An assessment of transgenomics as a tool for identifying genes involved in the evolutionary differentiation of closely related plant species

Raul Correa[1], **John Stanga**[1], **Bret Larget**[1,2], **Aaron Roznowski**[1], **Guoping Shu**[3], **Brian Dilkes**[4], and **David A. Baum**[1]

[1]University of Wisconsin-Madison, Department of Botany, 430 Lincoln Dr, Madison WI 53706, USA

[2]University of Wisconsin-Madison, Department of Statistics, 1300 University Ave, Madison WI 53706, USA

[3]Harvard University, Department of Organismic and Evolutionary Biology, 26 Oxford St, Cambridge MA 02138, USA

[4]Purdue University, Department of Horticulture and Landscape Architecture, 625 Agriculture Mall Dr, West Lafayette IN 47907, USA

Summary

- Transgenomics is the process of introducing genomic clones from a donor species into a recipient species and then screening the resultant transgenic lines for phenotypes of interest. This method might allow us to find genes involved in the evolution of phenotypic differences between species as well as genes that have the potential to contribute to reproductive isolation: potential speciation genes.

- More than 1,100 ~20 Kbp genomic clones from *Leavenworthia alabamica* were moved into *Arabidopsis thaliana* by transformation. After screening a single primary transformant for each line, clones associated with mutant phenotypes were tested for repeatability and cosegregation.

- We found 84 clones with possible phenotypic effects of which eight were repeatedly associated with the same phenotype. One clone, 11_11B, cosegregated with a short fruit phenotype. Further study showed that 11_11B affects seed development, with as much as one-third of the seeds aborted in some fruit.

- Transgenomics is a viable strategy for discovering genes of evolutionary interest. We identify methods to reduce false positives and false negatives in the future. 11_11B can

be viewed as a potential speciation gene, illustrating the value of transgenomics for studying the molecular basis of reproductive isolation.

Introduction

Several studies have identified genes that contributed to the differentiation of distinct species (e.g., Sucena & Stern, 2000; Busch & Zachgo, 2007; Werner *et al.*, 2010; Yamaguchi *et al.*, 2010). Such studies most commonly use developmental genetic data from model species to suggest candidate genes for phenotypes that differ between a pair of species. The most definitive tests of candidate gene hypotheses involve interspecies transformation experiments in which the candidate gene is introduced from a donor into a recipient species to see if it impacts the phenotype of interest (e.g., Wheeler *et al.*, 1991; Nasrallah *et al.*, 2002; Hanikenne *et al.*, 2008; Jeong *et al.*, 2008; Kimura *et al.*, 2008; Enard *et al.*, 2009). While the candidate-gene approach has been quite successful in elucidating the molecular genetic basis of species difference, it is inherently biased towards well-studied genes and cannot readily discover genes of previously unknown function.

A second commonly used approach for identifying species difference genes begins with quantitative trait locus (QTL) mapping and proceeds to clone genes using a variety of strategies (e.g., Bradshaw & Schemske, 2003; Briggs *et al.*, 2007; Hovav *et al.*, 2007; Lowry *et al.*, 2009). QTL approaches are not biased against uncharacterized genes, but they are limited to species pairs that can be hybridized to yield fertile offspring. Additionally, it often proves impractical to clone the causal gene underlying a QTL, especially when working with non-model species (Bradshaw & Schemske, 2003; Bowman, 2006).

Here, we utilized a third strategy, transgenomics, that has the potential to identify genes that have contributed to the phenotypic differentiation and/or reproductive incompatibility of even quite divergent species pairs. This method involves transforming plants of a recipient species with random genomic clones from a donor species and then screening the transformants for phenotypic effects. Transgenomics has complementary strengths and weaknesses to candidate gene and QTL methods. Like QTL analysis, it can identify causal genomic regions that would not be predicted based on prior work in model species. At the same time, like a candidate gene approach, transgenomics is not restricted to crossable species and leads directly to experimentation at the molecular level. Transgenomics therefore has the potential to be an important addition to the toolbox of the evolutionary developmental geneticist.

Transgenomics was proposed as a theoretical possibility almost a decade ago (Baum, 2002), but has not yet been implemented because of the technical challenge of conducting even a pilot screen. In this study, we assessed the viability of the approach by introducing large numbers of genomic DNA clones from Alabama gladecress (*Leavenworthia alabamica*) into *Arabidopsis thaliana*. We found eight clones that associate with repeatable phenotypic effects, representing diverse aspects of plant form and development. One of these clones, 11_11B, was shown through cosegregation analysis to cause seed abortion and shortened fruit. Based on this analysis we conclude that 11_11B has a dominant negative effect on plant fitness in an *A. thaliana* background. This shows that the causal gene within 11_11B has the potential to contribute to reproductive isolation between incipient species.

Materials and Methods

Donor species library construction

We partially digested total high-molecular weight DNA from *L. alabamica* Rollins with *Taq*I and fractionated the products on a 10–40% sucrose gradient. ~20–25 Kb fragments were ligated into the *Cla*I site of the binary cosmid vector pCLD04541 (Jones *et al.*, 1992). This vector (28 Kb) is based upon pRK290 and carries *cos* from bacteriophage lambda which allows highly efficient *in vitro* packaging. Additionally, it carries a polylinker with blue/white selection, T-DNA border sequences for insertion in the plant genome, and a kanamycin resistance gene (*NPTII*) for selection of transgenic plants. We packaged ligated products with the Gigapack III kit (Stratagene) and introduced the cosmid DNA library into XL1-Blue MR *E. coli* host cells. This library was determined to contain ~8.5×10^5 colony-forming units (CFU), which translates to ~34.5-fold coverage of the ~500 Mb *L. alabamica* genome. Aliquots of this library are available upon request to researchers interested in additional transgenomic screening.

96-well plate archives

We plated *Leavenworthia alabamica* clones in *E. coli* on 2X YT media with tetracycline selection (10 µg ml^{-1}). We picked single colonies of various sizes to reduce bias against clones that slow bacterial growth. Each colony was transferred into a different well of a 96-well plate containing 150 µl 2X YT media with 10 µg ml^{-1} tetracycline, and was grown overnight. 1.3 µl of each culture was used to inoculate l.3 ml 2X YT in 96 square-well block format. The remaining culture was mixed with 100 µl of 60% glycerol and frozen for long-term storage. We used the R.E.A.L. Prep 96 kit (Qiagen) to isolate cosmid DNA from cultures grown in 96 square-well format. We froze 16 µL of the purified cosmid for long-term storage in a 96-well plate, and used 4 µl of the DNA to transform 20 µl of freshly grown *Agrobacterium tumefaciens* strain GV3101::pMP90 using a high-throughput freeze-thaw transformation protocol (Weigel & Glazebrook, 2006) in 96-well format. Immediately following transformations, we mixed each *Agrobacterium* sample with 96 µl of liquid LB without selection and allowed 4 h of recovery with gentle rocking. We then further grew each *Agrobacterium* sample with an additional 60 µl of liquid LB with final concentrations of 75 µg ml^{-1} kanamycin, 50 µg ml^{-1} rifampicin, and 25 µg ml^{-1} gentamycin. We used kanamycin instead of tetracycline for selection because GV3101::pMP90 is capable of evolving spontaneous tetracycline resistance. Following growth, we mixed each *Agrobacterium* culture with 70 µl of 60% glycerol and froze it for long-term storage. We named each *L. alabamica* clone according to the well and 96-well archive in which it was stored. For example, the *L. alabamica* clone stored in the well in column 8 and row H of the fifth 96-well archive was given the name '05_08H'. Replicates of the fourteen 96-well plates of arrayed *E. coli* clones are available from the authors upon request.

Plant transformation

We grew three to five wildtype *Arabidopsis. thaliana* (L.) Columbia (T0) plants in a single 3.5" SVD pot (T.O. Plastics, Clearwater, MN, USA). We transformed up to 192 T0 pots (two plates' worth) per wk. Upon bolting of T0 plants, we clipped their primary shoots to induce early flowering of lateral shoots. We transferred 40 µl of each *Agrobacterium* culture from a 96-well archive to a 50 ml conical tube containing 10 ml of liquid LB with 75 µg ml^{-1} kanamycin, 50 µg ml^{-1} rifampicin, and 25 µg ml^{-1} gentamycin. These 10 ml *Agrobacterium* cultures were grown for 2 d at 28°C with 200 rpm shaking. Cells were pelleted and resuspended in 8 ml of infiltration medium (sucrose 50 g l^{-1}, MS salt 2 g l^{-1}, MES 0.4 g l^{-1}, 1.3 M N KOH, silwet 200 µl l^{-1}). We then dripped infiltration cultures onto the inflorescences of T0 plants. This protocol is based on Clough & Bent (1998) but is better suited to transgenomics because of the small volume of *Agrobacterium* and the relative

speed of dripping bacteria onto plants rather than dipping plants into bacterial cultures. After inoculation, we shaded T0 plants for 16 h and watered them for 3 wk before allowing them to dry. We harvested the seed of T0 plants in one pot and pooled the seed to form a T1 seed stock made with a single *L. alabamica* clone. All plants were grown under standard glasshouse conditions (16 h light at 22.5°C : 8 h dark at 18°C, 300–500 mmol m^{-2} s^{-1}, and air-conditioning) at the University of Wisconsin-Madison Biotron Facility.

Phenotypic screening of primary transformants

We applied a portion of each T1 seed stock to one kanamycin media plate (kanamycin 75 mg l^{-1}, MS salt 2.2 g l^{-1}, MES 0.4 g l^{-1}, agar 8.5 g l^{-1}, 0.8 mM KOH). We incubated plates at 4°C for 3 d to synchronize germination. Following cold treatment, we grew plates for 11 d in standard glasshouse conditions under shade cloth (50–100 mmol m^{-2} s^{-1}) to avoid excessive humidity inside plates. At the end of this period, we transferred one or two T1 seedlings per plate to a 3.25" SVD pot of soil with the goal of allowing one T1 plant to develop. T1 plants were visually screened during development and at maturity for various morphological phenotypes that were not seen in untransformed plants grown in parallel. We screened T1 plants for changes in plant architecture, stature, rosette and cauline leaf morphology, flowering time (indicated by the approximate number of total leaves on each plant), and reproductive traits such as flower and fruit morphology, including cases of low fertility as indicated by few or stunted fruit on plants. We screened flowers in the mornings when petals were open and observed the other traits throughout the day. Seed was harvested from all T1 plants (regardless of their phenotype) and those that were not used up in subsequent experiments are available from the authors upon request.

For each plant in which we identified a phenotypic abnormality, we grew four additional independent T1s and screened them for the phenotype. If phenotypic effects recurred in at least one of the four additional T1s, then further T1 plants were isolated to see if each phenotype continued to repeat. A 'repeatable phenotype' was defined as one that recurred in at least three T1 plants.

Cosegregation analysis

To infer the number of insert loci for each T1 line, we sowed approximately 500 T2 seeds on kanamycin medium. Seedlings were allowed to grow with 16 h light (~100 mmol m^{-2} s^{-1}) at 22.5°C : 8 h dark at 18°C. After 2 wk, we counted numbers of kanamycin resistant and sensitive offspring on each plate. A 99% confidence interval of the observed kanamycin resistant-to-sensitive (KRS) ratio was used to assess whether it was consistent with Mendelian expectations for 1 or 2 segregating loci. For those lines with a significant deficit of kanamycin resistant seedlings we also assessed segregation in T2 plants genotyped by PCR amplification of the *NPTII* gene using the primers (5'–3'): ATCCCATGGCTGATGCAATGCG and CCATGATATTCGGCAAGCAGGCAT.

For each repeatable clone, we selected on line whose segregation ratio suggested a single insert locus, namely lines for which the 99% confidence interval around the observed KRS included 3:1 (see Supporting Information Table S1). When there were multiple such lines for a clone, we chose the line with the clearest phenotypic effect in the T1 generation. For one clone (06_05C), all lines screened deviated significantly from a 3:1 ratio so we picked the line that was closest to a 3:1 ratio.

About 50 T2 plants from each selected line were grown in individual 3.5" pots. Each T2 family was grown to maturity for 42–56 d in standard glasshouse conditions. We then scored families for phenotypic variation. The length of fruit valves or entire fruit was determined by

scanning them at 4000 dpi on an Epson perfection V700 Photo scanner and using ImageJ 1.43u software (Rasband) to measure the length.

To obtain T2 genotypes, we harvested T3 seed produced by each T2 plant. About 100 seeds of each T3 plant were screened on kanamycin plates. The KRS ratios of T3 plates were then used to infer the transgene genotype of each T2 plant. In the family for clone 06_05C, we could not confidently genotype T2 plants due to deviations from Mendelian ratios, suggesting *trans*-gene silencing. Therefore, we scored T2 plants as 'transgene present' if they produced any resistant seedlings and 'transgene absent' if they produced only sensitive T3 seedlings.

Statistical analysis

For continuous phenotypes, we used ANOVA to assess whether there was a significant association between transgene genotype and phenotype. We used a one-way ANOVA for the initial round of experiments that tested one line per clone. When studying the four independent lines from clone 11_11B, we used a mixed-effects ANOVA model that explained fruit length variation with fixed effects for individual line, genotype, an interaction between them, location of fruit within the inflorescence (categorized with separate categories for the lowest five locations and one category for all other locations), an interaction between fruit location and line, and a random effect for individual plant. Categorization of fruit location was based on graphical data exploration. We considered models with additional interaction terms, but found none that significantly improved the fit to the data. We used the lme4 package in R to conduct the mixed-effects ANOVA analysis. ANOVA *P*-values < 0.05 were taken to indicate significant cosegregation. For clone 06_05C, cosegregation was assessed with a two-tailed homoscedastic Student's *t*-Test.

Results

Identifying appropriate model species for transgenomics

Transgenomics depends on being able to easily move donor species' genomic DNA into a recipient species. While reliable transformation methods exist for many species, we chose the model plant *A. thaliana* because of its great transformation efficiency and the immense body of available genetic and genomic data. An appropriate donor species would be phylogenetically close (i.e., within Brassicaceae), yet divergent from *A. thaliana* for as many distinct phenotypes as possible. *Leavenworthia* is a member of the Cardamineae clade, which is within 'Lineage I' of Brassicaceae, the same lineage that includes *Arabidopsis* (Beilstein *et al.*, 2006). The radiation of Lineage I Brassicaceae has been dated at 8–14 Ma (Franzke *et al.*, 2009) or 18–36 Ma (Couvreur *et al.*, 2010) or 35±6 Ma (Beilstein *et al.*, 2010). *Leavenworthia* and *Arabidopsis* differ in almost all visible traits that show discrete variation within Brassicaceae (Fig. 1), including leaf shape (pinnately compound vs simple), trichomes (absent vs present), shoot architecture (rosette flowering vs inflorescence flowering), flower size (>1cm diam. vs < 3mm diam.), fruit shape (strongly flattened vs terete), seed number per fruit (~9–10 vs ~50–70) and seed size (~3.1 mm vs ~0.5 mm). *Leavenworthia alabamica* has a modest genome size of *c*. 500 Mbp (M. A. Lysak & P. Bures, unpublished), 11 chromosomes (Lysak *et al.*, 2009), and is the target of an ongoing genome-sequencing project (http://biology.mcgill.ca/vegi/index.html). *A. thaliana* can be transformed with *L. alabamica* genomic clones in a high-throughput manner

As a first step in the transgenomic screen, we generated a plant transformation-competent genomic library of *Leavenworthia alabamica*, with inserts of 20–25 Kb. We conducted pilot analyses of alternative screening strategies and concluded that the optimal was a high-throughput clone-by-clone screening approach (Fig. 2). This method has the advantage that

one can readily determine whether a clone generates the same abnormal phenotype in multiple independent transformants. Because each independent transformant is likely to have a unique insertion site, one can avoid wasting effort following-up phenotypes that result from insertional mutagenesis or other artifacts of transformation.

We used this clone-by-clone transgenomic pipeline to conduct an initial screen of 14 96-well archives. We found that this workflow could be completed by a small team, comprising one graduate student and two to three undergraduates, in < 6 mo. with a peak rate of plant transformation of ~160 clones (grown from two 96-well plates) per week. 85% of the *E. coli* clones were successfully transferred to *Agrobacterium* and 99% of those were successfully introduced into plants (see below).

Analysis of primary transformants yielded clones associated with repeatable phenoypic abnormalities

Because *Agrobacterium*-mediated transformation of *A. thaliana* occurs in the female gametophyte, and because T-DNA integration is non-homologous, T1 plants are hemizygous for the transgenic fragment. A transgene is only expected to manifest a visible phenotype in the T1 generation if it is *trans*-dominant, meaning that its effects are not masked by endogenous homologs in the genome. Our aim was to screen for clones associated with such *trans*-dominant phenotypic effects.

We used a small portion of each T1 seed stock to isolate one kanamycin resistant T1 plant per *L. alabamica* clone (Fig. 2c). Of 1,147 *Agrobacterium* clones, 1,134 were successfully introduced into *A. thaliana* plants. For each clone, we visually screened at least one T1 individual during development and at maturity (~6 wk of age) for morphological phenotypes that differed from untransformed plants. 84 clones (7.4%) yielded initial T1 plants that were judged to deviate from wildtype (Table 1).

An abnormal phenotype in a single T1 plant could result from a *trans*-dominant effect of an *L. alabamica* clone, insertional mutagenesis, background mutation, perturbation caused by screening on kanamycin plates and transplanting, or environmental variability. Of these, only effects that are attributable to inserted *L. alabamica* DNA are of interest. Because independent transformants of the same clone are likely to have integrated into the genome at unique insertion sites, phenotypes that appear repeatably are are most likely to be due to the *L. alabamica* insert. Therefore, we excluded 70 clones that failed to repeat the abnormal phenotype after screening four additional, independent T1 plants for the same clone (Fig. 2c). We then screen additional T1 plants and excluded two further clones because of a failure to generate additional abnormal T1 plants.

We discarded four clones that were found, based on end-sequencing, to include Lambda viral DNA, probably introduced during library construction. These four clones were associated with low repeatability and involved phenotypes that occur not infrequently in wildtype plants.

After this winnowing process, we were able to identify 8 clones containing *L. alabamica* genomic DNA that were associated with repeatable, *trans*-dominant phenotypes in *A. thaliana* (Table 2). The levels of repeatability, defined as the number of T1 plants with phenotype divided by the number of T1s screened, varied from 16% to 60%, with an average of ~42% (Table 2). This collection of clones was associated with a broad range of phenotypes in *A. thaliana*: two were flagged primarily due to a change in fruit size/shape, three for plant architecture/stature, one for flower form, one for a leaf defect, and one for abnormal leaf and flower development (Figs 3, S1, S2).

For ten independent T1 plants from clone 05_01C and nine from 11_11B we measured valve length for 20–40 fruit. This confirmed that they do, indeed, tend to have shorter fruit than wildtype (Table 3). Because data were only collected for transgenic lines that had been visually flagged as having short fruit, meaningful statistical comparisons could not be made until later generations. However, the dramatic differences between wildtype and several T1 plants for clones 05_01C and 11_11B are consistent with these clones having a transdominant effect on fruit length.

Cosegregation analysis

To identify lines with one transgene locus, we screened T2 seeds for four to eight lines per clone (Table S1). For seven of the eight clones, we found at least two lines whose KRS ratio is consistent with a single transgene locus. Of the 44 lines screened, 21 were inferred to have one transgene locus, and two to have two unlinked transgene loci (Table S1). Of the remaining 21 lines, 12 had a deficit of resistant plants beyond that expected even for a single transgene locus. To help assess the causes of non-Mendelian segregation patterns, we used PCR amplification of *NPTII* to genotype 23–40 T2 plants from those lines that showed a deficit of kanamycin resistant seedlings (Table S2). In two of twelve lines (12_03E_1; 12_06G_1) we detected a significant deficit of *NPTII*-containing offspring. This could indicate an effect of the *L. alabamica* transgene on gamete, gametophyte or embryo development, or it could be due to insertional mutagenesis or chromosomal rearrangements during transformation (e.g., see Ray *et al.*, 1997). In three lines (11_11B_2; 11_11B_15; 12_06G_2) we found a significant difference between the KRS and *NPTII* ratios, where the latter is not significantly different from 3:1. This is most easily explained by Mendelian segregation of the transgene coupled with silencing of *NPTII* in many plants. In the remaining lines the ratio determined by PCR is consistent with both the KRS ratio and 3:1.

We selected one candidate line from each *L. alabamica* clone, choosing the line with the strongest T1 phenotype that was inferred to contain a single transgene locus. Because we found no lines for clone 06_05C that showed a 3:1 ratio, we selected a line (06_05C_03) that produced both about half sensitive and half resistant T2 seedlings.

For each selected line, we grew a T2 family on soil so that we would be able to score plant phenotypes while being blind to genotype. We allowed T2 plants to self and set T3 seed. A portion of T3 seed from each T2 plant was tested on kanamycin media to infer its transgene genotype (see the Materials and Methods section).

ANOVA could not reject the null hypothesis that variation in the scored phenotype was independent of transgene genotype in six of eight T2 families (Fig. 4). In the T2 family for clone 06_05C a two-tailed homoscedastic Student's *t*-Test did not detect a significant difference in fruit length between transgene-present and transgene-absent T2 plants. However, only 22 T2 plants were scored in this case. Combined with the distorted segregation ratios seen in this line, we judge this cosegregation test to be non-informative.

For clone 11_11B, transgene-hemizygous T2 plants were found to have shorter fruit, while transgene-homozygous and transgene-absent T2 plants had fruit of normal length (one-way ANOVA $P = 0.00046$; Fig. 4e). This result suggests that the transgene has a *trans*-overdominant effect: shortening fruit only when the transgene locus is hemizygous, and not when it is homozygous.

Cosegregation analysis suggests that 11_11B causes short fruit and ovule/seed abortion

The conclusion that 11_11B cosegregates with fruit length variation came from a single T1 line. To determine if this result could be replicated, we conducted a more thorough study of the same line (11_11B_1) plus three additional, independent 11_11B lines (5, 6, and 10). All

lines had been scored with KRS ratios that were close to 3:1 (Table S1). We scanned the inflorescences and used image analysis software to measure fruit length for all mature fruits along the main axis. Genotypes were then inferred by sowing batches of T3 seed on kanamycin plates.

Controlling for variation that was attributable to line, individual plant, individual fruit, and location of fruit in the inflorescence, the null hypothesis that genotype has no effect on fruit length was rejected. Once again, a significant transgene overdominance effect on fruit length was observed (Fig. 5). However, this effect was only detectable in lines 1, 5, and 6. Within these three lines transgene-absent and transgene-homozygous plants had indistinguishable fruit sizes, whereas transgene-hemizygous fruit were, on average, 1.6 ± 0.7 mm, 3.1 ± 0.8 mm, or 3.7 ± 0.7 mm shorter than the average transgene-absent fruit, respectively. The mean fruit length in line 10 hemizygotes was indistinguishable from wildtype.

Since it is well established that normal fruit elongation requires the production of signals from developing seeds (Gillaspy *et al.*, 1993; Chaudhury *et al.*, 1997; Vivian-Smith *et al.*, 2001), the fruit-length effect could be caused by defects during ovule/megagametophyte development, fertilization, or early embryogenesis. We observed aborted seeds in a subset of T2 plants. Aborted seeds were recognizable as short funicles with a knob-like ending, which persisted even after seed dispersal, whereas normal funicles were long and tapering after seed abscission. We scored 1–4 mature, dehisced fruit per plant from at least five individual plants for each genotypic class, for lines 1, 5, 6, and 10, all while blind to plant genotype.

As summarized in Table 4, plants carrying the 11_11B transgene showed significantly higher levels of seed abortion than wildtype plants in the same three lines (1, 5, and 6) that showed a fruit length effect. Whereas < 5% of seeds typically aborted in wildtype lines, the mean abortion rate in lines 1, 5, and 6 hemizygotes was 12%, 34%, and 31%, respectively. However, in contrast to the complete overdominance seen for fruit length, significantly elevated seed abortion was also seen in some transgene homozygotes. Examination of abortion rates for individual plants from lines 1, 5, and 6 suggested that among the transgene-containing lines there was a extensive variation in seed abortion rates (Fig. S3).

The aborted seeds seen in 11_11B-containing plants could be due to disruption of processes such as ovule development and pollen maturation in parental tissues. Alternatively, the transgene could act directly on gametophytes or embryos. In the latter case, we might expect the seeds of hemizygotes to show a deficit of transgenes homozygotes and hemizygotes because they would abort at a higher rate than wildtype gametophytes/embryos. However, for all four lines tested for cosegregation, the ratio of inferred wildtype, transgene-hemizygous, and transgene-homozygous T2 plants did not deviate significantly from 1:2:1. This corroborates that these four T1 lines each have a single transgene locus and suggests that embryos carrying the transgene do not have an elevated abortion rate.

Discussion

Transgenomic screens are feasible and can be improved to reduce the false negative and false positive rate

Our study succeeded in introducing 1,147 clones from an *L. alabamica* genomic library into *Agrobacterium*. For all but the 13 clones that did not yield transformant plants, we examined at least one primary transformant. Assuming a genome size for *L. alabamica* of 500 Mb, and an average insert size of 20 Kb, our study achieved just over 4.5% coverage of the *L. alabamica* genome. While this is still a small fraction of the donor genome, considering that the initial screen took < 6 mo. and was conducted by one graduate student with two to three undergraduate assistants, it is possible for a transgenomic screen to achieve 1X or higher

coverage of the donor genome. Even with a team twice the size as ours, and provided that the efforts were well-coordinated and that growth space were not limited, it should be possible to generate transformants for 2,000 clones per mo. With potential technical improvements, such as larger clone inserts, it could become feasible for even small research groups to undertake complete transgenomic screens of various donor species in an *A. thaliana* genetic background. Further, as transformation technology improves for many species, including more and more cases of drip/dip transformation (e.g., Bartholmes *et al.*, 2008); it may become possible to use recipient species other than *A. thaliana*.

In addition to generating a large number of transgenomic lines that can be investigated further, our research provides a baseline from which to optimize future transgenomic screens. While this study illustrates the feasibility of transgenomics, it also highlights some challenges. For example, we found a large number of false positives. Whereas 7.4% of the initial transgenic plants showed some deviation from wildtype morphology, the great majority did not reproduce that phenotype in additional lines. Most of these anomalous phenotypes are presumed to be caused by insertional mutagenesis, stress during kanamycin screening and transplantation, fluctuations in growth conditions, or disruption of T1 seedling development by *Agrobacterium*-mediated drip transformation of T0 plants.

We believe that growing a larger initial number of T1 plants per clone would be likely to reduce the number of false positives (and false negatives, see below). This would enhance the ability to quickly identify repeatable phenotypes and would not greatly slow down the screen. Thus, we predict that a significant reduction in the false positive rate can be achieved.

Of the 84 clones tested for repeatability, eight were capable of generating four or more independent T1 plants that shared a particular morphological effect (Table 2). However, these clones also yielded as many or more normal T1 plants (Table 2). If *L. alabamica* clones were responsible for the phenotypes, then potential explanations for low repeatability include sensitivity to position of insertion and/or transgene silencing. We demonstrated silencing of the selectable marker gene *NPTII* in some lines (Table S2), so it is reasonable to assume that silencing could also affect a phenotype-altering transgene derived from the donor species. Alternatively, repeatable phenotypes could reflect parental background mutations or non-genetic effects that recurred in plants in similar microenvironments.

The low repeatability associated with the eight clones suggests that by screening just one T1 plant per clone, we were likely to have experienced a considerable false negative rate. The one clone that was confirmed as inducing a phenotypic effect had a T1 repeatability of 56%. Assuming this is typical, we should expect to miss almost half of the phenotype-altering clones when conducting a screen based on only one initial T1 plant. The simplest way to reduce this false negative rate would be to screen several T1 plants per clone rather than just one.

Seven out of eight repeatable phenotypes are likely the result of background mutations in T0 plants or reflect environmental variability

Our finding that six of the eight repeatable phenotypes did not cosegregate with the transgene does not necessarily rule out the possibility that the transgenes do explain the observed T1 phenotypes. The lack of cosegregation could be due to a confounding factor such as post-transcriptional transgene silencing (Morel *et al.*, 2000) or environmental effects that prevented phenotype manifestation in T2 plants. To be sure that no lines are prematurely discarded, future cosegregation analyses should consider additional T2 lines from the same clones.

One possible explanation for the non-cosegregation of repeatable clones is that the repeatable phenotypes reflect genetic abnormalities of the T0 parental plants. Supposing that one or two of the T0 plants carried a background mutation or had a genotype that rendered it and its offspring prone to express a phenotypic abnormality, one could expect the phenotype to recur in T1 plants at a greater frequency than in the population as a whole. If one allows for uneven contributions of T0 plants to the T1 seed stock, then such an effect can readily explain traits occurring at a sufficient frequency to pass our test for repeatability (Table 2).

Understanding the potential sources of false positives, we can suggest methodological refinements for future transgenomic screens. For example, a phenotype due to a homozygous background mutation in a T0 plant should be inherited by all offspring of that plant, while the *L. alabamica* transgene can only occur in T1 plants that are resistant to kanamycin. Therefore, transgenomic screens should be able to distinguish potentially transgene-induced phenotypes from background effffects by growing several T1s from the same clone – half taken from a kanamycin plate and half from a kanamycin-lacking plate (the vast majority of plants from the latter plate will lack the transgene). Alternatively, T0 plants could be harvested separately and several T0 plants could be transformed. This would allow identification of repeatable phenotypes that are independent on T0 parent and, therefore, likely to be caused by the transgene.

L. alabamica clone 11_11B insert contains a transgene with overdominant effects on *A. thaliana* reproductive development

For three of the four 11_11B lines, the transgene genotype correlated with an overdominant phenotypic effect. Hemizygous fruit were significantly shorter than wildtype, whereas transgene homozygous fruit resembled wildtype. Likewise, mean seed abortion was highest in hemizygotes. Especially in lines 5 and 6, transgene homozygous plants showed elevated levels of seed abortion that were closer to hemizygous plants than to wildtype.

Overdominance is not uncommon in transgenic experiments (Nap *et al.*, 1997; Damgaard & Borksted, 2004). The most plausible hypothesis to explain overdominance of 11_11B is that transgene (but not *NPTII*) is silenced more frequently in homozygous lines than in hemizygous lines. This is suggested by the seed abortion data, which appears to show a bimodal distribution of seed abortion rates in the presence of the transgene (Fig. S3). The distribution of seed abortion rates (Fig. S3) suggests that the difference in the mean seed abortion rate between transgene homozygous and hemizygous plants may be due to differences in the probability of gene silencing rather than a difference in the severity of the phenotype when the transgene escapes silencing.

Taken together our data show that the 11_11B insert contains a gene that has a phenotypic effect when introduced as an extra copy into the genome of *A. thaliana*. There are two main ways that a transgene from one species can cause a phenotype when added to a foreign genome. First, if the causal transgene is similar to an endogenous gene in the recipient genome, the addition of an extra 1 or 2 allelic copies could cause a phenotypic abnormality by altering gene dosage. For example, the presence of the extra gene copy could elevate the expression level of a gene-product resulting in a gain-of-function phenotype or, conversely, it could trigger silencing of the endogenous homologs resulting in a loss-of-function phenotype. Second, coding or cis-regulatory sequences may have diverged from the orthologous gene in the recipient genome, resulting in novel biochemical functions or expression patterns. The genetic pattern of overdominance in 11_11B provides circumstantial evidence against an effect that is entirely due to gene dosage. Plausible dosage-based models do not predict overdominance, but either additivity, if each extra gene copy incrementally increases the effect, or simple dominance, if there is a threshold beyond which extra copies have no additional effect. This leads us to believe that our preliminary

transgenomic screen has succeeded in identifying at least one phenotype-altering gene whose effect is due to sequence evolution since *A. thaliana* and *L. alabamica* last shared a common ancestor.

The 11_11B clone causes a reduction in seed set of A. thaliana and this phenotypic effect is unlikely to be a dosage effect. Thus, it is reasonable to suppose that if an F1 hybrid were formed between a plant carrying the causal *L. alabamica* gene and a plant carrying interacting genes from *A. thaliana*, then these F1 hybrids would show reduced fertility. Thus, the causal sequence underlying the effect of 11_11B is a candidate for being a *potential* speciation gene, where a potential speciation gene is one that has the potential to undergo mutations that could contribute to hybrid infertility between the incipient species. We are not claiming that the 11_11B causal sequences played a role in the speciation event that gave rise (eventually) to the *Arabidopsis* and *Leavenworthia* lineages. Given how long ago the common ancestor lived, this is actually quite unlikely. Rather, we are suggesting that if the analogous changes were to occur in incipient species, the gene could contribute to their hybrid infertility. This shows that, in the same way that we can learn about the molecular basis of reproductive isolation by studying genes that cause reproductive isolation between genotypes within a single species (Bomblies *et al.*, 2007; Bomblies & Weigel, 2007, 2010), transgenomics can help identify genes that have the potential to participate in speciation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bartholmes C, Nutt P, Theissen G. Germline transformation of Shepherd's purse (*Capsella bursa-pastoris*) by the 'floral dip' method as a tool for evolutionary and developmental biology. Gene. 2008; 409:11–19. [PubMed: 18164559]

Baum, DA. Identifying the genetic causes of phenotypic evolution: a review of experimental strategies. In: Cronk, QCB.; Bateman, R.; Hawkins, JA., editors. Developmental genetics and plant evolution. London, UK: Taylor & Francis; 2002. p. 493-507.

Beilstein M, Al-Shehbaz I, Kellogg E. Brassicaceae phylogeny and trichome evolution. American Journal of Botany. 2006; 93:607–619. [PubMed: 21646222]

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences. 2010; 107:18724–18728.

Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PloS Biology. 2007; 5:1962–1972.

Bomblies K, Weigel D. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nature Reviews Genetics. 2007; 8:382–393.

Bomblies K, Weigel D. *Arabidopsis* and relatives as models for the study of genetic and genomic incompatibilities. Philosophical Transactions of the Royal Society B: Biological Sciences. 2010; 365:1815–1823.

Bowman JL. Molecules and morphology: comparative developmental genetics of the Brassicaceae. Plant Systematics and Evolution. 2006; 259:199–215.

Bradshaw HD, Schemske DW. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature. 2003; 426:176–178. [PubMed: 14614505]

Briggs WH, McMullen MD, Gaut BS, Doebley J. Linkage mapping of domestication loci in a large Maize-Teosinte backcross resource. Genetics. 2007; 177:1915–1928. [PubMed: 17947434]

Busch A, Zachgo S. Control of corolla monosymmetry in the Brassicaceae *Iberis amara*. Proceedings of the National Academy of Sciences. 2007; 104:16714–16719.

Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ. Fertilization-independent seed development in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences. 1997; 94:4223–4228.

Clough S, Bent A. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant Journal. 1998; 16:735–743. [PubMed: 10069079]

Couvreur TLP, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, Mummenhoff K. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Molecular Biology and Evolution. 2010; 27:55–71. [PubMed: 19744998]

Damgaard D, Borksted B. Transgenic insect resistant *Arabidopsis* may show chaotic population dynamic. Ecological Complexity. 2004; 1:261–265.

Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, Bruckner MK, Schreiweis C, Winter C, Sohr R, et al. A humanized version of *Foxp2* affects cortico-basal ganglia circuits in mice. Cell. 2009; 137:961–971. [PubMed: 19490899]

Franzke A, German D, Al-Shehbaz IA, Mummenhoff K. *Arabidopsis* family ties: molecular phylogeny and age estimates in Brassicaceae. Taxon. 2009; 58:425–437.

Gillaspy G, Bendavid H, Gruissem W. Fruits: a developmental perspective. Plant Cell. 1993; 5:1439–1451. [PubMed: 12271039]

Hanikenne M, Talke IN, Haydon MJ, Lanz C, Nolte A, Motte P, Kroymann J, Weigel D, Kraemer U. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of *HMA4*. Nature. 2008; 453:391–U344. [PubMed: 18425111]

Hovav R, Chehanovsky N, Moy M, Jetter R, Schaffer AA. The identification of a gene (*Cwp1*), silenced during *Solanum* evolution, which causes cuticle microfissuring and dehydration when expressed in tomato fruit. Plant Journal. 2007; 52:627–639. [PubMed: 17877702]

Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. Cell. 2008; 132:783–793. [PubMed: 18329365]

Jones JD, Shlumukov L, Carland F, English J, Scofield SR, Bishop GJ, Harrison K. Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. Transgenic Research. 1992; 1:285–297. [PubMed: 1338696]

Kimura S, Koenig D, Kang J, Yoong FY, Sinha N. Natural variation in leaf morphology results from mutation of a novel *KNOX* gene. Current Biology. 2008; 18:672–677. [PubMed: 18424140]

Lexer C, Lai Z, Rieseberg L. Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. New Phytologist. 2004; 161:255–233.

Lowry DB, Hall MC, Salt DE, Willis JH. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. New Phytologist. 2009; 183:776–788. [PubMed: 19549130]

Lysak MA, Koch MA, Beaulieu JM, Meister A, Leitch IJ. The dynamic ups and downs of genome size evolution in Brassicaceae. Molecular Biology and Evolution. 2009; 26:85–98. [PubMed: 18842687]

Morel J, Mourrain P, Beclin C, Vaucheret H. DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis*. Current Biology. 2000; 10:1591–1594. [PubMed: 11137011]

Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel J, Jouette D, Lacombe A, Nikic S, Picault N, et al. *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. Cell. 2000; 101:533–542. [PubMed: 10850495]

Nap J, Conner A, Mlynarova L, Stiekema W, Jansen R. Dissection of a synthesized quantitative trait to characterize transgene interactions. Genetics. 1997; 147:315–320. [PubMed: 9286691]

Nasrallah M, Liu P, Nasrallah J. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two *S* locus genes from *A. lyrata*. Science. 2002; 297:247–249. [PubMed: 12114625]

Rasband, WS. Image J. Bethesda, Maryland, USA: National Institutes of Health; 1997–2011. http://imagej.nih.gov/ij/.

Ray SM, Park SS, Ray A. Pollen tube guidance by the female gametophyte. Development. 1997; 124:2489–2498. [PubMed: 9199374]

Sucena E, Stern DL. Divergence of larval morphology between Drosophila sechellia and its sibling species caused by cis-regulatory evolution of *ovo/shaven-baby*. Proceedings of the National Academy of Sciences. 2002; 97:4530–4534.

Vivian-Smith A, Luo M, Chaudhury A, Koltunow A. Fruit development is actively restricted in the absence of fertilization in *Arabidopsis*. Development. 2001; 128:2321–2331. [PubMed: 11493551]

Weigel D, Glazebrook J. Transformation of Agrobacterium using the freeze-thaw method. Cold Spring Harbor Protocols. 2006

Werner T, Koshikawa S, Williams TM, Carroll SB. Generation of a novel wing colour pattern by the Wingless morphogen. Nature. 2010; 464:U1143–U1157.

Wheeler DA, Kyriacou CP, Greenacre ML, Yu Q, Rutila JE, Rosbash M, Hall JC. Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. Science. 1991; 251:1082–1085. [PubMed: 1900131]

Yamaguchi T, Yano S, Tsukaya H. Genetic framework for flattened leaf blade formation in unifacial leaves of *Juncus prismatocarpus*. Plant Cell. 2010; 22:2141–2155. [PubMed: 20647346]
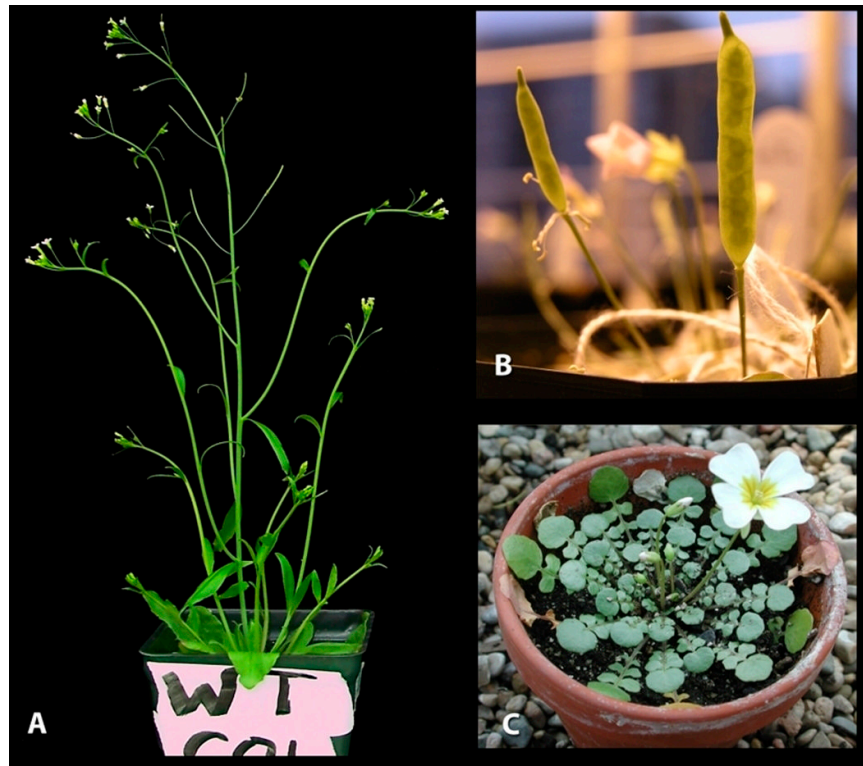
Fig. 1.
Comparison of *Arabidopsis thaliana* Columbia and *Leavenworthia alabamica*. Despite belonging to the same plant family (Brassicaceae), these species differ in many phenotypes. For example, *A. thaliana* Columbia (a) has simple leaves, an elongated primary inflorescence, short floral branches (pedicels), small flowers, narrow fruit, and an early flowering time in standard glasshouse conditions. By contrast, *L. alabamica* (b, c) has pinnately compound leaves, a suppressed primary inflorescence, elongated pedicels, large flowers, broader fruit, and a late flowering time in standard glasshouse conditions. Photos of *L. alabamica* kindly provided by J. Busch.
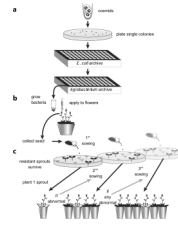
Fig. 2.
The clone-by-clone transgenomic strategy used to move *Leavenworthia alabamica* genomic DNA into *Arabidopsis thaliana* for phenotypic screening. (a) *L. alabamica* clones in *E. coli* are transferred 96-well format. Cosmids harboring *L. alabamica* inserts are then moved into *Agrobacterium tumefaciens* using high-throughput freeze thaw transformation. (b) A portion of each *Agrobacterium* stock in the 96-well plate is grown, then used to drip transform one pot of *A. thaliana* plants. Each T0 pot is then separately harvested for seed to identify primary transformant (T1) plants. (c) A portion of each T1 seed stock is then grown on kanamycin-containing plates. At least one kanamycin-resistant T1 plant is then transferred to a pot and allowed to grow to maturity. If the TI plant is abnormal, the transformed seed collection is sampled again to identify four additional independent T1 plants to see if the abnormal phenotype recurs. If the phenotype recurs, then a third seed sowing may be used to identify additional T1 plants.
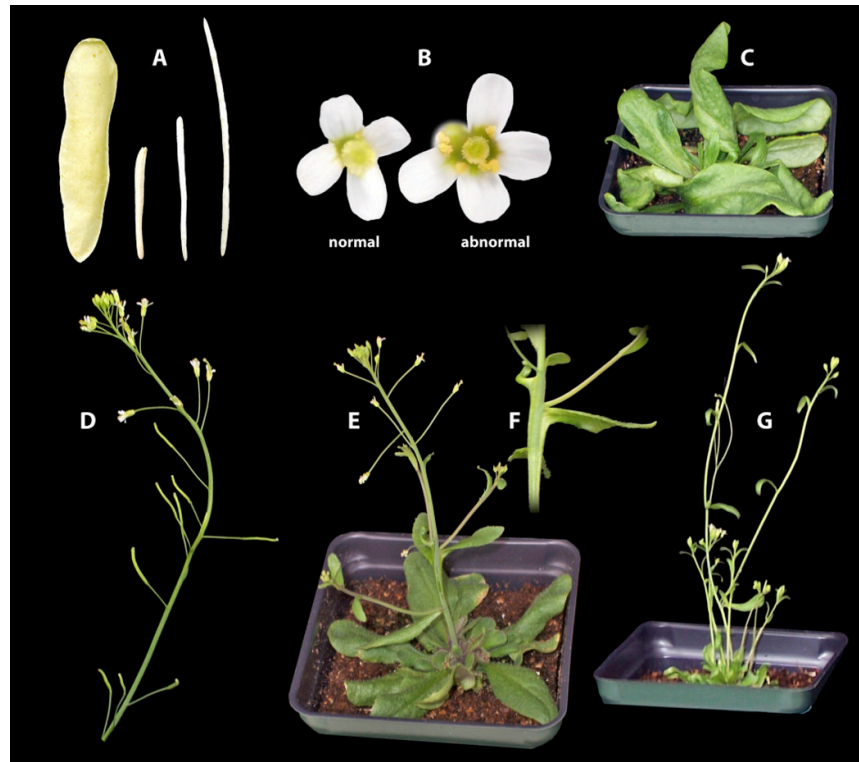
Fig. 3.
Repeatable morphological phenotypes identified in the transgenomic screen. Photographs of additional independent T1s with repeatable phenotypes are shown in Supporting Information Figs S1 and S2. (a) Fruits of *Leavenworthia alabamica* (left) and *Arabidopsis thaliana* (right) differ in their length-to-width ratio. The two stunted valves in the center were taken from two independent T1s from clone 11_11B and are representative of fruit from each plant. Clone 05_01C also yielded a repeatable short fruit phenotype. (b) Petals from *A. thaliana* flowers (left) are more evenly spaced than those from clone 09_09A (right). Although a small proportion of flowers on even wildtype plant have unevenly spaced petals, repeatable T1 plants from this clone were identified as having an elevated frequency of such flowers. (c) Abnormal twisting and contortion of rosette leaves on a T1 from clone 11_01D. Clone 11_01D also yielded a T1 plant with some lobed rosette leaves and two T1 plants with abnormal flowers (see Fig. S2). (d) Inflorescence with clustered fruit on a T1 plant from clone 12_03E. (e) A T1 plant from clone 12_05A displays cauline leaves developed decurrently along the primary axis. (f) Close-up picture of decurrent cauline leaf from the T1 of 12_05A (E). (g) A dwarf T1 plant from clone 12_06G.
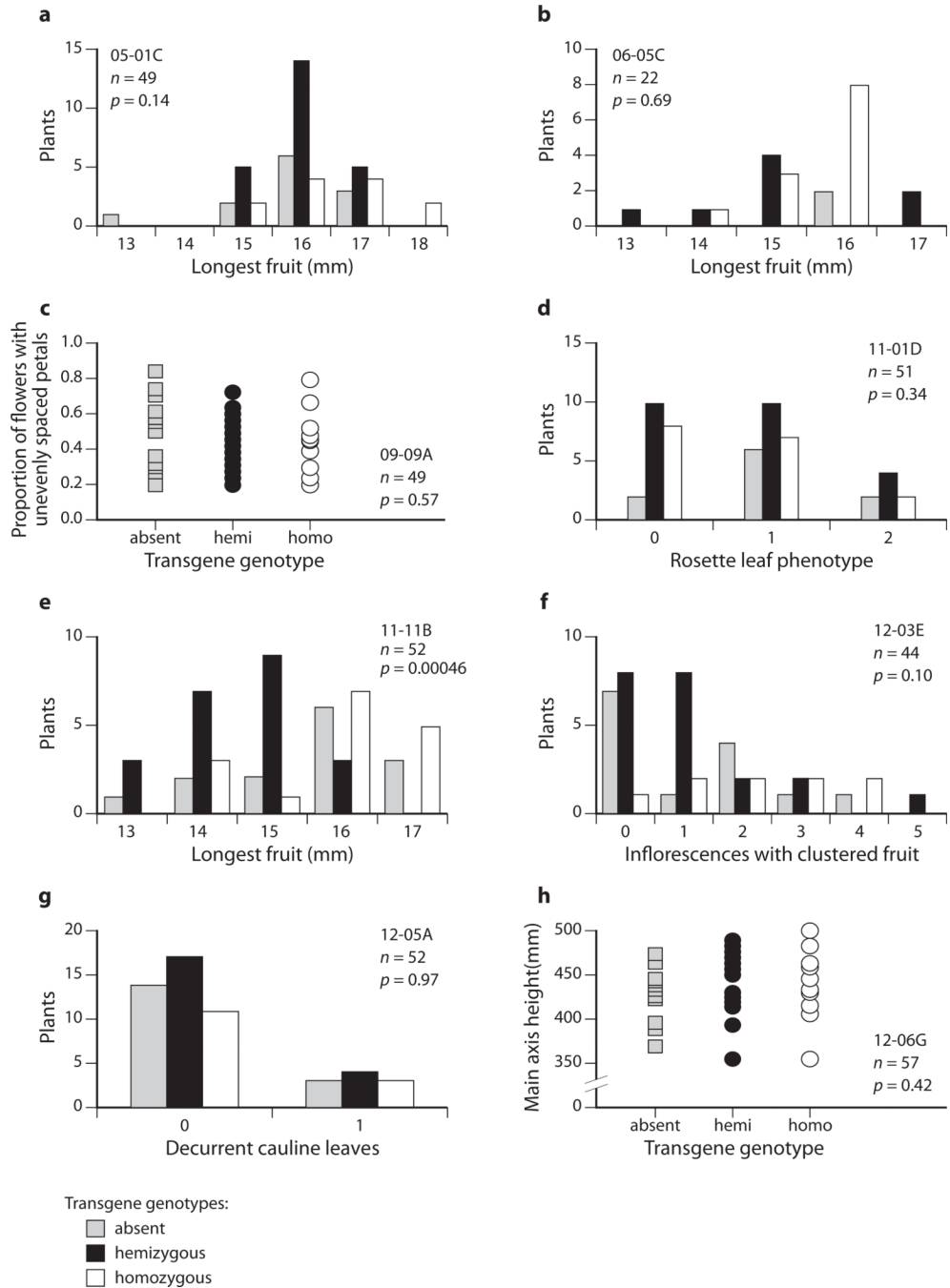
Fig. 4.

Cosegregation analyses of T2 families. Graphs (a–h) show phenotype distributions by genotype in T2 families. Each T2 family was derived from a T1 line that had shown a repeatable phenotype and was inferred to contain one transgene locus, except for 06_05C, which produced a smaller proportion of kanamycin resistant offspring than expected from a single-locus insertion line (see the Results section). Clone name, sample size of T2 family (*n*), and *P*-value result of a one-way ANOVA which tested genotypes for differences in phenotype are indicated in each graph. T2 plants in the family from clone 11_01D (D) were scored qualitatively for rosette leaf twisting as follows: 0, three or fewer leaves with slight twisting; 1, four or more leaves with slight twisting; 2, four or more with strong twisting.

Seven of eight cosegregation experiments, (a–d, f–h), did not produce a significant result. However the family derived from clone 11_11B (e) did reject the null hypothesis because hemizygous T2 plants had shorter fruit than either wildtype or transgene homozygote plants (corroborated by experiments shown in Fig. 5 and Supporting Information Fig. S3). Transgene genotypes: grey, absent; black, hemizygous; white, homozygous.
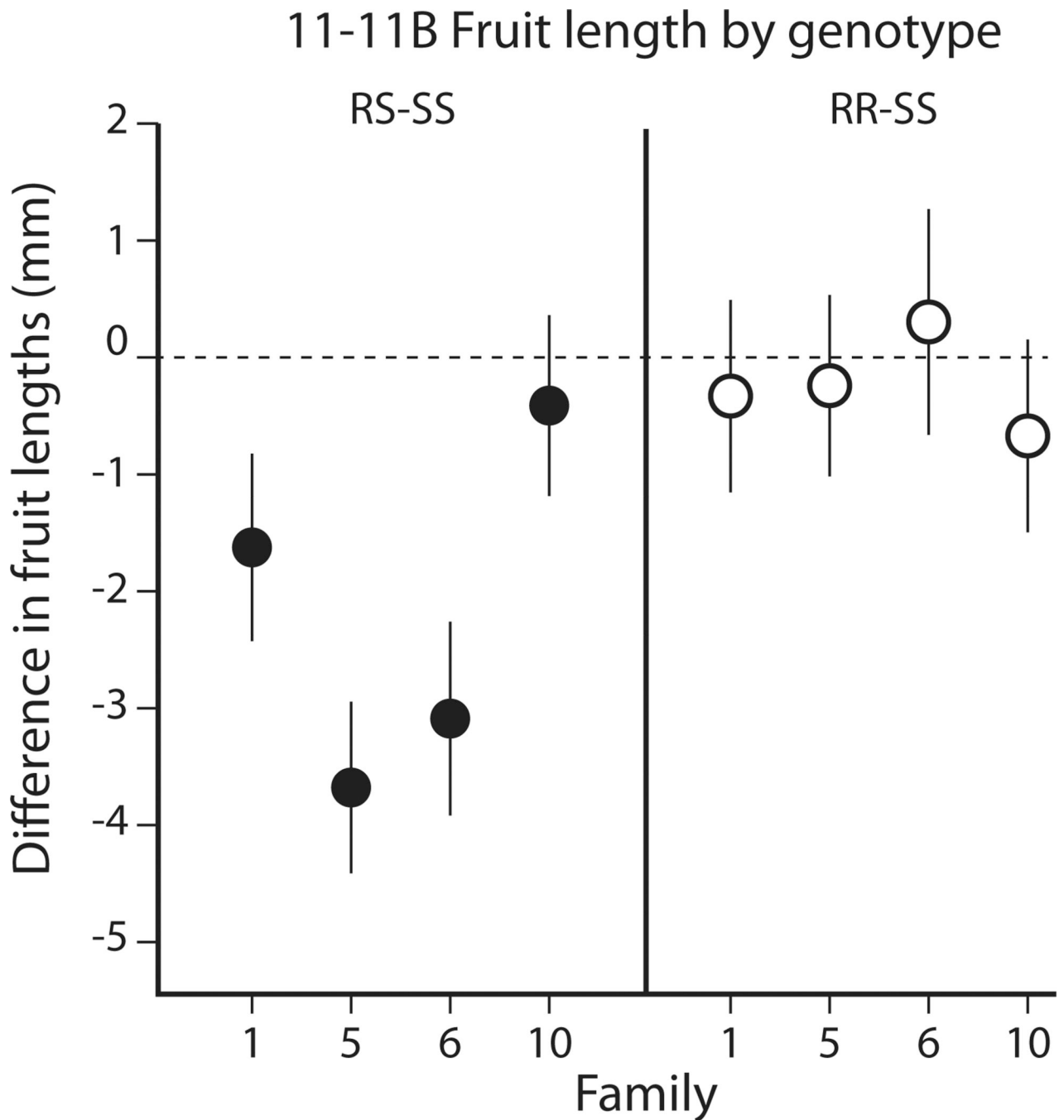
Fig. 5.
Cosegregation analysis of four independent T2 families verifies the overdominant short fruit effect of clone 11_11B. The plot compares transgene hemizygous (left panel) and transgene homozygous (right panel) mean fruit lengths versus transgene-absent mean fruit length. The mean fruit sizes are plotted with 95% confidence intervals. The four estimates per panel are from cosegregation data generated by four independent T2 families. Of the four independent families (mean $n = 56$ plants), three repeated the transgene overdominant effect observed for clone 11_11B in Fig. 4. For all four lines, the ratio of inferred wildtype, transgene-hemizygous, and transgene-homozygous T2 plants did not deviate significantly from 1:2:1, corroborating that these four T1 lines each have a single transgene locus.

Table 1

Effects of 1,147 *L. alabamica* clones

Clone effect	Frequency	Percent
No visible effect	1,050	91.5%
Non-repeatable phenotype	76	6.6%
Repeatable phenotype	8	0.7%
No transformants	13	1.1%
Total	1,147	---

Table 2

Clones assigned with repeatable phenotypes

Clone	Phenotype	Repeatability (T1s with phenotype/T1s screened)
05_01C	Short fruit	13/27 (48%)
06_05C	Short plants with reduced seed set	9/15 (60%)
09_09A	Petals unevenly spaced	5/16 (31%)
11_01D	Rosette leaves twisted	12/35 (34%)
11_11B	Short fruit	9/16 (56%)
12_03E	Fruit cluster with very reduced internodes	5/14 (36%)
12_05A	Cauline leaves decurrent to stem	4/7 (57%)
12_06G	Dwarfism	8/50 (16%)

Table 3

Fruit valve lengths

Line	# T1 plants	# valves per plant	Valve length (mean in mm ± SE)		
			Plant with shortest fruit	Plant with longest fruit	Grand mean
Wildtype	18	40	11.3 ± 1.0	14.2 ± 0.7	12.7 ± 0.2
05_01C	10	20	7.4 ± 2.5	13.5 ± 0.8	10.9 ± 4.5
11_11B	9	40	4.9 ± 0.6	12.5 ± 0.5	9.3 ± 0.3

Table 4

Proportion of seed aborted (mean ± SD) in the three genotypic classes in four independent 11_11B lines

	Transgenic		
T1 line	**Wildtype**	**Hemizygous**	**Homozygous**
11_11B_01	0.03±0.03	0.12±0.04	0.10±0.07
11_11B_05	0.03±0.02	0.34±0.14	0.27±0.19
11_11B_06	0.05±0.03	0.31±0.09	0.27±0.15
11_11B_10	0.04±0.02	0.03±0.04	0.06±0.02