

**Running Title:**

Metabolism and Biomass

**Corresponding author:**

Mark Stitt

Max Planck Institute of Molecular Plant Physiology,

Am Muehlenberg 1,

14476 Potsdam-Golm, Germany

Tel: (0049) (0) 331 567 8100

Email: [mstitt@mpimp-golm.mpg.de](mailto:mstitt@mpimp-golm.mpg.de)

**Suggested section:**

Biochemical Processes and Macromolecular Structures: Associate Editor John Ohlrogge

# **Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions**

Ronan Sulpice<sup>&\*</sup>, Zoran Nikoloski,<sup>\*</sup> Hendrik Tschoep, Carla Antonio, Sabrina Kleessen, Abdelhalim Larhlimi<sup>1</sup>, Joachim Selbig<sup>1</sup>, Hirofumi Ishihara, Yves Gibon, Alisdair R Fernie, Mark Stitt

Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam-Golm, Germany and <sup>1</sup> Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany

<sup>&</sup> Present address: NUIG, Plant Systems Biology Lab, Plant and AgriBiosciences Research Centre, Botany and Plant Science, Galway, Ireland

\* Joint first authors

## **Word count (characters with spaces).**

Abstract: 250

Introduction: 1152

Results: 4993

Discussion: 2296

Material and Methods: 995

Figures: 4

Tables: 3

Supplemental information: 10 Supplemental Tables, 4 Supplemental Figures

**Key words:** Arabidopsis; growth; environment; starch; proteins

**Financial support:**

This research was supported by the Max Planck Society, the German Ministry for Education and Research (GABI-GNADE 0315060E; OPTIMAL, 031958G) and the European Commission FP7 collaborative project TiMet under Contract 245143.

## Abstract

Natural genetic diversity provides a powerful tool to study the complex interrelationship between metabolism and growth. Profiling of metabolic traits combined with network-based and statistical analyses allow the comparison of conditions and identification of sets of traits that predict biomass. However, it often remains unclear why a particular set of metabolites is linked with biomass, and to what extent the predictive model is applicable beyond a particular growth condition. A panel of 97 genetically-diverse *Arabidopsis* accessions was grown in near-optimal C and N supply, restricted C supply and restricted N supply and analyzed for biomass and 54 metabolic traits. Correlation-based metabolic networks were generated from the genotype-dependent variation in each condition to reveal sets of metabolites that show coordinated changes across accessions. The networks were largely specific for a single growth condition. PLS regression from metabolic traits allowed prediction of biomass within and, slightly more weakly, across conditions (cross-validated Pearson correlations in the range 0.27-0.58 and 0.21-0.51; p-values in the range  $<0.001$ - $<0.13$ , and  $<0.001$ - $<0.023$ , respectively). Metabolic traits that correlate with growth or have a high weighting in the PLS regression were mainly condition-specific, and often related to the resource that restricts growth under that condition. Linear mixed model analysis using the combined metabolic traits from all growth conditions as an input indicated that inclusion of random effects for the conditions improves predictions of biomass. Thus, robust prediction of biomass across a range of conditions requires condition-specific measurement of metabolic traits to take account of environment-dependent changes of the underlying networks.

## Introduction

Plant biomass is the ultimate output of the interplay between metabolism and the cellular and developmental programs that control allocation (Poorter and Nagel 2000; Hermans et al 2006; Poorter 2011) and cell and organ growth (Krizek et al. 2009; Gonzales et al., 2009). A predictive understanding of these complex relationships would open up new perspectives in crop improvement. Given that an increase in the rate of growth must be underpinned by changes in metabolism, it should be possible to identify metabolic states that are associated with higher growth rates. One way to characterize metabolic states would be to measure fluxes. However, most flux measurements are in fact estimates, based on fitting labeling patterns of metabolites to a selected metabolic model. This is technically challenging in multicellular life forms like higher plants (Zamboni et al., 2011). Further, such estimates would need to be very precise because small changes in flux can result in large changes in biomass; plant growth is exponential with a typical increase in biomass of 10-25% per day, so a relatively small difference in fluxes and the momentary rate of growth will lead within 1-2 weeks to a large difference in biomass (Poorter 1989; Stitt and Zeeman, 2012). A complementary approach is to identify metabolic traits, such as the levels of metabolites, which are associated with higher rates of growth and biomass formation. The attractiveness of this approach has been enhanced by the development of increasingly powerful platforms to measure metabolite levels and sophisticated tools to analyze the resulting data sets (Lisec et al., 2006; Fernie et al., 2011; Saito and Masuda 2010).

Metabolite profiling of large populations of *Arabidopsis* natural accessions or inbred lines and the application of multivariate analysis tools such as canonical correlation analysis (CCA) and partial least squares (PLS) regression has allowed the identification of descriptor sets of metabolites that are predictive of biomass (Meyer et al., 2007; Sulpice et al., 2009; Steinfath et al., 2010a; Cuadros-Inostroza et al., 2010; Carreno-Quintera et al., 2012) as well as physiological traits like freezing tolerance (Korn et al., 2010) and herbivore resistance (Kleibenstein 2012; Züst et al., 2012). The advantage of surveying a wide range of metabolites is underlined by the fact that multivariate analysis allows predictions to be made from data matrices in which no individual metabolite significantly correlates with biomass (Meyer et al., 2007). This approach was recently extended to hybrid vigor. The relative density of networks based on correlations

extracted from metabolite profiles in *Arabidopsis* is modified in plants that show a strong degree of heterosis (Meyer et al., 2012). Further, metabolite profiles measured in parents allow prediction of hybrid vigor in their progeny both in *Arabidopsis* (Steinfath et al., 2010b) and maize (Reidelsheimer et al., 2012). In an analogous approach, robotized platforms can be used to profile large numbers of enzyme and search for relations between their maximum activities and growth (Sulpice et al., 2010).

This top-down approach nevertheless suffers from two major weaknesses. First, a statistical relationship between a set of metabolites and growth does not provide functional insights into how metabolism determines the rate of growth. Functional interpretation is compromised because the complexity of metabolic networks in primary metabolism makes it difficult to draw inferences about fluxes from changes in metabolite levels (Stitt et al. 2010; Sulpice et al., 2010; Fernie and Stitt, 2012), by the fact that current metabolite profiles only cover a small fraction of the total metabolome (Saito and Masuda 2010) and by the likelihood that many connections between metabolism and growth may be mediated by signaling pathways that impinge on physiological or developmental processes (LeClere et al., 2010; Lilley et al., 2012). The occurrence of a correlation between biomass and individual metabolites or linear combinations of metabolites also does not imply causality. Such correlations might arise if a given combination of metabolic traits supports increased biomass formation, but also if increased biomass formation resulted in a corresponding change in metabolite levels. Secondly, levels of metabolites in primary metabolism are dramatically influenced by the environment (Hannah et al., 2010; Caldana et al., 2011; Obata and Fernie 2012) including the irradiance regime (Gibon et al., 2006; 2009; Brautigam et al., 2009; Jankanpaa et al., 2012) and the nitrogen regime (Tschoep et al., 2009; Kusano et al., 2011; Amiour et al., 2012). It is not yet clear if the same sets of metabolites are predictive of biomass across different growth conditions.

Under short-day conditions biomass is strongly and negatively correlated with starch content at dusk and with the total protein content per unit fresh weight (FW) in a panel of *Arabidopsis* accessions (Sulpice et al., 2009). Multivariate data analysis using PLS revealed that biomass, starch and protein are predicted by overlapping sets of metabolites indicating that starch and protein concentration are integrative metabolic traits that capture information about the levels of many low molecular weight metabolites and are closely linked to biomass formation. Starch is a

transient carbon (C) store, which accumulates in leaves during the day and is remobilized to support metabolism and growth at night (Smith and Stitt, 2007; Stitt and Zeeman, 2012). A negative correlation between biomass and starch levels at dusk implies that faster-growing accessions convert C more efficiently to biomass, at least during the night. A negative correlation between protein concentration and biomass will result in a larger leaf area per unit of invested protein and hence absorption of more light per plant. This finding is consistent with comparative studies of different species, where there is often a negative correlation between leaf area mass (dry weight per unit leaf area) and the growth rate, especially in limiting irradiance (Poorter and Nagel 2000; Poorter et al., 2009). Furthermore, as protein synthesis is an energetically costly process (Piques et al., 2009; Raven 2012), it is possible that the lower protein concentration might contribute to the observed increased efficiency of C use. Subsequently, Sulpice et al. (2010) showed that large accessions invested a large proportion of their protein in enzymes of photosynthesis. This will allow photosynthetic capacity per unit leaf area to be maintained irrespective of the fact that the total leaf protein concentration decreases.

The studies of Sulpice et al. (2009; 2010) were carried out in short day conditions where growth is limited by C (Gibon et al., 2009). Plant growth is also often restricted by the supply of nutrients, especially nitrogen (N) (Poorter and Nagel 2000; Krapp et al., 2005; Hirel et al. 2007; Xu et al., 2012). In the following experiments, we profiled metabolites and enzyme activities in the same set of accessions in conditions where N was limiting for growth (Tschoep et al., 2009) and in conditions where N was saturating and C was close to saturating for growth. These data were combined with our previously published data for short day conditions, and analyzed to identify which, if any, features of the relationships and connectivity between metabolic traits and growth are shared across different growth conditions.

## **Results**

### **Experimental design**

In earlier studies, we established growth protocols for the reference accession Col0 in which a decreased N or C supply lead to compensatory changes in metabolism and a mild and sustained decrease in growth rate: (i) Growth with a full nutrient supply in a 12 h photoperiod (12hHN) allows near to maximal growth rates of Col0, and increasing the photoperiod does not lead to a

major further stimulation of growth (Gibon et al., 2009: Supplemental Figure SI). Longer photoperiods were avoided because they lead to early induction of flowering. (ii) A low-N growth regime (12hLN) was established under which there was a 20-25% decrease in the relative growth rate and a ~50% decrease in biomass after 29-35 days, compared to 12hHN (Tschoep et al., 2009). Protein levels were hardly altered and some amino acids even increased, revealing that metabolism and growth have adjusted in a coordinated manner to the decreased N supply. (iii) Similarly, in the 8 h photoperiod used by Sulpice et al. (2009, 2010) there was a ~30% decrease in the relative growth rate compared to a 12h photoperiod (Supplemental Table SII). Starch turnover was adjusted such that starch was almost but not completely exhausted at the end of the night, C was available throughout the 24 h cycle, and C-starvation marker genes were not induced until after a short extension of the night (Gibon et al., 2009; Usadel et al., 2008, reviewed in Stitt and Zeeman 2012).

A panel of 97 *Arabidopsis* accessions selected to maximize genotypic and geographic variation and biomass variation (Sulpice et al., 2009; 2010) was grown with an optimal supply of N in a 12h/12h light/dark regime (12hHN) and a suboptimal supply of nitrogen in a 12h/12h light/dark regime (12hLN) (Tschoep et al., 2009). Sets of plants from both growth conditions were analyzed for rosette biomass and the levels of metabolites and enzyme activities at dusk. The resulting data was combined with published data for the same accessions grown in an 8h/16h light/dark regime to more strongly limit growth by the C supply (8hHN; Sulpice et al., 2009; Sulpice et al., 2010). The combined dataset included information about rosette biomass and 54 metabolic traits in three growth conditions. The metabolic traits included the three structural components (protein, chlorophyll a chlorophyll b), the major transitory C store starch, 43 low molecular weight metabolites, including a range of sugars, amino acids, organics acids and other metabolites, and maximum activities of eight enzymes from central C and N metabolism (for a list of the metabolic traits and abbreviations see Supplemental Table SI). As measurements of nitrate, ornithine and spermidine were not available for the published 8hHN dataset, random numbers were introduced for these traits in the calculations of condition-specific correlation matrices. However, these traits were not used in the PLS and mixed model analyses.

### **Biomass involves an interaction between accession and growth condition**



Biomass differed between the 97 accessions by 3.1-fold, 2.8-fold and 4.2-fold in 12hHN, 12hLN and 8hHN, respectively, relative to the accession with the lowest biomass in that growth regime (Figure 1A-B, Supplemental Table SI). The impact of low-N and low-C differed between accessions, with some accessions showing a >70% decrease in biomass and others showing no decrease (Figure 1A-B). Accessions that had a large biomass in 12hHN tended to show a marked decrease in biomass in C-limiting or N-limiting conditions, while many of the accessions that had a small biomass in 12hHN showed little or no further decrease in C-limiting or N-limiting conditions. When individual accessions are inspected, some show a marked decrease in biomass in low-C and low-N conditions, some maintain biomass in low-C and low-N conditions, and others are especially sensitive to low-C (Mh1, Nok2, Lov5) or low-N (Bur0, Dijon5, Old1) (see Supplemental Table SI). Small sets of accessions ranked high (Bsch2, Da112, Dra0, Mt-0, Wei1) or low (Ang0, Bla11, Je-54, Pyl-1, RRS-10, TAMM-2) for biomass in all three conditions (Figure 1B). Overall, pairwise scatter plots revealed significant positive correlation ( $R = 0.47$ ;  $p = 1.19e^{-06}$ ) between biomass in 8hHN and 12hHN, and weaker relationships between biomass at 8hHN and 12LN ( $R = 0.31$ ;  $p = 0.001$ ) and biomass in 12hHN and 12hLN ( $R = 0.28$ ;  $p = 0.0046$ ) (Figure 1C).

Thus, the main trends are three-fold: (i) subsets of accessions produce higher or lower biomass than others in all three conditions, (ii) many accessions that produce high biomass in high-N and high-C conditions tend to show a larger decrease in biomass when N or C is decreased and (iii) many individual accessions respond differently to low N and low C.

### **Metabolic traits are subject to environmental and genotypic variation**

The absolute levels of structural components, metabolites and enzymes are provided in Supplemental Table I. ANOVA showed that all structural and metabolic traits except succinate showed highly significant changes ( $p < 0.0001$ ) in the growth condition term (Supplemental Table SII). Significant traits in the accession term included biomass, starch, protein (all  $p < 0.0001$ ) and many metabolites (including: fructose, glucose, malate, *myo*-inositol, proline, threonine and nicotinic acid at  $p < 0.0001$  and sucrose, raffinose, total amino acids and many individual amino acids at  $p < 0.05$ ) and enzyme activities (including: NR, PEPCx, AGPase and NAD-GIDH at  $p < 0.0001$  and GS and NAD-MDH at  $p < 0.05$ ).

Principle components analysis (PCA) generated three distinct groups corresponding to the three growth conditions (Figure 2A, Supplemental Table SIII). The major principle component (PC1, 41.3% of total variance) separated C-limited (8hHN) and N-limited (12hLN) conditions, with near-optimal conditions (12hHN) in an intermediate position. Principle component 2 (PC2; 14.3%) separated near-optimal conditions from the two limiting conditions. When axes are chosen that reflect the variance captured by each PC, the 97 accessions formed a fairly compact group in near-optimal conditions and a slightly more spread-out group, especially in PC1, in low-C and low-N conditions. In PC1, positive weightings were found for *myo*-inositol and traits related to nitrate assimilation (nitrate reductase activity, nitrate), ammonium assimilation (GOGAT) and organic acid metabolism (PEP carboxylase, malate, fumarate) and negative weightings were found for protein, sucrose, total amino acids and several minor amino acids. In PC2, positive weightings were found for GIDH activity, several sugars (sucrose, glucose, fructose), spermidine and ornithine, and negative weightings for shikimic acid, chlorophyll, starch, glutamate and nicotinic acid.

Figure 2B (see Supplemental Table SIV for details) summarizes the changes of individual metabolic traits in 12hLN and 8hHN compared to 12hHN. Some metabolic traits showed consistent changes across all 97 accessions in low-N. This included an increase of sucrose, several amino acids (e.g., leucine, isoleucine, lysine), urea, 4-aminobutyrate and NAD-GIDH activity, and a decrease in raffinose, *myo*-inositol, glycine, proline, spermidine, shikimate, malate, fumarate, dehydroascorbate, nitrate and nitrate reductase activity. Nevertheless, the extent of the change varied. Some traits showed large variation between accessions, with an increase in some and a decrease in others (e.g., maltose, trehalose, alanine, glutamate, asparagine, threonate). As previously reported for the reference accession Col0 (Tschoep et al., 2009) there was, perhaps against expectations, a slight but consistent increase in the protein concentration in 12hLN compared to 12hHN. This, and the maintenance or increase in most amino acids, shows that all accessions adjust to compensate for the decrease in N supply.

A different set of metabolic traits showed consistent changes across all accessions in low-C. This included an increase in sucrose, glucose, fructose, alanine and dehydroascorbate levels and GOGAT, PEPCx and NAD-GIDH activities, and a decrease of chlorophyll, several amino acids (including aspartate, glutamate, phenylalanine, asparagine, glutamine, arginine), shikimate and

nicotinic acid. As previously reported for the reference accession Col0 (Gibon et al., 2009, Hannemann et al., 2009) there was a slight but consistent decrease in protein in all accessions in 8hHN compared to 12hHN. The decrease in the protein concentration and the levels of many amino acids reflects the strong dependence of N metabolism on the C supply (Nunes-Nesi et al., 2010). As in the previous comparison, many metabolites showed quite varied changes between 12hHN and 8hHN, again pointing to genotypic variation in the response to the growth condition.

The coefficient of variation (CV, the standard deviation divided by the mean) was estimated to provide insights which metabolic traits show the largest genetic variation in a given growth condition (Supplemental Figure 2A). The average CV of all metabolic traits in 8hHN, 12hHN and 12hLN was 34, 33 and 31%, respectively. CV was generally low for structural components and higher for low molecular weight metabolites. Protein, Chla, Chlb, starch (CV<10%), sucrose, total amino acids, shikimate and most enzymes (<20%) had a low CV in all three conditions, while maltose, trehalose, raffinose, glutamine, asparagine, arginine and proline had a high CV in all conditions. Some metabolic traits showed a high CV in one condition, for example nitrate had a high CV in 12hLN. This may be because nitrate accumulates to high levels in N-replete conditions, but is used for growth in low N (see below).

### **Correlations between individual metabolites and biomass in each growth condition**

We next investigated if the same or different individual metabolic traits correlate to biomass in the three growth conditions. Biomass-metabolite trait correlations (Spearman correlation coefficient) that were significant at a false discovery (FD) rate <5% test are listed in Table I (for a full list, see Supplemental Table SV).

There was a highly significant negative correlation of starch with biomass in 8hHN ( $R = -0.54$ ) (see also Sulpice et al., 2009) and 12hHN ( $R = -0.49$ ), and a weaker non-significant negative correlation in 12hLN ( $R = -0.33$ ). The weakening of the negative correlation between biomass and starch in 12hLN is consistent with the hypothesis that allocation of C to transitory starch plays an especially important role when C limits growth. Alanine, valine and succinate were also negatively and significantly correlated with growth in all three conditions.

Protein showed a highly significant correlation with biomass in 8hHN ( $R = -0.39$ ; see Sulpice et al., 2009) that became weaker in 12hHN ( $R = -0.30$ ) and was not significant in 12hLN ( $R = -$

0.014) (Table I, see Supplemental Figure S3 for scatter plots of biomass against starch and protein). This is in agreement earlier reports that the negative relation between biomass and protein observed in short photoperiod conditions is lost when the photoperiod is longer than 12h (Hannemann et al., 2008). This negative correlation may be related to a possible link in low-C conditions between efficient use of C and increased biomass formation, possibly because N assimilation and protein synthesis represents a major cost for growth (Piques et al., 2009; Anthor 2010).

Sucrose, isoleucine, shikimic acid, malate and 4-hydroxyproline were negatively and significantly correlated with growth in both 12hHN and 12hLN, while aspartate, glutamate and glycine were negatively correlated with growth in both 12hHN and 8hHN, and raffinose was negatively and significantly correlated with growth in both 12hLN and 8hHN.

Other correlations were restricted to one growth condition. In 12HN, biomass was negatively correlated with xylose, tryptophan, and PEPC activity, and positively with spermidine. In 12hLN, biomass was negatively correlated to maltose, trehalose, *myo*-inositol, nitrate, leucine, threonate and nitrate reductase activity, and positively correlated to glutamic acid and asparagine. Nitrate is the major source of inorganic nitrogen and is assimilated via nitrate reductase, while asparagine is a major store for N. In 8hHN, biomass was negatively correlated with total amino acids, several individual amino acids including asparagine, dehydroascorbate and putrescine and positively correlated with PEPC and NAD-GIDH activity. Asparagine accumulates and NAD-GIDH activity is induced in C starvation (Melo-Oliviera et al., 1996; Gibon et al., 2004; Mayashita and Good 2008; Gibon et al., 2009). Some metabolic traits were negatively correlated with growth in one condition, and positively in another (glutamate, asparagine, PEPCx activity).

### **Comparison of metabolic networks in the three growth conditions**

We next analyzed connectivity between metabolic traits. To this end, matrices were generated from the variation in metabolic traits across 97 accessions in each growth condition to reveal which traits are subject to coordinated changes between accessions in a given growth condition (Supplemental Table VI).

Of a total of 1683 trait pairs, significant correlations at 10%, 5% and 1% FDR were found for 493, 303 and 293 trait pairs in 12hHN, for 493, 261 and 129 trait pairs in 12hLN and for 347, 261 and 129 trait pairs in 8hHN, respectively. The vast majority of the correlations were positive. This resembles earlier reports of high connectivity between metabolic traits in panels of cultivars, accessions or inbred lines (Carrari et al., 2006; Meyer et al., 2007; Schauer et al., 2008; Keurentjes et al., 2008; Sulpice et al., 2009; 2010; Keurentjes et al., 2008; Meyer et al., 2012).

A total of 893, 737 and 434 trait pairs showed a significant correlation in at least one condition at 10%, 5% and 1% FDR, respectively. These numbers are much higher than those for any single condition, indicating that there is considerable non-overlap between the correlation matrices in the three growth conditions. The RV coefficient can be used to compare matrices in high-dimensional data analysis studies (Robert and Escoufier 1976; Abdi 2007). It is a measure of the similarity between two matrices and varies between +1 (if the two compared matrices are identical) and zero (if the two matrices are completely different). The RV coefficients (Figure 3A) were between 0.35 and 0.26, which are rather low values, confirming that the metabolic networks are condition-dependent. The p-values were nonetheless significant, indicating there are some robustly shared features.

Figure 3B provides a visual overview of the correlation matrix in each condition (for original data and the full matrices see Supplemental Table SI). Color coding is used to distinguish positive and negative correlations, and to denote significance at  $p < 0.01$ ,  $p < 0.001$  and  $p < 0.0001$ . Some general features were conserved across all growth conditions: firstly, there were many more positive correlations than negative correlations and secondly, whilst there were many correlations between metabolites and many correlations between enzymes, there were relatively few correlations between enzymes and metabolites (see also Sulpice et al., 2010). However, closer inspection reveals that many correlations were condition-dependent. Indeed, the 12hHN dataset showed a relatively low connectivity between metabolites while enzymes were strongly correlated. In 12hLN correlations between metabolites were stronger, especially between amino acids and between organic acids. Nitrate assimilation is closely linked with organic acid synthesis, because organic acids act as counter-anions for nitrate and provide C skeletons for the synthesis of amino acids (Nunes-Nesi et al., 2010, Xu et al., 2012). In 8hHN, the matrix is dominated by positive correlations between amino acids and positive correlations between

enzymes. The positive correlations between amino acids reveals that the decrease in the levels of different amino acid levels noted above (Figure 2B) occurs in a coordinated manner and is larger in some accession than in others.

The extent of overlap of individual links (correlations) in the three correlation matrices is further explored in Figure 3C-D. Of the links that are significant at  $FDR < 0.01$ , only 19 were shared across all three growth conditions. These were restricted to metabolites that are immediately adjacent to each other in metabolic pathways or have very similar functionalities (glucose and fructose; aspartate and glutamate; the three basic amino acids lysine, asparagine and arginine, the three aliphatic amino acids valine, leucine, isoleucine and aminobutyric acid; aspartate, arginine, proline and 4-hydroxyproline), two closely adjacent enzymes that are involved in malate formation (PEPC, NADH-MDH), and the three structural components (protein, Chla, Chlb) (Supplemental Table SVI).

Testing for shared links at  $FDR < 1\%$  may result in false negatives because traits pairs that are significant in one growth condition may lie slightly below this stringent threshold in another. We therefore investigated how many conserved links are found for trait pairs that show a correlation at  $FDR < 0.01$  in at least one growth condition and a more relaxed significance level of  $FDR < 10\%$  for the other two conditions (Figure 3D). This analysis revealed that up to 83 (4.9% of all possible) links are conserved in all three conditions. The additional shared links include sucrose with protein, amino acids with Chla and Chlb, further pairs of amino acids, and a set of enzymes involved in starch and nitrogen metabolism (AGPase, GS, PEPC, NAD-MDH).

We also tested for links that were conserved in two of the three growth conditions. At the  $FDR < 0.01$  level, another 39, 19 and 32 pairwise correlations were significant in the 12hHN vs. 12hLN, 12hHN vs. 8hHN and 12LN vs. 8hHN comparison, respectively, rising to 115, 74 and 72 when the criteria were relaxed, as discussed above. We also asked for selected metabolic traits if the variation between accessions was conserved across different growth conditions. Pairwise plots revealed a weak but significant agreement for starch ( $R = 0.30$ ,  $p = 0.05$ ) and protein ( $R = 0.35$ ,  $p = 0.001$ ) when 8hHN was compared to 12hHN, and a non-significant correlation when 12hHN was compared to 12hLN (starch:  $R = 0.16$ ,  $p = 0.12$ ; protein:  $R = 0.15$ ,  $p = 0.15$ ). (Supplemental Figure S4).

Altogether these results point to a strong impact of the growth condition on the links in networks extracted from metabolic profiles. While there are a small proportion of conserved links, these are mainly for metabolites or enzymes that are closely related with respect to pathway topology or trait function.

### **Partial least squares regression of biomass, starch and protein on other metabolic traits**

As already noted, some individual metabolic traits correlate with biomass (Table 1). Predictive power can be increased by using multivariate analysis to predict biomass from a linear combination of a set of low-molecular-weight metabolites (Meyer et al., 2007; Sulpice et al., 2009). Therefore, we investigated whether multivariate analysis reveals shared features in the network linking metabolic traits and biomass formation that are not apparent at the level of pairwise comparisons.

In data sets like ours, where the number of predictors (52) is lower than the number of accessions (97), the predictive power of linear models can be improved by dimensionality-reduction methods like PLS regression. PLS identifies combinations of the original input traits (also termed variables or predictors) that have the maximum covariance with the output trait of interest. These orthogonal combinations of input traits, referred to as latent variables, are then used to predict the output trait. Sulpice et al. (2009) previously found for *Arabidopsis* accession growing in short day conditions that biomass, starch levels and protein concentration are correlated, and that each is predicted by a similar combination of low molecular weight metabolites. We repeated this analysis for all three growth conditions. The idea of PLS is to provide models of high predictive power while selecting a small number of latent variables, which might correspond to a reduced number of metabolic traits in the model. Selection of the number of latent variables was performed based on minimization of the residual mean squared prediction error after Leave-One-Out (LOO) cross validation. The predictive power of the model was evaluated in a second round of cross validation by determining the correlation between the predicted values and the respective response variable with the previously determined fixed number of latent variables; significance was evaluated by permutation of the data 5000 times (see Materials and Methods for details). Shuffling was conducted to allow consideration of permutations specific for a metabolic trait. The use of two separate rounds of cross-validation for robust estimation of the number of

latent variables and for evaluation of the predicting power of the model reduces the possibility of over-fitting.

In each growth condition, PLS regression using metabolite levels as an input allowed a significant prediction of biomass (Pearson correlation 0.36, 0.58, and 0.27, p-values < 0.05 for 12hHN, 12hLN, and 8hHN, respectively) and starch (Pearson correlation 0.67, 0.39, and 0.23, p-values < 0.05 for 12hHN, 12hLN, and 8hHN, respectively). It also allowed a significant prediction of protein concentration in 12hHN and 12hLN (Pearson correlation of 0.46 and 0.57, respectively, p-value > 0.05) but not in 8hHN (Table II, bold area, Supplemental Table SVII). The predictive power was improved compared to individual metabolites (Table I).

We also asked whether metabolite profiles measured in one growth condition allow prediction of biomass, starch or protein in a different growth condition. Whilst almost all PLS regressions were significant (except for the prediction of protein in 8hHN by metabolic traits from any condition), the cross-validated correlations were generally smaller (Table II). For significant regressions, the range and average of p-values was 0.001-0.003 and 0.013 for within-growth condition comparisons, and 0.001–0.030 and 0.023 for cross-growth condition comparisons, respectively.

We also analyzed two additional scenarios—in the first, we conducted PLS on the means of the traits across all three conditions, while in the second, we employed PLS on a combination of the three data sets ('Mean' and 'All', respectively, in Table II). More specifically, in the 'Mean' scenario, we used the mean input traits across the three conditions to build a PLS regression on the mean output traits across the three conditions; in the 'All' scenario, we conducted PLS analysis on the concatenated data matrices from the three conditions. Using the mean of the traits over the three conditions in the PLS analysis, we found that all regressions are significant and of greater predictive power than the average power of condition-specific PLS. This gain was even greater when we used all three data sets ('All'). However, although similar observations with much greater correlation values were obtained using the combined data sets, these regressions were based on a much larger number of latent variables, especially when we combined the data sets (as many as 32 from 52 available). Altogether, these findings suggest that PLS regression on metabolites to predict output traits like biomass and starch may be specific and more robust for comparisons within a given growth condition than for comparison across conditions.



The importance of individual input traits (variables) in the linear combination is given by the variable importance in the projection (VIP; Chong and Jung, 2005). The VIP for each individual metabolic trait as input in each growth condition is provided in Supplemental Table SVII. We next investigated the correlation between the VIP of the input traits from the PLS regressions on the pairs of output traits. This analysis led to two main conclusions. First, there was close agreement between the VIP of metabolic input traits in the prediction of the three output traits in 12hHN ( $p < 0.001$  in all pairwise comparisons of biomass, starch and protein) (Table III). The agreement was lower under 12hLN (especially for the comparison protein vs. biomass,  $p = 0.017$ ) and still weaker in 8hHN, when there was good agreement between biomass and protein ( $p < 0.0001$ ) but not between biomass and starch ( $p = 0.059$ ) or starch and protein ( $p = 0.28$ ). The latter differs from a previous report (Sulpice et al., 2009; see Discussion). Second, different metabolic traits were important in the PLS regression in different conditions. (assessments are based on  $VIP > 1$ ). Taking the PLS regressions on biomass as an example, in 12hHN xylose, several central amino acids (glutamine, alanine, aspartate, glycine), all three aromatic amino acids (phenylalanine, tryptophan, tyrosine), proline, hydroxyproline, malate and succinate had a high VIP, whilst in 8hHN raffinose, a similar set of central amino acids (glutamine, glutamate, alanine, aspartate, glycine), all N-rich amino acids (lysine, asparagine, arginine), fumarate, threonate and putrescine had a high VIP, and in 12hLN raffinose, maltose, trehalose, erythritol, a different set of amino acids (alanine, asparagine, arginine), succinate, glycerate, hydroxyproline, dehydroascorbate, threonate, putrescine, and nitrate reductase activity had a high VIP for biomass. Some metabolic traits (e.g., alanine, hydroxyproline) were represented in all three conditions, others (e.g., proline, asparagine, arginine) in two conditions and many in only one condition. With the exception of nitrate reductase, enzyme activities did not show high VIP. By comparing Table I and Supplemental Table SV, we observed that many of the metabolic traits with a high VIP show significant correlations with biomass in that condition. Indeed, the correlation between the VIP in the regression for biomass and the Spearman correlation coefficients of individual metabolic traits and biomass, with values of 0.65, 0.72 and 0.42 in 12hHN, 12hLN and 8hHN, respectively, provide statistical support for this observation.

### **Relation between biomass and nitrate and total N content in plants growing with a restricted N supply**

We next asked whether nitrate or other metabolic traits related to N adopt a more important role as a predictor for biomass in low-N conditions. These analyses were limited to 12hHN and 12hLN because values for nitrate were not available for the published 8hHN study. Nitrate levels were higher in 12hHN, where they typically accounted for about 20% of the N in the rosette, than in 12hLN (Figure 4A). Nitrate levels were negatively correlated to biomass in 12hLN but unrelated to biomass in 12hHN (Spearman's rank correlation coefficient,  $R = -0.5$ ,  $p = 8e^{-06}$  and  $-0.18$ ,  $p = 0.18$ , respectively Table I, Figure 4A-B). Total N content was estimated by summing N in nitrate, protein, amino acids and chlorophyll. Total N content was similar in both 12hLN and 12hHN, and was unrelated to biomass in both conditions (Figure 4A).

As already noted, accessions that maintained a relatively high biomass in low-N tended to show only a small increase in biomass in high-N, whereas accessions that showed a relatively small biomass in low-N showed a large (>3-fold) increase in biomass in high-N (Figure 4C). The ability of plants to grow with a low N supply, sometimes termed Nitrogen Use Efficiency (NUE) can be divided into two components; the ability to produce more biomass per unit N in the plant, and the ability to obtain N from the soil (Moll, 1982). The total N concentration in the rosette was unrelated to the biomass difference between low-N and high-N (Figure 4D). The N content (mg N per rosette) was strongly related to the response of accession to N; accessions that maintained biomass in low N contained more N in the rosette than accessions that showed a large gain in biomass in high N (Figure 4E). These results imply that accessions differ in the extent to which they can acquire N from low-N soil, and that this is far more important for the response of biomass to N supply than changes in the N content of the rosette.

### **Mixed model analysis**

The results presented in the previous sections were mainly based on per-condition PLS regressions without controlling for the effects of environmental conditions. This is usually referred to as by-group approach, where each group corresponds to a condition. A by-group approach does not detect relationships that are conserved across conditions, and may highlight very specific effects for the individual conditions. We therefore asked if a more generalizable model for each of the three output traits (i.e., biomass, starch, and protein) can be obtained by combining the data sets from the three conditions, using an approach based on linear mixed models.

Linear mixed models are a type of generalized linear mixed models (Breslow & Clayton 1993), which offer parsimonious ways to account for group level structure in data while simultaneously assessing effects within and across groups (i.e., conditions). In addition to individual level noise  $\varepsilon$ , linear mixed models allow for normally distributed group-level differences centered around the individual level parameters. Our analysis is based on a linear mixed model with random intercepts by condition (defined as a grouping factor), formulated as:

$$y_{i,cond} = \beta' + \sum_{k=1}^N \beta_k \log x_{i,k} + \varepsilon_i$$

$$\beta' = \beta_0 + \beta_{cond}$$

$$\varepsilon \sim N(0, \sigma\varepsilon), \beta_{cond} \sim N(0, \sigma_{cond}), \beta_{cond} \perp \varepsilon$$

In this model, the intercept ( $\beta'$ ) is the sum of the ordinary intercept (i.e., the global mean,  $\beta_0$ ) and the adjustment based on the group (condition,  $\beta_{cond}$ ) for each of the three output traits. The adjustment is assumed to be normally distributed and centered around zero as well as orthogonal to the individual level noise  $\varepsilon$ . This adjustment is termed the random intercept because it adjusts the overall intercept to reflect a randomly distributed condition-specific intercept.

Here we first ask whether we can remove the random intercept without sacrificing the power of the model. This is achieved by  $\chi^2$ -test (with degree of freedom (df) = 1) over the difference in deviance (defined as twice the log likelihood) between the model with a random intercept against the same model without a random per-condition intercept. This test aims at determining if the added number of parameters (due to the random intercepts) significantly improves the model quality. While inclusion of random intercepts increases the quality for the model of biomass ( $p$ -value < 0.05), this is not the case for starch and protein concentration (Supplemental Table SIX). The random intercepts in case of a linear mixed model for biomass were 35.75, 2.21, and -37.96 for the 12hHN, 12hLN, and 8hHN, respectively, indicating the condition-specificity. Moreover, analysis of deviance table reported that chlorophyll a, sucrose, myo-inositol, aspartate, glycine, serine, and nicotinic acid had significant coefficients in the regression for biomass at a significance level of 0.05. Whilst some of these metabolites (sucrose, glycine) had a significant correlation (Table 1) or a high VIP in the PLS regression on biomass (Supplemental Table SV)

in two of the growth conditions, others (e.g., serine, nicotinic acid) had not been uncovered in the previous analyses. This is due to the more important role of these metabolic traits in the generalized (cross-condition) model. Nevertheless, as for the PLS regressions, enzymes make only a weak contribution (none significant at  $p < 0.05$  and only one, GS, at  $p < 0.1$ ).

We next tested if random effects for the slope of these seven significant metabolic traits improve the quality of the model for biomass. Indeed,  $\chi^2$ -test ( $df = 7$ ) indicated that adding random slopes, presented in Supplementary Table IX, improves the predictive power of the model. Subsequent analysis of deviance table indicated that the combined effects (i.e., fixed and random) for sucrose and glycine are significant whilst alanine has a significant fixed effect ( $p < 0.05$ , Supplementary Table SIX). Alanine was one of the very few metabolites that in all three growth conditions correlated significantly ( $p < 0.05$ ) with biomass (Table I) and had a high VIP in the PLS regression on biomass (Supplemental Table SV).

## Discussion

Whilst it can be anticipated that metabolism will affect growth and that this dependence should be reflected in the values of metabolic traits, this connection is often masked due to the complexity of the network that links metabolism with growth (Forné and Stitt, 2012). Natural genetic diversity provides a powerful tool to analyze complex networks, because it allows the study of thousands of genetic perturbations that vary independently between different genotypes. Profiling of populations of *Arabidopsis* natural accessions or inbred lines and application of multivariate analysis tools has allowed sets of metabolites to be identified that are predictive of biomass (Meyer et al., 2007; Sulpice et al., 2009; Steinfath et al., 2010; Cuadros-Inostroza et al., 2010; Carreno-Quintera et al., 2012) and in some cases has allowed hypotheses to be formulated with respect to which aspects of metabolism play a key role in the determination of growth (Sulpice et al., 2009; 2010). However, metabolite levels depend on the growth condition (Caldana et al., 2011; Obata and Forné 2012, see Introduction for further references). We have investigated (i) whether metabolite profiles provide information that is predictive for biomass in three different growth conditions and (ii) whether the network connectivity is conserved or changes between growth conditions. To do this, a panel of 97 genetically-diverse *Arabidopsis* accessions was grown in three growth conditions; near-optimal C and N supply, restricted C

supply and restricted N supply. The growth protocols used to restrict C and N decreased biomass by, on average, about 2-fold compared to near-optimal C and N. This represents a small decrease in the rate of growth. Previous work in the reference accession Col0 has shown that *Arabidopsis* adjusts to these regimes to avoid an acute C-limitation (Gibon et al., 2009; Stitt and Zeeman 2012) or N-limitation of metabolism and growth (Tschoep et al., 2009).

The large genetic diversity in *Arabidopsis* for biomass is apparent with approximately 3-fold differences in biomass between the smallest and largest accessions in a given growth condition. Accessions vary in their response to the growth condition (Figure 1). Whilst there is a trend for accessions that are large in one condition to also be large in other conditions, this is modified by two further trends; firstly, accessions that develop a high biomass in near-optimal conditions show a larger decrease of biomass in limiting conditions and, secondly, many individual accessions show differing responses to low-C and low-N.

Whilst there are no other published studies of the response of biomass to low C in *Arabidopsis*, three earlier studies used a much smaller but partly overlapping panel of accessions to study the response to low N (<http://dbsgap.versailles.inra.fr/vnat/>; North et al., 2009; Chardon et al. 2010; see Supplemental Table SX). The trend for accessions that produce a high biomass in high-N to show a larger decrease in biomass in low-N is visible in these earlier studies. However, detailed comparison is difficult because of differences in accession ranking for biomass. In high N conditions there is very good agreement between biomass in our study and the VNAT database (20 shared accessions,  $R = 0.69$ ) and weak agreement with the study of Chardon et al. (2010) (18 shared accessions,  $R = 0.23$  due to two accession that show low biomass in our study and the VNAT database but have a high biomass in the study of Chardon et al., 2010). In low-N treatments the agreement between our data and that of the VNAT database and Chardon et al. (2010) breaks down ( $R = -0.03$  and  $0.07$ , respectively). This poor agreement may be due to different protocols in the low-N treatments. Whereas we used large pots containing soil with a high or a very low N content from the beginning of the experiment, earlier studies grew plants in small pots with sand and watered regularly with nutrient solution containing different amounts of N. In our growth protocol, N-restricted plants contain less nitrate and show a slower rate of growth but maintain rosette levels of amino acids and protein (Tschoep et al., 2009; see also Figures 2, 4 and Supplemental Table SI), whereas rosette N concentration decreased by 20-30%

in North et al. (2009) and Chardon et al. (2010). Further, while the variation in biomass is similar in high-N and low-N treatments in our growth protocol, in the other growth protocols there is less variation in biomass formation in low-N than in high-N treatments. Despite this variation between studies, our analysis confirms previous reports (Chardon et al., 2010, 2012) that Bur0 shows a large response to nitrogen, reveals that this accession is relatively insensitive to low-C, and identifies further accessions that show a similar response (Dijon5, Old1). Our study also identifies accessions that show a reverse response, with a large decrease in biomass in low-C and maintain biomass in low-N conditions (Mh1, Nok2, Lov5).

The response of metabolic traits is dominated by the growth condition (Figure 2) with low-C or low-N leading to marked and differing changes in many metabolic traits across all the accessions. There is nevertheless genetic variation for metabolic traits. This can be captured in each growth condition as a correlation matrix (Figure 3). These networks identify metabolic traits that are subject to coordinated changes between accessions in a given growth condition. The correlation matrices show some shared general features, in particular a predominance of positive correlations, and the presence of many correlations between metabolite levels, many correlations between enzyme activities and few correlations between metabolite levels and enzyme activities. As previously discussed (Sulpice et al., 2010), this may reflect the complexity of the network that links enzyme activities with metabolite levels. A small number of links are found in all three growth conditions, mainly between topologically adjacent or functionally similar metabolic traits. Nonetheless, the main feature emerging from our large study is that both the metabolic traits and the correlation network depend strongly on growth conditions. First, low-N and low-C lead to characteristic changes in metabolite levels that affect all accessions (Figure 2). In low-N this includes an increase in many amino acid levels, a decrease in organic acids and a decrease in nitrate reductase activity. In low-C, this includes an increase in sucrose and reducing sugars, a decrease in many amino acids with the exception of alanine which increases, and an increase in NAD-GIDH activity. Second, most of the individual links in the metabolic network are condition-specific (Figure 3). In low-N the correlation network is dominated by strong connectivity between amino acids and between organic acids, in low-C by strong connectivity between amino acids, and in near-optimal conditions by a less topologically-defined response.

The growth condition modifies the relation between metabolic traits and biomass. A different set of individual metabolic traits correlate to biomass in each growth condition (Table I). While PLS regression allows a highly significant prediction of biomass in each growth condition and often between growth conditions, when these analyses are made within a given growth condition the statistical significance for the predictive power tends to be stronger and a smaller number of latent variables is required (Table II). Application of linear mixed models highlighted that the inclusion of random (condition-dependent) effects for the intercept in the regression increases the quality of the model for biomass, but not for starch and protein concentrations. This further supports the condition-specificity of biomass prediction that is suggested by the PLS regressions. Additional analysis suggested that the inclusion of random slopes for metabolic traits that have significant coefficients in the linear mixed models could further improve the quality of these models.

A small number of individual metabolic traits are linked to biomass in all three growth conditions. For example, alanine correlated with biomass in all three growth conditions (Table I), had high VIP in PLS regressions on biomass in all conditions (Supplemental Table SVII) and, together with sucrose and glycine, was highlighted as important in the mixed linear model (Supplemental Table SIX). We previously reported that biomass is negatively correlated with starch and protein (Sulpice et al., 2009). This finding is confirmed for starch in all conditions used in the current study, and for protein in near-optimal and low-C conditions, but not in low-N (Table 1). We also reported that a similar set of metabolites have a high VIP in a PLS regression on all three traits and proposed that starch and protein concentration are integrative metabolic traits that capture information about the levels of many low molecular weight metabolites and are closely linked to biomass formation (Sulpice et al., 2009). This relationship with biomass is confirmed in near-optimal (12hHN) conditions for starch and for protein ( $p < 0.001$ ), in low-N for starch ( $p < 0.05$ ) but not for protein ( $p = 0.17$ ) and in low-C condition for protein ( $p < 0.001$ ) but not, or only very weakly, for starch ( $p = 0.059$ ) (Table III). The analysis in Sulpice et al. (2009) was carried out in short day (low-C) conditions; hence, there is a discrepancy in this particular condition. This may be due to use of a more stringent procedure for selection of the number of latent variables and validation of prediction in the current study, and because the published 8hHN dataset was obtained using a weaker experimental design than that used to obtain the 12hHN data set in the current study (see Methods).

The metabolic traits that adopt a major role in the network linking metabolism and growth in a given growth condition are often closely related to the metabolic resource that limits growth in that condition. In short day (low-C) conditions, low starch is the most powerful single predictor of biomass (Table 1). Protein is also negatively correlated to biomass, as are many amino acids (Table I). Further, protein and many amino acids decrease in short day conditions (Table II). As outlined in the Introduction, in low-C conditions, low protein concentration may increase the efficiency with which resources are used to generate biomass, which in turn may explain why starch reserves can be decreased (Sulpice et al., 2009; 2010). The second most strongly correlating individual metabolic trait is a positive correlation with NADH-GIDH activity. NAD-GIDH activity is induced by C starvation (Melo-Oliviera et al., 1996; Gibon et al., 2004; Mayashita and Good 2008). This prompts the hypothesis that large accessions, which contain less starch, operate with a lower margin of C than small accessions. When more C is available for growth in a 12h photoperiod (Gibon et al., 2009) the negative correlation between biomass and starch is retained, but the links to protein concentration, amino acid metabolism and NAD-GIDH activity are weakened or abolished. This is consistent with the idea that this link is driven by metabolic adjustment to low C, and that there is variation between accessions for way this interaction is regulated (Table I, Figure 3)

In contrast, in low-N conditions, the metabolic traits that correlate strongly with biomass include nitrate reductase activity and nitrate, with the latter being the most strongly correlating individual metabolic trait (Table I). A trivial explanation for the negative relation between biomass and nitrate would be that accessions with a larger biomass in low-N conditions exhaust nitrate; this, however, can be excluded because total nitrogen concentration was independent of accession biomass (Figure 4). A similar observation has been made in earlier studies with a small panel of *Arabidopsis* accessions (Chardon et al., 2010; 2012). Nitrate is typically taken up in the day and the night, but is mainly assimilated during the day when nitrate reductase is post-translationally activated and photosynthetic electron transport provides reducing equivalents of the reduction of nitrate and the subsequent reduction of nitrite (Lea et al., 2006; Lillo 2008). This results in a diurnal rhythm in which nitrate levels decrease in the light and recover during the night (Stitt and Krapp, 1999; Matt et al., 2001). The lower level of nitrate and higher activity of nitrate reductase at dusk in accessions that maintain a large biomass in low-N conditions is consistent with them assimilating more of the incoming nitrate during the day. Further, accessions that maintain a



larger biomass in low-N conditions absorb far more N from the soil (Figure 4). Earlier studies of small panels of *Arabidopsis* accessions indicate that differences in the root system may partly explain differences in N uptake (Loudet et al., 2005). It is possible that the lower rosette nitrate levels may promote root growth and N uptake, although more studies of root growth and transport activity will be needed to test this hypothesis.

In conclusion, while metabolic traits can be used to predict biomass in different growth conditions, this will require collection of data for the metabolic input traits in each growth condition. The growth condition has a large impact on the values of metabolic traits, on connectivity between metabolic traits, and influences the connectivity between metabolism and growth. While metabolic traits determined in one growth condition allow prediction of biomass in other conditions, the analysis is more robust when full input and output trait data is available for all conditions under study. Application of linear mixed models also reveals a marked condition-effect on the biomass prediction, and reveals that prediction can be improved when metabolic input data in all conditions is used as part of the model. Based on the growth conditions related to C and N availability analyzed in our study, in a given condition metabolic traits related to the limiting resource can adopt a more central role in the network that connects metabolism and growth. This implies that there is substantial natural variation in *Arabidopsis* for adjustment of metabolism to improve growth in low C and low N conditions. This variation, however, means that environmental conditions must be taken into account when searching for individual metabolites or sets of metabolites that act as biomarkers, and may compromise attempts to make predictions about genotype performance between different growth conditions.

## **Material and Methods**

### **Selection of the accessions and growth conditions**

*Arabidopsis thaliana* accessions used in this study were obtained from various sources as previously described (Sulpice et al., 2009; Sulpice et al., 2010). Geographical origin of the accessions is available at Vnat (<http://dbsgap.versailles.inra.fr/vnat/>). For the 8hHN plants were grown in multiple overlapping experiments as previously described (Sulpice et al., 2009; Sulpice et al., 2010). For the 12hHN and 12hLN treatments, plants were grown in large replicated experiments with all accessions. To eliminate effects due to seedling germination and

establishment, in all growth regimes seeds were germinated and grown for 7d with a 16h day length (irradiance  $145 \mu\text{mol m}^{-2} \text{s}^{-1}$ , temperature  $20^{\circ}\text{C}$  in the light and  $6^{\circ}\text{C}$  at night, humidity 75%) then in an 8-h-light/16-h dark regime for 7 days, ( $145 \mu\text{mol m}^{-2} \text{s}^{-1}$ , temperatures and humidities of  $20^{\circ}\text{C}$  and 60% during the day and  $16^{\circ}\text{C}$  and 75% at night). At 14 d, plants of average sizes were transferred to 6 cm diameter pots (five plants per pot). In all experiments, the position of the pots containing individual accessions was randomized.

For the 8hHN treatment the soil substrate was GS90 (composition: peat, clay, coconut fiber, 2 g/L salt, 160 mg/L N, 190 mg/L  $\text{P}_2\text{O}_5$ , 230 mg/L  $\text{K}_2\text{O}$ , pH 6, supplied by Werner Tantau Gmb & Co. KG) and vermiculite (Gebrueder Patzer) (Cross et al., 2006). For the 12hLN treatment the soil substrate was 50% (v/v) white peat (Gramoflor GmbH, Vechta, Germany) and 30% (v/v) fine and 20% (v/v) coarse grained vermiculite (AGRA – RHP, Kausek GmbH, Mittenwalde, Germany), fertilised with 260 mg  $\text{K}_2\text{HPO}_4$ , 396 mg GRANUKAL 85 (80%  $\text{CaCO}_3$  & 5%  $\text{MgCO}_3$  – Kreidewerke Dammann KG, Soehle, Germany), 1.6 mg Fetrilon-Combi micronutrient fertilizer (BASFAG, Ludwigshafen, Germany) and 30 mL tap water per 100 mL pot (Tschoep et al., 2009), For the 12hHN treatment the soil substrate was identical to 12hLN, except that it was additionally supplemented with 90 mg solid  $\text{NH}_4\text{NO}_3$  per 100ml pot. Prior to use, soils were stored for 2 weeks at  $10^{\circ}\text{C}$  to allow homogenization of nutrients. At 21 days, plants were transferred to a controlled small growth chamber ( $145 \mu\text{mol m}^{-2} \text{s}^{-1}$ ,  $20^{\circ}\text{C}$  day and night) for 2 more weeks in either an 8 h photoperiod (8hHN) or a 12 h photoperiod (12hHN, 12hLN). Plants were watered daily. Within each experiment, the position of the pots containing individual accessions was randomized. Harvests (five samples per accession, each consisting of three rosettes) were performed at the end of the light period. The entire sample was powdered under liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until its use.

### **Enzyme and Metabolite Assays**

Chemicals were purchased as described (Gibon et al., 2004). Total protein, starch, glucose, fructose, sucrose, and total amino acids were assayed as described (Cross et al., 2006). Malate and fumarate were assayed as described (Nunes-Nesi et al., 2007). For enzyme measurements, aliquots of 20 mg frozen FW were extracted by vigorous mixing with extraction buffer (Nunes-Nesi et al., 2007). AGPase, fumarase, GLDH, PEPC, INV, GOGAT, NR, and GS were assayed as described (Gibon et al., 2004b). NAD-MDH was assayed as described by Nunes-Nesi et al.

(2007). Metabolite extraction for GC-MS was performed as described previously (Schauer et al., 2006).

Derivatization and GC-MS analysis were performed as described previously (Lisec et al., 2006) starting from aliquots of 20 mg frozen FW. As measurements of nitrate, ornithine and spermidine were not available for the published 8hHN dataset, random numbers were introduced for these traits in the calculations of condition-specific correlation matrices. However, these were not used in the PLS and mixed model analyses.

### **Statistical analysis**

PLS regression is a dimensionality-reduction method which aims at determining predictor combinations with maximum covariance with the response variable (Eriksson et al., 2001; Wold et al., 1966). The identified combinations, called latent variables, are used to predict the response variable. Selection of the number of latent variables was performed based on minimization of the residual mean squared prediction error after Leave-One-Out (LOO) cross-validation. The predicted vector was correlated with the measured values to assess the predictive power of the predictor variables with the fixed number of latent variables. The significance of the prediction power was evaluated by permutation test with 5000 permutations of the data. We note that in every permutation, each row of the data matrix, corresponding to the profile of a metabolic trait, was shuffled independently of the others. Such permutation strategy is intended to break correlations in pairs of metabolic traits while maintaining the range that is specific for each metabolic trait. Then, for each permutation, a PLS model with the pre-determined number of latent variables was built to predict the randomized response variable and a Pearson correlation between the permuted response variable and in LOO cross-validation. The 5000 random correlations are compared to the performance of the PLS model which was used to predict the true response variable. The predictors were ranked according to their importance in projection (VIP) (Chong and Jung, 2005). The VIP measure of a predictor estimates its contribution in the PLS regression. The predictors having VIP values greater than one are considered important for the PLS prediction of the response variable. All procedures were applied after log-scaling the metabolic profiles. Our computations were carried out using the R package pls (Mevik and Wehrens, 2007). For the analysis based on linear mixed models, we used the lmer function from

the R package lme4 (Bates and Maechler 2010), while analysis of deviance table was carried out with the Anova function from the R package car (Fox and Weisberg, 2011).

The RV coefficient is a multivariate generalization of the *squared* Pearson correlation coefficient and provides a measure of the similarity between two squared symmetric positive semi-definite matrices and varies between +1 (two identical matrices) and zero (no similarity) (Robert and Escoufier 1976; Abdi 2007). It was calculated using the package <http://CRAN.R-project.org/package=FactoMineR>. The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean of all mean values obtained for a trait/accession pair.

**Acknowledgements.** This research was supported by the Max Planck Society, the German Ministry for Education and Research (GABI-GNADE 0315060E; OPTIMAL, 031958G) and the European Commission FP7 collaborative project TiMet under Contract 245143.

## Literature

**Abdi H** (2007) RV Coefficient and Congruence Coefficient. In: Neil Salkind (Ed.) (2007). Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage.

**Amiour N, Imbaud S, Clément G, Agier N, Zivy M, Valot B, Balliau T, Armengaud P, Quilleré I, Cañas R, Tercet-Laforgue T, Hirel B** (2010) The use of metabolomics integrated with transcriptomic and proteomic studies for identifying key steps involved in the control of nitrogen metabolism in crops such as maize. *J Exp Bot* **14**: 5017–5033

**Amthor JS** (2010) From sunlight to phytomass: on the potential efficiency of converting solar radiation to phyto-energy. *New Phytologist* **188**: 939-959

**Bates D, Maechler M** (2010.)Lme4: Linear mixed effects models using S4 classes. Lme4: Linear mixed effects models using S4 classes. R package version 0.999375-42. URL <http://CRAN.R-project.org/package=lme4>

**Breslow E, Clayton DG** (1993) Approximate inference in generalized linear mixed models. *Journal of American Statistical Association* **88**, pp. 9 -25.

- Brautigam K., Dietzel L., Kleine T., et al. (2009)** Dynamic plastid redox signals integrate gene expression and metabolism to induce distinct metabolic states in photosynthetic acclimation in Arabidopsis. *Plant Cell* **21**: 2715–2732
- Caldana C, Degenkolbe T, Cuadros-Inostroza A, Klie S, Sulpice R, Leisse A, Steinhauser D, Fernie AR, Willmitzer L, Hannah MA (2011)** High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions. *Plant Journal* **67**: 869-884
- Carrari F, Baxter C, Usadel B, Urbanczyk-Wochniak E, Zanon MI, Nunes-Nesi A, Nikiforova V, Centro D, Ratzka A, Pauly M, Sweetlove LJ, Fernie AR (2006)** Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol* **142**: 1380-1396
- Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CWB, Mumm R, Bouwmeester H, Visser RGF, Keurentjes JJB (2012)** Untargeted Metabolic Quantitative Trait Loci Analyses Reveal a Relationship between Primary Metabolism and Potato Tuber Quality. *Plant Physiology* **158**: 1306-1318
- Chardon F, Barthelemy J, Daniel-Vedele F, Masclaux-Daubresse C (2010)** Natural variation of nitrate uptake and nitrogen use efficiency in Arabidopsis thaliana cultivated with limiting and ample nitrogen supply. *Journal of Experimental Botany* **61**: 2293-2302
- Chardon F, Noel V, Masclaux-Daubresse C (2012)** Exploring NUE in crops and in Arabidopsis ideotypes to improve yield and seed quality. *Journal of Experimental Botany* **63**: 3401-3412
- Chon IG, Jung CH (2005)** Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* **78**: :103-112
- Cross JM, von Korff M, Altmann T, Bartzetko L, Sulpice R, Gibon Y, Palacios N, Stitt M (2006)** Variation of enzyme activities and metabolite levels in 24 arabidopsis accessions growing in carbon-limited conditions. *Plant Physiology* **142**: 1574-1588
- Cuadros-Inostroza A, Giavalisco P, Hummel J, Eckardt A, Willmitzer L, Pena-Cortes H (2010)** Discrimination of Wine Attributes by Metabolome Analysis. *Analytical Chemistry* **82**: 3573-3580

- Eriksson L, Joahnsson N, Kettane-Wold N, Tyrgg J, Wikstom C, Wold S (2001)** Multi- and Megavariate Data Analysis: Principles and Applications. Umetrics Academy, Umea, Sweden, 2001.
- Fernie AR, Aharoni A, Willmitzer L, Stitt M, Tohge T, Kopka J, Carroll AJ, Saito K, Fraser PD, DeLuca V (2011)** Recommendations for Reporting Metabolite Data. *Plant Cell* **23**: 2477-2482
- Fernie AR, Stitt M (2012)** On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and network interactions scientific correspondence. *Plant Physiology* **158**: 1139-1145
- Fox J, Weisberg S (2011)** An (R) Companion to Applied Regression, 2nd Edition, 2011 Thousand Oaks, CA USA. <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>
- Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M (2004)** A robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: Comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *Plant Cell* **16**: 3304-3325
- Gibon Y, Pyl ET, Sulpice R, Lunn JE, Hohne M, Gunther M, Stitt M (2009)** Adjustment of growth, starch turnover, protein content and central metabolism to a decrease of the carbon supply when Arabidopsis is grown in very short photoperiods. *Plant Cell and Environment* **32**: 859-874
- Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, Trethewey R, Stitt M (2006)** Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. *Genome Biol* **7**: R76
- Gonzalez N, Beemster GTS, Inzé D. (2009)** David and Goliath: what can the tiny weed Arabidopsis teach us to improve biomass production in crops? *Current Opinion in Plant Biology* **12**: 157-164
- Gonzalez , De Bodt S, Sulpice R, Jikumaru Y, Van Daele T, De Milde L, Weigel D, Kamiya Y, Stitt M, Beemster G, Inzé D (2010)** Increased Leaf Size: Different Means to an End. *Plant Physiology* **53**: 1261-1279
- Hannah MA, Caldana C, Steinhauser D, Balbo I, Fernie AR, Willmitzer L (2010)** Combined Transcript and Metabolite Profiling of Arabidopsis Grown under Widely Variant Growth Conditions Facilitates the Identification of Novel Metabolite-Mediated Regulation of Gene Expression. *Plant Physiology* **152**: 2120-2129

- Hanneman J, Poorter H, Usadel B, Tardieu F, Atkin O, Pons T, Stitt M, Gibon Y** (2009) XEML Lab: a tool for a standardised description of the growth environment of plants. *Plant Cell Environ* **32**: 1185-1200
- Hermans C, Hammond JP, White P, Verbruggen N** (2006) How do plants respond to nutrient shortage by biomass allocation. *Trends in Plant Science* **11**: 610-616
- Hirel B, Le Gouis J, Ney B, Gallais A** (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *Journal of Experimental Botany* **58**: 2369-2387
- Jankanpaa HJ, Mishra Y, Schröder WP, Jansson S** (2012) Metabolic profiling reveals metabolic shifts in Arabidopsis plants grown under different light conditions. *Plant Cell Environ* **35**: 1824–1836
- Keurentjes JJB, Sulpice R, Gibon Y, Steinhauser MC, Fu JY, Koornneef M, Stitt M, Vreugdenhil D** (2008) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. *Genome Biology* **9**: R129
- Kliebenstein DJ** (2012) Plant Defense Compounds: Systems Approaches to Metabolic Analysis. *Annual Review Plant Pathol* **50**: 155-173
- Korn M, Gartner T, Erban A, Kopka J, Selbig J, Hinch DK** (2010) Predicting Arabidopsis Freezing Tolerance and Heterosis in Freezing Tolerance from Metabolite Composition. *Molecular Plant* **3**: 224-235
- Krapp A, Saliba-Colombani V, Daniel-Vedele F** (2005) Analysis of C and N metabolisms and of C/N interactions using quantitative genetics. *Photosynthesis Research* **83**: 251–263
- Krizek BA**. (2009) Making bigger plants: key regulators of final organ size. *Current Opinion in Plant Biology* **12**: 17–22
- Kusano M, Fukushima A, Redestig H, Saito K** (2011) Metabolomic approaches toward understanding nitrogen metabolism in plants. *Journal of Experimental Botany* **62**: 1432–1453
- Lea US, Leydecker MT, Quillere I, Meyer C, Lillo C** (2006) Posttranslational regulation of nitrate reductase strongly affects the levels of free amino acids and nitrate, whereas transcriptional regulation has only minor influence. *Plant Physiology* **140**: 1085-1094
- LeClere S, Schmelz EA, Chourey PS** (2010) Sugar levels regulate tryptophan-dependent auxin biosynthesis in developing maize kernels. *Plant Physiology* **153**: 306-318

- Lilley JLS, Gee CW, Sairenen I, Ljung K, Nemhauser J** (2012) An Endogenous Carbon-Sensing Pathway Triggers Increased Auxin Flux and Hypocotyl Elongation. *Plant Physiology* **160**: 2261–2270,
- Lillo C** (2008) Signalling cascades integrating light-enhanced nitrate metabolism. *Biochemical Journal* **415**: 11-19
- Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR** (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature Protocols* **1**: 387-396
- Loudet O, Gaudon V, Trubuil A, Daniel-Vedele F** (2005) Quantitative trait loci controlling root growth and architecture in *Arabidopsis thaliana* confirmed by heterogeneous inbred family. *Theor Appl Genet* **110**: 742-753
- Matt P, Geiger M, Walch-Liu P, Engels C, Krapp A, Stitt M** (2001) Elevated carbon dioxide increases nitrate uptake and nitrate reductase activity when tobacco is growing on nitrate, but increases ammonium uptake and inhibits nitrate reductase activity when tobacco is growing on ammonium nitrate. *Plant Cell and Environment* **24**: 1119-1137
- Mayashita Y, Good AC** (2008) NAD(H)-dependent glutamate dehydrogenase is essential for the survival of *Arabidopsis thaliana* during dark-induced carbon starvation. *Journal of Experimental Botany*. doi:10.1093/jxb/erm340
- Melo-Oliveira R, Oliveira IC, Coruzzi GM.** (1996). *Arabidopsis* mutant analysis and gene regulation define a nonredundant role for glutamate dehydrogenase in nitrogen assimilation. *Proceeding of the National Academy USA* **93**: 4718–4723
- Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T** (2007) The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 4759-4764
- Meyer RC, Witucka-Wall H, Becher M, Blacha A, Boudichevskaia A, Dormann P, Fiehn O, Friedel S, von Korff M, Lisec J, Melzer M, Repsilber D, Schmidt R, Scholz M, Selbig J, Willmitzer L, Altmann T** (2012) Heterosis manifestation during early *Arabidopsis* seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant Journal* **71**: 669-683
- Mevik BH, Wehrens R** (2007) The pls package: Principal component and partial least squares regression in r. *Journal of Statistical Software* **18**: 1-24.



- Moll RH, Kamprath EJ, Jackson WA** (1982) Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. *Agronomy Journal* **74**: 562-564.
- North KA, Ehlting B, Koprivova A, Rennenberg H, Kopriva S** (2009) Natural variation in *Arabidopsis* adaptation to growth at low nitrogen conditions. *Plant Physiology and Biochemistry* **47**: 912-918
- Nunes-Nesi A, Carrari F, Gibon Y, Sulpice R, Lytovchenko A, Fisahn J, Graham J, Ratcliffe RG, Sweetlove LJ, Fernie AR** (2007) Deficiency of mitochondrial fumarase activity in tomato plants impairs photosynthesis via an effect on stomatal function. *Plant Journal* **50**: 1093-1106
- Nunes-Nesi A, Fernie AR, Stitt M.**(2010) Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Molecular Plant*. 3::973-96
- Obata T, Fernie AR** (2012) The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences* **69**: 3225-3243
- Piques M, Schulze WX, Höhne M, Usadel B, Gibon Y, Rohwer J, Stitt M** (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in *Arabidopsis*. *Molecular Systems Biology* **5**: doi:10.1038/msb.2009.1068
- Poorter H** (1989) Growth analysis: towards a synthesis of the classical and the functional approach. *Physiologia Plantarum* **75**: 237-244
- Poorter H, Nagel O** (2000) The role of biomass allocation in the growth response of plants to different levels of light, CO<sub>2</sub>, nutrients and water: a quantitative review (vol 27, pg 595, 2000). *Australian Journal of Plant Physiology* **27**: 1191-1191
- Poorter H, Niinemets U, Poorter L, Wright IJ, Villar R** (2009) Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytologist* **182**: 565-588
- Raven JA** (2012) Protein turnover and plant RNA and phosphorus requirements in relation to nitrogen fixation (vol 188, pg 25, 2012). *Plant Science* **190**: 131-131
- Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisek J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE** (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet* **44**: 217-220

- Robert P, Escoufier Y** (1976) Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **25**: 257-265
- Saito K, Matsuda F** (2010) Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. *Annual Review of Plant Biology*, Vol 61 **61**: 463-489
- Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR** (2008) Mode of inheritance of primary metabolic traits in tomato. *Plant Cell* **20**: 509-523
- Smith AM, Stitt M** (2007) Coordination of carbon supply and plant growth. *Plant Cell and Environment* **30**: 1126-1149
- Steinfath M, Strehmel N, Peters R, Schauer N, Groth D, Hummel J, Steup M, Selbig J, Kopka J, Geigenberger P, van Dongen JT** (2010a) Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. *Plant Biotechnology Journal* **8**: 900-911
- Steinfath M, Gartner T, Lisek J, Meyer RC, Altmann T, Willmitzer L, Selbig J** (2010b) Prediction of hybrid biomass in *Arabidopsis thaliana* by selected parental SNP and metabolic markers. *Theoretical and Applied Genetics* **120**: 239-247
- Stitt M, Krapp A** (1999) The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. *Plant Cell and Environment* **22**: 583-621
- Stitt M, Sulpice R, Keurentjes J** (2010) Metabolic Networks: How to Identify Key Components in the Regulation of Metabolism and Growth. *Plant Physiology* **152**: 428-444
- Stitt M, Zeeman SC** (2012) Starch turnover: pathways, regulation and role in growth. *Curr Opin Plant Biol* **15**: 282-292
- Sulpice R, Pyl ET, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC, Von Korff M, Steinhauser MC, Keurentjes JJB, Guenther M, Hoehne M, Selbig J, Fernie AR, Altmann T, Stitt M** (2009) Starch as a major integrator in the regulation of plant growth. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 10348-10353
- Sulpice R, Trenkamp S, Steinfath M, Usadel B, Gibon Y, Witucka-Wall H, Pyl ET, Tschoep H, Steinhauser MC, Guenther M, Hoehne M, Rohwer JM, Altmann T, Fernie AR, Stitt M** (2010) Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of *Arabidopsis* accessions. *Plant Cell* **22**: 2872-2893

- Tschoep H, Gibon Y, Carillo P, Armengaud P, Szecowka M, Nunes-Nesi A, Fernie AR, Koehl K, Stitt M** (2009) Adjustment of growth and central metabolism to a mild but sustained nitrogen-limitation in Arabidopsis. *Plant Cell and Environment* **32**: 300-318
- Usadel B, Blasing OE, Gibon Y, Retzlaff K, Hoehne M, Gunther M, Stitt M** (2008) Global transcript levels respond to small changes of the carbon status during progressive exhaustion of carbohydrates in Arabidopsis rosettes. *Plant Physiology* **146**: 1834-1861
- Wold H** (1966) Estimation of Principal Components and Related Models by Iterative Least squares, pages 391-420. Academic Press, NY, USA.
- Xu G, Fan X, Miller AJ** (2012) Plant Nitrogen Assimilation and Use Efficiency. *Annual Review of Plant Biology* **63**:153–82
- Zamboni N** (2011) C-13 metabolic flux analysis in complex systems. *Current Opinion in Biotechnology* **22**: 103-108
- Züst T, Heichinger C, Grossniklaus , Harrington R, Kliebenstein DJ, Turnbull LA** (2012) Natural Enemies Drive Geographic Variation in Plant Defenses. *Science* **338**: 116-119

**Figure 1.** Biomass of 97 accessions in the three growth conditions. A panel of 97 *Arabidopsis* accessions was grown in a 12h photoperiod with high nitrogen (12hHN), a 12h photoperiod with low nitrogen (12hLN) and a 8 h photoperiod with high nitrogen (8hHN). (A) Biomass in each conditions the accessions are ordered on the x-axis according to their biomass in the control treatment (12hHN). (B) Three-dimensional plot of the biomass of each accession in the three growth conditions; accessions that have a high and low ranking for biomass in all three conditions are indicated by green and red symbols, respectively. (C) Spearman's correlation between biomass in the three growth conditions. The original data is given in Supplemental Table SI.

**Figure 2.** Structural and metabolic traits. (A) Principle components analysis PCA based on z-scores values for 97 accessions grown in three contrasting conditions (12hHN, ●; 12hLN, ●; 8hHN, ●). In total, 58 traits were determined per accession. The full data set for each growth condition are given in Supplemental Table SI. VIP scores for the metabolic trait inputs are given in Supplemental Table SVII. (B) Heat maps of the metabolite changes under nitrogen deficiency (12hLN) or short days (8hHN) compared to control conditions (12hHN). Each square represents the  $\log_2$  ratio of the metabolite level using false color scale. Regions of red or blue indicate the metabolite level is decreased or increased, respectively. The full data set for each growth condition as given in Supplemental Table SIV.

**Figure 3.** Comparison of the correlation matrices for structural and metabolic traits in three growth conditions. Three separate networks were generated from the values of 50 metabolic traits, starch, protein and biomass in 97 *Arabidopsis* accessions grown in different growth conditions: 12hHN, 12hLN and 8hHN. (A) Analysis of the RV coefficient between the matrices. The lower left hand segment shows the RV coefficients and the upper right hand segment the p-values. The RV coefficient values varies between +1 (two identical matrices) and zero (no similarity). (B) Condensed heat map of correlation coefficients (Spearman R, FW basis) in the networks for 12hHN, 12hLN and 8hHN. The individual values are provided in

Supplemental Table SI.. Positive (blue) and negative (red) correlations are indicated at a level of  $p < 0.01$ ,  $< 0.001$  and  $< 0.0001$  (light, medium and dark, respectively). The order of traits is as in Supplemental Table SII. (C) Numbers of pairwise correlations between traits in one, two or all three growth conditions. The original matrices are provided in Supplemental Table SV. (D) Visualisation of the numbers of shared- and non-shared correlations ( $FD < 0.01$ ). The area of the circle depicts the number of shared correlations. The grey shell shows how overlap increases if  $FD$  is relaxed to  $FD < 0.1$  all except one of the shared growth conditions.

**Figure 4. N concentration and N content in accessions growing in low and optimal N supply.** (A) Rosette biomass and total N in nitrate, protein, amino acids and chlorophyll ( $\diamond \diamond$ ) and against N in nitrate ( $\triangle \triangle$ ). Total N contents are calculated in Supplemental Table VIII. (B) Relation between nitrate concentration and rosette biomass in 12hLN; data in panel A is re-plotted with a expanded y.-axis scale. (C-E) The x-axis ranks accessions according to their gain in biomass between low N and high-N growth conditions. The x-axis shows the values for a given accession for 12hHN ( $\circ$ ) and 12hLN ( $\triangle$ ). (C) Rosette biomass. (D) Summed N content in protein, amino acids, chlorophyll and nitrate. (E) Summed N content per rosette.

## Supplemental Material

**Supplemental Table SI.** Biomass, structural components, starch, low molecular weight metabolites and enzyme activities plus correlation matrices. Data are provided for 97 accessions grown optimal supply of nutrients and a 8h/16h light/dark regime and optimal supply of N (8hHN; Sulpice et al., 2009; Sulpice et al., 2010), a 12h/12h light/dark regime an optimal supply of N 12hHN) and a 12h/12h light/dark regime with suboptimal supply of N 12hLN), as well as the correlation matrices and for each condition.

**Supplemental Table SII.** ANOVA analysis. .the original data are given in Supplemental Table SI

**Supplemental Table SIII.** Principle components analysis.

**Supplemental Table SIV.** Heat map of changes in metabolic traits under low C and low N conditions compared to near optimal conditions.

**Supplemental Table SV.** Correlation of individual metabolic traits with biomass. Values were determined by Speamans R.

**Supplemental Table SVI.** Correlation matrices for metabolic traits in each of the three growth conditions

**Supplemental Table SVII.** PLS regression for biomass, starch or protein concentration on low molecular weight metabolites in three growth conditions. The inputs are the levels of all measured metabolites in that growth condition. The outputs are biomass, starch or protein in that growth condition. The Table lists, for each growth condition and output, the values of R and p in cross validation, the VIP (value in the prediction) of each metabolite for a given output in a given growth condition.

**Supplemental Table SVIII.** Calculation of nitrogen contents

**Supplemental Table SIX.** Linear mixed models for biomass, starch, and protein concentration: analysis of deviance tables. The first three tables include the results (fixed effects and statistical

significance) with random intercept only, while the fourth table includes the findings for a mixed model with not only random intercept but also random slopes for seven metabolites (chlorophyll a, sucrose, myo-inositol, aspartate, glycine, serine, nicotinic acid) which were identified as having significant coefficients in the regression for biomass (see the first table herein).

**Supplemental Table SX.** Comparison with published data for biomass in Arabidopsis accessions

**Supplemental Figure 1.** Relation between photoperiod and biomass in Col0 wildtype. The plants were germinated and seedling set was as described in the Materials and Methods until 19 days, when the plants were transferred to different photoperiods with a light intensity of  $145 \mu\text{mol m}^{-2} \text{s}^{-1}$ ,  $20^\circ\text{C}$  day and night. Light intensities were checked in each growth cabinet. Plants were harvested at 29 days to determine above-ground fresh weight. The results are the mean  $\pm$ SD (N = 5 replicates, each of 5 plants).

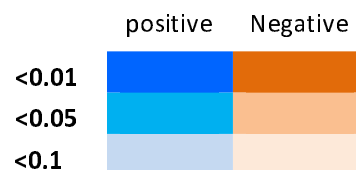
**Supplemental Figure 2.** Coefficient of variations (CV) for the metabolic traits analyzed between the accessions grown in a given growth condition (A) or averaged CV of all the accessions for the variation in metabolite amounts between 2 growth conditions (B). The full data sets for each growth condition are given in Supplemental Table SI.

**Supplemental Figure 3.** Relationship between biomass and selected metabolic traits in different growth conditions. The full data sets for each growth condition are given in Supplemental Table SI.

**Supplemental Figure 4.** Relationship between selected metabolic traits (starch, protein) in different growth conditions. The full data sets for each growth condition are given in Supplemental Table SI.

**Table I.** Spearman rank correlation coefficient between biomass and metabolic traits in different growth conditions. Adjusted p-values were calculated using the Benjamini-Hochberg correction. This display summarises metabolites that correlated at  $p < 0.05$  in at least one growth condition. A full set of correlations is provided in Supplemental Table SV. NA = not available

	12hHN	12hLN	8hHN
protein	-0.30	-0.14	-0.39
chl a	-0.27	-0.14	-0.27
chl b	-0.13	-0.29	-0.31
starch	-0.49	-0.33	-0.54
sucrose	-0.33	-0.30	-0.13
Maltose	-0.05	-0.33	0.05
Trehalose	-0.05	-0.50	-0.08
Raffinose	-0.06	-0.36	-0.37
Inositol. myo	-0.05	-0.28	0.16
Xylose	-0.25	-0.09	-0.06
NO3	-0.18	-0.50	NA
AA	-0.20	0.04	-0.33
Glutamic acid	-0.25	0.20	-0.27
Aspartic acid	-0.42	-0.12	-0.25
Alanine	-0.32	-0.34	-0.26
Glycine	-0.34	-0.16	-0.39
Arginine	-0.07	0.13	-0.30
Asparagine	-0.14	0.27	-0.24
Valine	-0.33	-0.38	-0.34
Isoleucine	-0.26	-0.33	-0.08
Leucine	-0.04	-0.32	0.02
Proline	-0.27	-0.14	-0.10
Phenylalanine	-0.34	0.01	-0.14
Tryptophan	-0.29	0.06	-0.18
Shikimic acid	-0.31	-0.28	0.03
Benzoic acid	-0.05	0.26	0.02
malate	-0.29	-0.39	-0.07
fumarate	-0.13	-0.33	-0.14
succinate	-0.32	-0.46	-0.32
Glyceric acid	-0.14	-0.42	-0.12
Proline. 4-hydroxy	-0.30	-0.32	0.01
Putrescine	-0.19	0.38	-0.26
Spermidine	0.28	0.17	NA
Threonate	-0.06	-0.28	-0.11
Dehydroascorbate	-0.06	-0.16	-0.26
NADGIDH	-0.11	0.15	0.34
NR_Vmax	-0.22	-0.31	-0.04
PEPCx	-0.37	0.09	0.27





**Table II.** Partial least squares (PLS) regression analysis of the relation between low molecular weight metabolites and biomass, starch level or protein concentration. PLS was performed using 52 metabolic traits from condition A as input and fresh weight (FW), starch or protein as output traits from condition B, for each of the nine combinations of conditions (A, B). The correlation between the predicted values and the response variables, denoted by R, was determined after cross-validated selection of the number of latent variables (n). The number of latent variables (n) and the p-values were determined in two successive and independent rounds of cross validation, the are determined by permutation test (see Materials and Methods).

		12hHN		12hLN		8hHN		Means across all three conditions		All individual values across all three conditions	
		R / n	p	R / n	p	R / n	p	R / n	p	R / n	p
12hHN	FW	<b>0.36 / 1</b>	0.002	<b>0.26 / 1</b>	0.006	<b>0.28 / 1</b>	0.002	<b>0.45 / 1</b>	0.001	<b>0.73 / 3</b>	0.001
	Starch	<b>0.67 / 1</b>	0.001	<b>0.47 / 1</b>	0.001	<b>0.68 / 1</b>	0.001	<b>0.61 / 2</b>	0.001	<b>0.77 / 31</b>	0.001
	Protein	<b>0.46 / 1</b>	0.001	<b>0.34 / 1</b>	0.001	<b>0.46 / 1</b>	0.001	<b>0.57 / 3</b>	0.001	<b>0.94 / 32</b>	0,001
12hLN	FW	<b>0.51 / 1</b>	0.001	<b>0.58 / 1</b>	0.001	<b>0.51 / 1</b>	0.001				
	Starch	<b>0.35 / 1</b>	0.001	<b>0.39 / 1</b>	0.001	<b>0.36 / 3</b>	0.001				
	Protein	<b>0.45 / 1</b>	0.001	<b>0.57 / 2</b>	0.001	<b>0.34 / 1</b>	0.002				
8hHN	FW	<b>0.27 / 1</b>	0.003	<b>0.21 / 1</b>	0.023	<b>0.27 / 1</b>	0.013				
	Starch	<b>0.26 / 1</b>	0.006	0.05 / 2	0.302	<b>0.23 / 1</b>	0.023				
	Protein	0.13 / 1	0.102	0.17 / 1	0.057	0.17 / 1	0.072				

**Table III. Correlation between the VIP of metabolic traits in the PLS regression on biomass, starch and protein.** The VIP of the metabolite traits in the PLS regressions on biomass, starch and protein concentration are provided in Supplemental Table SVII)

	Correlation (R) between VIP of inputs in the PLS regression					
	Biomass and starch		Biomass and protein		Starch and protein	
	R	p	R	p	R	p
12hHN	<b>0.81</b>	0.000	<b>0.43</b>	0.001	<b>0.65</b>	0.000
12hLN	<b>0.28</b>	0.042	0.19	0.172	<b>0.53</b>	0.000
8hHN	0.26	0.059	<b>0.55</b>	0.000	0.15	0.278

Figure 1. Biomass of 91 accessions in the three growth conditions. A panel of 92 *Arabidopsis* accessions was grown in a 12h photoperiod with high nitrogen (12hHN), a 12h photoperiod with low nitrogen (12hLN) and a 8 h photoperiod with high nitrogen (8hHN). (A) Biomass in each conditions the accessions are ordered on the x-axis according to their biomass in the control treatment (12hHN). (B) Three-dimensional plot of the biomass of each accession in the three growth conditions; accessions that have a high and low ranking for biomass in all three conditions are indicated by green and red symbols, respectively. (C) Correlation between biomass in the three growth conditions. The original data is given in supplemental Table SI, and scatter plots from which the regression coefficients were calculated in Supplemental Figure S1.

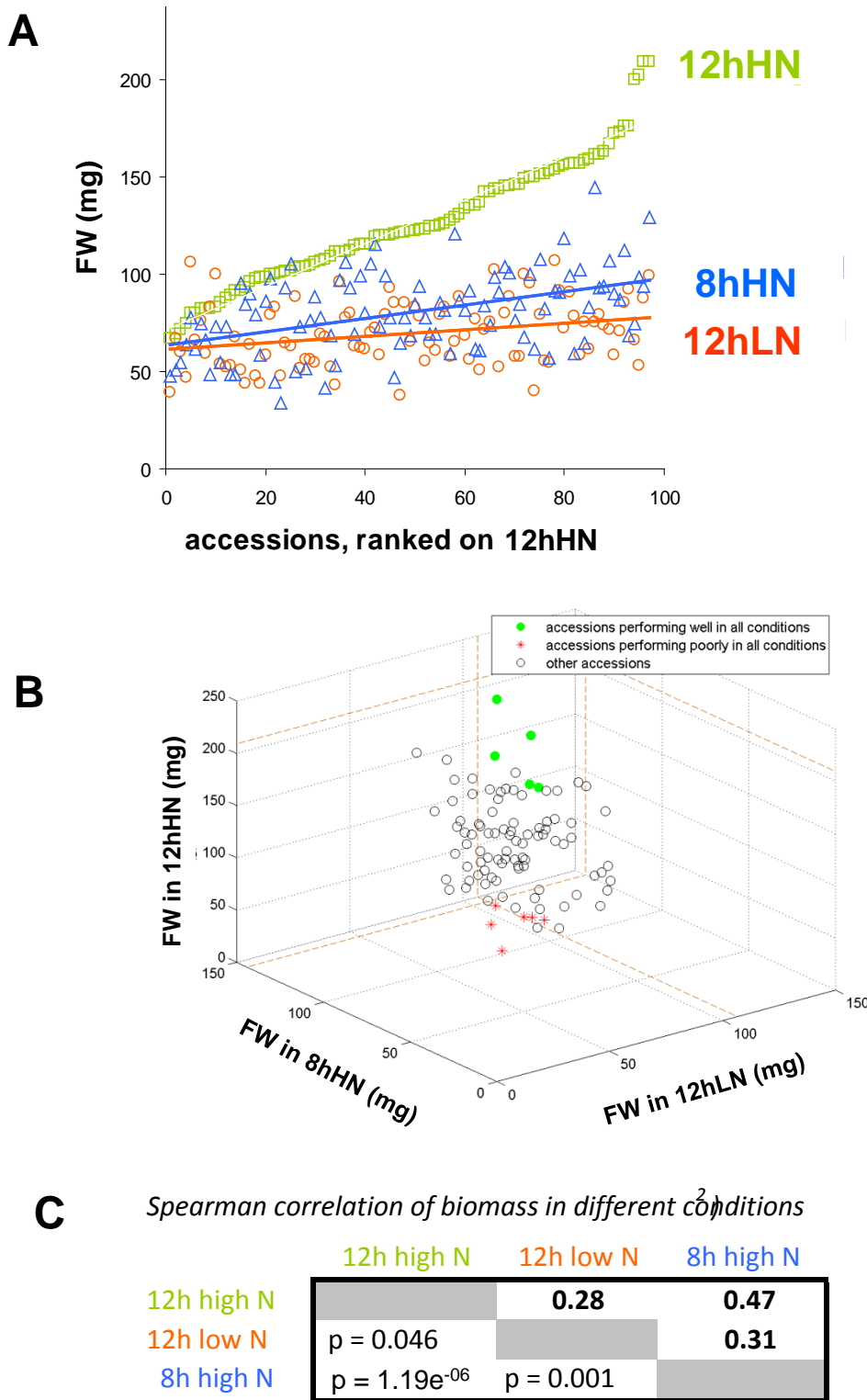
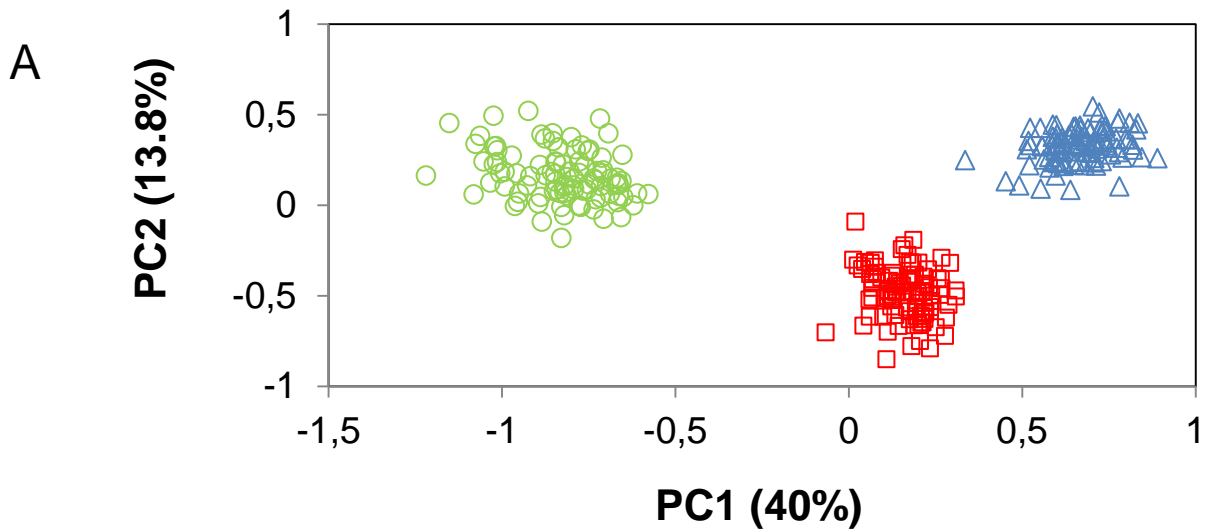
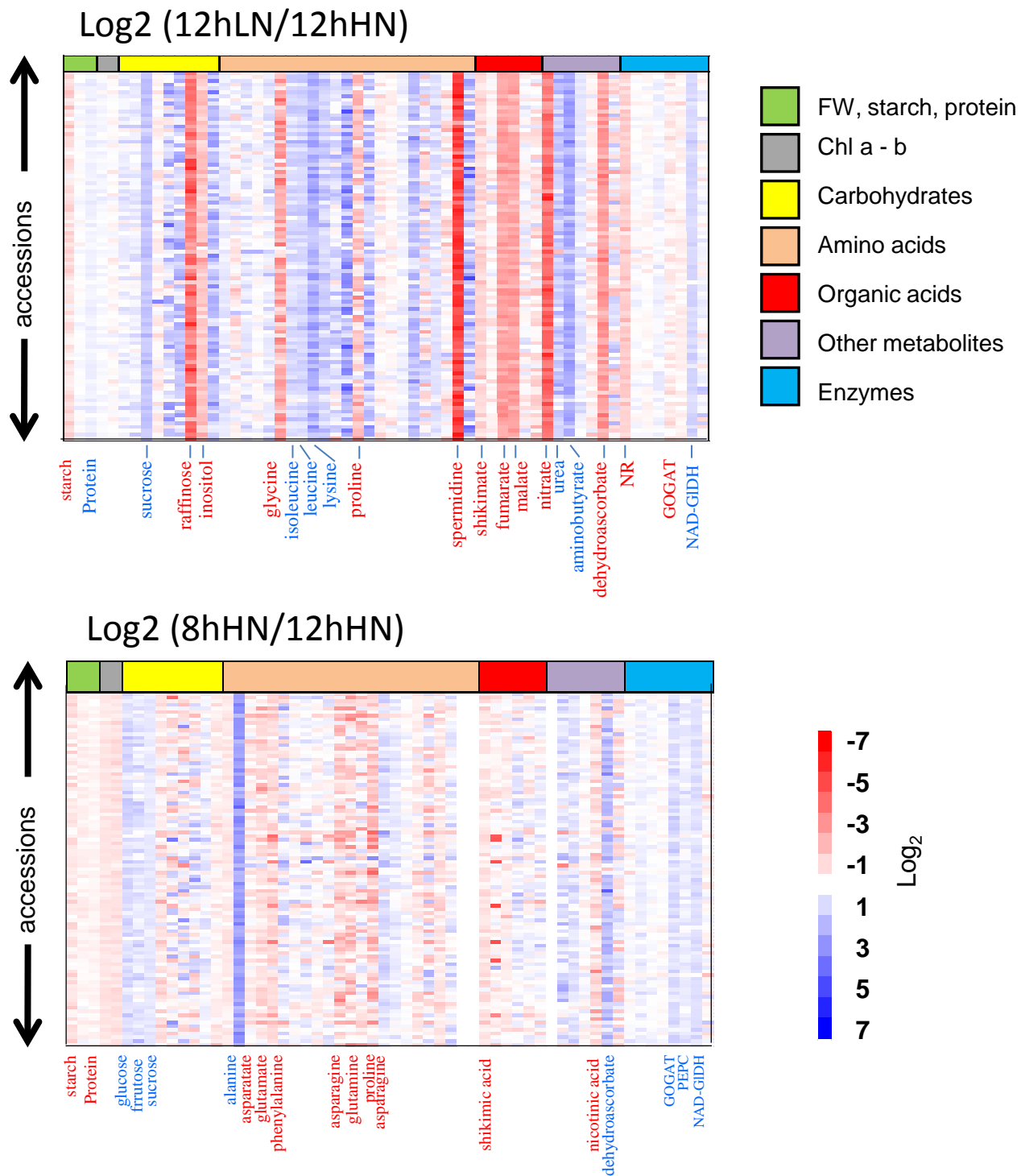


Figure 2. Structural and metabolic traits.

(A) Principle components analysis PCA based on z-scores values for 92 accessions grown in three contrasting conditions (12hHN, ●; 12hLN, ●; 8hHN, ●). In total, 58 traits were determined per accession. The full data set for each growth condition are given in Supplemental Table SI. VIP scores for the metabolic trait inputs are given in Supplemental Table SVII.



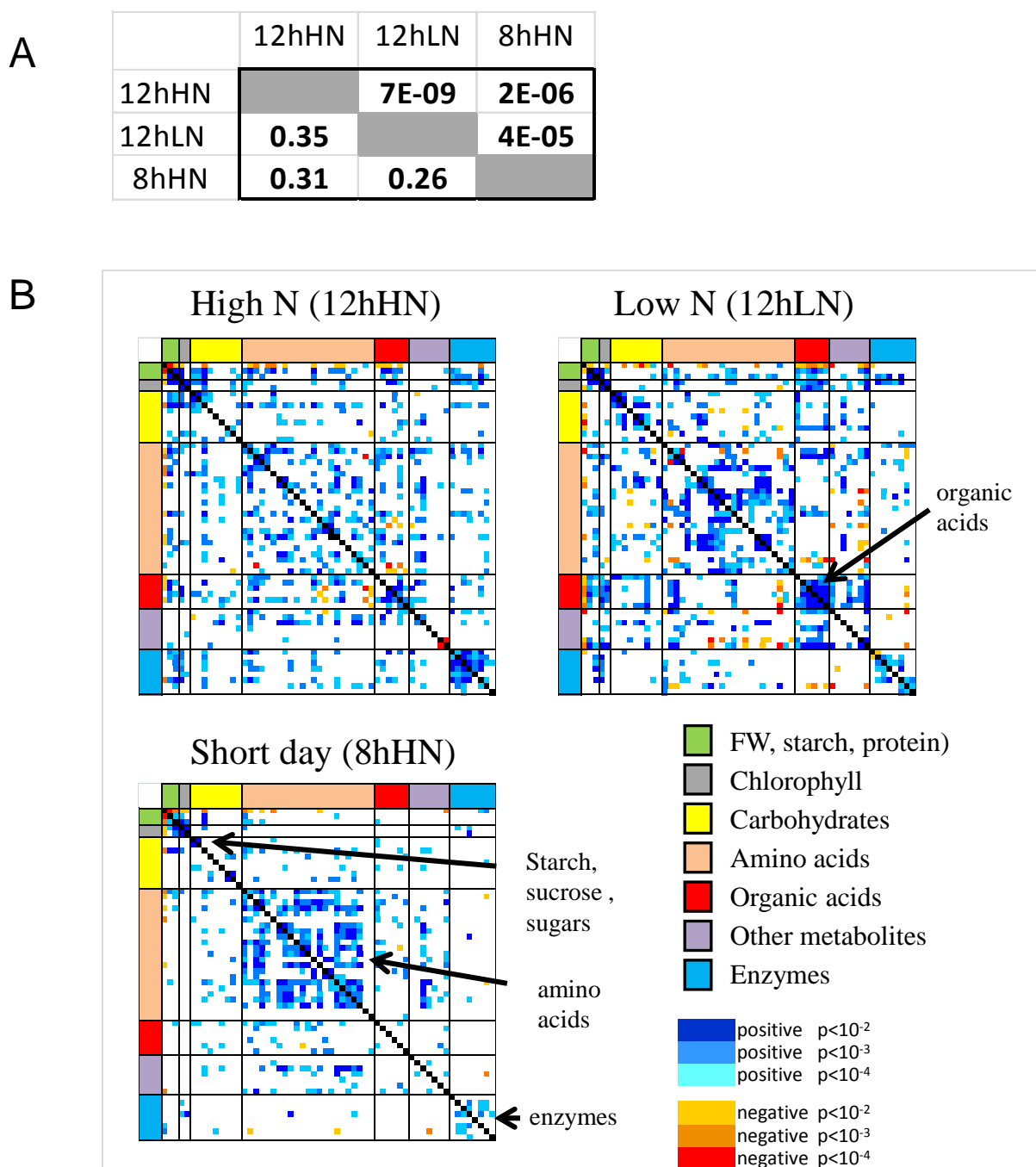
**Figure 2B** Heat maps of the metabolite changes under nitrogen deficiency (12hLN) or short days (8hHN) compared to control conditions (12hHN). Each square represents the log<sub>2</sub> ratio of the metabolite level using false color scale. Regions of red or blue indicate the metabolite level is decreased or increased, respectively. The full data set for each growth condition as given in Supplemental Table SIV.



**Figure 3.** Comparison of the correlation matrices for structural and metabolic traits in three growth conditions. Three separate networks were generated from the values of 50 metabolic traits, starch, protein and biomass in 92 *Arabidopsis* accessions grown in different growth conditions: 12hHN, 12hLN and 8hHN.

(A) Analysis of the RV coefficient between the matrices. The lower left hand segment shows the RV coefficients and the upper right hand segment the p-values. The RV coefficient values varies between +1 (two identical matrices) and zero (no similarity).

(B) Condensed heat map of correlation coefficients (Spearman R, FW basis) in the networks for 12hHN, 12hLN and 8hHN. The individual values are provided in Supplemental Table SI. Positive (blue) and negative (red) correlations are indicated at a level of  $p < 0.01$ ,  $< 0.001$  and  $< 0.0001$  (light, medium and dark, respectively. The order of traits is as in Supplemental Table SII.)



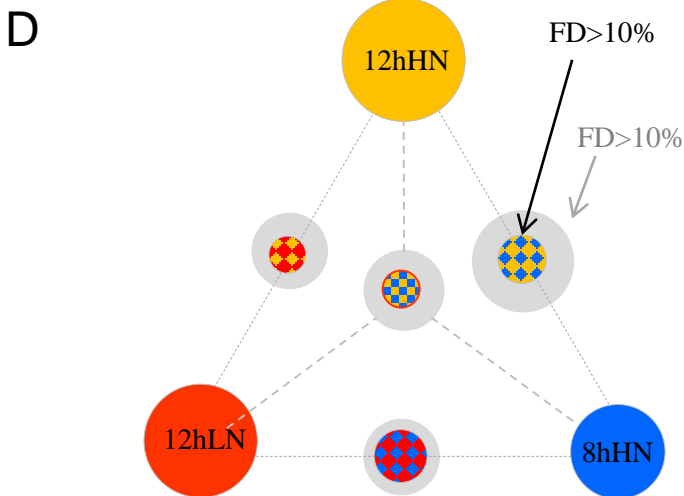
### Figure 3 continued

(C) Numbers of pairwise correlations between traits in one, two or all three growth conditions. The original matrices are provided in Supplemental Table SV.

(D) Visualization of the numbers of shared- and non-shared correlations (  $FD < 0.01$  ). The area of the circle depicts the number of shared correlations . The grey shell shows how overlap increases if FD is relaxed to  $FD < 10\%$  in all except one of the shared growth conditions.

**C**

	12hHN	12hLN	8hHN	number of significant correlations		
				10%	5%	1%
12hHN	418	8	32	418	374	248
12hLN	8	394	32	8	364	264
8hHN	32	32	216	32	36	14
12hHN & 12hLN	230	0	0	230	168	78
12hHN & 8hHN	144	2	4	144	114	64
12hLN & 8hHN	148	4	166	148	100	38
12hHN & 12hLN & 8hHN	0	0	0	0	0	0
12hHN & 8hHN (FD > 10%)	2	8	4	2	0	0
12hLN & 8hHN (FD > 10%)	4	4	6	4	4	2



**Figure 4.** A) Rosette biomass and total N in nitrate, protein, amino acids and chlorophyll ( $\diamond$   $\diamond$ ) and against N in nitrate ( $\triangle$   $\triangle$ ). Total N contents are calculated in Supplemental Table VIII. (B) Relation between nitrate concentration and rosette biomass in 12hLN; data in panel A is re-plotted with an expanded y-axis scale. (C-E) The x-axis ranks accessions according to their gain in biomass between low N and high-N growth conditions. The x-axis shows the values for a given accession for 12hHN ( $\circ$ ) and 12hLN ( $\triangle$ ). (C) Rosette biomass. (D) Summed N content in protein, amino acids, chlorophyll and nitrate. (E) Summed N content per rosette.

