Functional Divergence in Tandemly Duplicated *Arabidopsis thaliana* Trypsin Inhibitor Genes

M. J. Clauss[1] **and T. Mitchell-Olds**

Department of Genetics and Evolution, Max Planck Institute of Chemical Ecology, 07745 Jena, Germany

Manuscript received August 17, 2003
Accepted for publication November 24, 2003

ABSTRACT

In multigene families, variation among loci and alleles can contribute to trait evolution. We explored patterns of functional and genetic variation in six duplicated *Arabidopsis thaliana* trypsin inhibitor (*ATTI*) loci. We demonstrate significant variation in constitutive and herbivore-induced transcription among *ATTI* loci that show, on average, 65% sequence divergence. Significant variation in *ATTI* expression was also found between two molecularly defined haplotype classes. Population genetic analyses for 17 accessions of *A. thaliana* showed that six *ATTI* loci arranged in tandem within 10 kb varied 10-fold in nucleotide diversity, from 0.0009 to 0.0110, and identified a minimum of six recombination events throughout the tandem array. We observed a significant peak in nucleotide and indel polymorphism spanning *ATTI* loci in the interior of the array, due primarily to divergence between the two haplotype classes. Significant deviation from the neutral equilibrium model for individual genes was interpreted within the context of intergene linkage disequilibrium and correlated patterns of functional differentiation. In contrast to the outcrosser *Arabidopsis lyrata* for which recombination is observed even within *ATTI* loci, our data suggest that response to selection was slowed in the inbreeding, annual *A. thaliana* because of interference among functionally divergent *ATTI* loci.

NATURAL selection and neutral evolutionary processes can shape functionally important genetic variation at individual genes. In multigene families, we are further challenged to distinguish the contribution of variation among loci and among alleles to functional diversification (Ohno 1970; Force *et al.* 1999; Walsh 2003). After gene duplication, both the ecological function and the evolutionary fate of loci can diverge (*e.g.*, Zhang *et al.* 2002a; Ferrari *et al.* 2003). Duplicated loci that are retained for more than a few million years eventually experience strong purifying selection (Lynch and Conery 2003) and can undergo adaptive divergence via modification of the ancestral gene function and the acquisition of novel function (Force *et al.* 1999; Lynch and Force 2000; Lynch *et al.* 2001). For genes involved in biotic interactions where functional diversity is selectively favored (*e.g.*, plant-pathogen and plant-herbivore interactions, mate recognition), elevated gene copy retention associated with subfunctionalization and neofunctionalization may impart a selective advantage to the persistent multigene families observed in many plant genomes (Charlesworth *et al.* 2000; Rask *et al.* 2000; Bergelson *et al.* 2001). We are only now beginning to explore concerted functional evolution among duplicated loci and alleles at individual loci. Gene birth, death, and diversification dynamics are expected to interact in a complex manner with allelic diversity evolving in response to frequency-dependent selection, balancing selection, or positive selection in a coevolutionary "arms race" (Ellis *et al.* 1995; Meyers *et al.* 1998; Duda and Palumbi 2000; Bergelson *et al.* 2001; Tian *et al.* 2002; Van der Hoorn *et al.* 2002; Zhang *et al.* 2002b).

Evolutionary inferences based on patterns in molecular sequence data have been the topic of a large number of recent studies (Comeron and Kreitman 2002; Fay *et al.* 2002; Ford 2002; Kim and Stephan 2002, 2003; Navarro and Barton 2002; Tian *et al.* 2002; Nordborg and Innan 2003). Because tandemly duplicated loci share not only a functional and evolutionary origin, but also a common chromosomal context, identifying the signatures of selection at diverging loci can be confounded by positional nonindependence. Interference due to conflicting selection on linked sites is expected in asexual genomes (Miralles *et al.* 1999), but may also shape functional evolution in sexual genomes with local reductions in recombination (Hill and Robertson 1966; Kim and Stephan 2000, 2003; Comeron and Kreitman 2002). In Drosophila, where linkage disequilibrium (LD) typically decays within 1 kb (Long *et al.* 1998), local reductions in recombination have been suggested to limit adaptive evolution at linked sites (Kirby and Stephan 1996; Betancourt and Presgraves 2002). For the worldwide population of the inbreeding annual plant *Arabidopsis thaliana*, significant LD is found within individual loci (Hanfstingl *et al.* 1994; Kawabe

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AJ632250–AJ632267.

[1]*Corresponding author:* Department of Genetics and Evolution, Max Planck Institute of Chemical Ecology, Beutenberg Campus, Hans Knöll Str. 8, 07745 Jena, Germany. E-mail: clauss@ice.mpg.de

et al. 1997; Kawabe and Miyashita 1999; Aguadé 2001; Hauser *et al.* 2001) and on a genomic scale up to 250 kb (Nordborg *et al.* 2002). This potential nonindependence must be taken into consideration in functional genetic analyses for closely linked loci in *A. thaliana,* in particular for tandemly duplicated gene families.

Proteinase inhibitors (PIs) are widespread and highly diverse in the plant kingdom. The physiological function of plant proteinase inhibitors includes protection against the proteolytic enzymes of herbivores and pathogens, as well as the regulation of endogenous storage proteinases during seed dormancy and reserve protein mobilization (Green and Ryan 1972; Ryan 1990; Pautot *et al.* 1991; Koiwa *et al.* 1997; De Leo *et al.* 1998; Haruta *et al.* 2001; Glawe *et al.* 2003; Telang *et al.* 2003). A single plant may have proteinase inhibitors from several different functional classes (Laskowski and Kato 1980; Ryan 1990). On the basis of structural and biochemical properties, a novel class of serine trypsin inhibitors (TIs) encoded by a small gene family has been identified in the Brassicaceae (Menegatti *et al.* 1992; Ceciliani *et al.* 1994; Ruoppolo *et al.* 2000; Zhao *et al.* 2002). The specificity of the inhibited protease is determined in part by a single amino acid residue at the P1 position of the reactive site loop (Laskowski and Kato 1980; Ascenzi *et al.* 1999). Experimental randomization of P3-P3′ residues in the reactive site loop of the *Sinapis alba* TI (*MTI2*) and subsequent selection by phage display for elevated trypsin inhibitory activity demonstrated that the wild-type *MTI2* reactive site has optimal conformation for trypsin inhibition with very few degrees of freedom (Ceci *et al.* 2003). Trypsin inhibitory function has been maintained in orthologous TI loci in the Brassicaceae for >20 million years (MY), as suggested by a conserved reactive site loop (P3-P3′ APRIF/YP) in Brassica, Sinapis, and Arabidopsis (Koch *et al.* 2001; Zhao *et al.* 2002).

Functional investigation of Sinapis trypsin inhibitors has demonstrated expression in immature seeds and in wounded leaves (Ceci *et al.* 1995; De Leo *et al.* 2001a). The *MTI2* inhibitor interacts in a highly specific manner with insect gut proteinases, is induced upon feeding, and can be an effective defense against insect herbivores (De Leo *et al.* 1998, 2001b; Ceci *et al.* 2003). However, in response to a diet enriched in one TI, insects can preferentially express digestive proteinases insensitive to the dominant inhibitor and thereby render this plant defense less effective (Broadway 1995, 1997; Jongsma *et al.* 1995; Jongsma and Bolter 1997; De Leo *et al.* 1998; Volpicella *et al.* 2003). The fact that insect trypsin proteinases are encoded by members of large gene families (Gu *et al.* 2002) may contribute to this flexible counterstrategy. Thus, the diversity and expression of plant PIs may be shaped by an antagonistic coevolutionary dynamic favoring protein diversification as predicted by the arms-race model (Van Valen 1973; Broadway 1996; Jongsma and Bolter 1997; De Leo *et al.* 1998).

The loci encoding *A. thaliana* trypsin inhibitors (*ATTI*s) are members of a multigene family. Six *ATTI* loci are arranged in tandem within 10 kb on chromosome II and appear to have undergone duplication subsequent to a genome duplication event 24–40 million years ago (Blanc *et al.* 2003). Here, we test whether nucleotide polymorphism evolves independently among members of the *ATTI* tandem array in the selfing annual *A. thaliana*, with reference to the closely related outcrossing species, *A. lyrata*. We observed significant heterogeneity in the pattern of polymorphism among gene family members, including a peak of diversity flanking the presence/absence polymorphism for the *ATTI5* locus. We then analyze *ATTI* gene expression in response to herbivory and test for functional diversification both among loci and among alleles. Assessing the role of natural selection in shaping the significant functional diversification observed for members of this gene family was complicated by contrasting population genetic signatures for linked loci in a region of low recombination.

MATERIALS AND METHODS

***A. thaliana* trypsin inhibitors:** Six *ATTI* loci have been identified in the Columbia accession of *A. thaliana* on the basis of sequence homology to mustard trypsin inhibitor (*MTI2*; Menegatti *et al.* 1992) and rapeseed trypsin inhibitor (*RTI3*; Ceciliani *et al.* 1994). *ATTI1* is the ortholog to *MTI2* and *RTI3* (with on average 75% similarity) and has been structurally analyzed (Zhao *et al.* 2002). In our study *ATTI1*, *ATTI2*, *ATTI3*, *ATTI4*, and *ATTI6* occur in tandem on chromosome II and refer to At2g43510, At2g43520, At2g43530, At2g43535, and At2g43550, respectively. *ATTI4* is currently annotated in GenBank only as an expressed protein, not as a trypsin inhibitor (AC002335; March 11, 2002). *ATTI5* refers to a locus between *ATTI4* and *ATTI6* that was identified in several accessions other than Columbia on the basis of sequence homology (Figure 1). A locus with no homology to *ATTI* and of unknown function is located between *ATTI5* and *ATTI6* (At2g43540). A seventh locus in this gene family, *ATTI7*, is located on chromosome I (At1g47540). We focus here only on the *ATTI1–6* loci located within the tandem array on chromosome II.

The *ATTI1* locus encodes an 89-amino-acid precursor with 27 amino acids in the amino terminus not represented in the mature protein (Zhao *et al.* 2002). The 27-amino-acid N terminus of the precursor polypeptide most likely represents a signal peptide for secretion into the endoplasmic reticulum (SignalP v1.1). For *ATTI4* and *ATTI5*, a 5-amino-acid carboxy terminus was identified on the basis of sequence homology to *MTI2* (Volpicella *et al.* 2000). The position and length of a single intron is conserved among all *ATTI* loci.

Herbivory experiment and expression analysis: Seeds of seven *A. thaliana* accessions chosen randomly from two sequence-defined haplotype classes (A, Gö-0, Ler-0, Nd-1, and Rsch-0; B, Cvi-1, Fe-1a, and Wei-0; see below for haplotype classification) were sown onto a sterilized potting soil/vermiculite mix (3:1), vernalized at 4° for 1 week, placed in a randomized design into a short-day growth chamber, and reduced to a single individual per pot after germination. Four and 6 weeks after germination, six randomly selected plants per accession were placed into a separate growth chamber for 9 hr; three plants were subjected to the herbivory treatment and three individuals were control plants. The herbivory treatment con-

A

```
ATTI1 MAKAIV-SIFVVFFIFFLVISDVPE-IEAQGNE--CLKEYGGDVGFGFCAPRIFPTICYTRCRENKGAKGGRCRWGQG--SNVKCLCDFCDDTPQ
ATTI2 MAMKSV-STLAVFAILFLVIVEMPE-IKAQGSK--CLKEYGGNVGFSYCAPRIFPSFCYRNCRKNKGAKGGRCRSGGAGAGSMICLCDYCSDKP
ATTI3 MAMKSV-SNFAIFLILFLVTSEISE-IEAKDSE--CLKGY-IDAPLSYCMARIYPSLCYRNCRSYKGAKGGKC---DG----LKCFCDFCSDKPF
ATTI4 MATKSVLSIFAIFTILVLVIFEIPE-IEAHDSE--CLKEYGGDVGFGFCAPKIFPTICYRNCQKDKGANGGKCLWGEG--GNVKCLCDFCSKESFNQFISLT
ATTI5 MTTKSV-SIFAFFFILVLVIFEIPE-IEAQHDE--CLEEYGRDAGPALCAPEIVPTICYIKCRTEKEANGGRCLWGKD--GNAICLCEYCNEERSNQFTSLT
ATTI6 MK--SV-SVFFVLFIIFLVVLEAPEKIEAKDDKFICVVEYGGDVGPTFCNPKFFPTLCRQNCRSFKGAKGGKCVKQPK-HKHIKCFCDYCKDD
```

B

	ATTI2	*ATTI3*	*ATTI4*	*ATTI5*	*ATTI6*
ATTI1	63.2	53.1	66.7	57.3	50.0
ATTI2	-	57.9	61.1	53.7	46.8
ATTI3		-	59.4	46.9	46.8
ATTI4			-	67.0	48.9
ATTI5				-	43.6

FIGURE 1.—(A) Amino acid sequence of six *A. thaliana* trypsin inhibitor (*ATTI*) loci located on chromosome II. The signal peptide is indicated by light shading and the reactive site loop by dark shading, with P1 and P1′ indicated by dots. (B) Percentage pairwise sequence identity among *ATTI* coding sequences for one representative *A. thaliana* accession (Fe-1a).

sisted of one third instar *Plutella xylostella* (Leptidoptera) larva per plant. The larvae originate from line G88 (Cornell University) and were reared on an artificial diet (SHELTON *et al.* 1991), from which they were removed 16 hr prior to transfer to plants. Herbivore-induced transcription of homologous trypsin inhibitors is maximal after 9 hr of feeding in a related species of Brassicaceae (BAUKE 2002). After the experiment, up to six basal leaves were harvested from all plants and immediately placed into liquid nitrogen. All leaves harvested from herbivore-induced plants had feeding scars and thus had the potential for local as well as systemic induction.

For each of 12 plants per accession (two ages, two herbivore treatments, and three replicates), RNA was extracted from ≈100 mg of leaf tissue using a standard protocol employing cell lysis with TRIZOL reagent (GIBCO BRL, Gaithersburg, MD) and RNA purification with phenol-chloroform and ethanol precipitation. Approximately 1 μg of total RNA was used for cDNA synthesis, as described by FROHMAN *et al.* (1988). For reverse transcription polymerase chain reaction (RT-PCR), the amount of total cDNA added for each sample was adjusted so that the RT-PCR product within the linear phase for a housekeeping gene was ~5 ng/μl (*RAN*; Ras-related nuclear small GTP-binding protein; At5g55190). *RAN* primers were 5′ ACCAGCAAACCGTGGATTACC and 3′ CCACAAAG TGAAGATTAGCGTCC (57°; see RT-PCR conditions below). To assess trypsin inhibitor expression, one master mix was made for RT-PCR conducted with primers for all six *ATTI* loci and *RAN*. A standard PCR protocol was used with the following cycling scheme: 2 min at 94°; 30 cycles of 30 sec at 94°, 20 sec at primer-specific annealing temperature, and 40 sec at 72°; and a final extension at 72° for 10 min. Locus-specific primers were designed to span the intron, such that PCR products derived from the cDNA could be differentiated from genomic contaminants (primer sequences available upon request). The PCR products were visualized on a 2% agarose gel in a standardized manner. Band intensity provides a semiquantitative measure of the transcription level (ImageQuant 5.1; Molecular Dynamics, Sunnyvale, CA). Prior to analysis of transcription level, each RT-PCR band was standardized to the intensity of the 400-bp band of 2 μl of the Low Mass Ladder (Invitrogen, San Diego) loaded in every ninth well to control for gel effects. For the *ATTI* RT-PCR products, we standardized bands by the intensity of the *RAN* band for the given individual and by the PCR efficiency of the primers for that *ATTI* locus amplifying genomic DNA.

Statistical analysis of transcription: An analysis of variance was performed on the log-transformed standardized transcript level and, because transcription was sometimes zero, 100 was added to each standardized data point prior to log transformation (PROC GLM; version 8, SAS Institute). Haplotype class, herbivory, locus, age, and accession nested within haplotype class were fixed effects (*e.g.*, TEMPLETON *et al.* 1993). The full model including all interaction effects was analyzed. The locus × herbivory two-way interaction effect was further analyzed using linear contrasts within the context of the ANOVA to determine which loci deviated significantly from the average herbivore-induction response.

***ATTI* tandem array—amplification and sequencing:** A single individual was sampled from 17 accessions of *A. thaliana* (L.) Heynh. (Brassicaceae) originating from Eurasia, North Africa, and North America: Col-0, USA; Cvi-1, Cape Verde Islands; Di-0, France; Fe-1a, Germany; Gö-0, Germany; Ita-0, Morocco; Kas-1, India; Le-0, The Netherlands; Ler-0, Germany; Nd-1, Germany; Nok-0, The Netherlands; Rsch-0, Russia; Sah-0,

Spain; Ta-0, Czech Republic; Wei-0, Switzerland; Wil-2, Lithuania; and Ws-0, Belarus. All *A. thaliana* seeds were obtained from the Nottingham *Arabidopsis* Stock Center. After harvest, young leaf material from each accession was placed in liquid nitrogen and DNA extraction from 0.1 g of tissue followed the protocol for the Nucleon PhytoPure plant DNA extraction kit (Amersham, Arlington, Heights, IL). Primers for PCR amplification were placed in the exons of genes flanking the *ATTI* array in Columbia (AC002335; arrayF, 5′ GGACGGGTCGTTT CAGCTG, and arrayR, 5′ GACGTGAGCTTAGAGTTCATAC; 58°). PCR of up to 10 kb of chromosome II was conducted using ELONGase Enzyme mix (GIBCO BRL) with hotstart and an 8-min extension time. PCR products were gel purified and cloned using pCR-Blunt II TOPO vector (Invitrogen). Three clones per accession were sequenced on a 3700 ABI capillary sequencer using primers spaced approximately every 400 bp and designed from the Columbia accession (available upon request). For Ita-0, the *ATTI* region was amplified in two overlapping segments due to the presence of a large insertion. The *ATTI* array was amplified from one individual of *A. lyrata* ssp. *petraea* from Plech in Bavaria, Germany (49° 54′ 99″ N; 11° 30′ 64″ E; same as KOCH, Pfaffenhofen b. Neuhaus, Bavaria, Germany, leg. KOCH; Koch *et al.* 2001; Clauss and Mitchell-Olds 2003) using the primers 5′ CCAATCGGTTT GGTCCTAAAG and 3′ CATTCATTGAAGAACATCACATTG (55°). Three clones from *A. lyrata* were sequenced using the above primers from *A. thaliana* as well as additional species-specific primers where necessary.

Sequence analysis: Sequences were assembled with SeqMan 5.0 (DNASTAR) and all variable sites were checked manually during the construction of a consensus sequence from three clones for each *A. thaliana* accession and one *A. lyrata* individual. Only one allele per accession was included in the population genetic analysis of *A. thaliana* because individuals are derived from multiple generations of inbreeding in the laboratory. For the outgroup *A. lyrata*, we sequenced one of two alleles. All sequences were aligned with MegAlign 5.03 set to default gap penalty parameters (DNASTAR). Pairwise BLAST searches confirmed insertion/deletion (indel) breakpoints in regions of high polymorphism and repetitive nucleotide sequence. Coding and noncoding regions were inferred according to homology with *MTI2*, GenBank annotation, and sequenced cDNA clones in *A. thaliana* (M. J. Clauss, unpublished data). The DnaSP program version 3.84 (Rozas and Rozas 1999) was used for both intra- and interspecific analyses of nucleotide polymorphism. Nucleotide diversity, π, was estimated according to Nei (1987); θ according to Watterson (1975); and nucleotide divergence, *K*, between *A. thaliana* and *A. lyrata* according to Nei (1987). Linkage disequilibrium between variants at different polymorphic sites (Hill and Robertson 1968); Hudson and Kaplan's (1985) estimate of the minimum number of recombination events; the recombination parameter, *R*, per gene (Hudson 1987); and Wall's *B* (Wall 1999) were also calculated. Insertions and deletions are not considered in these algorithms. We assessed the spatial distribution of indels in the *ATTI* region by summing gapped sites in 17 nonoverlapping segments that were 400 bp in length without gaps.

Sequences were tested for departure from equilibrium-neutral expectations using statistics from Tajima (1989), Fay and Wu (2000), McDonald and Kreitman (1991), and Hudson *et al.* (1987). For Fay and Wu's *H*, coalescent simulations based on a neutral infinite-sites model assuming a large constant population size were employed to identify significant departure from neutral expectations. Input parameters for coalescent simulations were: (1) the observed θ per gene (estimated as the average number of nucleotide differences), (2) the sample size, (3) the estimated recombination parameter per gene (R), and (4) the estimated value of Fay and Wu's *H*. Homogeneity of diversity and divergence in silent sites along the tandem array was tested using a multilocus Hudson, Kreitman, and Aguadé (HKA) test for eight nonoverlapping, artificially defined loci each encompassing ~500 nongapped positions (multilocus HKA; distributed by Jody Hey through http://www.lifesci.rutgers.edu/heylab/).

RESULTS

Heterogeneous *ATTI* expression: Semiquantitative RT-PCR estimating mRNA levels of *ATTI* loci in rosette leaves for seven accessions of *A. thaliana* demonstrated that five of the six *ATTI* loci were transcribed. Although positive controls for three primer pairs with genomic DNA resulted in *ATTI5* amplification, we never observed *ATTI5* bands in RT-PCR analyses for any accession or tissue type (including basal leaves, flowers, fruits, and roots; results not shown). Consequently, the *ATTI5* locus was excluded from subsequent analyses. The *ATTI5* locus was identified here for the first time on the basis of sequence similarity and position within the tandem array from the accessions Cvi-1, Fe-1a, Nd-1, and Wei-0 (see below). The *ATTI4* locus, presently not annotated in the Col-0 sequence as a member of the trypsin inhibitor gene family, was transcribed in all accessions tested.

There was differential regulation among *ATTI* loci, among haplotype classes, and in response to herbivory, as seen by the significant main effects in the ANOVA of log-transformed transcript levels (Table 1). Gene regulation differed among loci with *ATTI3* and *ATTI6* having the highest transcript levels (Table 1; Figure 2A). The haplotype classes differed significantly in transcription; overall, haplotype A (HA) had higher transcript levels (Table 1; Figure 2B). Only *ATTI6* showed the opposite response, with haplotype B (HB) having more transcripts (Figure 2B). Accessions were assigned to haplotype class on the basis of the presence (HB) or absence (HA) of the *ATTI5* region, as well as >150 nucleotide polymorphisms in linkage disequilibrium throughout the array (see below). The effect of accession nested within haplotype class explained no additional variation in transcript levels (Table 1). The herbivory treatment accounted for the largest source of variation in the model, with herbivore-induced plants having, on average, 4.5 times greater transcript levels than uninduced control plants (Table 1; Figure 2C). Loci were induced differently by herbivory, such that transcription plasticity was greater than average for *ATTI3* and lower than average for *ATTI2* (Table 1, linear contrasts; Figure 2C). Plant age did not account for significant variation in expression. However, the two-way interaction of locus and age was significant, indicating the potential for age-specific regulation among loci (Table 1; Figure 2A). The remaining two-way as well as all three- and four-way interaction terms were not significant and, therefore, were removed from the model for clarity.

Nucleotide variation in the *ATTI* array: The alignment of 17 accessions of *A. thaliana* for the *ATTI* tandem array located on chromosome II spanned 9019 bp (from

TABLE 1

Analysis of variance of *ATTI* transcript level in *A. thaliana*

Source[a]	d.f.	Type III SS	Mean square	*F* value	*P*
Locus	4	251.17	62.79	34.20	<0.001
Haplotype class	1	33.50	33.50	18.25	<0.001
Herbivory	1	247.72	247.72	134.91	<0.001
Age	1	2.07	2.07	1.12	0.289
Ecotype (haplotype)	5	13.24	2.65	1.44	0.208
Locus × haplotype	4	132.53	33.13	18.04	<0.001
Locus × herbivory	4	22.54	5.64	3.07	0.016
Locus × age	4	41.28	10.32	5.62	<0.001
Herbivory × age	1	7.83	7.83	4.26	0.040
Linear contrasts[b]: locus × herbivory	d.f.	Contrast SS	Mean square	*F* value	*P*
ATTI1	1	0.01	0.01	0.01	0.943
ATTI2	1	7.26	7.26	3.95	0.047
ATTI3	1	19.04	19.04	10.37	0.001
ATTI4	1	1.82	1.82	0.99	0.320
ATTI6	1	0.05	0.05	0.03	0.869

SS, sum of squares.

[a] Model $R^2 = 0.52$.

[b] Linear contrasts were conducted within the ANOVA to determine which loci differ significantly in herbivory induction.

3′ of the preceding to 5′ of the succeeding gene). An analysis based on 7016 nongapped sites identified 239 polymorphic sites (including 59 singletons), 16 haplotypes, nucleotide diversity estimates of $\pi = 0.0107$ and $\theta = 0.0102$, and a minimum of six recombination events. Unless otherwise stated, analyses and references to position are based on a 10,352-bp global alignment of 17 *A. thaliana* accessions with *A. lyrata* as the outgroup.

Under neutrality, nucleotide sequences exhibiting high divergence among species are also predicted to evolve at high rates within species. In the *ATTI* region, however, diversity and divergence were uncorrelated ($r_{\text{Spearman}} = 0.137$; $P = 0.996$; $n = 17$ nonoverlapping 300-bp windows), and a multilocus HKA test rejected homogeneity of polymorphism and divergence (eight nonoverlapping loci of ~500 bp; $\chi^2 = 24.71$; $P < 0.001$). In relation to divergence, polymorphism was disproportionately lower in the 5′ and 3′ ends of the region and higher flanking *ATTI4*, in the middle of the array (Figure 3).

Significant linkage disequilibrium was evident throughout the entire tandem array (55% of comparisons by Fisher's exact test significant at $P < 0.001$; Figure 4), giving rise to two haplotype clades (Figure 5). In addition to the six recombination events identified via the Hudson and Kaplan (1985) algorithm applied to nucleotide polymorphisms for all accessions, visual inspection of indel and nucleotide polymorphisms together suggested two novel recombination events between positions 4470–4607 and 10009–10107 (Figure 4). Distribution of these eight recombination events was not uniform; events were concentrated in the 5′ and 3′ ends of the array (*a posteriori* Fisher's exact test of the number of recombination events in the first and last quarters of polymorphic sites combined *vs.* middle 50%; one-tail $P = 0.038$; Figure 4). Three of eight recombination events resulted in a switch of the downstream haplotype category for that accession (interhaplotype recombination; Figure 4). The common haplotype (HA) was represented by the Col-0 sequence. The less frequent haplotype (HB), typified by the presence of the *ATTI5* locus (see below), was found in 4 of 17 accessions in this study (frequency of HB is 0.24); HB was found at a frequency of 0.17 in a sample of 80 randomly chosen accessions of *A. thaliana* (results not shown).

Within the *ATTI* array, a peak in species-wide diversity centered on *ATTI5* was the result of a large number of fixed differences among haplotypes (Figure 6A; $\pi > 0.04$ in one 400-bp window; $n = 17$ accessions). Diversity within each haplotype was similarly low throughout the array (HA $\pi = 0.0032$; HB $\pi = 0.0027$) and showed no underlying peaks in polymorphism near *ATTI5* (Figure 6A). Using 2 accessions that are representative of HA and HB (Col-0 and Fe-1, respectively), we polarized each polymorphism in the array using the *A. lyrata* sequence as the outgroup. Equal numbers of derived nucleotide substitutions were found in the two lineages (62 in each Col-0 and Fe-1a). Even within the largest nonrecombining region in the array interior (Figure 4), derived and ancestral substitutions were evenly distributed among haplotypes (Table 2).

We tested whether the species-wide frequency spectrum of polymorphisms deviated from expectations under neutral equilibrium population dynamics. A sliding

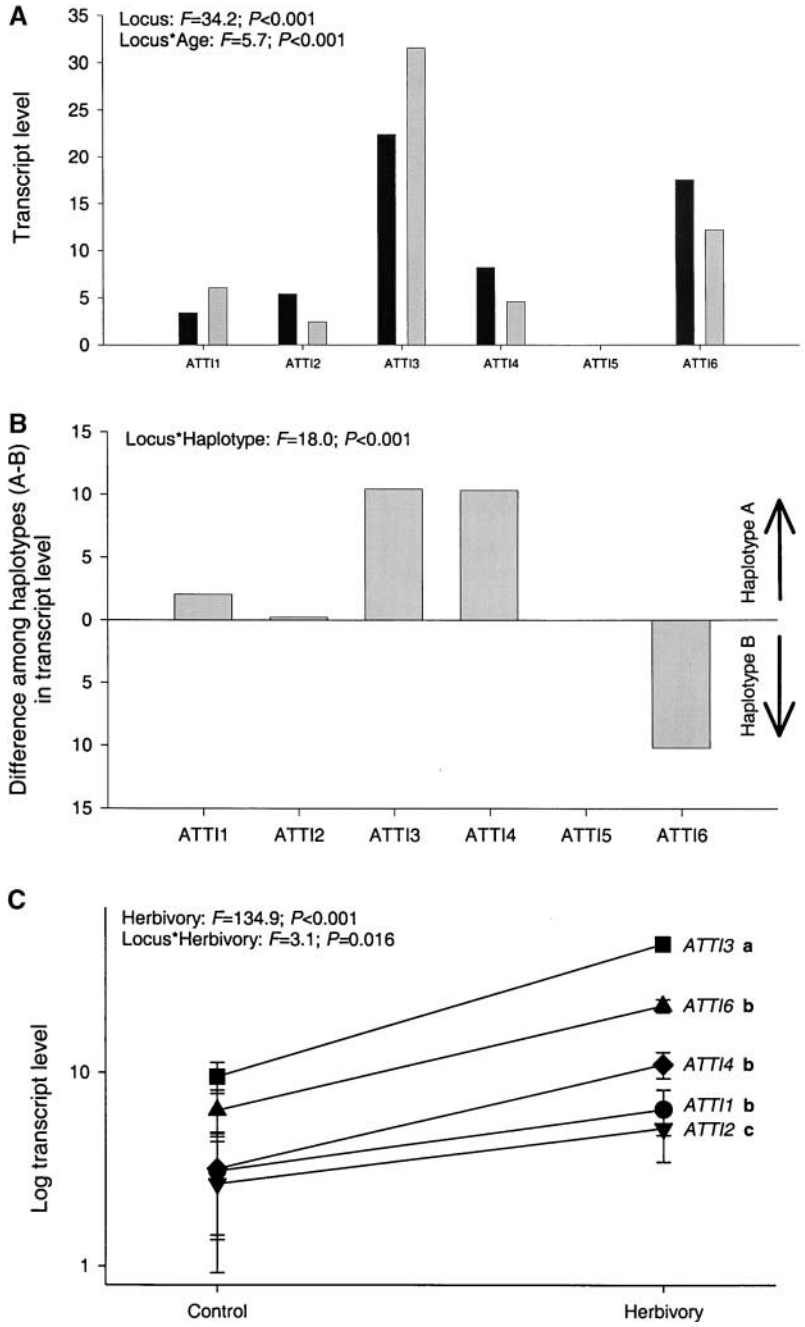

FIGURE 2.—Transcript levels of six *ATTI* loci for seven accessions of *A. thaliana* showing the effect of (A) locus and age (4 weeks, solid bars; 6 weeks, shaded bars), (B) locus and haplotype class, and (C) locus and herbivory. In B, the interaction between locus and haplotype class is graphically represented as expression of haplotype A minus the expression of haplotype B. Plotted values are least-square means from the full factorial ANOVA with untransformed data.

window analysis of the *ATTI* region demonstrated that, on average, Tajima's D was not different from zero ($D = 0.22$; $P > 0.10$). However, two 400-bp segments did deviate significantly from zero: $D = -2.08$ ($P < 0.05$) and $D = 2.26$ ($P < 0.05$), corresponding to the 5′ regions of ATTI2 and ATTI6, respectively (Figure 6B; see *Single-locus analysis* below). A sliding window analysis of Fay and Wu's H demonstrated that, within the context of an overall excess of high-frequency-derived polymorphisms ($H = -36.06$; $P < 0.001$), distinct deviations from neutrality were associated with *ATTI2* and *ATTI4* (Figure 6C).

Insertion/deletion variation in the *ATTI* array: The pattern of insertions and deletions contains valuable population genetic information and simultaneously can be the source of hidden biases in the analysis of nucleotide diversity. Alignment algorithms employing a high or low gap penalty can result in elevated or reduced diversity parameters within the resulting alignments, respectively. Indels were found throughout the *ATTI* array, with an indel peak roughly coinciding with the peak of polymorphism ($r_{Spearman} = 0.48$; $P = 0.049$; $n = 17$ nonoverlapping 400-bp windows; Figure 7). This positive correlation is not expected if high diversity were simply an alignment artifact. When we polarized 230 gapped sites in accessions representative of haplotypes A and B relative to the *A. lyrata* outgroup (Col-0 and Fe-1a, respectively), we observed approximately equal numbers of insertion and deletion events (20 and 23,

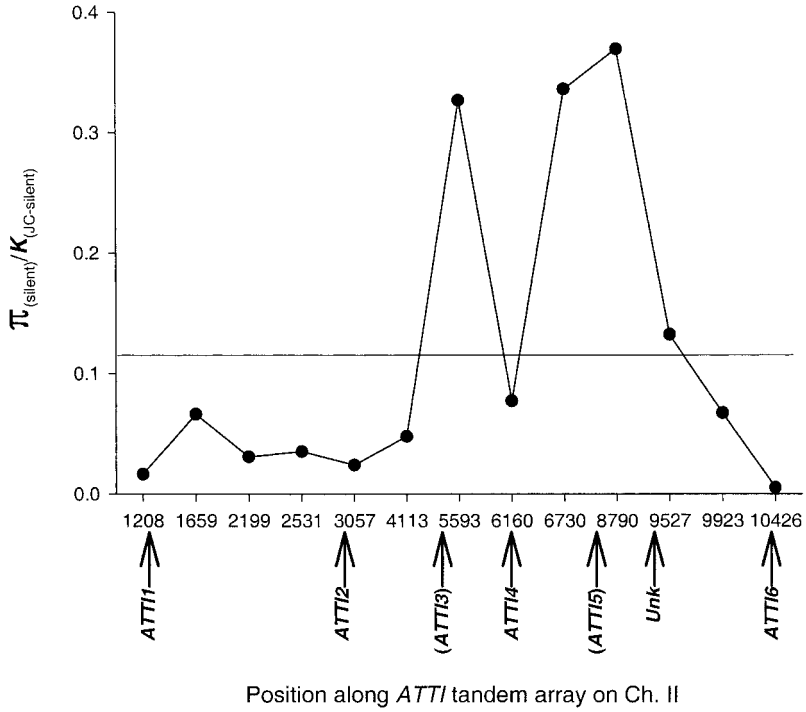

FIGURE 3.—The ratio of silent-site diversity (π) in *A. thaliana* to silent-site divergence (K) from *A. lyrata* ssp. *petraea* for nonoverlapping 400-bp windows. Approximate locations of *ATTI* loci are indicated. Loci in brackets fall within the indicated windows in *A. thaliana*, but were not included in the calculation because they were deleted from *A. lyrata* ssp. *petraea*. Average π/K = 0.12, indicated by horizontal line.

respectively), although 2.7 times more sites were deleted (163) than were inserted (67). Derived indels were distributed evenly among haplotype clades surrounding the ATTI5 deletion (Table 2).

In the Ita-0 accession, we identified a 4857-bp indel that appears to be a class II (DNA) transposable element insertion as evidenced by the presence of an 11-bp terminal inverted repeat, flanking 5-bp target site duplication, and sequence similarity to *A. thaliana* MuDR-like elements. The element was located within the 5′ untranslated region (UTR) of *ATTI4* (−85 bp of ATG, with no discernible effect on transcription; M. J. CLAUSS, unpublished data) and was lacking in the remaining 16 accessions as well as in the *A. lyrata* outgroup. The second largest indel was a 1505-bp deletion observed in 13 of 17 accessions, which includes the previously undescribed *ATTI5* locus (Figure 1). The *ATTI5*$^+$ allele was 43% diverged from the most similar locus; hence *ATTI5*$^-$ probably represents the recent deletion of an old gene family member. Consistent with this hypothesis, diversity within the *ATTI5* region (π = 0.0034 for positions 7273–8777) was not reduced relative to average diversity across the entire tandem array (π = 0.0027 for positions 197–10,548) when only the four accessions with *ATTI5*$^+$ (haplotype B; Figure 6A) were considered. Comparison with the outgroup *A. lyrata* was not informative because *ATTI5*, as well as *ATTI3*, were deleted in all alleles surveyed thus far (M. J. CLAUSS, unpublished data). Although the *ATTI5* deletion site in *A. thaliana* and *A. lyrata* was within 100 bp, homology of the deletion was uncertain because of ambiguous alignment in this highly repetitive and polymorphic region.

Single-locus analysis: The six *ATTI* loci differ in their level of diversity within *A. thaliana*, the frequency distribution of polymorphisms, and among-species divergence (Tables 3, 4, and 5; Figure 7). The consensus sequence of all *ATTI* loci was in frame, with no pseudogene signatures. Variation in sequence diversity among the five expressed *ATTI* loci was positively correlated with variability in transcript levels among accessions (Figure 8). For the nontranscribed *ATTI5* locus only, a GA microsatellite located 3–13 bp 5′ to the ATG in each TI locus was large and variable (up to 25 repeats), and there was a single missense replacement substitution in one accession (Cvi-1). As expected for functional genes, average diversity at *ATTI* loci (π = 0.0057) was lower than diversity across the ~10 kb spanning the array (π_{total} = 0.0107 or π_{silent} = 0.0135; Tables 3 and 4). Within this chromosomal context, diversity for *ATTI1* and *ATTI2* was extremely low (π = 0.0009), whereas π for *ATTI4* was high (0.0110), approaching that of silent-site diversity. Differences in the extent of haplotype structure were responsible for some of this diversity: *ATTI1* and *ATTI2* had only singleton polymorphisms, whereas the two dominant haplotypes could be identified in the remaining loci (*e.g.*, Wall's *B* for *ATTI4* and *ATTI6* was 0.54 and 0.44, respectively; Table 4). Among functional regions, exon 1, coding for most of the signal peptide, had the greatest average diversity, even higher than that of the 5′-UTR (Table 3). Significant deviation in the frequency spectra from equilibrium-neutral expectations was observed for the *ATTI2* locus (including flanking region; Tajima's D = −2.02; $P < 0.002$) and *ATTI6* (coding sequence only; D = 2.20; $P < 0.01$; Table 4). After a sequential Bonferroni correction (HOLM 1979) is applied to all 16 tests of

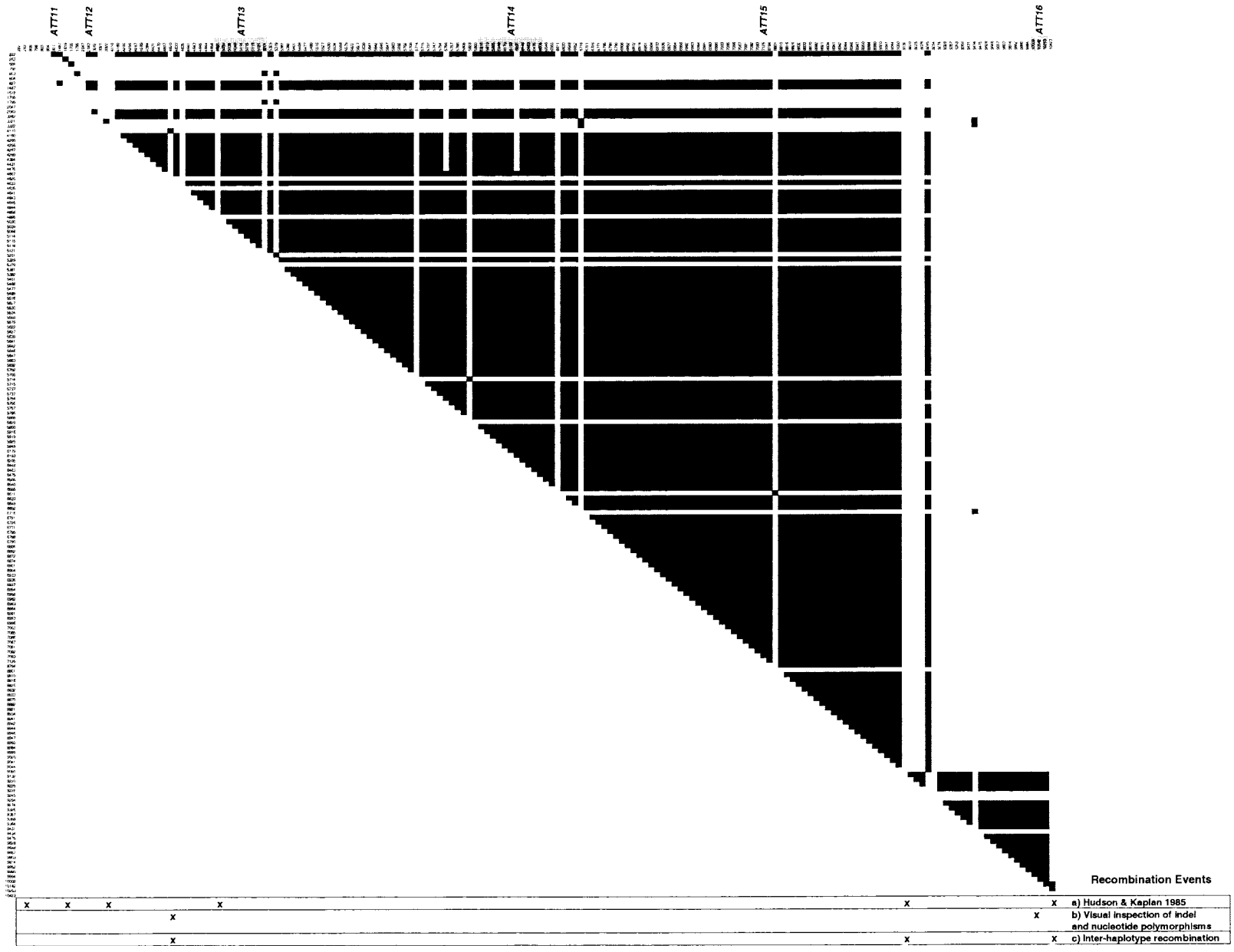

FIGURE 4.—Polymorphic sites among 17 accessions of *A. thaliana* showing significant linkage disequilibrium (Fisher's exact test; $P < 0.001$ for 8608 of 15,576 pairwise comparisons) across 10 kb on chromosome II. Locations of *ATTI* loci are indicated and shaded where polymorphic sites exist within the coding sequence. Although not included in the analysis due to indels, the approximate position of ATTI5 is also given. An unknown gene is located between 8958 and 9467. X identifies the approximate upper position of the inferred recombination events from (a) the HUDSON and KAPLAN (1985) algorithm based on nucleotide polymorphisms and (b) visual inspection of both nucleotide and indel polymorphisms, as well as (c) the interhaplotype recombination events.

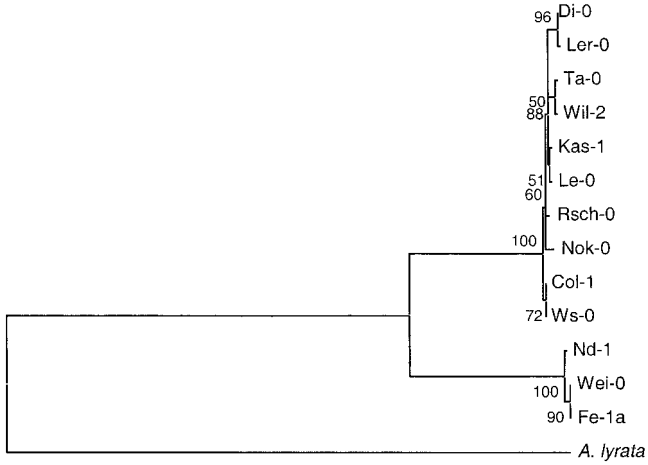

FIGURE 5.—Neighbor-joining analysis with 1000 bootstrap replicates showing nodes >50% for 3411 silent-site polymorphisms in *ATTI* array among the 13 noninterhaplotype recombining accessions of *A. thaliana* using one outgroup sequence from *A. lyrata* ssp. *petraea*.

site frequency spectra in Table 4, these results remain significant at the $P = 0.05$ level.

Although average *ATTI* divergence between *A. thaliana* and *A. lyrata* was similar to divergence across the entire chromosomal region ($K = 0.100$), K ranged from 0.0661 to 0.1671 among loci (Table 3). Among coding positions (~270 bp per gene) in the four *ATTI* loci for which an outgroup comparison was possible, divergence and diversity were positively correlated ($r_{S\text{pearman}} = 0.9487$; $P = 0.05$; $n = 4$; Table 3). Although the significant HKA test reported above rejected homogeneity for the entire 10-kb region, HKA was nonsignificant when only the coding regions were compared (results not shown). Among functional regions, the lowest divergence, on average, was seen for coding sequences (8%), whereas the 5′-UTR was on average 17% divergent (Table 3). High- frequency-derived polymorphisms were in significant excess for *ATTI2* (Fay and Wu's *H*; Table 4).

A comparison of synonymous (K_s) and amino acid replacement (K_a) substitutions can provide evidence of the evolutionary rate of functionally important changes on three levels in gene families: intraspecific diversity within loci, species divergence within loci, and divergence among loci. For *ATTI*, we found lower replacement than synonymous rates of evolution at each level. First, within *A. thaliana*, π_a/π_s varied from 0.29 (*ATTI1*) to 0.76 (*ATTI6*) along the array, suggesting different levels of purifying selection among loci (Table 5). Second, among species, the K_a/K_s ratio was lower than π_a/π_s within *A. thaliana* (Table 5). It is notable that K_a/K_s for *ATTI6* was the lowest observed (0.14). Third, comparisons among all pairs of *ATTI* loci for one representative accession (Fe-1a) also illustrated overall selective constraint in trypsin inhibitor evolution ($K_a/K_s \ll 1$): average K_a/K_s ratios for all comparisons including *ATTI1* through *ATTI6* were 0.33, 0.40, 0.44, 0.43, 0.61, and 0.21, respectively. Thus, *ATTI5* was under the least selective constraint (0.61), whereas amino acid evolution appears most constrained for *ATTI1* and *ATTI6*.

Of particular interest for understanding changes in inhibitory function are replacement changes in the reactive site loop (Figure 1). While there were synonymous changes between the P3 to P3′ positions in some loci, no within-species replacement polymorphisms were observed within this critical region. The most common residue at the P1 position for *ATTI*, arginine, was replaced by lysine in *ATTI4* and *ATTI6* (Figure 1).

DISCUSSION

We found a significant association between gene expression and haplotype class for loci of the *A. thaliana* trypsin inhibitor (*ATTI*) tandem gene family. Membership in two molecularly defined haplotype classes explained a significant proportion of the variation in transcription, whereas the contribution of accession nested within haplotype was not significant (Table 1). Haplotype classes differed up to sixfold in transcript level (Figure 2B). Within *A. thaliana*, linkage disequilbrium generally extends over <250 kb; hence a significant association between haplotype and function suggests *cis*-regulation at or near the tandemly arrayed *ATTI* loci, whereas unlinked *trans*-acting regulation is less likely (HAUBOLD *et al.* 2002; NORDBORG *et al.* 2002; SHEPARD and PURUGGANAN 2003).

An association between naturally occurring sequence haplotypes and function has been observed in *A. thaliana* for several qualitative phenotypes relating to pathogen resistance (STAHL *et al.* 1999; TIAN *et al.* 2002; MAURICIO *et al.* 2003). However, in an analysis of quantitative variation for a plant-herbivore defense trait, HAUSER *et al.* (2001) do not detect a significant relationship between nucleotide polymorphism and trichome density. Identifying significant functional differences among divergent alleles is critical for efforts to connect nucleotide sequence variation to ecologically important characteristics undergoing natural selection. In this study of transcription of herbivore-induced proteinase inhibitor genes, we identified significant associations between sequence haplotypes and quantitative functional variation in just one component of a complex phenotype. Additional components such as protein synthesis, inhibitory activity, and the fitness consequences of alternative alleles remain to be assessed. Population genetic analyses of polymorphisms can also provide evidence of the adaptive significance of functional variation. Below, we describe the chromosomal landscape of polymorphism surrounding the *ATTI* tandem array and consider the evidence for locus-specific functional evolution.

***ATTI* loci in the chromosomal landscape:** Tandemly duplicated genes located in close proximity to one another not only share a demographic and functional

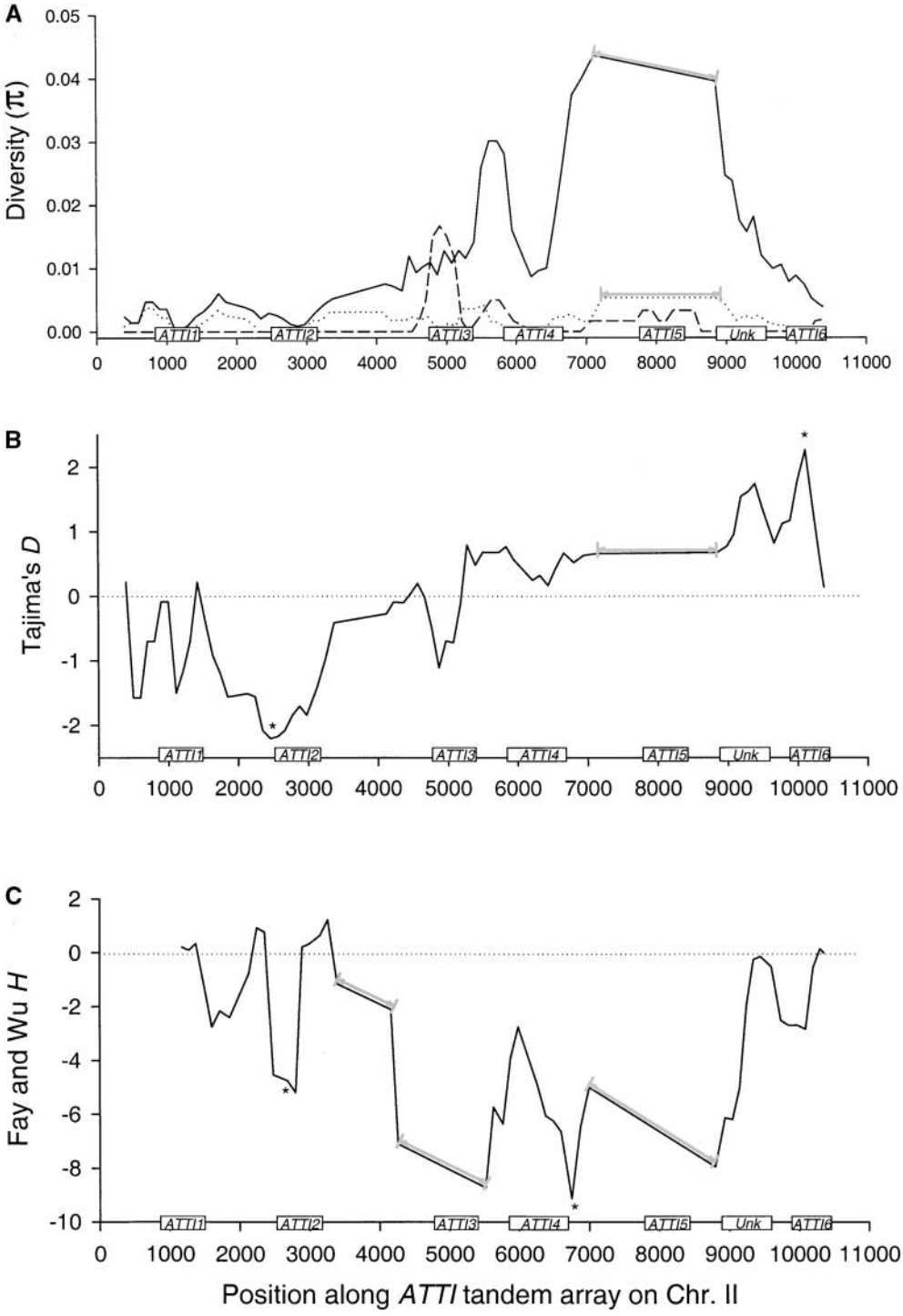

FIGURE 6.—Patterns of nucleotide polymorphism among 17 accessions of *A. thaliana* spanning 10 kb on chromosome II. Position is given from global alignment with *A. lyrata* ssp. *petraea*, and the locations of loci in the *ATTI* tandem array are indicated. All windows are 400 bp with a 100-bp step, shaded double arrows represent regions of indels, and asterisks indicate regions of significant deviation from neutrality for the respective test ($P < 0.01$). (A) Diversity (π_{total}) for all accessions (solid line; $n = 17$), haplotype A accessions not involved in interhaplotype recombination (dotted line; $n = 10$), and haplotype B accessions not involved in interhaplotype recombination (dashed line; $n = 3$). (B) Tajima's D ($n = 17$) and (C) Fay and Wu's H ($n = 17$ accessions and *A. lyrata* ssp. *petraea*).

history, but also exist within a shared chromosomal landscape. For 10,352 aligned nucleotide positions, we observed: (1) average levels of overall polymorphism within *A. thaliana* and divergence to *A. lyrata*, (2) two distinct haplotype classes in *A. thaliana* with similar nucleotide diversity within each class, (3) significant linkage disequilibrium, (4) abnormally low levels of diversity associated with *ATTI1* and *ATTI2*, and (5) a peak of polymorphism spanning the presence/absence polymorphism of the previously undescribed *ATTI5* locus. Significant heterogeneity of polymorphism relative to divergence allowed us to reject the hypothesis that the genomic region encompassing the *ATTI* tandem array has evolved according a homogeneous neutral model (Figure 3; multilocus HKA; $P < 0.001$).

The co-occurrence of a peak of polymorphism at one functional locus and distinct haplotype classes has previously been reported in several studies of *A. thaliana*

TABLE 2

Distribution of derived indel and nucleotide polymorphisms in two haplotype classes of *A. thaliana* regions flanking the *ATTI5* deletion

	5′ flanking: 6165–7149		3′ flanking: 8778–9050	
	Haplotype A: *ATTI5*$^{-}$	Haplotype B: *ATTI5*$^{+}$	Haplotype A: *ATTI5*$^{-}$	Haplotype B: *ATTI5*$^{+}$
No. derived nucleotide polymorphisms[a]	23	20	10	10
No. derived indel events[a]	8	11	2	3
No. derived indel sites	34	84	9	26

[a] Homogeneity tests identified no significant differences in the distribution among haplotypes for nucleotide polymorphisms ($G = 0.07$; $P = 0.790$) and the number of indel events ($G = 0.336$; $P = 0.562$).

molecular evolution and may indicate balancing or frequency-dependent selection (Stahl *et al.* 1999; Tian *et al.* 2002, 2003; Mauricio *et al.* 2003; but see Aguadé 2001). The peak of polymorphism documented for *ATTI* differs from previous reports in several ways.

First, the polymorphism peak in the *ATTI* tandem array spans >1.2 kb of nongapped sequence, including three genes, their promoter regions, and intergenic sequences (Figures 3 and 6). The genes are part of a tandem array of duplicated loci that are functionally differentiated and thus represent several linked targets of selection (*e.g.*, Parsch *et al.* 2001; Baines *et al.* 2002).

Second, the peak of nucleotide polymorphism in the *ATTI* array centers on a presence/absence polymor-

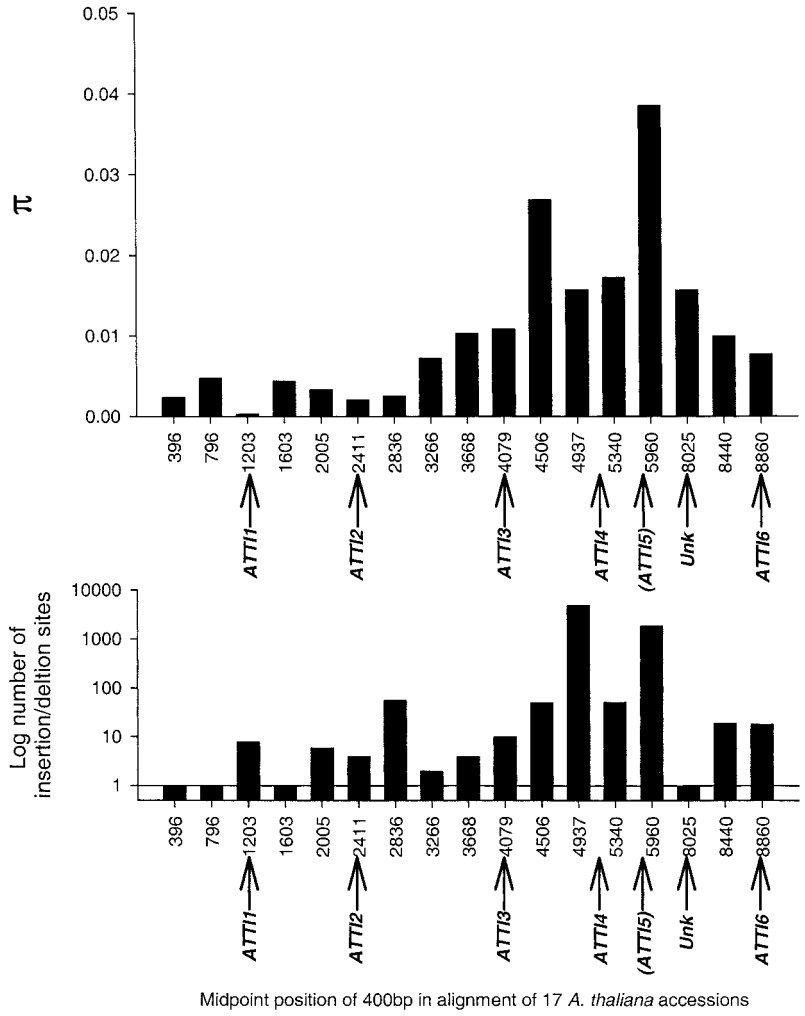

Figure 7.—Polymorphism among 17 accessions of *A. thaliana* over ~10 kb calculated for nonoverlapping windows 400 bp in length excluding gaps for (top) nucleotide polymorphisms and (bottom) insertion/deletion polymorphisms (log plus one). Correlation between nucleotide and indel polymorphisms was $r_{Spearman} = 0.48$, $P = 0.049$ for all data and $r_{Spearman} = 0.45$, $P = 0.069$ after removal of the two largest indels. Approximate locations of *ATTI* loci are indicated by arrows in an alignment of only *A. thaliana* sequences. *ATTI5* was not included in the calculations because it is located within an indel. Unk is At2g43540 (see materials and methods).

TABLE 3

Diversity (π) among 17 accessions of *A. thaliana* and divergence (K) to *A. lyrata* ssp. *petraea* for six *ATTI* loci

Loci	Diversity or divergence	5′-UTR	Exon 1	Intron 1	Exon 2	3′-UTR	CDS	Total
	Average bp π:	73.7	60.2	130.7	211.5	154.3	271.7	630.3
	Average bp K:	74.3	59.0	119.8	212.8	146.5	271.8	603.0
ATTI1	π	0.0000	0.0019	0.0000	0.0006	0.0017	0.0009	0.0009
	K	0.0561	0.0176	0.0647	0.0610	0.0905	0.0508	0.0661
ATTI2	π	0.0047	0.0019	0.0000	0.0006	0.0030	0.0009	0.0013
	K	0.2571	0.0676	0.1527	0.0497	0.0809	0.0537	0.1007
ATTI3	π	0.0142	0.0075	0.0148	0.0093	0.0117	0.0089	0.0115
	K	—	—	—	—	—	—	—
ATTI4	π	0.0148	0.0247	0.0105	0.0073	0.0175	0.0110	0.0127
	K	0.1893	0.1047	0.0803	0.1222	0.0919	0.1185	0.1214
ATTI5[a]	π	0.0061	0.0164	0.0047	0.0000	0.0060	0.0035	0.0047
	K	—	—	—	—	—	—	—
ATTI6	π	0.0068	0.0265	0.0000	0.0049	0.0033	0.0093	0.0058
	K	0.1816	0.1538	0.2024	0.0680	0.2591	0.0851	0.1671
Average	π	0.0078	0.0131	0.0050	0.0038	0.0072	0.0057	0.0061
	K	0.1710	0.0859	0.1250	0.0752	0.1306	0.0770	0.1138

CDS, coding sequence.
[a] *n*, four accessions.

phism for an old, apparently nonfunctional, gene copy (*ATTI5*). Whereas evidence in favor of an indel polymorphism maintained by selection is strengthened by the presence of at least one functional allele (*e.g.*, *RPM1*; Tian *et al.* 2003), our analysis indicated that both the complete gene deletion and the *ATTI5*$^+$ allele were nonfunctional. The lack of transcription, elevated K_a/K_s ratio, and a large polymorphic microsatellite immediately 5′ of the ATG all suggest that *ATTI5*$^+$ is a pseudogene and, hence, not a candidate for an extant balanced polymorphism. However, loss-of-function in *ATTI5*$^+$ appears to be relatively young, and thus this locus may have influenced past evolution in the array. Comparative data from *A. lyrata* do not shed light on this process, because *ATTI5* was also deleted in *A. lyrata* and the deletion site for both species is in the same repetitive, polymorphic region for which sequence alignment was ambiguous. The high level of divergence between *ATTI5* and all other *ATTI* genes (>43%) suggests that the indel represents the deletion of an old *ATTI5* allele, rather than a recent insertion via duplication from an existing TI locus. Recent gene conversion is also unlikely because (a) the most similar TI, *ATTI4*, was 43% divergent from *ATTI5*$^+$ and (b) ancestral and derived nucleotide polymorphisms were evenly distributed among *ATTI5*$^+$ and *ATTI5*$^-$ haplotypes in the nonrecombining region flanking the deletion (Table 2).

Third, although the peak of polymorphism reflects fixed differences among two haplotype classes spanning several *ATTI* loci, the allele frequency spectrum does

TABLE 4

Analysis of site frequency spectra for 17 accessions of *A. thaliana*

Loci	Tajima's *D*				Fay and Wu's *H*		
	Total	*P*	CDS	*P*	Total	*P*	Wall's *B*
ATTI1	−1.10	NS	−1.50	NS	0.46	NS	0.00
ATTI2	−2.02	0.002[a]	−1.50	NS	−5.07	0.01[a]	1.00
ATTI3	−0.37	NS	−0.59	NS	—	—	0.63
ATTI4	0.63	NS	0.44	NS	−7.27	0.066	0.54
ATTI5[b]	−0.31	NS	0.17	NS	—	—	0.60
ATTI6	1.05	NS	2.20	0.01[a]	−0.48	NS	0.44

NS, not significant.
[a] Significant after sequential Bonferroni correction for 16 tests at the 0.05 level.
[b] *n*, four accessions.

TABLE 5

Rates of synonymous (s) and replacement (a) substitution within *A. thaliana* for 17 accessions and between *A. thaliana* and *A. lyrata* ssp. *petraea*

	A. thaliana			*A. thaliana*-*A. lyrata*				
	π_s	π_a	π_a/π_s	K_s	K_a	K_a/K_s	π_s/K_s	π_a/K_a
ATTI1	0.0020	0.0006	0.29	0.1202	0.0322	0.27	0.02	0.02
ATTI2	0.0019	0.0006	0.30	0.1419	0.0292	0.21	0.01	0.02
ATTI3	0.0162	0.0068	0.42	—	—	—	—	—
ATTI4	0.0130	0.0092	0.71	0.2150	0.0987	0.46	0.06	0.09
ATTI5[a]	0.0000	0.0045	NA	—	—	—	—	—
ATTI6	0.0093	0.0071	0.76	0.2856	0.0407	0.14	0.03	0.17
Total CDS	0.0156	0.0066	0.43	0.1721	0.0489	0.28	0.09	0.14
ATTI locus average	0.0071	0.0048	0.68	0.1907	0.0502	0.26	0.0308	0.0761
All silent (n = 3782.8)	0.0135			0.1177			0.1148	

CDS, coding sequence. NA, not applicable.
[a] *n*, four accessions.

not provide statistical support for a balanced polymorphism. The worldwide frequency of the ancestral haplotype B (estimated by the frequency of the *ATTI5*$^+$ allele) was 0.17, and nucleotide polymorphisms in linkage disequilibrium with *ATTI5* did not deviate significantly from expectations under a neutral equilibrium model (Figure 6B). Diversity across the sampled 10 kb was similar for both *ATTI* haplotype classes (HA π = 0.0032 and HB π = 0.0027; or if accessions with interhaplotype recombination are excluded, HA π = 0.0017 and HB π = 0.0015), suggesting coexisting allele classes with segregating variation of similar age. The even distribution of derived sites among haplotypes also argues strongly against introgression of a divergent allele (Table 2). One possible explanation for this pattern is frequency-dependent selection for the maintenance of long-lived and divergent clades as proposed by Stahl *et al.* (1999) for the *RPM1* pathogen-defense gene. Although our peak spans an unlikely target of current selection, a simulation study by Nordborg and Innan (2003) argues that the peak of polymorphism need not always be centered on the site of selection (see discussion of *ATTI6* below). However, two dominant haplotype classes are also a likely outcome of a neutral coales-

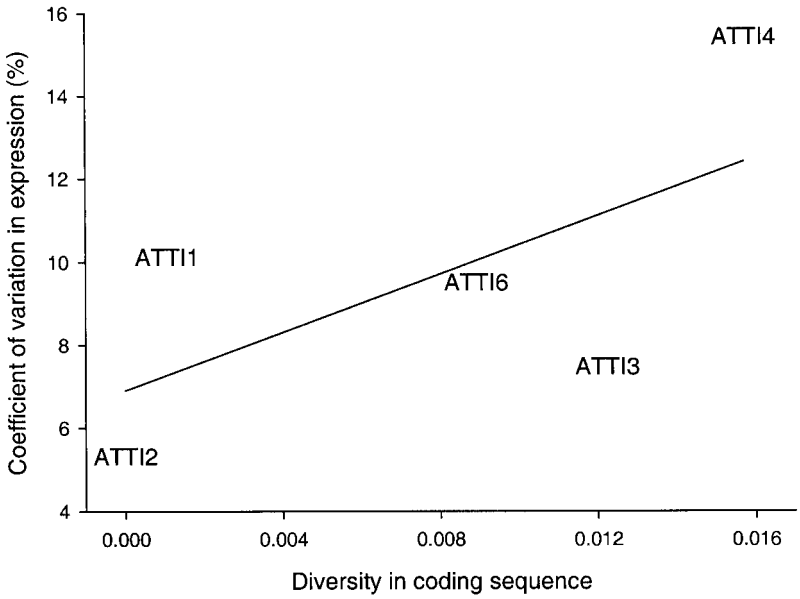

Figure 8.—Correlation between sequence diversity in the coding sequence and the coefficient of variation in expression among seven accessions of *A. thaliana* for members of the *ATTI* gene family (r^2 = 0.41).

cence process without recombination (*e.g.*, as seen within loci in inbreeding populations; Hudson 1990; Aguadé 2001; Charlesworth 2003).

Functional evolution in closely linked *ATTI* loci: *ATTI* loci exhibit substantial sequence divergence (65%), differences in reactive site residues determining functional specificity, and variation in constitutive as well as induced transcription (Figures 1 and 2). In *A. thaliana* there was an association of nucleotide polymorphism with *ATTI* transcript level as demonstrated by significant haplotype and locus × haplotype effects on transcription (Table 1; Figure 2) and by the positive correlation between sequence diversity and variability in transcription among loci (Figure 8). This functional and molecular diversity, together with multiple recombination events (Figure 4), suggests the potential for independent adaptive evolution among tandem duplicates in *A. thaliana*. However, the coding sequences of *ATTI* loci are in close physical linkage (neighboring genes are separated on average by only 716 bp), and we estimated significant linkage disequilibrium, particularly among polymorphisms in the interior of the tandem array (Figure 4). Below, we interpret the patterns of polymorphism and divergence in functionally differentiated TI loci within the constraints of their genetic backgrounds.

Unique patterns of sequence diversity and expression for the first two TI loci indicate an evolutionary history distinct from the remaining gene family members. *ATTI1* and *ATTI2* exhibit extremely low levels of intraspecific polymorphism in comparison to other *A. thaliana* loci (Tables 3 and 4; Figure 2A; Kawabe *et al.* 1997; Kuittinen and Aguadé 2000; Savolainen *et al.* 2000; Aguadé 2001; Hauser *et al.* 2001; Olsen *et al.* 2002; Tian *et al.* 2002; Wright *et al.* 2003; Ramos-Onsins *et al.* 2004), average divergence from *A. lyrata* (Table 5; Savolainen *et al.* 2000; Aguadé 2001; Ramos-Onsins *et al.* 2004), and low transcript levels (Figure 2A). Both loci appear to be under strong purifying selection; the nonsynonymous substitution rate was much less than the rate of synonymous changes (Table 5). Conservation of the amino acid sequence in the reactive site loop between *ATTI1*, *ATTI2*, and *ATTI1* orthologs in Sinapis (*MTI2*) and Brassica (*RTI3*) also indicates functional constraint (Zhao *et al.* 2002). Phage display experiments with *MTI2* have demonstrated that the P3-P3′ APRIFP reactive loop shared by *ATTI1* and *ATTI2* and, in particular, the arginine at the P1 position result in maximal inhibition of trypsin (Ceci *et al.* 2003). Nonetheless, a low π/K ratio (Figure 3) and an excess of high-frequency-derived polymorphisms concentrated in the signal peptide and 5′-UTR of *ATTI2* are consistent with recent positive selection (Figure 6, B and C; Table 4; Clauss and Mitchell-Olds 2003). An evolutionary trajectory independent from the rest of the tandem array was facilitated by several recombination events that reduced the correlation between polymorphisms in *ATTI1* and *ATTI2 vs.* downstream TI loci (Figure 4).

ATTI3, *ATTI4*, and *ATTI5* are located in a region with excess species-wide polymorphism, excess indels, and almost complete linkage disequilibrium (Figures 4, 6, and 7). Nonetheless, *ATTI3* and *ATTI4* were functional and show evidence of selective constraint ($\pi_a/\pi_s < 1$; Table 5). *ATTI3* has the common trypsin inhibitor P1 reactive site residue (arginine) and was the most highly transcribed *ATTI* locus in both control and herbivore-induced treatments (Table 1; Figure 2C). *ATTI4* has lysine at the P1 position and showed a low-to-intermediate transcription profile (Figure 2C). Lysine at the P1 position also results in trypsin inhibition (Polticelli *et al.* 1999) and is found in unrelated trypsin inhibitors (Ling *et al.* 1993). Thus, the change in reactive site at *ATTI4* from arginine to lysine is more likely to reflect modification of trypsin inhibition rather than an entirely novel function. We hypothesize that modification of the reactive site in *ATTI4* occurred within the last 5 MY (after divergence of *A. thaliana* and *A. lyrata*) because the *ATTI4* ortholog in *A. lyrata* has the putatively ancestral arginine at the P1 position. *ATTI4* also differs from the remaining functional *ATTI* array members in having a C-terminal peptide that may affect regulation of inhibitory activity toward endogenous proteases during sequestration (Figure 1A; Volpicella *et al.* 2000; De Leo *et al.* 2001a). Population genetic tests of ongoing evolution by natural selection in *ATTI4* were confounded by haplotype structure and linkage disequilibrium within the tandem array. Although Fay and Wu's *H* reached its nadir in the 3′ region of *ATTI4* (Figure 6C; Table 4), the statistic was also significantly negative for the entire nonrecombining region spanning *ATTI3*–*ATTI5* ($H = -24.7$; $P < 0.001$; $n = 1462$ aligned nongapped positions). Functional analyses also pointed to positive correlations across loci: transcription of both functional TI loci in this linkage group differed significantly among haplotypes (Figure 2B). At this juncture, we can only comment that HA for *ATTI4* had the greatest proportional increase in expression over HB observed for all loci (result not shown) and that this difference was associated with a high local concentration of derived polymorphisms (eight of nine polymorphisms derived in HA). One possible explanation for this pattern is ongoing positive selection for the high-expression *ATTI4* HA allele and hitchhiking at linked loci. Further functional experiments and population genetic analyses of a larger sample of accessions, including more naturally occurring recombinants, are needed to disentangle these correlated functional and evolutionary patterns.

The function of *ATTI6* in *A. thaliana* and *A. lyrata* appears to be modified in comparison to upstream TI loci. ATTI6 has lysine in place of arginine at the P1 position (as in *ATTI4*) and two additional amino acid substitutions in the P3-P3′ reactive site loop in comparison to *ATTI1* and *ATTI2*. The potential for independent functional evolution of *ATTI6* in *A. thaliana* was sug-

gested by three recombination events within and flanking this locus (Figure 4). *ATTI6* is functionally unique among *ATTI* loci in that transcript levels for haplotype B exceeded those of haplotype A (Figure 2B). An interhaplotype recombination event located ~880 bp 5′ of *ATTI6* was associated with a switch from haplotype A to B downstream and a concordant dramatic switch in expression. This result identifies candidate polymorphisms for regulatory control of haplotype-specific *ATTI6* expression and is consistent with promoter-deletion experiments that have identified the 520 bp 5′ of *MTI2* essential for gene expression (De Leo *et al.* 2001a). In addition to the distinctive pattern of expression for *ATTI6*, this locus also had a population genetic signature unique among *ATTI* loci. A significant excess of intermediate frequency polymorphisms was found in *ATTI6* (Tajima's $D = 2.20$; $P < 0.01$; Table 4). While we must be cautious of type I errors, we feel justified in further exploring the change in allele frequency distribution at this location in the array (Figure 6B) because of the highly significant change in expression pattern at this locus (Figure 2B). Several possible processes may give rise to an intermediate frequency polymorphism, including balancing selection at the *ATTI6* locus, directional selection at *ATTI6* constrained by selection at linked sites (*e.g.*, interference or traffic), or nonequilbrium population dynamics. We reject population admixture or other forms of population structure as the primary cause of intermediate frequency polymorphisms and elevated π_a/π_s observed at *ATTI6*, because demographic processes are expected to be pervasive throughout the genome and affect all site categories (Tables 3 and 4). The role of balancing selection or frequency-dependent selection favoring alternate alleles at *ATTI6* must be further explored via ecologically informed functional studies (*e.g.*, Tian *et al.* 2003).

Interference or traffic occurs when variants are positively selected but fail to go to fixation because of conflicts among multiple linked segregating sites (Kirby and Stephan 1996; Kim and Stephan 2003). If increased transcript levels are selectively favored, a balanced polymorphism at *ATTI6* could reflect interference between positive selection for haplotype B at *ATTI6* and positive selection for haplotype A upstream in the *ATTI3-ATTI5* linkage group (*e.g.*, Baines *et al.* 2002). Population subdivision and low outcrossing rates in *A. thaliana* reduce the probability of a recombination event coupling two putatively selected alleles separated by only 2100 bp (Nordborg *et al.* 2002). If positive selection to fix the rare beneficial recombinant is less effective in *A. thaliana* due to reduced N_e and other correlated characteristics of this inbreeding ruderal (Charlesworth 2003), the selective conflict and, thus, the alternate alleles can have long residence times in the global population. It is also possible that past interference with positive selection for haplotype A alleles at *ATTI2* or *ATTI1* may have contributed to the patterns of nucleotide and functional polymorphism at downstream loci. Our data are consistent with the hypothesis that interference plays an important role in limiting adaptive evolution in densely packed tandem genes for this highly selfing species. In the population genetic context of the closely related outcrossing species, *A. lyrata*, effective recombination appears to be widespread even within *ATTI* loci (Clauss and Mitchell-Olds 2003). Thus, constraints to functional evolution in *A. thaliana* may well be ephemeral and must not necessarily inhibit the long-term evolutionary potential of tandemly duplicated loci.

Population genetics of defense: Patterns of nucleotide polymorphism in plant defense-related genes can be divided into three categories. First, we find examples that are consistent with selective sweeps, as predicted by models of a coevolutionary arms race between plants and their enemies (Bittner-Eddy *et al.* 2000; Bergelson *et al.* 2001, Mondragon-Palomino *et al.* 2002; Zhang *et al.* 2002b; Clauss and Mitchell-Olds 2003; this study). Second, and contrary to the predictions emerging from the classic arms race model, several studies demonstrate balanced polymorphisms of ancient alleles, which have been interpreted in view of balancing selection or "trench warfare" (Stahl *et al.* 1999; Tian *et al.* 2002; Kroymann *et al.* 2003; Mauricio *et al.* 2003). Third, many loci display patterns that are complex and difficult to explain under simple evolutionary models of neutrality or selection (Kawabe *et al.* 1997; Tiffin and Gaut 2001; Zhang *et al.* 2002b; this study).

In gene-for-gene pathogen resistance, selection appears to result overwhelmingly in rapid adaptive evolution among both alleles and paralogs (Bergelson *et al.* 2001). However, in more diffuse interactions involving nonhost pathogen resistance or insect herbivory, selection may not follow classical gene-for-gene or race-specific predictions (Ellis *et al.* 2000; Collins *et al.* 2003). In the *ATTI* gene family we observed purifying selection for five of six functionally differentiated loci. These divergent TI loci may function as "fixed heterozygotes," providing an effective and ecologically flexible defense against trypsin proteinases of the diverse specialist and generalist herbivores and pathogens that attack Arabidopsis. An effective horizontal strategy for maintaining functional diversity through subfunctionalization may reduce the likelihood that allelic variation at any individual locus will be subject to positive selection. The opportunity for positive selection within *ATTI* loci may be further reduced by structural constraints in the mature protein. *A. thaliana* trypsin inhibitor sequences show convergence with entirely unrelated small proteins involved in protein-protein interactions whose basic structure is also given by four disulfide bonds (Ascenzi *et al.* 1999; Zang and Maizels 2001; Zhao *et al.* 2002). Empirical studies have demonstrated that copy number and intergenic divergence and exchange, as well as allele diversity, are all components of the evolutionary

response to selection pressures (Duda and Palumbi 2000; Bergelson *et al.* 2001). Understanding the relative contribution of these components under different selection regimes and in varying genomic backgrounds remains an exciting challenge in our quest to map genotype to phenotype for ecologically important traits.

Conclusions: The patterns of variability among tandemly duplicated *ATTI* loci capture a snapshot of the process of diversification through subfunctionalization. The identical P1 trypsin inhibitor reactive site, arginine, has been maintained in three *ATTI* loci that, nonetheless, differed in constitutive and herbivore-induced expression. A change in reactive site residue and an extended C-terminal sequence appear to modify function in two other *ATTI* loci. After being maintained for millions of years, a sixth *ATTI* locus has recently become a pseudogene and was segregating as an insertion/deletion polymorphism in *A. thaliana*. While physical linkage has not inhibited divergence over the evolutionary history of this gene family, linkage disequilibrium in the current population genetic context of *A. thaliana* may limit the opportunity for natural selection at some *ATTI* loci due to interference. The loci at the periphery of the tandem array (*e.g.*, *ATTI1* and *ATTI6*) were separated from the remaining *ATTI* loci by several recombination events and had greater opportunity for independent evolution by natural selection than did loci in the array interior. In the population genetic context of the self-incompatible, outcrossing perennial *A. lyrata*, recombination is widespread even within *ATTI* loci, suggesting that the constraints against functional evolution of the trypsin inhibitor gene family seen in *A. thaliana* may be ephemeral.

We thank S. Dietel, A. Figuth, and D. Schnabelrauch for technical assistance and J. Kroymann and H. Vogel for scientific support, as well as A. Lawton-Rauh and E. Schranz for useful comments on an earlier version of the manuscript. Financial support was provided by the Max-Planck Gesellschaft, European Union contract no. QLRT-2000-01097, the Bundesministerium fuer Bildung und Forschung, and grant DEB-9527725 from the U.S. National Science Foundation.

LITERATURE CITED

Aguadé, M., 2001 Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. Mol. Biol. Evol. **18:** 1–9.

Ascenzi, P., M. Ruoppolo, A. Amoresano, P. Pucci, R. Consonni *et al.*, 1999 Characterization of low-molecular-mass trypsin isoinhibitors from oil-rape (*Brassica napus* var. *oleifera*) seed. Eur. J. Biochem. **261:** 275–284.

Baines, J. F., Y. Chen, A. Das and W. Stephan, 2002 DNA sequence variation at a duplicated gene: excess of replacement polymorphism and extensive haplotype structure in the *Drosophila melanogaster* bicoid region. Mol. Biol. Evol. **19:** 989–998.

Bauke, A., 2002 Untersuchung zur Regulation von Chitinasen und Serin-Proteinase-Inhibitoren aus *Arabis drummondii*. Thesis, Friedrich-Schiller-Universitaet, Jena, Germany.

Bergelson, J., M. Kreitman, E. A. Stahl and D. C. Tian, 2001 Evolutionary dynamics of plant *R*-genes. Science **292:** 2281–2285.

Betancourt, A. J., and D. C. Presgraves, 2002 Linkage limits the power of natural selection in *Drosophila*. Proc. Natl. Acad. Sci. USA **99:** 13616–13620.

Bittner-Eddy, P. D., I. R. Crute, E. B. Holub and I. L. Beynon, 2000 RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in Peronospora parasitica. Plant J. **21** (2): 177–188.

Blanc, G., K. Hokamp and K. Wolfe, 2003 A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res. **13:** 137–144.

Broadway, R., 1995 Are insects resistant to plant proteinase inhibitors? J. Insect Physiol. **41:** 107–116.

Broadway, R., 1996 Dietary proteinase inhibitors alter complement of midgut proteases. Arch. Insect Biochem. Physiol. **32:** 39–53.

Broadway, R., 1997 Dietary regulation of serine proteinases that are resistant to serine proteinase inhibitors. J. Insect Physiol. **43:** 855–874.

Ceci, L., M. Volpicella, Y. Rahbe, R. Gallerani, J. Beekwilder *et al.*, 2003 Selection by phage display of a variant mustard trypsin inhibitor toxic against aphids. Plant J. **33:** 557–566.

Ceci, L. R., N. Spoto, M. Devirgilio and R. Gallerani, 1995 The gene coding for the mustard trypsin inhibitor-2 is discontinuous and wound-inducible. FEBS Lett. **364:** 179–181.

Ceciliani, F., F. Bortolotti, E. Menegatti, S. Rochi, P. Ascenzi *et al.*, 1994 Purification, inhibitory properties, amino acid sequence and identification of the reactive site of a new serine proteinase inhibitor from oil-rape (*Brassica napus*) seed. FEBS Lett. **342:** 221–224.

Charlesworth, D., 2003 Effects of inbreeding on the genetic diversity of populations. Philos. Trans. R. Soc. Lond. B Biol. Sci. **358:** 1051–1070.

Charlesworth, D., P. Awadalla, B. K. Mable and M. H. Schierup, 2000 Population-level studies of multiallelic self-incompatibility loci, with particular reference to Brassicaceae. Ann. Bot. **85:** 227–239.

Clauss, M., and T. Mitchell-Olds, 2003 Population genetics of tandem trypsin inhibitor genes in *Arabidopsis* species with contrasting ecology and life history. Mol. Ecol. **12:** 1287–1299.

Collins, N., H. Thordal-Christensen, V. Lipka, S. Bau, E. Kombrink *et al.*, 2003 SNARE-protein-mediated disease resistance at the plant cell wall. Nature **425:** 973–977.

Comeron, J. M., and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. Genetics **161:** 389–410.

De Leo, F., M. A. Bonade-Bottino, L. R. Ceci, R. Gallerani and L. Jouanin, 1998 Opposite effects on *Spodoptera littoralis* larvae of high expression level of a trypsin proteinase inhibitor in transgenic plants. Plant Physiol. **118:** 997–1004.

De Leo, F., L. Ceci, L. Jouanin and R. Gallerani, 2001a Analysis of mustard trypsin inhibitor-2 gene expression in response to developmental or environmental induction. Planta **212:** 710–717.

De Leo, F., M. Bonade-Bottino, L. R. Ceci, R. Gallerani and L. Jouanin, 2001b Effects of a mustard trypsin inhibitor expressed in different plants on three lepidopteran pests. Insect Biochem. Mol. Biol. **31:** 593–602.

Duda, T. F., and S. R. Palumbi, 2000 Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. Mol. Biol. Evol. **17:** 1286–1293.

Ellis, J., G. Lawrence, E. Finnegan and P. Anderson, 1995 Contrasting complexity of two rust resistance loci in flax. Proc. Natl. Acad. Sci. USA **92:** 4185–4188.

Ellis, J., P. Dodds and T. Pryor, 2000 Structure, function, and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. **3:** 279–284.

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

Fay, J. C., G. J. Wyckoff and C.-I Wu, 2002 Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature **415:** 1024–1026.

Ferrari, S., D. Vairo, F. M. Ausubel, F. Cervone and G. De Lorenzo, 2003 Tandemly duplicated Arabidopsis genes that encode polygalacturonase-inhibiting proteins are regulated coordinately by different signal transduction pathways in response to fungal infection. Plant Cell **15:** 93–106.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545.

Ford, M. J., 2002 Application of selective neutrality tests to molecular ecology. Mol. Ecol. **11:** 1245–1262.

Frohman, M. A., M. K. Dush and G. R. Martin, 1988 Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:** 8998–9002.

Glawe, G., J. Zavala, A. Kessler, N. Van Dam and I. Baldwin, 2003 Ecological costs and benefits correlated with trypsin protease inhibitor production in *Nicotiana attenuata*. Ecology **84:** 79–90.

Green, T. R., and C. A. Ryan, 1972 Wound-induced proteinase inhibitor in plant leaves—possible defense mechanism against insects. Science **175:** 776–777.

Gu, Z., A. Cavalcanti, F. Chen, P. Bouman and W. Li, 2002 Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. Mol. Biol. Evol. **19:** 256–262.

Hanfstingl, U., A. Berry, E. A. Kellog, J. T. Costa, W. Rudiger *et al.*, 1994 Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: Roles for both balancing and directional selection? Genetics **138:** 811–828.

Haruta, M., I. Major, M. Christopher, J. Patton and C. Constabel, 2001 A kunitz trypsin inhibitor gene family from trembling aspen (*Populus tremuloides* Michx.): cloning, functional expression, and induction by wounding and herbivory. Plant Mol. Biol. **46:** 347–359.

Haubold, B., J. Kroymann, A. Ratzka, T. Mitchell-Olds and T. Wiehe, 2002 Recombination and gene conversion in a 170-kb genomic region of *Arabidopsis thaliana*. Genetics **161:** 1269–1278.

Hauser, M. T., B. Harr and C. Schlotterer, 2001 Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene GLABROUS1. Mol. Biol. Evol. **18:** 1754–1763.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Holm, S., 1979 A simple sequentially rejective multiple test procedure. Scand. J. Stat. **6:** 65–70.

Hudson, R., 1990 Gene geneologies and the coalescence process. Oxf. Surv. Evol. Biol. **7:** 1–45.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

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

Jongsma, M. A., and C. Bolter, 1997 The adaptation of insects to plant protease inhibitors. J. Insect Physiol. **43:** 885–895.

Jongsma, M. A., P. L. Bakker, J. Peters, D. Bosch and W. J. Stiekema, 1995 Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. Proc. Natl. Acad. Sci. USA **92:** 8041–8045.

Kawabe, A., and N. Miyashita, 1999 DNA variation in the basic chitinase locus (ChiB) region of the wild plant *Arabidopsis thaliana*. Genetics **153:** 1445–1453.

Kawabe, A., H. Innan, R. Terauchi and N. T. Miyashita, 1997 Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Mol. Biol. Evol. **14:** 1303–1315.

Kim, Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155:** 1415–1427.

Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics **160:** 765–777.

Kim, Y., and W. Stephan, 2003 Selective sweeps in the presence of interference among partially linked loci. Genetics **164:** 389–398.

Kirby, D. A., and W. Stephan, 1996 Multi-locus selection and the structure of variation at the white gene of *Drosophila melanogaster*. Genetics **144:** 635–645.

Koch, M., B. Haubold and T. Mitchell-Olds, 2001 Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. Am. J. Bot. **88:** 534–544.

Koiwa, H., R. A. Bressan and P. M. Hasegawa, 1997 Regulation of protease inhibitors and plant defense. Trends Plant Sci. **2:** 379–384.

Kroymann, J., S. Donnerhacke, D. Schnabelrauch and T. Mitchell-Olds, 2003 Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc. Natl. Acad. Sci. USA **1073:** 1–10.

Kuittinen, H., and M. Aguadé, 2000 Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. Genetics **155:** 863–872.

Laskowski, M., Jr., and I. Kato, 1980 Protein inhibitors of proteinases. Annu. Rev. Biochem. **49:** 593–626.

Ling, M., H. Qi and C. Chi, 1993 Protein, cDNA, and genomic DNA sequences of the towel gourd trypsin inhibitor. A squash family inhibitor. J. Biol. Chem. **268:** 810–814.

Long, A. D., R. F. Lyman, C. H. Langley and T. F. C. Mackay, 1998 Two sites in the delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. Genetics **149:** 999–1017.

Lynch, M., and J. S. Conery, 2003 The evolutionary demography of duplicate genes. J. Struct. Funct. Genomics **3:** 35–44.

Lynch, M., and A. G. Force, 2000 The origin of interspecific genomic incompatibility via gene duplication. Am. Nat. **156:** 590–605.

Lynch, M., M. O'Hely, B. Walsh and A. Force, 2001 The probability of preservation of a newly arisen gene duplicate. Genetics **159:** 1789–1804.

Mauricio, R., E. A. Stahl, T. Korves, D. Tian, M. Kreitman *et al.*, 2003 Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. Genetics **163:** 735–746.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

Menegatti, E., G. Tedeschi, S. Ronchi, F. Bortolotti, P. Ascenzi *et al.*, 1992 Purification, inhibitory properties and amino acid sequence of a new serine proteinase inhibitor from white mustard (*Sinapis alba* L.) seed. FEBS Lett. **301:** 10–14.

Meyers, B., K. Shen, P. Rohani, B. Gaut and R. Michelmore, 1998 Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell **10:** 1833–1846.

Miralles, R., P. Gerrish, A. Moya and S. F. Elena, 1999 Clonal interference and the evolution of RNA viruses. Science **285:** 1745–1747.

Mondragon-Palomino, M., B. C. Meyers, R. W. Michelmore and B. S. Gaut, 2002 Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Res. **12:** 1305–1315.

Navarro, A., and N. H. Barton, 2002 The effects of multilocus balancing selection on neutral variability. Genetics **161:** 849–863.

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

Nordborg, M., and H. Innan, 2003 The genealogy of sequences containing multiple sites subject to strong selection in a subdivided population. Genetics **163:** 1201–1213.

Nordborg, M., J. Borevitz, J. Bergelson, C. Berry, J. Chory *et al.*, 2002 The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. **30:** 190–193.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer, Heidelberg, Germany.

Olsen, K. M., A. Womack, A. R. Garrett, J. I. Suddith and M. D. Purugganan, 2002 Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. Genetics **160:** 1641–1650.

Parsch, J., C. D. Meiklejohn and D. L. Hartl, 2001 Patterns of DNA sequence variation suggest the recent action of positive selection in the *janus-ocnus* region of *Drosophila simulans*. Genetics **159:** 647–657.

Pautot, V., F. Holzer and L. Walling, 1991 Differential expression of tomato proteinase inhibitor I and II genes during bacterial pathogen invasion and wounding. Mol. Plant-Microbe Interact. **4:** 284–292.

Polticelli, F., P. Ascenzi, M. Bolognesi and B. Honig, 1999 Structural determinants of trypsin affinity and specificity for cationic inhibitors. Protein Sci. **8:** 2621–2629.

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

Rask, L., E. Andréasson, B. Ekbom, S. Eriksson, B. Pontoppidan

et al., 2000 Myrosinase: gene family evolution and herbivore defense in Brassicaceae. Plant Mol. Biol. **42:** 93–113.

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

Ruoppolo, M., A. Amoresano, P. Pucci, S. Pascarella, F. Polticelli *et al.*, 2000 Characterization of five new low-molecular-mass trypsin inhibitors from white mustard (*Sinapis alba* L.) seed. Eur. J. Biochem. **267:** 6486–6492.

Ryan, C. A., 1990 Protease inhibitors in plants: genes for improving defenses against insects and pathogens. Annu. Rev. Phytopathol. **28:** 425–449.

Savolainen, O., C. H. Langley, B. P. Lazzaro and H. Freville, 2000 Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. Mol. Biol. Evol. **17:** 645–655.

Shelton, A. M., R. J. Cooley, M. K. Kroening, W. T. Wilsey and S. D. Eigenbrode, 1991 Comparative analysis of two rearing procedures for diamond-back moth (Lepidoptera: Plutellidae). J. Entomol. Sci. **26:** 17–26.

Shepard, K., and M. D. Purugganan, 2003 Molecular population genetics of the *Arabidopsis* CLAVATA2 region. The genomic scale of variation and selection in a selfing species. Genetics **163:** 1083–1095.

Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman and J. Bergelson, 1999 Dynamics of disease resistance polymorphism at the *RPM1* locus of *Arabidopsis*. Nature **400:** 667–671.

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

Telang, M., A. Srinivasan, A. Patankar, A. Harsulkar, V. Joshi *et al.*, 2003 Bitter gourd proteinase inhibitors: potential growth inhibitors of *Helicoverpa armigera* and *Spodoptera litura*. Phytochemistry **63:** 643–652.

Templeton, A. R., H. Hollocher and J. S. Johnston, 1993 The molecular through ecological genetics of abnormal abdomen in *Drosophila mercatorum*. V. Female phenotypic expression on natural genetic backgrounds and in natural environments. Genetics **134:** 475–485.

Tian, D., H. Araki, E. Stahl, J. Bergelson and M. Kreitman, 2002 Signature of balancing selection in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **99:** 11525–11530.

Tian, D. C., M. B. Traw, J. Q. Chen, M. Kreitman and J. Bergelson, 2003 Pleiotropic cost of *R*-gene mediated resistance in *Arabidopsis thaliana*. Nature **423:** 74–77.

Tiffin, P., and B. Gaut, 2001 Molecular evolution of the wound-induced serine protease inhibitor *wip1* in *Zea* and related genera. Genetics **18:** 2092–2101.

Van der Hoorn, R., P. De Wit and M. Joosten, 2002 Balancing selection favors guarding resistance proteins. Trends Plant Sci. **7:** 67–71.

Van Valen, L., 1973 A new evolutionary law. Evol. Theory **1:** 1–30.

Volpicella, M., A. Schipper, M. Jongsma, N. Spoto, R. Gallerani *et al.*, 2000 Characterization of recombinant mustard trypsin inhibitor 2 (MTI2) expressed in *Pichia pastoris*. FEBS Lett. **468:** 137–141.

Volpicella, M., L. Ceci, J. Cordewener, T. America, R. Gallerani *et al.*, 2003 Properties of purified gut trypsin from *Helicoverpa zea*, adapted to proteinase inhibitors. Eur. J. Biochem. **270:** 10–19.

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

Walsh, B., 2003 Population-genetic models of the fates of duplicate genes. Genetica **118:** 279–294.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

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

Zang, X., and R. Maizels, 2001 Serine proteinase inhibitors from nematodes and the arms race between host and pathogen. Trends Biochem. Sci. **26:** 191–197.

Zhang, J., Y. Zhang and H. Rosenberg, 2002a Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat. Genet. **30:** 411–415.

Zhang, J., A. Peek, D. Dunams and B. Gaut, 2002b Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). Genetics **162:** 851–860.

Zhao, Q., Y. Chae and J. Markley, 2002 NMR solution structure of ATTp, an *Arabidopsis thaliana* trypsin inhibitor. Biochemistry **15:** 12284–12296.

Communicating editor: J. Bergelson