Plant metabolism drives plant development and plant-environment responses, and data readouts from this cellular level could provide insights in the underlying molecular processes. Existing studies have already related key in vivo leaf gas-exchange parameters with structural traits and nutrient components across multiple species. However, insights in the relationships of leaf gas-exchange with leaf primary metabolism are still limited. We investigated these relationships through a multispecies meta-analysis approach based on data sets from 17 published studies describing net photosynthesis (A) and stomatal (gs) conductances, alongside the 53 data profiles from primary metabolism of 14 species grown in different experiments. Modeling results highlighted the conserved patterns between the different species. Consideration of species-specific effects increased the explanatory power of the models for some metabolites, including Glc-6-P, Fru-6-P, malate, fumarate, Xyl, and ribose. Significant relationships of A with sugars and phosphorylated intermediates were observed. While gs was related to sugars, organic acids, myo-inositol, and shikimate, gm showed a more complex pattern in comparison to the two other species. Some metabolites, such as malate and Man, appeared in the models for both conductances, suggesting a metabolic coregulation between gs and gm. The resulting statistical models provide the first hints for coregulation patterns involving primary metabolism plus leaf water and carbon balances that are conserved across plant species, as well as species-specific trends that can be used to determine new biotechnological targets for crop improvement.

Plant metabolism drives plant development alongside a considerable number of plant-environment interactions (Brunetti et al., 2013). This is largely accomplished through the joint modulation of enzyme and metabolite levels. Characterization of metabolites can be carried out by different profiling technologies, comprising the field of metabolomics (Brunetti et al., 2013). Recent advances in metabolic profiling technologies have led to the identification and quantification of 1,000 to 2,000 metabolites, while the number for the whole plant kingdom is estimated to be ~200,000, with a single plant species including up to a few thousand metabolites (e.g. ~5,000 for Arabidopsis [Arabidopsis thaliana]; Obata and Fernie, 2012). A metabolic profile comprises data about (changes in) the levels of metabolites in a given system. The data gathered via metabolic profiling may provide insights in the regulation of metabolic networks. Additionally, they can be used in elucidating gene functions, designing crop breeding strategies, understanding plant growth and plasticity, and characterizing plant-community responses to biotic and abiotic stresses (Meyer et al., 2007; Obata and Fernie, 2012; Brunetti et al., 2013).

In addition, metabolic profiles obtained from these technologies have been used to build statistical models of complex phenotypic traits. For instance, statistical models of good performance have been developed for biomass, one of the agronomically important plant traits (Meyer et al., 2007; Sulipce et al., 2013). The
success of the modeling approach ultimately depends on the ability to translate the statistical relationships in future crop breeding purposes (Fernie, 2012). While the first attempts of modeling complex agronomically relevant traits based on metabolic profiles were carried out in the model plant Arabidopsis (Timm et al., 2012; Sulpicie et al., 2013), recent approaches have expanded these efforts to other plant species, including important crops (Araújo et al., 2011b; Witt et al., 2012; Riedelsheimer et al., 2013).

Biomass depends on the underlying metabolic profile of the plant system, while this profile may reflect different physiological, structural, and compositional traits related to biomass. For instance, advances in the study of leaf photosynthesis has resulted in multispecies and biome data sets that allow us to prove the relationships between leaf mass area, specific leaf area, and nutrient content (e.g. leaf nitrogen and phosphorus concentrations; Reich et al., 1999; Wright et al., 2004; Kattge et al., 2011; Walker et al., 2014). These studies identified several clear trends, namely, maximum photosynthesis ($A_{\text{m}}$) correlates positively with nitrogen and phosphorus leaf content and negatively with leaf lifespan and leaf mass area (Reich et al., 1999), independently of biome, growth form, or plant functional type (Wright et al., 2004). Recently, it was reported that the maximum carboxylation velocity of Rubisco ($V_{\text{cmax}}$) was positively correlated with nitrogen content, while phosphorous increases the sensitivity of the relationship (Walker et al., 2014).

Despite these advances, there are only few studies that have examined the relationships between metabolic profiles and plant in vivo photosynthetic rates and conductances (Aranjuelo et al., 2011; Warren, 2011; Warren et al., 2011, 2012). Elucidating relations between metabolites and conductances is particularly relevant, since conductances are the diffusive physiological determinants that mainly drive the final plant productivity and plant’s relation with the environment. Identifying statistical relationships between metabolism and photosynthesis can provide important insights into how changes in particular metabolites, potentially achievable via metabolic engineering, may modulate photosynthesis and, in turn, improve plant growth, water use efficiency, and stress tolerance.

Carbon and water fluxes inside the leaf follow Fick’s First Law of diffusion for CO₂ assimilation rates ($A$) and transpiration ($E$): 

$$A = g_{sc} \cdot (C_a - C_i)$$

$$E = g_{sw} \cdot (W_i - W_a)$$

where $g_{sc}$ and $g_{sw}$ are the stomatal conductance to CO₂ and water vapor; $C_a$ and $C_i$ are the concentrations of CO₂ in the atmosphere and the substomatal cavity; and $W_i$ and $W_a$ are the concentrations of water and CO₂ vapor in the substomatal cavity and the atmosphere. Clearly, stomata are the common structures for carbon and water exchange ($g_{sc} = g_{sw}/1.6$) between the inner part of the leaf and the atmosphere, processes that ultimately drive photosynthesis. In addition to carbon and water exchange, there are other limiting processes inside the leaf. One of these involves the different consecutive barriers to CO₂ flux from the substomatal cavities to the carboxylation sites in the chloroplast stroma and is characterized by the mesophyll conductance ($g_{ms}$; Flexas et al., 2013). Concerning the gms at steady state conditions Equation 1 can be rearranged as follows:

$$A = g_m \cdot (C_i - C_c)$$

where $C_c$ is the CO₂ concentration at the carboxylation site of Rubisco. Taking into account the previous statements, $C_c$ highly depends on the CO₂ supply, limited by the diffusive parameters ($g_{sc}$ and $g_{sw}$), but also on the CO₂ consumption by Rubisco, driven by the maximum carboxylation rates ($V_{cmax}$) and the photosynthetic metabolism (Gago et al., 2014). Indeed, the widely used photosynthetic model (Farquhar et al., 1980) considers that photosynthesis is limited by the biochemical relations of the maximum carboxylation capacity of Rubisco ($V_{cmax}$), maximum rate of electron transport ($J_{max}$), and the use of triose-phosphates (Díaz-Espejo et al., 2012). The first limitation (operating mostly under CO₂ limiting conditions) is related to $V_{cmax}$ and depends on the kinetic properties of Rubisco (at saturating ribulose-bisphosphate concentrations) as well as O₂ and CO₂ concentrations. The second limitation (which is more important at CO₂ saturating conditions) is based on the assumption that photosynthesis depends on the rates of photosynthetic electron transport to produce NADPH and ATP, related to $J_{max}$. The third limitation can be considered as a product of feedback regulation and is related to the availability of inorganic phosphate needed for ribulose-bisphosphate regeneration, which depends on the relative rates of triose-phosphates production and inorganic phosphate release during Suc and starch synthesis (Díaz-Espejo et al., 2012).

Metabolite profiling has only recently been linked to in vivo leaf gas exchange (Nunes-Nesi et al., 2011; Warren et al., 2011, 2012; Araújo et al., 2011b; Araújo et al., 2011b; Medeiros et al., 2015). The stomatal pore is the first diffusional barrier for the influx of CO₂ for photosynthesis. The magnitude of the stomatal pore can directly affect net photosynthetic rate by limiting the amount of substrate (CO₂) for fixation. Conversely, the reduction of the internal CO₂ concentration ($C_i$) by photosynthesis and the transport and accumulation of photosynthesis-derived metabolites in the surrounding guard cells is known to regulate stomatal movements (Lawson and Blatt, 2014). Given this interplay between stomatal movements and photosynthesis, it is not surprising that both processes are tightly coregulated. However, it is important to note that guard cells are known to respond to a wide range of both endogenous and environmental signals (Kim et al., 2010). As a result, photosynthesis is not the only process that influences stomatal movements. It is also important to...
consider the complex interaction with mesophyll and biochemical limitations, which may explain why the relationship between \( A \) and \( g_s \) is not always unique (Messinger et al., 2006). Due to the high expression of ATP-dependent ion transporters (H\(^+\)-ATPase) found in guard cells (Bauer et al., 2013) and their high metabolic activity, guard cells have a relatively large energetic demand (Zeiger et al., 2002; Medeiros et al., 2015). However, even at high respiratory rates (Vani and Raghavendra, 1994; Araújo et al., 2011b), the amount of energy required for the regulation of stomatal movements seems to be greater than the synthesis capacity of guard cells. This fact connects the metabolism of mesophyll cells with stomatal movements since the former have been shown to support the metabolism of guard cells by providing metabolites and ATP (Zeiger et al., 2002; Wang et al., 2014). Furthermore, there are several pieces of evidence showing that accumulation of photosynthesis-derived metabolites, especially sugars and organic acids, in the apoplast space of guard cells can regulate stomatal movements (Lu et al., 1995; Kang et al., 2007; Lee et al., 2008; Araújo et al., 2011b; Fujita et al., 2013). Nevertheless, the mechanisms underlying the modulation of \( g_s \), \( g_m \), and \( A \) by sugars and organic acids are still unknown. Moreover, it is not clear whether these processes are conserved across plant species.

Mesophyll conductance, \( g_m \), a limiting factor for \( CO_2 \) diffusion to the carboxylation sites in the stroma, is usually tightly coregulated with the stomatal conductance, \( g_s \) (Flexas et al., 2013). However, in vivo measurements of \( g_m \) are not as straightforward as those of \( A \) and \( g_s \). The reasons for this lack of comparability is that \( g_m \) can be estimated based on three distinct models employing different theoretical approaches, methodologies, and assumptions (Pons et al., 2009; Flexas et al., 2012). These approaches comprise (1) the “Harley” method, combining chlorophyll fluorescence and gas-exchange data (Harley et al., 1992); (2) the “Ethier” curve-fitting method (Ethier and Livingston, 2004); and (3) the “Evans” method, employing an online carbon isotope discrimination methodology (Evans et al., 1986). There is another approach to estimate \( g_m \) in a manner based on leaf anatomical properties, which agrees with in vivo approaches in several angiosperms as well as in several fern species (Tosens et al., 2012; Tomás et al., 2013; Carriquí et al., 2015). However, several concerns were recently published about the reliability and uncertainties of the different in vivo methodologies (Tholen et al., 2012; Sun et al., 2014). Moreover, a wide range of parameters can largely affect \( g_m \) since, in addition to the leaf anatomical properties (leaf and cell wall thickness and chloroplast surface exposure to mesophyll air spaces), biochemical processes that are not yet fully understood are likely involved (e.g., related to aquaporins and carbonic anhydrase activities; Flexas et al., 2015). Furthermore, the usual tight coregulation between \( g_m \) and the other physiological traits (\( A \) and \( g_s \)) leads to an extremely complex scenario for the assessment of this trait (Flexas et al., 2013; Gago et al., 2014).

Therefore, photosynthesis is tightly linked to the rest of metabolism through coregulation. The challenge is then to rely on the existing publicly available data sets from metabolomic and ecophysiological studies to identify and further dissect the interactions between \( A \), \( g_s \), and \( g_m \). In this context, we carried out a multispecies meta-analysis aiming to identify and interpret statistical relationships conserved or species-specific through multivariate regression modeling (instead of the classical pairwise correlation approach traditionally employed in plant physiology literature) between leaf primary metabolites and \( A \), \( g_s \), and \( g_m \) using data on these parameters alongside metabolic profiles from several species grown under different conditions. The resulting models can be used to determine new biotechnological targets for crop improvement and can increase our understanding about metabolic network interactions and responses in leaf water and carbon balances conserved across plant species.

RESULTS

Relationship between Net Photosynthesis and Conductances

In our compiled data set, which includes 14 species grown under optimal and stress conditions (water stress and nutrient deficiency) and/or transgenic lines, each of the three simultaneously assessed physiological traits varied in a wide range: net photosynthetic rate, \( A \), from 0.8 to 30.6 \( \mu \)mol \( CO_2 \) m\(^{-2}\) s\(^{-1}\); stomatal conductance, \( g_s \), from 0.01 to 0.62 \( mol \) \( CO_2 \) m\(^{-2}\) s\(^{-1}\); and mesophyll conductance, \( g_m \), from 0.07 to 0.42 \( mol \) \( CO_2 \) m\(^{-2}\) s\(^{-1}\) (Supplemental Data Set S1). A statistically significant positive correlation between \( A \) and \( g_s \) (\( P \) value < 0.001) was observed after pooling all data (Fig. 1A). The correlation between \( A \) and \( g_m \) was not as strong when pooling all data; instead, only data from control treatments showed a significant correlation (Fig. 1B). Moreover, the relationship between \( g_s \) and \( g_m \) was very disperse and nonsignificant (Fig. 1C), suggesting uncoupling of the two traits that may depend on the experimental conditions, in the case of several \( Acacia \) and \( Eucalyptus \) species (Warren, 2011; Warren et al., 2011).

Photosynthesis and Conductances Driven by Primary Metabolism

Our assembled data set on net photosynthesis and both types of conductances contained the profiles of 53 metabolites. These data allowed us to build statistical models for the metabolites, as responses, in terms of the net photosynthesis and conductances as predictors. To this end, for each metabolite, we tested a linear model with \( A \), \( g_s \), and \( g_m \) as independent variables without grouping for the genetic background (i.e., species); this first modeling strategy coincides with classical multivariate regression. We also considered the six expanded models that, in addition to the fixed effects for the three
Figure 1. Relationships between net photosynthesis 
($A$) and stomatal conductance ($g_s$), net photosynthesis 
($A$) and mesophyll conductance ($g_m$), and both conductances ($g_s$ and $g_m$) in a multispecies 
data set (Supplemental Data Set S1). Data were 
plotted for different species under “control” (blue dots) or “treatment” (red dots) conditions. Species 
analyzed in the plots include Acacia maidenii, Acacia 
mangium, Eucalyptus areacea, Eucalyptus delegatensis, Eucalyptus globulus, Eucalyptus nitens, 
Eucalyptus regnans, and Eucalyptus socialis. Data 
were extracted from Warren (2011) and Warren 
et al. (2011).

The lack of good explanatory power for these models 
suggested that the genetic background may play a role, 
which is next examined by means of mixed-effects 
models. Models with random effects for both the 
treatment and the slope of $A$ were selected for nine of the 53 
analyzed metabolites (previously only one showed a 
correlation with this trait). These metabolites were His, 
Ile, Leu, Lys, Ser, gluconate, Xyl, Glc-6-P (G6P), and F6P. 
Including random effects considerably increased the 
marginal coefficients of determination in comparison to 
the simple models with fixed factors described above: 
the minimum was 23.44% (Leu) and the maximum 
was 89.69% (Ile), followed by G6P and F6P (>89%) 
with mean of 66.71% and median of 70.75% (Fig. 3; 
Supplemental Data Set S2), indicating that these 
relationships tend to be modulated in a species-specific 
manner. The largest variances in the random intercept 
and slope (higher species-specific role) for $A$ were 
observed in the models of His, Lys, and gluconate. The 
only negative significant fixed effects were found in the 
models for Ile ($g_s$ value = 0.0005; $g_m$ value = 0.016), 
while positive significant fixed effects were found for 
G6P ($g_s$ value = 0.0006; $g_m$ value = 0.019) and F6P ($g_s$ 
value = 0.002). Interestingly, $A$ did not show any 
significant fixed effect in the models with random slope 
and intercept for this variable (Supplemental Data Set S2).

Models with random effects for both the intercept 
and slope of $g_s$ were selected for 12 metabolites (just five 
were previously selected with fixed effects): Asp, Cys, 
GABA, Glu, Phe, Tyr, hydroxybenzoic acid, gluconate, 
malate, maleate, Man, and mannitol. However, the 
marginal coefficients of determination did not increase 
considerably, unlike the case of random effects for $A$, in 
comparison to the simple models considered above: 
The minimum was 0.07% (mannitol) and the maximum 
was 5.45% (malate), followed by Asp and hydroxybenzoic 
acid (>4.20%) with mean of 2.11% and median of 
1.28% (Fig. 3; Supplemental Data Set S2). The largest 
variances in the random intercept and slope (for $g_s$) 
were observed in the models of gluconate, Cys, Phe, 
and Tyr, which are highly species specific. Since all are

Gago et al.
closely related to oxidative stress protectants it seems reasonable to interpret that this finding reflects the different strategies for protection against reactive oxygen species in the studied plants. In all models, $g_m$ also had a significant fixed effect (except for Asp, gluconate, and mannitol), which was not the case for $A$ and $g_s$. This effect was negative in all discussed models, except for gluconate and mannitol (Supplemental Data Set S2).

Models with random effects for both the intercept and the slope of $g_m$ were selected for 28 metabolites (Arg, Cys, GABA, Glu, Gln, His, Ile, Leu, Lys, Pro, Ser, Thr, Tyr, Val, and fumarate), with random slopes for $A$ and $g_s$ for most of these metabolites, but showed significant fixed effects for $g_m$. However, the marginal coefficients of determination were the smallest (smaller than 0.07%) in comparison to those with random slope for $A$ and $g_s$. The largest variances in the random intercept and slope (for $g_m$) were observed in the models of Arg, Cys, His, and Thr. The significant fixed effect differed between metabolites. For instance, there was no significant fixed effect of $A$; fixed effect of $g_m$ was significant for malate, while $g_m$ exhibited significant fixed effect for Arg, GABA, Glu, His, Pro, Tyr, Val, hydroxybenzoic acid, malate, maleate, arabinose, Glc, Man, and mannitol (Supplemental Data Set S2).

Altogether, the analysis based on linear mixed-effects models indicated that the inclusion of random slope for $A$ considerably improved the explanatory models for a limited number of amino acids and sugars. In addition, $g_m$ exhibited significant fixed effects in the majority of the models only upon inclusion of random intercepts and slopes for $A$ and $g_s$. Furthermore, almost all of the metabolites whose models included a random slope for $A$ or $g_s$ could also be modeled with random slope for $g_m$, although at expense of explanatory power. This points to the interplay between $A$, $g_s$, and $g_m$; however, models with interactions could not be considered due to lack of data.

**Photosynthesis and Conductances Driven by Primary Metabolism**

In the following analysis, we examined to what extent the profiles of $A$, $g_s$, and $g_m$ as multiple responses (dependent variables) can be explained by the metabolites as predictors (i.e. independent variables). To this end, we used partial least squares (PLS) for building the statistical models. The model of good (root) mean squared error of prediction ([R]MSEP) and coefficient of multiple determination ($R^2$) bias-corrected via
Figure 3. A, Schematic diagram for CO₂ and H₂O diffusion pathways along a leaf cross section. H₂O\textsubscript{l} and H₂O\textsubscript{v}, water at liquid and gaseous state; C\textsubscript{a}, ambient CO₂ concentration; C\textsubscript{i}, CO₂ substomatal cavity concentration; C\textsubscript{c}, CO₂ concentration at the site of Rubisco carboxylation within the chloroplastic stroma; g\textsubscript{s}, stomatal conductance; g\textsubscript{m}, mesophyll conductance; W\textsubscript{i} and W\textsubscript{a}, concentration of water vapor at the substomatal cavity and the atmosphere. B, Relevant metabolic pathways within a guard cell and metabolites significantly related to g\textsubscript{s}. C and D, Relevant metabolic pathways within a mesophyll cell and metabolites significantly related to g\textsubscript{m} and A, respectively. Metabolites in red and blue were significantly related to physiological traits by our linear mixed-effects models with fixed and random effects, respectively. Metabolites with significant negative relationships are underlined. Metabolites found with both approaches were highlighted according to the model with fixed effects. Metabolites significantly related to physiological traits in the PLS modeling approach are marked with an asterisk. Metabolites in black were present in our data set but were not significantly related to any trait, and those in gray were absent from our data set. Dashed arrows indicate steps involving several reactions. CBC, Calvin-Benson cycle; TPT, triose-phosphate transporter; ST, Suc transporter; TMT, Suc transporter; PT, pyruvate transporter; OT, putative oxaloacetate transporter; AT, amino acid transporter; DT, dicarboxylate transporter; OGT, 2-oxoglutarate transporter; CLT, Cl\textsuperscript{-} transporter; KAT, K\textsuperscript{+} transporter; PEP, phosphoenolpyruvate; Pyr, pyruvate; HBA, hydroxybenzoic acid; OAA, oxaloacetate; Ac-CoA, acetyl coenzyme A; Cit, citrate; cis-Aco, cis-aconitate; Isoc, isocitrate; 2-OG, 2-oxoglutarate; Succ-CoA, succinyl coenzyme A; Succ, succinate; Fum, fumarate; Mal, malate; Put, putrescine; DehydroAsc, dehydroascorbate. Amino acids were abbreviated using the standard three-letter code. The green arrow represents the effect of malate on the vacuolar Cl\textsuperscript{-} transporter.
cross-validation scheme can be helpful in detecting which metabolites may affect the responses via various feedbacks (Supplemental Fig. S2). Inspection of the (R) MSEP indicated that the smallest value across the three dependent variables was obtained when using two components (Supplemental Fig. S2A). Therefore, we next inspected the coefficients of the PLS model built from two components explaining 19.74% of the variance in $g_m$, having a good compromise with $R^2$ (Supplemental Fig. S2B). The coefficients of the models for $A$ and $g_s$ were highly negatively correlated ($-0.99$) but weakly correlated to the coefficients in the model for $g_m$ (0.39 with $g_s$ and $-0.47$ with $A$; Supplemental Fig. S3). This implies that the variables of high (small) influence on $A$ have a negligible (large) effect on $g_s$ (Supplemental Data Set S2). This finding implies that metabolites that have a positive effect on $A$ tend to have a negative effect on $g_m$, a fact that must be considered when developing metabolic engineering strategies to improve photosynthesis. This is particularly interesting since, from Figure 1, $A$ and $g_s$ are positively correlated, but the metabolites used to explain them do not have coefficients of the same sign in the respective models.

For completeness, Figure 4 illustrates a heat map based on the values of the coefficients of individual metabolites in the models for each of the three traits. Interestingly, most of the sugars (Glc, Fru, Suc, Gal, and raffinose) and hexose-phosphates (G6P and F6P) had a positive effect on $A$; however, many amino acids (e.g., Asn, Gln, His, Ile, Leu, and Met) and organic acids (e.g., citrate, fumarate, glycerate, and maleate) entered the models with negative coefficients (Fig. 4A). In the case of $g_s$ and $g_m$ different trends can be observed, namely, several amino acids had positive coefficients in the models for $g_s$ and $g_m$, but some organic acids (e.g., maleate and shikimate) had coefficients of opposite sign in the models of $g_s$ and $g_m$ (Fig. 4B). In addition, it is interesting to note that both showed the same negative relationship with the main sugars (Glc, Fru, and Suc; Fig. 4B). The largest (positive) coefficients in the model of $g_m$ included glycerol, citrate, asparagine, maltose, and Met, while the smallest (negative) coefficients were associated with ethanolamine, hydroxybenzoic acid, Tyr, and Pro (Fig. 4B; Supplemental Data Set S2).

**DISCUSSION**

Our goal was to study the relationships between metabolite profiles from primary metabolic pathways and in vivo leaf photosynthesis and conductances. For this purpose, we collected published data from 17 articles, which cover 14 species analyzed at different developmental stages and environmental conditions. The range of the in vivo physiological traits was in agreement with previous photosynthetic multispecies meta-analysis (Wright et al., 2004; Flexas et al., 2013; Gago et al., 2014; Walker et al., 2014). Nevertheless, the amount of data available for modeling the relationship between $g_m$ and metabolite profiles is limited. Analysis of these parameters showed that $A$ is positively correlated with both $g_s$ and $g_m$ (Fig. 1, A and B), in agreement with the fact that both conductances determine the flux of CO$_2$ that reaches the Rubisco carboxylation sites in the chloroplast stroma (Flexas et al., 2013). However, the relationship between $g_s$ and $g_m$ was largely scattered and nonsignificant (Fig. 1C), suggesting a certain degree of "uncoupling" as shown, for instance, during drought acclimation and rewatering in different environments in herbaceous and woody plants (Galle et al., 2009, 2011). This is a similar case for most of the $g_m$ data
in our experimental data set, gathered from the drought recovery experiment of Warren et al. (2011) in *Eucalyptus* and *Acacia* sp. While the positive relationships of \( A \) with \( g_c \) and \( g_m \) suggest a direct positive relationship between both conductances, \( C_c \) and \( C_m \) are not just regulated by the conductances themselves. Moreover, these CO\(_2\) concentrations are also regulated by \( C_{av} \), the photosynthetic activity, light, leaf temperature, and Rubisco amounts and kinetics. For the same reasons, the relationship between \( A \) and \( g_m \) can be variable along different environments, particularly when considering different species (Farquhar et al., 1980; Flexas et al., 2013). In any case, the “supply” of CO\(_2\) (conductances) drives the final CO\(_2\) assimilation, as it was previously demonstrated in multispecies data sets; in fact, this is one of the main targets for the improvement of water use efficiency (i.e. the ratio of carboxylation versus water losses in transpiration) and the bottleneck for productivity in semi-arid areas, considering water scarcity and predictions of climate change (Flexas et al., 2013; Gago et al., 2014).

To the best of our knowledge, this is the first approach for multispecies primary metabolic profiling in relation to photosynthesis-limiting physiological traits, with overlap in the metabolic profiles between species and studies ranging from 33.9 to 98.1%. Previous multispecies meta-analysis only studied photosynthesis and/or its biochemical parameters in relation to the content of essential nutrients such as nitrogen and phosphorous, establishing the “leaf economics spectrum” through a global view approach (Wright et al., 2004; Walker et al., 2014). In addition, several studies used metabolomics to predict traits such as growth and/or biomass (which can be considered the final, composite outputs of metabolism; Meyer et al., 2007; Sulpice et al., 2013), while others targeted the role of some intermediates of the tricarboxylic acid (TCA) cycle in driving stomatal conductance and thus photosynthesis (Araújo et al., 2011b). In this work, we developed a more complex approach to leaf water and carbon balances in a multispecies data set, considering the feedback effects between the actors (in vivo gas-exchange and primary metabolic networks), in addition to the analysis of conserved and/or species-specific patterns captured in a variety of statistical models beyond the classical pairwise correlation approaches.

**Photosynthesis and Primary Metabolism: Major Sugars and Succinate as Markers of High Photosynthetic Activity**

In the modeling approaches employed in this work, soluble sugars (Glc, Fru, and Suc) were proportionally the most important parameters related to photosynthetic activity (Figs. 3 and 4; Supplemental Data Set S2), taking place mainly in the mesophyll cells. These statistical relationships could be easily explained given that the Calvin-Benson cycle produces triose-phosphates that are either used to replenish the CO\(_2\) acceptor (ribulose-bisphosphate) or to synthesize the major photosynthetic products, starch and Suc (Stitt et al., 2010).

When a conserved strategy was applied to model the metabolic response using \( A \) as input, it had significant effect (\( P \) value = 0.037) only for Fru, while the relationship was not significant for Suc or Glc (Fig. 3; Supplemental Data Set S2). Conversely, when metabolites were used to predict the photosynthetic response by means of PLS modeling, we found a positive effect of all major soluble sugars (Suc, Glc, and Fru) and succinate (Fig. 4; Supplemental Data Set S2). These results strongly agree with previously reported data showing that accumulation of Suc is related to the photosynthetic capacity (Stitt et al., 1983; Kruckeberg et al., 1989; Neuhaus et al., 1990). Considering our data set, these relationships probably respond to the wide range covered by \( A \) among different species (0.8–30.6 \( \mu \)mol CO\(_2\) m\(^{-2}\) s\(^{-1}\)). Thus, it seems likely that sugar metabolism can be used as a marker of the photosynthetic activity. Indeed, we found several metabolites of the Suc synthesis pathway that were positively related to \( A \). When linear mixed modeling was performed considering random effects (i.e. species specific), we found the variance of nine metabolites to be explained by \( A \) with positive effect including G6P and F6P (Fig. 3; Supplemental Data Set S2). It has been previously shown that the levels of these metabolites are related with \( A \) (Stitt et al., 1983; Gerhardt et al., 1987). Even though these metabolites rise marginally with increases in light intensity, the flux to Suc increases at a higher rate (Neuhaus et al., 1990). This is accomplished by relieving inhibition of the cytosolic Fru-1,6-bisphosphatase (by lowering Fru-2,6-bisphosphate levels) and activating Suc-6-P synthase (by reducing its phosphorylation). This feed-forward mechanism provides an effective way to increase carbon flux into Suc with minor changes in steady state levels of metabolites (Neuhaus et al., 1990) and explains the positive correlation observed for hexose-phosphates and Suc with \( A \).

In addition to soluble sugars, succinate was also positively related with \( A \) (Fig. 4; Supplemental Data Set S2). It has been reported that reduction of TCA cycle intermediates (malate and fumarate) in Arabidopsis plants overexpressing a heterologous NADP-malic enzyme led to reduction in leaf thickness, photosynthesis, and biomass. Malate and fumarate are accumulated during the day and used in the following dark period to fuel respiration. Therefore, lower levels of these organic acids in mutant plants promoted the use of reduced substrates for respiration, such as fatty acids and proteins (Zell et al., 2010). Whether succinate plays a similar role than malate and fumarate remains an open question. Recently, \(^{13}\text{CO}_2\) labeling of Arabidopsis whole rosettes demonstrated that biosynthesis of succinate and amino acids (such as Ile, Leu, and Lys) are directly related to carbon assimilation (Ishihara et al., 2015). Even though labeling was incomplete for these metabolites, a considerable portion of newly fixed carbon was directed to the synthesis of succinate and these amino acids during light hours. Our results are in close agreement with these observations, given that these molecules were positively related to \( A \) in the modeling approaches employed in our work. Little information is
available concerning the biotechnological potential of manipulating amino acids metabolism. Conversely, it is known that antisense inhibition of the succinate dehydrogenase iron-sulfur subunit enhances $A$ and $g_s$ in a mechanism that might be mediated by malate accumulation (Araújo et al., 2011b). Therefore, succinate metabolism may be an effective target for photosynthesis improvement.

Stomatal Conductance and Primary Metabolism: The Role of Organic Acids and Other Putative Regulators

It is known that the magnitude of stomatal opening can range according to the accumulation of metabolites from mesophyll in the symplast and apoplast of guard cells. These metabolites can induce stomatal opening or closure, depending on the concentration and the combination of stimuli (Lawson and Blatt, 2014). However, precise information about the mechanisms by which mesophyll cells influence stomatal movements is still lacking. Interestingly, the models used here were able to identify two groups of metabolites, namely, sugars and organic acids, which are known to regulate stomatal movement. While soluble sugars (Suc, Fru, and Glc) and succinate entered the model of $A$ with positive coefficients, they had opposite signs in the models for $g_s$. The production of sugars in the mesophyll cells and their transport to guard cells represent one of the main points of connection between mesophyll and guard cells and may present a central role in the tight regulation of $A$-$g_s$ (Lawson and Blatt, 2014). This hypothesis is supported by the fact that in periods of high photosynthetic rate, sugars produced in mesophyll cells are transported to the guard cell apoplast space by the transpiration stream, where they accumulate and induce stomatal closure by a process that remains not well understood (Lu et al., 1995; Kang et al., 2007). Alternatively, Suc, and probably Glc and Fru (given the presence of a presumed hexose carrier in guard cells; Ritte et al., 1999), can enter guard cells and induce stomatal closure in a mechanism dependent of hexokinase and mediated by abscisic acid (ABA; Kelly et al., 2013). This explains the opposite correlation we found for sugars with $A$ and $g_s$, and provides further evidence for a role of Suc and its derivatives, Glc and Fru, in connecting mesophyll and guard cells. Interestingly, our results showed that Man also is positively related with $g_s$. There is evidence showing that this sugar can act as an osmolyte and its accumulation in leaves inhibits the uptake of Suc, Fru, and Glc, while decreasing starch levels (Lucas and Wilson, 1987). It seems likely that Man may positively regulate stomatal opening, acting as an osmolyte in guard cells and/or by negatively regulating the accumulation of sugars and starch, which has been quantitatively related to stomatal aperture (Outlaw and Manchester, 1979; Lasceve et al., 1997; Kelly et al., 2013; Prasch et al., 2015). Taken together, these results reveal a complex regulatory network involving Suc and Man, and possibly starch, in regulating stomatal movements, a process where these metabolites seem to play a signaling role (apart from their osmotic potential).

Besides the role of the above-mentioned sugars, it is also known that organic acid accumulation can regulate stomatal movements. Malate, which showed a positive coefficient for $g_s$ in the conserved modeling, is the main organic acid related to stomatal movements (Fernie and Martinhoia, 2009). It seems that the flux of malate between vacuole, cytosol, and the apoplastic space of guard cells is pivotal to regulate stomatal movements (Chen et al., 2012; Hills et al., 2012). Indeed, it has been shown that this metabolite can act as a counter-ion of $K^+$ in the vacuole of guard cells or as a signaling molecule to activate guard cell vacular chloride channels during stomatal opening (Hedrich and Marten, 1993; De Angeli et al., 2013). Further evidence for the importance of malate and their products and precursors for guard cell metabolism and stomatal movements is demonstrated by the higher $g_s$ found in transgenic plants for enzymes related to organic acid metabolism (Laporte et al., 2002; Nunes-Nesi et al., 2007; Araújo et al., 2011b; Penfield et al., 2012) or guard cell malate transporters (Meyer et al., 2010; Medeiros et al., 2015). The findings of these works are reflected across our entire data compilation, which revealed a strong and positive correlation between the levels of malate and $g_s$ (Fig. 3; Supplemental Data Set S2).

It is well known that the TCA cycle intermediates are important for stomatal movements (Fernie and Martinhoia, 2009; Araújo et al., 2011b). Besides malate (discussed above), positive correlations between $g_s$ and fumarate, Asp, Glu, and Orn were also observed. Furthermore, succinate and citrate were negatively correlated with $g_s$ following the PLS and linear mixed effect models, respectively. Altogether, these results further support the idea that metabolites associated with the TCA cycle are involved in the regulation of stomatal movements. Given the similarity in the chemical structure of fumarate in comparison to malate, its pattern of accumulation throughout the day, and that both seem to be involved in the photosynthetic fluxes distribution during light-induced stomatal opening, fumarate is believed to exert the same function as malate in guard cells (Araújo et al., 2011a; Daloso et al., 2015). Indeed, plants that have higher accumulation of malate and fumarate show higher $g_s$ (Nunes-Nesi et al., 2007; Araújo et al., 2011b; Medeiros et al., 2015). In contrast, transgenic tobacco plants that accumulate less succinate and 2-oxoglutarate in leaves and less succinate and citrate in their guard cells also have higher $g_s$ (Daloso et al., 2016). These findings further support our results and suggest that the different parts of the TCA cycle, which is supposed to operate in a noncyclic mode in illuminated leaves (Cheung et al., 2014), have differential contribution to stomatal movements. It seems likely that malate and fumarate would be important osmolytes and signaling molecules in the $A$-$g_s$ relationship, while the roles of citrate and succinate may be related to maintenance of the energy status and
homeostasis. For instance, the degradation of succinate may be a mechanism to induce ATP production, while citrate could act as a substrate for the synthesis of Glu and Orn, which showed a positive correlation with \textit{g}_s.

However, the role of Glu and Orn for guard cell regulation has yet to be investigated.

Beyond sugars and organic acids, our modeling approach also determined that \textit{myo}-inositol and shikimate entered the models of \textit{g}_s with negative coefficients. The polyol \textit{myo}-inositol is synthesized from G6P, and it is known that its derivative \textit{myo}-inositol-3-phosphate (IP$_3$) has an important role in the signaling pathway of ABA-induced stomatal closure (Schroeder et al., 2001). The binding of ABA in G-receptors present in guard cell membranes (Liu et al., 2007; Wan and Liu, 2008) induces changes in phospholipase-C proteins, which in turn induce production of IP$_3$ (Hunt et al., 2003). IP$_3$ accumulated in the cytosol activates the efflux of Ca$^{2+}$ from the vacuole and endoplasmic reticulum, which induce the efflux of ions from guard cells and consequently stomatal closure (Fan et al., 2004).

Therefore, the negative correlation of \textit{myo}-inositol with \textit{g}_s found in our data could perhaps be anticipated from the literature. The other metabolite that was also negatively related to \textit{g}_s was shikimate, a known precursor of flavonols and phenylpropanoids. It was recently described that flavonols develop an important role buffering the ABA-induced oxidative response in the guard cells. Arabidopsis flavonol-less mutants showed a quicker stomatal response, demonstrating that they can act as dampeners regulating the oxidative burst mediated by ABA and the subsequent stomatal closure by the guard cells (Watkins et al., 2014). Overall, these results suggest that shikimate and flavonol/phenylpropanoid metabolism are also involved in regulating the guard cell metabolic network, and this mechanism could be highly conserved between different plant species.

Mesophyll Conductance and Primary Metabolism: A Complex Interplay

Results from our modeling approaches displayed a certain degree of agreement with the intrinsic complexity of \textit{g}_m and/or the concerns and assumptions of the different methodologies to estimate it, i.e. an important biological complexity driving CO$_2$ diffusion through different scales, from enzymatic activities (aquaporins and carbonic anhydrases) to the leaf structure and anatomical properties, like cell wall thickness and chloroplast surface exposed to the air spaces of the mesophyll (see “Introduction”). For instance, only one and five metabolites were significantly associated with \textit{A} and \textit{g}_s, respectively, when only fixed effects were considered; however, \textit{A} and \textit{g}_s had significant effect in the models for nine and 12 metabolites, respectively, in the species-specific modeling approach (i.e. when random effects were considered). Conversely, \textit{g}_m was found to have significant effect on 19 and 28 metabolites in the case of conserved and species-specific models, respectively (Fig. 3; Supplemental Data Set S2).

Interestingly, there were two main groups of metabolites (amino acids and sugars) in which \textit{g}_m was significant for the models with fixed effects, each of them representing 42% of the total significant metabolites found with our modeling approaches. Amino acids accounted for half of metabolites with significant effects, followed by organic acids. The latter included malate and fumarate, both intermediates of the TCA cycle, which were previously related with the regulation of guard cell turgor and stomatal opening (Araújo et al., 2011b; Daloso et al., 2015), as it was previously indicated in the models for \textit{g}_s (Figs. 3 and 4). In a certain sense, \textit{g}_m models showed an intermediate position between \textit{A} and \textit{g}_s (Fig. 3). These results fit well with the high degree of coupling between both conductances traditionally observed in wider multispecies data sets (Flexas et al., 2013; Gago et al., 2014) and reported elsewhere; however, under certain conditions, such as the recovery from a water stress period, both conductances can be uncoupled (Fig. 1C).

Surprisingly, some sugars with significant effect for \textit{g}_m did not have significant effect for \textit{A} or \textit{g}_s. These metabolites include Gal, arabinose, and raffinose for the conserved pattern, in addition to Xyl for the species-specific modeling (Fig. 3; Supplemental Data Set S2). Interestingly, these sugars and the hydroxybenzoic acid derivatives could be mostly related with the synthesis and maintenance of the cell wall structure and its pectin matrix (Schulzler et al., 1992; Alonso et al., 2010; Chen et al., 2013; Franková and Fry, 2013). Cell wall cannot be considered as a fixed element in cell metabolism. Plants invest an important amount of resources in the synthesis and turnover of the cell wall (mostly sugar polymers). For example, it has been described that its glycosidic components can account for 18% of the total dry weight biomass (Williams et al., 2008; Masakapalli et al., 2010; Chen et al., 2013; Franková and Fry, 2013). The role of sugars in relation with the cell wall can be highly dynamic. For instance, Arabidopsis \textit{mur}1 mutants with \textit{l}-Fuc deficiency showed a severe, stunted phenotype; however, when Fuc was exogenously applied, the plants recovered the wild-type phenotype (O’Neill et al., 2001).

In addition, different cell wall composition (including the pectin matrix) can affect the final thickness, structure, and porosity, thus determining the size of the molecules that can pass through it, as well as the tortuosity of the diffusion pathway (Carpita et al., 1979; Baron-Epel et al., 1988; Franková and Fry, 2013; Flexas et al., 2015). In our data set, we analyzed 14 different species, from trees (such as \textit{Eucalyptus} and \textit{Acacia} species) to herbs (like Arabidopsis, tomato \textit{Solanum lycopersicum}, and rice \textit{Oryza sativa}; Supplemental Data Set S1). It is well known that cell wall composition can strongly differ between plant species and groups, especially in those from the Graminaceae family, ferns, and equisetums (Fry et al., 2008; Franková and Fry, 2013). Additionally, it has been described pooling together 61 different species that the cell wall thickness shows a negative significant relationship with \textit{g}_m.
It is important to note that the composition of the plasma and chloroplastic membranes, as well as the chloroplast exposure to mesophyll air spaces, can largely affect $g_{\text{m}}$. It has been suggested that molecules from the raffinose family can help in stabilizing membranes and proteins of the thylakoid membrane under freezing temperatures, thus preserving chloroplast integrity. Additionally, these sugars might prevent cellular dehydration by deep insertion between lipid layers (Van den Ende, 2013). Ethanolamine is a known precursor of phosphatidylethanolamine and phosphatidylcholine, both major phospholipids in eukaryotic membranes (Kwon et al., 2012). Membrane properties and composition directly affect their permeability to solutes, water, and protons, potentially affecting protein membrane activities (Burgos et al., 2011). Moreover, lipidomes have previously been demonstrated to exhibit high predictive power with regard to agronomic traits in maize (*Zea mays*; Riedelsheimer et al., 2013). Considering previous multispecies leaf anatomical analysis, cell and chloroplast membrane limitations for CO$_2$ diffusion are not so important as cell wall thickness and chloroplast exposure but certainly impose a significant limitation to $g_{\text{m}}$ (Tomás et al., 2013; Flexas et al., 2014). However, their role and relationship with this trait is currently unknown, even considering their importance affected by environmental factors, as temperature that drives important permeability changes in the membranes that finally affect photosynthesis (Evans and von Caemmerer, 2013).

Finally, further efforts must be made to deeply explore and validate the knowledge extracted from the empirical modeling methodology employed in this study. The relationships observed can be confirmed by conducting an independent experiment, gathering the same metabolite profiles and gas-exchange measurements as in this study, and calculating the error in the prediction and/or through an experiment-based approach using metabolic engineering to obtain mutants (in comparison to the wild type) that target the modeled traits (e.g. increase in the level of a metabolite with positive coefficient and decrease in the metabolite entering the model with a negative coefficient). Subsequently, it must be tested if these mutants have affected photosynthesis, growth, water use efficiency, and stress tolerance in comparison to their wild types and the predicted response, in accordance with the built models.

**CONCLUSION**

In this work, we assessed the mutual relationships between photosynthesis and stomatal and mesophyll conductances with metabolite profiles from pathways of primary plant metabolism. Our modeling results from the multispecies meta-analysis showed some conserved patterns from herbaceous to woody plants for these traits. When a species-specific strategy was employed, models increased their explanatory power, suggesting certain patterns depend exclusively on each analyzed species. Photosynthesis was widely associated with sugars, phosphorylated intermediates, and organic amino acids that are linked to biosynthetic processes during the day. Our results for stomatal conductance together with evidence from past studies implied that the balance in the level of sugars, organic acids, myo-inositol, and shikimate may regulate stomatal movements through a complex metabolic network and, consequently, photosynthesis. What remains unclear, and thus deserves further exploration, is the identity of the transporters involved in importing these metabolites to guard cell. In addition, it is important to understand when (in which period of the day, which period of stress, and in response to which stimulus) and where (apoplast, symplasm, vacuole, chloroplast, and mitochondrion) these metabolites would be accumulated to regulate stomatal movements. In the case of $g_{\text{m}}$ the models showed a greater complexity when compared to the other traits. Interestingly, some metabolites (like malate) were selected for both $g_{\text{m}}$ and $g_{\text{m}}$, suggesting a tight coregulation between these parameters, as it has been suggested in the literature. In addition, the significance of some sugars mostly related with cell wall composition and structure (such as arabino, Xyl, and Gal) opens new biotechnological opportunities to regulate this trait and thus improve water use efficiency. More work is needed to extend this analysis to other species from different life forms and climate origin, in order to decipher the relationships between metabolism and in vivo carbon and water leaf exchanges. Linking ecophysiology and metabolomics offers a new perspective to improve our understanding of leaf water and carbon balances, providing opportunities to better define biotechnological targets related to water use efficiency, productivity, and plant stress tolerance.

**MATERIALS AND METHODS**

**Data Set Compilation**

Publically available data of $A$, $g_{\text{m}}$, and $g_{\text{c}}$ combined with metabolic profiles, collected in parallel in the same samples, were compiled from 17 published articles describing results for 14 different species. The data set comprised a maximum of 71 data points of $A$, while compilation of $g_{\text{m}}$ and $g_{\text{c}}$ was reduced to 43 and 21 data points, respectively. Metabolic profiles included data for 53 metabolites, covering amino acids, organic acids, sugars, sugar-alcohols, and phosphorylated intermediates, from the 14 genetic backgrounds (species) under transgenic or environmental (water stress and nutrient deficiency) perturbations. Metabolite overlap between all works ranged from 33.9 to 98.1%. The species included in the data set belong to herbs and woody plants (five herbs; eight trees). Plants were grown under different environments (greenhouses and growth chambers) at optimal environmental conditions (maximum photosynthetic levels were recorded) and under different treatments (such as drought, nutrient deficiency, and transgenic lines with some affected trait). The employed data are detailed in Supplemental Data Set S1. In the following sections, we describe the general methodologies employed in the reviewed papers for acquisition of gas-exchange, fluorescence, and metabolite data.

**Gas-Exchange and Fluorescence Measurements**

In vivo leaf water and carbon balances and fluorescence measurements were done by clamping fully expanded leaves into the chamber of an open infrared gas analyzer system. All gas-exchange measurements compiled in this data set were performed between 22 and 25°C, at saturating light intensity (250–2,000

μmol photon m⁻² s⁻¹), atmospheric CO₂ concentrations (380–420 ppm), and nonstressed conditions of vapor deficit pressure (1.5–2.5). When mesophyll conductance was calculated, authors followed the methods of variable J and isotopic discrimination (Harley et al., 1992; Warren, 2011).

Metabolite Analysis

Generally, metabolite analysis was performed in the same leaf type used for gas-exchange measurements. The samples were collected avoiding the midrib, at midday (to prevent the diurnal variations of metabolites), and immediately frozen in liquid nitrogen for subsequent analysis. Extraction procedures were developed by quick grinding of the tissues under liquid nitrogen and subsequent addition of the appropriate extraction buffer. Polar phase was dried, and subsequently the samples were derivatized and the relative levels of metabolites were established by gas chromatography-mass spectrometry following previously described methodologies for data acquisition, annotation, and interpretation (Lise et al., 2006). More information about the methodologies followed for photosynthetic characterization and metabolite extraction and quantification can be obtained from each article described in Supplemental Data Set S1.

Modeling Approach

Relationships between A, gs, gm, and the metabolic profile were analyzed with two different statistical approaches. In the first approach, we modeled the relationship of metabolic readouts, as individual responses, to A, gs, and gm as independent variables with linear models of different complexity. Since the differences due to the genetic background are expected to propagate and affect the remaining cellular levels, we used linear mixed-effects modeling approach. To this end, we compared models in which A, gs, and gm are fixed effects with those in which the intercept is a random effect and each of A, gs, and gm is treated as a random effect. In other words, if A, gs, and gm are treated as fixed effects, we are in the classical regression setting. If we allow that any of these regressors have slopes (i.e. regression coefficients), which can vary across the species, then we are in the mixed-effects setting. This strategy allowed us to address the issue of how robust the identified relationships are across species. The mixed-effects linear model approach is robust to missing values and does not suffer from unbalanced experimental design (see below), both present in our setting.

In the second approach, we modeled the relationship of A, gs, and gm as responses, to the metabolic readouts as independent variables, to investigate any relationships between metabolism and the three physiological responses. Since metabolic profiles tend to be correlated due to the structure and regulation of the underlying metabolic networks, and as there are more variables than observations, we used PLS regression. Since the available metabolite levels are relative (i.e. are expressed as ratios between wild-type/transgenic or ambient condition/water stress), the absolute measurements for the responses were first transformed into relative levels. To provide more robust estimates, the missing values were imputed using a recent random forest imputation method that outperforms other existing alternatives (Stekhoven and Bühlmann, 2012; Walpole et al., 2013; Gromski et al., 2014). All together, this resulted in a data matrix with 48 observations, nonuniformly distributed among 14 species, for each data profile.

Linear Mixed-Effects Models

For each metabolite, we tested several models of the type:

\[ y = X\beta + Z\alpha + e \]

where denotes the given vector with the relative levels of the metabolite across the considered species and experimental set-ups, stands for the vector of fixed effect to be determined, for the vector of random effects to be determined, for the observation error vector, under the assumption of normality for and and the assumption that CoV(ε) = 0, and and are appropriate design matrices, as detailed in the following. We considered one linear model with A, gs, and gm as independent variables without grouping for the genetic background (i.e., species). This is equivalent to classical multivariate regression in which A, gs, and gm are fixed effects whose regression coefficients we determined. We also considered the six expanded models which, in addition to the fixed effects for the independent variables, included either only random intercepts or both random intercept and random slope for each of the independent variables (i.e., A, gs, and gm). The random effects were with respect to the grouping for the considered species. Since we end up with multiple models for each metabolite based on the choice of fixed and random effects, we need to select the model for final inspection. For each metabolite, comparison based on ANOVA was conducted for model selection. For each of the selected models we determined its (marginal) coefficient of determination and performed additional tests for the fixed effects to aid the interpretation. The models were implemented in R by using the lme4 package (Nakagawa and Schielzeth, 2013; Bates et al., 2015).

PLS Regression

For PLS, we used A, gs, and gm as multiple responses (dependent variables) and the metabolite profiles as predictors (i.e. independent variables). The model of (RMSEP and coefficient of multiple determination (R²)) bias-corrected via cross-validation scheme can be helpful in detecting which metabolites may affect the responses via various feedbacks. We used the R function pls to obtain and analyze these models.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Distribution of the number of linear mixed models of the three independent variables and the coefficient of determination for each metabolite.

Supplemental Figure S2. (RMSEP and R² bias-corrected via cross-validation scheme of the models with different numbers of components through PLS modeling of the physiological traits.

Supplemental Figure S3. Coefficients of the metabolites in the models with one and two components for the three dependent variables.

Supplemental Data Set S1. Literature survey of photosynthetic characteristics in 17 different species.

Supplemental Data Set S2. Results obtained using different modeling approaches.

ACKNOWLEDGMENTS

We thank to Dr. Eva Miedes for critical reading of the manuscript and CFE, IRFS, and AFM for keeping our inspiration alive.

Received October 26, 2015; accepted March 10, 2016; published March 14, 2016.

LITERATURE CITED


278


Copyright © 2016 American Society of Plant Biologists. All rights reserved.

