DOI: 10.1534/genetics.106.064543

Cis-regulatory Evolution of *Chalcone-Synthase* Expression in the Genus Arabidopsis

Juliette de Meaux,*,†,[1] A. Pop† and T. Mitchell-Olds‡

**Genetics and Plant Breeding, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany,* †*Genetics and Evolution, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany and* ‡*Department of Biology, Duke University, Durham, North Carolina 27708-0338*

Manuscript received August 10, 2006
Accepted for publication September 26, 2006

ABSTRACT

The contribution of *cis*-regulation to adaptive evolutionary change is believed to be essential, yet little is known about the evolutionary rules that govern regulatory sequences. Here, we characterize the short-term evolutionary dynamics of a *cis*-regulatory region within and among two closely related species, *A. lyrata* and *A. halleri*, and compare our findings to *A. thaliana*. We focused on the *cis*-regulatory region of chalcone synthase (*CHS*), a key enzyme involved in the synthesis of plant secondary metabolites. We observed patterns of nucleotide diversity that differ among species but do not depart from neutral expectations. Using intra- and interspecific F_1 progeny, we have evaluated functional *cis*-regulatory variation in response to light and herbivory, environmental cues, which are known to induce *CHS* expression. We find that substantial *cis*-regulatory variation segregates within and among populations as well as between species, some of which results from interspecific genetic introgression. We further demonstrate that, in *A. thaliana*, *CHS cis*-regulation in response to herbivory is greater than in *A. lyrata* or *A. halleri*. Our work indicates that the evolutionary dynamics of a *cis*-regulatory region is characterized by pervasive functional variation, achieved mostly by modification of response modules to one but not all environmental cues. Our study did not detect the footprint of selection on this variation.

THE rhythm, dynamics, and location of gene expression are fundamentally important for development of phenotypes. Transcription is controlled in part by the interaction of regulatory proteins (*trans*-regulatory factors) with specific DNA regions (*cis*-regulatory DNA regions or promoters). In contrast with the explicit constraints imposed on protein-coding DNA by the genetic code, the functional architecture of *cis*-regulatory DNA is less apparent. Expression variation is widespread within and between species (Oleksiak *et al.* 2002; Becher *et al.* 2004; Khaitovich *et al.* 2004; Kliebenstein *et al.* 2006), and subtle changes in expression can significantly affect the phenotype (Wang *et al.* 1999; Gompel *et al.* 2005). Accordingly, *cis*-regulatory DNA is thought to play a prominent role in adaptive evolution (King and Wilson 1975; Wray *et al.* 2003). Rewiring the regulatory network through *cis*-changes may indeed allow the generation of phenotypic novelties while simultaneously preserving key physiological and developmental functions.

Our current understanding of *cis*-regulatory evolution is largely based on patterns of DNA conservation. There is now clear evidence that function in noncoding DNA (ncDNA) is broadly maintained. Whole-genome sequence comparisons between species have uncovered numerous segments of conserved noncoding DNA (Dermitzakis *et al.* 2004). Constraints on conserved segments are often experimentally related to functional conservation (Clifton *et al.* 2001; Koch *et al.* 2001; Boffelli *et al.* 2003). Interestingly, levels of constraint in ncDNA are found to vary across species (Keightley and Gaffney 2003; Keightley *et al.* 2005); they also can be larger than in protein-coding regions (Bejerano *et al.* 2004). However, opportunities for adaptive evolution and lineage-specific functional changes remain poorly understood.

Recently, a study of nucleotide polymorphism and divergence within and between two Drosophila species over a large number of noncoding DNA loci suggested that many ncDNA changes, mostly located in UTRs, may have undergone adaptive evolution (Andolfatto 2005). In humans, multiple instances of adaptive changes have been observed at specific *cis*-regulatory loci (Bamshad *et al.* 2002; Rockman *et al.* 2003; Hahn *et al.* 2004; Rockman *et al.* 2004), although some examples remain controversial (Sabeti *et al.* 2005). In other species, including Drosophila, examples of neutral nucleotide variation at *cis*-regulatory regions have been reported (Balhoff and Wray 2005; Fay and Benavides 2005; Macdonald and Long 2005).

Overall, the relationship between nucleotide and functional variation in *cis*-regulatory DNA has rarely been

[1]*Corresponding author:* Max Planck Institute for Plant Breeding Research, Carl-von-Linné Weg 10, 50829 Cologne, Germany.
E-mail: demeaux@mpiz-koeln.mpg.de

characterized, and little is known about the amount and type of variation to be expected within and among closely related species. In-depth characterization of *cis*-regulatory variation at both nucleotide and functional levels is required to gain some insight into the baseline evolutionary scenario of functional noncoding regulatory DNA. The most compelling models of *cis*-regulatory evolution are based on Drosophila developmental genes that are controlled by internal signals and whose misexpression is fatal to the organism (Phinchongsakuldit *et al.* 2004; Ludwig *et al.* 2005). So far, the generation of *cis*-regulatory novelties in a less constrained expression context has received little attention. In comparison to animals, plants continuously fine tune their development to prevailing environmental conditions. Thus, plant models of *cis*-regulatory evolution in genes controlled by environmental signals may shed light on the possible adaptive role of *cis*-regulatory variation.

Previously, we examined regulatory polymorphisms within *Arabidopsis thaliana*. We established a robust allele-specific assay to examine *cis*-regulatory variation in response to abiotic, biotic, or developmental changes (de Meaux *et al.* 2005). This assay of allele-specific expression in F_1 heterozygotes effectively controls for background variation and allows the detection of subtle *cis*-regulatory differences. We focused on the promoter region of the *chalcone synthase* (*CHS*) gene because it is among the best-characterized promoters in plants and its expression is induced by multiple cues (Hartmann *et al.* 1998; Koch *et al.* 2001; Logemann and Hahlbrock 2002). Among those cues, light and insect herbivory were shown to upregulate *CHS* expression in *A. thaliana* (Reymond *et al.* 2000; Jenkins *et al.* 2001; Wade *et al.* 2001). In addition, *CHS* is the branch-point enzyme of a pathway involved in the interaction between plants and their abiotic and biotic environments (Winkel-Shirley 2001). Hence this gene is likely to play a role in adaptive evolution (see Johnson and Dowd 2004). In *A. thaliana*, we found substantial functional *cis*-regulatory variation in *CHS* expression. However, patterns of nucleotide variation in the *A. thaliana CHS* promoter showed no evidence of non-neutral evolution in this inbreeding annual species (de Meaux *et al.* 2005).

In this article, we elucidate the evolutionary dynamics of *CHS cis*-regulation in Arabidopsis and report microevolutionary patterns of *cis*-regulation. These experiments were conducted with *A. halleri* and *A. lyrata*, two self-incompatible species that differ substantially in their ecology (Mitchell-Olds 2001). In central Europe, *A. halleri* grows in highly competitive open meadows, whereas *A. lyrata* is restricted to low-competition habitats on exposed rocks (Hoffmann 2005). We compared our findings to data from the model species *A. thaliana* (de Meaux *et al.* 2005), which differs from other species in the genus by its self-compatibility and low species-wide levels of diversity (Wright *et al.* 2003; Schmid *et al.* 2005).

We analyzed allele-specific expression levels in the progeny of intra- and interspecific crosses and evaluated functional polymorphism and divergence of *CHS cis*-regulation. We combined our functional assay with an analysis of polymorphism and divergence at the nucleotide level in the *CHS* 5′ upstream intergenic region and addressed the following questions: (i) How does *cis*-regulatory diversity in *A. lyrata* and *A. halleri* compare to that in *A. thaliana*?, (ii) What qualitative and quantitative divergence in *CHS cis*-regulation is seen among species?, and (iii) Is a footprint of selection detectable in the intergenic region containing the *CHS* promoter in any of the three Arabidopsis species examined?

At the nucleotide level, the evolutionary dynamics of the *CHS* promoter region differed among the Arabidopsis species examined, in agreement with their genomewide patterns of diversity. Thus, patterns of variation at the *CHS* promoter region show no indication of adaptive evolution. Nonetheless, patterns of functional diversity point to abundant *cis*-regulatory variation within and between species, most of which results from qualitative differences in the response to individual environmental cues. Our results reveal that *CHS cis*-regulation evolves mainly by modification of *cis*-regulatory response modules to one but not all environmental cues.

MATERIALS AND METHODS

Sequencing: The 5′ flanking region at the *CHS* gene (henceforth referred to as the "intergenic region") was sequenced from 15 and 8 individuals in *A. lyrata* and *Arabidopsis halleri*, respectively. All accessions are diploid and their geographic origin is described in Table 1. Each accession was named as follows: the two letters indicate the species, the first number indicates the population, the second number the individual in the population (if several individuals were studied in the population), and the number after the hyphen describes the allele (if more than one allele was uncovered in a given individual). All *A. lyrata* accessions were provided by M. J. Clauss (Max Planck Institute for Chemical Ecology, Jena, Germany) with the exception of AL11, AL12, and AL10, which were provided by T. Mitchell-Olds and T. Sharbel (Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany). *A. halleri* accessions were collected by M. J. Clauss (AH4 and AH5) and T. Mitchell-Olds. P. Saumitou-Laprade (University of Lille, Lille, France) provided the seeds for AH11, AH12, AH21, and AH22. Young leaves from each accession were ground in liquid nitrogen and DNA was subsequently purified following standard CTAB protocol. *CHS* is single copy in the *A. thaliana* genome (Koch *et al.* 2000). A Southern blot analysis confirmed that *CHS* occurs also as a single copy in *A. lyrata* and *A. halleri* (not shown). To amplify the *CHS* intergenic region in *A. lyrata* and *A. halleri*, we used the annotated *A. thaliana* genome to design a forward primer in the closest adjacent putative open reading frame (ORF) 5′ upstream from *CHS* (5′-AGGACAATCGTTGATCCAG-3′) and a reverse primer in the first *CHS* exon (5′-GTAGTCAGGATACTCCGC-3′). The adjacent ORF (annotated AT5G13920) belongs to the Zinc knuckle protein family (http://www.Arabidopsis.org). The PCR was conducted as previously described with the exception of the use of a 54° annealing temperature for PCR cycling (de Meaux *et al.* 2005). Two independent PCRs were

TABLE 1

Geographical origin of the individuals analyzed and levels of diversity found within populations

Accession	Origin	No. of alleles	Length of intergenic region	Genotypes used for *cis*-regulatory diversity assay	Level of nucleotide variation within population π_w
		A. *A. lyrata*			
AL11	North America	2	1825/1816		0.003
AL12	North America	1	1816	x	
AL21	Lilienfeld, Austria	1	3399		
AL22	Lilienfeld, Austria	1	2159	x	0.008
AL23	Lilienfeld, Austria	1	1254		
AL3	Schaeftal, Austria	1	1255	x	
AL41	Plech, Germany	1	1902	x	0.002
AL42	Plech, Germany	2	1913/1895		
AL51	Stolberg, Germany	1	1288		0.002
AL52	Stolberg, Germany	2	1288/1253	x	
AL6	Voesslauer Huete, Austria	2	1262/1273		0.004
AL7	Spitertuten, Sweden	1	1890	x	
AL8	Mjalöm, Sweden	1	1886		
AL9	Karhumäki, Russia	2	1248/1248		0.007
AL10	Isle of Sky, Scotland	1	1626	x	
Total of 15 accessions		19			
		B. *A. halleri*			
AH11	Belgium	1	1255		0.000
AH12	Belgium	1	1255	x	
AH21	Mortagne, France	1	1085		0.007
AH22	Mortagne, France	2	1091/1099	x	
AH3	Czech Republic	1	1704	x	
AH4	Rodacherbruenn, Germany	2	1098/1656	x	0.047
AH5	Sieber, Germany	1	1096	x	
AH6	Schierke, Germany	2	1093/1098		0.011
Total of 8 accessions		11			

performed and products were cloned using a TOPO TA cloning kit (Invitrogen Life Technologies, Paisley, UK). Six clones perPCR were sequenced on one strand with an ABI3700 capillary sequencer using primers placed approximately every 500 bp. Sequences were assembled with Seqman 5.0 (DNASTAR) and each variable site was checked by examining sequence chromatograms. Sites found to be polymorphic across clones obtained from separate PCRs indicated the segregation of two alleles in the PCR pool. Allele sizes were checked by RFLP in *A. halleri* (not shown). In *A. lyrata*, genotyping assays based on singletons detected a second allele in AL22 (henceforth called "unknown AL22 allele") but not in AL12. The *CHS* exons 1 and 2 were sequenced following Ramos-Onsins *et al.* (2004) in the 7 *A. lyrata* and 5 *A. halleri* parental genotypes used for the expression assay (Table 1). Sequences are available from the EMBL Nucleotide Sequence Database under accession nos. AM296511–AM296543.

Population genetic analyses: We assumed that individuals for which a single sequence was obtained carried two identical alleles. Sequences were aligned with Megalign 5.03 (DNASTAR). The DnaSP 4.0 program (Rozas and Rozas 1999) was used for both intra- and interspecific analyses of nucleotide polymorphism (Table 2). Deviations from panmixia, *e.g.*, population subdivision, violate basic assumptions of most neutrality tests. We investigated the existence of genetic differentiation using the S_{nn} estimator developed by Hudson *et al.* (1992) and implemented in DnaSP. This estimator is a nucleotide-sequence-based measure of genetic differentiation between populations. Significance of S_{nn} was tested using 1000 permutations. Following Ramos-Onsins *et al.* (2004), we chose one sequence per location if significant genetic differentiation was detected in our sample to perform neutrality tests on the basis of species-wide estimates of diversity. In *A. halleri*, two summary values of variation are reported (Table 2) because distinct subsamples sometimes yielded markedly different values. This was not the case for *A. lyrata*. In the reduced sample, per-site nucleotide diversity is described as π_t (between populations). Per-site nucleotide diversity was also computed within populations (π_w). Species-wide patterns of nucleotide polymorphism were summarized by

TABLE 2

Summary statistics

	A. lyrata		*A. halleri*[c]		
	5′ upstream intergenic region (1375 bp, 10 sequences)[a]	Coding region (exon 1, intron 1, exon 2): 1220 bp, 10 sequences	5′ upstream intergenic region (1117 bp, 6 sequences)		Coding region (exon 1, intron 1, and exon 2): 1220 bp, 14 sequences
			Set1	Set2	
Informative sites	20	20	12	36	2
Singletons	14	4	35	7	13
No. of nonsynonymous mutations	—	2	—	—	0
Haplotypes	9	8	5	5	4
Haplotype diversity (SE)	0.978 (0.0292)	0.933 (0.00597)	0.933 (0.015)	0.933 (0.015)	0.495 (0.02267)
Average no. of nucleotide differences	12.778	8.711	18.5	24.4	2.484
π_t	0.0107	0.00714	0.021	0.025	0.00204
θ_w	0.0102	0.00695	0.023	0.022	0.00387
Total no. of observed polymorphisms	34	24	47	43	15
Tajima's *D*	0.225 ($P > 0.1$)	0.12736 ($P > 0.5$)	−0.76 ($P > 0.2$)	0.96 ($P > 0.8$)	−1.93071 ($P = 0.009$)
Wall's *B*	0.121 ($P > 0.2$)	0.0625 ($P > 0.1$)	0.622 ($P > 0.9$)	0.667 ($P > 0.9$)	0.743 ($P = 0.967$)
No. of polymorphic sites	17[b]	23	31	26	14
No.of fixed differences from *A. thaliana*	63[b]	55	49	42	53
No. of base pairs compared for analyses with outgroup	774	1214	753	720	1214
Nucleotide divergence *vs.* *A. thaliana*, K_s (with Jukes and Cantor correction)	0.096	0.182	0.09	0.081	0.182
Fay and Wu's *H* (outgroup *A. thaliana*)	0.444 ($P > 0.3$)	−4.622 ($P > 0.1$)	−7.2 ($P = 0.09$)	−2.93 ($P > 0.1$)	−12.83 ($P = 0.001$)
HKA (*A. thaliana* outgroup), *P*-value	$\chi^2 = 0.413$, $P > 0.5$		$\chi^2 = 2.057$, $P > 0.1$	$\chi^2 = 2.140$, $P > 0.1$	

[a] Analysis after reducing sample size, following the exclusion of sequences obtained from the same population.
[b] Number of intraspecific polymorphism on the sequence portions alignable with *A. thaliana*.
[c] For *A. halleri* promoter, summary statistics were calculated for two subsamples containing one allele per population.

various test statistics: Tajima's *D*, based on the differences between two estimators of intraspecific diversity, and Fay and Wu's *H*, which makes use of an outgroup sequence to analyze the frequency of derived polymorphisms (Tajima 1989; Fay and Wu 2000). Associations among nucleotide variants can be summarized by Wall's *B* statistics, which evaluate the proportion of adjacent segregating sites that partition equally the sequence sample (Wall 1999). Such sites occur along the same branch of the coalescent tree and the test examines whether branch length is compatible with the neutral equilibrium model. The compatibility of *H* and *B* statistics with evolution under the standard neutral model was tested by 1000 coalescent simulations. The Hudson, Kreitman, and Aguadé (HKA) test is based on the prediction that, for a particular region of the genome, the rate of divergence between species is proportional to the levels of polymorphism within species (Hudson *et al.* 1987). This test compares the ratio of intraspecific polymorphism to interspecific divergence in two loci. HKA tests were performed for silent positions using silent segregating sites and the silent divergence value (Nei 1987) to compare the intergenic region and the *CHS* coding region. In the intergenic region, all positions were considered to be silent. These four neutrality tests (Tajima's *D*, Fay and Wu's *H*, Wall's *B*, and HKA) focus on different characteristics of nucleotide polymorphism and thus effectively summarize the evolutionary history of the *CHS* intergenic region. Intergenic sequences were examined for known transcription-factor-binding sites as previously described (de Meaux *et al.* 2005). Gene conversion tracts between alleles identified in the three different species were searched using the algorithm described by Betran *et al.* (1997) and implemented in DnaSP. This algorithm uses the frequency of a nucleotide at a site to determine if the site is informative to detect a conversion event between groups of sequences. The length of the conversion event is determined by the distance between informative sites (Betran *et al.* 1997). We investigated the gene genealogy among haplotypes using the program TCS, to account for population-level phenomena such as recombination or persistence of the ancestral haplotype (Clement *et al.* 2000).

Allele-specific quantification of *CHS* expression in F_1: *Plant material used for crosses:* The accessions used to perform crosses were chosen from different populations covering a representative part of the species ranges. Whenever possible, accessions found to be homozygous in the intergenic region were chosen.

Because both *A. halleri* and *A. lyrata* are self-incompatible, crosses were performed by simply rubbing stamen of the paternal genotype with the pistil of the maternal genotype. In *A. lyrata*, seven accessions were used to generate 17 F_1 progeny, 10 of which yielded enough individuals for statistical analysis (supplemental Table 1a at http://www.genetics.org/supplemental/). In *A. halleri*, five accessions were used to generate 8 F_1 progeny large enough to be analyzed statistically (supplemental Table 1b at http://www.genetics.org/supplemental/). Additional combinations could not be used, either because crosses remained unsuccessful (presumably due to self-incompatibility) or because parental alleles could not be differentiated by any polymorphism. When possible, reciprocal crosses were performed to control for maternal effects (see supplemental Table 1, a and b, at http://www.genetics.org/supplemental/).

We further obtained four *A. thaliana* × *A. lyrata* hybrid progeny from three *A. thaliana* (Ei-2, Kas-1, and Ag-0) and four *A. lyrata* parental genotypes (AL3, AL22, AL52, AL41) and three *A. thaliana* × *A. halleri* hybrid progeny from two *A. thaliana* (Kas-1, Ka-0) and two *A. halleri* parental genotypes (AH4, AH12) for a total of 81 and 25 hybrids of each type (see supplemental Table 1c at http://www.genetics.org/supplemental/). Only crosses using *A. thaliana* as a mother were successful and all crosses were not equally successful (see supplemental Table 3 at http://www.genetics.org/supplemental/).

Seeds were sown on humid filter paper in small petri dishes and vernalized at 4° in the dark for 2 weeks, followed by 2 weeks in Voetsch reach-in chambers (12-hr day, 20° day temperature, 16° night temperature, 70% humidity) for germination. The whole procedure was repeated until germination was successful. One-week-old seedlings were transplanted into single pots and assigned to random positions in York walk-in growth chambers in the following conditions: 11-hr day, 21° day temperature, 16° night temperature.

CHS expression experiments: In *A. thaliana*, *CHS* gene expression is repressed in the dark and strongly induced in the light (Jenkins *et al.* 2001). *CHS* expression is also induced upon feeding by *Plutella xylostella* larvae (H. Vogel and T. Mitchell-Olds, unpublished results). To assess the *cis*-regulatory diversity of *CHS* expression in response to these various environmental cues, plants either were placed for 48 hr in the dark followed by 8 hr of strong light or were challenged with *P. xylostella* larvae during 24 hr, following de Meaux *et al.* (2005).

Three- to 6-month-old plants were used for the *CHS* expression experiments. For intraspecific crosses, *CHS* expression experiments were performed in two independent trials separated by a 2-week interval. Half of the progeny of each cross were randomly attributed to one or the other trial. Within each trial, half of the progeny were randomly assigned to be sampled for *CHS* expression in the dark and the other half for *CHS* expression in the light. All plants, however, underwent the entire dark/light treatment. Insect-feeding experiments were carried out at least 4 weeks after the light experiment. Likewise, within each trial, half of the progeny were randomly assigned to be sampled for *CHS* expression in insect-challenged leaves and the other half for *CHS* expression in control leaves of insect-free plants.

For the interspecific progeny, *CHS* expression experiments were conducted in the same way but in a single trial. For the *A. thaliana–A. lyrata* F_1 progeny, flowers were harvested at the time point between the end of the dark period and the beginning of the light period on the plants that had been attributed to the dark treatment, to look at *CHS* expression independently from the influence of light. *CHS* is known to be specifically upregulated in *A. thaliana* flowers where flavonoids are produced abundantly (Burbulis *et al.* 1996). Due to delayed flowering, flower-specific *CHS* expression was not studied in the *A. thaliana–A. halleri* progeny.

Quantitative analysis of allele-specific *CHS* expression: Approximately 2 sq cm of leaf material (or two flower buds) was harvested. RNA extraction and cDNA synthesis were performed as described previously (de Meaux *et al.* 2005). Allele-specific *CHS* mRNA was quantified using the quantitative properties of pyrosequencing (Neve *et al.* 2002; de Meaux *et al.* 2005). To control for possible position effects in the thermocycler, cDNA samples together with DNA extracted from heterozygous plants were randomly distributed across 96-well plates prior to PCR. For the *A. lyrata* progeny, the number of samples allowed a hierarchical randomization of cDNA samples across plates within a given trial. For the *A. halleri* progeny, no hierarchical randomization was performed. The pyrosequencing reactions were performed using the PyrosequencerAB device (Biotage, Uppsala, Sweden) as previously described (de Meaux *et al.* 2005).

In each two species, three single nucleotide polymorphisms (SNPs) located in the *CHS* coding region were used to measure allele-specific *CHS* expression: SNP1008 (PCR primers 5′-TCGGTCAGGCTCTTTTCAGTG-3′and 5′-TGTCCGTCTATGGCACCATC-3′, sequencing primer 5′-GGGAGGATGGTCTGT-3′), SNP572 (PCR primers 5′-GGAAACGCCACATGCATCTG-3′ and 5′-TCCTTGATGGCCTTCACTGC-3′, sequencing primer 5′-TTAGGGACTTCAACC-3′), and SNP591 (PCR primers 5′-GGAAACGCCACATGCATCTG-3′ and 5′-TCCTTGATGGCCTTCACTGC-3′, sequencing primer 5′-CTGCCGCTTCTTTGCC-3′) in *A. lyrata*; SNPB6 (PCR primers 5′-GGAAACGCCACATGCATCTG-3′ and 5′-TCCTTGATGGCCTTCACTGC-3′, sequencing primer 5′-TAAGCGCACATGTGTGG-3′), SNPM6 (PCR primers 5′-GGAAACGCCACATGCATCTG-3′ and 5′-TCCTTGATGGCCTTCACTGC-3′, sequencing primer 5′-GATGTCCTGTCGGGTG-3′), and SNPCZ (PCR primers 5′-GACCGACCTCAAGGAGAAG-3′and 5′-TTGATGGCCTTCACTGCCG-3′, sequencing primer 5′-CTAGCTTAGGGACTTCA-3′) in *A. halleri*; for *A. thaliana–A. lyrata* hybrids, we used SNP1230 (PCR primers 5′-ACCTTCCATCTCCTCAAGG-3′ and 5′-CTCTTCCTTTAGTCCTAGC-3′, sequencing primer 5′-CCTTTAGTCCTAGCTT-3′) and SNP587 (PCR primers 5′-GACCGACCTCAAGGAGAAG-3′ and 5′-TTGATGGCCTTCACTGCCG-3′, sequencing primer 5′-CCTAGCTTAGGGAC-3′); for *A. thaliana–A. halleri* hybrids, we used SNP587 and SNP1370 (PCR primers 5′-AGGTGGAGATAAAGCTAGG-3′ and 5′-AAGACACCCCACTCCAACCC-3′, sequencing primer 5′-CTCCAACCCTTCTCCT-3′). Only half of the progeny of AL22xAL12, AL22xAL3, AH22xAH4, and AH22xAH5 was analyzed due to heterozygosity of either the AL22 parent for SNP591 or the AH22 parent for SNP M6. Promoter allele genotyping indicated that individuals analyzed in these progeny harbored either the sequenced AL22 or the AH22-1 alleles at the *CHS* intergenic region described below (Table 1). The genotyping assay is described below. For several progeny, as well as for interspecific hybrids, expression data were analyzed using two independent SNP assays. Data obtained with different SNP assays were all significantly correlated (minimum $P = 0.027$, Table 3). The strength of the correlation between SNP assays depends on the overall amount of *cis*-regulatory variation in the progeny.

Ratio of polymorphic over monomorphic sequencing peaks were deduced from pyrosequencing measurements, which provided an estimation of relative allelic concentration in mRNA pools. The ratios were calibrated as previously described by Wittkopp *et al.* (2004). For interspecific crosses, a marked PCR bias was observed for all three SNP assays, and the standard curve was better modeled by a second-degree polynomial equation. This might result from the higher sequence divergence of orthologous mRNAs and may partly explain the higher variance of the pyrosequencing measurement

TABLE 3

Correlation between SNP assays for parental allelic combinations harboring two SNP differences in the CHS coding region

	Species	Genotypes	R[a]	*P*
SN P591–SN P572	*A. lyrata*	AL41xAL22	0.609	0.027
SNP1008–SN P572	*A. lyrata*	AL41xAL7	0.510	<0.001
SNP1008–SN P572	*A. lyrata*	AL22xAL10	0.982	<0.001
SNPB6–SNPCZ	*A. halleri*	AH12xAH3	0.907	<0.001

[a] Pearson correlation coefficient.

observed in the quantification of species-specific *CHS* mRNA levels in interspecific F_1 hybrids (see below).

Evaluation of methylation in the *CHS* intergenic region: Approximately 2 g of leaves from natural genotypes Ei-2 (*A. thaliana*), AH4 (*A. halleri*), AL52 (*A. lyrata*), and from six and seven *A. thaliana*–*A. halleri* and *A. thaliana*–*A. lyrata* hybrids, respectively, were collected. DNA was extracted using the Midi-Prep DNA extraction kit (QIAGEN, Valencia, CA). Quantitative evaluation of methylation was performed at three CpG sites by M. Pettersson at Biotage (Premium CpG Methylation Service, Uppsala, Sweden) at two CpG sites located within the core promoter. The two CpG sites are located 2 bp upstream and within the A-box. The A-box is an essential element of the core promoter located between two equally essential elements (the MRE and ACE elements; Logemann and Hahlbrock 2002). These two CpG sites correspond to positions 1342 and 1346 on the alignment provided in the supplemental data at http://www.genetics.org/supplemental/.

Individual genotyping in progeny from heterozygous parents: In the AL22 parent, only one allele was detected in both intergenic and coding regions by the sequencing strategy described above. However, SNP591 revealed that the AL22 parent is heterozygous in the *CHS* coding region. Using a singleton carried by the AL22 individual at position 1367 (see alignment provided in supplemental data at http://www.genetics.org/supplemental/), we genotyped alleles in the intergenic region of all AL22 progeny by pyrosequencing, using PCR primers 5′-AAAGGGGGCTAACAACTAGCC-3′ and 5′-GAAAGATGGCGGAGAGTG-3′ and the SNP primer 5′-GGGAAAAAGGAGATG-3′. This analysis showed that SNP591 allowed the assessment of individuals carrying the identified AL22 allele. The AL22xAL7 progeny were assessed by the SNP1008 assay, which is based on a singleton carried by the AL7 cDNA allele. In these progeny, individuals carrying either one or the other alleles could be assessed. Similarly, we genotyped the AL52 intergenic allele of all individuals in the AL41xAL52 progeny (position 1333 in the alignment provided as supplemental data at http://www.genetics.org/supplemental/, PCR primers as above, SNP primer 5′-ATGGACGGGCGGATGAAG-3′). We further used singletons found in intergenic allele AL11-1 to confirm that this allele was not present in the AL12 individual used for crosses (position 914 in the alignment provided as supplemental data at http://www.genetics.org/supplemental/, PCR primers 5′-GAGTTAAGTATGCACGTG-3′ and 5′-TACGTACACCAACAAAAGGG-3′, SNP primer 5′-GGAGATTTCACTTCCC-3′). In *A. halleri*, RFLP blots confirmed the size and number of alleles obtained by sequencing (not shown). AH22 and AH4 alleles were genotyped at position 921 and 935, respectively (see alignment provided in supplemental data at http://www.genetics.org/supplemental/), using PCR primers 5′-GAGTTAAGTATGCACGTG-3′ and 5′-TACGTACACCAACAAAAGGGG-3′ and SNP primer 5′-GTAGAGTTTCTCCACC-3′.

Statistical analysis of expression data: We conducted the statistical analysis in three steps. In the first step, we investigated for trial effects without measurements made on heterozygous DNA and performed the following GLM analysis

$$x_{ijklm} = \mu + G_i + I_j + T_k + \mathrm{P}_l\,(\text{or } P_{kl}) + GI_{ij} + IT_{jk} + GT_{ik} + GTI_{ijk}\,(+ M_{im}) + C + \varepsilon_{ijklm}, \quad (1)$$

where μ is the grand mean, G_i is the effect of the ith genotypic combination or cross, I_j is the jth *CHS* expression environment or treatment (*i.e.*, dark-maintained, light-maintained, insect-damaged, and control leaves), T_k is the kth trial, P_l is the lth PCR and pyrosequencing plate (for *A. lyrata* data, P_{kl} is the lth PCR and pyrosequencing plate in the kth trial), C is a technical covariate following de Meaux *et al.* (2005) and GI_{ij}, IT_{jk}, GT_{ik}, and GTI_{ijk} represent interactions between cross × treatment, treatment × trial, cross × trial, and cross × treatment × trial, respectively. When possible, M_{im}, the effect of the mth mother in the ith genotype was added to the model. A significant cross × treatment effect indicates that *CHS cis*-regulatory alleles respond differently to the different expression conditions examined in this study. Within a genotypic combination or cross, several allelic combinations may be segregating if the parents are heterozygous. In this analysis, allelic combinations are taken together. This approach is conservative as the presence of different allelic combinations with different effects on *cis*-regulation will tend to increase the variance and consequently to decrease power to detect functional *cis*-regulatory differences between parents.

In the second step, we reincorporated the DNA measurements and investigated the existence of main effects or interaction. In this second analysis, the treatment source of variation included five treatments: cDNA samples from dark-maintained, light-maintained, insect-damaged, and control leaves as well as DNA samples from heterozygous individuals. In *A. halleri*, main and interaction effects involving trials were not significant. Subsequently, we included DNA samples and repeated the GLM analysis with a modified model (1) without trial effects. In *A. lyrata*, trial effect was significant in one of three SNP assays. Thus DNA samples were randomly attributed to one or the other trial and the GLM analysis was repeated using model (1).

In the third step, we dissected the main effects of genotypes and treatments as well as their interaction. For this, we conducted a separate GLM analysis for each progeny, with the following model

$$x_{jklm} = \mu + I_j + (T_k) + P_l\,(\text{or } P_{kl}) + (IT_{jk})\,(+ M_m) + C + \varepsilon_{ijklm}, \quad (2)$$

where μ is the grand mean, I_j is the jth *CHS* expression environment or treatment, P_l is the lth PCR and pyrosequencing plate (for *A. lyrata* data, P_{kl} is the lth PCR and pyrosequencing plate in the kth trial), T_k is the kth trial, IT_{jk} represents the interaction between treatment and trial, and C is a technical covariate following de Meaux *et al.* (2005). Effect trial and interaction trial × treatment were not included in the model for analysis in *A. halleri* because no significant trial effect was found in the global analysis (see above). When possible, M_m, the effect of the mth mother was added to the model. When data for a single genotype could be collected with more than one SNP assay, data obtained with both SNP assays were pooled, a SNP effect was added to the model, and the plate effect was nested within the SNP. If the size of the data set was too small, *i.e.*, if fewer than two repeated measurements were available per cell to test all effects described above, a model (3) without interaction was used. Three parental combination (AL22xAL7, AL41xAL52, and AH22xAH4) individuals of the

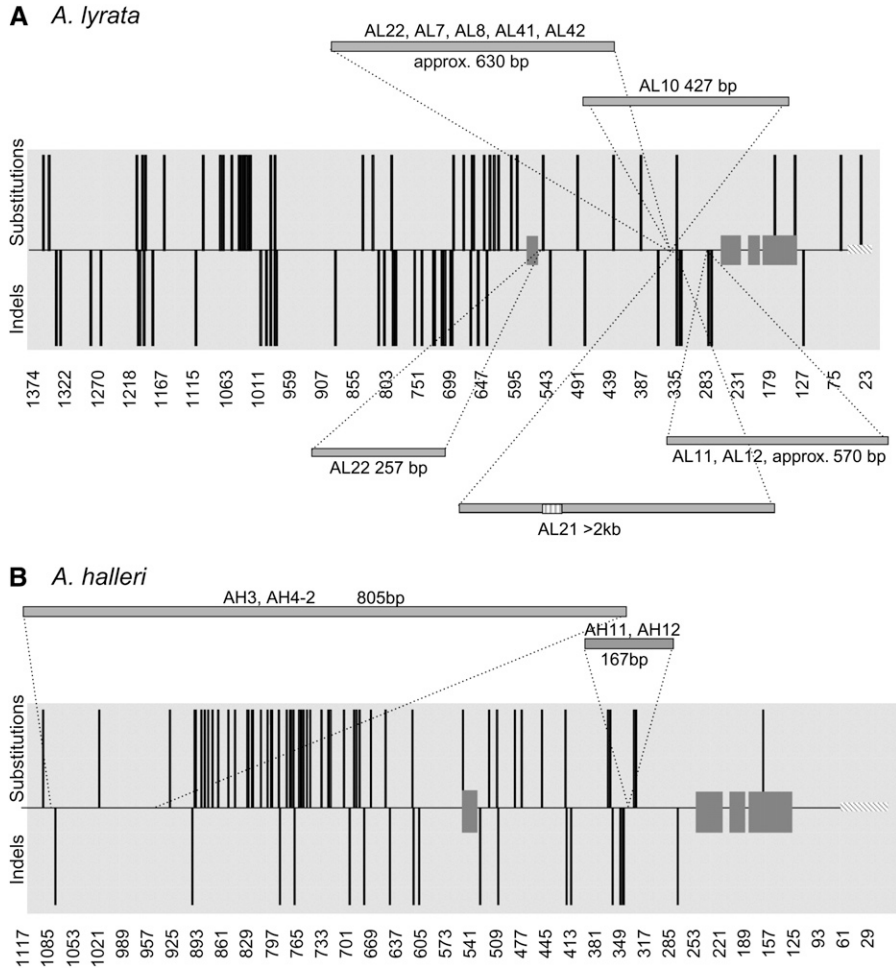

FIGURE 1.—Distribution of polymorphisms along the intergenic region upstream from the *CHS* open reading frame in (a) *A. lyrata* and (b) *A. halleri*. Bars on the top and bottom part of the graph indicate single nucleotide and insertion/deletion polymorphisms, respectively. Shaded bars along the sequence delineate the phylogenetic footprints found by KOCH *et al.* (2001) in the Brassicaceae. The hatched box indicates the 5′-UTR. Nucleotide positions along the sequence are indicated as base-pair distances from the ATG.

progeny were genotyped and an allele effect was added to the GLM model as well as an interaction between allele and treatment. Results for each GLM analysis are reported in supplemental Tables 4 and 6 at http://www.genetics.org/supplemental/.

Species-specific *CHS* expression in F_1 *A. thaliana–A. lyrata* hybrids was analyzed using a similar GLM model with modifications, depending on the sample size. For example, in the *A. thaliana–A. lyrata* hybrids, missing data prevented the analysis of a genotype × treatment interaction. For the *A. thaliana–A. halleri* F_1 hybrids, sample size was limited and *CHS* expression could not be studied in each genotype combination and each environment. In addition, all cDNA and DNA samples fit within one 96-well plate for one SNP assay. Therefore, for these F_1 hybrids, we used a GLM model that did not investigate either PCR plate or genotype effect.

To identify treatments in which relative allelic expression of a progeny was significantly different, we performed a post-hoc test using Tukey's honest significant difference (HSD) test, which compares each treatment least squares (LS) mean with every other treatment mean in a pairwise manner and controls the family-wise type I error to no >0.05. This test is suitable for pairwise comparisons performed without *a priori* on which pairs of average measurements may be different (QUINN and KEOUGH 2002). For the progeny in which parental allele effect was investigated, a Tukey's HSD test was performed to identify treatment × allele means that were significantly different (reported in supplemental Table 5 at http://www.genetics.org/supplemental/).

Fold-difference estimates: Calibrated pyrosequencing data provide a rough estimate of the amplitude of *cis*-regulatory differences between species or between genotypes within species. The mean of calibrated pyrosequencing data, or the mean relative allelic proportion of *CHS* in a biallelic *CHS* cDNA pool, was computed for each F_1 progeny and each *CHS* expression environment. The highest mean value identified by the GLM analysis as significantly different from allelic proportions in DNA was used to calculate an approximate maximum fold difference of *CHS* mRNA abundance due to *cis*-regulatory variation.

RESULTS

Nucleotide variation: We sequenced the 5′ flanking region at the *CHS* gene (the intergenic region) from 15 accessions in *A. lyrata* and from 8 accessions in *A. halleri*. Table 1 summarizes the number of accessions sequenced and the number of alleles obtained and their length. All accessions are diploid. The intergenic regions of *A. halleri* and *A. lyrata* are ~91% identical to the orthologous *A. thaliana* sequence (KOCH *et al.* 2001; DE MEAUX *et al.* 2005). In *A. lyrata*, the intergenic region was sequenced in 14 accessions from nine locations in Europe and one location in North America. Nineteen alleles varied in size from 1248 to 3399 bp, with large multiple independent insertions (Figure 1). Interestingly, all these insertions occurred between two regions, which were highlighted as strongly constrained in

the Brassicaceae (Koch *et al.* 2001). We found that one of these insertions contained a 200-bp fragment of a Mariner transposable element (Feschotte and Wessler 2002). The history of these insertions is likely to be complex. For example, in the AL10 allele, a 426-bp insertion was found, which is in the same position and alignable with the insertion found in alleles AL22, AL7, AL8, AL41, and AL42 (see Figure 1), but has modified nucleotides at the junction points. We removed these insertions from the sequences for the purpose of the alignment-based analysis of diversity (see supplemental data at http://www.genetics.org/supplemental/ for the alignments and the location of the large insertions). In the remaining alignment, we found 42 SNPs (20 singletons) and 47 indels (of which 26 were singletons). Indels ranged in size from 1 to 44 bp, with 28 of 47 affecting only one nucleotide position.

Significant population structure was detected ($S_{nn} = 0.685$, $P < 0.001$); hence one sequence was randomly chosen in each location, following Ramos-Onsins *et al.* (2004). Over this reduced sample the level of diversity (measured as the average pairwise number of differences per site, or π_t), reached 0.010 (Table 2). This value falls within the range of the silent diversity levels found at eight loci in *A. lyrata* (Ramos-Onsins *et al.* 2004). Interestingly, within one population, the level of diversity was comparable to species-wide diversity ($\pi_w = 0.008$ in Lilienfeld, Austria, Table 1). By contrast, individuals sampled in two Swedish populations were almost identical (only one 4-bp indel difference in a TA_n tract), pointing to heterogeneous distribution of diversity within this species. A nonsignificant Tajima's *D* indicated that the frequency distribution of SNP polymorphisms did not deviate from expectations under the neutral-equilibrium model ($D = 0.225$, $P > 0.1$). The value of *D* was typical of previously sampled loci in *A. lyrata* (Ramos-Onsins *et al.* 2004). Using *A. thaliana* as an outgroup, we analyzed the frequency distribution of derived mutations by Fay and Wu's *H* test. No excess of high-frequency-derived mutations was detected ($H = 0.413$, $P > 0.5$). Association patterns between adjacent sites indicated that branch length in the coalescent tree is compatible with an equilibrium neutral model (Wall's $B = 0.121$, $P > 0.2$ (Wall 1999). The HKA comparison of polymorphism-to-divergence ratios across loci indicates whether a given DNA region has an unusual rate of evolution or polymorphism. We compared the ratio of polymorphism to divergence in the *CHS* coding region to that in the intergenic region, using *A. thaliana* as outgroup. The HKA test results were nonsignificant ($\chi^2 = 0.413$, $P > 0.5$).

In *A. halleri*, the intergenic region was sequenced in eight accessions from six locations in Europe. Ten alleles were uncovered, varying in size from 1085 to 1704 bp (AH4-1 and AH6-2 are identical). Difference in allele length was mostly due to a large indel found in the 5′ part of the alleles AH4-2 and AH3 (Figure 1). In addition, two large indels (>40 bp) were observed at different positions along the sequence. A 42-bp deletion was observed only in allele AH4-2 and a 167-bp insertion was found AH11 and AH12 alleles. For alignment-based analysis of diversity, these regions were removed.

A total of 53 SNPs were observed, with five singletons. Eighteen indels of <40 bp were observed, with only one being a singleton. Population structure was apparent in our sample ($S_{nn} = 0.604$, $P < 0.001$); therefore we performed neutrality tests on samples containing one sequence for each population. The AH4 individual harbored two divergent alleles and the inclusion of one or the other allele in the subsample modified substantially the levels of diversity. Therefore neutrality tests were performed on two subsamples, one containing AH4-1 and the other AH4-2, to reflect the range of values that can be observed (Table 2). The level of diversity, π_t, reached 0.021 (with AH4-1) or 0.025 (with AH4-2), falling within the range of synonymous diversity at eight coding loci in *A. halleri* (Ramos-Onsins *et al.* 2004).

The haplotype structure of diversity in *A. halleri* differed from that found in *A. lyrata* (Figure 2). In contrast to *A. lyrata* alleles, *A. halleri* alleles formed three clades. In particular, one of these clades was closer to *A. lyrata* than to the other two clades (Table 4). Average levels of pairwise nucleotide differences between clade 3 and the other two clades reached $\pi = 0.04$, exceeding species-wide levels of diversity. This level of diversity is comparable to the level of nucleotide divergence observed in the intergenic region between *A. lyrata* and *A. halleri* (*K* per site = 0.037). To verify that *CHS* is a single-copy gene, we looked at allelic segregation in the progeny of two crosses: AH3xAH4 and AH4xAH22. The RFLP profiles of 10 progeny of each cross indicated that alleles segregated in a manner consistent with the expectations for alleles at a single locus (not shown). No significant deviation from neutrality was detected in the patterns of diversity in our sample (Table 2). A nonsignificant Tajima's *D* detected no deviation from expectations under the neutral-equilibrium model ($D = -0.76$ or $D = 0.96$, both $P > 0.2$). The coding region of *CHS* exhibits singular polymorphism features in *A. halleri* (Ramos-Onsins *et al.* 2004), with a highly significant Fay and Wu's *H*, presumably indicative of genetic introgression. In the intergenic region, Fay and Wu's *H* was not significant ($H = -7.2$ or $H = -2.93$, minimum $P = 0.09$), nor was the difference in the polymorphism-to-divergence ratio between the *CHS* intergenic and coding regions (HKA, $\chi^2 = 2.06$ or 2.14, minimum $P > 0.1$). Association patterns between adjacent sites indicated that branch length in the coalescent tree is compatible with a equilibrium neutral model (Wall's $B = 0.667$ or 0.622, minimum $P > 0.9$; Wall 1999). Each natural accession found to be heterozygous harbored alleles from different clades. Thus population subdivision is unlikely to explain this haplotype structure. Instead, in the two divergent alleles (AH3 and AH4-2),

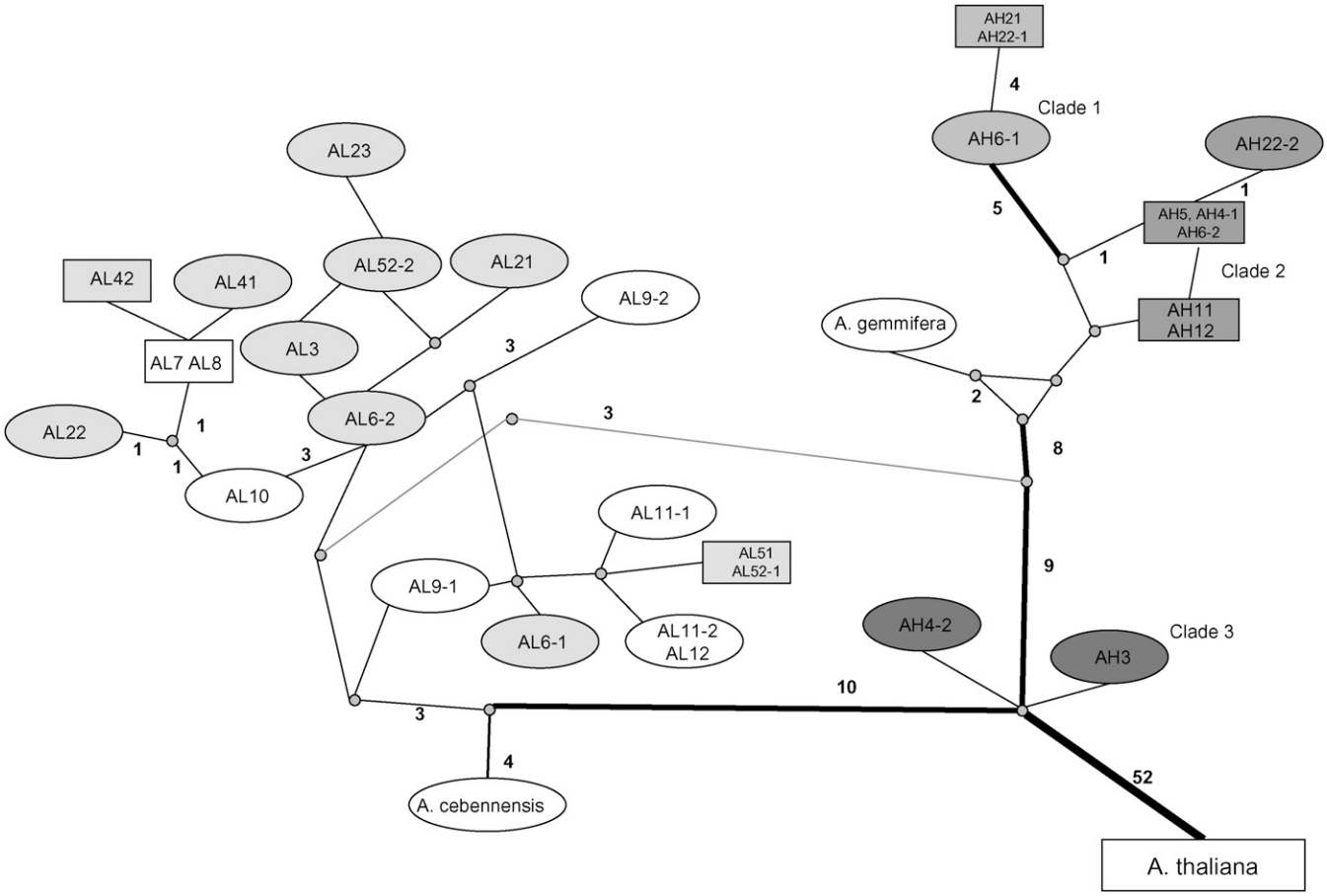

FIGURE 2.—Synthetic representation of the parsimony haplotype network built for the portion of the intergenic region that is alignable across species in Arabidopsis (see MATERIALS AND METHODS). Numbers indicate the number of intermediate changes along each branch. Thick solid lines indicate branches with larger number of changes that determine major clades. *A. lyrata* forms a separate clade. Three different haplotype groups are observed in *A. halleri*. For *A. lyrata*, haplotypes found in central Europe are indicated in gray and are interspersed in the network with haplotypes found in peripheral populations. The gray line reflects homoplasy between *A. halleri* and *A. lyrata*.

several gene conversion tracts were detected throughout the sequence (BETRAN *et al.* 1997). Three tracts involved a conversion from *A. lyrata* into *A. halleri* and one a conversion from *A. thaliana* into *A. halleri*. No similar pattern was found in the analysis of *A. lyrata* alleles. Introgression of related Arabidopsis species into *A. halleri* was previously suggested by a multilocus analysis (RAMOS-ONSINS *et al.* 2004). Therefore, the existence of divergent allelic lineages in *A. halleri* appeared compatible with genomewide patterns of variation. These alleles appear to have originated from the recombination of existing diversity and are not the result of a distinct evolutionary history along a separate branch of the genealogical tree. Haplotype networks, such as the one presented in Figure 2, allow the representation of recombination events and lineage mixtures and thus illustrate this phenomenon more accurately than a conventional phylogenetic tree.

TABLE 4

Average pairwise nucleotide divergence in *A. halleri*, within and among allelic clades defined by the haplotype network in Figure 2

	Clade 1	Clade 2	Clade 3	*A. halleri*
Clade 1	0.00307			
Clade 2	0.01408	0.00207		
Clade 3	0.04998	0.0459	0.00579	
A. lyrata[a]	0.04115	0.03067	0.04275	0.01099

[a] Analysis was made after reducing sample size, following the exclusion of sequences obtained from the same population.

Levels and patterns of polymorphism and divergence at noncoding DNA regions may reflect constraints on functionally important regulatory sites. The 5′ part of the intergenic region has substantially diverged across the genus Arabidopsis. The three species aligned poorly in this region (see alignment in supplemental data at http://www.genetics.org/supplemental/). Even within *A. halleri*, no reliable alignment could be obtained over this region, due to large insertions and deletions. By contrast, we found low levels of diversity in the 3′ part of the intergenic region, which contains the core promoter of *CHS*. We examined polymorphism within the

CHS promoter stretches found to be highly conserved throughout the Brassicaceae (Koch *et al.* 2001). One and two polymorphisms were found to segregate in *A. halleri* and *A. lyrata*, respectively, which occurred within the conserved region around the ACE and MRE regulatory elements. However, these polymorphisms did not affect any of the three elements required for expression of *CHS* in response to light or fungal elicitors (Hartmann *et al.* 1998; Logemann and Hahlbrock 2002). No segregating polymorphism was found in the three other conserved sequence blocks in the *CHS* promoter (Koch *et al.* 2001). Two fixed differences between *A. thaliana* and either *A. halleri* or *A. lyrata* were also observed in the ACE–MRE conserved block, one of them affecting the ACE element (see alignment in supplemental data at http://www.genetics.org/supplemental/), which is necessary for light-responsive *CHS* expression. No fixed differences were found between *A. lyrata* and *A. halleri* in any of the four phylogenetic footprints found by Koch *et al.* (2001) in the Brassicaceae. The low levels of nucleotide divergence and polymorphism in this region did not allow the application of an HKA test to examine whether its evolutionary rate is unusually low.

Expression diversity: In F_1 individuals obtained from intra- as well as interspecific crosses, parental *CHS* *cis*-regulatory regions are in perfect linkage with parental-coding regions and experience the same *trans*-regulatory background. Thus, the relative amount of parental *CHS* mRNA reflects the relative activity of parental *cis*-regulatory regions (Cowles *et al.* 2002). This approach allows us to evaluate the amount of *cis*-regulatory variation. In our assay, we used DNA from F_1 individuals to experimentally model the null hypothesis of the "no activity" difference between parental *cis*-regulatory alleles (de Meaux *et al.* 2005). Indeed, heterozygous DNA contains equal amounts of parental alleles. A total of 461 and 421 F1 progeny were obtained from intraspecific crosses between multiple genotypes in *A. lyrata* and *A. halleri*, respectively (summarized in supplemental Table 1 at http://www.genetics.org/supplemental/). Plants were submitted to four different *CHS* induction treatments (maintained for 48 hr in the dark and 8 hr in the light and submitted to herbivory by *P. xylostella* and control insect-free plants; see materials and methods). Leaf tissue was subsequently harvested from these plants to examine *CHS* *cis*-regulatory variation. Relative allelic amounts were determined in a total of 709 and 597 samples (summarized in supplemental Table 2 at http://www.genetics.org/supplemental/). Along with this, a total of 81 and 25 individuals were obtained from crosses between *A. thaliana* and either *A. lyrata* and *A. halleri*, respectively, which yielded a total of 361 measurements of relative allelic amounts (supplemental Table 3 at http://www.genetics.org/supplemental/). *CHS* expression was also examined in floral tissue in the *thaliana–lyrata* hybrids.

Expression diversity in *A. lyrata*: Using a GLM model, we investigated the effect of treatments (*i.e.*, DNA and cDNA pools) and genotypes (progeny of a parental combination; see materials and methods).

Three SNP assays were used to evaluate *cis*-regulatory diversity in *A. lyrata*. For each SNP assay, a significant effect of genotype and *CHS* induction treatments was detected ($P \leq 0.031$, Table 5). Likewise, the interactions between genotype and induction treatments were always highly significant ($P < 0.001$, Table 5). This reveals that relative allelic expression varies across the *CHS* expression environments in a way that depends on the genotype of the progeny.

We subsequently conducted a separate analysis of variance for each genotype (supplemental Table 5 at http://www.genetics.org/supplemental/). If the treatment effect was significant, we further performed a Tukey's post-hoc multiple mean comparison test to identify which *CHS* expression environment yielded differences in the relative ratios of parental *CHS* mRNA (Figure 3). All but two pairs of alleles exhibited significant functional differences in at least one *CHS* expression environment with respect to equal expression of parental alleles (Figure 3).

If parental genotypes used for the crosses are heterozygous, different combinations of intergenic alleles will segregate in the progeny. In particular, two singletons in the AL22 intergenic region and cDNA were found to segregate in all F_1 progeny of AL22, demonstrating that this individual is heterozygous. We genotyped each allele in the AL22xAL7 progeny and performed a GLM analysis with and without allele effect and allele × treatment interaction. An analysis not taking into account the allelic combination found no significant difference in expression between AL22 and AL7 alleles ($F_{4,98} = 1.16$, $P = 0.143$) and accounted poorly for variation ($R^2 = 0.166$). Instead, the analysis incorporating allele effect accounted for a much greater part of variation ($R^2 = 0.730$) and found significant effects of treatment ($F_{4,87} = 5.48$, $P = 0.001$), allele ($F_{1,87} = 86.68$, $P < 0.001$), and allele × treatment interaction ($F_{4,87} = 17.71$, $P < 0.001$). Post-hoc multiple mean comparison tests revealed that the identified AL22 allele was not significantly different from the AL7 allele, whereas the unknown AL22 allele differed markedly from the AL7 allele (Figure 4; see also supplemental Tables 4 and 5 at http://www.genetics.org/supplemental/). In individuals carrying the unknown AL22 allele, AL22 mRNA was overrepresented in the dark, as well as in both insect-damaged and control leaves but not in light-exposed leaves. We also genotyped the AL52 alleles in the AL52xAL41 progeny and found significant treatment and allele effects ($F_{4,56} = 8.469$, $P < 0.001$ and $F_{1,56} = 6.813$, $P = 0.012$, respectively; see also Figure 4). Post-hoc tests indicated that only the AL52-1 intergenic allele is significantly different from the AL41 allele, while the AL52-2 is not. However, the interaction between

TABLE 5

Global GLM analysis conducted separately for each SNP assay in *A. lyrata*

Source	Sum of squares	d.f.	Mean square	*F*-ratio	*P*
	SNP1008, model $R^2 = 0.458$				
Treatment	0.260	4	0.065	7.391	<0.001**
Genotype	0.062	2	0.031	3.534	0.031*
Genotype × treatment	0.394	8	0.049	5.606	<0.001**
Trial	0.027	1	0.027	3.035	0.083
Quality	0.017	1	0.017	1.891	0.171
Trial × genotype	0.114	2	0.057	6.494	0.002**
Treatment × trial	0.091	4	0.023	2.590	0.038*
Genotype × trial × treatment	0.200	8	0.025	2.851	0.005**
Pyroplate (trial)	0.067	2	0.034	3.833	0.023*
Error	1.669	190	0.009		
	SNP572, model $R^2 = 0.654$				
Treatment	0.198	4	0.049	15.871	<0.001**
Genotype	0.455	4	0.114	36.536	<0.001**
Genotype × treatment	0.447	16	0.028	8.980	<0.001**
Trial	0.051	1	0.051	16.447	<0.001**
Quality	0.156	1	0.156	50.282	<0.001**
Trial × genotype	0.038	4	0.010	3.088	0.016*
Treatment × trial	0.019	4	0.005	1.508	0.200
Genotype × trial × treatment	0.060	16	0.004	1.206	0.263
Pyroplate (trial)	0.016	4	0.004	1.251	0.290
Error	0.846	272	0.003		
	SNP591, model $R^2 = 0.474$				
Treatment	0.315	4	0.079	12.441	<0.001**
Genotype	0.045	1	0.045	7.044	0.009**
Genotype × treatment	0.135	4	0.034	5.312	0.001**
Trial	0.001	1	0.001	0.097	0.756
Quality	0.000	1	0.000	0.057	0.812
Trial × genotype	0.001	1	0.001	0.226	0.636
Treatment × trial	0.013	4	0.003	0.527	0.716
Pyroplate (trial)	0.017	2	0.009	1.360	0.262
Genotype × trial × treatment	0.018	4	0.004	0.710	0.587
Error	0.577	91	0.006		

For all three SNP assays, treatment effect, genotype effect, and the interaction genotype × treatment were significant. d.f., degrees of freedom. For SNP1008, the AL12xAL7, AL22xAL10, and AL22xAL7 progeny were included in the analysis. For SNP572, the AL12xAL41, AL12xAL3, AL41xAL22, AL41xAL52, and AL41xAL7 progeny were included in the analysis. For SNP591, the AL12xAL22 and AL22xAL3 progeny were included in the analysis. See MATERIALS AND METHODS. **Significant at $P < 0.01$; *significant at $P < 0.05$.

allele and treatment effect was not significant ($F_{4,56} = 1.891$, $P = 0.125$), indicating that the difference is tenuous (Figure 4; supplemental Tables 4 and 5 at http://www.genetics.org/supplemental/). This analysis demonstrates that distinct *cis*-regulatory alleles can segregate within populations. This is possible only if a second allele can be differentiated and if the progeny are big enough for an allele effect to be incorporated (for example, the AL22xAL10 progeny were too small for a similar analysis to be performed). In the AL22xAL12, AL22xAL41, and AL22xAL3 progeny, only individuals carrying the known AL22 alleles were analyzed (see MATERIALS AND METHODS) and we could not identify any second allele in AL41, AL7, AL12, or AL3. The results described above indicate that a statistical analysis that examines parental allelic combination in bulk tends to mask some of the *cis*-regulatory differences existing between parents. It remains possible that the larger variance observed for some measurements results from unknown allelic combinations segregating in the progeny.

A large insect-specific difference in *CHS cis*-regulation was detected in the cross between genotypes AL22 and AL10 (Figure 3). This difference was confirmed by two independent SNP assays and resulted from a relative decrease in AL10 *cis*-regulatory activity that was also apparent in a few plants obtained from a cross between genotypes AL10 and AL52 (not shown). The unknown AL22 allele was more active in both insect challenged and control leaves than the known AL22 allele. This may explain the larger variance observed in the AL22xAL10 progeny.

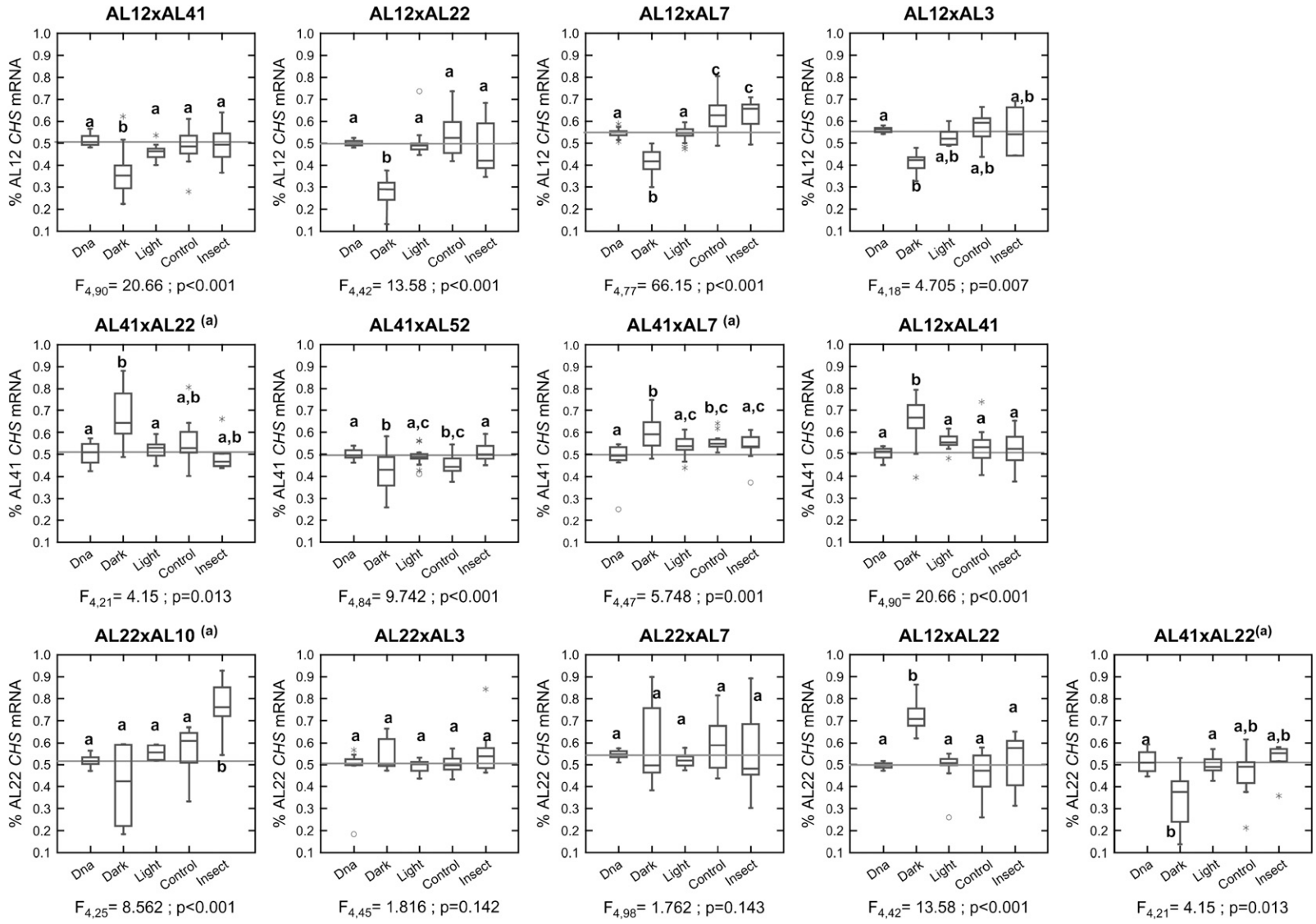

FIGURE 3.—Box plots reporting the relative *CHS cis*-regulatory activity in *A. lyrata* F_1 individuals from 10 parental combinations in response to dark, light, and insect feeding (with corresponding control). For each cross, the *y*-axis of the box plot indicates the relative expression level. In a box, the center horizontal line marks the median of the sample. The length of the box shows the range within which the central 50% of the values fall, with the box edges at the first and third quartiles. The whiskers show the range of observed values that fall within 1.5 times the midrange (or length of the box). The horizontal gray line indicates the expected value for equal promoter activity of both parental *cis*-regulatory regions in individuals of the progeny, as measured by relative allele abundance in DNA samples of the heterozygous individuals. For each genotype, an analysis of variance was conducted (see MATERIALS AND METHODS). Indicated here is the *F*-value of the treatment effect for each parental combination. Letters within the box plots indicate the result of the post-hoc multiple mean comparison (Tukey's test). The absence of a letter in common indicates significant differences in LS means. An "(a)" indicates samples that were analyzed with two independent SNP assays.

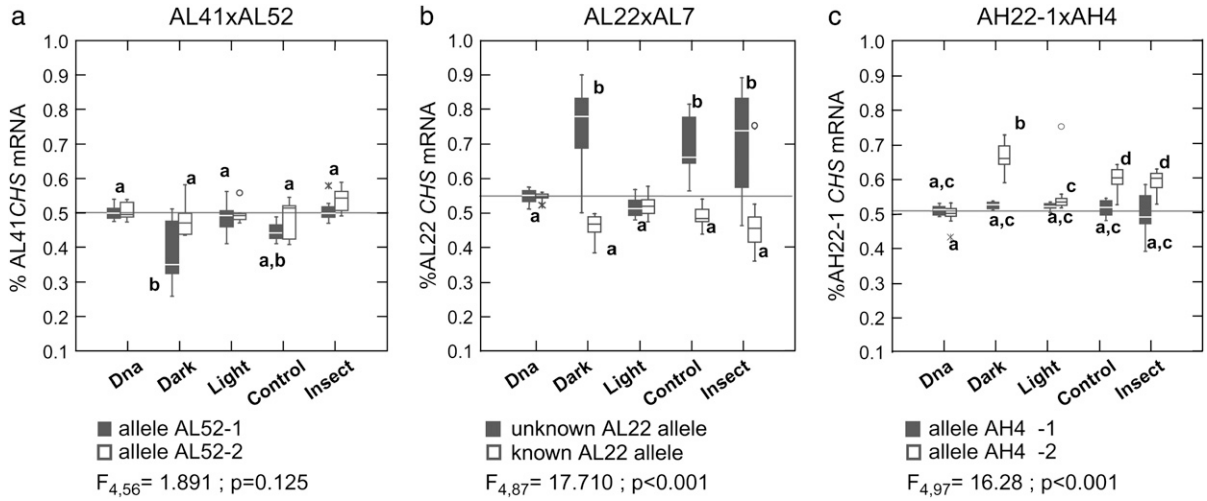

FIGURE 4.—*Cis*-regulatory variation in progeny in which different combinations of parental alleles are segregating. Box plots report the activity of one parental *cis*-regulatory allele relative to each of the two *cis*-regulatory alleles of the other parent in four different *CHS* expression environments. (a) Activity of each AL52 allele relative to the *cis*-regulatory allele of AL41. Only the AL52-1 allele is functionally different from the AL41 allele. (b) Activity of each AL22 allele relative to the *cis*-regulatory allele of AL7. Only one AL22 allele is functionally different from the AL7 allele. (c) Activity of each AH4 allele relative to the AH22-1 *cis*-regulatory allele. Only the AH4-2 allele is functionally different from the AH22-1 allele. *F*-value and associated *P*-values of the interaction between CHS expression environment (treatment) and allele are indicated below each box plot. Letters indicate significantly different pairs of treatment × allele means.

The analysis of the allelic differences in *CHS* expression in plants maintained 48 hr in the dark yielded three to four functional groups of *cis*-regulatory alleles (see Figure 3). The detailed analysis of the pairwise comparison of *cis*-regulatory activity indicates the following relationship in *cis*-regulatory activity: AL52-1 ≥ AL41 = AL52-2 ⩾ AL22/AL7/AL3/AL10 ⩾ AL12. The unknown AL22 intergenic allele could form an additional class but it is not known whether it is different from AL41. All *cis*-regulatory alleles showed equal activity after 8 hr of exposure to strong light as indicated by a nonsignificant difference between light-exposed-leaf cDNA samples and DNA samples for any of the genotypes. Thus these four functional groups respond differently to the onset of light, as they compensate in various degrees for the variable level of *CHS* expression in the dark.

Large-indel differences alone did not explain *cis*-regulatory differences in *A. lyrata*. For example, the known AL22 allele and AL3 had different large-indel content but no functional difference, whereas AL41 and AL7 were functionally different despite an identical large-indel content (Figures 2 and 3).

The average allelic proportion measured in our assay provided a rough estimate of the maximum fold difference in mRNA levels driven by *cis*-regulatory variation in each *CHS* expression environment (Table 6). In *A. lyrata*, maximums of 3.1- and 2.5-fold differences were observed in leaves maintained in the dark or challenged by herbivory, respectively.

Expression diversity in *A. halleri*: Three SNP assays were used to look at *cis*-regulatory variation in *A. halleri*. No trial effect was detected (see MATERIALS AND METHODS). Significant genotype and treatment effects were detected for two of three assays ($P < 0.001$, Table 7). For the third SNP assay (SNPM6), only the treatment effect was found to be significant ($P < 0.001$, Table 7) but a marginally significant interaction between genotype and treatment was found ($P = 0.048$).

We further conducted a separate GLM analysis for each genotype to identify statistically differentiated responses among genotypes (supplemental Table 6 at http://www.genetics.org/supplemental/). Figure 5 reports the pairwise comparison of *cis*-regulatory activity for each parental combination as well as the result of the post-hoc multiple mean comparison tests. Interestingly, the AH3 *cis*-regulatory allele was significantly less active than the alleles of AH5, AH12, or AH22 in all conditions although no clearly significant *cis*-regulatory differences were found in the progeny of AH3xAH4 (Figure 5; supplemental Table 6 at http://www.genetics.org/supplemental/). The AH3 intergenic alleles, as well as the AH4-2 allele, belong to clade 3. This clade is highly divergent from the other alleles segregating in *A. halleri*, in particular from the AH4-1 allele also carried by the AH4 individual (Figure 2). We genotyped individuals in the AH4xAH22 progeny for the AH4 allele that they inherited (note that the individuals that we analyzed in these progeny all harbored the same AH22-1 *CHS* intergenic region; see MATERIALS AND METHODS). A significant effect of the AH4 allele on expression was found as well as a significant interaction between treatment and the AH4 allele ($F_{1,97} = 84.04$, $P < 0.001$, $F_{1,97} = 16.28$, $P < 0.001$; supplemental Table 5 at http://www.genetics.

TABLE 6

Cis*-regulatory fold difference observed in *A. lyrata* and *A. halleri

	Dark	Light	Insect	Control
	A. lyrata			
Parental combination	AL22(unknown)xAL7	—	AL22xAL10	AL41xAL52
Mean fraction of the rarer *CHS* mRNA allele	0.244	—	0.261	0.456
Standard deviation	0.116	—	0.129	0.045
Approximate fold change	1:3.1	—	1:2.5	1:1.2
	A. halleri			
Parental combination	AH12xAH3	AH3xAH22	AH12xAH3	AH12xAH3
Mean fraction of the rarer *CHS* mRNA allele	0.207	0.469	0.275	0.272
Standard deviation	0.112	0.087	0.136	0.088
Approximate fold change	1:4	1:1.2	1:2.5	1:2.5

For each CHS expression environment, the highest *cis*-regulatory fold difference is indicated. The highest *cis*-regulatory fold difference was determined for the cross that showed an average allelic proportion most different from 0.5 and statistically significant from DNA measurements in post-hoc tests (see MATERIALS AND METHODS).

org/supplemental/). Only F_1 individuals harboring the promoter allele combination AH4-2 and AH22-1 showed significantly higher expression of the AH22-1 allele (Figure 4). The AH4xAH3 progeny was also genotyped. No significant effect of the AH4 allele was detected on expression data. However, restricted sample size limited our ability to detect a significant difference between the AH4-1/AH3 and AH4-2/AH3 promoter allele combinations. No significant *cis*-regulatory difference was observed between AH22-1 and AH5 (Figure 5).

Altogether, our comparative study of six alleles of the *CHS* intergenic region uncovered three functional groups. AH3 and AH4-2 constituted a set of alleles that are divergent at both nucleotide and functional levels. *CHS cis*-regulation in AH12 showed moderate but significant differences from AH5. And no *cis*-regulatory difference was detected among individuals harboring *CHS* intergenic alleles AH22-1, AH4-1, and AH5. In our study, *cis*-regulatory diversity in *A. halleri* controlled at most a fourfold difference in *CHS* mRNA level as observed in dark-maintained leaves of the progeny of AH12xAH3 (Table 6). High levels of nucleotide divergence in the intergenic region appeared to explain a large part of, but not all, *cis*-regulatory variation segregating in *A. halleri*.

Expression differences in interspecific hybrids: To evaluate the functional *cis*-regulatory divergence among Arabidopsis species, we crossed *A. thaliana* genotypes with both *A. lyrata* and *A. halleri*. Hybrid individuals have a haploid copy of each parental genome (*i.e.*, 13 chromosomes) and are sterile. They are morphologically similar to their non-*A. thaliana* parent. In total, five *CHS* expression environments were assessed (48 hr dark, 8 hr light, 24 hr insect feeding and respective control, expression in flowers after 48 hr in the dark). Altogether, 245 and 116 relative allelic measurements were performed for *A. thaliana*–*A. lyrata* and *A. thaliana*–*A. halleri* F_1 progeny, respectively (summarized in supplemental Table 3 at http://www.genetics.org/supplemental/).

***Cis*-regulatory differences between *A. thaliana* and *A. lyrata*:** In the *A. thaliana*–*A. lyrata* F_1 progeny, our assay did not detect *CHS* expression in either dark-maintained leaves or control non-insect-challenged leaves. Detection of *CHS* expression in the *A. thaliana*–*A. halleri* progeny with the same SNP assay (see below) suggested differences in transcription factor expression between hybrid types. The GLM analysis was conducted on a data set that included hybrid DNA samples and mRNA samples collected from three *CHS* expression environments (flowers, leaves after light exposure, and insect-damaged leaves; Table 8). The analysis examined the following sources of variation: *CHS* expression environment, parental genotype, SNP assay, and interactions of SNP × treatment and SNP × parental genotype, as well as a technical covariate (see MATERIALS AND METHODS).

A significant treatment effect was detected ($F_{3,218} = 173.393$, $P < 0.001$). Post-hoc multiple mean comparison tests indicated that this effect resulted mostly from the relative overexpression of the *CHS* mRNA of *A. thaliana* in insect-challenged leaves and, to a lesser degree, from slight overexpression of the *A. thaliana* mRNA in flowers (see Figure 6; supplemental Figure 1 at http://www.genetics.org/supplemental/). The first response is most likely insect specific because our assay failed to detect *CHS* expression in most control leaves but not in insect-challenged leaves. In addition, in the few control leaf samples where *CHS* expression could be detected, no skew toward one or the other parental *CHS* mRNA was apparent. Fold-change estimates indicated

TABLE 7

Global GLM analysis conducted separately for each SNP assay in *A. halleri*

Source	Sum of squares	d.f.	Mean square	*F*-ratio	*P*
SNPM6, model $R^2 = 0.326$					
Genotype	0.002	1	0.002	0.856	0.356
Treatment	0.047	4	0.012	5.021	0.001
Genotype × treatment	0.023	4	0.006	2.460	0.048
Mother (genotype)	0.011	2	0.006	2.387	0.095
Pyroplate	0.008	2	0.004	1.728	0.181
Quality	0.000	1	0.000	0.159	0.691
Error	0.368	156	0.002		
SNPB6, model $R^2 = 0.368$					
Genotype	0.260	2	0.130	27.892	<0.001
Treatment	0.226	4	0.056	12.079	<0.001
Genotype × treatment	0.124	8	0.015	3.319	0.001
Pyrolate	0.039	2	0.019	4.141	0.017
Quality	0.000	1	0.000	0.095	0.758
Error	1.079	2	0.005		
SNPCZ, model $R^2 = 0.571$					
Genotype	0.414	3	0.138	13.266	<0.001
Treatment	1.292	4	0.323	31.087	<0.001
Genotype × treatment	0.152	1	0.013	1.216	0.277
Pyroplate	0.012	1	0.012	1.129	0.290
Quality	0.041	1	0.041	3.910	0.050
Error	1.621	1	0.010		

For all three SNP assays, treatment effect, genotype effect, and the interaction genotype × treatment were significant. d.f., degrees of freedom. For SNPM6, the AH22xAH4 and AH22xAH5 progeny were included in the analysis. For SNPB6, the AH12xAH3, AH12xAH22, and AH12xAH5 progeny were included in the analysis. For SNPCZ, the AH12xAH3, AH3xAH22, AH3xAH4, and AH3xAH5 progeny were included in the analysis. See MATERIALS AND METHODS. **Significant at $P < 0.01$; *significant at $P < 0.05$.

that the *A. thaliana CHS* mRNA allele was four times more induced by insect feeding than its ortholog in *A. lyrata* (Table 9).

***Cis*-regulatory differences between *A. thaliana* and *A. halleri*:** Species-specific levels of *CHS* expression were also investigated in *A. thaliana*–*A. halleri* diploid hybrids. The GLM model incorporated variation attributable to the *CHS* expression environment, SNP assay, interaction between SNP and treatment, and a technical covariate (see MATERIALS AND METHODS). Due to limited sample size, *CHS* expression was not studied in all environments for some genotypes. Therefore, the effect of the parental genotypes was not incorporated into the analysis. The GLM analysis detected a significant treatment effect ($F_{4,105} = 23.125$, $P < 0.001$; Table 8). Post-hoc tests indicated a significant difference between insect-damaged leaves and all other treatments, including measurements made in hybrid DNA and control leaves (Figure 6; supplemental Figure 1 at http://www.genetics.org/supplemental/). Thus, the *A. thaliana CHS* gene was more induced by insect feeding than its ortholog in *A. halleri*. In addition, in the dark as well as in control non-insect-damaged leaves (which were collected in the early morning), the *A. halleri CHS* gene appeared to be more highly expressed (*i.e.*, presumably less repressed) than its ortholog in *A. thaliana*. The *A. thaliana CHS* gene transcript is twice as abundant in insect-damaged leaves as its *A. halleri* ortholog, whereas in the dark, the *A. halleri CHS* gene transcript is expressed three times more than its *A. thaliana* ortholog (Table 9).

Absence of large maternal effect and methylation: In each of the two species *A. lyrata* and *A. halleri*, four of nine crosses yielded individuals from both reciprocal crosses (supplemental Table 1 at http://www.genetics.org/supplemental/). In only two instances (two *A. halleri* progeny) was there any suggestion of reciprocal differences due to the direction of the cross (supplemental Table 6 at http://www.genetics.org/supplemental/; $P = 0.047$ for AH22xAH4 and $P = 0.032$ for AH12xSie).

Additionally, studies of newly formed allopolyploids suggest that interspecific hybrids may experience dramatic expression changes due to methylation of one or both parental copies (ADAMS *et al.* 2004; WANG *et al.* 2004). The interspecific hybrids obtained for this study were not polyploid. It seemed sensible, however, to evaluate the potential impact of methylation on the observed variation. We extracted DNA from leaves of the hybrids and from some of their parental genotypes. Levels of methylation were assessed at three potentially methylated CpG sites in the core promoter. No methylation could be detected at these sites in either parent or in the hybrid progeny. This suggests that bringing two distinct haploid genomes together in these interspecific hybrids did not alter dramatically the methylation at the *CHS* intergenic region.

No simple candidate mutation to explain functional variation: In *A. thaliana*, a light-responsive box was found to be polymorphic and to correlate with *cis*-regulatory differences in dark-maintained and light-exposed leaves (DE MEAUX *et al.* 2005). In both *A. halleri* and *A. lyrata*, this box is conserved. Thus, the differential *cis*-regulatory activity in the dark has to be found elsewhere. Association between polymorphisms and functional *cis*-regulatory differences has been successful in *A. thaliana*, where levels of nucleotide diversity are low. In *A. lyrata* and *A. halleri*, alleles instead differ on average at >13 positions ($\pi > 0.01$ in both species examined here). Therefore, the observed functional diversity, either within or between species, could not be tracked down to any single polymorphic sequence feature. Likewise, it was not possible to determine whether nucleotide differences in the ACE–MRE conserved regulatory element have functional consequences on *CHS cis*-regulation in *A. thaliana vs. A. lyrata* or *A. halleri*. We did not identify a candidate polymorphic motif to explain functional variation found within and between these species. It is interesting, however, to

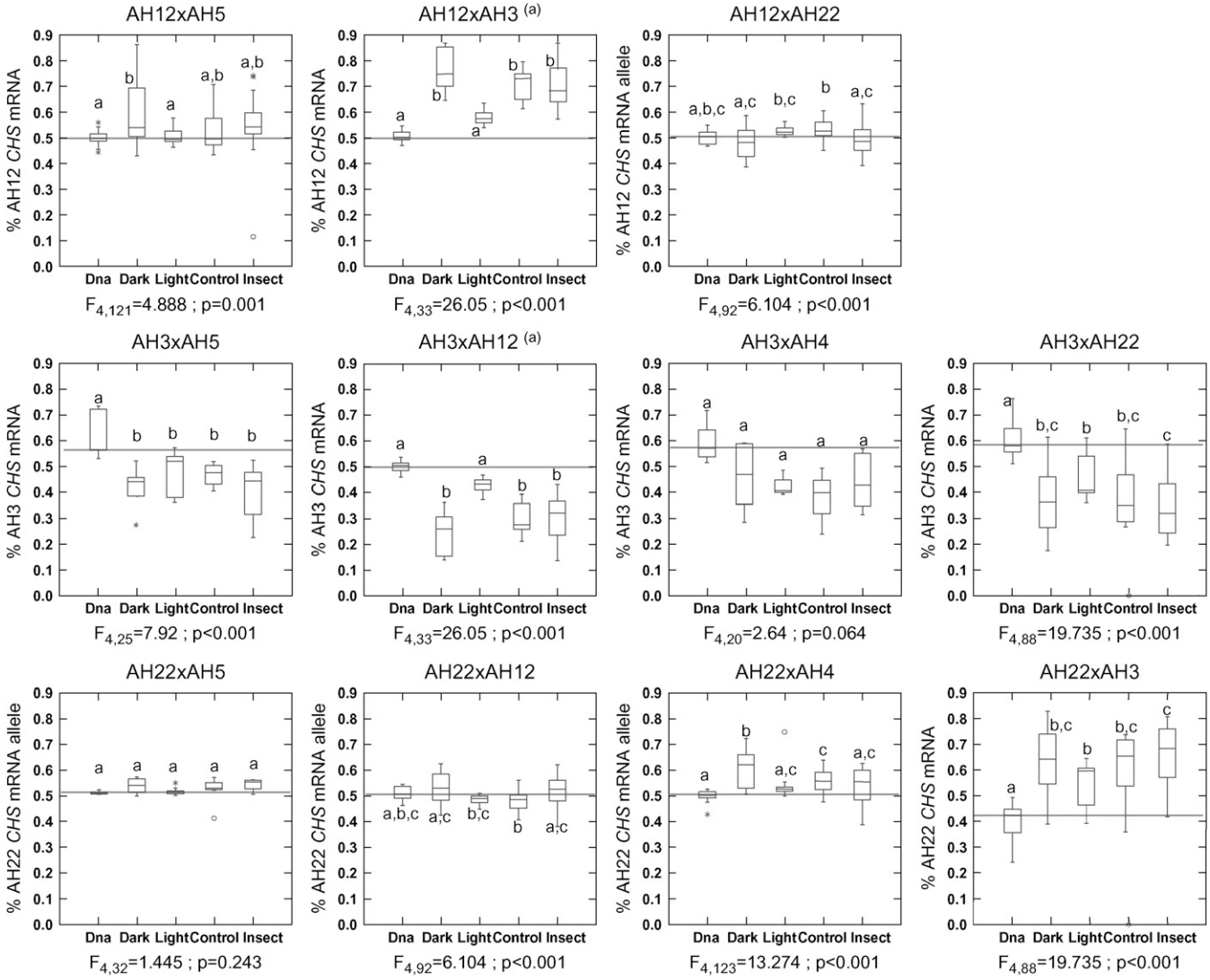

FIGURE 5.—Box plots reporting the relative CHS *cis*-regulatory activity in *A. halleri* F_1 individuals from eight parental combinations in response to dark, light, and insect feeding (with corresponding control). The horizontal gray line indicates the expected value for equal promoter activity of both parental *cis*-regulatory regions in the progeny. Indicated here is the *F*-value of the treatment effect for each progeny. Letters within the box plots indicate the result of post-hoc multiple mean comparison (Tukey's test; see MATERIALS AND METHODS). The absence of a letter in common indicates significant differences in LS means. An "(a)" indicates samples that were analyzed with two independent SNP assays.

note that a W-box was lost through introgression of a sequence fragment from *A. lyrata* into intergenic alleles AH4-2 and AH3 alleles. W-boxes are bound by WRKY transcription factors, involved in different types of stress and developmental responses (EULGEM *et al.* 2000). Whether this element is directly involved in the weaker *cis*-regulatory activity of the AH4-2 and AH3 alleles has to be tested experimentally.

DISCUSSION

Factors governing the diversity and evolution of *cis*-regulatory DNA are poorly understood. Here, we have characterized the standing variation of *cis*-regulation at the *CHS* locus at both nucleotide and functional levels in Arabidopsis. We show that large *cis*-regulatory differences segregate within species, both within and among populations. We further show that *CHS cis*-regulation has changed considerably among species, with alteration of the response to specific cues, which may be of ecological relevance. Our study reveals that *CHS cis*-regulation evolves in a modular fashion. In addition, we show that the patterns of nucleotide variation in the intergenic region upstream from *CHS* are complex and variable among species, yet they reveal no significant departure from neutrality. Interestingly, our study also documents some consequences of interspecific gene flow on *cis*-regulatory variation in *A. halleri*.

Modular *cis*-regulatory variation in Arabidopsis: To evaluate functional *cis*-regulatory variation at the species

TABLE 8

***Cis*-regulatory variation of CHS expression**

Source	Sum of squares	d.f.	Mean square	*F*-ratio	*P*
	Between *A. thaliana* and *A. lyrata*, model $R^2 = 0.797$				
Treatment	2.234	3	0.745	173.393	0.000
Genotype	0.172	4	0.043	10.03	0.000
Quality	0.006	1	0.006	1.312	0.253
SNP	0.166	1	0.166	38.565	0.000
SNP × treatment	0.136	3	0.045	10.571	0.000
SNP × genotype	0.091	4	0.023	5.273	0.000
Pyrolate (SNP)	0.109	4	0.027	6.321	0.000
Error	0.936	218	0.004		
Source	DNA	Flowers	Insect	Light	
	Between *A. thaliana* and *A. lyrata*, model $R^2 = 0.797$				
DNA	1.000				
Flowers	0.025	1.000			
Insect	0.000	0.000	1.000		
Light	0.778	0.476	0.000	1.000	
Source	Sum of squares	d.f.	Mean square	*F*-ratio	*P*
	Between *A. thaliana* and *A. halleri*, model $R^2 = 0.576$				
Treatement	1.439	4	0.360	23.124	0.000
SNP	0.116	1	0.116	7.481	0.007
Quality	0.054	1	0.054	3.442	0.066
Treatment × SNP	0.541	4	0.135	8.695	0.000
Error	1.633	105	0.016		
Source	Dark	Light	DNA	Insect	Control
	Between *A. thaliana* and *A. halleri*, model $R^2 = 0.576$				
Dark	1.000				
Light	0.000	1.000			
DNA	0.000	1.000	1.000		
Insect	0.000	0.043	0.006	1.000	
Control	0.040	0.008	0.000	0.000	1.000

Species-specific CHS expression analysis in interspecific hybrids. For *A. thaliana* and *A. lyrata*, GLM analysis was used, for *A. thaliana* and *A. halleri*, *P*-value associated with post-hoc multiple mean comparison tests was used.

level, we performed crosses between parental genotypes sampled in different locations throughout the native range of *A. lyrata* and *A. halleri*. By means of these crosses, we compared expression of different alleles within the same cells, and thus in the same *trans*-regulatory background. Because *cis*-regulatory and coding regions of each parent are linked, differences in the relative amount of allelic mRNA directly reflect allelic

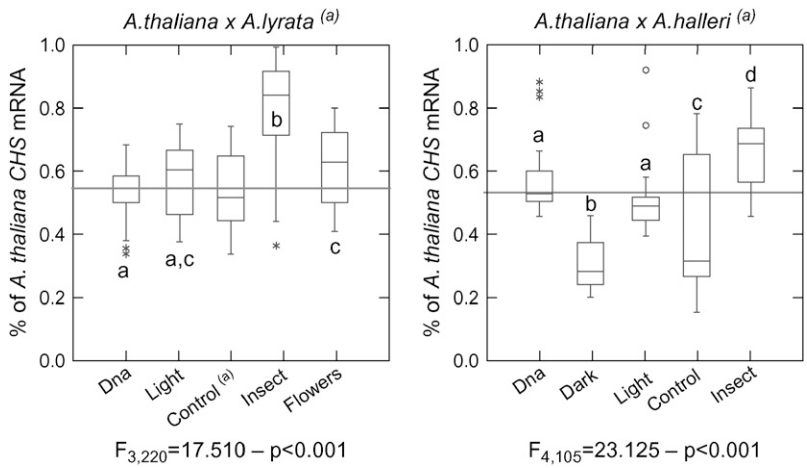

FIGURE 6.—Box plots reporting the relative *CHS cis*-regulatory activity in F_1 interspecific hybrids in response to different CHS expression environments. The *y*-axis of the box plots indicates the relative mRNA level of each parental species. The horizontal gray line indicates the expected value for equal promoter activity of both parental *cis*-regulatory regions in the progeny. Letters within the box plots indicate the result of post-hoc multiple mean comparison (Tukey's test; see MATERIALS AND METHODS). The absence of a letter in common indicates significant differences in LS means. Because in most samples *CHS* expression was not detectable, the *CHS* expression data for control non-insect-damaged leaves were excluded from the data analysis in *A. thaliana* × *A. lyrata*. An "(a)" indicates samples that were analyzed with two independent SNP assays.

TABLE 9

***Cis*-regulatory fold change observed in *A. thaliana* × *A. lyrata* and *A. thaliana* × *A. halleri* progenies**

	DNA	Dark	Light	Insect	Control	Flower
			Interspecific hybrid: *A. thaliana vs. A. lyrata*			
Mean % of *A. thaliana* allele	0.541	—	0.564	0.799	—	0.609
Standard deviation	0.069	—	0.108	0.159	—	0.129
Approximate fold change	1:1	—	1:1	4:1	—	1:1
			Interspecific hybrid: *A. thaliana vs. A. halleri*			
Mean % of *A. thaliana* allele	0.568	0.263	0.505	0.683	0.398	—
Standard deviation	0.093	0.1	0.113	0.14	0.262	—
Approximate fold change	1:1	1:3	1:1	2:1	1:1.5	—

For each CHS expression environment, the highest *cis*-regulatory fold change is indicated. The highest *cis*-regulatory fold change was determined for the cross that showed an average allelic proportion most different from 0.5 and statistically significant from DNA measurements in post-hoc tests (see MATERIALS AND METHODS).

cis-regulatory differences. Using this same approach, we also evaluated the functional divergence of *CHS cis*-regulation among species in the genus. Our data indicate that heritable DNA sequence differences likely are the molecular basis of the observed regulatory variation. Although methylation was shown to be one of several mechanisms that influence expression profiles in newly synthesized polyploids (ADAMS *et al.* 2004; WANG *et al.* 2004), no *CHS* methylation was found in natural *A. thaliana*, *A. lyrata*, or *A. halleri* individuals, nor in interspecific hybrids. Furthermore, all individuals generated for this study carry two haploid genomes. In addition, maternal effects on relative allelic expression were shown to be small and generally nonsignificant. These results suggest that differential methylation is unlikely to explain the expression differences observed in this study.

Two processes control mRNA abundance in the cell: transcription and degradation. Therefore differences in mRNA stability could also explain the differences found in our assay. However, we detected no functional differences between AH5 and AH22-1 *cis*-regulatory alleles (Figure 5), although the mRNAs differ by at least eight SNP polymorphisms (not shown). This absence of difference is especially interesting because the AH22-1 intergenic allele is associated with a *CHS* allele that is likely introgressed from *A. lyrata*, because it carries many *A. lyrata*-specific SNPs (not shown). This suggests that sequence differences do not dramatically influence mRNA stability. Instead, differences in *cis*-regulation likely cause variation in relative mRNA abundance, as suggested by the differences in relative mRNA abundance associated with AH4-1 and AH4-2 intergenic alleles (Figure 4). Indeed, both alleles are linked to coding sequences identical to the coding sequence carried by AH5 (supplemental Table 1b at http://www.genetics.org/supplemental/). Therefore, it is unlikely that the difference in relative mRNA abundance associated with AH4-2 and AH22-1 alleles is due to differences in mRNA stability. More generally, allele-specific differences in mRNA stability have rarely been documented whereas *cis*-regulatory differences seem to be widespread (KNIGHT 2004). Our results support the hypothesis that abundance differences observed in our assay result from variation in transcription rather than degradation.

Within *A. lyrata* and *A. halleri* we have quantified relative *cis*-regulatory activities under four environmental conditions (*i.e.*, 48 hr dark-maintained, 8 hr light-maintained, insect-damaged, and control leaves). Most pairwise *cis*-regulatory comparisons yielded significant differential expression in response to at least one of the environments used in this study. *Cis*-regulatory differences were often not correlated across expression environments (Figures 3–5). This suggests a modular evolution of *cis*-regulatory function, where a change in the response to one cue does not affect the response to other cues. For example, in *A. lyrata*, the AL10 *CHS cis*-regulatory allele was 2.5 times less responsive to insect feeding than the AL22 allele and both alleles showed equal activity in the control non-insect-challenged leaves. In *A. halleri*, the AH12 *CHS cis*-regulatory allele was four times more repressed in the dark than the AH3 allele, but both alleles showed equal activity in leaves exposed to light. In addition to modular cue-specific differences, functional differences that affected expression in all transcription environments of our study were also observed. In *A. halleri*, the progeny of AH4xAH22 (Figure 4), AH12xAH3, AH22xAH3, and AH5xAH3 (Figure 5) showed that individuals carrying an allele highly divergent in the intergenic region (as depicted in Figure 2) had a weaker *cis*-regulation of *CHS* expression in all expression environments, albeit at different degrees. In both species, we found functional differences between alleles carried by a single individual, suggesting that *cis*-regulatory variation is abundant between as well as within populations.

Our study cannot quantify the amount of *cis*-regulatory variation segregating within species, because pairwise *cis*-regulatory comparisons were not performed in a

common *trans*-regulatory background. However, several aspects of our results suggest that functional *cis*-regulatory diversity within species is higher in *A. lyrata* and *A. halleri* than in *A. thaliana*. First, we found *cis*-regulatory differences in all environments in *A. lyrata* and *A. halleri* but not in *A. thaliana*, although in the latter more *CHS* expression conditions were examined (de Meaux *et al.* 2005). Second, in these species up to 4-fold *cis*-regulatory differences were observed, whereas in *A. thaliana*, no difference >1.5-fold was detected. This is consistent with known levels of phenotypic and molecular diversity in these species (Clauss *et al.* 2002; Wright *et al.* 2003).

CHS cis-regulation also appears to have changed substantially since the *A. thaliana* lineage diverged from *A. lyrata* and *A. halleri*. We assessed *CHS cis*-regulatory divergence among species in multiple independent interspecific hybrid progeny. With this experimental design, differences likely reflect fixed *cis*-regulatory differences among species. We found that the *A. thaliana CHS cis*-regulatory region is four times more responsive to insect feeding than its ortholog in *A. lyrata*. Likewise, the *A. halleri–A. thaliana* hybrids revealed that the *A. thaliana CHS cis*-regulatory region is also approximately three times more responsive to insect feeding than its *A. halleri* ortholog (Table 9, Figure 6). Indeed, levels of *A. thaliana* mRNA were only two-thirds of the *A. halleri* mRNA amount in control plants, whereas after herbivory this increased to two times the level of *A. halleri* mRNA. Insect-specific *cis*-regulatory differences were observed in both hybrid types, further suggesting that it is a fixed interspecific difference. The magnitude of the *A. thaliana* insect response either decreased in the common ancestor of *A. halleri*/*A. lyrata* or increased in the *A. thaliana* lineage. In addition, the *A. halleri CHS* mRNA was more expressed in the dark than its ortholog in *A. thaliana*. In the *A. thaliana–A. lyrata* hybrids, *CHS* expression could not be detected by our assay, which suggests that mRNAs of both species were expressed at equally low levels. Therefore, the relatively higher expression of *CHS* in the dark is likely to be specific to the *A. halleri* lineage. Our study of functional divergence among species indicates that *cis*-regulation evolves largely by modification of regulatory modules that increase or decrease the response to one but not all environmental cues.

The constraints on *cis*-regulation seem to differ across *CHS* expression environments. For example, *cis*-regulation of *CHS* after 8 hr of light is conserved between species and rarely deviates from the null expectations within species (Figures 3, 4, and 6; de Meaux *et al.* 2005). By contrast, in each species, *CHS cis*-regulatory differences were observed for plants in the dark. *CHS* expression was shown to be strongly increased by light and severely reduced in the dark in *A. thaliana* (Zimmermann *et al.* 2005). Presumably, translation of *CHS* is unnecessary in the dark, which could lead to relaxed constraints on the control of *CHS* expression and in turn favor *cis*-regulatory change. A genomic study of expression variation reveals a similar trend in Drosophila (Rifkin *et al.* 2005). In our study, plants were submitted to 48 hr of darkness before exposure to strong light for 8 hr and leaves were collected at each of these two time points. This design was intended to study variation in the kinetic of light response from minimum to maximum expression levels. Indeed, we observed such variation in *A. thaliana* (Figure 6, de Meaux *et al.* 2005). But variation observed in dark-maintained leaves *per se* might not be ecologically relevant. *Cis*-regulatory differences found in control non-insect-challenged leaves instead may be more relevant because these leaf samples were collected in the early morning and thus may reflect naturally occurring low levels of *CHS* expression. Variation in aspects of *CHS* regulation related to plant defense is also likely to be ecologically relevant. In *A. thaliana*, *CHS cis*-regulation was more responsive to insect feeding, and response to this biotic stimulus was not variable within this species (Figure 6; de Meaux *et al.* 2005).

Molecular evolution in the intergenic region: In *A. thaliana* the influence of the *CHS* 5′ intergenic region on expression in response to light and fungal elicitors has been studied extensively (Hartmann *et al.* 1998; Logemann and Hahlbrock 2002). In addition, in multiple Brassicaceae species, including *A. lyrata* ssp. *petraea*, the 3′ portion of this region is sufficient to control light-responsive *CHS* expression (Koch *et al.* 2001). Association analysis did not yield any noteworthy candidate polymorphism to explain *cis*-regulatory variation, because variation was generally too high. We analyzed the frequency distribution of nucleotide polymorphisms in the complete 5′ intergenic region upstream from the *CHS* coding region in a sample of individuals representative of species-wide diversity in *A. halleri* and *A. lyrata*. The four tests of neutrality that we used are based on different characteristics of diversity and examine (i) how segregating sites were shared between individuals (Tajima's *D*), (ii) how derived mutations were distributed with respect to neutral expectations (Fay and Wu's *H*), (iii) whether patterns of association between adjacent sites conform to expectations of a standard neutral model, and (iv) whether the intergenic region has evolved at the same pace as the coding region (HKA test). In *A. thaliana*, these neutrality tests largely failed to reject neutral models (de Meaux *et al.* 2005). Likewise, in *A. lyrata* and *A. halleri*, these tests gave no indication that selection has influenced variation in this region. Levels of divergence from *A. thaliana* were lower in the *CHS* intergenic region than those observed on average at silent sites in either species (Table 2; Ramos-Onsins *et al.* 2004). Nonetheless, the values of all summary statistics were compatible with theoretical predictions under neutrality and previously reported patterns of genomewide variation (Ramos-Onsins *et al.* 2004).

Our analysis, however, detected an interesting phenomenon in *A. halleri*. In this species, strikingly divergent allelic lineages were found to cosegregate even within populations. These lineages appear to result from the recombination of interspecific diversity, which generated a patchwork sequence with a markedly different *cis*-regulatory activity (Figures 4 and 5). Our study therefore illustrates how rearrangements following genetic introgression can create functional diversity in *cis*-regulatory DNA. This is reminiscent of remarkable features of *cis*-regulatory evolution in Drosophila. The eve-stripe2 enhancer, which controls expression of developmental genes, has a conserved function in several species of the genus but has distinct coevolved regulatory elements. When experimentally dissociated, these regulatory elements loose their function (Ludwig *et al.* 2000). Our work shows that *cis*-regulatory elements from different species can get shuffled in natural populations, which in turn generates *cis*-regulatory variation. To our knowledge, this phenomenon has not been reported previously.

We conducted our population genetic study in three species to increase our chance of detecting selection on *CHS* regulatory variation. However, no significant selective effects were detected in this study. Nevertheless, the possibility that the observed *cis*-regulatory variation influences fitness should not be discarded. In the first place, our sampling of nucleotide variation could not infer whether observed variation has been the target of local (within-population) adaptive events. Second, the causal polymorphisms responsible for expression differences may reside outside of the intergenic region. Third, most population genetics tests can detect only selection events that have occurred within a certain time window and may be confounded by demographic processes, selection events acting on standing variation, or sweeps on recurrent mutations (Przeworski 2003; Przeworski *et al.* 2005; Pennings and Hermisson 2006).

Thus far, we have considered sequence variation in a *cis*-regulatory DNA region and expression phenotypes presumably associated with these polymorphisms. Natural selection, however, acts on biochemical phenotypes influenced by CHS, rather than on mRNA levels controlled by *cis*-regulatory DNA. To assess the consequences of *cis*-regulatory variation on plant phenotype, future work should examine the influence of mRNA expression on secondary metabolism and components of fitness. CHS controls a branch point in the phenylpropanoid pathway and quantitative variation of CHS can influence the output of more than one pathway. Higher CHS levels mediated by the overexpression of the flavonoid transcription factor PAP1 were shown to increase flavonoid production in *A. thaliana* (Borevitz *et al.* 2000) and the *tt1* mutant lacking a functional *CHS* gene has an increased sinapate content (Li *et al.* 1993). Thus, the observed fourfold mRNA expression variation in response to insect-feeding and light environment may alter flavonoid concentration. Indeed, between-species variation in flavonoid content can be readily observed in the greenhouse (J. de Meaux, personal communication). Whether this phenotypic variation is of adaptive importance remains to be established.

Substantial amounts of nucleotide variation at several *cis*-regulatory loci in the fruit fly, sea urchin, and yeast, which could not be related to selection, were also observed (Phinchongsakuldit *et al.* 2004; Balhoff and Wray 2005; Fay and Benavides 2005). However, these studies did not attempt to determine experimentally the existence of variation at the functional level. More generally, the extent of variation that causes neutral functional changes has been seldomly examined in either coding or noncoding DNA. Our study of *CHS* expression in Arabidopsis demonstrates that there is pervasive functional variation in *cis*-regulatory DNA. In addition, it suggests that the evolution of *cis*-regulatory DNA is modular. The influence of selective processes on this variation remains to be established. Follow-up studies will have to determine the extent to which *CHS cis*-regulatory evolution is typical for functional noncoding regions that mediate the physiological response to environmental signals in Arabidopsis.

We thank U. Goebel for careful scrutiny of the arrangement, polymorphism, and potential functional relevance of putative *cis*-regulatory motifs contained in the *CHS* intergenic sequences. We thank S. E. Ramos-Onsins for valuable advice concerning the population genetics analysis and M. J. Clauss, A. Lawton-Rauh, and two anonymous reviewers for helpful comments on the manuscript. This work was supported by the Max Planck Gesellschaft.

LITERATURE CITED

Adams, K. L., R. Percifield and J. F. Wendel, 2004 Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics **168:** 2217–2226.

Andolfatto, P., 2005 Adaptive evolution of non-coding DNA in Drosophila. Nature **437:** 1149–1152.

Balhoff, J. P., and G. A. Wray, 2005 Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. Proc. Natl. Acad. Sci. USA **102:** 8591–8596.

Bamshad, M. J., S. Mummidi, E. Gonzalez, S. S. Ahuja, D. M. Dunn *et al.*, 2002 A strong signature of balancing selection in the 5′ cis-regulatory region of CCR5. Proc. Natl. Acad. Sci. USA **99:** 10539–10544.

Becher, M., I. N. Talke, L. Krall and U. Kramer, 2004 Cross-species microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator Arabidopsis halleri. Plant J. **37:** 251–268.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent *et al.*, 2004 Ultraconserved elements in the human genome. Science **304:** 1321–1325.

Betran, E., J. Rozas, A. Navarro and A. Barbadilla, 1997 The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics **146:** 89–99.

Boffelli, D., J. McAuliffe, D. Ovcharenko, K. D. Lewis, I. Ovcharenko *et al.*, 2003 Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science **299:** 1391–1394.

Borevitz, J. O., Y. Xia, J. W. Blount, R. A. Dixon and C. Lamb, 2000 Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12:** 2383–2393.

Burbulis, I. E., M. Iacobucci and B. W. Shirley, 1996 A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in Arabidopsis. Plant Cell **8:** 1013–1025.

Clauss, M. J., H. Cobban and T. Mitchell-Olds, 2002 Cross-species microsatellite markers for elucidating population genetic structure in Arabidopsis and Arabis (Brassicaeae). Mol. Ecol. **11:** 591–601.

Clement, M., D. Posada and K. A. Crandall, 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657–1659.

Cliften, P. F., L. W. Hillier, L. Fulton, T. Graves, T. Miner *et al.*, 2001 Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis. Genome Res. **11:** 1175–1186.

Cowles, C. R., J. N. Hirschhorn, D. Altshuler and E. S. Lander, 2002 Detection of regulatory variation in mouse genes. Nat. Genet. **32:** 432–437.

de Meaux, J., U. Goebel, A. Pop and T. Mitchell-Olds, 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.

Dermitzakis, E. T., E. Kirkness, S. Schwarz, E. Birney, A. Reymond *et al.*, 2004 Comparison of human chromosome 21 conserved nongenic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment. Genome Res. **14:** 852–859.

Eulgem, T., P. J. Rushton, S. Robatzek and I. E. Somssich, 2000 The WRKY superfamily of plant transcription factors. Trends Plant Sci. **5:** 199–206.

Fay, J. C., and J. A. Benavides, 2005 Hypervariable noncoding sequences in *Saccharomyces cerevisiae*. Genetics **170:** 1575–1587.

Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Feschotte, C., and S. R. Wessler, 2002 Mariner-like transposases are widespread and diverse in flowering plants. Proc. Natl. Acad. Sci. USA **99:** 280–285.

Gompel, N., B. Prud'homme, P. J. Wittkopp, V. A. Kassner and S. B. Carroll, 2005 Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature **433:** 481–487.

Hahn, M. W., M. V. Rockman, N. Soranzo, D. B. Goldstein and G. A. Wray, 2004 Population genetic and phylogenetic evidence for positive selection on regulatory mutations at the factor VII locus in humans. Genetics **167:** 867–877.

Hartmann, U., W. J. Valentine, J. M. Christie, J. Hays, G. I. Jenkins *et al.*, 1998 Identification of UV/blue light-response elements in the Arabidopsis thaliana chalcone synthase promoter using a homologous protoplast transient expression system. Plant Mol. Biol. **36:** 741–754.

Hoffmann, M. H., 2005 Evolution of the realized climatic niche in the genus Arabidopsis (Brassicaceae). Evolution **59:** 1425–1436.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Hudson, R. R., M. Slatkin and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

Jenkins, G. I., J. C. Long, H. K. Wade, M. R. Shenton and T. N. Bibikova, 2001 UV and blue light signalling: pathways regulating chalcone synthase gene expression in Arabidopsis. New Phytol. **151:** 121–131.

Johnson, E. T., and P. F. Dowd, 2004 Differentially enhanced insect resistance, at a cost, in Arabidopsis thaliana constitutively expressing a transcription factor of defensive metabolites. J. Agric. Food Chem. **52:** 5135–5138.

Keightley, P. D., and D. J. Gaffney, 2003 Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. Proc. Natl. Acad. Sci. USA **100:** 13402–13406.

Keightley, P. D., M. J. Lercher and A. Eyre-Walker, 2005 Evidence for widespread degradation of gene control regions in hominid genomes. PloS Biol. **3:** 282–288.

Khaitovich, P., G. Weiss, M. Lachmann, I. Hellmann, W. Enard *et al.*, 2004 A neutral model of transcriptome evolution. PloS Biol. **2:** 682–689.

King, M. C., and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science **188:** 107–116.

Kliebenstein, D. J., M. A. L. West, H. van Leeuwen, K. Kim, R. W. Doerge *et al.*, 2006 Genomic survey of gene expression diversity in *Arabidopsis thaliana*. Genetics **172:** 1179–1189.

Knight, J. C., 2004 Allele-specific gene expression uncovered. Trends Genet. **20:** 113–116.

Koch, M. A., B. Haubold and T. Mitchell-Olds, 2000 Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol. Biol. Evol. **17:** 1483–1498.

Koch, M. A., B. Weisshaar, J. Kroymann, B. Haubold and T. Mitchell-Olds, 2001 Comparative genomics and regulatory evolution: conservation and function of the Chs and Apetala3 promoters. Mol. Biol. Evol. **18:** 1882–1891.

Li, J. Y., T. M. Oulee, R. Raba, R. G. Amundson and R. L. Last, 1993 Arabidopsis flavonoid mutants are hypersensitive to Uv-B irradiation. Plant Cell **5:** 171–179.

Logemann, E., and K. Hahlbrock, 2002 Crosstalk among stress responses in plants: pathogen defense overrides UV protection through an inversely regulated ACE/ACE type of light-responsive gene promoter unit. Proc. Natl. Acad. Sci. USA **99:** 2428–2432.

Ludwig, M. Z., C. Bergman, N. H. Patel and M. Kreitman, 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403:** 564–567.

Ludwig, M. Z., A. Palsson, E. Alekseeva, C. M. Bergman, J. Nathan *et al.*, 2005 Functional evolution of a cis-regulatory module. PloS Biol. **3:** 588–598.

Macdonald, S. J., and A. D. Long, 2005 Identifying signatures of selection at the enhancer of split neurogenic gene complex in Drosophila. Mol. Biol. Evol. **22:** 607–619.

Mitchell-Olds, T., 2001 Arabidopsis thaliana and its wild relatives: a model system for ecology and evolution. Trends Ecol. Evol. **16:** 693–700.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Neve, B., P. Froguel, L. Corset, E. Vaillant, V. Vatin *et al.*, 2002 Rapid SNP allele frequency determination in enomic DNA pools by pyrosequencing. Biotechniques **32:** 1138–1142.

Oleksiak, M. F., G. A. Churchill and D. L. Crawford, 2002 Variation in gene expression within and among natural populations. Nat. Genet. **32:** 261–266.

Pennings, P. S., and J. Hermisson, 2006 Soft sweeps II: molecular population genetics of adaptation from recurrent mutation or migration. Mol. Biol. Evol. **23:** 1076–1084.

Phinchongsakuldit, J., S. MacArthur and J. F. Y. Brookfield, 2004 Evolution of developmental genes: molecular microevolution of enhancer sequences at the Ubx locus in Drosophila and its impact on developmental phenotypes. Mol. Biol. Evol. **21:** 348–363.

Przeworski, M., 2003 Estimating the time since the fixation of a beneficial allele. Genetics **164:** 1667–1676.

Przeworski, M., G. Coop and J. D. Wall, 2005 The signature of positive selection on standing genetic variation. Evolution **59:** 2312–2323.

Quinn, G. P., and M. J. Keough, 2002 *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge, UK/London/New York.

Ramos-Onsins, S. E., B. E. Stranger, T. Mitchell-Olds and M. Aguade, 2004 Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. Genetics **166:** 373–388.

Reymond, P., H. Weber, M. Damond and E. E. Farmer, 2000 Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell **12:** 707–719.

Rifkin, S. A., D. Houle, J. Kim and K. P. White, 2005 A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. Nature **438:** 220–223.

Rockman, M. V., M. W. Hahn, N. Soranzo, D. B. Goldstein and G. A. Wray, 2003 Positive selection on a human-specific

transcription factor binding site regulating IL4 expression. Curr. Biol. **13:** 2118–2123.

Rockman, M. V., M. W. Hahn, N. Soranzo, D. A. Loisel, D. B. Goldstein *et al.*, 2004 Positive selection on MMP3 regulation has shaped heart disease risk. Curr. Biol. **14:** 1531–1539.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Sabeti, P. C., E. Walsh, S. F. Schaffner, P. Varilly, B. Fry *et al.*, 2005 The case for selection at CCR5-Delta 32. PloS Biol. **3:** 1963–1969.

Schmid, K. J., S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar and T. Mitchell-Olds, 2005 A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. Genetics **169:** 1601–1615.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Wade, H. K., T. N. Bibikova, W. J. Valentine and G. I. Jenkins, 2001 Interactions within a network of phytochrome, cryptochrome and UV-B phototransduction pathways regulate chalcone synthase gene expression in Arabidopsis leaf tissue. Plant J. **25:** 675–685.

Wall, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

Wang, J. L., L. Tian, A. Madlung, H. S. Lee, M. Chen *et al.*, 2004 Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics **167:** 1961–1973.

Wang, R. L., A. Stec, J. Hey, L. Lukens and J. Doebley, 1999 The limits of selection during maize domestication. Nature **398:** 236–239.

Winkel-Shirley, B., 2001 Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. **126:** 485–493.

Wittkopp, P. J., B. K. Haerum and A. G. Clark, 2004 Evolutionary changes in cis and trans gene regulation. Nature **430:** 85–88.

Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer *et al.*, 2003 The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. **20:** 1377–1419.

Wright, S. I., B. Lauga and D. Charlesworth, 2003 Subdivision and haplotype structure in natural populations of Arabidopsis lyrata. Mol. Ecol. **12:** 1247–1263.

Zimmermann, P., L. Hennig and W. Gruissem, 2005 Gene-expression analysis and network discovery using Genevestigator. Trends Plant Sci. **10:** 407–409.

Communicating editor: A. D. Long