MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 2844–2855 doi: 10.1111/j.1365-294X.2009.04206.x

Functional and geographical differentiation of candidate balanced polymorphisms in *Arabidopsis thaliana*

JENNIFER M. REININGA*, DAHLIA NIELSEN* and MICHAEL D. PURUGGANAN†

*Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC 27695, USA, †Department of Biology and Center for Genomics and Systems Biology, New York University, 1009 Silver, 100 Washington Square East, New York, NY 10003, USA

Abstract

Molecular population genetic analysis of three chromosomal regions in *Arabidopsis thaliana* suggested that balancing selection might operate to maintain variation at three novel candidate adaptive trait genes, including *SOLUBLE STARCH SYNTHASE I (SSI)*, *PLASTID TRANSCRIPTIONALLY ACTIVE 7(PTAC7)*, and *BELL-LIKE HOMEODOMAIN 10 (BLH10)*. If balanced polymorphisms are indeed maintained at these loci, then we would expect to observe functional variation underlying the previously detected signatures of selection. We observe multiple replacement polymorphisms within and in the 32 amino acids just upstream of the protein–protein interacting BELL domain at the *BLH10* locus. While no clear protein sequence differences are found between allele types in SSI and PTAC7, these two genes show evidence for allele-specific variation in expression levels. Geographical patterns of allelic differentiation seem consistent with population stratification in this species and a significant longitudinal cline was observed at all three candidate loci. These data support a hypothesis of balancing selection at all three candidate loci and provide a basis for more detailed functional work by identifying possible functional differences that might be selectively maintained.

Keywords: adaptation, balancing selection, population genetics

Received 26 November 2008; revision received 9 February 2009; accepted 11 February 2009

Introduction

Balancing selection maintains genetic variation within populations or species through mechanisms such as overdominance (Tishkoff *et al*. 2001; Aidoo *et al*. 2002), frequency-dependent selection (Charlesworth & Awadalla 1998), spatial-temporal selection (Tian *et al*. 2002; Charbonnel & Pemberton 2005), local adaptation to different environments (Schulte *et al*. 2000; Harr *et al*. 2002; Storz *et al*. 2004, 2007) and epistatic selection (Kroymann & Mitchell-Olds 2005). This expansive definition of balancing selection results from the recognition that these various modes of selection result in similar molecular signatures at the genomic level, including elevated levels of nucleotide diversity that decrease symmetrically with distance from the selected site(s). The intensity and breadth of this signal are highly dependent on the local rate of recombination, time since the selective event, and strength of selection (Charlesworth 2006).

Examples of balanced polymorphisms are best supported if one can establish the selective mechanism, provide molecular evidence for a signal of selection, and/or demonstrate functional or phenotypic effects of alternate putatively selected alleles. Expression phenotypes are often investigated for functional consequences of naturally occurring genetic variation in gene regulation that may contribute to adaptive differences both within and between species (Crawford *et al*. 1999; Schulte *et al*. 2000; Michalak *et al*. 2001; Bamshad *et al*. 2002; Rockman & Wray 2002; Lerman *et al*. 2003; Tian *et al*. 2003). For example, studies have identified molecular signatures of balancing selection associated with differential expression phenotypes among allelic variants of the baboon homolog of human class II MHC locus *DQA1* (Loisel *et al*. 2006), *Arabidopsis lyrata* self-incompatibility (SI) *S*-alleles (Prigoda *et al*. 2005) and among haplotypes of the *PDYN* regulatory region in humans (Rockman *et al*. 2005). Alternatively, amino acid replacement polymorphisms that result in protein sequence changes

Correspondence: Michael Purugganan, Fax: 1-(212)-995-4015; E-mail: mp132@nyu.edu

can be maintained by balancing selection when alternate protein forms are favoured at different points in space or time. (Storz *et al*. 2007; Schmidt *et al*. 2008).

The geographical distribution of alleles within and between populations can also provide further support for the hypotheses of balancing selection and in identifying mechanisms that might be responsible for the maintenance of diversity. For example, adaptive variation has been associated with latitudinal clines in *Fundulus heteroclitus* (Schulte *et al*. 2000), *Arabidopsis thaliana* (Caicedo *et al*. 2004), and *Drosophila melanogaster* (Schmidt *et al*. 2008). Alternatively, when frequency-dependent selection is the driving evolutionary force, multiple alleles are often observed co-segregating within local populations across a broad species range. Plant self-incompatibility loci are one of the best-documented examples of frequency-dependent selection. SI alleles can persist for long evolutionary timescales and occur at intermediate frequencies within populations of a species due to the increased fitness effects of an allele when it becomes rare (Charlesworth *et al.* 1998).

Strong cases for balancing selection have been made for several genes in *A. thaliana*. Some of the best examples come from genes with roles in disease resistance (Stahl *et al*. 1999; Tian *et al*. 2003) and protection against herbivory (Kroymann *et al*. 2003). In both cases, presence/absence of polymorphisms are thought to be maintained due to fitness trade-offs either due to fluctuating selection pressure across space and time, depending on patterns of pathogen occurrence (Tian *et al*. 2002, 2003), or due to differences in allelic performance against specialist vs. generalist herbivores (Kroymann *et al*. 2003). While spatial-temporal selection (Caicedo *et al*. 2004; Samis *et al*. 2008) and epistasis (Kroymann & Mitchell-Olds 2005) are well-supported mechanisms maintaining genetic and phenotypic diversity in *A. thaliana*, overdominance seems less likely, due to the fact that the species is primarily self-fertilizing.

Arabidopsis thaliana is a primarily self-fertilizing species, which results in a low effective rate of recombination (Abbot & Gomes 1989) and extensive linkage disequilibrium (Nordborg *et al*. 2005), thus providing a especially useful system for the identification of a broad set of balanced polymorphisms (Tian *et al*. 2002). Genome scanning has indeed identified the molecular signature of balancing selection at several candidate loci in this species (Cork & Purugganan 2005), including (i) increased levels of nucleotide diversity, and (ii) intermediate frequencies of divergent alleles in three genomic regions. Molecular population genetic analysis of short gene fragments spanning each of these candidate gene regions revealed one putative disease resistance locus and three novel genes as potential targets of balancing selection. These novel candidate genes include *SOLUBLE STARCH SYNTAHSE I* (*SSI*) (At5g24300) and *PLASTID TRANSCRIPTIONALLY ACTIVE CHROMOSOME 7 (PTAC7)* (also known as *PIGMENT DEFECTIVE EMBYRO 225* or *PDE225*) (At5g24314), which are linked in a common high-diversity region and the *BELL-LIKE HOMEODOMAIN 10* (*BLH10*) (At1g19700).

On the basis of homology, *SSI* was previously identified as a member of the starch synthetic pathway and is highly conserved throughout the plant kingdom (Ral *et al*. 2004). Aberrant chain length distributions in amylopectin from plants lacking functional *SSI* indicates that this gene is a determinant of branching structure of amylopectin starch molecules (Delvalle *et al*. 2005). The protein product of the second candidate gene, *PTAC7*, was purified from the *A. thaliana* chloroplasts in association with the nuclear-encoded single-subunit RNA polymerase (NEP), which suggests a function in the regulation of chloroplast gene expression (Pfalz *et al*. 2006). Consistent with this expectation, a pigment-defective knockout phenotype for this locus has been observed (Budziszewski *et al*. 2001). The final candidate-balanced polymorphism, *BLH10*, is a member of the BELL-like homeodomain family of transcription factors, which have conserved DNA-binding homeodomain and protein-interacting BELL domains (Becker *et al*. 2002). Members of this gene family have been shown to play important roles in various aspects of plant development, including ovule development (Ray *et al*. 1994), branching (Smith & Hake 2003) and internode patterning (Bao *et al*. 2004). Phylogenetic analysis indicates that *BLH10* is most closely related to an *Arabidopsis*-specific paralog, *BLH3* (Becker *et al*. 2002). Knockout mutant alleles of *BHLH10* and *BHLH3* have not yet been associated with any specific phenotypes, although these genes do show differences in expression patterns and no evidence for a relaxation of constraint (Duarte *et al*. 2006) suggesting the potential for nonredundant functions.

Here we expand our previous study (Cork & Purugganan 2005) by providing a more detailed analysis of these three candidate-balanced polymorphisms to investigate the possibility that functional genetic variation is maintained at these loci. We examine levels and patterns of nucleotide variation across each full candidate gene sequence and explore the geographical distribution of alleles at each locus. Protein-level variation and differences in levels and patterns of allele-specific gene expression are also investigated. Our findings further support a hypothesis of balancing selection at these candidate loci by demonstrating the potential for functional differences that may underlie the selective maintenance of allelic variants of these genes.

Materials and methods

Isolation and sequencing of alleles

Genomic DNA was extracted from young leaves of *Arabidopsis thaliana* accessions spanning the natural historic geographical range of the species (Table S1, Supporting information), using the plant DNeasy mini kit (QIAGEN). Polymerase chain reaction (PCR) and sequencing primers were designed using Primer3 software (http://fokker.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) with either the Col-0 genome sequence or conserved regions of existing partial gene alignments as a template (Text S1, Supporting information). PCR was performed using either *Taq* (Roche) or Ex*Taq* (TaKaRa) and PCR products were cleaned using QIAGEN PCR purification or gel extraction kits. PCR products were sequenced directly using BigDye terminators and were run on ABI 3700 96 capillary automated sequencers (Applied Biosystems) at North Carolina State University's Genome Research Laboratory. PHRED and PHRAP functions (Ewing *et al*. 1998; Ewing & Green 1998) of BioLign (Tom Hall, North Carolina State University) were used for base-calling and sequence contig formation. All polymorphisms were visually inspected and alignments were edited by hand.

Population genetic analysis

Parsimony trees used for haplotype network construction were generated using PAUP (Sinauer Associates). DnaSP version 4.10.7 (Rozas *et al*. 2003) was implemented to calculate summary statistics, measures of diversity and to conduct tests of selection. Nucleotide diversity was estimated using both total and silent sites as π (Tajima 1983) and as θ_W (Watterson 1975) for total sites only. Tajima's *D* (Tajima 1993), haplotype number, haplotype diversity, and the intragenic linkage disequilibrium coefficient, Z_{nS} (Kelly 1997), were also calculated. Significance of Tajima's *D*, haplotype number, haplotype diversity, and Z_{nS} was determined by coalescent simulation under the conservative assumption of no recombination, using 10 000 runs and conditioning on the number of segregating sites. Sliding window analysis of nucleotide diversity was performed for each gene and the proportion of nonsynonymous to synonymous nucleotide diversity was similarly investigated for *BLH10*. A window size of 100 bp and a step size of 20 bp were used for *BLH10*, while a window of 75 bp and a step size of 15 bp were used for *SSI* and *PTAC7/PDE225*. PolyPhen (http://genetics.bwh.harvard.edu/pph/) was used to assess the potential effects of replacement polymorphisms.

Allele-specific expression

For each candidate-balanced polymorphism, three heterozygous F_1 lines were generated by crossing three different B allele (Ler-0 like) accessions to Col-0 A allele plants. In the case of the tri-allelic *PTAC7*, two sets of three crosses were included to represent both A/B and the A/C allele comparisons (see Table S3, Supporting information). F_1 progeny were grown under long day (16-h day length, 20 °C) conditions in 16-cell flats with sub-irrigation following a 3-week vernalization period at 4 °C. For each cross, three individuals were sampled for leaf, mixed-stage bud, and mixed-stage silique tissues by flash-freezing in liquid nitrogen immediately after harvesting. Tissue samples were collected 21 days after planting or at the first point in development thereafter when all required tissues were available. mRNA extractions were performed for each of the three tissue samples for each individual using the QIAGEN Plant RNA Extraction kit. Thirty microlitres of each RNA preparation (ranging in concentration between 500 and 800 ng/μL) was DNase-treated using the DNA free kit according to the manufacturer's suggestions for rigorous treatment conditions (Ambion). Maximum amounts of RNA (10 μL) from each DNase treated sample were used in each of two replicate cDNA reactions using Ambion's Retroscript kit according to manufacturer specifications. Replicate cDNA samples were pooled and 1 μL of the combined cDNA preparation was used in each PCR with 7.5 pM of each, Ex*Taq* polymerase, and Ex*Taq* PCR reagents (TaKaRa). To test for the possibility of DNA contamination, several random DNase treated RNA samples were similarly subjected to PCR amplification. No product was observed in any of the tested samples, indicating the absence of DNA contamination in RNA samples subject to analysis. Four PCR and pyrosequencing reactions were included as technical replicates to assay the allele-differentiating single nucleotide polymorphisms (SNPs) of interest. Text S2, Supporting information lists the PCR and sequencing primers used in the analysis.

The proportion of alternate SNP nucleotides present in the sample was analysed using a PSQ96 MA Pyrosequencer and Pyro Gold chemistry (Biotage). SNP proportions were standardized by a within-sample monomorphic peak and then by an identical measure obtained from DNA. Since both alleles are present at 1:1 proportions in the DNA, any resulting differences in the observed quantity of alleles will reflect a bias in amplification during PCR amplification. By standardizing our cDNA measurements by a common DNA measure, we are assured that any observed differences in the proportion of transcripts are the result of a difference in expression only. Standardized measures of allele-specific expression were log-transformed to normalize the distri butions. Tests for allele-specific expression were conducted using analysis of variance (ANOVA), with the model $Y = \mu + T + L + A + T*L + L*A + A*T + A*L*T + I[L]random + A*I[L]random + T*I[L]random + A*T*I[L]random + error$, where Y is the log-transformed standardized allele expression level, T is the tissue type, L is the line or cross, A is the SNP allele, and I is the sampled individual.

Spatial distribution of alleles

Approximately 250 *A. thaliana* accessions were genotyped at allele-differentiating SNPs for each of the three candidate

Table 1 Measures of diversity for each candidate gene

Gene	*n*†	Length (bp)‡	S§	S (silent)¶	π_s¶	π††	θ_w††	Z_{nS}	Haplotype diversity	Tajima's *D*
BLH10	14	1856	62	53	0.042	0.016	0.011	0.82***	0.83*	+2.4083**
Copia-like transposon	9	684	35	35	0.017	0.017	0.019	nd	0.89	+0.4962
PTAC7/PDE225	15	1332	114	112	0.049	0.036	0.026	0.42	0.97	+1.6057*
SSI	16	3803	101	96	0.018	0.011	0.008	0.87***	0.97	+1.7100*

nd, not determined. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; † number of samples; ‡ length of sequence region excluding gaps; § number of segregating sites; ¶ estimates are based on silent sites; †† estimates are based on total sites.

genes (see Table S2, Supporting information). *A. thaliana* accessions from Siberia and Central Asia were provided by K. Schmidt (Leibniz-Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany). DNA was isolated from young plant leaves using the plant DNeasy mini kit (QIAGEN) and genotyped using cleaved amplified polymorphic sequences (CAPS) markers. CAPS markers were chosen by comparison of allele sequences in NEBCutter version 2.0 (http://tools.neb.com/NEBcutter2/index.php). Genotyping primers were designed using Primer3 software and gene products were subsequently digested with appropriate restriction enzymes (New England Biolabs). Allele-specific digest patterns were resolved by agarose gel electrophoresis. The genotyping primers and restriction enzymes are listed in Text S3, Supporting information.

Tests for geographical patterning of alternate alleles were performed by regressing allele identity on the latitude and longitude of each accession collection site origin. Population structure was controlled by including the most likely ancestry coefficients obtained from the Bayesian program Structure (Pritchard *et al.* 2000; Falush *et al.* 2003) ($K = 10$) (Ehrenreich *et al.*, submitted) as covariates in the analysis. Structure ancestry coefficients were log-contrasts transformed as previously described (Samis *et al.* 2008). The inverse distance weighted function of the Geographic Information System (GIS program, ArcMAP) was applied to allelic data according to the neighbourhood method using 10–15 points to generate maps depicting generalized predictions of allele frequencies across Europe. The mean value was used if more than two points occur at the same map position.

Results

Diversity across candidate genes

Elevated levels of nucleotide diversity, highly positive Tajima's *D*-values and Bonferroni-significant HKA (Hudson–Kreitman–Aguadé) tests were previously obtained for short (~1 kb) fragments of each of the three candidate genes, *BLH10*, *SSI* and *PTAC7* (see Cork & Purugganan 2005 for comparisons to empirical genome-wide distributions). To further investigate levels and patterns of nucleotide variation across each of the three putatively selected loci, full gene sequence data was collected from ~15 *Arabidopsis thaliana* accessions (see Table S2, Supporting information). Table 1 summarizes data on measures of nucleotide diversity and tests of selection based on the site frequency spectrum. In general, levels of silent site nucleotide diversity (π_S) are in accord with previous findings (Cork & Purugganan 2005) and are consistent with expectations for a balanced polymorphism: (i) increased polymorphism relative to the *A. thaliana* genome average (π~0.01), and (ii) a trend for variants to segregate at intermediate frequency (detected as a significantly positive Tajima's *D* value) were observed at all candidate loci.

The standardized intragenic linkage disequilibrium (LD) averaged across all pairwise comparisons, Z_{nS}, is expected to be significantly higher at loci subject to balancing selection (Kelly 1997). We observe significant elevations at both of the bi-allelic loci (*SSI* and *BLH10*), while the lack of significance at the trimorphic *PTAC7* likely reflects decreased power due to the small sample size and extreme variability of the locus. With the exception of Z_{nS}, deviations from neutrality (or from the genome average) are generally more extreme than reported in our previous analyses (Cork & Purugganan 2005), due in part to our current analysis being based on full-gene sequence data instead of only ~1 kb gene fragments.

Consistent with previous findings and with observations of high intragenic linkage disequilibrium, the intermediate frequency variants at *SSI* and *BLH10* form two highly distinct allele classes, herein referred to as alleles A and B, where A refers to the Col-0 type and B to the Ler-0 type allele. A trimorphic haplotype pattern was observed at the *PTAC7* locus and allele categories for this gene are as previously described with a third C-type allele category (see Fig. 1).

Interestingly, an ~810-bp polymorphic insertion/deletion (indel) in the *BLH10* gene, representing an insertion in

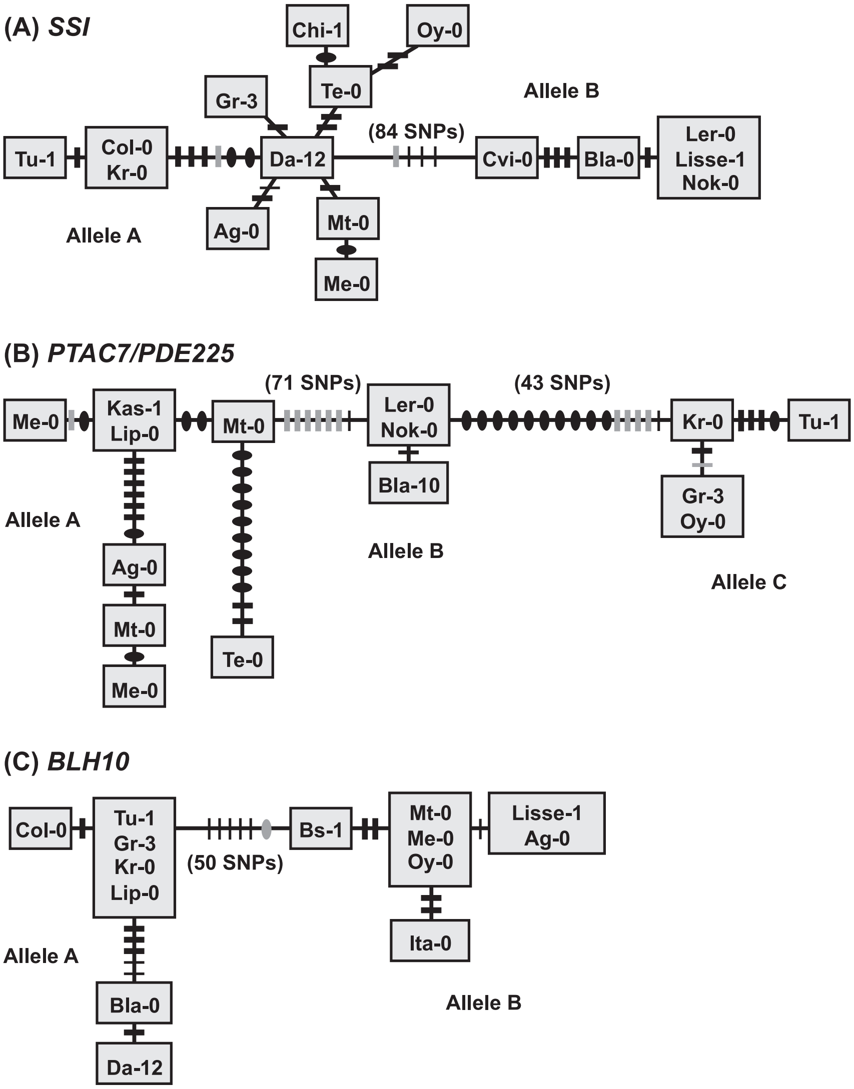

Fig. 1 Haplotype networks showing the relationship between allele classes among sequenced *A. thaliana* accessions at *SSI, PTAC7* and *BLH10*. Thick black bars mark noncoding single nucleotide polymorphisms (SNPs), thin bars for nonsynonymous changes, grey bars for tri-allelic SNPs, grey oval for nonsynonymous indels, and black ovals for homologise. Longer branches are marked by a solid line and all SNPs that are not silent or noncoding are indicated. The number of total allele-differentiating SNPs that contribute to these branches are also indicated.

several *A. thaliana* accessions relative to the Col-0 genome sequence, was identified in the intergenic region between BLH10 and its flanking gene, At1g19690. The indel was found to be associated with the *BLH10* B allele in all but two of the sampled accessions (Bs-1 from Switzerland and Ita-0 from Morocco). Sequence analysis reveals that this insertion shares 95% identity to a *copia*-like retrotransposon on chromosome 3 of the Col-0 reference sequence.

Variation in nucleotide polymorphism levels across gene regions

A sliding window analysis of nucleotide diversity was conducted to determine the distribution of nucleotide variation across the high-diversity region containing *SSI* and *PTAC7* (see Fig. 2). Patterns in fluctuating levels of nucleotide diversity across these genes do not differ notably when only silent sites are considered. While both genes display several characteristics suggestive of balancing selection, it is most reasonable to hypothesize that only one of these genes is the true target of selection in this genomic region. Indeed, significant intergenic linkage disequilibrium was previously observed between these two loci and suggests that they share correlated evolutionary histories, a finding consistent with expectations of hitchhiking of neutral mutations with selected variants (Cork & Purugganan 2005). Prior analysis revealed decreased levels of variation and a nonsignificant HKA test result at the locus At5g24310, which is located between these two candidate genes, and it was therefore excluded from the present analysis.

Nucleotide variation across *SSI* appears to be lowest at the 5′ end and increases to an approximately threefold higher level towards the 3′ end of the gene. Levels of nucleotide diversity are much higher at the downstream *PTAC7* locus, which has a silent site diversity level greater than fourfold higher than the genome-wide average (see Table 1). Two

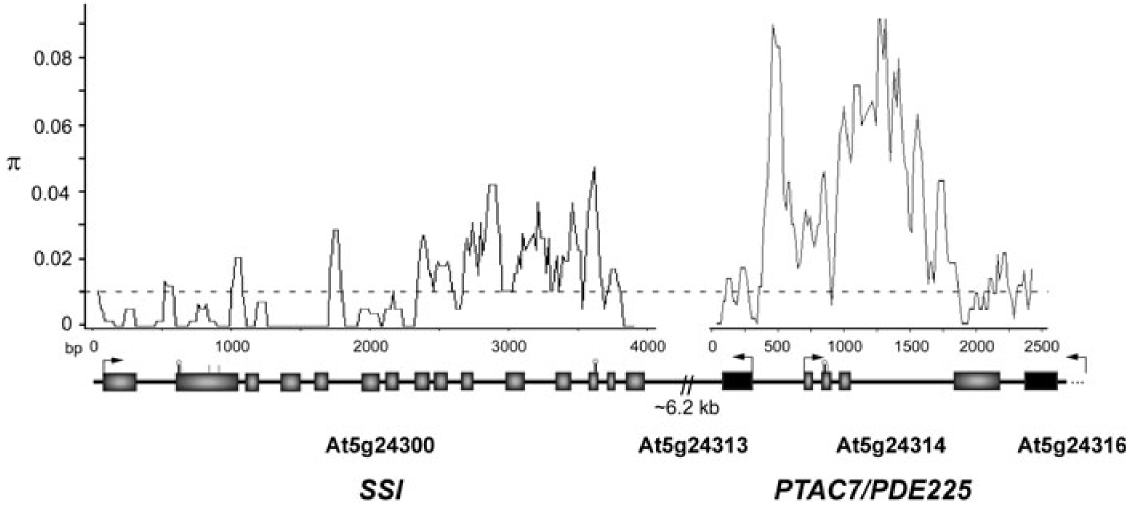

Fig. 2 Sliding window analysis of total nucleotide diversity (π) across *SSI* and *PTAC7* (*PDE225*). Focal genes are diagrammed in dark grey, while flanking genes are in black. The dashed line represents average genome-wide nucleotide diversity in *Arabidopsis thaliana*. All observed nonsynonymous variants at candidate loci are indicated as flags on the gene diagram as follows: grey flags represent singletons and knobbed black flags indicate allele-differentiating variants. Black flags indicate replacement polymorphisms that occur with a frequency > 10% that do not differentiate allele classes.

peaks of diversity were localized to the upstream region and within the third intron of the *PTAC7* locus. Together, these data implicate *PTAC7* as a more likely target of selection in this genomic region (see Fig. 2) and opens the possibility that selection might operate on *cis*-acting regulatory elements that could be located in the upstream and/or intronic regions of this gene.

The variability in levels of nucleotide diversity was also assessed across a ~4 kb region that includes the *BLH10* gene, the 3′ intergenic sequence including the identified retrotransposon insertion and the flanking gene, At1g19690 (see Fig. 3). Peaks of nucleotide diversity are localized to the *BLH10* locus within which two peaks of nucleotide polymorphism levels are observed, one located over the portion of the gene encoding the BELL domain and another centred in the second intron. As gaps are ignored in our sliding window analysis, nucleotide diversity is depicted as being zero across the section of the intergenic region corresponding to the large indel polymorphism representing the *copia*-like retroransposon. This insertion contains 35 segregating sites, and estimates of nucleotide diversity in this indel ($\pi = 0.017$) are slightly higher than the genome average (see Table 1) and suggest that this indel, although present at only intermediate frequencies in *A. thaliana*, does not represent a recent insertion.

Nonsynonymous variation at candidate genes

We examined nonsynonymous variation across each candidate gene to determine the potential for selective maintenance of differentially functioning proteins. Only

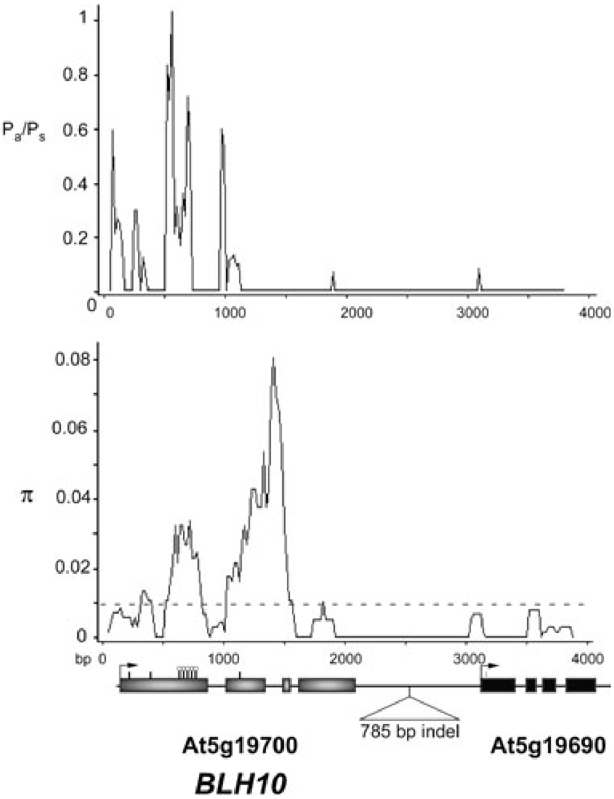

Fig. 3 Sliding window analysis of total nucleotide diversity (π) and the proportion of replacement to synonymous nucleotide diversity, Pa/Ps, across the *BLH10* gene region. The dashed line represents average genome-wide nucleotide diversity in *Arabidopsis thaliana*. Replacement polymorphisms are as described for Fig. 1.

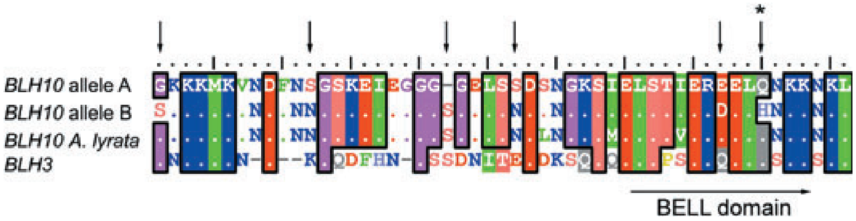

Fig. 4 Alignment of the region of the BLH10 protein showing the clustered replacement polymorphisms (arrows) observed within and just upstream of the BELL protein-interacting domain. Both A and B *BLH10* alleles are indicated and are aligned to the homologous *Arabidopsis lyrata BLH10* sequence and the *A. thaliana BLH3* paralog. The radical amino acid substitution predicted by PolyPhen is indicated by an asterisk.

two A/B allele-differentiating nonsynonymous mutations were observed in *SSI*: a Gln (Q) to Glu (E) change in exon 2, and a Thr (T) to Ala (A) change in exon 13 (see Fig. 1). To see if these variants represent potentially functional amino acid substitutions, we used the online prediction program, PolyPhen, which uses structural data and/or an alignment-based methodology to determine if a substitution is likely to affect protein structure and function. Both substitutions that differentiate the A and B alleles of *SSI* are predicted to have no functional effects and are therefore unlikely to produce differential phenotypes that might be selectively maintained. The single allele differentiating replacement polymorphism observed in *PTAC7*, a Gly (G) to Arg (R) polymorphism in exon 2 that differentiates allele B from alleles A and C, is also predicted to have no functional effects.

At the *BLH10* locus, however, there is a peak of non synonymous/synonymous nucleotiode diversity centred at exon 1. We observed five allele-differentiating nonsynonymous substitutions and one 3 bp in-frame indel clustered within and 32 amino acids upstream of the protein interacting BELL domain (see Fig. 4). Interestingly, one of these changes, a glutamine (Q) to histidine (H) substitution occurring within the conserved BELL domain is conserved among many (but not all) BELL-like homeodomain gene family members, and PolyPhen indicates that this may be a functional polymorphism. This suggests that alternate alleles of *BLH10* might possess variable functions that are dependent, in part, on protein structure/function differences.

Allele-specific expression of candidate genes

Differences in levels or patterns of expression among alternate alleles could be maintained by selection acting on divergent *cis*-regulatory elements; we thus tested for allele-specific expression of each of the three candidate genes. For each gene, three B allele-containing accessions (as well as three C allele accessions for the trimorphic *PTAC7* locus) were crossed to Col-0 (which had the A type allele) to provide three different heterozygous F_1 progeny backgrounds. By examining the relative expression levels of alternate alleles within multiple genetic backgrounds, we should be able to disentangle expression differences that result from *cis*- and *trans*-acting effects, since only *cis*-acting changes should yield reproducible differences in allelic abundance across F_1 lines. The proportion of transcripts attributed to each specific allele of a candidate locus in leaf, mixed-stage floral bud and silique tissues were assayed through detection of allele-differentiating SNPs via pyrosequencing.

No tissue-specific differences in allelic expression were observed, consistent with the ubiquitous expression patterns of these genes as reported by MPSS analysis (Meyers *et al.* 2004). Significant differences in transcript levels from each allele were detected, however, in two of the three candidate loci: *SSI* and *PTAC7*. The *SSI* locus shows a significant allele effect, with the Col-0 type A allele being expressed at higher levels throughout the plant than the Ler-0 type B allele ($P < 0.001$, ANOVA; see Table S3). Both investigated allelic combinations (A vs. B and A vs. C) at the *PTAC7* locus also displayed significant allele-specific differences in expression, with the A vs. B contrast revealing a significant line–allele interaction ($P < 0.017$, ANOVA; see Table S3) and the A vs. C comparison demonstrating an overall allele effect ($P < 0.019$, ANOVA; see Table S3). The significant line–allele interaction for the A vs. B allelic comparison at *PTAC7* indicates that alternate alleles may behave differently in different genetic backgrounds, suggesting epistatic interaction of alternate *cis*- alleles with *trans*-acting factors. For both genes, however, there is observed variability in allelic abundance that can be accounted for solely by *cis*-acting polymorphism at the candidate loci, indicating regulatory differences between alternate alleles that could be selectively maintained.

The geographical distribution of alleles

Geographical patterns in the distribution of alleles at selected loci can provide insights into the evolutionary forces that act on genes. In *A. thaliana*, the distribution of alleles within and between populations is confounded by human disturbance and its homogenizing effects on variation among populations, particularly in agricultural

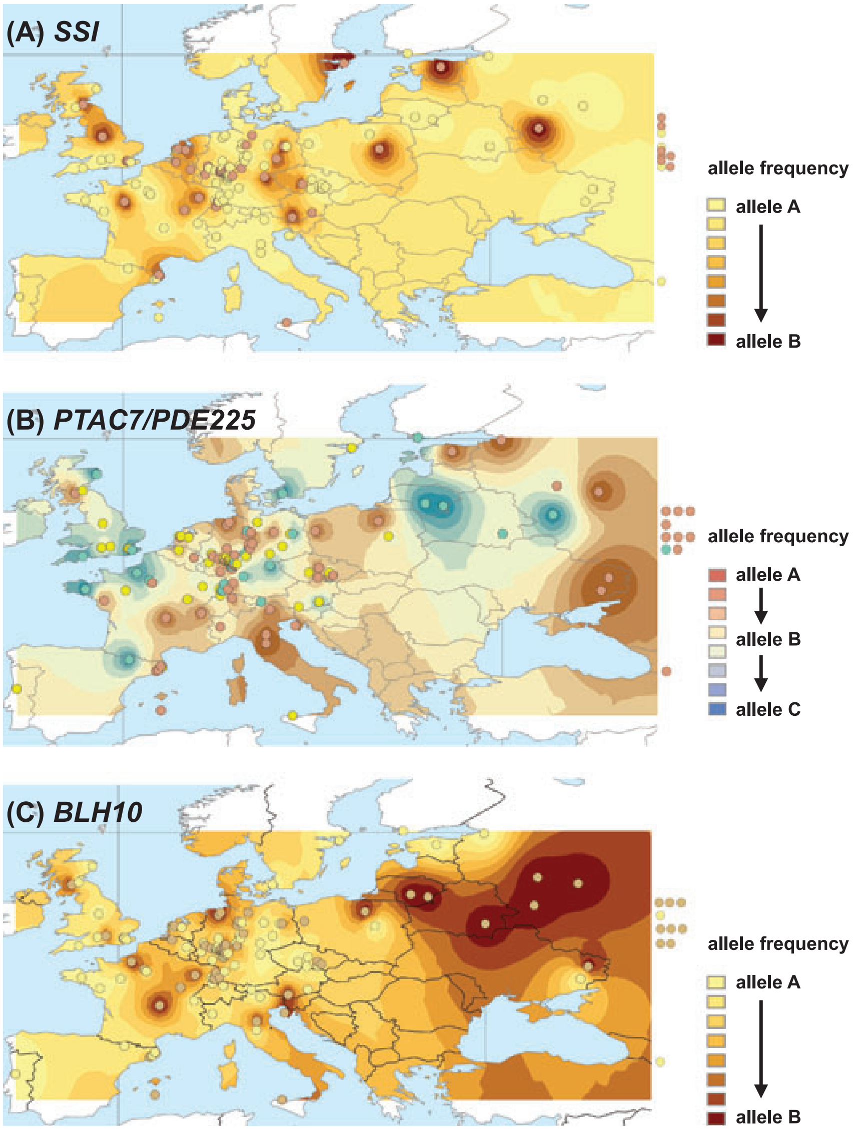

Fig. 5 Inverse distance weighted map of observed and predicted allele frequencies for each candidate locus across *Arabidopsis thaliana's* natural historic geographical range. Data points to the right of the map represent allele identities for ~30 accessions from eastern Siberia and Central Asia. Where geographical coordinates are monomorphic, only one data point describing the identity of accessions collected from that general locale is depicted.

regions of central Western Europe (Mitchell-Olds & Schmitt 2006). Despite the lack of historically distinct populations across most of the geographical range of this species, several studies have demonstrated extensive population structure and suggest that the present-day species range was populated through expansions from Asian and Mediterranean Pleistocene refugia (Sharbel *et al*. 2000; Nordborg *et al*. 2005; Schmid *et al*. 2005; Ostrowski *et al*. 2006).

We analysed the geographical distribution of alleles at *SSI*, *PTAC7* and *BLH10* by genotyping a set of ~250 *A. thaliana* accessions distributed largely across Eurasia. Logistic regression was used to test for allelic latitudinal and longitudinal clines while controlling for population structure, the latter inferred by Bayesian analyses (Falush *et al*. 2003; Pritchard *et al*. 2000; Ehrenreich *et al*., submitted). Evidence for population structure was found at all three candidate genes. While our geographical model did significantly explain allele identity among samples (*BLH10* $P < 0.0001$; *SSI* $P < 0.04$; *PTAC7* $P < 0.0001$), R^2 values are quite low (*BLH10* $R^2 = 0.167$; *SSI* $R^2 = 0.119$; *PTAC7* $R^2 = 0.129$), indicating that other unexplored factors are necessary to more adequately explain variation in allele identity among the investigated accessions. Interestingly, we observed a significant longitudinal effect on allele identity at the *BLH10* locus ($P < 0.02$) that is not accounted for by population structure. This finding may reflect historical adaptations to/in alternate refugial environments or local adaptation to longitudinally variable factors, as recently suggested for *PHYC* (Samis *et al*. 2008). Stepwise multiple logistic regression of allele type on average summer and winter temperature and precipitation shows no correlation of candidate gene alleles with these climatic variables (unpublished observations).

A GIS analysis allows us to map the distribution of alleles across the *A. thaliana* geographical range in Eurasia (see Fig. 5). In general, a longitudinal cline is evident across this geographical range, a pattern consistent with previous reports of population structure and a general trend for loss of allelic richness in the eastern portion of the geographical range for this species (Sharbel *et al*. 2000; Nordborg *et al*. 2005; Schmid *et al*. 2005). Commensurate with our statistical analysis, this pattern is especially marked for the *BLH10* locus, while less apparent for the linked *SSI* and *PTAC7*. Alternate alleles of *SSI* and *PTAC7* show an admixed pattern that broadly spans the investigated geographical range; both *SSI* alleles and two of the three *PTAC7* alleles, are observed across the full east–west sample distribution, and ~50% of Siberian samples collected from nearby locales contained multiple alleles of both genes even with the small sample sizes (2–5) investigated here. Since Siberian/Central Asian accessions tend to show an extreme loss of allelic richness relative to accessions across the rest of Eurasia (Schmid *et al*. 2006), it is intriguing to consider the possibility that some mechanism of selection that acts to preserve variation at a more local level is responsible for the maintenance of genetic variation in the *SSI*/*PTAC7* high-diversity region.

Discussion

High-diversity genes represent an important category of loci in organismal genomes since they may be targets of balancing selection. The relative importance of balancing selection in maintaining intraspecific variation is not yet well understood and is still debated among those who believe it may play a major role and those who favour a mutation–drift balance explanation (Ferreira & Amos 2006; Charlesworth *et al*. 2007). In *Arabidopsis thaliana*, several genomic regions with levels and patterns of sequence diversity consistent with the presence of a balanced polymorphism have been identified and suggest a role for selection in maintaining diversity at these genes (Kroymann *et al*. 2003; Cork & Purugganan 2005; Kroymann & Mitchell-Olds 2005; Bakker *et al*. 2006). If selection is indeed acting at these loci, then we would expect to find evidence for functional variation in addition to simply detecting molecular footprints of selection. In all three cases studied here, we observe either differences in allele-specific gene expression levels or differentiated protein variants that provide further support for the possibility of balancing selection acting at these candidate genes.

The *PTAC7* gene was identified as a high-diversity gene on chromosome 5 (Cork & Purugganan 2005) and our analysis indicates that it has high levels of variation in the promoter region and third intron that could potentially harbour *cis*-regulatory elements. Consistent with this analysis, our expression study reveals differences in the level of allele-specific expression at *PTAC7*. This gene encodes a protein that is associated with the plastid RNA polymerase, suggesting a role in chloroplast gene expression; mutant alleles of this gene produces a pigment-defective phenotype (Budziszewski *et al*. 2001). There is some evidence that *PTAC7* expression might be up-regulated in response to metabolite cues; MPSS signatures associated with this gene indicate a nearly twofold increase in expression in leaves following treatment with salicylic acid, which may also suggest a possible involvement in a disease resistance (Ward *et al*. 1991) or more generalized stress response. It would be attractive to consider a role for this gene in the transcriptional reprogramming that is known to occur during disease resistance (Nimchuk *et al*. 2003), especially given the pre-established role of balancing selection at other disease resistance loci.

We also found a difference in the level of expression of alternate alleles at the *SSI* locus, which is linked to the *PTAC7* gene and also displays high levels of nucleotide polymorphism. Although the possible functional consequence of this expression variation is unknown, it is conceivable that the amount of *SSI* product could impact amylopectin branching patterns and consequently alter aspects of resource availability in the plant. The molecular signal of selection seems stronger, however, at the downstream *PTAC7* locus; it may be that *SSI* shares regulatory elements with *PTAC7*, or that the observed variation in the level of expression is neutral, and thereby invisible to selection, but is the result of hitchhiking of neutral *cis*-regulatory variants at *SSI*.

In the case of *BHL10*, a sliding window analysis reveals a peak of diversity centred on multiple replacement polymorphisms located within and just upstream of the protein-protein interacting BELL domain. A recent yeast two-hybrid screen defined the protein interaction network involving the BELL-like gene family members and their interactors (Hackbusch *et al*. 2005). The *BLH10* protein was shown to be one of the most highly connected members of the network, interacting with 13 other known proteins. The binding affinity and specificity of these transcription factors are largely determined by the protein interactions they form. Our results, including identifying at least one radical glutamine to histidine amino acid substitution in the BELL domain, provide a basis for future investigations into allele-specific phenotypic effects once the precise function of *BLH10* is defined.

A polymorphic *copia*-like retrotransposon is located in the intergenic region downstream of the *BLH10* candidate locus, and it is known that large indel polymorphisms and repetitive elements can suppress recombination in local regions of the genome and facilitate the maintenance of differentiated alleles (Charlesworth *et al*. 1986; Jaarola *et al*.

1998; Casselman *et al*. 2000). This may produce a signature similar to that of a balanced polymorphism especially if alternate alleles are maintained in isolated populations for long evolutionary periods, as suggested in this case by the strong longitudinal cline at *BLH10*. It is unclear, however, why, in the absence of selection, so much excess variation would accumulate in this region and why it would be confined entirely to the *BLH10* locus. It is thought that transposition events often accompany gene duplication (Bailey *et al*. 2003) and can underlie adaptive evolution (Michalak *et al*. 2001). For this reason, transposable element insertions that occur at high frequency within a species are often considered candidate adaptive trait loci, particularly when they are observed in close proximity to known genes (Franchini *et al*. 2004).

The lack of discrete historical populations in *A. thaliana* precludes traditional population genetic analyses to compare levels of variation within and between populations. We were, however, able to investigate the Eurasian allelic distribution of alleles at each of our candidate loci. Our analysis indicates consistency with genome-wide population structure predictions (Nordborg *et al*. 2005; Schmid *et al*. 2005) at each investigated locus, with *BLH10* displaying a further significant longitudinal cline. It is unclear how to interpret this longitudinal cline in the absence of a corresponding phenotype to investigate. Taken together with the striking pattern of genetic dimorphism and clustered amino acid polymorphisms also observed at this locus, it seems most likely that an interplay among historical and selective mechanisms will best explain the maintenance of variation at *BLH10*.

Understanding the relative contribution of balancing selection to the adaptive process will be important if the maintenance of within-species diversity is to be adequately explained. Genome scanning may be a useful tool for identifying new sets of genes subject to balancing selection and this methodology is advanced if the functional mechanisms responsible can be further characterized. Determining a possible molecular basis for allelic differentiation of putative balanced polymorphisms and assessments of the spatial distribution of alleles are essential if we are to reveal the functional and ecological significance of these genes.

Acknowledgements

We would like to thank Trudy F. C. Mackay and Ian Dworkin for advice with statistical analysis and Lisa Goering for her pyrosequencing protocols. We are grateful to the North Carolina State University Genome Research Laboratory and Phytotron personnel for their assistance. We would also like to thank Karl Schmid for his generosity in sharing *A. thaliana* seed and Jennifer Modliszewski for her help with RNA isolation. This work was supported by a NIH predoctoral training grant fellowship to J.M.R and in part by a grant from the NSF Frontiers in Integrative Biological Research Program to M.D.P.

References

Abbot RJ, Gomes MF (1989) Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity*, **62**, 411–418.

Aidoo M, Terlouw DJ, Kolczak MS *et al*. (2002) Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet*, **359**, 1311–1312.

Bailey JA, Liu G, Eichler EE (2003) An *Alu* transposition model for the origin and expansion of human segmental duplications. *American Journal of Human Genetics*, **73**, 823–834.

Bakker EG, Toomajian C, Kreitman M, Bergelson J (2006) A genome-wide survey of *R* gene polymorphisms in Arabidopsis. *Plant Cell*, **18**, 1803–1818.

Bamshad MJ, Mummidi S, Gonzalez E *et al*. (2002) A strong signature of balancing selection in the 5′-*cis*-regulatory region of. *Ccr5 Proceedings of the National Academy of Sciences, USA*, **99**, 10539–10544.

Bao X, Franks RG, Levin JZ, Liu Z (2004) Repression of *AGAMOUS* by *BELLRINGER* in floral and inflorescence meristems. *Plant Cell*, **16**, 1478–1489.

Becker A, Bey M, Burglin TR, Saedler H, Theissen G (2002) Ancestry and diversity of *BEL1*-like homeobox genes revealed by gymnosperm (*Gnetum gnemon*) homologs. *Development Genes and Evolution*, **212**, 452–457.

Budziszewski GJ, Lewis SP, Glover LW *et al*. (2001) *Arabidopsis* genes essential for seedling viability: isolation of insertional mutants and molecular cloning. *Genetics*, **159**, 1765–1778.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences, USA*, **101**, 15670–15675.

Casselman AL, Vrebalov J, Conner JA *et al*. (2000) Determining the physical limits of the *Brassica S* locus by recombinational analysis. *Plant Cell*, **12**, 23–33.

Charbonnel N, Pemberton J (2005) A long-term genetic survey of an ungulate population reveals balancing selection acting on *MHC* through spatial and temporal fluctuations in selection. *Heredity*, **95**, 377–388.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *Public Library of Science, Genetics*, **2**, e64.

Charlesworth D, Awadalla P (1998) Flowering plant self-incompatibility: the molecular population genetics of *Brassica* S-loci. *Heredity*, **81**, 1–9.

Charlesworth B, Langley CH, Stephan W (1986) The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics*, **112**, 947–962.

Charlesworth B, Miyo T, Borthwick H (2007) Selection responses of means and inbreeding depression for female fecundity in *Drosophila melanogaster* suggest contributions from intermediate-frequency alleles to quantitative trait variation. *Genetics Research*, **89**, 85–91.

Cork JM, Purugganan MD (2005) High-diversity genes in the *Arabidopsis* genome. *Genetics*, **170**, 1897–1911.

Crawford DL, Segal JA, Barnett JL (1999) Evolutionary analysis of TATA-less proximal promoter function. *Molecular Biology and Evolution*, **16**, 194–207.

Delvalle D, Dumez S, Wattebled F *et al*. (2005) Soluble starch synthase I: a major determinant for the synthesis of amylopectin in *Arabidopsis thaliana* leaves. *Plant Journal*, **43**, 398–412.

Duarte JM, Cui L, Wall PK *et al*. (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular Biology and Evolution*, **23**, 469–478.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, **8**, 186–194.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, **8**, 175–185.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Ferreira AG, Amos W (2006) Inbreeding depression and multiple regions showing heterozygote advantage in Drosophila melanogaster exposed to stress. *Molecular Ecology*, **15**, 3885–3893.

Franchini LF, Ganko EW, McDonald JF (2004) Retrotransposon–gene associations are widespread among *D. melanogaster* populations. *Molecular Biology and Evolution*, **21**, 1323–1331.

Hackbusch J, Richter K, Muller J, Salamini F, Uhrig JF (2005) A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. *Proceedings of the National Academy of Sciences, USA*, **102**, 4908–4912.

Harr B, Kauer M, Schlotterer C (2002) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA*, **99**, 12949–12954.

Jaarola M, Martin RH, Ashley T (1998) Direct evidence for suppression of recombination within two pericentric inversions in humans: a new sperm-FISH technique. *American Journal of Human Genetics*, **63**, 218–224.

Kelly JK (1997) A test of neutrality based on interlocus associations. *Genetics*, **146**, 1197–1206.

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proceedings of the National Academy of Sciences, USA*, **100**, 14587–14592.

Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. *Nature*, **435**, 95–98.

Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME (2003) Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Molecular Biology and Evolution*, **20**, 135–144.

Loisel DA, Rockman MV, Wray GA, Altmann J, Alberts SC (2006) Ancient polymorphism and functional variation in the primate *MHC-DQA1* 5′-*cis*-regulatory region. *Proceedings of the National Academy of Sciences, USA*, **103**, 16331–16336.

Meyers BC, Lee DK, Vu TH *et al*. (2004) *Arabidopsis* MPSS. An online resource for quantitative expression analysis. *Plant Physiology*, **135**, 801–813.

Michalak P, Minkov I, Helin A *et al*. (2001) Genetic evidence for adaptation-driven incipient speciation of *Drosophila melanogaster* along a microclimatic contrast in 'Evolution Canyon', Israel. *Proceedings of the National Academy of Sciences, USA*, **98**, 13195–13200.

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolu tionary significance of natural variation in *Arabidopsis*. *Nature*, **441**, 947–952.

Nimchuk Z, Eulgem T, Holt BF, 3rd Dangl JL (2003) Recognition and response in the plant immune system. *Annual Review of Genetics*, **37**, 579–609.

Nordborg M, Hu TT, Ishino Y *et al*. (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *Public Library of Science, Biology*, **3**, e196.

Ostrowski MF, David J, Santoni S *et al*. (2006) Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Molecular Ecology*, **15**, 1507–1517.

Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmuller R (2006) *pTAC2-6*, and *-12* are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell*, **18**, 176–197.

Prigoda NL, Nassuth A, Mable BK (2005) Phenotypic and genotypic expression of self-incompatibility haplotypes in *Arabidopsis lyrata* suggests unique origin of alleles in different dominance classes. *Molecular Biology and Evolution*, **22**, 1609–1620.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Ral JP, Derelle E, Ferraz C *et al*. (2004) Starch division and partitioning. A mechanism for granule propagation and maintenance in the picophytoplanktonic green alga *Ostreococcus tauri*. *Plant Physiology*, **136**, 3333–3340.

Ray A, Robinson-Beers K, Ray S *et al*. (1994) *Arabidopsis* floral homeotic gene *BELL* (*BEL1*) controls ovule development through negative regulation of *AGAMOUS* gene (*AG*). *Proceedings of the National Academy of Sciences, USA*, **91**, 5761–5765.

Rockman MV, Hahn MW, Soranzo N *et al*. (2005) Ancient and recent positive selection transformed opioid *cis*-regulation in humans. *Public Library of Science, Biology*, **3**, e387.

Rockman MV, Wray GA (2002) Abundant raw material for *cis*-regulatory evolution in humans. *Molecular Biology and Evolution*, **19**, 1991–2004.

Rozas J, Sanchez-De I, Barrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Samis KE, Heath KD, Stinchcombe JR (2008) Discordant longitudinal clines in flowering time and *phytochrome C*. *Arabidopsis thaliana*. *Evolution*, **62**, 2971–2983.

Schmid KJ, Ramos-Onsins S, Ringys-Beckstein H, Weisshaar B, Mitchell-Olds T (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.

Schmid KJ, Torjek O, Meyer R *et al*. (2006) Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theoretical and Applied Genetics*, **112**, 1104–1114.

Schmidt PS, Zhu CT, Das J *et al*. (2008) An amino acid polymorphism in the *couch potato* gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA*, **105**, 16207–16211.

Schulte PM, Glemet HC, Fiebig AA, Powers DA (2000) Adaptive variation in *lactate dehydrogenase-B* gene expression: role of a stress-responsive regulatory element. *Proceedings of the National Academy of Sciences, USA*, **97**, 6597–6602.

Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology*, **9**, 2109–2118.

Smith HM, Hake S (2003) The interaction of two homeobox genes, *BREVIPEDICELLUS* and *PENNYWISE*, regulates internode patterning in the *Arabidopsis* inflorescence. *Plant Cell*, **15**, 1717–1727.

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature*, **400**, 667–671.

Storz JF, Payseur BA, Nachman MW (2004) Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Molecular Biology and Evolution*, **21**, 1800–1811.

Storz JF, Sabatino SJ, Hoffmann FG *et al*. (2007) The molecular basis of high-altitude adaptation in deer mice. *Public Library of Science, Genetics*, **3**, e45.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1993) Unbiased estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution*, **10**, 677–688.

Tian D, Araki H, Stahl E, Bergelson J, Kreitman M (2002) Signature of balancing selection in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA*, **99**, 11525–11530.

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. *Nature*, **423**, 74–77.

Tishkoff SA, Varkonyi R, Cahinhinan N *et al*. (2001) Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science*, **293**, 455–462.

Ward ER, Uknes SJ, Williams SC *et al*. (1991) Coordinate gene activity in response to agents that induce systemic acquired resistance. *Plant Cell*, **3**, 1085–1094.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Jennifer Reininga was a graduate student at North Carolina State University and is currently pursuing studies in evolutionary genomics as a postdoctoral fellow at Duke University. Dahlia Nielsen develops methods for statistical genetics and genomics as a professor at the Department of Genetics and Bioinformatics Research Center at North Carolina State University. Michael Purugganan is the Dorothy Schiff Professor of Genomics at the Department of Biology and Center for Genomics and Systems Biology at New York University, and investigates the genetic basis of adaptation in plants and *Dictyostelium*.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Table S1 Sequenced *A. thaliana* accessions

Table S2 Genotypes of candidate genes

Table S3 Mixed-Model ANOVA

Text S1 Gene PCR and sequencing primers.

Text S2 Pyrosequencing SNPs and Primers.

Text S3 Genotyping PCR primers and restriction enzymes.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.