Metabolic Network Constrains Gene Regulation of C₄ Photosynthesis: The Case of Maize

Semidán Robaina-Estévez and Zoran Nikoloski

Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, D-14476 Potsdam-Golm, Germany

*Corresponding author: E-mail: nikoloski@mpimp-golm.mpg.de; Fax, +49-331-567-8136.

(Received September 2, 2015; Accepted February 9, 2016)

Engineering C₃ plants to increase their efficiency of carbon fixation as well as of nitrogen and water use simultaneously may be facilitated by understanding the mechanisms that underpin the C₄ syndrome. Existing experimental studies have indicated that the emergence of the C₄ syndrome requires co-ordination between several levels of cellular organization, from gene regulation to metabolism, across two co-operating cell systems—mesophyll and bundle sheath cells. Yet, determining the extent to which the structure of the C₄ plant metabolic network may constrain gene expression remains unclear, although it will provide an important consideration in engineering C₄ photosynthesis in C₃ plants. Here, we utilize flux coupling analysis with the second-generation maize metabolic models to investigate the correspondence between metabolic network structure and transcriptomic phenotypes along the maize leaf gradient. The examined scenarios with publically available data from independent experiments indicate that the transcriptomic programs of the two cell types are co-ordinated, quantitatively and qualitatively, due to the presence of coupled metabolic reactions in specific metabolic pathways. Taken together, our study demonstrates that precise quantitative coupling will have to be achieved in order to ensure a successfully engineered transition from C₃ to C₄ crops.

Keywords: C₄ metabolism • Gene regulation • Metabolic networks • Systems biology.

Abbreviations: BS, bundle sheath; M, mesophyll; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PEPCK, PEP carboxykinase; 2PG, 2-phosphoglycerate; 3-PGA, 3-phosphoglycerate; Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase.

Introduction

The C₄ syndrome includes the combination of biochemical, physiological and anatomical traits that underpin C₄ carbon fixation (i.e. photosynthesis) (Laetsch 1974). C₄ photosynthesis has evolved from C₃ photosynthesis as an adaptation to high light intensities, high temperatures and dryness (Cowik and Westhoff 2011, Sage et al. 2012). Plants fix carbon dioxide (CO₂) by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), catalyzing the carboxylation of ribulose-1,5-bisphosphate (RuBP) to two molecules of 3-phosphoglycerate (3-PGA). In C₃ plants, this is the only carboxylation event during carbon fixation. However, in C₄ plants, carbon is first fixed by transforming CO₂ and phosphoenolpyruvate (PEP) into a four-carbon molecule—oxaloacetate (OAA). The utilization of OAA in C₄ photosynthesis entails the division of labor between two specialized cell types, i.e. the mesophyll (M) and bundle sheath (BS) cells (with some exceptions; Edwards et al. 2004). These cell types are arranged in a wreath structure called Kranz leaf anatomy (Hatch and Agostino 1992, Sage et al. 2012). OAA is transported from the M cells to the BS cells as a four-carbon acid (e.g. malate or aspartate). In the BS cells, CO₂ is released by one of three different decarboxylating enzymes, leading to the three (not necessarily fixed) subtypes of C₄ photosynthesis, namely NADP-dependent malic enzyme (NADP-ME), NAD-dependent malic enzyme (NAD-ME) and PEP carboxykinase (PEPCK), which may be developmentally and environmentally controlled (Furbank 2011). The released CO₂ is re-fixed by Rubisco, which exclusively operates in the BS cells of C₄ plants (Hatch 2002). The resulting three-carbon acid is transported back from the BS to M cells, where it is used to regenerate PEP.

Rubisco also catalyzes the oxygenation of RuBP to 3-PGA and the toxic 2-phosphoglycolate (2PG). The removal of 2PG via the photorespiratory pathway reduces the efficiency of photosynthesis under unfavorable conditions (Bauwe et al. 2010). The division of labor between the M and BS cells in C₄ plants facilitates the creation of a microenvironment of increased CO₂ concentration in the BS cells that largely suppresses photorespiration. This allows Rubisco to function near its maximal velocity, leading to higher rates of carbon fixation with a smaller amount of enzyme. Therefore, the co-ordination of carbon fixation between two cell types in C₄ plants is marked by higher efficiency of this process than in C₃ plants.

The agronomically important crops include both C₄ grasses (e.g. maize, sorghum and sugarcane) and C₃ grasses (e.g. rice, wheat and barley) (Leakey 2009). Due to advantageous characteristics of C₄ photosynthesis, significant efforts have been directed at engineering aspects of the C₄ syndrome in C₃ crops, albeit with little success (von Caemmerer et al. 2012, Li et al. 2015). Genetic evidence has indicated that the C₄ syndrome is a combination of spatio-temporal characteristics across different levels of cellular organization. Moreover, mounting evidence points out that it is probably governed by several interdependent gene regulatory programs (Brown and Bouton 1993, Weissmann and Brutnell 2012). Therefore, unraveling the
transcriptional changes that mark the functional differentiation and co-ordination of M and BS cell types is crucial to characterizing the C₄ syndrome. This is of particular interest for capitalizing on the idea of increasing efficiency of C₃ crops based on lessons from C₄ photosynthesis. This idea would be hardy tractable without a characterization of all determinants of the C₄ syndrome. This type of approach could also highlight the variety of cellular functions which need to be altered in a coordinated manner to arrive at the complex trait that is C₄ photosynthesis.

To this end, high-throughput technologies for transcriptome phenotyping (i.e. microarray and next-generation sequencing) were used to determine the cell-specific behavior of M cell and BS cell transcripts. For instance, Sawers et al. (2007) and Li et al. (2010) found that in maize approximately 18% and 21% of the transcripts, respectively, were differentially expressed between M and BS cells. Furthermore, Chang et al. (2012) characterized the cell type-specific expression, with a larger number of expressed genes in BS cells than in M cells. Finally, Tausta et al. (2014) also reported marked differences in the expression level of both cell types along a leaf developmental gradient. They found that either the M or the BS cell type contained an increased number of differentially expressed genes, dependent on the leaf section evaluated. These findings have been complemented with results from a proteomics study, highlighting the importance of transporters and biogenesis factors (Majeran et al. 2010). John et al. (2014) investigated the transcriptomic profiles of M and BS cells to determine the extent of their similarity between different C₄ lineages (maize and Setaria viridis were used as representative species); they found significant convergence (Pearson correlation coefficient of 0.89, and similar cell-specific differentially expressed transcription factors) in transcript accumulation in the core pathways of M and BS cells. Gowik et al. (2011) used next-generation sequencing to compare the transcriptomes of five closely related species in the genus Flaveria, including species with C₃ and C₄ photosynthesis as well as intermediates. In this way, they found which changes in gene expression were related to the C₄ syndrome rather than to the evolutionary distance of the two species (as was previously done with two species from the genus Cleome; Gowik et al. 2011). Finally, comparative transcriptomics and metabolomics studies between C₄ and C₃ grasses (with maize and rice as representatives, respectively) along the leaf development gradient (15 sections) have been recently conducted to identify candidate regulatory motifs putatively recruited in the evolution of C₄ photosynthesis, complementing an earlier study using a coarser gradient (i.e. 10 sections) (Pick et al. 2011, Wang et al. 2014).

In recent modeling attempts, C₄ photosynthesis has been embedded in genome-scale metabolic networks, which constitute a complementary way of understanding the differential behavior of C₄ metabolism. In particular, several models have recently been developed for the case of maize. The second-generation maize leaf model, an extension of the iRS1563 model, spans an additional 4,261 genes and 6,540 reactions (Simons et al. 2014). This second-generation maize model accounts for the M and BS cell types, with corresponding compartments based on maize-specific experimental proteomic and transcriptomic measurements. In contrast to the C₄GEM maize model (Dal’Molin et al. 2010), which focuses exclusively on maize primary metabolism, the second-generation model includes pathways from secondary metabolism, and also considers all characterized gene–protein reaction associations. With the help of the constraint-based modeling framework (Nikoloski et al. 2015), these models have been employed to analyze the flux changes under different nitrogen growth regimes as well as the role of C₃ photosynthesis in evolution of C₄ pathways (Mallmann et al. 2014, Simons et al. 2014). More recently, the evidence-based maize leaf model (Seaver et al. 2015) has been constructed using biochemical information derived from the PlantSEED database (Seaver et al. 2014). The evidence-based maize leaf model contains 2,361 reactions (localized in several subcellular compartments) and 2,869 genes. However, this model does not differentiate between M and BS cell types, and, like the maize leaf model, considers all characterized gene–protein reaction associations.

Existing experimental studies have indicated that the emergence of the C₄ syndrome requires co-ordination between several levels of cellular organization, from gene regulation to metabolism. We note that all of these studies were driven by analysis of differential behavior of single cellular components (e.g. genes, proteins and metabolites). Therefore, the contribution of the relationships between the cellular components to the C₄ syndrome remains unexplored and elusive. Moreover, the extent to which the structure of the C₄ plant metabolic network, incorporating the two co-operating metabolic subsystems, i.e. the M and BS cells, is reflected in the transcriptomics data remains unclear. These problems can be readily addressed in the constraint-based modeling framework by focusing on the coupled reactions with respect to their fluxes in steady state.

A pair of reactions (i, j) is called coupled if any of the following holds for the steady state fluxes $v_i$ and $v_j$: (i) full coupling, there exists $\lambda \neq 0$, such that $v_i = \lambda v_j$; (ii) partial coupling: $v_i = 0$ if and only if $v_j = 0$; and (iii) directional coupling: $v_j \neq 0$ if and only if $v_i \neq 0$ (Burgard et al. 2004). Here, we utilize the reaction couplings determined from the second-generation maize metabolic models to investigate two aspects of the correspondence between metabolic network structure and transcriptomic phenotypes of M and BS cells: (i) qualitatively, to determine whether any of the three coupling types is associated with particular strength of correlation between the genes coding for the corresponding enzymes; and (ii) quantitatively, to identify if the variability of expression values in the coupling types shows differences. To this end, we used the data from a coarse-grained scenario, whereby the transcript levels from the maize leaf developmental gradient without the separation of the two cell types were profiled. Altogether, our study demonstrates that precise quantitative coupling and effective co-ordination, reflected in the transcript co-expression, will have to be achieved to ensure a successfully engineered transition from C₃ to C₄ crops.
Results and Discussion

Flux coupling analysis of the maize leaf model

By applying the flux coupling analysis on the maize leaf model, we identified a total of 2,801 fully coupled and 9,366 directionally coupled pairs of reactions (Table 1). Due to its structure, the model did not contain any partial couplings between reactions. The full and directional coupling relationships are transitive, i.e. if \( v_i \) is fully (directionally) coupled to \( v_j \), and \( v_j \) is fully (directionally) coupled to \( v_k \), then \( v_i \) is fully (directionally) coupled to \( v_k \). As a result, fully (directionally) coupled reaction pairs can be combined into fully (directionally) coupled groups. In the case of a fully coupled group, a non-zero flux value in any reaction in the group implies a non-zero flux value in the entire group. However, in directionally coupled groups, only a non-zero flux value in the leading reaction implies a non-zero flux value in the rest of the reactions in the group. Here, a leading reaction of a group refers to the reaction to which the rest of reactions in the group are directionally coupled (Supplementary Fig. S5). We found a total of 631 fully coupled and 1,130 directionally coupled reaction groups in the model (Table 1).

We can further classify the coupled reaction groups with respect to the cellular localization of the participating reactions into three categories exclusive to the M cell type (referred to as M-groups), exclusive to the BS cell type (BS-groups) and those which have reactions from both cell types (referred to as M/BS-groups). In this sense, the number of fully coupled M-groups was greater than that of fully coupled BS-groups (309 vs. 269, respectively). Similarly, the number of directionally coupled M-groups was greater than that of the directionally coupled BS-groups (345 vs. 270, respectively). Interestingly, the number of fully coupled M/BS-groups is markedly smaller than that of the directionally coupled M/BS-groups (52 vs. 515, respectively). The latter finding suggests that flux coupling relationships between M and BS, as a realization of long-range spatial metabolic co-ordination, may be dominated by directionally coupled groups between the two cell types. In this scenario, a leading reaction in one cell type controls the activation of the reactions to which it is coupled in the other cell type. Supporting this observation, each directionally coupled M/BS-group contained a median of 46 reactions, which was markedly greater than the median number of reactions in any of the coupled sets found in the three previously described scenarios (Table 1). Of the leading reactions in the M/BS-groups which could be localized in the M or BS cell type, a 7-fold greater number were in the M cell in comparison with the BS cell. (i.e. 320 vs. 43, respectively, Table 1; we excluded the 36 reactions that could not be assigned to a specific cell type from this analysis). This finding pointed to the M cell type as a major player in ensuring long-range spatial co-ordination between the two cell-types.

Flux coupling analysis provided a different result when using the truncated version of the maize leaf model (i.e. without the reactions connecting the M and BS). In this case, the total numbers of fully and directionally coupled groups were decreased, from 631 to 426 fully coupled groups and from 1,130 to 668 directionally coupled groups. Interestingly, the reduction in coupling groups was due not only to the elimination of M/BS-groups (since connecting reactions are removed), but also to a general reduction in the number of fully and directionally coupled groups in both M and BS cell types (Table 1). This observation highlighted the importance of the connection between the M and BS to ensure metabolic co-ordination in the maize leaf model, not only at the intercellular level but also within each cell type.

We would like to emphasize that it is not trivial to assess which connecting reactions are independent of the M–BS configuration in the maize leaf. More specifically, the challenge is to determine which transport reactions from the fully connected model would exist if the two cells were disconnected. One way to perform such a comparison would require the isolation of M and BS cells from C3 plants. These cell types can in turn be compared with M and BS cells from maize. Such an analysis will allow the determination of intercellular transport reactions that are induced by the presence of both cell types in maize. However, the analysis of this type is beyond the scope of the present work and may be a study of interest to be considered elsewhere.

Main metabolic subsystems involved in the coupled groups

In all considered cases (i.e. fully and directionally coupled groups with respect to the localization, M, BS and M/BS), the number of coupled reactions across subsystems follows a heavy-tailed distribution. As a result, one or a very small number of subsystems contains the majority of fully or directionally coupled reactions.
We found that glycerolipid biosynthesis was on the first position in the ranking as the subsystem with the largest number of coupled reactions. However, the latter was also the subsystem with the greatest number of total reactions (i.e. irrespective of the participation in a coupled set) in the maize leaf model. Indeed, normalizing the number of coupled reactions by the total number of reactions per subsystem pushed glycerolipid biosynthesis to the bottom of the ranking; for instance, it contains 24.89%, 23.64% and 11.5% of the total number of reactions in the subsystem assigned to fully coupled M-, BS- and M/BS-groups, respectively (Supplementary Table S1).

Interestingly, the metabolic differentiation between M and BS cell types was reflected in the ranking of subsystems containing both fully and directionally coupled reactions. For instance, in the case of fully coupled groups, the ammonia assimilation cycle contains five out of the total six reactions (83.3%) in M, while it is not represented in the BS cell type. In addition, while the tRNA charging pathway includes 16 of the total 24 reactions (66.7%) in M, it contains only four reactions (16.7%) in the BS cell type (Supplementary Table S1). In the case of the M/BS-groups, subsystems such as glyoxylate and dicarboxylate (100%) and propanoate metabolism (75%) were prominently represented in fully coupled groups, while fatty acid beta-oxidation, ethanol degradation or alanine biosynthesis (all with 100% of the reactions) were key representatives of the directionally coupled groups (Supplementary Table S1). However, the degree of metabolic co-ordination between M and BS cells, as reflected by the involved metabolic subsystems, was wider than the few instances previously presented. In fact, 86 out of 220 total subsystems in the maize model were represented in fully and directionally coupled M/BS-groups, respectively (Supplementary Table S1). The latter observation points to a systemic metabolic co-ordination between the two cell types in the C₄ phenotype.

Correspondence between reaction coupling relationships and transcriptomics data profiles

We next asked if there was a correspondence between reaction coupling relationships and gene expression profiles. Observing such a correspondence would indicate that the coupling of reactions (which is based only on the metabolic network structure) imposes constraints on the underlying transcription regulation programs. This observation would be strongly supported if several, independent data sets and metabolic models are used in the analysis. Therefore, in the following, we analyzed gene expression levels over a maize leaf developmental gradient from two different studies: Leaf Data 1 with a spatial resolution of 15 leaf sections, obtained from Wang et al. (2014); and Leaf Data 2 from Pick et al. (2011), with a smaller spatial resolution of 10 leaf sections. In addition, we considered only the reactions to which gene expression levels can be mapped based on the gene–protein reaction rules defined in the metabolic models (see the Materials and Methods). Finally, we repeated the analysis using an alternative and independently reconstructed model: the evidence-based maize leaf model published in Seaver et al. (2015).

First, when evaluating the maize leaf model, we found that the set of fully coupled pairs showed increased correlation with respect to the total set of reaction pairs with both Leaf Data 1 and Leaf Data 2. This finding was supported by a bootstrapping analysis, whereby 10⁸ randomly generated correlation distributions were compared against the distribution of correlations corresponding to the fully coupled pairs (of the same sample size, see the Materials and Methods for a detailed description). For the comparison of distributions, we used the two-sample Kolmogorov–Smirnov test at a significance level of α = 0.05. We note that all results from Leaf Data 2 correspond to the background subtraction normalization of the microarray data. However, similar results were found when applying other normalization techniques (Supplementary Fig. S1).

Following the bootstrapping analysis, none of the randomly generated distributions significantly dominated the fully coupled correlation distribution when using both data sets (Fig. 1). The same approach was followed to analyze the distribution of correlations among the directionally coupled pairs of reactions. In this case, 0.7% of the randomly generated correlation distributions dominated that of the directionally coupled pairs in the case of Leaf Data 1, whereas 98.19% showed the same behavior in Leaf Data 2. Therefore, genes associated with directionally coupled pairs showed increased correlation only when using Leaf Data 1 (Fig. 1).

In many cases, a pair of reactions that is fully or directionally coupled is also adjacent in the metabolic network. Two reactions are considered adjacent if they share at least one metabolite as a substrate or a product (see the Materials and Methods). Therefore, the increased correlation in coupled pairs could be explained by the adjacency of the reactions in the network. To evaluate the validity of this claim, we considered two cases: in the first, we compared the correlation distribution corresponding to adjacent pairs with that of the total pairs of reactions. In the second case, we compared the distributions corresponding to the fully and directionally coupled pairs with that of adjacent reactions. We applied the same procedure as above: drawing random samples from the distribution of correlations of total pairs, in the first case, and from the distribution corresponding to adjacent pairs, in the second.

We found that none of the random distributions dominated that of the adjacent pairs when using Leaf Data 1 and 60.22% when using Leaf Data 2 (Supplementary Fig. S2). Therefore, the expression of genes associated with adjacent pairs of reactions tended to show increased correlation with respect to that of total pairs only when using Leaf Data 1. However, when comparing the distributions of fully and directionally coupled pairs with adjacent pairs, we found that none of the random correlation samples from adjacent pairs dominated the distribution corresponding to fully coupled pairs and 8.48% dominated that of directionally coupled pairs when using Leaf Data 1 (Supplementary Fig. S2). This indicates that genes associated with fully coupled and (to a lesser extent) directionally coupled pairs of reactions show increased correlation with respect to adjacent pairs. In the case of Leaf Data 2, only genes associated
Fig. 1 Complementary cumulative correlation distributions for Leaf Data 1 and Leaf Data 2. The correlation distributions obtained from the set of fully coupled (FC), directionally coupled (DC) and total reactions are depicted in blue, cyan and black, respectively. (A, B) and (C, D) correspond to the case where Leaf Data 1 and Leaf Data 2, respectively, were mapped to the maize leaf model. (E, F) and (G, H) correspond to the case where Leaf Data 1 and Leaf Data 2, respectively, were mapped to the evidence-based maize leaf model. In all cases, randomly generated samples from the total distribution are shown in dark gray, and the proportion of sampled distributions that significantly dominate (two-sample Kolmogorov–Smirnov test; $\alpha = 0.05$) the distribution evaluated is shown as a P-value.
with fully coupled pairs showed greater correlation than that of adjacent pairs, since 97.19% of the random distributions drawn from the set of correlations of adjacent pairs dominated that corresponding to directionally coupled pairs. This is in accord with the previous results, where genes associated with directionally coupled and adjacent pairs did not show increased correlation with respect to the total set of reactions (Fig. 1). As commented before, we repeated the correlation analysis using an independently constructed metabolic model—the evidence-based maize leaf model (Seaver et al. 2015) with the same expression data as evaluated before. Interestingly, genes associated with either fully or directionally coupled pairs of reactions showed significantly increased correlation with respect to the total set when using both data sets (indeed none of the random samples dominated the corresponding distribution in any case; Fig. 1). In addition, genes associated with adjacent pairs of reactions showed increased correlation with respect to the total set of reactions in both data sets (again none of the random samples dominated the correlation distribution). However, genes associated with either fully or directionally coupled pairs showed increased correlation with respect to that of adjacent pairs, here again when evaluating both data sets (Supplementary Fig. S2). These findings are in accord with the results observed when using the maize leaf model and Leaf Data 1, and partially match results when using Leaf Data 2. This is because, as commented before, genes associated with directionally coupled pairs did not show increased correlation when compared with the total set or adjacent pairs of reactions.

Altogether, our results from the data-driven analysis support the hypothesis that fully, and, to a lesser extent, directionally coupled pairs of reactions show increased correlation with respect to a random pair of reactions in the two leaf model. These findings match the results of a previous study by Notebaart et al. (2008), where increased correlation between fully, partially and directionally coupled pairs of reactions was reported when evaluating metabolic models of *Escherichia coli* and yeast. In addition, although genes associated with adjacent pairs also tend to show increased correlation, co-expression is stronger in the case of fully and directionally coupled pairs of reactions. Therefore, our findings imply that the metabolic coupling relationships between pairs of reactions may dictate the degree of co-expression between their associated genes, even in more complex, multicellular organisms such as maize.

We next investigated the mean correlation among reaction pairs in fully and directionally coupled groups in the maize leaf model. Our aim was to identify groups of reactions that showed a particularly increased mean correlation (see the Materials and Methods) when using both data sets. In this sense, 47 fully coupled and 48 directionally coupled groups showed high correlation when using Leaf Data 1 and 24 fully coupled and 10 directionally coupled groups showed high correlation when using Leaf Data 2. Interestingly, we found the Ammonia assimilation cycle II, Nitrate reduction II (assimilatory) and Glutamine biosynthesis among the most represented metabolic subsystems within the highly correlated fully and directionally coupled groups in both data sets (with five reactions, representing 83.3% of the total number of reactions in this subsystem). However, a total of 28 and 12 metabolic subsystems were shared in highly expressed fully and directionally coupled groups, respectively, across both data sets (Supplementary Table S2). For instance, the Calvin–Benson–Bassham cycle, Folate and Pyruvate metabolism were all among the shared subsystems associated with highly correlated fully coupled groups, while Chlorophyll a biosynthesis, Glutamine biosynthesis and C4 photosynthetic carbon assimilation cycle were among the shared subsystems encountered in highly correlated directionally coupled groups (Supplementary Table S2). An example of a highly expressed fully coupled group is given in Fig. 2, where four reactions are connected in a linear pathway. The group exhibited a mean correlation of 0.8933 when using Leaf Data 1 and of 0.8978 using Leaf Data 2.

We would like to stress that our study does not discriminate between cell types when evaluating the correlation when using the maize leaf model. This was due to the coarse-grained nature of the expression data sets evaluated, where the whole leaf was used to extract the mRNA. However, it would be of marked interest to repeat this study using M- and BS-specific expression data. Such an analysis will allow us to evaluate if the correlation between fully and directionally coupled M–BS pairs is also increased. As commented on in the Materials and Methods section, these pairs are assigned the same data values in this study and hence are eliminated from the sample. Unfortunately, to our knowledge, only one study, (Tausta et al. 2014) provides such a cell-specific spatial expression series for maize, although with only three sample points. The three sample points render these data set unsuitable to evaluate the correlation between gene pairs.

**Evaluation of reaction-associated expression data values across leaf sections**

In the preceding sections, we investigated the correspondence between expression profiles of genes across the leaf developmental gradient and the involvement of the corresponding reactions in full and directional couplings. This type of analysis provides qualitative investigations by using the spatial distribution of transcripts. However, it does not use the expression values for each section separately. In the following, we investigated if fully and directionally coupled reactions showed a particular magnitude of expression in comparison with the total set of reactions in each section across the leaf gradient. To this end, we calculated the median absolute deviation and the median for each distribution of expression values (i.e. the fully coupled, directionally coupled and uncoupled set of reactions; see the Materials and Methods) for each spatial data point. In addition, we performed this analysis using both the maize leaf model and the evidence-based maize leaf model.

In the case of Leaf Data 1, the median gene expression tended to decrease in all three groups from the basal part (section 1) to the tip of the leaf (section 15) when they were mapped to both the maize leaf and the evidence-based maize leaf model. Moreover, the median expression associated with
the fully and directionally coupled sets was smaller than that of the uncoupled set across all leaf sections when using the evidence-based maize leaf model (Fig. 3). We evaluated the significance of this difference by randomly generating $10^4$ samples from the fully and directionally coupled set of size equal to the uncoupled set (the uncoupled set was smaller, with 302 reactions with the associated data set vs. 903 and 422 reactions in the fully and directionally coupled set, respectively) (Materials and Methods). This analysis showed that the observed differences in medians were indeed significant (with a significant level of $\alpha = 0.05$, which is maintained in the following evaluations) (Supplementary Fig. S4). We next evaluated the median absolute deviations of the previously obtained medians, finding that, in each section of Leaf Data 1, the median absolute deviation values were significantly increased in the uncoupled set of reactions in comparison with the set of fully and directionally coupled reactions when using each of the two models (Fig. 3: Supplementary Fig. S4). Since the majority of the fully coupled pairs of reactions were in a constant ratio of value 1 (90.42% of the total fully coupled pairs), their transcript was expected to show similar abundance. The latter is true under the assumption of proportionality of flux and transcript levels for this type of reactions. Therefore, the decreased variability in the gene expression could be partly attributed to the reaction coupling relationships. Moreover, the maximum median absolute deviation value across the leaf gradient occurred in middle sections in Leaf Data 1 (sections 9–11; Fig. 3). It has been reported that a transition from $C_3$ (in the basal region) to $C_4$ metabolism (in the apical region) exists across the maize leaf developmental gradient (Pick et al. 2011, Wang et al. 2014a). Hence, the increased median absolute deviation in gene expression values found in the intermediate zone of the developmental gradient could be explained by the transition of the genetic program showing characteristics of both $C_3$ and $C_4$ metabolism.
In the case of Leaf Data 2, median expression values (mapped to either of the two models) tended to decrease from the basal part towards a minimum found at leaf section number 8 and to increase again at the tip in all three evaluated groups. Hence, they show the same pattern observed in Pick et al. (2011) before mapping the data to the models. Interestingly, genes associated with reactions in the fully and directionally coupled sets here again showed significantly smaller median expression values when using the evidence-based maize leaf model (Supplementary Figs. S3, S4). In contrast, only genes associated with reactions in the directionally coupled set showed significantly decreased median absolute deviations across all leaf sections in the case of the evidence-based maize leaf model (Supplementary Figs. S3, S4). The discrepancies observed when evaluating the median absolute deviations in both data sets could be explained by the general increase in the correlation values in Leaf Data 2 in comparison with Leaf Data 1. More specifically, the median correlation between any pair of reactions was 0.2016 and 0.0880 in Leaf Data 1 and 0.4682 and 0.3932 in Leaf Data 2 when using the maize leaf model and the evidence-based maize leaf model, respectively.

Conclusions
The aim of our study was to investigate and characterize the extent of co-ordination between the M and BS cells in maize, as a C₄ model plant. In addition, we set out to unravel the degree to which the co-expression of genes coding for metabolic enzymes across a developmental gradient of maize leaf was explained by coupling relationships among the correspondingly catalyzed reactions. To this end, we used the genome-scale metabolic network of maize first to determine the existing types of coupled reactions (i.e. full, partial and directional). In all cases, the P-value for the comparison was <1.2 x 10⁻³. 

Fig. 3 Median and median absolute deviation of the mapped expression values in Leaf Data 1. (A) Median and (B) median absolute deviation (mad) of gene expression values mapped to the maize leaf model and across the leaf gradient for Leaf Data 1. In both cases, results are shown for the set of uncoupled reactions (black) and for the sets of fully (blue) and directionally coupled (cyan) reactions (i.e. reactions that are found in at least one fully or directionally coupled pair in the maize leaf model, respectively). No significant differences (z = 0.05) were found between median values corresponding to fully and directionally coupled reactions and the set of uncoupled reactions. In the case of the median absolute deviations, both sets, fully and directionally coupled, show significantly smaller values as compared with the set of uncoupled reactions (P-values are shown in Supplementary Fig. S4). (C) Median and (D) median absolute deviations when Leaf Data 1 was mapped to the evidence-based maize leaf model. The sets of fully and directionally coupled reactions present associated median and median absolute deviations that are smaller than those corresponding to the set of uncoupled reactions. In all cases, the P-value for the comparison was <1.2 x 10⁻³.
categorized the pairs of coupled reactions based on whether they appeared in the same cell type or between the two cell types. Interestingly, we found that leading reactions in the directionally coupled pair between the two cell types were more prominent in the M in comparison with the BS type, indicating directional co-ordination of the activities between the two cell types.

We then examined the extent to which the coupling of reactions is reflected in the correlation of the expression levels of the corresponding genes. Our findings demonstrated that fully coupled reactions showed a larger magnitude of correlations in comparison with non-coupled reactions. This result indicates that the metabolic constraints, due to the network structure alone, shape the transcriptional regulation of the metabolic genes. This result held true on two independent data sets from different laboratories (Pick et al. 2011, Wang et al. 2014a) as well as with two independently constructed genome-scale metabolic network models. The observed correspondence between full coupling and respective transcriptomics data profiles shows that transcript levels tend to be adjusted to match the metabolic flux requirements dictated by the coupling relationships among the associated reactions. This observation may reflect the budgeting of resources by the plant. In the case of directional coupling, we considered only the correlations to the leading reaction. We demonstrated that the correlation between the respective transcripts, for directional coupling, was larger than that for the non-coupled reactions. This finding is yet another manifestation of the transcript distribution due to the dependence/coupling of reactions. Finally, we found that, in comparison with the total set of reactions, the variation in gene expression values (quantified by the median absolute deviation) was significantly smaller for the sets of fully and directionally coupled reactions.

Altogether, our findings demonstrated that precise quantitative and qualitative couplings of reactions are necessary for the C4 syndrome, and that these couplings are well reflected at the sets of fully and directionally coupled reactions. Interestingly, we found leading reactions involved in the C4 photosynthesis. More specifically, we categorized the pairs of coupled reactions based on whether they exclusively contained reactions located in the M or BS, respectively, or, in addition to M or BS, they contained reactions with unspecified location or assigned to ‘leaf’. Groups containing reactions located in the M and BS were categorized as M/BS-groups. The distribution of metabolic subsystems was evaluated for each fully and directionally coupled group category (i.e. M-, BS- and M/BS-groups). To this end, all fully or directionally coupled groups in each category were lumped together, and the number of reactions in each subsystem was counted. In addition, metabolic subsystems were ranked based on the normalized number of reactions in each category, obtained by dividing the number of reactions in all fully (directionally) coupled groups in a subsystem assigned to a location category by the total number of reactions assigned to this category in the maize leaf model.

Gene expression and metabolic data profiles
Gene expression levels over a maize leaf developmental gradient were taken from two different studies. One of the data sets was obtained from Wang et al. (2014), here named Leaf Data 1, and consists of a series of 15 leaf sections. The other data set was taken from Pick et al. (2011), here named Leaf Data 2, and consists of a series of decreased spatial resolution, including 10 leaf sections. Gene expression was mapped to each individual reaction following the gene–protein reaction rules defined in the model and our in-house-developed MATLAB code, provided in Supplementary Data File S1.

Fully (directionally) coupled groups were classified as M- or BS-groups if they exclusively contained reactions located in the M or BS, respectively, or, in addition to M or BS, they contained reactions with unspecified location or assigned to ‘leaf’. Groups containing reactions located in the M and BS were categorized as M/BS-groups. The distribution of metabolic subsystems was evaluated for each fully and directionally coupled group category (i.e. M-, BS- and M/BS-groups). To this end, all fully or directionally coupled groups in each category were lumped together, and the number of reactions in each subsystem was counted. In addition, metabolic subsystems were ranked based on the normalized number of reactions in each category, obtained by dividing the number of reactions in all fully (directionally) coupled groups in a subsystem assigned to a location category by the total number of reactions assigned to this category in the maize leaf model.

Materials and Methods

Genome-scale metabolic models and flux coupling analysis
Two genome-scale metabolic models were used in this study; the maize leaf metabolic model published in Simons et al. (2014) and the evidence-based maize leaf model published in Seaver et al. (2015). The maize leaf model was used to generate all results in this study (including the analysis on coupling relationships among cell types). In addition to the maize leaf model, the evidence-based maize leaf model was used when evaluating the correspondence between coupled pairs of reactions and the co-expression of associated genes, as well as the difference in expression values with respect to genes associated with uncoupled reactions. The use of an alternative metabolic model was intended to support the results found using the maize leaf model. However, the evidence-based maize leaf model does not consider M and BS cells and thus was not used during the first part of the study (concerning the description of coupling groups). The Systems Biology Markup Language (SBML) files were imported to MATLAB using the ‘readCbModel’ function of the Constraint Based Reconstruction Analysis (COBRA) toolbox (Schellenberger et al. 2011).

The flux coupling analysis on both models was performed using the F2C2 package in MATLAB (Larhlimi et al. 2012). In a pre-processing step, the F2C2 package automatically removes blocked reactions from the model (i.e. reactions that are not able to carry non-zero flux in any of the allowed steady-state flux distributions); these reactions, by definition, do not appear in any flux coupling relationship. Consequently, further analyses in this study were conducted on the reduced version of the original models, with blocked reactions removed. The reduced version of the model was extracted by using in-house-developed MATLAB code, provided in Supplementary Data File S1.

We then examined the extent to which the coupling of reactions is reflected in the correlation of the expression levels of the corresponding genes. Our findings demonstrated that fully coupled reactions showed a larger magnitude of correlations in comparison with non-coupled reactions. This result indicates that the metabolic constraints, due to the network structure alone, shape the transcriptional regulation of the metabolic genes. This result held true on two independent data sets from different laboratories (Pick et al. 2011, Wang et al. 2014a) as well as with two independently constructed genome-scale metabolic network models. The observed correspondence between full coupling and respective transcriptomics data profiles shows that transcript levels tend to be adjusted to match the metabolic flux requirements dictated by the coupling relationships among the associated reactions. This observation may reflect the budgeting of resources by the plant. In the case of directional coupling, we considered only the correlations to the leading reaction. We demonstrated that the correlation between the respective transcripts, for directional coupling, was larger than that for the non-coupled reactions. This finding is yet another manifestation of the transcript distribution due to the dependence/coupling of reactions. Finally, we found that, in comparison with the total set of reactions, the variation in gene expression values (quantified by the median absolute deviation) was significantly smaller for the sets of fully and directionally coupled reactions.

Altogether, our findings demonstrated that precise quantitative and qualitative couplings of reactions are necessary for the C4 syndrome, and that these couplings are well reflected at the level of transcriptional regulation. By examining the metabolic pathways in which coupled reactions take part, we showed that the co-ordination between M and BS cells goes beyond the reactions involved in C4 photosynthesis. More specifically, we found that reaction couplings between and within the two cell types mostly contribute to nitrogen recycling and assimilation, as a key pathway involved in the co-ordination between M and BS functioning. Since our results focused on the co-ordinated expression of metabolic genes, they pave the way to analyzing which gene regulators ensure the extent of qualitative and quantitative transcriptional co-ordination found and how this is ensured over developmental and environmental gradients. The findings from our study provide the first answers about the precise metabolic couplings which will have to be imposed for successful engineering of the C4 syndrome in C3 crops.
of the total of 3,438 reactions in the model. After the gene mapping was completed, 1,614 and 1,381 reactions were associated gene expression series for the leaf sections, when using Leaf Data 1 and Leaf Data 2, respectively.

Correlation of expression data within coupled reaction pairs and evaluation of significance

The Pearson correlation was computed for the associated spatial data series of all possible reaction pairs (with associated expression data) of the maize leaf and the evidence-based maize leaf model, generating a distribution of total correlation values. In addition, reaction pairs with the same associated gene expression data were removed from the distribution. To check if the fully and directionally coupled pairs presented increased correlation values, the distribution of total correlation values was compared against the correlation value subsets corresponding to fully and directionally coupled pairs. Following Notebaart et al. (2008), the set of correlation values corresponding to adjacent pairs of reactions was also extracted from the total correlation set of reaction pairs. A pair of reactions was defined as adjacent if they shared at least one metabolite acting as a substrate or product (Supplementary Fig. 5).

All the generated correlation subsets (i.e. fully coupled, directionally coupled and adjacent) were compared against the correlation set of total pairs of reactions. To this end, a two-sample Kolmogorov–Smirnov test was used (‘kstest2’ function in MATLAB) to evaluate if a given correlation distribution (i.e. fully coupled, directionally coupled and adjacent pairs) dominated the correlation distribution generated by the total pairs of reactions in the model. In addition, to take into account the differences in distribution size, the test was repeated over 10⁴ correlation samples of the total reaction pair distribution with sample size equal to the size of the distribution being compared. Thereafter the proportion of significant Kolmogorov–Smirnov tests ($\alpha = 0.05$) was taken as an overall $p$-value of the comparison. This procedure was also implemented for the two expression data sets previously mentioned. To generate the complementary cumulative distributions (Fig. 1: Supplementary Fig. S1), a sequence of correlation values (with step = 0.01) was first generated, and then, for each correlation value, the fraction of reaction pairs with equal or greater correlation was computed. This procedure was repeated for each reaction pair subset as well as for each sample generated.

Expression data correlation within coupled reaction pairs in fully and directionally coupled groups

Highly correlated fully and directionally coupled groups were identified as follows: for each fully (directionally) coupled pair, the mean correlation among its reaction pairs was first calculated and taken as a representative value of the gene expression correlation in the group. A total of $10^4$ random samples of $n$ reaction pairs (with $n$ equal to the number of reaction pairs in the group, i.e. the group size) were then generated and the mean correlation of each sample was calculated, hence giving a distribution of mean sample correlation values for each group size. A fully (directionally) coupled group was considered as highly correlated if its mean correlation value was over the 90th percentile of the previously generated distribution of mean sample correlation values. Additionally, highly correlated groups were also defined using the 80th and 70th percentile as thresholds. In these cases, highly correlated groups were, as expected, of larger size and, consequently, new subsystems were included in the ranking. However, the subsystems occupying the first positions in the ranking generated when using the 90th percentile maintained the first positions when using the 80th or the 70th percentile as the threshold value (Supplementary Table S2) additionally displays the ranked subsystems when using the 80th and 70th percentile and Leaf Data 1 as an example). In the case of directionally coupled groups, the correlation was computed among the pairs formed by the leading reaction and each of the remaining reactions in the group.

Evaluation of expression data values and variability in fully and directionally coupled reactions

The median absolute deviations as well as the median of the reaction-associated expression values were calculated for each spatial point in the leaf over the sets of uncoupled, fully coupled and directionally coupled reactions with associated data. The set of fully (directionally) coupled reactions was obtained by taking all unique reactions that were in a fully (directionally) coupled pair in the leaf model. Similarly, the set of uncoupled reactions was formed by reactions that did not participate in any coupling relationship. The medians and/or the median absolute deviation were tested for significant differences between the uncoupled set and the fully (directionally) coupled set of reactions. To this end, $10^4$ random samples of the fully (directionally) coupled set were generated, with sample size equal to the size of the uncoupled set of reactions. This is because the number of reactions with associated data in the uncoupled set was smaller in all cases than that corresponding to the fully and directionally coupled sets. To test if the median and/or median absolute deviation values in the fully (directionally) coupled set were smaller than in the total set of reactions, the proportion of the previously generated samples with an equal or greater value than that corresponding to the uncoupled set was taken as a $p$-value of the comparison. This procedure was applied to each of the metabolic models and data sets used in this study.

Construction of the truncated maize leaf model

A truncated version of the maize leaf model was constructed to perform the comparative analysis between a fully connected M-B5 model and an alternative scenario with the two unconnected cell types. To construct the modified network, all reactions connecting both cell types were first removed. Blocked reactions were then identified, and a reduced model version was extracted with blocked reactions removed, obtaining a smaller model with 2,173 reactions and 1,672 metabolites.

Availability of the code used in this study

All MATLAB code generated in this study as well as the metabolic models and the expression data used are available in Supplementary Data File S1.

Supplementary data

Supplementary data are available at PCP online.

Funding

This research was carried out without funding

Disclosures

The authors have no conflicts of interest to declare.

References

Bauwe, H., Hagemann, M. and Fernie, A.R. (2010) Photospiration: players, partners and origin. Trends Plant Sci. 15: 330–336.
Brown, H.R. and Bouton, J.H. (1993) Physiology and genetics of interspecific hybrids between photosynthetic types. Annu. Rev. Plant Physiol. Plant Mol. Biol. 44: 435–456.
Burgard, A.P., Nikolaev, E.V., Schilling, C.H. and Maranas, C.D. (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res. 14: 301–312.
Chang, Y.-M., Liu, W.-Y., Shih, A.C., Shen, M.-N., Lu, G.-H., Lu M.-Y., et al. (2012) Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiol. 160: 165–177.
DalMolin, C.G., Quek, L.E., Palfreyman, R.W., Brambley, S.M. and Nielsen, L.K. (2010) C4GEM, a genome-scale metabolic model to study C4 plant metabolism. Plant Physiol. 154: 1871–1885.
Edwards, G.E., Franceschi, V.R. and Voznesenskaya, E.V. (2004) Single-cell C(4) photosynthesis versus the dual-cell (Kranz) paradigm. Annu. Rev. Plant Biol. 55: 173–196.

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.

Gowik, U., Brautigam, A., Weber, K.L., Weber, A.P.M. and Westhoff, P. (2011) Evolution of C(4) photosynthesis in the genus Flaveria: how many and which genes does it take to make C4? Plant Cell 23: 2087–2105.

Gowik, U. and Westhoff, P. (2011) The path from C3 to C4 photosynthesis. Plant Physiol. 155: 56–63.

Hatch, M.D. and Agostino, A. (1992). Bilevel disulfide group reduction in the activation of C4 leaf nicotinamide adenine dinucleotide phosphate-malate dehydrogenase. Plant Physiol. 100: 360–366.

Hatch, M.D. (2002) C(4) photosynthesis: discovery and resolution. Photosynth. Res. 73: 251–256.

Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Clay, N.K., et al. (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat. Genet. 41: 258–263.

John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S. and Hibberd, J.M. (2014) Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. Plant Physiol. 165: 62–75.

Laetsch, W.M. (1974) The C4 syndrome: a structural analysis. Annu. Rev. Plant Physiol. 25: 27–52.

Larhlimi, A., David, L., Selbig, J. and Bockmayr, A. (2012) F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. BMC Bioinformatics 13: 57.

Leakey, A.D. (2009) Rising atmospheric carbon dioxide concentration and the future of C4 crops for food and fuel. Proc. R. Soc. B: Biol. Sci. 276: 2333–2343.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., et al. (2010) The developmental dynamics of the maize leaf transcriptome. Nat. Genet. 42: 1060–1067.

Li, Y., Ma, X., Zhao, J., Xu, J., Shi, J., Zhu, X.-G., et al. (2015) Developmental genetic mechanisms of C4 syndrome based on transcriptome analysis of C3 cotyledons and C4 assimilating shoots in Haloxylon ammodendron. PLoS One 10: e0117175.

Majeran, W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., et al. (2010) Structural and metabolic transitions of C4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. Plant Cell 22: 3509–3542.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P., et al. (2014) The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. eLife 3: e02478.

Nikoloski, Z., Perez-Storey, R. and Sweetlove, L.J. (2015) Inference and prediction of metabolic network fluxes. Plant Physiol. 169: 1443–1455.

Notebaart, R.A., Teusink, B., Siezen, R.J. and Papp, B. (2008) Co-regulation of metabolic genes is better explained by flux coupling than by network distance. PLoS Comput. Biol. 4: e26.

Pick, T.R, Bräutigam, A., Schlüter, U., Denton, A.K., Colmsee, C., Scholz, U., et al. (2011) Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. Plant Cell 23: 4208–4220.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W. Shi, W., et al. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43: e47.

Sage, R.F., Sage, T.L. and Kocacinar, F. (2012) Photorespiration and the evolution of C4 photosynthesis. Annu. Rev. Plant Biol. 63: 19–47.

Sawers, R.J.H., Liu, P., Anufrikova, K., Hwang J.T.G. and Brutnell, T.P. (2007) A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf. BMC Genomics 8: 12.

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.F., et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6: 1290–1307.

Seaver, S.M.D., Bradbury, L.M.T., Frelin, O., Zarecki, R., Ruppin, E., Hanson, A.D., et al. (2015) Improved evidence-based genome-scale metabolic models for maize leaf, embryo, and endosperm. Front. Plant Sci. 6: 142.

Seaver, S.M.D., Gerdes, S., Frelin, O., Lerma-Ortiz, C., Bradbury, L.M.T., Zallot, R., et al. (2014) High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. Proc. Natl. Acad. Sci. USA 111: 9645–9650.

Simons, M., Saha, R., Amiour, N., Kumar, A., Guillard, L., Clément, G., et al. (2014) Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. Plant Physiol. 166: 1659–1674.

Tausta, S.L., Li, P., Si, Y.T., Gandotra, N., Liu, P., Sun, Q., et al. (2014) Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. J. Exp. Bot. 65: 3543–3555

von Caemmerer, S., Quick, W.P. and Furbank, R.T. (2012) The development of C4 rice: current progress and future challenges. Science 336: 1671–1672.

Wang, L., Czedik-Eysenberg, A., Mertz, R.A., Si, Y., Tohge, T., Nunes-Nesi, A., et al. (2014) Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. Nat. Biotechnol. 32: 1158–1165.

Weissmann, S. and Brutnell, T.P. (2012) Engineering C4 photosynthetic regulatory networks.Curr. Opin. Biotechnol. 23: 298–304.