



Stoichiometric Correlation Analysis: Principles of Metabolic Functionality from Metabolomics Data

Kevin Schwahn^{1,2}, Romina Beleggia³, Nooshin Omranian^{1,2,4} and Zoran Nikoloski^{1,2,4*}

¹ Systems Biology and Mathematical Modeling Group, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany, ² Bioinformatics Group, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany,

³ Consiglio per la Ricerca in Agricoltura e L'analisi Dell'economia Agraria, Centro di Ricerca per la Cerealicoltura e le Colture Industriali (CREA-CI), Foggia, Italy, ⁴ Center of Plant Systems Biology and Biotechnology, Plovdiv, Bulgaria

OPEN ACCESS

Edited by:

Ping Ma,
University of Georgia, United States

Reviewed by:

Xiaowei Wu,
Virginia Tech, United States
Wenxuan Zhong,
University of Georgia, United States

Jingyi Zhang and Honghe Jin
contributed to the review of Wenxuan
Zhong

*Correspondence:

Zoran Nikoloski
nikoloski@mpimp-golm.mpg.de

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Plant Science

Received: 21 September 2017

Accepted: 05 December 2017

Published: 18 December 2017

Citation:

Schwahn K, Beleggia R, Omranian N
and Nikoloski Z (2017) Stoichiometric
Correlation Analysis: Principles of
Metabolic Functionality from
Metabolomics Data.
Front. Plant Sci. 8:2152.
doi: 10.3389/fpls.2017.02152

Recent advances in metabolomics technologies have resulted in high-quality (time-resolved) metabolic profiles with an increasing coverage of metabolic pathways. These data profiles represent read-outs from often non-linear dynamics of metabolic networks. Yet, metabolic profiles have largely been explored with regression-based approaches that only capture linear relationships, rendering it difficult to determine the extent to which the data reflect the underlying reaction rates and their couplings. Here we propose an approach termed Stoichiometric Correlation Analysis (SCA) based on correlation between positive linear combinations of log-transformed metabolic profiles. The log-transformation is due to the evidence that metabolic networks can be modeled by mass action law and kinetics derived from it. Unlike the existing approaches which establish a relation between pairs of metabolites, SCA facilitates the discovery of higher-order dependence between more than two metabolites. By using a paradigmatic model of the tricarboxylic acid cycle we show that the higher-order dependence reflects the coupling of concentration of reactant complexes, capturing the subtle difference between the employed enzyme kinetics. Using time-resolved metabolic profiles from *Arabidopsis thaliana* and *Escherichia coli*, we show that SCA can be used to quantify the difference in coupling of reactant complexes, and hence, reaction rates, underlying the stringent response in these model organisms. By using SCA with data from natural variation of wild and domesticated wheat and tomato accession, we demonstrate that the domestication is accompanied by loss of such couplings, in these species. Therefore, application of SCA to metabolomics data from natural variation in wild and domesticated populations provides a mechanistic way to understanding domestication and its relation to metabolic networks.

Keywords: metabolism, systems biology, maximal correlation, correlation analysis, domestication

INTRODUCTION

Metabolomics profiling technologies are increasingly used for phenotyping of biological systems to understand the contribution of metabolism to complex phenotypes, including growth and diseases (Sumner et al., 2003; Schauer and Fernie, 2006; Kaddurah-Daouk et al., 2008). They have been used to assess the relative and absolute levels of different metabolites after perturbation or

over time (Fiehn et al., 2000). The resulting metabolic data profiles manifest the joint effect of the rates of multiple biochemical reactions interrelated in metabolic networks. Reaction rates are themselves subjected to different types of regulation, often carried out by altering the concentration of metabolites (Koshland, 1970).

Regulation of reaction rates is necessary to ensure that the activities attributed to different parts of the system are coordinated. The simplest way to capture the coordination of reaction rates is through their coupling, whereby the ratio of the reaction rates is maintained in a narrow range (Millard et al., 2017), resulting in high positive correlation values between the coupled reaction rates over different experiments (e.g., environments). The principle questions in analyzing the data from metabolomics technologies are then to determine the extent to which the metabolite levels reflect the coupling of the underlying biochemical reactions as well as any differences in these characteristics between experimental scenarios (e.g., comparison of genotypes or treatments).

Despite the apparent non-linearities due to the metabolic structure and regulation, metabolic data profiles are usually analyzed by regression-based approaches that can only capture linear relationships. Ever since the seminal work of Vance et al. (2002), which used partial correlations to analyze the dependence between metabolites and reconstruct the reactions in which they participate, the existing analyses of metabolic data profiles rely on applying various similarity measures to given metabolic profiles (Çakir et al., 2009; Krumsiek et al., 2016). Since correlation, like other similarity measures, results in bilateral relationships between metabolites, the resulting metabolite-metabolite relationships have been represented and analyzed in the framework of metabolic correlation network analysis (MCNA) (Toubiana et al., 2013). This has led to the usage of MCNA to compare data from different scenarios based on the concept of differential networks (Chen et al., 2009; Ideker and Krogan, 2012). However, the principle question about coupling of biochemical reactions reflected in the metabolic profiles remains unresolved.

Assuming random fluctuations around a given steady state, metabolic correlations have been related to the Jacobian of the system of ODEs that describe the change in metabolite concentrations (van Kampen, 2007). In a series of studies, this relation has been employed for reconstructing the Jacobian of simplified metabolic networks and for comparison of different treatments (Steuer et al., 2003; Sun et al., 2015; Nägele et al., 2016). While this approach places metabolic correlations on strong theoretical basis, it is not applicable for analysis of instationary data. In another network-driven approach (Hackett et al., 2016), metabolic profiles have been fitted to steady-state compatible fluxes [extracted under optimality assumption of the flux balance analysis (Orth et al., 2010)] with different functional form for the reaction rates $v(x, k)$. This approach has allowed the elucidation of novel regulators of reaction rates.

Here we take a principally different approach motivated by biochemically reasonable assumptions which often hold in realistic biological scenarios. Since biological systems sense and respond to environmental perturbations, they achieve normal

functionality in face of these perturbations. To this end, various feedbacks and mechanisms based on network structure have evolved to maintain coupling of reaction rates. Based on this idea and under the assumption that elementary biochemical reactions can be modeled via mass action kinetics (without neglecting the effect of enzymes), here we propose a novel means to analyze metabolic profiles based on the concept of constrained maximal correlation coefficient. We use this approach to analyze and characterize the role of metabolites in a network that captures the reaction rate coupling. First, by using a paradigmatic model of the tricarboxylic acid (TCA) cycle, we investigate the effect from departures of the assumption of mass action on the identified reaction coupling and couplings of reactant complexes. We then show that Stoichiometric Correlation Analysis (SCA) can be employed to perform cross-species comparison of the TCA cycle and amino acid synthesis pathways. In addition, we demonstrate that the proposed approach can be used to mechanistically understand the agronomically important process of domestication, here, in the case of wheat as well as in tomato and strawberry.

MATERIALS AND METHODS

Description of the Approach with the Underlying Assumptions and Principles Maximal Correlation

Modern applications, particularly in computational biology, often consider a large number of variables involved in nonlinear (pairwise) relationships. The maximal correlation coefficient ρ , between a pair of random variables W and Z , introduced by Gebelein (1941) and already extensively studied by Lancaster (1957) and Rényi (1959), is defined as:

$$\rho = \sup \left\{ \frac{\text{cov}(f(W), g(Z))}{\sqrt{V(f(W))V(g(Z))}} \mid V(f(W)) > 0, V(g(Z)) > 0 \right\}, (1)$$

where the supremum is taken over all functions f of W and g of Z with finite variances, i.e., $V(f(W)) > 0$ and $V(g(Z)) > 0$. Maximal correlation then infers (non-linear) transformations of two random variables by maximizing their pairwise correlation (see **Figure 1** for illustration). We note that W and Z are independent if and only if $\rho = 0$, relating maximal correlation to mutual information (see Introduction).

There exist efficient algorithms to compute maximal correlation for both discrete (Breiman and Friedman, 1985) and continuous (Lancaster, 1957) random variables. Direct application of these algorithms for calculation of maximal correlation to time-resolved metabolic profiles is hampered since: (1) metabolic profiles are quantitative (i.e., continuous variables), as they capture the content of metabolic pools in biological systems; therefore, any decision to move to a range of values (e.g., small, medium, large, as it is done in discretization), will lead to drastic simplification, and (2) time-resolved metabolic profiles include relatively few time points, rendering the calculation of maximal correlation based on contingency

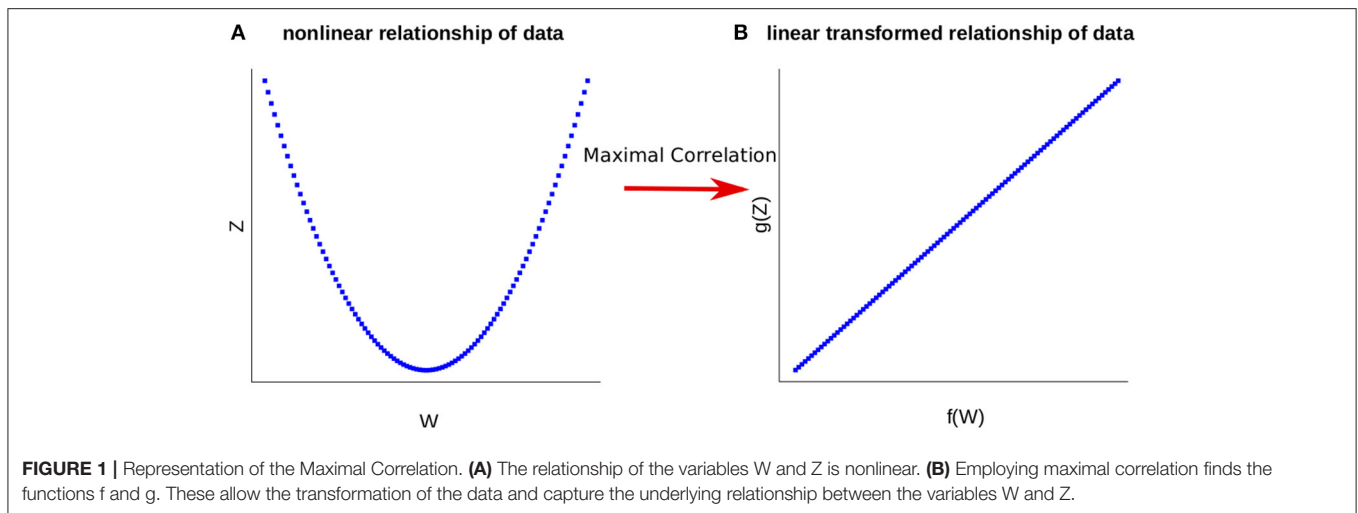


table challenging (e.g. Nguyen et al. (2014) analyzed maximal correlation with at least 100 data points which is still not available for metabolomics data).

Stoichiometric Correlation Analysis and the Principle of Metabolic Network Robustness

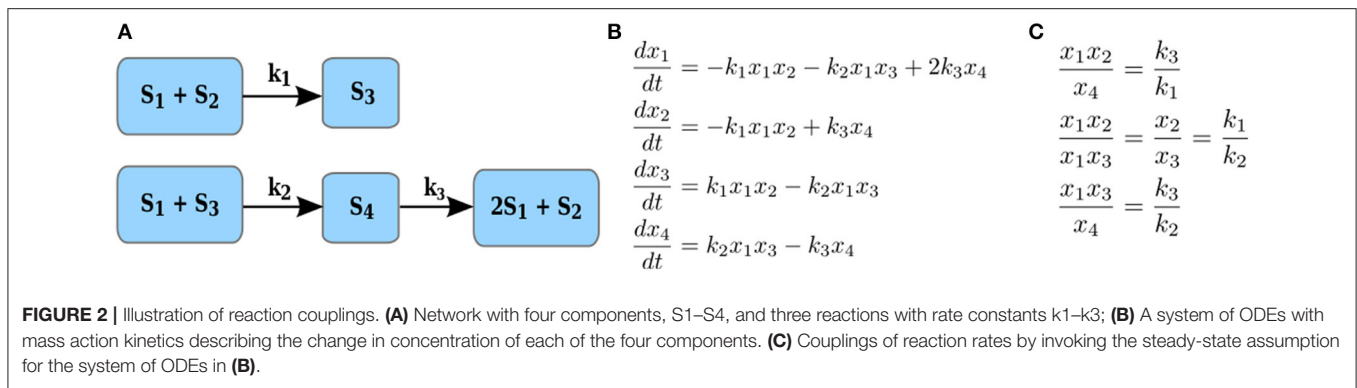
Here we define a constrained version of the maximal correlation coefficient which is motivated by modeling of metabolic networks and the principles of their operation. A metabolic network is a collection of metabolites and biochemical reactions through which they are transformed and/or exchanged with the environment. For instance, the network on **Figure 2A** transforms four metabolites, S_1 to S_4 via three reactions. Each reaction takes a non-negative linear combination of reactants metabolites, called substrate complex, and transforms it into a product complex, i.e., a non-negative combination of product metabolites. The coefficients in the non-negative linear combination denote the stoichiometry with which a metabolite enters a reaction as a substrate and/or product. For instance, in **Figure 2A**, $S_1 + S_2$ is the substrate complex of reaction r_1 and $2S_1 + S_2$ is the product complex of reaction r_3 . The difference between the stoichiometry of the product and substrate complexes defines a reaction vector stoichiometry gathered in the stoichiometric matrix N . In other words, the entry α_{ij} of the stoichiometric matrix N contains the molarity (integer number) with which metabolite i is involved as a substrate or product in the reaction j (Heinrich and Schuster, 1996).

The change in the levels of n metabolites, x_1, \dots, x_n can then be described by an ordinary differential equation (ODE), $\frac{dx}{dt} = N * v(x, k, t)$ where N denotes the stoichiometric matrix with dimensions $m \times n$, with m the number of metabolites and n the number of reactions, v denotes the reaction rate functions, x , the concentrations of the considered metabolites, k , the parameters on which the reaction rates depend, and t stands for time (Nägele et al., 2016). Even in the simple case of mass action kinetics for a network of

bimolecular reactions, the reaction rates, gathered in the time-dependent vectors, $v(x, k, t)$, are described by a non-linear function (Horn and Jackson, 1972). Metabolic reactions usually are not spontaneous and are catalyzed by enzymes. Every enzymatic reaction can in turn be divided into elementary reactions. Elementary reactions consider the formation and dissociation of enzyme-substrate complexes and provide the possibility for modeling variety of regulatory mechanisms (Segel, 1975). Elementary reactions can be effectively modeled with mass action, since they can be cast to explicitly consider the action of the enzyme (as in the derivation of the Michaelis-Menten kinetic). This was the approach taken in the large-scale model of *E. coli* (Khodayari and Maranas, 2016) and some of the subsystems in the models of photosynthesis (Arnold and Nikoloski, 2011). Therefore, due to the combined effect of multiple reactions and their regulation, metabolic data profiles can be regarded as observations from non-linear dynamics of metabolic networks.

Let x_i denote the concentration of a substrate component S_i (i.e., metabolite or enzyme). The rate of reaction j with a substrate complex $\sum_i \alpha_{ij} S_i$ with $\alpha_{ij} > 0$ where α_{ij} is the stoichiometry with which S_i enters the substrate complex of reaction j , under mass action kinetics is then expressed as $k_j \prod_i x_i^{\alpha_{ij}}$, where k_j denotes a rate constant. For instance, the rate of reaction with rate constant k_1 in **Figure 2A** is given by $k_1 x_1 x_2$, since S_1 and S_2 enter this reaction as substrates, each with stoichiometric coefficients of one; similarly, the rate of reaction with rate constant k_3 is given by $k_3 x_4$, since S_4 enters the reaction as a substrate with a stoichiometric coefficient of one.

To arrive at our approach termed Stoichiometric Correlation Analysis (SCA) we rely on the observation that metabolic networks, as part of inter-related cellular systems (e.g. transcription, translation, and signaling), operate toward providing robust functionality (Kitano, 2007; Wilson, 2013). We translate the robust functionality in the ability to ensure coupling of reaction rates (Millard et al., 2017). To formalize SCA, we provide the following definitions:



Definition 1: Two elementary reactions, p and q , have coupled rates under mass action kinetics if for any steady-state concentration of the participating components, gathered in x ,

$$\frac{k_p \prod_i x_i^{\alpha_{ip}}}{k_q \prod_i x_i^{\alpha_{iq}}} = \frac{k_p}{k_q} \prod_i x_i^{\alpha_{ip} - \alpha_{iq}} = \gamma_{pq}, \text{ where } \gamma_{pq} \text{ is a constant.}$$

For instance, at any steady state for the network in **Figure 2A**, whereby the equations in **Figure 2B** all equal 0, i.e., $\frac{dx_i}{dt} = 0$, reactions r_1 and r_2 , r_1 and r_3 , as well as r_3 and r_2 have coupled rates (see **Figure 2C**). We note that the same definition can be extended to hold in states which are not necessarily equilibrium points, allowing the treatment of time-series data. We would like to note that the coupling of reaction rates may lead to coupling of component concentrations which are not apparent by directly inspecting the reaction networks. For instance, due to the coupling of reactions r_1 and r_2 , the concentration of components x_1 and x_2 also are coupled, i.e., are proportional to each other.

Since the non-zero stoichiometric coefficients α_{ij} are integers in the set $I = \{1, \dots, 4\}$ (Basler et al., 2012), given two disjoint sets U_p and U_q of random variables denoting the data profiles for the metabolites, we next define the stoichiometric correlation.

Definition 2: Given two disjoint sets of random variables U_p and U_q , denoting two sets of metabolic profiles, the stoichiometric correlation is given by:

$$\sup \left\{ \frac{\text{cov}(f(U_p), g(U_q))}{\sqrt{V(f(U_p))} \sqrt{V(g(U_q))}} \mid V(f(U_p)) > 0, V(g(U_q)) > 0 \right\}, \quad (2)$$

with

$$f(U_p) = \sum_{i=1}^{|U_p|} \beta_{ip} \log(x_i), \beta_{ip} \in I$$

and

$$g(U_q) = \sum_{i=1}^{|U_q|} \eta_{iq} \log(x_i), \eta_{iq} \in I.$$

If U_p and U_q include the random variables corresponding to the metabolite levels in the substrate complexes of reaction p and q , respectively, the proposed definition of the stoichiometric correlation is a direct consequence of Definition 1, where the functions $f(U_p)$ and $g(U_q)$ are the logarithm of the rate of the reactions p and q under mass action kinetics, respectively. The presence of coupled rates in mass action for reactions p and q , after taking the logarithm, leads to stoichiometric correlation of value one for U_p and U_q . This observation pinpoints the main principle on which SCA relies.

If there exist multiple vectors β and η , yielding the same value of the stoichiometric correlation, we consider the one of smallest magnitude $\|\beta + \eta\|$. Therefore, stoichiometric correlation can be regarded as constrained maximal correlation, where the constraints pertain to the limited set of values that the entries of β and η are allowed to take following the stoichiometry of reactants. The transformation used in the constrained maximal correlation is explicitly non-linear, since the function f involves logarithms.

Clearly, the reverse direction also holds and can be used to draw hypotheses about the couplings in reaction rates and substrate complexes in a given metabolic network. To this end, we focus on the statistically significant stoichiometric correlations larger than a threshold value of 0.8 (to account for effects of noise and small deviations from coupling of reaction rates, per Definition 1). Note that since U_p and U_q are disjoint sets of random variables denoting the data profiles of metabolites, the entries of β and η are positive. For instance, given several steady-state measurements for the components in network on **Figure 2A**, the stoichiometric correlations with $U_p = \{S_1, S_2\}$ and $U_q = \{S_4\}$ is one with coefficients in β and η equal to one. Similar conclusions can be drawn for all components involved in the coupled reaction rates given in **Figure 2C**. The two definitions provide the basis for SCA: Since majority of reactions in real-world metabolic networks are mono- or bi-molecular (i.e., include one or two substrates), we determine the stoichiometric correlation, per Definition 2, between any two disjoint subsets of random variables of cardinality at most two. The implementation can either be achieved by: (1) solving a non-linear program with constraints for the coefficients $\beta, \eta \in I$ or (2) generating all subsets of at most two variables with different contribution due to stoichiometry, and determining the Pearson correlation coefficient only between the disjoint subsets. Since

the apparent couplings of elementary reaction rates, assumed to obey mass action kinetic. In addition, the findings reflect the additional couplings which cannot be directly related to reaction rates but are direct consequence of them (e.g., components S_1 and S_2 in the network on **Figure 2A** are coupled due to the coupling of the rates of reactions r_1 and r_2). We note that every enzymatic reaction $\sum_i \alpha_{ij} S_i \rightarrow \sum_i \alpha'_{ij} S_i$ (α_{ij}/α'_{ij} are the stoichiometry with which S_i enters the substrate/product complex of reaction j , respectively) can be rewritten to include the action of an enzyme $\sum_i \alpha_{ij} S_i + E \rightleftharpoons SE \rightarrow \sum_i \alpha'_{ij} S_i + E$ (E denotes the enzyme and SE the substrate-enzyme complex), so that the elementary reaction can be still modeled with mass action kinetic. Therefore, SCA can also include the effect of enzyme action. However, while this approach provides a way to model Michaelis-Menten kinetic which accounts for enzyme saturation, it does not explicitly consider the Michaelis-Menten form for the reaction kinetic.

To investigate the effects of the departure from the mass action kinetic for the considered reactions (with and without accounting for enzyme action), we considered three models of the tricarboxylic acid (TCA) cycle. All three models include the same metabolites, and differ only with respect to whether or not they include the effect of enzyme action and if they

use mass action kinetic or the more involved functional forms of the Michaelis-Menten kinetic. All reactions are considered reversible, and they are split into irreversible reactions in the cases in which mass action kinetic was employed. We used the TCA cycle model embedded in the kinetic model of *E. coli* (Khodayari et al., 2014). There are two parameterized variants for this model, one that includes mass action kinetic without enzyme action, and the other one which explicitly considers the formation of substrate-enzyme complexes. In addition we used a model of the TCA cycle with reversible Michaelis-Menten kinetic of the reaction rates (Singh and Ghosh, 2006).

To conduct the comparative analysis, we simulated the models metabolite concentrations with physiologically relevant randomly chosen initial values. The simulation time ranged from 0 to 1,280 min and metabolite concentrations were obtained at 21 time points identical to those used in the study of Caldana et al. (2011) (which we employ later in the empirical analysis). The simulated metabolite concentrations were used to calculate the stoichiometric correlations for 11 metabolites for each simulation and model separately. The distribution of the total number of stoichiometric correlations over 10 repetitions

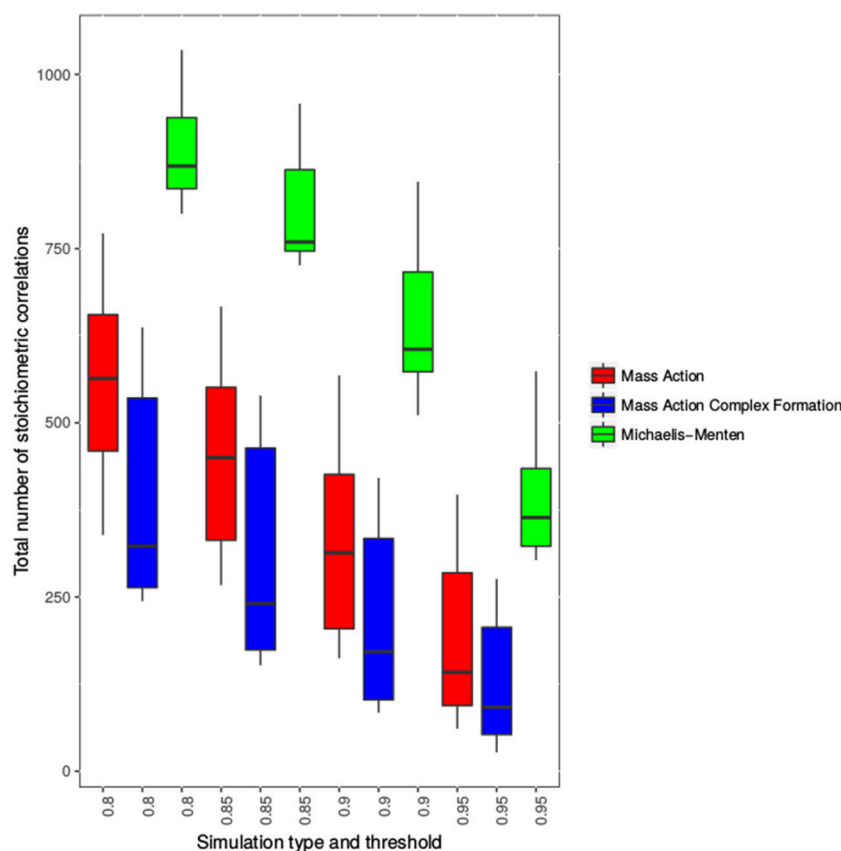


FIGURE 3 | Distribution of the number of stoichiometric correlations for three models of the TCA cycle. Shown are the distributions of the total number of stoichiometric correlations at four thresholds 0.8, 0.85, 0.9, and 0.95. The distributions for the mass action simulation are shown in red, the distributions for the substrate-enzyme complex mass action simulation are shown in blue, whereas the Michaelis-Menten simulation of the TCA cycle is shown in green.

of the procedure is shown in **Figure 3**, and all stoichiometric correlations (pairs, triplets and quadruples) are provided in Supplemental Table S1.

We found that the total number of stoichiometric correlations between the models with mass action kinetic was more similar with the increase in the considered threshold. In fact, at a threshold of 0.95, the distributions of the total number of stoichiometric correlations between the mass action models with and without the consideration of enzyme action largely overlapped. However, the consideration of reversible Michaelis-Menten kinetic results, on average, in at least three-fold increase in the total number of stoichiometric correlations (see **Figure 3**). These findings were supported by the results of the empirical cumulative distribution function (see **Figure 4**). The distribution of the Michaelis-Menten simulations are shifted to the right and show a higher proportion of correlations above 0.8. In addition, we report the quintiles of the correlation values in Supplemental Table S2.

Therefore, in the case of the TCA cycle models, we concluded that the findings from the assumption that the network is composed of elementary reactions modeled with mass action do not differ upon consideration of enzyme action. In these cases, the couplings corresponding to the stoichiometric correlations reflect the underlying reaction couplings. In contrast, the usage of Michaelis-Menten kinetic results in a considerably larger number of stoichiometric correlations, which cannot be brought in direct correspondence to the coupling of reaction rates and are challenging to mechanistically explain.

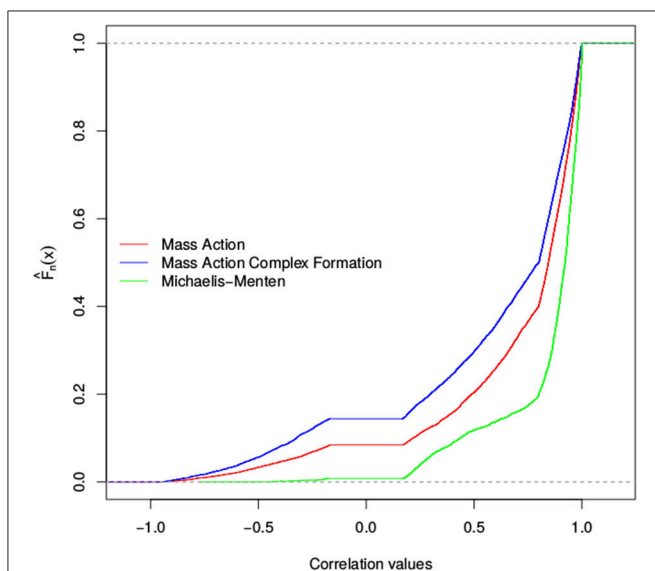


FIGURE 4 | Empirical cumulative distribution function of stoichiometric correlations for three models of the TCA cycle. Shown is the Empirical cumulative distribution of the total number of stoichiometric correlations of simulations of the TCA cycle. The distribution for the mass action simulation is shown in red, the distribution for the substrate-enzyme complex mass action simulation is shown in blue, whereas the Michaelis-Menten simulation is shown in green.

SCA Demonstrates Differences in the Stringent Response between *E. coli* and *A. thaliana*

The stringent response is one of the most important regulatory systems used by bacteria to adapt to environmental stresses. Upon sensing the environmental change, like nutrient limitation, the organism starts a series of reactions to redirect its metabolic fluxes. The stringent response is mediated by guanosine 3',5'-bis(pyrophosphate) (ppGpp) whose level is controlled by two enzymes, RelA and SpoT (Traxler et al., 2008). ppGpp has overall a large influence on several metabolic pathways and transcription and translation (Gallant, 1979; Mizusawa et al., 2008). The effect in metabolism involves the pathways of nucleotides, glycolytic intermediates, carbohydrates, lipids and fatty acid synthesis. It has been reported that there is evidence that the stringent response is evolutionary conserved from bacterial to photosynthetic bacterial to higher plants (Sugliani et al., 2016). Four homologs of these RelA and SpoT have been identified in Arabidopsis, and their role in green tissues and flower development has been well characterized (Masuda et al., 2008; Mizusawa et al., 2008). Since all these plant proteins are targeted to the chloroplast, it has been suggested that they control the stringent response in photosynthesizing organisms through mechanisms that may mimic those in bacteria. However, it remains unclear to what extent the molecular role of the homologs in *A. thaliana* are equivalent to those in *E. coli*.

To help answer this question, we analyzed the set of metabolites from the TCA cycle and the amino acid synthesis pathways from the two model organisms. We used these metabolites since ppGpp controls transcription and translation, which is ultimately reflected in the levels of amino acids. Moreover, a comparative analysis between the two organisms is only meaningful for the same set of metabolites. Altogether, we used the publicly available data profiles of three metabolites from the TCA cycle (i.e., malate, succinate, and fumarate) as well as 16 amino acid measured over seven and five conditions in *A. thaliana* and *E. coli* (see Materials and Methods).

The degree of coupling for metabolite *S* can be defined as the number of stoichiometric correlations above a given threshold τ in which the metabolite *S* participated. Based on the derivation of SCA, a higher degree of coupling on the same set of metabolites then implies maintenance of more coupled reaction rates over a set of studied conditions in one organism in comparison to another. We considered the significant stoichiometric correlations ($p \leq 0.05$, Benjamini-Hochberg corrected) 0.8, 0.85, 0.9, and 0.95, and compared them to classical Pearson correlations (**Table 1**). The quintiles of correlation values were additionally reported in Supplemental Table S2.

For the purpose of comparison, at all threshold values and in both species, we observed a decrease on the number of significant stoichiometric correlations for pairs of metabolites, compared to Pearson correlation (i.e., $|U_p| = |U_q| = 1$). The reduction in the number of significant stoichiometric correlations for metabolite pairs can be explained by the monotonic transformation of metabolite profiles. We would like to emphasize that the

TABLE 1 | Overview of the number of significant stoichiometric correlations at the considered thresholds for metabolic profiles of the stringent response in *E. coli* and *A. thaliana*.

Threshold	Organism	Stoichiometric Correlation				Pearson Correlation
		Total	Pairs	Triplets	Quadruples	Pairs
0.80	<i>A. thaliana</i>	3,419	13	579	2,827	24
	<i>E. coli</i>	3,301	9	517	2,775	10
0.85	<i>A. thaliana</i>	2,500	8	398	2,094	18
	<i>E. coli</i>	1,921	6	285	1,630	7
0.90	<i>A. thaliana</i>	1,821	6	285	1,530	15
	<i>E. coli</i>	597	1	76	520	2
0.95	<i>A. thaliana</i>	1,137	6	188	943	7
	<i>E. coli</i>	2	0	0	2	0

The total number of stoichiometric correlations is divided into three groups based on whether they involve pairs, triples, or quadruples of metabolites. Additionally, the number of significant Pearson correlations found in the dataset is shown.

result does not suggest that metabolites are linearly related, which would be contrast to what is expected from mechanistic understanding of metabolism.

However, SCA allows the analysis of stoichiometric correlations due to triples and quadruples of metabolites, which provides information about the presence of non-linear relationships via the couplings of reaction rates. For all considered thresholds, applying SCA with the *E. coli* data set resulted in a smaller number of stoichiometric correlations on triples and quadruples than the data set of *A. thaliana* (Table 1). For instance, at a threshold of $\tau = 0.85$, *E. coli* yielded 285 significant stoichiometric correlations due to triples while *A. thaliana* resulted in 398 such correlations; similarly, *A. thaliana* contained three-fold the number of stoichiometric correlations resulting from quadruple at $\tau = 0.9$ in comparison to *E. coli*. Therefore, based on these results we concluded that there was a stronger coupling of reaction rates of *A. thaliana* in comparison to *E. coli* during the stringent response.

Additionally, we can investigate overlapping pairs, triple and quadruple for each threshold. The small similarity of SCA findings was reflected in 65 and 442 stoichiometric correlations due to triple and quadruple, respectively, shared between the two species at a threshold value of 0.8 (see Supplemental Table S3). In line with this observation, the participation of metabolites in the stoichiometric correlations largely differed between the two species, as manifested in the lack of association between the metabolite coupling degrees. For instance, at a threshold value of 0.85, the metabolites with the largest coupling degrees in *E. coli* were: phenylalanine, threonine, proline and lysine, while in *A. thaliana* they included: isoleucine, leucine, tyrosine and lysine (see Supplemental Table S4). It must be noted that these results and interpretations warrant caution since the metabolite profiles from *A. thaliana* were obtained from entire Arabidopsis rosette rather than from isolated chloroplast, which may bias the drawn conclusions. The analysis can be conducted with compartment-specific metabolic profiles once they become available.

SCA Shows That Domestication in Wheat Is Associated with Loss of Regulatory Couplings

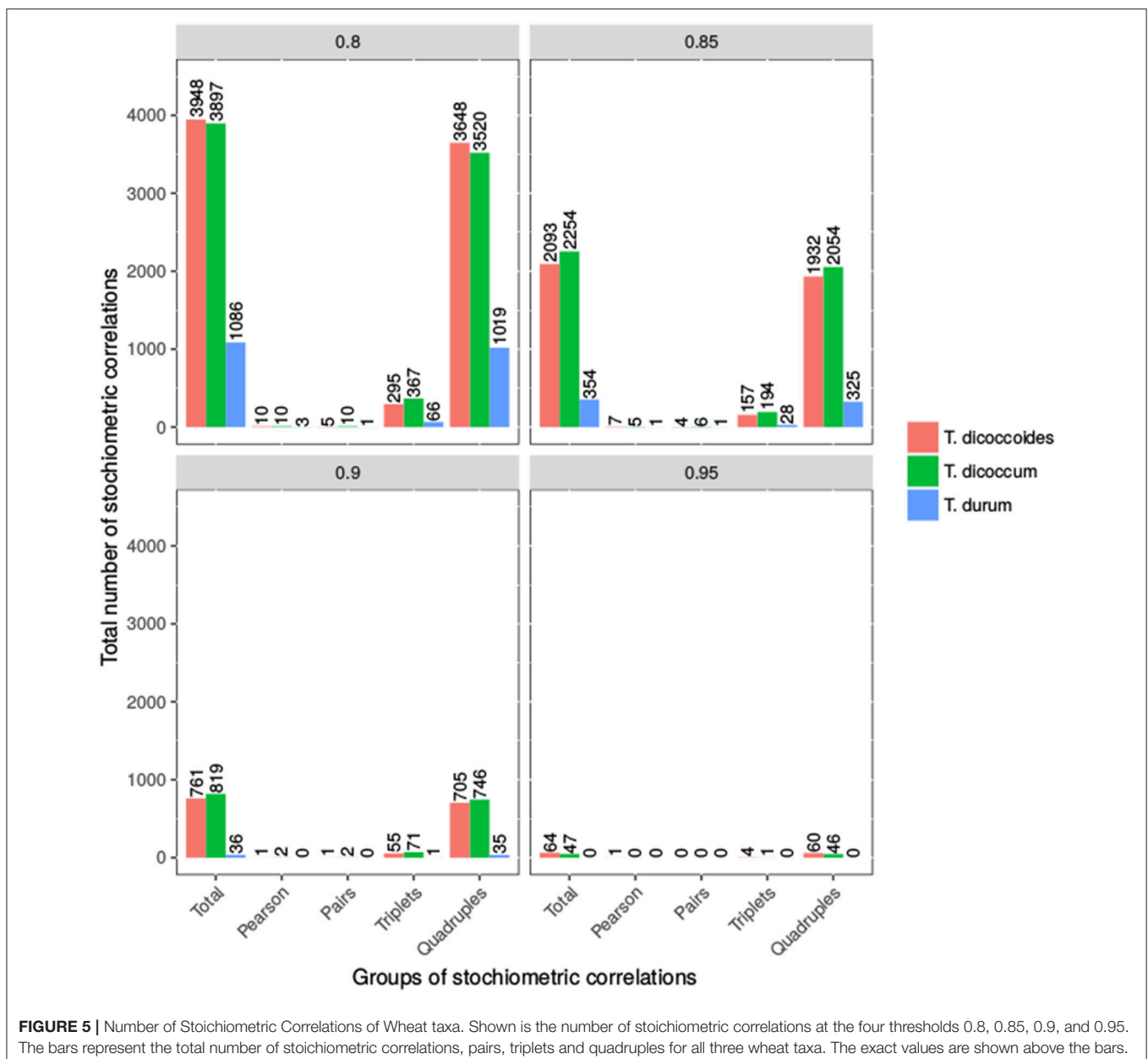
Domestication of tetraploid wheats, *Triticum turgidum* L., is an important evolutionary event for the human development. Emmer (*T. turgidum* ssp. *dicoccum*) was domesticated from wild emmer (*T. turgidum* ssp. *dicoccoides*) around 12,000 years ago (Nesbit and Samuel, 1998). Free-threshing tetraploid wheats (*T. turgidum* ssp. *turgidum*) subsequently originated from emmer, followed by the selection of durum wheat (*T. turgidum* ssp. *turgidum* convar. *durum*). Therefore, it has been suggested that the evolution of tetraploid wheats consists of at least two steps: primary domestication, from wild emmer to emmer, and secondary domestication, from emmer to durum wheat (Gioia et al., 2015).

Since important domestication-associated traits (e.g., the increase in seed size, the loss of dormancy Gepts and Papa, 2002) often necessitate alteration of metabolic process, we asked if application of SCA to metabolic profiles can be used to quantify the effect of domestication with respect to loss or gain of regulatory couplings. To this end, we used recently analyzed data about the phenotypic variation of primary metabolites in the kernels from three *T. turgidum* populations that represent both the primary and secondary domestication (Beleggia et al., 2016). Beleggia et al. (2016) determined that there were changes in content of specific metabolites, particularly amino acids and unsaturated fatty acids, associated with the primary and secondary domestication events. The resulting metabolic profiles of accessions within each taxon were also employed to construct Pearson correlation networks. Based on various properties of the correlation networks (e.g., shared correlations, centrality of metabolites) it was concluded that the difference between wild emmer and emmer was larger than the difference between wild emmer and durum wheat. In addition, it was found that durum wheat contained a larger number of significant correlations, followed by wild emmer and emmer. Therefore, surprisingly, the results from Pearson correlation analysis captured contrasting

findings in comparison to the evolutionary distance between the three analyzed taxa.

We applied SCA to contents of 22 metabolites from four compound classes (i.e., amino acids, sugars, organic acids, and alcohols) within each population at four threshold values (see Materials and Methods). These metabolites were selected based on their presence in every of the analyzed accessions to allow comparative analysis of the populations without the need of imputation as well as assumptions about the reasons for absence of detected metabolite. The number of stoichiometric correlations due to triples shared between emmer and wild emmer was the highest, followed by that between durum and wild emmer (at threshold of 0.8). At a threshold value

of 0.85, both emmer and wild emmer had one overlapping triple with durum. However, at a threshold value of 0.8, durum wheat shared more stoichiometric correlations due to quadruples with wild emmer than emmer. At thresholds of 0.85 and 0.9, durum shared the same number of quadruples with emmer and wild emmer. In all cases, only stoichiometric correlations due to quadruples were shared between all three populations (e.g., stoichiometric correlations at threshold of 0.85, Supplemental Table S6). Overall, we observed more triples and quadruples in *T. dicoccum* and *T. dicoccoides* in comparison to *T. durum* (see Figure 5). The observation was supported by the quintiles of the correlation values shown in Supplemental Table S2.



This finding implied that the loss of traits due to domestication and increase in seed size were associated with an overall loss of reaction couplings reflected in the smaller number of stoichiometric correlations in durum wheat in comparison to (wild) emmer (Supplemental Tables S5–S7). The metabolites involved in the largest number of stoichiometric correlations above a threshold value of 0.85 in wild emmer included glycine, threonine, aspartate, serine and glutamate; in emmer, these metabolites included serine, leucine, threonine, and glutamate, while in durum wheat they consisted of fructose, glucose, glutamate, and asparagine (see Supplemental Table S7). Altogether, the application of SCA identified a shift in importance of regulatory role of sugars in comparison of organic and amino acids which is in line with the increase in seed size due to the need for more cell wall components.

To further validate our results from the three wheat taxa, we included data of six different tomato species into the analysis. We compared the domesticated *S. lycopersicum* (M82) to the group of the other five species, as their fruits drastically differ from those of M82 (Schauer et al., 2005). However, it has to be noted that there is no clear lineage from the undomesticated plants to the M82. Additionally, the combination of the different tomato species might result in an inclusion of additional noise. Nevertheless, it is a necessary step to have the needed amount of replicates per metabolite. Overall, the tomato data set contains 43 metabolites common to the analyzed species. In line with the results of wheat, we observed fewer stoichiometric correlations for M82 than for the undomesticated wildtype tomato (Table 2, Supplemental Table S8). The exception is the threshold of 0.8;

in this case, the M82 has roughly 4,000 more pairs, triples and quadruples than the wild type tomato. At a threshold value of 0.85 the M82 has still around 300 stoichiometric correlations more than the wild type species. With increasing threshold, however, the number of significant stoichiometric correlations decreases in M82 more than in the wild type. This finding was reflected in the different quintiles of the correlation values for the two species (Supplemental Table S2). The metabolites with the largest number of stoichiometric correlations above a value of 0.9 in wildtype tomato are erythritol, cysteine, succinic acid and beta-alanine, while in M82, they include: leucine, putrescine, dehydroascorbic and sucrose (Supplemental Table S4).

A very similar scenario was considered with the strawberry accessions *F. vesca* (wild) and *F. ananassa* (domesticated and commercially available) without direct domestication lineage between the two species. In contrast to our observations in wheat and tomato, the domesticated strawberry exhibits a higher number of stoichiometric correlations above all thresholds (Table 2, Supplemental Table S9). The reason for this finding may lay in the different ploidy of the investigated organisms, namely, *F. ananassa* is an octaploid organism, whereas *F. vesca* is diploid with a rather small genome.

The application of SCA to metabolomics data from domestication implies a new principle which underlies this agronomically and evolutionary important process; namely, optimizing a given trait could be accomplished by breaking the existing regulatory mechanisms, reflected in the coupling of

TABLE 2 | Overview of number of significant stoichiometric correlations at different thresholds for the considered tomato and strawberry species.

Threshold	Organism	Stoichiometric Correlation				Pearson Correlation
		Total	Pairs	Triplets	Quadruples	Pairs
0.80	Tomato wildtype	19,519	8	1,245	18,266	27
	M82	23,571	15	1,824	21,732	12
	<i>F. vesca</i>	1,346	6	204	1,136	6
	<i>F. ananassa</i>	2,374	12	433	1,929	5
0.85	Tomato wildtype	9,291	5	588	8,698	20
	M82	9,539	5	688	8,846	4
	<i>F. vesca</i>	504	2	73	429	3
	<i>F. ananassa</i>	2,075	10	366	1,699	5
0.90	Tomato wildtype	3,741	3	255	3,483	11
	M82	1,493	1	112	1,380	0
	<i>F. vesca</i>	135	1	22	112	1
	<i>F. ananassa</i>	1,153	2	185	966	1
0.95	Tomato wildtype	818	1	76	741	3
	M82	21	0	0	21	0
	<i>F. vesca</i>	1	0	0	1	0
	<i>F. ananassa</i>	423	1	56	366	1

The total number of stoichiometric correlations is divided into three groups based on whether they involve pairs, triples, or quadruples of metabolites. Additionally, the number of significant Pearson correlations found in the dataset is shown.

the biochemical reaction rates, which in turn provides a greater space of possibilities on which selection can operate.

CONCLUSION

Here we proposed a constrained extension to the concept of maximal correlation, based on the concept of reaction rate coupling in networks of metabolic reactions. The concept of reaction couplings forms the core of the stoichiometric correlation analysis. The constraints in the maximal correlation are due to the values which the linear combinations of log-transformed metabolic profiles are allowed to take. SCA facilitates the comparison of data sets on the same metabolites between two scenarios with the idea of comparing and contrasting the degree of coupling. By determining the stoichiometric correlations of metabolic profiles from the TCA cycle and amino acid synthesis, we showed that *E. coli* stringent response is differently (and less strongly) controlled than that of *A. thaliana*. Therefore, while the enzymes underlying the stringent response are preserved in these two model organisms, their integration in the metabolic networks may have evolved different regulatory action. In addition, SCA can be used to investigate the differences between undomesticated and domesticated species, and to determine if the difference can be ascribed to alterations in metabolic couplings brought about by various regulatory mechanisms. Based on this idea, we demonstrate that stoichiometric correlations from metabolic profiles from natural variation in wild and domesticated species indicate that domestication is associated with loss of regulatory control. Therefore, our findings provide the basis for future flux-oriented studies toward mechanistic understanding of this important evolutionary process.

REFERENCES

- Arnold, A., and Nikoloski, Z. (2011). A quantitative comparison of calvin-benson cycle models. *Trends Plant Sci.* 16, 676–83. doi: 10.1016/j.tplants.2011.09.004
- Basler, G., Grimbs, S., Ebenhoh, O., Selbig, J., and Nikoloski, Z. (2012). Evolutionary significance of metabolic network properties. *J. R. Soc. Interface* 9, 1168–1176. doi: 10.1098/rsif.2011.0652
- Beleggia, R., Rau, D., Laidò, G., Platani, C., Nigro, F., Fragasso, M., et al. (2016). Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol. Biol. Evol.* 33, 1740–1753. doi: 10.1093/molbev/msw050
- Breiman, L., and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation: rejoinder. *J. Am. Stat. Assoc.* 80, 614–619. doi: 10.2307/2288477
- Çakir, T., Hendriks, M. M. W. B., Westerhuis, J. A., and Smilde, A. K. (2009). Metabolic network discovery through reverse engineering of metabolome data. *Metabolomics* 5, 318–329. doi: 10.1007/s11306-009-0156-4
- Caldana, C., Degenkolbe, T., Cuadros-Inostroza, A., Klie, S., Sulpice, R., Leisse, A., et al. (2011). High-density kinetic analysis of the metabolomic and transcriptomic response of *Arabidopsis* to eight environmental conditions. *Plant J.* 67, 869–884. doi: 10.1111/j.1365-313X.2011.04640.x
- Chen, Y., Zhang, R., Song, Y., He, J., Sun, J., Bai, J., et al. (2009). RRLC-MS/MS-based metabolomics combined with in-depth analysis of metabolic correlation network: finding potential biomarkers for breast cancer. *Analyst* 134, 2003–2011. doi: 10.1039/b907243h

AUTHOR CONTRIBUTIONS

ZN conceived the project and wrote the article with contribution of all authors. KS performed most of the analysis and analyzed the data. RB provided data and helped interpret the findings related to domestication. NO designed parts of the algorithm and helped with its implementation.

FUNDING

KS was funded by the International Max Planck Research School on Plant Growth at the Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany. NO and ZN acknowledge the support of the Max Planck Society and PlantaSYST funded by the EU's Horizon 2020 Programme.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Roberto Papa at the Universita Politecnica delle Marche, Ancona, Italy, for kindly providing the metabolomics data from the three wheat taxa. The authors would also like to thank Dr. Detlef Ulrich from the Julius Kühn-Institut (JKI) - Bundesforschungsanstalt für Kulturpflanzen, Institut für ökologische Chemie, Pflanzenanalytik und Vorratsschutz, for providing the strawberry data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02152/full#supplementary-material>

- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161. doi: 10.1038/81137
- Gallant, J. A. (1979). Stringent control in *E. Coli*. *Annu. Rev. Genet.* 13, 393–415. doi: 10.1146/annurev.ge.13.120179.002141
- Gebelein, H. (1941). Das statistische problem der korrelation als variations- und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *ZAMM Zeitschrift für Angewandte Mathe Matik Mechanik* 21, 364–379. doi: 10.1002/zamm.19410210604
- Gepts, P., and Papa, R. (2002). “Evolution during domestication,” in *eLS* (Chichester: John Wiley & Sons Ltd.). doi: 10.1038/npg.els.0003071
- Gioia, T., Nagel, K. A., Beleggia, R., Fragasso, M., Ficco, D. B. M., Pieruschka, R., et al. (2015). The impact of domestication on the phenotypic architecture of durum wheat under contrasting nitrogen fertilisation. *J. Exp. Bot.* 66, 5519–30. doi: 10.1093/jxb/erv289
- Hackett, S. R., Zanotelli, V. R. T., Xu, W., Goya, J., Park, J. O., Perlman, D. H., et al. (2016). Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* 354, 6311. doi: 10.1126/science.aaf2786
- Harrell, F. E. Jr. with contributions from Charles Dupont and many others. (2015). *Hmisc: Harrell Miscellaneous*. R package version 3.17-0. Available online at: <http://CRAN.R-project.org/package=Hmisc>
- Heinrich, R., and Schuster, S. (1996). *The Regulation of Cellular Systems*. Boston, MA: Springer.
- Horn, F., and Jackson, R. (1972). General mass action kinetics. *Arch. Ration. Mech. Anal.* 47, 81–116. doi: 10.1007/BF00251225

- Ideker, T., and Krogan, N. J. (2012). Differential network biology. *Mol. Syst. Biol.* (Basel). 8, 565. doi: 10.1038/msb.2011.99
- Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhäuser, D., et al. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* (Basel). 6, 364. doi: 10.1038/msb.2010.18
- Kaddurah-Daouk, R., Kristal, B. S., and Weinshilboum, R. M. (2008). Metabolomics: a global bio-chemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.* 48, 653–683. doi: 10.1146/annurev.pharmtox.48.113006.094715
- Khodayari, A., and Maranas, C. D. (2016). A genome-scale *Escherichia coli* kinetic metabolic k-ecoli457 satisfying flux data for multiple mutant strains. *Nat. Commun.* 7:13806. doi: 10.1038/ncomms13806
- Khodayari, A., Zomorrodi, A. R., Liao, J. C., and Maranas, C. D. (2014). A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data. *Metab. Eng.* 25, 50–62. doi: 10.1016/j.ymben.2014.05.014
- Kitano, H. (2007). Towards a theory of biological robustness. *Mol. Syst. Biol.* (Basel). 3, 137. doi: 10.1038/msb4100179
- Koshland, D. (1970). The molecular basis for enzyme regulation. *Enzymes* 1, 341–396. doi: 10.1016/S1874-6047(08)60170-5
- Krumsiek, J., Bartel, J., and Theis, F. J. (2016). Computational approaches for systems metabolomics. *Curr. Opin. Biotechnol.* 39, 198–206. doi: 10.1016/j.copbio.2016.04.009
- Lancaster, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* 44, 289–292. doi: 10.1093/biomet/44.1-2.289
- MATLAB (2015). *MATLAB, 2015a The MathWorks, Inc.*, Natick, MA.
- Masuda, S., Mizusawa, K., Narisawa, T., Tozawa, Y., Ohta, H., and Takamiya, K.-I. (2008). The bacterial stringent response, conserved in chloroplasts, controls plant fertilization. *Plant Cell Physiol.* 49, 135–141. doi: 10.1093/pcp/pcm177
- Millard, P., Smallbone, K., and Mendes, P. (2017). Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*. *PLoS Comput. Biol.* 13:e1005396. doi: 10.1371/journal.pcbi.1005396
- Mizusawa, K., Masuda, S., and Ohta, H. (2008). Expression profiling of four RelA/SpoT-like proteins, homologues of bacterial stringent factors, in *Arabidopsis thaliana*. *Planta* 228, 553–562. doi: 10.1007/s00425-008-0758-5
- Nägele, T., Furtauer, L., Nagler, M., Weismann, J., and Weckwerth, W. (2016). A strategy for functional interpretation of metabolomic time series data in context of metabolic network information. *Front. Mol. Biosci.* 3:6. doi: 10.3389/fmolb.2016.00006
- Nesbit, M., and Samuel, D. (1998). Wheat domestication: archaeobotanical evidence. *Science* 279:1431. doi: 10.1126/science.279.5356.1431e
- Nguyen, H. V., Müller, E., Vreeken, J., Efros, P., and Böhm, K. (2014). “Multivariate maximal correlation analysis,” in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)* (Beijing), 775–783.
- Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614
- Rényi, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungaricae* 10, 441–451. doi: 10.1007/BF02024507
- Schauer, N., and Fernie, A. R. (2006). Plant metabolomics: towards biological function and mechanism. *Trends Plant Sci.* 11, 508–516. doi: 10.1016/j.tplants.2006.08.007
- Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* 24, 447–454. doi: 10.1038/nbt1192
- Schauer, N., Zamir, D., and Fernie, A. R. (2005). Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J. Exp. Bot.* 56, 297–307. doi: 10.1093/jxb/eri057
- Schwahn, K., Omranian, N., and Nikoloski, Z. (2017). *KSchwahn/Stoichiometric-Correlation V.1.0*.
- Segel, I. (1975). *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems*. New York, NY: John Wiley.
- Singh, V. K., and Ghosh, I. (2006). Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in *Mycobacterium tuberculosis*, and its application to assessment of drug targets. *Theor. Biol. Med. Model.* 3:27. doi: 10.1186/1742-4682-3-27
- Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W. (2003). Observing and interpreting correlations in metabolomics networks. *Bioinformatics* 19, 1019–1026. doi: 10.1093/bioinformatics/btg120
- Sugliani, M., Abdelkefi, H., Ke, H., Bouveret, E., Robaglia, C., Caffarri, S., et al. (2016). An ancient bacterial signaling pathway regulates chloroplast function to influence growth and development in *Arabidopsis*. *Plant Cell* 28, 661–679. doi: 10.1105/tpc.16.00045
- Sumner, L. W., Mendes, P., and Dixon, R. A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62, 817–836. doi: 10.1016/S0031-9422(02)00708-2
- Sun, X., Langer, B., and Weckwerth, W. (2015). Challenges of inversely estimating Jacobian from metabolomics data. *Front. Bioeng. Biotechnol.* 3:188. doi: 10.3389/fbioe.2015.00188
- Toubiana, D., Fernie, A. R., Nikoloski, Z., and Fait, A. (2013). Network analysis: tackling complex data to study plant metabolism. *Trends Biotechnol.* 31, 29–36. doi: 10.1016/j.tibtech.2012.10.011
- Traxler, M. F., Summers, S. M., Nguyen, H.-T., Zacharia, V. M., Hightower, G. A., Smith, J. T., et al. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol. Microbiol.* 68, 1128–1148. doi: 10.1111/j.1365-2958.2008.06229.x
- Ulrich, D., and Olbricht, K. (2013). Diversity of volatile patterns in sixteen *Fragaria vesca* L. accessions in comparison to cultivars of *Fragaria xananassa*. *J. Appl. Bot. Food Qual.* 86, 37–46. doi: 10.5073/JABFQ.2013.086.006
- Vance, W., Arkin, A., and Ross, J. (2002). Determination of causal connectivities of species in reaction networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5816–5821. doi: 10.1073/pnas.022049699
- van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry, 3rd Edn*. Amsterdam: Elsevier.
- Wilson, D. F. (2013). Regulation of cellular metabolism: programming and maintaining metabolic homeostasis. *J. Appl. Physiol.* (1985). 115, 1583–8. doi: 10.1152/jappphysiol.00894.2013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer WZ and handling Editor declared their shared affiliation.

Copyright © 2017 Schwahn, Beleggia, Omranian and Nikoloski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.