The Footprint of Polygenic Adaptation on Stress-Responsive *Cis*-Regulatory Divergence in the *Arabidopsis Genus*

Fei He,[1,2] Agustin L. Arce,[1] Gregor Schmitz,[1] Maarten Koornneef,[3] Polina Novikova,[4] Andreas Beyer,[2] and Juliette de Meaux*[,1]

[1]Institute of Botany, Biozentrum, University of Cologne, Cologne, Germany

[2]CECAD, University of Cologne, Cologne, Germany

[3]Max Planck Institute for Plant Breeding Research, Cologne, Germany

[4]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna Biocenter, Vienna, Austria

***Corresponding author:** E-mail: jdemeaux@uni-koeln.de.

Associate editor: John Parsch

Abstract

Adaptation of a complex trait often requires the accumulation of many modifications to finely tune its underpinning molecular components to novel environmental requirements. The investigation of *cis*-acting regulatory modifications can be used to pinpoint molecular systems partaking in such complex adaptations. Here, we identify *cis*-acting modifications with the help of an interspecific crossing scheme designed to distinguish modifications derived in each of the two sister species, *Arabidopsis halleri* and *A. lyrata*. Allele-specific expression levels were assessed in three environmental conditions chosen to reflect interspecific ecological differences: cold exposure, dehydration, and standard conditions. The functions described by Gene Ontology categories enriched in *cis*-acting mutations are markedly different in *A. halleri* and *A. lyrata*, suggesting that polygenic adaptation reshaped distinct polygenic molecular functions in the two species. In the *A. halleri* lineage, an excess of *cis*-acting changes affecting metal transport and homeostasis was observed, confirming that the well-known heavy metal tolerance of this species is the result of polygenic selection. In *A. lyrata*, we find a marked excess of *cis*-acting changes among genes showing a transcriptional response to cold stress in the outgroup species *A. thaliana*. The adaptive relevance of these changes will have to be validated. We finally observed that polygenic molecular functions enriched in derived *cis*-acting changes are more constrained at the amino acid level. Using the distribution of *cis*-acting variation to tackle the polygenic basis of adaptation thus reveals the contribution of mutations of small effect to Darwinian adaptation.

***Key words:* *cis*-regulation, transcriptome, polygenic selection, *Arabidopsis*, cold stress, drought stress.**

Article

Introduction

Understanding the molecular mechanisms that allow phenotypic diversification in response to natural selection is one of the major challenges of modern biology (Dean and Thornton 2007; Barrett and Hoekstra 2011). While a majority of the changes at the DNA level are thought to result from neutral processes (Kimura 1968; Lynch 2007), natural selection is considered a major force behind the emergence of phenotypes tightly tuned to the challenges imposed by the environments.

The theory of adaptation predicts that initial adaptive steps toward a new fitness optimum will often be achieved through the quick fixation of a (few) mutation(s) of large effect (Orr 2002). Supporting this hypothesis, the few adaptive mutations that have been tracked at the molecular level commonly show a large effect on the phenotype (Colosimo et al. 2005; Clark et al. 2006; Steiner et al. 2007; Linnen et al. 2009; Reed et al. 2011; Prasad et al. 2012).

Mutations of small effect are more difficult to track at the molecular level in the lab, but they are also necessary to reach the fitness optimum. Simulations and experimental work have shown that when adaptation occurs in a slowly changing environment, the adaptive path may even exclude large effect mutations and only rely on mutations of small effect (Collins et al. 2007; Collins and de Meaux 2009; Kopp and Hermisson 2009; Yeaman 2015). Natural variation in fitness is often detected in natural populations (Kingsolver et al. 2001) and positive selection on standing variation is weak but pervasive (Morrissey and Hadfield 2012). As a result, many natural systems seem to display no limit to the response to selection (Barton and Keightley 2002). These observations are in line with Fisher's infinitesimal model, a model validated for example by the successful prediction of maize hybrid performance based on the allelic composition of the parents (Bulmer 1985; Hill 2009; Riedelsheimer et al. 2012; Falke et al. 2013).

The role played by small effect mutations for adaptation results from their "number" because their cumulative effect can be large. Clearly, dissecting the molecular and phenotypic consequences of a small-effect mutation is difficult. Therefore, it has been proposed to analyze their cumulative effect across sets of genes participating in a given molecular system,

pathway, or functional module, to unravel the fast evolving functional modules that underpin polygenic adaptation (Bullard et al. 2010; Daub et al. 2013; Fraser 2013). Considering all mutations irrespective of their effect size has indeed the potential to significantly improve our power to detect traits subjected to natural selection (Berg and Coop 2014).

Cis-regulatory mutations are particularly useful to study the targets of polygenic selection, and this is for five reasons. First, mutations causing *cis*-regulatory change are in linkage with the expressed transcript, so that their location in the genome is approximately that of the transcript. Second, *cis*-regulatory changes can be readily identified by monitoring allele-specific expression (ASE) differences genome-wide, in F1 heterozygotes (de Meaux et al. 2006; He et al. 2012). Third, genome-wide analyses of *cis*-regulatory variation have demonstrated that there are a large number of such mutations (Wittkopp et al. 2008; Zhang and Borevitz 2009; He et al. 2012). It is thus possible to study the genomic distribution of mutations causing *cis*-regulatory changes. Fourth, *cis*-regulatory mutations have a phenotypic effect (they alter gene expression level) and because they are found in a large number, it is reasonable to assume that taken together, they constitute a bona fide collection of small effect mutations. Fifth, *cis*-regulatory regions are believed to play a significant role in adaptive evolution (King and Wilson 1975; Wray et al. 2003; Prud'homme et al. 2006; Wray 2007; Stern and Orgogozo 2008; Emerson and Li 2010) because they limit deleterious pleiotropy by allowing modular changes. The analysis of genome-wide distributions of *cis*-regulatory modifications has indeed proven a useful tool in various species to pinpoint the action of natural selection (Bullard et al. 2010; Fraser et al. 2012; Chang et al. 2013; Lemmon et al. 2014; Steige et al. 2015).

Here, we apply this approach to the analysis of organismic responses to environmental stresses in the two sister species *Arabidopsis halleri* and *A. lyrata*. Stress responses define to a large extent the abundance and distribution of divergent species (Stebbins 1952). For example, the species of *A. halleri* differs from its sister species *A. lyrata* in its ecological preferences. *Arabidopsis halleri* populations are found in highly competitive meadows of Central Europe and accumulate high levels of heavy metal in the leaf, thereby tolerating high soil metal concentrations (Clauss and Koch 2006; Clauss and Mitchell-Olds 2006; Courbot et al. 2007). Its sister species, *A. lyrata*, is found in poor and stressful soils (sand, dolomitic outcroppings), but also colonized Northern areas of the globe (Clauss and Koch 2006; Clauss and Mitchell-Olds 2006). The congeneric species *A. thaliana* has again a distinct ecology. It is a pioneer species thriving in disturbed habitats over a broad geographic distribution (Hoffmann 2005; Clauss and Koch 2006). Variation in the response to organismic stresses revealed the adaptive potential of gene expression variants segregating within species (de Meaux et al. 2005, DesMarais et al. 2012, Cubillos et al. 2014).

We generated F1 hybrids of each of the sister species *A. halleri* and *A. lyrata* with their outgroup relative *A. thaliana* and monitored allele-specific levels of expression in standard growth conditions, in response to dehydration or cold exposure. This data allowed us to map the genome-wide distribution of *cis*-regulatory mutations active in three distinct environments reflecting divergent adaptations of the two species. Because the sister species were both crossed to an outgroup species, it was possible to assign a phylogenetic origin to *cis*-acting changes. *Cis*-acting changes observed in only one of the two hybrids were likely to be caused by derived mutations, whereas those observed in both hybrids either predated the split between the two species or arose along the *A. thaliana* lineage. By contrasting the distribution of *cis*-regulatory mutations derived in *A. halleri* to those derived on the *A. lyrata* lineage, we could establish relative rates of *cis*-acting evolution across polygenic molecular functions (MFs) and detect lineage-specific polygenic adaptation to environmental challenges.

Results and Discussion

Arabidopsis Interspecific Hybrids Respond Similarly to Stress

We produced hybrids of *A. thaliana* with each of the two sister species *A. lyrata* and *A. halleri*, named AthxAly and AthxAha hybrids, respectively, using one parental genotype in each species (see Materials and Methods). The leaf transcriptome of the two hybrid types was characterized by RNAseq in three conditions (standard, cold, dehydration) with three biological replicates. The number of reads ranged from 29 to 364 million per sample (supplementary table S1, Supplementary Material online). Reads were mapped to a hybrid genome, concatenating the *A. thaliana* and *A. lyrata* reference genomes, with the help of a specifically tailored pipeline accounting for allelic differences within and between hybrids in their divergence to the reference genome (see Materials and Methods). We compared expression levels between samples (measured as FPKM, Fragments Per Kilobase of transcript per Million mapped reads) by computing a Heatmap of Spearman correlations. Gene expression correlations built three main clusters reflecting the three environmental conditions, each of which then subdivided into genotypes (supplementary fig. S1A, Supplementary Material online). This demonstrated, that despite their genetic differences, the hybrids displayed comparable reactions to the stresses. Between 1,465 and 3,225 genes changed significantly their expression level in response to cold exposure or dehydration, compared with standard growth conditions (DESeq2 analysis, see Materials and Methods and supplementary fig. S2, Supplementary Material online). This number was lower in samples collected from dehydrating samples compared with cold samples in AthxAha hybrids, presumably because the variance between biological replicates was greater.

In interspecific hybrids, genes may show aberrant expression patterns (Barreto et al. 2014). For most genes, this is not the case in these hybrids. Indeed, despite their sterility, the hybrids form healthy and long-lived plants in the greenhouse. In addition, genes changing their expression in response to stress overlapped significantly with genes observed to respond to the same stresses in *A. thaliana* (supplementary table S2, Supplementary Material online; Matsui et al. 2008).

The interspecific hybrids therefore displayed a stress response comparable to the response observed in the outgroup species. The genome-wide distribution of allelic expression in these hybrids can thus bring biologically meaningful insights into differences between species.

To quantify allele-specific levels of expression, we designed a bioinformatics pipeline that efficiently controlled for variation in sequencing depth as well as allelic mapping biases caused by different quality of parental reference genomes (see Materials and Methods and supplementary fig. S3, Supplementary Material online). After scaling, distributions of either FPKM or allelic ratios were similar across samples (supplementary fig. S4, Supplementary Material online). We applied a generalized linear model (GLM) model with a quasi-binomial distribution of error to detect orthologous genes with significant ASE pattern, hereafter named ASE genes (He et al. 2012). In this model, allelic ratios observed in RNAseq were contrasted to allelic ratios observed in hybrid DNA sequencing, to account for possible locus-specific allelic biases that were not corrected in our algorithm. Significant ASE was independent from allelic ratios in DNA as well as from gene expression level for a large range of FPKM values (supplementary figs. S5 and S6, Supplementary Material online).

Among 16,419 orthologous genes, between 1,721 and 2,454 genes displayed significant ASE in at least one hybrid and one environment (supplementary fig. S7 and tables S3A and B, Supplementary Material online). As expected, allele-specific expression variance clustered by genetic background (supplementary fig. S1*B*, Supplementary Material online). Most significant ASE genes were observed in all three conditions. Only 130 and 122 genes showed a significant change in allelic ratio in cold-exposed versus standard plants at false discovery rate (FDR) 0.05 in AthxAha and AthxAly hybrids, respectively. Similar numbers of genes changed allelic ratios in dehydrating versus standard plants (205 and 197 at FDR 0.05 in AthxAha and AthxAly hybrids, respectively).

Three lines of evidence indicated that our identification of ASE is reliable. First, genes displaying ASE in standard conditions significantly overlapped with ASE reported for AthxAly hybrids grown in similar conditions but involving a different pair of parents (Hypergeometric test, $P < 2e\text{-}5$, He et al. 2012). Second, allelic ratios in this study correlated significantly with allelic ratios measured by pyrosequencing at 12 loci in this previous data set ($R^2 = 0.64$, $P = 0.003$, He et al. 2012). We further quantified allele-specific expression by pyrosequencing at three additional loci in each hybrid and treatments using the same RNA samples used for RNAseq. In this case, ASE correlated strongly with ASE measured by RNAseq ($R^2 = 0.90$, $P < 2e\text{-}16$). Third, we quantified ASE in standard conditions in the AthxAly hybrids in an independent trial, with three replicates. We observed a strong correlation in allelic expression ratios across experiments (Pearson correlation, $r = 0.92$, $P < 2.2e\text{-}16$) and 1,458 significant ASE genes (75.3%) were recovered in both experiments.

Phylogenetic Assignment of Derived *Cis*-Acting Changes Increasing Allelic Expression

Since *A. thaliana* is an outgroup species to *A. halleri* and *A. lyrata*, *cis*-regulatory changes that were specific to one of the two hybrid types could be assigned as derived in the corresponding lineage (fig. 1, orange and blue dots). Instead, *cis*-acting changes observed in both hybrids were assumed to have either occurred before the split between *A. halleri* and *A. lyrata* or evolved along the *A. thaliana* lineage. For simplicity, we will name these *cis*-acting changes "ancestral" here (fig. 1, green dots). Genes showing ASE in opposite directions in one and the other hybrids were rare, supporting the assumption that genes that have undergone more than one *cis*-acting change were infrequent (fig. 1, top left and bottom right corners of the graphs). Heterozygosity in the *A. halleri* parent probably reduced our power to detect significant ASE in the AthxAha hybrid, compared with the AthxAly hybrid. This may in turn have inflated the number of ASE genes derived in *A. lyrata* or decreased the number of genes showing derived ASE in the *A. halleri* lineage. To mitigate this problem, it was important to 1) give more weight to genes showing a

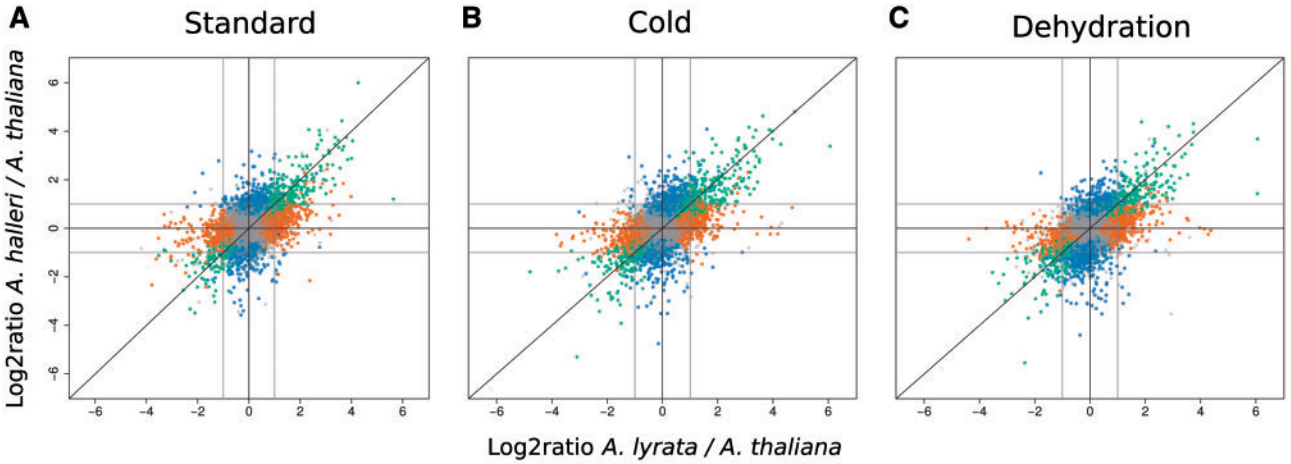

FIG. 1. Log2 allelic ratios in AthxAha hybrids plotted against log2 allelic ratios in AthxAly hybrids in standard conditions (*A*), in cold-exposed plants (*B*) and in dehydrating plants (*C*). ASE genes with an ancestral mutations are on the diagonal (green), ASE genes with a derived changes in *Arabidopsis lyrata* are on the horizontal central lines (orange) and ASE genes with a derived change in *A. halleri* are on the vertical line (blue). Compensatory changes are rare, which supports the phylogenetic interpretation of ASE genes as ancestral or derived modifications.

stronger allelic expression difference, 2) avoid an arbitrary definition of derived and ancestral ASE changes but 3) capture well the phylogenetic origin that can be derived by overlapping significant ASE genes in each hybrid. We thus designed phylogenetic indices to construct three gene rankings reflecting the likelihood that allele-specific expression is being derived in *A. lyrata* (or *A. halleri*) or the likelihood of being ancestral. For example, genes with ASE derived in one or the other lineage should have the largest difference in allelic ratio between hybrids, that is, a high log ratio in one of the hybrid and a log ratio close to 0 in the other one. We thus ranked derived genes by decreasing allelic ratio difference between hybrids (see Materials and Methods). We removed from the ranking all genes that did not have significant ASE in one or the other hybrid, to reduce the impact of outlier mean ratios, as well as genes that were not expressed in both hybrids (scaled coverage below 10). Examples of this ranking are shown in supplementary fig. S8, Supplementary Material online. We designed a similar index ranking first the genes showing ancestral *cis*-acting changes shared between both hybrids (see Materials and Methods for details).

Reasoning that positive selection should favor an accumulation of derived *cis*-acting changes in the molecular systems underpinning adaptive phenotypes, we searched for the polygenic MFs that accumulated an excess of *cis*-acting mutations specifically in the *A. lyrata* or *A. halleri* lineages. We applied Gene Set Enrichment Analysis (GSEA), a test based on a Kolmogorov–Smirnov-statistics first introduced by Subramanian et al. (2005), which is designed to identify functional enrichments in a ranked set of genes, and tested functional enrichments among the top ASE genes in each phylogenetic ranking and each environment (supplementary fig. S8, Supplementary Material online, for details see Materials and Methods). To describe polygenic MFs, we used 1) gene ontology (GO) annotations for "biological processes" (BP) and "MF" 2) published sets of stress-responsive genes in the outgroup species *A. thaliana*, and 3) a combination of (1) and (2). GO enrichments were also performed on randomized rankings to define a *P* value threshold below which less than 10% of the enriched GO categories are false positives (see Materials and Methods).

We first focused on GO enrichment detectable among all ASE changes, irrespective of the direction of the change. Some polygenic MFs, may certainly adapt via a combination of up- and downregulating *cis*-acting changes (Picotti et al. 2013). In total, 22 and 15 GO categories were significantly enriched among genes with a changed *cis*-activity derived in *A. halleri* and *A. lyrata*, respectively, at a *P* value threshold set so that the FDR approximates 0.1 (table 1, see Materials and Methods). Among ASE genes displaying lower *cis*-activity of the *A. thaliana* allele in both hybrids, that is, ancestral *cis*-acting mutations, ten GO categories were significantly enriched at FDR 0.1 (supplementary tables S4 and S5, Supplementary Material online). These findings demonstrate that ASE changes accumulate in distinct functions in each lineage.

The relationship between expression level and fitness is complex (Rest et al. 2013). Therefore, we also analyzed separately genes showing a derived *cis*-acting change leading to higher allelic activity in the derived lineage (by removing derived ASE genes with a significantly negative allelic ratio, FDR 0.05, from the ranking) or those leading to lower allelic activity (by removing derived ASE genes with a significantly positive allelic ratio, FDR 0.05, from the ranking). This was possible because our crossing design allows identifying the direction of derived *cis*-regulatory changes. Coordinated *cis*-acting changes increasing allelic activity may be required for the higher activity of selected molecular systems. Alternatively, *cis*-acting changes decreasing gene expression level may be selected to decrease molecular costs when evolutionary constraints are released. *Arabidopsis thaliana* pseudogenes, for example, tend to be less expressed than active *A. lyrata* orthologues (He et al. 2012). This analysis of directional *cis*-regulatory divergence yielded another partially overlapping set of significantly enriched GO categories (supplementary tables S4 and S5, Supplementary Material online). Finally, we also asked whether specific MFs were enriched among the genes showing a significant change in allelic ratio in stress versus standard conditions, which yielded another set of 13 GO categories mostly enriched in the *A. halleri* lineage (fig. 2 and table 2).

To unify all observed enrichments, we constructed a hierarchical clustering dendrogram based on measures of similarity between BP GO categories represented by at least five genes expressed in both hybrids in at least one treatment (fig. 2, see Materials and Methods for details). This graph described the functional "landscape" of expressed genes in our samples and showed that significant GO categories target distinct functional pathways in the *A. lyrata* and *A. halleri* lineages. If mutation rate alone explained accumulation of *cis*-regulatory mutations in specific molecular systems, enrichments would be expected to highlight related GO categories in all lineages. The GO categories enriched among derived ASE genes were generally not enriched among ancestral ASE genes, confirming that the accumulation of *cis*-acting changes occurred in the derived lineage. This pattern suggests the action of divergent natural selection by which distinct polygenic functions were progressively modified in the two lineages.

Signature of Polygenic Selection in Genes Controlling Heavy Metal Homeostasis in the *A. halleri* Lineage

Arabidopsis halleri stands out in the *Arabidopsis* genus as the only species that tolerates and accumulates heavy metals, especially Zn and Cd (Krämer 2010). This tolerance allows it to grow in heavily polluted areas (Pauwels et al. 2005). In agreement with this ecological preference, we discovered consistent evidence that *cis*-acting mutations accumulated in the *A. halleri* lineage to increase the expression of genes functioning in metal transport and homeostasis (fig. 2 and supplementary table S5, Supplementary Material online). The MF "zinc ion transmembrane transporter activity" was significantly enriched among *A. halleri*-derived ASE genes ($P = 0.0052$) under standard conditions, as well as among *A. halleri*-derived ASE genes with increased allelic expression ($P = 0.0025$). Under dehydration conditions, the BP "zinc ion homeostasis" and

Table 1. GO Categories Showing a Significant Enrichment ($P < 0.009$) among Genes Showing Lineage Specific *Cis*-Acting Changes.

	GO.ID	GO.Domain	Treatment	Term	*P* Value
Arabidopsis halleri derived	GO:0003735	MF	Standard	Structural Constituent Of Ribosome	0.00054
	GO:0010310	BP	Standard	Regulation of hydrogen peroxide metabolism	0.00059
	GO:0001510	BP	Standard	RNA methylation	0.00074
	GO:0043085	BP	Dehydration	Positive regulation of catalytic activity	0.0011
	GO:0002237	BP	Cold	Response to molecule of bacterial origin	0.0013
	GO:0019288	BP	Dehydration	Isopentenyl diphosphate biosynthetic process	0.0017
	GO:0010106	BP	Cold	Cellular response to iron ion starvation	0.0021
	GO:0010027	BP	Dehydration	Thylakoid membrane organization	0.0021
	GO:0046482	BP	Cold	Para-aminobenzoic acid metabolic process	0.0031
	GO:0016226	BP	Dehydration	Iron-sulfur cluster assembly	0.0031
	GO:0009658	BP	Dehydration	Chloroplast organization	0.0033
	GO:0043241	BP	Cold	Protein complex disassembly	0.0041
	GO:0006468	BP	Cold	Protein phosphorylation	0.0052
	GO:0019252	BP	Dehydration	Starch biosynthetic process	0.0052
	GO:0007165	BP	Standard	Signal transduction	0.00524
	GO:0005385	MF	Standard	Zinc ion transmembrane transporter activity	0.00527
	GO:0010359	BP	Cold	Regulation of anion channel activity	0.0059
	GO:0046482	BP	Dehydration	Para-aminobenzoic acid metabolic process	0.0064
	GO:0000902	BP	Cold	Cell morphogenesis	0.0076
	GO:0016226	BP	Cold	Iron-sulfur cluster assembly	0.0077
	GO:0009651	BP	Cold	Response to salt stress	0.0081
	GO:0019825	MF	Cold	Oxygen binding	0.0081
Arabidopsis-lyrata derived	GO:0016706	MF	Cold	Oxidoreductase activity, acting on paired donors, with 2-oxoglutarate as one donor. . .	0.00035
	GO:0006366	BP	Cold	Transcription from RNA polymerase II promoter	0.0015
	GO:0046463	BP	Dehydration	Acylglycerol biosynthetic process	0.0023
	GO:0000394	BP	Cold	RNA splicing, via endonucleolytic cleavage and ligation	0.0028
	GO:0016556	BP	Standard	mRNA modification	0.0033
	GO:0048598	BP	Cold	Embryonic morphogenesis	0.0034
	GO:0004806	MF	Dehydration	Triglyceride lipase activity	0.0047
	GO:0016798	MF	Cold	Hydrolase activity, acting on glycosyl bonds	0.00722
	GO:0006457	BP	Cold	Protein folding	0.0075
	GO:0008170	MF	Standard	N-methyltransferase activity	0.0079
	GO:0007389	BP	Cold	Pattern specification process	0.0083
	GO:0009086	BP	Cold	Methionine biosynthetic process	0.0083
	GO:0006084	BP	Standard	Acetyl-CoA metabolic process	0.0086
	GO:0004857	MF	Cold	Enzyme inhibitor activity	0.00867
	GO:0016132	BP	Standard	Brassinosteroid biosynthetic process	0.0089

NOTE.—The enrichment was computed against a universe comprising all ASE genes using the GSEA test.

"cellular transition metal ion homeostasis" were enriched ($P = 0.0032$ and 0.001, respectively). Additional BP such as "iron ion transport" or "cellular response to iron ion starvation" were also enriched in cold-exposed plants. To confirm the biological significance of these enrichments, we used a list of 26 genes with proven relevance for heavy metal tolerance and accumulation (supplementary table S6, Supplementary Material online see also Talke et al. 2006). These genes were enriched among *A. halleri*-derived *cis*-acting changes increasing allelic expression in all three conditions (FDR corrected *P* value = 0.035, 0.007, and 0.007, for standard, cold-exposure and dehydration respectively, supplementary tables S6 and S7, Supplementary Material online). For example, the gene *HMA4*, whose duplication and stronger *cis*-acting region was shown to cause a significant increase in Zn accumulation and tolerance to high levels of Zn and Cd (Courbot et al. 2007; Hanikenne et al. 2008) was among *A. halleri*-derived ASE genes. We should note here again that ASE detection is independent of gene duplication events because RNA allelic ratios are compared with DNA allelic ratios (supplementary fig. S5, Supplementary Material online). We also observed higher expression of the *A. halleri* allele for the metal transporters MTP1/ZAT1 (Dräger et al. 2004) and ZIP3 (Talke 2006). The genes *IRT3* (Lin et al. 2009), *NRAMP3* (Oomen et al. 2009), and *HMA1* (Kim et al. 2009), involved in Zn detoxification in *A. thaliana*, also showed a derived and constitutively higher activation of the *A. halleri* allele.

Quantitative-trait mapping approaches described nine QTLs for heavy metal tolerance, pointing to its polygenic basis in *A. halleri* (Willems et al. 2007, 2010). The QTL *HMA4* plays a key role in heavy metal tolerance, but the signature left by positive selection on this individual gene was difficult to investigate due to the complex structure of a triplicated gene subjected to frequent gene conversions (Hanikenne et al. 2008, 2013). By contrasting the relative enrichment in *cis*-acting changes across the two sister lineages, our approach now provides one of the strongest evidence that tolerance to heavy metal did not evolve neutrally, but arose as a response to natural selection in the *A. halleri* species. Our study shows that an excess of *cis*-acting changes affecting heavy metal

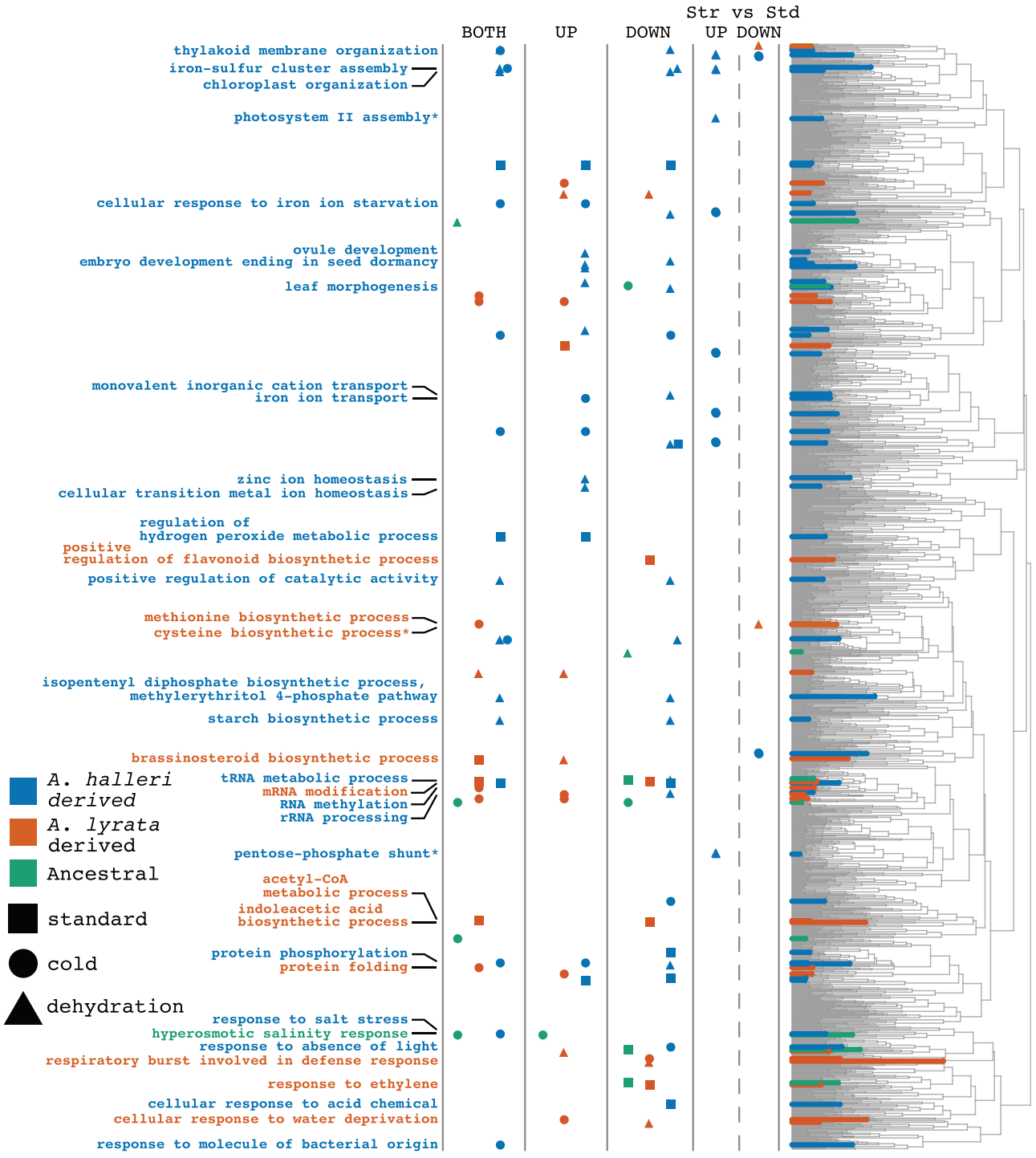

Fig. 2. The functional distribution of GO categories showing an enrichment in up- and/or downregulating derived or ancestral *cis*-acting mutations, as well as in cold- or dehydration-responsive *cis*-acting changes. The hierarchical clusters are based on measures of semantic similarity among GO terms with at least five expressed genes in our data set. Enriched GO categories are highlighted in orange for in *Arabidopsis lyrata*-derived *cis*-acting changes, blue for *A. halleri*-derived *cis*-acting changes, and in green for ancestral *cis*-acting changes. Names are given for GO categories that are enriched in both *A. thaliana* responsive genes (Hannah et al. 2005, Des Marais et al. 2012) and among lineage-specific ASE genes or cold- or dehydration-responsive *cis*-acting changes. The name of categories related to metal transport and homeostasis are also mentioned on the left. The position of GO categories enriched among stress-specific ASE genes is given in the columns "versus standard," which are subdivided into those enriched among "up" and "down"-regulating stress-responsive ASE changes. A disc and a triangle are used for categories enriched among ASE changes responsive to cold or dehydration, respectively. To avoid confusion, the names of the GO categories enriched among stress-responsive ASE genes are highlighted on the left with an asterisk (*).

Table 2. Significant Enrichments among Genes Displaying a Significant ASE Change in Response to Environmental Stress.

	GO.ID	GO.Domain	Treatment	Enriched GO Term	Enriched ASE Subset	*P* Value
Enriched GO categories	GO:0006098	BP	Dehyd. vs Std	Pentose-phosphate shunt	Aha-derived up	0.0026
	GO:0010207	BP	Dehyd. vs Std	Photosystem II assembly	Aha-derived up	0.0058
	GO:1902589	BP	Dehyd. vs Std	Single-organism organelle organization	Aha-derived up	0.0062
	GO:0009657	BP	Dehyd. vs Std	Plastid organization	Aha-derived up	0.0096
	GO:0019344	BP	Dehyd. vs Std	Cysteine biosynthetic process	Aly-derived down	0.0025
	GO:0032535	BP	Dehyd. vs Std	Regulation of cellular component size	Aly-derived down	0.0082
	GO:0072594	BP	Cold vs Std	Establishment of protein localization to organelle	Aha-derived up	0.0045
	GO:0016482	BP	Cold vs Std	Cytoplasmic transport	Aha-derived up	0.0074
	GO:0006605	BP	Cold vs Std	Protein targeting	Aha-derived up	0.0097
	GO:1902580	BP	Cold vs Std	Single-organism cellular localization	Aha-derived up	0.0097
	GO:0046148	BP	Cold vs Std	Pigment biosynthetic process	Aha-derived down	0.0046
	GO:1902589	BP	Cold vs Std	Single-organism organelle organization	Aha-derived down	0.0098
	GO:0016787	MF	Cold vs Std	Hydrolase activity	Aly-derived up	0.0089
Stress response genes			Cold vs Std	Long term cold response (Hannah et al. 2005)	Aly-derived	0.0009
			Cold vs Std	Down regulated in long term cold response (Hannah et al. 2005)	Aly-derived	0.0004

NOTE.—GO terms as well as cold-response (Hannah et al. 2005) and drought-response genes (Des Marais et al. 2012) were tested. Enrichments were also investigated among genes where the expression of the derived allele was increased or decreased in the stress (Aha- or Aly- derived "up"/"down").

homeostasis was derived in the *A. halleri* lineage, and not in the *A. lyrata* or the ancestral lineages. Lineage-specific acceleration in substitution rates in specific genes or gene sets are indeed the hallmark of natural selection (Nielsen 2005; Daub et al. 2013).

Recent work suggests that herbivores and other biotic challenges may have been the cause of selection pressure for heavy metal accumulation (Kazemi Dinan et al. 2014, 2015; Plaza et al. 2015). This may have contributed to the adaptive rewiring of stress reaction networks in *A. halleri*. In cold-exposed plants, we detected an excess of *cis*-acting mutations modifying the expression of genes involved in the "response to salt stress" or "response to molecular of bacterial origin" (table 1 and fig. 2). The adaptation of stress responses may also have required a modified regulation of photosynthesis. Indeed, genes with *A. halleri*-derived ASE change between cold-exposed or dehydrating plants and plants growing in standard conditions were enriched in gene functions related to plastid organization (tables 1 and 2, fig. 2).

Cis-Regulatory Variation Derived in the *A. lyrata* Lineage

The signature of polygenic selection on a function of proven ecological relevance in the *Arabidopsis halleri* lineage provides a proof-of-principle that the distribution of *cis*-acting mutations can be efficiently used to detect targets of polygenic selection in natural plant systems. We thus further asked whether specific molecular systems might have been the target of polygenic selection in the *A. lyrata* lineage (tables 1 and 2). *Arabidopsis lyrata* grows on relatively stressful and unprotected habitats exposed to sharp climatic fluctuations (Clauss and Mitchell-Olds 2006). In the greenhouse, *A. lyrata* is strikingly more robust than *A. halleri*, maintaining growth at high temperatures and low soil moistures. We thus expected that ASE derived in *Arabidopsis lyrata* would be specifically enriched among drought responsive genes if this function had been targeted by polygenic selection. In addition, in a synthetic allopolyploid containing the *A. halleri* and the *A. lyrata* genomes, the cold response was largely due to the specific activation of *A. lyrata* homeologs, indicating that *A. lyrata* alleles may have provided the ability of the *A. lyrata*/*A. halleri* polyploid *A. kamchatica* to colonize areas reaching the pre-Arctic region (Akama et al. 2014). Indeed, *A. lyrata* was able to colonize regions of high latitude (Hoffmann 2005). If cold-tolerance had evolved under polygenic selection, ASE derived in *A. lyrata* should also be enriched in genes important for cold tolerance.

Our analysis highlighted 43 GO categories enriched among *A. lyrata*-derived ASE genes, 20 of which were enriched among ASE genes expressed in cold-exposed plants (tables 1 and 2, supplementary table S5, Supplementary Material online). To refine the interpretation of these functional enrichments, we used two sets of genes whose transcription is modified by cold- and drought-stress in *A. thaliana* (Hannah et al. 2005; Des Marais et al. 2012, supplementary table S8, Supplementary Material online). These gene sets helped describe the molecular basis of the stress response in the outgroup species *A. thaliana* so that patterns of ASE could be used to monitor the extent to which this response may have been modified by natural selection in each of the two derived lineages.

We found a marked enrichment of genes responsive to short-, and long-term responses to cold exposure among *A. lyrata*-derived ASE genes. Genes downregulated in the short-term cold exposure and upregulated in long-term cold exposure were enriched among upregulating *A. lyrata*-derived *cis*-acting changes ($P = 0.006$ and $P = 0.002$, respectively). This enrichment was detected only in ASE genes expressed in standard conditions and suggests that some genes important for cold acclimation have evolved to be constitutively activated in *A. lyrata*, which in turn may have allowed the evolution of a decreased short-term stress response. Interestingly, genes showing an ASE ratio that changed significantly in cold-exposed versus standard growing AthxAly hybrids were 6-fold enriched in genes downregulated after long exposure to cold ($P = 0.00036$, table 2). There is therefore in *A. lyrata* a strong excess of *cis*-acting changes among genes

downregulated by cold stress in the outgroup *A. thaliana*. This excess seems to target a subsample of GO categories that are enriched both in *A. thaliana* cold-responsive genes and in *A. lyrata*-derived ASE genes, such as BP "brassinosteroid biosynthetic process" or "acetyl-coA metabolic process" (fig. 2, supplementary table S5 and S9, Supplementary Material online). In contrast, *A. halleri*-derived ASE genes did not show enrichments in cold-responsive genes. This suggests that the response to cold was adaptively remodeled in the *A. lyrata* lineage.

The response to drought, instead, was globally not more rewired in one or the other lineage. Genes downregulated by drought were enriched among *A. lyrata*-derived ASE decreasing allelic expression ($P = 0.02$), but this enrichment was detected in cold-exposed plants and not in dehydrating plants. In addition, genes upregulated by drought were enriched among *A. halleri*-derived ASE changes decreasing gene expression in dehydrating plants ($P = 0.03$). The patterns of GO enrichments seemed to be less marked with only ten GO categories enriched in *A. lyrata*-derived ASE genes expressed in dehydrating plants, and two categories enriched among genes changing allelic ratio in dehydration versus standard conditions in the *A. lyrata* lineage (tables 1 and 2, supplementary table S5, Supplementary Material online). Nevertheless, some GO enrichments suggested that derived ASE may contribute to the optimization of the dehydration response. GO category "response to chitin" is the most strongly enriched category among genes downregulated in response to drought in *A. thaliana* (Fisher test, $P = 6.10e\text{-}16$, Des Marais et al. 2012, supplementary table S9, Supplementary Material online). It contains several typical drought stress-responsive genes such as ethylene-response factors. Interestingly, this category and a related one "cellular response to ethylene stimulus" were marginally enriched among *A. lyrata* ASE downregulating allelic activity in dehydrating plants ($P = 0.011$, $P = 0.034$, respectively, supplementary table S5 and S9, Supplementary Material online). On the other hand, the GO category "isopentenyl diphosphate biosynthetic process," which is among the most strongly enriched categories among genes upregulated in response to drought in *A. thaliana* (Fisher test, p1e-30, Des Marais et al. 2012, supplementary tables S5 and S9, Supplementary Material online) was clearly enriched among *A. halleri*-derived ASE genes with lower allelic transcription in dehydrating plants ($P = 0.0012$, supplementary table S5, Supplementary Material online). Therefore, this data did not bring clear support for the specific evolution of the response to drought in the *A. lyrata* lineage. Instead, distinct functional categories evolved an excess of *cis*-acting changes in each lineage. Improved biological annotations, along with better models predicting the collective effects of *cis*-acting changes, will be needed to improve the interpretation of patterns of enrichment among GO categories that in *A. thaliana* are activated in response to drought stress.

Cis-Regulatory Change Accumulate in Evolutionary Constrained Functions

Numerous studies have documented the important contribution of *cis*-regulation to regulatory and phenotypic change (King and Wilson 1975; Prud'homme et al. 2007; Wray 2007; Stern and Orgogozo 2008; Wittkopp et al. 2008; Emerson et al. 2010; McManus et al. 2010; Coolon et al. 2014). Contrasting patterns of regulatory divergence to other sources of molecular or phenotypic information thought to reflect fitness has further contributed to document the adaptive relevance of this source of variation. For example, in maize, the 4% of genes showing fixed *cis*-acting differences between teosinte and maize were enriched in genome regions carrying the signature of recent selection, suggesting that they played a predominant role in the adaptation of gene regulation during domestication (Lemmon et al. 2014). In the *Capsella* genus, *cis*-acting changes between the outcrossing species *Capsella grandiflora* and its inbred relative *C. rubella* were enriched in regions underlying QTLs involved in flower morphology, a trait that was selectively remodeled after *C. rubella* evolved selfing (Steige et al. 2015). In yeast, the analysis of *cis*-acting differences between a pathogenic strain and the wild type uncovered a marked excess of *cis*-acting changes in genes coding for interacting proteins involved in endocytosis. The functional importance of the regulation of these proteins in pathogenicity was clearly validated experimentally (Fraser et al. 2012).

In this study, we directly contrasted the number and effect of *cis*-acting mutations accumulating on separate lineages within each polygenic functional module, thereby controlling for differences in evolutionary constraints during divergence. This approach was thus akin to the relative-rate tests commonly used in molecular evolution to describe the molecular basis of adaptation (see Nielsen 2005 for a review). We detect a signature of polygenic selection on several functions of known relevance in the *A. halleri* lineage and possible indication that the cold response was adaptively rewired in *A. lyrata*. Interestingly, the functional molecular systems highlighted here do not stand out as fast evolving at the amino acid level. Contrary to other reports of ASE between *A. lyrata* and *A. thaliana* or *C. grandiflora* and *C. rubella* (He et al. 2012, Steige et al. 2015), the rate of nonsynonymous to synonymous mutations among derived ASE genes does not differ significantly from the rest of the genome (supplementary fig. S9, Supplementary Material online). Instead, we observe that GO categories enriched in derived *cis*-acting changes tend to be more constrained at the amino acid level (maximum $P = 0.0027$ for enriched GO categories in cold-exposed, dehydration and standard plants, in *A. halleri* and *A. lyrata*). This is particularly strong in enriched GO categories in cold-exposed and standard conditions in *A. halleri* ($P = 2.2e\text{-}16$, $P = 2e\text{-}16$ in cold-exposed and standard plants, respectively). This effect is not due to the amino acid divergence of ASE genes themselves within each category (Kolmogorov–Smirnov test, $P > 0.1$, not shown). With the exception of GO categories enriched in *A. halleri*-derived ASE in dry conditions, this effect remains when gene expression level is included as a covariate (expression level measured as mean FPKM, $P = 0.49$, $P = 2.53e\text{-}5$, $P = 2.19e\text{-}5$ in dehydrating, cold-exposed and standard plants, respectively). The patterns of *cis*-regulatory divergence described here provide an example of how polygenic selection targets functions that tend to be more constrained at the amino acid level. It shows that

adaptive changes revealed by lineage-specific *cis*-regulatory divergence cannot be predicted by classical signatures of diversifying selection on coding sequence (Nielsen 2005; Storz 2005; Fournier-Level et al. 2011). We conclude that the analysis of *cis*-acting changes delivers insights into the wealth of molecular changes that accumulate as lineages adapt to novel niches and promises to find applications in a broader range of natural systems.

Materials and Methods

Plant Materials, Stress Treatments and RNA Isolation

Arabidopsis thaliana accession Col-0 was obtained from the Arabidopsis Biological Resource Center (ABRC, USA). *Arabidopsis lyrata* ssp. *lyrata* genotype MN47 and *A. halleri* h2-2 (Gorno, Italy) were obtained from Pierre Saumitou-Laprade (University of Lille, France). The MN47 line is a selfing *A. lyrata* line, h2-2, instead, is strictly outcrossing. Parental lines were crossed by rubbing *A. lyrata* or *A. halleri* pollen on the pistils of emasculated *A. thaliana* flowers, as described in de Meaux et al. (2006), and interspecific F1 progenies were generated, named AthxAly and AthxAha hybrids, respectively. Reciprocal crosses using *A. thaliana* as a father were tried with many parental genotypes but were not successful. Note that analyses of reciprocal crosses within *A. thaliana* and *A. lyrata* showed that imprinting is not affecting allele-specific expression in these species (Cubillos et al. 2014; Videvall et al. 2016).

For each treatment, samples were collected for three independent individuals, grown each in 1-week intervals in the same growth chamber. AthxAly and AthxAha F1 hybrid seeds were germinated and grown on germination medium containing Murashige and Skoog salts, 1% sucrose, and 0.8% agar. The plants were stratified for 5 days at 4 °C, and then grown for 4 weeks in a growth chamber at 22 °C/16 °C under 16 h light/8 h dark. Dehydration and cold exposure were applied following a published protocol (Seki et al. 2002). For dehydration, plants were removed from the agar and dehydrated in plastic dishes for 1 h at 22 °C under dim light (0.7 ± 0.8 mmol s^{-1} m^{-2}). For cold exposure, plants were grown under dim light (0.7 ± 0.8 mmol s^{-1} m^{-2}) at 4 °C for 1 h. Leaf samples of plants growing in nonstressful conditions (standard treatment) were collected on 4 week-old plants grown at 22 °C. To control for circadian changes in gene expression, all samples were collected at 12:00 pm. The samples collected in "standard" conditions did not form a proper "control" for the stress reactions because, for technical reasons, they were not collected in the same trial as the samples collected in stressful conditions.

Preparation of cDNA and DNA Libraries

Total RNA was extracted from whole aerial part of one hybrid individual in 1 ml of TRIZOL Reagent (Thermo Fisher Scientific). DNA was cleaned up using the DNA-free kit (Thermo Fisher Scientific). RNA quality and quantity was examined with the Bioanalyzer 2100 (Agilent) and Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). Two micrograms of total RNA were used for library preparation. The library preparation followed the TruSeq® RNA Sample Preparation v2 Guide (Illumina). DNA was extracted from two individuals of each interspecific hybrid type. DNA library preparation followed the Illumina Paired-End Sample Preparation Guide. Sequencing was performed on Illumina HiSeq2000 following the manufacturer's protocols. Paired-end 100 bp long reads were obtained. Raw reads and allele-specific coverage data can be accessed on the Gene Expression Omnibus (ref: GSE80462)

Data Analysis

After FastQC quality check (www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed May 18, 2016), the FastX-toolkit was used for sequence trimming and filtering. Low quality nucleotides were removed from the 3′-ends of the sequences (quality threshold of phred score (t): 20, minimum length (l): 50). Sequences were reverse complemented by fastx_reverse_complement, and the other end was cut following the parameter described above before being reversely complemented to the original direction of reads. Reads with less than 90% bases above quality threshold and paired-end reads with a single valid end were removed. Trimmed and filtered reads were mapped to the hybrid genome, that is, the concatenation of the *A. thaliana* Col-0 reference genome (TAIR10, www.arabidopsis.org, last accessed May 18, 2016) and the *A. lyrata* MN47 reference genome (Araly1, [Nordberg et al. 2014], using tophat2 with the built-in Bowtie2 mapping program [Trapnell et al. 2010]) as the following parameter: -p 5 (number of threads), -N 5 (read mismatches), –read-edit-dist 5 (Final read alignments having more than 5 edit distance were discarded), –read-gap-length 5 (Final read alignments having more than 5 total length of gaps were discarded). Uniquely and high-quality mapping reads were selected by "samtools view -q 10" (-q 10, selecting genome fractions were 90% possibility of the mapping is correct). We focused our analysis on orthologues present in each of the three species, *A. thaliana*, A. *lyrata*, and *A. halleri*. The Araly1 genome annotation detects 17,846 genes orthologous between *A. thaliana* and *A. lyrata*. The sequences of these orthologues were blasted against the *A. halleri* draft genome sequences of Aha1.1 (Phytozome 10.1, *Arabidopsis* halleri v1.1). Criteria for orthologous genes were set as 1) hit E-value $<$ 1 e -20 and 2) pairs of genes with reciprocal best BLAST hits were defined to be orthologues. A total of 528 genes had no detectable orthologue in the *A. halleri* draft genome, resulting in an initial set of 17,318 orthologous genes in the three species. After initial mapping, we observed that many *A. halleri* DNA reads did not map to the *A. lyrata* reference genome causing a distortion of *A. halleri* to *A. thaliana* allelic ratios (supplementary fig. S3, Supplementary Material online).

We therefore used DNA of hybrid individuals to optimize our mapping pipeline. In DNA of hybrids, biases in allelic amounts can only result from read mapping biases. First, unmapped *A. halleri* reads were trimmed down to 30 bp single reads and remapped to the reference of *A. lyrata*. The orthologous cDNA sequences of *A. thaliana* and *A. lyrata* were aligned using MAFFT (Katoh

and Standley 2013) and 1.34 million divergent sites (hereafter single nucleotide polymorphisms [SNPs]) were located. The parental origin of each read had to be reassigned, based on Samtools SNP callings, because the bowtie2 algorithm sometimes maps reads on the wrong parental allele. To avoid the misassignment of an *A. halleri* allele due to mismatches with the *A. lyrata* reference genotype, we decided to rely only on SNPs fixed in the common ancestor of *A. lyrata* and *A. halleri* for calling the parental origin of each read in both hybrids. For this, we used data collected by whole-genome sequencing of 22 and 72 representative genotypes of *A. halleri* and *A. lyrata*, respectively (Novikova, Nordborg, GMI, Vienna, unpublished). The reads of these 92 genomes were mapped on the *A. lyrata* reference genome (Hu et al. 2011) using the BWA-MEM algorithm from BWA (version 0.7.4) with an increased penalty for unpaired read pairs of 15. Duplicated reads were removed with Samtools (version 0.1.18) rmdup function. Local realignment were performed with Genome Analysis ToolKit (GATK, version 2.5.2), IndelRealigner (McKenna et al. 2010; De Pristo et al. 2011), and filtered for primary alignment of the reads. To reduce erroneous variant callings caused by undetected gene duplications, we calculated coverage distribution using GATK Pileup and genomic regions with coverage lower or higher than the 3rd and 97th percentiles were removed. SNPs and short indels were called with GATK UnifiedGenotyper with default quality thresholds at chosen intervals. From all detected SNPs, we used the 1.02 M SNPs that were fixed in the common ancestor of *A. lyrata* and *A. halleri* and called the parental origin of each mapped read in both hybrids.

We also observed that regions close to intron and/or highly divergent segments often failed to map properly. Our estimation of allelic abundance thus excluded gene regions that were less than 50 bp away from introns, or within regions with more than 10 fixed SNPs in 200 bp (greater than the mapping parameter of mismatch number 5 in 100 bp). This minimized the impact of possible allelic mapping biases caused by high divergence (and/or misannotated gene regions). We further excluded any SNP position with a coverage of less than 5 reads in the DNA samples from estimation of parental allelic ratios. Thus, 252,454 SNPs distributed in 16,419 orthologues remained for the estimation of allelic ratios in each of the two hybrid types (15.4 SNPs per orthologue). Final parental read counts at each SNP included both mapped full and remapped 30 bp reads and the allele ratio for each gene was determined as the median of the SNP ratio computed for all the SNPs of the respective gene. Finally, each parental read count was divided by a size factor (total read count for each orthologue/median total read counts for all orthologues following [Dillies et al. 2013]). supplementary fig. S3, Supplementary Material online shows that our algorithm effectively corrects for mapping biases: log allelic ratios in hybrid DNA samples show a distribution centered very close to 0, and the distinct hybrid types show similar variance (supplementary fig. S3, Supplementary Material online). This distribution fits well with a model of gene expression evolution based on infrequent mutations of moderate effect subjected to moderate selective constraints as described in Hodgins-Davis et al. (2015).

Nevertheless, due to segregating heterozygosity among AthxAha individuals, the number of genes showing ASE is smaller as there was probably lower power to detect ASE in these hybrids (supplementary fig. S5, Supplementary Material online).

We identified genes showing ASE with a GLM model including a quasi-binomial distribution of error. Under this model, the null hypothesis H_0 assumes no difference between RNA and DNA samples (Allele ratio of $\mu_{RNA} = \mu_{DNA}$), the alternative hypothesis H_1 assumes ASE at the gene considered (ASE, Allele ratio of $\mu_{RNA} \neq \mu_{DNA}$). Significant ASE was determined in a single GLM model nesting treatments within hybrid type. Loci with a mean final read count less than 10 were discarded from the analysis to exclude low-expressed genes. Significant ASE genes were defined by contrasting treatment-specific allelic ratios to DNA allelic ratios. An FDR threshold of 5% was fixed to call significant ASE genes.

To validate the accuracy of our RNAseq estimation of ASE, we used the pyrosequencing technology to measure allele-specific expression on the same RNA samples, at loci AT2G17260, AT3G28340, and AT1G65960, following the protocol described in de Meaux et al. (2005).

To quantify expression levels, reads were mapped on the *A. thaliana* genome only using the pipeline and parameters described above. This yielded an accurate mapping position for 90% of reads. Read counts were computed as number of FPKM. Genes were ranked by relative read count and a Spearman correlation coefficient was computed for each gene, to quantify the among-sample correlation in expression level. Genes displaying a significant change in expression level between treatments or hybrids were identified with the package DESeq2 (Love et al. 2014), testing for a treatment effect nested within hybrid type. *P* values were adjusted for controlling the FDR correction (Benjamini and Hochberg 1995).

Functional Enrichments among Derived and Ancestral ASE Mutations

Genes were first filtered for being expressed in both hybrids (scaled coverage >10) and then ranked according to ad hoc sorting values (SVs) designed to reflect the phylogenetic origins of the *cis*-acting mutations causing ASE. Genes with ASE derived in *A. lyrata* should show an allelic ratio in the AthxAly hybrid that differs significantly from the allelic ratio observed in the AthxAha hybrid. Using SV=|log2 AthxAly ratio|-|log2 AthxAha ratio|, *A. lyrata*-derived ASE genes ranked first. In contrast, using SV=|log2 AthxAha ratio|-|log2 AthxAly ratio|, ranked first the *A. halleri*-derived ASE genes. To rank genes that were most likely to have evolved before the split between the two species (or along the branch leading to *A. thaliana*), which we called ancestral here, we reasoned that ancestral ASE changes should show allelic ratios that are similar and

both differ from 1. The SV for ancestral *cis*-acting changes was designed using the following formula:

$$SR = \log_2(\text{AthxAly ratio}) + \log_2(\text{AthxAha ratio})$$

$$DR = \log_2(\text{AthxAly ratio}) - \log_2(\text{AthxAha ratio})$$

$$SV = \frac{|SR|}{\max(|SR|)} - \frac{|DR|}{\max(|DR|)}$$

In these rankings, all genes expressed in both hybrids formed the universe in which enrichment could be computed. A filter was further applied to each ranking to test functional enrichment against the alternative universe formed by all genes displaying ASE in at least one hybrid. For this, we removed all genes that did not display ASE in at least one hybrid (FDR adjusted *P* value $< 5\%$). This analysis removed potential functional enrichment that might be driven by the presence of genes subjected to strong constraints at the transcription level. Finally, we reasoned that polygenic evolution may require coordinated evolution of *cis*-acting changes increasing (or decreasing) the activity of genes involved in a common function. For this, we removed ASE genes with derived ASE change downregulating allelic expression in the *A. lyrata* lineage (i.e., genes with a significantly negative log2 AthxAly ratio, FDR adjusted *P* value $< 5\%$), to focus on derived *A. lyrata* ASE changes "upregulating" gene activity. Similarly, we removed ASE genes with derived ASE change upregulating allelic expression in the *A. lyrata* lineage (i.e., genes with a significantly positive log2 AthxAly ratio, FDR adjusted *P* value $< 5\%$), to focus on derived *A. lyrata* ASE changes "downregulating" gene activity. The same filters were also applied to detect enrichment among *A. halleri*-derived and ancestral ASE changes (supplementary table S5, Supplementary Material online).

Finally, we also analyzed functional enrichment among genes that showed a significant change in allelic ratio between treatments. In this case, ranking was not possible, so we analyzed enrichment within genes showing 1) changing allelic ratio between stress and standard in the AthxAly hybrid only—*A. lyrata* derived, 2) changing allelic ratio between stress and standard in the AthxAha hybrid only—*A. halleri*-derived, and 3) changing allelic ratio between stress and standard in both hybrid—ancestral change.

The GO enrichment analysis was performed with the GSEA test akin to nonparametric Kolmogorov–Smirnov tests, first described by Subramanian et al. 2005, and implemented in the "topGO" R package (Alexa and Rahnenfuhrer, 2010). We further applied the "elim" procedure, available in this package, which calculates enrichment significance of parent nodes after eliminating genes of significant children nodes. This controls for the dependency among nested parent–child GO categories so that the significance of each enrichment can be interpreted without overconservative *P* value corrections for multiple-testing (Alexa et al. 2006). We generated 100 permuted gene ranks to set a *P* value cutoff for calling significantly enriched GO categories. The *P* value distribution of significantly enriched GO categories was similar for all rankings and a *P* value cutoff was set at 0.009 to limit the FDR to 10%.

The GO annotations were obtained from the "org.At.tair.db" R package (version 3.1.2, http://bioconductor.org/). We excluded electronically inferred annotations, which are less reliable, by removing all annotations with the IEA evidence code. Enrichment was evaluated only for categories with five or more annotated genes in the universe. For heavy metal homeostasis, we used the list of experimentally supported genes described in supplementary table S6, Supplementary Material online proposed by Talke et al. 2006. For *A. thaliana* drought-responsive genes, we used the list of genes showing a significant change in expression in at least five *A. thaliana* ecotypes acclimated to low water availability (Des Marais et al. 2012). Finally, for cold-responsive genes, we used genes that were detected to change their expression in at least three independent studies classified as plant responses to short-, medium-, and long-term exposure to cold (Hannah et al. 2005). The list of genes are given in supplementary table S8, Supplementary Material online, along with a list of GO categories enriched among these stress response genes obtained by running fisher tests using the topGO package (Alexa and Rahnenfuhrer, 2010) and the *elim* method (Alexa et al. 2006) with the same filtered annotation package used on ASE rankings (supplementary table S9, Supplementary Material online). The GSEA test of heavy metal and stress-responsive genes on derived and ancestral gene ranks was performed using the clusterProfiler R package (Yu et al. 2012) and 1,000 random permutations. For stress-responsive genes, we analyzed enrichments separately for three subclasses: 1) all responsive genes, 2) downregulated genes, and 3) upregulated genes. We report in the text the subclass that is most significantly enriched.

Similarity Analysis of Enriched Functional Categories

The similarity between GO terms was calculated with the R package GOSemSim (Yu 2010); using the Wang method of graph-based semantic similarity (Wang et al. 2007). We calculated pairwise similarities between all GO terms harboring at least five genes expressed in both hybrids in at least one treatment (i.e., all GO terms that could have displayed an excess of differential ASE in our data). These similarity measures were converted into distances by subtracting them from one. Distances were used to perform a hierarchical clustering with average linkage as agglomerative method. The dendrogram of clustered GO categories was plotted using the R package ggplot2 (Wickham 2009). The significantly enriched GO categories that were present among the GO categories with a *P* value below 0.01 in any of the two stress-related gene sets analyzed (Hannah et al. 2005; Des Marais et al. 2012) were highlighted in this dendrogram.

Computation of Amino Acid Divergence (dN/dS)

The *A. halleri* genome sequence was obtained by correction of *A. lyrata* genome (Hu et al 2011) with the fixed 1.02 M SNPs described previously. The coding sequences (CDS) of 16,419 orthologues among three parents were selected according to genome annotation (TAIR10, Araly1). The

orthologous CDS sequences were aligned using MAFFT (Katoh and Standley 2013). The number of nonsynonymous mutations per nonsynonymous site (dN), the number of synonymous mutations per synonymous site (dS) and their ratio (dN/dS) were calculated with the formulae described in Li (1993) implemented in the "kaks" function of R package seqinr (Charif and Lobry 2007). We used the glm procedure implemented in R to test for the effect of dN/dS ratio and expression level (measured as standardized read number FPKM) on the following dependent variables: the probability to belong or not to an ASE enriched GO category in *A. lyrata* or *A. halleri*, or the probability to be ASE derived in *A. lyrata* or *A. halleri*. Because the dependent variables took the values 0 or 1, we tested the effect of dN/dS ratio and expression level on these dependent variables with a quasibinomial error distribution.

Supplementary Material

Supplementary figures S1–S10 and tables S1–S9 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank M. Lercher and U. Krämer for valuable advice and Christa Lanz from the Max Planck Institute of Developmental Biology in Tübingen for sequencing of hybrid DNA. This research was supported by the Deutsche Forschungsgemeinschaft (DFG), through programs SPP1529 "Adaptomics" (J.d.M.) and SFB 680 "Molecular Basis of Evolutionary Innovations" (to A.B.) and by the European Research Council (ERC-Co Grant "Adaptoscope") to J.d.M. Data are accessible on Gene Expression Omnibus GSE80462.

References

Akama S, Shimizu-Inatsugi R, Shimizu KK, Sese J. 2014. Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid Arabidopsis. *Nucleic Acids Res.* 42:e46.

Alexa A, Rahnenfuhrer J. 2010. topGO: enrichment analysis for Gene Ontology. R package version 2.20.0.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22:1600–1607.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.

Barreto FS, Pereira RJ, Burton RS. 2014. Hybrid dysfunction and physiological compensation in gene expression. *Mol Biol Evol.* 32:613–622.

Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat Rev Genet.* 12:767–780.

Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nat Rev Genet.* 3:11–21.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B-Methodol.* 57:289–300.

Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet.* 10:e1004412.

Bullard JH, Mostovoy Y, Dudoit S, Brem RB. 2010. Polygenic and directional regulatory evolution across pathways in Saccharomyces. *Proc Natl Acad Sci U S A.* 107:5058–5063.

Bulmer MG. 1985. The mathematical theory of quantitative genetics. Oxford (UK): Oxford University Press.

Chang J, Zhou Y, Hu X, Lam L, Henry C, Green EM, Kita R, Kobor MS, Fraser HB. 2013. The molecular mechanism of a cis-regulatory adaptation in yeast. *PLoS Genet.* 9:e1003813.

Charif D, Lobry JR. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Eduardo RomanH, Vendruscolo M, editors. Structural approaches to sequence evolution: molecules, networks, Populations. New York: Springer Verlag. p. 207–232.

Clark RM, Wagler TN, Quijada P, Doebley J. 2006. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat Genet.* 38:594–597.

Clauss MJ, Koch MA. 2006. Poorly known relatives of *Arabidopsis thaliana*. *Trends Plant Sci.* 11:449–459.

Clauss MJ, Mitchell-Olds T. 2006. Population genetic structure of *Arabidopsis lyrata* in Europe. *Mol Ecol.* 15:2753–2766.

Collins S, de Meaux J, Acquisti C. 2007. Adaptive walks toward a moving optimum. *Genetics* 176:1089–1099.

Collins S, de Meaux J. 2009. Adaptation to different rates of environmental change in *Chlamydomonas*. *Evolution* 63:2952–2965.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307:1928–1933.

Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* 24:797–808.

Courbot M, Willems G, Motte P, Arvidsson S, Roosens N, Saumitou-Laprade P, Verbruggen N. 2007. A major quantitative trait locus for cadmium tolerance in *Arabidopsis halleri* colocalizes with HMA4, a gene encoding a heavy metal ATPase. *Plant Physiol.* 144:1052–1065.

Cubillos FA, Stegle O, Grondin C, Canut M, Tisné S, Gy I, Loudet O. 2014. Extensive cis-regulatory variation robust to environmental perturbation in *Arabidopsis*. *Plant Cell* 26:4298–4310.

Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L. 2013. Evidence for polygenic adaptation to pathogens in the human genome. *Mol Biol Evol.* 30:1544–1558.

de Meaux J, Goebel U, Pop A, Mitchell-Olds T. 2005. Allele-specific assay reveals functional variation in the chalcone synthase promoter of *Arabidiopsis thaliana* that is compatible with neutral evolution. *Plant Cell* 17:676–690.

de Meaux J, Pop A, Mitchell-Olds T. 2006. Cis-regulatory evolution of chalcone-synthase expression in the genus *Arabidopsis*. *Genetics* 174:2181–2202.

De Pristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43:491–498.

Dean AM, Thornton JW. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet.* 8:675–688.

Des Marais DL, McKay JK, Richards JH, Sen S, Wayne T, Juenger TE. 2012. Physiological genomics of response to soil drying in diverse *Arabidopsis* accessions. *Plant Cell* 24:893–914.

Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, et al. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform.* 14:671–683.

Dräger DB, Desbrosses Fonrouge AG, Krach C, Chardonnens AN, Meyer RC, Saumitou-Laprade P, Krämer U. 2004. Two genes encoding Arabidopsis halleri MTP1 metal transport proteins co-segregate with zinc tolerance and account for high MTP1 transcript levels. *Plant J.* 39:425–439.

Emerson JJ, Hsieh LC, Sung HM, Wang TY, Huang CJ, Lu HHS, Lu MYJ, Wu SH, Li W-H. 2010. Natural selection on cis and trans regulation in yeasts. *Genome Res.* 20:826–836.

Emerson JJ, Li W-H. 2010. The genetic basis of evolutionary change in gene expression levels. *Philos Trans R Soc Lond B Biol Sci.* 365:2581–2590.

Falke KC, Glander S, He F, Hu J, de Meaux J, Schmitz G. 2013. The spectrum of mutations controlling complex traits and the genetics of fitness in plants. *Curr Opin Genet Dev*. 23:665–671.
Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* 334:86–89.
Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Res*. 23:1089–1096.
Fraser HB, Levy S, Chavan A, Shah HB, Perez JC, Zhou Y, Siegal ML, Sinha H. 2012. Polygenic cis-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Res*. 22:1930–1939.
Hanikenne M, Kroymann J, Trampczynska A, Bernal M, Motte P, Clemens S, Krämer U. 2013. Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genet*. 9:e1003707.
Hanikenne M, Talke IN, Haydon MJ, Lanz C, Nolte A, Motte P, Kroymann J, Weigel D, Kramer U. 2008. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature* 435:391–395.
Hannah MA, Heyer AG, Hincha DK. 2005. A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. *PLoS Genet*. 1:e26.
He F, Zhang X, Hu J, Turck F, Dong X, Goebel U, Borevitz J, de Meaux J. 2012. Genome-wide analysis of cis-regulatory divergence between species in the *Arabidopsis* genus. *Mol Biol Evol*. 29:3385–3395.
Hill WG. 2009. Understanding and using quantitative genetic variation. *Philos Trans R Soc Lond B Biol Sci*. 365:73–85.
Hodgins-Davis A, Rice DP, Townsend JP. 2015. Gene expression evolves under a house-of-cards model of stabilizing selection. *Mol Biol Evol*. 32:2130–2140.
Hoffmann MH. 2005. Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* 59:1425–1436.
Hu TT, Pattyn P, Bakker E, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet*. 43:476–481.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.
Kazemi Dinan A, Sauer J, Stein RJ, Krämer U, Müller C. 2015. Is there a trade-off between glucosinolate-based organic and inorganic defences in a metal hyperaccumulator in the field? *Oecologia* 178:369–378.
Kazemi Dinan A, Thomaschky S, Stein RJ, Krämer U, Müller C. 2014. Zinc and cadmium hyperaccumulation act as deterrents towards specialist herbivores and impede the performance of a generalist herbivore. *New Phytol*. 202:628–639.
Kim YY, Choi H, Segami S, Cho HT, Martinoia E, Maeshima M, Lee Y. 2009. AtHMA1 contributes to the detoxification of excess Zn(II) in Arabidopsis. *Plant J*. 58:737–753.
Kimura M. 1968. Evolutionary rate at the Molecular Level. *Nature* 217:624–626.
King MC, Wilson AC. 1975. Evolution at 2 levels in humans and chimpanzees. *Science* 188:107–116.
Kingsolver J, Hoekstra H, Hoekstra J, Berrigan D, Vignieri S, Hill C, Hoang A, Gibert P, Beerli P. 2001. The strength of phenotypic selection in natural populations. *Am Nat*. 157:245–261.
Kopp M, Hermisson J. 2009. The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. *Genetics* 182:233–249.
Krämer U. 2010. Metal hyperaccumulation in plants. *Annu Rev Plant Biol*. 61:517–534.
Lemmon ZH, Bukowski R, Sun Q, Doebley JF. 2014. The role of cis regulatory evolution in maize domestication. *PLoS Genet*. 10:e1004745.
Li WH. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol*. 36:96–99.
Lin Y-F, Liang H-M, Yang S-Y, Boch A, Clemens S, Chen C-C, Wu J-F, Huang J-L, Yeh K-C. 2009. Arabidopsis IRT3 is a zinc-regulated and plasma membrane localized zinc/iron transporter. *New Phytol*. 182:392–404.
Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009. On the origin and spread of an adaptive allele in deer mice. *Science* 325:1095–1098.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.
Lynch M. 2007. The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet*. 8:803–813.
Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M, Nambara E, Nakajima M, Kawashima M, et al. 2008. Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol*. 49:1135–1149.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20:1297–1303.
McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res*. 20:816–825.
Morrissey MB, Hadfield JD. 2012. Directional selection in temporally replicated studies is remarkably consistent. *Evolution* 66:435–442.
Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet*. 39:197–218.
Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev IV, Dubchak I. 2014. The genome portal of the department of energy joint genome institute: 2014 updates. *Nucleic Acids Res*. 42:D26–D31.
Oomen RJFJ, Wu J, Lelièvre F, Blanchet S, Richaud P, Barbier Brygoo H, Aarts MGM, Thomine S. 2009. Functional characterization of NRAMP3 and NRAMP4 from the metal hyperaccumulator *Thlaspi caerulescens*. *New Phytol*. 181:637–650.
Orr HA. 2002. The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 56:1317–1330.
Pauwels M, Saumitou-Laprade P, Holl AC, Petit D, Bonnin I. 2005. Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. *Mol Ecol*. 14:4403–4414.
Picotti P, Clément-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW, Röst H, Sun Z, Rinner O, Reiter L, et al. 2013. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 494:266–270.
Plaza S, Weber J, Pajonk S, Thomas J, Talke IN, Schellenberg M, Pradervand S, Burla B, Geisler M, Martinoia E, et al. 2015. Wounding of *Arabidopsis halleri* leaves enhances cadmium accumulation that acts as a defense against herbivory. *Biometals* 28:521–528.
Prasad KVSK, Song B-H, Olson-Manning C, Anderson JT, Lee C-R, Schranz ME, Windsor AJ, Clauss MJ, Manzaneda AJ, Naqvi I, et al. 2012. A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* 337:1081–1084.
Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A*. 104(Suppl 1): 8605–8612.
Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* 440:1050–1053.
Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Chen R, et al. 2011. Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333:1137–1141.
Rest JS, Morales CM, Waldron JB, Opulente DA, Fisher J, Moon S, Bullaughey K, Carey LB, Dedousis D. 2013. Nonlinear fitness consequences of variation in expression level of a eukaryotic gene. *Mol Biol Evol*. 30:448–456.
Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet*. 44:217–220.
Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, et al. 2002. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-

salinity stresses using a full-length cDNA microarray. *Plant J.* 31:279–292.
Stebbins GL. 1952. Aridity as a stimulus to plant evolution. *Am Nat.* 86:33–44.
Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T. 2015. Cis-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella*. *Mol Biol Evol.* 32:2501–2514.
Steiner CC, Weber JN, Hoekstra HE. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* 5:e219.
Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155–2177.
Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol.* 14:671–688.
Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 102:15545–15550.
Talke IN, Hanikenne M, Krämer U. 2006. Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator *Arabidopsis halleri*. *Plant Physiol.* 142(1):148–167.
Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 28:516–520.
Videvall E, Sletvold N, Hagenblad J, Ågren J, Hansson B. 2016. Strong maternal effects on gene expression in *Arabidopsis lyrata* hybrids. *Mol Biol Evol.* 33:984–994.
Wang JZ, et al. 2007. A new method to measure the semantic similarity of GO terms. *Bioinformatics* 23(10):1274–1281.
Wickham H. 2009. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.
Willems G, Dräger DB, Courbot M, Godé C, Verbruggen N, Saumitou-Laprade P. 2007. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. *Genetics* 176:659–674.
Willems G, Frérot H, Gennen J, Salis P, Saumitou-Laprade P, Verbruggen N. 2010. Quantitative trait loci analysis of mineral element concentrations in an *Arabidopsis halleri* x *Arabidopsis lyrata* petraea F2 progeny grown on cadmium-contaminated soil. *New Phytol.* 187:368–379.
Wittkopp PJ, Haerum BK, Clark AG. 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet.* 40:346–350.
Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 8:206–216.
Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.* 20:1377–1419.
Yeaman S. 2015. Local adaptation by alleles of small effect. *Am Nat.* 186:S74–S89.
Yu G. 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26(7):976–978.
Yu G, Wang L, Han Y, He Q. 2012. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* 16:284–287.
Zhang X, Borevitz JO. 2009. Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics* 182:943–954.