



# Inheritance patterns in metabolism and growth in diallel crosses of *Arabidopsis thaliana* from a single growth habitat

Andrés Eduardo Rodríguez Cubillos<sup>1</sup> · Hao Tong<sup>1,2</sup> · Saleh Alseekh<sup>1</sup> · Francisco de Abreu e Lima<sup>1</sup> · Jing Yu<sup>1</sup> · Alisdair R. Fernie<sup>1</sup> · Zoran Nikoloski<sup>1,2</sup> · Roosa A.E. Laitinen<sup>1</sup>

Received: 23 August 2017 / Revised: 9 October 2017 / Accepted: 30 October 2017 / Published online: 13 December 2017  
© The Genetics Society 2018

## Abstract

Metabolism is a key determinant of plant growth and modulates plant adaptive responses. Increased metabolic variation due to heterozygosity may be beneficial for highly homozygous plants if their progeny is to respond to sudden changes in the habitat. Here, we investigate the extent to which heterozygosity contributes to the variation in metabolism and size of hybrids of *Arabidopsis thaliana* whose parents are from a single growth habitat. We created full diallel crosses among seven parents, originating from Southern Germany, and analysed the inheritance patterns in primary and secondary metabolism as well as in rosette size in situ. In comparison to primary metabolites, compounds from secondary metabolism were more variable and showed more pronounced non-additive inheritance patterns which could be attributed to epistasis. In addition, we showed that glucosinolates, among other secondary metabolites, were positively correlated with a proxy for plant size. Therefore, our study demonstrates that heterozygosity in local *A. thaliana* population generates metabolic variation and may impact several tasks directly linked to metabolism.

## Introduction

Phenotypic traits can be inherited in either an additive or non-additive manner (Falconer and Mackay 1996). For traits that are inherited in an additive manner, the phenotypes observed in the progeny are the average of the effect of parental gene(s)

and do not differ from the average phenotype of the parents (i.e., the mid-parent phenotype). The traits that deviate from the mid-parent phenotype are inherited in a non-additive way and can be either beneficial or disadvantageous for hybrid fitness in comparison to the parents (Charlesworth and Willis 2009; Davenport 1908; Schwartz and Laughner 1969). There are three contending explanations for non-additive inheritance, namely: dominance, over-dominance and epistasis (Charlesworth and Willis 2009; Davenport 1908; Hull 1945; Powers 1944). According to the dominance theory, the phenotypic effect of a deleterious allele from one parent can be complemented in a hybrid with a dominant allele from the other parent (Davenport 1908). By the over-dominance explanation, heterozygosity results in trait scores that are higher in the hybrid than in either of the parents. In addition, non-additive inheritance can be due to epistasis, whereby in hybrids a locus from one parent influences another locus from the other parent. Irrespective of the explanation used, non-additive inheritance is expected to have large implications for adaptation to changing environments, particularly relevant for sessile organisms, since it provides strategies which expand beyond those of the range provided by the parents.

Understanding the basis of non-additive inheritance has prospects to further improve hybrid breeding, and the effects of non-additive inheritance patterns have been

---

Andrés Eduardo Rodríguez Cubillos and Hao Tong have equally contributed to this work and are first authors

---

Zoran Nikoloski and Roosa A. E. Laitinen have equally contributed to this work and are corresponding authors

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41437-017-0030-5>) contains supplementary material, which is available to authorized users.

---

✉ Zoran Nikoloski  
Nikoloski@mpimp-golm.mpg.de

✉ Roosa A. E. Laitinen  
Laitinen@mpimp-golm.mpg.de

<sup>1</sup> Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany

<sup>2</sup> Bioinformatics Group, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany

traditionally studied in several crop species (Alseekh et al 2015; Bestfleisch et al 2013; Boer et al 2007; Riedelsheimer et al 2012a; Schauer et al 2008). So far, studies on natural variation among global *Arabidopsis thaliana* accessions have helped us understand genetic mechanisms underlying genes that cause differential fitness in hybrids in comparison to the parents (Weigel 2012) together with the genetic basis of non-additive inheritance patterns (Seymour et al 2016). While most plant species cross-fertilize, *A. thaliana* is mainly self-fertilizing. The outcrossing rates of *A. thaliana* on the field vary according to growth habitat, but are on average ~0–3% and can obtain values of up to 10% (Abbott and Gomes 1989; Bomblies et al 2010; Platt et al 2010), indicating that heterozygosity could play a role in adding genetic variation in *A. thaliana* growing in one local habitat.

We hypothesized that increased variation due to heterozygosity in a single natural growth habitat could be a beneficial strategy for highly homozygous plants to rapidly respond to sudden changes in their surroundings. To test the impact of heterozygosity on variability in local *A. thaliana*, we studied non-additive inheritance of metabolic variation and growth in a set of individuals coming from a single growth habitat. The metabolome of a plant directly reflects the physiological responses leading to traits influencing fitness, such as: stress resistance, yield, and growth (Hirai et al 2004; Parker et al 2009; Sulpice et al 2009). Therefore, analysis of metabolic variation gives us information about the different metabolic strategies plants can use to cope with changing environments (Obata and Fernie 2012). Here, we created a full diallel cross among seven genetically monitored individuals originating from a single habitation in Tübingen, Germany in 2007 (Swiadek et al 2017) and monitored the metabolism and growth of hybrids and parents to investigate the extent of variation in these traits.

Our analysis showed that the metabolite levels and rosette size (as a proxy for growth) of hybrids exhibited higher variability in comparison to those of the parents and that non-additive mode of inheritance is typically observed for secondary metabolism. The correlation analyses highlighted the role of non-additive inheritance of secondary metabolism in modulating growth. Therefore, our study provides novel insights on the influence of heterozygosity in the genetic and phenotypic landscape of plants growing at a single growth habitat.

## Materials and methods

### Growth conditions and phenotyping

Accessions from different collection sites collected in 2007, including the Altenriet (Alt) individuals, were already described in (Bomblies et al 2010). For full diallel crosses,

two plants were used to create two sets of crosses in order to control the biological variation of individual plants. To avoid possible heterotic effects due to manual fertilization (Meyer et al 2004) and effects arising from different individuals, the same parent was used for all crosses in one replicate and the parental seeds were also produced by manual crosses. Plants were grown in randomized individual pots under long day conditions (16 h light/ 8 h dark) at 21 °C/17 °C in growth chambers. To control the local effects of the growth chambers, the trays (each containing 30 pots) were moved and turned every second day. Rosette radii (from the middle of the rosette to the end of the leaf tip) were measured from photos taken from plants that reached a 10-leaf stage using imageJ (version 1.48). For metabolic profiling, whole rosettes were harvested at the 10-leaf stage in the middle of the day (between 12:00–14:00) to avoid any bias due to circadian effects. All analyses were conducted with at least four biological replicates.

### Metabolic profiling

Extraction and analysis by gas chromatography mass spectrometry were performed using the same equipment set up and exact same protocol as described in Lisec et al (2006). Chromatograms and mass spectra were evaluated using the Chroma TOF 4.5 (Leco) and TagFinder 4.2 softwares (Roessner et al 2001; Schauer et al 2005). Secondary metabolite analysis was performed as previously described by Tohge and Fernie (2010) using a high-performance liquid chromatography (HPLC; Surveyor; Thermo Finnigan, USA) coupled to a Finnigan LTQ-XP system (Thermo Finnigan, USA). Metabolites were evaluated on the basis of the peak area of parental ion peaks processed using Xcalibur 2.1 software (Thermo Fisher Scientific). The obtained relative peak areas were normalized by comparison to an internal standard (isovitexin; CAS29702-25-8) and the fresh weight of the sample used for extraction.

### Statistical analyses

We used Bonferroni correction for multiple hypotheses testing throughout the manuscript where applicable, and reported significance at the level of 0.05 if not otherwise stated.

### Correlation analyses

For the correlation matrix between SNPs, 721 non-polymorphic SNPs in seven parents were excluded from the original 1985 SNPs. Since hybrid SNPs can arise from multiple possibilities if the SNP from one of the parents is heterozygous, all SNPs that were heterozygous in parents were excluded as well; hence, 1149 SNPs were used for the correlation analysis. The Pearson correlation based on

genotype was performed among each pair of alleles for parents or hybrids separately. SNP pairs were coded with the same number independent of the direction of the cross. For the correlation matrix between metabolites, the mean values over the replicates were used. The Mantel correlation coefficient was then determined by using the correlation matrices between genotypes based on primary metabolites, secondary metabolites, and SNPs.

### Analysis of metabolic variation within hybrids and parents

Metabolite data was first normalized for differences in fresh weight and machine performance using an internal standard. All metabolite intensities were log<sub>10</sub>-transformed to render normally distributed data. Each metabolite value was then scaled by its standard deviation and all values with more than five standard deviations away from the mean were treated as missing values. Consequently, all metabolites with more than 20% of missing values were eliminated from posterior analyses and the remaining missing values were imputed using the missForest package available for the free software statistical environment R (Stekhoven and Bühlmann 2012).

Principal component analyses were conducted by using the function `prcomp` within the R statistical environment. To determine the significance of the observed difference between variance explained by the first two principal components between secondary and primary metabolites, i.e., 72.8%–57.7% = 15.1% (see Results), we performed the following permutation test: (1) shuffle each data set metabolite-wise, (2) determine the difference of the variance explained by the two components of the shuffled data, (3) count the value in step 2 if it is larger than the observed one, and store the value in *ctr* and (4) determine *p*-value as  $(ctr + 1)/(B + 1)$ , where *B* is the number of shuffling repetitions.

To identify the classes of metabolites that were deviating most from the average mid-parent value, the relative percentage of deviation from the mid-parent value (rMPD) was calculated for each metabolite within each hybrid using the formula  $rMPD = 100d/a$ , where *d* is the difference between the hybrid and parental mean and *a* is the parental mean.

The coefficient of variation was calculated by dividing the standard deviation of each metabolite by its mean, or across all hybrids (CV<sub>h</sub>) or all parents (CV<sub>p</sub>). Hence, the log<sub>2</sub> CV-ratios were calculated as  $\log_2(CV_h/CV_p)$ . The mean log<sub>2</sub> CV-ratios for primary and secondary metabolites were calculated independently and compared against a random mean CV-ratio. The random mean CV-ratio was generated by permutation of parent and hybrid labels. This process was repeated 10000 times and a mean CV-ratio was calculated each time to produce a random distribution of

mean CV-ratios (separately for primary and secondary metabolites). A shift in the observed mean CV-ratio with respect to the random mean CV-ratio was then assessed.

### Analysis of inheritance patterns

To identify the different inheritance patterns across hybrids, the deviations from the mid-parent values (MPVs) were calculated for each metabolite in every hybrid. The deviation of each hybrid for a given metabolite was obtained by subtracting the observed value from the predicted mid-parent value. To determine empirical *p*-values, the resulting deviations were then contrasted individually against a random distribution generated for each metabolite by resampling.

### Broad-sense and narrow-sense heritability

We used a variant of the linear mixed model used for analysis of diallel crossing (Mohring et al 2011) where the additive paternal and maternal effects, as well as the non-additive hybrid effect are considered as random effects (i.e., we do not obtain specific combining ability of each parent). The random effects follow the normal distribution with zero mean and respective variances. The model used was

$$y = X_m\beta_m + X_p\beta_p + Z\gamma + \varepsilon,$$

where, *y* is the metabolic profile for each cross, *X<sub>m</sub>* and *X<sub>p</sub>* are the design matrix for maternal and paternal effect, respectively,  $\beta_m$  and  $\beta_p$  are the maternal and paternal effect value, *Z* is the design matrix representing the non-additive effect and  $\gamma$  is the non-additive effect value, and  $\varepsilon$  is the residual error. The error and the random effects were estimated with the REML method (Mohring et al 2011). Using the variance components of this model, we determined the narrow sense heritability (corresponding to the percentage of additive variance from the phenotypic variance) and the broad sense heritability representing the percentage of total (additive and non-additive) genetic variance from the phenotypic variance.

### Diallel analysis

For epistasis, Hayman's diallel approach was implemented (Atnaf et al 2014; Hayman 1954). An array in Hayman's analysis corresponds to a group of crosses which involve one particular parent. Then, *W<sub>r</sub>* is the covariance of parents and their offspring in each array, while *V<sub>r</sub>* is the variance of means of an array. In absence of epistasis, the regression of *W<sub>r</sub>* with respect to *V<sub>r</sub>* gives a line of unit slope. As a result, the deviation from the slope of this regression line is an indication for existence of epistasis. Besides, we also used

**Table 1** Pairwise comparison of genetic similarity of the Alt1 to Alt7 genotypes

	Heterozygosity	Alt2	Alt3	Alt4	Alt5	Alt6	Alt7
Alt1	1.66%	66.50%	65.50%	72.70%	73.60%	97.50%	60.30%
Alt2	1.21%		68.80%	66.40%	67.00%	65.60%	68.80%
Alt3	2.37%			64.30%	67.50%	65.20%	67.50%
Alt4	1.01%				79.70%	72.60%	65.00%
Alt5	1.21%					73.20%	69.50%
Alt6	2.67%						60.30%
Alt7	5.39%						

Amount of heterozygosity in each parent is indicated. Altogether, 1985 SNPs were used in the genetic similarity analysis.

ANOVA on Wr-Vr to indicate the presence of epistasis for each metabolite.

## Results

### Analysed genotypes and phenotypes

Our experiment is based on a full diallel crossing of seven parental lines in two replicates, collected in 2007 from one location in Altenriet, Tübingen (Southern Germany). Altogether, we assessed 38 out of the 42 hybrids since the remaining four could not be phenotyped due to technical issues. The parents were genotyped using RAD sequencing (Swiadek et al 2017). Based on the 1985 informative markers (Materials and Methods), parents showed more than 95.6% homozygosity (Table 1). Pairwise comparison of the SNPs showed that similarity among the parents varied from 60.3 to 97.5%, with an average similarity of 69% (Table 1). Both the parents and hybrids were phenotyped for their metabolism and rosette size. For metabolic profiling, rosettes of at least four plants from the analysed hybrids and parents were harvested at the 10-leaf stage to avoid changes due to different developmental stages in the plants. Altogether, 100 metabolites, including 66 analytes from primary metabolism and 34 from secondary metabolism (Table S1A), were identified and quantified using gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) (Lisec et al 2006; Tohge and Fernie 2010 and Materials and Methods). The rosette radius was measured from photos taken at a 10-leaf stage.

### Similarity between genotype and metabolism

First, we investigated if more similar genotype (i.e., parents and hybrids) display more similar phenotypes. To answer this question, we created relationship matrices of both genotype and metabolic profiles for both hybrids and parents. Primary and secondary metabolic profiles were analysed separately.

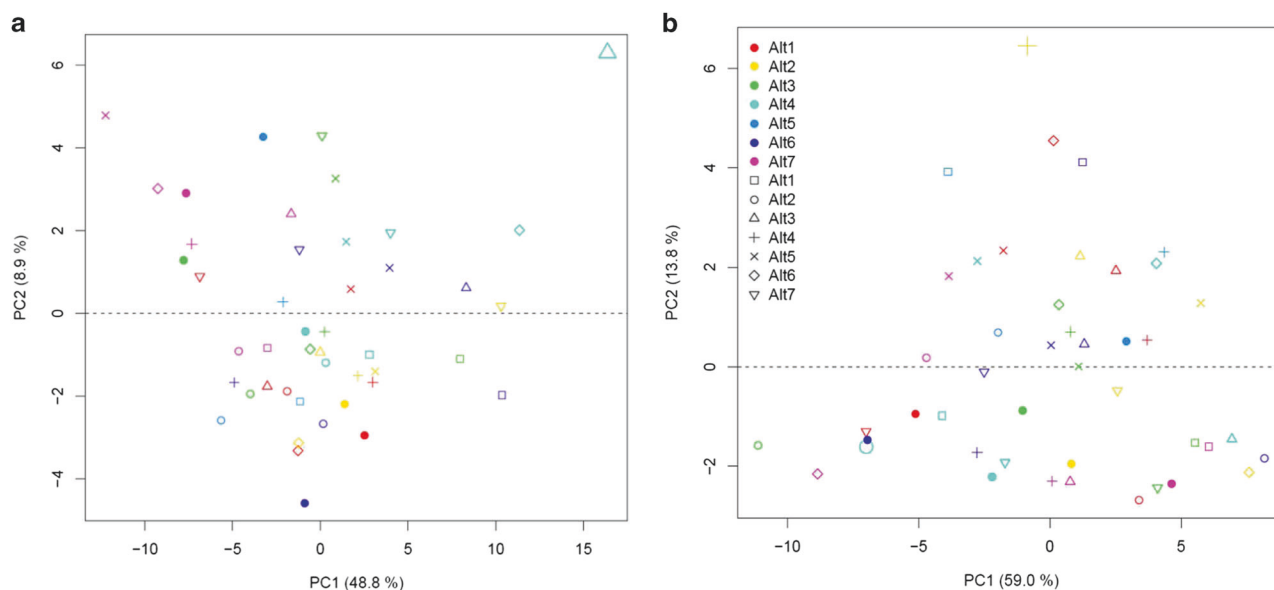
Based on the relationship matrix calculated by the correlations of informative SNPs in the parents, parents Alt1 and Alt6 were the closest, while parents Alt1 and Alt7 were the most genetically distant (Fig. S1A). This was in line with the genetic similarities of the parents (Table S2A). We also determined the relationship matrix based on the correlation of SNPs for the hybrids visualizing the genetic closeness of all genotypes used for analyses (Fig. S1D, Table S2B).

Similarity matrix of the primary metabolic profiles of parents indicated that this group of genotypes is highly similar over all pairs, with the smallest correlation value of 0.98 (Fig. S1B). For the secondary metabolic profiles, we identified that Alt4 was the most discrepant in comparison to the other parents (which exhibit mutually concordant secondary metabolic phenotypes) (Fig. S1C).

To assess if similar genotypes displayed similar primary and secondary metabolic profiles we determined the Mantel correlation coefficient between the similarity matrices of genotypes and metabolic profiles (see Fig. S1). The concordance between the similarity of genotypes and similarity of primary metabolic profiles was more pronounced for the parents (Mantel correlation of 0.75,  $p$ -value 0.04,  $n = 7 \times 7$ ) in comparison to the hybrids (Mantel correlation of 0.45,  $p$ -value 0.002,  $n = 38 \times 38$ ). In contrast, we did not find significant concordance between the similarity of the secondary metabolic phenotype and the genetic relationships of the parents, as well as the hybrids (Mantel correlations 0.67,  $p$ -value 0.09 and 0.3,  $p$ -value 0.05, respectively). This finding indicated that similar parents have similar primary metabolism, but that this observation does not extend to secondary metabolism.

### Variability of metabolic profiles of parents and hybrids

Next, we performed a principal component (PC) analysis with the two metabolic phenotypes in order to investigate the parental effects on hybrid metabolism. We found that the first two PCs captured a smaller percentage of the total variance of primary metabolites (PC1, 48.8% and PC2, 8.9%) in comparison to secondary metabolites (PC1, 59%



**Fig. 1** Principal component analysis (PCA) of the parents and hybrids. PCA was conducted based on **a** primary metabolic profiles and **b** secondary metabolic profiles. Hybrids are coded based on the genotype of the parents. Colors indicate the mother and the shapes indicate the father plant, included in the legend of panel **a**. Solid points present parents

and PC2, 13.8%). We used a permutation test (see Materials and Methods) with 100 repetitions and found that in the first two PCs in secondary metabolism capture significantly larger variance than the first two PCs in primary metabolism ( $p$ -value < 0.05).

PCA did not reveal clear separations of certain parents with their respective hybrids (Fig. 1). However, based on their secondary metabolism (Fig. 1b), the second PC located all parents to the lower half of the plot, and more than half of the hybrids were not grouped with the parents. A similar pattern was not identified for primary metabolism (Fig. 1a), where parents and hybrids were distributed throughout the projection space. In most cases, the reciprocal hybrids were plotted in close proximity to each other, indicating no clear parental effect. There were a few cases, all in secondary metabolism, where reciprocal hybrids showed differences, but only in the reciprocal hybrids between Alt2 and Alt4 parents. For these parents, the corresponding points of the reciprocal hybrids in Fig. 1b were the most distant in the plot with respect to the second PC. Additionally, it is notable that Alt2xAlt4 was separated by PC2 from the rest of the hybrids and its difference in secondary metabolism was already observed in the correlation analysis (Fig. S1F).

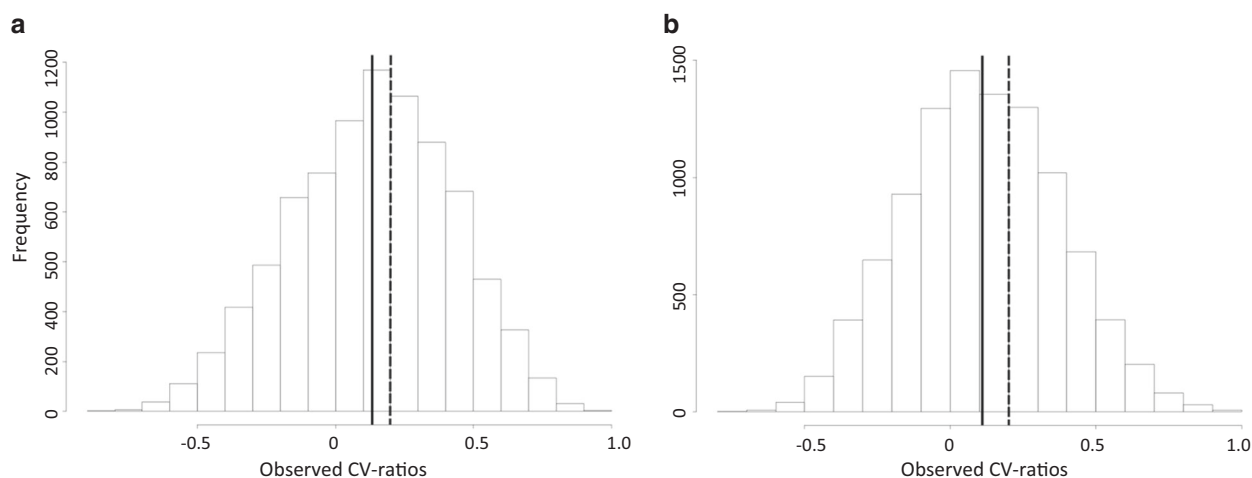
In this sense, parents with the most dissimilar secondary metabolism (Alt2 and Alt4) also gave rise to the most dissimilar hybrid in secondary metabolism (Alt2xAlt4), although this behavior was not seen in its reciprocal counterpart (Alt4xAlt2). This could be due to a strong parental effect in the specific combination of having Alt2 as a mother and Alt4 as a father; however, we did not observe similar effect in other crosses made with Alt2 or Alt4 as

parent. Further evidence is needed to find out the cause for the specificity of the observed parental effect in Alt2xAlt4 hybrids. Interestingly, the most divergent hybrid in primary metabolism was Alt4xAlt3 rather than the hybrid from the most divergent parents, suggesting a considerable contribution from non-additivity (Fig. 1a). In addition based on the primary metabolic phenotypes, the PC plot allowed us to identify groups of hybrids that grouped based on the closeness of the respective maternal genotypes (proximity of shapes with same color, Fig. 1a).

### Metabolic phenotypes of hybrids vary more than those of the parents

Although the PC analyses provided insights into the metabolic variation of the different hybrids in relation to the parental genotypes, it did not tell us about the overall metabolic variation between the groups of hybrid and parental genotypes. To address this question, we next calculated the log-transformed ratio of the coefficients of variation (CV) for the level of every compound of the two groups of genotypes. This analysis normalizes the effect of variability due to the differences in metabolite levels (Lisek et al 2011).

Both primary and secondary metabolism showed a positive non-significant shift in the CV ratio from the empirical mean of a permutation test, indicating the similarity in robustness of the metabolism of parents and hybrids (Fig. 2a–b). For completeness, we also provide graphical information about the CV of secondary and primary metabolites in hybrids and parents (Fig. S2).



**Fig. 2** Analysis of the variation in hybrids in comparison to the parents. **a** The coefficient of variation (CV) ratio between hybrids and the parents for primary metabolites from permutation test. **b** The

coefficient of variation (CV) ratio between hybrids and the parents for secondary metabolites from permutation test. Black solid line indicates random mean and dashed line indicates observed mean CV ratio

### The metabolic traits in hybrids differ in the deviation from the mid-parent value

Motivated by the observed differences between hybrids and parents based on the metabolic profiles, we next wanted to find out the sources of the higher variability and thus investigated the inheritance patterns for each metabolic compound separately over all hybrids. The distribution of deviations from the mid-parent value (MPV) in the investigated crosses, which indicate the amount of non-additive inheritance across all hybrids, showed that the compounds from secondary metabolism, comprising glucosinolates, phenylpropanoids, a flavonoid and several unknown analytes, exhibited the largest median deviation from the respective MPV (Fig. 3a). While glucosinolates showed both positive and negative median deviations from the MPV, the two detected phenylpropanoids exhibited only a negative median deviation (Fig. 3a). For the compound classes in primary metabolism, the median deviations from the MPV were negative and considerably smaller than those for the majority of secondary metabolites. We also found that the variability of deviations from MPV was larger for compounds from secondary compared to those from primary metabolism. On inspecting the significant deviations from MPV at level 0.05, we found that for almost all metabolites the respective median deviations were negative (Fig. S3). This indicates that some of the analysed hybrids showed particularly different levels of secondary metabolites in comparison to the average of the parents, and strongly suggests that non-additive inheritance contributes in shaping secondary metabolism.

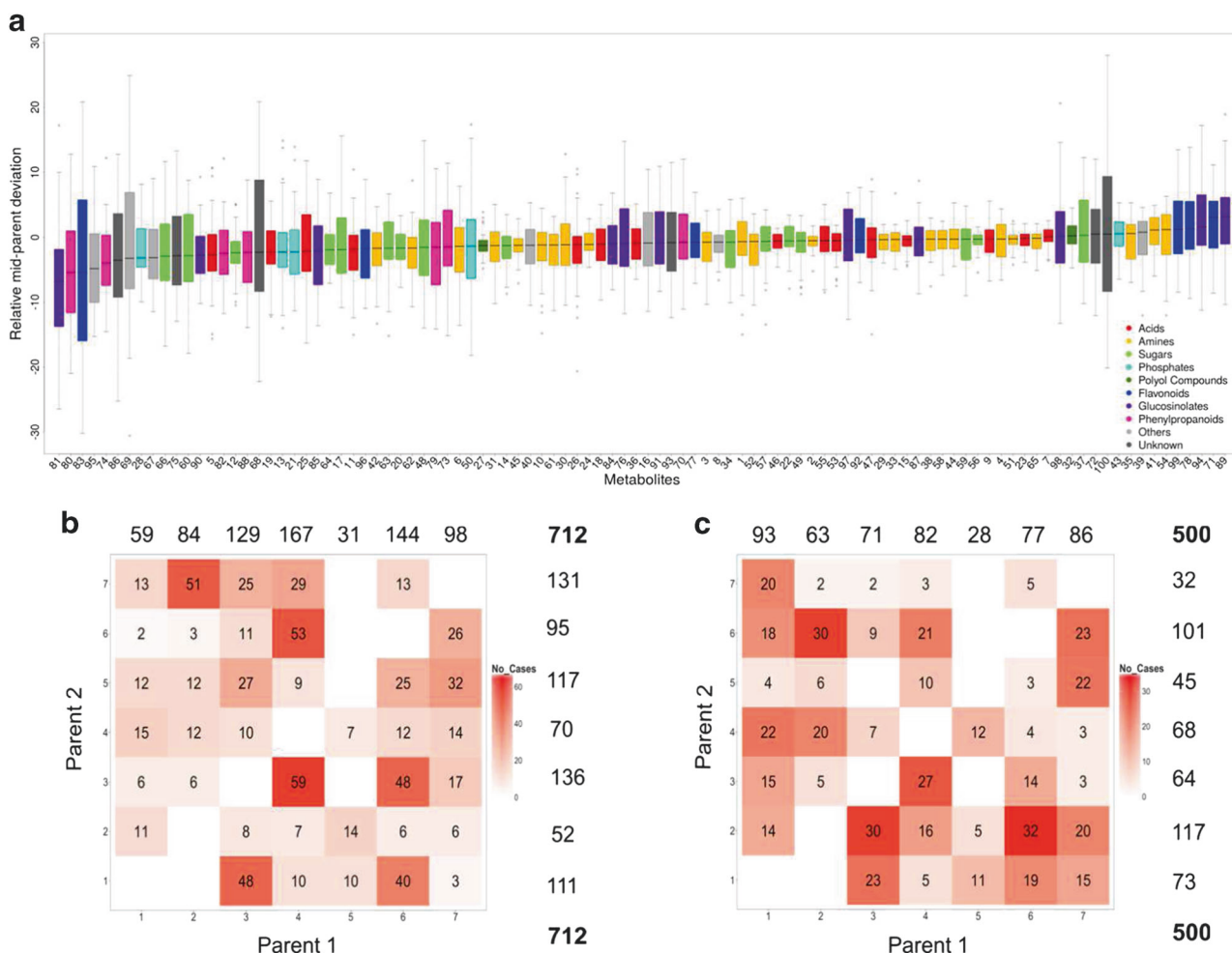
We next wanted to find out if certain crosses exhibited a higher degree of non-additive inheritance in metabolism

than others. Six crosses showed non-additive inheritance in more than 50% (i.e., 33) of the primary metabolites (Fig. 3b, Table S3B) while at least 50% (i.e., 17) of secondary metabolites showed non-additive inheritance in 14 crosses (Fig. 3c, Table S3B). From these, only two were reciprocal (Alt6xAlt1 and Alt6xAlt2), indicating that here the extent of non-additive inheritance depended on the direction of the cross (Fig. 3c, Table S3B). In addition, the crosses with Alt2 and Alt6 as a father were, on average, involved in more than 40% of the non-additive inheritance patterns for secondary metabolites (Fig. 3c, Table S3B) while seven secondary metabolites (i.e., coniferin, *trans*-sinapoyl malate, phenylpropanoid sinapoyl malate, a structurally undefined aliphatic glucosinolate, a structurally undefined benzenoid, a structurally undefined flavonoid and an unknown compound) were significantly inherited in a non-additive way across all hybrid crosses (Table S3B).

Altogether, we could not identify a particular pattern for the sign of deviations from MPV for the examined compound classes. Nevertheless, the levels of compounds from secondary metabolism showed predominantly non-additive inheritance patterns in comparison to primary metabolites. Moreover, some crosses were more likely to show non-additive inheritance than others. These findings were in line with the observed higher variance in secondary metabolism when compared to primary metabolism.

### Non-additive variance is a determinant of secondary metabolism

Motivated by the identified presence of non-additive inheritance, we next analysed the contribution of non-additive variance to the variation of metabolic traits in a



**Fig. 3** Non-additive inheritance patterns for metabolites in hybrids. **a** Relative deviations of all measured metabolites from the mid-parent value. The indices correspond to the metabolite annotation provided in Table S1A and the metabolites are ordered according to their relative deviations from negative to positive. **b** The number of primary metabolites inherited in non-additive way in different hybrids; **c** The number of secondary metabolites inherited in non-additive way in

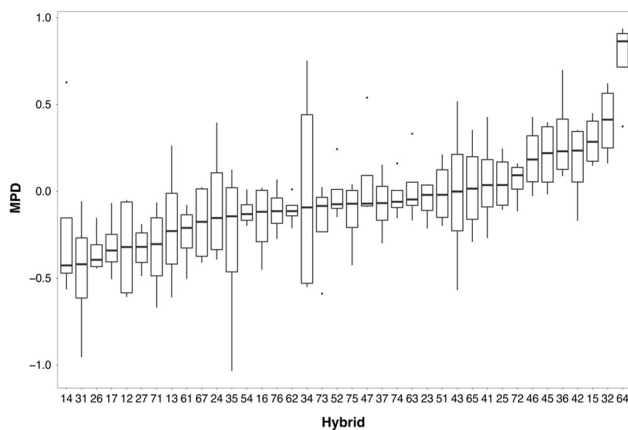
different hybrids. In **b** and **c** non-additive inheritance was based on significant deviation from the parental mean at level  $\alpha=0.01$ . The total number of analysed primary metabolites was 66 (numbered 1–66) and the total number of secondary metabolites was 33 (numbered 67–100). The four first groups (i.e. acids, amines, sugars, phosphates and polyol compounds) belong to primary metabolism and flavonoids, glucosinolates and phenylpropanoids to secondary metabolism

greater detail. We used the quantitative genetics analysis of full diallel designs to partition the genetic variance into additive and non-additive variance. We calculated narrow sense heritability and broad sense heritability to find out the additive and non-additive variance that the different hybrids show in the metabolites analysed here (see Materials and Methods).

The narrow sense heritability for the primary metabolites was not larger than 50.8% with a mean of 14.3% over all primary metabolites (Table S4A). For 30 of the primary metabolites we found that the broad and narrow sense heritabilities coincided, indicating the genetic variance for these metabolic traits was only additive. In addition, from the primary metabolites, only 2-oxoglutaric acid showed more than 10% of non-additive variance.

The narrow sense heritability for the secondary metabolites was smaller than the primary metabolites, with the largest value of 48% and a mean of 4.8% (Table S4B). However, the narrow sense heritability for only one secondary metabolite coincided with the value of the broad sense heritability. The non-additive variance was larger than 10% of their phenotypic variance for 19 metabolites, with aliphatic glucosinolate exhibiting maximum non-additive variance (27.1%).

Our results indicated that secondary metabolites showed more non-additive inheritance than primary metabolites in line with the observed larger deviation from MPV in secondary metabolites than primary metabolites. However, we also found compounds in both metabolic profiles for which non-additive variance contributed a portion larger than 10% of genetic variance.



**Fig. 4** Non-additive inheritance patterns in the final rosette radius of the hybrids. Deviations of observed final rosette radius (cm) from the mid-parent value (MPD) in each hybrid are presented. The hybrids are ordered according to their median MPD, from negative dominant to positive dominant. The two-digit number for each hybrid appearing on the *x*-axis denotes the combination of the respective parents (e.g., hybrid 14 denotes the Alt1xAlt4 cross)

### Contribution of dominance and epistasis in non-additive variance

The non-additive genetic variance can be further divided into dominance and interaction (epistatic) variance (Falconer and Mackay 1996). Our results suggested, that non-additive genetic variance plays a role in shaping secondary metabolism, with some isolated effects on metabolites from primary metabolic pathways. As a next step, we asked if the non-additive inheritance is due to dominance or to epistasis. To dissect the contribution of epistasis for specific metabolic traits, we used Hayman's diallel analysis and ANOVA on  $W_r$ - $V_r$  (see Materials and Methods).

For the primary metabolic phenotype, we found significant differences at level of 0.05 between  $W_r$  and  $V_r$  for four metabolites, namely raffinose, GABA, phosphoric acid and urea (Table S5A). With the exception of urea, these metabolites showed only additive variance (see section above) and, therefore, there was no contribution of epistasis for these traits.

For the secondary metabolic phenotype, we found significant differences at level of 0.05 between  $W_r$  and  $V_r$  for three metabolites, namely, aliphatic glucosinolate, neoglucobrassicin and kaempferol 3-galactoside-7-rhamnoside (two with non-additive variance larger than 10%) (Table S5B). Therefore, our findings suggested that the larger non-additive variance in secondary metabolism may be attributed to epistasis rather than dominance effects.

### Relationship between growth and metabolism

Studies on plant metabolism and growth-related traits have revealed a tight relationship especially between primary

metabolism and biomass (Stitt 2013; Sulpice et al 2009). To find out if in our case the changes in metabolism are reflected to changes in growth-related traits, we measured rosette radius at the 10-leaf stage of in both hybrids and parents. These measurements showed that rosette radii had both positive and negative non-additive inheritance in comparison to the mid-parent value (Fig. 4). From the 12 hybrids with a significant deviation from the MPV, seven had larger and five smaller rosette radii than the mid-parent value indicating that the parents contribute to hybrid vigor (i.e., heterosis) and hybrid incompatibility (Fig. 4). From all hybrids, Alt6xAlt4 showed the highest positive deviation from the MPV, and its reciprocal counterpart also followed a similar behavior with a significant positive deviation. Alt3xAlt1 and its reciprocal hybrid had the two highest significant negative deviations in rosette radius (Fig. 4). However, many of the reciprocal hybrids did not show similar inheritance patterns, indicating parental effects. For example, Alt2xAlt3 and Alt5xAlt1 did not show significant deviations from the MPV although their reciprocal hybrids, Alt3xAlt2 and Alt1xAlt5, were among the three with the largest significant positive deviation from the MPV.

We also investigated the relationship between the metabolism and growth by looking at the correlation of primary and secondary metabolites with the final size measured as rosette radius. We found 23 metabolites significantly correlated (at level  $\alpha = 0.05$ ) with the final rosette radius (Table S6), with all positive correlations for secondary metabolites. Glucosinolates were the predominant group of secondary metabolites correlated with the analysed proxy for growth. In addition, only two primary metabolite analytes, namely, 1,6-anhydrobetaglucose (a hydroxylated form of glucose which forms on the pyrolysis of cellulose and as such can be regarded as a proxy for cellulose content (Sasaki et al 2008), and spermidine were positively correlated to the rosette radius.

### Discussion

Here, we presented a full diallel analysis of plant metabolism and growth-related traits among seven *A. thaliana* parents collected from a single growth habitat. These parents were mostly inbred with heterozygosity ranging from 1–5.4%. A recent study highlighted the power of using the  $F_1$  hybrids generated by a diallel crossing scheme to investigate the non-additive inheritance patterns underlying hybrid vigor in *A. thaliana* inbred lines (Seymour et al 2016). However, less is known about the role of heterozygosity in the variation of plant metabolism in nature. In order to unravel the inherited differences in metabolism, we grew the analysed hybrids and parents in constant growth conditions.



We showed that primary and secondary metabolite abundances, as well as growth, exhibited on average higher fluctuations in hybrids compared to the parents. This finding shows an interesting contrast with that from a similar experimental design with six maize inbred lines, in which the resulting heterotic hybrids (in terms of biomass) showed, on average, a more robust primary metabolism than parents (Liseć et al 2011). In our data set, we have a mixture of hybrids which showed positive and negative inheritance patterns in respect to size. Therefore, our finding that the hybrids and parents show similar robustness can serve as a support of the principle underlying heterosis proposed by Liseć et al. (2011), whereby hybrids exhibiting heterosis tend to show similar metabolic profiles associated to better growth. Our results suggest that larger variation due to heterozygosity in individuals adapted to a single growth habitat results in both positive and negative dominance.

Secondary metabolites varied more than primary metabolites and contributed more to the non-additive inheritance patterns. Interestingly, Hayman's analysis revealed epistasis in hybrids for some of the studied traits. The secondary metabolites showing a significant non-additive inheritance across all hybrids are known to be involved in plant stress responses suggesting that the increased genetic and phenotypic variation due to heterozygosity could be beneficial for homozygous plant populations to rapidly adapt to fluctuating environments. Fewer non-additive inheritance in primary metabolites, and hence variability, across individuals go hand in hand with their essential role in normal growth and development, since drastic modifications in their levels might hinder survival.

We demonstrated here that a diallel design can be used to assess the heritability of traits. That heritabilities of primary and secondary metabolism differ has been reported previously in a study of introgression lines of tomato (Alseekh et al 2015; Schauer et al 2008); however, in these cases, primary metabolites displayed a greater degree of non-additive inheritance than secondary metabolites. Yet, heritability of secondary metabolites was considerably lower in both species; reinforcing the fact that secondary metabolism is subject to greater environmental fluctuation than primary metabolism (Schillmiller et al 2012). Interestingly, we also observed that in some cases the same genetic background yielded different metabolic phenotypes in reciprocal hybrids. It is notable, that those reciprocal hybrids that showed the most different metabolic profiles included the four biggest hybrids of this study; however, additional experiments are needed to find out if the heterosis in terms of plant size plays a role in causing stronger parental effect.

Not only were we able to infer important information concerning the heritability of metabolic phenotypes but also to determine metabolites that were strongly associated with

plant size. Amongst these we found a likely marker for cellulose, which according to both textbook physiology and a recent gene-metabolite-growth analysis in maize, demarcates the cell wall during growth (Riedelsheimer et al 2012b), and the polyol spermidine whose levels have similarly often been shown to correlate with plant growth (Paschalidis and Roubelakis-Angelakis 2005). In addition, glucosinolates, secondary metabolites known to be associated with plant stress, were positively correlated with the size of the plant. Glucosinolates are compounds known to respond to both biotic and abiotic stress in plants, also influencing the size of the plant (Chan et al 2010; Halkier and Gershenzon 2006). However, several recent studies have suggested that the trade-off between defense and growth which is illustrated by the production of these compounds is grossly oversimplified (Joseph et al 2013; Kliebenstein 2016; Mauricio 1998) and this may explain the fact that we observed a positive relationship between glucosinolate levels and plant size which was in the opposite direction to what we initially anticipated. However, it is important to note that such a positive relation has previously been observed in the field and the simple model of trade-off between investing in glucosinolate production for defense or investing in growth is thus unlikely to hold under all conditions (Joseph et al 2013; Kliebenstein 2016; Mauricio 1998). These results, furthermore, add intriguing insights to our understanding of the nature of non-additive inheritance in a natural *A. thaliana* population, which may also hold in major crop species.

To conclude, we found that hybrids of *A. thaliana* individuals coming from the same collection site showed an increased metabolic variability when compared to their parents at controlled growth conditions. The greater variation in hybrids than in parents goes in contrast to previous studies on hybrids among genetically more diverse parents in crop plants. In nature, *A. thaliana* plants are subjected to many environmental fluctuations and the increased variation due to heterozygosity provides a source for genetic and phenotypic variation under changing conditions in a specific growth habitat. While the parental effects were visible in both primary and secondary metabolism, the pronounced non-additive inheritance in some secondary metabolites could be statistically attributed to epistasis. Nevertheless, the proportion of the contribution of epistasis in non-additive inheritance in these metabolites is not known. Therefore, these findings provide insights for plant breeding by adding relevant information on how heterozygosity influences metabolism and growth in a single growth habitat of *A. thaliana*.

**Acknowledgements** We thank Karin Köhl and the Green Team of the Max Planck Institute of Molecular Plant Physiology for taking excellent care of plants. Aditya Sharma, Christian Jorzig and

Magdalena Świadek are thanked for assistance in initial screening of the hybrids. This work was supported by the Max Planck Society (R. A.E.L.).

**Author contributions** R.A.E.L. generated the diallel crosses, A.E.R.C. and H.T. analysed the data with contributions from F. A. and J.Y., S. A. did metabolic profiling, A.R.F., Z.N., and R.A.E.L. supervised the experiments, R.A.E.L. and Z.N. planned the experiments and conceived the original research plans. R.A.E.L., Z.N., A.E.R.C., drafted the manuscript, and H.T. and A.R.F. made a substantial contribution to editing of the manuscript; all authors approved the final version of the manuscript.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no competing interests.

## References

- Abbott RJ, Gomes MF (1989) Population genetic-structure and out-crossing rate of *Arabidopsis thaliana* (L.) heynh. *Heredity* 62:411–418
- Alseekh S, Tohge T, Wendenberg R, Scossa F, Omranian N, Li J et al. (2015) Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* 27(3):485–512
- Atnaf M, Tesfaye K, Dagne K, Mohammed H (2014) Hyman's diallel analysis to study genetic parameters of phenological traits in common bean (*Phaseolus vulgaris*). *Int J Agric Sci Nat Resour* 1(4):65–71
- Bestfleisch M, Hofer M, Hanke MV, Flachowsky H, Richter K, Schulte E (2013) Breeding of resistant strawberry cultivars for organic fruit production - preliminary results with *Botrytis cinerea*. *Acta Hort* 976:87–90
- Boer MP, Wright D, Feng LZ, Podlich DW, Luo L, Cooper M et al. (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics* 177(3):1801–1813
- Bombles K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N et al. (2010). Local-scale patterns of genetic variability, out-crossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *Plos Genet* 6(3):e1000890
- Chan EKF, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185(3):991–1007
- Charlesworth D, Willis JH (2009) Fundamental concepts in genetics The genetics of inbreeding depression. *Nat Rev Genet* 10(11):783–796
- Davenport CB (1908) Degeneration, albinism and inbreeding. *Science* 28:454–455
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Addison Wesley Longman, Harlow, UK
- Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* 57:303–333
- Hayman BI (1954) The theory and analysis of diallel crosses. *Genetics* 39(6):789–809
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuwara M et al. (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 101(27):10205–10210
- Hull FH (1945) Recurrent selection for specific combining ability in corn. *J Am Soc Agron* 37(2):134–145
- Joseph B, Corwin JA, Zust T, Li BH, Iravani M, Schaeppman-Strub G et al. (2013) Hierarchical nuclear and cytoplasmic genetic architectures for plant growth and defense within *Arabidopsis*. *Plant Cell* 25(6):1929–1945
- Kliebenstein DJ (2016) False idolatry of the mythical growth versus immunity tradeoff in molecular systems plant pathology. *Physiol Mol Plant P* 95:55–59
- Lisec J, Romisch-Margl L, Nikoloski Z, Piepho HP, Giavalisco P, Selbig J et al. (2011) Corn hybrids display lower metabolite variability and complex metabolite inheritance patterns. *Plant J* 68(2):326–336
- Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 1(1):387–396
- Mauricio R (1998) Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *Am Nat* 151(1):20–28
- Meyer RC, Torjek O, Becher M, Altmann T (2004) Heterosis of biomass production in arabidopsis. establishment during early development. *Plant Physiol* 134(4):1813–1823
- Mohring J, Melchinger AE, Piepho HP (2011) REML-based diallel analysis. *Crop Sci* 51(2):470–478
- Obata T, Fernie AR (2012) The use of metabolomics to dissect plant responses to abiotic stresses. *Cell Mol Life Sci* 69(19):3225–3243
- Parker D, Beckmann M, Zubair H, Enot DP, Caracul-Rios Z, Overy DP et al. (2009) Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by *Magnaporthe grisea*. *Plant J* 59(5):723–737
- Paschalidis KA, Roubelakis-Angelakis KA (2005) Spatial and temporal distribution of polyamine levels and polyamine anabolism in different organs/tissues of the tobacco plant. Correlations with age, cell division/expansion, and differentiation. *Plant Physiol* 138(1):142–152
- Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW et al. (2010) The Scale of Population Structure in *Arabidopsis thaliana*. *Plos Genet* 6:2
- Powers L (1944) An expansion of Jones' theory for the explanation of heterosis. *Am Nat* 78:275–280
- Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R et al (2012a) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet* 44(2):217–220
- Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C et al (2012b) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc Natl Acad Sci USA* 109(23):8872–8877
- Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L et al (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13(1):11–29
- Sasaki M, Takahashi K, Haneda Y, Satoh H, Sasaki A, Narumi A et al (2008) Thermochemical transformation of glucose to 1,6-anhydroglucose in high-temperature steam. *Carbohydr Res* 343(5):848–854
- Schauer N, Semel Y, Balbo I, Steinfath M, Reipsilber D, Selbig J et al. (2008) Mode of inheritance of primary metabolic traits in tomato. *Plant Cell* 20(3):509–523
- Schauer N, Steinhäuser D, Strelkov S, Schomburg D, Allison G, Moritz T et al (2005) GC-MS libraries for the rapid identification of metabolites in complex biological samples. *Febs Lett* 579(6):1332–1337
- Schillmiller AL, Pichersky F, Last RL (2012) Taming the hydra of specialized metabolism: how systems biology and comparative

- approaches are revolutionizing plant biochemistry. *Curr Opin Plant Biol* 15(3):338–344
- Schwartz D, Laughner WJ (1969) A molecular basis for heterosis. *Science* 166(3905):626. -&
- Seymour DK, Chae E, Grimm DG, Pizarro CM, Habring-Muller A, Vasseur F et al. (2016) Genetic architecture of nonadditive inheritance in *Arabidopsis thaliana* hybrids. *Proc Natl Acad Sci USA* 113(46):E7317–E7326
- Stekhoven DJ, Buhlmann P (2012) MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28(1):112–118
- Stitt M (2013) Systems-integration of plant metabolism: means, motive and opportunity. *Curr Opin Plant Biol* 16(3):381–388
- Sulpice R, Pyl ET, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H et al (2009) Starch as a major integrator in the regulation of plant growth. *Proc Natl Acad Sci USA* 106(25):10348–10353
- Swiadek M, Proost S, Sieh D, Yu J, Todesco M, Jorzig C et al (2017) Novel allelic variants in ACD6 cause hybrid necrosis in local collection of *Arabidopsis thaliana*. *New Phytol* 213(2):900–915
- Tohge T, Fernie AR (2010) Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat Protoc* 5(6):1210–1227
- Weigel D (2012) Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol* 158(1):2–22