Mathematical modelling of metabolism applied to the evolution of photosynthesis

Inaugural-Dissertation

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

David Heckmann
aus Bergisch Gladbach

Düsseldorf, 24. Juli 2014

aus dem Institut für Informatik
der Heinrich-Heine Universität Düsseldorf

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

Referent: Prof. Dr. M.J. Lercher
Korreferent: Prof. Dr. A.P.M. Weber

Tag der mündlichen Prüfung: 17. Juni 2014

Erklärung

Ich versichere an Eides statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine Universität Düsseldorf" erstellt worden ist. Die Dissertation habe ich in dieser oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen oder erfolgreichen Promotionsversuche unternommen.

Düsseldorf, den 24. Juli 2014

David Heckmann

Contents

1 Abbreviations

3PGA	3-phosphoglycerate
A	net CO_2 assimilation rate
A_c	net CO_2 assimilation rate at saturating amounts of RuBP
A_j	CO_2 assimilation rate at limiting amounts of RuBP
BEP	Bambusoideae, Ehrhartoideae, Pooideae
C_m	partial pressure of CO_2 in the mesophyll chloroplasts
C_s	bundle sheath partial pressure of CO_2
E_{tot}	Rubisco active site concentration
FBA	flux balance analysis
iFBA	integrated flux balance analysis
FvCB	Farquhar, von Caemmerer, Berry
GDC	glycine decarboxylase
g_s	bundle sheath conductance for CO_2
K_C	affinity of Rubisco for CO_2
k_{ccat}	maximum turnover rate of Rubisco carboxylation
K_O	affinity of Rubisco for O_2
K_p	Michaelis-Menten constant of PEPC for bicarbonate
mya	million years ago

O_m	partial pressure of O_2 in the mesophyll chloroplasts
PACMAD	Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthoideae
PEPC	phosphoenolpyruvate carboxylase
PEP	phosphoenolpyruvate
PGly	2-phosphoglycolate
PPDK	pyruvate, phosphate dikinase
Rubisco	ribulose-1,5-bisphosphate carboxylase-oxygenase
RuBP	ribulose-1,5-bisphosphate
$S_{C/O}$	Rubisco specificity
V_{om}	rate of Rubisco oxygenation reaction in the mesophyll
V_{pmax}	PEPC-limited C_4 cycle activity
β	fraction of Rubisco expressed in the mesophyll
ξ	fraction of mesophyll-derived glycine decarboxylated in the mesophyll

2 List of Figures

3 Preface

This document was prepared as a cumulative thesis according to 'Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf vom 06.12.2013' §6 (4). Two manuscripts are presented along with an introduction that puts them into context with current literature. Additionally, an explanation about contributions and a discussion on future research directions is provided with each manuscript.

Manuscript 1 describes the application of a mathematical model of leaf metabolism to the evolution of C_4 photosynthesis and the comparison of model predictions to experimental data. It was published as Heckmann et al., 2013, Cell 153, 1579–1588.

Manuscript 2 follows up on this work and uses a mathematical model of whole leaf metabolism to provide a mechanistic explanation for the phylogenetic co-occurrence of C_2 and C_4 photosynthesis in the light of recent data from the genus *Flaveria*. It was published as Mallmann, Heckmann et al., 2014, eLife, 10.7554/eLife.02478.

4 Summary

The majority of plant species belong to the paraphyletic group of C_3 plants, which fix atmospheric CO_2 into three-carbon compounds. C_4 photosynthesis expands the ancestral photosynthetic mode of C_3 photosynthesis, allowing for improved CO_2 assimilation and nitrogen and water-use efficiency. These properties render it a promising trait to be engineered into C_3 crops like rice, where a yield increase of up to 50 % is projected (Mitchell and Sheehy, 2006). Hindering this endeavour is the fact that C_4 photosynthesis is a multigenic complex trait that involves changes in gene expression, enzyme kinetics, regulation, and leaf anatomy (Gowik et al., 2011). Despite this complexity, C_4 photosynthesis has multiple independent evolutionary origins (Sage et al., 2011). Understanding why and how a trait of such complexity has evolved in a variety of genetic backgrounds can support engineering and directed breeding efforts (Sage and Sage, 2007; Slewinski, 2013). While the evolvability of a genetic state is determined by its underlying fitness landscape (Orr, 2005), the experimental evaluation of realistic fitness landscapes is infeasible and the biological systems that allow for *in silico* prediction of fitness landscapes are sparse (Orr, 2005).

The work presented in *Manuscript 1* shows that mathematical models of photosynthesis allow the investigation of the phenotypic fitness landscape underlying the complex trait C_4 photosynthesis. This analysis is feasible because a clear correlate of fitness, the net CO_2 assimilation rate, can be predicted from parameters that are known to vary between C_3 and C_4 species. The resulting landscape shows that C_4 photosynthesis is evolutionarily accessible through individually adaptive steps from any intermediate state, even though it exhibits extensive sign epistasis. This finding provides an explanation for the more than 60 independent origins of C_4 photosynthesis. Evolutionary trajectories on the landscape were calculated through stochastic simulation. These clustered significantly, suggesting that C_4 evolution is repeatable in its details. Furthermore, patterns described in earlier conceptual models (Gowik et al., 2011; Sage,

2013) were rediscovered in the trajectories, while modularity of traits along the simulated paths extends those previous views. Experimental evolution studies repeatedly found signs of diminishing returns (Lenski and Travisano, 1994; Elena and Lenski, 2003); the predicted trajectories did not show this decelerated evolution.

In order to validate the model, a diverse dataset of biochemical, kinetic, immuno-labelling, and expression data on C_3–C_4 intermediate species from the genera *Flaveria*, *Moricandia*, and *Panicum* was compared to the predicted trajectories. These data points fall on predicted trajectories, indicating that (1) the model agrees with data and (2) that the species represent true intermediates that are expected to further evolve towards C_4 photosynthesis in a suitable environment. The model presented in *Manuscript 1* predicts the establishment of a photorespiratory CO_2 pump ('C_2 photosynthesis') to occur during the first steps in C_4 evolution. This result is sensitive to the assumption of a high mutational probability of the photorespiratory CO_2 pump (*Manuscript 1* Fig. S3.B), and a mechanistic explanation of the phylogenetic proximity of C_2 and C_4 species is still lacking.

In *Manuscript 2*, metabolic *in silico* modelling of C_2 and C_4 photosynthesis is combined with transcript and protein abundance data from closely related C_3, C_3–C_4, and C_4 *Flaveria* species to understand the role of C_2 photosynthesis in C_4 evolution. The coupling of a kinetic model of C_2 photosynthesis with a stoichiometric genome-scale metabolic model revealed the most parsimonious shuttles that can be employed to balance the nitrogen budget between mesophyll and bundle sheath tissue. The predicted shuttles utilize components associated with C_4 photosynthesis in *Flaveria*, indicating a case of pre-adaptation. In cases where a low activity of the C_4 cycle occurs in parallel with C_2 photosynthesis, the two cycles are predicted to share shuttled metabolites in order to minimise plasmodesmatal flux between mesophyll and bundle sheath. These findings are in agreement with transcript and protein data from *Flaveria*

and enhance our understanding of the role of pre-adaptation in the evolution of this complex trait.

References

Elena, S. F. and Lenski, R. E. (2003). Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4(6):457–469.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? *The Plant Cell Online*, 23(6):2087–2105.

Lenski, R. E. and Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences*, 91(15):6808–6814.

Mitchell, P. L. and Sheehy, J. E. (2006). Supercharging rice photosynthesis to increase yield. *New Phytologist*, 171(4):688–693.

Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, 6(2):119–127.

Sage, R. and Sage, T. (2007). Learning from nature to develop strategies for the directed evolution of C_4 rice. *Charting new pathways to C_4 rice. World Scientific Publishing*, pages 195–216.

Sage, R. F. (2013). Photorespiratory compensation: A driver for biological diversity. *Plant Biology*, 15(4):624–638.

Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C_4 plant lineages of planet Earth. *Journal of Experimental Botany*, 62(9):3155–3169.

Slewinski, T. L. (2013). Using evolution as a guide to engineer Kranz-type C_4 photosynthesis. *Frontiers in Plant Science*, 4(212).

5 Zusammenfassung

Die meisten Pflanzenspezies gehören zur paraphyletischen Gruppe der C_3-Pflanzen, welche CO_2 primär in C_3-Körpern fixieren. C_4-Photosynthese stellt eine Erweiterung zur ursprünglichen C_3-Photosynthese dar und erlaubt erhöhte CO_2-Assimilationsraten, sowie verbesserte Stickstoff- und Wassernutzungseffizienz. Daher ist die C_4-Photosynthese ein vielversprechendes Ziel für die Verbesserung von C_3-Nutzpflanzen wie Reis, wobei Ertragssteigerungen von bis zu 50 % erwartet werden (Mitchell and Sheehy, 2006). Dass die C_4-Photosynthese Veränderungen in der Genexpression, Enzymkinetik, metabolischen Regulation und Blattanatomie erfordert (Gowik et al., 2011), erschwert dieses Vorhaben. Trotz dieser Komplexität hat sich die Evolution der C_4-Photosynthese mehrmals unabhängig voneinander ereignet (Sage et al., 2011). Ein besseres Verständnis davon, wie und warum diese komplexe Ausprägung von unterschiedlichen genetischen Ausgangspunkten aus evolviert ist, kann die Verbesserung der C_3-Nutzpflanzen maßgeblich unterstützen (Sage and Sage, 2007; Slewinski, 2013). Die Evolvierbarkeit eines Genotyps wird durch die zugrundeliegende Fitness-Landschaft (*fitness landscape*) bestimmt (Orr, 2005). Allerdings ist die experimentelle Ermittlung von realistischen *fitness landscapes* nicht möglich, und Beispiele für erfolgreiche *in silico* Simulation sind selten (Orr, 2005).

Die in *Manuskript 1* präsentierten Ergebnisse zeigen, dass die phänotypische *fitness landscape*, die der C_4-Evolution zugrunde liegt, mithilfe mathematischer Modelle der Photosynthese untersucht werden kann. Diese Analyse wird dadurch ermöglicht, dass ein Fitness-relevanter Parameter, die CO_2-Assimilationsrate, in Abhängigkeit von Parametern vorhergesagt werden kann, die sich während der C_4-Evolution verändern.

Die simulierte *landscape* zeigt, dass die C_4-Photosynthese durch adaptive Schritte von jedem Zwischenzustand aus evolvieren kann, obwohl an vielen Stellen der Landschaft Vorzeichen-Epistasis (*sign epistasis*) auftritt. Dieses Ergebnis erklärt, warum die C_4-Photosynthese in mehr als

60 unabhängigen Fällen evolvieren konnte. Weiterhin wurden mithilfe stochastischer Simulation evolutionäre Pfade durch die Landschaft berechnet. Diese Pfade wiesen eine statistisch signifikante Ähnlichkeit zueinander auf, was auf eine detaillierte Wiederholbarkeit der C_4-Evolution schließen lässt. Des Weiteren zeigten diese Pfade die Muster, die zuvor in konzeptionellen Modellen angenommen wurden (Gowik et al., 2011; Sage, 2013), wobei die gefundene Modularität der Pfade diese Modelle erweitert. Arbeiten zur experimentelle Evolution haben wiederholt das Auftreten von *diminishing returns* gefunden (Lenski and Travisano, 1994; Elena and Lenski, 2003); die vorhergesagten Pfade der C_4 Evolution weisen keine sich verlangsamende Evolution auf.

Um die Modellvorhersagen zu prüfen, wurden die vorhergesagten Pfade mit einem diversen Datensatz aus biochemischen, kinetischen, Immunmarkierungs- und Expressionsdaten von C_3–C_4 Zwischenspezies aus den Gattungen *Flaveria*, *Moricandia* und *Panicum* verglichen. Diese Datenpunkte liegen auf den vorhergesagten Pfaden, was darauf hindeutet, dass (1.) die Daten mit dem Modell vereinbar sind, und dass (2.) die Spezies tatsächlich Zwischenspezies darstellen, die in einer passenden Umgebung C_4-Photosynthese evolvieren würden.

Das in *Manuskript 1* präsentierte Modell sagt die Entstehung einer photorespiratorischen CO_2-Pumpe (auch „C_2-Photosynthese“ genannt), als frühen Schritt der C_4-Evolution voraus. Dieses Ergebnis hängt von der Annahme ab, dass die photorespiratorische CO_2-Pumpe eine hohe Mutationswahrscheinlichkeit aufweist; eine mechanistische Erklärung für die phylogenetische Nachbarschaft von C_2- und C_4-Spezies konnte aber noch nicht geliefert werden.

In *Manuskript 2* wird metabolische *in silico* Modellierung der C_2- und C_4-Photosynthese mit Daten zu Transkript- und Proteinabundanzen aus nah verwandten C_3-, C_3–C_4- und C_4-Spezies kombiniert, um die Rolle der C_2-Photosynthese in der C_4-Evolution zu verstehen. Die Kopplung eines kinetischen Modells der C_2-Photosynthese mit einem genomweiten stöchiometrischen Modell ergab die einfachsten Stoffwechsel-

wege, durch die die Pflanze den Stickstoffhaushalt zwischen Mesophyll und Bündelscheide ausgleichen kann. Die vorhergesagten Stoffwechselwege benutzen Bestandteile der C_4 Photosynthese, was auf Präadaption durch den Stickstoffstoffwechsel hindeutet. Wenn zusätzlich eine geringe Aktivität des C_4-Zyklus simuliert wird, zeigt sich eine Überlappung der Stoffwechselwege, die der Minimierung von Fluss über die Plasmodesmata dient.

Diese Ergebnisse werden durch die Protein- und Transkriptdaten aus *Flaveria* bestätigt und erweitern unser Verständnis der Rolle von Präadaption für die Evolution des komplexen Merkmals C_4 Photosynthese.

Literatur

Elena, S. F. and Lenski, R. E. (2003). Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4(6):457–469.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? *The Plant Cell Online*, 23(6):2087–2105.

Lenski, R. E. and Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences*, 91(15):6808–6814.

Mitchell, P. L. and Sheehy, J. E. (2006). Supercharging rice photosynthesis to increase yield. *New Phytologist*, 171(4):688–693.

Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, 6(2):119–127.

Sage, R. and Sage, T. (2007). Learning from nature to develop strategies for the directed evolution of C_4 rice. *Charting new pathways to C_4 rice. World Scientific Publishing*, pages 195–216.

Sage, R. F. (2013). Photorespiratory compensation: A driver for biological diversity. *Plant Biology*, 15(4):624–638.

Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C_4 plant lineages of planet Earth. *Journal of Experimental Botany*, 62(9):3155–3169.

Slewinski, T. L. (2013). Using evolution as a guide to engineer Kranz-type C_4 photosynthesis. *Frontiers in Plant Science*, 4(212).

6 Introduction

6.1 Evolution of complex traits and underlying fitness landscapes

Organs and tissues of multicellular organisms exhibit an astonishing complexity, caused by the interplay of multiple genes. Since each part of such complex systems appears to be essential for the adaptive function of the system as a whole, the question arises of how these complex systems were able to appear through adaptive evolution. For instance, Charles Darwin discussed the counter-intuitive way by which 'an optic nerve merely coated with pigment' can evolve into the eye of an eagle. He stated that this seeming paradox can be explained '[. . .] if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist [. . .]' (Darwin, 1859).

The existence of such evolutionary trajectories depends on the availability of adaptive mutations and the way emerging mutations interact to influence fitness through epistatic effects (Wagner and Altenberg, 1996). These properties of a biological system can be modelled in the form of a genetic fitness landscape (Orr, 2005). The high-dimensional space of possible mutations is mapped to respective fitness values, thus resulting in a landscape that constrains the possible evolutionary trajectories that connect two genotypes. Whether a genotype is accessible from a second one depends on the topographic structure of the fitness landscape spanned by the mutations separating the two. Smooth, single-peaked ('Mount Fuji-type') landscapes allow adaptive changes from each state in the landscape, while rugged topographies imply the existence of local optima that render additional changes deleterious and result in 'dead end'-trajectories (Franke et al., 2011). Rugged landscapes were shown to exhibit reciprocal sign epistasis, a form of epistatic interaction between two loci where individual mutations change fitness in a different direction than the co-occurrence of both genetic changes (Poelwijk et al., 2011).

Experimental mapping of fitness landscapes can be achieved by selecting a small set of n mutations that are known to be relevant for fitness *a priori*, and measuring the fitness of all $2n$ combinations of mutated and wild type loci. This approach was applied to single-gene fitness landscapes of bacterial β-lactamase and isopropylmalate dehydrogenase (Lunzer et al., 2005; Weinreich et al., 2006), as well as a landscape spanned by five genes that were found during an evolutionary experiment (Khan et al., 2011). All of these systems exhibited a smooth, Mount Fuji-type fitness landscape.

Another way of experimentally exploring subsets of genome-scale landscapes are long-term evolutionary experiments. Replicated populations of microorganisms are grown under controlled conditions for tens to tens of thousands of generations in order to study evolution in the laboratory (Elena and Lenski, 2003; Hindré et al., 2012). While these experiments potentially do not restrict the fitness landscape to a few loci, they are limited to exploring a small number of evolutionary trajectories and thus only a tiny fraction of the genetic landscape. A unifying property of the trajectories found in long-term evolutionary experiments is the occurrence of diminishing returns (Lenski and Travisano, 1994; Elena and Lenski, 2003): the fitness gain of novel mutations decreases along the trajectory, indicating the presence of epistatic interactions (Chou et al., 2011). Another interesting result of these experiments is the high level of parallel evolution, both on the level of phenotype and genotype (Lenski and Travisano, 1994; Bull et al., 1997; Barrick et al., 2009).

Because of the high dimensionality and the combinatorial explosion of genetic states, the experimental exploration of fitness landscapes is limited to minute subsets of biological landscapes. This confines the exploration of complex fitness landscapes to the evaluation of theoretical models. These are in many cases of purely statistical character (e.g.: Kauffman and Levin, 1987; Franke et al., 2011), while some studies included the modelling of biological systems: Fontana and Schuster (1998) simulated RNAs that evolve towards a defined secondary structure and

found that the evolutionary trajectories exhibited sudden increases in 'fitness'. Kauffman and Levin (1987) simulated evolutionary trajectories of model gene-regulatory networks and showed that they agree with their theory of long-jump adaptation.

While these examples capture the underlying mechanisms of biological systems and serve to understand the general rules of adaptive walks on fitness landscapes, they define fitness in relation to an arbitrary state that does not correspond to a real biological function.

6.1.1 Contribution of *Manuscript 1* to understanding the evolution of complex traits and underlying fitness landscapes

The research presented in *Manuscript 1* provides a novel approach to the modelling of fitness landscapes. The net CO_2 assimilation rate is used as a proxy for plant fitness and allows the investigation of the phenotypic fitness landscape of a specific complex trait.

In the light of complex trait evolution, the investigation of the fitness landscape structure shows how the evolution of a complex trait can be realised in Darwin's framework of small adaptive steps. Like in the experimental genetic landscapes mentioned in Section 6.1 (Lunzer et al., 2005; Weinreich et al., 2006; Khan et al., 2011), a smooth, Mount Fuji-type landscape was found to govern the evolution of C_4 photosynthesis. Only about 0.04 % of states in the landscape exhibited reciprocal sign epistasis for neighbouring phenotypic changes, providing a partial explanation for this smooth topography (Poelwijk et al., 2011).

Surprisingly, no sign of diminishing returns and decelerating evolution was found when simulating evolutionary trajectories on the C_3–C_4 fitness landscape. On the contrary, initial steps were found to yield the smallest absolute gains in fitness, while subsequent steps exhibited constant, high gains.

Evolutionary trajectories were simulated through stochastic simulations and exhibited significant clustering. This shows that the landscape

structure results in repeatability of evolution, resembling the parallel evolution found in long-term evolutionary experiments (Lenski and Travisano, 1994; Bull et al., 1997; Barrick et al., 2009).

6.2 C_4 photosynthesis

6.2.1 Photorespiration

In the mesophyll cells of C_3 plants, the primary fixation of atmospheric CO_2 into three-carbon compounds is catalysed by the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco). After Rubisco is activated by CO_2 and Mg^{2+}, it catalyses the carboxylation of ribulose-1,5-bisphosphate (RuBP) into two molecules of 3-phosphoglycerate (3PGA). Owing to the chemical similarity of CO_2 and O_2, a fraction of Rubisco-bound RuBP is oxygenated into one molecule each of 2-phosphoglycolate (PGly) and 3PGA. 3PGA can be reduced to triose-phosphates in the Calvin-Benson Cycle, while PGly needs to be converted to 3PGA in the photorespiratory cycle. This biochemical pathway is situated in the plastids, mitochondria, and peroxisomes, and serves important roles in plant metabolism in addition to the recycling of 3PGA (e.g.: Keys et al., 1978; Ros et al., 2013). Apart from a significant energy usage, photorespiration involves the release of CO_2 by the glycine decarboxylase (GDC) complex, decreasing photosynthetic capacity by up to 30 % (Ehleringer et al., 1991; Sage et al., 2012; Sage, 2013).

6.2.2 Augmenting C_3 photosynthesis with an efficient carbon concentrating mechanism

C_4 photosynthesis is an expansion of the ancestral C_3 mechanism and avoids photorespiration by concentrating CO_2 around Rubisco. In order to be operational, this energy-dependent CO_2 pump requires changes in gene expression, leaf anatomy, and biochemical features. Leaves of C_4 species exhibit a higher vein density than those of C_3 plants, with prominent bundle sheath cells surrounding the veins; a structural feature termed Kranz anatomy. Rubisco is located in these cells, which possess thick cell walls that are assumed to reduce CO_2 leakage (Kiirats et al., 2002). Phosphoenolpyruvate carboxylase (PEPC), an enzyme lacking an oxygenation function, catalyses the primary fixation of CO_2 as bicarbonate

in the mesophyll. Resulting C_4 acids enter the bundle sheath cells and are decarboxylated, releasing CO_2 in proximity to Rubisco (Fig. 1). The biochemical implementation of the CO_2 pump differs between species: different decarboxylating enzymes are used and different metabolites are shuttled (Furbank, 2011; Pick et al., 2011). The metabolic subtypes of C_4 photosynthesis are classified by the decarboxylating enzyme that is dominant in a species: NADP-malic enzyme, NAD-malic enzyme, and PEP carboxykinase. The result is in all cases a high ratio of CO_2 to O_2 in the bundle sheath cells that suppresses the oxygenation function of Rubisco and thus efficiently decreases photorespiration.

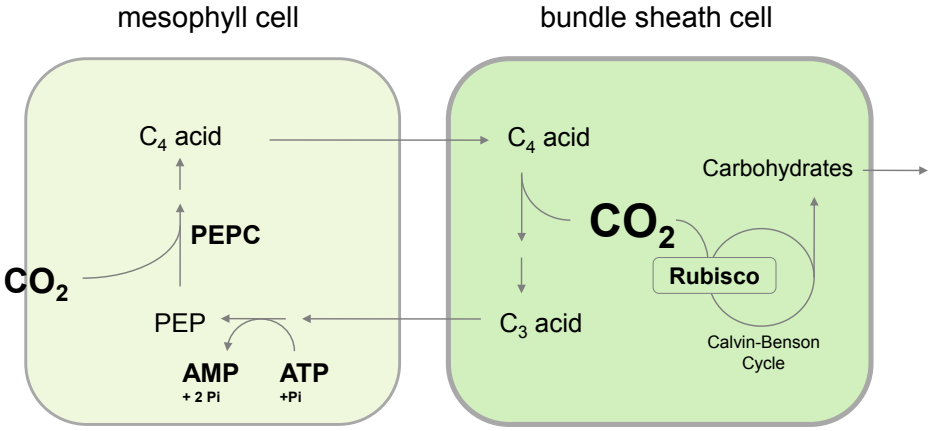

Figure 1: The mechanism of C_4 photosynthesis: Primary fixation of CO_2 in the form of bicarbonate is catalysed by phosphoenolpyruvate carboxylase (PEPC) in the cytosol of mesophyll cells. Resulting C_4 acids diffuse into the bundle sheath, where ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco) and the Calvin-Benson cycle is located. Decarboxylation of C_4 acids creates an environment of high CO_2 concentration that suppresses the oxygenation reaction of Rubisco. C_3 acids diffuse back to the mesophyll and are used to regenerate PEP through pyruvate, phosphate dikinase (PPDK). Regeneration of PEP requires hydrolysis of ATP to AMP; a reaction that serves as the energy source for the C_4 cycle. The C_4 and C_3 acids and the decarboxylating enzymes that are utilised in the pathway differ between C_4 species.

6.2.3 Advantages of C_4 photosynthesis

Although C_4 plants represent only 3 % of all land plant species, they contribute about a quarter of primary biomass production on land (Still et al., 2003; Sage, 2004). High temperatures and low atmospheric CO_2 promote photorespiration and consistently C_4 plants are most abundant in tropical environments (Brooks and Farquhar, 1985; Sage et al., 2012). Under such conditions, the ability to suppress photorespiration results in higher rates of photosynthesis and increased photosynthetic water and nitrogen-use efficiency (Brown, 1978; von Caemmerer, 2000; Taylor et al., 2010; Vogan and Sage, 2011), while the quantum yield is comparable to that of C_3 plants (Ehleringer and Björkman, 1977). Increased water and nitrogen-use efficiency are both results of a high photosynthetic capacity that allows fixation of CO_2 with less Rubisco and lower stomatal conductivity when compared to C_3 plants. The most productive crop species are C_4 plants. C_4 metabolism allows improved biomass production rates also in temperate regions (Beale and Long, 1995), making the engineering of the C_4 trait into major C_3 crops a highly promising route towards meeting the growing demands on food production (Hibberd et al., 2008; von Caemmerer et al., 2012). Since major details about the genetic blueprint of C_4 photosynthesis are still lacking, combining directed breeding with genetic engineering approaches is a promising option to achieving this goal (Sage and Sage, 2007). Once the evolutionary dynamics that produced C_4 photosynthesis in nature are understood, they can prove a valuable guide in this endeavour (Sage and Sage, 2007; Slewinski, 2013).

6.3 The evolution of C_4 photosynthesis

6.3.1 C_4 photosynthesis: A textbook example of convergent evolution

Despite the complexity of the trait, C_4 photosynthesis has evolved independently in more than 60 angiosperm lineages (Sage et al., 2011). This high level of polyphyly constitutes a striking example of convergent evolution and suggests a combination of a low evolutionary trough towards expression of the trait and a high selection pressure. The enzymes participating in the C_4 cycle can be found in all C_3 plants, and many C_3 species possess C_4-like anatomical features, providing a partial explanation for the observed polyphyly (Kinsman and Pyke, 1998; Aubry et al., 2011).

Evolutionary origins of C_4 photosynthesis tend to cluster phylogenetically. For example, the Grasses contain multiple C_4 origins in the Panicoideae, Arundinoideae, Chloridoideae, Micraroideae, Aristidoideae, Danthoideae (PACMAD) clade, while the Bambusoideae, Ehrhartoideae, Pooideae (BEP) clade lacks C_4 species (Grass Phylogeny Working Group II, 2012). One explanation for the observed phylogenetic clustering is the apparent necessity of anatomical preconditioning events that need to precede the evolution of C_4 photosynthesis (Christin et al., 2013). A small interveinal distance in combination with larger bundle sheath cells appears to increase the probability of C_4 evolution (Christin et al., 2013). The adaptive advantage of such an anatomy is probably unrelated to carbon concentrating mechanisms; instead structural support, improved leaf hydraulics, and cavitation repair are discussed as functions of the C_3 bundle sheath (Sage, 2004; Griffiths et al., 2013).

6.3.2 Environmental factors promoting C_4 evolution

In addition to anatomical drivers, environmental factors will have a strong influence on the evolvability of the C_4 syndrome. During the Oligocene epoch, about 34 to 23 mya, atmospheric levels of CO_2 dropped

from about 1500 ppm to about 250 ppm (Pagani et al., 2005). The resulting low ratios of atmospheric CO_2 to O_2 were probably associated with a major increase in photorespiration, and these conditions rendered carbon concentrating mechanisms an adaptive trait (Ehleringer et al., 1991; Sage, 2004). In agreement with this notion, molecular dating of C_4 lineages in the grasses suggests that early origins of C_4 photosynthesis coincided with the drop of atmospheric CO_2 concentration during the Oligocene (Christin et al., 2008). In addition to the CO_2/O_2 ratio, paleoclimate data and the habitats of recent C_4 species indicate heat, aridity, and high light as important abiotic factors promoting C_4 evolution (Ehleringer et al., 1991; Edwards et al., 2010; Sage et al., 2012).

6.3.3 Conceptual models of C_4 evolution are based on intermediate species

Genera that comprise recent C_3, C_4, and intermediate C_3–C_4 species, like *Flaveria* and *Moricandia*, are studied to derive hypotheses on the evolutionary path to C_4 photosynthesis. Sage et al. (2012) give a conceptual model for C_4 photosynthesis evolution:

A general pre-adaptation in the form of gene duplications is assumed to be the first step (Monson, 2003; Wang et al., 2009). This expanded genetic repertoire allows neofunctionalisation of genes and the distinct expression patterns that are found in C_4 plants. Additionally, a high vein density with short intercellular diffusion paths is assumed to be a prerequisite for efficient carbon-concentrating mechanisms (McKown and Dengler, 2009).

Although bundle sheath cells occur as a distinct cell type in C_3 plants (Kinsman and Pyke, 1998), a high metabolic capacity of the C_3 bundle sheath is required for an efficient CO_2 pump. Large bundle sheath cells with high abundances of plastids and mitochondria, features termed 'proto-Kranz' anatomy in C_3 plants, meet this demand (Muhaidat et al., 2011; Christin et al., 2013). When the proto-Kranz anatomy is in place, it allows the establishment of a photorespiratory CO_2 pump, termed

'C_2 photosynthesis' (Sage et al., 2012). This carbon-scavenging mechanism was found in C_3–C_4 intermediate species and involves a loss of expression of the GDC complex in the mesophyll tissue (e.g.: Rawsthorne et al., 1988b). Consequently, photorespiratory glycine is decarboxylated in the bundle sheath and can result in higher photosynthetic assimilation rates than those of C_3 plants when environmental conditions cause high rates of photorespiration (von Caemmerer, 1989). This preliminary CO_2 pump then allows increased expression of PEPC to improve the existing pump. PEP regeneration is at this point possibly achieved through 3PGA mutase and enolase activity (Monson and Moore, 1989), and experiments in tobacco indicate that the expression of NADP-malic enzyme is already high in the surrounding vascular tissue of stems and petioles of C_3 plants (Hibberd and Quick, 2002).

According to Sage et al. (2012), the subsequent step involves a shift of the Calvin-Benson cycle and the corresponding metabolism to the bundle sheath, an increased activity of the C_4 cycle, and higher expression of carbonic anhydrase in the mesophyll cells. Final steps of C_4 evolution include fine-tuning of metabolism in form of regulatory (e.g.: Engelmann et al., 2003) and kinetic (e.g.: Sage, 2002) enzyme properties.

In a recent study, Williams et al. (2013) used a dataset on binary features of lineages that contain C_3, C_3–C_4, and C_4 species to infer evolutionary trajectories through a Bayesian approach. While these trajectories generally mirror the conceptual model presented above, they exhibit a bimodal distribution of the bundle sheath specific expression of GDC, indicating that this shift of GDC expression is not necessarily an early event in all cases of C_4 evolution.

6.3.4 Contribution of *Manuscript 1* to understanding the evolution of C_4 photosynthesis

In *Manuscript 1*, the evolution of C_4 photosynthesis from an ancestral C_3 species is simulated *in silico*. Based on the assumption that gains in photosynthetic capacity are proportional to a gain in fitness, a biochem-

ical model of C_3–C_4 intermediate photosynthesis is used to investigate the shape of the fitness landscape that is framed by the limiting cases of C_3 and C_4 photosynthesis. The analysis in *Manuscript 1* shows that this landscape exhibits a smooth shape with the C_4 state as its single peak (Fig. 2). Conceptual models of C_4 evolution rely on the assumption that intermediate species do not represent dead ends to C_4 evolution (Monson and Moore, 1989; Sage, 2004; Gowik and Westhoff, 2011; Sage et al., 2012). In the context of fitness landscapes, this means that species are not found on intermediate peaks in the landscape, and the work in *Manuscript 1* supports this notion.

Furthermore, the shape of the fitness landscape provides an explanation for the multiple independent evolutionary events resulting in C_4 photosynthesis (Sage et al., 2011): Because of the smooth, single-peaked shape, the C_4 state is evolvable from any intermediate point in the landscape.

This finding raises the question of why we do not observe even more evolutionary origins of C_4 photosynthesis, especially in arid tropical regions. The answer lies in the fact that the model in *Manuscript 1* makes assumptions about the anatomy and genetic constitution of the ancestral C_3 plant. As discussed above, the clustering of C_4 origins on the phylogenetic tree of the angiosperms can be explained by the occurrence of a potentiating 'proto-Kranz' anatomy in some lineages, which allows efficient diffusion of metabolites and sufficient metabolic capacity of the bundle sheath (Muhaidat et al., 2011; Sage et al., 2012; Christin et al., 2013). The ancestral C_3 plant in the model is assumed to possess such an anatomy in the form of an adaptive trait that is unrelated to photosynthesis (Griffiths et al., 2013).

The genetic basis for the changes of phenotypic model parameters along evolutionary paths is mostly unknown. It seems obvious though that the neofunctionalisation of C_3 enzymes for use in the C_4 pathway required preconditioning in the form of gene duplications to avoid interference with the ancestral metabolic functions (see Section 6.3.3). The

simulations described in *Manuscript 1* thus implicitly assume preconditioning through gene duplication, and a lack thereof might pose a genetic constraint to the phenotypic evolution and render parts of the fitness landscape inaccessible.

Stochastic simulations in *Manuscript 1* revealed evolutionary trajectories that overlap significantly with experimental data points from C_3–C_4 intermediate species, indicating that these species represent true intermediates that are expected to further evolve towards C_4 photosynthesis (Fig. 2). The predicted trajectories cluster significantly with a mean path, indicating that C_4 evolution is repeatable and the proposed model is robust. The common pattern of phenotypic changes in these evolutionary paths mirrors the order of evolutionary events in the conceptual model described in Section 6.3.3. Moreover, Sage's model is extended by the observed modularity of traits along evolutionary trajectories. Given the similar fitness gains along the greedy evolutionary trajectories, the observed pattern is probably caused by epistasis. The evolutionary trajectories inferred by Williams et al. (2013) from a binary dataset generally agree with those predicted in *Manuscript 1*. Interestingly, Williams et al. (2013) found a bimodal distribution of the bundle sheath specific expression of GDC along the trajectories, which is also apparent in the trajectories in *Manuscript 1* (see *Manuscript 1* Fig. S3).

6.3.5 C_2 photosynthesis: An evolutionary link between C_3 and C_4 photosynthesis

Species with gas exchange characteristics intermediate between C_3 and C_4 photosynthesis were identified in about 20 lineages of vascular plants (Sage et al., 2011). The intermediate CO_2 compensation points and O_2 sensitivities of these intermediate species are, in addition to incomplete C_4 cycles in some species, a result of efficient recycling of photorespiratory CO_2 in the bundle sheath during C_2 photosynthesis. This mechanism is based on a confinement of GDC to mostly centripetally positioned mitochondria in the bundle sheath cells (Hylton et al., 1988; Morgan

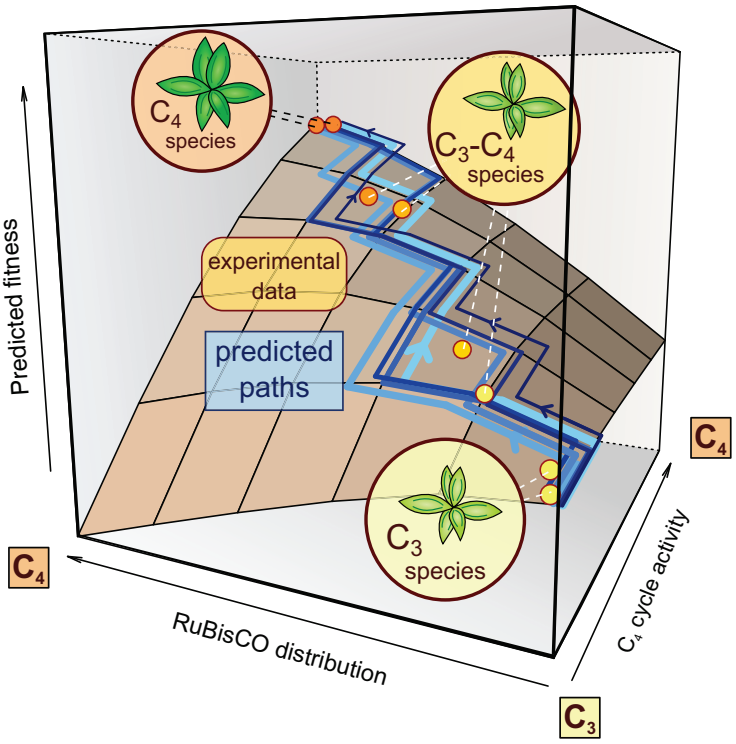

Figure 2: Graphical abstract for *Manuscript 1*: net CO_2 assimilation rate at saturating amounts of RuBP (A_c) is shown as a function of fraction of Rubisco expressed in the mesophyll (β) and PEPC-limited C_4 cycle activity (V_{pmax}) (brown surface). Seven trajectories resulting from Monte Carlo simulations are plotted in blue. Experimental data from *Flaveria* is shown as orange points. Note the occurence of sign-epistasis for β.

et al., 1993; Marshall et al., 2007; Schulze et al., 2013, also see Section 6.3.3). Such differential expression results in an exclusive decarboxylation of photorespiratory glycine in the bundle sheath and thus a carbon-scavenging mechanism referred to as 'C_2 photosynthesis' (Sage et al., 2012). The molecular basis for this evolutionary change in physiology was recently elucidated for *Flaveria* (Schulze et al., 2013). Two isoforms of the P-subunit of the GDC complex, one that is expressed throughout the chlorenchyma and one that is bundle sheath specific, can be found in C_3 *Flaveria* species (Schulze et al., 2013). Decreased expression of

the chlorenchyma form results in bundle sheath specificity of the GDC P-subunit in intermediate species (Schulze et al., 2013).

The current model for C_2 photosynthesis assumes diffusion of glycine from the mesophyll to the bundle sheath mitochondria, where it is decarboxylated and the resulting serine can diffuse back (Rawsthorne et al., 1988b; von Caemmerer, 2000; Sage et al., 2012). This model results from reports about the distribution of photorespiratory enzymes (Rawsthorne et al., 1988a), elevated concentrations of glycine and serine in *Moricandia* (Rawsthorne and Hylton, 1991), and the incorporation of ^{14}C label into glycine and serine in pulse chase experiments with *Flaveria* (Monson et al., 1986).

About half of the lineages operating C_2 photosynthesis possess sister clades that utilize full C_4 photosynthesis, indicating that C_2 species are ancestral to the C_4 forms (Sage et al., 2011). This finding led to the hypothesis that C_2 photosynthesis is an intermediate step in C_4 evolution (see Section 6.3.3 and Sage et al., 2012), but the mechanism of how the photorespiratory pump increases the probability of C_4 evolution is still unclear. Sage (2004) hypothesizes that an initial increase in PEPC activity might have served to scavenge CO_2 that leaks from the bundle sheath as a result of decarboxylation by GDC. In a more recent review, Sage et al. (2012) discuss a possible selection pressure towards Kranz anatomy, caused by the operation of C_2 photosynthesis.

6.3.6 Contribution of *Manuscript 2* to understanding the mechanistic link between C_2 and C_4 photosynthesis evolution

In *Manuscript 2*, a novel explanation for the phylogenetic association of C_2 and C_4 plants is provided. The current model of C_2 photosynthesis implies a net transport of nitrogen to the bundle sheath at a rate of 0.5 V_{om} (rate of Rubisco oxygenation reaction in the mesophyll). In order to investigate the impact of this distributed C_2 cycle onto leaf nitrogen metabolism, a genome-scale metabolic model of C_4 photosynthesis (de Oliveira Dal'Molin et al., 2010b) is coupled with a mechan-

istic model of C_2 photosynthesis (von Caemmerer, 2000). Constraint based simulations using optimality principles suggest the operation of nitrogen shuttles that compensate for the net import of nitrogen into the bundle sheath. These involve known components of the C_4 cycle and thus provide a mechanistic explanation of how the nitrogen shuttle serves as a pre-adaptation for C_4 photosynthesis. Additionally, modelling low activities of the C_4 cycle in concert with C_2 photosynthesis, as found in intermediate species of *Flaveria*, shows how both cycles overlap in order to minimise flux over the mesophyll bundle sheath interface. The trade-off between producing a gas-tight cellular environment and efficient diffusion of metabolites between the tissues thus results in a close interaction between the two cycles.

6.4 Mathematical models of leaf photosynthesis

6.4.1 The role of a metabolic model in *Manuscript 1*

The adaptive advantages of C_4 photosynthesis are considered to be high rates of net CO_2 assimilation (A) and high water and nitrogen-use efficiency (Brown, 1978; von Caemmerer, 2000; Taylor et al., 2010; Vogan and Sage, 2011). Water-use efficiency is mostly defined as CO_2 assimilation rate per stomatal conductance and is thus closely connected to A. Nitrogen-use efficiency is defined as CO_2 assimilation rate per leaf nitrogen and thus likewise connected to A. A suitable proxy for the fitness gain provided by C_4 photosynthesis is thus given by the net rate of CO_2 assimilation A.

The aim of the work presented in *Manuscript 1* was to model evolutionary dynamics of leaf photosynthesis. This required a mathematical model that is able to predict the fitness-relevant trait A from biochemical and anatomical parameters that are known to change during the evolution of C_4 photosynthesis. Further prerequisites include an appropriate resolution that captures the evolutionary parameters and omits physiological details that we lack data for, as well as biologically interpretable parameters that allow for validation of model predictions.

6.4.2 Metabolic models of leaf gas exchange

Since the 1970s, mathematical models of photosynthesis were used in a variety of contexts like unravelling the inhibition of photosynthesis through O_2 (Laing et al., 1974), explaining the carbon isotope discrimination of carbon concentration mechanisms (von Caemmerer, 1989), understanding photosynthetic oscillations (Laisk et al., 1989), modelling global carbon cycles (Lloyd and Farquhar, 1994), biotechnological optimization efforts (Zhu et al., 2007), and understanding optimization of metabolism on a genome-wide scale (de Oliveira Dal'Molin et al., 2010a,b; Saha et al., 2011). These models differ in scope and resolution, as well as mathematical foundation.

At the heart of most models of photosynthesis are the reactions catalysed by Rubisco. Once the kinetics of the ordered reaction mechanism were understood, steady state net CO_2 assimilation rate at saturating amounts of RuBP (A_c) could be explained for C_3 plants by a Rubisco model alone (Peisker, 1974; Farquhar, 1979; Farquhar et al., 1980). Considering the stoichiometry of light reactions as well as the energy demand of Calvin-Benson and photorespiratory cycle, the RuBP limited rate A_j can be deduced. The minimum of A_c and A_j yields the net CO_2 assimilation rate A. Although these algebraic Farquhar, von Caemmerer, Berry (FvCB) models of C_3 photosynthesis focus only on the main rate limiting steps, they were used successfully in a variety of applications (Yin and Struik, 2009).

The FvCB model describing C_3 photosynthesis was adapted to model the carbon concentrating mechanism of C_4 plants by including the PEPC reaction in the form of irreversible Michaelis-Menten kinetics (Berry and Farquhar, 1978; Peisker, 1979). This extension involves describing the bundle sheath partial pressure of CO_2 (C_s) based on a new parameter for bundle sheath conductance for CO_2 (g_s) and yields a quadratic equation in A.

Further extensions of the original FvCB model by von Caemmerer (1989, 1992, 2000) included a description of the amount of C_2 photosynthesis as a function of photorespiration in the mesophyll. This model was successfully used to describe the carbon isotope patterns and the curvilinear response of the CO_2 compensation point to varying O_2 concentrations in C_2 species (von Caemmerer, 1989, 1992, 2000).

The model used in the evolutionary simulations in *Manuscript 1* is based on the C_3–C_4 intermediate model presented in von Caemmerer (2000). It predicts A_c at steady state from the following parameters: the partial pressure of CO_2 in the mesophyll chloroplasts (C_m), the partial pressure of O_2 in the mesophyll chloroplasts (O_m), the Rubisco active site concentration (E_{tot}), the fraction of Rubisco expressed in the mesophyll (β), the maximum turnover rate of Rubisco carboxylation (k_{ccat}),

the PEPC-limited C_4 cycle activity (V_{pmax}), the Michaelis-Menten constant of PEPC for bicarbonate (K_p), the bundle sheath conductance for CO_2 (g_s), and the fraction of mesophyll-derived glycine decarboxylated in the mesophyll (ξ). in the original model given in von Caemmerer (2000), the Rubisco affinities for CO_2 (K_C) and O_2 (K_O), and the related specificity ($S_{C/O}$) are also free parameters. These kinetic parameters are subject to trade-offs (Savir et al., 2010) and can thus not be treated as independent in an evolutionary model. In *Manuscript 1*, an empirical model for Rubisco evolution published by Savir et al. (2010) was employed, expressing affinities and specificity as functions of k_{ccat}.

The resulting model thus captures the six main phenotypic parameters that are known to change during the evolution of C_4 photosynthesis from a C_3 ancestor (β, k_{ccat}, V_{pmax}, K_p, g_s, ξ). Experimental estimation in C_3 , C_3–C_4, and C_4 species is tractable for β, k_{ccat}, V_{pmax}, and ξ (Edwards and Gutierrez, 1972; Ku et al., 1976; Winter et al., 1982; Bauwe, 1984; Holaday et al., 1988; Moore et al., 1988, 1989; Wessinger et al., 1989; Schulze et al., 2013), while measurements of K_p and g_s are challenging and not available in literature. The feasibility of estimating four of the six parameters allowed for the evaluation of model predictions in *Manuscript 1*.

Finally, the model can be solved in a closed form. This allows even large numbers of evaluations like the sensitivity analysis presented in the supplementary material of *Manuscript 1* to remain computationally manageable.

6.4.3 The role of a metabolic model in *Manuscript 2*

The work presented in *Manuscript 2* seeks to understand nitrogen metabolism in a leaf operating C_2 photosynthesis. Since the conceptual model of C_2 metabolism results in a net transport of nitrogen into the bundle sheath (see Section 6.3.5), a balancing shuttle is necessary during steady state metabolism. In order to predict possible shuttles that follow optimality criteria, a genome-scale metabolic reconstruction of leaf photo-

synthesis was coupled to the mechanistic model representing C_2 photosynthesis described in Section 6.4.2.

6.4.4 Constraint-based modelling of metabolism via flux balance analysis

Flux balance analysis (FBA) is a widely used constraint-based modelling approach that is able to predict metabolic flux in large-scale metabolic networks (Price et al., 2004; Orth et al., 2010). FBA requires only sparse information about the metabolic system. The stoichiometry of reactions is combined with biological and thermodynamic constraints to define a space of possible solutions. In order to pinpoint a more specific set of solutions, an evolutionary argument can be applied: The system is assumed to have optimized an objective function under the given the constraints. Linear programming is employed to find a solution of such an optimization problem. This solution is not necessarily unique, and there might be additional flux distributions yielding the same value for the objective function.

On the one hand, FBA allows simulation of genome-scale metabolism without the need for kinetic data or knowledge about underlying regulatory networks, but on the other hand it will in some cases ignore essential properties of the system. The ability of FBA to predict the metabolic flux distribution was initially tested in microorganisms. In bacteria, the maximization of biomass accumulation is widely used as an objective function and, given certain constraints on nutrient uptake, translates to an optimization of metabolic yield (i.e. max(Biomass/substrate-uptake)). Although ideas from evolutionary game theory challenge the proposed optimization of metabolic yield (Schuster et al., 2008), this objective was successfully applied to predict intracellular flux and exchange rates of microorganisms (Varma and Palsson, 1994; Ibarra et al., 2002; Segrè et al., 2002; Schuetz et al., 2007).

While the cellular objective of biomass production can be used to predict metabolic flux in bacteria, multicellularity and compartmenta-

tion render the application of FBA to plant metabolism less straightforward. Nevertheless, a multitude of metabolic reconstructions were developed in the recent years. Poolman et al. (2009) constructed the first FBA-compatible model for *Arabidopsis thaliana*, and used a measure for enzymatic efficiency as the objective function to predict flux in a heterotrophic cell culture. Subsequent studies provided models for autotrophic tissue in Arabidopsis, and improved on compartmentation, gene-reaction associations, and tissue specificity (de Oliveira Dal'Molin et al., 2010a; Mintz-Oron et al., 2011). While these models focused on leaf metabolism in a single cell type, genome-scale metabolic reconstructions of C_4 metabolism, C4GEM and *i*RS1563 (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011), were developed to describe the interaction between mesophyll and bundle sheath cells. C4GEM was successfully validated using cell type-specific proteomics data, while *i*RS1563 predicted the phenotype of *brown midrib* (*bm*) mutants in maize (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011).

Although these models predict the operation of the C_4 cycle in FBA simulations and successfully explain experimental data, the fact that FBA does not model metabolite concentrations results in an inability to predict the C_4 cycle without the use of artificial constraints. To overcome this lack of mechanistic detail in constraint-based modelling of carbon concentrating mechanisms, the model presented in *Manuscript 2* combines the kinetics-based model described in Section 6.4.2 with C4GEM. Mathematically, this is achieved by calculating fluxes using the mechanistic model and constraining the respective reactions in C4GEM. This approach is similar to that of an integrated flux balance analysis (iFBA) (Covert et al., 2008), where fluxes predicted by a system of ordinary differential equations describing kinetics and regulation of a subset of *Escherichia coli* metabolism are used to constrain the solution space of an FBA problem.

The coupled model of C_2 metabolism was then used to predict nitrogen shuttles between mesophyll and bundle sheath cells that follow

optimality criteria. The objective function was applied in a two-step process. First, maximization of leaf biomass production was applied as previously applied for FBA by Saha et al. (2011). In order to find the nitrogen shuttles that involve the least amount of total plasmodesmatal transport, a minimization of the sum of fluxes with special weight on plasmodesmatal reactions was subsequently performed while fixing the optimal biomass production.

Estimating whether a certain metabolite is suitable for maintaining a diffusional gradient between mesophyll and bundle sheath is an unsolved problem. The impact on regulatory mechanisms and homeostasis of the C_3 leaf may render some metabolites unsuitable. A hierarchy of differentially optimal cycles was proposed by gradually constraining transport reactions over the mesophyll-bundle sheath interface.

6.4.5 Conclusion

In summary, the computational modelling of plant metabolism is still hampered by the lack of data on kinetic parameters, cell specificity, compartmentation, and regulation.

The various successful applications of the FvCB models show how a thorough identification of the system's limiting factors allows the development of predictive models. Although the nature of metabolic limitations is less clear in hypothetical intermediate states, the work presented in *Manuscript 1* demonstrates that these models can be applied to answer current evolutionary questions and contribute to understand the evolution of C_4 photosynthesis.

Constraint-based approaches to modelling leaf metabolism account for the lack of data by applying optimality assumptions. While this technique was applied successfully to explain experimental data from C_3 and C_4 plants, various objective functions were used, and a unifying systematic understanding of the optimized features in different photosynthetic networks is still lacking. The modelling work in *Manuscript 2* shows how

additional constraints can be obtained from the detailed kinetic FvCB models to reduce the volume of the solution space when modelling C_2 photosynthesis. This allowed the production of testable hypotheses on the nature of nitrogen shuttles. In order to understand photosynthesis and its evolution on a systems scale, future experimental work will have to reveal additional constraints on C_2 and C_4 metabolism and focus on the nature of the evolutionary forces that shaped photosynthetic networks.

References

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. *Journal of Experimental Botany*, 62(9):3049–3059.

Barrick, J., Yu, D., Yoon, S., Jeong, H., Oh, T., Schneider, D., Lenski, R., and Kim, J. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461(7268):1243–1247.

Bauwe, H. (1984). Photosynthetic enzyme activities and immunofluorescence studies on the localization of ribulose-1,5-biphosphate carboxylase/oxygenase in leaves of C_3, C_4, and C_3–C_4 intermediate species of *Flaveria* (*Asteraceae*). *Biochemie und Physiologie der Pflanzen*, 179(4):253–268.

Beale, M., C. V. and Long, S. P. (1995). Can perennial C_4 grasses attain high efficiencies of radiant energy conversion in cool climates? *Plant, Cell & Environment*, 18(6):641–650.

Berry, J. and Farquhar, G. (1978). The CO_2 concentrating function of C_4 photosynthesis: a biochemical model. In *Proceedings of the Fourth International Congress on Photosynthesis. Biochemical Society, London*, pages 119–131.

Brooks, A. and Farquhar, G. (1985). Effect of temperature on the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light. *Planta*, 165(3):397–406.

Brown, R. (1978). A difference in N use efficiency in C_3 and C_4 plants and its implications in adaptation and evolution. *Crop Science*, 18(1):93–98.

Bull, J. J., Badgett, M. R., Wichman, H. A., Huelsenbeck, J. P., Hillis, D. M., Gulati, A., Ho, C., and Molineux, I. J. (1997). Exceptional convergent evolution in a virus. *Genetics*, 147(4):1497–1507.

Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D., and Marx, C. J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*, 332(6034):1190–1192.

Christin, P., Besnard, G., Samaritani, E., Duvall, M., Hodkinson, T., Savolainen, V., and Salamin, N. (2008). Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology*, 18(1):37–43.

Christin, P.-A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., Garrison, L. M., Vorontsova, M. S., and Edwards, E. J. (2013). Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences*, 110(4):1381–1386.

Covert, M., Xiao, N., Chen, T., and Karr, J. (2008). Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics*, 24(18):2044.

Darwin, C. (1859). *The origin of species by means of natural selection, or, the preservation of favored races in the struggle for life*. J. Murray, London, UK.

de Oliveira Dal'Molin, C., Quek, L., Palfreyman, R., Brumbley, S., and Nielsen, L. (2010a). AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiology*, 152(2):579.

de Oliveira Dal'Molin, C., Quek, L., Palfreyman, R., Brumbley, S., and Nielsen, L. (2010b). C4GEM, a genome-scale metabolic model to study C_4 plant metabolism. *Plant Physiology*, 154(4):1871–1885.

Edwards, E., Osborne, C., Strömberg, C., and Smith, S. (2010). The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*, 328(5978):587.

Edwards, G. and Gutierrez, M. (1972). Metabolic activities in extracts of mesophyll and bundle sheath cells of *Panicum miliaceum* (L.) in

relation to the C_4 dicarboxylic acid pathway of photosynthesis. *Plant Physiology*, 50(6):728.

Ehleringer, J. and Björkman, O. (1977). Quantum yields for CO_2 uptake in C_3 and C_4 plants: Dependence on temperature, CO_2, and O_2 concentration. *Plant Physiology*, 59(1):86.

Ehleringer, J., Sage, R., Flanagan, L., and Pearcy, R. (1991). Climate change and the evolution of C_4 photosynthesis. *Trends in Ecology & Evolution*, 6(3):95–99.

Elena, S. F. and Lenski, R. E. (2003). Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4(6):457–469.

Engelmann, S., Bläsing, O., Gowik, U., Svensson, P., and Westhoff, P. (2003). Molecular evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Flaveria*—a gradual increase from C_3 to C_4 characteristics. *Planta*, 217(5):717–725.

Farquhar, G., Caemmerer, S., and Berry, J. (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta*, 149(1):78–90.

Farquhar, G. D. (1979). Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Archives of Biochemistry and Biophysics*, 193(2):456–468.

Fontana, W. and Schuster, P. (1998). Continuity in evolution: On the nature of transitions. *Science*, 280(5368):1451–1455.

Franke, J., Klözer, A., de Visser, J. A. G. M., and Krug, J. (2011). Evolutionary accessibility of mutational pathways. *PLoS Computational Biology*, 7(8):e1002134.

Furbank, R. T. (2011). Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *Journal of Experimental Botany*, 62(9):3103–3108.

Gowik, U. and Westhoff, P. (2011). The path from C_3 to C_4 photosynthesis. *Plant Physiology*, 155(1):56–63.

Grass Phylogeny Working Group II (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*, 193(2):304–312.

Griffiths, H., Weller, G., Toy, L. F. M., and Dennis, R. J. (2013). You're so vein: Bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant, Cell & Environment*, 36(2):249–261.

Hibberd, J. and Quick, W. (2002). Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. *Nature*, 415(6870):451–454.

Hibberd, J. M., Sheehy, J. E., and Langdale, J. A. (2008). Using C_4 photosynthesis to increase the yield of rice—rationale and feasibility. *Current Opinion in Plant Biology*, 11(2):228–231.

Hindré, T., Knibbe, C., Beslon, G., and Schneider, D. (2012). New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nature Reviews Microbiology*, 10(5):352–365.

Holaday, A., Brown, R., Bartlett, J., Sandlin, E., and Jackson, R. (1988). Enzymic and photosynthetic characteristics of reciprocal F1 hybrids of *Flaveria pringlei* (C_3) and *Flaveria brownii* (C_4-like species). *Plant physiology*, 87(2):484.

Hylton, C., Rawsthorne, S., Smith, A., Jones, D., and Woolhouse, H. (1988). Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3–C_4 intermediate species. *Planta*, 175(4):452–459.

Ibarra, R., Edwards, J., and Palsson, B. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, 420(6912):186–189.

Kauffman, S. and Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128(1):11–45.

Keys, A., Bird, I., and Cornelius, M. (1978). Photorespiratory nitrogen cycle. *Nature*, 275:741–743.

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., and Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332(6034):1193–1196.

Kiirats, O., Lea, P., Franceschi, V., and Edwards, G. (2002). Bundle sheath diffusive resistance to CO_2 and effectiveness of C_4 photosynthesis and refixation of photorespired CO_2 in a C_4 cycle mutant and wild-type *Amaranthus edulis*. *Plant physiology*, 130(2):964.

Kinsman, E. and Pyke, K. (1998). Bundle sheath cells and cell-specific plastid development in *Arabidopsis* leaves. *Development*, 125(10):1815.

Ku, S., Edwards, G., and Kanai, R. (1976). Distribution of enzymes related to C_3 and C_4 pathway of photosynthesis between mesophyll and bundle sheath cells of *Panicum hians* and *Panicum milioides*. *Plant and Cell Physiology*, 17(3):615–620.

Laing, W. A., Ogren, W. L., and Hageman, R. H. (1974). Regulation of soybean net photosynthetic CO_2 fixation by the interaction of CO_2, O_2, and ribulose 1,5-diphosphate carboxylase. *Plant Physiology*, 54(5):678–685.

Laisk, A., Eichelmann, H., Oja, V., Eatherall, A., and Walker, D. A. (1989). A mathematical model of the carbon metabolism in photosynthesis. difficulties in explaining oscillations by fructose 2,6-bisphosphate regulation. *Proceedings of the Royal Society of London. B. Biological Sciences*, 237(1289):389–415.

Lenski, R. E. and Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial popula-

tions. *Proceedings of the National Academy of Sciences*, 91(15):6808–6814.

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

Lunzer, M., Miller, S. P., Felsheim, R., and Dean, A. M. (2005). The biochemical architecture of an ancient adaptive landscape. *Science*, 310(5747):499–501.

Marshall, D. M., Muhaidat, R., Brown, N. J., Liu, Z., Stanley, S., Griffiths, H., Sage, R. F., and Hibberd, J. M. (2007). *Cleome*, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. *The Plant Journal*, 51(5):886–896.

McKown, A. D. and Dengler, N. G. (2009). Shifts in leaf vein density through accelerated vein formation in C_4 *Flaveria* (Asteraceae). *Annals of Botany*, 104(6):1085–1098.

Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A., and Shlomi, T. (2011). Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proceedings of the National Academy of Sciences*, 109(1):339–344.

Monson, R. (2003). Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences*, 164(3):43–54.

Monson, R. and Moore, B. (1989). On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. *Plant, Cell & Environment*, 12(7):689–699.

Monson, R., Moore, B., Ku, M., and Edwards, G. (1986). Co-function of C_3-and C_4-photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. *Planta*, 168(4):493–502.

Moore, B., Monson, R., Ku, M., and Edwards, G. (1988). Activities of principal photosynthetic and photorespiratory enzymes in leaf mesophyll and bundle sheath protoplasts from the C_3–C_4 intermediate *Flaveria* ramosissima. *Plant and cell physiology*, 29(6):999–1006.

Moore, B. D., Ku, M. S. B., and Edwards, G. E. (1989). Expression of C_4-like photosynthesis in several species of *Flaveria*. *Plant, Cell & Environment*, 12(5):541–549.

Morgan, C. L., Turner, S. R., and Rawsthorne, S. (1993). Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and of serine hydroxymethyltransferase in leaves of C_3-C_4 intermediate species from different genera. *Planta*, 190:468–473.

Muhaidat, R., Sage, T. L., Frohlich, M. W., Dengler, N. G., and Sage, R. F. (2011). Characterization of C_3–C_4 intermediate species in the genus *Heliotropium* l. (Boraginaceae): Anatomy, ultrastructure and enzyme activity. *Plant, Cell & Environment*, 34(10):1723–1736.

Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, 6(2):119–127.

Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature biotechnology*, 28(3):245–248.

Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B., and Bohaty, S. (2005). Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science*, 309(5734):600–603.

Peisker, M. (1974). A model describing the influence of oxygen on photosynthetic carboxylation. *Photosynthetica*, 8:47–50.

Peisker, M. (1979). Conditions for low, and oxygen-independent CO_2 compensation concentrations in C_4 plants as derived from a simple model. *Photosynthetica*, 13:198–207.

Pick, T. R., Bräutigam, A., Schlüter, U., Denton, A. K., Colmsee, C., Scholz, U., Fahnenstich, H., Pieruschka, R., Rascher, U., Sonnewald, U., and Weber, A. P. (2011). Systems analysis of a maize leaf developmental gradient redefines the current C_4 model and provides candidates for regulation. *The Plant Cell Online*, 23(12):4208–4220.

Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J., and Tans, S. J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology*, 272(1):141 – 144.

Poolman, M. G., Miguet, L., Sweetlove, L. J., and Fell, D. A. (2009). A genome-scale metabolic model of Arabidopsis and some of its properties. *Plant Physiology*, 151(3):1570–1581.

Price, N. D., Reed, J. L., and Palsson, B. Ø. (2004). Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11):886–897.

Rawsthorne, S., Hylton, C., Smith, A., and Woolhouse, H. (1988a). Distribution of photorespiratory enzymes between bundle-sheath and mesophyll cells in leaves of the C_3–C_4 intermediate species *Moricandia arvensis* (L.) DC. *Planta*, 176(4):527–532.

Rawsthorne, S., Hylton, C., Smith, A., and Woolhouse, H. (1988b). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C_3 and C_3–C_4 intermediate species of *Moricandia*. *Planta*, 173(3):298–308.

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

Ros, R., Cascales-Miñana, B., Segura, J., Anoman, A. D., Toujani, W., Flores-Tornero, M., Rosa-Tellez, S., and Muñoz-Bertomeu, J. (2013). Serine biosynthesis by photorespiratory and non-photorespiratory pathways: An interesting interplay with unknown regulatory networks. *Plant Biology*, 15(4):707–712.

Sage, R. (2002). Variation in the k_{cat} of Rubisco in C_3 and C_4 plants and some implications for photosynthetic performance at high and low temperature. *Journal of Experimental Botany*, 53(369):609–620.

Sage, R. and Sage, T. (2007). Learning from nature to develop strategies for the directed evolution of C_4 rice. *Charting new pathways to C_4 rice. World Scientific Publishing*, pages 195–216.

Sage, R. F. (2004). The evolution of C_4 photosynthesis. *New Phytologist*, 161(2):341–370.

Sage, R. F. (2013). Photorespiratory compensation: A driver for biological diversity. *Plant Biology*, 15(4):624–638.

Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C_4 plant lineages of planet Earth. *Journal of Experimental Botany*, 62(9):3155–3169.

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*, 63(1):19–47.

Saha, R., Suthers, P. F., and Maranas, C. D. (2011). *Zea mays* *i*RS1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS ONE*, 6(7):e21784.

Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010). Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proceedings of the National Academy of Sciences*, 107(8):3475–3480.

Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3(1).

Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., Gowik, U., and Westhoff, P. (2013). Evolution of C_4 pho-

tosynthesis in the genus *Flaveria*: Establishment of a photorespiratory CO_2 pump. *The Plant Cell Online*, 25(7):2522–2535.

Schuster, S., Pfeiffer, T., and Fell, D. A. (2008). Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, 252(3):497–504.

Segrè, D., Vitkup, D., and Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117.

Slewinski, T. L. (2013). Using evolution as a guide to engineer Kranz-type C_4 photosynthesis. *Frontiers in Plant Science*, 4(212).

Still, C. J., Berry, J. A., Collatz, G. J., and DeFries, R. S. (2003). Global distribution of C_3 and C_4 vegetation: Carbon cycle implications. *Global Biogeochemical Cycles*, 17(1):1006.

Taylor, S., Hulme, S., Rees, M., Ripley, B., Ian Woodward, F., and Osborne, C. (2010). Ecophysiological traits in C_3 and C_4 grasses: a phylogenetically controlled screening experiment. *New Phytologist*, 185(3):780–791.

Varma, A. and Palsson, B. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60(10):3724.

Vogan, P. J. and Sage, R. F. (2011). Water-use efficiency and nitrogen-use efficiency of C_3–C_4 intermediate species of *Flaveria* Juss. (Asteraceae). *Plant, Cell & Environment*, 34(9):1415–1430.

von Caemmerer, S. (1989). A model of photosynthetic CO_2 assimilation and carbon-isotope discrimination in leaves of certain C_3–C_4 intermediates. *Planta*, 178(4):463–474.

von Caemmerer, S. (1992). Carbon isotope discrimination in C_3–C_4 intermediates. *Plant, Cell & Environment*, 15(9):1063–1072.

von Caemmerer, S. (2000). *Biochemical models of leaf photosynthesis*. Csiro Publishing, Collingwood, Australia.

von Caemmerer, S., Quick, W. P., and Furbank, R. T. (2012). The development of C_4 rice: Current progress and future challenges. *Science*, 336(6089):1671–1672.

Wagner, G. P. and Altenberg, L. (1996). Perspective: Complex adaptations and the evolution of evolvability. *Evolution*, pages 967–976.

Wang, X., Gowik, U., Tang, H., Bowers, J., Westhoff, P., and Paterson, A. (2009). Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses. *Genome Biology*, 10(6):1–18.

Weinreich, D. M., Delaney, N. F., DePristo, M. A., and Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, 312(5770):111–114.

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

Williams, B. P., Johnston, I. G., Covshoff, S., and Hibberd, J. M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *eLife*, 2.

Winter, K., Usuda, H., Tsuzuki, M., Schmitt, M., Edwards, G. E., Thomas, R. J., and Evert, R. F. (1982). Influence of nitrate and ammonia on photosynthetic characteristics and leaf anatomy of *Moricandia arvensis*. *Plant Physiology*, 70(2):616–625.

Yin, X. and Struik, P. (2009). C_3 and C_4 photosynthesis models: An overview from the perspective of crop modelling. *NJAS-Wageningen Journal of Life Sciences*, 57(1):27–38.

Zhu, X., deSturler, E., and Long, S. (2007). Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: A numerical simulation using an evolutionary algorithm. *Plant physiology*, 145(2):513–526.

7 Manuscripts

7.1 *Manuscript 1*: Predicting C_4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape

Theory

Predicting C_4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape

David Heckmann,[1] Stefanie Schulze,[2] Alisandra Denton,[3] Udo Gowik,[2] Peter Westhoff,[2,4] Andreas P.M. Weber,[3,4] and Martin J. Lercher[1,4,*]
[1]Institute for Computer Science
[2]Institute for Plant Molecular and Developmental Biology
[3]Institute for Plant Biochemistry
Heinrich Heine University, 40225 Düsseldorf, Germany
[4]Cluster of Excellence on Plant Sciences (CEPLAS)
*Correspondence: lercher@cs.uni-duesseldorf.de
http://dx.doi.org/10.1016/j.cell.2013.04.058

SUMMARY

An ultimate goal of evolutionary biology is the prediction and experimental verification of adaptive trajectories on macroevolutionary timescales. This aim has rarely been achieved for complex biological systems, as models usually lack clear correlates of organismal fitness. Here, we simulate the fitness landscape connecting two carbon fixation systems: C_3 photosynthesis, used by most plant species, and the C_4 system, which is more efficient at ambient CO_2 levels and elevated temperatures and which repeatedly evolved from C_3. Despite extensive sign epistasis, C_4 photosynthesis is evolutionarily accessible through individually adaptive steps from any intermediate state. Simulations show that biochemical subtraits evolve in modules; the order and constitution of modules confirm and extend previous hypotheses based on species comparisons. Plant-species-designated C_3-C_4 intermediates lie on predicted evolutionary trajectories, indicating that they indeed represent transitory states. Contrary to expectations, we find no slowdown of adaptation and no diminishing fitness gains along evolutionary trajectories.

INTRODUCTION

To predict the evolution of biological systems, it is necessary to embed a systems-level model for the calculation of fitness into an evolutionary framework (Papp et al., 2011). However, explicit theories to predict strong correlates of fitness exist for very few complex model systems (Papp et al., 2011; Stern and Orgogozo, 2008). A major example is the stoichiometric metabolic network models of microbial species, which have been used to predict bacterial adaptation to nutrient conditions in laboratory experiments (Fong and Palsson, 2004; Hindré et al., 2012; Ibarra et al., 2002). On a macroevolutionary timescale, related methods have been applied to predict the outcome and temporal order of reductive genome evolution in endosymbiotic bacteria (Pál et al., 2006; Yizhak et al., 2011). These studies on microbial evolution have employed metabolic yield of biomass production as a correlate of fitness, an approach that cannot be transferred directly to multicellular organisms.

However, it is likely that the efficiency with which limiting resources are converted into biomass precursors is under strong selection across all domains of life. For multicellular eukaryotes, this trait may be most easily studied in plants, which use energy provided by solar radiation to build sugars from water and CO_2. To fix carbon from CO_2, plants use the enzyme RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase). RuBisCO has a biologically relevant affinity for O_2, resulting in a toxic product that must be recycled in the energy-consuming metabolic repair pathway known as photorespiration (Maurino and Peterhansel, 2010). The decarboxylation of glycine—a key metabolite within this pathway—by the glycine decarboxylase complex (GDC) releases CO_2. About 30 million years ago, photorespiration increased to critical levels in many terrestrial ecosystems due to the depletion of atmospheric CO_2. To circumvent this problem, C_4 photosynthesis evolved to concentrate CO_2 around RuBisCO in specific cell types (Edwards et al., 2010; Sage et al., 2012).

CO_2 first enters mesophyll (M) cells, where most RuBisCO is located in C_3 plants. In contrast, C_4 plants have shifted RuBisCO to neighboring bundle sheath (BS) cells. In the M of C_4 plants, PEPC (phospho*enol*pyruvate carboxylase, which does not react with oxygen) catalyzes the primary fixation of CO_2 as bicarbonate. The resulting C_4 acids enter the BS and are decarboxylated, releasing CO_2 in proximity to RuBisCO. BS cells are surrounded by thick cell walls, believed to reduce CO_2 leakage (Kiirats et al., 2002). Such an energy-dependent biochemical CO_2-concentrating pump is the defining feature of C_4 plants; species differ in the decarboxylating enzyme employed and in the metabolites shuttled between cell types (Drincovich et al., 2011; Furbank, 2011; Pick et al., 2011).

Despite the complexity of C_4 photosynthesis, this trait constitutes a striking example of convergent evolution: it has evolved

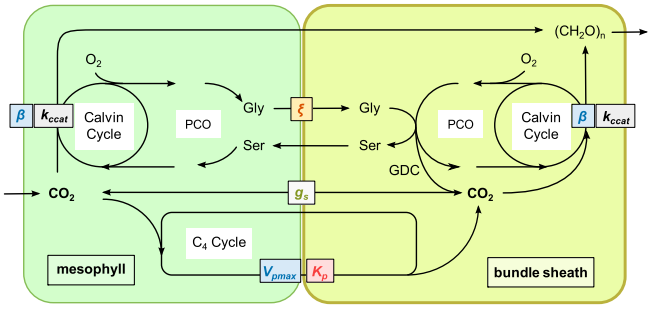

Figure 1. Overview of C_3-C_4 Biochemistry, Modeled as Two Interacting Cell Types

CO_2 enters the M and is either fixed by RuBisCO in the M or shuttled to the BS through the C_4 cycle and fixed by RuBisCO there. The resulting C_3 acids are fed into the Calvin cycle. Deleterious fixation of O_2 by RuBisCO leads to photorespiration (PCO). Model parameters are β, the fraction of RuBisCO active sites in the M; k_{ccat}, the maximal turnover rate of RuBisCO; ξ, the fraction of M derived glycine decarboxylated by GDC in the BS (note that for $\xi < 1$, decarboxylation of glycine also takes place in the M); V_{pmax}, the activity of the C_4 cycle; K_p, the Michaelis-Menten constant of PEPC for bicarbonate; and g_s, the BS conductance for gases. See also Figure S2 and Table S2.

independently in more than 60 angiosperm lineages from the ancestral C_3 photosynthesis (Sage et al., 2011). The leaf anatomy typical for C_4 plants—close vein spacing and prominent BS cells, designated "Kranz" anatomy—is also adaptive for C_3 species in environments associated with C_4 evolution (Brodribb et al., 2010). A rudimentary Kranz anatomy was thus likely already present in the C_3 ancestors of C_4 species (Sage et al., 2012), forming a "potentiating" anatomical state (Christin et al., 2011, 2013). Furthermore, all enzymes required for C_4 photosynthesis have orthologs in C_3 species, where they perform unrelated functions. In the evolution of C_4 biochemistry, these enzymes required concerted changes in their cell-type-specific gene expression as well as adjustment of their kinetic properties (Aubry et al., 2011; Gowik and Westhoff, 2011; Sage, 2004).

Some plant species have biochemistry that is intermediate between C_3 and C_4 (Edwards and Ku, 1987). These species possess a rudimentary Kranz anatomy and divide RuBisCO between M and BS cells. Often, however, photorespiratory glycine decarboxylation by GDC is largely shifted to the BS (see Figure 1), resulting in a moderate increase in the CO_2 concentration in BS cells (Sage et al., 2012).

C_4 plants make up 3% of today's vascular plant species but account for ~25% of terrestrial photosynthesis (Edwards et al., 2010; Sage et al., 2012). How C_4 photosynthesis evolved and why it evolved with such repeatability, are two fundamental questions in plant biology (Sage et al., 2012). Low atmospheric CO_2/O_2 ratio, heat, aridity, and high light are discussed as important factors promoting C_4 evolution, explaining the abundance of C_4 plants in tropical and subtropical environments (Edwards et al., 2010; Ehleringer et al., 1991). However, C_4 metabolism also allows higher biomass production rates in temperate regions (Beale and Long, 1995). The resulting accelerated growth makes engineering of the C_4 trait into major crops a promising route toward meeting the growing demands on food production (Hibberd et al., 2008). Rational strategies to approach this challenge require a detailed understanding of not only the C_4 state but also the fitness landscape connecting it with the ancestral C_3 biochemistry.

Here, we map the biochemical fitness landscape on which evolution from C_3 to C_4 photosynthesis occurs. Inserting the fitness estimates into a population genetic framework, we then explore the probability distribution of evolutionary trajectories leading from C_3 to C_4 systems. We thereby predict biochemical evolution in a multicellular eukaryote on macroevolutionary timescales (Hindré et al., 2012; Papp et al., 2011). Our results show that C_4 evolution is repeatable and predictable in its details. Importantly, experimentally determined parameter sets for C_3-C_4 intermediates fall well within the clustered distribution of predicted evolutionary trajectories. This agreement not only validates the model but also further provides important insights into the evolutionary nature of these species as transitory states in the evolution toward full C_4 photosynthesis.

RESULTS

A Biochemical Model for C_3-C_4 Evolution

RuBisCO is the most abundant protein on earth, responsible for up to 30% of nitrogen investment and 50% of total protein investment in plants (Ellis, 1979). C_4 plants typically contain lower amounts of RuBisCO per leaf area than C_3 plants (Ghannoum et al., 2011), explaining their lower nitrogen requirements (Brown, 1978). Reduced RuBisCO production is facilitated by higher CO_2 assimilation per RuBisCO protein, allowing C_4 plants to channel protein investment into other processes. In addition, C_4 plants do not need to open their stomata as much as C_3 plants to ensure sufficient internal CO_2 partial pressure, and they thus lose less water in hot and arid environments (Ghannoum et al., 2011). We assume that the overall fitness gain associated with C_4 photosynthesis is proportional to the amount of CO_2 that can be fixed using a given quantity of RuBisCO per leaf area (A_c).

To predict the steady-state enzyme-limited net CO_2 assimilation rate, A_c, from phenotypic parameters, we modified a mechanistic biochemical model developed by von Caemmerer (2000) to describe C_3-C_4 intermediates (Figure 1 and Experimental Procedures; see also Peisker, 1986). The underlying von Caemmerer model is itself based on models describing gas exchange in C_3 and in C_4 plants (Berry and Farquhar, 1978; Farquhar et al., 1980; von Caemmerer, 1989, 2000); these models have been used and validated in a variety of contexts (Yin and Struik, 2009). An extensive discussion of the model's generality and the choice of parameters can be found in the von Caemmerer book (2000).

C_3 and C_4 metabolisms represent limiting cases of the model, and representative parameter ranges were derived from C_3 and

C_4 species (Experimental Procedures). Evolution is modeled via changes in the following parameters: β, the fraction of RuBisCO active sites in the M, which ranges from ~95% in C_3 to 0% in some C_4 plants (where all RuBisCO is shifted to the BS); k_{ccat}, the maximal turnover rate of RuBisCO, which is lower in C_3 plants due to a trade-off with CO_2 specificity (Savir et al., 2010); ξ, the fraction of glycine derived from unwanted fixation of O_2 in M cells that is decarboxylated by GDC in the BS, ranging from 0 in C_3 to 1 in many C_3-C_4 intermediates (i.e., activity of the photorespiratory CO_2 pump); V_{pmax}, quantifying the activity of the C_4 cycle (i.e., the PEPC-dependent CO_2 pump); K_p, the Michaelis-Menten constant of PEPC (the core protein of the C_4 cycle) for bicarbonate; and g_s, the BS gas conductance (which quantifies the combined effects of cell geometry and cell wall properties).

Other kinetic parameters for RuBisCO were shown to be strongly linked to k_{ccat} (Savir et al., 2010) and are modeled accordingly (Extended Experimental Procedures and Figure S1 available online). The model describes the core steps of carbon fixation in communicating M and BS cells (Figure 1). CO_2 and O_2 enter M cells, with diffusion into and out of BS cells (g_s). CO_2 can be fixed in both cell types at rates characterized by the allocation (β) and kinetics (k_{ccat}) of RuBisCO. Alternatively, CO_2 may initially be fixed into a C_4 acid through the action of the C_4 cycle in M cells, characterized by the activity (V_{pmax}) and the kinetics (K_p) of its rate-limiting enzyme, PEPC. The C_4 acids then diffuse into the BS cells, where they are decarboxylated to free CO_2. We assume PEPC to be rate limiting (von Caemmerer, 2000), and thus neither this part of the C_4 cycle nor the recycling of the CO_2 carrier to the M is modeled explicitly. Finally, due to downregulation of GDC in the M, a fraction of the glycine resulting from the fixation of O_2 in the M is decarboxylated by GDC in BS cells (ξ).

The C_3 ancestors of C_4 species likely possessed a potentiating anatomy, characterized by decreased vein spacing and increased BS size (Christin et al., 2011, 2013). These anatomical features enable efficient diffusion of photorespiratory and C_4 cycle metabolites between compartments. C_3 plants that are closely related to C_4 species were further shown to exhibit a specific localization of chloroplasts and mitochondria in the BS cells. This ''proto-Kranz'' anatomy (Muhaidat et al., 2011) may be necessary for the establishment of a photorespiratory CO_2 pump by allowing the loss of GDC activity in the M to be compensated by the BS (Sage et al., 2012). Accordingly, our model starts from a C_3 state with proto-Kranz anatomy. This morphology can evolve further toward full C_4 Kranz anatomy (McKown and Dengler, 2007) via two main processes: (1) a reduction in the relative number of M cells and (2) an increase of BS cell size. Both processes influence our model exclusively by changing the proportion of RuBisCO allocated to BS cells instead of M cells (i.e., by decreasing β).

All parameters were normalized to total leaf area. At environmental conditions relevant for the evolution of C_4 photosynthesis and the constant RuBisCO concentration assumed in the model, C_3 and C_4 parameterizations lead to A_c values of 15.5 and 83.8 μmol m^{-2} s^{-1}, respectively. These hypothetical A_c values are assumed to reflect fitness gains during C_4 evolution, even if these fitness gains are in fact partially realized by the channeling of resources from RuBisCO production into other processes.

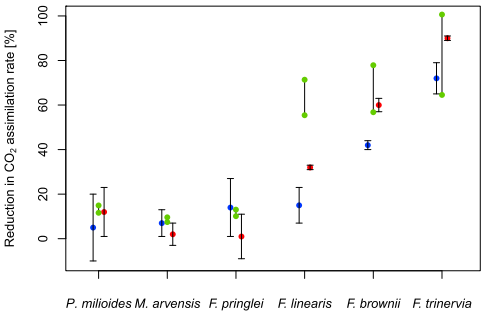

Figure 2. The Model Predicts the Reduction in Carbon Fixation Rate when the C_4 Cycle Is Reduced by Inhibiting PEPC

Blue and red dots show A_c reduction at 1 mM and 4 mM DCDP, respectively, with error bars indicating SD (Brown et al., 1991). Green dots show the range of predicted A_c reduction at 80%–100% inhibition of the C_4 cycle. See Extended Experimental Procedures for details.

C_4 species have been categorized into three subtypes, depending on the predominant decarboxylating enzyme (NAD malic enzyme, NAD-ME; NADP malic enzyme, NADP-ME; or phospho*enol*pyruvate carboxykinase, PEPCK) (Hatch et al., 1975). Our model is compatible with the stoichiometry of all three of these pathways under excess light. This agrees with experimental observations, which show that fitness-relevant traits are independent of C_4 subtype (Ehleringer and Pearcy, 1983; Ghannoum et al., 2001).

One major reason for the generality of our modeling approach is that carbon fixation is largely decoupled from other parts of plant metabolism. When light and nitrogen are available in excess, we thus expect that biomass production is strictly proportional to the carbon fixation rate, A_c. To confirm this, we coupled our C_3/C_4 model to a full plant metabolic network (Dal'Molin et al., 2010). The full model can be modified to reflect the different subtypes of C_4 metabolism (NAD-ME, NADP-ME, PEPCK). We sampled the parameter space of our C_3/C_4 model, using the predicted metabolite fluxes to constrain flux-balance analyses (FBA) of the full model (Oberhardt et al., 2009). For each of the three C_4 subtypes, we demonstrated that biomass production is indeed directly proportional to A_c (Figure S2; Pearson's $R^2 > 0.999$). These results support the robustness of our model to differences in the metabolism of different plant lineages.

As long as RuBisCO is active in both M and BS ($0 < \beta < 1$), our model predicts that CO_2 assimilation increases with decreasing M GDC expression (i.e., decreasing ξ). This prediction is consistent with experimental data from crosses between C_3-C_4 intermediate *Moricandia* and C_3 *Brassica* (Hylton et al., 1988). Furthermore, the model predicts the quantitative influence of experimentally suppressed C_4 cycles in phylogenetically diverse C_3-C_4 intermediates and C_4 plants (Brown et al., 1991) (Figure 2). A discrepancy between model and experiments is observed only for *F. linearis*. In this species, PEPC activity appears to be a suboptimal predictor for C_4 cycle activity, likely because of insufficient activity of PPDK (pyruvate, Pi dikinase) (Ku et al., 1983).

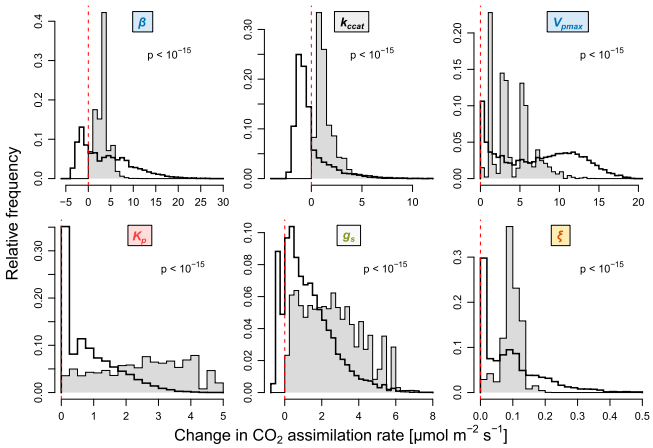

Figure 3. Realized Fitness Gains Are More Narrowly Distributed Than Potential Fitness Gains

White bars show potential fitness gains when one parameter is changed towards the C4 value. Gray bars show fitness gains realized in the evolutionary simulations. Negative values (to the left of the dashed red lines) indicate fitness reductions. Fitness is approximated by CO_2 assimilation rate. Although potential fitness gains vary widely, realized fitness gains are comparable between parameters. The distributions of potential and of realized fitness gains are significantly different ($p < 10^{-15}$ for each parameter, median tests). See also Figure S4.

Changes of the model parameters are ultimately caused by DNA mutations of protein coding or regulatory regions, and hence occur in discrete steps. Although each model parameter is known to show genetic variation, we currently lack a detailed understanding of the genotype-phenotype relationships. We thus divided each parameter range into six equidistant phenotypic states, with C_3 and C_4 states as endpoints. Choosing different discretizations did not change the observed patterns (Figure S3), except for ξ (see Discussion).

Despite Extensive Epistasis, the C_4 State Is Accessible from Every Point in the Fitness Landscape

The phenotypic parameters that distinguish C_3 from C_4 metabolism span a six-dimensional fitness landscape. Due to functional dependencies between the parameters, this landscape shows strong epistasis: fitness effects of changes in one parameter vary widely depending on the values of other parameters (Figure 3). Parameters differ in their potential influence on fitness. Whereas any individual increase in ξ raises A_c by at most 0.5 μmol m^{-2} s^{-1} (and never decreases fitness), a single increase in β can boost A_c by as much as 27 μmol m^{-2} s^{-1} or diminish A_c by as much as 3.7 μmol m^{-2} s^{-1}.

For half of the parameters (β, k_{ccat}, g_s), the same parameter change toward C_4 can both increase and decrease fitness, depending on the background provided by the remaining parameter values. This type of interaction has been termed sign epistasis (Weinreich et al., 2005) and affects 5.5% of the discretized fitness landscape (25,145 out of 486,000 pairwise combinations of parameter changes). Sign epistasis can be further classified as reciprocal if changing either of two parameters modifies fitness in one direction, while subsequently adding the second change modifies fitness in the opposite direction (Poelwijk et al., 2011). Reciprocal sign epistasis is a necessary (though not sufficient) condition for the existence of multiple fitness maxima (Poelwijk et al., 2011). The discrete C_3/C_4 fitness landscape contains only 20 points with reciprocal sign epistasis. All 20 involve an interaction between β and k_{ccat} at intermediate activity of the C_4 cycle (V_{pmax}). At these points, changes toward C_4 of β or k_{ccat} individually increase fitness. However, the C_4 cycle is not sufficiently active to compensate for the associated reduction in M photosynthetic efficiency when both parameters change simultaneously.

Maximal fitness is achieved when all parameters reach their C_4 values. Despite strong and often sign-changing epistasis, there is always at least one parameter change (median four changes) toward the C_4 state that increases fitness (Figure S4). Thus, the global fitness optimum is evolutionary accessible (Weinreich et al., 2005) from every position in the landscape. It immediately follows that there are no local maxima, giving the biochemical fitness landscape an exceedingly simple, smooth, "Mount (Mt.) Fuji-like" structure.

Modular Evolution of a Complex Trait

To evolve from C_3 to C_4 metabolism, our model requires 30 individual mutational changes (five steps in each of the six parameters). Parameters change with unequal probabilities. For example, the mutational target for inactivation of M GDC (increasing ξ) is large (Sage, 2004). Active GDC is a multienzyme system consisting of four distinct subunits, and downregulation of any of these will result in reduced GDC activity (Engel et al., 2007). Furthermore, M expression of each subunit is likely regulated by several transcription factor binding sites, each with several nucleotides important for binding. Random mutations at any of these sites are likely to downregulate M GDC expression. This inactivation is sufficient to establish a photorespiratory CO_2 pump, as we assume a low diffusional distance between M and BS cells, as well as a specific subcellular distribution of organelles in the BS (proto-Kranz anatomy). Due to this photorespiratory pump, any RuBisCO present in the BS will operate under increased CO_2 pressure, thereby increasing organismal fitness. Conversely, reduced GDC activity in BS cells would lead to decreased CO_2 pressure in the BS and hence would reduce organismal fitness. Thus, while random mutations may be equally likely to diminish GDC activity in M and in BS cells, only reductions in M activity are likely to be fixed in a population.

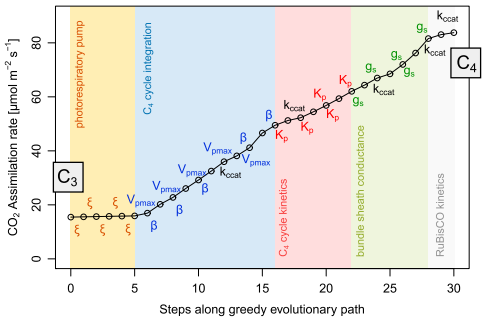

Figure 4. Fitness Changes along the "Greedy" Path through the Fitness Landscape from C_3 to C_4

This trajectory always chooses the most likely parameter change, combining mutation and fixation probabilities. The label centered above or below each edge indicates the mutation connecting two states. Evolution along the greedy path is modular (colored areas), except for the RuBisCO turnover rate k_{ccat}. CO_2 assimilation rate is used as a proxy for fitness. See also Figures S3 and S5.

In contrast to the large mutational target for the reduction of M GDC expression, other parameter changes involve increases in tissue-specific gene expression or changes in enzyme kinetics, which require specific mutations, restricted to only a few potential target nucleotides. Specifically, mutations that increase C_4 cycle activity appear much less likely, as different enzymes need to be upregulated in BS and in M cells, respectively. In the absence of precise estimates, we used plausible relative mutational probabilities for the model parameters (Extended Experimental Procedures). The general evolutionary patterns were found to be robust over a wide range of mutational probabilities and discretizations (Figure S3B).

Once a mutation that changes a model parameter occurs, its probability of fixation in the evolving plant population is determined by the associated change in fitness. Our simulations assume a "strong selection, weak mutation" regime, such that beneficial mutations are fixed in the population before the next mutation occurs (Gillespie, 1983). We estimated the fixation probability using a population genetic model first derived by Kimura (1957), assuming a constant population size of 100,000 individuals.

Each sequence of evolutionary changes linking the C_3 to the C_4 state defines an adaptive trajectory (or path) through the biochemical fitness landscape. The probability of individual steps is estimated as a combination of mutation and fixation probabilities. Figure 4 shows fitness changes associated with a unique "greedy" path, which always realizes the most likely parameter change. Here, changes for all but one of the six parameters are strictly clustered in modules (Figure 4). First, photorespiration is shifted to the BS ($\xi \uparrow$). Next, the C_4 cycle is established ($V_{pmax} \uparrow$), while RuBisCO is simultaneously shifted to the BS ($\beta \downarrow$). Then, the Michaelis-Menten constant of PEPC is adjusted ($K_p \downarrow$). Finally, gas diffusion is reduced ($g_s \downarrow$) in order to avoid leakage of CO_2 from the BS. The only parameter whose changes are not modular in this scenario is the maximal turnover rate of RuBisCO ($k_{ccat} \uparrow$), which is continuously adjusted along the greedy evolutionary trajectory, reflecting a shifting optimum due to the different CO_2 concentrations in M and BS.

Evolution is not deterministic, and the greedy path shown in Figure 4 represents only one of more than 10^{19} possible sequences of changes from C_3 to C_4. To more realistically characterize the evolution of C_4 biochemistry, we thus performed Monte Carlo simulations. At each step, we chose one parameter at random, weighted by the relative mutational probabilities. Using the biochemical model (Figure 1), we calculated the fitness change associated with adjusting the chosen parameter one step toward C_4. The change was accepted with a corresponding probability, derived from the population genetics model.

Despite the strong influence of chance, our Monte Carlo simulations support the same qualitative succession of modular changes in C_4 evolution (Figures S3A and S5). As observed in the greedy path, k_{ccat} is the only parameter that is continuously adjusted along the evolutionary trajectory, whereas ξ, V_{pmax} combined with β, K_p, and g_s tend to cluster with themselves ($p < 10^{-15}$ for dispersion higher than random of k_{ccat} and for modularity of ξ, V_{pmax} combined with β, K_p, and g_s; median tests for the distance between changes in the same parameter compared to random model).

Changes Early and Late in Adaptation Lead to Similar Fitness Increases

Strikingly, the greedy path through the fitness landscape (Figure 4) shows an almost linear fitness increase toward the C_4 state, with each evolutionary step resulting in a similar fitness increase. The only exceptions are the early establishment of a photorespiratory pump (ξ), the initial establishment of the C_4 cycle (V_{pmax}), and the two last adjustments of k_{ccat}. Thus, *realized* fitness gains along the greedy evolutionary path are very similar among the different parameters. This finding is in stark contrast to the broad distribution of *potential* fitness changes across the landscape (Figure 3).

Again, the stochastic evolutionary simulations support the result for the greedy path. Figure 3 shows that the distributions of realized fitness changes are much narrower than those of possible fitness changes. Furthermore, the median of realized fitness gains is similar across parameters, and lies around 2 μmol m^{-2} s^{-1} for all parameters except ξ. Accordingly, the time needed until the next parameter change is fixed in the population remains similar along evolutionary trajectories (Figure S6).

Repeatability of Evolution

The observed modularity and the narrow distributions of realized fitness gains demonstrate that the order of evolutionary changes toward C_4 is not arbitrary. Thus, evolution of this biochemical system is expected to repeat itself qualitatively in different species. Simulated evolutionary trajectories indeed cluster narrowly around a "mean path" ($p < 10^{-15}$; Figures 5 and S7).

Experimental Data from C_3-C_4 Intermediates Validate the Model

Our model of C_4 evolution is based on a number of simplifying assumptions and uses rough estimates of relative mutational

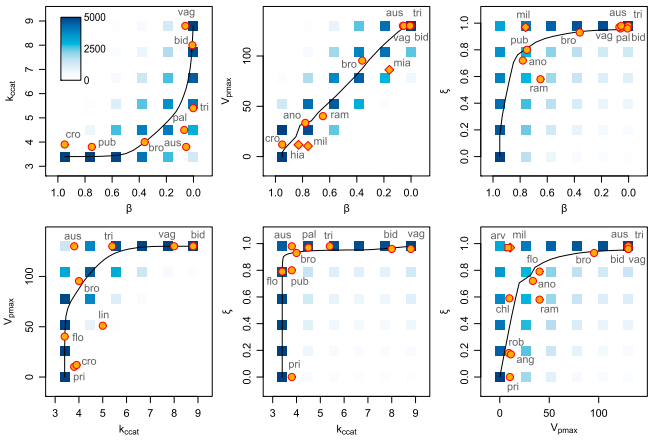

Figure 5. Projections of Trajectories through the Six-Dimensional Fitness Landscape Predicted by the Combined Biochemical and Stochastic Populations Genetics Model

Density of blue dots is proportional to the number of times a given parameter combination was crossed by a simulated trajectory. Black lines show the mean path of the set of trajectories. Orange dots are the *Flaveria* data described in the text, except for V_{pmax}, which was capped at 130 μmol m^{-2} s^{-1}. Abbreviations of species names: ang, *F. angustifolia*; ano, *F. anomala*; aus, *F. australasica*; bid, *F. bidentis*; bro, *F. brownii*; chl, *F. chloraefolia*; cro, *F. cronquistii*; flo, *F. floridana*; lin, *F. linearis*; pal, *F. palmeri*; pri, *F. pringlei*; pub, *F. pubescens*; ram, *F. ramosissima*; rob, *F. robusta*; tri, *F. trinervia*; vag, *F. vaginata*. Diamonds correspond to *Panicum* species: mil, *P. milioides*; hia, *P. hians*; mia, *P. miliaceum*. The square corresponds to *Moricandia arvensis*. See also Figures S6 and S7.

probabilities and population size. To assess its ability to quantitatively describe the evolution of real plants, we compared the model predictions to experimental data from the genera *Flaveria*, *Moricandia*, and *Panicum*. The experimental parameter sets for four plants and one plant correspond to the C_3 and C_4 endpoints, respectively. In addition, our data set included 15 species that have measured biochemical parameters intermediate between C_3 and C_4 (Figure 5); some of these species were previously classified as either C_3 or C_4 based on other criteria (McKown et al., 2005). Each of the intermediate species constitutes a separate point on evolutionary trajectories that started at C_3 biochemistry.

We collected experimental estimates of the biochemical model parameters for each of the 20 species from the literature, and we extended this data set by experimentally determining V_{pmax} and ξ for several *Flaveria* species (Experimental Procedures). With few exceptions, the experimentally determined parameter sets indeed lie very close to the predicted mean path through the fitness landscape (Figure 5). The model predicts experimental parameter combinations much better than a null model assuming a random order of evolutionary changes (Figure 6; $p < 10^{-15}$, median test).

DISCUSSION

The evolution of C_4 photosynthesis represents a rare opportunity to predict the functional evolution of a complex system: a closed six-parameter model calculates a phenotypic variable (A_c) of high relevance to fitness. The comparison to experimental data from diverse C_3-C_4 intermediates confirms the model's ability to quantitatively predict biochemical evolution over a timescale of several million years (Sage et al., 2012). While the majority of the data describe the genus *Flaveria*, the model also correctly predicts data from two phylogenetically distant genera (Figure 5). Comparisons to additional C_3-C_4 intermediates are currently limited by the availability of species-specific protocols for the separation of BS and M cells.

A hypothesis for the evolutionary succession of biochemical and morphological changes in the evolution of the C_4 syndrome was previously derived from phylogenetically informed analyses of C_3-C_4 intermediates (Sage et al., 2012). This hypothesis assumes modular biochemical changes, starting with a shift of photorespiration to the BS, followed by the establishment of a C_4 cycle in conjunction with a shift of RuBisCO to the BS, and finally an optimization stage in which parameters are fine-tuned. Our simulations support this scenario, narrowing it further by indicating that upregulation of the C_4 cycle usually precedes a shift of RuBisCO to the BS (Figure S3) even after previous establishment of a photorespiratory pump.

As expected due to the stochastic nature of evolution, the simulations indicate that modules are not strict and that the order of events may vary between independently evolving species. In particular, we find that the initial establishment of a photorespiratory pump (or C_2 cycle) is typical of evolutionary trajectories toward C_4 photosynthesis but may not be mandatory, as suggested previously (Sage et al., 2012).

Model Assumptions

While our model tracks changes in a phenotypic biochemical space, evolution is ultimately based on genomic mutations. We used qualitative reasoning when choosing relative mutational probabilities and the distribution of discrete steps linking C_3 and C_4 states. The sensitivity analysis (Figure S3B) demonstrates that other parameterizations lead to qualitatively very similar results. The only exception is the early establishment of a photorespiratory pump (ξ), which occurs with high probability only when the large mutational target for deactivation of the M GDC is taken into account.

The full C_4 cycle requires expression shifts in at least four separate enzymes. At each point in evolution, one of the enzymes that constitute the C_4 cycle will be rate limiting, making it the next target for fitness-enhancing upregulation. Distinct implementations of the C_4 cycle were shown to overlap in a

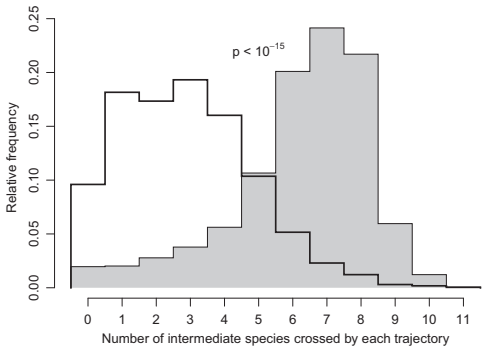

Figure 6. Distribution of the Number of Different C_3-C_4 Intermediate Species Whose Experimental Parameter Combinations Are Crossed by Each Single Predicted Trajectory

The combined biochemical and population genetics model (gray) fits the experimental data much better than a random model that ignores fitness effects (white) ($p < 10^{-15}$, median test). The parameter sets for *F. robusta*, *F. pringlei*, *F. cronquistii*, *F. angustifolia*, *and F. vaginata* are located at the C_3 or C_4 endpoints and hence crossed by every trajectory; they were excluded from this analysis.

single species (Furbank, 2011; Pick et al., 2011), potentially increasing the size of the mutational target. Our model uses the central enzyme PEPC to represent the complete pathway, accounting for the complexity of the C_4 cycle by using a low relative mutational probability.

A Simple, Mt. Fuji-like Biochemical Fitness Landscape

We found that the biochemical fitness landscape is exceedingly smooth: there are no local maxima besides the C_4 endpoint, as there is always at least one parameter change toward the C_4 value that increases the CO_2 fixation rate.

Comparison to experimental data from C_3-C_4 intermediate species indicates that our model indeed captures their evolutionary dynamics. The single-peaked fitness landscape suggests that these species are transitory states rather than evolutionary dead ends, continuously evolving toward the full C_4 syndrome as long as selective environmental conditions persist. The origin of *Flaveria* C_4 traits in the past 5 million years, together with the unusually large number of C_3-C_4 intermediate species in this genus (Sage et al., 2012), is consistent with this notion.

Half of the parameters in our model exhibit sign epistasis (Figure 3). Certain evolutionary trajectories thus involve reductions in fitness and are deemed not accessible (Weinreich et al., 2005); their inaccessibility contributes to the clustering of evolutionary trajectories. The paucity of reciprocal sign epistasis provides a partial explanation for the smooth landscape structure (Poelwijk et al., 2011).

Fitness landscapes resulting from interactions of mutations within the same gene can be rough and multipeaked (Weinreich et al., 2006). However, experimental fitness landscapes spanned by independently encoded functional units are similar in structure to the biochemical fitness landscape observed here: interactions among alleles of different genes rarely exhibit sign epistasis and often lead to simple, single-peaked landscapes (Chou et al., 2011; Khan et al., 2011; but see Kvitek and Sherlock, 2011).

Evolutionary Trajectories

Due to extensive sign epistasis among mutations within the same coding sequence, it was concluded that protein evolution may be largely reproducible and even predictable (Lozovsky et al., 2009; Weinreich et al., 2006). Despite the relatively low incidence of sign epistasis, we find that the same is true for the evolution of a complex biochemical system. Thus, different plants that independently "replay the tape of evolution" toward C_4 photosynthesis tend to follow similar trajectories of phenotypic changes (Figure 5). This resembles the high level of phenotypic and often genotypic parallelism in microbial evolution observed in experiments (Hindré et al., 2012) and predicted based on stoichiometric metabolic modeling (Fong and Palsson, 2004; Ibarra et al., 2002).

To explain the polyphyly of the C_4 syndrome, it has been hypothesized that each evolutionary step comes with a fitness gain (Gowik and Westhoff, 2011; Sage, 2004). We found that reality may be even more extreme: the fitness gain achieved by each individual change remained comparable along evolutionary trajectories (Figure 4). Accordingly, realized fitness advantages were much more similar across parameters than expected for random trajectories (Figure 3). This differs markedly both from theoretical expectations (Fisher, 1930; Orr, 2005) and from experimental observations in some genetic landscapes (Chou et al., 2011; Khan et al., 2011), which find diminishing fitness increases and a slowdown of adaptation along adaptive trajectories.

In the case of C_4 evolution, late-changing parameters (C_4 cycle kinetics, BS conductance) benefit from an already optimized background provided by previous evolution. Because everything else required for C_4 photosynthesis is already in place, their potential to contribute favorably to fitness is increased. Accordingly, we find no clear pattern of decelerated evolution along simulated trajectories, except for the last steps in PEPC kinetics and for late-occurring fixations of the now-superfluous photorespiratory pump (Figure S6). Conversely, the first few steps in C_4 evolution (initial establishment of CO_2 pumps) are only weakly selected, as only little RuBisCO is available in the BS at this time (Figure 4). Their fixation thus takes substantially longer than later changes (Figure S6): the first step is the most difficult one, also in C_4 evolution.

Why do C_3 plants still dominate many habitats, despite the simple, single-peaked fitness landscape and the substantial fitness gains resulting from individual evolutionary changes toward C_4 metabolism? A partial explanation is provided by weak selection on the first mutations. Furthermore, C_4 metabolism is strongly favored by selection only under specific environmental conditions, such as drought, high temperatures, and high light (excluding, for example, plants in woodlands). Finally, the potentiating Kranz-like anatomy in the C_3 ancestors of C_4 lineages (Christin et al., 2011, 2013; Sage et al., 2012) is not present in many other lineages, making the evolution of C_4 metabolism in these species unlikely.

The evolutionary dynamics uncovered above may shed light onto plans for experimental evolution of C_4 photosynthesis in C_3 plants through the application of increased selection pressure (Sage and Sage, 2007). Our results indicate that this endeavor may be accelerated by genetically engineering the first, slow steps of C_4 evolution. In particular, it may be advisable to pre-establish a photorespiratory CO_2 pump by knocking out M-specific GDC expression.

EXPERIMENTAL PROCEDURES

Biochemical Model and Fitness Landscape

The steady-state enzyme-limited net CO_2 assimilation rate (A_c) was used as a proxy for fitness of C_3, C_4, and intermediate evolutionary phenotypes. To predict A_c from phenotypic parameters, we slightly modified a mechanistic biochemical model for C_3-C_4 intermediates developed by von Caemmerer (2000) (Figure 1).

The CO_2 assimilation rates in the M and in the BS are calculated from the respective rates of carboxylation, oxygenation, and mitochondrial respiration (in addition to photorespiration). We assume constant concentrations of CO_2 (250 μbar) and O_2 (200 mbar) in M cells. Carboxylation and oxygenation are modeled as inhibitory Michaelis-Menten kinetics. RuBisCO kinetic parameters were shown to be subject to trade-offs (Savir et al., 2010); accordingly, we model these parameters as a function of RuBisCO maximal turnover rate (k_{ccat}). Activity of the C_4 cycle is assumed to be limited by PEPC activity and to follow Michaelis-Menten kinetics. The parameterization corresponds to a temperature of 25°C. The resulting set of equations can be solved for A_c in closed form. Equations, parameters, and further details are given in Extended Experimental Procedures and Tables S1 and S2.

For each evolving model parameter, we obtained representative C_3 and C_4 values (see below). The resulting range was subdivided into equidistant steps, leading to a discrete six-dimensional phenotype space. Based on the biochemical model, we calculated A_c for each parameter combination.

Calculation of Evolutionary Trajectories

We simulated a set of 5,000 evolutionary trajectories on the discrete fitness landscape, starting with the C_3 state. At each step, a trait (parameter) to be changed was chosen at random, with relative probabilities derived from current qualitative knowledge about the genetic complexity of the trait (Extended Experimental Procedures). We estimated selection coefficients (s) as the relative difference in A_c between ancestral and derived state, calculated using the biochemical model. We assumed a randomly mating population of diploid hermaphrodites, with incomplete dominance of mutations. The derived state was accepted with its probability of fixation, estimated using a formula first derived by Kimura (1957). We repeated the simulation process until reaching the C_4 parameter set.

To calculate a mean path from the set of 5,000 simulated trajectories, we averaged each parameter at each step (i.e., β at the first step of the mean path is the average of β values across the first steps of all simulated trajectories, etc.). Parameters were normalized to the interval [0,1]. Clustering of trajectories was quantified by calculating for each trajectory the mean of the normalized point-wise Manhattan distances to this mean trajectory. This measure is closely related to the recently introduced mean path divergence (Lobkovsky et al., 2011).

To estimate evolutionary modularity for each parameter, we used a distance measure defined as the number of other fixation events that occurred between two subsequent fixation events of the same parameter. V_{pmax} and β evolve together and were treated as a joint parameter in this context.

To determine a greedy trajectory through the landscape, we changed at each step the parameter that maximized the product of mutational probability and probability of fixation.

Comparison to Experimental Data

Data for the partitioning of RuBisCO between M and BS cells (β) and RuBisCO turnover rates (k_{ccat}), as well as PEPC activities (V_{pmax}) and decarboxylation of M-derived glycine in the BS (ξ) for *Moricandia* and *Panicum*, were obtained from the literature. We assayed PEPC activity in leaf extracts (summarized by Ashton et al., 1990) from 14 *Flaveria* species as a proxy for V_{pmax}. ξ was estimated for 14 *Flaveria* species by comparing the transcript levels of glycine decarboxylase P subunit genes that are expressed specifically in the BS (*gldpA*) to those expressed in all inner leaf tissues (*gldpD*). GldP transcript levels in leaves of 14 *Flaveria* species were determined by RNA sequencing. Data on K_p and g_s in intermediate species were not available. See Extended Experimental Procedures for more details on experimental data. We mapped experimental parameter values to the closest point in the discrete space of the model fitness landscape.

Random Null Model and Statistical Methods

To assess the statistical significance of our findings, we used a random null model to predict evolutionary trajectories. In this model, each trajectory starts with the C_3 state and evolves randomly, i.e., with equal probability for each directed parameter change, until the C_4 state is reached.

All simulations and statistical analyses were performed in the R environment (R Development Core Team, 2010). Statistical significance was assessed using Fisher's exact test and the median test implemented in the coin package (Hothorn et al., 2006).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.04.058.

ACKNOWLEDGMENTS

We thank Veronica Maurino, Itai Yanai, Eugene Koonin, Joachim Krug, and Andrea Bräutigam for helpful discussions. Computational support and infrastructure was provided by the "Center for Information and Media Technology" (ZIM) at the Heinrich Heine University Düsseldorf. This work was supported by the Deutsche Forschungsgemeinschaft (IRTG 1525 to D.H. and A.D.; FOR 1186 to S.S.; EXC 1028 to M.J.L., A.P.M.W., and P.W.; and CRC 680 to M.J.L.).

Received: December 18, 2012
Revised: March 21, 2013
Accepted: April 23, 2013
Published: June 20, 2013

REFERENCES

Ashton, A.R., Burnell, J.N., Furbank, R.T., Jenkins, C.L.D., and Hatch, M.D. (1990). The enzymes in C_4 photosynthesis. In Enzymes of Primary Metabolism, P.M. Dey and J.B. Harboene, eds. (London, UK: Academic Press), pp. 39–72.

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

Beale, C.V., and Long, S.P. (1995). Can perennial C_4 grasses attain high efficiencies of radiant energy-conversion in cool climates. Plant Cell Environ. *18*, 641–650.

Berry, J.A., and Farquhar, G.D. (1978). The CO_2 concentrating function of C_4 photosynthesis: a biochemical model. In Proceedings of the Fourth International Congress on Photosynthesis Biochemical Society, London, pp. 119-131.

Brodribb, T.J., Feild, T.S., and Sack, L. (2010). Viewing leaf structure and evolution from a hydraulic perspective. Funct. Plant Biol. *37*, 488–498.

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

Brown, R.H., Byrd, G.T., and Black, C.C. (1991). Assessing the degree of $c_{(4)}$ photosynthesis in c_3-c_4 species using an inhibitor of phosphoenolpyruvate carboxylase. Plant Physiol. *97*, 985–989.

Chou, H.H., Chiu, H.C., Delaney, N.F., Segrè, D., and Marx, C.J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science *332*, 1190–1192.

Christin, P.A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R., and Sage, R.F. (2011). Complex evolutionary transitions and the significance of c_3-c_4 intermediate forms of photosynthesis in Molluginaceae. Evolution *65*, 643–660.

Christin, P.A., Osborne, C.P., Chatelet, D.S., Columbus, J.T., Besnard, G., Hodkinson, T.R., Garrison, L.M., Vorontsova, M.S., and Edwards, E.J. (2013). Anatomical enablers and the evolution of C_4 photosynthesis in grasses. Proc. Natl. Acad. Sci. USA *110*, 1381–1386.

Dal'Molin, C.G., Quek, L.E., Palfreyman, R.W., Brumbley, S.M., and Nielsen, L.K. (2010). C4GEM, a genome-scale metabolic model to study C_4 plant metabolism. Plant Physiol. *154*, 1871–1885.

Drincovich, M.F., Lara, M.V., Andreo, C.S., and Maurino, V.G. (2011). Evolution of C_4 decarboxylases: Different solutions for the same biochemical problem: provision of CO_2 in Bundle Sheath Cells. In C_4 photosynthesis and related CO_2 concentration mechanisms, A.S. Raghavendra and R.F. Sage, eds. (Dordrecht: Springer), pp. 277–300.

Edwards, G.E., and Ku, M.S.B. (1987). Biochemistry of C_3-C_4 intermediates. In The biochemistry of plants, *Volume 10* (New York: Academic Press, Inc.), pp. 275–325.

Edwards, E.J., Osborne, C.P., Strömberg, C.A.E., Smith, S.A., Bond, W.J., Christin, P.A., Cousins, A.B., Duvall, M.R., Fox, D.L., Freckleton, R.P., et al.; C4 Grasses Consortium. (2010). The origins of C_4 grasslands: integrating evolutionary and ecosystem science. Science *328*, 587–591.

Ehleringer, J., and Pearcy, R.W. (1983). Variation in Quantum Yield for CO_2 Uptake among C_3 and C_4 Plants. Plant Physiol. *73*, 555–559.

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

Ellis, R.J. (1979). Most abundant protein in the world. Trends Biochem. Sci. *4*, 241–244.

Engel, N., van den Daele, K., Kolukisaoglu, U., Morgenthal, K., Weckwerth, W., Pärnik, T., Keerberg, O., and Bauwe, H. (2007). Deletion of glycine decarboxylase in Arabidopsis is lethal under nonphotorespiratory conditions. Plant Physiol. *144*, 1328–1335.

Farquhar, G.D., Caemmerer, S., and Berry, J.A. (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta *149*, 78–90.

Fisher, R.A. (1930). The Genetical Theory of Natural Selection (Oxford: Oxford Univ. Press).

Fong, S.S., and Palsson, B.O. (2004). Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat. Genet. *36*, 1056–1058.

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

Ghannoum, O., von Caemmerer, S., and Conroy, J.P. (2001). Carbon and water economy of Australian NAD-ME and NADP-ME C_4 grasses. Funct. Plant Biol. *28*, 213–223.

Ghannoum, O., Evans, J.R., and von Caemmerer, S. (2011). Nitrogen and water use efficiency of C_4 plants. In C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R.F. Sage, eds. (Dordrecht, The Netherlands: Springer), pp. 129–146.

Gillespie, J.H. (1983). A simple stochastic gene substitution model. Theor. Popul. Biol. *23*, 202–215.

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

Hatch, M.D., Kagawa, T., and Craig, S. (1975). Subdivision of C_4-pathway species based on differing C_4 acid decarboxylating systems and ultrastructural features. Funct. Plant Biol. *2*, 111–128.

Hibberd, J.M., Sheehy, J.E., and Langdale, J.A. (2008). Using C_4 photosynthesis to increase the yield of rice-rationale and feasibility. Curr. Opin. Plant Biol. *11*, 228–231.

Hindré, T., Knibbe, C., Beslon, G., and Schneider, D. (2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nat. Rev. Microbiol. *10*, 352–365.

Hothorn, T., Hornik, K., van de Wiel, M.A., and Zeileis, A. (2006). A Lego system for conditional inference. Am. Stat. *60*, 257–263.

Hylton, C.M., Rawsthorne, S., Smith, A.M., Jones, D.A., and Woolhouse, H.W. (1988). Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. Planta *175*, 452–459.

Ibarra, R.U., Edwards, J.S., and Palsson, B.O. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. Nature *420*, 186–189.

Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E., and Cooper, T.F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. Science *332*, 1193–1196.

Kiirats, O., Lea, P.J., Franceschi, V.R., and Edwards, G.E. (2002). Bundle sheath diffusive resistance to CO_2 and effectiveness of C_4 photosynthesis and refixation of photorespired CO_2 in a C_4 cycle mutant and wild-type *Amaranthus edulis*. Plant Physiol. *130*, 964–976.

Kimura, M. (1957). Some problems of stochastic processes in genetics. Ann. Math. Stat. *28*, 882–901.

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

Kvitek, D.J., and Sherlock, G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet. *7*, e1002056.

Lobkovsky, A.E., Wolf, Y.I., and Koonin, E.V. (2011). Predictability of evolutionary trajectories in fitness landscapes. PLoS Comput. Biol. *7*, e1002302.

Lozovsky, E.R., Chookajorn, T., Brown, K.M., Imwong, M., Shaw, P.J., Kamchonwongpaisan, S., Neafsey, D.E., Weinreich, D.M., and Hartl, D.L. (2009). Stepwise acquisition of pyrimethamine resistance in the malaria parasite. Proc. Natl. Acad. Sci. USA *106*, 12025–12030.

Maurino, V.G., and Peterhansel, C. (2010). Photorespiration: current status and approaches for metabolic engineering. Curr. Opin. Plant Biol. *13*, 249–256.

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

McKown, A.D., Moncalvo, J.-M., and Dengler, N.G. (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. Am. J. Bot. *92*, 1911–1928.

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

Oberhardt, M.A., Palsson, B.O., and Papin, J.A. (2009). Applications of genome-scale metabolic reconstructions. Mol. Syst. Biol. *5*, 320.

Orr, H.A. (2005). The genetic theory of adaptation: a brief history. Nat. Rev. Genet. *6*, 119–127.

Pál, C., Papp, B., Lercher, M.J., Csermely, P., Oliver, S.G., and Hurst, L.D. (2006). Chance and necessity in the evolution of minimal metabolic networks. Nature *440*, 667–670.

Papp, B., Notebaart, R.A., and Pál, C. (2011). Systems-biology approaches for predicting genomic evolution. Nat. Rev. Genet. *12*, 591–602.

Peisker, M. (1986). Models of carbon metabolism in C_3-C_4 intermediate plants as applied to the evolution of C_4 photosynthesis. Plant Cell Environ. *9*, 627–635.

Pick, T.R., Bräutigam, A., Schlüter, U., Denton, A.K., Colmsee, C., Scholz, U., Fahnenstich, H., Pieruschka, R., Rascher, U., Sonnewald, U., and Weber, A.P. (2011). Systems analysis of a maize leaf developmental gradient redefines the

current C_4 model and provides candidates for regulation. Plant Cell *23*, 4208–4220.

Poelwijk, F.J., Tănase-Nicola, S., Kiviet, D.J., and Tans, S.J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J. Theor. Biol. *272*, 141–144.

R Development Core Team. (2010). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

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

Sage, R.F., and Sage, T.L. (2007). Learning from nature to develop strategies for directed evolution of C_4 rice. In Charting New Pathways to C_4 Rice, J.E. Sheehy, P.L. Mitchell, and B. Hardy, eds. (Hackensack, NJ, USA: World Scientific Publishing), pp. 195–216.

Sage, R.F., Christin, P.A., and Edwards, E.J. (2011). The C_4 plant lineages of planet Earth. J. Exp. Bot. *62*, 3155–3169.

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

Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010). Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. Proc. Natl. Acad. Sci. USA *107*, 3475–3480.

Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? Evolution *62*, 2155–2177.

von Caemmerer, S. (1989). A model of photosynthetic CO_2 assimilation and carbon-isotope discrimination in leaves of certain C_3-C_4 intermediates. Planta *178*, 463–474.

von Caemmerer, S. (2000). Biochemical Models of Leaf Photosynthesis (Collingwood, Australia: Csiro Publishing).

Weinreich, D.M., Watson, R.A., and Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution *59*, 1165–1174.

Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. Science *312*, 111–114.

Yin, X., and Struik, P.C. (2009). C_3 and C_4 photosynthesis models: an overview from the perspective of crop modelling. NJAS-Wagen. J. Life Sci. *57*, 27–38.

Yizhak, K., Tuller, T., Papp, B., and Ruppin, E. (2011). Metabolic modeling of endosymbiont genome reduction on a temporal scale. Mol. Syst. Biol. *7*, 479.

Supplemental Information

EXTENDED EXPERIMENTAL PROCEDURES

Biochemical Model

In the following, we list the equations for the biochemical model. The model is slightly modified from the model of enzyme-limited C_3-C_4 intermediate photosynthesis given by von Caemmerer (2000). Parameter dimensions are listed in Table S1.

The maximal RuBisCO activity per leaf area in the mesophyll (V_{mmax}) and the bundle sheath (V_{smax}) are given as a function of the fraction of RuBisCO active sites in the mesophyll (β), the total leaf RuBisCO concentration (E_{tot}), and the maximal RuBisCO turnover rate (k_{ccat}):

$$V_{mmax} = \beta E_{tot} k_{ccat}$$

$$V_{smax} = (1 - \beta) E_{tot} k_{ccat}$$

The CO_2 assimilation rate in the mesophyll (A_m) comprises the rate of carboxylation (V_{cm}) and oxygenation (V_{om}) and the mitochondrial respiration other than photorespiration (R_m) in the mesophyll. V_{cm} and V_{om} are modeled as inhibitory Michaelis-Menten kinetics:

$$V_{cm} = \frac{C_m V_{mmax}}{C_m + K_C\left(1 + \frac{O_m}{K_O}\right)}$$

$$V_{om} = \frac{O_m V_{mmax}}{O_m + K_O\left(1 + \frac{C_m}{K_C}\right)}$$

$$A_m = V_{cm} - 0.5 V_{om} - R_m = \frac{(C_m - \gamma_* O_m) V_{mmax}}{C_m + K_c + O_m \frac{K_C}{K_O}} - R_m$$

C_m and O_m represent the CO_2 and O_2 partial pressure in the mesophyll chloroplasts, respectively. K_c and K_o are Michaelis-Menten constants of RuBisCO for CO_2 and O_2, respectively. γ^* is a function of the RuBisCO specificity ($S_{c/o}$):

$$\gamma_* = \frac{0.5}{S_{C/O}}$$

Activity of the C_4 cycle is assumed to be limited by PEPC activity and is given by Michaelis-Menten kinetics:

$$V_p = \frac{C_m V_{pmax}}{C_m + K_p}$$

CO_2 assimilation in the bundle sheath (A_s) is also catalyzed by RuBisCO, so kinetics as for A_m are applied. A_s can also be expressed as a function of the activity of the photorespiratory pump (ξ), the amount of photorespiration in the mesophyll (V_{om}), the C_4 cycle, and the rate of CO_2 leakage from the bundle sheath (L):

$$A_s = \frac{(C_s - \gamma_* O_s) V_{smax}}{C_s + K_c + O_s \frac{K_C}{K_O}} - R_s = \xi(0.5 V_{om}) + V_p - L,$$

where L is given by

$$L = g_s(C_s - C_m),$$

and g_s is the bundle sheath conductance for CO_2. As C_s is given by

$$C_s = \frac{V_p + 0.5\xi V_{om} - A_s}{g_s} + C_m,$$

and O_s is

$$O_s = \frac{A_s}{0.047 g_s} + O_m,$$

a second degree polynomial with respect to A_s is obtained. The smaller solution for A_s is chosen and given by:

$$a = 1 - \frac{1}{0.047}\frac{K_C}{K_O}$$

$$b = -V_p - 0.5\xi V_{om} - g_s C_m - V_{smax} + R_s - g_s\left(K_C + O_m\frac{K_C}{K_O}\right) - \frac{\gamma_* V_{smax} + \frac{K_C}{K_O}R_s}{0.047}$$

$$c = (V_{smax} - R_s)(V_p + 0.5\xi V_{om} + g_s C_m) - V_{smax} g_s \gamma_* O_m - R_s g_s\left(K_C + O_m\frac{K_C}{K_O}\right)$$

$$A_s = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

The enzyme limited net CO_2 assimilation rate A_c equals the sum of net assimilation in mesophyll and bundle sheath:

$$A_c = A_s + A_m$$

E_{tot} was set to 19.35 μmol m^{-2} and mitochondrial respiration was scaled to RuBisCO activity as suggested by von Caemmerer (2000). The mesophyll CO_2 and O_2 partial pressures (C_m, O_m, respectively) in the model were set to 250 μbar and 200 mbar, respectively; parameterization corresponds to a temperature of 25°C. Heat and a high O_2/CO_2 ratio promote photorespiration in an exponential manner (e.g., Ehleringer et al., 1991), so extreme environmental conditions may further increase the benefit of CO_2 concentration mechanisms.

RuBisCO Kinetic Constants

Savir et al. (2010) showed that constraints on the evolution of RuBisCO allow the description of its kinetic parameters through simple power laws. Thus it would not be adequate to treat the maximal carboxylation rate (k_{ccat}), the Michaelis-Menten constants for CO_2 (K_c) and O_2 (K_O), and the specificity ($S_{C/O}$) as independent evolutionary parameters in the model. Data from Savir et al. (2010) excluding form II RuBisCOs and the extreme *Synechococcus 6301* form were used to deduce power laws that are more suitable for land plants (Figure S1):

$$K_C = 16.07 k_{ccat}^{2.36}$$

$$\frac{K_C}{K_O} = 3.7 \cdot 10^{-4} k_{ccat}^{1.16}$$

$$\gamma_* = \frac{0.5}{S_{C/O}} = \frac{0.5}{5009.76 k_{ccat}^{-0.6}}$$

Inserting the resulting power laws into the model described above reduces the number of evolutionary parameters to six, namely β, V_{pmax}, K_p, g_s, ξ, and k_{ccat}. The resulting model thus spans a six-dimensional fitness landscape.

Population Genetics Model

The selection coefficient (s) is calculated using the net CO_2 assimilation rate of the ancestral state (A_{C1}) and the net CO_2 assimilation rate of the derived state (A_{C2}). We assume that fitness is proportional to net CO_2 assimilation rate:

$$S = \frac{A_{C2} - A_{C1}}{A_{C2}}$$

The probability of fixation (π) of the derived state in a population of randomly mating diploid hermaphrodites, where mutations are incompletely dominant (i.e., heterozygous effect h = 1/2), is given by (Kimura, 1957):

$$\pi = \begin{cases} \frac{1}{2N}, & s = 0 \\ \frac{1 - e^{-s}}{1 - e^{-2Ns}}, & s \neq 0 \end{cases}$$

Comparison of Model Predictions to Data from Experimental Inhibition of PEPC

Brown et al. (1991) evaluated the effect of the PEPC inhibitor DCDP (3,3-dichloro-2-dihydroxyphosphinoylmethyl-2-propenoate) on steady state net photosynthesis in C_3, C_4 and C_3-C_4 species from the genera *Flaveria*, *Panicum* and *Moricandia*. DCDP is expected to inhibit PEPC activity by 80% to 100% (Jenkins et al., 1989). In order to validate our model of steady state photosynthesis, we parameterized it for the species used in Brown et al. (1991) and evaluated the effect on A_c when reducing PEPC activity (V_{pmax}) by 80% and by 100%. Where experimental parameters were unavailable (see section ''Comparison to experimental data''), we used C_3 parameters for C_3 and C_3-C_4 intermediates, and C_4 parameters for C_4 species. Where β was not available, the value that maximizes A_c (given the remaining parameters) was used.

Coupling the Mechanistic Model with a Genome-Scale Metabolic Reconstruction

In order to show that the choice of biochemical model operates at the right resolution, we coupled the mechanistic model presented above with a genome scale metabolic reconstruction of C_4 metabolism, C4GEM (Dal'Molin et al., 2010). C4GEM accounts for 1,755 metabolites and 1,588 unique reactions and contains a complex biomass reaction including carbohydrates, cell wall components, amino acids, and nucleotides (Dal'Molin et al., 2010). Flux Balance Analysis (FBA) was conducted using the C4GEM model:

$$\text{maximize } cv$$

$$\text{subject to } Sv = 0$$

$$v_{\min} \leq v \leq v_{\max}$$

where c is the vector of coefficients in the objective function, here the leaf biomass production. v is the vector of fluxes through the networks reaction, S is the stoichiometric matrix of the metabolic network, and v_{min} and v_{max} represent constraints on the respective fluxes. In addition to the constraints used in C4GEM, the following reactions were constrained using the values predicted by the mechanistic model: net CO_2 uptake, RuBisCO carboxylation and oxygenation in mesophyll and bundle sheath, CO_2 leakage from the bundle sheath, PEPC activity in the mesophyll, activity of the respective decarboxylating enzyme in the bundle sheath, plasmodesmatal flux of glycine and serine and decarboxylation by the GDC complex.

We sampled the parameter space given by the mechanistic model 1,000 times, each time calculating the solution for A_c, constrained the FBA model using the predicted values and optimized biomass production under these constraints (Figure S2). This procedure was repeated for NADPME, NADME and PEPCK subtype constraints.

Analysis of the Fitness Landscape

In order to analyze the model, the six evolutionary parameters were constrained to ranges given by representative C_3 and C_4 values (Table S2). For β, k_{ccat}, and ξ, parameter ranges were chosen based on the data set from the genera *Flaveria*, *Moricandia* and *Panicum* presented below. Comparison of measurements for V_{pmax} with data on other proxies for C_4 cycle activity in *Flaveria* (such as $\delta^{13}C$ [Apel et al., 1988; Monson et al., 1988; Sudderth et al., 2007], CO_2 compensation point [Vogan and Sage, 2011], % ^{14}C in C_4 acids after 8-10 s pulse [Vogan and Sage, 2011]) showed saturation above PEPC activities of about 130 μmol m^{-2} s^{-1}, and the parameter range for V_{pmax} was thus chosen from zero to 130 μmol m^{-2} s^{-1}.

Data on bundle sheath conductivity are very sparse. We used 3 mmol m^{-2} s^{-1} for the C_4 value (as suggested by von Caemmerer [2000]) and a 15-fold higher value for the C_3 state, although this parameter was to our knowledge never measured for C_3 plants.

(Bauwe, 1986) used kinetic progress curves to estimate K_p in different species, and these results were used to estimate ranges for this parameter.

All parameter ranges were divided into five equidistant steps.

Analysis of Evolutionary Trajectories

The ultimate cause of evolutionary phenotypic changes are genomic mutations. As we currently lack a precise genotype-phenotype map for this system, we used qualitative reasoning when choosing relative mutational probabilities. This yielded the following hierarchy of mutational probabilities μ:

$$\mu(\xi) > \mu(k_{ccat}) > \mu(K_p) = \mu(g_s) = \mu(\beta) > \mu(V_{pmax})$$

As discussed above, loss of the chlorenchymatous isoforms of GLDP are sufficient to divert glycine decarboxylation to the bundle sheath specific forms, a comparatively minor molecular change (Sage, 2004). We thus placed the highest mutational probability on the activity of the photorespiratory pump, ξ (see discussion in the main text).

It was shown that a single mutation in the *rbc*L gene can act as a switch between C_3-like and C_4-like catalytic properties in *Flaveria* RuBisCO (Whitney et al., 2011). Although the underlying mechanism to gain C_4-like kinetics seems to differ between species (Whitney et al., 2011), this result suggests a rather high mutational probability for k_{ccat}. A large mutational target for changes in ξ is further supported by the fact that active GDC is a multi-enzyme system consisting of four distinct subunits, and downregulation of any of these will result in reduced GDC activity (Engel et al., 2007). Furthermore, M expression of each subunit is likely regulated by several transcription factor binding sites, each with several nucleotides important for binding. Random mutations at any of these sites are likely to downregulate M GDC expression. This inactivation is sufficient to establish a photorespiratory CO_2 pump, as we assume a low diffusional distance between M and BS cells, and a specific subcellular distribution of organelles in the BS (proto-Kranz anatomy).

We assigned the lowest mutational probability to V_{pmax}. Implementation of the C_4 cycle can vary between species (Furbank, 2011), and incomplete C_4 cycles can be operational (Monson and Moore, 1989). This increases the size of the C_4 cycle as a mutational target. Nevertheless, increased and localized expression of the respective rate-limiting gene is required. In the case of *Flaveria*, two *cis*-regulatory elements responsible for C_4-like expression of the *ppcA* gene coding for PEPC were identified (Crona et al., 2013). This suggests a higher complexity of changes needed when compared to loss of expression of an isoform or change in kinetic properties of an enzyme.

Although there is some insight into the coordinated expression of RuBisCO subunits (Rodermel, 2001), the molecular mechanisms for changes in β, g_s, and K_p are largely unknown. We set the corresponding mutational probabilities to equal values intermediate between those of k_{ccat} and V_{pmax}.

To rule out that wrong assumptions about the probabilities of changes and number of equidistant steps affect our results, we ran a sensitivity analysis against these factors. The simulation of 1,000 evolutionary trajectories was repeated 30,000 times with randomly chosen sets of parameters for probabilities of changes and number of equidistant steps. Mutational probabilities were each drawn uniformly between 0 and 1, and then normalized to sum up to one. Numbers of steps for each parameter were drawn uniformly between 1 and 10.

For each parameter in the biochemical model, the normalized mean of the step numbers at which fixation occurred was used to characterize each simulation run (Figure S3). The qualitative patterns of our specific parameter set are reproduced for almost all biochemical parameters. The only exception is ξ, which is a very late change in most scenarios, indicating that the photorespiratory pump needs a high probability of change in order to play a role in the evolutionary process. As discussed above, the underlying mechanism for increasing ξ justifies this high probability in our assumptions.

Comparison to Experimental Data

The dicotyledonous genera *Flaveria* (Asteraceae) and *Moricandia* (Brassicaceae), as well as the monocotyledonous *Panicum* (Poaceae), each contain C_3-C_4 intermediate species. In order to validate the evolutionary model we obtained data on species from these genera from the literature and complemented it with further measurements.

PEPC activity in leaf extracts was used as a proxy for C_4 cycle activity (V_{pmax}). *F. robusta*, *F. chloraefolia*, *F. pringlei*, *F. angustifolia*, *F. cronquistii*, *F. anomala*, *F. floridana*, *F. ramosissima*, *F. linearis*, *F. brownii*, *F. vaginata*, *F. trinervia*, *F. bidentis*, and *F. australasica* were grown in 17 cm pots on soil (C-400 with Cocopor [Stender Erden, Schermbeck, Germany] fertilized with 3 g/l Osmocote exact standard 3 – 4 M [Scotts, Nordhorn, Germany]) in May 2012 in the greenhouse. Additional light was given 16h per day. The first and second youngest fully expanded leaves were harvested from about 2 month old plants of comparable sizes. Four biological replicates were used per species, each containing material of three individuals. PEPC activity was determined as summarized by Ashton et al. (1990).

Additional PEPC activities for one *Moricandia* and three *Panicum* species were obtained from Winter et al. (1982) and Ku et al. (1976). The values from Ku et al. (1976) were converted to leaf area basis using data from Ku and Edwards (1978).

Data on RuBisCO distribution (β) for six *Flaveria* species and three *Panicum* species were obtained from cell separation experiments (Edwards and Gutierrez, 1972; Holaday et al., 1988; Ku et al., 1976; Moore et al., 1988, 1989), and in the case of the data from Ku et al. (1976) and Holaday et al. (1988), corrected for mesophyll to bundle sheath area ratio (Hattersley, 1984; McKown and Dengler, 2007; Wilson et al., 1983). For four *Flaveria* species, β was estimated from immunofluorescence studies (Bauwe, 1984). Immunofluorescence data were evaluated visually and corrected for mesophyll to bundle sheath cell ratio (McKown and Dengler, 2007).

RuBisCO turnover rate (k_{ccat}) for 11 *Flaveria* species was taken from Wessinger et al. (1989).

The fraction of mesophyll derived photorespirational glycine decarboxylated in the bundle sheath (ξ) in *Flaveria* was estimated from transcriptome data. The transcriptomes of photosynthetically active leaves from 14 *Flaveria* species (see above) were analyzed by RNA-seq via Illumina sequencing. The resulting reads (from one to four RNaseq experiments per species with 30 to 51 million reads per experiment) were mapped to the sequences of the *F. trinervia gldpA* and *gldpD* genes (GenBank accession: Z99767.1 and Z99768.1) with the software package CLC Genomic Workbench using standard settings and allowing nonambiguous mapping only. The P subunit of the glycine decarboxylase is an essential component of glycine decarboxylation. While the *gldpA* gene is

known to be transcribed exclusively in the bundle sheath in *Flaveria pringlei* (C_3) and *F. trinervia* (C_4), *gldpD* is transcribed throughout all inner leaf tissues in *F. pringlei*. ξ was calculated according to:

$$c = \frac{A + D}{2} - D$$

$$\xi = \frac{c}{c + D}$$

where *A* is the sum of reads mapped to *gldpA* and *D* is the sum of reads mapped to *gldpD*.

Estimates for ξ in one *Moricandia* and one *Panicum* species were obtained from immunogold labeling experiments (Hylton et al., 1988), corrected for mesophyll to bundle sheath distribution of mitochondria (Brown and Hattersley, 1989).

Bundle sheath conductance was estimated in some C_4 species using inhibitors of PEPC (Brown, 1997; Jenkins et al., 1989). These methods rely on the assumption that RuBisCO activity is confined to the bundle sheath, and g_s has to our knowledge never been measured for C_3-C_4 intermediates or C_3 species, where this assumption does not hold.

We used data from Bauwe (1986) to define the parameter range for K_p, but further data were not available.

The data set was compared to the predicted set of trajectories. Data points were mapped to the closest point in the discrete 6-dimensional space given by the model. This allowed counting the number of species that are crossed by each predicted path. Results were compared to the random null model described above.

SUPPLEMENTAL REFERENCES

Apel, P., Bauwe, H., Bassüner, B., and Maass, I. (1988). Photosynthetic properties of *Flaveria cronquistii*, *F. palmeri*, and hybrids between them. Biochem. Physiol. Pflanz. *183*, 291–299.

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

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

Brown, R.H. (1997). Analysis of bundle sheath conductance and C4 photosynthesis using a PEP-carboxylase inhibitor. Aust. J. Plant Physiol. *24*, 549–554.

Brown, R.H., and Hattersley, P.W. (1989). Leaf anatomy of C3-C4 species as related to evolution of C4 photosynthesis. Plant Physiol. *91*, 1543–1550.

Crona, K., Greene, D., and Barlow, M. (2013). The peaks and geometry of fitness landscapes. J. Theor. Biol. *317*, 1–10.

Edwards, G.E., and Gutierrez, M. (1972). Metabolic activities in extracts of mesophyll and bundle sheath cells of Panicum miliaceum (L.) in relation to the C_4 dicarboxylic acid pathway of photosynthesis. Plant Physiol. *50*, 728–732.

Hattersley, P.W. (1984). Characterization of C_4 type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. Ann. Bot. (Lond). *53*, 163–180.

Holaday, A.S., Brown, R.H., Bartlett, J.M., Sandlin, E.A., and Jackson, R.C. (1988). Enzymic and photosynthetic characteristics of reciprocal F1 hybrids of Flaveria pringlei (C_3) and Flaveria brownii (C_4-like species). Plant Physiol. *87*, 484–490.

Jenkins, C.L.D., Furbank, R.T., and Hatch, M.D. (1989). Inorganic carbon diffusion between C_4 mesophyll and bundle sheath cells: direct bundle sheath CO_2 assimilation in intact leaves in the presence of an inhibitor of the C_4 pathway. Plant Physiol. *91*, 1356–1363.

Ku, S.B., and Edwards, G.E. (1978). Photosynthetic efficiency of *Panicum hians* and *Panicum milioides* in relation to C_3 and C_4 plants. Plant Cell Physiol. *19*, 665–675.

Ku, S.B., Edwards, G.E., and Kanai, R. (1976). Distribution of enzymes related to C_3 and C_4 pathway of photosynthesis between mesophyll and bundle sheath cells of *Panicum hians* and *Panicum milioides*. Plant Cell Physiol. *17*, 615–620.

Monson, R.K., and Moore, B.D. (1989). On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. Plant Cell Environ. *12*, 689–699.

Monson, R.K., Teeri, J.A., Ku, M.S.B., Gurevitch, J., Mets, L.J., and Dudley, S. (1988). Carbon-isotope discrimination by leaves of *Flaveria* species exhibiting different amounts of C_3- and C_4-cycle co-function. Planta *174*, 145–151.

Moore, B.D., Monson, R.K., Ku, M.S.B., and Edwards, G.E. (1988). Activities of principal photosynthetic and photorespiratory enzymes in leaf mesophyll and bundle sheath protoplasts from the C_3-C_4 intermediate *Flaveria ramosissima*. Plant Cell Physiol. *29*, 999–1006.

Moore, B.D., Ku, M.S.B., and Edwards, G.E. (1989). Expression of C_4-like photosynthesis in several species of *Flaveria*. Plant Cell Environ. *12*, 541–549.

Rodermel, S. (2001). Pathways of plastid-to-nucleus signaling. Trends Plant Sci. *6*, 471–478.

Sudderth, E.A., Muhaidat, R.M., McKown, A.D., Kocacinar, F., and Sage, R.F. (2007). Leaf anatomy, gas exchange and photosynthetic enzyme activity in *Flaveria kochiana*. Funct. Plant Biol. *34*, 118–129.

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

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

Whitney, S.M., Sharwood, R.E., Orr, D., White, S.J., Alonso, H., and Galmés, J. (2011). Isoleucine 309 acts as a C4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) carboxylation rate in Flaveria. Proc. Natl. Acad. Sci. USA *30*, 14688–14693.

Wilson, J.R., Brown, R.H., and Windham, W.R. (1983). Influence of Leaf Anatomy on the Dry Matter Digestibility of C_3, C_4, and C_3/C_4 Intermediate Types of *Panicum* Species. Crop Sci. *23*, 141–146.

Winter, K., Usuda, H., Tsuzuki, M., Schmitt, M., Edwards, G.E., Thomas, R.J., and Evert, R.F. (1982). Influence of Nitrate and Ammonia on Photosynthetic Characteristics and Leaf Anatomy of *Moricandia arvensis*. Plant Physiol. *70*, 616–625.

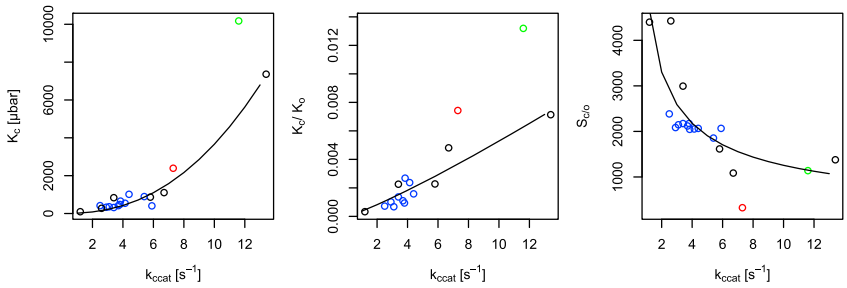

Figure S1. Nonindependence of RuBisCO Kinetic Constants, Related to Figure 1 and Extended Experimental Procedures
The figure shows two-dimensional fits to RuBisCO kinetic constants obtained from Savir et al. (2010). Least-squares fitting of power laws was conducted using the optim() function of the R environment. The resulting power laws reflect trade-offs, and were used to predict the other RuBisCO kinetic parameters from k_{ccat}. Blue, Land plants; red; Form II RuBisCO from *Rhodospirillum rubum*, not used for fitting; green, *Synechococcus 6301*, not used for fitting.

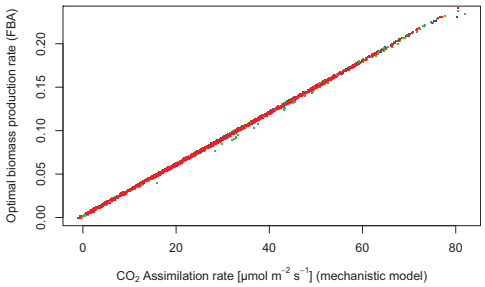

Figure S2. The Biomass Production Rate Predicted from Genome-wide Flux-Balance Analysis Is Directly Proportional to the Rate of Carbon Fixation, A_c, Related to Figure 1

C_4 subtypes are shown in different colors: green, NAD malic enzyme (NAD-ME); blue, NADP malic enzyme (NADP-ME); and red, phospho*enol*pyruvate carboxykinase. The slopes obtained from linear regressions for the three C_4 subtypes were statistically indistinguishable ($p = 0.38$, ANCOVA), demonstrating the robustness of the model.

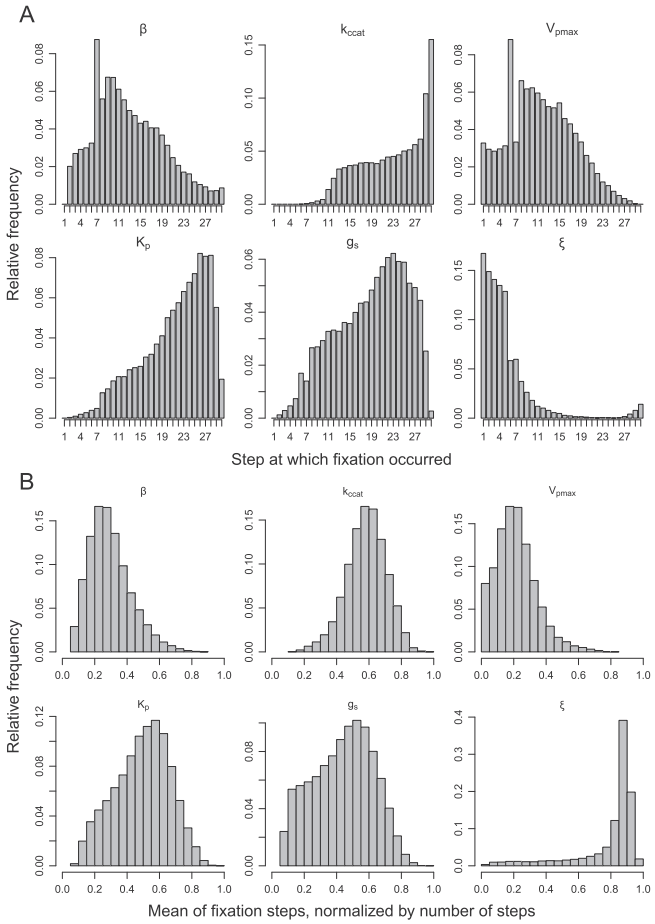

Figure S3. The Distribution of Fixation Times for Each Model Parameter, Related to Figure 4

(A and B) In most simulations, establishment of the photorespiratory pump (ξ) is the first change to occur. The C_4 cycle (V_{pmax}) and shift of RuBisCO activity to the bundle sheath (β) are also fixed in early stages. In our simulations, reduction of the conductance (g_s) is adaptive as soon as one of the pumps is established, but mainly occurs in later stages when the C_4 cycle is fully operating. K_p also changes late. Except for the last two changes, k_{ccat} shows the most uniform distribution along evolutionary trajectories. The same general pattern is seen with the discretizations and relative mutational probabilities assumed in our simulations (A) and in a sensitivity analysis that combines results from 1,000 simulations each of 30,000 randomly chosen parameter combinations (B). The only exception is the early establishment of the photorespiratory pump (ξ), which only happens in our simulations because the respective mutational probability is high.

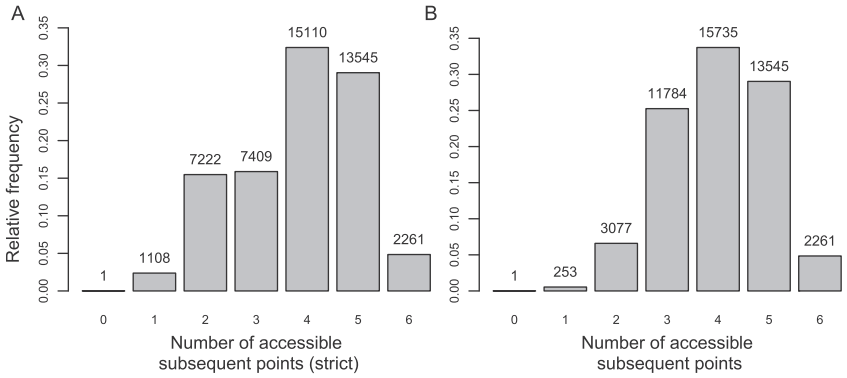

Figure S4. Evolutionary Accessibility of Subsequent Points in the Fitness Landscape, Related to Figure 3

(A and B) Subsequent points are defined as accessible if they come with a fitness change that is strictly positive (A) or at least zero (B); *i.e.*, for a point with *n* accessible subsequent points, *n* different parameters can be increased alternatively while increasing (A) or not decreasing (B) fitness. The only location lacking accessible subsequent points is the global maximum, the C_4 state.

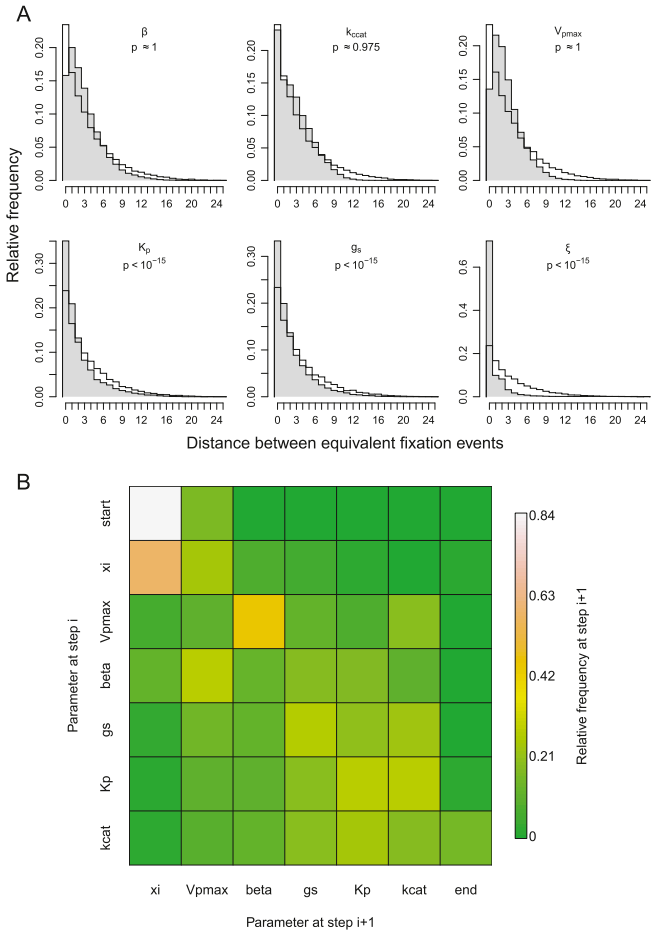

Figure S5. Relationships between Changes in Individual Parameters in the Stochastic Simulations, Related to Figure 4

(A) Distributions of distances between two changes in the same parameter. A distance of zero indicates two immediately successive steps. For the parameters in the bottom row (K_p, g_s, ξ), immediately successive steps (distance=0) are much more common than expected by chance ($p < 10^{-15}$ in each case, Fisher's exact test); the same is true for β and V_{pmax} when treated as a combined parameter set. The only trait that does not evolve in a modular fashion is thus k_{ccat}, which is significantly more dispersed than expected by chance ($p < 10^{-15}$, median test).

(B) Transition matrix for evolutionary trajectories in stochastic simulations. Colors indicate the relative frequency with which a change in parameter *Y* at step *i* is followed by a change in parameter *X* at step *i* + 1.

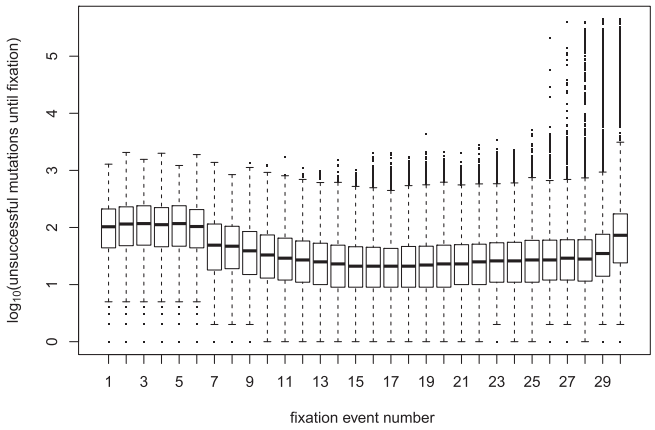

Figure S6. No Systematic Slowdown of Evolution, Related to Figure 5

Boxplot for the number of parameter changes that were attempted before a change was fixed according to the population genetic model, based on 5,000 simulated evolutionary trajectories from C_3 to C_4. The first six steps – mostly shifts in photorespiration to the bundle sheath (ξ) and the first establishment of C_4 cycle activity (V_{pmax}) – take substantially longer than later steps. Except for the very last steps, there is no clear trend of decelerating evolution, contrasting previous observations in experimental studies and theoretical expectations (see Discussion in the main text).

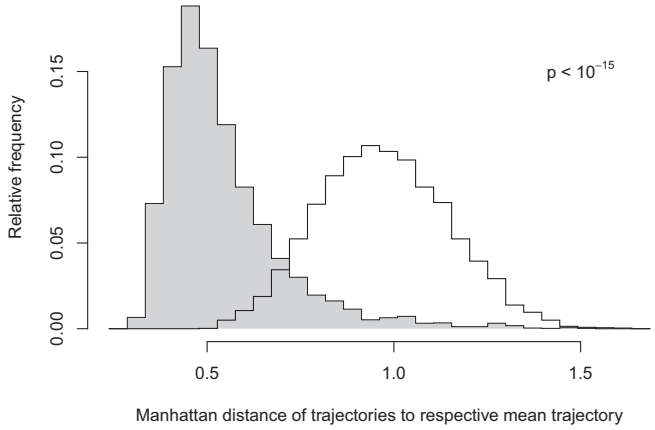

Figure S7. Simulated Paths Cluster, Related to Figure 5
Histograms of pointwise distances of simulated trajectories to the mean path. Gray: Evolutionary model. White: Random model. The two distributions are significantly different ($p < 10^{-15}$, Wilcoxon rank sum test).

Table S1. Parameter Dimensions in the Biochemical Model, Related to Extended Experimental Procedures

Parameters	Dimension
V_{mmax}, V_{smax}, V_{pmax}	μmol m^{-2} s^{-1}
C_m, O_m, C_s, O_s	μbar
β	-
E_{tot}	μmol m^{-2}
K_c, K_o, K_p	μbar
k_{ccat}	s^{-1}
R_m, R_s	μmol m^{-2} s^{-1}
ξ	-
$S_{c/o}$	-
g_s, g_o	μmol m^{-2} s^{-1}

Parameter descriptions: V_{mmax}, V_{smax}: maximal RuBisCO activity per leaf area in the mesophyll and bundle sheath, respectively; V_{pmax}: Activity of the C_4 cycle; C_m, C_s: CO_2 partial pressure in the mesophyll and bundle sheath chloroplasts, respectively; O_m, O_s: O_2 partial pressure in the mesophyll and bundle sheath chloroplasts, respectively; β: fraction of RuBisCO active sites in the mesophyll; E_{tot}: total leaf RuBisCO concentration; K_c, K_o,: Michaelis-Menten constants of RuBisCO for CO_2 and O_2, respectively; K_p: Michaelis-Menten constant of PEPC for bicarbonate; k_{ccat}: maximal rate of carboxylation for RuBisCO; R_m, R_s: mitochondrial respiration other than photorespiration in the mesophyll and the bundle sheath, respectively; ξ: activity of the photorespiratory pump; $S_{c/o}$: RuBisCO specificity for CO_2; g_s, g_o: bundle sheath conductance for CO_2 and O_2, respectively.

Table S2. Ranges and Discretization of Evolving Parameters, Related to Figure 1

Parameter	C_3 value	C_4 value	Dimension	Number of steps	Mutational probability
V_{pmax}	0	130	µmol m^{-2} s^{-1}	5	1/75
β	0.95	$2.0 \cdot 10^{-3}$	-	5	2/75
K_p	200	80	µbar	5	2/75
k_{ccat}	3.4	8.8	s^{-1}	5	4/75
ξ	0	0.98	-	5	64/75
g_s	$1.5 \cdot 10^{-2}$	$1.0 \cdot 10^{-3}$	µmol m^{-2} s^{-1}	5	2/75

Sources for parameter values are given in the text. Parameter descriptions: V_{pmax}: activity of the C_4 cycle; β: fraction of RuBisCO active sites in the mesophyll; K_p: Michaelis-Menten constant of PEPC for bicarbonate; k_{ccat}: maximal rate of carboxylation for RuBisCO; ξ: activity of the photorespiratory pump; g_s: bundle sheath conductance for CO_2.

7.1.1 Contributions

DH conducted all modelling and data integration, and drafted the manuscript; DH and MJL completed the manuscript; SS mapped the transcript data; DH and AD quantified PEPC activities; MJL, APMW and DH designed the research.

7.1.2 Outlook

In *Manuscript 1*, knowledge on the biochemistry, physiology, gene expression, leaf anatomy, molecular biology, and phylogeny of C_3, C_3–C_4, and C_4 plants is integrated—explicitly and in the form of assumptions—into a mathematical systems model. The model explains experimental data on intermediate species from three distantly related genera and makes predictions on future evolutionary trajectories that are shaped by the underlying fitness landscape. At present, the analysis is focusing on the dynamics in one specific environment exhibiting a low CO_2 to O_2 ratio and an intermediate temperature of 25°C. Evolutionary dynamics depend on the surrounding environment, and the topography of the fitness landscape will change as a function of environmental conditions. It will thus be of high interest to create a model that contains CO_2 and O_2 partial pressure, temperature, and light as environmental variables and to investigate the shape of the fitness landscape and patterns of evolutionary trajectories under different conditions.

Work on the correlation between environment and abundance of C_4 species was conducted using paleoclimate data (Ehleringer et al., 1991; Christin et al., 2008; Edwards et al., 2010; Sage et al., 2012) as well as data on present ecosystems (e.g.: Teeri and Stowe, 1976; Ehleringer et al., 1997). These studies indicate that temperature and CO_2 to O_2 ratio are important abiotic factors that render C_4 photosynthesis advantageous over the C_3 type in a given environment. Furthermore, they suggest that fitness landscapes of cool and high CO_2 environments will lose the single optimum of the C_4 state.

In the light of these results, it will not only be interesting to simulate landscapes under different static conditions, but to also predict trajectories that result from dynamically changing environments. This enables the investigation of fascinating questions on the reversibility of C_4 evolution (Bull and Charnov, 1985). In the face of ongoing climate change that will on the one hand increase CO_2 to O_2 ratios, but on the other hand increase mean day temperatures and most likely frequency of droughts (Sherwood and Fu, 2014), a systematic understanding of underlying fitness landscapes will aid in predicting the future of plant ecosystems.

References

Bull, J. and Charnov, E. (1985). On irreversible evolution. *Evolution*, pages 1149–1155.

Christin, P., Besnard, G., Samaritani, E., Duvall, M., Hodkinson, T., Savolainen, V., and Salamin, N. (2008). Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology*, 18(1):37–43.

Edwards, E., Osborne, C., Strömberg, C., and Smith, S. (2010). The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*, 328(5978):587.

Ehleringer, J., Cerling, T., and Helliker, B. (1997). C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia*, 112(3):285–299.

Ehleringer, J., Sage, R., Flanagan, L., and Pearcy, R. (1991). Climate change and the evolution of C_4 photosynthesis. *Trends in Ecology & Evolution*, 6(3):95–99.

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*, 63(1):19–47.

Sherwood, S. and Fu, Q. (2014). A drier future? *Science*, 343(6172):737–739.

Teeri, J. and Stowe, L. (1976). Climatic patterns and the distribution of C_4 grasses in North America. *Oecologia*, 23(1):1–12.

7.2 *Manuscript 2*: The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*

eLIFE elifesciences.org

RESEARCH ARTICLE

The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*

Julia Mallmann[1†], David Heckmann[2†], Andrea Bräutigam[3], Martin J Lercher[2,4], Andreas PM Weber[3,4], Peter Westhoff[1,4], Udo Gowik[1*]

[1]Institute for Plant Molecular and Developmental Biology, Heinrich-Heine-Universität, Düsseldorf, Germany; [2]Institute for Computer Science, Heinrich-Heine-Universität, Düsseldorf, Germany; [3]Institute of Plant Biochemistry, Heinrich-Heine-Universität, Düsseldorf, Germany; [4]Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-Universität, Düsseldorf, Germany

Abstract C_4 photosynthesis represents a most remarkable case of convergent evolution of a complex trait, which includes the reprogramming of the expression patterns of thousands of genes. Anatomical, physiological, and phylogenetic and analyses as well as computational modeling indicate that the establishment of a photorespiratory carbon pump (termed C_2 photosynthesis) is a prerequisite for the evolution of C_4. However, a mechanistic model explaining the tight connection between the evolution of C_4 and C_2 photosynthesis is currently lacking. Here we address this question through comparative transcriptomic and biochemical analyses of closely related C_3, C_3–C_4, and C_4 species, combined with Flux Balance Analysis constrained through a mechanistic model of carbon fixation. We show that C_2 photosynthesis creates a misbalance in nitrogen metabolism between bundle sheath and mesophyll cells. Rebalancing nitrogen metabolism requires anaplerotic reactions that resemble at least parts of a basic C_4 cycle. Our findings thus show how C_2 photosynthesis represents a pre-adaptation for the C_4 system, where the evolution of the C_2 system establishes important C_4 components as a side effect.
DOI: 10.7554/eLife.02478.001

***For correspondence:** gowik@uni-duesseldorf.de

†These authors contributed equally to this work

Competing interests: The authors declare that no competing interests exist.

Funding: See page 19

Received: 06 February 2014
Accepted: 14 June 2014
Published: 16 June 2014

Reviewing editor: Detlef Weigel, Max Planck Institute for Developmental Biology, Germany

Introduction

The dual-specific enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes two opposing reactions—the carboxylation and the oxygenation of ribulose 1,5-bisphosphate. The former reaction yields 3-phosphoglycerate (3-PGA), whereas the latter produces 2-phosphoglycolate (2-PG). 3-PGA is reduced to carbohydrates in the Calvin–Benson cycle and incorporated into biomass. However, 2-PG is toxic, which requires its removal by a metabolic repair pathway called photorespiration (***Anderson, 1971***; ***Bowes et al., 1971***; ***Ogren, 1984***; ***Leegood et al., 1995***). In the photorespiratory cycle, 2-PG is regenerated to 3-PGA, but it involves the release of formerly assimilated CO_2 and NH_3, entails energy costs for the plants and reduces the efficiency of photosynthesis by up to 30% (***Ehleringer et al., 1991***; ***Bauwe et al., 2010***; ***Raines, 2011***; ***Fernie et al., 2013***). Eight core enzymes are required for photorespiration, which in higher plants are located in the chloroplast, the peroxisome, and the mitochondrion (***Bauwe et al., 2010***; ***Figure 1A***). The pathway rescues ¾ of the carbon, which would otherwise be lost through the oxygenase activity of Rubisco (***Peterhansel et al., 2010***; ***Fernie et al., 2013***). Ammonia refixation in the chloroplast by the combined activities of glutamine synthase (GS) and glutamine oxoglutarate aminotransferase (GOGAT) is an integral part of photorespiration.

In hot and dry environments and under low atmospheric CO_2 conditions, when the oxygenation activity of Rubisco is increased, the high rate of photorespiration becomes unfavorable for the plants

eLife digest Environmental pressures sometimes cause different organisms to independently evolve the same traits. A dramatic example of this phenomenon, which is called convergent evolution, can be seen in the modes used by plants to convert carbon dioxide from the air into starch during photosynthesis.

Early plants existed in an environment with high levels of carbon dioxide in the air. Over time, carbon dioxide levels decreased, so plants evolved more efficient types of photosynthesis to cope. A very efficient type of photosynthesis, called C_4 photosynthesis essentially represents a carbon dioxide concentration mechanism. It has evolved at least 62 times independently in 19 different families of flowering plants.

Scientists have shown that a less advanced, low-efficiency version of photosynthetic carbon dioxide concentration, called C_2 photosynthesis, is a stepping-stone to C_4 photosynthesis. It is also known that the evolution of C_4 photosynthesis required changes to the expression patterns of thousands of genes, but the exact mechanism that leads from C_2 photosynthesis to C_4 photosynthesis is not clear.

To explore this in greater detail, Mallmann, Heckmann et al. studied plants from the genus *Flaveria*, which belongs to the same family as sunflowers and asters. Under identical greenhouse conditions, plants that use three different photosynthetic pathways—C_3 photosynthesis, C_4 photosynthesis, or an intermediate between the two—were grown and their gene expression patterns were compared. Computer simulations were used to model the metabolism of plants that relied on C_2 photosynthesis.

Based on the modeling, it appears that C_2 photosynthesis shifts the balance of nitrogen metabolism between two types of cell that are critical to photosynthesis. To rebalance the nitrogen, several genes are expressed to trigger an ammonia recycling mechanism. The same genes are turned on during C_4 photosynthesis, and this recycling mechanism include parts of the C_4 process.

The findings of Mallmann, Heckmann et al. suggest that the initial steps in C_4 photosynthesis evolved to prevent nitrogen imbalance. Over time, this mechanism was co-opted to become part of a more efficient form of photosynthesis, which may explain why so many different plants evolved from C_2 to C_4 photosynthesis.

DOI: 10.7554/eLife.02478.002

(***Sage, 2001***, ***2013***). C_4 plants possess a mechanism that minimizes the oxygenase function of Rubisco and thereby reduces photorespiration and decreases the loss of carbon. C_4 photosynthesis is based on a division of labor between two different cell types, mesophyll and bundle sheath cells, which are organized in a wreath-like structure called 'Kranz Anatomy' (***Haberlandt, 1904***; ***Dengler and Nelson, 1999***). Atmospheric CO_2 is initially fixed in the mesophyll by phosphoenolpyruvate carboxylase (PEPC), and the resulting four-carbon compound is transported to the bundle sheath cells and decarboxylated by NADP/NAD malic enzyme or phosphoenolpyruvate carboxykinase (***Hatch et al., 1975***). Thereby CO_2 is concentrated at the site of the Rubisco in the bundle sheath cells (***Hatch, 1987***), outcompeting the molecular oxygen. As a consequence, photorespiration is drastically reduced as compared to C_3 plants, and C_4 plants are characterized by a high photosynthetic efficiency (***Figure 1B***).

C_4 plants have evolved multiple times independently from C_3 ancestors. The evolution of C_4 photosynthesis occurred at least 62 times in 19 different families of the angiosperms (***Sage et al., 2011***), implying a low evolutionary barrier towards expression of this trait. The analysis of recent intermediate species (***Bauwe and Kolukisaoglu, 2003***; ***Sage, 2004***; ***Bauwe, 2011***; ***Sage et al., 2012***, ***2013***; ***Schulze et al., 2013***) indicates that establishing a photorespiratory CO_2 pump was an early and important step in the evolution towards C_4 photosynthesis (***Figure 1C***). Since the two-carbon compound glycine serves as a transport metabolite, this photorespiratory CO_2 concentrating mechanism is also termed C_2 photosynthesis. Computational modeling of the evolutionary trajectory from C_3 to C_4 photosynthesis indicated C_2 photosynthesis represented an evolutionary intermediate state (***Heckmann et al., 2013***; ***Williams et al., 2013***) as well suggesting that C_2 photosynthesis is a prerequisite for the evolution of C_4. However, it remained unclear if the evolution of C_2 photosynthesis fosters the evolution of C_4 photosynthesis beyond providing a selection pressure to reallocate Rubisco to the bundle sheath.

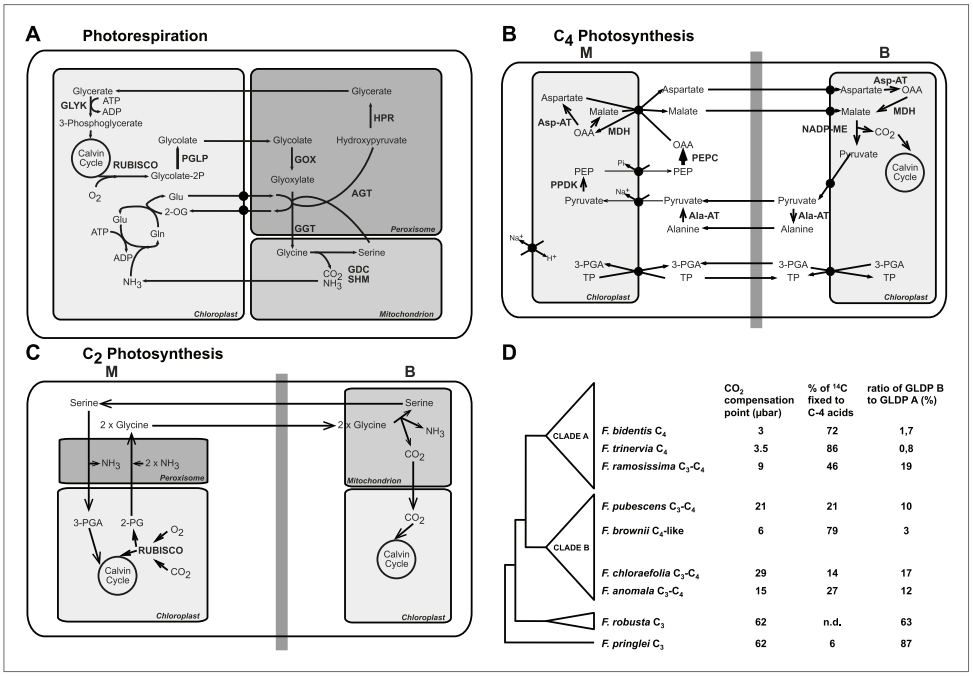

Figure 1. The genus *Flaveria* as a model organism to study C_4 evolution. Schematic view of the photorespiratory pathway (**A**), the NADP-ME type C_4 pathway as it can be found in C_4 Flaveria species (**B**) and the C_2 photosynthesis pathway (**C**). (**D**) Phylogeny and physiological properties of selected Flaveria species. The phylogeny was redrawn according to ***McKown et al. (2005)***, CO_2 compensation points are taken from ***Ku et al. (1991)***, incorporation of $^{14}CO_2$ is from ***Moore et al. (1987)*** and the ratios of GLDP B (expressed in all chlorenchyma cells) and GLDP A (expressed in bundle sheath cells only) are from ***Schulze et al. (2013)***. (Abbreviations: AGT: serine glyoxylate aminotransferase; AlaAT: alanine aminotransferase; AspAT: aspartate aminotransferase; GDC: glycine decarboxylase complex; GGT: glutamate, glyoxylate-aminotransferase; GLYK: D-glycerate 3-kinase; GOX: glycolate oxidase; HPR: hydroxypyruvate reductase; MDH: malate dehydrogenase; NADP-ME: NADP dependent malic enzyme; PEPC: phosphoenolpyruvate carboxylase; PGLP: 2-phosphoglycerate phosphatase; PPDK pyruvate, phosphate-dikinase; RUBISCO: Ribulose-1,5-bisphosphat-carboxylase/-oxygenase; SHM: serine hydroxymethyltransferase; 2-OG: oxoglutarate; 2-PG 2-phosphoglycolate; 3-PGA: 3-phosphoglycerate; Gln: glutamine; Glu: glutamate; OAA: oxaloacetate; PEP: phosphoenolpyruvate; TP: triosephosphate).
DOI: 10.7554/eLife.02478.003

In the present study, we have used the genus *Flaveria* as a model system for investigating the transition from C_2 to C_4 photosynthesis. To this end, we study a phylogenetic framework consisting of C_3, C_3–C_4 intermediate, and C_4 species (***Powell, 1978***; ***Edwards and Ku, 1987***; ***Ku et al., 1991***) of this genus which rather recently evolved C_4 (***Christin et al., 2011***), focusing on genes encoding photorespiratory enzymes and other components of C_2 photosynthesis. The genus *Flaveria* contains three main phylogenetic groups, of which the first diverging group includes all C_3 *Flaveria*. Clade B contains seven C_3–C_4 intermediate species and the C_4-like species *F. brownii*. All C_4 *Flaveria* species belong to clade A, which also contains several C_4-like species and the C_3–C_4 intermediate *F. ramosissima* (***McKown et al., 2005***; ***Figure 1D***). We hypothesized that the analysis of species in the genus *Flaveria* combined with *in silico* modeling elucidates the evolutionary changes accompanying and following the establishment of the C_2 pathway. To this end we simulated the metabolism of C_2 plants by coupling a mechanistic model of C_3–C_4 intermediate photosynthesis (***von Caemmerer, 2000***; ***Heckmann et al., 2013***)

with a detailed modified stoichiometric model of C_4 photosynthesis (***Dal'Molin et al., 2010***), and investigated the evolution of C_4 photosynthesis and photorespiration by following the changes in mRNA and protein abundance along the evolutionary path.

RNA and protein amounts of the majority of the photorespiratory enzymes were reduced in C_4 as compared to C_3 species. In contrast, photorespiratory mRNA and protein amounts did not decrease in the C_3–C_4 intermediate species but were mostly equal or even higher than in the C_3 species, demonstrating that the establishment of the photorespiratory CO_2 pump in the genus *Flaveria* relies on coordinated changes in the expression of all core photorespiratory enzymes. Metabolic modeling in combination with comparisons of transcript abundances in the different *Flaveria* species strongly indicates that introduction of C_2 photosynthesis has a direct impact on the nitrogen metabolism of the leaf. Its implementation necessitates the parallel establishment of components of the C_4 cycle to cope with these changes in refixation of photorespiratory nitrogen. Based on these results, we predict a mechanistic interaction between C_4 and C_2 photosynthesis.

Results

Selection of *Flaveria* species, cultivation of plant material and experimental design

To study the evolution of the expression of photorespiratory and C_4 cycle genes during the transition from C_3 to C_4 photosynthesis in the genus *Flaveria*, nine species reflecting the evolutionary trajectory taken were selected, including two C_3 (*F. robusta* and *F. pringlei*), two C_4 (*F. bidentis* and *F. trinervia*), and five C_3–C_4 intermediate species (***Figure 1D***). According to their CO_2 compensation points and the percentage of carbon initially fixed into malate and aspartate, *F. chloraefolia* and *F. pubescens* were earlier classified as type I C_3–C_4 intermediates. *F. anomala* and *F. ramosissima* belong to the type II C_3–C_4 intermediates and *F. brownii* is classified as a C_4-like species (***Edwards and Ku, 1987***; ***Moore et al., 1987***; ***Cheng et al., 1988***; ***Ku et al., 1991***). Type I C_3–C_4 intermediates are defined as solely relying on the photorespiratory CO_2 concentration cycle whereas a basal C_4 cycle activity is present in type II C_3–C_4 intermediates species. C_4-like species exhibit much higher C_4 cycle activities but lack complete bundle sheath compartmentation of Rubisco activity (***Edwards and Ku, 1987***).

Four independent experiments with plants grown during different seasons were performed to identify differences between the species that are dependent on their different modes of photosynthesis and independent of environmental influences. For each experiment the plants were seeded concurrently and grown side-by-side under greenhouse conditions. The second and fourth visible leaves from the top of all nine species were harvested at noon on the same day for transcript and protein analysis. Plants for experiment one were harvested in September 2009, for experiment two in June 2010, for experiment three in October 2010 and for experiment four in April 2011. The amounts of the core photorespiratory and C_4 enzymes were assessed by immunoblotting using specific antibodies raised against synthetic peptides or recombinant proteins. The abundances of the corresponding RNAs as well of C_4 cycle associated transcripts were quantified by total transcriptome sequencing.

The transcript profiles of the individual *Flaveria* species were comparable throughout all four experiments

The transcriptomes of the different *Flaveria* species were sequenced by Illumina technology following standard procedures. In total, close to 200 Gb of raw sequence data were produced. After filtering of low quality reads 30 to 58 million reads per species and experiment were quantified (***Figure 2—source data 1***). In a cross species approach, we mapped the sequences onto the minimal set of *Arabidopsis thaliana* coding sequences using the BLAST-like alignment tool BLAT (***Kent, 2002***) as described previously (***Gowik et al., 2011***) (***Figure 2—source data 2***, data available from the Dryad Digital Repository: 10.5061/dryad.q827h). We were able to align approx. 50% of our reads to the *Arabidopsis* transcripts. This is lower as compared to a similar approach using 454 sequencing (***Gowik et al., 2011***) and likely due to the shorter read length of the Illumina compared to the 454 reads. To overcome the low mapping efficiency, the leaf transcriptomes of *Flaveria* species were assembled de novo based on 454 (***Gowik et al., 2011***) and Illumina reads (this study). Among the contigs from *F. robusta*, we identified full-length transcripts for all photorespiratory and C_4 genes in the focus of the present study and used these for further read mapping and detailed analysis.

To evaluate the variation between the four independent experiments, we performed hierarchical sample clustering and a principal component analysis of the transcript profiles derived from read mapping on the minimal set of *Arabidopsis* coding sequences. Hierarchical sample clustering using Pearson correlation and average linkage clustering shows that the transcript profiles of all *Flaveria* species were quite similar in all four experiments since the samples cluster strictly species-wise (***Figure 2A***). The transcriptome patterns are influenced by the photosynthesis type and the phylogenetic relationships of the different species. The two C_4 species, both belonging to clade A, cluster together as do the two C_3 species that belong to the basal *Flaveria* species. Within the C_3–C_4 intermediates the two more advanced intermediates *F. ramosissima* and *F. anomala* cluster together, the only pattern which contradicts phylogenetic proximity since *F. ramosissima* belongs to clade A and *F. anomala* belongs to clade B. The last cluster consists out of the C_3–C_4 intermediates *F. chloraefolia* and *F. pubescens*, and the C_4-like species *F. brownii*.

Principle component analysis supports the results of the hierarchical clustering. The samples are mainly separated by photosynthesis type and phylogenetic relationships with the two intermediate species from different phylogenetic trajectories again forming a tight cluster (***Figure 2B***). The first three components, shown in ***Figure 2B***, explain only 27% of the total variance. This is in good accordance with earlier results where it was shown that about 16% of all analyzed genes showed photosynthesis type related expression changes when the transcriptomes of the C_4 species *F. trinervia* and *F. bidentis*, the C_3 species *F. robusta* and *F. pringlei* and the C_3–C_4 intermediate species *F. ramosissima* were compared (***Gowik et al., 2011***).

Amounts of photorespiratory transcripts and proteins indicate that the C_2 pathway was established early during C_4 evolution in *Flaveria* and is present also in the C_4-like species *F. brownii*

Photorespiratory genes are expressed in all species and photorespiratory proteins are detected in all species. To visualize the differences in transcript and protein abundance heat maps were plotted (***Figure 3***). The transcription of all photorespiratory genes except the transport proteins DIT1 and DIT2 and one isoform of GLDH was downregulated in the C_4 species *F. bidentis* and *F. trinervia* compared to the C_3 species *F. pringlei* und *F. robusta* (***Figure 3A***, ***Figure 3—source data 1***). Both dicarboxylate transporters play an important role in generating the transfer acids in the C_4 pathway of NADP-ME plants such as *F. trinervia* and *F. bidentis* (***Renne et al., 2003***; ***Gowik et al., 2011***; ***Kinoshita et al., 2011***). This may explain why their expression pattern is more similar to the C_4 genes than to the other photorespiratory genes.

The amounts of photorespiratory transcripts did not decrease gradually from C_3 to C_4 but the expression levels in the C_3–C_4 intermediate species *F. chloraefolia*, *F. pubescens*, *F. anomala* and *F. ramosissima* were mostly equal or higher than in the C_3 species. An exception are the transcripts of one GLDP, one GLDH and one SHM isoform which are drastically down-regulated also in the C_3–C_4 intermediate species. It was shown earlier that the down-regulation of this GLDP isoform is tightly associated with the establishment of the C_2 pathway in *Flaveria* (***Schulze et al., 2013***). The down-regulation of the GLDH and SHM isogenes might have similar reasons since both enzymes are also involved in glycine decarboxylation. Only the C_4-like species *F. brownii* is intermediate with respect to photorespiratory transcripts. 19 of 27 transcripts are reduced compared to the C_3–C_4 intermediate and C_3 species but have higher levels than the true C_4 species *F. bidentis* and *F. trinervia*. Exceptions are the components of the glycine decarboxylase complex as the respective transcripts levels are equal to these in the C_3 and C_3–C_4 intermediate species (***Figure 3A***).

The expression patterns described above were not only found for the genes encoding the core enzymes of photorespiration but also for the genes responsible for recycling of ammonia set free during photorespiration, GS/GOGAT. Also the genes of recently discovered transporters associated with photorespiration, PLGG1 and BOU (***Eisenhut et al., 2013***; ***Pick et al., 2013***), behave accordingly.

To test whether transcript abundance reflects protein abundance, amounts of core photorespiratory proteins in the leaves of all nine species were quantified by protein gel blots. To this end we generated antibodies against conserved peptides from *Flaveria* GLDP, GLDT, GLDL, SHM, HPR, PGLP and GLYK proteins. Total leaf proteins were extracted from plant material harvested together with the material used for RNA isolation and equal amounts of protein were separated via SDS gel-electrophoresis prior to blotting (***Figure 3—figure supplement 1***). The changes of protein amounts essentially reflected the changes of the amounts of the corresponding transcripts (***Figure 3B***,

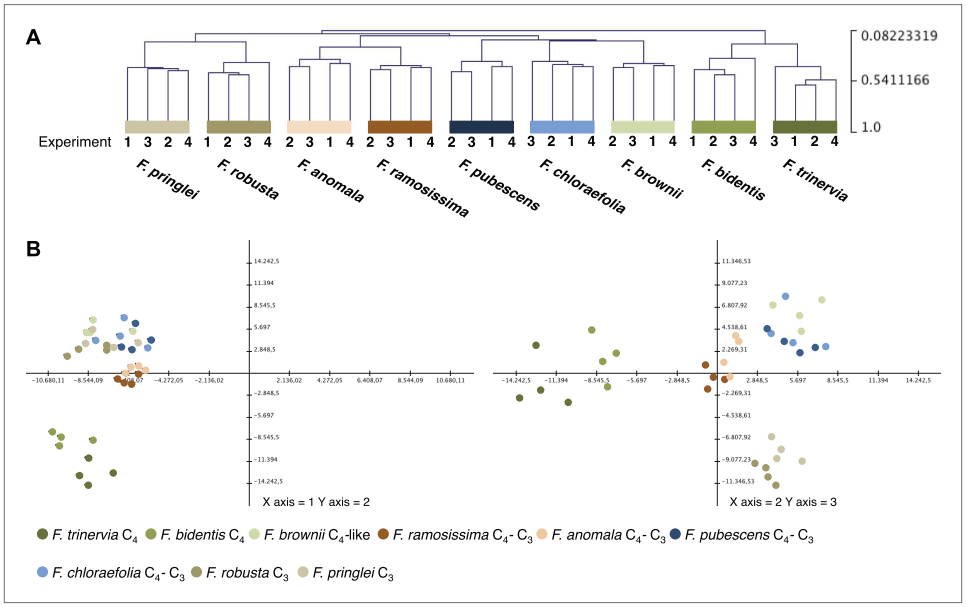

Figure 2. Variation of transcript profiles of the individual *Flaveria* species between the four experiments. (**A**) Hierarchical sample clustering of all expressed transcripts. The tree was calculated with the MEV program using the HCL module with Pearson correlation and the average linkage method. (**B**) Principal component analysis of transcript levels. The first three components explain 27% of the total variance.

DOI: 10.7554/eLife.02478.004

The following source data are available for figure 2:

Source data 1. Results of the Illumina sequencing and cross species read mapping.

DOI: 10.7554/eLife.02478.005

Source data 2. Quantitative information for all reads mapped in a cross species approach onto the reference transcriptome from *Arabidopsis thaliana*.

DOI: 10.7554/eLife.02478.006

Figure 3—figure supplement 2, Figure 3—source data 2). The amounts of core photorespiratory proteins in the C_3–C_4 intermediates were equal to the amounts in the C_3 species. A clear reduction of these proteins can be observed only for the true C_4 species and the C_4–like species. *F. brownii* exhibits intermediate amounts of most photorespiratory proteins. This indicates that the regulation of photorespiratory genes mainly occurs on the transcriptional level and that our approach to analyze the photorespiratory activity by comparative transcriptomics is reasonable.

While the overall patterns remain similar between all independent experiments, individual proteins and transcripts vary between the four experiments. This likely reflects adjustments of photorespiratory gene expression to the different light and temperature conditions in our green house in the different seasons of the year.

We conclude that the four experiments support the establishment of a photorespiratory C_2 cycle early during C_4 evolution in *Flaveria* and that this C_2 cycle was maintained until Rubisco activity was constricted to the bundle sheath cells in the true C_4 *Flaveria* species.

An integrated model of C_2 photosynthesis

While the principal physiological differences between C_3 and C_4 leaves are widely understood, knowledge about the metabolic reconfiguration required to implement a functional C_2 pathway into a C_3 leaf is

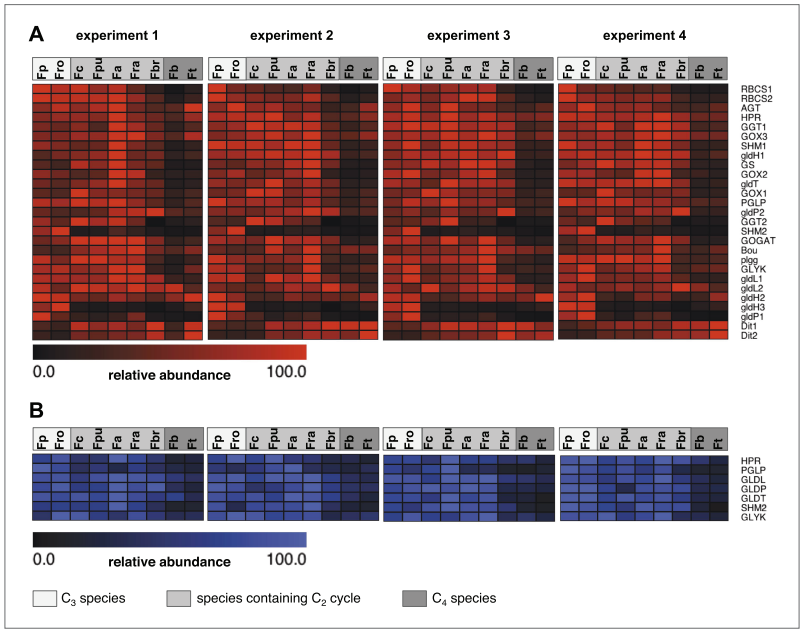

Figure 3. Abundance of photorespiratory transcripts and proteins in leaves of individual *Flaveria* species. Normalized transcript (**A**) and protein (**B**) amounts are plotted as heat maps. Transcript amounts were determined by Illumina sequencing of the leaf transcriptomes and read mapping on selected *F. robusta* full length transcript sequences. Protein amounts were determined by protein gel blots. See ***Figure 3—source data 1*** for absolute transcript levels, ***Figure 3—source data 2*** for protein quantification and ***Figure 3—figure supplements 1 and 2*** for immunoblots. Fp: *F. pringlei* (C_3); Fro: *F. robusta* (C_3); Fc: *F. chloraefolia* (C_3–C_4); Fpu: *F. pubescens* (C_3–C_4); Fa: *F. anomala* (C_3–C_4); Fra: *F. ramosissima* (C_3–C_4); Fbr: *F. brownii* (C_4-like); Fb: *F. bidentis* (C_4); Ft: *F. trinervia* (C_4).

DOI: 10.7554/eLife.02478.007

The following source data and figure supplements are available for figure 3:

Source data 1. Transcript abundance of photorespiratory genes determined by read mapping on *F. robusta* full length transcript sequences.

DOI: 10.7554/eLife.02478.008

Source data 2. Quantification of photorespiratory proteins by protein gel blot.

DOI: 10.7554/eLife.02478.009

Figure supplement 1. Results of the protein analyses.

DOI: 10.7554/eLife.02478.010

Figure supplement 2. Results of the protein analyses.

DOI: 10.7554/eLife.02478.011

incomplete. In particular, moving glycine from mesophyll to bundle sheath cells (***Hylton et al., 1988***; ***Morgan et al., 1993***) does not only translocate carbon, it also transports one nitrogen atom per two carbon atoms. Evidently, implementing the C_2 carbon pump requires balancing of metabolic routes to maintain homeostasis of both carbon and nitrogen metabolism (***Monson and Rawsthorne, 2000***). How this can be achieved is non-intuitive and it thus requires a systematic analysis by metabolic modeling. To this end, we simulated the leaf metabolism of a C_2 plant using an integrated model. We coupled a mechanistic model of C_3–C_4 intermediate photosynthesis (***von Caemmerer, 2000***; ***Heckmann et al., 2013***) with a modified genome-scale stoichiometric model of C_4 photosynthesis that was designed to describe the entire metabolic interactions of mesophyll and bundle sheath cells in C_4 leaves (***Dal'Molin et al., 2010***).

We used the mechanistic model to predict constraints for the stoichiometric model. It provided values for net CO_2 uptake, Rubisco carboxylation as well as oxygenation in mesophyll and bundle sheath, CO_2 leakage from the bundle sheath, PEPC activity in the mesophyll, activity of NADP-ME in the bundle sheath, plasmodesmatal flux of glycine and serine, and decarboxylation by the GDC. Given specific activities of the C_2 and C_4 cycles in the mechanistic model, we used flux balance analysis (FBA) to predict detailed flux distributions that follow biologically realistic optimality criteria (***Varma and Palsson, 1994***). We employed a maximization of leaf biomass production, followed by a minimization of the sum of absolute fluxes including transport processes. In the minimization of total flux, we allocated higher weights to plasmodesmatal fluxes in order to account for the trade-off between CO_2 leakage and diffusion of metabolites between the cells. This framework allows us to investigate the most parsimonious implementation of C_2 and C_4 cycles, given a hypothesis about which metabolites are suitable for plasmodesmatal transport.

The first outcome of simulating the photorespiratory CO_2 pump was that the establishment of the C_2 pathway has indeed a direct impact on the nitrogen metabolism of the leaf. It transports two molecules of glycine from the mesophyll to the bundle sheath, where one molecule each of serine, CO_2, and ammonium are produced. CO_2 is fixed by bundle sheath Rubisco and serine is transferred back to the mesophyll, where it is used for the regeneration of phosphoglycerate and photorespiratory glycine. This results in a net transport of CO_2 but also ammonia from the mesophyll to the bundle sheath. To create a noticeable CO_2 enrichment in the bundle sheath, the C_2 cycle must run with an appreciable capacity; indeed, the mechanistic model of C_3–C_4 intermediate photosynthesis predicted an oxygenation rate of Rubisco of about one third of its carboxylation rate. Running at such rates, the C_2 cycle will create a massive nitrogen imbalance between mesophyll and bundle sheath cells, as was also predicted earlier by ***Monson and Rawsthorne (2000)***. Within the stoichiometric model, the free diffusion of ammonia between the two cell types was not allowed, since ammonia is toxic and known to effectively uncouple electrochemical gradients (***Krogmann et al., 1959***). Thus, ammonia must be refixed in the bundle sheath cells and transferred back to the mesophyll in the form of amino acids. According to the intergrated model, ammonia is fixed by glutamine synthetase and glutamine oxoglutarate aminotransferase (GS/GOGAT) in the bundle sheath cells (***Figure 4***). Consistent with this prediction, we found that GS/GOGAT transcripts were upregulated in the C_3–C_4 intermediate species (***Figure 3***).

Estimating whether a certain metabolite is suitable for maintaining a diffusional gradient between mesophyll and bundle sheath is an unsolved problem. The impact on regulatory mechanisms and homeostasis of the C_3 leaf may render some metabolites unsuitable to serve as transport metabolites. We address this problem by modeling multiple scenarios that assume different transport metabolites.

If major amino acids and the corresponding oxoacids and dicarbonic acids are allowed to freely diffuse between cells in an integrated model representing a C_2 cycle, glutamate is predicted to be transferred to the mesophyll, where it is deaminated by GGT, regenerating the photorespiratory glycine. The resulting 2-oxoglutarate is transferred back to the bundle sheath cells (***Figure 4A***). The model preference for glutamate/2-oxoglutarate reflects the minimization of total flux in the FBA model, as this effectively minimizes the number of active enzymatic reactions and holds the plasmodesmatal flux for ammonia balance at one acceptor and one transport metabolite.

To elucidate if alternative solutions exist that contain more steps but retain the same biomass output, the 2-oxoglutarate transfer between mesophyll and bundle sheath was constrained to prevent the glutamate/2-oxoglutarate exchange. The integrated model then predicts an alanine/pyruvate shuttle (***Figure 4B***). The glutamate produced by GS/GOGAT activity in the bundle sheath cells is used by alanine aminotransferase (Ala-AT) to aminate pyruvate. The resulting alanine is transferred to the mesophyll and trans-aminated by Ala-AT resulting in pyruvate and glutamate. The glutamate is used to regenerate photorespiratory glycine and pyruvate is transferred back to the bundle sheath.

If alanine and pyruvate transfer are also constrained, the model predicts an aspartate/malate shuttle (***Figure 4C***). This includes the oxidation of malate in the bundle sheath. The resulting oxaloacetate (OAA) is aminated by aspartate aminotransferase (Asp-AT) and aspartate moves to the mesophyll. Here aspartate is trans-aminated by Asp-AT and malate is regenerated by reduction of the resulting OAA and transferred to the bundle sheath.

In all these scenarios, further increasing the weights on plasmodesmatal flux leads to transporter metabolites with increased N carrying capacity such as asparagine (***Figure 4—source data 1***).

In a restrictive scenario, all nitrogen containing compounds were excluded from plasmodesmatal transport, except for glycine and serine, which are used by the C_2 cycle itself. In this case, the model

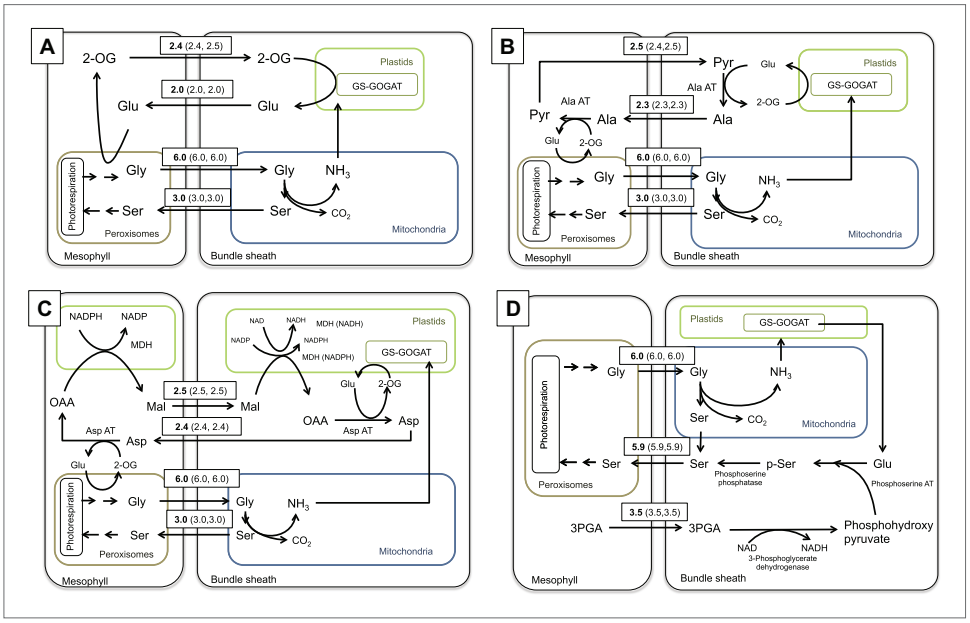

Figure 4. Flux Balance Analysis of the C_2 photosynthetic pathway. Predicted fluxes if (**A**) major amino acids and the corresponding oxoacids and dicarbonic acids are allowed to freely diffuse between cells, (**B**) the α-ketoglutarate and glutamate transfer between mesophyll and bundle sheath was constrained (**C**) additionally the transfer of alanine and pyruvate between mesophyll and bundle sheath was constrained (**D**) transfer of all nitrogen containing compounds except for glycine and serine, which are used by the C_2 cycle were constrained. Fluxes are given in μmol s^{-1} m^{-2}. Values in brackets show minimum and maximum of flux resulting from flux variability analysis. Flux of dissolved gasses, sucrose, inorganic compounds and processes that carry flux below 1 μmol s^{-1} m^{-2} are not shown. The sums of absolute fluxes over the plasmodesmata for the different variants were (**A**): 17.8 μmol s^{-1} m^{-2}; (**B**): 18.4 μmol s^{-1} m^{-2}; (**C**): 19.0 μmol s^{-1} m^{-2}; (**D**): 22.1 μmol s^{-1} m^{-2}. See ***Figure 4—source data 1*** for plasmodesmatal fluxes.
DOI: 10.7554/eLife.02478.012
The following source data are available for figure 4:

Source data 1. Fluxes over plasmodesmata depending on the weight on plasmodesmatal fluxes including flux variability analysis.
DOI: 10.7554/eLife.02478.013

predicts that bundle sheath derived ammonia is transferred from glutamate to phosphohydroxypyruvate by phosphoserine aminotransferase to yield phosphoserine; phosphoserine is then converted to serine by phosphoserine phosphatase. Finally, the serine moves to the mesophyll. This variant includes the transfer of 3-phosphoglycerate from the mesophyll to the bundle sheath, where it is converted to phosphohydroxy pyruvate by 3-phosphoglycerate dehydrogenase (***Figure 4D***).

The model predicts a mechanistic interaction between C_2 and C_4 cycle

In C_3 plants, basal activities of the typical C_4 cycle enzymes are present (***Aubry et al., 2011***). When our integrated model is parameterized to include an active C_4 cycle, it predicts that a contingent of the bundle sheath ammonia will be transferred to the mesophyll cells by the C_4 cycle as a biomass neutral alternative to the 2-OG/Glu shuttle or as the unique solution when additional weight on plasmodesmatal fluxes is applied (***Figure 5—source data 1***). In this solution malate is decarboxylated in the bundle sheath cells. CO_2 is refixed by Rubisco, and the resulting pyruvate is aminated by Ala-AT. Alanine moves to the mesophyll cells, where ammonia is fed into the photorespiratory cycle by Ala-AT

and GGT. The resulting pyruvate is converted back to malate by PPDK, PEPC, and NADPH-dependent MDH (***Figure 5A***). Flux variability analysis shows that only marginal variability in the fluxes of the shuttle is possible (***Figure 5—source data 1***). According to our model predictions, the cycle is active even at low PEPC activities, such as those measured in C_3 *Flaveria* species (***Gowik et al., 2011***; ***Heckmann et al., 2013***). When the C_4 cycle runs with low capacity, according the model, the surplus of bundle sheath ammonia is transferred back to the mesophyll by the glutamate/2-oxoglutarate shuttle. Once the capacity of the C_4 cycle gradually increases, the recirculation of nitrogen is shifted from the glutamate/2-oxoglutarate shuttle towards the C_4 cycle (***Figure 5B***). The predicted biomass production increases linearly with C_4 cycle activity (***Figure 5C***). Thus, our model predicts a strong interaction between C_2 and C_4 photosynthesis.

Analysis of C_4 cycle gene expression in C_3–C_4 intermediate *Flaveria* species

When the C_2 cycle is running with high capacity, our integrated modeling approach predicts the necessity of auxiliary metabolite fluxes between mesophyll and bundle sheath cells to prevent a massive nitrogen imbalance. Among those auxiliary fluxes were the pyruvate/alanine and the malate/aspartate exchanges. The metabolites used in these shuttles also serve as transport metabolites in C_4 photosynthesis. Furthermore, the model highlights the possibility that a low capacity C_4 cycle balances part of the C_2 cycle ammonia production. Therefore we analyzed in detail the expression of C_4 cycle related genes in our dataset. True C_4 *Flaverias*, such as *F. bidentis* or *F. trinervia*, are believed to use a NADP-ME type C_4 cycle (***Moore et al., 1984***; ***Ku et al., 1991***; ***Meister et al., 1996***; ***Gowik et al., 2011***). All genes associated with this type of C_4 photosynthesis are gradually upregulated in the analyzed C_3–C_4 intermediate species in line with their degree of 'C_4-ness'. This is true for the typical C_4 enzymes like PEPC, PPDK, MDH, NADP-ME, Ala-AT and a plastidic aspartate aminotransferase (Asp-AT), as well as for several C_4 associated transporters, such as the pyruvate transporter BASS2, the H^+/Na^+ exchanger NHD, the PEP translocator CUE1 and the putative malate and aspartate transporters DIT1 and DIT2 (***Weber and von Caemmerer, 2010***; ***Brautigam et al., 2011***; ***Furumoto et al., 2011***; ***Gowik et al., 2011***). The regulators of the C_4 enzymes (like PEPC kinase or the PPDK regulatory protein) and enzymes with auxiliary functions of C_4 enzymes (like pyrophosphatases or adenosinmonophosphatases) show a similar pattern (***Figure 6***, ***Figure 6—source data 1***). To corroborate the results of the transcript abundance measurements, selected C_4 cycle enzymes (PEPC, PPDK, and NADP-ME) were measured by immunoblotting. The protein abundance correlates well with the transcript abundance (***Figure 6***, ***Figure 6—source data 2***).

The expression changes of C_4 cycle genes do not all follow the same quantitative pattern (***Figure 6C***). Although all of these genes gradually increase in expression when plants gain C_4 properties, as judged, for example, by the percentage of $^{14}CO_2$ directly fixed into C_4 acids (***Vogan and Sage, 2011***), the quantitative changes in gene expression are quite different. PEPC and PPDK transcript amounts increase slowly in the C_3–C_4 intermediates *F. chloraefolia*, *F. pubescens*, and *F. anomala*, more steeply in the advanced C_3–C_4 intermediate *F. ramosissima* and the C_4-like species *F. brownii* before reaching the highest transcript abundances in the true C_4 species (***Figure 6C***). In contrast, NADP-ME and Ala-AT gene expression already increase in expression in the more C_3-like intermediate species. Their expression rises more linearly in the further advanced intermediates and plateaus in the C_4-like and C_4 species. If one uses the different *Flaveria* species as evolutionary proxies as suggested by the results of ***Heckmann et al. (2013)***, these results suggest that NADP-ME and Ala-AT are strongly upregulated earlier in evolution than other C_4 core enzymes like PEPC or PPDK.

F. chloraefolia is classified as a type I C_3–C_4 intermediate species, and no enhanced C_4 cycle activity should be present in this species based on the classification. We detected upregulation of all NADP-ME type associated C_4 genes, with some of the genes showing comparable small increases in expression (***Figure 6***). This is in line with the results of $^{14}CO_2$ uptake studies that indicate about 14% of CO_2 is directly incorporated into C_4 acids in *F. chloraefolia*, whereas only 6% goes into C_4 acids directly in the C_3 species *F. pringlei* (***Moore et al., 1987***). We think therefore that a basal C_4 cycle activity is present in *F. chloraefolia* and its classification as type I C_3–C_4 intermediate is questionable.

A gene encoding a mitochondrial NAD dependent malate dehydrogenase as well as several cytosolic and especially one mitochondrial Asp-AT were upregulated exclusively in the C_3–C_4 intermediate species and the C_4-like *F. brownii* (***Figure 6***). Often, high activities of these genes are associated with the NAD-ME or PEP-CK type of C_4 photosynthesis. NAD dependent malic enzyme and PEP

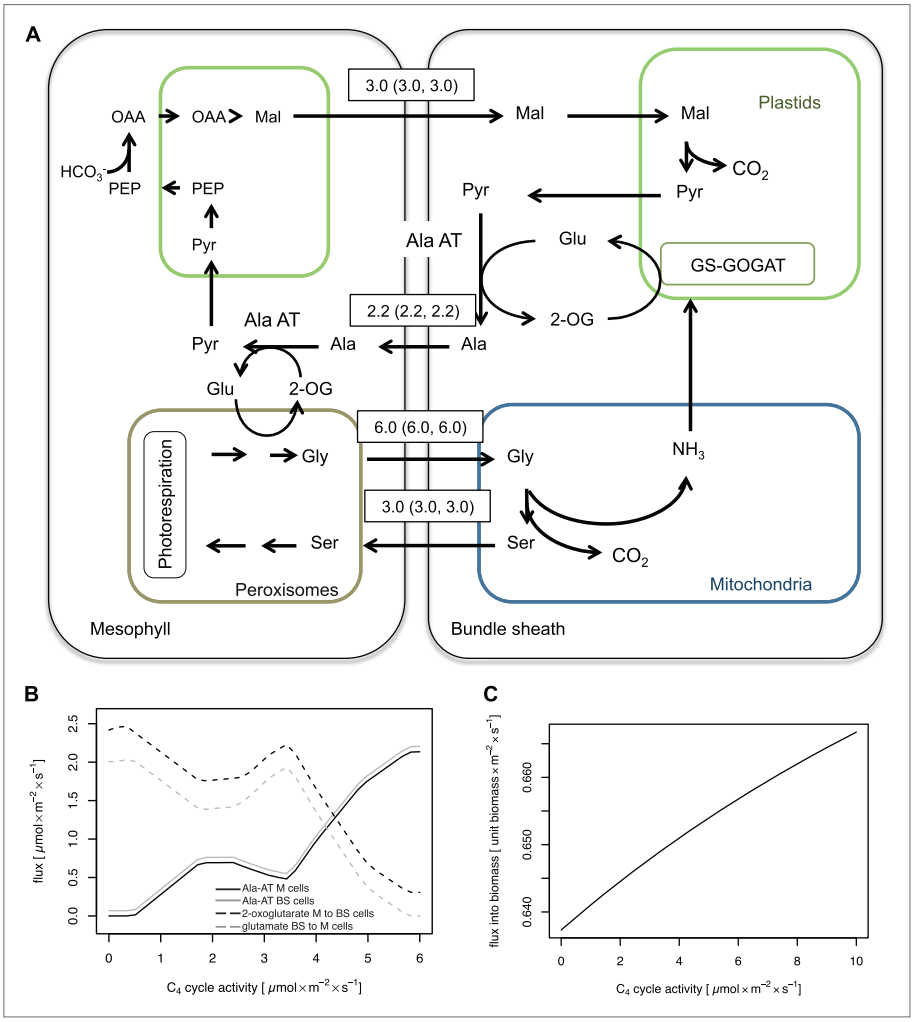

Figure 5. Mechanistic interaction between C_2 and C_4 cycle. (**A**) Predicted fluxes when the model is parameterized to include activity of the C_4 cycle enzymes. Fluxes are given in μmol s^{-1} m^{-2}. Values in brackets show minimum and maximum of flux resulting from flux variability analysis. The sum of absolute flux over plasmodesmata was 21.9 μmol s^{-1} m^{-2}. Flux of dissolved gasses, sucrose, inorganic compounds and processes that carry flux below 1 μmol s^{-1} m^{-2} are not shown. See ***Figure 5—source data 1*** for plasmodesmatal fluxes. (**B**) Predicted activities of Ala-AT in mesophyll (black line) and bundle sheath (gray line) cells and predicted transfer of α-ketoglutarate from mesophyll to bundle sheath cells (black dashed line) and glutamate from

Figure 5. Continued on next page

Figure 5. Continued

bundle sheath to mesophyll cells (gray dashed line) at low C_4 cycle activities. (**C**) Changes in biomass production with varying (low) activity of the C_4 cycle in a C_2 plant.

DOI: 10.7554/eLife.02478.014

The following source data are available for figure 5:

Source data 1. Fluxes over plasmodesmata depending on the weight on plasmodesmatal fluxes including flux variability analysis.

DOI: 10.7554/eLife.02478.015

carboxykinase genes were only very lowly expressed in all analyzed *Flaverias*, and no obvious differences between the C_3–C_4 intermediates and the other species could be found (***Figure 2—source data 2***, data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.q827h).

We found no transcriptomic evidence that ammonia is recirculated by the phosphoserine pathway predicted by the model that restricts the free diffusion of all amino acids except serine and glycine. The amounts of transcripts for all three enzymes of this pathway, i.e., phosphoserine aminotransferase, phosphoserine phosphatase, and 3-phosphoglycerate dehydrogenase, were found to be very low in all analyzed *Flaveria* species (***Figure 2—source data 2***, data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.q827h) (***Mallmann et al., 2014***).

Taken together, these data imply that the anaplerotic ammonia shuttle, required to maintain the nitrogen homeostasis in mesophyll and bundles sheath cells of plants performing C_2 photosynthesis, is active in all analyzed C_3–C_4 *Flaveria* species, as predicted by the computer simulations. Furthermore, it appears that even the most C_3-like C_3–C_4 intermediate species analyzed within the present study, *F. chloraefolia*, exhibits low level C_4 cycle activity. This activity is again in accordance with the *in silico* model, which predicts the C_4 cycle to be a highly efficient ammonia recirculation pathway.

Discussion

Photorespiration is mainly seen as a wasteful process, which arises from a malfunction of Rubisco and reduces photosynthetic efficiency (***Ogren, 1984***). In a high CO_2 atmosphere, Rubisco can operate efficiently. But the current atmospheric CO_2 concentration, combined with heat and drought, leads to an enhanced oxygenase activity and thereby the photosynthetic efficiency decreases (***Raines, 2011***). Up to 30% of the initially fixed CO_2 may be lost by photorespiration (***Bauwe et al., 2010***). C_4 plants avoid this problem by enriching CO_2 at the site of Rubisco. CO_2 is prefixed in the mesophyll and released in the bundle sheath cells, where Rubisco is operating (***Hatch, 1987***). The establishment of the photorespiratory CO_2 pump, which relocates the release of photorespiratory CO_2 to the bundle sheath cells, appears to be an important intermediate step towards the C_4 cycle and our detailed study of *Flaveria* intermediate species suggests that genes associated with C_4 photosynthesis also played a role in the C_2 cycle.

Implementation of the C_2 pathway leads to high expression of photorespiratory genes in C_3-C_4 intermediate *Flaveria* species

The expression of photorespiratory genes, including all genes encoding the core enzymes of the pathway, most of the transporters, and the enzymes involved in ammonia refixation, is not downregulated in the analyzed intermediate species; the transcript and protein amounts remain constant or in some cases are even higher compared to C_3 species. A significant drop in photorespiratory gene expression is only observed in the C_4-like species *F. brownii* and is decreased further in the C_4 species. Together with earlier results (***Schulze et al., 2013***), this indicates that indeed a C_2 photosynthetic cycle is active in all these C_3-C_4 intermediate *Flaveria* species and that a reduction in photorespiratory transcripts and proteins only occurs once the amounts of Rubisco have been reduced in the mesophyll as was described for the C_4-like species *F. brownii* (***Bauwe, 1984***; ***Holaday et al., 1988***). Rubisco reduction in the mesophyll is thus a late step of C_4 evolution, which in the *Flaveria* series appears to not occur gradually but rather abruptly towards the end of the evolutionary trajectory. It is followed by a strong increase of C_4 cycle activity, as can be deduced from the upregulation of PEPC and PPDK genes in the real C_4 species (***Figure 6C***), when the primary CO_2 fixation is completely taken over by PEPC. In the intermediate species C_2 and C_4 cycles operate in parallel leading to similar or higher photorespiratory gene expression compared with the C_3 species.

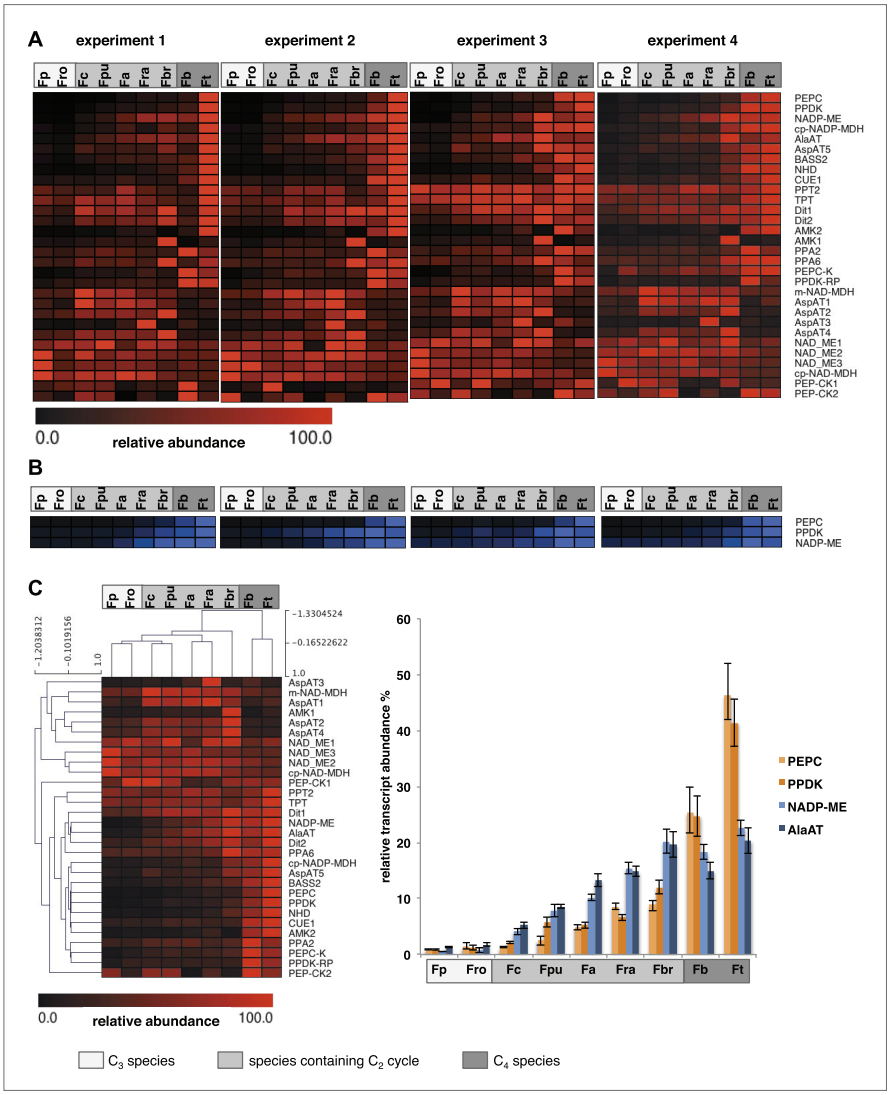

Figure 6. Abundance of C_4 related transcripts and proteins in leaves of individual Flaveria species. Normalized transcript (**A**) and protein (**B**) levels are plotted as heat maps. Transcript amounts were determined by Illumina sequencing of the leaf transcriptomes and read mapping on selected *F. robusta* full length transcript sequences. Protein amounts were determined by protein gel blots. See ***Figure 6—source data 2*** for absolute transcript level,

Figure 6. Continued on next page

Figure 6. Continued

Figure 6—source data 2 for protein quantification and ***Figure 3—figure supplement 1*** for immunoblots. (**C**) Mean values of transcript levels from all four experiments were clustered by hierarchical using the HCL module of MEV program with Pearson correlation and the average linkage method. The relative transcript abundance for PEPC, PPDK, NADP-ME and Ala-AT (mean values from all four experiments) are plotted for all nine species. Fp: *F. pringlei* (C_3); Fro: *F. robusta* (C_3); Fc: *F. chloraefolia* (C_3–C_4); Fpu: *F. pubescens* (C_3–C_4); Fa: *F. anomala* (C_3–C_4); Fra: *F. ramosissima* (C_3–C_4); Fbr: *F. brownii* (C_4–like); Fb: *F. bidentis* (C_4); Ft: *F. trinervia* (C_4).

DOI: 10.7554/eLife.02478.016

The following source data and figure supplements are available for figure 6:

Source data 1. Transcript abundance of C_4 cycle genes determined by read mapping on *F. robusta* full length transcript sequences.

DOI: 10.7554/eLife.02478.017

Source data 2. Quantification of C_4 proteins by protein gel blots.

DOI: 10.7554/eLife.02478.018

Figure supplement 1. Results of the protein analyses.

DOI: 10.7554/eLife.02478.019

Analysis of C_4 cycle gene expression supports the predictions of the C_2 model for C_3–C_4 intermediate *Flaveria* species and implies the early establishment of a complete C_4 pathway

The model of the C_2 cycle and the underlying metabolism proposes GS/GOGAT, Ala-AT, and Asp-AT to be involved in balancing the amino groups during C_2 cycle operation (***Figure 4***). The transcriptome data from the C_3-C_4 intermediate *Flaveria* species largely support the results of our integrated model for the C_2 pathway (***Figure 6***). In these species we found an upregulation of genes involved in the three most likely mechanisms for the recovery of ammonia predicted by the model. GS/GOGAT, which catalyzes the primary refixation of ammonia in the bundle sheath cells, is important for all three versions of ammonia shuttles (***Figure 4***) and is upregulated in the intermediate species. Transcripts for the glutamate/2-oxoglutarate shuttle, the alanine/pyruvate shuttle, and the aspartate/malate shuttle are enriched in all C_3–C_4 intermediates compared to the C_3 and C_4 *Flaverias*. For the alanine/pyruvate shuttle, Ala-AT is needed in the bundle sheath and the mesophyll cells. Ala-AT is upregulated already in the least advanced C_3–C_4 intermediates *F. chloraefolia* and *F. pubescens*, but also in all the other C_3–C_4 intermediates. Ala-AT transcripts are also highly abundant in the true C_4 *Flaverias* since Ala-AT is directly involved in the C_4 cycle when alanine is used as transport metabolite.

We found several Asp-AT and two MDH genes upregulated in the C_3–C_4 intermediate species (***Figure 6***). The chloroplast-located MDH and Asp-AT genes are involved in the C_4 cycle of C_4 *Flaverias*, in which malate and aspartate are used concurrently as C_4 transport metabolites (***Meister et al., 1996***). Two further Asp-AT genes and another MDH gene were found to be upregulated exclusively in the C_3–C_4 intermediates, including the C_4-like species *F. brownii*. The most likely reason for upregulation of these genes is their involvement in the recirculation of photorespiratory ammonia by a malate/aspartate shuttle.

The pathways of ammonia recirculation between mesophyll and bundle sheath foreshadow the establishment of a true C_4 cycle (***Figure 4***). All variants described above need the establishment of inter- and intra-cellular transport capacities for amino acids and small organic acids, which are also needed for a functional C_4 cycle (***Weber and von Caemmerer, 2010***). The existence of an aspartate/malate and an alanine/pyruvate shuttle anticipates important components of a functional C_4 pathway. Our transcript data imply that both of these shuttles are active in C_3–C_4 intermediate *Flaverias*. Only a few additions would be required to convert these pathways of ammonia recirculation into a C_4-like CO_2 concentration mechanism, that is, malate would have to be decarboxylated in the bundle sheath cells and pyruvate would have to be converted to malate in the mesophyll. Our transcript data implies that this conversion of the photorespiratory ammonia recirculation pathway into a C_4-like CO_2 concentrating pump must have been an early event in C_4 evolution of *Flaveria* since already in the least advanced intermediates such as *F. chloraefolia* and *F. pubescens*, NADP-ME transcripts are elevated and their amounts increase in parallel with Ala-AT and Asp-AT transcript levels.

To extend the pathways of ammonia recirculation into a rudimentary C_4 cycle, a capacity to regenerate malate from pyruvate in the mesophyll is required. As deduced from the transcriptome data, the enzymatic functions required are also already enhanced in the least advanced C_3–C_4 intermediates, since we observe a low but consistent upregulation of PEPC and PPDK genes in these species

compared to the C_3 *Flaverias*. Measurements of radiolabeled CO_2 incorporation support the view that a rudimentary C_4 cycle is already operating in intermediate *Flaveria* species (**Rumpho et al., 1984**; **Monson et al., 1986**; **Moore et al., 1987**; **Chastain and Chollet, 1989**). *F. chloraefolia* as well as *F. pubescens* incorporate a higher percentage of $^{14}CO_2$ into the C_4 compounds malate and aspartate (11.3% and 24.9%) than the C_3 species *F. pringlei* and *F. cronquistii* (4.1% and 7.7%) (**Vogan and Sage, 2011**). Thus even the least advanced intermediates analyzed in this study run already a low-level C_4 cycle, which assists in recycling the ammonia liberated by GDC in the bundle sheath cells.

The question arises whether amino group transfer initially exclusively happened via amino acid/oxoacid pairs or whether the enzymatic content of C_3 plants immediately supported a shuttle that also involved decarboxylation and carboxylation reactions. C_3 plants have considerable capacity for the decarboxylation of four-carbon organic acids in their bundle sheath cell (**Hibberd and Quick, 2002**; **Brown et al., 2010**) and measurements of total leaf NAD-ME and NADP-ME activity in C_3 plants repeatedly demonstrated basal activities for various C_3 species (**Wheeler et al., 2005**; **Aubry et al., 2011**; **Maier et al., 2011**). C_3 plants also accumulate high amounts of organic C_4 acids like malate or fumarate during the day (**Zell et al., 2010**), which are produced by PEPC, the only enzyme capable of producing C_4 acids de novo. It is tempting to hypothesize that plants use a malate/alanine shuttle to recycle parts of the ammonia liberated by glycine decarboxylation from the very beginning of the C_2 cycle.

Elevating the C_4 cycle activity in a C_2 plant enhances the CO_2 fixation capacity

If the C_4 cycle is superimposed onto a C_2 cycle operating in a C_3–C_4 intermediate plant, the C_2 photosynthesis model predicts a mechanistic interaction between the C_2 and C_4 cycles (**Figure 5**). When the C_4 cycle is running, the photorespiratory ammonia is recirculated from the bundle sheath to the mesophyll cells by moving malate from the mesophyll to the bundle sheath and transferring alanine back to the mesophyll. This malate/alanine cycling leads to a net transport of ammonia from the bundle sheath into the mesophyll cells. In contrast to the other mechanisms of ammonia recirculation described above, the C_4 cycle does not only lead to a net transport of ammonia from the bundle sheath to the mesophyll but additionally also to a net transport of CO_2 in the opposite direction. Thus CO_2 is transferred from the mesophyll to the bundle sheath without increasing the number of transport processes between the cells. By elevating the CO_2 concentration in the bundle sheath cells the C_4 cycle acts cooperatively with the C_2 cycle. The bundle sheath Rubisco would work under a more elevated CO_2 concentration and thus operate more effectively compared to a pure C_2 plant, leading to an increased biomass production. The C_4 cycle thus has a dual beneficial effect: an efficient nitrogen shuttle is combined with a CO_2 concentrating pump.

To investigate the possible interaction with regard to biomass, a C_4 cycle at the enzyme capacities of C_3 plants was allowed and tested for biomass changes (**Figure 5 C**). When the C_4 cycle is running with PEPC activities comparable to those found in C_3 *Flaveria* species, the model already predicts a gain in biomass production compared to the C_2 cycle on its own. Under these conditions, the bulk of photorespiratory ammonia is recycled through a rudimentary C_4 cycle limited by the C_4 cycle flux capacity. The model predicts that biomass production will be further enhanced with higher activity of the C_4 cycle. Consequently, there is permanent positive selection on enhancing the activity of the currently rate limiting enzyme once a C_4 cycle is running.

The evolutionary scenario described above is in good agreement with the *Flaveria* transcriptome data. We observe gradual increases in the amounts of C_4 transcript with increasing 'C_4-ness' of the C_3–C_4 intermediates until the most advanced species *F. brownii*. The abundance of NADP-ME and Ala-AT transcripts increases faster than the transcript abundance of the other core C_4 genes like PEPC, PPDK, MDH or Asp-AT. This implies that these evolutionary changes were driven by selection on high bundle sheath decarboxylation capacity, consistent with the idea that the C_4 cycle began as an auxiliary pathway to the C_2 cycle to recirculate photorespiratory ammonia. Hence, in this early phase, the main purpose of the C_4 cycle was to provide the ammonia acceptor pyruvate. The C_2 model and its evolutionary implications are consistent with the properties of the C_3–C_4 intermediate *Flaveria* species including *F. brownii*, which possess mesophyll Rubisco activity and consequently the C_2 photosynthetic pathway. The next iteration during C_4 evolution in *Flaveria* must have been the restriction of Rubisco activity to the bundle sheath, making the C_2 cycle obsolete, as observed for the true C_4 *Flaveria* species.

A scenario for C_4 evolution in the genus *Flaveria*—a general blueprint for the evolution of C_4 photosynthesis?

The establishment of a photorespiratory CO_2 pump, termed C_2 photosynthesis, is thought to be an important step in C_4 evolution. Recent work has shown how C_3 *Flaveria*s were preconditioned for the evolution of the C_2 pathway and how the C_2 cycle was implemented on the molecular level (***Sage et al., 2013***; ***Schulze et al., 2013***). Together with the present work, this gives us a detailed picture of what happened in the early and intermediate stages during C_4 evolution in *Flaveria*.

We have argued that the establishment of the C_2 cycle requires the implementation of at least components of the C_4 pathway, if not the whole pathway. This fact might be a partial explanation for the polyphyletic evolution of C_4 photosynthesis. Only the C_2 cycle has to evolve to set a system on a slippery slope towards C_4 photosynthesis. Nature seems to confirm this idea. So far, 66 independent origins of C_4 photosynthesis could be identified. In contrast, there are only seven known groups with independent origins of C_2 plants and no direct ancestry to C_4 species (***Sage et al., 2012***). If one assumes that all recent C_4 lineages evolved via C_2 intermediates, which appears likely (***Sage et al., 2012***; ***Heckmann et al., 2013***; ***Williams et al., 2013***), this would mean that the C_2 pathway evolved 73 times independently and that over 90% of these C_2 plant containing lineages proceeded to C_4 photosynthesis. This indicates that the C_2 photosynthetic pathway must indeed be a strong enabler of C_4 photosynthesis. It will be highly enlightening to analyze these C_2 groups without ancestry to C_4 species, like *Moricandia*, *Steinchisma* or *Mollugo*, to find out in how far they differ from groups that evolved the C_4 pathway and why C_4 evolution may have been hampered in these groups.

The close evolutionary interconnection of the C_2 and the C_4 pathway could be seen as an example of metabolic exaptation (***Barve and Wagner, 2013***). Exaptation or pre-adaptation was defined as an adaptation involving the co-option of traits that originally evolved for a different purpose (***Gould and Vrba, 1982***). While both C_2 and C_4 act as carbon shuttles to the bundle sheath cells, the two systems achieve this goal through different biochemical processes. In particular, the amino acid shuttle in the C_2 system evolved to transport nitrogen, and its later use in C_4 photosynthesis to shuttle carbon thus represents a molecular exaptation. Our findings therefore corroborate the general idea that the evolution of complex traits may be accelerated through exaptations (***Darwin, 1872***; ***Gould and Vrba, 1982***; ***Barve and Wagner, 2013***).

We do not know if the scenario on the early and intermediate stages of evolution described above is limited to the genus *Flaveria* or if it is valid for C_4 evolution in general. Our prediction of the C_2 pathway being a strong facilitator of C_4 evolution should apply to all C_4 origins, as the integrated model is not specific to *Flaveria*.

Materials and methods

Plant material

F. pringlei, *F. robusta*, *F. chloraefolia*, *F. pubescens*, *F. anomala*, *F. ramosissima*, *F. brownii*, *F. bidentis and F. trinervia* plants were grown in the green house at University of Duesseldorf side-by-side and harvested at four different points of time over the year. The plants were grown in 17-cm pots on soil (C-400 with Cocopor [Stender Erden, Schermbeck Germany] fertilized with 3 g/l Osmocote exact standard 3 to 4 M [Marysville, USA]) with additional light for 16 hr per day until 50 to 60 cm height and before the onset of flowering.

Plants for experiment one were harvested in September, for experiment two in June, for experiment three in October and for experiment four in April. The plant material was immediately frozen in liquid nitrogen, stored at −80°C and used for the following analyses.

RNA isolation, transcriptome sequencing and analysis

Total RNA was isolated from the second and fourth leaves according to (***Westhoff et al., 1991***) followed by a DNAse treatment. After phenol/chloroform extraction and precipitation with NaAc and isopropyl alcohol the RNA was dissolved in H_2O. The RNA quality was tested with the Agilent 2100 bioanalyzer. 1 µg of total RNA was used for cDNA library generation, which was accomplished with the TruSeq RNA Sample Preparation Kit (Illumina Inc., San Diego, USA) via the Low-Throughput Protocol (TruSeq RNA Sample Preparation Guide, Illumina Proprietary Catalog # RS-930-2001, Part # 15008136 Rev. A, November 2010). Clusters were generated with the TruSeq SR Cluster Kit v2 according to the

Reagent Preparation Guide with the Illumina cBot device. The single read sequencing was performed with the Illumina HiSeq2000.

Sequences of transcripts from genes involved in photorespiraton, C_4 photosynthesis and refixation and recirculation of photorespiratory ammonia were identified among de novo assembled transcripts of *F. robusta*. De novo assembly was performed with either CLC Genomics Workbench (CLC-Bio, Aaarhus, Denmark) or the Velvet/Oases software package (**Schulz et al., 2012**) using *F. robusta* 454 (**Gowik et al., 2011**) and Illumina reads (this study).

After quality control and processing, Illumina reads were aligned to the *F. robusta* transcript sequences with the CLC Genomics Workbench using standard parameters. Read mapping against a minimal set of coding sequences (**Brautigam et al., 2011**) of the TAIR 9 release of the *Arabidopsis thaliana* genome (http://www.Arabidopsis.org/) was performed using BLAT (**Kent, 2002**) as described in (**Gowik et al., 2011**).

The MEV software package (http://www.tm4.org/mev.html) was used for plotting heat maps, hierarchical clustering and principal component analysis.

Protein isolation and quantification

Total proteins were isolated from plant material harvested together with the material for RNA isolation according to **Shen et al. (2007)** and quantified using the RC-DC protocol (Bio-Rad Laboratories, Hercules, USA). 30 µg of total protein was electrophoresed on polyacrylamide-SDS gels (**Schägger and von Jagow, 1987**) and electrophoretically transferred to nitrocellulose membranes (Protran BA85, 0.45 μm; Schleicher & Schuell, Dassel, Germany) for 1 hr with 0.8 mA per cm^2. Specific primary antibodies were raised against conserved *Flaveria* peptides (Agrisera Vännäs, Sweden). For the detection of specific proteins the nitrocellulose membranes were incubated with the primary antibodies and a Horseradish peroxidase-conjugated secondary antibody (Sigma-Aldrich, St. Louis, USA). An enhanced chemiluminescent Horseradish peroxide substrate was added and signals were recorded using a Fuji LAS-4000 mini CCD camera system. The signals were quantified with the Multi Gage analysis software (Fujifilm, Tokyo, Japan). As loading control a gel was stained for 45 min with 0.25% Coomassie blue, 50% methanol, 7% acetic acid, and destained in 50% methanol, 7% acetic acid.

Coupling a mechanistic model with a genome-scale metabolic reconstruction

In order to model the metabolic integration of C_2 and C_4 cycle in the context of leaf metabolism, we conducted Flux Balance Analysis (FBA) based on a genome-scale metabolic reconstruction of C_4 metabolism, C4GEM (**Dal'Molin et al., 2010**). This reconstruction contains a complex biomass reaction including carbohydrates, cell wall components, amino acids and nucleotides (**Dal'Molin et al., 2010**).

FBA is a powerful tool to understand the adaptation of metabolism on a genomic scale. Since metabolite concentrations are not modeled explicitly, fluxes related to carbon concentration mechanisms (CCMs) cannot be captured by this constraint-based approach alone. To account for this issue, we coupled the FBA model with a mechanistic model of C_3–C_4 photosynthesis (**von Caemmerer, 2000**; **Heckmann et al., 2013**).

C4GEM representing NADP-ME types was provided by the authors and FBA was conducted using this model:

$$\text{Maximize } c^T v$$

subject to $Sv = 0$.

$$v_{min,i} \leq v_i \leq v_{max,i}$$

where c is the vector of coefficients in the objective function, here the leaf biomass production. v is the vector of fluxes through the network reactions, S is the stoichiometric matrix of the metabolic network, and v_{min} and v_{max} represent constraints on the respective fluxes.

In order to test hypotheses concerning nitrogen metabolism in C_3–C_4 intermediate plants, S had to be modified. The plasmosdesmatal transport reactions in the original C4GEM model include malate, pyruvate, 3-phosphoglycerate, trioses, phosphates, sucrose, aspartate, alanine, phosphoenolpyruvate, CO_2, and O_2. Reactions were added to S in order to include transport of serine, glycine, glutamate, glutamine, asparagine, threonine, 2-oxoglutarate and water over the mesophyll/bundle sheath interface. Furthermore, the lack of photosystem II in the bundle sheath of certain C_4 plants does not

hold in our scenario (**Nakamura et al., 2013**) and we added a reaction for linear electron transport to the bundle sheath. C4GEM does not contain a reaction for a plastidal NADP-dependent malate dehydrogenase in the bundle sheath; we added this reaction to *S*.

In addition to the stoichiometric matrix *S*, the constraints used in C4GEM were modified:

The original constraint on leaf sucrose production was changed to result in an output ratio of sucrose to amino acids of about 5 (**Riens et al., 1991**). Fixed constraints on production of starch and fatty acids are not appropriate in the coupled framework. Since we are not aware of data that explains how these fluxes scale with net CO_2 assimilation rate, the constraints were removed from the model. Reactions belonging to the GS/GOGAT system were assumed to be irreversible. Nitrogen is available in the form of nitrate as opposed to NH_3 in the original model. Since there is no evidence suggesting mesophyll specificity of PEPC in intermediate *Flaveria* species, we unconstrained PEPC flux in the bundle sheath.

To couple the genome-scale FBA model with the mechanistic model of carbon fixation, the following reactions were constrained using the values predicted by the mechanistic model: net CO_2 uptake, Rubisco carboxylation and oxygenation in mesophyll and bundle sheath, CO_2 leakage from the bundle sheath, PEPC activity in the mesophyll, activity of NADP-ME in the bundle sheath, plasmodesmatal flux of glycine and serine and decarboxylation by the GDC complex. The lower bound on glycine diffusion ($V_{min,Gly}$), serine diffusion ($V_{min,Ser}$), and GDC reaction ($V_{min,GDC}$) can be obtained from the rate of Rubisco oxygenation in the mesophyll (V_{om}) and the fraction of photorespiratory CO_2 in the bundle sheath derived from mesophyll oxygenations (ξ):

$$v_{min,Gly} = \xi V_{om},\ v_{min,Ser} = 0.5\xi V_{om},\ v_{min,GDC} = 0.5\xi V_{om}$$

The mechanistic model was parameterized to the C_3 state as given in **Heckmann et al. (2013)**, with the exception of the parameter ξ, which was set to a value of 0.98 (*i.e.*, the majority of GDC activity was restricted to the bundle sheath. Derivation from transcriptome data is given in **Heckmann et al. (2013)**). These constraints on the reactions of the photorespiratory pump are necessary to adequately predict C_2 photosynthesis because of the inability of FBA alone to model CCMs (see discussion above).

In the FBA part of the model, a minimization of total flux (MTF) analysis was conducted in order to narrow down the space of optimal solutions:

$$\text{Minimize} \sum_{i=1}^{n} w_i |v_i|$$

subject to: $Sv = 0$.

$$v_{min,i} \leq v_i \leq v_{max,i}$$

$$c^T v = c^T v_{FBA}$$

where v_{FBA} is the flux distribution of the FBA optimization described above. *w* denotes a vector of weights, where plasmodesmatal flux received a higher weighting factor (1.1 for plasmodesmatal exchange, 1 for the remaining reactions). This method implements a simple minimization of protein costs for a given optimal biomass production. The higher weights on plasmodesmatal fluxes account for the trade-off between CO_2 containment in the bundle sheath and metabolite diffusion between the cells. Since this trade-off is difficult to quantify, we conducted a sensitivity analysis by varying the weight on plasmodesmatal transport reactions.

In order to investigate the possible range that fluxes can take while yielding an optimal solution, flux variability analysis was conducted:

For each v_i:

Maximize or Minimize v_i.

Subject to: $Sv = 0$.

$$v_{min,i} \leq v_i \leq v_{max,i}$$

$$c^T v = c^T v_{FBA}$$

$$\sum_{i=1}^{n} w_i|v_i| = s_{opt}$$

Where s_{opt} is the minimum for the weighted sum of absolute flux found in the MTF optimization.

All simulations were conducted in the R environment for statistical computing (***R Core Team, 2013***) using the sybil library (***Gelius-Dietrich et al., 2013***).

Accession numbers

The read data have been submitted to the National Center for Biotechnology Information Short Read Archive under accession numbers SRP036880 (*F. bidentis*), SRP036881 (*F. anomala*), SRP036883 (*F. brownii*), SRP036884 (*F. chloraefolia*), SRP036885 (*F. pringlei*), SRP037526 (*F. pubescens*), SRP037527 (*F. ramosissima*), SRP037528 (*F. robusta*) and SRP037529 (*F. trinervia*).

Acknowledgements

We thank the 'Genomics and Transcriptomics laboratory' of the 'Biologisch-Medizinischen Forschungszentrum' (BMFZ) at the Heinrich-Heine-University of Duesseldorf (Germany) for technical support and conducting the Illumina sequencing.

Additional information

Funding

Funder	Grant reference number	Author
Deutsche Forschungsgemeinschaft	FOR 1186	Andreas PM Weber, Peter Westhoff
Deutsche Forschungsgemeinschaft	IRTG 1525	David Heckmann
Deutsche Forschungsgemeinschaft	CRC 680	Martin J Lercher
Deutsche Forschungsgemeinschaft	EXC 1028	Martin J Lercher, Andreas PM Weber, Peter Westhoff
European Union	EU Project 3to4	Andreas PM Weber, Peter Westhoff

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

JM, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; DH, In silico modeling, Conception and design, Analysis and interpretation of data, Drafting or revising the article; AB, UG, Conception and design, Analysis and interpretation of data, Drafting or revising the article; MJL, APMW, PW, Conception and design, Drafting or revising the article

Additional files

Major datasets

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset ID and/or URL	Database, license, and accessibility information
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria bidentis leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP036880	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria anomala leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP036881	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria brownii leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP036883	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria chloreafolia leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP036884	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria pringlei leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP036885	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria pubescens leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP037526	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria ramosissima leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP037527	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria robusta leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP037528	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Flaveria trinervia leaf transcriptomes	http://www.ncbi.nlm.nih.gov/sra/?term=SRP037529	Publicly available at NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra).
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U	2014	Data from: the role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria	http://dx.doi.org/10.5061/dryad.q827h	Available at Dryad Digital Repository under a CC0 Public Domain Dedication.

References

Anderson LE. 1971. Chloroplast and cytoplasmic enzymes. II. Pea leaf triose phosphate isomerases. *Biochimica et Biophysica Acta* **235**:237–244. doi: 10.1016/0005-2744(71)90051-9.

Aubry S, Brown NJ, Hibberd JM. 2011. The role of proteins in C(3) plants prior to their recruitment into the C(4) pathway. *Journal of Experimental Botany* **62**:3049–3059. doi: 10.1093/jxb/err012.

Barve A, Wagner A. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* **500**:203–206. doi: 10.1038/nature12301.

Bauwe H. 1984. Photosynthetic enzyme activities and immunofluorescence studies on the localization of ribulose-1, 5-bisphosphate carboxylase/oxygenase in leaves of C3, C4, and C3–C4 intermediate species of Flaveria (Asteraceae). *Biochemie und Physiologie der Pflanzen* **179**:253–268. doi: 10.1016/S0015-3796(84)80041-4.

Bauwe H. 2011. Photorespiration: the bridge to C4 photosynthesis. In: Raghavendra AS, Sage RF, editors. *C4 Photosynthesis and related CO2 concentrating mechanisms*. Dordrecht, Netherlands: Springer. p. 81–108.

Bauwe H, Hagemann M, Fernie AR. 2010. Photorespiration: players, partners and origin. *Trends in Plant Science* **15**:330–336. doi: 10.1016/j.tplants.2010.03.006.

Bauwe H, Kolukisaoglu Ü. 2003. Genetic manipulation of glycine decarboxylation. *Journal of Experimental Botany* **54**:1523–1535. doi: 10.1093/jxb/erg171.

Bowes G, Ogren WL, Hageman RH. 1971. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and Biophysical Research Communications* **45**:716–722. doi: 10.1016/0006-291X(71)90475-X.

Brautigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Mass J, Lercher MJ, Westhoff P, Hibberd JM, Weber AP. 2011. An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. *Plant Physiology* **155**:142–156. doi: 10.1104/pp.110.159442.

Brown NJ, Palmer BG, Stanley S, Hajaji H, Janacek SH, Astley HM, Parsley K, Kajala K, Quick WP, Trenkamp S, Fernie AR, Maurino VG, Hibberd JM. 2010. C4 acid decarboxylases required for C4 photosynthesis are active in

the mid-vein of the C3 species *Arabidopsis thaliana*, and are important in sugar and amino acid metabolism. *The Plant Journal* **61**:122–133. doi: 10.1111/j.1365-313X.2009.04040.x.
Chastain CJ, Chollet R. 1989. Interspecific variation in assimilation of (14)CO2 into C4 acids by leaves of C3, C4 and C3-C4 intermediate Flaveria species near the CO2 compensation concentration. *Planta* **179**:81–88. doi: 10.1007/BF00395774.
Cheng SH, Moore BD, Edwards GE, Ku MSB. 1988. Photosynthesis in Flaveria brownii, a C4-like species. Leaf anatomy, characteristics of CO2 exchange, compartmentation of photosynthetic enzymes, and metabolism of 14CO2. *Plant Physiology* **87**:867–873. doi: 10.1104/pp.87.4.867.
Christin PA, Osborne CP, Sage RF, Arakaki M, Edwards EJ. 2011. C4 eudicots are not younger than C4 monocots. *Journal of Experimental Botany* **62**:3171–3181. doi: 10.1093/jxb/err041.
Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. 2010. C4GEM, a genome-scale metabolic model to study C4 plant metabolism. *Plant Physiology* **154**:1871–1885. doi: 10.1104/pp.110.166488.
Darwin C. 1872. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray. 6th Edn.
Dengler NG, Nelson T. 1999. Leaf structure and development in C4 plants. In: Rowan FS, Russell KM, editors. *C4 plant biology*. San Diego: Academic Press. p. 133–172.
Edwards GE, Ku MS. 1987. Biochemistry of C3-C4 intermediates. In: Hatch MD, Boardman NK, editors. *The biochemistry of plants*. New York: Academic Press. p. 275–325.
Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW. 1991. Climate change and the evolution of C4 photosynthesis. *Trends in Ecology & Evolution* **6**:95–99. doi: 10.1016/0169-5347(91)90183-X.
Eisenhut M, Planchais S, Cabassa C, Guivarc'h A, Justin AM, Taconnat L, Renou JP, Linka M, Gagneul D, Timm S, Bauwe H, Carol P, Weber AP. 2013. Arabidopsis A BOUT DE SOUFFLE is a putative mitochondrial transporter involved in photorespiratory metabolism and is required for meristem growth at ambient CO(2) levels. *The Plant Journal* **73**:836–849. doi: 10.1111/tpj.12082.
Fernie AR, Bauwe H, Eisenhut M, Florian A, Hanson DT, Hagemann M, Keech O, Mielewczik M, Nikoloski Z, Peterhänsel C, Roje S, Sage R, Timm S, von Cammerer S, Weber APM, Westhoff P. 2013. Perspectives on plant photorespiratory metabolism. *Plant Biology* **15**:748–753. doi: 10.1111/j.1438-8677.2012.00693.x.
Furumoto T, Yamaguchi T, Ohshima-Ichie Y, Nakamura M, Tsuchida-Iwata Y, Shimamura M, Ohnishi J, Hata S, Gowik U, Westhoff P, Brautigam A, Weber AP, Izui K. 2011. A plastidial sodium-dependent pyruvate transporter. *Nature* **476**:472–475. doi: 10.1038/nature10250.
Gelius-Dietrich G, Desouki A, Fritzemeier C, Lercher M. 2013. sybil - efficient constraint-based modelling in R. *BMC Systems Biology* **7**:125. doi: 10.1186/1752-0509-7-125.
Gould SJ, Vrba ES. 1982. Exaptation-a missing term in the science of form. *Paleobiology* **8**:4–15.
Gowik U, Brautigam A, Weber KL, Weber APM, Westhoff P. 2011. Evolution of C4 photosynthesis in the genus Flaveria: how many and which genes does it take to make C4? *The Plant Cell* **23**:2087–2105. doi: 10.1105/tpc.111.086264.
Haberlandt G. 1904. *Physiologische Pflanzenanatomie*. Leipzig, Germany: Verlag von Wilhelm Engelmann. 3rd edn.
Hatch MD. 1987. C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta* **895**:81–106. doi: 10.1016/S0304-4173(87)80009-5.
Hatch M, Kagawa T, Craig S. 1975. Subdivision of C4-pathway species based on differing C4 acid decarboxylating systems and ultrastructural features. *Functional Plant Biology* **2**:111–128. doi: 10.1071/PP9750111.
Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ. 2013. Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount fuji fitness landscape. *Cell* **153**:1579–1588. doi: 10.1016/j.cell.2013.04.058.
Hibberd JM, Quick WP. 2002. Characteristics of C4 photosynthesis in stems and petioles of C3 flowering plants. *Nature* **415**:451–454. doi: 10.1038/415451a.
Holaday AS, Brown RH, Bartlett JM, Sandlin EA, Jackson RC. 1988. Enzymic and photosynthetic characteristics of reciprocal F(1) hybrids of Flaveria pringlei (C(3)) and Flaveria brownii (C(4)-Like species). *Plant Physiology* **87**:484–490. doi: 10.1104/pp.87.2.484.
Hylton C, Rawsthorne S, Smith A, Jones DA, Woolhouse H. 1988. Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C3,âiC4 intermediate species. *Planta* **175**:452–459. doi: 10.1007/bf00393064.
Kent WJ. 2002. BLAT–the BLAST-like alignment tool. *Genome Research* **12**:656–664. doi: 10.1101/gr.229202.
Kinoshita H, Nagasaki J, Yoshikawa N, Yamamoto A, Takito S, Kawasaki M, Sugiyama T, Miyake H, Weber AP, Taniguchi M. 2011. The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism. *The Plant Journal* **65**:15–26. doi: 10.1111/j.1365-313X.2010.04397.x.
Krogmann DW, Jagendorf AT, Avron M. 1959. Uncouplers of spinach chloroplast photosynthetic Phosphorylation. *Plant Physiology* **34**:272–277. doi: 10.1104/pp.34.3.272.
Ku MSB, Wu J, Dai Z, Scott RA, Chu C, Edwards GE. 1991. Photosynthetic and photorespiratory characteristics of Flaveria species. *Plant Physiology* **96**:518–528. doi: 10.1104/pp.96.2.518.
Leegood RC, Lea PJ, Adcock MD, Häusler RE. 1995. The regulation and control of photorespiration. *Journal of Experimental Botany* **46**:1397–1414. doi: 10.1093/jxb/46.special_issue.1397.
Maier A, Zell MB, Maurino VG. 2011. Malate decarboxylases: evolution and roles of NAD (P)-ME isoforms in species performing C4 and C3 photosynthesis. *Journal of Experimental Botany* **62**:3061–3069. doi: 10.1093/jxb/err024.
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U. 2014. Data from: the role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. *Dryad Digital Repository* doi: 10.5061/dryad.q827h.

McKown AD, Moncalvo JM, Dengler NG. 2005. Phylogeny of Flaveria (Asteraceae) and inference of C4 photosynthesis evolution. *American Journal of Botany* **92**:1911–1928. doi: 10.3732/ajb.92.11.1911.
Meister M, Agostino A, Hatch MD. 1996. The roles of malate and aspartate in C4 photosynthetic metabolism of Flaveria bidentis (L.). *Planta* **199**:262–269. doi: 10.1007/BF00196567.
Monson R, Rawsthorne S. 2000. CO2 assimilation in C3-C4 intermediate plants. In: Leegood R, Sharkey T, Caemmerer S, editors. *Photosynthesis*. Netherlands: Springer. **Vol 9**. p. 533–550.
Monson RK, Moore BD, Ku MS, Edwards GE. 1986. Co-function of C3-and C4-photosynthetic pathways in C3, C 4 and C3-C4 intermediate Flaveria species. *Planta* **168**:493–502. doi: 10.1007/BF00392268.
Moore B, Ku M, Edwards G. 1987. C4 photosynthesis and light-dependent accumulation of inorganic carbon in leaves of C3–C4 and C4 Flaveria species. *Australian Journal of Plant Physiology* **14**:657–668.
Moore BD, Ku MS, Edwards GE. 1984. Isolation of leaf bundle sheath protoplasts from C4 dicot species and intracellular localization of selected enzymes. *Plant Science Letters* **35**:127–138. doi: 10.1016/0304-4211(84)90186-X.
Morgan CL, Turner SR, Rawsthorne S. 1993. Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and of serine hydroxymethyltransferase in leaves of C3-C4 intermediate species from different genera. *Planta* **190**:468–473. doi: 10.1007/bf00224785.
Nakamura N, Iwano M, Havaux M, Yokota A, Munekage YN. 2013. Promotion of cyclic electron transport around photosystem I during the evolution of NADP–malic enzyme-type C4 photosynthesis in the genus Flaveria. *New Phytologist* **199**:832–842. doi: 10.1111/nph.12296.
Ogren WL. 1984. Photorespiration: pathways, regulation, and Modification. *Annual Review of Plant Physiology* **35**:415–442. doi: 10.1146/annurev.pp.35.060184.002215.
Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kürkcüoglu S, Kreuzaler F. 2010. Photorespiration. *The Arabidopsis Book* **8**:e0130. doi: 10.1199/tab.0130.
Pick TR, Brautigam A, Schulz MA, Obata T, Fernie AR, Weber AP. 2013. PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proceedings of the National Academy of Sciences of the United States of America* **110**:3185–3190. doi: 10.1073/pnas.1215142110.
Powell M. 1978. Systematics of Flaveria (Flaveriinae-Asteraceae). *Annals of the Missouri Botanical Garden* **65**:590–636. doi: 10.2307/2398862.
R Core Team. 2013. R: a language and environment for statistical computing. *R foundation for statistical computing*, Vienna, Austria. URL http://www.R-project.org/.
Raines CA. 2011. Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant Physiology* **155**:36–42. doi: 10.1104/pp.110.168559.
Renne P, Dressen U, Hebbeker U, Hille D, Flugge UI, Westhoff P, Weber AP. 2003. The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2. *The Plant Journal* **35**:316–331. doi: 10.1046/j.1365-313X.2003.01806.x.
Riens B, Lohaus G, Heineke D, Heldt HW. 1991. Amino acid and sucrose content determined in the cytosolic, chloroplastic, and vacuolar compartments and in the Phloem sap of spinach leaves. *Plant Physiology* **97**:227–233. doi: 10.1104/pp.97.1.227.
Rumpho ME, Ku MS, Cheng SH, Edwards GE. 1984. Photosynthetic characteristics of C(3)-C(4) intermediate Flaveria species : III. Reduction of photorespiration by a limited C(4) pathway of photosynthesis in Flaveria ramosissima. *Plant Physiology* **75**:993–996. doi: 10.1104/pp.75.4.993.
Sage RF. 2001. Environmental and evolutionary preconditionsfor the origin and diversification of the C4 photosyntheticsyndrome. *Plant Biology* **3**:202–213. doi: 10.1055/s-2001-15206.
Sage RF. 2004. The evolution of C4 photosynthesis. *New Phytologist* **161**:341–370. doi: 10.1111/j.1469-8137.2004.00974.x.
Sage RF. 2013. Photorespiratory compensation: a driver for biological diversity. *Plant Biology* **15**:624–638. doi: 10.1111/plb.12024.
Sage RF, Christin P-A, Edwards EJ. 2011. The C4 plant lineages of planet Earth. *Journal of Experimental Botany* **62**:3155–3169. doi: 10.1093/jxb/err048.
Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the evolution of C4 Photosynthesis. *Annual Review of Plant Biology* **63**:19–47. doi: 10.1146/annurev-arplant-042811-105511.
Sage TL, Busch FA, Johnson DC, Friesen PC, Stinson CR, Stata M, Sultmanis S, Rahman BA, Rawsthorne S, Sage RF. 2013. Initial events during the evolution of C4 photosynthesis in C3 species of Flaveria. *Plant Physiology* **163**:1266–1276. doi: 10.1104/pp.113.221119.
Schägger H, von Jagow G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry* **166**:368–379. doi: 10.1016/0003-2697(87)90587-2.
Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**:1086–1092. doi: 10.1093/bioinformatics/bts094.
Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P. 2013. Evolution of C4 photosynthesis in the genus flaveria: establishment of a photorespiratory CO2 pump. *The Plant Cell* **25**:2522–2535. doi: 10.1105/tpc.113.114520.
Shen Y, Khanna R, Carle CM, Quail PH. 2007. Phytochrome induces rapid PIF5 phosphorylation and degradation in response to red-light activation. *Plant Physiology* **145**:1043–1051. doi: 10.1104/pp.107.105601.
Varma A, Palsson BO. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology* **60**:3724–3731.

Vogan PJ, Sage RF. 2011. Water-use efficiency and nitrogen-use efficiency of C3-C4 intermediate species of Flaveria Juss. (Asteraceae). *Plant, Cell & Environment* **34**:1415–1430. doi: 10.1111/j.1365-3040.2011.02340.x.
von Caemmerer S. 2000. *Biochemical models of leaf photosynthesis*. Collingwood, Australia: Csiro Publishing.
Weber AP, von Caemmerer S. 2010. Plastid transport and metabolism of C3 and C4 plants–comparative analysis and possible biotechnological exploitation. *Current Opinion in Plant Biology* **13**:257–265. doi: 10.1016/j.pbi.2010.01.007.
Westhoff P, Offermann-Steinhard K, Höfer M, Eskins K, Oswald A, Streubel M. 1991. Differential accumulation of plastid transcripts encoding photosystem II components in the mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C4 plants. *Planta* **184**:377–388. doi: 10.1007/bf00195340.
Wheeler MCG, Tronconi MA, Drincovich MF, Andreo CS, Flügge U-I, Maurino VG. 2005. A comprehensive analysis of the NADP-malic enzyme gene family of Arabidopsis. *Plant Physiology* **139**:39–51. doi: 10.1104/pp.105.065953.
Williams BP, Johnston IG, Covshoff S, Hibberd JM. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *eLife* **2**:e00961. doi: 10.7554/eLife.00961.
Zell MB, Fahnenstich H, Maier A, Saigo M, Voznesenskaya EV, Edwards GE, Andreo C, Schleifenbaum F, Zell C, Drincovich MF, Maurino VG. 2010. Analysis of Arabidopsis with highly reduced levels of malate and fumarate sheds light on the role of these organic acids as storage carbon molecules. *Plant Physiology* **152**:1251–1262. doi: 10.1104/pp.109.151795.

7.2.1 Contributions

Computational modelling Conception and design: DH, AB, UG; Computer modelling: DH; Analysis and interpretation of data: DH, JM, AB, UG; Drafting or revising the article: DH, JM, AB, MJL, APMW, PW, UG.

Experimental part Conception and design: JM, AB, APMW, PW, UG; Acquisition of data: JM; Analysis and interpretation of data: JM, DH, AB, UG; Drafting or revising the article: JM, DH, AB, MJL, APMW, PW, UG.

7.2.2 Outlook

In *Manuscript 2*, *in silico* modelling, transcript and protein data are presented to describe a novel model of how nitrogen shuttling that supports the photorespiratory pump can serve as a preadaptation for the C_4 cycle in *Flaveria*.

To test whether the proposed mechanism is a general phenomenon, future analysis will have to include additional genera that contain C_3–C_4 intermediates like *Cleome*, *Mollugo*, *Moricandia*, and *Panicum* (*Steinchisma*). This could improve our understanding of the similarity of evolutionary trajectories and the mechanisms of convergent evolution.

The close phylogenetic relationship of C_2 and C_4 species supports the model presented in *Manuscript 2*, but C_4 species tend to have a more distinct isotopic signature, making the identification of C_4 species feasible on a large scale (Smith and Epstein, 1971). Identification of C_2 species is experimentally more complex (Kennedy and Laetsch, 1974; von Caemmerer, 1992; Sage et al., 2012), and phylogenetic sampling might be biased towards species related to C_4 groups. Unbiased phylogenetic screening for C_2 species would help to understand the extent to which C_2 photosynthesis is only a transient state that continues to evolve towards the C_4 syndrome.

Furthermore, while comparative studies of transcript and protein abundances can serve to elucidate metabolic functions, a direct measure of flux will have to be determined using isotope labeling techniques. This approach would reveal the nature of the nitrogen cycle and its diversity in the different independently evolved forms of C_2 photosynthesis.

References

Kennedy, R. A. and Laetsch, W. M. (1974). Plant species intermediate for C_3, C_4 photosynthesis. *Science*, 184(4141):1087–1089.

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*, 63(1):19–47.

Smith, B. N. and Epstein, S. (1971). Two categories of $^{13}C/^{12}C$ ratios for higher plants. *Plant Physiology*, 47(3):380–384.

von Caemmerer, S. (1992). Carbon isotope discrimination in C_3–C_4 intermediates. *Plant, Cell & Environment*, 15(9):1063–1072.

8 Acknowledgements

Ich möchte allen herzlich danken die zu dieser Arbeit direkt oder indirekt beigetragen haben / Sincere thanks to everyone who contributed to this work directly or indirectly:

Prof. Martin Lercher für die Initiierung dieses spannenden Projektes, die Möglichkeit daran zu forschen, die hervorragende Betreuung, die Fähigkeit Biologie und speziell Evolution zu abstrahieren, das Vertrauen in meine Fähigkeiten und den Erfolg des Projektes, sowie mindestens tausend Tassen Kaffee.

Prof. Andreas Weber für die Initiierung dieses spannenden Projektes, seine hervorragenden Ideen zum biologisch/biochemischen Teil, die Möglichkeit an der MSU zu arbeiten, sowie die Austragung der iGRAD-Plant Retreats in führenden deutschen Weinanbaugebieten.

Prof. Yair Shachar-Hill for hosting my visit at MSU; through our discussions I learned a lot about handling omics data, how to think about metabolism on a systems level, and how to make a good Margharita.

Der gesamten Bioinformatik für eine tolle Zeit; Dr. Christian Esser für seine Hilfe bei tausend Kleinigkeiten, sowie dass er mich stets auf dem neuesten Stand im Bereich Filmkunstkinos und Einzelhandels- und Elektronikversandangebote gehalten hat; Janina Maß für Wortspiele durch die Blume.

Al Bay for being the best host you can hope for when going abroad.

Rebecca Roussey, Rahul Deshpande, Nathan Prashan, Lisa Carey, Tibo Tagleb and everyone at the Shachar-Hill Lab for all the support and a great time.

Prof. Barb Sears from MSU for giving me a great place to stay, many equally great dinners, and letting me drive her amazing truck.

Ali Denton for keeping my time in the wet lab short and successful.

Dr. Udo Gowik und Dr. Andrea Bräutigam für viele produktive Diskussionen zur C_4-Evolution.

Dr. Sigrun Wegener-Feldbrügge dafür, dass sie alles Organisatorische sehr erleichtert hat.

Dr. Erwin Clef, Dr. Gabriel Gelius-Dietrich, Prof. Gerhard Steger und Dr. Birger Marin dafür, dass sie gute Lehrer sind.

Rajen Piernikarczyk, Julia Kleinmanns, Dr. Christian Esser und Jonathan Fritzemeier für die hilfreichen Korrekturen zu dieser Arbeit.

Pia van Ackern, Lea Kuhl und Viviana Correa dafür, dass ich mich selbst als Exil-Kölner in Düsseldorf zuhause fühlen konnte.

Meinen Eltern und meiner Familie für unerschöpfliche Unterstützung und Rückhalt.

Julia Kleinmanns für ihre liebevolle Unterstützung, Geduld, Verständnis, und dass sie sich mindestens zwanzig Vorträge zur C_4-Evolution anhören musste.

Der DFG und dem iGRAD Plant für das Stipendium.