A systems biology approach and single gene analysis to identify transporters and regulatory proteins in the plant carbon cycle

Kumulative Dissertation

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

vorgelegt von

Thea Renate Pick

aus Aachen

Düsseldorf, September 2012
Aus dem
Institut Biochemie der Pflanzen
der Heinrich-Heine-Universität Düsseldorf

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

Referent: Prof. Dr. Andreas P. M. Weber
Korreferent: Prof. Dr. Peter Westhoff

Tag der mündlichen Prüfung: 08.10.2012
Eidesstattliche Versicherung und Selbstständigkeitserklärung


Düsseldorf, den 10.10.2012

________________________
Thea Pick
Für meine Familie...
Summary

Green plants fix inorganic carbon dioxide (CO₂) by Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) in a process called photosynthesis. This carboxylation reaction of Rubisco generates two molecules of 3-phosphoglycerate (3-PGA) and finally leads to biomass production. Hence plants account for a significant share of the global organic carbon pool. While most plants use the original C₃ type of photosynthesis for CO₂ assimilation some of the most important crop plants on earth like maize (Zea mays) use a highly efficient type of photosynthesis, namely C₄ photosynthesis.

In C₄ photosynthesis CO₂ is prefixed by an oxygen (O₂) insensitive enzyme into a C₄ acid and concentrated in close vicinity of Rubisco. This leads to a high efficiency of photosynthesis through a reduced rate of photorespiration that is caused by the oxygenation reaction of Rubisco. Although Rubisco strongly favors CO₂, the enzyme can also accept O₂ as a substrate. The product of the oxygenation reaction is one molecule of 3-PGA and one molecule of 2-phosphoglycolate (2-PG). The latter can neither enter the Calvin cycle nor be converted to carbohydrates. This can lead to a dramatic decrease in biomass production since 2-PG has to be recycled at the cost of energy and reduction equivalents in the photorespiratory cycle.

The photorespiratory cycle is a highly compartmentalized pathway involving metabolic reactions in the chloroplasts, peroxisomes, mitochondria, and the cytosol. The mass flow through this pathway is only excelled by photosynthesis and therefore requires a tight and fast connection between the different compartments. This is obtained through highly specific metabolite transporters that account for the shuttle of precursors, intermediates and products across the membranes. It has previously been shown that short-circuiting of the photorespiratory cycle led to enhanced biomass production in C₃ plants. This important finding verifies the significance of research for molecular plant engineering as a significant share of the components of the pathway, such as metabolite transporters and regulatory proteins are still unknown.

Therefore one aim of this thesis was the identification of photorespiratory metabolite transporters since all genes encoding enzymes required for appropriate function of the core cycle are known to date while no transporter has been characterized on the molecular level. In a reverse genetics screen Plastidic glycolate glycerate transporter 1 (PLGG1) was designated as a promising candidate. Molecular characterization of the candidate was used to identify expression pattern and subcellular localization. An A. thaliana (Arabidopsis thaliana) T-DNA insertional knockout mutant (plgg1-1) was isolated and characterized concerning its phenotype and steady state and dynamic metabolite accumulation. Biochemical characterization of the candidate was used to identify transport capacity and to define the role of PLGG1 in plant metabolism. The molecular characterization revealed that PLGG1 is expressed in all green
tissues and that it is located at the chloroplast envelope. Analysis of the knockout mutant revealed that \textit{plgg1-1} plants show a photorespiratory phenotype and the photorespiratory metabolites glycolate and glycerate accumulate in the mutant. Finally the biochemical characterization indicated that glycolate and glycerate transport are impaired in the mutant. These analyses lead to the discovery of PLGG1 as the chloroplastidic glycolate glycerate transporter and thereby to the identification of the first transporter of the core photorespiratory pathway.

Surprisingly PLGG1 was found to be high abundant in maize chloroplast what is somehow counter intuitive as PLGG1 is a photorespiratory transporter and the rate of photorespiration is reduced in the C₄ plant maize. Until today the role of PLGG1 in C₄ photosynthesis is still unresolved and emphasizes the question how the C₄ cycle is organized and how the establishment of C₄ photosynthesis is orchestrated since the C₄ cycle is obviously not as linear and simple as it has been assumed.

Therefore the second aim of this thesis was to set up a systems biology approach to find candidates for unidentified components of the C₄ pathway, such as metabolite transporters and regulatory proteins. The third leaf of the C₄ model plant maize was analyzed along a base-to-tip developmental gradient. Ten continuous leaf slices were analyzed individually and transcriptome analysis, oxygen sensitivity of photosynthesis, photosynthetic rate measurements, and chlorophyll and protein measurements revealed a gradual sink-to-source transition for the leaf without a binary on-off switch. Analysis of transcription factors exhibiting a similar expression pattern to key C₄ enzymes along the leaf gradient designated in a list of putative regulatory proteins orchestrating the establishment of C₄ photosynthesis. Finally, transcriptome and metabolome analysis, enzyme activity measurements, and quantification of selected metabolites showed that the C₄ cycle is not linear but rather a branched cycle. These results led to a revised model of maize C₄ photosynthesis.
Zusammenfassung


Daher war ein Ziel dieser Arbeit photorespiratorische Metabolit Transporter zu identifizieren, da alle Gene, die für unabdingbare Enzyme des Hauptzyklus kodieren, bekannt sind, während bisher kein Transporter molekular charakterisiert werden konnte. Mit Hilfe reverser Genetik konnte Plastidic glycolate glycerate transporter 1 (PLGG1) als ein vielversprechender Kandidat isoliert werden. Die molekulare Charakterisierung des Kandidaten wurde zur Untersuchung des Expressionsmusters und der zellulären Lokalisation verwendet.

Überraschender Weise wurde zudem festgestellt, dass PLGG1 sehr stark in Mais Chloroplasten vertreten ist, was eigentlich widersprüchlich ist, da PLGG1 ein photorespiratorischer Transporter ist und die Photoresurationsrate in der C₄ Pflanze Mais stark reduziert ist. Bisher ist die Rolle von PLGG1 in der C₄ Photosynthese noch ungeklärt und verstärkt die Frage wie der C₄ Zyklus organisiert ist und wie die Ausbildung des C₄ Syndroms gesteuert wird, da der C₄ Zyklus offenbar nicht so linear und schlicht ist, wie es bisher angenommen wurde.

Table of contents

1. Abbreviations ............................................................................................................................................................... 2
2. Introduction .................................................................................................................................................................. 4
   2.1 Motivation .............................................................................................................................................................. 4
   2.2 Photorespiration ..................................................................................................................................................... 4
      2.2.1 Photorespiratory transporters ..................................................................................................................... 7
      2.2.2 The chloroplastidic glycolate glycerate transporter ................................................................................. 7
   2.3 A. thaliana mutants impaired in the photorespiratory pathway ................................................................. 8
   2.4 C_4 photosynthesis ................................................................................................................................................. 8
   2.5 Reverse and forward genetics as a tool to find missing links in plant metabolism and development ......................................................................................................................................................... 9
3. Discussion and Outlook ........................................................................................................................................ 11
4. Theses ........................................................................................................................................................................... 13
5. References .................................................................................................................................................................. 14
6. Manuscripts ................................................................................................................................................................ 19
   6.1 First publication: Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C_4 Model and Provides Candidates for Regulation ................................................................. 20
   6.2 Second publication: PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a new class of metabolite transporters ................................................... 45
1. Abbreviations

2-OG  2-oxoglutarate  
2-PG  2-phosphoglycolate  
3-PGA  3-phosphoglycerate  
\textit{A. thaliana}  \textit{Arabidopsis thaliana}  
ATP  adenine triphosphate  
C\textsubscript{3}, C\textsubscript{4}  three-, four-carbon molecule  
CAT  catalase  
CO\textsubscript{2}  carbon dioxide  
DiT1  2-oxoglutarate/malate translocator  
DiT2  glutamate/malate translocator  
G2  Golden 2 transcription factor  
GDC  glycine decarboxylase  
GGT  glutamate:glyoxylate aminotransferase  
Glk  G2-like  
Glu  glutamate  
GLYK  glycerate kinase  
GO  glycolate oxidase  
GOX  flavin-dependent glycolate oxidase  
GS/GOGAT  glutamine synthetase/glutamate synthase  
H\textsubscript{2}O  water  
H\textsubscript{2}O\textsubscript{2}  hydrogen peroxide  
HPR  hydoxypyruvate reductase  
mMDH  mitochondrial malate dehydrogenase  
MS  malate synthase  
NAD  nicotinamide adenine dinucleotide  
NADH  reduced nicotinamide adenine dinucleotide
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH₄⁺</td>
<td>ammonium ion</td>
</tr>
<tr>
<td>O₂</td>
<td>oxygen</td>
</tr>
<tr>
<td>Pᵢ</td>
<td>orthophosphoric acid</td>
</tr>
<tr>
<td>PEPC</td>
<td>Phosphoenolpyruvate carboxylase</td>
</tr>
<tr>
<td>PGLP</td>
<td>2-phosphoglycolate phosphatase</td>
</tr>
<tr>
<td>PGP</td>
<td>phosphoglycolate phosphatase</td>
</tr>
<tr>
<td>PLGG1</td>
<td>Plastidic glycolate glycerate transporter 1</td>
</tr>
<tr>
<td>pMDH</td>
<td>peroxisomal malate dehydrogenase</td>
</tr>
<tr>
<td>Rubisco</td>
<td>ribulose 1,5-bisphosphate carboxylase/oxygenase</td>
</tr>
<tr>
<td>RuBP</td>
<td>ribulose-bisphosphate</td>
</tr>
<tr>
<td>SHMT</td>
<td>serine hydroxymethyl transferase</td>
</tr>
<tr>
<td>SGT</td>
<td>serine:glyoxylate aminotransferase</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydofolate</td>
</tr>
<tr>
<td>Z. mays</td>
<td>Zea mays</td>
</tr>
</tbody>
</table>
2. Introduction

Research on photorespiration and C₄ photosynthesis has been of major interest for molecular plant breeding in the last decades. To date especially regulatory proteins and molecular transporters are of special interest as a significant share of components involved in those pathways is still unknown. This thesis is concerned with the construction of a data set to generate candidates for regulatory proteins and molecular transporters of the C₄ photorespiratory cycle as well as single candidate analysis for a molecular transporter of the photorespiratory cycle.

2.1 Motivation

Plants fix inorganic carbon in a process called photosynthesis. Hence plants are responsible for a considerable share of the world’s carbon pool and energy supply. However, in future we are facing severe problems concerning the conflict of growth of population and a boosted demand for biofuels and crops (Cassman and Liska, 2007). In this context approaches to generate (a) highly productive plants and (b) plants that are able to grow in harsh habitats and cope with critical environmental factors such as heat, salinity and drought are of major interest and current focus of research (Kajala et al., 2011; von Caemmerer et al., 2012). Approximately 15 to 40 million years ago, long after C₃ photosynthesis was established, a highly efficient type of photosynthesis, namely C₄ photosynthesis, evolved (Sage, 2004). Plants performing C₄ photosynthesis are able to grow in harsher habitats than C₃ plants. Therefore C₄ plants like maize are ideal models to study the adaptation to harsh environments and to use the knowledge to boost photosynthesis in C₃ plants like rice (Oryza sativa) (Kajala et al., 2011). A further approach to successfully enhance biomass production is short-circuiting of photorespiration (Kebeish et al., 2007; Maier et al., 2012). Hence identification of regulatory proteins and factors involved in both C₄ photosynthesis and photorespiration are of major interest.

2.2 Photorespiration

Photosynthesis is the light dependent process in which inorganic carbon is incorporated into carbohydrates and is thereby made accessible to heterotrophic organisms. Plants performing oxygenic photosynthesis are therefore indispensable for the land-based biosphere. Fixation of inorganic carbon is catalyzed by Rubisco that feeds CO₂ into the Calvin cycle finally leading to the production of the C₃ acid 3-PGA. Although Rubisco is the most abundant protein on earth, it is inefficient due to its catalytic activity to both carboxylate and oxygenate the acceptor ribulose-
2. Introduction

1,5-bisphosphate. The photosynthetic process evolved 2 billion years ago when the atmosphere was free of O₂ (Kirschvink and Kopp, 2008). Hence, the oxygenation reaction played a minor role during land plant evolution while it became significant with increasing atmospheric oxygen concentration 1 billion years ago. The product of this oxygenation reaction in one molecule 3-PGA and one molecule 2-PG. 2-PG is toxic to plants and can neither enter the Calvin cycle nor be converted to carbohydrates and inhibits enzymes of the Calvin cycle (Husic and Tolbert, 1987). Hence, 2-PG has to be detoxified in the photorespiratory cycle. The first enzymatic steps take place in the chloroplast stroma where 2-PG is dephosphorylated to glycolate by phosphoglycolate phosphatase (PGP). The glycolate is transported out of the chloroplast in counter-exchange with glyceraldehyde. Glycolate enters the peroxisomes and is oxidized to glyoxylate by flavin-dependent glycolate oxidase (GOX). Glyoxylate is transaminated to glycine by the two aminotransferases Ser:glyoxylate and Glu:glyoxylate aminotransferase (SGT and GGT, respectively). Glycine is further transported into the mitochondria where it is deaminated and decarboxylated by a multienzyme complex. This complex is called the glycine decarboxylase complex (GDC). One product of this reaction, the 5,10-methylene tetrahydrofolate (Collakova et al., 2008), is further converted by serine hydroxymethyltransferase (SHMT) to yield one molecule of serine from two molecules of Glycine (Somerville and Ogren, 1981, 1982). Serine is transported back into the peroxisomes and converted to glyceraldehyde by serine glyoxylate aminotransferase (SGT) and hydroxypropyruvate reductase (HPR). The glyceraldehyde is transported back into the chloroplasts and finally phosphorylated by glycokinate kinase (GLYK) to yield 3-PGA that is fed back to the Calvin cycle. The complete cycle is depicted in Figure 1.

Because previously fixed carbon is lost and energy consumed photorespiration has been considered as a wasteful process. Besides those disadvantageous effects photorespiration can also be beneficial for plants under certain stress conditions (Migge et al., 1999). For example under high light stress when the production of photochemical energy exceeds the consumption photorespiration can work as a valve and protect plants from photoinhibition (Kozaki and Takeba, 1996). In addition photorespiration can serve as a source for metabolites like glycine and serine that are used by the plant for other processes (e.g. Madore and Grodzinski, 1984). For instance glycine can be further converted to glutathione that is used in the plant antioxidant system (Gillham and Dodge, 1986; Noctor et al., 1999). Furthermore H₂O₂ that is produced in the peroxisomes during photorespiration is discussed to be involved in plant defense mechanisms (Taler et al., 2004).

It has previously been successfully shown in two independent approaches that short-circuiting of the photorespiratory cycle led to enhanced biomass production of C₃ plants (Kebeish et al., 2007; Maier et al., 2012).
Figure 1: The photorespiratory carbon cycle. Enzymatic reactions take place in the compartments chloroplasts, peroxisomes, and mitochondria; RubisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; 3-PGA, 3-phosphoglycerate; Glu, glutamate; RuBP, ribulose-bisphosphate; THF, tetrahydofolate. Black boxes symbolize transport proteins. The transport proteins represented in this scheme have not been characterized on the molecular level. The scheme was adapted from [Linka, 2008].
2. Introduction

2.2.1 Photorespiratory transporters

As described above the photorespiratory cycle is a highly compartmentalized pathway with metabolic reactions occurring in the chloroplast, the leaf peroxisomes, the mitochondria and the cytosol. Each of the compartments is surrounded by one (peroxisomes) or two membranes (chloroplast and mitochondria) that circumvent unregulated diffusion or leakage of metabolites. However, to allow the metabolic interaction of the different compartments in a controlled manner, transporters account for the shuttle of precursors, intermediates and products across the membranes. To control the selective flux of metabolites most transporters are considered to be highly specific concerning their transport substrate (Weber, 2004; Weber et al., 2004; Weber et al., 2005; Weber and Fischer, 2007). In C3 plants the ratio of carboxylation to oxygenation of Rubisco is approximately 3:1 when grown in ambient air with a CO₂ concentration of 380 ppm and O₂ concentration of 210,000 ppm (Sharkey, 1988). This results in a high flux rate of metabolites through the photorespiratory cycle. Hence, transporters involved in photorespiration have to cope with an immense metabolite flux and have to have distinct properties. Between 14 and 18 transmembrane transport processes are required for a functional photorespiratory carbon and nitrogen cycle. However, only DiT1 and DiT2 (transport proteins for 2-oxoglutarate and glutamate), both transporters of the nitrogen cycle, are characterized at a molecular level to date (Somerville and Ogren, 1983; Renne et al., 2003). Characterization of transporters of the carbon cycle of photorespiration still remained elusive.

2.2.2 The chloroplastidic glycolate glycerate transporter

In the 1980s in the lab of Howitz and McCarty specific transport activity in extracts of chloroplast envelope fractions was analyzed (Howitz and Mccarty, 1985a, b). Glycolate is produced in the chloroplast and transported out of this compartment. One of the final steps of the photorespiratory cycle is the production of glycerate and its transport into the chloroplast. Howitz and McCarty biochemically analyzed a chloroplastidic protein, which promotes the transport of glycolate and glycerate across the envelope membrane with rates required to cope with the flux through the photorespiratory pathway (Howitz and McCarty, 1991). Additionally, they postulated that the single protein is a transporter for both substrates, glycolate and glycerate, and that the transport mode is not necessarily stoichiometric. Hence, a proton-substrate-symport was measured and it was observed that the presence of the second substrate on the opposite side of the membrane stimulates transport (Howitz and McCarty, 1986, 1991). This is an important finding as for two molecules glycolate exported only one molecule glycerate is imported (Figure 1). Although these findings were completed decades ago the transporter was never characterized on the molecular level.
2. Introduction

2.3 *A. thaliana* mutants impaired in the photorespiratory pathway

The majority of molecular components of the photorespiratory pathway have been identified using *A. thaliana* mutants impaired in photorespiration. Prominent examples are mutants impaired in 2-phosphoglycolate phosphatase (PGLP) the enzyme catalyzing the first reaction of the photorespiratory cycle (Somerville and Ogren, 1979), or the mitochondrial SHMT (Somerville and Ogren, 1981; Voll et al., 2006).

A significant similarity of photorespiratory mutants is reduced or no growth under ambient air but normal wild type-like growth under CO₂ enriched air. Only peroxisomal HPR mutants show no typical photorespiratory growth limitation in ambient air due to an alternative cytosolic pathway that suppresses the effect of the mutation (Timm et al., 2008). An additional common observation is the accumulation of specific photorespiratory metabolites in the mutant plants and differences in chlorophyll fluorescence parameters between WT and mutants plants. These shared characteristics of photorespiratory mutants can be used as a tool to identify yet unknown factors involved in the photorespiratory cycle, as it was recently shown for the chlorophyll fluorescence imaging approach (Badger et al., 2009).

2.4 C₄ photosynthesis

As described above high rates of photorespiration can lead to a decrease in biomass production. During evolution several carbon concentrating mechanisms evolved to circumvent the oxygenation reaction of Rubisco and thereby inhibiting the negative effects of photorespiration. One of the most prominent mechanisms is C₄ photosynthesis. This special type of photosynthesis evolved 15 to 40 million years ago (reviewed in (Sage, 2004). In C₄ photosynthesis, CO₂ is prefixed by the O₂ insensitive enzyme phosphoenolpyruvate carboxylase (PEPC) into a C₄ acid and released by decarboxylation of the C₄ acid in close vicinity to Rubisco (Figure 2A). The pathway as it occurs in maize involves enzymatic reactions taking place in the cytosol and chloroplast of two morphological different cell types, namely mesophyll and bundle sheath cells (Figure 2B). This spatial separation results in an increased number of required membrane transport proteins compared to C₃ photosynthesis. Despite detailed understanding of the biochemical reactions in the pathway (Furbank, 2011) we still know comparatively little about the membrane proteins mediating the shuttle of metabolites. As a long term goal of plant researchers is to introduce the highly efficient C₄ photosynthetic cycle in C₃ plants it is also of interest to identify the regulatory proteins that orchestrate the establishment and onset of C₄ photosynthesis.
2.5 Reverse and forward genetics as a tool to find missing links in plant metabolism and development

As a tool to find regulatory and transport proteins involved in C₄ photosynthesis, photorespiration and other metabolic pathways in plants reverse and forward genetics have been of great advantage. For instance the common features of mutants impaired in the photorespiratory pathway (see 2.3) were used for successful forward genetic screens (Somerville and Ogren, 1979, 1980b, 1981; Wallsgrove et al., 1987; Voll et al., 2006).

In the past co-expression analysis as a reverse genetics approaches has been a well studied and successful method to find candidates for missing links in pathways. This analysis is based on the assumption that genes that are co-expressed function in the same pathway and was developed by (Eisen et al., 1998). Hence, unknown genes that are co-regulated with genes of a particular metabolic pathway are probably involved in the same biological process. By this method a wide range of genes have been functional characterized in non-plant organisms like yeast (Wu et al., 2002) or humans (Lee et al., 2004). In Arabidopsis thaliana a co-expression analysis in combination with reverse genetics has been a successfully strategy to find genes for example involved in flavonoid metabolism (Yonekura-Sakakibara et al., 2007), cellulose and...
(Persson et al., 2005) lignin biosynthesis (Ehlting et al., 2005; Alejandro et al., 2012) and aliphatic glucosinolate biosynthesis (Hansen et al., 2007).

An additional reverse genetics tool is the analysis of the whole transcriptome of an organism or different tissues of an organism. It has been successfully used by comparing the transcriptome of closely related C₃ and C₄ plants of the genus *Cleome* (Bräutigam et al., 2011). Furthermore, comparative transcriptomics were used in a recent approach to shed light on the question: “Evolution of C₄ photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C₄?” (Gowik et al., 2011). In this approach the transcriptome of five species within the genus *Flaveria*, including C₃ plants, C₃- C₄ intermediates and plants performing C₄ photosynthesis were analyzed to isolate new candidates for yet unknown factors in C₄ photosynthesis, as well as identify C₄ associated pathways.

A different approach is the analysis of the whole proteome of an organism, specific tissues or even specific organelles. For example, comparative proteomics of chloroplasts was used to identify the characteristic composition of proteins in chloroplast of C₄ plants and thereby finding candidates for yet unidentified proteins involved in C₄ photosynthesis. Here two different and independent proteomic approaches were used: on the one hand the set of proteins characteristic for chloroplasts of C₃ and C₄ plants were compared (Bräutigam et al., 2008) and on the other hand chloroplasts of the functionally different cell types bundle sheath and mesophyll cells (Figure 2) in the C₄ plant *Z. mays* (Majeran et al., 2008).

Hence reverse and forward genetics are important tools to find missing links, like regulatory and transport proteins in C₄ photosynthesis and photorespiration.
3. Discussion and Outlook

Plants are able to fix inorganic carbon in photosynthesis and thereby contribute to a significant share of the worldwide carbon cycle what makes them indispensable and important for the land-based biosphere. Oxygenic photosynthesis evolved over 2 billion years ago when the atmosphere was free of O$_2$ (Kirschvink and Kopp, 2008). When oxygenic photosynthesis emerged the level of O$_2$ in the atmosphere 'rapidly' increased to relatively high levels about 1 billion years ago. This event resulted in a significant competition of O$_2$ and CO$_2$ for the active site of Rubisco, as Rubisco functions as carboxylase as well as oxygenase (Smith, 1976). The oxygenation reaction generates one molecule 2-PG that is toxic to the plant and has to be recycled in the photorespiratory pathway that, in the end, leads to a decrease in biomass production. 15 to 40 million years ago C$_4$ photosynthesis evolved, what is a highly efficient type of photosynthesis that minimizes the rate of photorespiration. In C$_4$ photosynthesis CO$_2$ is concentrated around Rubisco and therefore the oxygenation reaction is reduced.

Both, C$_4$ photosynthesis and photorespiration are major subjects of plant research that aims to develop highly efficient and robust C$_3$ plants. The aim of this thesis was to contribute to the knowledge about the unknown factors of these pathways, such as regulatory proteins and metabolite transporters. The long-term goal of studying factors involved in the establishment and the processing of C$_4$ photosynthesis and photorespiration is a "second green evolution" (von Caemmerer et al., 2012). In this project the C$_4$ cycle is inserted into the C$_3$ plant rice to increase the yield. An obstacle on the way to achieve this goal is the observation that the C$_4$ cycle is not a linear cycle but rather branched and involves several levels of regulation and organization (Bräutigam et al., 2011; Gowik et al., 2011).

Therefore a systems biology approach was used in this thesis to identify candidates for regulatory proteins that orchestrate the establishment of photosynthesis and to finally provide a comprehensive understanding of the dynamic interactions that take place between different cell types and cellular compartments (Pick et al., 2011). As carbon fixation by the oxygen insensitive enzyme PEPC and the fixed carbon entering the Calvin cycle is separated between different cell types a different set of proteins is required in each cell. Hence a highly coordinated expression in the specific cell type is important. The transcription factor Golden 2 (G2) was shown to be a bundle sheath specific transcription factor in maize (Roth et al., 1996). Additionally a whole family of G2-like (Glk) factors was identified in maize offering promising candidates for further research (Rossini et al., 2001). A second attempt to investigate cell specific expression is the analysis of cis-regulatory elements responsible for cell-specific gene expression. The promoter elements responsible for bundle sheath specific expression were identified for the P- subunit of the GDC in the C$_4$ plant *Flaveria trinervia* (Engelmann et al., 2008; Wiludda et al., 2012) offering an alternative approach. Still, one major gap in the knowledge about the C$_4$ photorespiratory
cycle are important metabolite transporters that facilitate the transport of metabolites across organellar membranes. Hence, the systems analysis approach presented in this thesis also provides candidates for putative transporters. The next step will be the molecular and biochemical analysis of single candidates to fill the gaps and to complete the knowledge about the establishment and the processing of the C₄ photosynthetic cycle.

In this thesis the single gene analysis of a candidate originating from a reverse genetics screen to identify photorespiratory transporters was shown to be successful and PLGG1 was identified as the photorespiratory glycolate glycerate transporter. As photorespiration leads to a decrease in biomass production it is also an interesting field for plant research.

Therefore a different approach to increase biomass production in C₃ plants deals with photorespiration. A biotechnological approach was shown to successfully short-cut photorespiration by inducing an alternative pathway in A. thaliana to recycle 2-PG and increase biomass (Kebeish et al., 2007). In this approach the Escherichia coli glycolate catabolic pathway was introduced into the plastid in A. thaliana. As the chloroplastidic glycolate transporter is finally identified it needs to be investigated whether biomass production could be further improved by enhancing the flux through the biotechnologically introduced prokaryotic glycolate pathway by trapping the glycolate in the chloroplast as it occurs in plgg1-1 plants. In this context it is also important to further investigate the transport characteristics and get a detailed view of the structure of the transporter PLGG1. By inserting PLGG1 into the biotechnologically designed plants PLGG1 could be used as a regulated valve to control the amount of glycolate in the chloroplast. A very recent attempt followed an alternative approach by introducing an artificial complete glycolate catabolic cycle in chloroplasts of A. thaliana comprising glycolateoxidase (GO), malatesynthase (MS), and catalase (CAT) (Maier et al., 2012).

A surprising and interesting finding was that PLGG1 was found to be highly abundant in maize chloroplasts (Bräutigam et al., 2008). But why is a photorespiratory transporter highly abundant in C₄ photosynthesis that shows very reduced rates of photorespiration? This observation shows that there are still major gaps in our knowledge about the C₄ photosynthetic cycle. The objectives of further research will be to answer the question if PLGG1 took over another function in C₄ photosynthesis or if it is still the chloroplastidic glycolate glycerate transporter. If it is still the glycolate glycerate transporter the functional relevance of its abundance in C₄ needs to be investigated. The observations from these studies will bring further clarity in the C₄ cycle and will help to shed light on yet unknown parts of C₄ photosynthesis.
4. Theses

Despite their importance for economy and plant breeding research many important factors involved in the metabolic processes photorespiration and the highly efficient C_4 photosynthesis are still unknown. Among those factors are regulatory proteins orchestrating the onset of C_4 photosynthesis or transport proteins mediating the shuttling of specific metabolites across organellar membranes in the photorespiratory cycle.

1) The model C_4 plant *Zea mays* displays a developmental gradient along the leaf with young and immature cells at the base of the leaf and old and mature cells at the leaf tip. This developmental gradient is mirrored in the C_4 photosynthetic activity, which is highest at the mature leaf tip and very weak at the young base of the leaf. Against previous hypothesis there is no phase of C_3 photosynthesis at the base of the leaf, where the rate of C_4 photosynthesis is low. Along the gradient specific patterns of transcript and metabolite accumulation appear. These specific patterns can be used to refine the current C_4 model as it occurs in *Z. mays* and to generate candidates for yet unknown regulatory and transport proteins (Manuscript 1: Pick et al., 2011).

2) The photorespiratory cycle is a highly compartmentalized process with enzymatic reactions taking place in the chloroplasts, the peroxisomes, the mitochondria and the cytosol. To ensure the metabolic connection of the different compartments highly selective transporters account for the shuttling of specific metabolites. Despite their importance only two transporters involved in the photorespiratory nitrogen cycle could be identified to date. A reverse genetics approach followed by single gene analysis can be used to identify candidates for yet unknown transport proteins. *Arabidopsis thaliana* T-DNA-insertional k.o. mutants of candidate genes can be used to analyze the role in photorespiration. Metabolite analysis in combination with molecular and biochemical characterization are used to define transport substrates and transport characteristics. With this approach Plastidic glycolate glycerate transporter 1 (PLGG1) could be identified as the chloroplastidic glycolate glycerate transporter (Manuscript 2: Pick et al., 2013).
5. References


6. Manuscripts


2. Thea R. Pick, Andrea Bräutigam, Matthias A. Schulz, Toshihiro Obata, Alisdair R. Fernie, and Andreas P.M. Weber (2013). PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a new class of metabolite transporters. Submitted for publication to PNAS.
Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C₄ Model and Provides Candidates for Regulation
We systematically analyzed a developmental gradient of the third maize (Zea mays) leaf from the point of emergence into the light to the tip in 10 continuous leaf slices to study organ development and physiological and biochemical functions. Transcriptome analysis, oxygen sensitivity of photosynthesis, and photosynthetic rate measurements showed that the maize leaf undergoes a sink-to-source transition without an intermediate phase of C3 photosynthesis or operation of a photorespiratory carbon pump. Metabolome and transcriptome analysis, chlorophyll and protein measurements, as well as dry weight determination, showed continuous gradients for all analyzed items. The absence of binary on–off switches and regulons pointed to a morphogradient along the leaf as the determining factor of developmental stage. Analysis of transcription factors for differential expression along the leaf gradient defined a list of putative regulators orchestrating the sink-to-source transition and establishment of C4 photosynthesis. Finally, transcriptome and metabolome analysis, as well as enzyme activity measurements, and absolute quantification of selected metabolites revised the current model of maize C4 photosynthesis. All data sets are included within the publication to serve as a resource for maize leaf systems biology.

INTRODUCTION

The mechanisms underlying organ development and function are fundamental questions of biology. In plants, grass leaves represent an excellent model in which the establishment of various functions can be followed in a base-to-tip developmental gradient in a single leaf. Cells at the tip of the leaf are the oldest and most mature cells, while cells at the base are the youngest (Nelson and Langdale, 1992). We chose maize (Zea mays) to follow the establishment of photosynthetic functions during leaf development. Maize employs the highly efficient C4 type of photosynthesis, which concurrently evolved in multiple seed plant families 15 to 40 million years ago, long after C3 photosynthesis had been established (Edwards and Smith, 2010). It has been previously proposed that the evolutionary progression from C3 to C4 can also be detected in maize leaves along a spatial gradient (Nelson and Langdale, 1992, and references therein), very much like Haekel suggested that ontogeny recapitulates phylogeny during embryo development in animals (Haekel, 1866).

C4 photosynthesis has been considered a possible route for spawning a second green revolution in C3 crop plants, such as rice (Oryza sativa) (Hibberd et al., 2008). Plants using C4 photosynthesis are capable of producing biomass at faster rates than C3 plants, or, alternatively, these plants can inhabit harsher habitats with limited resources (Sage, 2004, and references therein). The key limitation for more productive photosynthesis is the concentration of carbon dioxide at the site of its assimilation, the reductive pentose phosphate pathway (rPPP) in plant chloroplasts. The enrichment of carbon dioxide around ribulose-1,5-bisphosphate carboxylase/oxygenase minimizes the oxygenation of ribulose-1,5-bisphosphate, which leads to a reduced rate of photorespiration. From an engineering standpoint, C4 photosynthesis, similar to a supercharged combustion engine, enriches the limiting factor, carbon dioxide, via a biochemical cycle operating between the site of initial fixation and final assimilation. Although the C4 cycle as described below appears
deceptively simple, differences between C₃ and C₄ photosynthesis go beyond just the addition of the C₄ cycle on top of the rPPP and include photorespiration, protein translation, cellular and tissue architecture, electron transfer adaptations, cell–cell connections, and likely other still unknown adaptations (Bräutigam et al., 2011; Gowik et al., 2011). In C₄ plants, carbon dioxide is enriched by affixing it to an acceptor, transferring it to the site of final assimilation, liberating it, and returning and recycling the acceptor for a new round. This system is referred to as the C₄ cycle. To avoid a futile cycle, the site of initial fixation, the mesophyll, is spatially separated from the site of assimilation, the bundle sheath. Canonically, maize operates a linear C₄ cycle (Hatch, 1987; Furbank, 2011): In the compartment of initial fixation, the mesophyll, the carbon dioxide acceptor phosphoenolpyruvate (PEP) is formed in the chloroplast from pyruvate by pyruvate:phosphate dikinase (PPDK) and then exported to the cytosol. There, carbon dioxide in the form of bicarbonate is fixed by PEP carboxylase (PEPC), creating the dicarboxylic C₄ acid oxaloacetate (OAA) from PEP. OAA is subsequently transferred to the chloroplast and reduced to malate, which is then exported to the cytosol of mesophyll cells. Malate is transported by mass flow to the bundle sheath, the compartment of final assimilation, where it is imported into chloroplasts and decarboxylated by the NADP-dependent malic enzyme (NADP-ME), yielding carbon dioxide, pyruvate, and NADPH. Pyruvate is exported from the chloroplast and returned to the mesophyll for regeneration of the acceptor PEP (Hatch, 1987). Whereas this canonical model of NADP-ME C₄ photosynthesis is depicted in many textbooks, several reports question its simplicity; however, an alternative model has not yet been formulated. For example, bundle sheath strands can efficiently decarboxylate not only malate but also the amino acid Asp (Chapman and Hatch, 1981). Older maize leaves, at least, harbor a second decarboxylation enzyme, PEP carboxykinase (PEP-CK), which releases carbon dioxide from OAA, producing PEP (Wingler et al., 1999). Furthermore, approximately one-quarter of radioactively labeled carbon dioxide that was fed to maize leaves was found to be rapidly incorporated into Asp (Hatch, 1971). Such side routes to the canonical NADP-ME C₄ pathway would require alternative transfer metabolites between mesophyll and bundle sheath cells, such as Asp or Ala, and alternative decarboxylation pathways would alter the demands on the remaining enzymes and the intracellular (Bräutigam and Weber, 2011a) and intercellular (Sowinski et al., 2008) transport systems. Understanding both the intracellular transport system between chloroplasts and cytosol and the intercellular transport between the mesophyll and bundle sheath cells is still in its infancy (Bräutigam et al., 2008; Sowinski et al., 2008; Bräutigam and Weber, 2011a, 2011b; Weber and von Caemmerer, 2010; Weber and Linka, 2011). Finally, understanding the regulatory circuits controlling C₄ photosynthesis is an ongoing quest in plant biology. Although limited information is available, such as the light dependence of C₄ enzyme expression (Chollet et al., 1996), the transcription factors mediating the abundant, cell-specific expression patterns remain unknown.

Recent work demonstrates that the maize leaf displays a gradient with regard to proteins (Majeran et al., 2010) and that large-scale transcriptional changes between four leaf areas can be detected (Li et al., 2010). In this work, we set out to generate a comprehensive systems level picture of the changes in metabolite, enzyme activity, and transcript amounts occurring along a developmental gradient of a growing maize leaf. Using this systems biology data set, we addressed the questions of (1) how photosynthesis is organized along the developmental gradient of the light-exposed leaf with special regard to the presence of C₃ photosynthesis, (2) whether the biochemistry of the C₄ cycle changes along this developmental gradient, and (3) which regulatory modules define the developmental progression in the gradient.

### RESULTS AND DISCUSSION

#### Organization of the Light-Exposed Third Maize Leaf

The transcript and metabolite amounts, as well as protein and chlorophyll contents, displayed characteristic and continuous changes along a tip-to-base gradient of the light-exposed part of the third leaf of maize (Figures 1A and 1B). The relative expression or metabolite contents were most distinct at the distal parts of the leaf compared with relatively minor changes in the center of the leaf. A principal components analysis of transcript and metabolite amounts along the leaf gradient demonstrated clear separation of the leaf slices. The principal components determining this pattern were the distance from the leaf base (component 1) and the distance from the leaf center (component 2). The complete data set is available in readable form as Supplemental Data Set 1 online.

If the leaf was divided from top to bottom into slices, with slice 1 being the tip, gene expression patterns reflecting biochemical pathways could be followed through the development of the leaf (Figure 2A). Since previous work has demonstrated good correlation between transcript and protein abundance in maize (Li et al., 2010), we took transcript amounts as proxies for the corresponding protein amounts. Steady state amounts of transcripts encoding the classical NADP-ME C₄ proteins PEPC, PPDK, and NADP-ME were low toward the leaf base and increased until they reached a maximum around slice 2 or 3 for PEPC and slice 10 for the decarboxylating enzymes (Figure 2A). Transcripts representing subunits of photosystems I and II and of the rPPP had a similar pattern, but their increase was much less pronounced than that of the C₄ transcripts. The pattern of the photosynthetic transcripts mirrored that of the photosystems and of the rPPP (Figure 2A). No peak of photosynthetic transcripts was observed where expression of the C₄ transcripts was low. Photosynthesis, measured as carbon fixation per leaf area, steadily increased between the bottom and the top of the leaf (Figure 2B). Finally, the oxygen sensitivity of photosynthesis was measured to determine whether C₃ photosynthesis or inefficient C₄ photosynthesis would occur in the light-exposed leaf, which should be reflected by a major increase in the apparent photosynthetic rate at low oxygen partial pressure. However, the ratio of photosynthetic rates measured at high and low oxygen partial pressures did not change along the leaf gradient (Figure 2C). Maize leaves were previously hypothesized to undergo a C₃-to-C₄ transition. That is, the program initiating C₄ photosynthesis was proposed being switched on in a particular region of the leaf.
This switch, if existent, was an important target for understanding C4 genesis and thus replicating it in making C4 rice. Our systems-level analysis does not support this hypothesis: Photorespiratory transcripts do not peak in the presumed area of C3-ness (Figure 2A) and there is no evidence for C3 photosynthesis or leaky C4 photosynthesis, as oxygen sensitivity of photosynthesis did not change along the leaf gradient (Figure 2C). We thus conclude that the maize leaf undergoes a gradual sink-to-source transition without a distinct intermediary C3 phase.

Metabolite Clusters

Although the leaf did not contain a zone of C3 photosynthesis, it clearly displayed a gradient along its length (Figure 1; Majeran et al., 2010). To investigate the nature of the gradient in detail, extractable metabolites were analyzed by clustering algorithms. For K-means clustering, a figure of merit analysis determined five clusters as the best compromise between cluster formation with limited information loss (Friedman and Stuetzle, 1981) (see Supplemental Figure 1 online). Metabolites in the pyruvate cluster with 23 members, cluster 1, were low at the bottom of the leaf and increased until slice 3 where the increase leveled off (Figure 3A). Cluster 2, with 15 members, contained metabolites that were high at the very bottom and at the tip, while cluster 3, with 10 members, contained metabolites that were level until the middle of the leaf and then increased toward the tip. The building block cluster, cluster 4, was the largest cluster with 60 members. The metabolites in this cluster started high at the bottom and decreased toward the middle of leaf from where the level stabilized. Cluster 5 was the malate cluster whose metabolites had the highest level between slice 3 and slice 7 and lower levels at the tip and the bottom (Figure 3A; condensed list of metabolites in Table 1).

Apart from pyruvate, cluster 1 contained Ala, glycerate, Glu, and citrulline. Five carotenoids, α-tocopherol, glycerol, and Gal of the lipid fraction as well as digalactosylglycerol and 3-O-galactosylglycerolipids, four fatty acids, and five other metabolites were also members of cluster 1 (Table 1). Surprisingly, the C4 acids formed a distinct cluster, the malate cluster 5, with Asp, fumarate, citrate, glyoxylate, γ-tocopherol, 3-O-galactosylglycerol, and three other metabolites. Since the C3 and C4 acids clearly separated into

![Figure 1.](image-url)
distinct clusters, there must be a major shift in the C4 cycle. One likely explanation was the observed gradient in PEPC activity. At the point where PEPC activity started to decrease in slice 3 (Figure 2A), the C4 acid pool sizes also sloped downwards. This may indicate that the balance between C3 and C4 acids pools shifted toward C3 acids because carboxylation activity decreased while the sum of decarboxylation activities increased relative to each other (Figure 2A). Some of the pool sizes of tricarboxylic acid (TCA) cycle acids connected to malate also shifted with alterations in PEPC activity. The malate pattern extended to fumarate, citrate, and isocitrate but not to succinate and 2-oxoglutarate. The pool sizes of TCA cycle intermediates were thus only partially isolated from the C4 cycle. The remaining metabolites shadow the buildup of the chloroplasts and thylakoids, including their pigments. The galactolipids, which dominate the chloroplast membranes (Dörmann and Benning, 2002), were mostly members of the pyruvate cluster, as were the accessory pigments carotenoids that can diffuse excess light energy via the xanthophyll cycle (Bilger and Bjorkman, 1990).

Cluster 2 contained Man, galactitol, diethylene glycol, salicylic acid, five fatty acids, and six other metabolites, while the raffinose cluster 3 contains raffinose, stachyose, galactinol, and seven other metabolites (Table 1). Metabolites from both clusters elevated toward the tip, although metabolites in cluster 2 also elevated at the very bottom (Figure 3A). The abundance of metabolites of the raffinose family from cluster 2 and cluster 3 (myo-inositol-2-P, raffinose, stachyose, and galactinol) pointed toward a drought response (Seki et al., 2007) at the tip, but, strikingly, cluster 2 and cluster 3 did not include amino acids such as Pro. The dry weight-to-fresh weight ratio increased toward the tip, with only half the water content at the tip (Figure 3B). However, the third maize leaf analyzed in this study did not show any apparent signs of cell death at the tip (see Supplemental Figures 2A and 2D online). We thus hypothesize that despite the low water content at the tip (Figure 3B), the accumulation of compatible solutes at the tip allows photosynthesis to operate efficiently (Figure 2B). On mature field-grown maize plants, the majority of leaf tips are completely dry with only dead cells remaining. We hypothesize that maize leaves undergo a constitutive innate drought response toward the tip of each leaf to continue photosynthesizing (Figure 2B) until water content gets too low to maintain metabolism and cells undergo cell death. Considering the parallel venation pattern of grasses, any drought stress will likely initially manifest in the leaf tip. In leaf tips of the third maize leaf, chlorophyll content was already reduced (Figure 3C); however, the tissue was likely not senescent since the protein content was high (Figure 3D), photosynthesis was highly efficient (Figure 2B), and senescence markers were not highly expressed (see Supplemental Data Set 1 online).

Cluster 4 was termed the building block cluster. It contained 15 proteinogenic amino acids but not Asp, Ala, and Glu, which were part of the pyruvate and malate clusters. In addition to the amino acids, four precursors (shikimate, quinate, homoserine, and S-adenosylhomoserine) were part of this cluster. The major sugars Glc and Fru as well as the minor sugars Rib and Fru had elevated amounts at the leaf base. Ten sphingolipids, four sterols, and six fatty acids were part of cluster 4. Finally, coumaric and ferulic acid, isopentenylpyrophosphate, glucosephosphates, free

Figure 2. Photosynthetic Transitions in the Maize Leaf.

(A) Average relative expression levels for the transcripts encoding photosystem I, photosystem II, the RPPP and three key C4 proteins.
(B) Photosynthetic rate along the light-exposed leaf. Error bars depict SD of three biological replicates. nd, not determined since not exposed to light.
(C) Ratio of photosynthesis at 20% and 2% O2 concentration along the leaf gradient. Error bars depict SD; nd, not determined since not exposed to light, two technical replicates.
phosphate, and 12 other metabolites finished the cluster (Table 1). The leaf base represents a sink tissue (Evert et al., 1996) with minimal photosynthetic activity (Figure 2B). Chloroplasts began to develop at the leaf base (Evert et al., 1996) and chlorophyll content increased (Figure 3C). Genes encoding components of the rPPP and the electron transfer chain were highly expressed. Consequently, proteinogenic amino acids were in high demand and thus present in large amounts. The major sugars likely reflected transferred carbon, while the minor sugars and lignin precursors pointed to active cell wall synthesis. Membrane buildup was in process. Transcript analysis of four distinct zones of the maize leaf found increased transcript amounts for cell wall, lipids, secondary metabolism, and chloroplast targeting for the area at the bottom of the gradient (Li et al., 2010), thereby corroborating the analysis of metabolites. Proteins involved in lipid synthesis also peak toward the base of the gradient (Majeran et al., 2010).

In summary, four clusters defined the leaf gradient, the building block cluster defined by elevated metabolites at the bottom end of the gradient, which was followed by the C4 malate cluster with increased C4 acids and TCA cycle intermediates, which in turn was followed by the C4 pyruvate cluster with high C3 acids, carotenoids, and galactolipids. The tip of the leaf contained elevated amounts of drought-related metabolites of the raffinose family, which are included in clusters 2 and 3 (Figure 3E).

**Modules of the Leaf: Transcripts**

The changes along the maize leaf were recently investigated using four leaf segments sampled from different parts of the leaf (Li et al., 2010). Li et al. identified 938 transcription factors that showed a differential expression pattern between at least two of the segments. In our study, we followed three different strategies to assess the dynamics and extent of the reprogramming of the transcriptome along the maize leaf developmental gradient: (1) K-means clustering to identify patterns of expression along the leaf, (2) hierarchical clustering to identify transcripts with similar patterns as the C4 transcripts, and (3) a comparison of this leaf gradient with previously published data from four distinct leaf segments (Li et al., 2010).

The K-means clustering was prefaced by a figure of merit analysis, which prompted us to choose six clusters as a good solution (see Supplemental Figure 3 online). Four distinct patterns were evident in the clusters: Clusters 1 and 3 contained transcripts that are either very low (1450) or low (8521) toward the

---

**Figure 3. Changes along the Leaf Gradient.**

(A) K-means clusters of metabolites. Cluster 1 is the pyruvate cluster with 23 members; cluster 2 contains 15 metabolites; cluster 3 is the raffinose cluster with 10 members; cluster 4 is the building block with 60 members; cluster 5 is the malate cluster with 10 members.

(B) The fresh weight (FW)–to–DW ratio indicating a low water content at the leaf tip.

(C) The chlorophyll content.

(D) and (E) The protein content (D) and K-means clustering of transcripts (E). Error bars indicate SD of four biological replicates.
light is necessary for transcription of C4 genes (Langdale et al., 1988) but not a sufficient cue to alter gene expression in the leaf (Figures 1A and 3E). Thus, light was a factor determining the metabolite accumulation pattern was set up by transcripts that reflected positional information, only 68 transcription factors increasing in expression toward the leaf tip. The plurality, 52%, did increase in expression but dipped slightly at the tip similar to chlorophyll content and PEPC activity (Figure 4B). This change in pattern was visible only with a continuous gradient and cannot be detected with segmental analysis. Only 10% showed patterns that were not congruous with earlier data. If the morphogradient was set up by transcripts that reflected positional information, only 68 transcription factors increasing in expression would be on the short list of candidate transcription factors at or near the core of the morphogradient. By contrast, 276 transcription factors decrease more or less continuously. Adding this second analysis reduced the list of potential transcription factors from 938 in the earlier study down to 344 in our work. Additional analyses have the potential to reduce the list to where single-gene functional analyses become feasible. Three important pieces of information are missing: (1) Which of these factors, if any, display a similar gradient in older and bigger leaves, (2) does this gradual behavior extend throughout the leaf to the point of emergence from the apical meristem, and (3) which factors have a similar gradient in other grass species? In older maize leaves of 40-cm length, enzyme activity measurements clearly show gradients for the C4 marker enzymes (see Supplemental Figure 4 online), which are similar to those in the

Table 1. Condensed List of Metabolites within Each Cluster

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyruvate</td>
<td>Man</td>
<td>Raffinose</td>
<td>Comaric acid</td>
<td>Asp</td>
</tr>
<tr>
<td>Ala</td>
<td>Galactitol</td>
<td>Galactitol</td>
<td>Ferulic acid</td>
<td>Malate</td>
</tr>
<tr>
<td>Glycerate</td>
<td>Diethylene glycol</td>
<td>Stachyose</td>
<td>Phosphate</td>
<td>Fumarate</td>
</tr>
<tr>
<td>Glu</td>
<td>Salicylic acid</td>
<td>Tryptamine</td>
<td>Glucosephosphat</td>
<td>Citrate (additional: isocitrate)</td>
</tr>
<tr>
<td>Citrulline</td>
<td>Ribonic acid</td>
<td>Trp</td>
<td>Glyoxylate</td>
<td></td>
</tr>
<tr>
<td>Galactose, lipid fraction</td>
<td>Cys</td>
<td>α-Ketoglutarate</td>
<td>15 proteinogenic amino acids</td>
<td>γ-Tocopherol</td>
</tr>
<tr>
<td>Glycerol, lipid fraction</td>
<td>NAD</td>
<td>Nicotinamide</td>
<td>10 sphingolipids</td>
<td>Threonic acid</td>
</tr>
<tr>
<td>Digalactosylglycerol</td>
<td>Five fatty acids</td>
<td>Gluconic acid</td>
<td>Four sterols</td>
<td>3-O-galactosylglycerol</td>
</tr>
<tr>
<td>3-O-galactoylglycerolipids</td>
<td>Three other metabolites</td>
<td>Myristic acid</td>
<td>Six fatty acids</td>
<td>Two other metabolites</td>
</tr>
<tr>
<td>Five carotenoids</td>
<td></td>
<td>UDP-glucose</td>
<td>Four precursors</td>
<td></td>
</tr>
<tr>
<td>Four fatty acids</td>
<td></td>
<td></td>
<td>12 other metabolites</td>
<td></td>
</tr>
<tr>
<td>Five other metabolites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Metabolites were K-means clustered. A figure of merit analysis determined five clusters as the best compromise between cluster formation with limited information loss (Friedman and Stuetzle, 1981).

bottom of the gradient, clusters 4 and 6 contained transcripts that are either very high (1067) or high (4287) at the bottom, while the largest cluster 5 (10,935) displayed little change (Figure 3E). The rate of change in transcript abundance (Figures 1A and 3E) agrees with those published earlier for differences between noncontinuous leaf segments (Li et al., 2010) and is comparatively modest, especially for the average changes of each cluster (Figure 3E). Our continuous gradient revealed that virtually all transcripts with changes in the transition zone (Li et al., 2010) and is comparable to slice 10 in this study showed steady declines or increases throughout the remainder of the gradient rather than an on-off behavior. These patterns indicated that it is highly likely that the transcriptional changes and functional changes were set up by a morphogradient along the leaf, which may be defined as the point where single-gene functional analyses become feasible. Three important pieces of information are missing: (1) Which of these factors, if any, display a similar gradient in older and bigger leaves, (2) does this gradual behavior extend throughout the leaf to the point of emergence from the apical meristem, and (3) which factors have a similar gradient in other grass species? In older maize leaves of 40-cm length, enzyme activity measurements clearly show gradients for the C4 marker enzymes (see Supplemental Figure 4 online), which are similar to those in the
young leaves. This at least indicated that older, more mature leaves still display a gradient.

Systematic analyses of dicot C4 species showed that all C4 enzyme activities except for malate dehydrogenase are, at least to some degree, regulated at the transcriptional level (Bräutigam et al., 2011; Gowik et al., 2011). Even if the C4-related transcripts are piggybacking on the developmental gradient, the direct regulators of their transcription would be expected to be coexpressed with or just predating their targets. We identified transcripts encoding putative regulators that were tightly coexpressed with the major C4 transcripts. The transcript for the major isoform of PEPC accumulated slowly throughout the gradient, reached a plateau between slices 3 and 5, and dipped toward the tip (Figure 4C). In the hierarchical clustering (see Supplemental Data Set 2 online, readable with MeV, www.tm4.org/mev), 16 transcripts representing regulatory functions were identified (see Supplemental Table 1 online). Two of these transcripts related to calcium signaling, one EF hand protein, and a calreticulin. An SnRK1 subunit implicated in sugar and nitrogen signaling (Rolland et al., 2006) had the same pattern as PEPC. In addition, a mitogen-activated protein kinase phosphatase and PP2C, which is involved in abscisic acid (ABA) and drought signaling in Arabidopsis thaliana (Kuhn et al., 2006), were coregulated with PEPC. Finally, 13 transcription factors, one orphan, four APETALA2 (AP2)-ETHYLENE RESPONSE FACTORS (ERFs), two G2-like myb transcription factors, one Auxin Response Factor (ARF), and one CCAAT-type transcription factor tightly correlate with PEPC throughout the gradient. The Arabidopsis homologs of these transcription factors are involved in ABA signal transduction and ethylene signal transduction (see Supplemental Table 1 online). Possibly, an ABA and/or ethylene-driven regulon was used in evolution of C4 photosynthesis. Neither DOF1 nor DOF2, which are known to bind the PEPC promoter region (Yanagisawa and Sheen, 1998), are tightly coexpressed with PEPC. For maize nuclear factor and PEP-I, no sequences were deposited at the National Center for Biotechnology Information (NCBI); hence, they could not be compared with the current data. Notably, none of the tightly correlated transcripts are known to be involved in light signaling, underscoring that light

Figure 4. Targeted Expression Analysis of Regulatory Functions.

(A) and (B) Expression pattern of transcripts detected as low at the tip (A) and high at the tip (B) in a previous analysis. Patterns in orange confirm the expectation based on Li et al. (2010), patterns in blue partially confirm, and patterns in gray have different patterns.

(C) Transcripts coexpressed with PEPC (blue), NADP-ME (green), and PPDK (yellow) major isoforms.
Manuscript 1

System-Wide Analysis of a Maize Leaf

28

is necessary but not sufficient to drive expression (see above). The two major NADP-ME isoforms showed the same pattern as PEPC up to slice 3 but lacked the dip at the tip. Only nine transcripts encoding regulatory functions tightly correlate (see Supplemental Table 1 online): a dicer homolog, one orphan transcription factor, two AP2-ERFs, a zinc-finger transcription factor, a PLATZ transcription factor, and two coactivators. The PPDK transcript behaved quite differently; it accumulated from very low levels toward slice 5 and then mirrored NADP-ME. Comparatively few regulatory transcripts mirror this more extreme pattern (see Supplemental Table 1 online): a kinase, one phosphatase, a phosphorelay transmitter similar to AHP4 of Arabidopsis, a LOB-type transcription factor, and a regulatory protein similar to a flowering regulator from Arabidopsis. Selected transcript abundance patterns were confirmed by quantitative RT-PCR (see Supplemental Figure 5 online). The differences in pattern between the three key C4 transcripts pointed to the fact that a simple generic C4 regulon may not exist. Rather, additional data sets taken during leaf development of maize and other grasses will increase the resolution of co-variation analyses and lead to the identification of the leaf morphoradiants and ultimately of regulons that induce expression of the separate C4 genes.

C4 Photosynthesis along the Developmental Gradient of the Leaf

It was recently proposed that C4 plants undergo changes in their mode of C4 photosynthesis based on developmental stage and in response to environmental cues (Furbank, 2011). The well-defined maize leaf developmental gradient analyzed in our study represented a unique opportunity to test this hypothesis.

The classical C4 genes PEPC, PPDK, and NADP-ME were identified from the literature, and their abundance and expression pattern was used to identify transcripts with similar abundance and pattern. The C4 genes were among the transcripts that occupy more than one per thousand of the total (see Supplemental Data Set 3 online). Aside from PEPC, PPDK, and NADP-ME, the list of abundant transcripts contained mainly transcripts that encode the chloroplast electron transfer chain and the rPPP (see Supplemental Data Set 3 online). Surprisingly, genes for a plastid-localized Asp aminotransferase (AspAT) and PEP-CR were also members of this group of 147 transcripts. Coexpressed transcripts frequently act in the same or in connected pathways (Eisen et al., 1998; Reumann and Weber, 2006). Hence, all transcripts were clustered to identify transcripts that are coexpressed with known C4 transcripts. The known C4 transcripts are low toward the bottom of the leaf and increase toward the tip. The very tip portion is slightly lower in expression compared with middle of the leaf blade (Figure 2A). If transcripts, which might be active in any of the C4 types, were plotted, one Ala aminotransferase (AlaAT) and a plastidic AspAT as well as a PEP-CR would display a comparable pattern (see Supplemental Figure 6 online). Taken together with the observation that in maize, 25% of the carbon label initially was located in Asp (Hatch, 1971) and the observation that Asp is a carbon donor to the bundle sheath (Chapman and Hatch, 1981), we decided to investigate the seemingly simple C4 cycle of maize at the levels of transcripts, metabolites, and enzyme activity.

We initiated the analysis by testing whether the leaf had reached C4 configuration at the point where the analysis commenced. In slice 10, the bottom of the gradient, the leaf was already differentiated into a vein, bundle sheath, mesophyll, bundle sheath, vein configuration (see Supplemental Figure 2 online). PEPC activity increased from the bottom, reached a maximum at slice 3, and decreased only slightly toward the tip of the leaf. Maximal activity was 18 milli units/mg dry weight (DW). The major decarboxylation enzyme NADP-ME increased from the bottom toward the top and reached its maximal activity at the leaf tip with close to 15 mU/mg DW. Both AspAT and AlaAT had similar patterns compared with NADP-ME and reached activities of 25 mU/mg and 15 mU/mg DW. Although PEP-CR activity was only a quarter of NADP-ME, it had a comparable pattern and reached up to 4 mU/mg DW (Figure 5A). The pattern for all enzymes except PEPC was similar; that is, the activity was low at the bottom of the leaf and increased toward the tip.

On the basis of transcript abundance, the major enzyme isoform of PEPC mirrored the pattern of PEPC activity in the leaf and also peaked around slice 3 (Figure 5B). Transcripts of the major isoforms of NADP-ME, PEP-CR, AspAT, AlaAT, and PPDK, displayed a pattern comparable to that of PEPC but different than the extractable activities of the enzymes. Hence, the total activity was likely composed of multiple isoforms of NADP-ME, PEP-CR, AspAT, and AlaAT and/or subject to posttranscriptional regulation. Indeed, there were other isoforms that were of appreciable transcript abundance (see Supplemental Data Set 3 online) and patterns unlike that displayed by the major isoform (Figure 5B). The enzyme activity of AspAT was sufficient to support the carboxylation and decarboxylation activity, while AlaAT fell short for the majority of the leaf. PEP-CR activity was appreciable. While certainly not the major decarboxylation activity, its activity was high enough to catalyze at least one-fifth of the decarboxylation reactions. This was almost certainly an underestimation since PEP-CR was assayed in the unfavorable reverse reaction (Ashton et al., 1990). If amino acids carried part of the carbon flow in the C4 cycle, their abundance should mirror that of the canonical C4 cycle acids malate and pyruvate. Ala and Asp mirrored the accumulation pattern of pyruvate and malate, respectively (Figure 5C). In addition, not only their pattern but also their absolute abundance should be comparable to that of malate. The absolute abundance of Asp and Ala were about one-fourth of the abundance of malate (see Supplemental Figure 7 online).

Taken together, these results suggested a revised model of the C4 cycle in maize (Figure 5D): After PEP is carboxylated to OAA, it is moved to the chloroplast, either in exchange with malate through DiT1 (Kinoshita et al., 2011) or in exchange with Asp through DiT2 (Renne et al., 2003), which are produced by malate dehydrogenase and AspAT, respectively, in the chloroplast. The major AspAT in maize is predicted to be chloroplast localized. Labeling experiments by Hatch (1971) indicated that as much as 25% of the carbon initially labels Asp, not malate. Both C4 acids diffuse to the bundle sheath, reducing the necessary mass flow compared with either C4 acid carrying the full load. In the bundle
sheath, malate is taken up by a currently unknown mechanism into the chloroplast and decarboxylated. Asp may have two fates: It may be transaminated to OAA and decarboxylated by PEP-CK to PEP, or it may enter the chloroplast by an as yet unknown mechanism and be transformed via OAA to malate to serve as the substrate for NADP-ME. It has long been assumed that malate transfer is preferable to Asp transfer since malate carries a reducing equivalent while Asp does not. However, it has been shown that triosephosphate translocator is one of the most abundant chloroplast envelope proteins (Bräutigam et al., 2008) and that the reduction of 3-phosphoglycerate is almost entirely located in the mesophyll (Majeran et al., 2005), making the generation of reducing equivalents unnecessary in the bundle sheath. The regeneration of the carbon acceptor and its transfer to the mesophyll may be dissected by analysis of metabolite compartmentation, flux, and gradients and may occur as PEP, pyruvate, or Ala (Figure 5D).

We propose that independent of environmental or developmental cues, the core \(C_4\) cycle in maize is set up already as a branched rather than a linear cycle. In addition to the scheme presented (Figure 5D), the branched core \(C_4\) cycle is also connected to basal metabolism (for example, see Leegood and von Caemmerer, 1988). A distribution of carbon between two \(C_4\) acids and three \(C_3\) acids reduces the diffusion requirements for
any one molecule between mesophyll and bundle sheath. This distribution becomes especially important considering that distribution by diffusion is by no means proven (Sowiński et al., 2008; Bräutigam and Weber, 2011a). It remains to be investigated whether the distribution of carbon to malate and Asp is fixed at 3:1 as reported (Hatch, 1971) or whether this ratio is adjusted by the plant during its life cycle (Furbank, 2011). Within the age gradient in a single leaf, there is no evidence in the enzyme activities, transcript abundance, or metabolite accumulation pattern to suggest that operation of the cycle switches from one transfer acid to another (Figure 5). The presence of higher PEP-CK activity in older maize plants with older leaves (Wingler et al., 1999), however, points to a developmental regulation between leaves rather than within a leaf, similar to what has been recently observed in the dicotyledonous C₄ plant Cleome gynandra (Sommel et al., 2012). Environmental adaptation of Asp metabolism based on N availability in maize leaves with regard to pool size and turnover has also been demonstrated (Khamis et al., 1992). Hence, the C₄ cycle is apparently quite flexible.

Conclusion

On the basis of a comprehensive systems biology data set, we conclude that C₄ photosynthesis is established from sink tissue without an intermediate phase of C₂ or C₃ photosynthesis. That is, the likely evolutionary events are not recapitulated during ontogeny. No binary on–off switches were detected within the leaf gradient, pointing to gradual onset of features and, therefore, morphogradients as the determinants for leaf development. Finally, the biochemistry of C₄ photosynthesis is more complex than anticipated but stays constant throughout the leaf.

METHODS

Plant Growth and Harvest

Maize (Zea mays) plants of the ecotype B73 were grown in the greenhouse for 14 to 15 d in clay pots in Floraton soil. Natural light, a shading system, and artificial light were used to extend the daylight period to 16 h at a photon flux density of ~500 μmol m⁻² s⁻¹. The humidity in the greenhouse was between 75 and 90%. The greenhouse’s ventilation system kept the temperature at 24°C.

The third leaf was harvested at 18-cm length measured from tip to emergence from the stem. Leaves were harvested by placing them atop a custom-made leaf guillotine where they were snap frozen (see Supplemental Figure 8 online). By closing the lid, the leaf is cut into 10 pieces of 2-cm width each, the last of which had not yet emerged (see Supplemental Figure 2 online). Twenty plants were pooled for each biological replicate. The sections were ground to a fine powder in a porcelain mortar cooled with liquid nitrogen. Frozen powder was used for each biological replicate. The mRNA was isolated after the method of Logemann et al. (1987) and from the same plant material in which the enzyme activities and metabolites were measured. The isolated RNA was purified with the RNeasy purification kit according to the manufacturer’s instructions (Qiagen). The quality was checked with the Agilent 2100 Bioanalyzer using the RNA 6000 Nano kit. The cDNA and following antisense cRNA synthesis was performed according the one-color microarray-based gene expression analysis protocol (Agilent Technologies). An aliquot of 1.65 μg of this RNA was loaded on one-color microarrays with custom-designed oligonucleotide probes (Agilent 025271). Transcripts were normalized to the 75th percentile within each array using the Agilent Gene spring program. Arrays can be accessed under submission number GSE33861 in the NCBI Gene Expression Omnibus database. Quantitative RT-PCR was performed with three biological replicates using the SYBR-green technique (MESA GREEN qPCR MasterMix Plus; Eurogentec) and gene-specific primers (see Supplemental Table 2 online) as described by Schmittgen and Livak (2008). Relative expression values were calculated with the 2⁻ΔΔCt (cycle threshold) method after Pfaffl (2001) with threshold values normalized to expression of 18S rRNA.

Data Analysis

For data analysis, the maize transcript list was downloaded from www.maizesequence.org. For each transcript, a best BLAST hit was produced with Sorghum bicolor and Arabidopsis thaliana as databases (Altschul et al., 1997). Gene Ontology terms were added based on the S. bicolor annotation. Information about putative and known transcription factors (Pérez-Rodríguez et al., 2010) and transport proteins (http://membranetransport.org/) were added based on the Arabidopsis annotation. A Mapman annotation was downloaded from http://mapman.gabipd.org; Thimm et al., 2004). Protein localization was predicted based on amino acid sequence (Emanuelsson et al., 2000). For each maize transcript, an annotation was created based on the Arabidopsis TAIR10 description (Swarbreck et al., 2008) and, if not available, manually added based on the Sorghum data. Transcripts without known or predicted functions were labeled POUF (for protein of unknown function). Based on all information, transcripts were grouped into classes in a

Metabolite Profiling

Lyophilized tissue equivalent to 200 mg of fresh weight was used for metabolite profiling. Metabolites were extracted with the use of accelerated solvent extraction with polar (methanol + water, 80 + 20 by volume) and nonpolar (methanol + dichloromethane, 40 + 60 by volume) solvents. Subsequent analyses of metabolites by gas chromatography–mass spectrometry (GC-MS) were performed as described elsewhere (Roessner et al., 2000; Walk et al., 2007). In addition, liquid chromatography–tandem mass spectrometry (Niessen, 2003) analyses were performed with the use of an Agilent 1100 capillary LC system (Agilent Technologies) coupled with an Applied Biosystems/MDS SCIEX API-4000 triple quadrupole mass spectrometer (AB SCIEX). After reverse-phase HPLC separation, detection and quantification of metabolites were performed in the multiple reaction monitoring and full scan mode (Gergov et al., 2003). Absolute Ala, Asp, and malate contents were estimated by GC-MS (Fiehn et al., 2000), which included an external complex standard and were quantified by coupled enzymatic assays (Bergmeyer, 1974).

Transcript Profiling

The mRNA was isolated after the method of Logemann et al. (1987) and from the same plant material in which the enzyme activities and metabolites were measured. The isolated RNA was purified with the RNeasy purification kit according to the manufacturer’s instructions (Qiagen). The quality was checked with the Agilent 2100 Bioanalyzer using the RNA 6000 Nano kit. The cDNA and following antisense cRNA synthesis was performed according the one-color microarray-based gene expression analysis protocol (Agilent Technologies). An aliquot of 1.65 μg of this RNA was loaded on one-color microarrays with custom-designed oligonucleotide probes (Agilent 025271). Transcripts were normalized to the 75th percentile within each array using the Agilent Gene spring program. Arrays can be accessed under submission number GSE33861 in the NCBI Gene Expression Omnibus database. Quantitative RT-PCR was performed with three biological replicates using the SYBR-green technique (MESA GREEN qPCR MasterMix Plus; Eurogentec) and gene-specific primers (see Supplemental Table 2 online) as described by Schmittgen and Livak (2008). Relative expression values were calculated with the 2⁻ΔΔCt (cycle threshold) method after Pfaffl (2001) with threshold values normalized to expression of 18S rRNA.
Manuscript 1

hierarchical manner. Arabidopsis information was given precedence over other information given that Arabidopsis annotations are currently the best within the plant genomes. Group and functional assignments throughout the publication are based on this annotation table. The complete annotation table, including all raw data, can be accessed as Supplemental Data Set 1 online. The major isoform of C4 enzymes were determined by read mapping of raw data from Li et al. (2010) on the maize transcriptome since analyzed data were not included in the original publication. Read mappings were normalized to reads per million without any further correction factors applied. The data are included in Supplemental Data Set 1 online.

All large-scale data analyses were performed with the MultiExperiment Viewer (http://www.tm4.org/mev/; Saeed et al., 2003). Average metabolite contents were expressed as z-scores (the number of standard deviations the value is different from the mean of all values), resulting in mean centered values. Only metabolites detectable in all biological replicates of eight or more slices were analyzed. Transcripts were normalized to the 75th percentile within each array; the mean of replicates was calculated for each slice, followed by mean centering along each row. For K-means cluster analysis, the ideal number of clusters was determined by figure of merit analysis as implemented in MeV (Saeed et al., 2003). Metabolites and transcripts were clustered by Euclidian average linkage clustering and visualized in MeV. For comparison with the Li et al. (2010) data set, transcription factors of different groups were extracted from Li et al. (2010) supplemental data and visualized in MeV (see Supplemental Data Sets 4 and 5 online). Transcripts coexpressed with major C4 enzymes were determined by hierarchical clustering followed by list extraction from MeV. All raw data, including the MeV readable files, are provided as supplemental material accompanying the publication (see Supplemental Data Sets 1 and 2 online).

Accession Numbers

Microarray data from this article can be found in the NCBI Gene Expression Omnibus database under accession number GSE33861.

Supplemental Data

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

- Supplemental Figure 1. Figure of Merit Analysis of Metabolite Clustering.
- Supplemental Figure 2. Configuration of the Leaf Anatomy along the Stem.
- Supplemental Figure 3. Figure of Merit Analysis of Transcript Clustering.
- Supplemental Figure 4. Enzyme Activity for Three C4 Marker Enzymes in Maize Leaves of 40-cm Length.
- Supplemental Figure 5. qRT-PCR Results for Selected Regulatory Transcripts.
- Supplemental Figure 6. Expression Pattern for Enzymes Likely Involved in C4 Photosynthesis.
- Supplemental Figure 7. Absolute Concentrations of Malate, Aspartate, and Alanine along the Leaf Gradient.
- Supplemental Figure 8. The Guillotine Used for Sampling the Gradient.
- Supplemental Table 1. List of Maize Identifiers of Regulatory Transcripts Coregulated with PEPC, PPDK, or NADP-ME.
- Supplemental Table 2. Primers Used for qRT-PCR.
- Supplemental Data Set 1. The Complete Data Set in Human-Readable Form.

Supplemental Data Set 2. MeV Readable Hierarchical Clustering of All Data to Identify Transcripts Coregulated with Major C4 Enzymes.

Supplemental Data Set 3. Maize Transcripts with the Highest Number of Read Mappings Based on SRR039507 and SRR039508 Originally Published by Li et al. (2010).

Supplemental Data Set 4. MeV Readable Microarray Data for G1 and G2 Upregulated Transcription Factors Published as a Supplemental File by Li et al. (2010).

Supplemental Data Set 5. MeV Readable Microarray Data for G3 Upregulated Transcription Factors Published as a Supplemental File by Li et al. (2010).

ACKNOWLEDGMENTS

This study was supported by a grant of the German Federal Ministry of Education and Research (BioEnergy 2021, OPTIMAS) and grants from the German Research Foundation (IRTG 1525 and FOR 1186 PROMICS to A.P.M.W.). We thank Katrin L. Weber for support with GC-MS analyses.

AUTHOR CONTRIBUTIONS

T.R.P. sampled the gradient, measured all parameters except transcripts and relative metabolite content, analyzed the data, and cowrote the article. A.B. designed the research, produced the custom annotation, analyzed data, and wrote the article. U. Schlüter measured transcripts. A.K.D. measured absolute metabolite contents and photosynthesis rates. C.C. and U. Scholz provided tables for the custom annotation. H.F. measured relative metabolite contents. R.P. and U.R. supported the oxygen sensitivity measurements. U. Sonnewald designed the research and analyzed data. A.P.M.W. designed the research, analyzed data, and cowrote the article.

Received September 2, 2011; revised November 23, 2011; accepted December 1, 2011; published December 20, 2011.

REFERENCES


Supplemental Figure 1. Figure of merit analysis of metabolite clustering. The algorithm tries 20 times to cluster the data and calculates the information loss based on the cluster pattern compared with the patterns of the items in the cluster. Ideally, at some point, additional clusters no longer lead to more information in the clusters. In practice, the information gain for an additional cluster will reach a constant level. This point was chosen for k-means clustering.
Supplemental Figure 2. Configuration of the leaf anatomy along the stem. (A) position of cut sites marked by arrows (B) cross section of the maize stem; red arrows mark veins; in contrast to the first and second leaf, the third leaf is already in C₄ configuration (C) UV illuminated cross section of the third leaf; two bundles are surrounded in red, the distance between them allows not more than two intervening mesophyll cells; (D) position of the leaf slices along the third maize leaf.
Supplemental Figure 3. Figure of merit analysis of transcript clustering. The algorithm tries 20 times to cluster the data and calculates the information loss based on the cluster pattern compared with the patterns of the items in the cluster. Ideally, at some point, additional clusters no longer lead to more information in the clusters. In practice, the information gain for an additional cluster will reach a constant level. This point was chosen for k-means clustering.
Supplemental Figure 4. Enzyme activity for three C₄ marker enzymes in maize leaves of 40 cm length. All C₄ marker enzymes are more active in the tip region of the leaf compared with the point of emergence into the light. Error bars depict standard deviation; n=3.
Supplemental Figure 5. qRT-PCR results for selected regulatory transcripts; 1 = leaf slice at the tip; 10 = leaf slice at the leaf base; expression values were normalized to expression levels of 18srRNA. Error bars indicate standard error of three technical replicates of three individual experiments.
Supplemental Figure 6. Expression pattern for enzymes likely involved in C₄ photosynthesis. Candidate C₄ transcripts follow a ‘PEPC’ pattern, low at the base, peaking in the middle of the leaf, equal of dipping at the very tip. Abbreviations: AlaAT alanine aminotransferase, AMK adenosin monophosphate kinase, AspAT aspartate aminotransferase; Dit1 dicarboxylate transporter 1, NAD-ME NAD dependent malic enzyme, NADP-ME NADP dependent malic enzyme, PEPC phosphoenolpyruvate carboxylase, PEP-CK phosphoenolpyruvate carboxykinase, PP pyrophosphorylase, PPDK pyruvate phosphate dikinase, PPDK-RP regulatory protein of PPDK, PPT phosphoenolpyruvate phosphate translocator.
Supplemental Figure 7. Absolute concentrations of malate, aspartate and alanine along the leaf gradient (determined in a separate experiment). The absolute concentration of alanine and aspartate are in the same order of magnitude compared to that of malate. Error bars indicate standard error; n=3.
Supplemental Figure 8. The guillotine used for sampling the gradient. The guillotine is filled with liquid nitrogen. A leaf is placed on top of the open guillotine, snap frozen and cut by closing the lid. Each leaf slice falls into its own compartment and can be retrieved on its ledger.
### Supplemental Table 1. List of maize identifiers of regulatory transcripts co-regulated with PEPC, PPDK and NADP-ME, respectively.

#### Coregulated with PEPC

<table>
<thead>
<tr>
<th>Parent Gene</th>
<th>best hit AGI</th>
<th>gene name of corresponding AGI (where available)</th>
<th>transcription factor class</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRMZM2G467184</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRMZM2G305115</td>
<td>AT1G08450</td>
<td>CALRETICULIN 3 (CRT3)</td>
<td></td>
</tr>
<tr>
<td>GRMZM2G025459</td>
<td>AT5G21170</td>
<td>(AKINBETA1)</td>
<td></td>
</tr>
<tr>
<td>GRMZM2G465287</td>
<td>AT2G30020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRMZM2G177693</td>
<td>AT3G11410</td>
<td>ARABIDOPSIS THALIANA PROTEIN PHOSPHATASE 2CA (PP2CA)</td>
<td></td>
</tr>
<tr>
<td>GRMZM2G062218</td>
<td>AT5G41380</td>
<td></td>
<td>Orphans</td>
</tr>
<tr>
<td>GRMZM2G363908</td>
<td>AT4G25480</td>
<td>DEHYDRATION RESPONSE ELEMENT B1A (DREB1A)</td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G307152</td>
<td>AT1G71520</td>
<td></td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G369912</td>
<td>AT1G53170</td>
<td>ETHYLENE RESPONSE FACTOR 8 (ERF8)</td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G125460</td>
<td>AT2G47520</td>
<td>HYPOXIA RESPONSIVE ERF (ETHYLENE RESPONSE FACTOR 2 (HRE2))</td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G384386</td>
<td>AT5G19790</td>
<td>RELATED TO AP2 11 (RAP2.11)</td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G382379</td>
<td>AT3G48590</td>
<td>NUCLEAR FACTOR Y, SUBUNIT C1 (NF-YC1)</td>
<td>CCAAT</td>
</tr>
<tr>
<td>GRMZM2G386337</td>
<td>AT5G64750</td>
<td>ABA REPRESSOR1 (ABR1)</td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G390641</td>
<td>AT4G30080</td>
<td>AUXIN RESPONSE FACTOR 16 (ARF16)</td>
<td>ARF</td>
</tr>
<tr>
<td>GRMZM2G159119</td>
<td>AT1G25550</td>
<td></td>
<td>G2-like</td>
</tr>
<tr>
<td>GRMZM2G016370</td>
<td>AT1G25550</td>
<td></td>
<td>G2-like</td>
</tr>
<tr>
<td>GRMZM2G114389</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRMZM2G009598</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Coregulated with PPDK

<table>
<thead>
<tr>
<th>Parent Gene</th>
<th>best hit AGI</th>
<th>gene name of corresponding AGI (where available)</th>
<th>transcription factor class</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRMZM2G330751</td>
<td>AT4G21390</td>
<td>(B120)</td>
<td></td>
</tr>
<tr>
<td>GRMZM2G039246</td>
<td>AT3G16360</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRMZM2G070315</td>
<td>AT1G05000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRMZM2G079768</td>
<td>AT3G13850</td>
<td>LOB DOMAIN-CONTAINING PROTEIN 22 (LBD22)</td>
<td>LOB</td>
</tr>
<tr>
<td>GRMZM2G450273</td>
<td>AT5G24860</td>
<td>FLOWERING PROMOTING FACTOR 1 (FPF1)</td>
<td></td>
</tr>
</tbody>
</table>
coregulated with NADP-ME

<table>
<thead>
<tr>
<th>Parent Gene</th>
<th>best hit AGI</th>
<th>gene name of corresponding AGI (where available)</th>
<th>transcription factor class</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRMZM2G040762</td>
<td>AT1G01040</td>
<td>DICER-LIKE 1 (DCL1)</td>
<td></td>
</tr>
<tr>
<td>GRMZM2G346466</td>
<td>AT2G21320</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRMZM2G052667</td>
<td>AT1G53910</td>
<td>RELATED TO AP2 12 (RAP2.12)</td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G110333</td>
<td>AT3G14230</td>
<td>RELATED TO AP2 2 (RAP2.2)</td>
<td>AP2-EREBP</td>
</tr>
<tr>
<td>GRMZM2G092483</td>
<td>AT1G66140</td>
<td>ZINC FINGER PROTEIN 4 (ZFP4)</td>
<td>C2H2</td>
</tr>
<tr>
<td>GRMZM2G427618</td>
<td>AT4G35270</td>
<td></td>
<td>RWP-RK</td>
</tr>
<tr>
<td>GRMZM2G086403</td>
<td>AT4G17900</td>
<td></td>
<td>PLATZ</td>
</tr>
<tr>
<td>GRMZM2G074821</td>
<td>AT1G29810</td>
<td>CELLULOSE SYNTHASE LIKE B5</td>
<td></td>
</tr>
<tr>
<td>GRMZM2G074821</td>
<td>AT1G29810</td>
<td>CELLULOSE SYNTHASE LIKE B5</td>
<td></td>
</tr>
</tbody>
</table>

**Supplemental Table 2: Primer used for qRT-PCR.** All sequences in 5’-3’ orientation.

<table>
<thead>
<tr>
<th></th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>18SrRNA</td>
<td>CTGTCGGCCAAGGCTATATACT</td>
<td>CAGAACATCTAAGGGCATCACAG</td>
</tr>
<tr>
<td>CALRETICULIN 3 (CRT3)</td>
<td>GGCACGGGAGGAAGGG</td>
<td>GGTAGTCGTAGTGCCTATGACGC</td>
</tr>
<tr>
<td>Phosphotyrosine protein phosphatases superfamily protein</td>
<td>CCTCAAGTCCCTCAACCTCC</td>
<td>GAGCCTGATCCCATTTCTTTTC</td>
</tr>
<tr>
<td>Plant regulator RWP-RK family protein</td>
<td>CGCCTTCTTCGTATCGTC</td>
<td>CAGGCTTGGGCCACCA</td>
</tr>
</tbody>
</table>
6.1 First publication: Systems Analysis of a Maize Leaf Developmental Gradient

Redefines the Current C₄ Model and Provides Candidates for Regulation

Status: Published (December 2011)

Thea R. Pick, Andrea Bräutigam, Urte Schlüter, Alisandra K. Denton, Christian Colmsee, Uwe Scholz, Holger Fahnenstich, Roland Pieruschka, Uwe Rascher, Uwe Sonnewald, and Andreas P.M. Weber

Journal: "Plant Cell"

Impact factor: 10.224

1. Co-Author

Own contribution: 75%

- Preliminary experiments
- Sampling of the gradient
- Measured all parameters except transcripts and relative metabolite content
- Absolute metabolite content
- Data analysis
- Co-writing the manuscript
PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a new class of metabolite transporters.
**PLGG1**, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a new class of metabolite transporters


*both authors contributed equally to this work

a Institute of Plant Biochemistry, Heinrich Heine University Düsseldorf, Center of Excellence on Plant Sciences (CEPLAS), 40225 Duesseldorf, Germany

b Max-Planck Institute for Molecular Plant Physiology, 14476, Potsdam-Golm, Germany

Photorespiratory carbon flux reaches up to a third of photosynthetic flux and thus contributes massively to the global carbon cycle. The pathway recycles glycolate-2-phosphate, the most abundant byproduct of RubisCO reactions. This oxygenation reaction of RubisCO and subsequent photorespiration limit significantly the biomass gains of many crop plants. Although photorespiration is a compartmentalized process with enzymatic reactions in the chloroplast, the peroxisomes, the mitochondria and the cytosol, no transporter required for the core photorespiratory cycle has been identified at the molecular level to date. Using transcript co-expression analyses, we identified Plastidal glycolate glycerate translocator 1 (PLGG1) as a candidate core photorespiratory transporter. Related genes are encoded in the genomes of archaea, bacteria, fungi, and all Archaeplastida and have previously been associated with a function in programmed cell-death. A mutant deficient in PLGG1 shows WT-like growth only in an elevated carbon dioxide atmosphere. The mutant accumulates glycolate and glycerate, leading to the hypothesis that PLGG1 is a glycolate/glycerate transporter. This hypothesis was tested and supported by in vivo and in vitro transport assays and $^{18}$O$_2$-metabolic flux profiling. Our results indicate that PLGG1 is the chloroplastidic glycolate/glycerate transporter, which is required for the function of the photorespiratory cycle. Identification of the PLGG1 transport function will facilitate unraveling the role of similar proteins in bacteria, archaea, and fungi in the future.
**Introduction**

Carbon flux through photorespiration is second in magnitude only to photosynthesis and thus this metabolic pathway constitutes a major component of the global carbon cycle. Photorespiration is essential since the enzyme RubisCO, which assimilates CO\(_2\) from the atmosphere into biomass, also catalyzes a futile reaction, the oxygenation of the CO\(_2\) acceptor, ribulose 1,5-bisphosphate (RuBP). This latter reaction leads to the formation of the toxic metabolite 2-phosphoglycolate, which is detoxified and recycled to RuBP by a complex metabolic pathway called photorespiration. Photorespiration became an essential requirement for photosynthesis after the carbon composition of the atmosphere changed to an oxygen rich atmosphere, as a consequence of oxygenic photosynthesis by cyanobacteria, algae, and plants. Large gains in photosynthetic efficiency can be achieved if photorespiration is suppressed by enriching CO\(_2\) in the vicinity of RubisCO. For example, plants carrying a metabolic bypass for photorespiration indeed produce more biomass (1, 2) providing a promising approach for increasing the productivity of some the most important crop plants, such as rice (*Oryza sativa*) or wheat (*Triticum aestivum*) that have to cope with high rates of photorespiration.

In plants, the photorespiratory cycle is a highly compartmentalized process with enzymatic reactions in chloroplasts, peroxisomes, and mitochondria as well as in the cytosol. In the chloroplast stroma 2-PG resulting from the RubisCO oxygenation reaction is dephosphorylated to glycolate by 2-phosphoglycolate phosphatase (PGLP). Glycolate is exported from the chloroplasts to the peroxisomes where it is oxidized to glyoxylate by glycolate oxidase (GOX) and transaminated to glycine by Ser:glyoxylate and Glu:glyoxylate aminotransferase (SGT and GGT, respectively). Glycine leaves the peroxisomes and enters the mitochondria where two molecules of glycine are deaminated and decarboxylated by the glycine decarboxylase complex (GDC) and serine hydroxymethyltransferase (SHMT) to form one molecule of each, serine, ammonia, and carbon dioxide. Serine is exported from the mitochondria to the peroxisomes where it is predominantly converted to glyceraldehyde by SGT and hydroxypyruvate reductase (HPR). Glycerate leaves the peroxisomes and is taken up into the chloroplast where it is phosphorylated by glyceraldehyde kinase (GLYK) to yield 3-PGA. In essence, one of the four carbon atoms contained in two molecules of 2-PG entering the pathway is lost as CO\(_2\), whereas three are shuttled back into the Calvin-Benson cycle. Thus, photorespiration constitutes a metabolic repair cycle that is required to detoxify 2-PG, at the expense of energy and loss of carbon dioxide and ammonia. Most of the components of the photorespiratory cycle have been identified by forward genetic analyses starting in the 1980s. For example the first enzyme of the photorespiratory cycle PGLP (3), or the mitochondrial SHMT (4, 5) have been discovered by this approach. The hallmark of photorespiratory mutants is reduced or no growth under ambient air but normal WT-like growth under CO\(_2\) enriched air. Only peroxisomal HPR mutants show no typical photorespiratory growth limitation in ambient air due to an alternative cytosolic pathway that
suppresses the effect of the mutation (6). Photorespiratory mutants additionally contain elevated pools of photorespiratory metabolites.

The photorespiratory pathway intermediates have to be transported across multiple cellular membranes at high flux rates, which is facilitated by metabolite transporters residing in the organellar membranes. In total two molecules of glycolate and one molecule of glycerate cross the chloroplast envelope and the peroxisomal membrane in one turn of the photorespiratory cycle. Two molecules of glycine and one molecule of serine cross the peroxisomal membrane and the mitochondrial envelope. These six transport steps in the carbon cycle of photorespiration together with the transport steps required for the nitrogen cycle result in 18 postulated transport processes (see (7), Eisenhut et al, 2012 for review). While all soluble enzymes of the photorespiratory cycle are identified on the molecular level, this is only true for the chloroplastidic transporters DiT1 and DiT2 that are involved in nitrogen recycling during photorespiration (8-11). Transporters required for the photorespiratory carbon cycle are still unknown, which represents a major gap in the knowledge about this important pathway.

While most enzymes required for photorespiration have been identified by forward genetics, only one transporter associated with photorespiration has been found by this approach to date (8, 10). An alternative means for the identification of candidate genes is transcript co-expression analysis, which was developed by Eisen et al. (12) using microarray data for yeast. Co-expression analysis is based on the assumption that genes that function in the same pathway tend to display similar expression patterns. Hence, unknown genes that are co-regulated with genes of a particular metabolic pathway are hypothesized to be involved in the same biological process. By this method a wide range of genes were functionally characterized in yeast (13) or humans (14). In Arabidopsis thaliana a co-expression analysis in combination with reverse genetics has been a successfully strategy to find genes involved in, e.g., flavonoid metabolism (15), cellulose and (16) lignin biosynthesis (17, 18), and aliphatic glucosinolate biosynthesis (19).

In this work we have used co-expression analysis in combination with reverse genetics to identify a transporter involved in the photorespiratory cycle. Co-expression analysis revealed that the genes encoding photorespiratory enzymes are co-expressed with each other and with the candidate transporter Plastidal glycolate glycerate translocator 1 (PLGG1), which was previously identified in proteomics of chloroplasts (20), and was assumed to be involved in programmed cell death (21, 22). By analyzing an A. thaliana T-DNA knockout mutant deficient in PLGG1 (plgg1-1), we could demonstrate that the mutant has a strong photorespiratory phenotype and is no longer able to transport glycolate and glycerate across the chloroplast envelope. The chloroplastidic glycolate/glycerate transporter PLGG1 thus defines a new class of metabolite transporters that is present in Archaeplastida, fungi, bacteria, and archaea.
Results

PLGG1 is co-expressed with known photorespiratory enzymes. To identify transporters involved in the photorespiratory pathway a co-expression analysis with eleven known photorespiratory pathway enzymes was conducted using publicly available co-expression databases (AtGenExpress & NASC Array, CSB.DB (http://csbdb.mpimp-golm.mpg.de/)). Photorespiratory genes were significantly co-expressed with approx. 100 to 150 genes, depending on the co-expression matrix used. These included genes involved in photorespiration, photosynthesis, and chloroplast function (23). The Spearman and Pearson co-expression coefficients for ten photorespiratory genes exceeded 0.9 for 73 out of 98 tested cases (Table 1). PGLP is co-expressed with all genes involved in the pathway with the exception of GLYK (Table 1). For some of the enzymes several isoforms of the enzymes exist. In these cases, only one distinct isoform was found to be co-expressed with other photorespiratory genes. For example, of the five isoforms encoding enzymes for glycolate oxidation, only AtGOX1 is strongly correlated (AtGOX2 is indistinguishable due to the probe on the ATH1 chip) (Table 1). Likewise only GGT1 but not GGT2, only one isoform each of the subunits of GDC, and only the SHM1 isoform is co-expressed in the context of photorespiration (Table 1). Only GLYK is not co-expressed with the other genes in the pathway (Table1).

To identify unknown transporters in the photorespiratory cycle, proteins of unknown function with a co-expression coefficient of at least 0.9 with the majority of photorespiratory genes were tested for the presence of predicted membrane spanning helices. One protein of unknown function previously identified by proteomics and named PLGG1 (20) was both co-expressed with genes involved in photorespiration and had twelve predicted membrane spanning helices (ARAMEMNON, (24)). We hypothesized that PLGG1 is a transporter involved in photorespiration.

The plgg1-1 insertion mutant exhibits a photorespiratory phenotype

To test the hypothesis that PLGG1 is involved in photorespiration, we isolated an A. thaliana T-DNA insertional mutant. The plgg1-1 mutant carries the T-DNA insertion in the first intron of the respective gene At1g32080 (SI1). PLGG1 cDNA can only be detected in WT but not in plgg1-1 mutant plants (SI1). Under ambient CO₂ conditions (380 ppm) the plgg1-1 mutant developed yellow and bleached lesions on the leaf lamina but not along the veins (Fig. 1A). This phenotype was suppressed under high CO₂ (3000 ppm) conditions (Fig. 1B). The phenotype was complemented when PLGG1 was expressed under its own promoter in plgg1-1 plants (Fig. 1C). We conclude that plgg1-1 displays a bleached leaf phenotype in ambient CO₂ that can be suppressed by elevated CO₂, which is consistent with a function of PLGG1 in photorespiration.

Photorespiratory metabolites accumulate in plgg1-1 plants
To determine the position of PLGG1 in the photorespiratory pathway, we analyzed metabolites in WT and plgg1-1 plants under high and ambient CO$_2$ conditions. The chosen time points represent metabolite levels under: (i) high CO$_2$ (0 days), conditions under which the rate of photosynthesis was nearly identical in WT and mutant (14.56 ± 0.42 and 13.87 ± 0.23 µmol CO$_2$ m$^{-2}$ s$^{-1}$, respectively, SI2); (ii) after shift to ambient CO$_2$ conditions of plants with no or very weak visible phenotype but already decreased photosynthetic capacity in the mutant (2 days, 6.7 ± 0.76 µmol CO$_2$ m$^{-2}$ s$^{-1}$); and (iii) after a pronounced visible phenotype and low rates of photosynthesis were observed in the mutant (5 days, 4.62 ± 0.47 µmol CO$_2$ m$^{-2}$ s$^{-1}$). The largest and most significant differences between plgg1-1 mutant plants and WT were found for glycolate, glycine, serine, hydroxypyruvate, and glycerate, which are all photorespiratory intermediates. Already in CO$_2$ enriched air (0 days) glycolate and glycerate were significantly elevated in the mutant compared to WT (Fig. 2A and F). This accumulation escalated when plants were shifted from elevated CO$_2$ to ambient air. Glycine, serine, and hydroxypyruvate did not accumulate in mutant plants under high CO$_2$ but when shifted to ambient air (Fig. 2C-E). Steady state glyoxylate levels in mutant plants did not differ significantly from those in WT. After shift from high CO$_2$ to ambient CO$_2$ conditions, only serine levels changed in WT plants (Fig. 2C).

Since photorespiration occurs only during the day when RubisCO is active, we expected day-time dependent metabolite accumulation patterns in the mutant. During the night in elevated CO$_2$, in WT and plgg1-1 plants, photorespiratory metabolite levels with the exception of glycerate did not differ strongly (Fig. 3). During the day in elevated CO$_2$, only glycerate and glycolate accumulated in a time dependent manner (Fig. 3A and E). After shift to ambient air, all five metabolites showed a light dependent accumulation that became more pronounced throughout the light period. During the night in ambient air glycolate and glycine levels in plgg1-1 plants dropped to WT levels. Serine and hydroxypyruvate levels did not mirror the values during the night in CO$_2$ enriched air, but stayed elevated (Fig. 3C-D). The metabolite profile of the plgg1-1 mutant is consistent with a role of PLGG1 in photorespiration. The pronounced accumulation of glycolate and glycerate and the localization of this transporter at the chloroplast envelope membrane ((25), SI2) point to PLGG1 as a glycolate/glycerate transporter.

**Glycerate- and light-dependent O$_2$ evolution in intact chloroplasts**

Transport experiments with isolated chloroplasts showed that a single transporter transports glycolate and glycerate across the envelope (26, 27). Therefore we expressed PLGG1 heterologously and measured active $^{14}$C-glycerate uptake in an in vitro uptake system. A Michaelis-Menten-type saturation kinetics of glycerate uptake was observed when liposomes were preloaded with glycerate or glycolate. However, a high background signal was detected in the liposome system due to high rates of unspecific diffusion of glycerate in the liposome system (SI3). Therefore to corroborate the result of the liposome uptake assays, we employed an in vivo
system that was successfully used before (26) to test whether the plgg1-1 mutant was affected in the transport of glycerate. Chloroplasts show light-dependent oxygen evolution in the presence of glycerate (26) because imported glycerate is converted by glycerate kinase to 3-PGA, which is further reduced to triosephosphate. The reduction consumes NADPH and the photosynthetic regeneration of NADPH produces oxygen. Hence, feeding of glycerate to isolated chloroplasts drives oxygen evolution (for a detailed scheme see SI4).

To test whether plgg1-1 chloroplasts were physiologically intact and transport competent, 3-PGA-dependent oxygen evolution was measured. The rates of 3-PGA-dependent oxygen evolution observed for WT and comp chloroplasts (20.32 μmol/mg Chl*h) did not differ significantly from the rates observed for plgg1-1 chloroplasts (Table 2). With WT and comp chloroplasts O₂ evolution rates of 22.54 and 18.77 μmol/mg Chl*h, respectively, were observed with 1mM glycerate. The rates observed with plgg1-1 chloroplasts did not exceed the background signal and differed significantly (P < 0.0001) from the WT rates. Hence chloroplasts of the plgg1-1 mutant do not display glycerate-dependent oxygen-evolution, which supports PLGG1’s function as a glycerate transporter.

**Glycolate and glycerate flux are impaired in plgg1-1 plants**

PLGG1’s function as a glycolate transporter was further tested by [¹⁸Ο]oxygen flux analysis. Transfer of label from glycolate to downstream metabolites would be impaired in plgg1-1 if PLGG1 was involved in catalyzing the efflux of glycolate from the chloroplast. To follow photorespiratory flux in plgg1-1 in relation to WT plants we incubated plants in an [¹⁸Ο]oxygen-atmosphere and followed the incorporation of the label into metabolites, a method that was previously used to identify metabolites and the kinetics of the photorespiratory cycle (28). The label was incorporated into glycolate within 5 sec and into glycerate after 3 min (28). Therefore we chose timepoints 30 sec, 1 min, 2 min and 5 min for a kinetic analysis. All values were expressed relative to the initial label (30 sec). For glycolate, the values observed with WT and comp plants were similar to those obtained by Berry et al. (Fig. 4). After 30 sec exposure to ¹⁸Ο, the incorporation reached a plateau and no further increase in label incorporation was detectable after 1 (0.93 fold for WT and 0.86 fold for comp), 2 (0.88 fold for WT and 0.81 fold for comp), and 5 min (0.88 fold for WT and 0.88 fold for comp), respectively, indicating that a steady state was achieved. For plgg1-1 plants, no increase in label incorporation into glycolate was obtained after 1 min as compared to the value measured after 30 sec. In contrast to WT, the label in glycolate continued to accumulate after 2 min (1.11 fold) and 5 min (1.86 fold) in the mutant, which indicates a slower removal of label from the glycolate pool. This was consistent with a reduced export of glycolate from the chloroplast. In WT and comp plants, the ¹⁸Ο label incorporation in glycerate increased slowly. This was not true for the mutant plants. Here no increase in label incorporation into glycerate could be observed with continuous low values for
all time points. Again, this data indicates that the transfer of label from the glycolate to the glycerate pool is impaired, which can be explained by reduced export of glycolate from the chloroplast. Thus, *plgg1-1* is impaired in the transport of both glycolate and glycerate.

**Discussion**

Photorespiration is a highly compartmentalized pathway that requires multiple transmembrane transport steps across organelar membranes to enable the high flux of metabolites in this process. Despite their importance, no transporter involved in the carbon cycle of photorespiration has been identified to date. Here we report the identification of the chloroplastidic glycolate/glycerate transporter PLGG1 as the first transporter of the photorespiratory carbon cycle.

Co-expression analysis showed that PLGG1 is co-regulated with photorespiratory enzymes (Table 1) and an *A. thaliana* T-DNA insertional mutant deficient in *plgg1-1* (SI1) is only able to grow WT-like and exhibit WT-like rates of photosynthesis in elevated CO$_2$ (Fig. 1A-B, SI2). In previous studies this CO$_2$-dependent phenotype was also observed for other photorespiratory mutants (4, 29-31). Those plants grow normally in CO$_2$ enriched air but have reduced growth or are even inviable under ambient CO$_2$. Compared to other photorespiratory mutants, the phenotype observed for *plgg1-1* plants is relatively mild because plants can still grow, albeit more slowly and with visible symptoms (Fig. 1), in ambient air, which is likely the reason it was not identified in previous photorespiratory mutant screens. Glycolate, a small organic acid, is probably able to diffuse out of chloroplasts, as do other small organic acids (32) once it accumulates to high levels (Fig. 2). In the absence of high photorespiratory flux during nighttime, the glycerate and glycolate pools approach wild type levels (Fig. 3), which likely alleviated the photorespiratory symptoms in the *plgg1-1* mutant. A shift in night time metabolism at least partially metabolized extra-plastidial glycerate, likely towards serine synthesis which explains why, during the night in ambient air, serine and hydroxyxypyruvate levels continue to rise while glycerate levels fall but do not reach WT levels. Accumulation of the photorespiratory metabolites glycine, serine and hydroxyxypyruvate is likely driven by feedback through cytoplasmatic and peroxisomal glycerate accumulation and was previously observed in other photorespiratory mutants with elevated glycerate levels (6, 30).

Glycerate and glycolate are transported across the chloroplast envelope in the photorespiratory cycle by the same transport protein (26). The localization of PLGG1 in the chloroplast envelope (SI 3) together with the metabolite accumulation patterns (Fig. 2 and 3) identify PLGG1 as the chloroplastidic glycolate/glycerate transporter, that was characterized biochemically the 1980s and 1990s in McCarthy's laboratory (26, 27, 33-35). It was shown that the transport rates observed with isolated chloroplasts are sufficient to cope with the high
photorespiratory carbon flux, that the transporter exhibits a proton/substrate symport activity, and that glycolate and glycerate are transported through the same transport protein.

To verify that PLGG1 is the glycolate/glycerate transporter, we expressed PLGG1 heterologously and measured \(^{14}\text{C\)-glycerate/glycolate} counter-exchange activity in a reconstituted liposome assay. While a saturable and preloading-dependent uptake kinetics could be observed, also high rates of unspecific diffusion were detected (SI3). This high rate of diffusion is typical for small organic acids in artificial membrane systems and was detected before (27). We therefore next employed an in vivo approach that is more suitable to verify glycolate/glycerate transport. To this end, we used an in vivo approach with isolated chloroplasts that was developed by Howitz and McCarty (26) and \(^{18}\text{O\)-flux analysis} in plgg1-1 plants. WT chloroplasts evolved oxygen when provided with 3-PGA or glycerate, proving that they were transport competent and biochemically active. PLGG1-1 chloroplasts were also transport competent and biochemically active as they were capable of evolving oxygen when supplied with 3-PGA (Table 2). They were only unable to transport glycerate as no glycerate-dependent oxygen evolution was observed with plgg1-1 chloroplasts (Table 2). Flux analysis further supported PLGG1’s role as the glycolate/glycerate transporter since the \(^{18}\text{O\)-label} accumulation in glycolate increased in the plgg1-1 plants over time indicating that glycolate is trapped in the chloroplast and is not accessible to enzymes that process glycolate outside of the chloroplast (Fig. 3). In contrast, the label accumulation in glycolate in WT plants already reached a plateau after 30 sec and did not increase, indicating that glycolate was processed outside of the chloroplast and can thus proceed in the photorespiratory cycle. In WT plants label incorporation into glycerate increased over time as it takes a few minutes until turnover of the majority of the glycerate pool is achieved (28). In plgg1-1 plants the transfer of the label from glycolate to other metabolites is mostly blocked and thus no increase in label in the glycerate pool is detectable.

Two recent publications hypothesized that PLGG1 is involved in chloroplast development or functions against cell death (21, 22) but in neither publication was the molecular function of PLGG1 identified. The conclusions were drawn from the visible observations that true leaves develop chlorotic regions in which chloroplasts are destroyed. We demonstrate that that elevated CO\(_2\) alleviates the symptoms in all developmental stages (Fig. 1, SI5), that the accumulation of photorespiratory metabolites predates the occurrence of visible symptoms (Fig. 2), and that PLGG1 is the chloroplast glycerate/glycolate carrier (Fig. 4, Table 2). Since the phenotype can be suppressed with high CO\(_2\) (Fig. 1) and biochemically active chloroplasts can be isolated from plgg1-1 mutant plants grown in high CO\(_2\) (Table 2) we posit that the visible symptoms are due to the accumulation of toxic concentrations of glycolate and glycerate. Indeed photorespiratory intermediates including glycerate can be toxic to chloroplasts at high concentrations (36) explaining chloroplast and cell disruption in the plgg1-1
mutant. Thus, the chloroplastic glycolate/glycerate transporter Plgg1 defines a new class of metabolite transporters, which is present in Archaeplastida, fungi, bacteria, and archaea (SI6).

**Material and Methods**

**Plant growth and conditions**

*A. thaliana* ecotype Columbia (Col-0) was used as wild-type reference (WT). The SALK line SALK_053469 (*plgg1-1*) was obtained from the Nottingham Arabidopsis Stock Centre (37). Complemented *plgg1-1* mutant plants (*comp*) were used as control for the SALK line. Unless stated otherwise, plants were grown in normal air (380 ppm CO₂) and in air with elevated CO₂ (3000 ppm; high CO₂) at a 12-h-light/12-h-dark cycle (22/18°C) in growth chambers (150 μmol m⁻² sec⁻¹ light intensity) on soil (mixture of 1/4 Floraton and ¾ Arabidopsis root substrate).

**Isolation of the T-DNA insertion line**

PCR-based screening was used to isolate a homozygous T-DNA insertion line for *Plgg1*. Primers P1, P2 and P3 were used for the genomic DNA screening (for primer sequences see SI Table 1). P1 and P2 for amplification of the WT gene and P1 and P3 for the T-DNA/gene junction. The effect of the T-DNA insertion on the amount of *PLGG1* transcript amounts was tested using qPCR with cDNA of WT and *plgg1-1* plants as template and primers P4 and P5 for amplification. As a positive control *ACTIN7* (AtACT7, At5g09810) was amplified using P6 and P7.

**Statistical analysis**

Curve fits and Student’s t-tests were performed with PRISM5.0a (GraphPad, http://www.graphpad.com/prism/prism.htm). Results were called as 'extremely significant’ if the *P* value was < 0.01% (*P*<0.0001) and 'very significant' if the *P* value was < 0.1% (*P*<0.001), and is indicated by three and two asterisks, respectively.

**beta-glucuronidase (GUS) expression and establishment of a complementation line**

To assess the expression profile of the *PLGG1* gene, a 3.5 kb gDNA fragment upstream of the ATG-start site including the first nine bases of the *PLGG1* gene was amplified using P8/P9 and cloned into vector pCAMBIA3301 for a C-terminal GUS fusion. For complementation analysis a 9 kb gDNA fragment, including 3.5 kb upstream of the ATG-start site, the full genomic *PLGG1* sequence und 0.5 kb 3’-UTR was amplified using P10/P11 and cloned into vector pCAMBIA3301. *A. thaliana* plants were stably transformed using the floral dip method (38). GUS staining of two week old seedlings was performed using the method described in (39).

**Metabolite Extraction, Gas Chromatography–Time of Flight–Mass Spectrometry Analysis**

Methanolic extraction of leaf material was performed according to the method described by
Fiehn et al. (40) and GC/MS was performed according to Lee & Fiehn (41). Analysis of metabolites was performed by GC/MS (Agilent Technologies 5973 (Santa Clara, CA, USA)). Results were analyzed using the MassLynx software package supplied with the instrument (Waters). As an internal standard ribitol was added and relative metabolite levels were determined from the ratio of the area of each metabolite and the corresponding ribitol area. A detailed description of the methods is available in the supplement.

**Transient expression of GFP fusions in tobacco protoplast**

For localization studies a C-terminal GFP fusion construct was cloned using Gateway vectors to insert PLGG1 into pMDC83 (42) via pDONR207 (Invitrogen, Gateway®). For amplification primer P12/P13 were used. Vectors were transformed into *Nicotiana benthamiana* leaves via infection with *Agrobacterium tumefaciens* strain GV3101 (43, 44), protoplasts were isolated after three days and localization was visualized using a Zeiss Laser scanning microscope 510 Meta (Zeiss, http://www.zeiss.com) as described in detail in Breuers et al. (25).

**Chloroplast isolation and glycerate dependent oxygen evolution**

Chloroplasts were isolated according to the method described by Aronsson and Jarvis (45) using a two-step-Percoll gradient, with the modification that plants were grown on soil for three weeks in high CO₂. Intactness of chloroplasts was determined using the Hill-reaction (46). 3-PGA and glycerate dependent oxygen evolution was measured according to the method described by Howitz and McCarty (26).

**18O₂ feeding**

*Arabidopsis thaliana* WT and *plgg1-1* plants were fed with 18O₂ according to the method described in (28). Plants were grown on soil for four weeks in ambient air in 8-h-light/16-h-dark cycle (20/16°C). A single plant in a pot was placed in a 1 L plastic bag with a plastic seal (Padan GmbH, Henfenfeld). Air was removed by vacuum and replaced by air-mixture of 0.03% CO₂, 78.97% N₂ and 21% 18O₂. After 0.5, 1, 2 and 5 min plants were freeze-quenched within less than 2 s in liquid nitrogen and processed for metabolite analysis. Metabolites were extracted by the method optimized for photorespiratory metabolites. A detailed description of the methods is available in the supplement.

**Acknowledgments:**

We thank D. Weits and J. van Dongen for providing and helping with the gas mixer equipment.
References:


Figure legends:

**Fig. 1.** Photorespiratory phenotype of *plgg1-1* plants; (A) *plgg1-1*, (C) WT and (D) *comp* plants grown in high CO$_2$ for four weeks, shifted to ambient CO$_2$ for one week. (B) *Plgg1-1* plant grown in high CO$_2$ for five weeks.

**Fig. 2.** Accumulation of photorespiratory metabolites in WT and *plgg1-1* plants. WT and *plgg1-1* plants were grown in 3000ppm CO$_2$ conditions for four weeks and then shifted to ambient air. Steady state metabolite levels were measured for plants kept at high CO$_2$ and after two and five day shift to ambient air; all values are measured in µmol/g fresh weight except hydroxypyruvate which is shown as arbitrary units; glycolate (A), glyoxylate (B), glycine (C), serine (D), hydroxypyruvate (E) and glycerate (F). Error bars indicate Standard deviation, n = 3.

**Fig. 3.** Time course of photorespiratory metabolite accumulation in WT and *plgg1-1* plants. WT and *plgg1-1* plants were grown in 3000ppm CO$_2$ conditions for four weeks and shifted to ambient air (300ppm, at 24 hours, indicated by black arrow); all values are measured in µmol/g fresh weight except hydroxypyruvate which is shown as arbitrary units; glycolate (A), glycine (B), serine (C), hydroxypyruvate (D) and glycerate (E). Error bars indicate Standard deviation, n = 3.

**Fig. 4.** Revised model of the photorespiratory cycle and $^{18}$O$_2$ incorporation in glycolate and glycerate; (A) The photorespiratory cycle including the new chloroplastidic glycolate/glycerate transporter PLGG1. $^{18}$O$_2$ incorporation into glycolate (A) and glycerate (B) was determined in WT, complemented mutants (*comp*) and *plgg1-1* mutant plants. Plants were grown in ambient air for four weeks, incubated in $^{18}$O$_2$ air for 0.5, 1, 2 and 5 min and metabolite levels were analyzed using GC/MS. Metabolite levels were normalized to the relative abundance after 0.5 min in order to calculate the enrichment of $^{16}$O$_2$ incorporation in glycolate and glycerate. 3-PGA: 3-Phosphoglycerate; ETC: electron transport chain; 2-OG: 2-Oxoglutarate; TP: triose phosphates. N = 3.
Table legends:

Table 1: Co-expression Spearman coefficients for genes involved in photorespiration and for the new candidate gene PLGG1; only the isoform with the best correlation coefficient is shown; correlations with p<0.1 are printed boldface.

Table 2. 3-PGA and glycerate dependent O₂ evolution in isolated intact WT, plgg1-1 and comp chloroplasts [µmol/mg Chl*h]. ± indicates Standard deviation, n = 3.

Supporting Information:

SI Fig. 1: T-DNA insertion leads to full k.o. of the gene At1g32080. (A) The T-DNA is inserted in the first intron of the gene At1g32080. The arrow symbolizes the transcription start side. Grey boxes stand for exons, lines for introns, dark gray boxes for 5′-UTR, light gray boxes for 3′-UTR. (B) Expression control of the k.o. gene. WT and mutant gene show actin expression. Only WT shows PLGG1 expression.

SI Fig. 2. PLGG1 is located at the chloroplast envelope. (A) GFP fluorescence; (B) chlorophyll auto-fluorescence and (C) merge of both signals; GFP was fused C-terminal translationally to the full-length PLGG1 protein; the construct was transiently expressed in Nicotiana benthamiana; protoplasts were used for microscopic detection of the GFP signal.

SI Fig. 3. PLGG1 shows an active and time-dependent uptake of [¹⁴C]Glycerate, that can be inhibited by increasing concentrations of unlabeled glycerate.

(A) Time-dependent uptake of radiolabelled [¹⁴C]Glycerate (50 µM external concentration) in the presence (filled circles) or absence of glycolate (open rectangles) as internal substrate (20 mM). In the presence of a suitable counter-exchange substrate inside the liposomes (e.g., glycolate) a Michaelis-Menten-type saturation kinetics was observed. In control liposomes preloaded with equal amounts of buffer or in liposomes that did not contain reconstituted PLGG1, only unspecific diffusion of the labeled substrate was observed.

(B) Concentration-dependent inhibition of active transport in the presence (black bars) or absence of glycolate (white bars) as internal substrate (20 mM). Active transport is inhibited with increasing external glycerate concentrations, while passive diffusion is independent on external glycerate concentration.

Error bars represent SD of three independent uptake experiments.
Recombinant At1g32080 protein was expressed in vitro using the 5PRIME wheat germ expression kit (5PRIME GmbH, Hamburg, Germany) as described previously (1) and reconstituted into liposomes using reconstitution and uptake buffers as described in Howitz et al. (2).

**SI Fig. 4. Scheme of glycerate dependent oxygen evolution.** Glycerate is transported by PLGG1 into the chloroplast in a proton-dependent manner, converted to 3-PGA under ATP consumption. 3-PGA is converted to TP under NADPH consumption. NADP is regenerated to NADPH at the photosynthetic electron transport chain under water consumption and oxygen evolution (red arrow). 3-PGA: 3-Phosphoglycerate; TP: triose phosphates.

**SI Fig. 6. Phylogenetic Tree of PLGG1.** AtPLGG1 was used as the query to Blast Explorer (http://www.phylogeny.fr/version2.cgi/one_task.cgi?task_type=blast) and the sequences suggested by the program were manually curated and fed to an “à la carte” phylogeny analysis (http://www.phylogeny.fr/version2.cgi/phylogeny.cgi). Briefly, sequences were aligned with Muscle and all positions with gaps were removed. The phylogenetic tree was calculated with PhyML and bootstrapped (100 repetitions). For display, all branches with <30% branch support were collapsed.

The tree divides in two large branches, a bacterial and an eukaryotic branch. The sole cyanobacterium *Acaryochloris* spec is nested within the bacteria. In eukaryotes, the tree divides the fungi (more precisely the dikarya) from the archaeaplantida. Within the archaeaplantida, *Cyanophora paradoxa* separates from the red and green lineages which themselves form separate branches. In the green lineage, the green algae separate from the streptophyta in which the moss branches before the fern. Monocots and dicots form separate branches.

Since the fungi branch with the archaeaplantida with a bootstrap support of 62%, the gene likely evolved before the archaeaplantida and the fungi split and it was lost from the other branches of life. Had the gene been acquired by a lateral gene transfer, one would expect it to branch within the bacteria. In higher plants PLGG1 is nuclear encoded. For transport proteins of the chloroplasts, at least three different phylogenetic roots have been demonstrated: rooted in the Chlamydiae, rooted in the cyanobacteria and a host origin. PLGG1-like proteins do not appear in the Chlamydiae but in Cyanobacteria. The phylogenetic tree indicates that the gene is likely not acquired from the cyanobacteria since the single cyanobacterium PLGG1 is nested within the bacteria. Therefore it is more likely that the cyanobacterium acquired it by lateral gene transfer from a bacterium. Hence PLGG1 likely originated from the host.


### Tables

Table 1: Co-expression Spearman coefficients for genes involved in photorespiration and for the new candidate gene PLGG1; only the isoform with the best correlation coefficient is shown; correlations with $p<0.1$ are printed boldface.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>AGI</th>
<th>Abbreviation</th>
<th>PGLP1</th>
<th>GOX1</th>
<th>AGT1</th>
<th>GGT1</th>
<th>GLDP1</th>
<th>GLDH1</th>
<th>GLDH3</th>
<th>GLDT1</th>
<th>SHM1</th>
<th>HPR1</th>
<th>GlyK</th>
<th>PLGG1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-PG phosphatase</td>
<td>At5g3 6790</td>
<td>PGLP1 1</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glycolate oxidase</td>
<td>At3g1 4420</td>
<td>GOX1</td>
<td>0.93</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ser:glyoxylate aminotransferase</td>
<td>At2g1 3360</td>
<td>AGT1</td>
<td>0.91</td>
<td>0.95</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glu:glyoxylate aminotransferase</td>
<td>At1g2 3310</td>
<td>GGT1</td>
<td>0.92</td>
<td>0.96</td>
<td>0.96</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GDC P-protein</td>
<td>At4g3 3010</td>
<td>GLDP1 1</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.94</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GDC H-protein</td>
<td>At2g3 5370</td>
<td>GLDH1</td>
<td>0.93</td>
<td>0.92</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GDC H-protein</td>
<td>At1g3 2470</td>
<td>GLDH3</td>
<td>0.91</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>0.94</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gly decarboxylase T-protein</td>
<td>At1g1 1860</td>
<td>GLDT1</td>
<td>0.92</td>
<td>0.92</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>0.92</td>
<td>0.95</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Serine:hydroxymethyltransferase</td>
<td>At4g3 7930</td>
<td>SHM1</td>
<td>0.93</td>
<td>0.96</td>
<td>0.94</td>
<td>0.96</td>
<td>0.95</td>
<td>0.91</td>
<td>&lt;0.9</td>
<td>1.00</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hydroxypyruvate reductases</td>
<td>At1g6 8010</td>
<td>HPR1</td>
<td>0.93</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
<td>0.94</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>0.91</td>
<td>0.96</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>Glycerate kinase</td>
<td>At1g8 0380</td>
<td>GlyK</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>&lt;0.9</td>
<td>1.00</td>
</tr>
<tr>
<td>AtPLGG1</td>
<td>At1g3 2080</td>
<td>PLGG1 1</td>
<td>0.93</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>&lt;0.9</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.91</td>
<td>0.93</td>
<td>0.93</td>
<td>&lt;0.9</td>
</tr>
</tbody>
</table>
Table 2. 3-PGA and glycerate dependent O$_2$ evolution in isolated intact WT, *plgg1-1* and *comp* chloroplasts [µmol/mg Chl*hr*]. ± indicates Standard deviation, n = 3.

<table>
<thead>
<tr>
<th></th>
<th>glycerate</th>
<th>3-PGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>22.54 (± 1.11)</td>
<td>22.12 (± 3.58)</td>
</tr>
<tr>
<td><em>plgg1-1</em></td>
<td>1.61 (± 1.72)</td>
<td>19.91 (± 1.72)</td>
</tr>
<tr>
<td><em>comp</em></td>
<td>18.77 (± 0.66)</td>
<td>20.32 (± 2.22)</td>
</tr>
</tbody>
</table>
Figures

Fig. 1. Photorespiratory phenotype of \textit{plgg1-1} plants; (A) \textit{plgg1-1}, (C) WT and (D) \textit{comp} plants grown in high CO\textsubscript{2} for four weeks, shifted to ambient CO\textsubscript{2} for one week. (B) \textit{Plgg1-1} plant grown in high CO\textsubscript{2} for five weeks.
Fig. 2. Accumulation of photorespiratory metabolites in WT and plgg1-1 plants. WT and plgg1-1 plants were grown in 3000ppm CO₂ conditions for four weeks and then shifted to ambient air. Steady state metabolite levels were measured for plants kept at high CO₂ and after two and five day shift to ambient air; all values are measured in µmol/ g fresh weight except hydroxypyruvate which is shown as arbitrary units; glycolate (A), glyoxylate (B), glycine (C), serine (D), hydroxypyruvate (E) and glycerate (F). Error bars indicate Standard deviation, n = 3.
Fig. 3. Time course of photorespiratory metabolite accumulation in WT and plgg1-1 plants. WT and plgg1-1 plants were grown in 3000ppm CO₂ conditions for four weeks and shifted to ambient air (300ppm, at 24 hours, indicated by black arrow); all values are measured in µmol/g fresh weight except hydroxypyruvate which is shown as arbitrary units; glycolate (A), glycine (B), serine (C), hydroxypyruvate (D) and glycerate (E). Error bars indicate Standard deviation, n = 3.
Fig. 4. Revised model of the photorespiratory cycle and $^{18}$O$_2$ incorporation in glycolate and glycerate; (A) The photorespiratory cycle including the new chloroplastidic glycolate/glycerate transporter PLGG1. $^{18}$O$_2$ incorporation into glycolate (A) and glycerate (B) was determined in WT, complemented mutants (comp) and plgg1-1 mutant plants. Plants were grown in ambient air for four weeks, incubated in $^{18}$O$_2$ air for 0.5, 1, 2 and 5 min and metabolite levels were analyzed using GC/MS. Metabolite levels were normalized to the relative abundance after 0.5 min in order to calculate the enrichment of $^{18}$O$_2$ incorporation in glycolate and glycerate. 3-PGA: 3-Phosphoglycerate; ETC: electron transport chain; 2-OG: 2-Oxoglutarate; TP: triose phosphates. N = 3.
Supporting Information

SI Fig. 1: T-DNA insertion leads to full k.o. of the gene At1g32080. (A) The T-DNA is inserted in the first intron of the gene At1g32080. The arrow symbolizes the transcription start side. Grey boxes stand for exons, lines for introns, dark gray boxes for 5’-UTR, light gray boxes for 3’-UTR. (B) Expression control of the k.o. gene. WT and mutant gene show actin expression. Only WT shows PLGG1 expression.
SI Fig. 2. PLGG1 is located at the chloroplast envelope. (A) GFP fluorescence; (B) chlorophyll auto-fluorescence and (C) merge of both signals; GFP was fused C-terminal translationally to the full-length PLGG1 protein; the construct was transiently expressed in *Nicotiana benthaminana*; protoplasts were used for microscopic detection of the GFP signal.
SI Fig. 3. PLGG1 shows an active and time-dependent uptake of [14C]Glycerate, that can be inhibited by increasing concentrations of unlabeled glycerate.

(A) Time-dependent uptake of radiolabelled [14C]Glycerate (50 µM external concentration) in the presence (filled circles) or absence of glycolate (open rectangles) as internal substrate (20 mM). In the presence of a suitable counter-exchange substrate inside the liposomes (e.g., glycolate) a Michaelis-Menten-type saturation kinetics was observed. In control liposomes preloaded with equal amounts of buffer or in liposomes that did not contain reconstituted PLGG1, only unspecific diffusion of the labeled substrate was observed.

(B) Concentration-dependent inhibition of active transport in the presence (black bars) or absence of glycolate (white bars) as internal substrate (20 mM). Active transport is inhibited with increasing external glycerate concentrations, while passive diffusion is independent on external glycerate concentration.

Error bars represent SD of three independent uptake experiments.

Recombinant At1g32080 protein was expressed in vitro using the 5PRIME wheat germ expression kit (5PRIME GmbH, Hamburg, Germany) as described previously (1) and reconstituted into liposomes using reconstitution and uptake buffers as described in Howitz et al. (2).
SI Fig. 4. Scheme of glycerate dependent oxygen evolution. Glycerate is transported by PLGG1 into the chloroplast in a proton-dependent manner, converted to 3-PGA under ATP consumption. 3-PGA is converted to TP under NADPH consumption. NADP is regenerated to NADPH at the photosynthetic electron transport chain under water consumption and oxygen evolution (red arrow). 3-PGA: 3-Phosphoglycerate; TP: triose phosphates.
**SI Fig. 5. Photorespiratory phenotype of plgg1-1 seedlings.** Seedlings from *Arabidopsis thaliana* Col-0 (WT), the photorespiratory mutant *shm1-1* (impaired in mitochondrial SHMT) and the T-DNA insertion line *plgg1-1*. Plants were grown on MS-media under elevated CO2 concentrations (3000 ppm) for ten days (left panel) or for four days and shifted to ambient CO2 (300 ppm) for six days (right panel).
SI Fig 6. Phylogenetic Tree of PLGG1. AtPLGG1 was used as the query to Blast Explorer (http://www.phylogeny.fr/version2_cgi/one_task.cgi?task_type=blast) and the sequences suggested by the program were manually curated and fed to an “à la carte” phylogeny analysis (http://www.phylogeny.fr/version2_cgi/phylogeny.cgi). Briefly, sequences were aligned with Muscle and all positions with gaps were removed. The phylogenetic tree was calculated with PhyML and bootstrapped (100 repetitions). For display, all branches with <30% branch support were collapsed.

The tree divides in two large branches, a bacterial and an eukaryotic branch. The sole cyanobacterium Acaryochloris spec is nested within the bacteria. In eukaryotes, the tree divides the fungi (more precisely the dikarya) from the archaeplastida. Within the archaeplastida, Cyanophora paradoxa separates from the red and green lineages which themselves form separate branches. In the green lineage, the green algae separate from the streptophyta in which the moss branches before the fern. Monocots and dicots form separate branches.

Since the fungi branch with the archaeplastida with a bootstrap support of 62%, the gene likely evolved before the archaeplastida and the fungi split and it was lost from the other branches of life. Had the gene been acquired by a lateral gene transfer, one would expect it to branch within the bacteria. In higher plants PLGG1 is nuclear encoded. For transport proteins of the chloroplasts, at least three different phylogenetic roots have been demonstrated: rooted in the Chlamydiae, rooted in the cyanobacteria and a host origin. PLGG1-like proteins do not appear in the Chlamydiae but in Cyanobacteria. The phylogenetic tree indicates that the gene is likely not acquired from the cyanobacteria since the single cyanobacterium PLGG1 is nested within the bacteria. Therefore it is more likely that the cyanobacterium acquired it by lateral gene transfer from a bacterium. Hence PLGG1 likely originated from the host.


**SI Table 1:** Primer used for PCR amplification. All sequences are displayed in 5'-3' orientation.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>CGTGTGTGTCTCCATACCCAT</td>
</tr>
<tr>
<td>P2</td>
<td>GTTTTGCCATAGGTCCGCTT</td>
</tr>
<tr>
<td>P3</td>
<td>GCCTGGACGCTTTGCTGCAACT</td>
</tr>
<tr>
<td>P4</td>
<td>CACCATGGCTACTCTTTTAGCCACTCC</td>
</tr>
<tr>
<td>P5</td>
<td>GACACCTGAAGCAGCCGG</td>
</tr>
<tr>
<td>P6</td>
<td>TCAATGTCCTGCGGAGCCTGTA</td>
</tr>
<tr>
<td>P7</td>
<td>TGAACCAATCGATCGGACCTG</td>
</tr>
<tr>
<td>P8</td>
<td>CACCGGAGCATGAGTACCCGTTACAATTG</td>
</tr>
<tr>
<td>P9</td>
<td>GACACCTGAAGCAGCCGG</td>
</tr>
<tr>
<td>P10</td>
<td>CACCGGAGCATGAGTACCCGTTACAATTG</td>
</tr>
<tr>
<td>P11</td>
<td>ACTAACGGCCTTCGAGGAT</td>
</tr>
<tr>
<td>P12</td>
<td>CACCACCTGAACCCGGACCTCC</td>
</tr>
<tr>
<td>P13</td>
<td>GACGGAGCTAGCAGGACTCTG</td>
</tr>
</tbody>
</table>

**SI Table 2:** Plants were germinated and grown under different light regimens and both the size difference compared to wildtype and the percentage of plants with lesions were scored.
While very young plants up to two weeks of age benefit from extended light periods which alleviate the phenotype, plants from week three on suffer from extended light periods. nd not determined

<table>
<thead>
<tr>
<th></th>
<th>8h L, 16h D</th>
<th>16h L, 8 hD</th>
<th>24h L</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>14 days</strong> size ratio WT/ mutant</td>
<td>1.58</td>
<td>1.48</td>
<td>1.36</td>
</tr>
<tr>
<td><strong>percentage of mutant plants with lesions</strong></td>
<td>50%</td>
<td>10%</td>
<td>nd</td>
</tr>
<tr>
<td><strong>size ratio WT/ mutant</strong></td>
<td>1.94</td>
<td>1.94</td>
<td>1.49</td>
</tr>
<tr>
<td><strong>21 days</strong> percentage of mutant plants with lesions</td>
<td>60%</td>
<td>90%</td>
<td>nd</td>
</tr>
</tbody>
</table>
Metabolite extraction.
For metabolite analysis, 100 mg (fresh weight) of powdered tissue was extracted with 700 µL of methanol for 15 minutes at 70°C. 700 µL of water and 375 µL of chloroform were added and the samples were incubated in a rotating shaker for 30 minutes at 4°C (1). The phases were separated and the methanol/water phase retained for further experiments. Amino acid contents were determined directly from the extract exactly as described in (2) except that the samples were extracted in 1:1 (v/v) methanol/water. Polar metabolites were analyzed as described by (1). Briefly, 50 ul of the extract was dried in a speed-vac and carbonyl moieties were protected by methoximation using a 20 mg/mL solution of methoxyamine hydrochloride in pyridine at 30°C for 90 min. Acidic protons were modified with N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) at 37°C for 30 min. Samples were injected at 230°C, separated on HP5 columns for 30 minutes and analyzed from m/z=50 to m/z=800 in total ion scans in an Agilent technologies 5973 Inert mass spectrometer. Metabolites were identified by comparison to the retention time and the fractionation pattern of standards and quantified with external standard curves of complex standards. The complex standard for LC/MS contained all amino acids. The complex standard for GC/MS contained glucose, fructose, glycolate, succinate, glycerate, fumarate, malate, 2-oxoglutarate, citrate, isocitrate, myo-inositol, sucrose, glycine, serine, pyruvate, maltose and ribitol at concentrations of 0.1, 0.5, 1, 5, 10, 50 and 100 µM. The complex standard was injected prior to the randomized samples, after injection of 21 samples and at the end of the measurement. Retention times and peak size for the complex standards remained stable.

Metabolite extraction optimized for photorespiratory metabolites.
The extraction procedure was conducted at 4°C. Fifty-milligram of powdered material was extracted with 600 µl of ice-cold N,N-dimethylformamide. Following the addition of 400 µl of water, the samples were shaken for 10 min and separated into phases by centrifugation for 8 min at 22,000xg. The upper aqueous phase was mixed with 600 µl of xylene for 10 min and centrifuged for 3 min. Upper organic phase was discarded and 300 µl of lower aqueous phase was dried for the analysis. The GC-MS analysis was conducted exactly as described by Lisec et al. (2006) (3). Chromatograms and mass spectra were evaluated by Chroma TOF® 1.6 (Leco, St Joseph, MI) and TagFinder 4.0 (4) for the quantification and annotation of the peaks. The 18O enrichment was calculated according to Berry et al. (5)


6.2 Second publication: *PLGG1*, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a new class of metabolite transporters.

Status: **Submitted** (August 2012)

Thea R. Pick, Andrea Bräutigam, Matthias A. Schulz, Toshihiro Obata, Alisdair R. Fernie, and Andreas P.M. Weber

Journal: "Proceedings of the National Academy of Sciences of the United States of America"

Impact factor: 9.681

1. Co-Author

*Own contribution: 60%*

- Visualization of the phenotypes
- GFP Localization
- Data analysis
- Oxygen evolution
- $^{18}$O$_2$ labeling and metabolite measurement and analysis
- Manuscript writing
Danksagungen

Zu guter Letzt möchte ich den Menschen danken, ohne die ich meine Promotion nicht hätte abschließen können und denen ich zu tiefstem Dank verpflichtet bin.

Vielen Dank...

... Prof. Dr. Andreas Weber für Möglichkeit an Ihrem Institut promovieren zu können und ein Teil des Weber-Universums zu sein. Sie haben mir ermöglicht an begeisternden und spannenden Forschungsthemen zu arbeiten, die keine „Bad Projects“ waren und im Rahmen der PROMICS und OPTIMAS meetings interessante Forscher zu treffen, an lehrreichen Diskussionen teilzunehmen und Amerika zu entdecken. In Ihrem Institut haben Sie mir ermöglicht mich als Wissenschaftler weiter zu entwickeln und weiterhin so viel Spaß an der Wissenschaft zu haben.

... Prof. Dr. Peter Westhoff für die Übernahme des Koreferats, hilfreiche Denkanstöße und Ratschläge als Committee member.

... Andrea, dass du mir das interessante Mep1 Thema überlassen hast, so viel Geduld mit meinen holprigen Schreib-Versuchen hattest und meine Doktorarbeit betreut hast.

... allen iGRAD plant Mitgliedern und ganz besonders Dr. Sigrun Wegener-Feldbrügge für anregende Diskussionen und sehr unterhaltsame Retreats.

... der PROMICS-Crew Marion, Christian und Daria für schöne gemeinsame Meetings und Konferenzen.

... Frau Nöcker für Ihre unkomplizierte Hilfe bei organisatorischen Angelegenheiten.

... den Gärtnern für die Aufzucht und Pflege der Pflanzen.

... Katja für die großartigen und unendlich lustigen Geschichten, einmaligen Parodien, unterhaltsamen Progress-Reports, die lustigen Kniffel-Spiele zwischendurch und vor allem die SOS E-Mail Hilfe, die du so oft leisten musstest und wahrscheinlich auch noch weiter musst.

... Katrin, meiner Lauf-Partnerin, für die schönen Gespräche. Das Joggen mit dir macht einfach riesigen Spaß und ich bin froh, dass wir uns als Lauf-Team gefunden haben.

... Daria, aka Chrobonaut, Chrobofant, Chrobosaurier oder einfach Chrobo, für ihre lustigen, unterhaltsamen Aktionen und Geschichten und die entspannte Zeit als Zimmer-Partnerin in Warnemünde.

... meiner Kaffee-Familie. Ihr seid einfach die Besten und die Pausen mit euch sind einfach wunderschön und keeelasse!
... Jan, Christian, Sarah, Ali, Nadine, Chris, Sam, Alex, Nicole, Fabio und allen anderen, die die Mittagspausen immer zu einem sehr unterhaltsamen Erlebnis gemacht haben und hoffentlich weiterhin machen werden.

... allen weiteren aktuellen und ehemaligen „Weber-Knechten“ für die tolle, unvergessliche Zeit, das angenehme Arbeitsklima und produktive Miteinander, die tollen Betriebsausflüge und die schönsten Geburtstags-Ständchen, die man bekommen kann.

... Lisa für ihre unkomplizierte, lockere, selbstlose und unendlich liebe Art. Ich hoffe, dass unsere Freundschaft noch lange so schön bleibt wie sie ist, egal wo auf der Welt wir uns grade rumtreiben sollten. Bleib so, wie du bist, denn so bist du einfach super!

... Dominik, für so vieles. Deinen Kaffee-Service, die für dich so selbstverständliche Hilfe (die überhaupt nicht selbstverständlich ist), das Rumhampeln beim Sporti, die Spieleabende, das Rumalbern im Labor, die gemeinsamen Skate-Nights, die leckeren Rezepte, die ständigen peinlichen Aktionen, die mich so herzhaft lachen lassen.... Du bist einfach immer so lustig und herzlich und gleichzeitig auch immer für ein gutes Gespräch zu haben. Ich hoffe, dass das noch lange so schön bleibt.

... Mareike, für deine sonnige und optimistische Art. Wenn ich an dich denke muss ich irgendwie immer lächeln, weil deine unerschütterlich gute Laune so ansteckend ist. Die Zeit mit dir ist immer so schön – egal auf welcher Welthalbkugel wir grade zusammen sind :-) und du bist meine absolute Postkarten-Königin! Ich hoffe, dass wir weiterhin so viel zusammen erleben und lachen wie bisher.

... Manuel, weil du einfach das Beste bist, was mir je passiert ist. Du hilfst mir so sehr in allen Lebenslagen ruhig zu bleiben und die richtigen Prioritäten zu setzen. Mit dir ist das Leben so relaxt und schön und aufregend und spannend zugleich. Und vor allem lustig... :-) mit dir werde ich immer ein Lächeln auf den Lippen haben...

... meiner Mutter Hanne, meinem Vater Rolf und meiner Schwester Lena, weil ihr mich zu dem Menschen gemacht habt, der ich jetzt bin. Mit euch konnte ich mich frei entfalten und wusste immer, dass ich einen Rückzugsort und sicheren Hafen habe in dem ich geliebt werde. Ich freue mich schon auf ganz ganz viele gemeinsame Familienurlaube in Frankreich und alles was wir in der Zukunft zusammen erleben werden.

Danksagungen