Molecular Ecology (2007) **16**, 3050–3062 doi: 10.1111/j.1365-294X.2007.03298.x

Sequence diversity and haplotype associations with phenotypic responses to crowding: *GIGANTEA* affects fruit set in *Arabidopsis thaliana*

MARCUS T. BROCK, PETER TIFFIN and CYNTHIA WEINIG
Department of Plant Biology, University of Minnesota, 1445 Gortner Avenue, Saint Paul, Minnesota 55108, USA

Abstract

Identifying the molecular genetic basis of intraspecific variation in quantitative traits promises to provide novel insight into their evolutionary history as well as genetic mechanisms of adaptation. In an attempt to identify genes responsible for natural variation in competitive responses in *Arabidopsis thaliana*, we examined DNA sequence diversity at seven loci previously identified as members of the phytochrome B signalling network. For one gene, *GIGANTEA* (*GI*), we detected significant haplotype structure. To test for *GI* haplogroup–phenotype associations, we genotyped 161 *A. thaliana* accessions at *GI* and censused the same accessions for total fruit set and the expression of three phenotypic traits (days to flowering, petiole length, and inflorescence height) in a greenhouse experiment where plants were grown in crowded and uncrowded environments. We detected a significant association between *GI* and total fruit set that resulted in a 14% difference in average fruit set among *GI* haplogroups. Given that fruit set is an important component of fitness in this species and given the magnitude of the effect, the question arises as to how variation at this locus is maintained. Our observation of frequent and significant epistasis between *GI* and background single nucleotide polymorphisms (SNP), where the fitness ranking of the *GI* allele either reverses or does not differ depending on the allele at the interacting SNP, suggests that epistatic selection may actively maintain or at least slow the loss of variation at *GI*. This result is particularly noteworthy in the light of the ongoing debate regarding the genetic underpinnings of phenotypic evolution and recent observations that epistasis for phenotypic traits and components of fitness is common in *A. thaliana*.

Keywords: *Arabidopsis thaliana*, association mapping, epistasis, fitness, *GIGANTEA*, sequence diversity

Received 19 August 2006; revision received 28 November 2006; accepted 16 January 2007

Evolutionary biologists have traditionally described the inheritance and evolution of complex traits using purely statistical techniques, without identifying or characterizing the underlying genes (Falconer & Mackay 1997; Lynch & Walsh 1997). Through the use of experimental genetic approaches such as mutant screens and transgenic overexpression, molecular biologists are rapidly characterizing the genes that act in diverse developmental pathways. This rapid and ongoing annotation of developmental loci can be used to formulate a list of candidate loci that contribute to phenotypic variation in natural populations. Of the many genes identified in any given pathway, only a subset is likely to contribute to phenotypic and fitness differences among populations, i.e. only a subset is likely to be affected by phenotypic selection. Analysis of DNA sequence diversity and its association with phenotypic variation provides a means to narrow the field of genes likely to affect the expression of variation in phenotypic traits (Templeton 1995; Olsen *et al.* 1997; Cardon & Bell 2001; Thornsberry *et al.* 2001; Stinchcombe *et al.* 2004). Identification of these genes then allows for further investigation into the evolutionary genetics of quantitative traits, including how selection acts on genes' underlying complex traits (Fenster *et al.* 1997; Gillespie & Turelli 1989) and the role of both spatial variation (Levene 1953; Felsenstein 1976; Gillespie *et al.* 1989) and epistatic interactions in maintaining genetic variation (Shook & Johnson 1999; Leips & Mackay 2000; Wade 2001; Weinig *et al.* 2003; Kroymann & Mitchell-Olds 2005; Malmberg *et al.* 2005). Here, we sample *Arabidopsis*

Correspondence: Marcus Brock, Fax: 612 625-1738; E-mail: brockmt@umn.edu

thaliana accessions for molecular variation at seven loci involved in plant competitive responses and test for associations between molecular and phenotypic variation.

Plant competitive responses to variation in light quantity and quality are ecologically important quantitative traits that have been the subject of ecological, evolutionary, and molecular genetic investigation. Ecologically, shifts in the ratio of red:far-red (R:FR) wavelengths are reliably associated with neighbour proximity and provide a cue of future competition for light (Smith 1982, 2000; Ballaré *et al*. 1990; Schmitt & Wulff 1993); in particular, the amount of red light diminishes in crowded stands as chlorophyll absorbs red light but transmits far-red light. Perception of these cues elicits a range of phenotypic responses, including acceleration of flowering, increased petiole elongation, decreased area of leaf blades, elevated leaf angles, and dramatic increases in stem elongation (Smith 1982; Smith & Whitelam 1990). Experimental investigations in contemporary populations have shown that variation in shade avoidance can significantly affect fitness (Schmitt *et al*. 1995; Dudley & Schmitt 1996; Weinig 2000). In high-density stands, shade-avoidance responses to R:FR cues often occur in tandem with additional phenotypic responses to increased crowding and competition for above- (e.g. light quantity) and below-ground resources (e.g. nutrients and water); these latter responses may or may not be adaptive and, although further molecular characterization is needed, they are likely regulated by different developmental pathways (Hodge 2004; Malamy 2005).

Although the specific genes underlying intraspecific variation in shade-avoidance responses have not been identified, molecular genetic analyses of *A. thaliana* mutants have identified a set of photoreceptors and downstream signalling molecules involved in sensing and responding to the light environment. *PHYTOCHROME B* is one of five phytochrome genes (*PHYA–PHYE*; Sharrock & Quail 1989) present in the *A. thaliana* genome and is the primary photoreceptor involved in sensing neighbour proximity (Aukerman *et al*. 1997; Whitelam & Devlin 1997; Devlin *et al*. 1999). Genetic and or biochemical analyses have identified a large number of genes that are involved in *PHYB*-mediated signalling in *A. thaliana*. These include three genes, *NUCLEOSIDE DIPHOSPHATE KINASE 2* (*NDPK2*) and the bHLH transcription factors *PHYTOCHROME INTERACTING FACTOR 3* and *4* (*PIF3* and *PIF4*), which encode proteins that bind directly to the active form of phytochrome (Huq & Quail 2002; Ni *et al*. 1998, 1999; Soh *et al*. 1999). A putative transcription factor induced by reduced R:FR has also been identified: *ARABIDOPSIS THALIANA HOMEOBOX 2* (*ATHB-2*; Steindler *et al*. 1999). When we began our studies, *PHYTOCHROME KINASE SUBSTRATE 1* (*PKS1*) had been shown to bind to *PHYA* and *PHYB* and is phosphorylated by activated phytochrome (Fankhauser *et al*. 1999), although it now appears to be more important as a regulator of *PHYA* signalling (Lariguet *et al*. 2003). Finally, *GIGANTEA (GI)* encodes a novel protein that has been implicated as a positive factor in *PHYB* signalling (Huq *et al*. 2000). The *GI* locus is pleiotropic and affects a variety of phenotypic traits, including photoperiodic induction of flowering (Fowler *et al*. 1999), starch accumulation (Eimert *et al*. 1995), diurnal regulation of transpiration (Sothern *et al*. 2002), and cold and oxidative stress responses (Cao *et al*. 2005, 2006). Mutant screens and studies of natural allelic variation also indicate that *GI* is required for proper functioning of the circadian clock and helps maintain circadian rhythms across a range of temperatures (Yanovsky & Kay 2002; Edwards *et al*. 2005; Gould *et al*. 2006).

Our primary objective was to test for evidence that any of these seven candidate genes contribute to naturally occurring variation in response to increased crowding or in fitness. To accomplish this objective, we sequenced a small sample of *A. thaliana* accessions with extreme differences in the expression of one shade-avoidance phenotype (hypocotyl elongation) and evaluated sequence diversity and haplotype structure. From this initial screen, significant haplogroup structure was detected in one gene (*GIGANTEA*), which we then tested for genotype–phenotype associations using a much larger sample of ecotypic diversity.

Materials and methods

Study system and accession sampling

Arabidopsis thaliana (Brassicaceae) originated in Eurasia and is distributed across a wide latitudinal gradient. Of particular relevance to the current study, plants of *A. thaliana* in natural populations are variably crowded by conspecifics. Due to the passive mode of seed dispersal in which seeds drop from dehiscing siliques, a dense seed shadow typically forms around the seed parent; however, the rare seed that falls far from the parent may grow under noncompetitive conditions (Donohue *et al*. 2005; Weinig *et al*. 2007). Thus, the natural history of this species is suitable to studying phenotypic responses to crowding and genetic adaptation to competition.

We chose 15 accessions for analysis of sequence diversity at the candidate loci. Thirteen of the 15 were chosen because they exhibited extreme phenotypes in prior surveys of hypocotyl elongation responses to experimental red light environments (Maloof *et al*. 2001; Botto & Smith 2002). Of ~150 sampled accessions, five accessions (Table 1) showed much lower than average responsiveness of elongation to low R:FR (An-2, Col-0, Di^{-1}, Li-5–3, Pi-0), while six had much higher than average elongation responses (Bla-6, Bu-2, Ma-2, Pa-2, Et-0, Kl-1) (Botto, Smith 2002). Two accessions (Br-0, Su-0) were chosen because hypocotyl elongation was not suppressed by red wavelengths, suggesting disabling of *PHYB*-mediated signal transduction (Maloof *et al*. 2001).

Table 1 Natural accessions of *Arabidopsis thaliana* sequenced at *GIGANTEA* and the neighbouring up- and downstream loci (*PAB3* and *PFL*). Location of origin and stock number at The Arabidopsis Information Resource are shown, as are hypocotyl sensitivity to R:FR and genotype at *GI*. Asterisks denote accessions also sequenced at *ATHB-2*, *PHYB*, *NDPK-2*, *PIF3*, *PIF4*, and *PKS1*

	Geographical origin	TAIR accession identifier	Hypocotyl sensitivity to R/FR	*GI* allele
An-2*	Antwerpen, Belgium	cs6604	Less responsive	—
Bch-3	Büchen, Germany	cs6610	Less responsive	C
Bla-4	Blanes, Spain	cs6619	Less responsive	A
Bla-6*	Blanes, Spain	cs6621	Responsive	C
Br-0*	Brunn, Czech Republic	cs6626	Less responsive	B
Bs-2	Basel, Switzerland	cs6628	Responsive	B
Bu-2*	Buchschlag, Germany	cs6633	Responsive	C
Co-3	Coimbra, Portugal	cs6671	Responsive	A
Col-0*	Columbia, USA	cs6673	Less responsive	C
Di-1*	Dijon, France	cs6681	Less responsive	A
Et-0*	Etraygues, France	cs6702	Responsive	A
Kl-1*	Köln, Germany	cs6757	Responsive	A
Li-5 : 3*	Limburg, Germany	cs6776	Less responsive	A
Ll-2	Llagostera, Spain	cs6783	Less responsive	A
Ma-2*	Marburg, Germany	cs6790	Responsive	A
Mh-1	Mühlen, Germany (origin)	cs6793	Responsive	C
Mt-0*	Martuba, Libya	cs6799	NA	C
Old-2	Oldenburg, Germany	cs6821	Less Responsive	B
Pa-2*	Palermo, Italy	cs6826	Responsive	—
Pi-0*	Pitztal, Austria	cs6832	Less responsive	B
Rsch-0	Rschew, Russia	cs6848	Responsive	C
Su-0*	Southport, UK	cs1540	Less responsive	C
Tsu-1*	Tsu, Japan	cs6926	NA	A

The remaining two accessions (Tsu-1, Mt-0) were chosen because they are parents in two sets of recombinant inbred lines being developed for quantitative trait loci (QTL) mapping (Maloof and Weinig, unpublished data). We identified accessions with lower and higher than average responsiveness to light quality by regressing genotypic values of elongation expressed under low R:FR on those expressed by the same accession under high R:FR; accessions with negative residuals in this analysis are comparatively unresponsive to low R:FR given their level of hypocotyl elongation in high R:FR, while accessions with positive residuals are comparatively more responsive. By focusing primarily on accessions with extreme phenotypes, we aimed to maximize the probability of detecting sequence diversity in loci influencing traits that commonly respond to plant density.

Sequence collection and analyses

DNA used for polymerase chain reaction (PCR) amplifications was extracted from one individual from each of the 15 accessions using QIAGEN DNeasy (QIAGEN). We initially PCR-amplified and sequenced ~500–1600 bases from each of seven genes (*PHYB*, *NDPK2*, *PIF3*, *PIF4*, *ATHB-2*, *PKS1*, *GI*) from each accession (PCR primer sequences and information on the region of each gene sequence are provided in Table S1, Supplementary material). With the exception of *PHYB* and *GI*, these regions contained 50–90% of the coding regions of each gene (see Table S1, Supplementary material). We also amplified and sequenced five of these genes from *Arabidopsis lyrata* ssp. *petrea*; despite multiple attempts, we were unable to obtain sequences for *A. lyrata* ssp. *petrea PHYB* or *PKS1*. *A. lyrata* ssp. *petrea* PCR products, as well as two *A. thaliana* PCR products (*NDPK2* from accessions Di^{-1} and An-2) were cloned into TA vectors (Promega) prior to sequencing. Singleton sequence variation of cloned products was confirmed by recloning and resequencing samples. All DNA sequences have been submitted to GenBank (Accession nos: *PHYB*, EF193566–EF193580; *PIF3*, EF193481–EF193496; *PIF4*, EF193513–EF193527; *NDPK2*, EF193497–EF193512; *PKS*, EF193581–EF193595; *GI*, EF193528–EF193549; *ATHB2*, EF193550–EF193565).

Sequences from each locus were aligned visually using BIOEDIT version 5.0.9 (Hall 1999). For each locus, we estimated nucleotide diversity using the number of haplotypes (H), the number of segregating sites (S), the average number of segregating sites, θ (Watterson 1975), and the average number of pairwise differences between sequences, π (Nei 1987). S and π were calculated separately on replacement sites as well as on total and silent (synonymous and intron) sites. Deviations from a neutral-equilibrium model were tested using Tajima's *D* (Tajima 1989), Fu and Li's *D* with and without an outgroup (Fu & Li 1993), and ZnS (Kelly 1997). Confidence intervals around estimates of π, the probability of obtaining the number of sampled haplotypes (H), and the significance of H and ZnS were estimated by running 1000 coalescent simulations of the neutral model. These simulations were conditioned on the number of segregating sites and conducted assuming no recombination. Estimates of polymorphism, Tajima's *D*, ZnS, and coalescent simulations were conducted using DNASP version 3.53 (Rozas *et al*. 2003).

To assess the phylogenetic relationships among sequences and to test for the presence of distinct haplogroups at each locus, we constructed genealogical phylogenies using both maximum parsimony and maximum likelihood (ML) criteria in PAUP* version 4.0 beta10 (Swofford 2002). The evolutionary model used for maximum-likelihood trees was identified using MODELTEST 3.06 (Posada & Crandall 1998). In all cases, the ML tree was among the most parsimonious trees. Support for the phylogenetic relationships was assessed by bootstrapping the data 1000 times. We defined groups of > 3 sequences that had > 80% bootstrap support as being genealogically distinct from the other sampled sequences, and these are referred to as haplogroups. Insertion–deletion polymorphisms were excluded from genealogical reconstructions, as well as the estimates of sequence diversity and tests of neutral evolution described above.

Because we were primarily interested in identifying genes that may contribute to commonly segregating phenotypic variation, we restricted further analyses to the one gene,

GIGANTEA, which segregated distinct haplogroups. A small pilot study (n = 20 accessions) also suggested a significant association between *GI* and petiole elongation responses to low R:FR (Brock, Tiffin and Weinig, unpublished data). We extended our analyses of *GI* sequence polymorphism in two directions. First, to increase the robustness of our phylogenetic analyses and to identify all single nucleotide polymorphisms (SNPs) that differentiated the defined haplogroups, we sequenced the entire *GI* coding region (~5100 bases) and the 60 upstream bases from the original 15 accessions as well as from an additional eight accessions that were sampled by Botto & Smith (2002). Four of these additional accessions were highly responsive (Bs-2, Co-3, Mh-1, Rsch-0) to changes to low R:FR, while the remaining four exhibited low responsiveness (Bch-3, Bla-4, Ll-2, Old-2), according to the criteria described for our initial selection of accessions (see above). Second, we sequenced 500–1500 bases from adjacent up- and downstream loci [*polyadenylate-binding protein 3* (*PAB3*) located ~6.0 kb upstream from *GI* start site and *pointed first leaf* (*PFL*) located ~550 bases downstream from *GI* stop codon] to evaluate the extent of linkage disequilibrium (LD) around *GI* (GenBank Accession nos: *PAB3*, EF193596–EF193616; *PFL*, EF193617–EF193637). We tested all pairwise combinations of polymorphic sites for LD across the entire *PAB3-GI-PFL* region and used a two-tailed Fisher's exact test to determine significance of r^2 (DNASP version 3.53). A Bonferroni correction was applied to adjust for multiple comparisons; however, this is a conservative correction and could exclude existing LD. We therefore present LD values that were significant ($P < 0.05$) with and without a Bonferroni correction.

From the full-length *GI* sequence, we identified two replacement polymorphisms that differentiated three haplogroups at *GI*. We designed one cleaved amplified polymorphic sequence (CAPs) and one derived cleaved amplified polymorphic sequence (dCAPs) assay to genotype 161 accessions at these two polymorphisms (see Table S2, Supplementary material for primer sequences and methods). Because the 161 accessions were primarily European and were sampled from a wide latitudinal distribution (see Table S3, Supplementary material), we tested for a cline in *GI* haplogroups using a contingency test (Proc FREQ, SAS version 8.02), in which the accessions were binned into four groups across the latitudinal gradient (30–45°, 45–50°, 50–55°, 55–60°). We also used logistic regression to test for a latitudinal cline in *GI* (results not presented); interpretation of results did not differ between the two methods.

Phenotypic measurements

We measured fruit production, which is highly correlated with fitness in *A. thaliana* (Westerman & Lawrence 1970; Mauricio *et al*. 1997), and three phenotypic traits [days to flower, inflorescence height, and proportion petiole length (petiole length/total leaf length)] in a greenhouse experiment involving each of the 161 accessions that were genotyped at *GI*. Focal plants were grown in either an uncrowded treatment (one focal plant/pot) or crowded treatment (one focal plant and four neighbours of the same accession/pot evenly spaced 1 cm from the focal plant). Plant responses to increased neighbour density have been demonstrated to be strongly correlated with phytochrome-mediated responses (Donohue & Schmitt 1999); however, genetic pathways influencing below-ground resource acquisition likely play an additional important role that is not evaluated here. Pots were stratified in the dark at 4 °C for 4 days. We distributed one crowded and one uncrowded pot of each accession among eight spatial blocks (2608 total plants). Replicates within each block were randomized among 11 trays, which were rotated weekly within blocks. Plants were grown in Metro-Mix 200 (Scott's-Sierra Horticultural Products) in ($6.5 \times 6.5 \times 8.5$ cm) pots under natural light conditions in February–June 2005. Greenhouse temperature was set to 21 °C, and plants were watered weekly.

Focal plants were censused daily for the first open flower. One week after flowering, we collected the longest leaf in the southeast quadrant of each pot. Leaves were digitally scanned, and we measured total leaf length (petiole and blade) and petiole length on the scanned images using IMAGEI software version 1.31 (Wayne Rasband, National Institutes of Health). Following plant senescence, we measured total inflorescence height, and counted the number of fruits produced by each plant.

Testing genetic architecture and haplogroup–phenotype associations

For each phenotypic trait, we used a mixed-model analysis of variance (ANOVA) to partition variance among the fixed treatment effect and the following random effects: block, accession, and accession × treatment (PROC GLM; SAS version 8.02). In the analyses of inflorescence height, we included leaf length as a covariate in order to lessen the effects of overall plant size and enable a more direct test of the effects of R:FR on inflorescence elongation. To meet assumptions of ANOVA, we square-root transformed proportion petiole length and fruit set, and inverse-transformed days to flower. Sample sizes were reduced by focal plant mortality and removal of pots in the crowded treatment with only one neighbour, resulting in a final sample size of 1550 focal plants representing all 161 accessions.

To test for significant associations between *GI* haplogroup and phenotype (fruit production, proportion petiole length, inflorescence height, and days to flower), we first estimated least-square means of each accession by treatment combination from the ANOVA model above. These estimates were then used as the response variable in the following ANCOVA: trait = haplogroup + treatment + haplogroup ×

treatment +5 structure covariates + error. The five population structure covariates were included in the model to minimize the potential for cryptic population structure to produce spurious associations between *GI* haplogroup and phenotype (Pritchard *et al*. 2000b; Cardon & Palmer 2003). These covariates estimate per cent relatedness of an accession to inferred ancestral populations, and were calculated by running STRUCTURE 2.0 simulations (Pritchard *et al*. 2000a, b) on 115 SNPs assayed from all 161 accessions by Schmid *et al*. (2005). The STRUCTURE analyses identified six ancestral populations ($k = 6$ produced the highest estimated log-likelihood value for sampled models $k = 1 - 7$; Korves *et al*. 2007); however, the sixth population estimate is collinear with the sum of the other ancestries and is therefore excluded from the model. We also evaluated the significance of *GI* haplogroup–phenotype associations using a mixed-model analysis (TASSEL; 1.9.4; www.maizegenetics.net; Yu *et al*. 2006) that incorporates a kinship matrix (K-matrix) estimated from the SNP data in addition to the population structure covariates from STRUCTURE 2.0 (Q-matrix). The significant *GI* associations did not differ between the ANCOVA and TASSEL analyses, and we therefore present only the ANCOVA results.

For the one trait, fruit production, that showed a significant association with *GI* haplogroup, we also estimated the false positive rate using genome-wide SNP markers (Schmid *et al*. 2005). A significant *GI*–fruit set association could have any of several sources: (i) a functional or quantitative difference among the *GI* haplogroups in their effect on fruit set , (ii) LD with a neighbouring locus that causally affects fruit set, or (iii) widespread background LD with a causal locus (loci) resulting from population structure. Determining the frequency of SNPs with significant associations with fruit set not only provides a false positive metric (i.e. an estimate of whether the association is due to chance), but also qualitatively evaluates the hypothesis that the association results from LD between *GI* and other loci. We evaluated the frequency of SNP–fruit set associations using 84 SNPs where each alternative nucleotide was present at a minimum frequency of 5% in the 161 accessions. The model was identical to the model used to test for *GI* haplogroup–phenotype associations, with the exception that each of the 84 SNPs was serially substituted for *GI*, i.e. trait = SNP + treatment + SNP × treatment + structure covariate 1 + structure covariate 2 + … structure covariate 5 + error. If cryptic population structure (not accounted for by the structure covariates) underlies the significant *GI*–fruit set association, then a large proportion of random SNPs would be expected to show associations with fruit set as significant as (or more significant than) the one with *GI*.

Epistatic selection

We tested for epistatic interactions that involve *GI* and contribute to variation in fitness using the following ANCOVA model: fitness = treatment + *GI* + SNP + *GI* × SNP interaction + the 5 structure covariates + error (PROC GLM; SAS version 8.02). Because of limited degrees of freedom and the lack of a significant *GI* × treatment interaction in the initial analyses, we did not include any interaction terms involving treatment. This analysis was carried out using only two of the three *GI* haplogroups because the third *GI* haplogroup (GI_B) was infrequent (21.5%) and the within-level replication was low (i.e. the sample size was often less than eight for the $GI_B \times$ SNP interaction). To minimize imbalance among groups and the effect of phenotypic outliers, we also excluded *GI* × SNP interactions from the tally of significant epistatic interactions when the within-level frequency was less than eight; a total of 84 SNPs were used in this analysis.

To estimate a false positive rate for epistatic interactions, we randomly selected 50 SNPs from the full set of 84. We evaluated the frequency of significant pairwise interactions between each of these 50 'focal' SNPs and the remaining 83 SNPs. In effect, we ran the preceding model for *GI* 50 times, and substituted one the 50 'focal' SNPs for *GI* in each new model. As above, only SNP × SNP interactions with a sample size greater than or equal to eight were included.

Results

Patterns of sequence variation

We sequenced 500 to > 5000 bases of seven genes from each of 15 accessions of *Arabidopsis thaliana* (a total of ~14 400 bases from each accession) as well as the entire *GI* gene (> 5000 bases) from an additional eight accessions. The number of haplotypes, segregating sites, and estimates of pairwise differences among sequences at these loci varied considerably (Table 2).

Measures of intraspecific diversity reveal some evidence for deviations from neutral-equilibrium expectations for three genes *PIF3*, *NDPK2*, and *PHYB*. At both *PIF3* and *PHYB*, two distinct alleles were segregating as revealed by significant linkage disequilibrium (LD) that extended across the entire length of *PIF3* (> 600 bases, ZnS = 1.0, $P < 0.001$) and nearly the entire length of the sequenced region of *PHYB* (> 600 bases, ZnS = 0.90, $P < 0.001$). The significant linkage disequilibrium at *PIF3* and *PHYB* may reflect some form of balancing selection or local adaptation. This result for *PIF3* is also supported by significantly positive Fu and Li's *D* (Table 2). In contrast, Fu and Li's *D* is negative for *NDPK2* ($D = -2.35$, $P < 0.05$) indicating an excess of rare alleles, which can be characteristic of a selective sweep. However, departures from neutral expectations are common in *A. thaliana* (Hanfstingl *et al*. 1994; Kawabe *et al*. 1997; Purugganan & Suddith 1999; Aguade 2001; Nordborg *et al*. 2005). An excess of rare alleles in *NDPK2* could also result from demographic history or population structure.

Table 2 Number of sequences (N), total segregating sites (S_T), replacement segregating sites (S_R), total length, the number of haplotypes (H), number of coding sites, diversity at silent (π_{silent}) and replacement (π_{rep}), Tajima's *D*, Fu & Li's *D**, Fu & Li's *D*, and ZnS values for each of the seven genes sampled

Gene	N	S_T	S_R	H	Length	Coding	π_{silent}	$\pi_{rep.}$	Tajima's *D*	Fu & Li's *D**	Fu & Li's *D*	ZnS
ATHB2	15	61	2	14	1022	596	0.029	0.0021	–0.38	–0.74	–0.24	0.21
GI	21	42	7	14	5066	3517	0.004	0.0005	–0.23	–0.24	–0.62	0.23
NDPK2	15	20	3	12	1612	609	0.0031	0.0013	–1.45	–1.75	–2.35*	0.22
PHYB	15	23	3	6	850	850	0.024	0.0011	–0.96	0.96	NA	0.90***
PIF3	15	18	7	2	875	784	0.01	0.0029	–0.79	1.54*	1.74*	1.0***
PIF4	14	4	1	3	1053	753	0.0015	0.0009	–0.16	0.31	0.25	0.13
PKS1	15	20	7	10	876	876	0.019	0.0026	–0.75	0.85	NA	0.26
PAB3	21	26	1	7	589	81	0.018	0.0088	1.39	0.99	NA	0.57*
PFL	21	14	0	10	1404	456	0.0034	0	–0.32	–1.07	NA	0.27

*,*** significant by coalescent simulations, $P < 0.05$, 0.001, respectively.

Similarly, these last two factors could account for the strong LD at *PIF3* and *PHYB* as well as the location of these loci in regions of low recombination.

Despite finding significant LD across the length of *PIF3* and *PHYB*, at only one locus, *GI*, did we find evidence for common haplogroups. Specifically, at *GI* we detected three common haplogroups segregating among the sampled accessions (Figure S1, Supplementary material) and differing from one another at an average of 14.2 (14.0–14.5) sites and an average of 4.6 (4.4–5.0) coding sites. There were two fixed replacement sites that differentiated GI_A from GI_B, one fixed replacement that differentiated GI_A from GI_C, and one fixed replacement that differentiated GI_B from GI_C.

Tests of LD across the region including *GI* and adjacent loci (*PAB3* and *PFL*) reveal strong LD within *PAB3*. LD between polymorphic sites within *GI* as well as within *PFL* was weaker and less common than LD within *PAB3* (Fig. 1). The LD matrix also illustrates that strong linkage disequilibrium between *GI* and *PAB3* declines by the fourth polymorphic site of *GI* and does not include the two replacement polymorphisms used to distinguish the haplogroups (Fig. 1), limiting the potential that significant haplogroup–phenotype associations arise from LD between *GI* and upstream loci. In contrast, we detected significant LD between the sequenced *PFL* region and *GI*, and cannot discount the possibility that phenotypic associations with *GI* result from LD with *PFL* or other downstream loci.

We detected no significant change in the frequency of *GI* haplogroups across latitudes ($\chi^2 = 7.28$, d.f. = 6, $P = 0.30$), suggesting that variation at *GI* does not confer adaptation to selective agents such as climate and season duration that vary across latitudes. The lack of a cline also suggests that the observed *GI* haplogroup–fruit set association (see below) does not result from linkage disequilibrium between *GI* and loci under selection for adaptation to latitudinal climate parameters.

Phenotypic variation and genotype–phenotype associations

Both accession and density treatment accounted for significant variation in all traits (Table 3). Relative to plants in the uncrowded treatment, those in the crowded treatment accelerated days to flowering [60.8 ± 0.5 days (mean ± 95% CL) and 60.0 ± 0.5 days, respectively], had greater proportion petiole length (29.2 ± 0.6% and 36.8 ± 0.7%, respectively), and produced shorter inflorescences (32.8 ± 0.4 mm and 26.1 ± 0.5 mm, respectively). Increased plant density also decreased fruit set by an average of 69.1%. We detected significant genetic variation for responsiveness to crowding for proportion petiole length, inflorescence height, and fruit production, as shown by the significant accession × treatment interaction (Table 3).

GI haplogroup was not associated with variation in average number of days to flowering, proportion petiole length, or inflorescence height (all *P* values > 0.20; Table S4 and Figure S2, Supplementary material), nor did we detect any evidence for significant interactions between *GI* haplogroup and treatment (all *P* values > 0.20). By contrast, *GI* haplogroup explained significant variation in average fruit set (Table 4), even after statistically controlling for population structure. Post-hoc Tukey's tests indicate that GI_C has significantly higher fruit production in comparison with GI_A (Fig. 2, Tukey's Test, $P = 0.026$) and is marginally higher than GI_B (Tukey's Test, $P = 0.065$). Haplogroups GI_A and GI_B do not differ significantly from each other (Fig. 2, Tukey's Test, $P = 0.89$). Allelic variation at *GI* explained ~5% of the total variation in fruit set in the crowded treatment and ~1.3% of the variation in the uncrowded treatment. Only four of the 84 SNPs had a significant effect on fruit set, indicating a false positive rate of < 5% and the effect of only one of these four SNPs was more significant than the observed effect of *GI*. If the observed *GI*–fruit set association was due to cryptic population structure and genome-wide

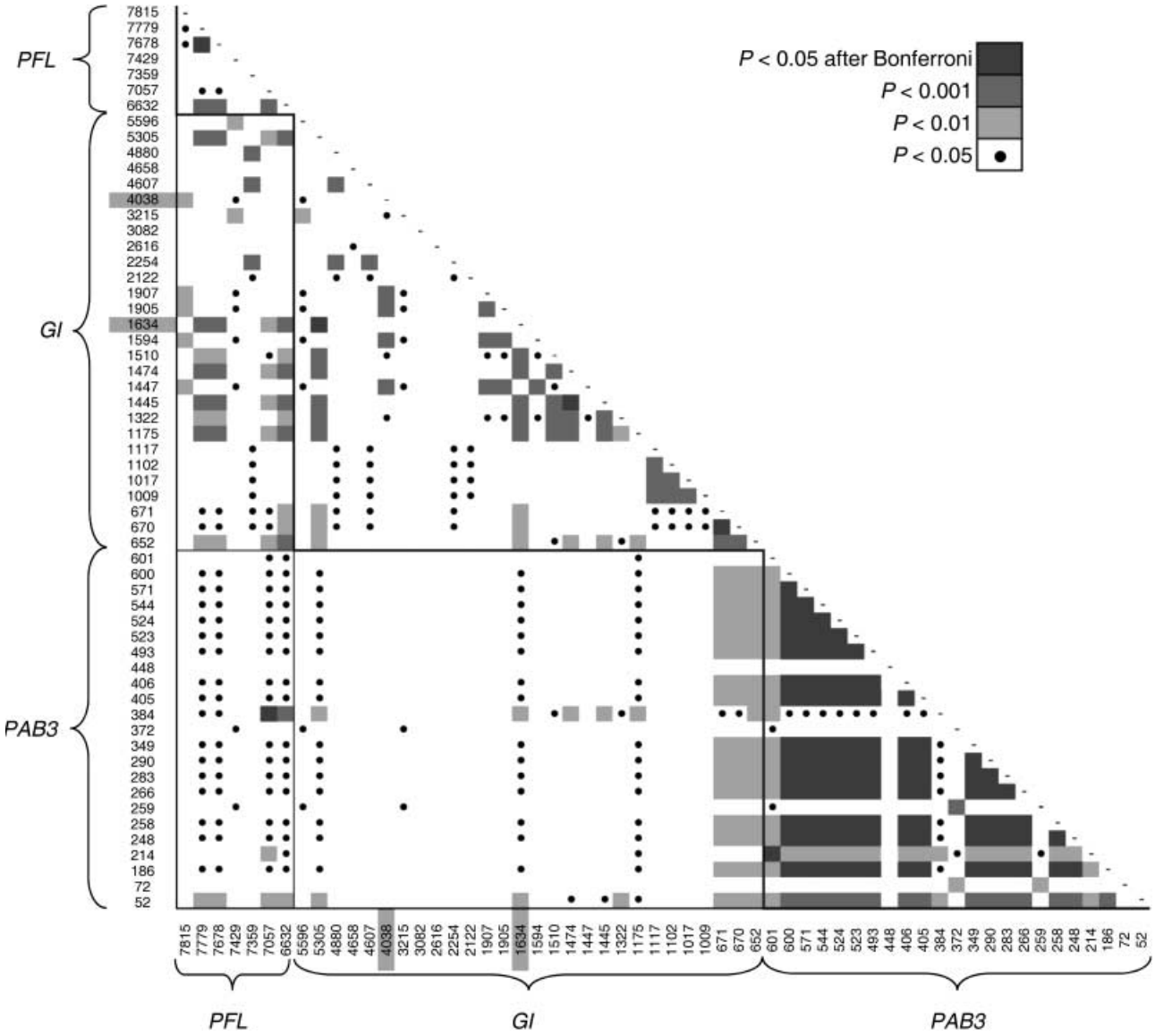

Fig. 1 Linkage disequilibrium matrix between polymorphic sites across *GIGANTEA* and sequenced regions of adjacent genetic loci, *PAB3* and *PFL*. Colour-coded significance levels from two-tailed Fisher's exact tests of nonrandom association between polymorphic sites are indicated in the legend. Highlighted sites within *GI* indicate locations of replacement polymorphisms that distinguish haplogroups.

LD, the rate of significant SNP–fruit set associations should be greater than that predicted by chance alone (5%). Our results support the hypothesis that functional differences among *GI* alleles, rather than population structure, account for the observed effect of *GI* on fruit set.

Tests for *GI* × SNP epistatic interactions on fruit set revealed six of 84 significant interactions (all P values < 0.05). Depending on the SNP marker involved, we detected both reversals in the ranking of *GI* alleles across SNP backgrounds as well as conditional neutrality of *GI* rankings (i.e. epistatic interactions where the effect of *GI* on fruit set is significant in only one of the two possible SNP backgrounds; Fig. 3). Only three of the 50 SNP markers (6%) we tested for epistasis had = 6 significant SNP × SNP interactions. Although this false positive rate is slightly greater than that predicted by chance (5%), our results suggest that population ancestry does not account for observed *GI* × SNP epistasis.

Discussion

Advances in sequencing techniques are facilitating the exploration of molecular diversity as well as identification of loci that contribute to variation in quantitative traits. We sampled genetic diversity in 15 *Arabidopsis thaliana* accessions at seven genes previously implicated in *PHYB*-mediated signal transduction as well as diversity in 23 accessions

Table 3 Results from analysis of (co)variance testing the effect of density treatment (uncrowded and crowded) on fruit set and three phenotypic traits in *Arabidopsis thaliana*. Leaf length was included in the analysis of inflorescence height (B) to reduce the influence of plant vigor on inflorescence elongation

(A)		Days to flower		Proportion petiole length		Fruit set	
Source	d.f.	*F*	*P*	*F*	*P*	*F*	*P*
Block	7	45.78	0.0001	3.02	0.0038	1.41	0.1225
Accession (A)	160	18.68	0.0001	2.50	0.0001	1.94	0.0001
Treatment (T)	1	6.35	0.0118	218.58	0.0001	1307.27	0.0001
A × T	160	1.02	0.4317	1.39	0.0017	1.41	0.0011

(B)		Inflorescence height	
Source	d.f.	*F*	*P*
Leaf Length	1	185.45	0.0001
Block	7	13.34	0.0001
Accession (A)	160	2.82	0.0001
Treatment (T)	1	277.23	0.0001
A × T	160	1.64	0.0001

Table 4 Analysis of covariance in fruit set of 161 *Arabidopsis thaliana* accessions testing the effects of competition treatment (uncrowded and crowded treatments) and *GIGANTEA* haplogroup. Five estimates of relatedness to predicted historical populations (spop1, spop2, … spop5) were included to account for confounding effects of population structure

		Fruit set	
Source	d.f.	*F*	*P*
spop1	1	0.13	0.7207
spop2	1	0.42	0.5171
spop3	1	0.65	0.4219
spop4	1	0.04	0.8454
spop5	1	1.08	0.2994
Treatment (T)	1	800.4	0.0001
GI	2	4.43	0.0127
GI × T	2	0.27	0.7602

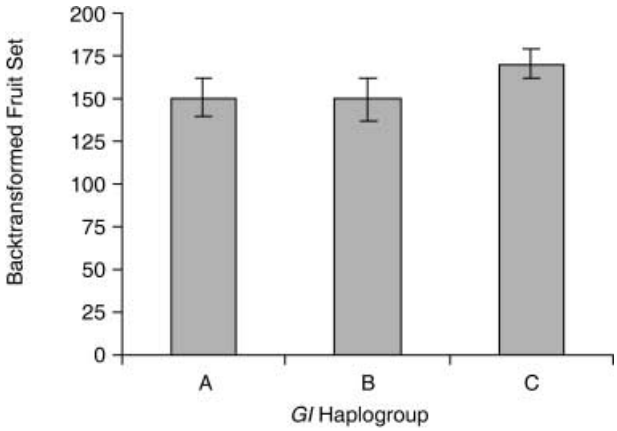

Fig. 2 Average backtransformed fruit production of *Arabidopsis thaliana* accessions in three identified *GI* haplogroups (GI_A $N = 45$, GI_B $N = 34$, and GI_C $N = 82$) that were subjected to uncrowded and crowded treatments in a greenhouse experiment. The GI_C haplogroup produced significantly more fruit than GI_A ($P = 0.026$) and marginally more than GI_B ($P = 0.065$); GI_A and GI_B did not differ significantly from each other ($P = 0.89$). Error bars indicate 95% confidence intervals.

at one locus, *GI*. Only *GI* showed evidence of haplogroup structure. In a large phenotypic screen, *GI* haplogroups were not associated with variation in plasticity of days to flowering, proportion petiole length, or inflorescence height to variable density, suggesting that variation at other loci accounts for differences among natural populations of *A. thaliana* in responsiveness to crowding. It is, however, possible that competition among crowded plants for above- and below-ground resources outweighs the phenotypic effects of changes in R:FR in *A. thaliana*. Experimental manipulations of R:FR independent of density might therefore reveal an effect of the sampled loci on shade-avoidance traits, and variation at these loci might contribute to the expression of naturally occurring variation in crowding responses in other species. With regard to fitness, *GI* haplogroup was significantly associated with fruit set. We also find evidence for epistatic interactions between *GI* and genetic background that may act to preserve, or at least slow the loss of, haplogroup variation (Leips, Mackay 2000; Wade 2001). These results are consistent with the increasingly frequent observation that epistatic variance contributes to natural variation in quantitative traits, and stress the need to explore the role of epistasis in maintaining diversity.

Molecular diversity

Recent examination of nucleotide diversity across the *A. thaliana* genome suggests that levels of silent pairwise diversity of ~0.7% on average (Yoshida *et al*. 2003; Nordborg *et al*. 2005; Schmid *et al*. 2005). We observe a slightly higher

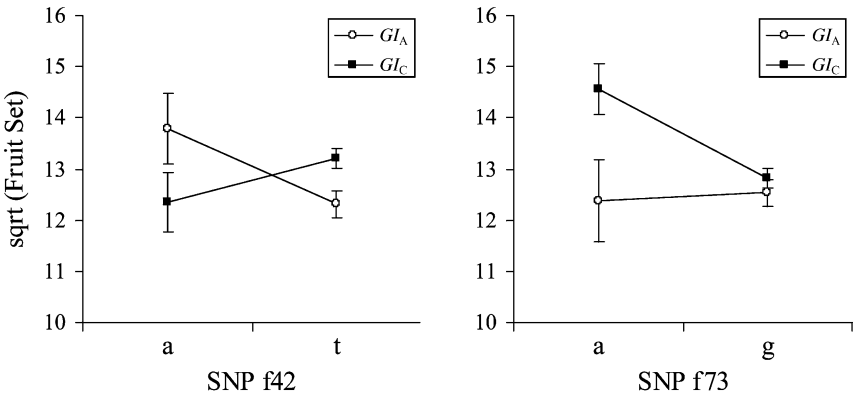

Fig. 3 Epistatic effects of *GI* × SNP markers on square-root transformed fruit set (± 1 SE) illustrating (A) an epistatic reversal (SNP f42) and (B) conditional neutrality (SNP f73).

average level of silent site diversity (1.3%) in the seven loci we sequenced. The observed range of diversity (0.15 to 2.9%, Table 2) is nevertheless consistent with the range observed in other genes that significantly influence phenotypic traits (Cork & Purugganan 2005). Although the diversity estimates for most loci sampled here fell within the range of previously reported estimates of species-wide diversity in *A. thaliana*, our estimate for *PHYB* diversity ($\pi_{silent} = 0.0256$) is considerably higher than that observed by McBreen and Mathews (unpublished data; $\pi_{silent} = 0.0023$). Our higher estimate appears to be due to the inclusion of two sequences (Bla-6 from Spain and Li-5–3 from Germany) that differ from each of the other 13 sequences at 20 or more sites. Excluding these two diverged sequences resulted in $\pi_{silent} = 0.0015$, similar to McBreen and Mathews's estimate. The comparatively high diversity estimate we obtained for *PHYB* does not appear to be due to our having sequenced accessions with extreme R:FR responses; one of the two accessions responsible for the high diversity estimate (Bla-6) exhibited higher than average responsiveness to low R:FR relative to a large sample of accessions (Botto, Smith 2002), while the other (Li-5–3) exhibited lower than average responsiveness. Rather, the effect these two sequences have on diversity estimates suggests a need for caution when comparing diversity from different samples, especially for species like *A. thaliana* that appear to harbour a high proportion of genes segregating two distinct haplotypes (Aguade 2001).

For one gene, *GI*, we detected significant haplogroup structure. Because genes that contribute to phenotypic variation in quantitative traits in natural populations are likely to have commonly segregating alleles (Falconer *et al.* 1997), we focused on *GI* to test for haplotype–phenotype associations. The three haplogroups we identified at *GI* are distinguished by two nonsynonymous polymorphisms. The GI_A haplogroup differed from both GI_B and GI_C at amino acid 162 (R vs. S, respectively) and at one fixed silent site. At the second nonsynonymous polymorphism, the B haplogroup differed from both the A and C at four fixed silent sites and a single fixed replacement site (amino acid 672 R vs. G, respectively), which occurs within a nuclear localization region of the protein (Huq *et al*. 2000). Based on the proximity to a functional region and on the fact that substitution matrices of amino acid biochemical properties (e.g. hydrophilic vs. hydrophobic nature) suggest both replacement polymorphisms may alter protein structure (Henikoff & Henikoff 1992), these sites may be promising for further investigation of SNPs that underlie significant haplotype–phenotype associations.

Quantitative genetics and haplogroup–phenotype associations

We detected statistically significant among-accession variation in all of the phenotypic traits we measured, days to flower, proportion petiole length, inflorescence height, and fruit production. We also detected significant effects of density on these four traits; plants experiencing crowding flowered earlier, had longer petioles, shorter inflorescences and reduced fruit set compared to those grown in an uncrowded setting. Accessions also differed in the relative responsiveness of proportion petiole length, inflorescence height, and total fruit set to the density treatments.

Haplogroup at *GI* was not significantly associated with either the average expression of days to flower, proportion petiole length, or inflorescence height or responsiveness of these traits to density. In contrast, we detected a strong and significant association between *GI* and total fruit set. Plants harbouring GI_C produced 14% more fruit on average than those with GI_A or GI_B, which did not differ significantly from each other (Fig. 2). Although density significantly reduced total fruit set and the accessions differed in the sensitivity of fruit set to density, the association between *GI* haplogroups and fruit set did not differ across treatments, suggesting that *GI* explains variation in fruit set across variable density environments in natural populations.

One complicating factor in association mapping is cryptic population structure, which can lead to the identification of spurious candidate gene–trait associations (Pritchard *et al*. 2000b; Cardon, Palmer 2003); however, two analyses

suggest population structure is not responsible for the *GI* effects we detected. First, we applied TASSEL and STRUCTURE to data at 115 SNPs (Schmid *et al*. 2005) to estimate population ancestry, and the resulting ancestry estimates were included as covariates in all association analyses (main effect and epistatic) to remove potential biases due to cryptic structure (e.g. Aranzana *et al*. 2005; Korves *et al*. 2007). Second, we used the SNPs to establish a false positive threshold (e.g. Thornsberry *et al*. 2001; Korves *et al*. 2007). The latter analyses are useful in estimating the potentially confounding effects of population structure, although they cannot control for the effects of population structure if found. Our analysis of SNP–fruit set associations detected a significant association with fruit set for < 5% of the SNPs, and only one SNP was more significantly associated with fruit set than *GI*, suggesting the *GI*–fruit set association was not explained by genome-wide linkage disequilibrium.

Although the association between *GI* and fruit set does not appear to result from population structure, we cannot exclude the possibility that the apparent effect of *GI* on fruit set is due to a locus in tight physical linkage. To examine the extent of linkage disequilibrium in the vicinity of *GI*, we sequenced regions from neighbouring loci and tested for LD between SNPs. We detected strong linkage disequilibrium within the *PAB3* locus and to a lesser extent within intragenic regions of *GI* and *PFL* (Fig. 1). Pairwise tests of intergenic regions suggest that significant LD between *GI* and upstream *PAB3* is limited to initial ~400 bases of *GI*; however, the *PFL-GI* region shows stronger evidence of intergenic linkage disequilibrium (Fig. 1). *PFL* (pointed first leaf) is one of three copies of ribosomal protein S18 that partially makes up the 40S small ribosomal subunit, and a *PFL*-null mutant results in reduced plant fresh weight (Van Lijsebettens *et al*. 1994), which may in turn affect fruit set. Further downstream of *PFL* are several other candidates including: a member of the ethylene response factor (*ERF*) family of transcription factors (~6.6 kb), regions with similarity to mammalian transforming growth factor (*TGF*) Beta receptors (~15.8 kb), and a region with similarity to Glycosyl Hydrolase family 9 proteins (28.2 kb). In sum, LD falls off quickly between *GI* and upstream regions; however, the possibility remains that *GI* is in linkage disequilibrium with a downstream locus responsible for observed fitness differences. Quantitative or transgenic complementation tests will be necessary to conclusively determine whether it is *GI* or a tightly linked locus that is responsible for the fruit set differences we detected.

If *GI* is the causal locus, the question arises as to what phenotypic traits account for the observed differences in fruit set among haplogroups. Molecular genetic analyses have characterized multiple functions for *GI*. This nuclear localized protein is an upstream regulator of *CONSTANS* and *FLOWERING LOCUS T*, which influence flowering time primarily under long days (Suarez-Lopez *et al*. 2001; Tseng *et al*. 2004). We did not detect a *GI*–flowering time association, however, suggesting that the *GI*–fitness association we detected was not mediated through an effect on flowering time. Similarly, the effect of *GI* haplogroup on hypocotyl elongation ($GI_B > GI_A = GI_C$; Maloof, Brock and Weinig unpublished data) differs from our observed *GI*-fitness pattern ($GI_C > GI_A = GI_B$), suggesting the fitness effects are not mediated through seedling growth. The *GI* protein is also an integral component of the circadian clock; *GI* mutants differ from wild-type plants in the expression of other circadian-controlled loci (e.g. *CIRCADIAN CLOCK ASSOCIATED1*), leaf movement, and transpiration (Schaffer *et al*. 1998; Fowler *et al*. 1999; Sothern *et al*. 2002), and *GI* is a QTL for natural variation in circadian rhythm (Edwards *et al*. 2005). Because circadian rhythms can affect a plant's ability to match the timing of biological process (e.g. carbon fixation, transpiration) to critical daily events (e.g. sunrise; Dodd *et al*. 2005), the effect of *GI* on fruit set may be mediated by its effects on circadian rhythm. *GI* has also been shown to influence leaf starch accumulation (Eimert *et al*. 1995), which could influence plant size and, ultimately, fruit production. It is noteworthy that the effect of *GI* on starch accumulation is not mediated through its effect on flowering time (Eimert *et al*. 1995); therefore the absence of *GI* effect on flowering time does not preclude a role for this locus in regulating other phenotypic traits and fitness. Currently, we are completing additional association studies and quantitative complementation analyses to explore possible *GI* effects on circadian rhythm and leaf starch accumulation.

Regardless of whether *GI* or a tightly linked locus accounts for the observed fitness differences among haplogroups, the question arises as to how variation in this region is maintained in populations of *A. thaliana*. One common hypothesis is that spatial heterogeneity may contribute to the maintenance of genetic variation (Levene 1953; Lively 1986; Gillespie *et al*. 1989), e.g. that alternative *GI* haplogroups are most well-adapted to different microsites. We did not detect a significant *GI* × density treatment interaction, suggesting that variation in plant density does not help maintain variation at *GI*. Nor did we detect a significant latitudinal cline in *GI*, suggesting that the *GI* haplogroups are not differentially adapted to seasonal or climatic conditions that vary with latitude (e.g. day length, winter temperature, etc.). We cannot exclude, however, the possibility that adaptation to other spatially variable factors acts to maintain variation at *GI*.

Our results do, however, support the role of another mechanism hypothesized to maintain allelic variation, namely epistasis. When genetic backgrounds are mixed by migration among populations, epistasis for fitness can limit the fixation of allelic variants at specific loci, resulting in higher levels of standing variation (Wade 2001). Our analyses of epistasis revealed six significant *GI* × SNP interactions for fitness. This should be considered a minimum estimate

of epistasis, because low $GI \times$ SNP group size limited comparisons to only 84 SNP markers with only GI_A and GI_C haplogroups. Despite these limitations, the effect of *GI* haplogroup on fruit set either reversed across SNP backgrounds or was 'conditionally neutral' (i.e. allelic substitutions at *GI* affected fitness only in one of two possible SNP backgrounds; Fig. 3). As a largely selfing species, rare outcrossing events in *A. thaliana* may bring together favourable combinations of alleles at *GI* and background loci, which are subsequently maintained by generations of selfing. From an alternative viewpoint, the infrequency of outcrossing may limit the occurrence of GI_C in backgrounds that result in differences in fruit set and in exposure of this locus to selection. Such epistatic interactions may facilitate the maintenance of variation at *GI* among natural populations, the level at which molecular polymorphisms are usually assessed in *A. thaliana*. These findings are particularly interesting in the light of the growing number of studies demonstrating the importance of epistatic variance to quantitative trait variation in *A. thaliana* (Borevitz *et al*. 2002; Weinig *et al*. 2003; Juenger *et al*. 2005; Malmberg *et al*. 2005; Korves *et al*. 2007).

Acknowledgements

We thank four anonymous reviewers, H. Smith, and D. Moeller for their insightful comments which greatly improved this manuscript, N. Schult and J. Stringer for assistance in collecting the sequence data, O. Savolainen for providing seeds of *Arabidopsis lyrata*, J. Maloof for discussions about the molecular genetics of phytochrome signalling in *Arabidopsis thaliana*, and L. Hartweck for discussions about *GIGANTEA*. We are also grateful to the National Science Foundation for grants (DEB-0235027 and IBN-0417094 to P.T. and DBI-0227103 to C.W.) that support work in our laboratories.

References

Aguade M (2001) Nucleotide sequence variation at two genes in the phenylpropanoid pathway, the *FAH1* and *F3H* genes. *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **18**, 1–9.

Aranzana MJ, Kim S, Zhao K *et al*. (2005) Genome–wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *Public Library of Science Genetics*, **1**, 531–539.

Aukerman MJ, Hirschfeld M, Wester L *et al*. (1997) A deletion in the *PHYD* gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. *Plant Cell*, **9**, 1317–1326.

Ballaré CL, Scopel AL, Sánchez RA (1990) Far-red radiation reflected from adjacent leaves: an early signal of competition in plant canopies. *Science*, **247**, 329–332.

Borevitz JO, Maloof JN, Trainer G *et al*. (2002) QTL controlling light and hormone responses in two accessions of *Arabidopsis thaliana*. *Genetics*, **160**, 683–696.

Botto JF, Smith H (2002) Differential genetic variation in adaptive strategies to a common environmental signal in *Arabidopsis* accessions: phytochrome-mediated shade avoidance. *Plant Cell Environment*, **25**, 57–67.

Cao S, Ye M, Jiang S (2005) Involvement of *GIGANTEA* gene in the regulation of the cold stress response in *Arabidopsis*. *Plant Cell Reports*, **24**, 683–690.

Cao S, Jiang S, Zhang R (2006) The role of *GIGANTEA* gene in mediating the oxidative stress response and in *Arabidopsis*. *Plant Growth Regulation*, **48**, 261–270.

Cardon LR, Bell JI (2001) Association study designs for complex diseases. *Nature Reviews Genetics*, **2**, 91–99.

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet*, **361**, 598–604.

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

Devlin PF, Robson PR, Patel SR *et al*. (1999) Phytochrome *D* acts in the shade–avoidance syndrome in *Arabidopsis* by controlling elongation growth and flowering time. *Plant Physiology*, **119**, 909–915.

Dodd AN, Salathia N, Hall A *et al*. (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science*, **309**, 630–633.

Donohue K, Polisetty CR, Wender NJ (2005) Genetic basis and consequences of niche construction: plasticity-induced genetic constraints on the evolution of seed dispersal in *Arabidopsis thaliana*. *American Naturalist*, **165**, 537–550.

Donohue K, Schmitt J (1999) The genetic architecture of plasticity to density in *Impatiens capensis*. *Evolution*, **53**, 1377–1386.

Dudley SA, Schmitt J (1996) Testing the adaptive plasticity hypothesis: density-dependent selection on manipulated stem length in *Impatiens capensis*. *American Naturalist*, **147**, 445–465.

Edwards KD, Lynn JR, Gyula P, Nagy F, Millar AJ (2005) Natural allelic variation in the temperature-compensation mechanisms of the *Arabidopsis thaliana* circadian clock. *Genetics*, **170**, 387–400.

Eimert K, Wang S-M, Lue W-L, Chen J (1995) Monogenic recessive mutations causing both late floral initiation and excess starch accumulation in *Arabidopsis*. *Plant Cell*, **7**, 1703–1712.

Falconer DS, Mackay TFC (1997) *Introduction to Quantitative Genetics*. Addison Wesley Longman, Essex, UK.

Fankhauser C, Yeh K-C, Lagarias JC *et al*. (1999) *PKS1*, a substrate phosphorylated by phytochrome that modulates light signaling in *Arabidopsis*. *Science*, **284**, 1539.

Felsenstein J (1976) The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, **10**, 253–280.

Fenster CB, Galloway LF, Chao L (1997) Epistasis and its consequences for the evolution of natural populations. *Trends in Ecology & Evolution*, **12**, 282–286.

Fowler S, Lee K, Onouchi H *et al*. (1999) *GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *The EMBO Journal*, **18**, 4679–4688.

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Gillespie JH, Turelli M (1989) Genotype–environment interactions and the maintenance of polygenic variation. *Genetics*, **121**, 129–138.

Gould PD, Locke JCW, Larue C *et al*. (2006) The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. *Plant Cell*, **18**, 1177–1187.

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hanfstingl U, Berry A, Kellogg EA *et al*. (1994) Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol deydrogenase locus: roles for both balancing and directional selection? *Genetics*, **138**, 811–828.

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences, USA*, **89**, 10915–10919.

Hodge A (2004) The plastic plant: root responses to heterogeneous supplies of nutrients. *New Phytologist*, **162**, 9–24.

Huq E, Quail PH (2002) PIF4, a phytochrome-interacting BHLH factor, functions as a negative regulator of phytochrome B signaling in *Arabidopsis*. *EMBO*, **21**, 2441–2450.

Huq E, Tepperman JM, Quail PH (2000) *GIGANTEA* is a nuclear protein involved in phytochrome signaling in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA*, **97**, 9789–9794.

Juenger TE, Sen S, Stowe KA, Simms EL (2005) Epistasis and genotype–environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica*, **123**, 87–105.

Kawabe A, Innan H, Terauchi R, Miyashita NT (1997) Nucleotide polymorphism in the acidic chitinase locus (ChiA) region of the wild plant *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **14**, 1303–1315.

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

Korves TM, Schmid KJ, Caicedo AL *et al.* (2007) Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *American Naturalist*, **169**, E141–E157.

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

Lariguet P, Boccalandro HE, Alonso J *et al.* (2003) A growth regulatory loop that provides homeostasis to phytochrome signaling. *Plant Cell*, **15**, 2966–2976.

Leips J, Mackay TFC (2000) Quantitative trait loci for life span in *Drosophila melanogaster*: interactions with genetic background and larval density. *Genetics*, **155**, 1773–1788.

Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *American Naturalist*, **87**, 331–333.

Lively CM (1986) Canalization vs. developmental conversion in a spatially variable environment. *American Naturalist*, **128**, 561–572.

Lynch M, Walsh B (1997) *Genetics and the Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, Massachusetts.

Malamy JE (2005) Intrinsic and environmental response pathways that regulate root system architecture. *Plant Cell and Environment*, **28**, 67–77.

Malmberg RL, Held S, Waits A, Mauricio R (2005) Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics*, **171**, 2013–2027.

Maloof JN, Borevitz JO, Dabi T *et al.* (2001) Natural variation in light sensitivity of *Arabidopsis*. *Nature Genetics*, **29**, 441–446.

Mauricio RMD, Rausher Burdick DS (1997) Variation in the defense strategies of plants: are resistance and tolerance mutually exclusive? *Ecology*, **78**, 1301–1311.

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

Ni M, Tepperman JM, Quail PH (1998) PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell*, **95**, 657–667.

Ni M, Tepperman JM, Quail PH (1999) Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature*, **400**, 781–784.

Nordborg M, Hu TT, Ishino Y, *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *Public Library of Science Biology*, **3**, 1289–1299.

Olsen JE, Junttila O, Nilsen J *et al.* (1997) Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. *Plant Journal*, **12**, 1339–1350.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Yu J, Pressoir G, Briggs WH *et al.* (2006) A unified mixed–model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, **38**, 203–208.

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

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. *Journal of Human Genetics*, **67**, 170–181.

Purugganan MD, Suddith JI (1999) Molecular population genetics of floral homeotic loci. Departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. *Genetics*, **151**, 839–848.

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

Schaffer R, Ramsay N, Samach A *et al.* (1998) The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell*, **93**, 1219–1229.

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.

Schmitt J, Wulff RD (1993) Light spectral quality, phytochrome and plant competition. *Trends in Ecology & Evolution*, **8**, 47–51.

Schmitt J, McCormac AC, Smith H (1995) A test of the adaptive plasticity hypothesis using transgenic and mutant plants disabled in phytochrome-mediated elongation responses to neighbors. *American Naturalist*, **146**, 937–953.

Sharrock RA, Quail PH (1989) Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes and Development*, **3**, 1745–1757.

Shook DR, Johnson TE (1999) Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype–environment interactions, pleiotropy and epistasis. *Genetics*, **153**, 1233–1243.

Smith H (1982) Light quality, photoreception, and plant strategy. *Annual Review of Plant Physiology*, **33**, 481–518.

Smith H (2000) Phytochromes and light signal perception by plants — an emerging synthesis. *Nature*, **407**, 585–591.

Smith H, Whitelam GC (1990) Phytochrome, a family of photoreceptors with multiple physiological roles. *Plant, Cell and Environment*, **13**, 695–707.

Soh MS, Song P-S, Choi G (1999) Phytochrome signalling is mediated through nucleoside diphosphate kinase 2. *Nature*, **401**, 610–613.

Sothern RB, Tseng T-S, Orcutt SL, Olszewski NE, Koukkari WL (2002) *GIGANTEA* and *SPINDLY* genes linked to the clock pathway that controls circadian characteristics of transpiration in *Arabidopsis*. *Chronobiology International*, **19**, 1005–1022.

Steindler C, Matteucci A, Sessa G *et al.* (1999) Shade avoidance responses are mediated by the HD-zip protein, a negative regulator of gene expression. *Development*, **126**, 4235–.

Stinchcombe JR, Weinig C, Ungerer M *et al.* (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by

the flowering time gene. *FRIGIDA*. *Proceedings of the National Academy of Sciences, USA*, **101**, 4712–4717.
Suarez-Lopez P, Wheatley K, Robson F *et al.* (2001) CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature*, **410**, 1116–1120.
Swofford DL (2002) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–596.
Templeton AR (1995) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apoprotein E locus. *Genetics*, **140**, 403–409.
Thornsberry JM, Goodman MM, Doebley J *et al.* (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nature Genetics*, **28**, 286–289.
Tseng T-S, Salome PA, McClung CR, Olszewski NE (2004) SPINDLY and GIGANTEA interact and act in *Arabidopsis thaliana* pathways involved in light responses, flowering, and rhythms in cotyledon movements. *Plant Cell*, **16**, 1550–1563.
Van Lijsebettens M, Vanderhaeghen R, De Block M *et al.* (1994) An S18 ribosomal protein gene copy at the *Arabidopsis* PFL locus affects plant development by its specific expression in meristems. *EMBO*, **13**, 3378–3388.
Wade MJ (2001) Epistasis, complex traits, and mapping genes. *Genetica*, **112–113**, 59–69.
Watterson GA (1975) On the number of segregation sites. *Theoretical Population Biology*, **7**, 256–276.
Weinig C (2000) Differing selection in alternative competitive environments: shade-avoidance responses and germination timing. *Evolution*, **54**, 124–136.
Weinig C, Dorn LA, Kane NC, *et al.* (2003) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics*, **165**, 321–329.
Weinig C, Johnston JA, Willis C, Maloof JN (2007) Antagonistic multilevel selection on size and architecture in variable density settings. *Evolution*, **61**, 58–67.
Westerman JM, Lawrence MJ (1970) Genotype–environment interaction and developmental regulation in *Arabidopsis thaliana*. I. Inbred lines; description. *Heredity*, **25**, 609–627.
Whitelam GC, Devlin PF (1997) Roles of different phytochromes in *Arabidopsis* photomorphogenesis. *Plant, Cell and Environment*, **20**, 752–758.
Yanovsky M, Kay SA (2002) Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature*, **419**, 308–313.
Yoshida K, Kamiya T, Kawabe A, Miyashita NT (2003) DNA polymorphism at the ACAULIS5 locus of the wild plant *Arabidopsis thaliana*. *Genes and Genetic Systems*, **78**, 11–21.

Marc Brock is a Post-doctoral researcher working in collaboration with Cynthia Weinig to explore the quantitative genetics of plant competitive and shade-avoidance responses in both segregating progenies and natural populations. Peter Tiffin uses molecular population genetic analyses and experiments in contemporary populations to investigate biotic interactions, species range expansion, and evolutionary responses to global climate change.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Maximum-likelihood phylogram of *GIGANTEA* sequences from 21 accessions of *Arabidopsis thaliana*. Bootstrap values > 70 as well as haplogroups, designated GI_A, GI_B, and GI_C, used to test for haplogroup-phenotype associations are shown. Nodes with less than 50% bootstrap support were collapsed, the *Arabidopsis lyrata* sequence was designated as an outgroup.

Fig. S2 Average backtransformed days to flower, backtransformed proportion petiole length, and inflorescence height of *Arabidopsis thaliana* accessions in *GI* haplogroups (GI_A N = 45, GI_B N = 34, and GI_C N = 82) that were subjected to uncrowded and crowded treatments in a greenhouse experiment. *GI* haplogroup did not explain significant amounts of variation in any of the three traits. Error bars indicate 95% confidence intervals.

Table S1 Sequence of forward (F) and reverse (R) PCR primers, information on the region and proportion of total coding length for each gene we sequenced, and the GenBank Accession numbers of the sequences from which we obtained information on the intron/exon structure. All PCR programs were 35 cycles at 94 °C for 1 min, 50 °C or 55 °C for 1 min, and 72 °C for 2 min, 1 M Betaine was added to all reactions. *GI* X, *GI* Y, and *GI* Z, refer to three sets of primers that amplified overlapping regions, used to obtain the full length *GI* sequence

Table S2 Primers and associated annealing temperatures used to genotype 161 *Arabidopsis* accessions for *GIGANTEA* haplogroup. Extraction and amplification were performed with Sigma's Extract-N-Amp plant PCR kit. PCR programs were 37 cycles at 94 °C for 30 sec, Annealing for 1 min, and 72 °C for 1.5 min extension. Restriction digests were resolved on 4% SFR agarose gels (Solon, Ohio, USA)

Table S3 List of *Arabidopsis thaliana* accessions used in a greenhouse experiment testing association between the genetic locus, *GIGANTEA*, and the following phenotypes: fruit set, days to flower, inflorescence height, and proportion petiole length (petiole length/total leaf length)

Table S4 Analysis of covariance in three phenotypic traits of 161 *Arabidopsis thaliana* accessions testing the effects of competition treatment (uncrowded and crowded treatments) and *GIGANTEA* haplogroup. Estimates of relatedness to predicted historical populations (spop1, spop2, … spop5) were included to remove potential confounding effects of population structure

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03298.x
(This link will take you to the article abstract).

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