon. Annual Review of Plant Biology 56, 187–220.

Burgess AJ, Gibbs JA, Murchie EH. 2019. A canopy conundrum: can wind-induced movement help to increase crop productivity by relieving photosynthetic limitations? Journal of Experimental Botany 70, 2371–2380.

Busch FA, Sage TL, Cousins AB, Sage RF. 2013. C_3 plants enhance rates of photosynthesis by reassimilating photorespired and respired CO_2. Plant, Cell & Environment 36, 200–212.

Cruz JA, Emery C, Wüst M, Kramer DM, Lange BM. 2008. Metabolite profiling of Calvin cycle intermediates by HPLC-MS using mixed-mode stationary phases. The Plant Journal 55, 1047–1060.

Dietz KJ, Heber U. 1984. Rate-limiting factors in leaf photosynthesis. I. Carbon fluxes in the Calvin cycle. Biochimica et Biophysica Acta 767, 432–443.

Dingkuhn M, Schnier HF, de Datta SK, Wijangco E, Dörffling K. 1990. Diurnal and developmental changes in canopy gas exchange in relation to growth in transplanted and direct-seeded flooded rice. Australian Journal of Plant Physiology 17, 119–134.

Driever SM, Simkin AJ, Alotaibi S, Fisk SJ, Madgwick PJ, Sparks CA, Jones HD, Lawson T, Parry MAJ, Raines CA. 2017. Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. Philosophical Transactions of the Royal Society B: Biological Sciences 372, 20160384.

Eisenhut M, Bräutigam A, Timm S, Florian A, Tohge T, Fernie AR, Bauwe H, Weber APM. 2017. Photorespiration is crucial for dynamic response of photosynthetic metabolism and stomatal movement to altered CO_2 availability. Molecular Plant 10, 47–61.

Evans JR. 1989. Photosynthesis and nitrogen relationships in leaves of C_3 plants. Oecologia 78, 9–19.

Farquhar GD, Von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO_2 assimilation. Planta 90, 78–90.

Flügel F, Timm S, Arrivault S, Florian A, Stitt M, Fernie AR, Bauwe H. 2017. The photorespiratory metabolite 2-phosphoglycolate regulates photosynthesis and starch accumulation in Arabidopsis. The Plant Cell 29, 2537–2551.

Foyer CH, Bloom AJ, Queval G, Noctor G. 2009. Photorespiratory metabolism: genes, mutants, energetics, and redox signaling. Annual Review of Plant Biology 60, 455–484.

Franks PJ, Farquhar GD. 2007. The mechanical diversity of stomata and its significance in gas-exchange control. Plant Physiology 143, 78–87.

Fridlyand LE, Backhausen JE, Scheibe R. 1999. Homeostatic regulation upon changes of enzyme activities in the Calvin cycle as an example for general mechanisms of flux control. What can we expect from transgenic plants? Photosynthesis Research 61, 227–239.

Fridlyand LE, Scheibe R. 1999. Regulation of the Calvin cycle for CO_2 fixation as an example for general control mechanisms in metabolic cycles. Bio Systems 51, 79–93.

Galmés J, Andralojc PJ, Kapralov MV, Flexas J, Keys AJ, Molins A, Parry MA, Conesa MA. 2014. Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C_3 genus *Limonium* (Plumbaginaceae). New Phytologist 203, 989–999.

Gardemann A, Schimkat D, Heldt HW. 1986. Control of CO_2 fixation regulation of stromal fructose-1,6-bisphosphatase in spinach by pH and Mg^{2+} concentration. Planta 168, 536–545.

Gardemann A, Stitt M, Heldt HW. 1983. Control of CO_2 fixation: regulation of spinach ribulose-5-phosphate kinase by stromal metabolite levels. Biochimica et Biophysica Acta 6, 115–126.

Gerhardt R, Heldt HW. 1984. Measurement of subcellular metabolite levels in leaves by fractionation of freeze-stopped material in nonaqueous media. Plant Physiology 75, 542–547.

Gerhardt R, Stitt M, Heldt HW. 1987. Subcellular metabolite levels in spinach leaves: regulation of sucrose synthesis during diurnal alterations in photosynthetic partitioning. Plant Physiology 83, 399–407.

Gibbs JA, Pound M, French AP, Wells DM, Murchie E, Pridmore T. 2018. Plant phenotyping: an active vision cell for three-dimensional plant shoot reconstruction. Plant Physiology 178, 524–534.

Gibon Y, Vigeolas H, Tiessen A, Geigenberger P, Stitt M. 2002. Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system. The Plant Journal 30, 221–235.

Gontero B, Maberly SC. 2012. An intrinsically disordered protein, CP12: jack of all trades and master of the Calvin cycle. Biochemical Society Transactions 40, 995–999.

Haake V, Geiger M, Walch-Liu P, Engels C, Zrenner R, Stitt M. 1999. Changes in aldolase activity in wild-type potato plants are important for acclimation to growth irradiance and carbon dioxide concentration, because plastid aldolase exerts control over the ambient rate of photosynthesis across a range of growth condition. The Plant Journal 17, 479–489.

Haake V, Zrenner R, Sonnewald U, Stitt M. 1998. A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants. The Plant Journal 14, 147–157.

Hagemann M, Kern R, Maurino VG, Hanson DT, Weber AP, Sage RF, Bauwe H. 2016. Evolution of photorespiration from cyanobacteria to land plants, considering protein phylogenies and acquisition of carbon concentrating mechanisms. Journal of Experimental Botany 67, 2963–2976.

Hasunuma T, Harada K, Miyazawa S, Kondo A, Fukusaki E, Miyake C. 2010. Metabolic turnover analysis by a combination of in vivo ^{13}C-labelling from $^{13}CO_2$ and metabolic profiling with CE-MS/MS reveals rate-limiting steps of the C_3 photosynthetic pathway in *Nicotiana tabacum* leaves. Journal of Experimental Botany 61, 1041–1051.

Heldt HW, Piechulla B, Heldt F. 2005. Plant biochemistry. Cambridge, MA: Academic Press.

Herz E, Antonovsky N, Bar-On Y, *et al.* 2017. The genetic basis for the adaptation of *E. coli* to sugar synthesis from CO_2. Nature Communications 8, 1705.

Herzog B, Stitt M, Heldt HW. 1984. Control of photosynthetic sucrose synthesis by fructose 2,6-bisphosphate: III. Properties of the cytosolic fructose 1,6-bisphosphatase. Plant Physiology 75, 561–565.

Hetherington AM, Woodward FI. 2003. The role of stomata in sensing and driving environmental change. Nature 424, 6951–6951.

Jelitto T, Sonnewald U, Willmitzer L, Hajirezeai M, Stitt M. 1992. Inorganic pyrophosphate content and metabolites in potato and tobacco plants expressing *E. coli* pyrophosphatase in their cytosol. Planta 188, 238–244.

Kromdijk J, Głowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, Long SP. 2016. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science 354, 857–861.

Laing WA, Stitt M, Heldt HW. 1981. Changes in the activity of ribulosephosphate kinase and fructose- and sedoheptulose-bisphosphatase in chloroplasts. Biochimica et Biophysica Acta 637, 348–359.

Laisk A, Eichelmann H, Oja V. 2006. C_3 photosynthesis in silico. Photosynthesis Research 90, 45–66.

Lawson T, Kramer DM, Raines CA. 2012. Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. Current Opinion in Biotechnology 23, 215–220.

Lefebvre S, Lawson T, Zakhleniuk OV, Lloyd JC, Raines CA, Fryer M. 2005. Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. Plant Physiology 138, 451–460.

Lennicke C, Rahn J, Heimer N, Lichtenfels R, Wessjohann LA, Seliger B. 2016. Redox proteomics: methods for the identification and enrichment of redox-modified proteins and their applications. Proteomics 16, 197–213.

Long SP, Marshall-Colon A, Zhu XG. 2015. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell 161, 56–66.

Long SP, Zhu XG, Naidu SL, Ort DR. 2006. Can improvement in photosynthesis increase crop yields? Plant, Cell & Environment 29, 315–330.

López-Calcagno PE, Howard TP, Raines CA. 2014. The CP12 protein family: a thioredoxin-mediated metabolic switch? Frontiers in Plant Science 5, 9.

Lorimer GH. 1981. The carboxylation and oxygenation of ribulose 1,5-bisphosphate: the primary events in photosynthesis and photorespiration. Annual Review of Plant Physiology 32, 349–382.

Lunn JE, Feil R, Hendriks JH, Gibon Y, Morcuende R, Osuna D, Scheible WR, Carillo P, Hajirezaei MR, Stitt M. 2006. Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in *Arabidopsis thaliana*. The Biochemical Journal 397, 139–148.

Ma F, Jazmin LJ, Young JD, Allen DK. 2014. Isotopically nonstationary ^{13}C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. Proceedings of the National Academy of Sciences, USA 111, 16967–16972.

Makino A, Mae T, Ohira K. 1985. Photosynthesis and ribulose-1,5- bisphosphate carboxylase/oxygenase in rice leaves from emergence through senescence. Quantitative analysis by carboxylation/oxygenation and regeneration of ribulose 1,5-bisphosphate. Planta 166, 414–420.

McClain AM, Sharkey TD. 2019. Triose phosphate utilization and beyond: from photosynthesis to end product synthesis. Journal of Experimental Botany 70, 1755–1766.

Merlo L, Geigenberger P, Hajirezaei M, Stitt M. 1993. Changes of carbohydrates, metabolites and enzyme activities in potato tubers during development, and within a single tuber along astolon–apex gradient. Journal of Plant Physiology 142, 392–402.

Mettler T, Mühlhaus T, Hemme D, *et al.* 2014. Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism *Chlamydomonas reinhardtii*. The Plant Cell 26, 2310–2350.

Moore BD, Isidoro E, Seemann JR. 1993. Distribution of 2-carboxyarabinitol among plants. Phytochemistry 34, 703–707.

Murchie EH, Chen Yz, Hubbart S, Peng S, Horton P. 1999. Interactions between senescence and leaf orientation determine in situ patterns of photosynthesis and photoinhibition in field-grown rice. Plant Physiology 119, 553–564.

Ort DR, Merchant SS, Alric J, *et al*. 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proceedings of the National Academy of Sciences, USA 112, 8529–8536.

Osmond CB. 1981. Photorespiration and photoinhibition: some implications for the energetics of photosynthesis. Biochimica et Biophysica Acta 639, 77–98.

Parry MA, Keys AJ, Madgwick PJ, Carmo-Silva AE, Andralojc PJ. 2008. Rubisco regulation: a role for inhibitors. Journal of Experimental Botany 59, 1569–1580.

Pearcy R. 1990. Sunflecks and photosynthesis in plant canopies. Annual Review of Plant Physiology and Plant Molecular Biology 41, 421–453.

Perchorowicz JT, Raynes DA, Jensen RG. 1981. Light limitation of photosynthesis and activation of ribulose bisphosphate carboxylase in wheat seedlings. Proceedings of the National Academy of Science, USA 78, 2985–2989.

Pettersson G, Ryde-Pettersson U. 1988. A mathematical model of the Calvin photosynthesis cycle. European Journal of Biochemistry 175, 661–672.

Poolman MG, Fell DA, Thomas S. 2000. Modelling photosynthesis and its control. Journal of Experimental Botany 51, 319–328.

Portis AR Jr, Li C, Wang D, Salvucci ME. 2008. Regulation of Rubisco activase and its interaction with Rubisco. Journal of Experimental Botany 59, 1597–1604.

Portis AR Jr, Parry MA. 2007. Discoveries in Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. Photosynthesis Research 94, 121–143.

Prins A, Orr DJ, Andralojc PJ, Reynolds MP, Carmo-Silva E, Parry MA. 2016. Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. Journal of Experimental Botany 67, 1827–1838.

Raines CA, Harrison EP, Ölçer H, Lloyd JC. 2000. Investigating the role of the thiol-regulated enzyme sedoheptulose-1,7-bisphosphatase in the control of photosynthesis. Physiologia Plantarum 110, 303–308.

Raissig MT, Abrash E, Bettadapur A, Vogel JP, Bergmann DC. 2016. Grasses use an alternatively wired bHLH transcription factor network to establish stomatal identity. Proceedings of the National Academy of Sciences, USA 113, 8326–8331.

Raissig MT, Matos JL, Ximena M, *et al*. 2017. Supplementary materials for mobile MUTE specifies subsidiary cells to build physiologically improved grass stomata. Science 355, 1215.

Raven JA, Beardall J, Sánchez-Baracaldo P. 2017. The possible evolution and future of CO_2-concentrating mechanisms. Journal of Experimental Botany 68, 3701–3716.

Sage RF. 2017. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. Journal of Experimental Botany 68, e11–e28.

Sage TL, Sage RF. 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 Physiology 50, 756–772.

Scheibe R. 1991. Redox-modulation of chloroplast enzymes: a common principle for individual control. Plant Physiology 96, 1–3.

Schimkat D, Heineke D, Heldt HW. 1990. Regulation of sedoheptulose- 1,7-bisphosphatase by sedoheptulose-7-phosphate and glycerate, and of fructose-1,6-bisphosphatase by glycerate in spinach chloroplasts. Planta 181, 97–103.

Seemann JR, Sharkey TD. 1986. Salinity and nitrogen effects on photosynthesis, ribulose-1,5-bisphosphate carboxylase and metabolite pool sizes in *Phaseolus vulgaris* L. Plant Physiology 82, 555–560.

Seemann JR, Sharkey TD. 1987. The effect of abscisic acid and other inhibitors on photosynthetic capacity and the biochemistry of CO_2 assimilation. Plant Physiology 84, 696–700.

Servaites JC, Geiger DR, Tucci MA, Fondy BR. 1989. Leaf carbon metabolism and metabolite levels during a period of sinusoidal light. Plant Physiology 89, 403–408.

Sharkey TD. 1985a. Photosynthesis in intact leaves of C_3 plants. The Botanical Review 5, 53–105.

Sharkey TD. 1985b. O_2-insensitive photosynthesis in C_3 plants: its occurrence and a possible explanation. Plant Physiology 78, 71–75.

Sharkey TD. 1988. Estimating the rate of photorespiration in leaves. Physiologia Plantarum 73, 147–152.

Sharkey TD, Seemann JR. 1989. Mild water stress effects on carbonreduction- cycle intermediates, ribulose bisphosphate carboxylase activity, and spatial homogeneity of photosynthesis in intact leaves. Plant Physiology 89, 1060–1065.

Sharwood RE, Sonawane BV, Ghannoum O, Whitney SM. 2016. Improved analysis of C_4 and C_3 photosynthesis via refined *in vitro* assays of their carbon fixation biochemistry. Journal of Experimental Botany 67, 3137–3148.

Simkin AJ, Lopez-Calcagno PE, Davey PA, Headland LR, Lawson T, Timm S, Bauwe H, Raines CA. 2017. Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphophate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO_2 assimilation, vegetative biomass and seed yield in Arabidopsis. Plant Biotechnology Journal 15, 805–816.

South PF, Cavanagh AP, Liu HW, Ort DR. 2019. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. Science 363, eaat9077.

Stitt M. 1990. Fructose-2,6-bisphosphate as a regulatory molecule in plants. Annual Review of Plant Physiology and Plant Molecular Biology 41, 153–185.

Stitt M. 1999. The first will be last and the last will be first: non-regulated enzymes call the tune? In: Burrell JA, Bryant MM, Kruger NJ, eds. Plant carbohydrate biochemistry. Oxford: BIOS Scientific Publishers, 1–16.

Stitt M, Herzog B, Heldt HW. 1984a. Control of photosynthetic sucrose synthesis by fructose 2,6-bisphosphate. I. Coordination of CO_2 fixation and sucrose synthesis. Plant Physiology 75, 548–553.

Stitt M, Huber S, Kerr P. 1987. Control of photosynthetic sucrose formation. In: Hatch MD, Boardman NK, eds. Photosynthesis, a comprehensive treatise. Academic Press, 327–409.

Stitt M, Kürzel B, Heldt HW. 1984*b*. Control of photosynthetic sucrose synthesis by fructose 2,6-bisphosphate: II. Partitioning between sucrose and starch. Plant Physiology 75, 554–560.

Stitt M, Lilley RM, Heldt HW. 1982a. Adenine nucleotide levels in the cytosol, chloroplasts, and mitochondria of wheat leaf protoplasts. Plant Physiology 70, 971–977.

Stitt M, Lunn J, Usadel B. 2010. Arabidopsis and primary photosynthetic metabolism— more than the icing on the cake. The Plant Journal 61, 1067–1091.

Stitt M, Mieskes G, Soling HD, Heldt HW. 1982b. On a possible role of fructose 2,6-bisphosphate in regulating photosynthetic metabolism in leaves. FEBS Letters 145, 217–222.

Stitt M, Schulze D. 1994. Does Rubisco control the rate of photosynthesis and plant growth? An exercise in molecular ecophysiology. Plant, Cell & Environment 17, 465–487.

Stitt M, Sonnewald U. 1995. Regulation of metabolism in transgenic plants. Annual Review of Plant Physiology and Plant Molecular Biology 46, 341–368.

Stitt M, Sulpice R, Gibon Y, Whitwell A, Skilbeck R, Parker S, Ellison R. 2007. Cryogenic grinder system. German Patent No. 08146.0025U1. MPG/ SFX Link Resolver.

Stitt M, Wirtz W, Heldt HW. 1980. Metabolite levels during induction in the chloroplast and extrachloroplast compartments of spinach protoplasts. Biochimica et Biophysica Acta 593, 85–102.

Stitt M, Wirtz W, Heldt HW. 1983. Regulation of sucrose synthesis by cytoplasmic fructose bisphosphatase and sucrose phosphate synthase during photosynthesis in varying light and carbon dioxide. Plant Physiology 72, 767–774.

Stitt M, Zeeman SC. 2012. Starch turnover: pathways, regulation and role in growth. Current Opinion in Plant Biology 15, 282–292.

Szecowka M, Heise R, Tohge T, *et al.* 2013. Metabolic fluxes in an illuminated Arabidopsis rosette. The Plant Cell 25, 694–714.

Taylor SH, Long SP. 2017. Slow induction of photosynthesis on shade to sun transitions in wheat may cost at least 21% of productivity. Philosophical Transactions of the Royal Society B: Biological Sciences 372, 20160543.

von Caemmerer S, Edmondson DL. 1986. Relationship between steady state gas exchange, in vivo ribulose bisphosphate carboxylase activity and some carbon reduction cycle intermediates in *Raphanus sativus*. Functional Plant Biology 13, 669–688.

von Caemmerer S, Farquhar GD. 1981. Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. Planta 153, 376–387.

Voss I, Sunil B, Scheibe R, Raghavendra AS. 2013. Emerging concept for the role of photorespiration as an important part of abiotic stress response. Plant Biology 15, 713–722.

Weise SE, Schrader SM, Kleinbeck KR, Sharkey TD. 2006. Carbon balance and circadian regulation of hydrolytic and phosphorolytic breakdown of transitory starch. Plant Physiology 141, 879–886.

Wirtz W, Stitt M, Heldt HW. 1982. Light activation of Calvin cycle enzymes as measured in pea leaves. FEBS Letters 142, 223–226.

Wullschleger SD. 1993. Biochemical limitations to carbon assimilation in C_3 plants—a retrospective analysis of the A/C_i curves from 109 species. Journal of Experimental Botany 44, 907–920.

Yeoh HH, Badger MR, Watson L. 1980. Variations in k_m (CO_2) of ribulose-1,5-bisphosphate carboxylase among grasses. Plant Physiology 66, 1110–1112.

Zanella M, Borghi GL, Pirone C, Thalmann M, Pazmino D, Costa A, Santelia D, Trost P, Sparla F. 2016. β-Amylase 1 (BAM1) degrades transitory starch to sustain proline biosynthesis during drought stress. Journal of Experimental Botany 67, 1819–1826.

Zhang T, Zhu M, Zhu N, Strul JM, Dufresne CP, Schneider JD, Harmon AC, Chen S. 2016. Identification of thioredoxin targets in guard cell enriched epidermal peels using cysTMT proteomics. Journal of Proteomics 133, 48–53.

Zhu XG, de Sturler E, Long SP. 2007. Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. Plant Physiology 145, 513–526.

Manuscript three

Metabolic characterization of the transition from C_3 to C_4 photosynthesis in the genus *Flaveria*

METABOLIC CHARACTERIZATION OF THE TRANSITION FROM C_3 TO C_4 PHOTOSYNTHESIS IN THE GENUS *Flaveria*

G.L. Borghi [a], S. Arrivault [a], M. Günther [a], David B. Medeiros [a], E. Dell'Aversana [b], G.M. Fusco [b], P. Carillo [b], M. Ludwig [c], A.R. Fernie [a], J.E. Lunn [a], M. Stitt [1-a]

[a] Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany (G.L.B., S.A., M.G., D.B.M., A.R.F., J.E.L., M.S.)

[b] Universitá degli Studi della Campania, Dipartimento di Scienze e Tecnologie Ambientali, Biologiche e Farmaceutiche, Via Vivaldi 43, 81100 Caserta, Italy (E.D.A., G.M.F., P.C.)

[c] The University of Western Australia, School of Molecular Sciences, 35 Stirling Highway, 6009 Perth, Australia (M.L.)

[1] Corresponding author

ORCID: GLB: 0000-0002-6933-2615; SA: 0000-0003-0516-6950; MG: 0000-0002-2448-7322; DBM: 0000-0001-9086-730X; PC: 0000-0003-3723-0398; ML: 0000-0002-0324-7602; ARF: 0000-0001-9000-335X; JEL: 0000-0001-8533-3004; MS: 0000-0002-4900-1763.

For correspondence: Mark Stitt, Tel: (+49) 331 567-8100; E-mail: mstitt@mpimp-golm.mpg.de

Authors emails: Gian Luca Borghi (Borghi@mpimp-golm.mpg.de); Stéphanie Arrivault (Arrivault@mpimp-golm.mpg.de); David Barbosa Medeiros (Medeiros@mpimp-golm.mpg.de); Emilia Dell'Aversana (emilia.dell'aversana@unicampania.it); Giovanna Marta Fusco (giovannamarta.fusco@unicampania.it); Petronia Carillo (petronia.carillo@unicampania.it); Martha Ludwig (martha.ludwig@uwa.edu.au); Alisdair R. Fernie (Fernie@mpimp-golm.mpg.de); John E. Lunn (Lunn@mpimp-golm.mpg.de); Mark Stitt (mstitt@mpimp-golm.mpg.de)

Date of submission: Prepared for submission
Number of tables: 0
Number of figures: 6 (6 in colour for print)
Word count: 12901
Supplementary Material: 5 Figures, 4 Tables, 4 Supplemental Data files

Key words: C_4 photosynthesis, evolution, *Flaveria*, metabolites, Calvin Benson cycle, photorespiration

ABBREVIATIONS

Abbreviations in figures and supplementary material are not included. For all the abbreviations related to metabolites measured in this study please refer to Supplementary Table S1.

This list contains most of the abbreviations present in the main text, ordered alphabetically:

BS = bundle sheath

CA = carbonic anhydrase

CBC = Calvin Benson cycle

CO_2 = carbon dioxide

CCM = CO_2-concentrating mechanism

FBA = flux balance analysis

FW = fresh weight

GC-MS = gas chromatography coupled with mass spectrometry

GDC = glycine decarboxylase

GOGAT = glutamine-oxoglutarate aminotransferase

HPLC = High-performance liquid chromatography

LC-MS/MS = liquid chromatography coupled with tandem mass spectrometry

N_2 = molecular nitrogen

NADP-ME = NADP-dependent malic enzyme

NH_3 = ammonia

NOAA = National Oceanic and Atmospheric Administration

MSO = Murashige-Skoog medium

MSTFA - N-methyl-N-[trimethylsilyl]trifluoroacetamide

O_2 = molecular oxygen

OAA = oxaloacetate

PCA = principal component analysis

PCK = phospho*enol*pyuvate carboxykinase

PEPC = phospho*enol*pyuvate carboxylase

RH = relative humidity

TCA cycle = tricarboxylic acid cycle

AUTHOR CONTRIBUTIONS

G.L.B., S.A., M.L., J.L. and M.S. conceived and planned the experiments. G.L.B. grew plants and harvested samples. G.L.B., S.A., M.G. and D.B.M. extracted the samples for downstream analyses. G.L.B., S.A. and M.G. processed and analyzed LC-MS/MS and enzymatically assayed samples. D.B.M. processed and analyzed GC-MS samples. E.D., G.M.F. and P.C. processed and analyzed HPLC samples. G.L.B. performed statistical analyses and graphs on obtained data. G.L.B. and M.S. wrote the manuscript. M.L., A.F., J.L. and M.S. supervised the project.

ABSTRACT

C_4 plants independently evolved over 60 times from C_3 ancestors. As a complex trait, C_4 evolution required the alteration of the leaf anatomy, leaf physiology and expression of thousands of genes. Under current atmospheric CO_2 and in combination with high light and temperature, the C_4 pathway results is more efficient than the ancestral C_3 pathways due CO_2 concentration around the main carbon assimilating enzyme, Rubisco. This process increases the rate of carboxylation over oxygenation, strongly reducing wasteful photorespiration costs for C_4 plants. The intermediate steps that brought C_3 plants to fully evolve C_4 photosynthesis can still be observed in different genera, of which the genus *Flaveria* (Asteraceae) contains the complete evolutionary spectrum of intermediates and was elected as a model for C_4 evolution.

In this paper, we tackle the evolution of C_4-ness within the genus *Flaveria* by a metabolic perspective. We measured absolute amounts for 52 different primary metabolites in nine different species and ^{13}C enrichments in a smaller selection of metabolites and species. Our results point toward an early recruitment of an aspartate shuttle in C_3-C_4 intermediates and a later switch to malate and pyruvate shuttles in C_4-like and C_4 species. This study provides solid metabolic insights along the evolutionary path and corroborates the current model of evolution, by now mostly supported by anatomical and transcriptomic data. Furthermore, our data give hints on possible hidden steps along the evolutionary trajectory and pose as valid support for subsequent modelling efforts.

INTRODUCTION

During photosynthesis, atmospheric carbon dioxide (CO_2) is assimilated by the enzyme Rubisco, which fuses CO_2 with a five-carbon acceptor ribulose 1,5-bisphosphate (RuBP), to generate two molecules of 3-phosphoglycerate (3PGA). However, in the presence of atmospheric oxygen (O_2), Rubisco catalyzes a side-reaction leading the formation of one molecule of 3PGA and one molecule of 2-phosphoglycolate (2PG). This toxic by-product of Rubisco oxygenation is transformed back to 3PGA at the expense of energy, fixed carbon and nitrogen in a cycle called photorespiration (Lorimer and Andrews, 1973; Bauwe et al., 2010). Some photosynthetic species have evolved different CO_2-concentrating mechanisms (CCMs) to concentrate CO_2 around Rubisco to minimize O_2 interference, thus reducing the wasteful cycle of oxygenation and photorespiration. One of these CCMs is called "C_4" because the first metabolite after the carbon assimilation reaction contains four carbons, rather than the three carbons of 3PGA (Kortschak et al., 1965; Hatch and Slack, 1966; Hatch, 1987; Edwards et al., 2001). In the leaves of C_4 plants, carbonic anhydrase (CA) in the mesophyll cells converts atmospheric CO_2 into HCO_3^- that combines with PEP to form oxaloacetate (OAA). This reaction is catalyzed by PEP carboxylase (PEPC), which has a high affinity for HCO_3^- and no side reaction with O_2. OAA is transformed into other four-carbon metabolites like malate and aspartate (Asp). These four-carbon molecules diffuse to the bundle sheath (BS) cells where they are decarboxylated to generate a high concentration of CO_2 around Rubisco, which together with the remainder of the Calvin Benson cycle (CBC), is localized in the same cell type. The last step is the movement of a three-carbon metabolite back to the mesophyll and the regeneration of the acceptor PEP. C_4 photosynthesis is more efficient than C_3 photosynthesis for several reasons. First, as already mentioned, the wasteful process of oxygenation and photorespiration is strongly reduced (Osmond and Harris, 1971; Sage et al., 2012). This results in a lower CO_2 compensation point, which is a useful diagnostic tool to distinguish C_4 and C_3 species (Moss, 1971). Second, water use efficiency is increased because stomata can close more at a given rate of CO_2 entry to the leaf (Ghannoum, 2009)., Third, C_4 photosynthesis allows increased nitrogen use efficiency (Ghannoum et al., 2011). In the high CO_2 internal environment, Rubisco can evolve to a lower-specificity higher K_m form, allowing the abundance of Rubisco protein to be reduced compared to C_3 species (Brown, 1978; Schmitt and Edwards, 1981). On the

other hand, extra light energy, compared to CBC alone, is needed to drive C_4 CCM (Zhu et al., 2008).

C_4 photosynthesis is a complex trait whose evolution required not only the emergence of a CCM, but also changes in leaf anatomy and cell specification. Most C_4 plants employ what is known as "Kranz" anatomy (Haberlandt, 1918); sparse mesophyll cells, the site of PEPC carboxylation, interspersed between vascular strands with conspicuous BS cells, the site of CO_2 release and re-fixation by the CBC. Thus, compared to C_3 species, C_4 species show an increase in vein density, a decrease in the number of mesophyll cells between veins, reprogramming of the BS cell from an essentially non-photosynthetic cell into a photosynthetic cell, and loss of the Rubisco and most of the CBC, although not of thylakoids and light reactions, from the mesophyll cells (Muhaidat et al., 2007; Lundgren et al., 2014). There are only a few rare cases of single cell C_4 plants, in which assimilation and re-fixation of carbon happens in different intercellular compartment of the same cell (Voznesenskaya et al., 2001).

Under conditions of low CO_2 and high O_2, combined with high irradiance and temperature, photorespiration rates are high and seriously hamper photosynthetic efficiency (Bauwe et al., 2010). Such conditions create an evolutionary pressure to evolve coping mechanisms like the C_4 cycle (Sage et al., 2012; Bräutigam and Gowik, 2016). In fact this CCM evolved more than 60 times independently in different plant lineages (Sage et al., 2011; Sage, 2017), starting from around 30-25 Mya ago with a crucial atmospheric CO_2 drop from values above 1000ppm to below 300ppm (Christin et al., 2008; Zachos et al., 2008). In the current atmospheric CO_2 concentration (global average of 407.4 ppm in 2018; Source NOAA), C_4 plants can maintain higher photosynthetic rates in environmental conditions where higher photorespiration rates are expected, outcompeting C_3 plants in terms of primary productivity (Still et al., 2003). This is particularly true for tropical and temperate grassland habitats, in which C_4 grasses and sedges are most prevalent (Christin and Osborne, 2014; Sage, 2017; Sage et al., 2018).

C_4 photosynthesis is a complex trait, whose evolution requires changes in leaf anatomy, cell specification, chloroplast development and changes in the expression level and/or spatial patterns of probably hundreds of genes (Nelson and Langdale, 1992; Hibberd and Covshoff, 2010; Gowik et al., 2011; Gowik and Westhoff, 2011a; Schlüter and Weber, 2020). The fact that not all plant lineages evolved C_4 photosynthesis suggests that its evolution involved a series of pre-conditioning steps, which have been hypothesized to include close vein spacing, acquisition of genetic regulatory elements and extensive

genomic duplication and size (Langdale, 2011; Sage et al., 2012). The full Kranz anatomy was most likely preceded by a pre-conditioned C_3 ancestor, leading to proto-Kranz species having a more metabolically active BS cells containing bigger and centripetally organized organelles, which can serve as scavenging system for CO_2 released by photorespiration (Mckown and Dengler, 2007; Gowik and Westhoff, 2011a; Sage et al., 2013).

In most genera, the intermediate steps in this evolutionary progression no longer exist. However, the anatomical shift can be observed microscopically in some genera that contain so-called C_3-C_4 intermediates species like *Flaveria*, *Heliotropium*, *Salsola*, *Steinchisma* and *Neurachne* (Ku et al., 1983; Muhaidat et al., 2011; Khoshravesh et al., 2016; Schüssler et al., 2017; Khoshravesh et al., 2019). This shift in anatomy is accompanied by a small decrease in the CO_2 compensation point (Monson et al., 1984). Such genera provide the possibility to explore and test ideas about the evolution of C_4 photosynthesis. Bayesian modelling and meta-analysis of several traits in a large collection of intermediates species in different phylogenies led to the proposal that that the various anatomical, ultrastructural and metabolic traits required for C_4 photosynthesis may have been acquired in different time sequences (Williams et al., 2013). These authors suggested that this flexibility may have facilitated convergent evolution of this complex trait. A complementary analysis modelled the fitness landscape for the evolution of C_3 into C_4 photosynthesis, focusing on metabolic traits (Heckmann et al., 2013). This study concluded that the biochemical sub traits have evolved as modules, and that evolution of each module was associated with a gain in fitness, defined as the amount of CO_2 that can be fixed by a given amount of Rubisco per leaf area. The authors described this as a "Mount Fuji" evolutionary landscape (Heckmann et al., 2013; Heckmann, 2016).

It is widely accepted that the first metabolic step on a continuum between C_3 and C_4 photosynthesis involved a change in the location of glycine decarboxylation, and the development of intercellular shuttles for amino groups. Glycine decarboxylase (GDC) is located in mitochondria and catalyzes the key step in the photorespiratory pathway: the conversion of two molecules of glycine into one molecule of serine plus CO_2 and ammonia (NH_3). Early studies of C_3-C_4 intermediate species showed that GDC activity is decreased or abolished in their mesophyll cells and increased in their BS cells (Hylton et al., 1988; Rawsthorne et al., 1988a; Rawsthorne and Hylton, 1991). Based on the distribution of GDC, it was predicted that C_3-C_4 intermediates operate a photorespiration-driven CCM. In this cycle (termed the C_2 cycle, as glycine has two carbon atoms), Glycine

moves from the mesophyll to the BS cells, where it is decarboxylated leading to up to 3-fold increase of the CO_2 concentration within BS cells (Keerberg et al., 2014), and Ser moves back to the mesophyll. It was proposed that the establishment of this C_2 cycle was of immediate selective advantage in a falling-CO_2 world, as it allows more efficient recapture of photorespired CO_2 and would lower the photosynthetic compensation point. Even more importantly, operation of the C_2 pathway may have created a context in which the next steps towards C_4 could evolve (Bauwe and Kolukisaoglu, 2003; Sage, 2003; Bauwe et al., 2010; Sage et al., 2012; Busch et al., 2013; Heckmann et al., 2013; Schulze et al., 2013; Williams et al., 2013). Firstly, there might be selective pressure to shift Rubisco from the mesophyll into the BS cells, where the CO_2 concentration is enhanced. Secondly, considerations of pathway stoichiometry indicate that C_2 photosynthesis will require large fluxes of organic acids and amino acids between the mesophyll and BS cells. Monson and Rawsthorne (2000) noted that movement of glycine from the mesophyll to the BS cells creates a nitrogen imbalance; for two nitrogen atoms that move (in two glycine molecules) into the BS, only one nitrogen (in one Ser) moves back to mesophyll, whilst the other is released as NH_3 in the BS cells. The consequence of this imbalance were explored in Mallmann et al., 2014. They first used comparative transcriptomics to show that the expression of genes for photorespiration and the glutamine-oxoglutarate aminotransferase (GOGAT) pathway remained as high in C_3-C_4 intermediate species as in C_3 species, and only fell in C_4 species. This is consistent with a high flux through these pathways in C_3-C_4 species. It also underlined that operation of C_2 photosynthesis will only be possible if the asymmetry of nitrogen exchange between glycine and serine is compensated by other intercellular metabolite shuttles. They employed flux balance analysis (FBA) modelling to predict that nitrogen re-fixation by the GOGAT pathway in the BS cells will require large scale movement of 2-oxoglutarate (2OG) to the BS cells and glutamate (Glu) back to the mesophyll or, if their movement is constrained, exchange of pyruvate (Pyr) and alanine (Ala), or malate and aspartate. They also noted that fluxes could be balanced by allowing a C_4 cycle flux, with exchange of malate and alanine, and with malate being decarboxylated in the BS cells. This theoretical analysis pointed to a possible causal relationship between the C_2 cycle and the emergence of C_4 photosynthesis.

C_3 plants already contain C_4 related genes and the associated proteins (Aubry et al., 2011). Current opinion is that during the transition from C_2 to C_4 photosynthesis these pre-existing genes were co-opted under the selective pressure exerted by photorespiration, in

the context of the metabolic preconditioning provided by the C_2 pathway (Sage, 2003; Mallmann et al., 2014; Schlüter and Weber, 2016; Schlüter and Weber, 2020). The non-optimized status reached by early C_4 plants (also called C_4-like) comprised a fully functional C_4 cycle, but still retaining a small fraction of their carbon assimilation by Rubisco in the mesophyll and non-optimal distribution of C_4 enzyme between the tissues. This is exemplified by some members of the *Flaveria* genus (Cheng et al., 1988; Moore et al., 1989) and *Sesuvium sesuvioides* (Bohley et al., 2019). The final step to achieve a complete C_4 status is thought to have included fine tuning of the tissue specific compartmentation of enzymes and their kinetic properties (e.g. K_m and V_{max}) and regulatory properties (Heckmann et al., 2013; Williams et al., 2013). For example, Rubisco enzymes from C_4 plants generally have a lower affinity for CO_2 than the C_3 enzyme, but a higher V_{max} (Kubien et al., 2008; Kapralov et al., 2011), and C_4 PEPC enzymes have lower sensitivity to malate inhibition than C_3 PEPC (Westhoff and Gowik, 2004; Gowik and Westhoff, 2011b).

As already mentioned, our ideas about the evolution of C_4 photosynthesis are based on experimental data from a few genera like *Flaveria*, in which species with putative intermediate stages are still extant. *Flaveria* is a relatively young genus (2-3 Mya; Sage, 2017) belonging to the Asteraceae family, which finds its geographical origin in the Mesoamerican region (Powell, 1978). *Flaveria* genus include 23 species, comprising both C_3 and NADP-dependent malic enzyme (NADP-ME) subtype C_4 species, as well as many intermediates between these two photosynthetic types (Apel and Maass, 1981; Ku et al., 1983; Nakamoto et al., 1983; Sudderth et al., 2007). The phylogenetic tree comprises three C_3 species at the bottom of the tree, and then a division into two major clades: Clade A includes all of the fully C_4 species as well as several advanced intermediates, while Clade B includes C_3-C_4 intermediates, and one species (*F. brownii*) that has reached the status of C_4-like (Kopriva et al., 1996; McKown et al., 2005; Lyu et al., 2015). C_3 *Flaveria* ancestral species already possessed pre-conditioning elements that helped this genus proceed down the path of C_4 evolution, like the duplication of the chlorenchima GLDP gene and the reduction of one of these isoforms in mesophyll cells (Schulze et al., 2013; Schulze et al., 2016). Furthermore, the attainment of the status of proto-Kranz quite early in the evolutionary tree (see *F. robusta* and F. *pringlei*; Sage et al., 2012; Sage et al., 2013; Sage et al., 2014) favored the emergence of C_2 photosynthesis in ancestors of both clades: this idea is also strengthened by the presence at the tree base of *F. sonorensis*, a primitive C_3-C_4 intermediate (Moore et al., 1987). Elements of the C_4 pathway were

already present in early C_3-C_4 *Flaveria* intermediates like *F. anomala*, however not as carbon assimilation-relevant as in advanced C_3-C_4 species like *F. ramosissima*, which assimilate around 50% of their carbon through PEPC (Moore et al., 1987; Chastain and Chollet, 1989). Apart from *F. brownii*, which has an ancestral C_4-like phenotype (Monson et al., 1987; Cheng et al., 1988), all the other C_4-like *Flaveria* species are more advanced and fix >90% of their carbon using a C_4 cycle (Moore et al., 1989).

The basis for our current understanding of the carbon assimilation strategies in different *Flaveria* species was laid down by studies of enzyme activities and kinetic properties, followed by genetic sequence characterizations. These studies targeted GDC (Hylton et al., 1988; Kopriva et al., 1996; Schulze et al., 2013), PEPC (Nakamoto et al., 1983; Bauwe, 1984; Bauwe and Chollet, 1986; Ku et al., 1991; Svensson et al., 1997; Engelmann et al., 2003; Westhoff and Gowik, 2004), Rubisco (Bauwe, 1984; Wessinger et al., 1989; Kubien et al., 2008; Kapralov et al., 2011; Perdomo et al., 2015) and several others enzymes of the C_4 cycle (Ku et al., 1983; Bauwe, 1984; Ku et al., 1991; Drincovich et al., 1998; Ludwig, 2011). These data provided a basis for formulating draft metabolic pathways in the various species. These pathways fit well with results from comparative studies of physiological traits, like the CO_2 compensation point (Apel and Maass, 1981; Ku et al., 1983; Monson et al., 1987; Cheng et al., 1989; Monson, 1989; Ku et al., 1991; Leegood and von Caemmerer, 1994; Dai et al., 1996; Sudderth et al., 2007; Kocacinar et al., 2008; Vogan and Sage, 2011) and carbon isotope signatures (Smith and Turner, 1975; Apel and Maass, 1981; Smith and Powell, 1984; Monson et al., 1988; Sudderth et al., 2007). Most of the more recent support for the current *Flaveria* evolutionary model comes from further comparative studies of anatomy (McKown and Dengler, 2007; Kocacinar et al., 2008; Sage et al., 2013; Stata et al., 2016; Billakurthi et al., 2018) and transcriptome profiles (Gowik et al., 2011; Schulze et al., 2013; Mallmann et al., 2014; Stata et al., 2016; Billakurthi et al., 2018). There have also been several studies of labelling kinetics, usually showing that label incorporation into malate and aspartate in C_3-C_4 intermediates is much higher than in C_3 species but lower than in C_4 species (Bassüner et al., 1984; Rumpho et al., 1984; Monson et al., 1986; Byrd et al., 1992). This is consistent with the idea (Mallmann et al., 2014) that a primitive low-level C_4 cycle may contribute to maintaining stoichiometry in the intercellular shuttles between the mesophyll and BS cells.

There have been few studies of metabolite pools: one pioneering study investigated how intracellular CO_2 impacts on metabolite levels in one C_4-like and two C_4 species (Leegood and von Caemmerer, 1994). More recent studies investigated a small set of metabolites

at fixed irradiance in a C_3, a C_3-C_4 and two C_4 species (Gowik et al., 2011), and CBC intermediates in two C_4 species at one irradiance level (Arrivault et al., 2019). There is however, to our knowledge, no study in which a broad range of metabolites have been compared across many *Flaveria* species. Such information is needed to test the idea that intercellular metabolite movement is a key feature of C_3-C_4 photosynthesis and to learn which metabolites may be moving. On the assumption that intercellular movement is by diffusion, it will require buildup of large overall pools of the metabolites that are participating in intercellular shuttles (see Leegood, 1985; Stitt and Heldt, 1985; Arrivault et al., 2017). Comparative profiling of metabolite levels may provide insights into how C_3-C_4 photosynthesis evolved, and into the steps involved in the progression to C_4-like and then to complete C_4 photosynthesis. In this paper we provide quantitative measurements of 52 primary metabolites in two C_3, two C_3-C_4 intermediates, three C_4-like and two C_4 *Flaveria* species, together with the ^{13}C-enrichment in a selected subgroup of species and metabolites. The main aim is to test current ideas about different photosynthetic pathways, support the elucidation of the *Flaveria* genus evolutionary path, and provide a basis for future metabolic modelling.

MATERIAL AND METHODS

Chemicals

Solvents and other chemicals for metabolite analyses were obtained from Merck (www.merckmillipore.com), Roche Applied Science (www.lifescience.roche.com) and Sigma-Aldrich (www.sigmaaldrich.com). Labelled carbon dioxide ($^{13}CO_2$, isotopic purity 99%) was from Campro Scientific GmbH (www.campro.eu), while N_2, O_2, and unlabeled CO_2 from Air Liquide (www.industrie.airliquide.de/).

Plant material and harvests for metabolite measurements

Eight *Flaveria* species (*F. robusta*, *anomala*, *ramosissima*, *palmeri*, *vaginata*, *brownii*, *trinervia* and *bidentis*) seeds were germinated on moist soil under 16/8 h day/night cycle, 150 µmol m^{-2} s^{-1} irradiance, 20/18 °C and relative humidity (RH) of 70%. After two weeks, each seedling was transferred into a 6 cm diameter pot. One week later they were repotted in 18 cm diameter pots and grown under 16/8 h day/night cycle, average

irradiance of 350 µmol m^{-2} s^{-1}, 26/22 °C, 70% RH, repotting in 30 cm diameter pots two weeks before the harvest.

One species (*Flaveria cronquistii*) due to seed sterility, was propagated by sterile cuttings (sterilization media: 0.5% MSO + 5% PPM® (Plant Cell Technologies Inc.)) grown in tissue culture (growth media: 1% MSO + 6.8% agarose + 1% PPM® (Plant Cell Technologies Inc.). Well-rooted cuttings were transferred to 10 cm diameter pots containing soil after approximately 30 days. After two weeks of adaptation with a 16/8 h day/night cycle, 150 µmol m^{-2} s^{-1}, 20/18 °C, 70% RH, plantlets were repotted into 18 cm diameter pots and moved to a 16/8 h day/night cycle, 350 µmol m^{-2} s^{-1} irradiance, 26/22 °C, 70% RH. The plants were repotted in 30 cm diameter pots two weeks before the harvest.

Harvest was conducted for most of the species 55-56 days after sowing (with the exception of the more slow-growing *F. ramosissima* and *F. trinervia*, harvested between 79-80 days after sowing). *F. cronquistii* plants were harvested 55 days after being transferred to soil. First fully expanded leaves were carefully cut under an average irradiance of 350 µmol m^{-2} s^{-1} and quickly plunged in a bath of liquid N_2, without shading or turning the leaf in the process. Four biological replicates, each from different plants, were harvested for each *Flaveria* species. For a resume of growing and harvesting conditions, refer to Supplementary Data file S4.

Plant material and harvests for $^{13}CO_2$ labelling

Species for performing labelling were only four: *F. robusta* (C_3), *F. ramosissima* (C_3-C_4), *F. palmeri* (C_4-like) and *F. bidentis* (C_4), chosen among the others for photosynthetic diversity, as well as growing habits and leaf shapes suitable for the labelling experiments. Seeds of the above-mentioned species were germinated on moist soil under 16/8 h day/night cycle, 150 µmol m^{-2} s^{-1}, 20/18 °C and 70% RH. After 3 weeks, each seedling was transferred into 10 cm diameter pots and grown under 16/8 h day/night cycle, average irradiance of 500 µmol m^{-2} s^{-1}, 26/22 °C, 70% RH, until the labelling harvest happening 49/50 days after sowing.

Labelling was conducted by lodging a single leaf per plant (i.e. biological replicate) within custom-made transparent gas-tight acrylic glass chamber connected with a gas flowmeter, infrared CO_2 analyzer (LI-800 Gashound; LI-COR Biosciences), humidifier bottle and gas bottles. The overall labelling system was adapted from similar systems

used in previous publications (Arrivault et al., 2009; Arrivault et al., 2017). Within the chamber, the irradiance at leaf level was exactly 500 µmol m^{-2} s^{-1}. At first, every leaf was acclimated for 2-3 minutes with an artificial atmospheric gas mixture (~420 ppm $^{12}CO_2$, 21% O_2, 78% N_2), then subjected to either 40 or 60 min $^{13}CO_2$ pulses (~420 ppm) using the same gas concentrations. Time point zero was harvested at the end of the acclimation phase, without introducing labelled CO_2. The harvest was conducted by rapidly flooding the chamber with liquid N_2 while avoiding shading. For a resume of growing and harvesting conditions, refer to Supplementary Data file S4.

Quantitative metabolite analyses

Leaf samples were ground into a fine powder at liquid N_2 temperatures using metal marbles and a mixer mill at 25 Hz speed (Retsch GmbH) and stored at -80 °C. Most of organic acids, sugar phosphates and cofactors were measured after methanol/chloroform extraction from 15 mg fresh weight (FW) aliquots using an established reverse phase liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) platform (Arrivault et al., 2009). Stable Isotopic Labelled Internal Standards (SIL-IS) were added to rule out matrix effects for a subset of metabolites, as in Arrivault et al., (2015). Pyruvate, PEP, 3PGA and ATP were measured enzymatically on 50 mg FW aliquots extracted with trichloroacetic acid (Merlo et al., 1993), due to loss of these metabolites in the methanol/chloroform extraction method, or poor detection with our LC-MS/MS platform. Most of the amino acids were measured by HPLC on 20 mg FW aliquots after ethanolic extraction and OPA-derivatization (Carillo et al., 2005).

^{13}C-enrichment analyses

$^{13}CO_2$ labelled samples were ground in a fine powder as done for other plant material in this paper (see above). Metabolite extraction for gas chromatography coupled with mass spectrometry (GC-MS) was carried out as described previously (Lisec et al., 2006) using 50 mg FW samples. Ribitol (0.2 mg mL^{-1} in H_2O) was added during extraction as an internal standard. A dry aliquot (150 µL of the polar phase) was resuspended in methoxyamine hydrochloride (20 mg ml^{-1} in pyridine) and derivatized using N-methyl-N-[trimethylsilyl]trifluoroacetamide (MSTFA). The GC-TOF-MS system comprised a CTC CombiPAL autosampler, an Agilent 6890N gas chromatograph, and a LECO Pegasus III TOF-MS running in EI+ mode. A volume of 1 µL of derivatized metabolite

solution was used for injection. Peaks were annotated to metabolites by comparing mass spectra and GC retention times to database entries and those of authentic standards available in a reference library from the Golm Metabolome Database (Kopka et al., 2005). For mass spectral analysis, the relative isotopomer abundance and the determination of the total ^{13}C enrichment was performed using Xcalibur® 2.1 software (Thermo Fisher Scientific) and the CORRECTOR software (www.mpimp-golm.mpg.de/719693/Bioinformatik-Tools) was used to correct for the distorting effect of naturally occurring stable isotopes (Huege et al., 2014).

Statistical analyses

Clustering and principal component analyses (PCAs) were performed in R Studio vers. 1.2.5033 (www.rstudio.com) integrated with R vers. 3.6.1 (www.r-project.org/). One-way ANOVA with Holm-Sidak post-hoc tests were computed using SigmaPlot vers. 14.0 (Systat Software, Inc.).

RESULTS

Global analysis of metabolite levels in nine *Flaveria* species

The nine *Flaveria* species chosen for this study represent all of the modes of photosynthesis found within the genus. They include two basal C_3 species (*F. robusta* and *F. cronquistii*), two C_3-C_4 intermediates (*F. anomala* and *F. ramosissima*; clade B and A, respectively), three C_4-like species (*F. brownii*, *F. palmeri* and *F. vaginata*; clade B, A and A respectively), and two complete C_4 species (*F. bidentis* and *F. trinervia*; both clade A) (Supplementary Fig. S1). Each species was grown in the same controlled conditions and harvested under growth irradiance by rapidly quenching the youngest fully expanded leaves in a bath of liquid nitrogen (see Material and Methods and Supplementary Data file S4 for details).

We measured 52 primary metabolites, mostly by LC-MS/MS, GC-MS and HPLC in nine *Flaveria* species. They encompassed sugar-phosphates, organic acids and amino acids, and covered most of the CBC, the C_4 pathway (from here onward termed "CCM-related" metabolites), photorespiration, the TCA cycle, precursors for sucrose, starch and cell wall synthesis, and several cofactors (listed with abbreviation and method of analysis in

Supplementary Table S1). Absolute metabolite amounts are provided in Supplementary Data file S1.

To provide an overview of the metabolite profile, metabolite amounts per unit FW were Z-score normalized and used to perform two way clustering (Fig. 1, results displayed as heat map) and PCAs (Fig. 2A-B; Supplementary Fig. S2A-B). PCA was also performed on data that was transformed to give each metabolite as the amount of carbon in that metabolite normalized on total carbon in all measured metabolites (Fig. 2C-D, S2C-D; termed a 'dimensionless' data set, see Supplementary Data file S1 for details, method applied also in Arrivault et al., 2019). This normalization excludes possible bias due to species differences in leaf protein and/or water content.

All of these analyses separated the species based on their photosynthetic type and on the phylogenetic tree. In the clustered heatmap (Fig. 1) basal C_3 species were separated on the left side, clade A species (C_3-C_4, C_4-like and C_4) in the middle and clade B (*F. anomala* and *F. brownii*) species on the far right. Within each clade, C_3-C_4, C_4-like and (for clade A) C_4 species grouped separately. In clade A, the C_3-C_4 species *F. ramosissima* grouped closer to the basal C_3 species (*F. cronquistii* and *F. robusta*) than to the C_4-like (*F. palmeri* and *F. vaginata*) and C_4 species (*F. bidentis* and *F. trinervia*). In the PCA, PC1 and PC2 captured 22 and 18% of the total variance in the FW-normalized dataset (Fig. 2A-B), and 42 and 14% of the total variance in the dimensionless dataset (Fig. 2C-D) (for more details regarding PC contributions see Supplementary Table S2). Both PC analyses separated species based their mode of photosynthesis and their phylogeny. The true C_4 species (*F. bidentis* and *F. trinervia*) took positive scores in PC1 and separated on this axis from the other photosynthesis types. C_4-like species separated from C_3-C_4 and C_3 species mainly in PC2. In clade A, the C_3-C_4 species *F. ramosissima* was slightly separated from the two C_3 species (*F. cronquistii* and *F. robusta*) and, curiously, in an opposite direction to the clade A C_4-like species (*F. vaginata*, *F. palmeri*). This separation of the basal and clade A species indicates that the evolutionary path from C_3 through C_3-C_4 and C_4-like to true C_4 does not involve a progressive change in metabolite levels, but instead the establishment of discrete states at each stage in the evolutionary process. In clade B, the C_3-C_4 species *F. anomala* is well separated from the basal C_3 species, again mainly in PC2, and in this case lies closer to the corresponding C_4-like species *F. brownii*. In the analysis with FW-normalized data, PC1 and PC2 captured only 39% of the total variance. We therefore also inspected PC3 and PC4 combined with PC1 (Supplementary Fig. S2; 12 and 11% of total variance, respectively). There was no clear pattern in PC3.

C_3-C_4 species separated clearly from C_3 species in PC4, and this separation reverted slightly for the C_4-like and C_4 species.

With regard to which metabolites drive the separations, both the clustered heatmap and the PCAs reveal similarities and differences between photosynthetic types. Starting with C_3 species, the separation from other species (except clade A C_3-C_4 species, see below) was driven by 2PG, glycine, serine and glycerate, several CBC intermediates (e.g., RuBP, Ru5P+Xu5P, R5P, S7P), and some amino acids (proline, threonine, histidine, glutamine, arginine, glutamate, also the aromatic amino acid biosynthesis intermediate shikimate). As already mentioned, the PC analysis separated C_3-C_4 species in a clade-dependent manner. *F. anomala*, an early C_3-C_4 in clade B, has a unique amino acidic profile with high tryptophan, tyrosine, methionine, glutamine and total amino acids as well as high SBP and FBP. The clade B C_4-like species *F. brownii* partly shared this pattern. *F. ramosissima*, which is an advanced C_3-C_4 species in clade A, lay close to the basal C_3 species. Its separation from other species is driven by high levels of some CBC (mainly S7P, F6P, R5P) and photorespiratory (2PG, glycine and glycerate) intermediates. Separation of the two clade A C_4-like species from C_3-C_4 intermediate (and C_3) species was driven by higher levels of CCM-related metabolites (malate, pyruvate) and some amino acids (alanine, asparagine). Separation of true C_4 species from C_4-like species was driven mainly by high TCA cycle intermediates like 2OG, citrate, isocitrate and aconitate and, to a lesser extent (only clear in the analysis with FW-normalized data set, see below within the Discussion) pyruvate and malate.

Summarizing, clade B species differ from clade A species mainly by high levels of some amino acids (e.g., alanine, asparagine, tryptophan, tyrosine and methionine) and sugar phosphates (FBP, SBP, G1P and G6P). Within clade A, C_3-C_4 species tend to have higher CBC and photorespiratory metabolites than basal C_3 species, C_4-like species have higher CCM intermediates, and the move from C_4-like to true C_4 is mainly associated with higher levels of several more organic acids.

Changes of individual metabolites between species

To provide more information about the responses of metabolites highlighted by our clustering and PCAs, and to search for further metabolite-specific responses that might be masked in these global analyses, we inspected the responses of individual metabolite (Fig. 3A-C – Supplementary Fig. S3 A-D).Fig. 3A shows malate, aspartate, pyruvate,

alanine and PEP, which are directly involved in the C_4 CCM. It also shows glutamine, glutamate and 2OG, which might potentially facilitate aminotransferase reactions in C_4 photosynthesis and/or be involved in shuttles that return -NH_2 groups from the BS to the mesophyll in C_3-C_4 photosynthesis (see Introduction). All of these metabolites were relatively low in the *Flaveria* C_3 species, except for aspartate and glutamate whose levels in the two C_3 species resembled those in most of the other *Flaveria* species. The only major difference between the two C_3 species was for glutamine, which was high in *F. cronquistii* and significantly lower in *F. robusta*. The transition from C_3 to C_3-C_4 intermediate was not accompanied by a significant increase of any CCM-related metabolite. The only noticeable difference between the two C_3-C_4 species was for glutamine, which was significantly higher in *F. anomala* than *F. ramosissima*. In the transition between C_3-C_4 and C_4-like species, some metabolite levels started to rise, especially when viewed as a photosynthesis-type trend, with a trend to higher levels of malate, pyruvate and alanine in C_4-like species than C_3-C_4 species. That said, the C_4-like species were quite diverse in terms of metabolite levels: for example, the three species had differing levels of malate, alanine, glutamate, glutamine and PEP. Part of these differences might be attributed to *F. brownii* (low glutamate, high alanine and low PEP), possibly due to it belonging to a different clade. In the transition from C_4-like to true C_4, several CCM-related metabolites showed a significant rise: very clearly 2OG (Fig. 3A) and several other TCA cycle intermediates like citrate, isocitrate and aconitate (see in Supplementary Fig. S3) and, as a trend, malate, pyruvate and PEP. On the other hand, aspartate and glutamate levels were lower levels C_4 species than other species. Within the clade A species, alanine did not change between C_4-like to true C_4 species.

Levels of CBC intermediates (Fig. 3B) were less dependent on photosynthesis mode. Comparing the two C_3 species, many CBC metabolites showed higher levels in *F. cronquistii* than *F. robusta* (significant for RuBP, DHAP and F6P, non-significant for FBP, S7P, SBP and R5P): an exception is 3PGA, in which amounts are comparable. Comparing the C_3 species (*F. cronquistii* and *F. robusta*) and the C_3-C_4 species (*F. anomala* and *F. ramosissima*), in most cases there were no significant or consistent changes across all four species. Compared to *F. robusta*, the C_3-C_4 species had significantly higher RuBP, DHAP, F6P and S7P. Compared to *F. cronquistii,* there were no significant changes. Within the C_3-C_4 species, only F6P and R5P were different, due to higher levels in *F. ramosissima*. Comparing C_4-like species with C_3-C_4 species, there was a trend to lower RuBP (significant in the comparison with *F. anomala* but not *F.*

ramosissima). In between the three C_4-like species, the most different is the clade B species *F. brownii*, which had significantly higher FBP, F6P and SBP than the clade A C_4-like species. Comparison of C_4-like clade A species and true C_4 species revealed some changes in CBC metabolites, with a trend to lower RuBP, FBP and SBP, although this was only significant for some metabolite and pairwise comparisons. The two C_4 species had very similar level of CBC intermediates, except that *F. bidentis* had significantly higher DHAP and F6P levels. Across entire set of C_3-C_4, C_4-like and true C_4 species, there was a consistent trend to lower levels of RuBP, FBP and SBP in true C_4 species *F. bidentis* and *F. trinervia*. This contrasts with 3PGA or DHAP, which did not change.

The next panel shows metabolites involved in photorespiration; the first intermediate 2PG and glycine, serine and glycerate. (Fig. 3C). These metabolites could potentially be involved in C_3-C_4 intermediate CCM shuttling (check C_2 photosynthesis part in the Introduction). None of these photorespiratory metabolites did show any significant differences between the C_3 species (*F. cronquistii* and *F. robusta*). If we move to C_3-C_4 intermediate species things start to change: *F. anomala* had the highest levels of serine of all nine species, that holds for *F. ramosissima* with 2PG (significant) and glycine (although not significant) higher amounts than in C_3 species. Compared to C_3-C_4 species (*F. anomala* and *F. ramosissima*), C_4-like species showed lower levels of glycine (significant in some pairwise comparisons), serine (significant in all pairwise comparisons) and glycerate (significant in some pairwise comparisons). Compared to C_4-like species, true C_4 species had similar levels of 2PG, consistently but non-significantly higher glycine and serine, and higher and in some pairwise comparisons significantly higher glycerate. Curiously, whereas C_4-like species tended to have low levels of glycerate, the two true C_4 species had the highest glycerate levels of all nine species; this might indicate a role for glycerate in C_4 photosynthesis that is not directly related to photorespiration.

Changes of classes of metabolites between species

After seeing that many individual metabolites show marked changes between species belonging to C_3, C_3-C_4, C_4-like and C_4 photosynthetic types, we wondered if there were any marked trends for the summed content of metabolites from different sectors of metabolism. We first normalized the metabolite data set and expressed the amount of carbon in each metabolite as a fraction of the summed carbon in all analyzed metabolites

(see Fig. 2 and Supplementary Data file S1 for calculation method details and Supplementary Table S1 for the list of all metabolites measured). We then grouped metabolites based on pathways or sectors of metabolism: CBC, photorespiration, CCM-related, TCA cycle and amino acid metabolism (Fig. 4). As already explained the class termed 'CCM-related' contained metabolites that are involved in intercellular shuttles in C_4 photosynthesis, as well as metabolites that could potentially be involved in shuttles that carry -NH_2 groups back to the mesophyll cells in C_3-C_4 species. In some cases, metabolites were assigned to more than one class (for example, malate and 2OG were assigned to both 'CCM-related" and "TCA cycle"). As metabolite levels are normed on the total carbon in all detected metabolites, this plot reveals not only changes in the relative levels of individual metabolites but also in allocation of carbon between different sectors of central metabolism.

Relatively little carbon was allocated to the CBC in C_3 species (2.4-3%). This reflects the small size and rapid turnover of CBC metabolites (see also Szecowka et al., 2013; Ma et al., 2014; Arrivault et al., 2017). There was marginally more (3.0-3.8%) in C_3-C_4 species, falling to 1.1-2.0% in C_4-like species and 0.4-0.8% in true C_4 species. Inspection of individual metabolites reveals a fairly homogeneous reduction of all CBC metabolites across this evolutionary axis although, for example, RuBP tended to show a larger relative decline.

Photorespiration accounted for about 5-10% of the carbon in central metabolism in C_3 species. This reflects the large size and relatively slow turnover of glycine, serine and glycerate (see also Szecowka et al., 2013; Ma et al., 2014; Arrivault et al., 2017). Allocation to photorespiratory metabolites rose to 13-20% in C_3-C_4 intermediate species, and fell to under 2.5% in C_4-like and C_4 species. These marked changes are due to large changes in a small number of metabolites. In particular, serine contribution rose between C_3 and C_3-C_4 species and fell dramatically in C_4-like and C_4 species. Glycerate also decreased dramatically between C_3 or C_3-C_4 species and C_4-like species. The changes of these carbon-normalized values differ slightly from the FW-related values discussed in Fig. 3C; nevertheless, there is a trend to higher relative levels of glycerate in true C_4 than C_4-like species.

Metabolites assigned to the CCM-related category accounted for 44-47% of all carbon in C_3 and C_3-C_4 species, slightly more (53-58%) in C_4-like species and 38-44% in true C_4 species. It should be noted that in any given species, many metabolites in this set may not be involved in intercellular shuttles (see Introduction) and that part of the pool may not

be directly involved in photosynthesis (see below). It is noticeable that carbon allocation to malate starts to rise in the transition between C_3-C_4 intermediates and C_4-like species, while carbon in glutamate and aspartate starts to decline at the same transition. Similar to what was seen by comparing absolute levels of metabolite between species (Fig. 3A), carbon allocation to 2OG rose only in true C_4 species *F. bidentis* and *F. trinervia*. alanine shows a unique pattern: it tends to be high in all Clade B species (between 5 and 15%) and one clade A species, *F. palmeri* (just below 5%).

Moving to TCA cycle intermediates, C_3 species and C_3-C_4 species have relatively high carbon allocation (20-45%) with marked differences between the single species, rising further to 45-70% in C_4-like species and above 75% in true C_4 species. Regarding single metabolites, citrate, similar to 2OG (already highlighted in the CCM-related category), only rose to high level (over 30%) in true C_4 species and one advanced C_4-like species (*F. vaginata*).

We restricted our analysis of amino acids to major ones, defined as those amino acids contributing to >5% of total carbon allocation in at least one species. The trend of carbon allocation to amino acids was largely inverse to that for the TCA cycle. In C_3 species, amino acids accounted for >50% of total carbon in *F. cronquistii* and slightly less than 40% in *F. robusta*. In C_3-C_4 species, carbon allocation to amino acids was marginally higher, with slightly over 60% in *F. anomala* and slightly above 45% in *F. ramosissima*. In C_4-like species, carbon allocation in amino acids was considerably lower, with values slightly above 40% in *F. brownii* and about 20% in *F. vaginata*. True C_4 species had the lowest carbon allocation into amino acids (about 10% in *F. bidentis* and 5% in *F. trinervia*). The contribution of some specific amino acids was discussed above in the context of photorespiration or CCM-related categories. Others included asparagine, proline and the aromatic amino acid precursor, shikimate. Asparagine reaches a carbon allocation above 10% only in C_4-like species. Proline and shikimate, normally not very carbon relevant in other *Flaveria* species, made together around 20% of total carbon allocation in the C_3 species *F. cronquistii*.

$^{13}CO_2$ labelling experiments in four selected species

Many metabolites are compartmented, with large pools in vacuoles or non-photosynthetic cells. Consequently, the total pool may not provide an accurate reflection of the pool involved in active metabolism; this issue is especially acute for organic and amino acids

(see Introduction). One way to deconvolute compartmentation is by $^{13}CO_2$ labelling. Experiments in C_3 tobacco (Hasunuma et al., 2010), C_3 Arabidopsis (Szecowka et al., 2013; Ma et al., 2014), C_3 cassava (Arrivault et al., 2017) and the C_4 plant maize (Wang et al., 2014; Arrivault et al., 2017) have shown that CBC intermediates usually label to high enrichment within seconds to minutes. Enrichment in metabolites involved in photorespiration rise more slowly, but usually plateaus by 60 min. In the C_4 species maize, pools involved in the CCM plateau by 15-20 min. The labeled portion is thought to reflect the pool that is actively involved in photosynthesis, i.e. which is directly downstream of carbon assimilation.

It is challenging to perform $^{13}CO_2$ labeling experiments with many pulse durations on a large number of species. We carried $^{13}CO_2$ labelling on four representative *Flaveria* species belonging to C_3, C_3-C_4, C_4-like and C_4 photosynthetic categories in clade A, to avoid clade-specific differences. We chose pulse durations of 40 or 60 minutes, by which time we expected that most enrichment in most metabolites would be at, or approaching, a plateau (see last paragraph). An unlabeled control (0 min) was included for all species, except for *F. ramosissima*. Enrichment was determined using GC-MS in nine key metabolites, mainly belonging to CCM-related or photorespiratory pathways (Supplementary Table S1 for GC-MS measured metabolites, highlighted in red). $^{13}CO_2$ enrichment values are displayed in Supplementary Data file S2.

$^{13}CO_2$ enrichment values are plotted against time in Fig. 5, and as histograms with the 40 and 60 min time points averaged in Supplementary Fig. S4 (here it is also indicated if a plateau was reached). Malate was weakly labelled (10% or lower enrichment) in the C_3 *F. robusta* and true C_4 species *F. bidentis* (resembling other C_3 and true C_4 species; Hasunuma et al. 2010, Arrivault et al., 2013, Ma et al., 2014; Weismann et al., 2016; Arrivault et al., 2017) and the C_3-C_4 species *F. ramosissima*, and slightly more strongly (about 25%) in the C_4-like *F. palmeri*. Enrichment in pyruvate reached over 50% in C_4-like and true C_4 *Flaveria* species (like maize, see Weismann et al., 2016; Arrivault et al., 2017) but was lower (around 30%) in C_3 and C_3-C_4 *Flaveria* species. Enrichment rose to quite high levels in aspartate (about 60%) in the C_3-C_4 and C_4-like species, but was lower in the C_3 and true C_4 species. Enrichment in aspartate is lower in the C_4 *F. bidentis* species than in earlier studies of maize (Weismann et al., 2016; Arrivault et al., 2017), a NADP-ME species with dimorphic chloroplasts. Enrichment in alanine rose to 60-80% in C_3-C_4, C_4-like and C_4-like species, and about 50% in the C_3 *F. robusta* species. Considering the photorespiration pathway metabolites, labelled carbon enrichment rose to high levels in

glycine (75-90%) and serine (70-80%) but was lower in glycerate (20-60%; enrichment was lowest in the C_3 *F. robusta* species). Slower labeling and incomplete labelling of glycerate has been seen in studies with other species and may reflect slow movement of label through the preceding large glycine and serine pools. glutamate and 2OG were weakly and incompletely labelled in all four *Flaveria* species, as seen in earlier studies of other C_3 and C_4 species (Szecowka et al., 2013; Ma et al., 2014; Ishihara et al., 2015; Arrivault et al., 2017). Enrichment was highest for 2OG in the C_3-C_4 species (about 35%) and for glutamate in the true C_4 *F. bidentis* species (about 20%). It is hard to explain the high enrichment in most controls (0 min) for metabolites like 2OG and glycerate; the most likely hypothesis is that they represent artifacts of the GC-MS data processing.

Estimation of minimum active pools in *Flaveria* species, and comparison with C_4 maize

We multiplied the ^{13}C enrichment values by pool size to estimate a minimum value for the active pool for each metabolites and species (Supplementary Table S3) and displayed it in a single histogram (Supplementary Fig. S5). For comparison, this displays contain active pools in the model C_4 species *Zea mays* (maize), estimated from published data from Arrivault et al. (2017). In cases where a clear plateau was not reached, the active pool sizes in the *Flaveria* species are minimum values. Nevertheless, our estimates highlight differences in active pool size between the four *Flaveria* species, indicate the extent to which true C_4 *Flaveria* species resemble maize, and reveal at what stage in the evolutionary process from C_3 to C_4 photosynthesis major changes in metabolism occur.
Compared to the small active malate pool in the C_3 *Flaveria robusta* (85 nmol/g FW), the active pool of malate rose marginally (by about 75%) in the C_3-C_4 *F. ramosissima*, and dramatically (over ten-fold) in the C_4-like *F. palmeri* and true C_4 *F. bidentis*, reaching values almost as high as those previously estimated for maize. The active pool of pyruvate was very low in the C_3 *F. robusta* (4 nmol/g FW), rose only slightly (about 80%) in the C_3-C_4 *F. ramosissima* and dramatically (20-40-fold) in the C_4-like *F. palmeri* and C_4 *F. bidentis*, again to values similar to those previously estimated for maize. aspartate was relatively high in the C_3 *F. robusta* (180 nmol/g FW), rose about 3-fold in the C_3-C_4 *F. ramosissima*, and declined in the C_4-like *F. palmeri* and especially in the true C_4 *F. bidentis*, whose estimated active aspartate pool resembled that previously estimated for maize. The active pool of alanine in the C_3 *F. robusta* (80 nmol/g FW) is about half that

of aspartate, remained low (like pyruvate) in the C_3-C_4 *F. ramosissima* and rose dramatically (over 15-fold) in the C_4-like *F. palmeri* and true C_4 *F. bidentis* to values like those in maize.

Among metabolites involved in photorespiration, the active pool of glycine was low in the C_3 *F. robusta* (about 50 nmol/g FW), rose almost 2-fold in the C_3-C_4 intermediate *F. ramosissima* and declined again the C_4-like *F. palmeri* and true C_4 *F. bidentis*, with values 2-fold lower than previously estimated for maize. The active pool of serine was large in the C_3 *F. robusta* (710 nmol/g FW), rose another 70% in the C_3-C_4 *F. ramosissima*, and fell dramatically in the C_4-like *F. palmeri* and true C_4 *F. bidentis* to values like those estimated for maize. It is noteworthy that the active pool of serine in the C_3-C_4 species *F. ramosissima* is 12-fold higher than that for glycine, and was also 5-10-fold higher than that of any other analyzed metabolite in this species, except for aspartate. The estimated active pool of glycerate in the C_3 *F. robusta* (about 170 nmol/g FW) rose almost 3-fold in the C_3-C_4 *F. ramosissima*, was very low in the C_4-like species *F. palmeri* and high in the true C_4 species *F. bidentis* (this reflected both the relatively larger pool and enrichment).The active pools of 2OG and glutamate were low in the C_3 *F. robusta*, rose about 2-fold in the C_3-C_4 species *F. ramosissima* and C_4-like species *F. palmeri*, then (glutamate only) rose by another 2-fold in the true C_4 species *F. bidentis*. The estimated active pool sizes of 2OG and glutamate in the true *C_4 Flaveria* species resembled those previously estimated for maize.

Cross–species comparison of the CBC intermediate profile in the *Flaveria* genus with two further true C_4 species and five further C_3 species

Recently, Arrivault et al. (2019) presented a cross-species comparison of CBC intermediates and 2PG in five C_3 plants and four NADP-ME subtype C_4 plants, including the two C_4 *Flaveria* species used in the current study. In these CBC-focused PCAs, the C_3 species separated from the C_4 species, and there was also divergence between different C_3 species, and between different C_4 species. These analyses pointed to changes in CBC operation between C_3 and C_4 plants, likely as a result of the CBC operating in a high-CO_2 environment in the latter. They also pointed to inter-species variation in CBC function in C_3 plants. We were interested to learn how *Flaveria* C_3 species, C_3-C_4 intermediate species and C_4-like species mapped onto this landscape. To do this, we performed PCA on a combined data set that contained information on the levels of CBC metabolites and

2PG in the nine species used by Arrivault et al. (2019) and the seven new C_3, C_3-C_4 and C_4-like *Flaveria* species from the current study (Fig. 6). The dataset combining FW-normalized values, as well as the dimensionless dataset, are available as Supplementary Data file S3. The dimensionless normalization within was obtained by expressing the carbon in a given metabolite as a fraction of the summed carbon in all CBC metabolites and 2PG (for details see Supplementary Data file S3). We chose this normalization because some of the species in the dataset from Arrivault et al (2019) had high protein or chlorophyll, leading to a skewed view when metabolite levels are expressed on FW basis (for details see Arrivault et al., 2019).

PC1 and PC2 (Fig. 6) captured 30.4% and 24.5% of the total variance in this large data set (see also Supplementary Table S4 for the first ten PC contributions). Strikingly, the overall sample distribution of the nine species in the study of Arrivault was largely retained in the analysis of this extended species compendium (compare Fig.6 to Fig. 4D in Arrivault et al. (2019)). Of the seven added *Flaveria* species, the two C_3 species lie close to other C_3 species especially *Triticum aestivum* and *Oryza sativa*. All the C_3-C_4 species and the clade B C_4-like species *F. brownii* fall into a diagonal between true C_4 and C_3 species. The clade A C_4-like species *F. palmeri* and *F. vaginata* cluster very close to the true C_4 species *F. bidentis* and *F. trinervia*, which are also in clade A. The clade B C_3-C_4 species *F. anomala*, except for one data point, falls really close to other intermediate species, demonstrating a similar CBC asset even if possessing an early C_3-C_4 intermediacy status. This separation is seen along PC1, and is driven in particular by high 3PGA and lower levels of 2PG and several CBC intermediates like S7P, F6P, SBP, FBP and, especially, RuBP (Fig. 6B).

The clade B C_4-like *F. brownii* is clearly separated from all of the other species, irrespective of whether they are C_4, C_4-like and C_3-C_4 intermediate species. This separation is driven by high FBP and SBP (Fig. 6B). *F. brownii* deviation from other C_4-like and C_4 species shows how the CBC pathway is not constrained into one single evolutionary direction when C_4 photosynthesis is evolving at its side, and might even indirectly support the supposedly independent origin of C_4-ness in clade B *Flaveria*.

DISCUSSIONS

C_4 photosynthesis first appeared as an adaptation to an abrupt global climate change (including an atmospheric CO_2 drop) around 25-30 million years ago in relatively short span of time (Christin et al., 2008). From that time on, evolutionary selective pressure in specific environments was high enough to promote the independent convergent evolution of C_4 photosynthesis (and its intermediate forms) in over 60 vascular plant genera (Sage, 2017), despite the higher energy cost this CCM brings with it (Zhu et al., 2008). In current atmospheric conditions, C_4 photosynthesis still represents an improvement compared to C_3 photosynthesis in warm, dry environments, allowing higher rates of photosynthesis and higher water and nitrogen use efficiencies. Therefore, there were and currently are several engineering efforts to introduce it into C_3 crops (Matsuoka et al., 2001; Häusler et al., 2002; Miyao et al., 2011; Ermakova et al., 2019). These efforts depend on a good understanding of the operation of C_4 photosynthesis, but may also gain from a better understanding of how C_4 photosynthesis evolved.

In this light, *Flaveria* represent a very interesting plant genus due to its photosynthetic diversity. The genus includes basal C_3 species, as well as C_3-C_4 intermediate, C_4-like and C_4 species in clade A, while only C_3-C_4 intermediate and a C_4-like species in clade B (see Supplemental Fig. S1 and McKown et al., 2005; Lyu et al., 2015). The genus was extensively characterized both photosynthetically and enzymatically during the 80's and 90's (Ku et al., 1983; Nakamoto et al., 1983; Smith and Powell, 1984; Moore et al., 1987; Wessinger et al., 1989; Ku et al., 1991; Dai et al., 1996; Drincovich et al., 1998). However, except for some limited studies of a small group of metabolites and species (Leegood and von Caemmerer, 1994; Gowik et al., 2011; Arrivault et al., 2019), a comprehensive study of metabolite levels in multiple *Flaveria* species was lacking. The goal of our study was to quantify the absolute amounts of many primary metabolites by chromatography and mass spectrometry in set of photosynthetically different *Flaveria* species.

Our study reveals substantial metabolic diversity between *Flaveria* species. Importantly, metabolite profiles on their own suffice to differentiate species based on clade and on photosynthetic type, both in clustering analyses and PCAs (Fig. 1 - 2; Supplementary Fig. S2), providing intrinsic support for our approach. That said, the total pool might not reflect the active pool involved in anabolic processes related to or following photosynthesis (Szecowka et al., 2013; Ma et al., 2014; Arrivault et al., 2017; Allen and

Young, 2020). To untangle active pool from pools that are not involved in photosynthesis, we performed 40-60 min $^{13}CO_2$ pulse labelling experiments in one basal C_3 species (*F. robusta*) and three species from clade A representing one C_3-C_4 intermediate (*F. ramosissima*), one C_4-like (*F. vaginata*) and one true C_4 (*F. bidentis*) species. This allowed us, for a subset of the metabolites, to estimate the size of the active pools involved in photosynthesis. In general, absolute and active pools showed similar trends; however in some cases (malate, 2OG and glutamate) the active pools were noticeably smaller than the total absolute pools, as previously shown for other species (Hasunuma et al., 2010; Szecowka et al., 2013; Weissmann et al., 2016; Arrivault et al., 2017).

Overall, our results provide new insights into the evolutionary progression from C_3 though C_3-C_4, to C_4-like and complete C_4 photosynthesis. In particular, we asked which metabolites show changes in their levels in the transitions between photosynthesis types, and if these match the changes that are expected as new reactions are co-opted to photosynthetic carbon fixation. We also asked if there are increases in the contents of metabolites that are involved in the predicted intercellular shuttles, as might be expected if intercellular movement occurs by diffusion driven by intercellular concentration gradients. The follow up discussion proceeds by tackling each evolutionary transitions in a dedicated paragraph.

From C_3 to C_3-C_4 intermediacy

The two C_3 *Flaveria* in our study were slightly different in terms of anatomy: *F. cronquistii* has the classical C_3 anatomy with abundant palisade mesophyll and inconspicuous BS cells, while *F. robusta* has the so called Proto-Kranz anatomy, with slightly larger BS cell and richer organelle content within those (Sage et al., 2013). Despite these anatomical differences, they showed similar metabolic profiles (Fig. 1 – 2), and the same was true for the C_3-C_4 Clade A *F. ramosissima* which harbors a completely developed Kranz anatomy. At a molecular levels, the main transition from C_3 to C_3-C_4 in *Flaveria* involves a restriction of GDCH activity to the BS cells and the associated establishment of a photorespiratory carbon pump (hence C_3-C_4 photosynthesis is also termed C_2 photosynthesis; Schulze et al., 2013). Following the formation of 2PG by the oxygenation reaction of Rubisco in the mesophyll, it is proposed that glycine moves to the BS cells where it is decarboxylated to generate an elevated concentration of CO_2, which diffuses back to the mesophyll. The serine formed by GDC activity moves back to

the mesophyll. The NH_3 released during GDC activity is assimilated by the GOGAT pathway in the BS cells, and a further intercellular shuttle is required to return an amino acids to the mesophyll and recycle an organic acid to the BS (Rawsthorne et al., 1988b; Rawsthorne and Hylton, 1991; Gowik et al., 2011; Mallmann et al., 2014).

Our metabolite analysis reveals a noticeable increase in the absolute (both) and active (available only for *F. ramosissima*) pool of serine in C_3-C_4 intermediates compared to basal C_3 species (Fig. 3C – Supplementary Fig. S5). This is compatible with serine moving from the BS back to the mesophyll in C_3-C_4 photosynthesis. The metabolically active serine is quite high already in *F. robusta*, which could potentially suggest an imbalanced serine between mesophyll and BS cells in this species, or even an early establishment of photorespiratory shuttles; to confirm these hypotheses would require further labelling studies with a non-Proto-Kranz species, like *F. cronquistii*.

We expected to find an increase in glycine levels to support a fast diffusion from the mesophyll to the BS cells. However, glycine levels were relatively low in C_3 species and neither the active nor the total pool increased in C_3-C_4 species compared to C_3 species. This might be explained by the rapid turnover of glycine in C_3-C_4 *Flaveria* species (Keerberg et al., 2014), or by active transport of glycine, for example in the BS mitochondria. It is also possible that other metabolites that were not analyzed in our study participate to the C_2 shuttle, like glycolate.

F. ramosissima is an advanced C_3-C_4 species with several C_4-related enzymatic activities being higher than early C_3-C_4 intermediates, including PEPC and AspAT (Ku et al., 1983; Bauwe, 1984; Ku et al., 1991). Indeed, it has been reported that it assimilates up to 50% of its carbon through PEPC (Moore et al., 1987; Chastain and Chollet, 1989). Slightly increased PEPC activity in the mesophyll could potentially allow a faster buildup of carbon skeletons and associated amino acids. Whilst the total malate pool even decreased between C_3 and C_3-C_4 species, the active pool showed a slight increase in C_3-C_4 species *F. ramosissima* compared to the C_3 species *F. robusta*. The total pool of aspartate did not change between C_3 and C_3-C_4 species, but the active pools was almost 3-fold higher in *F. ramosissima* than *F. robusta*. Alanine and pyruvate levels were low in C_3 species and in *F. ramosissima*. Our results indicate that aspartate could act as nitrogen equilibrator between BS and mesophyll cells; this actually fits with a C_3-C_4 photosynthesis model in which alanine and pyruvate movement between cells was constrained (Mallmann et al., 2014). A way in which aspartate could act as nitrogen equilibrator would be that *F. ramosissima* runs a partial NAD-ME cycle, where aspartate represent the shuttle

metabolite from mesophyll to BS, in parallel with C_3-C_4 photosynthesis: this latter hypothesis is supported by transcriptomic and enzymatic evidence in this specific *Flaveria* species by Gowik et al., (2011). Another supporting evidence on why aspartate was preferred than malate in C_3-C_4 intermediates is that malate inhibits the C_3-form of PEPC, which is the isoform present in C_3-C_4 species (Gowik and Westhoff, 2011b).

F. anomala, which is an early C_3-C_4 intermediate in clade B (Moore et al., 1987), contained higher absolute pool of 2PG, glycine and glycerate than *F. ramosissima* and both C_3 species. This is consistent with movement of glycine in this clade B species, however definitive conclusions would require further labelling experiments to provide information about active pool sizes. *F. anomala* also contains an elevated pool of alanine, which in this species might be involved in nitrogen shuttling and might support the same theory of C_3-C_4 *Flaveria* species running a parallel C_4 pathway (see above for *F. ramosissima*).

To conclude, aspects of our data are consistent with the idea that aspartate and serine are involved in shuttling of carbon and nitrogen between mesophyll and BS cells. However, these two metabolites alone are insufficient to achieve a balanced exchange between the mesophyll and BS cells, so further carriers like glycine, malate, pyruvate, alanine, 3PGA, 2OG or glutamate are probably also involved. It is particularly striking that there is no clear increase of glycine between C_3 and the C_3-C_4 species *F. ramosissima*, raising the question whether earlier metabolites in the photorespiration pathway, like glycolate, also move from the mesophyll to the BS cells, which in turn would decrease the amount of amino groups that need to be returned to the mesophyll cells. Overall, our data indicate that that multiple metabolites may have been co-opted as carriers during the C_3 to C_3-C_4 intermediate transition, in which case the overall level of single metabolite pools would not need to increase sensibly. This fits with the overall metabolic closeness between these two photosynthetic subtypes, as seen in clustering or PCAs.

From C_3-C_4 intermediacy to C_4-like

In the genus *Flaveria* it is already proven that advanced C_3-C_4 species (like *F. ramosissima*) possess elements of a C_4-like CCM, with higher activity of PEPC, NADP-ME and PPDK than in early C_3-C_4 species (like *F. anomala*) (Ku et al., 1983). It is thought that the incomplete C_4 cycle initially supported C_3-C_4 intermediate photosynthesis, but ended up reducing photorespiration in advanced intermediates like *F. ramosissima*

(Rumpho et al., 1984) and eventually took over from the C_3-C_4 cycle (aka C_2 cycle) at a certain point in the evolutionary line to become the main route of carbon assimilation. This transition to C_4-like species most likely involved boosts in enzymatic activity in the later part of the cycle, especially PPDK (Ku et al., 1983), which is involved in the regeneration of PEP. In advanced C_3-C_4 species the latter is provided mainly by the limited capacity of glycolysis and pentose phosphate pathway (Monson and Moore, 1989).

Our metabolite data reveal a substantial change in the primary metabolome of C_4-like species compared to C_3-C_4 intermediates, which is clearly displayed in both the clustering and especially the PCA (Fig. 1 – 2 – Supplementary Fig. S2). C_4-like species differ from C_3-C_4 intermediate species by having larger pools of CCM-relevant malate, pyruvate and alanine, while generally sporting lower pools of photorespiratory intermediates like serine and glycerate. The analyses with *F. palmeri* confirmed that this involves changes in active pools (Figure 5; Supplementary Fig. S5). Malate represents a better forward shuttle metabolite than aspartate, because it can provide reducing power to the BS cells after being decarboxylated by NADP-ME, which can power reactions in the CBC downstream of Rubisco carboxylation. The switch from aspartate to malate most likely happened quickly after PEPC transitioned from a C_3-like form in C_3-C_4 intermediates to a C_4-form in C_4-like species (Engelmann et al., 2003).

Interestingly, there is an abrupt increase in asparagine in all C_4-like species compared to C_3 and C_3-C_4 intermediates, and these high levels are retained in complete C_4 species; asparagine represents a potential nitrogen-rich compounds for shuttling (Mallmann et al., 2014), possibly in combination with aspartate or malate. Labelling data would be required to provide direct evidence for its involvement.

Our labelling analyses on the C_4-like species *F. palmeri* suggest that, in this evolutionary step, the active pools of not only malate, but also other key C_4 metabolites (pyruvate, alanine) are higher than in C_3-C_4 intermediates; this strongly suggest that an expanded NADP-ME C_4 cycle with integration from aminotransferases (Furbank, 2011; Schlüter et al., 2018) is up and running. Active aspartate level in *F. palmeri* is ~25% below *F. ramosissima*, which is slightly lower, but likely still sufficient to support some intercellular movement of aspartate. The active pools of glycine, serine and glycerate are consistently smaller in *F. palmeri* than *F. ramosissima*, indicating both that actual C_4-like species do not employ a C_2 cycle anymore and that they benefit from lower photorespiration rates (Ku et al., 1991).

Summarizing, our results point to a major reorganization of metabolism between C_3-C_4 and C_4-like photosynthesis, including large increases in the pools of malate, pyruvate and alanine, retention of a large pool of aspartate that is consistent with the establishment of a C_4-like CCM, and decreased levels of photorespiratory metabolites consistent with a substantial decrease in photorespiration.

From C_4-like to complete C_4

In this last phase, it is thought that several aspects of C_4 photosynthesis were optimized, especially regarding compartmentation and enzymatic properties of key CCM enzymes, like Rubisco and PEPC (Kubien et al., 2008; Gowik and Westhoff, 2011b; Kapralov et al., 2011). Our metabolite analysis captures further changes that occur in parallel with optimization of enzymes.

These include an increase in the levels of several TCA cycle intermediates (citrate, isocitrate, aconitate, 2OG) in true C_4 species compared to C_4-like species (Fig. 1 – 2 – 3A – Supplementary Fig. S2 – S3). These findings, coupled with labelling data (Fig. 5 – Supplementary Fig. S4 – S5) in the C_4 plant *F. bidentis*, show that, although the total glutamate pool did not increase, the active pool of glutamate doubles compared to C_4-like *F. palmeri*, suggest increased flux from the TCA cycle to this amino acid. Interestingly 2OG and glutamate levels in *F. bidentis* were similar to those in the model NADP-ME C_4 species maize (Supplementary Fig. S5). One possible role for 2OG and glutamate in C_4 photosynthesis might be to support flux through aspartate aminotransferase or alanine aminotransferase in the mesophyll cells or the BS cells. That said, while glutamate active pools tend to increase in the transition from C_4-like to true C_4, the active pool of aspartate drops in *F. bidentis* five times compared to C_4-like *F. palmeri*. Our data also reveals a much smaller total pool of aspartate in the C_4 species *F. bidentis* compared to C_4-like *F. palmeri*. In the same comparison between C_4-like *F. palmeri* and true C_4 *F. bidentis*, only malate absolute level increase in the latter, however by less than 20%. These findings point toward a further shift from aspartate shuttling, picked up partially by increase shuttling toward malate in C_4 species. In fact, it is known that, for *F. bidentis*, the label in aspartate increases slower in a pulse with labelled CO_2 and decreases faster in a change to normal CO_2 than is the case for malate (Meister et al., 1996).

We also observed a small drop in the alanine active pool and 2-fold increase in the pyruvate active pool between *F. palmeri* and *F. bidentis*, as well as an increase in the total

pyruvate pool in the whole set of C_4 compared to C_4-like species. These observations might indicate a refinement of NADP-ME C_4 cycle toward one that is less reliant on aminotransferase reactions. However, these changes would also be explained by a more efficient usage of aspartate and alanine in the shuttles, which is consistent with our suggestion that higher 2OG and Glu may facilitate fluxes through the respective aminotransferases.

Another interesting change in the transition from C_4-like to C_4 concerns glycerate: absolute pools were generally higher in C_4 species than in C_4-like species, and there is an almost 7-fold increase in active pools from C_4-like *F. palmeri* to C_4 *F. bidentis*. It is difficult to foresee what might be the role of glycerate in C_4 species, but it should have different roles from that in C_3 and C_3-C_4 intermediate, since photorespiration is severely reduced in C_4 plants. One possibility is that glycerate acts as a carbon reserve that can be used via glycerate kinase to replenish the levels of CBC and CCM metabolites (Usuda and Edwards, 1980; Weber and von Caemmerer, 2010), for example during transitions from different carbon availability statuses (Levey et al., 2019).

Overall, our results indicate that, in addition to optimization of the characteristics of enzymes, the final stages in the optimization of C_4 photosynthesis also involved shifts towards increased use of malate/pyruvate exchange over aspartate/alanine exchange in the CCM. At the same time increased levels of 2OG and glutamate might aid aminotransferase in one or both cell types, and increased levels of TCA acids and glycerate which might serve as a reservoir of carbon to buffer C_4 metabolism against fluctuations in the environment.

Clade-dependent metabolite differences and considerations

As already mentioned, our study included nine species, which are divided in three phylogenetic groups: two basal (*F. cronquistii* and *F. robusta*), five from clade A (*F. ramosissima*, *F. palmeri*, *F. vaginata*, *F. trinervia* and *F. bidentis*) and two from clade B (*F. anomala* and *F. brownii*) species. We included almost all existent species of the basal *Flaveria* group; *F. pringlei* was not considered available due to European seed stock resulted contaminated with hybrids between a C_3 and a C_3-C_4 intermediate species (Kopriva et al., 1996). Our inclusion of species was skewed toward clade A species, which according to many authors was the earlier clade in which C_4 photosynthesis evolved within *Flaveria* genus (McKown et al., 2005; Lyu et al., 2015). Our coverage of

clade B species, which most likely evolved C_4 photosynthesis a second time within the genus, was very incomplete and this represents a limitation of our study. Nevertheless, even if species distribution was uneven within the genus, some subtle differences emerged between the two clades.

Absolute metabolite amounts are able to separate clade B species *F. anomala* and *F. brownii*, from the corresponding photosynthetic types in clade A, both in cluster analysis and in PC analysis (Fig. 1 - 2). This separation is mainly driven by differences in absolute pools of sugar phosphates (F6P, FBP, SBP, R5P, G1P) and several amino acids, some of which containing a substantial amount of carbon (glutamine, alanine, asparagine, aspartate and serine; for carbon impact see Fig. 5). A similar trend is seen for sugar phosphates, but with decreased magnitude, for *F. brownii* and the two C_4-like species in clade A. However, further studies are needed to understand the significance of the clade-dependent changes in CBC metabolites.

The observation of clade differences for C_4-related amino acids might point to slightly different carbon assimilation patterns. However, both clade B species in this study are considered early intermediates in their respective photosynthetic categories (Moore et al., 1987; Ku et al., 1991), which might partly explain these metabolic differences. Moreover, *F. brownii* has an interesting photosynthetic flexibility, which dampens C_4 photosynthesis at lower light intensities (Cheng et al., 1989); our growth conditions were compatible with the low light condition used in this earlier study, which might explain part of the metabolite differences with other C_4-like species.

Summarizing, our metabolic data suggest there are differences between how CCM operates in clade A and B *Flaveria* species. That said, they do not confirm or disprove the idea that there was independent evolution of C_4 photosynthesis; extensive molecular phylogeny on key CCM enzymes could prove more useful for answering this aspect.

Changes in CBC operation along the evolutionary trajectory from C_3 to C_4 photosynthesis

It is well established that CBC operation differs between C_3 and C_4 species, for example due to the relaxation of selection pressure on the selectivity factor and the increase in k_{cat} of Rubisco (Wessinger et al., 1989; Sage, 2002; Kubien et al., 2008). Recently, profiling of CBC metabolites revealed more widespread changes in CBC metabolite levels between different C_4 species, between C_4 and C_3 species, and between different C_3 species

(Arrivault et al., 2019; Borghi et al., 2019). Such studies point on the one hand to more widespread adjustment of the CBC during the evolution of C_4 photosynthesis, and also to inter-species diversity in how the CBC operates in different C_3 species. The former is relevant to current attempts to engineer C_4 photosynthesis into C_3 crops, and the latter to the exploitation of photosynthetic diversity for improving crops (Lawson et al., 2012; Driever et al., 2014; Simkin et al., 2019).

In order to get a specific view on how operation of the CBC was modified during the evolutionary progression from C_3 to C_4 photosynthesis in the *Flaveria* genus, we compared CBC metabolite profiles from the nine studied *Flaveria* species with previously published data for several C_3 and C_4 species from outside the genus. To do this we focused on a 'dimensionless' dataset in which each metabolite normalized on the total carbon in all CBC metabolites (Arrivault et al., 2019). This allowed comparison across species even though their leaf composition might differ.

In the resulting PCAs (Fig. 6) showed that it is possible to separate plants with differing photosynthetic strategies by just looking at CBC metabolites. The C_3 *Flaveria* species have a CBC profile close to other C_3 species, especially rice and wheat. Intermediates *Flaveria* species lie on the trajectory between C_3 and C_4 species, with C_3-C_4 intermediate species being closer to C_3 species and C_4-like species closer to true C_4 species. Further, diversity of clade B *Flaveria* species can be observed in the case of *F. brownii*, probably suggesting that the CBC operates differently in this species, indirectly supporting the different origin of C_4 CCM in the clade A and B. Furthermore, and more generally, these results indicate that CBC operation has most likely started to adapt to enhanced internal CO_2 status in the C_3-C_4 *Flaveria* ancestor species (shared by both clades) and that this adaptation has progressed furthest in C_4-like and C_4 species in clade A. Further, it appears that this modification of CBC operation occurs along an axis that is shared between C_4 species in three phylogenies and C_3 species in multiple phylogenies.

Separation along the transition from C_3 to C_4 photosynthesis was mainly in PC1 and was driven by increasing 3PGA and decreasing levels of 2PG and several CBC intermediates, like S7P, F6P, SBP, FBP and, especially, RuBP. In PC2, species were separated by 2PG, S7P and R5P in a more phylogeny-dependent way (including all *Flaveria* C_3-C_4 intermediate and C_4-like species, except *F. brownii*), and DHAP, SBP and FBP (followed mainly by *Setaria viridis*, and secondarily by *Zea mays*, *Nicotiana tabacum* and *Flaveria brownii*). The association of high 2PG with C_3 or C_3-C_4 intermediates likely reflects the higher rates of photorespiration in these species. C_4 photosynthesis is associated with

lower RuBP contents. Indeed, we observe it directly in *Flaveria* species, where RuBP is highest in C_3 and C_3-C_4 species, and decreases through C_4-like to C_4 species (Fig. 3B). Evolution of C_4 photosynthesis is accompanied by a decrease in Rubisco abundance, made possible by relaxed selection for catalytic fidelity for CO_2 and an associated increase in k_{cat} (Sage, 2002; Ghannoum, 2005; Kapralov et al., 2011). As most RuBP is bound in the active site of Rubisco (Salvucci, 1989), the lower RuBP may in part reflect the lower Rubisco abundance.

The higher 3PGA concentrations in most C_4-like and C_4 *Flaveria* species (and partially *Zea mays*) can be associated to improved Rubisco carboxylation over oxygenation rates, but also with carbon shuttles between mesophyll and BS cells, together with DHAP (Leegood, 1985; Stitt and Heldt, 1985; Arrivault et al., 2017). DHAP is downstream of 3PGA in the CBC, and its levels are representative of triose-phosphates, due to rapid equilibration between DHAP and GAP (Arrivault et al., 2017). It is interesting to notice that high levels of DHAP are characteristic for *Setaria viridis* and *Zea mays* alone, while *Flaveria* species have lower level that are only marginally above those in C_3 *Flaveria* species; this could mean that C_4 *Flaveria* have probably only a limited 3PGA/triose phosphate shuttle (Leegood and von Caemmerer, 1994). This shuttle is associated to NADP-ME C_4 species, like *Zea mays*, that have dimorphic chloroplast and little or no PSII activity in the BS cells (Woo et al., 1970; Andersen et al., 1972; Nakajima Munekage, 2016), while both C_4 *Flaveria* species used in this study still partially retain these features in the BS (Höfer et al., 1992; Nakamura et al., 2013). It still unproven if *Setaria viridis* makes use of this shuttle and has the same chloroplastic properties as *Zea mays*, but our metabolite data suggest this is the case for this species. That said, weak non-significant trend to high 3PGA (see also Fig. 3B) in *Flaveria* C_4 and C_4-like species might reflect restricted energy for 3PGA reduction in the BS cells, and be a first step towards establishment of an intercellular 3PGA/triose phosphates shuttle.

General conclusions

In this study, we carried out a comparative analysis of metabolite levels in nine species along the evolutionary trajectory from C_3 to C_4 photosynthesis in the *Flaveria* genus. Our results are broadly consistent with current ideas that C_4 photosynthesis evolved progressively; starting with preferential location of GDC activity to the BS cells and the associated appearance of intercellular shuttles of photorespiration metabolites, allowing

the C_2 CCM to run. At the same time, shuttles to balance nitrogen between the mesophyll and BS cells emerged, possibly accompanied and certainly followed by the rising role of PEPC in the mesophyll and partial relocation of Rubisco to the BS. The increased need for reducing power in the BS, coupled with the emergence of a non-wasteful CCM from the co-opted nitrogen-balancing system, led to the establishment of a C_4-like CCM to concentrate CO_2 in the BS cells, without the burden of photorespiration. This last step was followed by optimization to reach a status of full C_4 photosynthesis.

The high levels of serine we found in C_3-C_4 *Flaveria* species support the idea that this metabolite moves back from the BS to the mesophyll cells, but we did not see substantial increase in the levels of other metabolites that are potentially involved in intercellular shuttles, indicating that most intercellular fluxes may be partitioned between two or more metabolites. The transition from C_3-C_4 to C_4-like was associated with clear increases of malate and most other metabolites that are required to shuttle CO_2 into the BS cells. The transition from C_4-like to C_4 was associated with changes in metabolites indicative of a further shift in the balance between malate/pyruvate and aspartate/alanine shuttles, together with improved operation of aminotransferase reactions that may support intercellular shuttles and an increased carbon reservoir to buffer C_4 metabolism against fluctuations in the environment. In addition, focused analysis of CBC metabolites revealed progressive adaptation of the CBC as the CCM became increasingly effective.

SUPPLEMENTARY DATA

- Supplementary Fig. S1 - Phylogenetic tree of the 16 *Flaveria* species
- Supplementary Fig. S2 – Principal component (PC) analyses of metabolite profiles in nine *Flaveria* species: PC1 combination with PC3 or PC4.
- Supplementary Fig. S3 – Absolute amounts of remaining metabolites in nine Flaveria species.
- Supplementary Fig. S4 – Summary of ^{13}C labelling in key metabolites in four Flaveria species.
- Supplementary Fig. S5 – Estimated size of the active pool of nine metabolites potentially involved in CCMs, transfer of amino groups and/or photorespiration in four *Flaveria* species + *Zea mays.*
- Supplementary Table S1 – List of all measured metabolites, their abbreviations and analytical methods used to detect them.
- Supplementary Table S2 – Summaries of first 10 PCs for Principal Component Analyses on *Flaveria*-only whole dataset.
- Supplementary Table S3 – Use of ^{13}C labelling data to estimate active and inactive pools in four *Flaveria* species, and comparison with *Zea mays*, a model NADP-ME C_4 species.
- Supplementary Table S4 – Summaries of first 10 PCs for Principal Component Analyses on dimensionless multispecies dataset (CBC+2PG only).
- Supplementary Data file S1 – Absolute amounts of all metabolites measured in nine *Flaveria* species, expressed on FW and as a dimensionless data sets.
- Supplementary Data file S2 – ^{13}C enrichment in a selection of metabolites in four *Flaveria* species, expressed as % over the total ^{13}C-enrichment.
- Supplementary Data file S3 – Absolute amounts of CBC + 2PG metabolites measured in nine *Flaveria* species combined with a subset of those in Arrivault et al. (2019), expressed on FW and dimensionless data sets.
- Supplementary Data file S4 – Germination, growing and harvesting conditions for absolute metabolite and $^{13}CO_2$ labelling harvests.

ACKNOWLEDGEMENTS

We wish to thank Christin Abel and Sandra Stegemann (MPI-MP) for the *Flaveria* cultivation and propagation help and advices. Many thanks go to Hirofumi Ishihara (MPI-MP) for helping during the $^{13}CO_2$ labelling system setup, and to lots of labelling helping hands from AG Stitt at the MPI-MP: Jana Verbančič, Thiago Moraes, Rubén Vicente, Haim Trebes, Armin Schlereth, Virginie Mengin, Sandra Kerbler and Chiara Baccolini. This work was financially supported by the Max Planck Society (G.L.B., S.A., D.B.M., A.R.F., J.E.L., M.S.), Bill & Melinda Gates Foundation (G.L.B., S.A., J.E.L., M.S.), the FullThrottle project (031B0205B) sponsored by the German Federal Ministry of Education and Research (S.A., D.B.M., A.R.F., M.S.), the VALERE grant from the Università degli Studi della Campania Luigi Vanvitelli (E.D.A., G.M.F., P.C.) and by the Australian Research Council Discovery Project Grant DP150101037 (M.L., J.E.L., M.S.).

MANUSCRIPT THREE – MAIN FIGURES

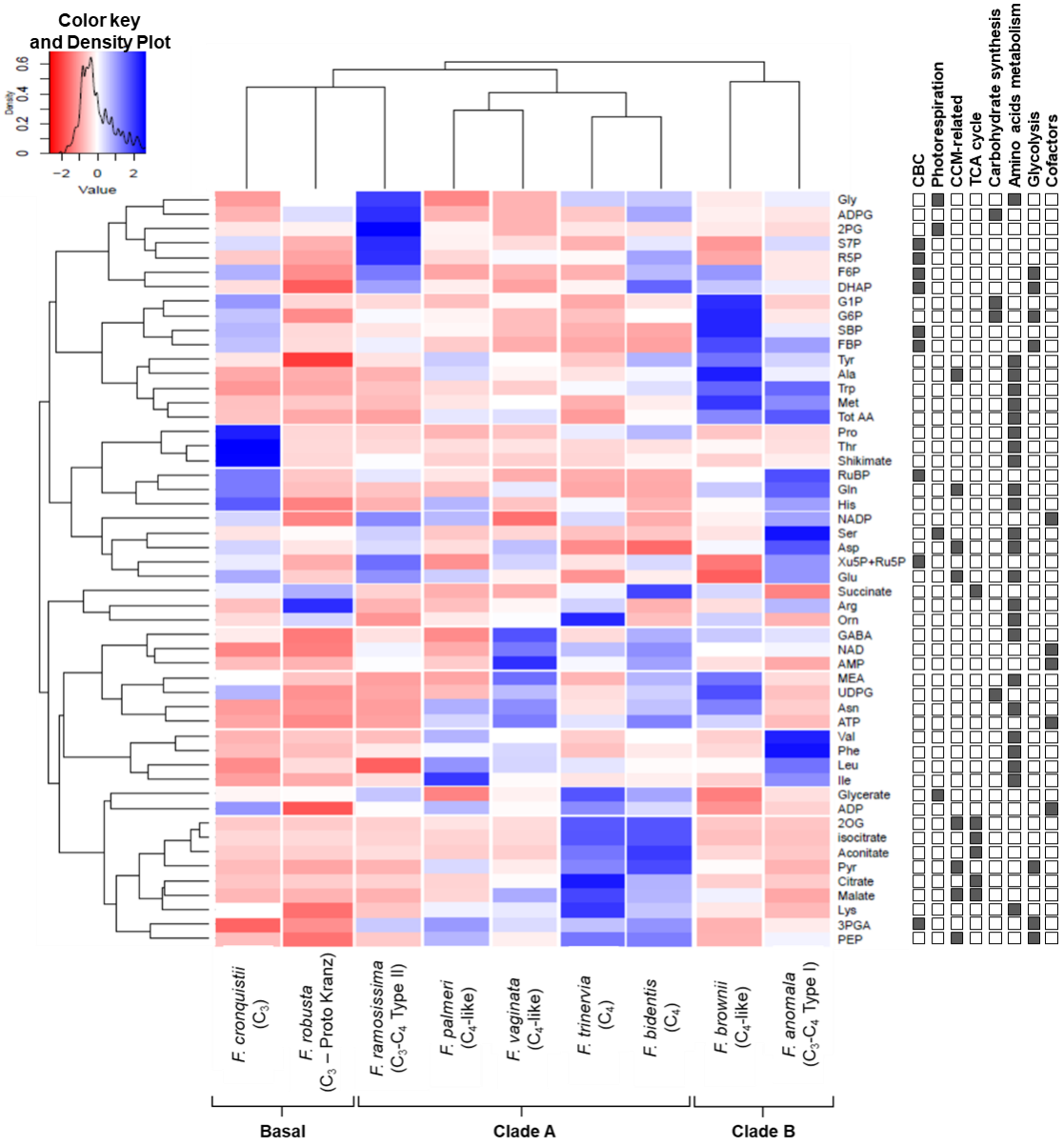


Fig. 1 – Heatmap and dendrograms of absolute metabolite levels in nine *Flaveria* species. The heatmap was generated by two-way clustering: each column represents a *Flaveria* species, and each row represents a metabolite. The *Flaveria* species is indicated at the bottom of each column, as well as its photosynthesis mode and the taxonomical grouping, according to current knowledge. Species (top) and metabolite (right hand side) dendrograms were calculated using a Z-scores clustering. Data were Z-scored by expressing the average level (of 4 biological replicates) in a given species as a fraction of the average across all species. The heatmap cell color intensity represents the average Z-score for the metabolite in that species. The color key and data distribution plot for the heat map are shown in the insert in the upper-left corner. For metabolite abbreviations, refer to Supplementary Table S1. Each biological replicate was obtained by harvesting several newly fully expanded leaves from one *Flaveria* plant. Detailed growth and harvest conditions can be found in the Material & Methods section and Supplementary Data file S4. The right-hand sub-panel indicates which sector of metabolism a given metabolite is involved in, as indicated by the grey box. The metabolic sectors are indicated at the top of the sub-panel. In some cases, a metabolite is assigned to more than one sector. For the raw data used for creating this figure, refer to Supplementary Data file S1.

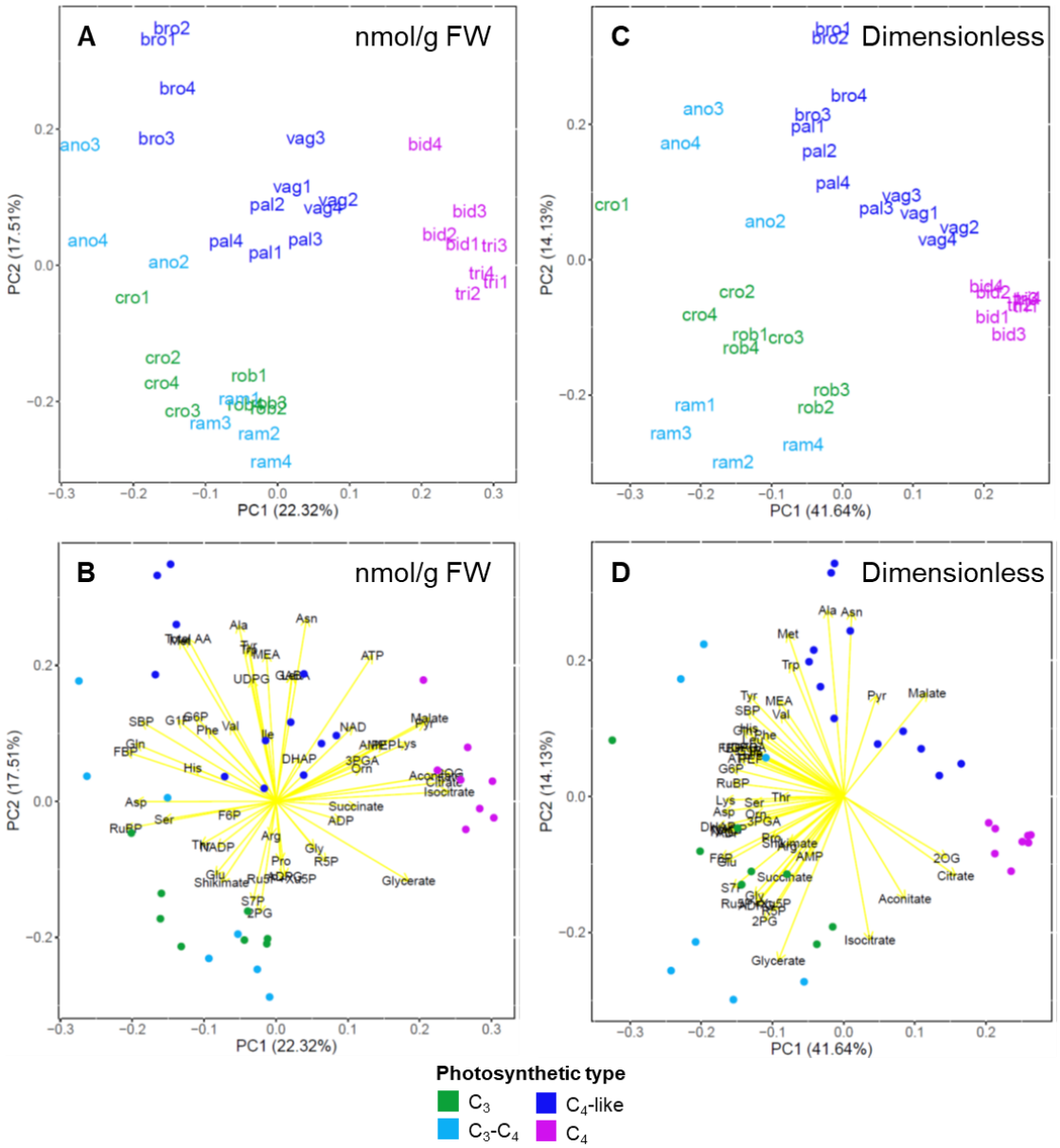


Fig. 2 – **Principal component (PC) analyses of metabolite profiles in nine *Flaveria* species: PC1 and PC2 visualization**. Plots A and B show PC1 and PC2 of the FW-normalized dataset after Z-scoring, while C and D plots show the same PCs using the dimensionless dataset. The FW-normalized data set was transformed into a dimensionless data set using the following procedure: I) in a given sample, the level of each metabolite was transformed to carbon equivalent values by multiplying the absolute amount by the number of carbon atoms in a molecule of that metabolite; II) the carbon equivalent amounts of all metabolites measured were summed for that sample; III) the carbon equivalent value of a given metabolite was divided by the summed carbon equivalent value, giving the % of total carbon in that metabolite in that sample. (A, C) Distribution of samples along the two first two principal components, with each sample being represented by a colored label indicating the species and biological replicate number. (B, D) Metabolite eigenvectors driving sample separation are shown in yellow, while single samples appear as colored dots. Colors represent the different photosynthetic types, as indicated by the legend on the right. Species abbreviations are, alphabetically: ano, *F. anomala*; bid, *F. bidentis*; bro, *F. brownii*; cro, *F. cronquistii*; pal, *F. palmeri*; ram, *F. ramosissima*; rob, *F. robusta*; tri, *F. trinervia*; vag, *F. vaginata*. For metabolite abbreviations, refer to Supplementary Table S1. Detailed growth and harvest conditions can be found in the Material & Methods section and Supplementary Data file S4. For the raw data used for creating this figure, refer to Supplementary Data file S1.

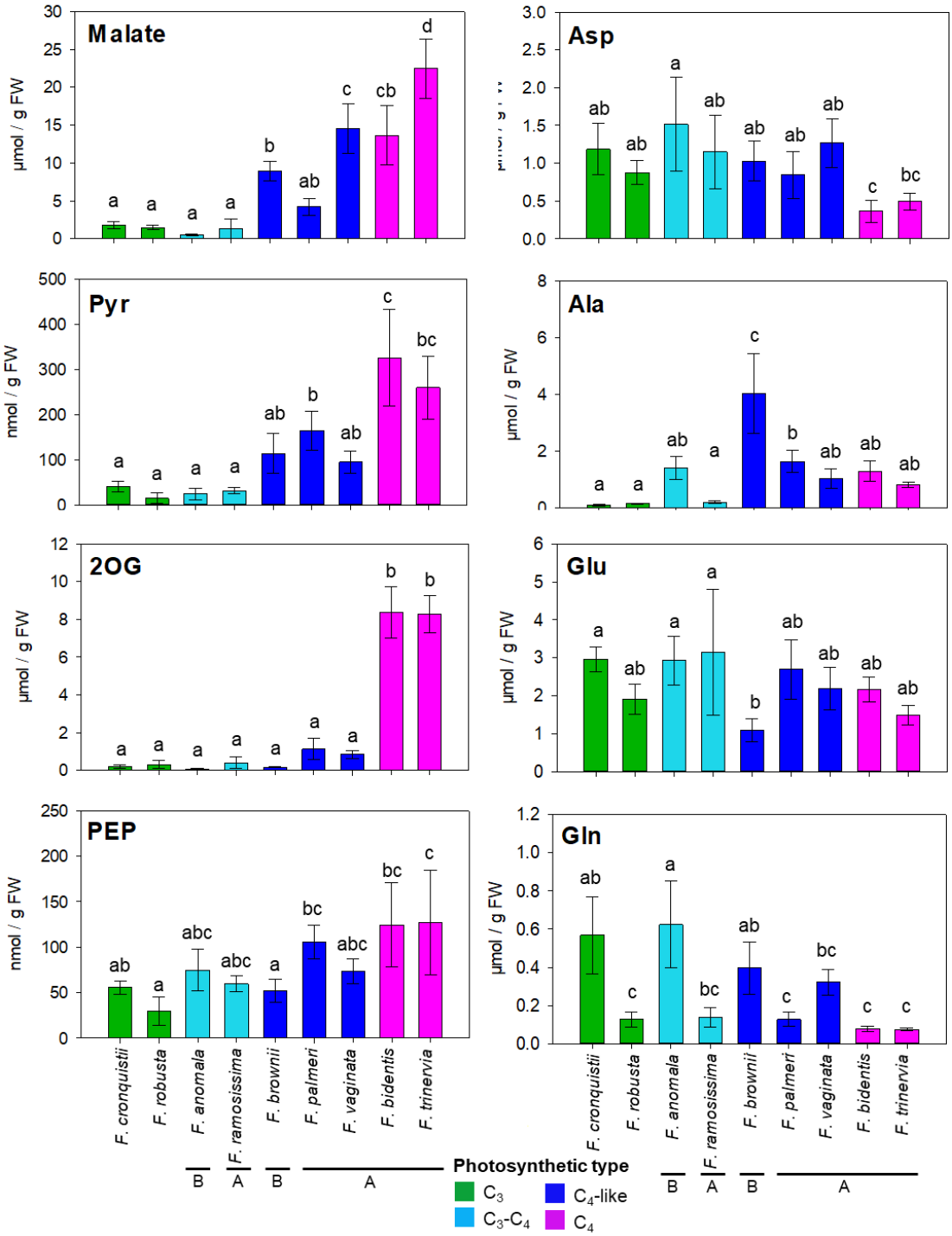


Fig. 3A – Absolute amounts of metabolites in nine *Flaveria* species. The color of each bar represents the photosynthetic type of the corresponding species, as indicated by the color legend at the bottom of each panel. In the sets of C_3-C_4 and C_4-like species, the clade B species (*F.anomala* and *F. brownii,* respectively) are placed to the left of the clade A species; clade is also indicated as 'A' or 'B' below the species name. **(**A) Absolute amounts of CCM-related metabolites. The amounts are plotted as average (nmol/g FW or µmol/g FW) ± SD (n=4). Each biological replicate was obtained by harvesting several newly fully expanded leaves from one *Flaveria* plant. Letters above each bar represent post-hoc pairwise comparison grouping (Holm-Sidak method, $p<0.05$). For metabolite abbreviations, refer to Supplementary Table S1. Detailed growth and harvest conditions can be found in the Material & Methods section and Supplementary Data file S4. For the raw data used for creating this figure, refer to Supplementary Data file S1.

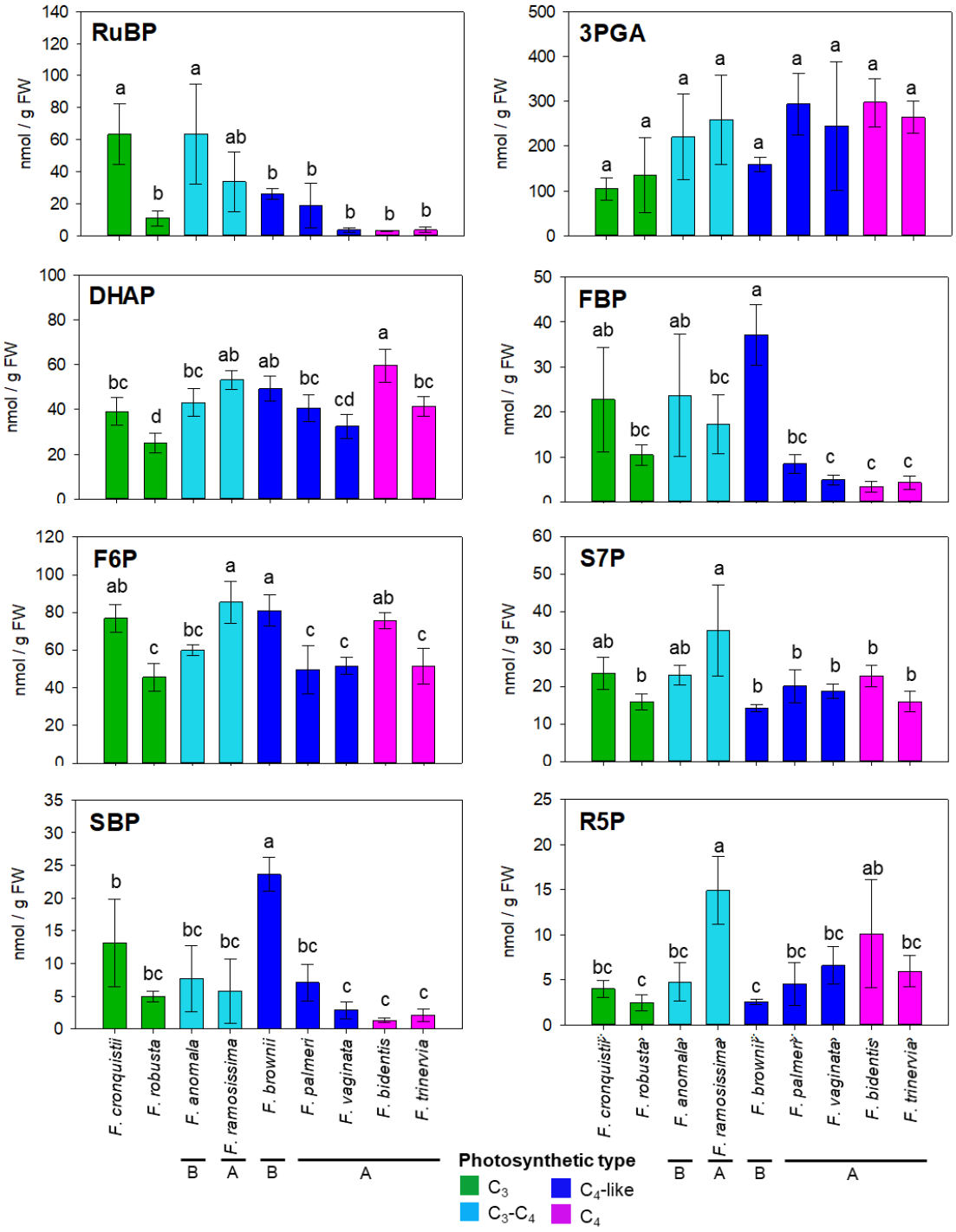


Fig. 3B – Absolute amounts of metabolites in nine *Flaveria* species. The color of each bar represents the photosynthetic type of the corresponding species, as indicated by the color legend at the bottom of each panel. In the sets of C_3-C_4 and C_4-like species, the clade B species (*F. anomala* and *F. brownii,* respectively) are placed to the left of the clade A species; clade is also indicated as 'A' or 'B' below the species name. (B) Absolute amounts of selected CBC metabolites.The amounts are plotted as average (nmol/g FW or µmol/g FW) ± SD (n=4). Each biological replicate was obtained by harvesting several newly fully expanded leaves from one *Flaveria* plant. Letters above each bar represent post-hoc pairwise comparison grouping (Holm-Sidak method, $p<0.05$). For metabolite abbreviations, refer to Supplementary Table S1. Detailed growth and harvest conditions can be found in the Material & Methods section and Supplementary Data file S4. For the raw data used for creating this figure, refer to Supplementary Data file S1.

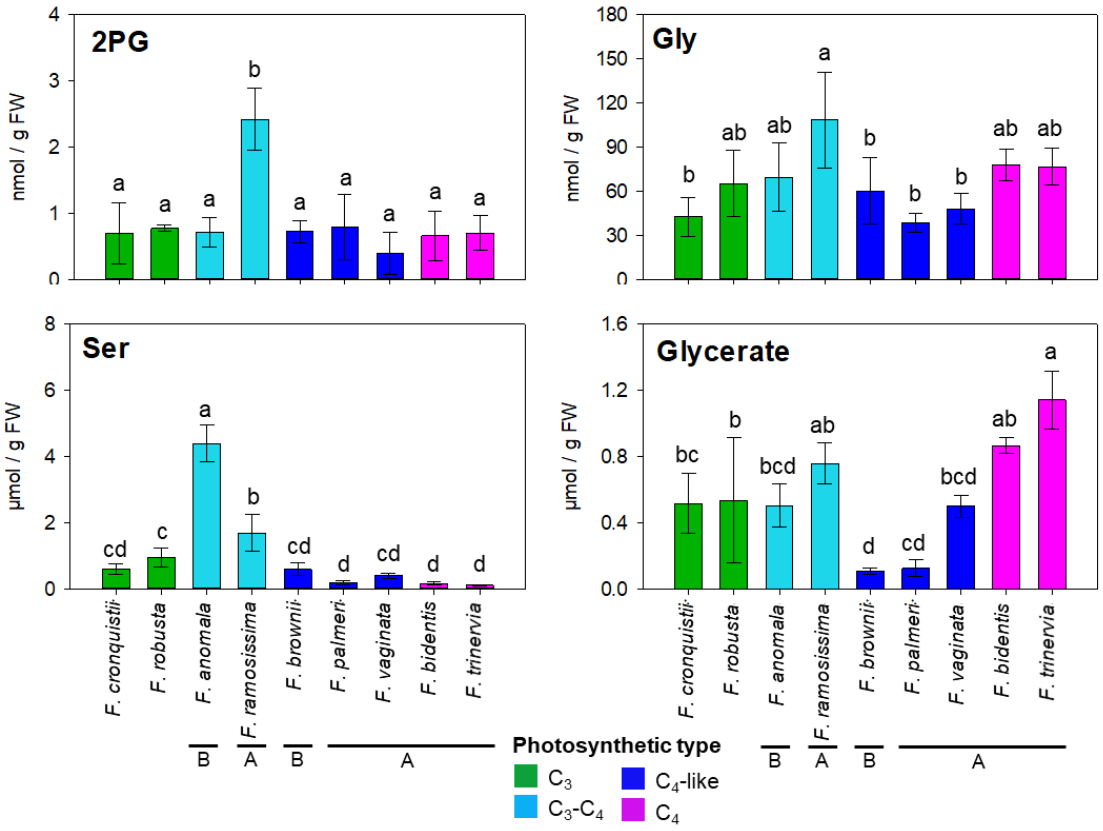


Fig. 3C – Absolute amounts of metabolites in nine *Flaveria* species. The color of each bar represents the photosynthetic type of the corresponding species, as indicated by the color legend at the bottom of each panel. In the sets of C_3-C_4 and C_4-like species, the clade B species (*F. anomala* and *F. brownii,* respectively) are placed to the left of the clade A species; clade is also indicated as 'A' or 'B' below the species name. **(**C) Absolute amounts of photorespiratory metabolites**.** The amounts are plotted as average (nmol/g FW or µmol/g FW) ± SD (n=4). Each biological replicate was obtained by harvesting several newly fully-expanded leaves from one *Flaveria* plant. Letters above each bar represent post-hoc pairwise comparison grouping (Holm-Sidak method, $p<0.05$). For metabolite abbreviations, refer to Supplementary Table S1. Detailed growth and harvest conditions can be found in the Material & Methods section and Supplementary Data file S4. For the raw data used for creating this figure, refer to Supplementary Data file S1.

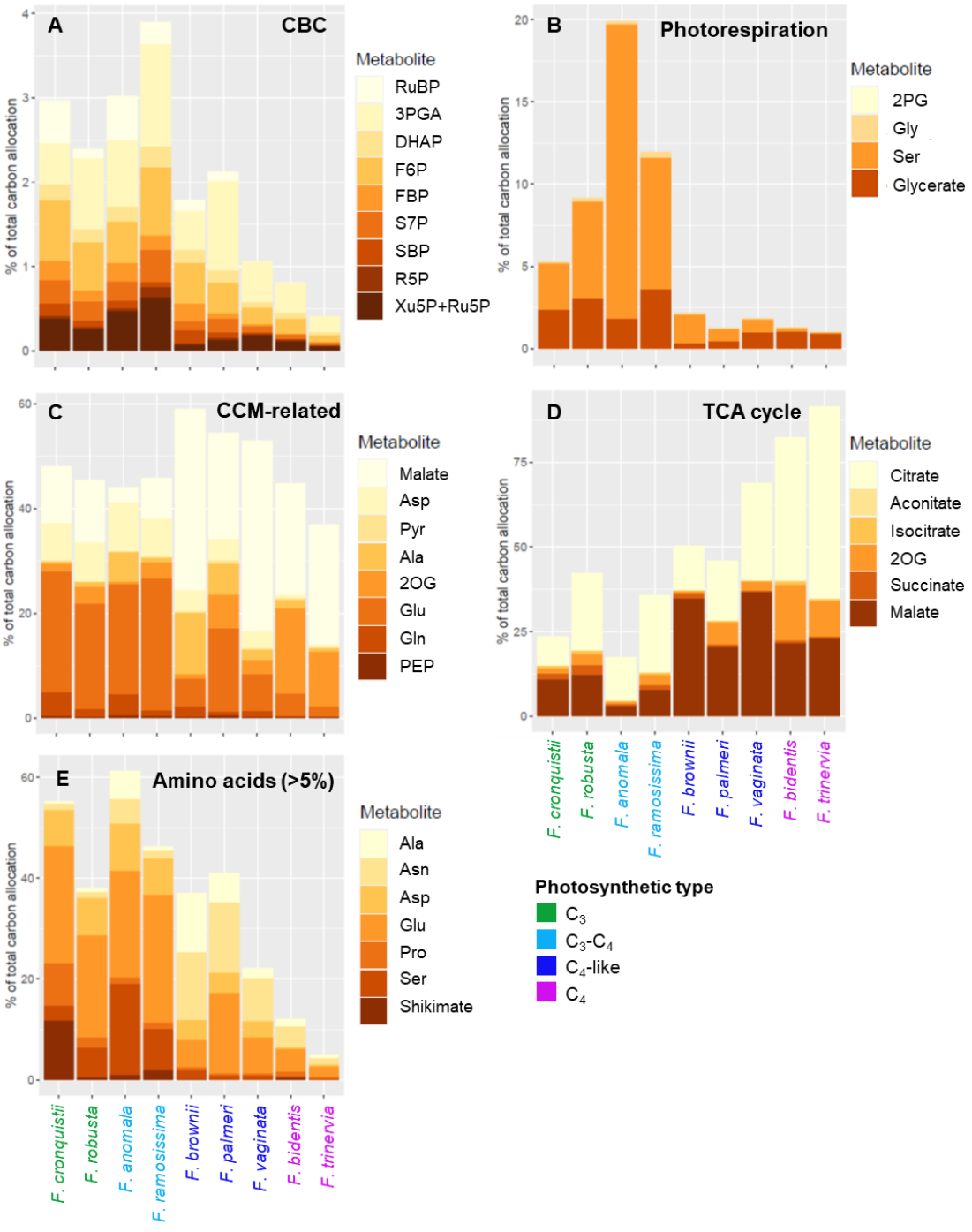


Fig. 4 – **Carbon allocation in nine *Flaveria* species** – Representation of the fraction of total carbon allocated to selected metabolites, grouped by biochemical pathway. Carbon allocation was calculated by transforming the FW-normalized data set into a dimensionless data set using the procedure described in Fig. 2. The % of total carbon in a given metabolite is depicted in the stacked bar diagrams as average (n=4); metabolites are ordered according to their sequence in the metabolic pathway or sector. Metabolite can appear multiple times in different panels. The color assigned to each metabolite is indicated in the insert of each panel: (A) CBC intermediates; (B) Photorespiration intermediates; (C) CCM-related intermediates; (D) TCA cycle intermediates; (E) amino acids (only amino acids corresponding to more than 5% of the total carbon pool). For metabolite abbreviations refer to Supplementary Table S1. The color of each species' name represents the corresponding photosynthetic type (see "Photosynthetic type" legend at the bottom right). Detailed growth and harvest conditions can be found in the Material & Methods section and Supplementary Data file S4. For the raw data used for creating this figure, refer to Supplementary Data file S1.

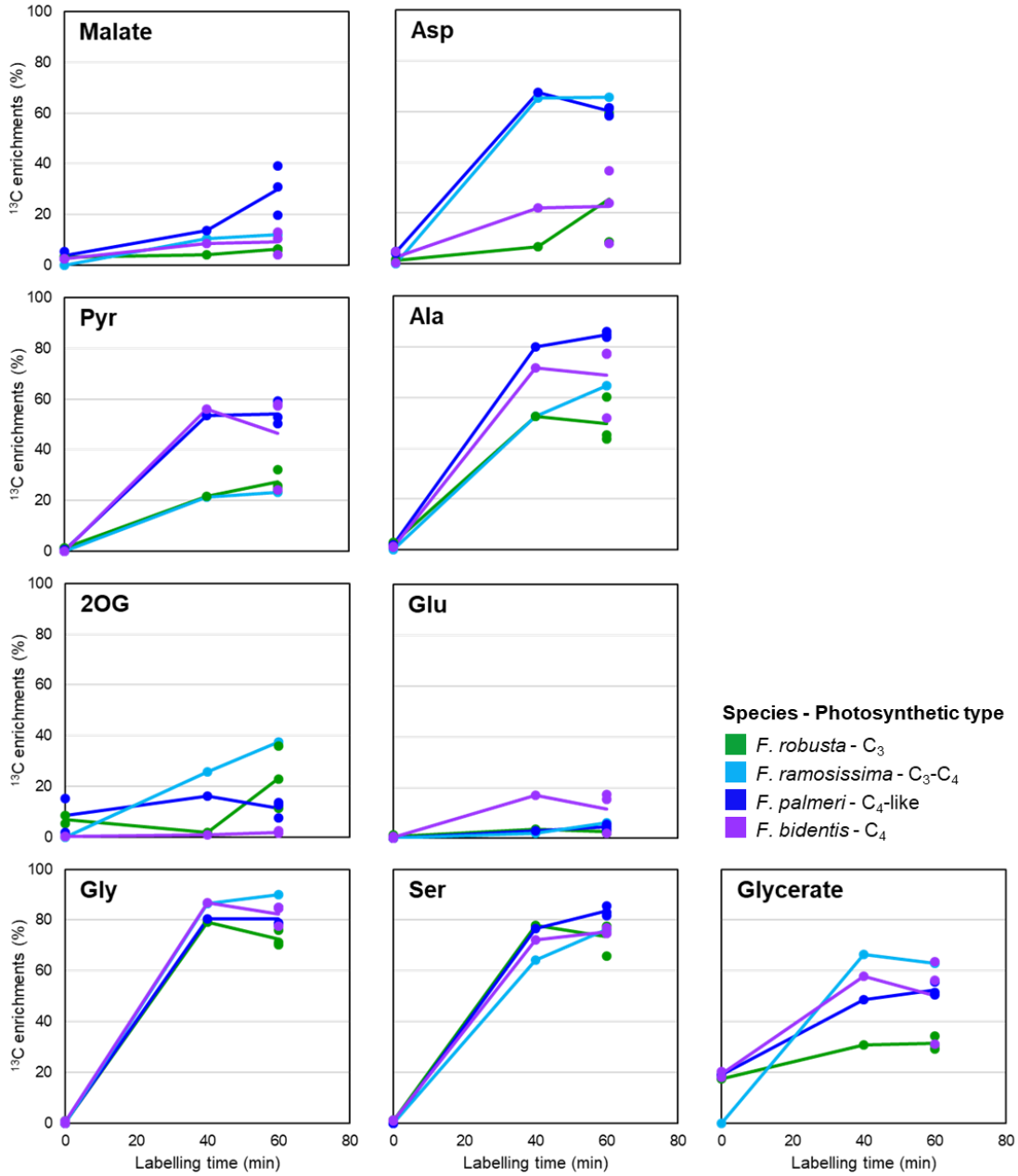


Fig. 5 – Time course of ^{13}C enrichment in key metabolites in four *Flaveria* species with different modes of photosynthesis. *F. robusta* (C_3), *F. ramosissima* (C_3-C_4), *F. palmeri* (C_4-like) and *F. bidentis* (C_4) plants were encased in a transparent labelling chamber, acclimated for 2 minutes in a ~420 ppm $^{12}CO_2$ atmosphere (+ 21% O_2, 78% N_2) and then switched to a ~420 ppm $^{13}CO_2$ atmosphere (+ 21% O_2, 78% N_2), left under growth irradiance and harvested after 40 min or 60 min for analysis by GC-MS. Controls (0 min) were harvested at the end of the acclimation phase. ^{13}C enrichments were calculated from the relative abundance of individual isotopomers, ad detailed in Material and Methods. The density of sampling was restricted by availability of plant material, with n=2 samples for 0 min, n=1 for 40 min and n=3 for 60 min except for *F. ramosissima*, were only 40 and 60 min pulses were performed (both n=1), and the unlabeled control set to 0% for graphing purposes. In the plots, each point represents a single sample, while the line represents the average trend. The color of the dot indicates the corresponding photosynthetic type (color coded according to the bottom right legend). For metabolite abbreviations refer to Supplementary Table S1. Growth conditions and labelling details are found in Material & Methods section and Supplementary Data file S4. For the raw data used for creating this figure, refer to Supplementary Data file S2.

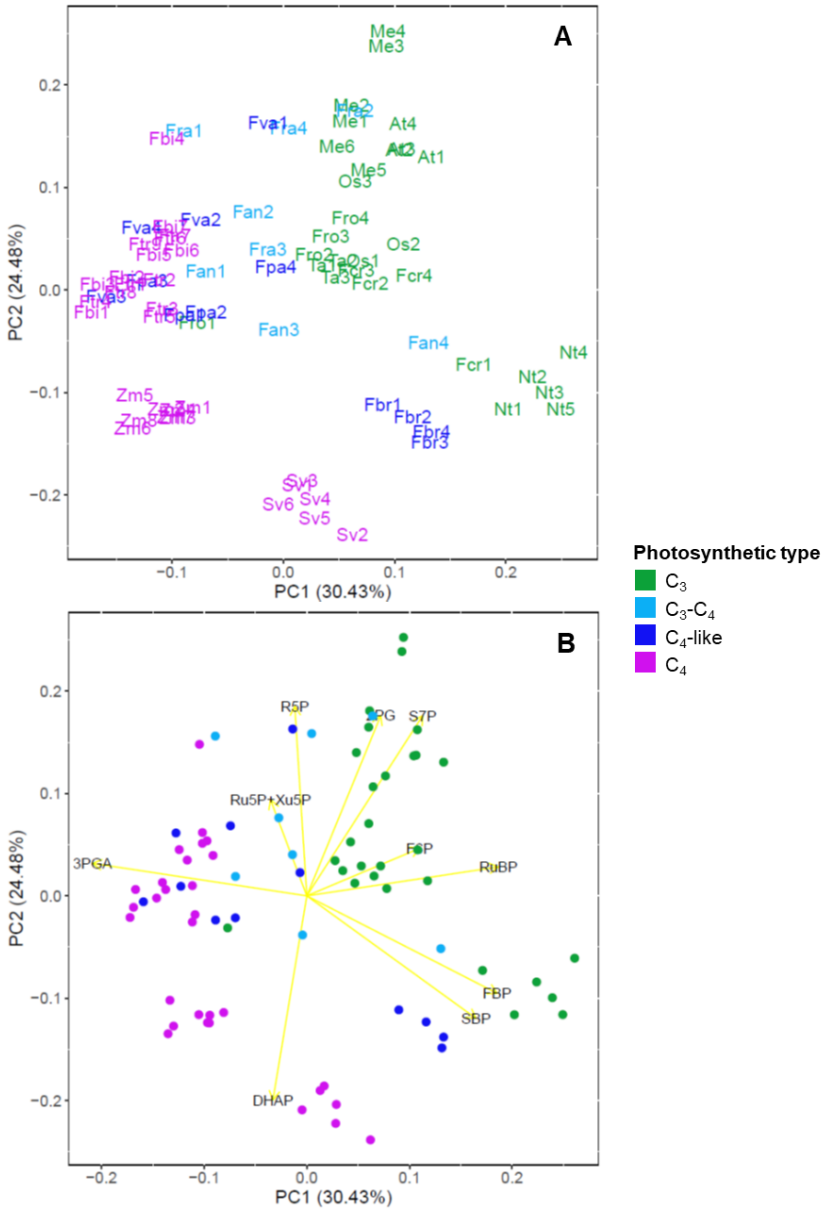


Fig. 6 – Principal component (PC) analysis on levels of CBC metabolites and 2PG, combining data from the current study of nine *Flaveria* with data for five further C_3 species and two further NADP-ME subtype C_4 species. PC analysis was performed on dimensionless normalized datasets, combining data from this paper and from Arrivault et al. (2019). The dimensionless transformation was carried out in an analogous manner to the described in the legend of Fig. 2, except that the data set contained only CBC metabolites and 2PG, and precisely as described in Arrivault et al. (2019). (A) Distribution of samples in the Euclidean space created by PC1 and PC2 with each sample being represented by a colored label indicating the species and replicate number. (B) Metabolite eigenvectors driving samples separation; here samples are depicted as colored dots. In both panels, the color of the dot and label indicates the photosynthetic types, as indicated by the legend on the right. Species abbreviations (and photosynthesis mode) are, alphabetically: At, *Arabidopsis thaliana* (C_3); Fan, *F. anomala* (C_3-C_4); Fbi, *F. bidentis* (C_4); Fbr, *F. brownii* (C_4-like); Fcr, *F. cronquistii* (C_3); Fpa, *F. palmeri* (C_4-like); Fra, *F. ramosissima* (C_3-C_4); Fro, *F. robusta* (C_3); Ftr, *F. trinervia*; Fva, *F. vaginata* (C_4-like); Me, *Manihot esculenta* (C_3); Nt, *Nicotiana tabacum* (C_3); Os, *Oryza sativa* (C_3); Sv, *Setaria viridis* (C_4); Ta, *Triticum aestivum* (C_3); Zm, *Zea mays* (C_4). For metabolite abbreviations refer to Table S1. Detailed growth and harvest conditions can be found in the Material & Methods section, Supplementary Data file S4 or Arrivault et al. (2019). For the raw data used for creating this figure, refer to Supplementary Data file S3.

REFERENCES

Allen DK, Young JD (2020) Tracing metabolic flux through time and space with isotope labeling experiments. Curr Opin Biotechnol **64**: 92–100

Andersen KS, Bain JM, Bishop DG, Smillie RM (1972) Photosystem II activity in agranal bundle sheath chloroplasts from *Zea mays*. Plant Physiol **49**: 461–466

Apel P, 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 und Physiol der Pflanz **176**: 396–399

Arrivault S, Alexandre Moraes T, Obata T, Medeiros DB, Fernie AR, Boulouis A, Ludwig M, Lunn JE, Borghi GL, Schlereth A, et al (2019) Metabolite profiles reveal interspecific variation in operation of the Calvin-Benson cycle in both C_4 and C_3 plants. J Exp Bot **70**: 1843–1858

Arrivault S, Guenther M, Ivakov A, Feil R, Vosloh D, Van Dongen JT, Sulpice R, Stitt M (2009) Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. Plant J **59**: 824–839

Arrivault S, Obata T, Szecówka M, Mengin V, Guenther M, Hoehne M, Fernie AR, Stitt M (2017) Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation. J Exp Bot **68**: 283–298

Arrivault SS, Guenther M, Fry SC, Fuenfgeld MMFFFF, Veyel D, Mettler-Altmann T, Stitt M, Lunn JE (2015) Synthesis and use of stable-isotope-labeled internal standards for quantification of phosphorylated metabolites by LC-MS/MS. Anal Chem **87**: 6896–6904

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

Bassüner B, Keerberg O, Bauwe H, Pyarnik T, Keerberg H (1984) Photosynthetic CO_2 metabolism in C_3-C_4 intermediate and C_4 species of *Flaveria* (Asteraceae). Biochem und Physiol der Pflanz **179**: 631–634

Bauwe H (1984) Photosynthetic enzyme activities and immunofluorescence studies on the localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in leaves of C_3, C_4, and C_3−C_4 intermediate species of *Flaveria* (Asteraceae). Biochem und Physiol der Pflanz **179**: 253–268

Bauwe H, Chollet R (1986) Kinetic properties of phosphoenolpyruvate 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, Fernie AR (2010) Photorespiration: players, partners and origin. Trends Plant Sci **15**: 330–336

Bauwe H, Kolukisaoglu Ü (2003) Genetic manipulation of glycine decarboxylation. J Exp Bot **54**: 1523–1535

Billakurthi K, Wrobel TJ, Bräutigam A, Weber APM, Westhoff P, Gowik U (2018) Transcriptome dynamics in developing leaves from C_3 and C_4 *Flaveria* species reveal determinants of Kranz anatomy. doi: 10.1101/473181

Bohley K, Schröder T, Kesselmeier J, Ludwig M, Kadereit G (2019) C_4-like photosynthesis and the effects of leaf senescence on C_4-like physiology in *Sesuvium sesuvioides* (Aizoaceae). J Exp Bot **70**: 1553–1565

Borghi GL, Moraes TA, Günther M, Feil R, Mengin V, Lunn JE, Stitt M, Arrivault S, Raines C (2019) Relationship between irradiance and levels of Calvin-Benson cycle and other intermediates in the model eudicot Arabidopsis and the model monocot rice. J Exp Bot **70**: 5809–5825

Bräutigam A, Gowik U (2016) Photorespiration connects C_3 and C_4 photosynthesis. J Exp Bot **67**: 2953–2962

Brown RH (1978) A difference in N use efficiency in C_3 and C_4 plants and its implications in adaptation and evolution. Crop Sci **18**: 93–98

Busch FA, Sage TL, Cousins AB, Sage RF (2013) C_3 plants enhance rates of photosynthesis by reassimilating photorespired and respired CO_2. Plant, Cell Environ **36**: 200–212

Byrd GT, Brown RH, Bouton JH, Bassett CL, Black CC (1992) Degree of C_4 photosynthesis in C_4 and C_3-C_4 *Flaveria* species and their hybrids: I. CO_2 assimilation and metabolism and activities of phosphoeno/pyruvate carboxylase and NADP-malic enzyme. Plant Physiol **100**: 939–946

Carillo P, Mastrolonardo G, Nacca F, Fuggi A (2005) Nitrate reductase in durum wheat seedlings as affected by nitrate nutrition and salinity. Funct Plant Biol **32**: 209–219

Chastain CJ, 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

Cheng S-H, Moore B d., Wu J, Edwards GE, Ku MSB (1989) Photosynthetic Plasticity in *Flaveria brownii* . Plant Physiol **89**: 1129–1135

Cheng S-H, Moore BD, Edwards GE, Ku MSB (1988) Photosynthesis in *Flaveria brownii*, a C_4-Like Species. Plant Physiol **87**: 867–873

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N (2008) Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr Biol **18**: 37–43

Christin PA, Osborne CP (2014) The evolutionary ecology of C_4 plants. New Phytol **204**: 765–781

Dai Z, Ku MSB, Edwards GE (1996) Oxygen sensitivity of photosynthesis and photorespiration in different photosynthetic types in the genus *Flaveria*. Planta **198**: 563–571

Driever SM, Lawson T, Andralojc PJ, Raines CA, Parry MAJ (2014) Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. J Exp Bot **65**: 4959–4973

Drincovich MF, Casati P, Andreo CS, Chessin SJ, Franceschi VR, Edwards GE, Ku MSB (1998) Evolution of C_4 photosynthesis in *Flaveria* species. Plant Physiol **117**: 733–744

Edwards GE, Furbank RT, Hatch MD, Osmond CB (2001) What does it take to be C_4? Lessons from the evolution of C_4 photosynthesis. Plant Physiol **125**: 46–49

Engelmann S, Bläsing OE, Gowik U, Svensson P, Westhoff P (2003) Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria* - A gradual increase from C_3 to C_4 characteristics. Planta **217**: 717–725

Ermakova M, Danila FR, Furbank RT, Caemmerer S Von, von Caemmerer S (2019) On the road to C_4 rice: advances and perspectives. Plant J 1–11

Furbank RT (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? J Exp Bot **62**: 3103–3108

Ghannoum O (2005) Faster Rubisco is the key to superior nitrogen-use efficiency in NADP-malic enzyme relative to NAD-malic enzyme C_4 grasses. Plant Physiol **137**: 638–650

Ghannoum O (2009) C_4 photosynthesis and water stress. Ann Bot **103**: 635–644

Ghannoum O, Evans JR, von Cammerer S (2011) Nitrogen and water use efficiency of C_4 plants. *In* A Raghavendra, RF Sage, eds, C_4 Photosynth. Relat. CO2 Conc. Mech. Dordrecht: Springer, pp 129–146

Gowik U, Bräutigam A, Weber KL, Weber APM, 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

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

Gowik U, Westhoff P (2011b) C_4 Phosphoenolpyruvate carboxylase. C_4 photosynth. Relat. CO_2 Conc. Mech. pp 257–275

Haberlandt G (1918) Physiologische pflanzenanatomie. W. Engelmann

Hasunuma T, Harada K, Miyazawa SI, Kondo A, Fukusaki E, Miyake C (2010) Metabolic turnover analysis by a combination of in vivo ^{13}C-labelling from $^{13}CO_2$ and metabolic profiling with CE-MS/MS reveals rate-limiting steps of the C_3 photosynthetic pathway in Nicotiana tabacum leaves. J Exp Bot **61**: 1041–1051

Hatch MD (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim Biophys Acta - Rev Bioenerg **895**: 81–106

Hatch MD, Slack CR (1966) Photosynthesis by sugar-cane leaves. a new carboxylation reaction and the pathway of sugar formation. Biochem J **101**: 103–11

Häusler RE, Hirsch HJ, Kreuzaler F, 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 (2016) C_4 photosynthesis evolution: the conditional Mt. Fuji. Curr Opin Plant Biol **31**: 149–154

Heckmann D, Schulze S, Denton AK, Gowik U, Westhoff P, Weber APM, Lercher MJ (2013) Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell **153**: 1579–88

Hibberd JM, Covshoff S (2010) The regulation of gene expression required for C_4 photosynthesis. Annu Rev Plant Biol **61**: 181–207

Höfer MU, Santore UJ, Westhoff P (1992) Differential accumulation of the 10-, 16- and 23-kDa peripheral components of the water-splitting complex of photosystem II in mesophyll and bundle-sheath chloroplasts of the dicotyledonous C_4 plant *Flaveria trinervia* (Spreng.) C. Mohr. Planta **186**: 304–312

Huege J, Goetze J, Dethloff F, Junker B, Kopka J (2014) Quantification of stable isotope label in metabolites via mass spectrometry. Methods Mol Biol **1056**: 213–223

Hylton CM, Rawsthorne S, Smith AM, Jones DA, Woolhouse HW (1988) Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. Planta **175**: 452–459

Ishihara H, Obata T, Sulpice R, Fernie AR, Stitt M (2015) Quantifying protein synthesis and degradation in Arabidopsis by dynamic $^{13}CO_2$ labeling and analysis of enrichment in individual amino acids in their free pools and in protein. Plant Physiol **168**: 74–93

Kapralov M V., Kubien DS, Andersson I, Filatov DA (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

Keerberg O, Pärnik T, Ivanova H, Bassüner B, Bauwe H (2014) C_2 photosynthesis generates about 3-fold elevated leaf CO_2 levels in the C_3-C_4 intermediate species *Flaveria pubescens*. J Exp Bot **65**: 3649–3656

Khoshravesh R, Stata M, Busch FA, Saladie M, Castelli JM, Dakin N, Hattersley PW, Macfarlane TD, Sage RF, Ludwig M, et al (2019) The evolutionary origin of C_4 photosynthesis in the grass subtribe Neurachninae. Plant Physiol **182**: pp.00925.2019

Khoshravesh R, Stinson CR, Stata M, Busch FA, Sage RF, Ludwig M, Sage TL (2016) C_3-C_4 intermediacy in grasses: organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C_2 photosynthesis. J Exp Bot **67**: 3065–3078

Kocacinar F, McKown AD, Sage TL, Sage RF (2008) Photosynthetic pathway influences xylem structure and function in *Flaveria* (Asteraceae). Plant, Cell Environ **31**: 1363–1376

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, Dörmann P, Weckwerth W, Gibon Y, Stitt M, et al (2005) GMD@CSB.DB: The Golm metabolome database. Bioinformatics **21**: 1635–1638

Kopriva S, Chu CC, 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 Environ **19**: 1028–1036

Kortschak HP, Hartt CE, Burr GO (1965) Carbon dioxide fixation in sugarcane leaves. Plant Physiol **40**: 209–213

Ku MSB, Monson RK, Littlejohn RO, Nakamoto H, Fisher DB, Edwards GE (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 **75**: 993–996

Ku MSB, Wu J, Dai Z, Scott R a, Chu C, Edwards GE (1991) Photosynthetic and photorespiratory characteristics of *Flaveria* species. Plant Physiol **96**: 518–528

Kubien DS, Whitney SM, Moore P V., Jesson LK (2008) The biochemistry of Rubisco in *Flaveria*. J. Exp. Bot. pp 1767–1777

Langdale JA (2011) C_4 cycles: past, present, and future research on C_4 photosynthesis. Plant Cell **23**: 3879–3892

Lawson T, Kramer DM, Raines CA (2012) Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. Curr Opin Biotechnol **23**: 215–220

Leegood RC (1985) The intercellular compartmentation of metabolites in leaves of *Zea mays* L. Planta **164**: 163–171

Leegood RC, von Caemmerer S (1994) Regulation of photosynthetic carbon assimilation in leaves of C_3-C_4 intermediate species of *Moricandia* and *Flaveria*. Planta **192**: 232–238

Levey M, Timm S, Mettler-Altmann T, Borghi GL, Koczor M, Arrivault S, Weber APM, Bauwe H, Gowik U, Westhoff P (2019) Efficient 2-phosphoglycolate degradation is required to maintain carbon assimilation and allocation in the C_4 plant *Flaveria bidentis*. J Exp Bot **70**: 575–587

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc **1**: 387–396

Lorimer GH, Andrews TJ (1973) Plant photorespiration - an inevitable consequence of the existence of atmospheric oxygen. Nature **243**: 359–360

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

Lundgren MR, Osborne CP, Christin PA (2014) Deconstructing Kranz anatomy to understand C_4 evolution. J Exp Bot **65**: 3357–3369

Lyu MJA, Gowik U, Kelly S, Covshoff S, Mallmann J, Westhoff P, Hibberd JM, Stata M, Sage RF, Lu H, et al (2015) RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria* (Asteraceae) with some modifications evolutionary developmental biology and morphology. BMC Evol Biol **15**: 116

Ma F, Jazmin LJ, Young JD, Allen DK (2014) Isotopically nonstationary ^{13}C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. Proc Natl Acad Sci **111**: 16967–16972

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U (2014) The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. Elife **2014**: 1–23

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

Mckown AD, Dengler NG (2007) Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). Am J Bot **94**: 382–389

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

Meister M, Agostino A, Hatch MD (1996) The roles of malate and aspartate in C_4 photosynthetic metabolism of *Flaveria bidentis* (L.). Planta **199**: 262–269

Merlo L, Geigenberger P, Hajirezaei M, Stitt M (1993) Changes of carbohydrates, metabolites and enzyme activities in potato tubers during development, and within a single tuber along a stolon-apex gradient. J Plant Physiol **142**: 392–402

Miyao M, Masumoto C, Miyazawa SI, Fukayama H (2011) Lessons from engineering a single-cell C_4 photosynthetic pathway into rice. J Exp Bot **62**: 3021–3029

Monson RK (1989) The relative contributions of reduced photorespiration, and improved water- and nitrogen-use efficiencies, to the advantages of C_3-C_4 intermediate photosynthesis in *Flaveria*. Oecologia **80**: 215–221

Monson RK, Edwards GE, Ku MSB (1984) C_3-C_4 intermediate photosynthesi plants. Bioscience **34**: 563–566

Monson RK, Moore B d. (1989) On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. Plant Cell Environ **12**: 689–699

Monson RK, Moore B d., Ku MSB, Edwards GE (1986) Co-function of C_3-and C_4-photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. Planta **168**: 493–502

Monson RK, Rawsthorne S (2000) CO_2 assimilation in C_3-C_4 intermediate plants. photosynthesis. pp 533–550

Monson RK, Schuster WS, Ku MSB (1987) Photosynthesis in *Flaveria brownii*, a C_4-like C_3-C_4 intermediate. Plant Physiol **85**: 1063–1067

Monson RK, Teeri JA, Ku MSB, Gurevitch J, Mets LJ, Dudley S (1988) Carbon-isotope discrimination by leaves of *Flaveria* species exhibiting different amounts of C_3-and C_4-cycle co-function. Planta **174**: 145–151

Moore B, Ku MSB, Edwards GE (1989) Expression of C_4-like photosynthesis in several species of *Flaveria*. Plant Cell Environ **12**: 541–549

Moore BD, Ku MSB, Edwards GE (1987) C_4 photosynthesis and light-dependent accumulation of inorganic carbon in leaves of C_3-C_4 and C_4 *Flaveria* species. Aust J Plant Physiol **14**: 657–68

Moss DN (1971) Carbon dioxide compensation in plants with C_4 characteristics. *In* MD Hatch, CB Osmond, RO Slatyer, eds, Photosynth. photorespiration. New York, USA, Wiley-Interscience, pp 120–123

Muhaidat R, Sage RF, Dengler NG (2007) Diversity of Kranz anatomy and biochemistry in C_4 eudicots. Am J Bot **94**: 362–381

Muhaidat R, Sage TL, Frohlich MW, Dengler NG, Sage RF (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

Nakajima Munekage Y (2016) Light harvesting and chloroplast electron transport in NADP-malic enzyme type C_4 plants. Curr Opin Plant Biol **31**: 9–15

Nakamoto H, Ku MSB, Edwards GE (1983) Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria* species II. Kinetic properties of phosphoenolpyruvate carboxylase from C_3, C_4 and C_3-C_4 intermediate species. Plant Cell Physiol **24**: 1387–1393

Nakamura N, Iwano M, Havaux M, Yokota A, Munekage YN (2013) Promotion of cyclic electron transport around photosystem I during the evolution of NADP-malic enzyme-type C_4 photosynthesis in the genus *Flaveria*. New Phytol **199**: 832–842

Nelson T, Langdale J (1992) Developmental genetics of C_4 photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 25–47

Osmond CB, Harris B (1971) Photorespiration during C_4 photosynthesis. BBA - Bioenerg **234**: 270–282

Perdomo JA, Cavanagh AP, Kubien DS, Galmés J (2015) Temperature dependence of *in vitro* Rubisco kinetics in species of *Flaveria* with different photosynthetic mechanisms. Photosynth Res **124**: 67–75

Powell AM (1978) Systematics of *Flaveria* (Flaveriinae-Asteraceae). Missouri Bot Gard Press **65**: 590–636

Rawsthorne S, Hylton CM (1991) The relationship between the post-illumination CO_2 burst and glycine metabolism in leaves of C_3 and C_3-C_4 intermediate species of *Moricandia*. Planta **186**: 122–126

Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW (1988a) Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C_3 and C_3-C_4 intermediate species of *Moricandia*. Planta **173**: 298–308

Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW (1988b) Distribution of photorespiratory enzymes between bundle-sheath and mesophyll cells in leaves of the C_3-C_4 intermediate species *Moricandia arvensis* (L.) DC. Planta **176**: 527–532

Rumpho ME, Ku MSB, Cheng S-H, Edwards GE (1984) Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria* species. Plant Physiol **75**: 993–996

Sage RF (2017) A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. J Exp Bot **68**: e11–e28

Sage RF (2003) The evolution of C_4 photosynthesis. New Phytol **161**: 341–370

Sage RF (2002) Variation in the k_{cat} of Rubisco in C_3 and C_4 plants and some implications for photosynthetic performance at high and low temperature. J Exp Bot **53**: 609–620

Sage RF, Christin P-A, Edwards EJ (2011) The C_4 plant lineages of planet Earth. J Exp Bot **62**: 3155–3169

Sage RF, Khoshravesh R, Sage TL (2014) From proto-Kranz to C_4 Kranz: Building the bridge to C_4 photosynthesis. J Exp Bot **65**: 3341–3356

Sage RF, Monson RK, Ehleringer JR, Adachi S, Pearcy RW (2018) Some like it hot: the physiological ecology of C_4 plant evolution. Oecologia **187**: 941–966

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

Sage TL, Busch FA, Johnson DC, Friesen PC, Stinson CR, Stata M, Sultmanis S, Rahman BA, Rawsthorne S, Sage RF (2013) Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. Plant Physiol **163**: 1266–1276

Salvucci ME (1989) Regulation of Rubisco activity in vivo. Physiol Plant **77**: 164–171

Schlüter U, Bräutigam A, Droz JM, Schwender J, Weber APM (2018) The role of alanine and aspartate aminotransferases in C_4 photosynthesis. Plant Biol **21**: 64–76

Schlüter U, Weber APM (2020) Regulation and evolution of C_4 photosynthesis. Annu Rev Plant Biol **71**: 1–33

Schlüter U, Weber APM (2016) The road to C_4 photosynthesis: evolution of a complex trait via intermediary states. Plant Cell Physiol **57**: 881–889

Schmitt MR, Edwards GE (1981) Photosynthetic capacity and nitrogen use efficiency of maize, wheat, and rice: a comparison between C_3 and C_4 photosynthesis. J Exp Bot **32**: 459–466

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P (2013) Evolution of C_4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 Pump. Plant Cell **25**: 2522–2535

Schulze S, Westhoff P, Gowik U (2016) Glycine decarboxylase in C_3, C_4 and C_3-C_4 intermediate species. Curr Opin Plant Biol **31**: 29–35

Schüssler C, Freitag H, Koteyeva N, Schmidt D, Edwards G, Voznesenskaya E, Kadereit G (2017) Molecular phylogeny and forms of photosynthesis in tribe Salsoleae (Chenopodiaceae). J Exp Bot **68**: 207–223

Simkin AJ, López-Calcagno PE, Raines CA (2019) Feeding the world: improving photosynthetic efficiency for sustainable crop production. J Exp Bot **70**: 1119–1140

Smith B, Turner B (1975) Distribution of Kranz syndrome among Asteraceae. Am J Bot **62**: 541–545

Smith BN, Powell AM (1984) C_4-like F1-hybrid of C_3 x C_4 *Flaveria* species. Naturwissenschaften **71**: 217–218

Stata M, Sage TL, Hoffmann N, Covshoff S, Ka-Shu Wong G, Sage RF (2016) Mesophyll chloroplast investment in C_3, C_4 and C_2 species of the genus *Flaveria*. Plant Cell Physiol **57**: 904–918

Still CJ, Berry JA, Collatz GJ, DeFries RS (2003) Global distribution of C_3 and C_4 vegetation: carbon cycle implications. Global Biogeochem Cycles **17**: 6-1-6–14

Stitt M, Heldt HW (1985) Generation and maintenance of concentration gradients between the mesophyll and bundle sheath in maize leaves. Biochim Biophys Acta **808**: 400–414

Sudderth EA, Muhaidat RM, McKown AD, Kocacinar F, Sage RF (2007) Leaf anatomy, gas exchange and photosynthetic enzyme activity in *Flaveria kochiana*. Funct Plant Biol **34**: 118–129

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

Szecowka M, Heise R, Tohge T, Nunes-Nesi A, Vosloh D, Huege J, Feil R, Lunn J, Nikoloski Z, Stitt M, et al (2013) Metabolic fluxes in an illuminated Arabidopsis rosette. Plant Cell **25**: 694–714

Usuda H, Edwards GE (1980) Localization of glycerate kinase and some enzymes for sucrose synthesis in C_3 and C_4 plants. Plant Physiol 1017–1022

Vogan PJ, Sage RF (2011) Water-use efficiency and nitrogen-use efficiency of C_3-C_4 intermediate species of *Flaveria* Juss. (Asteraceae). Plant, Cell Environ **34**: 1415–1430

Voznesenskaya E V., Franceschi VR, Kiirats O, Freitag H, Edwards GE (2001) Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. Nature **414**: 543–546

Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, et al (2014) Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. Nat Biotechnol **32**: 1158–1165

Weber APM, von Caemmerer S (2010) Plastid transport and metabolism of C_3 and C_4 plants-comparative analysis and possible biotechnological exploitation. Curr Opin Plant Biol **13**: 256–264

Weissmann S, Ma F, Furuyama K, Gierse J, Berg H, Shao Y, Taniguchi M, Allen DK, Brutnell TP (2016) Interactions of C_4 subtype metabolic activities and transport in maize are revealed through the characterization of DCT2 mutants. Plant Cell **28**: 466–484

Wessinger ME, Edwards GE, Ku MSB (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, Gowik U (2004) Evolution of C_4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus *Flaveria*. Ann Bot **93**: 13–23

Williams BP, Johnston IG, Covshoff S, Hibberd JM, Abramoff M, Magalhaes P, Ram S, Adams C, Leung F, Sun S, et al (2013) Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. Elife **2**: e00961

Woo KC, Anderson JM, Boardman NK, Downton WJS, Osmond CB, Thorne SW (1970) Deficient photosystem II in agranal bundle sheath chloroplasts of C_4 plants. Proc Natl Acad Sci **67**: 18–25

Zachos JC, Dickens GR, Zeebe RE (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature **451**: 279–283

Zhu XG, Long SP, Ort DR (2008) What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Curr Opin Biotechnol **19**: 153–159

General Discussion and Conclusions

Photosynthetic carbon assimilation through Rubisco and the CBC is an ancient pathway that emerged around 2.5 billion years ago (Rasmussen et al., 2008) and since then the basic structure of the pathway has not changed much. However, the atmosphere around Rubisco did change over the eons, from high CO_2 and almost absent O_2 levels at its emergence, through several fluctuations until stabilizing to relatively low CO_2 and high O_2 levels in the last million years (Zachos et al., 2008). In this time, plants established themselves on land, initially benefitting from increased access to CO_2, but also starting to deal with problems associated with water uptake and retention. Both in algae and land plants, falling CO_2 exerted a strong evolutionary pressure to circumvent the O_2 side reaction of Rubisco and avoid related photorespiratory costs. In terrestrial plants, the impact of falling atmospheric CO_2 was exacerbated by the need to minimize stomatal conductance when water was in short supply, and especially when temperature was high. Together, these evolutionary pressures led to the evolution of different CCMs. The most recent of these CCMs are CAM and C_4 photosynthesis, appeared for the first time in vascular plants respectively around 200 and 30 million years ago (Keeley and Rundel, 2003). In particular, C_4 photosynthesis grants the plants that run this mode of photosynthesis higher rates of photosynthesis, improved biomass accumulation (Still et al., 2003) and better light, water and nitrogen use efficiencies (Ghannoum et al., 2011) compared to C_3 plants. This is still so today in tropical and sub-tropical climates.

For the abovementioned reasons, introduction of C_4 photosynthesis into C_3 crops was theorized as a possible solution to solve productivity thresholds reached by conventional breeding in staple crops (Mitchell and Sheehy, 2006; Hibberd et al., 2008), and is currently being actively pursued in rice (von Caemmerer et al., 2012; Ermakova et al., 2019). However, introducing a CCM it is not the only way in which carbon assimilation could be improved in plants (Kubis and Bar-Even, 2019): a better understanding of the CBC diversity and adaptation to different environmental conditions could also be the key for better future crops (Raines, 2006; Lawson et al., 2012).

The aim of this thesis, with the three papers hereby included, was to investigate the metabolic diversity in the CBC residing within two common carbon assimilation strategies: C_3 photosynthesis, running only the CBC cycle, and C_4 photosynthesis, running both CBC coupled with the C_4 CCM. The genus *Flaveria*, containing several intermediate species between C_3 and C_4 photosynthesis, is investigated in the last paper,

in this case considering CBC and C_4 cycles intermediates, as well as related metabolic pathways downstream of carbon assimilation.

Metabolite levels: why they are important and what we can learn from them

Measurements of metabolite levels provide information about the poise between different reactions in a pathway, and provide insights into how it is regulated and how this may vary between conditions or genotypes. As an example, it was realized fifty years ago that the changes in metabolite levels between two steady states can be used to infer which reactions are stimulated and which are inhibited in moving from one state to another (Rolleston, 1972; Newsholme and Start, 1973). In a transition in which pathway flux increases, a fall in the level of a metabolite indicates that the enzyme that uses this metabolite has been activated, especially if the product also declines. Conversely, in a transition where pathway flux decreases, an increase in the level of a metabolite indicates that the enzyme that uses it has been inhibited. This is especially so for the enzymes that catalyze reactions that are removed from thermodynamic equilibrium. An analogous approach can be taken to compare a pathway between two different species; a lower level of a metabolite in species A compared to species B may indicate that the enzyme that catalyzes the reaction have a higher activity or is positively regulated in a more effective way in species A than in species B. To put it another way, metabolite profiling provides a top-down strategy to investigate the operational balance in a pathway.

This approach requires comprehensive and quantitative analysis of as many metabolites in a pathway or metabolic sequence as possible. Measurements of primary metabolites linked to photosynthesis (both CBC and CCM-related metabolites) have been performed on whole leaf material from various species since the 80's, albeit with limited coverage and scattered among multiple papers (Stitt et al., 1982; Stitt et al., 1983; Badger et al., 1984; Dietz and Heber, 1984; Stitt et al., 1984a; Stitt et al., 1984b; Stitt and Heldt, 1985b; Usuda, 1985; Seemann and Sharkey, 1986; Usuda, 1986; Usuda, 1987; Leegood and von Caemmerer, 1988; Leegood and von Caemmerer, 1989; Servaites et al., 1989). Most of these papers included only a small subset of metabolites due conceptual and technical reasons. Each paper also targeted only one or at the most a limited number of species.

In the last two decades, large scale measurements of primary metabolites in plants became easier thanks to the development of several techniques based on chromatography coupled

with mass spectrometry (Sato et al., 2004; Lisec et al., 2006; Lunn et al., 2006; Antonio et al., 2007; Huege et al., 2007; Cruz et al., 2008; Arrivault et al., 2009; Hasunuma et al., 2010; Ma et al., 2014). These techniques were able to detect absolute amounts of many metabolites at the pmol range.

Some of the platforms were also able to detect isotopomer distribution in these compounds following isotopic labelling experiments. Isotopomer distributions are important, together with amounts, in order to calculate metabolic fluxes downstream of carbon assimilation (Ratcliffe and Shachar-Hill, 2006), and this principle was applied several times in the model organism *Arabidopsis thaliana* (Huege et al., 2007; Williams et al., 2010; Szecowka et al., 2013).

In the three manuscripts that constitute this thesis, I was able to obtain quantitative information about the levels of CBC intermediates using reverse phase LC-MS/MS technique. This information provides a top-down snapshot of the balance between the different reaction steps in the CBC pathway. These data served the two main aims of the manuscripts: firstly, to focus on CBC metabolites and investigate the extent to which operation of the CBC is maintained or diverge between different photosynthesis modes (manuscript 1 and 3) and species (all manuscripts). Secondly, as part of a broader set of metabolites that I investigated to learn not only how the CBC, but also other sectors of metabolism, were modified along the evolutionary axis from C_3 to C_3-C_4 intermediate to C_4-like to C_4 photosynthesis in the *Flaveria* genus. Here, I added measurement with HPLC and GC-MS to access a wider set of metabolites involved in the evolving CCM, like key amino acids.

In analyzing these data sets, I used both integrative analyses like hierarchical clustering and principal component analysis (PCA) to provide a high-level overview, and detailed inspection of the contents of individual metabolites to provide first glimpses into underlying mechanisms. Despite the lack of detailed metabolic flux analyses or modelling within this thesis, the substantial metabolic information provided within constitute a solid base for future studies going in this direction.

Differences in CBC operation between plants with different photosynthetic modes

In all the papers included in this thesis the CBC metabolite profile varied between species, in particular we uncovered differences in the relative levels of several CBC metabolites.

This points toward differences in the CBC operation between species. Although not studied further, it presumably reflects changes in the balance between enzymatic activities and regulation between species.

The CBC metabolic differences analyzed in PCA graphs in manuscript one and three separate clearly those species that utilize the C_4 cycle from those which do not; this separation was driven mainly by a decline in RuBP absolute level and an increased level of 3PGA and triose-P in C_4 plants. The lower level of RuBP in C_4 species was expected, and reflects the higher k_{cat} and lower Rubisco abundance in C_4 species compared to C_3 species (see Introduction). At the same time, other differences emerge for other CBC intermediates, resulting in species-specific separation even between species sharing the same carbon assimilation mode. This, and especially the divergence of C_3 species, was one of the most unexpected results in my thesis (discussed below).

If we introduce several intermediates between C_3 and C_4 photosynthesis (like in manuscript three of this thesis), the overall CBC-driven separation of C_3 and C_4 species is maintained, and the intermediate species are located on the trajectory between the two. This last observation points toward progressive adaptation of the CBC in parallel with the evolution of a CCM. To confirm this finding, it would be interesting in the future to extend these multispecies comparisons to other genera that contain C_3-C_4 intermediates in order to make a clearer generalization about CBC evolution; this might include genera like *Panicum* (Sternberg et al., 1986) and *Neurachne* (Khoshravesh et al., 2019), which are particularly interesting because their phylogenetically closeness to well-studied C_4 grasses, like maize and *Setaria viridis*.

Variation between CBC operation in different C_3 species

As already mentioned, it was especially striking that there we were able to detect differences in the profile of CBC intermediates between species running the ancestral C_3 cycle: this included Arabidopsis, tobacco, cassava, two *Flaveria* basal species, wheat and rice. There were many differences in their metabolite profiles affecting, for example, the relation between the levels of fructose 1,6-bisphosphate (FBP) and sedoheptulose 1,7-bisphosphate (SBP), and the levels of other metabolites downstream of them in the CBC. This finding was initially surprising for us, as it indicated that the CBC operates in a different way in species sharing the same carbon assimilation pathway.

We considered possible alternative explanations for our results. The interspecies differences were largely robust against different ways of normalizing the metabolite data before performing PCAs, indicating that they are not a trivial secondary effect of changes in leaf composition, for example water, protein or chlorophyll content (Arrivault et al., 2019). On the other hand, it is possible that in some cases the changes in CBC metabolite levels might be linked to changes in leaf anatomy, in the sense that they reflect an adaption to, or a mode of operation that is allowed by a special anatomical or morphological feature in that species. For example, this is discussed in relation to the generally high levels of protein and metabolites in the leaves of rice and cassava compared to most other C_3 species. In rice these may be linked to special leaf anatomical features that allow more effective recapture of photorespired CO_2 (Sage and Sage, 2009). For cassava, more research will be needed to identify what changes in leaf anatomy, ultrastructure of physiological state allow photosynthesis operate efficiently with a high leaf protein content.

Another possible explanation for the differing CBC metabolite profiles between species might be that the plants were grown and harvested in different conditions. Indeed, given that plant species differ in their photosynthetic rate and its dependence on light, temperature, and the availability of water, nutrients and CO_2, it is challenging to design experiments where such effects could be totally ruled out. We chose to grow and harvest plants in a light regime that was somewhat limiting for that species, rather than using identical conditions for all species. In the conditions we used, RuBP regeneration is likely to be limiting, and effects of light stress are avoided. In a detailed comparison of the response of Arabidopsis and rice to different ambient light intensities, it emerged that the CBC in the two species responds differently to increasing irradiances (i.e. different photosynthetic plasticity), with Arabidopsis favoring 3PGA reduction and rice favoring RuBP regeneration (Borghi et al., 2019). Thus, it can be concluded that there is indeed species variance in CBC operation, with different strategies being used to increase CBC flux to use the available energy as light intensity rises.

These more detailed analyses in Arabidopsis and rice at differing irradiance, and differing CO_2 levels, revealed that these two species do show some CBC responses in common. One example is high FBP and SBP in low light, pointing to inhibition of both FBPase and sedoheptulose-1,7-bisphosphatase (SBPase; EC 3.1.3.37) not only on the dark but also at irradiances close to the compensation point. Our FBPase and SBPase findings are compatible with these enzymes being redox regulated to inhibit them in the dark and to

promote CBC flux at higher light intensities (Buchanan and Balmer, 2005; Heldt et al., 2005). They indicate that these enzymes are inactivated not only in the dark, but at low light intensity. This will restrict or avoid futile cycling, which would affect seriously the quantum yield and photosynthetic efficiency in low light. Another shared feature was the maintenance of quite high CBC metabolite levels in very low light and in low CO_2. This observation indicates that withdrawal of carbon from the CBC is strongly restricted when CO_2 fixation is slow, supporting the idea that, for example, sucrose synthesis is regulated according to a "threshold" principle (Stitt, 1990; Stitt et al., 2010). According to this principle, an interplay between fructose 2,6-bisphosphate, substrate availability and the cooperative substrate saturation kinetics of cytosolic FBPase leads to this enzyme being inactive until a threshold level of triose-P is reached, and then strongly activated as the triose-P level rises further. Meta-analysis of published data for many other species from the 1980s and 1990s revealed that both of the above are general features, although only evident in retrospect based on our more comprehensive analysis in Arabidopsis and rice. The following question arose; why is there such large variance between CBC operations in different C_3 species, given that the pathway is over 2 million years old, long pre-dating the divergence of the monocot and dicot lineages? Here we advanced an hypothesis that, although not tested, is plausible. It is well known that C_4 photosynthesis evolved from C_3 ancestors following specific evolutionary pressures like low CO_2/O_2 ratios, seasonal droughts and high irradiances in recent geological time (Sage, 2003; Sage et al., 2012; Sage, 2017). However, the same pressures were in place for plant lineages which never evolved C_4 CCM, and such pressures are in place in many environments where C_3 plants still thrive. These evolutionary pressures have acted and will be acting independently on each plant lineage. Therefore, other mechanisms linked to photosynthetic efficiency may have evolved in C_3 species to cope with these unfavorable conditions. These mechanisms would have evolved independently from lineage to lineage, and the same trajectory may not have been followed in each lineage, either by chance or because they had different starting points or different driving conditions. Indeed, it is known that C_3 species differ for many aspect of photosynthesis, including Rubisco kinetics (Galmés et al., 2014; Prins et al., 2016) and Rubisco regulation (Parry et al., 2008), leaf anatomy and intracellular organelle distribution (Sage and Sage, 2009; Busch et al., 2013; Sage et al., 2013b) and tuning of stomatal behavior (Tardieu and Simonneau, 1998; Lin et al., 2015; Lawson and Vialet-Chabrand, 2018). All of these adaptations could represent ways of improving photosynthetic efficiency (Lawson et al., 2012), and could be behind the observed

dazzling diversity in photosynthetic efficiencies in C_3 species (Evans, 1989; Wullschleger, 1993).

The different CBC metabolite profiles that we found could be the result of these and other molecular changes, including differences in the relation between the activities or regulation of enzymes in the CBC, or the relation between the two bisphosphatases and other CBC enzymes. Such differences may be important to consider when planning strategies to improve photosynthesis in certain crop species.

Investigating intercellular metabolite shuttles involved in CCMs

The operation of C_4 photosynthesis and of related intermediate modes of photosynthesis requires one or more intercellular metabolite shuttles. Movement of metabolites in these shuttles is thought to occur mainly by diffusion, thus requiring gradients between cell types and, hence, large total pools of the involved metabolites (Bräutigam and Weber, 2010). These high pools could also function as buffers against fluctuating environmental conditions, like irradiance (Stitt and Zhu, 2014).

The basic C_4 shuttle is the one involving 4-carbon acids, like malate, moving from the mesophyll to the BS cells, and pyruvate moving back to the mesophyll after decarboxylation, to allow in loco regeneration of PEP. A supplementary version of this cycle involves the conversion of organic acids into the corresponding amino acids by aminotransferase reactions, allowing the movement of both carbon and nitrogen at the same time. The employment of this parallel amino acid cycle is paramount for the functioning of NAD-ME and PEPCK decarboxylation subtypes (Furbank, 2011; Wang et al., 2014), but it is employed by NADP-ME subtype as well, although probably with different levels of relevance between species. An example of these shuttle differences can be observed by comparing NADP-ME species like maize (Pick et al., 2011; Arrivault et al., 2017) and *Flaveria* C_4 species (Meister et al., 1996, manuscript three of this thesis); active malate pools in maize are ten times higher than aspartate, while in *F. bidentis* only two times, indicating that maize relies less on aspartate shuttling, and more on malate than *F. bidentis*. Together with being less reliant on aspartate shuttling, maize also appears to have higher rates of 3PGA/triose-P shuttling compared to C_4 *Flaveria*. These findings are compatible with different chloroplastic features of maize vs *Flaveria*, in which the first one has dimorphic chloroplasts and a lack of PSII activity in BS cells (Woo

et al., 1970; Andersen et al., 1972; Nakajima Munekage, 2016), while the second has not (Höfer et al., 1992; Nakamura et al., 2013).

The appearance of amino acid-based shuttles, at least in *Flaveria* genus, most likely emerged during the C_3-C_4 Type I and II intermediate phases, in order to assist the nitrogen-imbalanced glycine/serine shuttle linked to photorespiration (Mallmann et al., 2014). This is corroborated by our labelling data in the *Flaveria* paper included in this thesis; active pools of aspartate, glutamate and alanine start to rise already in the intermediate *F. ramosissima*, which employs C_2 photosynthesis as main CCM.

It seems that C_4 species, despite converging toward a series of traits that ultimately confer them a similar CCM mechanism, differ in many aspects, both physiological and morphological. Based on literature and our data, it seems that this is the case for the C_4 *Flaveria* species, maize and *Setaria*. Together with the apparent morphological differences, they have different metabolite profiles that give hints on possible differences in several shuttles, and differences in co-opted metabolites for shuttles. Which shuttle is utilized and related metabolic differences, at least in the genus *Flaveria*, reflect the evolutionary history of the C_4 pathway.

Metabolic perspectives on C_4 photosynthesis evolution in the genus *Flaveria*

Regarding the evolution of C_4 photosynthesis in the genus *Flaveria*, which was investigated in manuscript three of this thesis, we compared absolute levels of metabolites in several species, encompassing all the photosynthetic types. In this analysis we also included a small set of $^{13}CO_2$ labelling experiments. These were necessary to delineate an approximate size for the so-called "active" pool, defined as the pool of a specific metabolite that is actually involved in active photosynthetic metabolism. Whilst this was not essential for studies of the CBC, it was for study of metabolites like malate, aspartate, alanine and pyruvate that are involved in intercellular shuttles; these metabolites usually have substantial pools that are not involved in active photosynthesis (inactive pools), due to being partly located in the vacuole or, possibly, within non-photosynthetic cell types.

Only relatively small changes in metabolite levels occurred between basal C_3 species and C_3-C_4 intermediate species, especially Ser (absolute and active pool) and Asp (active pool only). The latter finding supports the increased carboxylation through PEPC (Chastain and Chollet, 1989) and the emergence of an aspartate-based shuttle to help equilibrate

nitrogen between BS and mesophyll cells (Mallmann et al., 2014), at least in advanced C_3-C_4 species like *F. ramosissima*.

On the other hand, large changes happened in the transition between C_3-C_4 and C_4-like species; we observed an increased content of CCM-related metabolites like malate, alanine and pyruvate (both total and active pools) and decrease contents in several photorespiratory metabolites like serine and glycerate (both total and active pools). These results are compatible with this transition, in which the C_4 cycle became the main carbon assimilation route and photorespiration gets heavily dampened (Ku et al., 1991).

The changes between C_4-like and full C_4 species were less marked, and affected especially levels of several organic acids and amino acids. We observed a moderate increase in the content of malate, glutamate (both total and active pools), 2-oxoglutarate and citrate (only absolute pools) and decreases in the content of aspartate (both absolute and active pools) and Ala (only active pool). The majority of these metabolites are directly involved in the C_4 CCM, however some others may support these cycles in an indirect manner by facilitating aminotransferase reactions (Mallmann et al., 2014; Schlüter et al., 2018), or providing a reservoir of carbon to allow faster build-up of the pools of metabolites that are directly involved in the CCM (Stitt and Zhu, 2014).

Concluding remarks

To conclude, the aims of this thesis were largely fulfilled: CBC metabolism has been systematically compared in different species by taking a targeted metabolite profiling approach, giving a much broader perspective than in previous literature. Moreover, metabolic differences were uncovered in one of the C_4 evolution-related most studied genera, *Flaveria*, revealing interesting details about single evolutionary steps. However, there are several aspects of our research that could be investigated further.

In terms of both CBC and C_4 evolution metabolic differences, it must be pointed out that a better coverage of species would definitely improve the picture that is currently forming. Including more C_3 species that are belonging to different plant families and have different growth optimums could tell us if the CBC trajectories we observe are family-specific or species-specific, and/or if these differences depend on the environment in which that species have evolved. More fine-grained studies of closely related species or even of genotypes within a species (like in Driever et al., 2014) could provide information about the speed of these evolutionary changes. Expanding the number of genera in C_4 evolution

studies will solidify or dismiss the somewhat general hypothesis which have been formulated based on the genus *Flaveria*, which might not be a common model for this CCM evolution (Brown et al., 2005; Williams et al., 2013).

We extracted lots of information by absolute metabolite measurements and active pools estimation by stable isotope labelling alone. However, these techniques could be brought to another level of precision by analyzing labelling kinetics over pulse and pulse+chase experiments, providing even more solid data for metabolic modelling, like it was done in our group for Arabidopsis (Szecowka et al., 2013) and maize (Arrivault et al., 2017). This level of detail, coupled with the detection of extra metabolites, is what we hope further studies will reach in order to have a clearer picture, especially regarding C_3-C_4 and C_4-like intermediates. Metabolite analyses could be integrated in the future with analyses of the related enzymatic activities. Indeed, our metabolite data points to which enzymes it might be interesting to compare across species. In the case of CBC, it would be interesting to measure activity and the regulatory properties of the pathway enzymes, especially FBPase and SBPase activities, in the species and conditions included in the first two manuscripts of this thesis, to confirm the metabolite results and our advanced hypothesis. Regarding C_4 evolution in the *Flaveria* genus, enzymatic activities of main C_4 pathway enzymes have been investigated extensively (Bauwe, 1984; Ku et al., 1991; Kubien et al., 2008; Gowik et al., 2011), however activities of key enzymes for intermediate species like GDC and aminotransferases have been neglected, especially in the light of different expression of these enzymes between mesophyll and BS tissues. In all cases, it would be also fascinating to tie any such changes in enzymatic activities or properties back to changes in the DNA sequence, either within the promoter (for changes in expression) or the open reading frame, like it was detailed in *Flaveria* for Rubisco (Kapralov et al., 2011) and PEPC (Westhoff and Gowik, 2004).

Bibliography

This bibliography section does not include the references from the three manuscripts included in this doctoral thesis. For manuscript-related bibliography, refer to that specifc manuscript References section.

Andersen KS, Bain JM, Bishop DG, Smillie RM (1972) Photosystem II activity in agranal bundle sheath chloroplasts from *Zea mays*. Plant Physiol **49**: 461–466

Anderson LE (1971) Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases. Biochim Biophys Acta **235**: 237–244

Andreou AI, Nakayama N (2018) Mobius assembly: a versatile golden-gate framework towards universal DNA assembly. PLoS One **13**: 1–18

Antonio C, Larson T, Gilday A, Graham I, Bergström E, Thomas-Oates J (2007) Quantification of sugars and sugar phosphates in *Arabidopsis thaliana* tissues using porous graphitic carbon liquid chromatography-electrospray ionization mass spectrometry. J Chromatogr A **1172**: 170–178

Arrivault S, Alexandre Moraes T, Obata T, Medeiros DB, Fernie AR, Boulouis A, Ludwig M, Lunn JE, Borghi GL, Schlereth A, et al (2019) Metabolite profiles reveal interspecific variation in operation of the Calvin-Benson cycle in both C_4 and C_3 plants. J Exp Bot **70**: 1843–1858

Arrivault S, Guenther M, Ivakov A, Feil R, Vosloh D, Van Dongen JT, Sulpice R, Stitt M (2009) Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. Plant J **59**: 824–839

Arrivault S, Obata T, Szecówka M, Mengin V, Guenther M, Hoehne M, Fernie AR, Stitt M (2017) Metabolite pools and carbon flow during C_4 photosynthesis in maize: $^{13}CO_2$ labeling kinetics and cell type fractionation. J Exp Bot **68**: 283–298

Arrivault SS, Guenther M, Fry SC, Fuenfgeld MMFFFF, Veyel D, Mettler-Altmann T, Stitt M, Lunn JE (2015) Synthesis and use of stable-isotope-labeled internal standards for quantification of phosphorylated metabolites by LC-MS/MS. Anal Chem **87**: 6896–6904

Badger MR, Andrews TJ, Whitney SM, Ludwig M, Yellowlees DC, Leggat W, Price GD (1998) The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. Can J Bot **76**: 1052–1071

Badger MR, Sharkey TD, von Caemmerer S (1984) The relationship between steady-state gas exchange of bean leaves and the levels of carbon-reduction-cycle intermediates. Planta **160**: 305–313

Bauwe H (1984) Photosynthetic enzyme activities and immunofluorescence studies on the localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in leaves of C_3, C_4, and C_3−C_4 intermediate species of Flaveria (Asteraceae). Biochem und Physiol der Pflanz **179**: 253–268

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

Bauwe H, Hagemann M, Kern R, Timm S (2012) Photorespiration has a dual origin and manifold links to central metabolism. Curr Opin Plant Biol **15**: 269–275

Bender MM (1971) Variations in the $^{13}C/^{12}C$ ratios of plants in relation to the pathway of photosynthetic carbon dioxide fixation. Phytochemistry **10**: 1239–1244

Blätke MA, Bräutigam A (2019) Evolution of C_4 photosynthesis predicted by constraint-based modelling. Elife **8**: 1–25

Borghi GL, Moraes TA, Günther M, Feil R, Mengin V, Lunn JE, Stitt M, Arrivault S, Raines C (2019) Relationship between irradiance and levels of Calvin-Benson cycle and other intermediates in the model eudicot Arabidopsis and the model monocot rice. J Exp Bot **70**: 5809–5825

Botha CEJ (1992) Plasmodesmatal distribution, structure and frequency in relation to assimilation in C_3 and C_4 grasses in southern Africa. Planta **187**: 348–358

Bowes G, Ogren WL, Hageman RH (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. Biochem Biophys Res Commun **45**: 716–722

Bräutigam A, Gowik U (2016) Photorespiration connects C_3 and C_4 photosynthesis. J Exp Bot **67**: 2953–2962

Bräutigam A, Weber APM (2010) Transport processes: connecting the reactions of C_4 photosynthesis. Adv. Photosynth. Respir. vol 32. Springer, Dordrecht, pp 199–219

Brown NJ, Parsley K, Hibberd JM (2005) The future of C_4 research - Maize, *Flaveria* or *Cleome*? Trends Plant Sci **10**: 215–221

Brown RH (1978) A difference in N use efficiency in C_3 and C_4 plants and its implications in adaptation and evolution. Crop Sci **18**: 93–98

Brown RH, Bouton JH (1993) Interspecific hybrids between photosynthetic types. Annu Rev Plant Physiol Plant Mol Biol **44**: 435–356

Buchanan BB, Balmer Y (2005) Redox regulation: a broadening horizon. Annu Rev Plant Biol **56**: 187–220

Busch FA, Sage TL, Cousins AB, Sage RF (2013) C_3 plants enhance rates of photosynthesis by reassimilating photorespired and respired CO_2. Plant, Cell Environ **36**: 200–212

von Caemmerer S, Furbank RT (2003) The C_4 pathway: an efficient CO_2 pump. Photosynth Res **77**: 191–207

von Caemmerer S, Quick WP, Furbank RT (2012) The development of C_4 rice: current progress and future challenges. Science (80-) **336**: 1671–1672

Cassman KG (1999) Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. Proc Natl Acad Sci **96**: 5952–9

Cassman KG, ed (1994) Breaking the yield barrier. work. rice yield potential favor. environ. IRRI - International Rice Research Institute, p 152

Chastain CJ, 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

Collatz GJ, Berry JA, Clark JS (1998) Effects of climate and atmospheric CO_2 partial pressure on the global distribution of C_4 grasses: present, past, and future. Oecologia **114**: 441–454

Covshoff S, Hibberd JM (2012) Integrating C_4 photosynthesis into C_3 crops to increase yield potential. Curr Opin Biotechnol **23**: 209–214

Cruz JA, Emery C, Wüst M, Kramer DM, Lange BM (2008) Metabolite profiling of Calvin cycle intermediates by HPLC-MS using mixed-mode stationary phases. Plant J **55**: 1047–1060

Danila FR, Quick WP, White RG, Furbank RT, von Caemmerer S (2016) The metabolite pathway between bundle sheath and mesophyll: quantification of plasmodesmata in leaves of C_3 and C_4 monocots. Plant Cell **28**: 1461–1471

Dietz KJ, Heber U (1984) Rate-limiting factors in leaf photosynthesis. I. Carbon fluxes in the calvin cycle. BBA - Bioenerg **767**: 432–443

Driever SM, Lawson T, Andralojc PJ, Raines CA, Parry MAJ (2014) Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. J Exp Bot **65**: 4959–4973

Ehleringer J, Björkman O (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, Pearcy RW (1983) Variation in quantum yield for CO_2 uptake among C_3 and C_4 plants. Plant Physiol **73**: 555–559

Eisenhut M, Bräutigam A, Timm S, Florian A, Tohge T, Fernie AR, Bauwe H, Weber APM (2017) Photorespiration is crucial for dynamic response of photosynthetic metabolism and stomatal movement to altered CO_2 availability. Mol Plant **10**: 47–61

Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A, Hagemann M (2008) The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proc Natl Acad Sci U S A **105**: 17199–17204

Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JDG, Patron NJ, Marillonnet S (2014) A Golden Gate modular cloning toolbox for plants. ACS Synth Biol **3**: 839–843

Ermakova M, Danila FR, Furbank RT, Caemmerer S Von, von Caemmerer S (2019) On the road to C_4 rice: advances and perspectives. Plant J 1–11

Evans JR (2013) Improving photosynthesis. Plant Physiol **162**: 1780–1793

Evans JR (1989) Photosynthesis and nitrogen relationships in leaves of C_3 plants. Oecologia **78**: 9–19

FAO (2018) FAOSTAT - Statistical Database. http://www.fao.org/faostat/en/#home

Farquhar GD, Sharkey TD (1982) Stomatal conductance and photosynthesis. Annu Rev Plant Physiol **33**: 317–345

Feller U, Anders I, Mae T (2007) Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. J Exp Bot **59**: 1615–1624

Fernie AR, Bauwe H, Eisenhut M, Florian A, Hanson DT, Hagemann M, Keech O, Mielewczik M, Nikoloski Z, Peterhänsel C, et al (2013) Perspectives on plant photorespiratory metabolism. Plant Biol **15**: 748–753

Fisher AE, McDade LA, Kiel CA, Khoshravesh R, Johnson MA, Stata M, Sage TL, Sage RF (2015) Evolutionary history of *Blepharis* (Acanthaceae) and the origin of C_4 photosynthesis in section Acanthodium. Int J Plant Sci **176**: 770–790

Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA (2002) GLK gene pairs regulate chloroplast development in diverse plant species. Plant J **31**: 713–727

Fouracre JP, Ando S, Langdale JA (2014) Cracking the Kranz enigma with systems biology. J Exp Bot **65**: 3327–3339

Furbank RT (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? J Exp Bot **62**: 3103–3108

Galmés J, Kapralov M V., Andralojc PJ, Conesa MÀ, Keys AJ, Parry MAJ, Flexas J (2014) Expanding knowledge of the Rubisco kinetics variability in plant species: Environmental and evolutionary trends. Plant, Cell Environ **37**: 1989–2001

Ghannoum O (2009) C_4 photosynthesis and water stress. Ann Bot **103**: 635–644

Ghannoum O, Evans JR, von Cammerer S (2011) Nitrogen and water use efficiency of C_4 plants. *In* A Raghavendra, RF Sage, eds, C4 Photosynth. Relat. CO_2 Conc. Mech. Dordrecht: Springer, pp 129–146

Gowik U, Bräutigam A, Weber KL, Weber APM, 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

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

Haberlandt G (1918) Physiologische pflanzenanatomie. W. Engelmann

Hagemann M, Kern R, Maurino VG, Hanson DT, Weber APM, Sage RF, Bauwe H (2016) Evolution of photorespiration from cyanobacteria to land plants, considering protein phylogenies and acquisition of carbon concentrating mechanisms. J Exp Bot **67**: 2963–2976

Hasunuma T, Harada K, Miyazawa SI, Kondo A, Fukusaki E, Miyake C (2010) Metabolic turnover analysis by a combination of *in vivo* ^{13}C-labelling from $^{13}CO_2$ and metabolic profiling with CE-MS/MS reveals rate-limiting steps of the C_3 photosynthetic pathway in *Nicotiana tabacum* leaves. J Exp Bot **61**: 1041–1051

Hatch MD, Slack CR (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. Biochem J **101**: 103–11

Hatch MD, Slack CR (1970) Photosynthetic CO_2-fixation pathways. Annu Rev Plant Physiol **21**: 141–162

Häusler RE, Hirsch HJ, Kreuzaler F, 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 (2016) C_4 photosynthesis evolution: the conditional Mt. Fuji. Curr Opin Plant Biol **31**: 149–154

Heckmann D, Schulze S, Denton AK, Gowik U, Westhoff P, Weber APM, Lercher MJ (2013) Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell **153**: 1579–88

Heldt HW, Piechulla B, Heldt F (2005) Plant Biochemistry. Plant Biochem. doi: 10.1016/B978-0-12-088391-2.X5000-7

Hibberd JM, Quick WP (2002) Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. Nature **415**: 451–454

Hibberd JM, Sheehy JE, Langdale JA (2008) Using C_4 photosynthesis to increase the yield of rice-rationale and feasibility. Curr Opin Plant Biol **11**: 228–231

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. Plant J **6**: 271–282

Höfer MU, Santore UJ, Westhoff P (1992) Differential accumulation of the 10-, 16- and 23-kDa peripheral components of the water-splitting complex of photosystem II in mesophyll and bundle-sheath chloroplasts of the dicotyledonous C_4 plant *Flaveria trinervia* (Spreng.) C. Mohr. Planta **186**: 304–312

Huege J, Sulpice R, Gibon Y, Lisec J, Koehl K, Kopka J (2007) GC-EI-TOF-MS analysis of in vivo carbon-partitioning into soluble metabolite pools of higher plants by monitoring isotope dilution after $^{13}CO_2$ labelling. Phytochemistry **68**: 2258–2272

Ibrahim DG, Burke T, Ripley BS, Osborne CP (2009) A molecular phylogeny of the genus *Alloteropsis* (Panicoideae, Poaceae) suggests an evolutionary reversion from C_4 to C_3 photosynthesis. Ann Bot **103**: 127–136

Ingram AL, Christin PA, Osborne CP (2011) Molecular phylogenies disprove a hypothesized C_4 reversion in *Eragrostis walteri* (Poaceae). Ann Bot **107**: 321–325

Kapralov M V., Kubien DS, Andersson I, Filatov DA (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

Keeley JE, Rundel PW (2003) Evolution of CAM and C_4 carbon-concentrating mechanisms. Int J Plant Sci **164**: 54–77

Keerberg O, Pärnik T, Ivanova H, Bassüner B, Bauwe H (2014) C_2 photosynthesis generates about 3-fold elevated leaf CO_2 levels in the C_3-C_4 intermediate species *Flaveria pubescens*. J Exp Bot **65**: 3649–3656

Kelly GJ, Latzko E (1976) Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycolate. FEBS Lett **68**: 55–58

Khoshravesh R, Stata M, Busch FA, Saladie M, Castelli JM, Dakin N, Hattersley PW, Macfarlane TD, Sage RF, Ludwig M, et al (2019) The evolutionary origin of C_4 photosynthesis in the grass subtribe Neurachninae. Plant Physiol **182**: pp.00925.2019

Kiniry JR, Jones CA, O'toole JC, Blanchet R, Cabelguenne M, Spanel DA (1989) Radiation-use efficiency in biomass accumulation prior to grain-filling for five grain-crop species. F Crop Res **20**: 51–64

Kopriva S, Chu CC, 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 Environ **19**: 1028–1036

Kortschak HP, Hartt CE, Burr GO (1965) Carbon dioxide fixation in sugarcane leaves. Plant Physiol **40**: 209–213

Ku MSB, Wu J, Dai Z, Scott R a, Chu C, Edwards GE (1991) Photosynthetic and photorespiratory characteristics of *Flaveria* species. Plant Physiol **96**: 518–528

Kubien DS, Whitney SM, Moore P V., Jesson LK (2008) The biochemistry of Rubisco in *Flaveria*. J. Exp. Bot. pp 1767–1777

Kubis A, Bar-Even A (2019) Synthetic biology approaches for improving photosynthesis. J Exp Bot **70**: 1425–1433

Lacoste-Royal G, Gibbs SP (1987) Immunocytochemical localization of ribulose-1,5-bisphosphate carboxylase in the pyrenoid and thylakoid region of the chloroplast of *Chlamydomonas reinhardtii*. Plant Physiol **83**: 602–606

Laetsch WM, Price I (1969) Development of the dimorphic chloroplasts of sugar cane. Am J Bot **56**: 77–87

Lawson T, Kramer DM, Raines CA (2012) Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. Curr Opin Biotechnol **23**: 215–220

Lawson T, Vialet-Chabrand S (2018) Speedy stomata, photosynthesis and plant water use efficiency. New Phytol 93–98

Leegood RC, von Caemmerer S (1988) The relationship between contents of photosynthetic metabolites and the rate of photosynthetic carbon assimilation in leaves of *Amaranthus edulis* L. Planta **174**: 253–262

Leegood RC, von Caemmerer S (1989) Some relationships between contents of photosynthetic intermediates and the rate of photosynthetic carbon assimilation in leaves of *Zea mays* L. Planta **178**: 258–266

Li G, Jain R, Chern M, Pham NT, Martin JA, Wei T, Schackwitz WS, Lipzen AM, Duong PQ, Jones KC, et al (2017) The sequences of 1504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. Plant Cell **29**: 1218–1231

Li L, Qu R, de Kochko A, Fauquet C, Beachy RN (1993) An improved rice transformation system using the biolistic method. Plant Cell Rep **12**: 250–255

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al (2010) The developmental dynamics of the maize leaf transcriptome. Nat Genet **42**: 1060–1067

Lin YS, Medlyn BE, Duursma RA, Prentice IC, Wang H, Baig S, Eamus D, De Dios VR, Mitchell P, Ellsworth DS, et al (2015) Optimal stomatal behaviour around the world. Nat Clim Chang **5**: 459–464

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc **1**: 387–396

Long SP (1983) C_4 photosynthesis at low temperatures. Plant Cell Environ **6**: 345–363

Long SP, Zhu XG, Naidu SL, Ort DR (2006) Can improvement in photosynthesis increase crop yields? Plant, Cell Environ **29**: 315–330

López-Calcagno PE, Fisk S, Brown KL, Bull SE, South PF, Raines CA (2019) Overexpressing the H-protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants. Plant Biotechnol J **17**: 141–151

Lundgren MR, Osborne CP, Christin PA (2014) Deconstructing Kranz anatomy to understand C_4 evolution. J Exp Bot **65**: 3357–3369

Lunn JE, Feil R, Hendriks JHM, Gibon Y, Morcuende R, Osuna D, Scheible W-R, Carillo P, Hajirezaei M-R, Stitt M (2006) Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADP-glucose pyrophosphorylase and higher rates of starch synthesis in *Arabidopsis thaliana*. Biochem J **397**: 139–148

Lyu MJA, Gowik U, Kelly S, Covshoff S, Mallmann J, Westhoff P, Hibberd JM, Stata M, Sage RF, Lu H, et al (2015) RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria* (Asteraceae) with some modifications: evolutionary developmental biology and morphology. BMC Evol Biol **15**: 116

Ma F, Jazmin LJ, Young JD, Allen DK (2014) Isotopically nonstationary ^{13}C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. Proc Natl Acad Sci **111**: 16967–16972

Makino A, Mae T, Ohira K (1985) Photosynthesis and ribulose-1,5-bisphosphate carboxylase/oxygenase in rice leaves from emergence through senescence. Quantitative analysis by carboxylation/oxygenation and regeneration of ribulose 1,5-bisphosphate. Planta **166**: 414–420

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U (2014) The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. Elife **2014**: 1–23

Marshall DM, Muhaidat R, Brown NJ, Liu Z, Stanley S, Griffiths H, Sage RF, Hibberd JM (2007) *Cleome*, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. Plant J **51**: 886–896

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

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

Meister M, Agostino A, Hatch MD (1996) The roles of malate and aspartate in C_4 photosynthetic metabolism of *Flaveria bidentis* (L.). Planta **199**: 262–269

Mitchell PL, Sheehy JE (2006) Supercharging rice photosynthesis to increase yield. New Phytol **171**: 685–687

Miyao M, Masumoto C, Miyazawa SI, Fukayama H (2011) Lessons from engineering a single-cell C_4 photosynthetic pathway into rice. J Exp Bot **62**: 3021–3029

Monson RK, Edwards GE, Ku MSB (1984) C_3-C_4 intermediate photosynthesis. Plants. Bioscience **34**: 563–566

Monson RK, Moore B d. (1989) On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. Plant Cell Environ **12**: 689–699

Moore B, Ku MSB, Edwards GE (1989) Expression of C_4-like photosynthesis in several species of *Flaveria*. Plant Cell Environ **12**: 541–549

Morell MK, Paul K, Kane HJ, Andrews TJ (1992) Rubisco: maladapted or misunderstood? Aust J Bot **40**: 431–441

Muhaidat R, Sage TL, Frohlich MW, Dengler NG, Sage RF (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

Nakajima Munekage Y (2016) Light harvesting and chloroplast electron transport in NADP-malic enzyme type C_4 plants. Curr Opin Plant Biol **31**: 9–15

Nakamura N, Iwano M, Havaux M, Yokota A, Munekage YN (2013) Promotion of cyclic electron transport around photosystem I during the evolution of NADP-malic enzyme-type C_4 photosynthesis in the genus *Flaveria*. New Phytol **199**: 832–842

Nelson T, Langdale J (1992) Developmental genetics of C_4 photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 25–47

Newsholme EA, Start C (1973) Regulation in metabolism.

Norman EG, Colman B (1991) Purification and characterization of phosphoglycolate phosphatase from the cyanobacterium *Coccochloris peniocystis*. Plant Physiol **95**: 693–698

Oaks A (1994) Efficiency of nitrogen utilization in C_3 and C_4 cereals. Plant Physiol 407–414

Osmond CB (1978) Crassulacean Acid Metabolism: a curiosity in context. Annu Rev Plant Physiol **29**: 379–414

P'Yankov VI, Voznesenskaya E V., Kondratschuk A, Black CC (1997) A comparative anatomical and biochemical analysis in *Salsola* (Chenopodiaceae) species with and without a Kranz type leaf anatomy: a possible reversion of C_4 to C_3 photosynthesis. Am J Bot **84**: 597–606

Parry MAJ, Andralojc PJ, Scales JC, Salvucci ME, Carmo-Silva AE, Alonso H, Whitney SM (2013) Rubisco activity and regulation as targets for crop improvement. J Exp Bot **64**: 717–730

Parry MAJ, Keys AJ, Madgwick PJ, Carmo-Silva AE, Andralojc PJ (2008) Rubisco regulation: a role for inhibitors. J Exp Bot **59**: 1569–1580

Pearcy RW, Ehleringer JR (1984) Comparative ecophysiology of C_3 and C_4 plants. Plant Cell Environ **7**: 1–13

Pick TR, Bräutigam A, Schlüter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R, Rascher U, Sonnewald U, et al (2011) Systems analysis of a maize leaf developmental gradient redefines the current C_4 model and provides candidates for regulation. Plant Cell **23**: 4208–4220

Powell AM (1978) Systematics of *Flaveria* (Flaveriinae-Asteraceae). Missouri Bot Gard Press **65**: 590–636

Price GD, Coleman JR, Badger MR (1992) Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942. Plant Physiol **100**: 784–793

Prins A, Orr DJ, Andralojc PJ, Reynolds MP, Carmo-Silva E, Parry MAJ (2016) Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. J Exp Bot **67**: 1827–1838

Raines CA (2006) Transgenic approaches to manipulate the environmental responses of the C_3 carbon fixation cycle. Plant, Cell Environ **29**: 331–339

Ramazanov Z, Rawat M, Henk MC, Mason CB, Matthews SW, Moroney J V. (1994) The induction of the CO_2-concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of *Chlamydomonas reinhardtii*. Planta **195**: 210–216

Ranson SL, Thomas M (1960) Crassulacean Acid Metabolism. Annu Rev Plant Physiol **11**: 81–110

Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. Nature **455**: 1101–1104

Ratcliffe RG, Shachar-Hill Y (2006) Measuring multiple fluxes through plant metabolic networks. Plant J **45**: 490–511

Rawsthorne S (1992) C_3–C_4 intermediate photosynthesis: linking physiology to gene expression. Plant J **2**: 267–274

Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA (2012) Recent patterns of crop yield growth and stagnation. Nat Commun **3**: 1–7

Rolleston FS (1972) A theoretical background to the use of measured concentrations of intermediates in study of the control of intermediary metabolism. Curr. Top. Cell. Regul. pp 47–75

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

Sage RF (2002) Variation in the k_{cat} of Rubisco in C_3 and C_4 plants and some implications for photosynthetic performance at high and low temperature. J Exp Bot **53**: 609–620

Sage RF (2017) A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. J Exp Bot **68**: e11–e28

Sage RF (2003) The evolution of C_4 photosynthesis. New Phytol **161**: 341–370

Sage RF, Adachi S, Hirasawa T (2017) Improving photosynthesis in rice: from small steps to giant leaps. 1–32

Sage RF, Khoshravesh R, Sage TL (2014) From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. J Exp Bot **65**: 3341–3356

Sage RF, Kubien DS (2007) The temperature response of C_3 and C_4 photosynthesis. Plant, Cell Environ **30**: 1086–1106

Sage RF, Monson RK, eds (1999) C_4 Plant Biology. Academic Press

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

Sage TL, Busch FA, Johnson DC, Friesen PC, Stinson CR, Stata M, Sultmanis S, Rahman BA, Rawsthorne S, Sage RF (2013a) Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. Plant Physiol **163**: 1266–1276

Sage TL, Busch FA, Johnson DC, Friesen PC, Stinson CR, Stata M, Sultmanis S, Rahman BA, Rawsthorne S, Sage RF (2013b) Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. Plant Physiol **163**: 1266–1276

Sage TL, Sage RF (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

Sato S, Soga T, Nishioka T, Tomita M (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. Plant J **40**: 151–163

Schlüter U, Bräutigam A, Droz JM, Schwender J, Weber APM (2018) The role of alanine and aspartate aminotransferases in C_4 photosynthesis. Plant Biol **21**: 64–76

Schlüter U, Weber APM (2020) Regulation and evolution of C_4 photosynthesis. Annu Rev Plant Biol **71**: 1–33

Schlüter U, Weber APM (2016) The road to C_4 photosynthesis: evolution of a complex trait via intermediary states. Plant Cell Physiol **57**: 881–889

Schuler ML, Mantegazza O, Weber APM (2016) Engineering C_4 photosynthesis into C_3 chassis in the synthetic biology age. Plant J **87**: 51–65

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P (2013) Evolution of C_4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 pump. Plant Cell **25**: 2522–2535

Seemann JR, Sharkey TD (1986) Salinity and nitrogen effects on photosynthesis, ribulose-1,5-bisphosphate carboxylase and metabolite pool sizes in *Phaseolus vulgaris* L. Plant Physiol **82**: 555–560

Servaites JC, Geiger DR, Tucci M a, Fondy BR (1989) Leaf carbon metabolism and metabolite levels during a period of sinusoidal light. Plant Physiol **89**: 403–408

Shively JM, Ball F, Brown DH, Saunders RE (1973) Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. Science (80-) **182**: 584–586

Simkin AJ, Lopez-Calcagno PE, Davey PA, Headland LR, Lawson T, Timm S, Bauwe H, Raines CA (2017) Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphophate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO_2 assimilation, vegetative biomass and seed yield in Arabidopsis. Plant Biotechnol J **15**: 805–816

Slewinski TL, Anderson AA, Zhang C, Turgeon R (2012) Scarecrow plays a role in establishing Kranz anatomy in maize leaves. Plant Cell Physiol **53**: 2030–2037

South PF, Cavanagh AP, Liu HW, Ort DR (2019) Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. Science (80-) **363**: eaat9077

Sternberg LDSL, Deniro MJ, Sloan ME, Black CC (1986) Compensation point and isotopic characteristics of C_3-C_4 intermediates and hybrids in *Panicum*. Plant Physiol **80**: 242–245

Still CJ, Berry JA, Collatz GJ, DeFries RS (2003) Global distribution of C_3 and C_4 vegetation: carbon cycle implications. Global Biogeochem Cycles **17**: 6-1-6–14

Stitt M (1990) Fructose 2,6-bisphosphate as a regulatory molecule in plants. Annu Rev Plant Physiol Plant Mol Biol **41**: 153–185

Stitt M, Heldt HW (1985a) Control of photosynthetic sucrose synthesis by fructose 2,6-Bisphosphate : VI. regulation of the cytosolic fructose 1,6-bisphosphatase in spinach leaves by an interaction between metabolic intermediates and fructose 2,6-bisphosphate. Plant Physiol **79**: 599–608

Stitt M, Heldt HW (1985b) Generation and maintenance of concentration gradients between the mesophyll and bundle sheath in maize leaves. Biochim Biophys Acta **808**: 400–414

Stitt M, Herzog B, Heldt HW (1984a) Control of photosynthetic sucrose synthesis by fructose 2,6-bisphosphate: I. Coordination of CO_2 fixation and sucrose synthesis. Plant Physiol **75**: 548–553

Stitt M, Kurzel B, Heldt HW (1984b) Control of photosynthetic sucrose synthesis by fructose 2,6-bisphosphate: II. Partitioning between sucrose and starch. Plant Physiol **75**: 554–560

Stitt M, Lunn J, Usadel B (2010) Arabidopsis and primary photosynthetic metabolism - more than the icing on the cake. Plant J **61**: 1067–91

Stitt M, Mieskes G, Soling HD, Heldt HW (1982) On a possible role of fructose 2,6-bisphosphate in regulating photosynthetic metabolism in leaves. FEBS Lett **145**: 217–222

Stitt M, Wirtz W, Heldt HW (1983) Regulation of sucrose synthesis by cytoplasmic fructosebisphosphatase and sucrose phosphate synthase during photosynthesis in varying light and carbon dioxide. Plant Physiol **72**: 767–774

Stitt M, Zhu XG (2014) The large pools of metabolites involved in intercellular metabolite shuttles in C_4 photosynthesis provide enormous flexibility and robustness in a fluctuating light environment. Plant, Cell Environ **37**: 1985–1988

Suzuki Y, Fujimori T, Kanno K, Sasaki A, Ohashi Y, Makino A (2012) Metabolome analysis of photosynthesis and the related primary metabolites in the leaves of transgenic rice plants with increased or decreased Rubisco content. Plant, Cell Environ **35**: 1369–1379

Suzuki Y, Miyamoto T, Yoshizawa R, Mae T, Makino A (2009) Rubisco content and photosynthesis of leaves at different positions in transgenic rice with an overexpression of RBCS. Plant, Cell Environ **32**: 417–427

Suzuki Y, Wada S, Kondo E, Yamori W, Makino A (2019) Effects of co-overproduction of sedoheptulose-1,7-bisphosphatase and Rubisco on photosynthesis in rice. Soil Sci Plant Nutr **65**: 36–40

Szecowka M, Heise R, Tohge T, Nunes-Nesi A, Vosloh D, Huege J, Feil R, Lunn J, Nikoloski Z, Stitt M, et al (2013) Metabolic fluxes in an illuminated Arabidopsis rosette. Plant Cell **25**: 694–714

Tardieu F, Simonneau T (1998) Variability among species of stomatal control under fluctuating soil water status and evaporative demand: modelling isohydric and anisohydric behaviours. J Exp Bot **49**: 419–432

Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T (2014) Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C_4-related processes. J Exp Bot **65**: 3543–3555

Tcherkez GGB, Farquhar GD, Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proc Natl Acad Sci **103**: 7246–7251

Trudeau DL, Edlich-Muth C, Zarzycki J, Scheffen M, Goldsmith M, Khersonsky O, Avizemer Z, Fleishman SJ, Cotton CAR, Erb TJ, et al (2018) Design and *in vitro* realization of carbon-conserving photorespiration. Proc Natl Acad Sci U S A **115**: E11455–E11464

Usuda H (1987) Changes in levels of intermediates of the C_4 cycle and reductive pentose phosphate pathway under various light intensities in maize leaves . Plant Physiol **84**: 549–554

Usuda H (1985) Changes in levels of intermediates of the C_4 cycle and reductive pentose phosphate pathway during induction of photosynthesis in maize leaves. Plant Physiol **78**: 859–864

Usuda H (1986) Non-autocatalytic build-up of ribulose 1,5-bisphosphate during the initial phase of photosynthetic induction in maize leaves. Plant Cell Physiol **27**: 745–749

Voznesenskaya E V., Franceschi VR, Kiirats O, Freitag H, Edwards GE (2001) Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. Nature **414**: 543–546

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

Wallsgrove R, Keys AJ, Lea PJ, Miflin BJ (1983) Photosynthesis, photorespiration and nitrogen metabolism. Plant Cell Environ **6**: 301–309

Wang L, Peterson RB, Brutnell TP (2011) Regulatory mechanisms underlying C_4 photosynthesis. New Phytol **190**: 9–20

Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, et al (2013) Evolution of GOLDEN2-LIKE gene function in C_3 and C_4 plants. Planta **237**: 481–495

Wang P, Karki S, Biswal AK, Lin HC, Dionora MJ, Rizal G, Yin X, Schuler ML, Hughes T, Fouracre JP, et al (2017) Candidate regulators of early leaf development in maize perturb hormone signalling and secondary cell wall formation when constitutively expressed in rice. Sci Rep **7**: 1–15

Wang Y, Bräutigam A, Weber APM, Zhu XG (2014) Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. J Exp Bot **65**: 3567–3578

Waters MT, Moylan EC, Langdale JA (2008) GLK transcription factors regulate chloroplast development in a cell-autonomous manner. Plant J **56**: 432–444

Weber APM, von Caemmerer S (2010) Plastid transport and metabolism of C_3 and C_4 plants-comparative analysis and possible biotechnological exploitation. Curr Opin Plant Biol **13**: 256–264

Weiner H, Burnell JN, Woodrow IE, Heldt HW, Hatch MD (1988) Metabolite diffusion into bundle sheath cells from C_4 plants. Plant Physiol **88**: 815–822

Weissmann S, Ma F, Furuyama K, Gierse J, Berg H, Shao Y, Taniguchi M, Allen DK, Brutnell TP (2016) Interactions of C_4 subtype metabolic activities and transport in maize are revealed through the characterization of DCT2 mutants. Plant Cell **28**: 466–484

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

Williams BP, Johnston IG, Covshoff S, Hibberd JM, Abramoff M, Magalhaes P, Ram S, Adams C, Leung F, Sun S, et al (2013) Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. Elife **2**: e00961

Williams TCR, Poolman MG, Howden AJM, Schwarzlander M, Fell DA, Ratcliffe RG, Sweetlove LJ (2010) A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. Plant Physiol **154**: 311–323

Woo KC, Anderson JM, Boardman NK, Downton WJS, Osmond CB, Thorne SW (1970) Deficient photosystem II in agranal bundle sheath chloroplasts of C_4 Plants. Proc Natl Acad Sci **67**: 18–25

Wullschleger SD (1993) Biochemical limitations to carbon assimilation in C_3 plants—a retrospective analysis of the A/C_i curves from 109 species. J Exp Bot **44**: 907–920

Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nat Rev Microbiol **6**: 681–691

Zachos JC, Dickens GR, Zeebe RE (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature **451**: 279–283

Zelitch I, Schultes NP, Peterson RB, Brown P, Brutnell TP (2009) High glycolate oxidase activity is required for survival of maize in normal air. Plant Physiol **149**: 195–204

Zhu X-G, de Sturler E, Long SP (2007) Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. Plant Physiol **145**: 513–526

Zhu XG, Govindjee, Baker NR, DeSturler E, Ort DR, Long SP (2005) Chlorophyll a fluorescence induction kinetics in leaves predicted from a model describing each discrete step of excitation energy and electron transfer associated with Photosystem II. Planta **223**: 114–133

Zhu XG, Long SP, Ort DR (2008) What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Curr Opin Biotechnol **19**: 153–159

Zhu XG, Wang Y, Ort DR, Long SP (2013) e-photosynthesis, a comprehensive dynamic mechanistic model of C_3 photosynthesis: from light capture to sucrose synthesis. Plant, Cell Environ **36**: 1711–1727