Pervasive Under-Dominance in Gene Expression as Unifying Principle of Biomass Heterosis in *Arabidopsis*

Wei Yuan, Fiona Beitel, Thanvi Srikant, Ilja Bezrukov, Sabine Schäfer, Robin Kraft, Detlef Weigel*

Department of Molecular Biology, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany

*Corresponding author. Email: weigel@tue.mpg.de

Heterosis, the generally superior performance in hybrids compared to their inbred parents, is one of the most enigmatic biological phenomena. Many different explanations have been put forward for heterosis, which begs the question whether common principles underpinning it do exist at all. We performed a systematic transcriptomic study in *Arabidopsis thaliana* involving 141 random crosses, to search for the general principles, if any, that heterotic hybrids share. Consistent additive expression in F_1 hybrids was observed for only about 300 genes enriched for roles in stress response and cell death. Regulatory rare-allele burden affects the expression level of these genes but does not correlate with heterosis. Non-additive gene expression in F_1 hybrids is much more common, with the vast majority of genes (over 90%) being expressed below parental average. These include genes that are quantitatively correlated with biomass accumulation in both parents and F_1 hybrids, as well as genes strongly associated with heterosis. Unlike in the additive genes, regulatory rare allele burden in this non-additive gene set is strongly correlated with growth heterosis, even though it does not covary with the expression level of these genes. Together, our study suggests that while additive complementation is an intrinsic property of F_1 hybrids, the major driver of growth in hybrids derives from the quantitative nature of non-additive gene expression, especially under-dominance and thus lower expression in hybrids than predicted from the parents.

Introduction

Heterosis, or hybrid vigor, describes the phenomenon that F_1 hybrids show superior growth, stress resistance or fertility compared to their inbred parents (Hochholdinger and Baldauf, 2018). Defined as mid- or best-parent heterosis, where hybrids are either better than the parental average or even superior to both parents, it was first described by Darwin (Darwin, 1876), with the term heterosis later coined by East and Shull (East, 1908; Shull, 1912). It remains one of the most fascinating phenomena in genetics, even though there is still no unifying theory for heterosis. One recurring question is whether best-parent heterosis, which is of particular interest to breeders, is an emergent property resulting from mid-parent heterosis in component traits (Griffing, 1990). Results from different species have often led to very different explanations.

Despite extensive exploitation in breeding and agriculture, we are only beginning to understand the molecular basis and mechanisms behind heterosis, and results from different species have often led to very different explanations. In tomato and rice, there is good evidence for single genes, sometimes with overdominant effects, making major contributions to heterosis (Huang et al., 2016; Krieger et al., 2010; Xue et al., 2008), while heterosis in maize appears to result from reciprocal complementation at a very large number of loci (Paschold et al., 2012; Kremling et al., 2018), consistent with heterosis positively correlating with genome-wide inter-parental genetic distance (Hochholdinger and Baldauf, 2018). There must, however, be a role for overdominance in maize, since the continuous purging of deleterious variants in inbred parents through breeding has not decreased heterosis in F_1 hybrids (Birchler et al., 2003).

A key area of debate for the study of heterosis is whether true heterozygote advantage exist by means of overdominance at a few heterozygous loci, or whether it is merely phenotypic emergence resulting from additive complementation of numerous mildly unfavorable recessive alleles (Lippman and Zamir, 2007). The search for overdominant loci has proven to be exceedingly difficult, due to the genetic load of (domesticated) species as well as limitations in quantitative genetics mapping approaches (Mitchell-Olds, 1995). Species that are naturally inbred, such as *Arabidopsis thaliana*, provide an alternative for the study of heterosis, because repeated selfing purges strongly deleterious variants (Oakley et al., 2019). That heterosis, which is common in *A. thaliana* intraspecific hybrids, shows little correlation with inter-parental genetic distance (Yang et al., 2017), suggests overdominance as an important factor for heterosis in this species. Studies in *A. thaliana* have already provided important insights into the underlying mechanisms, such as mitigation of defense-growth tradeoffs in superior hybrids (Miller et al., 2015). Superior performance of hybrids when it comes to such tradeoffs is, however, by no means ubiquitous in *A. thaliana*, with prominent exceptions where F_1 hybrids are greatly compromised in growth because of excessive activation of defense (Chae et al., 2014).

Given the diversity of proposed mechanisms for heterosis, systematic comparisons of a large number of crosses can help to find general principles. To this end, 6,925 knockout mutants in *S. cerevisiae* were crossed to a single wild-type strain of a close relative (Herbst et al., 2017). In *A.*

thaliana, about 200 natural accessions were crossed to a single lab-adapted accession (Yang et al., 2017), while another study in this species investigated heterosis in a 30 x 30 diallel cross (Seymour et al., 2016). In rice, over 10,000 F_2 individuals from 17 F_1s were analyzed to map the genes responsible for heterosis (Huang et al., 2016). While each of the studies came to different conclusions, they aptly demonstrate the power of a systematic design. Combining the strengths of previous approaches, we designed a study in *A. thaliana* that surveyed not only a broad range of the species' genetic diversity, but also allowed for the detection of interactions between an exceptionally large number of alleles. We find non-additivity in gene expression in F_1 hybrids to be common, with non-additive genes being much more commonly expressed below the mid-parental value than above it. Expression close to the mid-parental value in turn is rare, with a substantial fraction of such genes having a role in biotic defense pathways, suggesting that defense is particularly well buffered.

Results

Biomass heterosis is tunable in response to defense induction

For our work, we drew on resources from the 1001 Genomes Project for this species (1001 Genomes Consortium, 2016), crossing resequenced, naturally inbred accessions to generate a panel of F_1 hybrids. To broadly survey the prevalence of heterosis in *A. thaliana*, and to evaluate whether consistent patterns of additive and non-additive gene expression exist, we carried out a first experiment ("SHB1" in Fig. 1A, Methods) in which we measured whole-rosettes of 82 parent-F_1 trios, i.e., each F_1 hybrid and their inbred parents. As described in Methods, rosette size was confirmed to be a good proxy of biomass. We evaluated the robustness of heterosis in an altered environment in a second experiment ("SHB2", Methods), in which we applied BTH (acibenzolar-S-methyl), an analog of the defense hormone salicylic acid (SA) ("SHB2" in Fig. 1B, Methods), to 40 parent-F_1 trios, because it has been postulated that heterosis comes with the cost of defense activation due to the shifted balance in defense-growth trade-offs (Groszmann et al., 2015; Miller et al., 2015).

Heterosis was widespread, with F_1s (107.1±66.1 mm^2, n=82) being on average considerably larger than the parents (67.2±42.5 mm^2, n=124, Fig. 1C). Neither the distribution of the rosettes of the parents nor those of the F_1 plants were normal, with the F_1 population having significantly more large individuals (Two-tailed Kolmogorov-Smirnov, $p=2x10^{-5}$). In a comparison of randomly chosen F_1s and inbreds, the F_1 hybrid was twice as likely than the inbred to be the larger individual (Cliff's delta=0.33). After BTH treatment, considerable variation in responses were observed (Fig. 1 G-I, Fig. S3), with a stronger size reduction in F$_1$s (mock: 79.2±3.0 mm^2, BTH: 39.5±4.2 mm^2, One-way ANOVA, $p<2x10^{-16}$, Fig. 1D-F, Fig. S4), but still more larger hybrids than inbreds (Two-tailed Kolmogorov-Smirnov, p=0.0002, Cliff's delta=0.32). Most trios showed similar patterns after both mock- and BTH-treatment (Fig. 1E), with the majority of F$_1$s remaining larger than the MPV.

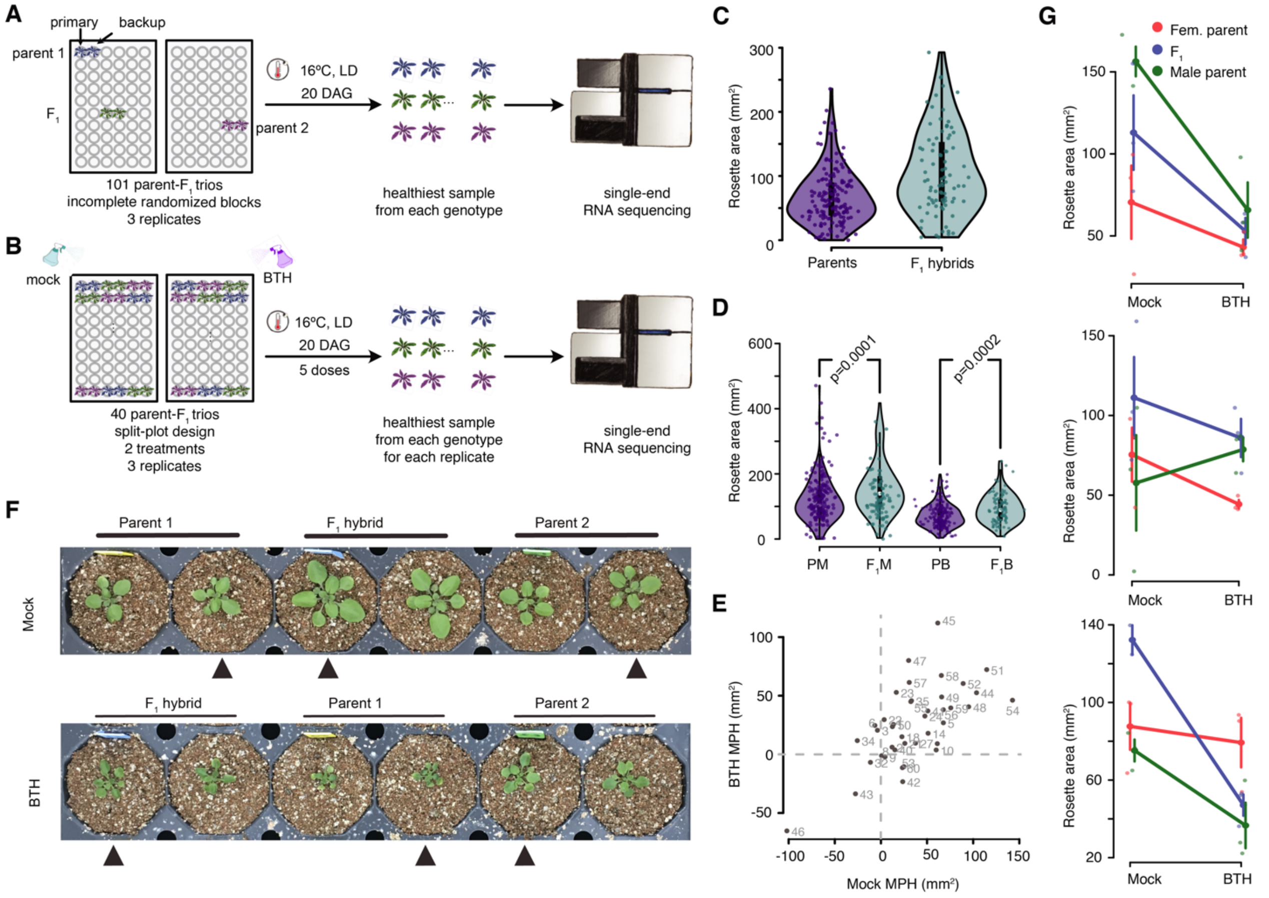


Fig. 1. Experimental design and growth measurements. (A,B) Schematic view of SHB1 and SHB2 experiments. Note that not all trios could be converted into complete measurements. **(C,D)** Final rosette area distribution of sequenced SHB1 (C) and SHB2 (D) samples. PM: parent mock, F$_1$M: F$_1$ hybrid mock, PB: parent BTH treated, F$_1$B: F$_1$BTH treated. **(E)** Heterotic F$_1$s under mock condition mostly remained heterotic after BTH treatment and *vice versa*. Numbered labels indicate the ID of the SHB2 trios. **(F)** Typical rosette phenotype of a trio. **(G-I)** Diverse response of three example trios to BTH treatment. Reaction norm lines connect the mean±SD rosette area of each genotype under both treatments.

If heterosis was the sole result of reciprocal complementation, we would expect a positive correlation between inter-parental genetic distance with heterosis, which has been generally observed in domesticated species (Fujimoto et al., 2018; Hochholdinger and Baldauf, 2018), and to a lesser degree in natural species including *A. thaliana* (Oakley et al., 2019; Yang et al., 2017). In our panel, whole-genome genetic differences (Hamming distance) between parental accessions showed a very modest positive correlation with the degree of rosette size mid-parent heterosis

(MPH) of the F_1s (Fig. 2E, Pearson correlation coefficient R=0.32 and p=0.08 for mock, and R=0.25 and p=0.16 for BTH,), consistent with the notion that additive complementation cannot fully account for heterosis observed. These results not only validate the usefulness of our panel for detailed transcriptome studies, but also indicate that heterosis remains robust in the face of environmental perturbation.

Rare-allele burden affects additively expressed genes but has no effect on F_1 biomass

Altered gene expression is a major component of heterosis, and it has been postulated that exceptionally low or high expression of genes is mostly deleterious (Kremling et al., 2018). Since most deleterious alleles are rare, it follows that gene expression in F_1s will typically not be as extreme as in inbred parents due to complementation. Such additive expression in turn could provide an explanation for the generally superior performance of F_1s. Given that it has been difficult in the past to identify unifying mechanisms for heterosis, we asked whether there are genes that are almost always additively expressed in F_1s across all trios. There were close to 900 consistently additively expressed genes, 300 of which were found in both experiments (Fig. 2A, Data S3, Methods). For these 300 genes (hereafter, additive genes), the relationship between expression MPV and observed expression in F_1s was individually examined (Fig. 2B, Fig. S5).

GO-term analysis of these additive genes revealed enrichment in cell death- and stress response-related processes (Fig. 2D), suggesting that these pathways are systematically buffered in F_1s. As expected, the additively expressed genes from the second experiment were more enriched for defense response than the genes from the first experiment without BTH treatment (Fig. S6). To learn whether common sequence polymorphisms within the additive genes are more likely to affect growth than the genome background, we examined the association between inter-parental genetic distance at these genes and heterosis. Correlation between the two was even weaker when only this subset rather than genome-wide polymorphisms was used (Fig. 2F, R=0.23, p=0.21 for mock, and R=0.18, p=0.32 for BTH). Therefore, inter-parental genetic distance was not a major explanatory factor for heterosis, in agreement with simple reciprocal genetic complementation explaining only some of the observed heterosis (Birchler et al., 2003).

Next, we looked into the relationship between gene expression level and the average number of regulatory rare SNPs, which are more likely to have a deleterious effect on gene expression than common SNPs, within 1 kb upstream of the additive genes, in both inbred parents and F_1s

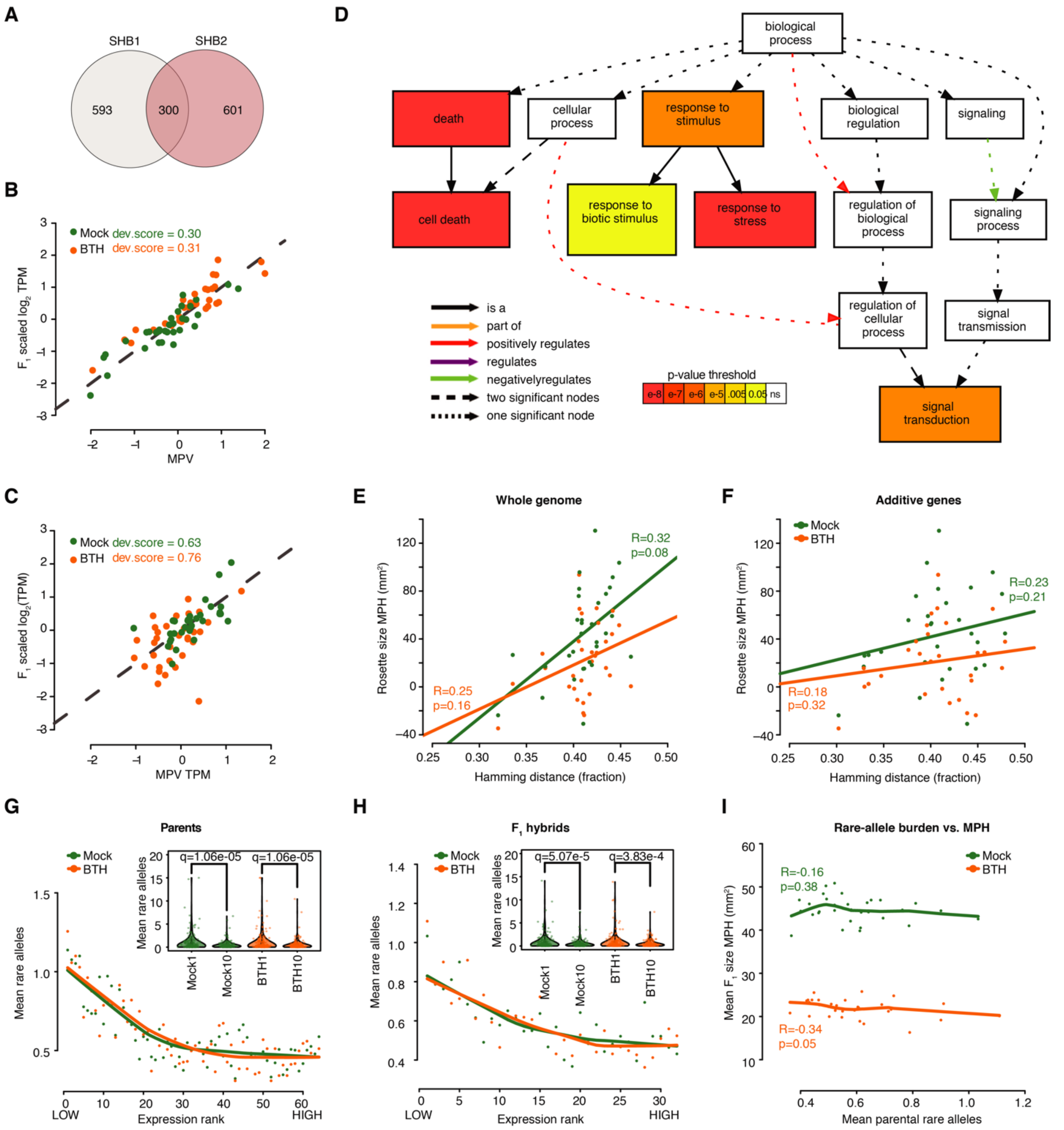


Fig. 2. Neither genetic distance nor rare-allele burden in additively expressed genes affect biomass in F$_1$ hybrids. **(A)** Overlap in additively expressed genes in the two experiments. **(B,C)** At4G11830, a common additive gene (B), and At2G41980, a randomly chosen gene that does not behave in a consistent manner (C). **(D)** GO enrichment of the common additive genes. **(E,F)** Inter-parental Hamming distance calculated with whole-genome SNPs (E) or SNPs of additive genes significantly correlated with heterosis. **(G,H)** Expression of additive genes and upstream-rare-allele counts in parents (G) and F$_1$s (H). Samples were ranked by expression value for every gene within the gene list. Points: average upstream rare allele counts of all samples sharing the same rank; lines: LOWESS trend lines. Insets show the upstream rare-allele count of samples in the top (Mock 10, BTH10) and bottom decile (Mock 1, BTH 1) of expression ranks. **(I)** Rosette size MPH varies independently of mean parental rare allele burden in additive genes. For each additive gene, F$_1$s were ranked by average number of rare alleles in their parents. Points: average rosette size MPH of all F$_1$s sharing the same rank; lines: LOWESS trend lines.

(Fig. 2G-H, Methods). Regardless of the treatment, we observed on average a significantly higher rare allele count associated with low expression ranks in both inbred parents (Wilcoxon signed rank sum test with Benjamin-Hochberg FDR: mock, **α**=1.1x10^{-5}, BTH: **α**=1.1x10^{-5}) and their F$_1$s (mock: **α**=5.1x10^{-5}, BTH: **α**=3.8x10^{-4}). Upstream rare-allele burden therefore tends to quantitatively reduce the expression of these genes, and the mid-parent-like nature of F$_1$ gene expression is a likely result of quantitative complementation. However, an excess number of rare alleles upstream of these additive genes did not affect rosette size MPH in F$_1$s, as there was no correlation between expression MPV in the F$_1$s and the average count of parental rare alleles (Fig. 2I, Pearson correlation coefficient R=-0.16, p=0.38). Quantitative complementation may therefore be an inherent property of hybrids regardless of the degree of heterosis.

Biomass heterosis is associated with large-scale low-expression dominance

Because complementation and dominance are common explanations for heterosis, we wanted to learn whether there is a set of genes for which the F$_1$ expression level is particularly likely to be close to the level in the high-expression parent. We first identified 1,805 non-additively expressed genes whose RNA abundance distribution in the F$_1$s significantly deviated from the expected distribution of MPV (Data S4, Methods). These included both high- and low-abundance transcripts as well as ones that varied little across samples or ones that varied substantially (Fig. 3A-B), indicating that expression level and variance did not greatly bias our ability to discover non-additive genes. Only 150 (~8%) of these were consistently expressed above MPV in the F$_1$s. For the great majority of genes, low expression level was dominant, such that most genes were expressed in the F$_1$s at a level between the lower parent and the MPV. Note that dominance here only refers to the observed, combined expression level of both alleles, and that no interaction between alleles is implied (Fig. 3C).

To discover groups of genes with a similar behavior across samples, we performed k-means clustering (k=13) of the 1,805 non-additive genes (Fig. 3D-E). We examined how well the behavior of different clusters across samples was correlated (Fig. 3D), finding that one particular cluster (cluster 6, n=116) behaved in a manner that was opposite to that of all other clusters. Nine clusters had similar trends in F$_1$s and inbreds, where mean expression values were either positively or negatively correlated with rosette size, but the details of the relationships between gene expression and rosette size differed for F$_1$s and inbred parental lines (Fig. 3E).

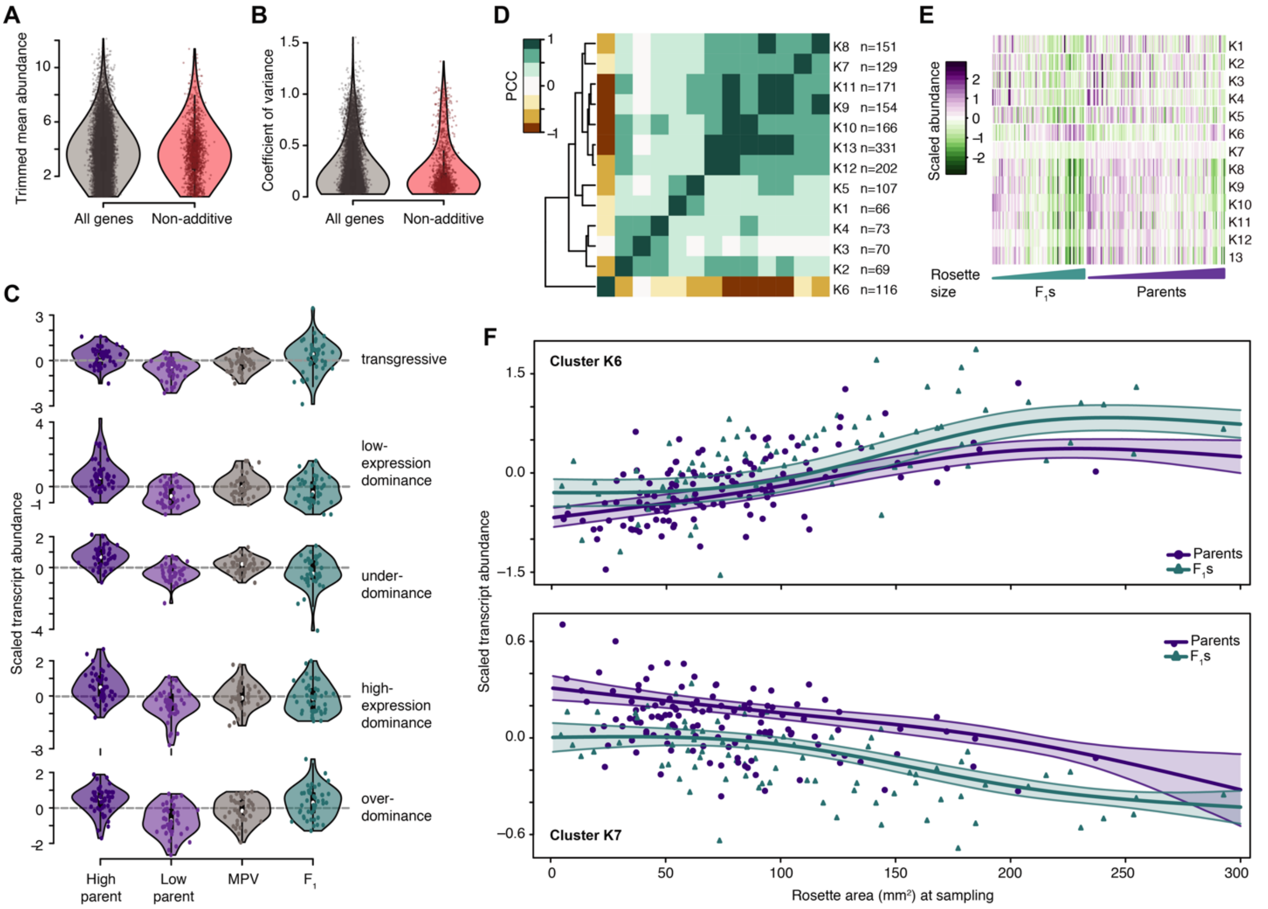


Fig. 3. Expression levels of non-additive genes are systematically different between inbreeding accessions and F_1 hybrids. (A-B) Non-additive genes are unbiased for transcript abundance and expression variance compared to the background. **(C)** Five example genes (from top: AT5G07960, AT3G61170, AT2G17670, AT2G27510, AT4G12970) showing that non-additive genes exhibit various forms of dominance in F_1s. **(D)** Correlation among all non-additive gene clusters. Pearson correlation coefficients were calculated using cluster average. **(E)** Heatmap showing the average expression of each non-additive cluster in each sample, sorted into F_1s and inbred parental lines, and arranged by ascending final rosette size. **(F)** Linear-Mixed-Model spline fitting of cluster 6 (top) and 7 (bottom). Points: cluster mean expression in each sample; shaded area: 95% Bayesian credible intervals.

To obtain further insight into the above observations, we formally investigated the relationship between average gene expression of each cluster and rosette size, by performing Bayesian linear-mixed-model spline fitting (Fig. 3F-G, Fig. S7, Methods). Clusters 1, 2, 3 and 5 showed no clear trend of expression and size, cluster 6 showed expression increasing in parallel with rosette size (Fig. 3F), and the remaining clusters showed expression decreasing with increasing rosette size. For these remaining eight clusters, F_1 hybrids tended to have lower expression than the inbreds across the entire range of rosette size (Fig. 3G, Fig. S7). Gene Ontology (GO) analysis did not indicate that clusters were specific for particular biological processes.

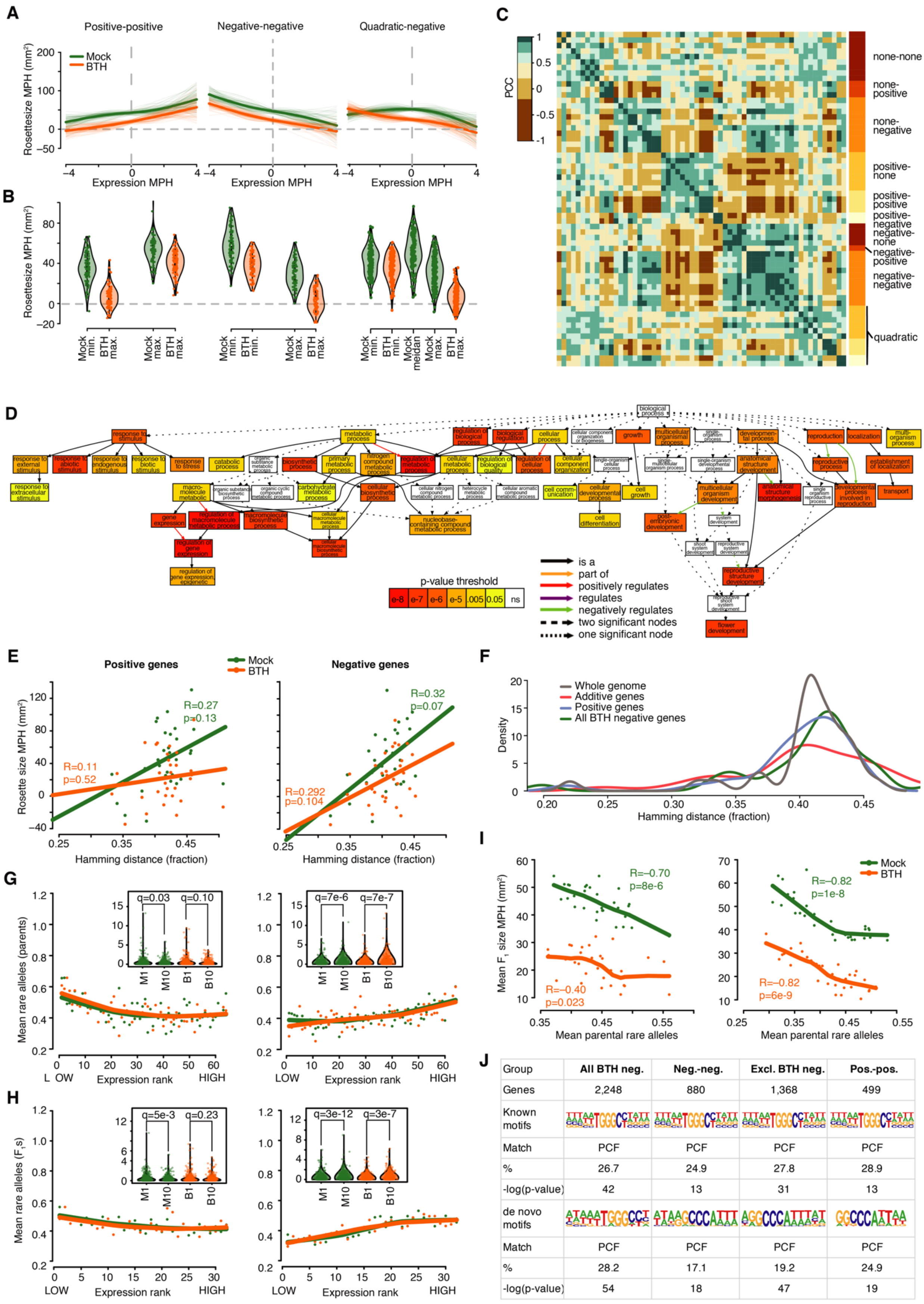
Non-additive gene expression is thus pervasive in F_1 hybrids, with the overwhelming majority of genes expressed below parental average level. Expression levels of the non-additive genes, involved in multiple biological processes, covaries with rosette size, with a consistent trend of average expression level in F_1s in the direction that correlates with larger size.

Targets of a common transcription factor are quantitatively associated with heterosis

Having established that non-additively expressed genes are systematically associated with plant size, we investigated whether the degree of non-additive expression in F_1s may affect heterosis. We focused on BTH responsive genes and asked whether deviations of F_1 expression values from the MPV (i.e., expression MPH) in individual trios exhibited were related to rosette size MPH. Clear correlations could be observed for many genes, which could be broadly categorized as (monotonic) positive, (monotonic) negative, or quadratic, while in some cases no correlation was observed (Fig. 4A). We defined 61 groups of genes that fell into these different categories by k-means clustering. Some clusters shared similar relationships between rosette size and expression MPH, and we therefore sorted the clusters further into 12 classes reflecting the pattern of correlation under mock and BTH treatment (e.g., "negative-negative" means negative correlation under both treatments, Fig. 4C, Methods). Correlation often changed in response to treatment, with the majority of genes exhibiting negative correlation with rosette size MPH under at least one condition, a trend that increased after BTH treatment (Fig. S8). This observation confirms our finding that negative dominance is important for heterosis in *A. thaliana*.

The genes with negative size-expression correlation after BTH or after BTH and mock treatment (Data S5, "negative genes" hereafter) are enriched for GO terms "regulation of gene expression", "floral organ development", and "response to (abiotic) stimuli" (Fig. 4D, Fig. S10). Genes with positive correlation under both treatments (Data S6, "positive genes" hereafter) were moderately enriched for photosynthesis (Fig. S9), suggesting that biomass heterosis is related to repressed reproductive growth and (abiotic) stress response as well as increased photosynthesis in F_1 hybrids.

Similar to whole-genome genetic distance between parents (Fig. 2E), common polymorphisms in genes from either the positive (Fig. 4E, left panel) or the negative clusters (Fig. 4E, right panel) were at best moderately correlated with biomass MPH after mock treatment, and even less so after BTH treatment, again emphasizing that genetic complementation alone has limited success in explaining heterosis. Moreover, the distribution of polymorphisms across the entire genome or in the focal gene groups was similar (Fig. 4F), implying that the focal genes are not characterized by a distinct spectrum of common polymorphisms.



Group	All BTH neg.	Neg.-neg.	Excl. BTH neg.	Pos.-pos.
Genes	2,248	880	1,368	499
Known motifs				
Match	PCF	PCF	PCF	PCF
%	26.7	24.9	27.8	28.9
-log(p-value)	42	13	31	13
de novo motifs				
Match	PCF	PCF	PCF	PCF
%	28.2	17.1	19.2	24.9
-log(p-value)	54	18	47	19

Fig. 4. Genes whose degree of non-additive expression in trios correlates with hybrid performance. (A) Exemplary clusters of "positive-positive" (left), "negative-negative" (middle), and "quadratic-negative" (right) genes. Thick solid line: spline fitting of the cluster means; thin lines: spline fitting of individual cluster members. **(B)** Average biomass MPH for rosette samples with low- vs. high- expression of genes in the same

clusters as in (A). Each violin depicts the distribution of cluster gene expression averaged across the top and the bottom (and the middle for the quadratic relationship) deciles of samples. **(C)** Pearson correlation coefficients (PCC) of all 61 clusters based on LMM-spline modeling. The clusters are further sorted into 12 classes labeled on the right according to the relationship between gene expression and biomass under mock or BTH treatment. **(D)** GO enrichment for genes from the negative cluster; see Fig. S10 for more details. **(E)** Interparental Hamming distance using SNP subsets of genes from the positive (left panel) and negative (right panel) clusters and correlation with rosette size MPH. **(F)** Kernel density plots of inter-parental Hamming distance as the fraction of all filtered diallelic SNPs. **(G,H)** Association between expression rank and rare-allele burden upstream of positive (left) and negative(right) genes in parents (G) and F$_1$s (H). Insets show the upstream rare allele counts of samples in the top and bottom decile of expression ranks. M1, mock bottom decile; M10, mock top decile; B1, BTH bottom decile; B10, B top decile. **(I)** Mean upstream rare-allele burden between parents in genes from both positive and negative clusters significantly affects heterosis. **(J)** Regulatory regions of genes from both positive and negative clusters are enriched for a PCF binding motif.

In the parents, a larger number of mean upstream rare alleles is significantly associated with increased expression of negative genes (Fig. 4G, mock: $\boldsymbol{\alpha}=7\text{x}10^{-6}$, BTH: $\boldsymbol{\alpha}=7\text{x}10^{-7}$, Wilcoxon signed rank sum test, Benjamin-Hochberg FDR), and less associated with reduced gene expression in the positive genes (mock: $\boldsymbol{\alpha}=0.03$, BTH: $\boldsymbol{\alpha}=0.10$). These trends are also seen in the F$_1$ hybrids (Fig. 4H, $\boldsymbol{\alpha}=0.005$, positive-mock; $\boldsymbol{\alpha}=0.23$, positive BTH; $\boldsymbol{\alpha}=3\text{x}10^{-12}$, negative-mock; $\boldsymbol{\alpha}=3\text{x}10^{-7}$, negative-BTH). Considering the implication of the expression of these genes in biomass heterosis, this observation is consistent with rare-allele burden disrupting normal gene expression. In contrast, the number of mean upstream rare alleles in both positive and negative genes is a strong negative predictor of rosette size heterosis in F$_1$ hybrids (Fig. 4I). This suggests that heterosis tends to be greater in hybrids whose parents have fewer rare alleles in the upstream regions of genes from the positive and negative clusters.

To begin to discover potential regulatory mechanisms, we performed motif enrichment analysis among the heterosis associated genes (Methods). PROLIFERATING CELL FACTOR (PCF) and c-Myc transcription factor binding motifs are highly enriched in the promoters of negative genes (569/2,248 genes, $p=10^{-42}$, and 433/2,248 genes, $p=10^{-21}$, Fig. 4J, Fig. S10). PCF-binding motifs are also highly enriched in the positive genes (144/499 genes, $p=10^{-13}$, Fig. 4J). These findings were corroborated by *de novo* motif searches (Fig. 4J). PCF/TCP proteins constitute a conserved plant-specific transcription factor family that includes several regulators of cell cycle, growth, and disease resistance (Gonzalez, 2015; Li, 2015). That both positive and negative genes were enriched for PCF motifs points to these factors as a potential central toggle for global re-modeling of hybrid transcriptomes.

Discussion

Heterosis by definition must be a highly heterogeneous phenomenon, as it describes the performance of F_1 hybrids in a variety of traits that are of interest to humans, whether related to fitness in the wild or not. Such heterogeneity calls for systematic approaches to capture the emergent properties of a broad spectrum of heterotic hybrids. Systematically sampling the genetic diversity of the species' natural range, our study confirmed the pervasiveness of heterosis in *A. thaliana*, consistent with previous studies in which smaller populations were used (Oakley et al., 2019). Incorporating size measurements into our analyses enabled us to quantitatively assess the relationships between gene expression and plant growth (and hybrid performance), which sets our analysis apart from other -omics studies that evaluated heterosis as a binary feature or that did not link -omics data and growth phenotypes. By associating transcriptome changes and phenotypic variation, we identified specific physiological processes as a signature of heterosis in our system.

Previous work has produced strong evidence for a role of circadian gene expression in heterosis among certain *A. thaliana* F_1 hybrids. While we did not find an enrichment of binding motifs for core circadian regulators in genes whose non-additive expression affects heterosis in our set of hybrids, we did find potential finding sites for PCF/TCP factors, several of which have been linked to the circadian clock (Li, 2015; Pruneda-Paz et al., 2009). Studies with individual hybrids have suggested up-regulation of photosynthetic function (Fujimoto et al., 2012) and down-regulation of stress-response (Miller et al., 2015) in *A. thaliana*. We observed the same, with the new insight that the degree of non-additivity in these pathways correlates with the degree of heterosis.

Additive (i.e., near-MPV) expression itself does not correlate with the larger size of F_1 hybrids, hence probably does not directly contribute to heterosis. Additive expression is an intrinsic trait for certain genes, mainly enriched in cell-death and stress-response pathways, in F_1 hybrids, largely independent of specific parental combinations. Whether tighter control of biotic defense responses capacitates hybrid vigor requires further investigation.

In our system, biomass heterosis is driven in large part by non-additive gene expression. For many genes, when parent-hybrid trios are considered, the degree of non-additivity in F_1 expression can either positively or negatively predict heterotic performance. GO enrichment points to repression of reproductive development and abiotic stress response, as well as increased photosynthesis as factors that support more vigorous growth in F_1 hybrids. The rewiring of transcriptional networks is likely to occur in *trans*, as binding motifs for a small number of transcription factors were enriched in the promoters of the focal genes. This further suggests that beyond reciprocal complementation, unifying factors may exist that determine heterotic potentials of individual parental combinations.

An important question is whether heterosis is more likely to be associated with up- or down-regulation of genes (Birchler et al., 2003). Previous studies, while pointing to a global modulation of gene regulatory networks in F_1 hybrids, did not provide conclusive answers (Jeffrey Chen, 2013). Here, we observe a consistent species-wide trend of non-additive expression leading to F_1 hybrids being more similar to the low-expression parents. Considering that stabilized or increased gene expression is generally favored by selection (Fraser et al., 2010; Kita et al., 2017), the coordinated down-regulation of genes in hybrids suggests that heterosis is not an emergent property resulting from selection benefits. Consistent with our finding that transcriptional programs related to cell proliferation seem to be negatively regulated in heterotic *A. thaliana* hybrids, a recent study in yeast showed that inter-specific F_1 hybrids can grow more vigorously by relaxing cell-cycle checkpoints and tolerating excessive genetic mutations (Bar-Zvi et al., 2017; Herbst et al., 2017). It is possible, therefore, that heterosis arises as an emergent consequence of how gene regulatory networks are hardwired, and with its long-term beneficial effect (such as supporting gene flow and adaptive introgression) being merely an indirect consequence.

We found that neither inter-parental whole-genome genetic distance nor the genetic distance of focal genes groups adequately explained the performance of the F_1 hybrids. While this seems to contradict insights from crops, it agrees with recent findings in non-domesticated species including *A. thaliana* (Oakley et al., 2019) and *S. cerevisiae* (Plech et al., 2014). It highlights the importance of searching for mechanisms of heterosis outside of domesticated species of agronomic interest, in which a higher genetic load and therefore a stronger role of complementation is expected from their population history.

We observed putative deleterious effects (i.e., strong deviation of gene expression from that of the population mean) of rare alleles in upstream regulatory sequences, primarily in inbred parental accessions, and to a lesser degree in F_1 hybrids, suggesting that rare-allele burden has common effects in inbreds and hybrids alike. The correlation between rare-allele burden and gene expression is more pronounced in additively than non-additively expressed genes. However, while the (milder) rare-allele burden in non-additive genes is strongly associated with phenotypic consequences in F_1 hybrids, this is not reflected in the behavior of the additive genes. We conclude that concerted non-additive gene expression, rather than canalization via additive gene expression, is a main driver of heterosis. It needs to be tested independently whether heterosis can be predicted from expression and genetic data of a focused set of parental genes.

Materials and Methods

Generation of genetic resource: Accessions covering the entire species range were chosen from the *A. thaliana* 1001 Genomes Project (1001 Genomes Consortium, 2016). F_1 hybrids used in this study were generated via random crosses, either by randomly crossing individual accessions that reached flowering stage at the same time (SHB1) (Vasseur et al., 2019), or according to a pre-generated randomized crossing scheme after subjecting seedlings to a saturating (12 weeks) vernalization under 4°C short-day conditions (SD, 8/16 hr photoperiod) to synchronize flowering (SHB2).

Experimental design: The first experiment (Fig. 1A, "SHB1") initially included 101 parent-F_1 trios of altogether 286 distinct genotypes. Single plants were grown in triplicate following an incomplete randomized block design, with each tray as a block within which each genotype was sown as an adjacent pair. The second experiment (Fig. 1B, "SHB2") included 40 parent-F_1 trios. Single plants were grown in triplicates, following a split-block design where each block held 10 parent-F_1 trios with each row consisting of one trio with duplicates of plants in adjacent pots. The trios within each block, and the relative positions of genotypes within each trio were randomized. Plants were subjected to either a mock or an artificial defense hormone treatment (see below). After accounting for germination, survival, and initial filtering of RNA-seq outputs, 82 hybrids and 124 inbred parents in SHB1, and 32 trios in SHB2 remained for downstream analyses.

To minimize circadian bias, sowing for both experiments was scheduled in batches to ensure that harvesting could be finished within a 30-minute window at the same hour for a number of consecutive days. At 21 days after sowing, the healthiest appearing plant of each genotype was used for RNA-seq, to ensure any sampling bias is systematically towards the same direction for both inbreds and hybrids. Meanwhile rosette size measurements were obtained for the same individual plants for which RNA-seq were performed.

For a list of genotypes analyzed in both experiments, see Data S1.

Plant culture, treatment, and sampling: Single plants were grown in a 1:1 mixture of calcined clay media (Diamond Pro, Arlington, TX, USA) and vermiculite (Floragard, Oldenburg, Germany) supplemented with liquid growth media (Conn et al., 2013). Plants were not vernalized, to ensure that they remained in the vegetative growth phase. As a proxy of vegetative biomass ((Vasseur et al., 2018) and Fig. S1), the rosette area of 21-day old plants was measured. The full rosettes grown under 16°C long-day (LD,16/8 photoperiod) were harvested and flash-frozen at 21 day-after-germination (DAG).

Defense hormone treatment started 14 days after sowing. An analog of the defense hormone salicylic acid (SA), BTH (acibenzolar-S-methyl, Sigma-Aldrich) was used, with optimal dose and treatment scheme of the defense hormone that had been established in a pilot experiment on multiple accessions (Fig. S2). Each 10×6-pot-tray block was the unit of treatment and plants were treated by topical spraying every other day with either a mock solution (20 mL; ddH_2O, 0.1% DMSO, 0.006% Silwet) or a BTH solution (20 mL; 100 mM acibenzolar-S-methyl, 0.1% DMSO, 0.006% Silwet), and covered for 1 hour with transparent plastic lids after spraying. A total of five treatments were administered. Full rosettes were harvested and flash frozen at 21 DAG.

Growth image analysis: Plant growth was monitored by daily image capture from the top of the trays using the RAPA system (Capovilla et al., 2018). Rosette areas were acquired by automatic image segmentation and counting of green pixels, supplemented with manual curation. The rosette size estimates were then converted from pixel counts to mm^2 by multiplication with a calibration factor.

RT-qPCR: To establish defense hormone treatment and dosage, the effect of salicylic acid and BTH application was tested by treating 18 Accessions with Mock (ddH_2O, 0.1% DMSO, 0.006% Silwet), 350 mM SA and 100 mM BTH in 3 replicates, each with duplicated plants for phenotyping and qPCR. After 5 treatments the rosettes were harvested in one set of plants to compare their sizes, while the other set of plants were used for qPCR to compare the effect of 350 mM SA and 100 mM BTH treatments. Specifically, RNA was extracted and reverse transcribed. qPCR was performed using SYBER green (Thermo Scientific Maxima SYBR Green qPCR Master Mix (2x)) and primers for *ACTIN2*, *UBC21*, *PR1* and *NPR1*. Normalization across plates was performed using the same set of samples featured on all plates. The data were analyzed by calculating ΔΔCq (Fig. S2).

RNA-seq: RNA-seq libraries were constructed as described (Cambiagno et al., 2021), using 750 ng total RNA from full rosettes as input. All libraries, each carrying a unique barcode combination were pooled and sequenced in multiple single-end lanes on an Illumina HiSeq 3000 platform for a target coverage of 5M reads per sample.

RNA-seq read mapping and post-processing: FASTQ files from multiple lanes were merged and mapped to TAIR10 transcriptome using *RSEM* (*bowtie2*) with default parameters. Libraries with more than 8M mapped reads were subsampled to 8M with *seqtk* prior to mapping. Transcripts mapped to chloroplast, mitochondria, rDNA clusters, transposable elements (TEs) and pseudogenes, as well as transcripts with effective length less than 150 nt were removed from the raw *RSEM* count file. TPM (transcripts per million) counts were then re-estimated for the rest of the genes. Libraries with less than 2M mapped reads and those identified as extreme outliers following a principal-component analysis (PCA) of the whole-transcriptome $\log_2$(TPM) values were excluded from further analysis. Gene lists were further filtered for average transcript abundance (trimmed mean of $\log_2$(TPM)>0.3) and coefficient of variance >0.15.

Additive gene calling: For SHB1 data, MPV of each gene was calculated for all complete Parent-F_1 trios by taking the arithmetic mean of the parental $\log_2$(TPM). Linear regression was then performed between corresponding F_1 expression value and the MPV. For SHB2, a linear-mixed model was used to correct for treatment and batch effects. Genes were filtered for regression coefficient >0.5 and R^2 >0.4 for SHB1, and regression coefficient >0.4 and sigma <0.6 for SHB2. All thresholds were determined by quantiles. Genes called in both SHB1 and SHB2 were taken as common additive genes.

Non-additive gene calling: With SHB1, a population-wide MPV distribution was established for each gene by calculating arithmetic means of $\log_2$(TPM) between all possible pairwise combinations of inbred accessions. Two-sided Kolmogorov-Smirnov test was performed per gene to test if the $\log_2$(TPM) from the F_1 hybrids were drawn from the MPV distribution. Genes with $q<0.001$ (Benjamin-Hochberg FDR) were considered as non-additively expressed genes.

Bayesian modelling of non-additive expression and plant size: Non-additive genes from SHB1 were clustered by K-means, with the optimal K determined by the elbow method. A linear-mixed-model (LMM) spline was fitted using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2021) for gene expression:

GeneExpression~IsHybrid+Size+Size:IsHybrid+(IsHybrid|LibraryBatch),

in which *GeneExpression* is the z-scaled $\log_2$TPM, *IsHybrid* is a binary code of the hybrid/inbred identity, *Size* is the z-score of rosette size at sampling. Natural cubic splines were modeled for *Size* and *Size-IsHybrid* interaction. The 95% credible intervals for the parameter estimates were established with 10,000 iterations of Bayesian simulation using the *arm* package (Andrew Gelman, Yu-Sung Su, Masanao Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman, Tian Zheng, Vincent Dorie, 2021).

Expression-plant size MPH correlation: Expression-MPH and size-MPH were calculated per trio by calculating the per-gene expression and rosette area difference between F_1 and the MPV in corresponding treatments and replicates. Size MPH-to-expression MPH regression spline was acquired separately for both treatments. An initial round of K-means clustering was performed on the resulting spline coefficients, with the optimal K determined as the division with the highest Dunn index which allows no more than 25% of the clusters carrying less than 5% of the genes. Resulting clusters were inspected and removed if size and expression MPH do not co-vary. The remaining genes (n=6,371) were re-clustered with the same criteria, and the resulting clusters were manually sorted based on size-expression covariation into 12 general categories (Data S2, Fig. 4D, Fig. S10).

Deviation score: A deviation score was calculated for each common additive gene:

d=i=1n(yi-yfit)2(ymax-ymin)2 ,

in which, y_i is the observed F_1 expression value of a given individual, and y_{fit} is the fitted F_1 expression value assuming a perfect additivity ($y=x$).

BTH responsive genes: The effect of BTH treatment on gene expression was identified by LMM:

*GeneExpression~Treatment+IsHybrid+Treatment*IsHybrid+(1|PlantBatch),*

to establish a significant threshold, 10,000 permutation was performed per gene, and the empirical p-value was corrected with Benjamin-Hochberg FDR. Genes with q<0.001 were kept as BTH responsive genes (n=8797), and examined for their size-MPH correlation.

Size-MPH covariation test: To establish the significance of the size-expression correlation, we performed a Wilcoxon signed-rank test on the genes in each of the 61 clusters (Fig. 4B). For each "none" cluster, each gene within the cluster was used as a data point, and the mean rosette size of 4 plants having the lowest and the highest expression MPH of the gene was calculated. The average rosette size corresponding to the two extremes of expression MPH were then compared with a two-sided Wilcoxon signed-rank test. Likewise, for "positive" and "negative" clusters, one-sided tests were used to test for significant differences between average rosette size corresponding to the two extremes of expression MPH within individual clusters. For "quadratic" clusters, separate one-sided tests were performed comparing the samples with extreme expression MPH against those with median expression MPH. Bonferroni correction was used to control for multiple hypothesis testing.

The test revealed that our sorting procedure erred on the conservative side: while the top and bottom deciles were significantly different for 17 clusters assigned to the "none" category, only 6 of the "none" clusters were misassigned as "positive", and none were misassigned as "negative" (Bonferroni corrected α<0.001, Data S2). The evidence for truly quadratic correlations was less clear.

Genetic distance: Pairwise SNP Hamming distances were acquired for parental combinations using *PLINK* v.1.90b (Purcell et al., 2007). For genome-wide genetic distance, all biallelic SNPs with minor allele frequency (MAF) >0.2 were used. For sub-genome genetic distance, SNPs were further subset by the genome coordinates (gene body + 1 kb upstream) of target features using *vcftools* v4.2 (Danecek et al., 2011). Pearson correlation coefficients were calculated between rosette size MPH and genetic distance either using all common polymorphisms in the genome, or those within the target features of interest.

GO enrichment Ontology enrichment was performed using the *Agrigo* v2 platform (Tian et al., 2017), with all gene IDs that passed our initial filtering (n=14,067, TAIR10 annotation) as background against plant GOslim database. Fisher's Exact Test was used, and the enrichment p-value was corrected using Yekutieli FDR. The enrichment results were visualized with the built-in DAG-drawer of *Agrigo* v2.

Rare allele analysis: Rare (MAF<0.05), biallelic SNPs 1 kb upstream of gene features were subset from the SNP annotations of the 1001 Genomes project (1001 Genomes Consortium, 2016). Genotype information at these SNPs were acquired for the accessions used in SHB2, and the sum of these rare SNPs upstream of each gene were calculated per accession. For F_1 hybrids, upstream rare-allele count was determined by the mean of the rare-allele counts of both parents.

Samples, separated by inbred parents/F_1 hybrids and treatments, were ranked for their expression values for each gene within a gene list of interest. For each given rank, a gene-list mean upstream rare-allele count was acquired by averaging across all samples received the same rank in any of the genes within the gene list. Relationships between gene-list mean upstream rare-allele count and expression rank were examined by LOWESS regression and tested with Wilcoxon signed rank sum test between the top and bottom decile of the expression rank.

Likewise, an average rosette-size MPH for each rank was calculated for the gene list, and the Pearson correlation was acquired between average rosette-size MPH and upstream rare-allele count.

Motif enrichment and *de novo* motif finding: Motif enrichment and *de novo* motif finding was carried out using *HOMER* v4.10.4 (Heinz et al., 2010) with TAIR10 reference genome and gene annotation. For every set of candidate genes, genomic sequences 1 kb upstream from the transcription start site (TSS) and 1 kb downstream from the transcription termination site (TTS) were indexed from the strand-specific gene coordinates. Both assays were performed by using the *findMotifsGenome.pl* function and *HOMER*'s in-built plant promoter motif database as reference:

```
findMotifsGenome.pl [geneset_promoter_coord.bed] [TAIR10_customref]
[motif_outputdir] -size given -S 15 -preparse -mcheck
[homer_database/data/knownTFs/plant/all.motifs]
```

The top known motif hit (in all cases $p \leq 10^{-10}$) from each candidate set was then used for a second motif enrichment step, where the promoters and downstream sequences were searched for the significant motif using *HOMER*'s *annotatePeaks.pl* function

```
annotatePeaks.pl [geneset_promoter_downstream.bed] [TAIR10_customref] -m
[motif_outputdir]/knownResults/known1.motif -size given -nmotifs -mbed [output1.bed] >
[output1.txt],
```

to generate corresponding genomic coordinates, and subsequently associated back to the genes containing the motif of interest in their regulatory regions (*bedtools* v2.26.0 intersect) (Quinlan and Hall, 2010).

References

1001 Genomes Consortium. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell **166**: 481–491.

Andrew Gelman, Yu-Sung Su, Masanao Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman, Tian Zheng, Vincent Dorie (2021). Data Analysis Using Regression and Multilevel/Hierarchical Models [R package arm version 1.12-2].

Bar-Zvi, D., Lupo, O., Levy, A.A., and Barkai, N. (2017). Hybrid vigor: The best of both parents, or a genomic clash? Current Opinion in Systems Biology **6**: 22–27.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, Articles **67**: 1–48.

Birchler, J.A., Auger, D.L., and Riddle, N.C. (2003). In search of the molecular basis of heterosis. Plant Cell **15**: 2236–2239.

Cambiagno, D.A., Giudicatti, A.J., Arce, A.L., Gagliardi, D., Li, L., Yuan, W., Lundberg, D.S., Weigel, D., and Manavella, P.A. (2021). HASTY modulates miRNA biogenesis by linking pri-miRNA transcription and processing. Mol. Plant **14**: 426–439.

Capovilla, G., Delhomme, N., Collani, S., Shutava, I., Bezrukov, I., Symeonidi, E., de Francisco Amorim, M., Laubinger, S., and Schmid, M. (2018). PORCUPINE regulates development in response to temperature through alternative splicing. Nat Plants **4**: 534–539.

Chae, E. et al. (2014). Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. Cell **159**: 1341–1351.

Conn, S.J., Hocking, B., Dayod, M., Xu, B., Athman, A., Henderson, S., Aukett, L., Conn, V., Shearer, M.K., Fuentes, S., Tyerman, S.D., and Gilliham, M. (2013). Protocol: optimising hydroponic growth systems for nutritional and physiological analysis of Arabidopsis thaliana and other plants. Plant Methods **9**: 4.

Danecek, P. et al. (2011). The variant call format and VCFtools. Bioinformatics **27**: 2156–2158.

Darwin, C. (1876). The effects of cross and self fertilisation in the vegetable kingdom (London: John Murray).

East, E.M. (1908). Inbreeding in corn. Rep. Conn. Agric. Exp. Stn **1907**: 419–428.

Fraser, H.B., Moses, A.M., and Schadt, E.E. (2010). Evidence for widespread adaptive evolution of gene expression in budding yeast. Proc. Natl. Acad. Sci. U. S. A. **107**: 2977–2982.

Fujimoto, R., Taylor, J.M., Shirasawa, S., Peacock, W.J., and Dennis, E.S. (2012). Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. Proc. Natl. Acad. Sci. U. S. A. **109**: 7109–7114.

Fujimoto, R., Uezono, K., Ishikura, S., Osabe, K., Peacock, W.J., and Dennis, E.S. (2018). Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. Breed. Sci. **68**: 145–158.

Gonzalez, D.H. (2015). Plant Transcription Factors: Evolutionary, Structural and Functional Aspects (Academic Press).

Griffing, B. (1990). Use of a controlled-nutrient experiment to test heterosis hypotheses. Genetics **126**: 753–767.

Groszmann, M., Gonzalez-Bayon, R., Lyons, R.L., Greaves, I.K., Kazan, K., Peacock, W.J., and Dennis, E.S. (2015). Hormone-regulated defense and stress response networks contribute to heterosis in Arabidopsis F1 hybrids. Proc. Natl. Acad. Sci. U. S. A. **112**: E6397–406.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell **38**: 576–589.

Herbst, R.H., Bar-Zvi, D., Reikhav, S., Soifer, I., Breker, M., Jona, G., Shimoni, E., Schuldiner, M., Levy, A.A., and Barkai, N. (2017). Heterosis as a consequence of regulatory incompatibility. BMC Biol. **15**: 38.

Hochholdinger, F. and Baldauf, J.A. (2018). Heterosis in plants. Curr. Biol. **28**: R1089–R1092.

Huang, X. et al. (2016). Genomic architecture of heterosis for yield traits in rice. Nature **537**: 629–633.

Jeffrey Chen, Z. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. Nat. Rev. Genet. **14**: 471–482.

Kita, R., Venkataram, S., Zhou, Y., and Fraser, H.B. (2017). High-resolution mapping of cis-regulatory variation in budding yeast. Proc. Natl. Acad. Sci. U. S. A. **114**: E10736–E10744.

Kremling, K.A.G., Chen, S.-Y., Su, M.-H., Lepak, N.K., Romay, M.C., Swarts, K.L., Lu, F., Lorant, A., Bradbury, P.J., and Buckler, E.S. (2018). Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature **555**: 520–523.

Krieger, U., Lippman, Z.B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nat. Genet. **42**: 459–463.

Lippman, Z.B. and Zamir, D. (2007). Heterosis: revisiting the magic. Trends Genet. **23**: 60–66.

Li, S. (2015). The Arabidopsis thaliana TCP transcription factors: A broadening horizon beyond development. Plant Signal. Behav. **10**: e1044192.

Miller, M., Song, Q., Shi, X., Juenger, T.E., and Chen, Z.J. (2015). Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of Arabidopsis. Nat. Commun. **6**: 7453.

Mitchell-Olds, T. (1995). Interval mapping of viability loci causing heterosis in Arabidopsis. Genetics **140**: 1105–1109.

Oakley, C.G., Lundemo, S., Ågren, J., and Schemske, D.W. (2019). Heterosis is common and inbreeding depression absent in natural populations of Arabidopsis thaliana. J. Evol. Biol.

Paschold, A., Jia, Y., Marcon, C., Lund, S., Larson, N.B., Yeh, C.-T., Ossowski, S., Lanz, C., Nettleton, D., Schnable, P.S., and Hochholdinger, F. (2012). Complementation contributes to transcriptome complexity in maize (Zea mays L.) hybrids relative to their inbred parents. Genome Res. **22**: 2445–2454.

Plech, M., de Visser, J.A.G.M., and Korona, R. (2014). Heterosis is prevalent among domesticated but not wild strains of Saccharomyces cerevisiae. G3 **4**: 315–323.

Pruneda-Paz, J.L., Breton, G., Para, A., and Kay, S.A. (2009). A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. Science **323**: 1481–1485.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**: 559–575.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Seymour, D.K., Chae, E., Grimm, D.G., Martín Pizarro, C., Habring-Müller, A., Vasseur, F., Rakitsch, B., Borgwardt, K.M., Koenig, D., and Weigel, D. (2016). Genetic architecture of nonadditive inheritance in Arabidopsis thaliana hybrids. Proc. Natl. Acad. Sci. U. S. A. **113**: E7317–E7326.

Shull, A.F. (1912). The Influence of Inbreeding on Vigor in Hydatina Senta. Biol. Bull. **24**: 1–13.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. **45**: W122–W129.

Vasseur, F., Bresson, J., Wang, G., Schwab, R., and Weigel, D. (2018). Image-based methods for phenotyping growth dynamics and fitness components in Arabidopsis thaliana. Plant Methods **14**: 63.

Vasseur, F., Fouqueau, L., de Vienne, D., Nidelet, T., Violle, C., and Weigel, D. (2019). Nonlinear phenotypic variation uncovers the emergence of heterosis in Arabidopsis thaliana. PLoS Biol. **17**: e3000214.

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X., and Zhang, Q. (2008). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat. Genet. **40**: 761–767.

Yang, M., Wang, X., Ren, D., Huang, H., Xu, M., He, G., and Deng, X.W. (2017). Genomic architecture of biomass heterosis in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A.

Acknowledgements

We thank members of the Weigel lab for assistance in sample collection. François Vasseur, Ulrich Lutz, Rebecca Schwab, and Yiliang Ding for helpful discussions. This work was supported by the Max Planck Society.

Author Contributions

Conceptualization – WY, DW; Methodology – WY, IB; Experiment – WY, FB, SS, RK; Data analysis – WY, FB, TS; Writing – original draft: WY, FB; Writing – review & editing – WY, DW.

Competing Interests

D.W. holds equity in Computomics, which advises breeders.

Data and Material Availability

Raw sequencing data is available at the ENA under the accession ERA9420648 and ERA9420737. Code to generate the results is available at: https://github.com/weigelworld/SigHeterosis.

Supplementary Data Sets

DataS1.xlsx	Sequenced and analyzed genotypes from SHB1 and SHB2
DataS2.xlsx	Wilcoxon rank sum test on the size-expression MPH spline clusters
DataS3.txt	300 common additive genes
DataS4.txt	Non-additive genes called in SHB1, and their cluster identity
DataS5.txt	Genes negatively associated with heterosis under BTH or mock + BTH conditions (“negative genes”)

DataS6.txt	Genes positively associated with heterosis under both mock and BTH conditions (“positive genes”)

Supplementary Figures

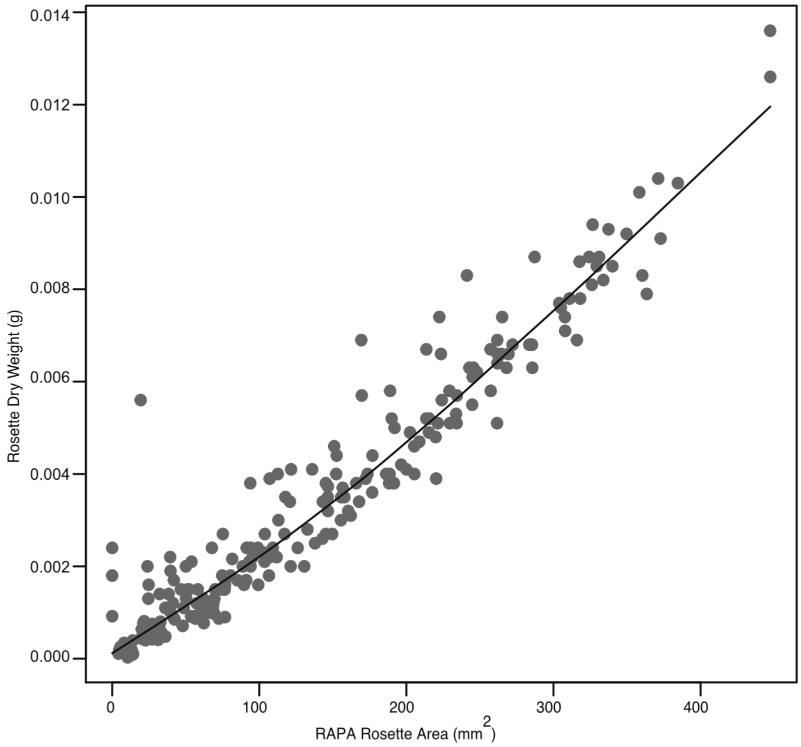


Fig. S1. Rosette area serves as a good predictor of rosette biomass. Pearson correlation coefficient (R= 0.96, p<2.2e-16) of biomass (g) with rosette area (mm^2) as measured on the RAPA system at 21 days after sowing of 221 individuals of mixed genotypes (grey dots), and LOWESS trendline.

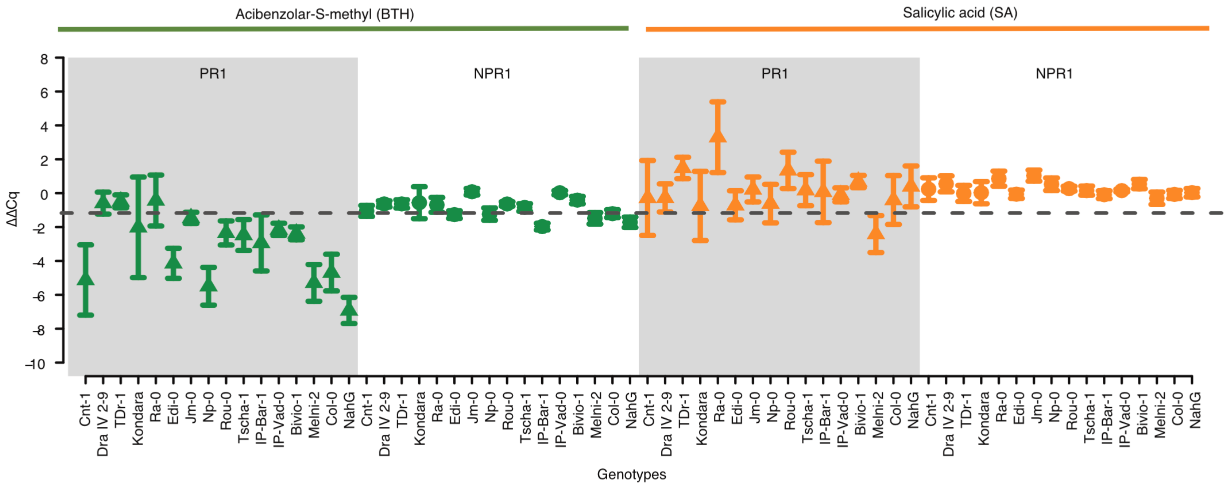


Fig. S2. Efficient induction of defense responses in *A. thaliana* accessions by BTH dosage used. qRT-PCR of two defense marker genes *PR1* and *NPR1* in 16 natural accessions after BTH and SA treatment. Each dot represents the mean ΔΔCq value against housekeeping genes in mock-treated plants, with error bars indicating the standard deviation of the biological replicates.

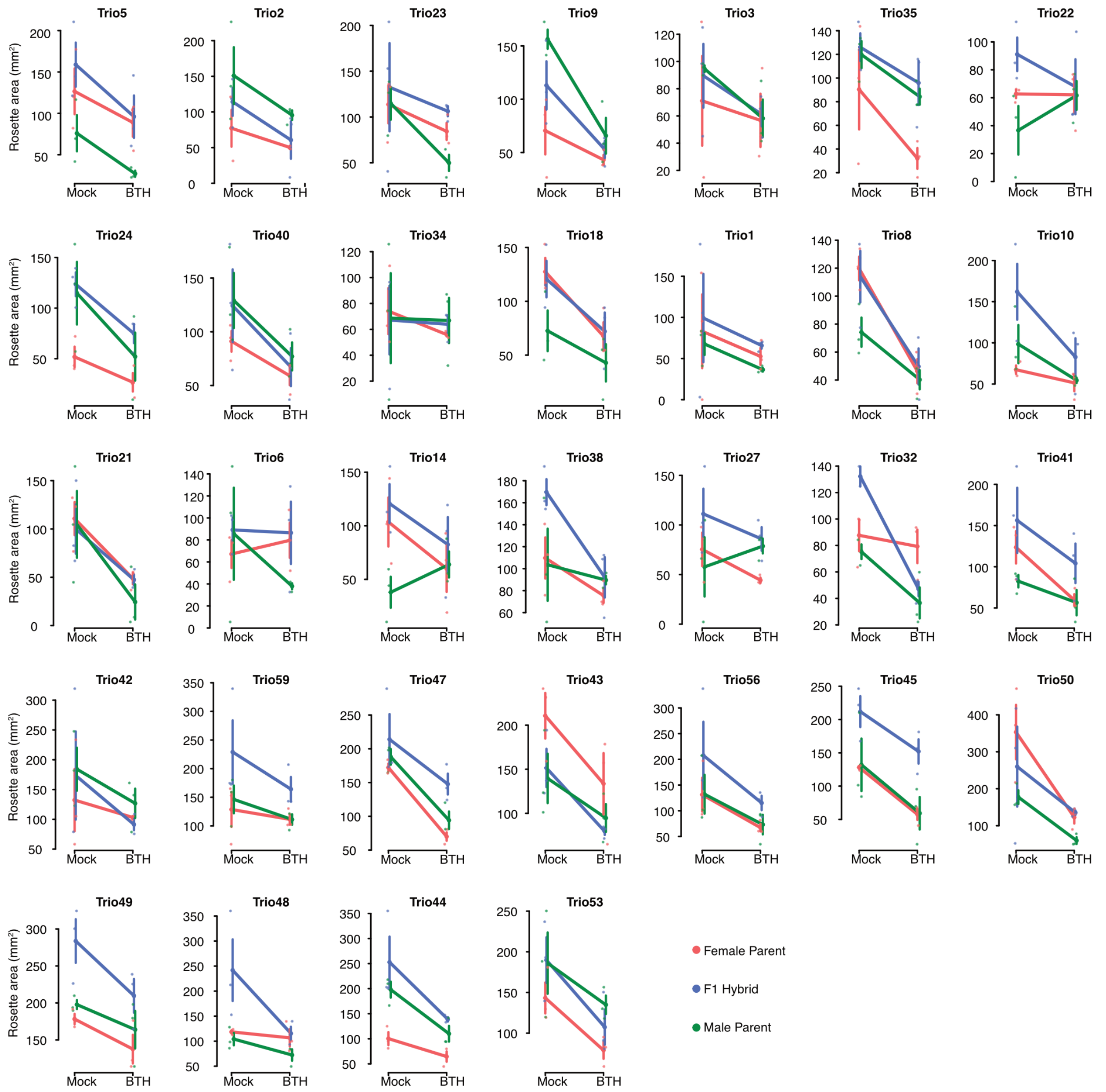


Fig. S3. Reaction norm of rosette area (mm^2) after mock and BTH treatments for all inbred parents and F$_1$ hybrid trios. Solid lines connected the mean rosette area under the two treatments, with the dots illustrating individual rosette area of each plant and the error bars showing standard deviation of the biological replicates.

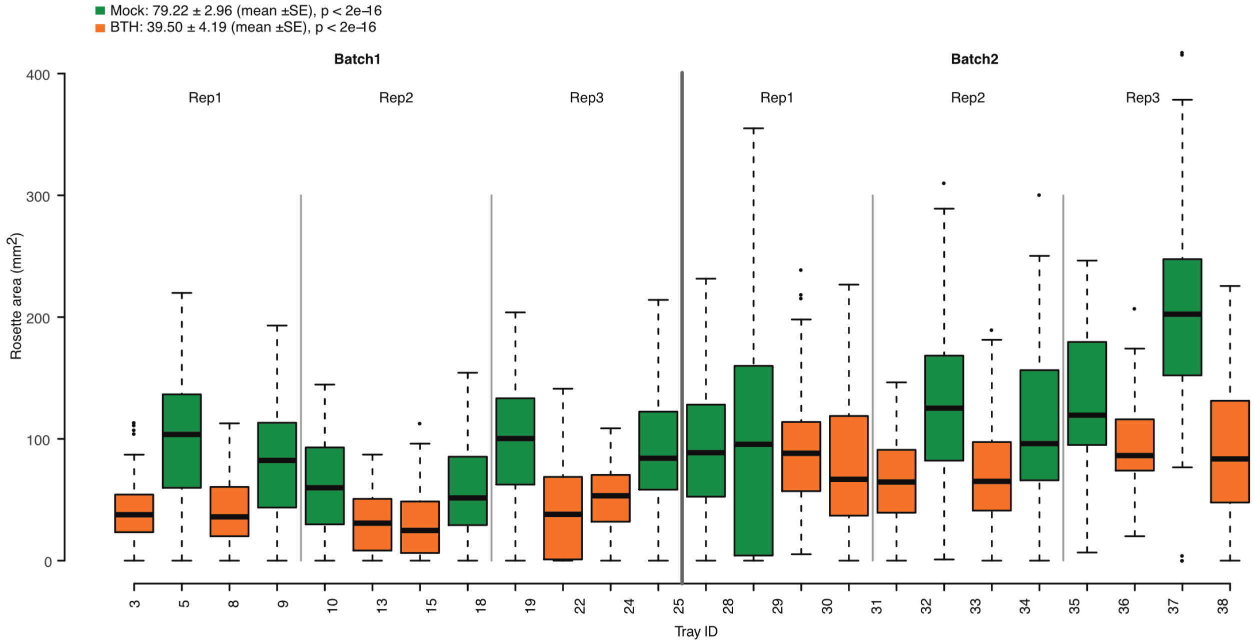
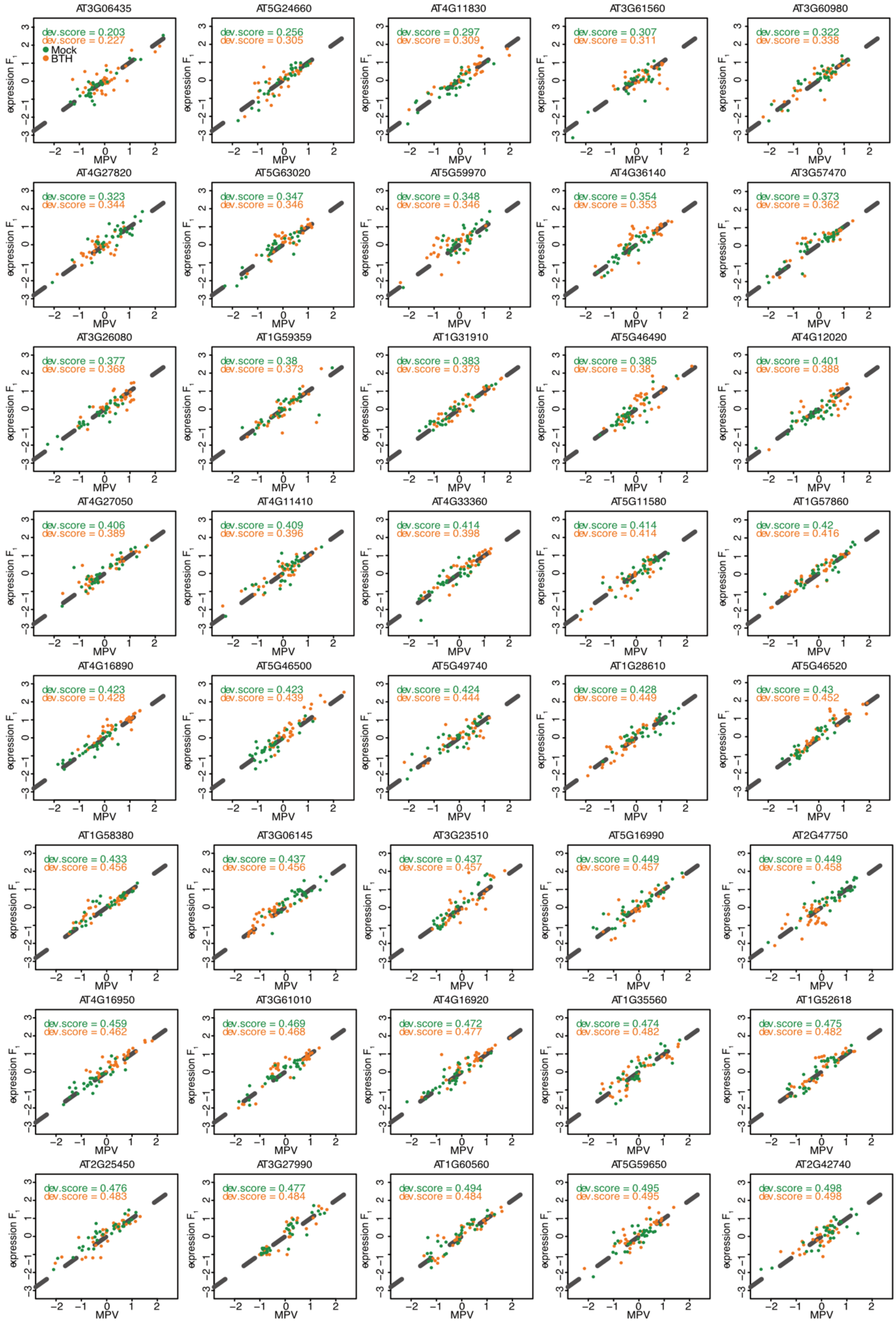
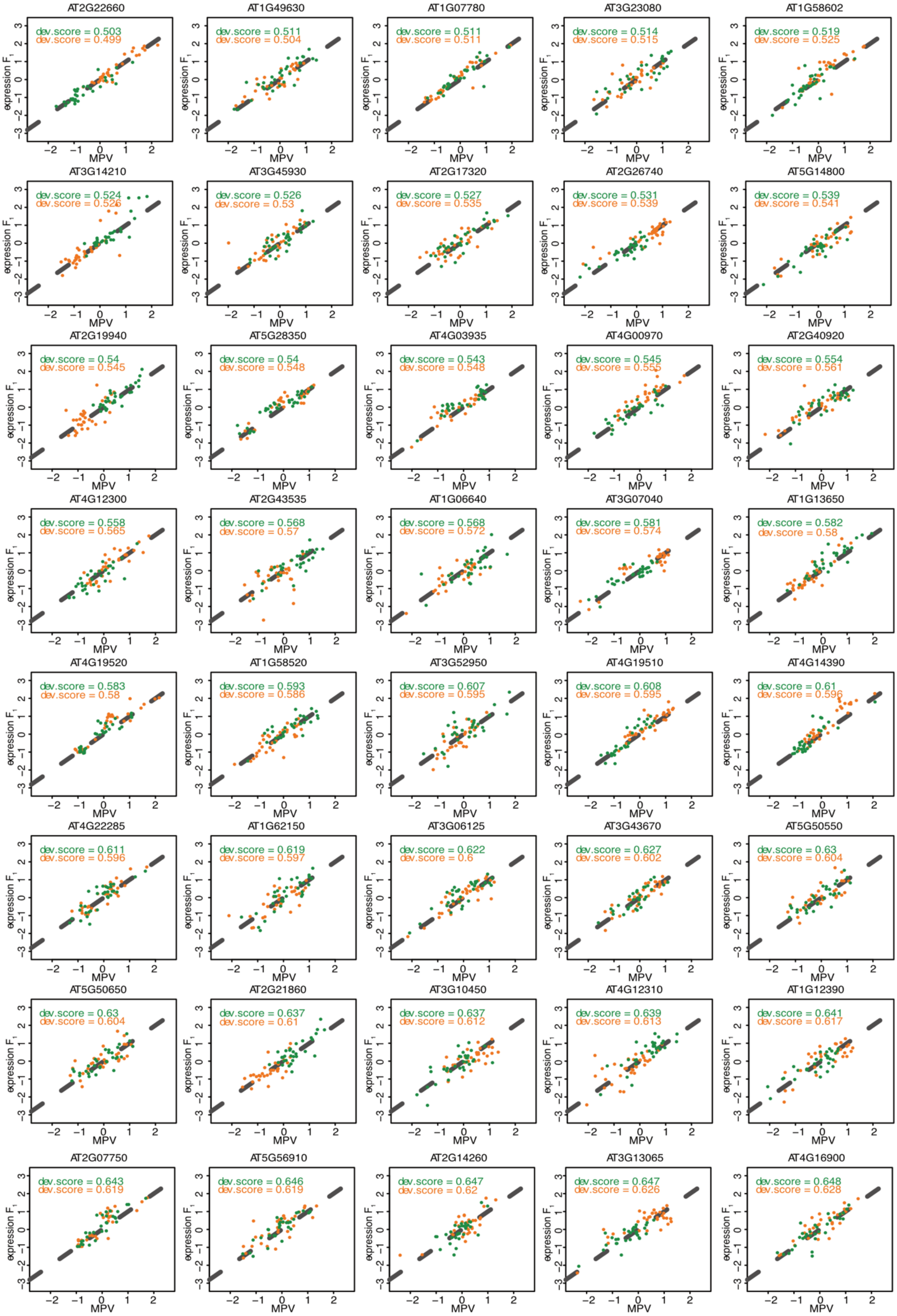
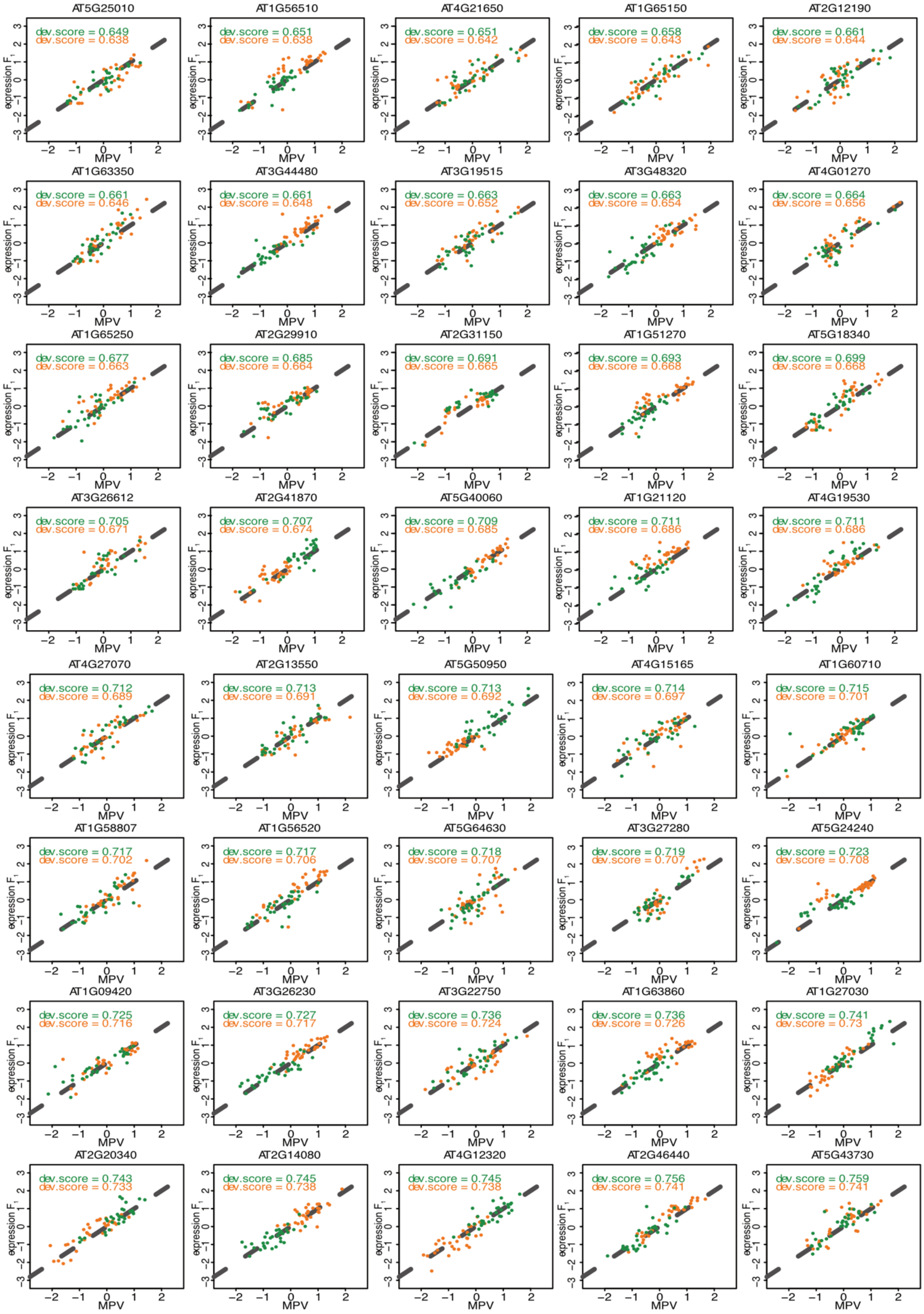
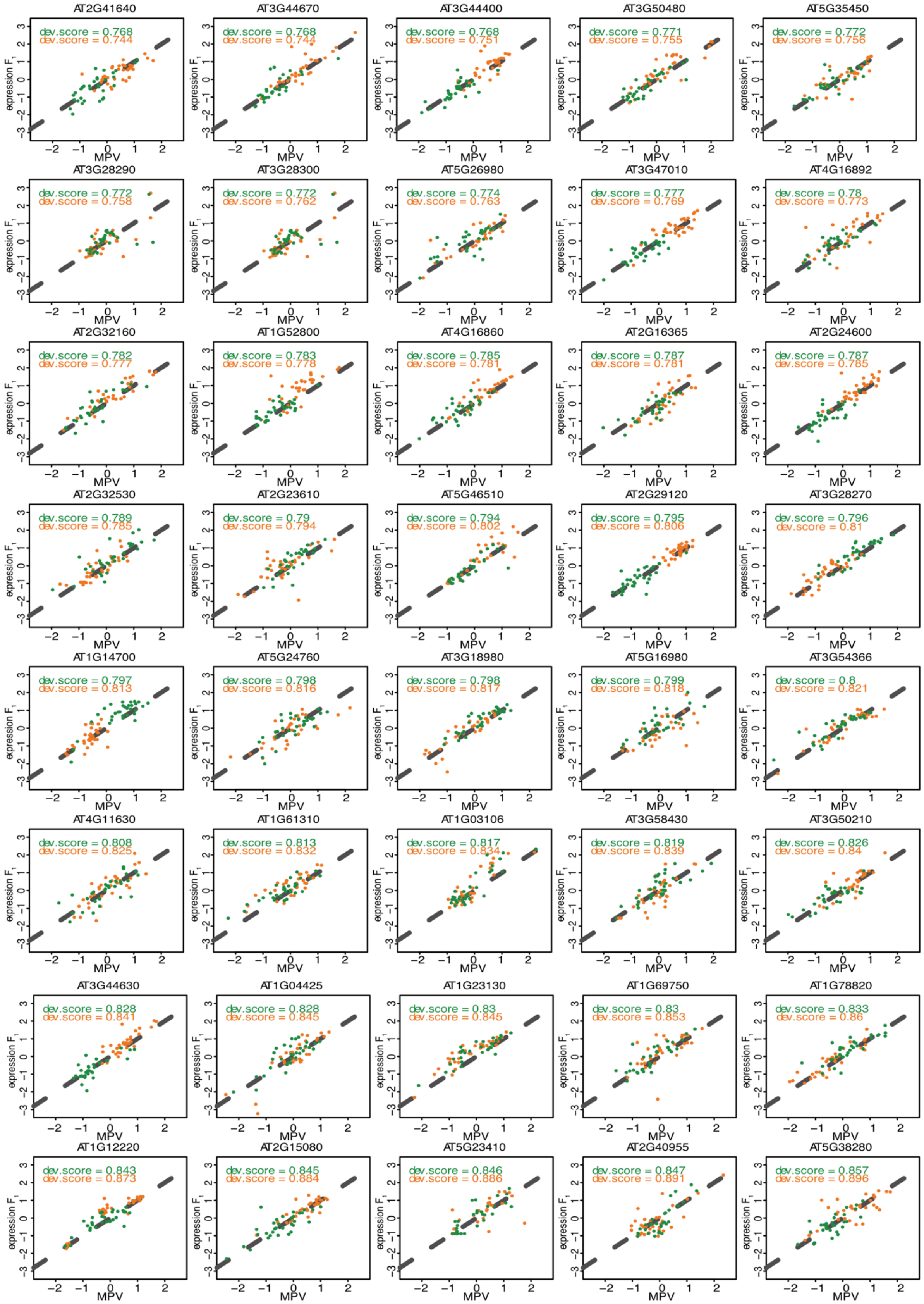


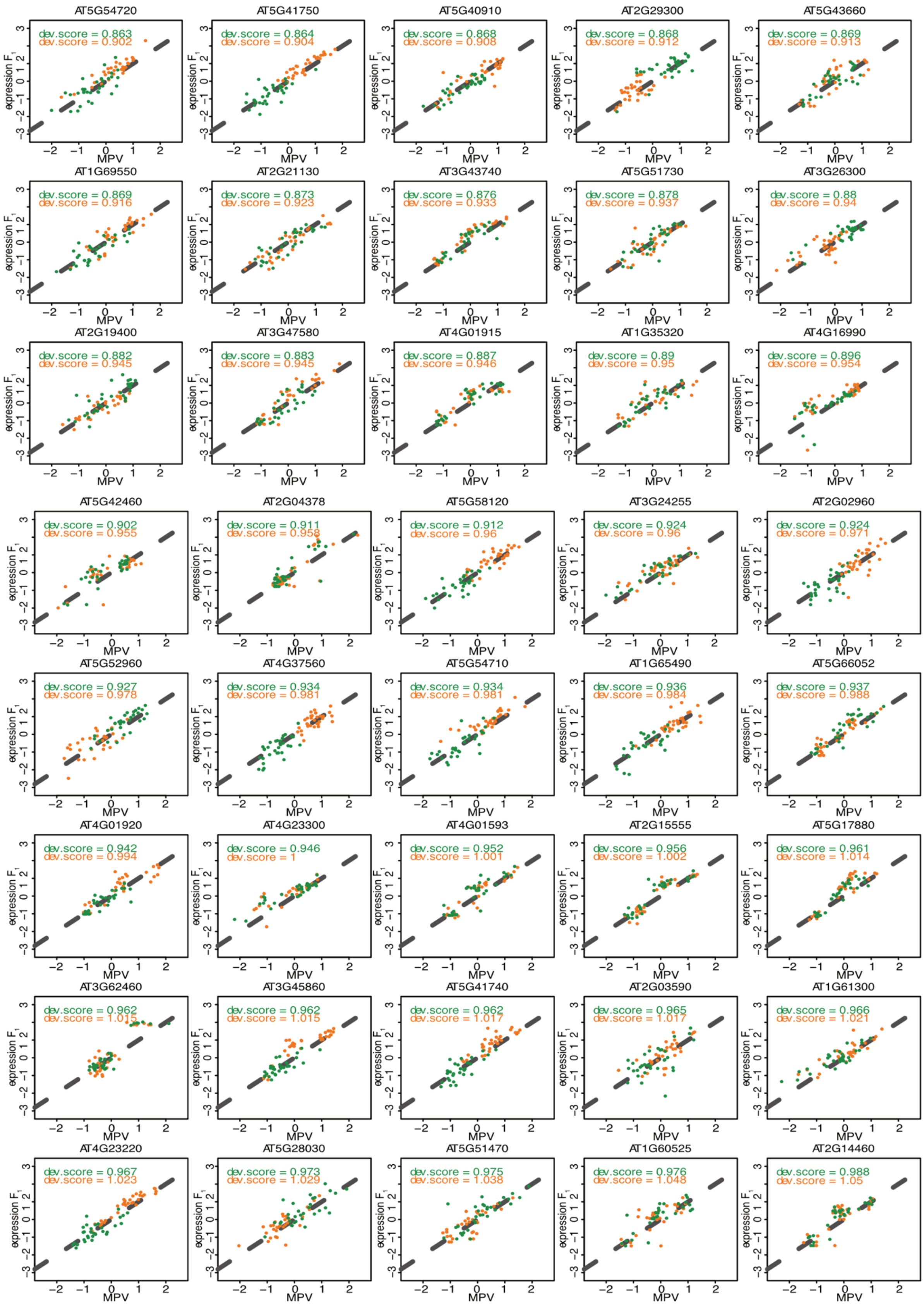
Fig. S4. Consistent and significant reduction in rosette area by BTH in both batches of the SHB2 experiment. Each tray was a treatment unit. Shown are boxplots of rosette area (mm^2) in each tray with mock (green) or BTH treatment (orange). Only trays with parent-hybrid trios further used for transcriptome analysis are included. The rosette area after treatment with BTH was reduced compared to mock treatment (2-way nested ANOVA, mock: 79.2 ± 3.0 mm^2, BTH treatment: 39.5 ± 4.2 mm^2; $p < 10^{-16}$).

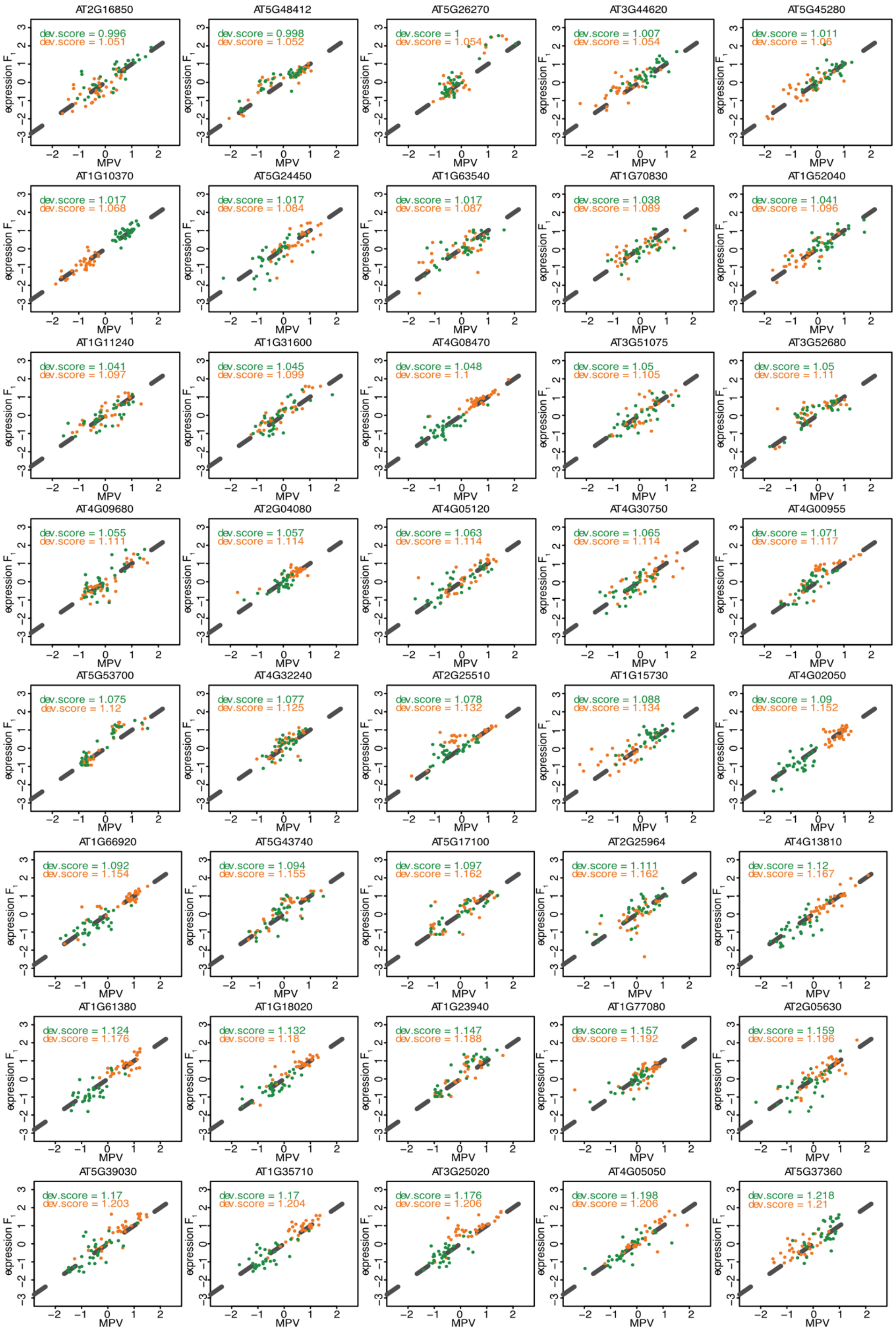

AT3G06435
dev.score = 0.203
dev.score = 0.227
Mock
BTH
AT5G24660
dev.score = 0.256
dev.score = 0.305
AT4G11830
dev.score = 0.297
dev.score = 0.309
AT3G61560
dev.score = 0.307
dev.score = 0.311
AT3G60980
dev.score = 0.322
dev.score = 0.338
AT4G27820
dev.score = 0.323
dev.score = 0.344
AT5G63020
dev.score = 0.347
dev.score = 0.346
AT5G59970
dev.score = 0.348
dev.score = 0.346
AT4G36140
dev.score = 0.354
dev.score = 0.353
AT3G57470
dev.score = 0.373
dev.score = 0.362
AT3G26080
dev.score = 0.377
dev.score = 0.368
AT1G59359
dev.score = 0.38
dev.score = 0.373
AT1G31910
dev.score = 0.383
dev.score = 0.379
AT5G46490
dev.score = 0.385
dev.score = 0.38
AT4G12020
dev.score = 0.401
dev.score = 0.388
AT4G27050
dev.score = 0.406
dev.score = 0.389
AT4G11410
dev.score = 0.409
dev.score = 0.396
AT4G33360
dev.score = 0.414
dev.score = 0.398
AT5G11580
dev.score = 0.414
dev.score = 0.414
AT1G57860
dev.score = 0.42
dev.score = 0.416
AT4G16890
dev.score = 0.423
dev.score = 0.428
AT5G46500
dev.score = 0.423
dev.score = 0.439
AT5G49740
dev.score = 0.424
dev.score = 0.444
AT1G28610
dev.score = 0.428
dev.score = 0.449
AT5G46520
dev.score = 0.43
dev.score = 0.452
AT1G58380
dev.score = 0.433
dev.score = 0.456
AT3G06145
dev.score = 0.437
dev.score = 0.456
AT3G23510
dev.score = 0.437
dev.score = 0.457
AT5G16990
dev.score = 0.449
dev.score = 0.457
AT2G47750
dev.score = 0.449
dev.score = 0.458
AT4G16950
dev.score = 0.459
dev.score = 0.462
AT3G61010
dev.score = 0.469
dev.score = 0.468
AT4G16920
dev.score = 0.472
dev.score = 0.477
AT1G35560
dev.score = 0.474
dev.score = 0.482
AT1G52618
dev.score = 0.475
dev.score = 0.482
AT2G25450
dev.score = 0.476
dev.score = 0.483
AT3G27990
dev.score = 0.477
dev.score = 0.484
AT1G60560
dev.score = 0.494
dev.score = 0.484
AT5G59650
dev.score = 0.495
dev.score = 0.495
AT2G42740
dev.score = 0.498
dev.score = 0.498
expression F$_1$
MPV
−3 −2 −1 0 1 2 3
−2 −1 0 1 2

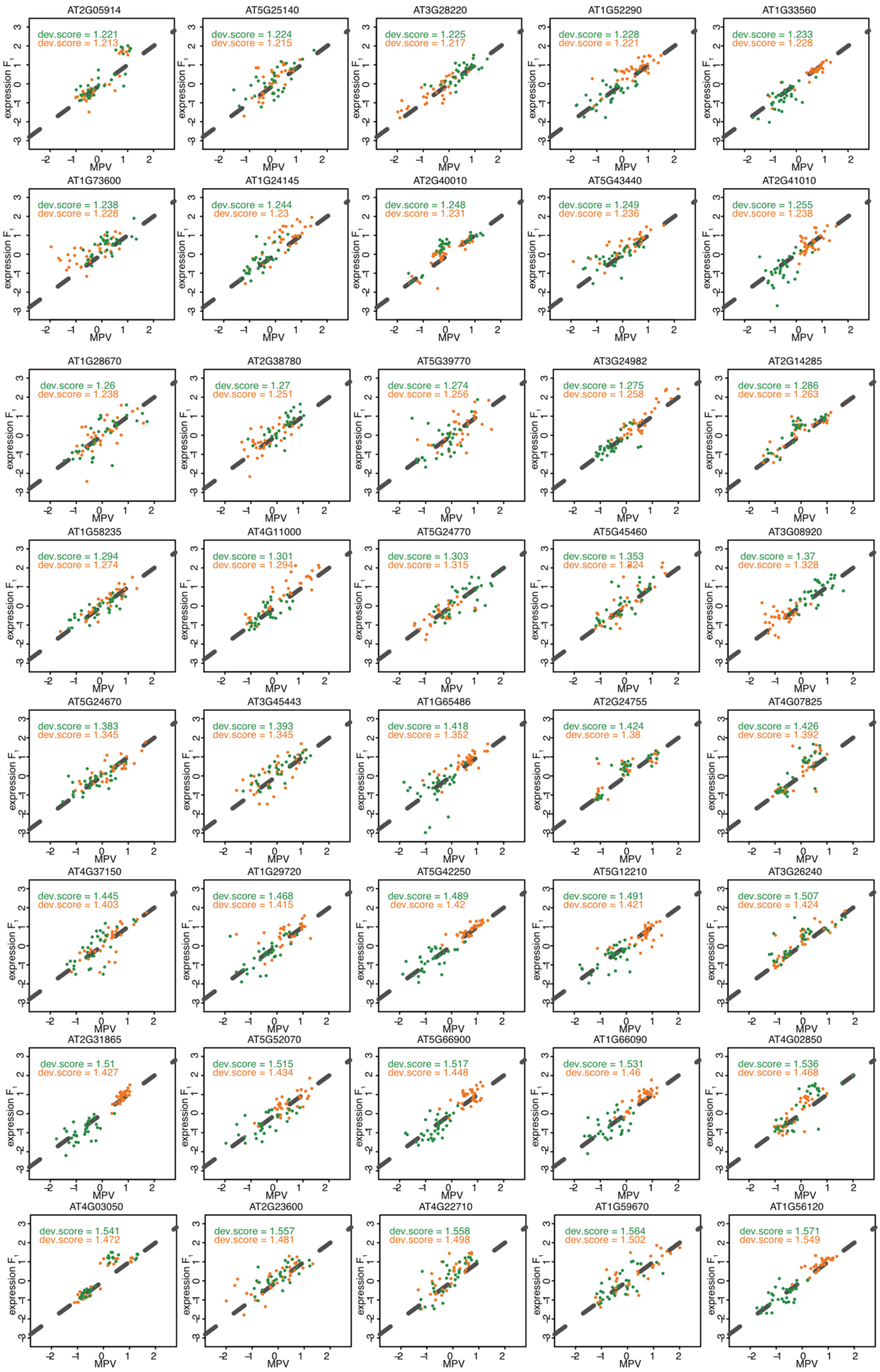

AT2G22660
dev.score = 0.503
dev.score = 0.499
AT1G49630
dev.score = 0.511
dev.score = 0.504
AT1G07780
dev.score = 0.511
dev.score = 0.511
AT3G23080
dev.score = 0.514
dev.score = 0.515
AT1G58602
dev.score = 0.519
dev.score = 0.525
AT3G14210
dev.score = 0.524
dev.score = 0.526
AT3G45930
dev.score = 0.526
dev.score = 0.53
AT2G17320
dev.score = 0.527
dev.score = 0.535
AT2G26740
dev.score = 0.531
dev.score = 0.539
AT5G14800
dev.score = 0.539
dev.score = 0.541
AT2G19940
dev.score = 0.54
dev.score = 0.545
AT5G28350
dev.score = 0.54
dev.score = 0.548
AT4G03935
dev.score = 0.543
dev.score = 0.548
AT4G00970
dev.score = 0.545
dev.score = 0.555
AT2G40920
dev.score = 0.554
dev.score = 0.561
AT4G12300
dev.score = 0.558
dev.score = 0.565
AT2G43535
dev.score = 0.568
dev.score = 0.57
AT1G06640
dev.score = 0.568
dev.score = 0.572
AT3G07040
dev.score = 0.581
dev.score = 0.574
AT1G13650
dev.score = 0.582
dev.score = 0.58
AT4G19520
dev.score = 0.583
dev.score = 0.58
AT1G58520
dev.score = 0.593
dev.score = 0.586
AT3G52950
dev.score = 0.607
dev.score = 0.595
AT4G19510
dev.score = 0.608
dev.score = 0.595
AT4G14390
dev.score = 0.61
dev.score = 0.596
AT4G22285
dev.score = 0.611
dev.score = 0.596
AT1G62150
dev.score = 0.619
dev.score = 0.597
AT3G06125
dev.score = 0.622
dev.score = 0.6
AT3G43670
dev.score = 0.627
dev.score = 0.602
AT5G50550
dev.score = 0.63
dev.score = 0.604
AT5G50650
dev.score = 0.63
dev.score = 0.604
AT2G21860
dev.score = 0.637
dev.score = 0.61
AT3G10450
dev.score = 0.637
dev.score = 0.612
AT4G12310
dev.score = 0.639
dev.score = 0.613
AT1G12390
dev.score = 0.641
dev.score = 0.617
AT2G07750
dev.score = 0.643
dev.score = 0.619
AT5G56910
dev.score = 0.646
dev.score = 0.619
AT2G14260
dev.score = 0.647
dev.score = 0.62
AT3G13065
dev.score = 0.647
dev.score = 0.626
AT4G16900
dev.score = 0.648
dev.score = 0.628
expression F$_1$
MPV
−3 −2 −1 0 1 2 3
−2 −1 0 1 2


AT5G25010
dev.score = 0.649
dev.score = 0.638
AT1G56510
dev.score = 0.651
dev.score = 0.638
AT4G21650
dev.score = 0.651
dev.score = 0.642
AT1G65150
dev.score = 0.658
dev.score = 0.643
AT2G12190
dev.score = 0.661
dev.score = 0.644
AT1G63350
dev.score = 0.661
dev.score = 0.646
AT3G44480
dev.score = 0.661
dev.score = 0.648
AT3G19515
dev.score = 0.663
dev.score = 0.652
AT3G48320
dev.score = 0.663
dev.score = 0.654
AT4G01270
dev.score = 0.664
dev.score = 0.656
AT1G65250
dev.score = 0.677
dev.score = 0.663
AT2G29910
dev.score = 0.685
dev.score = 0.664
AT2G31150
dev.score = 0.691
dev.score = 0.665
AT1G51270
dev.score = 0.693
dev.score = 0.668
AT5G18340
dev.score = 0.699
dev.score = 0.668
AT3G26612
dev.score = 0.705
dev.score = 0.671
AT2G41870
dev.score = 0.707
dev.score = 0.674
AT5G40060
dev.score = 0.709
dev.score = 0.685
AT1G21120
dev.score = 0.711
dev.score = 0.686
AT4G19530
dev.score = 0.711
dev.score = 0.686
AT4G27070
dev.score = 0.712
dev.score = 0.689
AT2G13550
dev.score = 0.713
dev.score = 0.691
AT5G50950
dev.score = 0.713
dev.score = 0.692
AT4G15165
dev.score = 0.714
dev.score = 0.697
AT1G60710
dev.score = 0.715
dev.score = 0.701
AT1G58807
dev.score = 0.717
dev.score = 0.702
AT1G56520
dev.score = 0.717
dev.score = 0.706
AT5G64630
dev.score = 0.718
dev.score = 0.707
AT3G27280
dev.score = 0.719
dev.score = 0.707
AT5G24240
dev.score = 0.723
dev.score = 0.708
AT1G09420
dev.score = 0.725
dev.score = 0.716
AT3G26230
dev.score = 0.727
dev.score = 0.717
AT3G22750
dev.score = 0.736
dev.score = 0.724
AT1G63860
dev.score = 0.736
dev.score = 0.726
AT1G27030
dev.score = 0.741
dev.score = 0.73
AT2G20340
dev.score = 0.743
dev.score = 0.733
AT2G14080
dev.score = 0.745
dev.score = 0.738
AT4G12320
dev.score = 0.745
dev.score = 0.738
AT2G46440
dev.score = 0.756
dev.score = 0.741
AT5G43730
dev.score = 0.759
dev.score = 0.741
expression F_1
MPV
−3
−2
−1
0
1
2
3


AT2G41640
dev.score = 0.768
dev.score = 0.744
AT3G44670
dev.score = 0.768
dev.score = 0.744
AT3G44400
dev.score = 0.768
dev.score = 0.751
AT3G50480
dev.score = 0.771
dev.score = 0.755
AT5G35450
dev.score = 0.772
dev.score = 0.756
AT3G28290
dev.score = 0.772
dev.score = 0.758
AT3G28300
dev.score = 0.772
dev.score = 0.762
AT5G26980
dev.score = 0.774
dev.score = 0.763
AT3G47010
dev.score = 0.777
dev.score = 0.769
AT4G16892
dev.score = 0.78
dev.score = 0.773
AT2G32160
dev.score = 0.782
dev.score = 0.777
AT1G52800
dev.score = 0.783
dev.score = 0.778
AT4G16860
dev.score = 0.785
dev.score = 0.781
AT2G16365
dev.score = 0.787
dev.score = 0.781
AT2G24600
dev.score = 0.787
dev.score = 0.785
AT2G32530
dev.score = 0.789
dev.score = 0.785
AT2G23610
dev.score = 0.79
dev.score = 0.794
AT5G46510
dev.score = 0.794
dev.score = 0.802
AT2G29120
dev.score = 0.795
dev.score = 0.806
AT3G28270
dev.score = 0.796
dev.score = 0.81
AT1G14700
dev.score = 0.797
dev.score = 0.813
AT5G24760
dev.score = 0.798
dev.score = 0.816
AT3G18980
dev.score = 0.798
dev.score = 0.817
AT5G16980
dev.score = 0.799
dev.score = 0.818
AT3G54366
dev.score = 0.8
dev.score = 0.821
AT4G11630
dev.score = 0.808
dev.score = 0.825
AT1G61310
dev.score = 0.813
dev.score = 0.832
AT1G03106
dev.score = 0.817
dev.score = 0.834
AT3G58430
dev.score = 0.819
dev.score = 0.839
AT3G50210
dev.score = 0.826
dev.score = 0.84
AT3G44630
dev.score = 0.828
dev.score = 0.841
AT1G04425
dev.score = 0.828
dev.score = 0.845
AT1G23130
dev.score = 0.83
dev.score = 0.845
AT1G69750
dev.score = 0.83
dev.score = 0.853
AT1G78820
dev.score = 0.833
dev.score = 0.86
AT1G12220
dev.score = 0.843
dev.score = 0.873
AT2G15080
dev.score = 0.845
dev.score = 0.884
AT5G23410
dev.score = 0.846
dev.score = 0.886
AT2G40955
dev.score = 0.847
dev.score = 0.891
AT5G38280
dev.score = 0.857
dev.score = 0.896
expression F_1
MPV
−3 −2 −1 0 1 2 3
−2 −1 0 1 2


AT5G54720
dev.score = 0.863
dev.score = 0.902
AT5G41750
dev.score = 0.864
dev.score = 0.904
AT5G40910
dev.score = 0.868
dev.score = 0.908
AT2G29300
dev.score = 0.868
dev.score = 0.912
AT5G43660
dev.score = 0.869
dev.score = 0.913
AT1G69550
dev.score = 0.869
dev.score = 0.916
AT2G21130
dev.score = 0.873
dev.score = 0.923
AT3G43740
dev.score = 0.876
dev.score = 0.933
AT5G51730
dev.score = 0.878
dev.score = 0.937
AT3G26300
dev.score = 0.88
dev.score = 0.94
AT2G19400
dev.score = 0.882
dev.score = 0.945
AT3G47580
dev.score = 0.883
dev.score = 0.945
AT4G01915
dev.score = 0.887
dev.score = 0.946
AT1G35320
dev.score = 0.89
dev.score = 0.95
AT4G16990
dev.score = 0.896
dev.score = 0.954
AT5G42460
dev.score = 0.902
dev.score = 0.955
AT2G04378
dev.score = 0.911
dev.score = 0.958
AT5G58120
dev.score = 0.912
dev.score = 0.96
AT3G24255
dev.score = 0.924
dev.score = 0.96
AT2G02960
dev.score = 0.924
dev.score = 0.971
AT5G52960
dev.score = 0.927
dev.score = 0.978
AT4G37560
dev.score = 0.934
dev.score = 0.981
AT5G54710
dev.score = 0.934
dev.score = 0.981
AT1G65490
dev.score = 0.936
dev.score = 0.984
AT5G66052
dev.score = 0.937
dev.score = 0.988
AT4G01920
dev.score = 0.942
dev.score = 0.994
AT4G23300
dev.score = 0.946
dev.score = 1
AT4G01593
dev.score = 0.952
dev.score = 1.001
AT2G15555
dev.score = 0.956
dev.score = 1.002
AT5G17880
dev.score = 0.961
dev.score = 1.014
AT3G62460
dev.score = 0.962
dev.score = 1.015
AT3G45860
dev.score = 0.962
dev.score = 1.015
AT5G41740
dev.score = 0.962
dev.score = 1.017
AT2G03590
dev.score = 0.965
dev.score = 1.017
AT1G61300
dev.score = 0.966
dev.score = 1.021
AT4G23220
dev.score = 0.967
dev.score = 1.023
AT5G28030
dev.score = 0.973
dev.score = 1.029
AT5G51470
dev.score = 0.975
dev.score = 1.038
AT1G60525
dev.score = 0.976
dev.score = 1.048
AT2G14460
dev.score = 0.988
dev.score = 1.05
expression F1
MPV
−3 −2 −1 0 1 2 3
−2 −1 0 1 2


AT2G16850
dev.score = 0.996
dev.score = 1.051
AT5G48412
dev.score = 0.998
dev.score = 1.052
AT5G26270
dev.score = 1
dev.score = 1.054
AT3G44620
dev.score = 1.007
dev.score = 1.054
AT5G45280
dev.score = 1.011
dev.score = 1.06
AT1G10370
dev.score = 1.017
dev.score = 1.068
AT5G24450
dev.score = 1.017
dev.score = 1.084
AT1G63540
dev.score = 1.017
dev.score = 1.087
AT1G70830
dev.score = 1.038
dev.score = 1.089
AT1G52040
dev.score = 1.041
dev.score = 1.096
AT1G11240
dev.score = 1.041
dev.score = 1.097
AT1G31600
dev.score = 1.045
dev.score = 1.099
AT4G08470
dev.score = 1.048
dev.score = 1.1
AT3G51075
dev.score = 1.05
dev.score = 1.105
AT3G52680
dev.score = 1.05
dev.score = 1.11
AT4G09680
dev.score = 1.055
dev.score = 1.111
AT2G04080
dev.score = 1.057
dev.score = 1.114
AT4G05120
dev.score = 1.063
dev.score = 1.114
AT4G30750
dev.score = 1.065
dev.score = 1.114
AT4G00955
dev.score = 1.071
dev.score = 1.117
AT5G53700
dev.score = 1.075
dev.score = 1.12
AT4G32240
dev.score = 1.077
dev.score = 1.125
AT2G25510
dev.score = 1.078
dev.score = 1.132
AT1G15730
dev.score = 1.088
dev.score = 1.134
AT4G02050
dev.score = 1.09
dev.score = 1.152
AT1G66920
dev.score = 1.092
dev.score = 1.154
AT5G43740
dev.score = 1.094
dev.score = 1.155
AT5G17100
dev.score = 1.097
dev.score = 1.162
AT2G25964
dev.score = 1.111
dev.score = 1.162
AT4G13810
dev.score = 1.12
dev.score = 1.167
AT1G61380
dev.score = 1.124
dev.score = 1.176
AT1G18020
dev.score = 1.132
dev.score = 1.18
AT1G23940
dev.score = 1.147
dev.score = 1.188
AT1G77080
dev.score = 1.157
dev.score = 1.192
AT2G05630
dev.score = 1.159
dev.score = 1.196
AT5G39030
dev.score = 1.17
dev.score = 1.203
AT1G35710
dev.score = 1.17
dev.score = 1.204
AT3G25020
dev.score = 1.176
dev.score = 1.206
AT4G05050
dev.score = 1.198
dev.score = 1.206
AT5G37360
dev.score = 1.218
dev.score = 1.21
expression F$_1$
MPV
−3 −2 −1 0 1 2 3
−2 −1 0 1 2


AT2G05914
dev.score = 1.221
dev.score = 1.213
AT5G25140
dev.score = 1.224
dev.score = 1.215
AT3G28220
dev.score = 1.225
dev.score = 1.217
AT1G52290
dev.score = 1.228
dev.score = 1.221
AT1G33560
dev.score = 1.233
dev.score = 1.228
AT1G73600
dev.score = 1.238
dev.score = 1.228
AT1G24145
dev.score = 1.244
dev.score = 1.23
AT2G40010
dev.score = 1.248
dev.score = 1.231
AT5G43440
dev.score = 1.249
dev.score = 1.236
AT2G41010
dev.score = 1.255
dev.score = 1.238
AT1G28670
dev.score = 1.26
dev.score = 1.238
AT2G38780
dev.score = 1.27
dev.score = 1.251
AT5G39770
dev.score = 1.274
dev.score = 1.256
AT3G24982
dev.score = 1.275
dev.score = 1.258
AT2G14285
dev.score = 1.286
dev.score = 1.263
AT1G58235
dev.score = 1.294
dev.score = 1.274
AT4G11000
dev.score = 1.301
dev.score = 1.294
AT5G24770
dev.score = 1.303
dev.score = 1.315
AT5G45460
dev.score = 1.353
dev.score = 1.324
AT3G08920
dev.score = 1.37
dev.score = 1.328
AT5G24670
dev.score = 1.383
dev.score = 1.345
AT3G45443
dev.score = 1.393
dev.score = 1.345
AT1G65486
dev.score = 1.418
dev.score = 1.352
AT2G24755
dev.score = 1.424
dev.score = 1.38
AT4G07825
dev.score = 1.426
dev.score = 1.392
AT4G37150
dev.score = 1.445
dev.score = 1.403
AT1G29720
dev.score = 1.468
dev.score = 1.415
AT5G42250
dev.score = 1.489
dev.score = 1.42
AT5G12210
dev.score = 1.491
dev.score = 1.421
AT3G26240
dev.score = 1.507
dev.score = 1.424
AT2G31865
dev.score = 1.51
dev.score = 1.427
AT5G52070
dev.score = 1.515
dev.score = 1.434
AT5G66900
dev.score = 1.517
dev.score = 1.448
AT1G66090
dev.score = 1.531
dev.score = 1.46
AT4G02850
dev.score = 1.536
dev.score = 1.468
AT4G03050
dev.score = 1.541
dev.score = 1.472
AT2G23600
dev.score = 1.557
dev.score = 1.481
AT4G22710
dev.score = 1.558
dev.score = 1.498
AT1G59670
dev.score = 1.564
dev.score = 1.502
AT1G56120
dev.score = 1.571
dev.score = 1.549
expression F1
MPV
-3 -2 -1 0 1 2 3
-2 -1 0 1 2

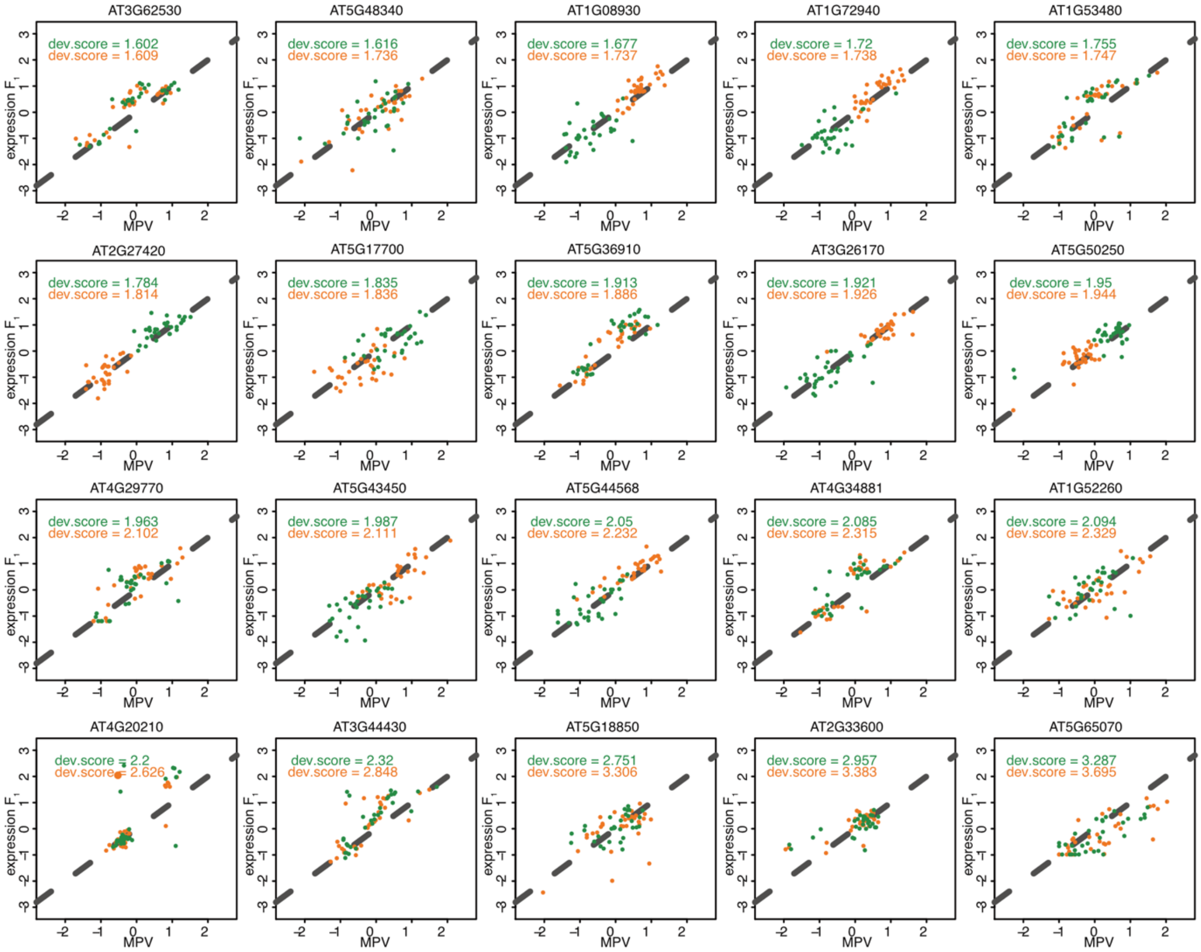


Fig. S5. Hypothetical MPV expression from all trios reliably predicts expression level in F_1 hybrids for the 300 additive genes called in both SHB1 and SHB2 datasets. Mean expression values of replicates for all F_1 hybrids are plotted against the corresponding means of MPV for mock (green) and BTH treatments (orange). The dashed line represents unity of y=x, tight scatter of data points along which indicates a good prediction of MPV to F_1 expression. Graphs are sorted according to their deviation score (Methods).

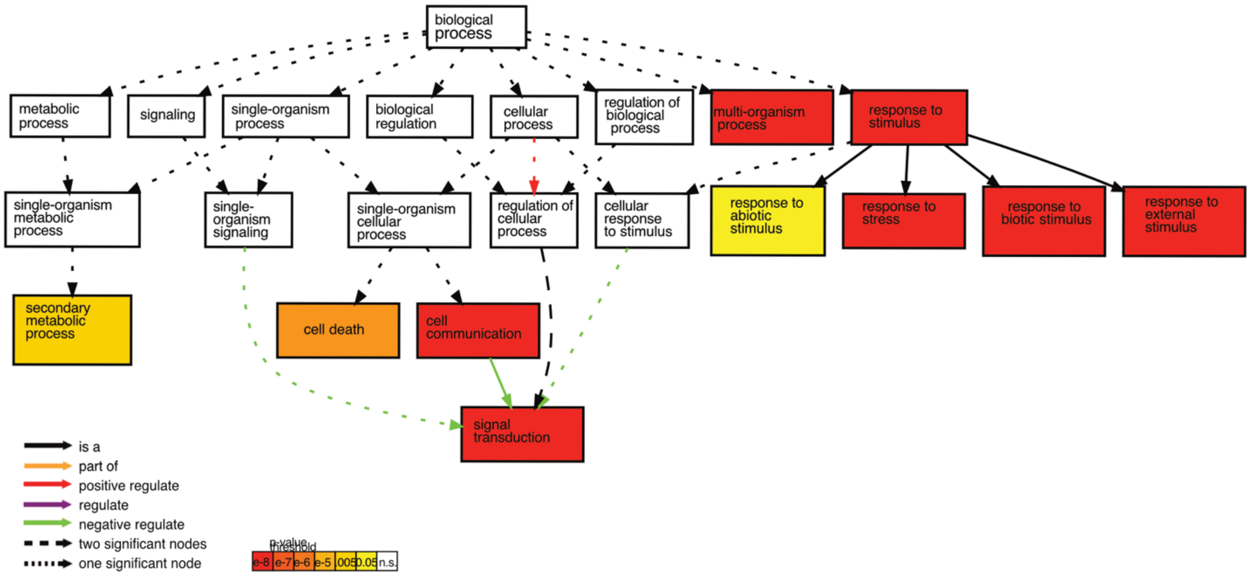

biological process
metabolic process
signaling
single-organism process
biological regulation
cellular process
regulation of biological process
multi-organism process
response to stimulus
single-organism metabolic process
single-organism signaling
single-organism cellular process
regulation of cellular process
cellular response to stimulus
response to abiotic stimulus
response to stress
response to biotic stimulus
response to external stimulus
secondary metabolic process
cell death
cell communication
signal transduction
is a
part of
positive regulate
regulate
negative regulate
two significant nodes
one significant node
p-value threshold
e-8
e-7
e-6
e-5
0.005
0.05
n.s.

Fig. S6. Additive genes in SHB2 (n=901) are enriched for Gene Ontology terms of response to stress and (biotic) stimuli, cell death, and secondary metabolic process.

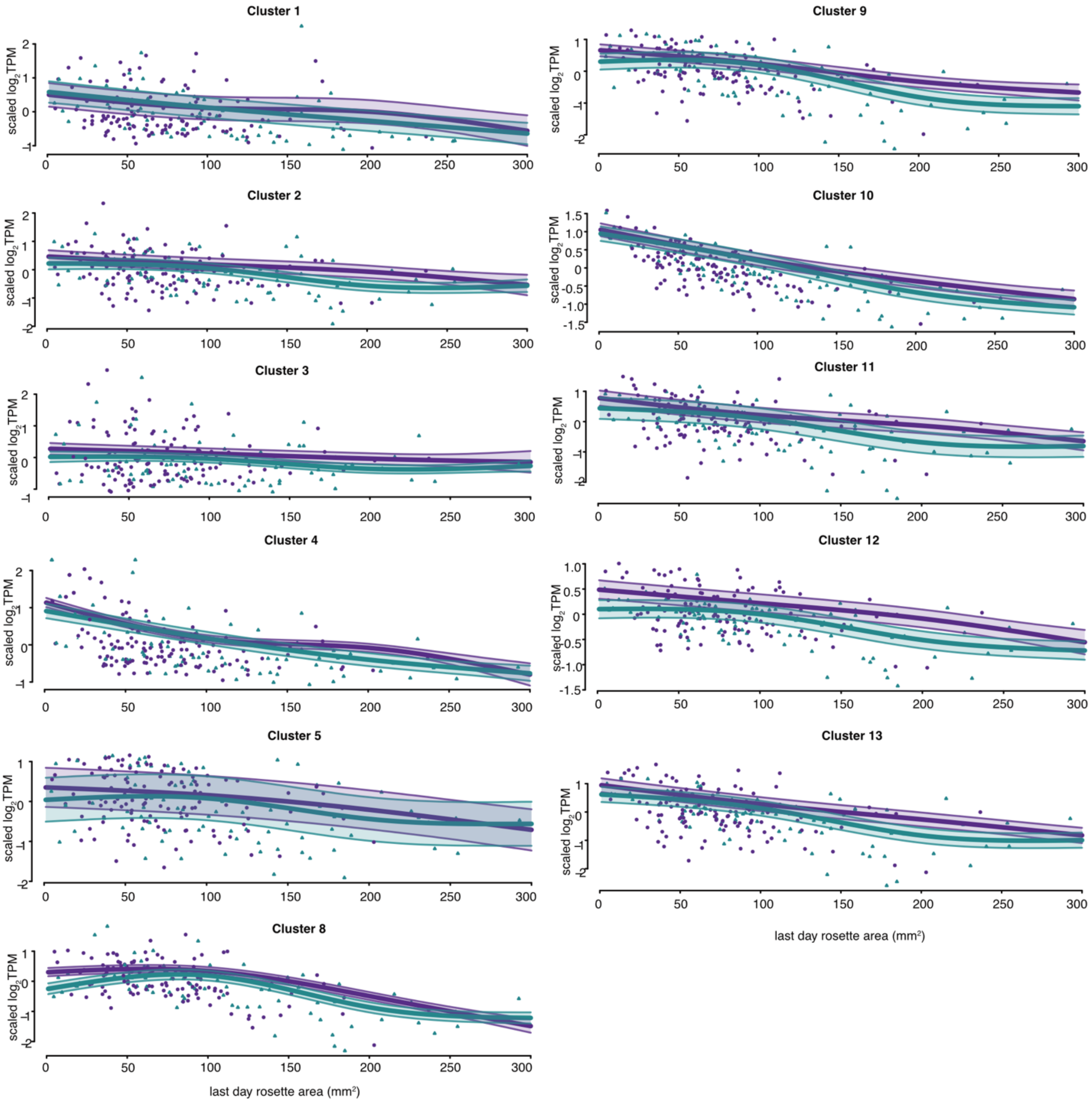


Fig. S7. LMM spline fitting of cluster mean expression levels with 95% Bayesian credible intervals for non-additive gene clusters not shown in Fig. 3. Mean expression level across all genes within a cluster for inbred parents (purple dots) and F_1 hybrids (turquoise dots) against last day rosette area (mm^2) were plotted. Clusters 1-5 showed little rosette size-expression level association, while clusters 8-13 showed monotonic decrease of cluster mean expression level with increased plant size. While F_1 hybrids exhibited the same trend as the inbred parents, the mean expression levels are consistently lower in F_1 hybrids across the entire rosette size range for cluster 8-13.

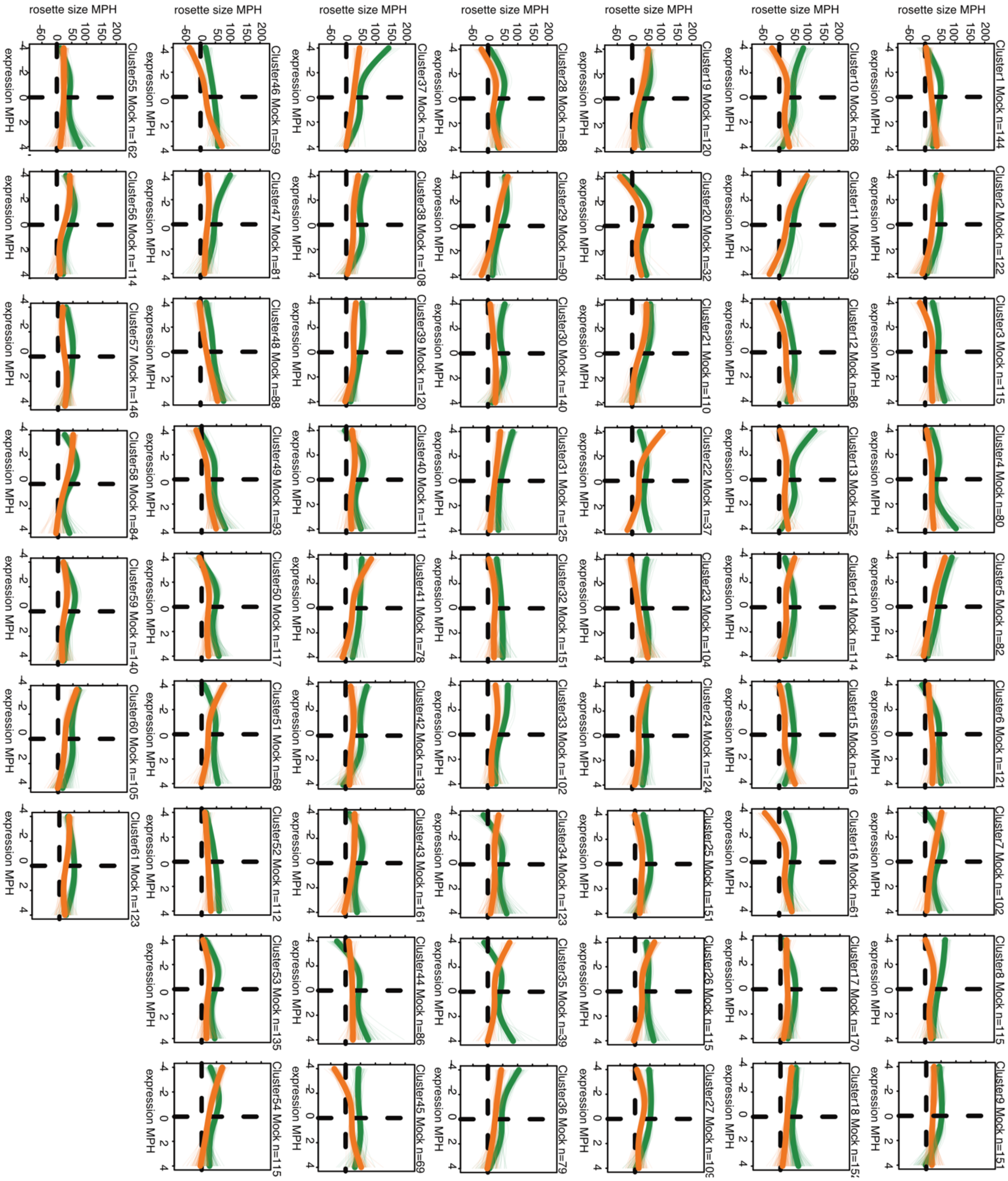


Fig. S8. BTH-responsive genes sorted into 61 clusters based on spline regression of rosette area MPH (mm^2) to their expression MPH across all trios. Shown are a graphic representation of the clusters. Each thin line represents a gene, and the thick line represents cluster mean (green: mock, orange: BTH). The 61 clusters were subsequently sorted into 12 general categories based on the regression trends in mock and BTH treatments.

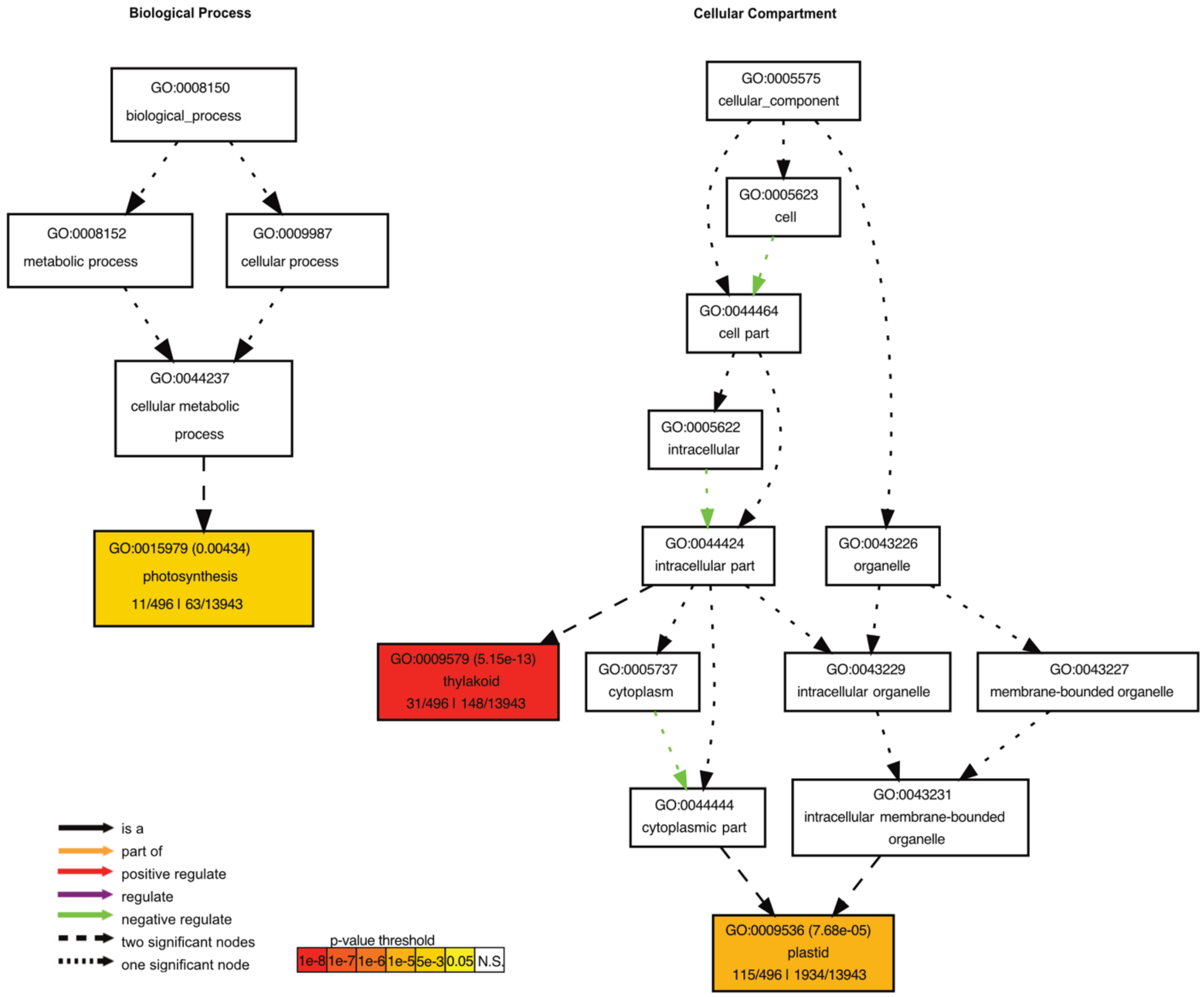

Biological Process
GO:0008150
biological_process
GO:0008152
metabolic process
GO:0009987
cellular process
GO:0044237
cellular metabolic process
GO:0015979 (0.00434)
photosynthesis
11/496 I 63/13943
Cellular Compartment
GO:0005575
cellular_component
GO:0005623
cell
GO:0044464
cell part
GO:0005622
intracellular
GO:0044424
intracellular part
GO:0043226
organelle
GO:0009579 (5.15e-13)
thylakoid
31/496 I 148/13943
GO:0005737
cytoplasm
GO:0043229
intracellular organelle
GO:0043227
membrane-bounded organelle
GO:0044444
cytoplasmic part
GO:0043231
intracellular membrane-bounded organelle
GO:0009536 (7.68e-05)
plastid
115/496 I 1934/13943
is a
part of
positive regulate
regulate
negative regulate
two significant nodes
one significant node
p-value threshold
1e-8
1e-7
1e-6
1e-5
5e-3
0.05
N.S.

Fig. S9. Positive genes are enriched for genes encoding thykaloid-localized proteins that are involved in photosynthetic process. Fisher's Exact Test was used against the background list of filtered expressed genes in SHB2 dataset, using plant GOslim based on TAIR10 annotation.

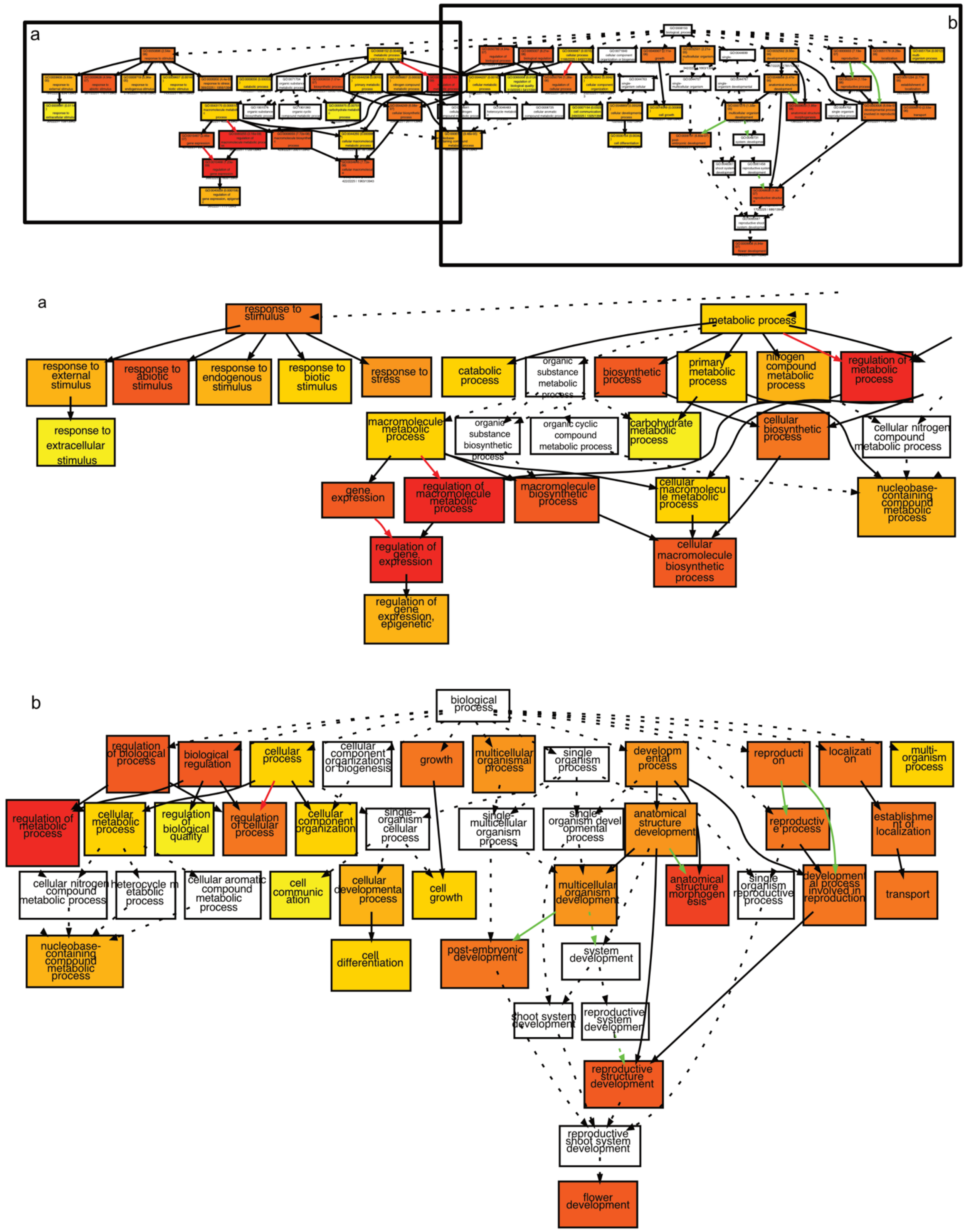

a
b
a
response to stimulus
metabolic process
response to external stimulus
response to abiotic stimulus
response to endogenous stimulus
response to biotic stimulus
response to stress
catabolic process
organic substance metabolic process
biosynthetic process
primary metabolic process
nitrogen compound metabolic process
regulation of metabolic process
response to extracellular stimulus
macromolecule metabolic process
organic substance biosynthetic process
organic cyclic compound metabolic process
carbohydrate metabolic process
cellular biosynthetic process
cellular nitrogen compound metabolic process
gene expression
regulation of macromolecule metabolic process
macromolecule biosynthetic process
cellular macromolecule metabolic process
nucleobase-containing compound metabolic process
regulation of gene expression
cellular macromolecule biosynthetic process
regulation of gene expression, epigenetic
b
biological process
regulation of biological process
biological regulation
cellular process
cellular component organization or biogenesis
growth
multicellular organismal process
single organism process
developmental process
reproduction
localization
multi-organism process
regulation of metabolic process
cellular metabolic process
regulation of biological quality
regulation of cellular process
cellular component organization
single-organism cellular process
single-multicellular organism process
single-organism developmental process
anatomical structure development
reproductive process
establishment of localization
cellular nitrogen compound metabolic process
heterocycle metabolic process
cellular aromatic compound metabolic process
cell communication
cellular developmental process
cell growth
multicellular organism development
anatomical structure morphogenesis
single organism reproductive process
developmental process involved in reproduction
transport
nucleobase-containing compound metabolic process
cell differentiation
post-embryonic development
system development
shoot system development
reproductive system development
reproductive structure development
reproductive shoot system development
flower development

Fig. S10. Enlarged rendering of Fig. 4D, GO enrichment for the negative genes. The diagram was into two parts (a and b) for improved legibility.