



Evolutionary transgenomics: prospects and challenges

Raul Correa¹ and David A. Baum^{2*}

¹ Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA, ² Department of Botany, University of Wisconsin-Madison, Madison, WI, USA

Many advances in our understanding of the genetic basis of species differences have arisen from transformation experiments, which allow us to study the effect of genes from one species (the donor) when placed in the genetic background of another species (the recipient). Such interspecies transformation experiments are usually focused on candidate genes - genes that, based on work in model systems, are suspected to be responsible for certain phenotypic differences between the donor and recipient species. We suggest that the high efficiency of transformation in a few plant species, most notably Arabidopsis thaliana, combined with the small size of typical plant genes and their cisregulatory regions allow implementation of a screening strategy that does not depend upon a priori candidate gene identification. This approach, transgenomics, entails moving many large genomic inserts of a donor species into the wild type background of a recipient species and then screening for dominant phenotypic effects. As a proof of concept, we recently conducted a transgenomic screen that analyzed more than 1100 random, large genomic inserts of the Alabama gladecress Leavenworthia alabamica for dominant phenotypic effects in the A. thaliana background. This screen identified one insert that shortens fruit and decreases A. thaliana fertility. In this paper we discuss the principles of transgenomic screens and suggest methods to help minimize the frequencies of false positive and false negative results. We argue that, because transgenomics avoids committing in advance to candidate genes it has the potential to help us identify truly novel genes or cryptic functions of known genes. Given the valuable knowledge that is likely to be gained, we believe the time is ripe for the plant evolutionary community to invest in transgenomic screens, at least in the mustard family Brassicaceae where many species are amenable to efficient transformation.

Keywords: developmental system drift, evolution, evo-devo, genetic screens, speciation genes, transgenomics, transformation

INTRODUCTION

At its most general, evolutionary developmental biology, "evo-devo," seeks to understand how development, the translation of a genotype to a phenotype in a given environment, constrains, or enables phenotypic evolution (Stern, 2000; Wagner et al., 2000; Arthur, 2002; Carroll et al., 2004; Lee et al., 2014). The core data that are needed to achieve such understanding are genetic and developmental changes that have been shown experimentally to cause particular evolutionary transitions from ancestral to derived phenotypes. While much evo-devo research can focus on phenotypic variation within living populations, there has long been an interest in also studying

OPEN ACCESS

Edited by:

Verónica S. Di Stilio, University of Washington, USA

Reviewed by:

Marcelo Carnier Dornelas, Universidade Estadual de Campinas, Brazil William Oki Wong, Institute of Botany, Chinese Academy of Sciences, China

> *Correspondence: David A. Baum dbaum@wisc.edu

Specialty section:

This article was submitted to Plant Evolution and Development, a section of the journal Frontiers in Plant Science

Received: 01 August 2015 Accepted: 28 September 2015 Published: 20 October 2015

Citation:

Correa R and Baum DA (2015) Evolutionary transgenomics: prospects and challenges. Front. Plant Sci. 6:858. doi: 10.3389/fpls.2015.00858

1

characters that differ between living species, where one species manifests the ancestral character state and the other manifests the derived. How, then, can we experimentally determine the genetic and developmental basis of traits that differ between species?

Until now, the search for genes responsible for species differences has mainly exploited either candidate gene or quantitative trait locus (QTL) approaches. Candidate gene methods use information from genetic model systems to hypothesize that a certain genetic change caused the transition from the ancestral to the derived character state and then set about to test that hypothesis using comparative studies of two (or more) species that differ for the trait. The test usually involves comparative expression studies combined with various functional studies, which might include knocking down gene expression and/or moving the candidate gene among species either by crossing or transgenic methods. However, candidate gene approaches are limited to phenotypes whose developmental basis is well understood in model genetic systems. While much of what we know about evo-devo comes from candidate gene studies, they have limitations. In particular, when a phenotype is caused by unpredictable genes, as can happen due to neofunctionalization [e.g., (Vlad et al., 2014)], candidate gene approaches will come-up empty. Indeed, using only a candidate gene approach would make it very hard to answer one of the key questions in evo-devo: how many different genetic pathways are available for the evolution of a new trait?

Currently, the main alternative to candidate gene approaches is QTL analysis. This involves crossing two species with contrasting traits and then looking for cosegregation of the trait with genetic markers in the F2 or later generations. The goal is to positionally clone genes causing phenotypic differences and, eventually, home in on the sequence difference causing the trait difference. However, QTL analysis is limited to cases where species are capable of being crossed. Further, it is notoriously difficult to clone the gene underlying a QTL in non-model species.

In this paper we will argue that evolutionary transgenomics (Baum, 2002) represents a third method, complementary to the other two, that could be used to identify the genes responsible for species' differences. An evolutionary transgenomic screen involves introducing genomic fragments of a donor species into the genome of a recipient species (or perhaps a divergent ecotype) and screening the resulting transgenic lines for phenotypic effects. Such screens have the advantage of not being limited to crossable species and yet being able to find genes that would not have been predicted *a priori*. We will suggest that while transgenomics poses practical challenges, it has great potential value for plant evo-devo research in those taxa that are readily transformed and may help us identify many genes of evolutionary and developmental interest that would otherwise be difficult to discover.

Systematic screens in yeast, including screens of plant cDNA libraries, have successfully been used to identify genes controlling cell-level phenotypes [e.g., (Papoyan and Kochian, 2004; Liu et al., 2007)]. However, there have been few attempts to screen the genome of one multicellular eukaryote in that of another. Because there are many plants that are closely related, show

many phenotypic differences, have compact genes, and are readily transformed, plants are better suited to transgenomics than are most animals. But are the benefits to be gained likely to outweigh the work entailed in conducting evolutionary transgenomic experiments? In this paper we use evolutionary theory and data from prior interspecies transformation experiments and our published, pilot transgenomic screen (Correa et al., 2012) to assess the approach and how it could best be implemented. We conclude that the time is ripe to develop transgenomics resources at least in Brassicaceae Burnett, a clade of flowering plants, many of whose species can now be transformed with high efficiency.

PRINCIPLES OF TRANSGENOMICS

In an evolutionary transgenomic screen, fragments of genomic DNA from a donor species are added to the wildtype genome of a recipient species. We then screen primary transformants (T1s) to look for phenotypic effects that might be due to the inserted DNA. Since plant transformation usually entails inserting an extra piece of DNA rather than homologous replacement of endogenous sequences, inserts will only cause phenotypes in T1s if they act in a *transdominant* manner. That is to say, one copy of the foreign gene must manifest a phenotypic effect even in the presence of two functional copies of that gene (if any exist in the recipient genome). Genetic theory suggests two primary causes of a *transdominant* phenotype.

The first potential cause of a *transdominant* phenotypic effect is the addition of a supernumerary gene copy to the genome, a gene dosage effect. It is well documented that changes in gene dosage can have phenotypic effects [reviewed by (Birchler and Veitia, 2007)]. This is most obvious when aneuploids (e.g., trisomics) yield distinct phenotypes, including lethality or sterility. Dosage effects presumably result from additional copies altering the balance of expression of genes in regulatory pathways. Addition of a single additional gene (as would occur in a hemizygous T1 transgenic plant) might be expected to increase expression level by approximately 50%, or higher if multiple insertions of the transgene occur. However, the actual effect on expression will vary subject to position effects and whether or not transgene silencing is triggered.

It is not known how often dosage alone will yield a visible phenotype in a transgenic line, and this might vary depending on the phylogenetic distance between donor and recipient species. However, based on prior transgenic data it seems likely that many dosage effects will primarily be quantitative. For example, it has been found that adding extra *Arabidopsis thaliana* (L.) Heynh. *LFY* transgenes to a wildtype *A. thaliana* background has a dosage dependent effect on flowering time (Blazquez et al., 1997).

Dosage effects do not depend on sequence divergence between donor and recipient species. Quite the contrary – a dosage effect depends upon conservation of molecular function between the endogenous and exogenous gene copies. This leads to a powerful test for discriminating dosage effects from other mechanisms: if the phenotypic effect can be replicated by introducing the homologous fragment of the recipient species back into the

Evolutionary transgenomics

recipient species, then dosage is likely to be the cause of the observed phenotype. If, there is no homologous region, or if the homologous region fails to cause the phenotype, then dosage is unlikely to be responsible.

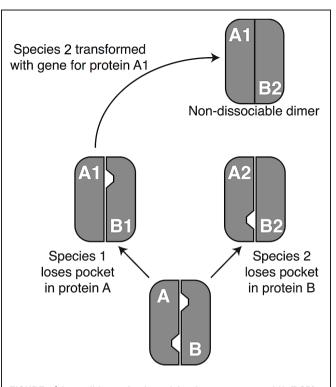
The second potential cause of a *transdominant* phenotype is evolutionary divergence between the donor and recipient genes. This could arise through one of two mechanisms, developmental system drift (DSD) or phenotypic divergence. DSD occurs when proteins and/or regulatory DNA/RNA sequences coevolve without altering visible phenotypes (True and Haag, 2001). The "drift" in underlying molecular mechanisms can cause a gene from the donor species to malfunction in the recipient genome in a such a way that a phenotype is seen, analogous to transgressive segregation, which is often seen in QTL studies (Rieseberg et al., 1999). To make the concept more concrete, Figure 1 shows a hypothetical example involving a protein with two subunits that must be disassembled for proper development, with disassembly requiring at least one "pocket" of low attraction between the two subunits. Reciprocal loss of the pocket in the two subunits could result in a case in which subunit A from species 1 yields a dominant-negative phenotypic effect when placed in the genome of species 2. It is worth noting the similarities between this DSD model and Dobzhansky-Muller interactions (Kondrashov et al., 2002; Bomblies et al., 2007). Indeed, one exciting aspect of transgenomics is its potential to identify potential hybrid inviability genes.

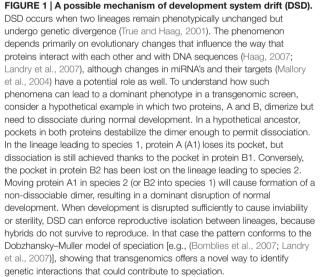
The alternative explanation of a *transdominant* phenotype is that phenotypic evolution has been driven by sequence evolution (whether in coding or regulatory regions) at genes of large effect. In this case, a gene can carry the donor species' phenotype into the recipient species' genome (**Figure 2**). This is expected to happen when a difference in phenotype is due to a fully or partially dominant mutation on the lineage leading to the donor species or to a recessive mutation on the lineage leading to the recipient species (**Figure 3**). Assuming that the relative frequency of dominant and recessive mutations is about equal on the two evolutionary lineages, first principles would suggest that a complete transgenomic screen would be able to detect about 50% of the major genes explaining phenotypic differences between the donor and recipient species.

If we knew how many of the phenotypic differences between species were due to genes of large effect we could predict the frequency with which transgenomic lines will manifest a phenotype that resembles the donor species. Unfortunately, we are not aware of any relevant quantitative data. Indeed, one of the most compelling reasons to conduct transgenomic screens is because they will help quantify the frequency of evolution via genes of large effect, something that has long been a source of controversy (Gottlieb, 1984; Doebley and Lukens, 1998; Hoekstra and Coyne, 2007).

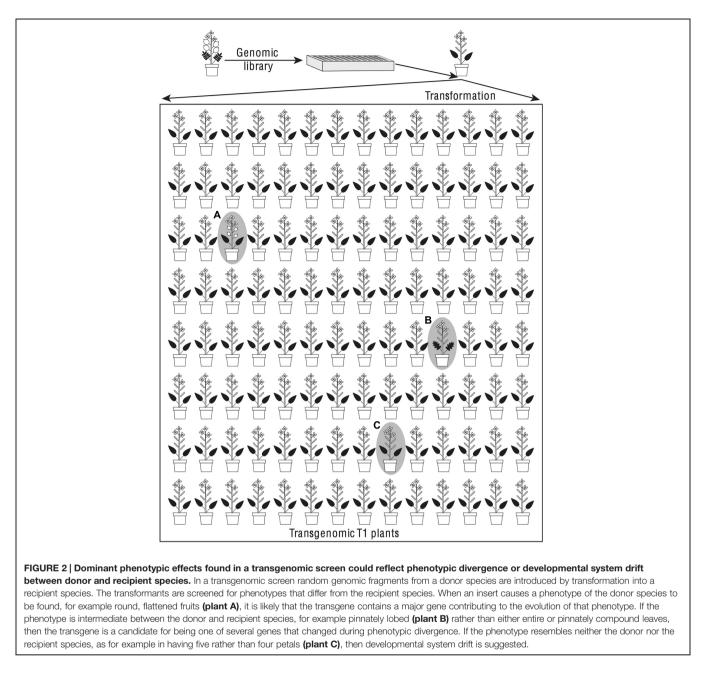
ALTERNATIVE TRANSGENOMIC STRATEGIES

In considering transgenomics two main experimental approaches suggest themselves. One approach is a shotgun strategy, where



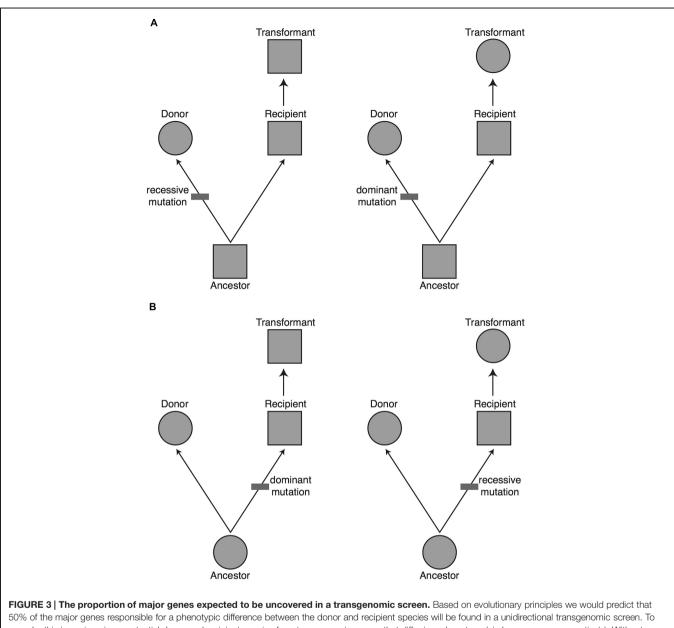


we generate a genomic library of a donor species in a suitable bacterium, e.g., *Agrobacterium tumefaciens* (Smith et Town.) Conn, and then introduce this *en masse* (or perhaps in pools) into a population of recipient plants (**Figure 4**). T1s would be screened for phenotypes of interest and, when a phenotype is observed, we would determine *post hoc* what insert had been introduced into the recipient's genome. Alternatively, a clone-by-clone strategy can be followed in which we isolate individual clones from a genomic library and use each clone for multiple transformations of the same recipient species to identify repeatable phenotypic effects (**Figure 4**).



A shotgun strategy has the advantage of quickly generating large populations of transformant plants to screen for phenotypes of interest. Nonetheless, it does have some serious drawbacks. (1) When interesting phenotypes are seen it might not be trivial to isolate the responsible genomic fragment, especially for large inserts or for inserts that cause sterility meaning that one could not obtain much transgenic plant tissue. (2) Because each clone will be introduced into only one recipient, a phenotypic effect could be a false positive due to genetic or microenvironmental differences among transformed plants. (3) Clones that cause dominant early lethality will not be identified so their frequency in the genome could not be assessed. (4) A shotgun-transformed pool could not readily interface with existing genomic information to yield a durable resource for other researchers to utilize.

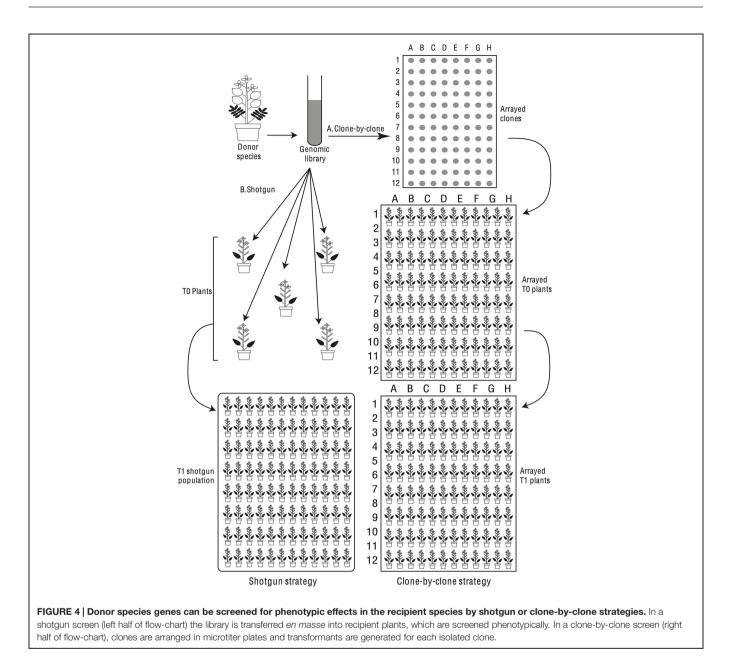
A clone-by-clone strategy has the virtue that the identity of inserts is easily determined (by sequencing the clone) and one can obtain multiple T1 plants per clone, reducing the false positive problem. Furthermore, inserts causing early lethality can, at least theoretically, be identified by their inability to generate mature transgenic plants and, once a population of transgenomic lines has been made, it represents a durable resource that could be screened repeatedly for different phenotypes in a diversity of growth conditions. On the other hand, a clone-by-clone strategy requires more work to separate and bulk-up individual



50% of the major genes responsible for a phenotypic difference between the donor and recipient species will be found in a unidirectional transgenomic screen. To see why this is so, imagine a potential donor and recipient species for a transgenomic screen that differ in a phenotype (circle vs. square, respectively). Without further information it is equally likely that: (A) the donor species has the derived phenotype, with a change having occurred on the lineage from the common ancestor to the donor species (upper panels), or (B) the donor has the ancestral phenotype, with a change having occurred on the lineage from the common ancestor to the recipient species (lower panels). The mutation that gave rise to the derived phenotype could have been fully recessive or at least partially dominant. If the donor has a derived phenotype that is dominant (top right), or it has an ancestral phenotype that is dominant phenotype. In approximately 50% of cases (left panels) the causal gene will not yield a dominant phenotype when moved from the donor to the recipient. The major genes missed in a unidirectional transgenomic screen could theoretically be found with a reciprocal screen in which the genome of the former recipient species.

clones and requires one to do many more individual plant transformations.

Whether one uses a shotgun or clone-by-clone approach, once transgenomics lines are found to have phenotypic effects a number of different downstream experiments can be undertaken. Before, or in parallel with, standard methods for studying gene-function (e.g., isolation of T-DNA insert lines, double mutant assays, expression studies) experiments should be conducted to assess the role of sequence divergence between donor and recipient species in explaining the phenotype. Some critical experiments will include: (1) repeating transformation to confirm that the phenotype is caused by the insert; (2) subcloning the insert to identify the causal gene region; (3) introducing the homologous gene from the recipient species as an extra copy to assess if the result is due to gene dosage; (4) generating chimeric constructs between the donor and recipient genes to



locate the causal differences, and; (5) isolating the homologous gene from additional species to assess the correlation between the phenotype of the donor species and the ability of the gene to confer an effect in transgenic lines of the recipient species. Through such experiments, there is every reason to hope that we could eventually arrive at a clear understanding of the molecular and developmental basis of the transgene's effect on phenotype and, in some cases, shed light on the evolution of phenotypic differences between species.

A PILOT CLONE-BY-CLONE SCREEN

Correa et al. (2012) piloted a clone-by-clone transgenomic screen using genomic clones with \sim 20 kb inserts from the gladecress

Leavenworthia alabamica Rollins introduced into *A. thaliana. Leavenworthia alabamica*, and *A. thaliana* differ in almost all visible morphologies and yet they are both members of Brassicaceae Lineage I (Beilstein et al., 2008, 2010; Franzke et al., 2011). While this was a large experiment, high-throughput "drip" transformation allowed one graduate student and 2–3 undergraduate assistants to screen T1s for as many as 750 clones per month.

Of the 1134 *L. alabamica* clones screened, 84 produced an initial T1 that deviated from "normal". However, in only eight cases was the initial "abnormal" phenotype repeated in additional independent T1s. Correa et al. (2012) focused on one clone that was shown to cause stunted fruit and increased seed abortion in a *transdominant* manner. Sequencing of the clone insert suggests that this effect is most likely explained by a gene region that

shows homology to the *A. thaliana SLOW-WALKER2 (SWA2)* gene. *SWA2* encodes a protein that is important for ribosome biogenesis, and *A. thaliana swa2* mutants manifest shortened fruit and seed abortion (Li et al., 2009). Follow-on experiments have confirmed that a subclone containing the region of *SWA2* homology is sufficient to cause these phenotypes and have detected *L. alabamica SWA2-like* mRNA in developing fruit (Wu and Baum, unpublished data).

What lessons were learned thanks to this pilot transgenomic screen? On the plus side, we established that a clone-by-clone strategy is feasible, at least when *A. thaliana* is the recipient, and showed that it is possible to use this strategy to identify a gene that can alter the phenotype of a recipient in a *transdominant* manner. Back-of-the-envelope calculation would suggest that if 750 clones can be screened per month by a small team piloting the approach for the first time, a larger and more experienced team could realistically screen more than 10,000 clones per year making it plausible that one could screen a moderate size genome to saturation.

COULD TRANSGENOMICS IDENTIFY MAJOR GENES EXPLAINING SPECIES DIFFERENCES?

The single gene identified by Correa et al. (2012) is likely to reflect DSD rather than phenotypic divergence. The question this raises is whether this result indicates that the method has limited utility for scientists whose goal is to identify genes that explain species differences.

The first fact to emphasize is that the screen conducted by Correa et al. (2012) covered less than 5% of the donor species' genome. Furthermore, considering that with a false negative rate of at least 50%, only 2.5% of the donor genome was effectively screened. This means that it would be grossly premature to take the results of Correa et al. (2012) as indicating that transgenomics cannot find species-differentiating genes. Thus, the best we can do is to look at prior experiments in which single genes have been moved between closely species to evaluate how often we might expect to find a dominant phenotypic effect in a transgenomic screen that is indicative of some functional role in explaining species differences (Correa et al., 2012).

A majority of experiments in which a gene (coding region or cDNA) from one species is moved into a different species use either a broadly active promoter such as 35S or the homologous promoter from the recipient species. The most common result from such studies is that the exogenous gene functions equivalently to the endogenous gene [e.g., (Whipple et al., 2004; Maizel et al., 2005; Dornelas and Rodriguez, 2006; Busch and Zachgo, 2007)]. Sometimes, especially with distantly related donors, the exogenous gene shows reduced functionality resembling a partial loss-of-function allele [e.g., (Tzeng and Yang, 2001; Maizel et al., 2005)]. Some studies using 35S have combined transformation data with evidence on comparative gene expression to show that altered expression of a single functionally conserved gene likely contributed to the evolution of plant phenotypes (Doczi et al., 2005; He and Saedler, 2005; Lee et al., 2005; Maizel et al., 2005; Zahn et al., 2005; Hay and Tsiantis, 2006; Busch and Zachgo, 2007; Hovav et al., 2007). In one case, the dominant effect of a full-length transgene was demonstrated using introgression, rather than transformation (Hovav et al., 2007). Likewise, some experiments using 35S have yielded novel phenotypes, with examples including pathogen resistance (Lee and Lee, 2005; Lee et al., 2005; Zahn et al., 2005; Carlson et al., 2006), stress-tolerance (Hsieh et al., 2002; Liu et al., 2007; Wang et al., 2007), the production of novel secondary metabolites (de Majnik et al., 2000; Mietkiewska et al., 2004; Mori et al., 2004), and even donor species-like morphology (Fourquin et al., 2013) that hint at evolutionary divergence in protein function sufficient to cause a dominant phenotype in a transgenomics screen.

Another class of interspecies transformation experiments move homologous *cis*-regulatory ("promoter") regions with reporters from one species to another to see if expression patterns are conserved [e.g., (Lin et al., 1993; Burger et al., 2006)] or divergent [e.g., (Doczi et al., 2005; Hay and Tsiantis, 2006)]. When a donor species' promoter drives expression in developmental stages or tissue types where the recipient species' promoter is inactive [e.g., (Zahn et al., 2005; de Martino et al., 2006)] there is a potential for a dominant phenotypic effect to be found in a transgenomics screen.

More direct evidence comes from those few studies that have moved genomic fragments with both cis-regulatory and coding regions between species. One example comes from work on the evolution of self-compatibility in A. thaliana. Genomic fragments from the S-locus of the self-incompatible A. lyrata (L.) O'Kane et Al-Shehbaz, when used to transform wildtype plants of the self-compatible A. thaliana, converted the latter to self-incompatibility - the effect varying with ecotype of the recipient plants (Nasrallah et al., 2002, 2004). This effect appears to be due to divergence in the coding region (loss of gene function in A. thaliana) rather than the evolution of regulatory regions. Similarly, Vlad et al. (2014) used transformation to show that the presence of the gene REDUCED COMPLEXITY (RCO) in the genomes of A. lyrata and Cardamine hirsuta L., but its loss in A. thaliana, largely explains the dissected leaf shape of the former species and the simple leaves of the latter. Another example involves the introduction of LFY and TFL1 genes from different Brassicaceae species into A. thaliana. In all three cases involving LFY (Yoon and Baum, 2004; Sliwinski et al., 2007) and the one case involving TFL1 (Liu et al., 2011), the transgene resulted in a novel phenotype and these were shown to also occur in a wildtype background showing transdominance (Sliwinski et al., 2006, 2007; Liu et al., 2011). Similarly, studies of REPLUMLESS (RPL), a gene that promotes fruit dehiscence, showed that introducing the Arabidopsis RPL gene into Brassica is sufficient to induce Arabidopsis-like fruit dehiscence in the recipient (Arnaud et al., 2011). As was also found in the LFY and TFL1 experiments, the transgene effect is not replicated when using the endogenous gene copy, showing that the effect is due to sequence divergence (specifically in *cis*-regulatory regions) rather than to a gene dosage effect.

Taken together, evolutionary theory and prior candidate gene interspecies transformation experiments suggest that a transgenomics screen has a high potential to uncover *transdominant* phenotypic effects. Given this we should assess how best to design future transgenomic screens to maximize their efficiency.

PRACTICALITIES OF SCREENING

One of the striking findings of Correa et al. (2012) was the high false positive rate. Specifically, out of 84 cases where an abnormal phenotype was seen in an initial T1, only eight recurred in further T1s from the same clone, and only one was definitively shown to be due to a *L. alabamica* insert. It is not surprising that false positives arise since there is likely to be genetic variation among the plants used for transformation and even wildtype plants occasionally manifest phenotypic abnormalities (Hempel and Feldman, 1995). Furthermore, transformation and *Agrobacterium* infection are both potentially mutagenic, so some abnormalities likely reflect *de novo* mutation. However, although some false positives are inevitable, it is worth considering strategies for reducing their frequency so as to avoid wasting time and effort following up phenotypes that are not caused by the transgene.

One strategy that we have explored is to immediately growup and screen \sim 5 T1s per clone instead of just one. Once one is screening seedlings on plates for a selective resistance trait, it is not much more difficult to transplant and retain five rather than one T1. Requiring that phenotypes recur in at least 2–3 of the independent T1s from the same clone would go a long way toward excluding phenotypic effects that are due to a position effect or an insertional mutation in an endogenous gene, since independent T1s are not expected to have their inserts integrated at the same *A. thaliana* locus. Furthermore, if the locations of plants used to test a particular clone are randomized within growth rooms, there is little chance that multiple T1s would have the same phenotype because of a shared microenvironment.

An additional benefit of screening several T1s per clone is that it should also reduce the number of false negatives: overlooking a causally important clone due to a lack of a visible phenotype in the initial T1 screened. Despite clear evidence that the clone containing the *SWA2*-like region causes increased seed abortion and reduced fruit size, Correa et al. (2012) noted that about one half of the T1s containing the clone showed a wild type morphology. This matches other experiments that have reported that many clones fail to manifest a phenotype in *A. thaliana* due to transgene silencing (Morel et al., 2000; Schubert et al., 2004). However, while there are benefits to screening several rather than just one T1, such an approach significantly increases the space needed to conduct the initial screen. So, if space limitations rather than labor limitations are paramount, this strategy might not be worth deploying.

We have also tried a further embellishment to reduce the impact of genetic differences among the transformed (T0) plants. This involved growing \sim 5 independent T1 plants germinated as usual on selective agar plates and in parallel growing the same number of plants from the same seed stock but isolated from non-selective plates. Based on *Arabidopsis* transformation efficiency, non-selectively grown plants are much (ca. two orders

of magnitude) more likely to lack donor species DNA than to contain it. This means that phenotypic effects that recur in the T1 plants but are absent in non-selectively grown siblings are much more likely to be due to a *transdominant* donor species gene. However, this represents a further doubling of the space required and additional work for making plates, transplanting seedlings, and scoring plants for abnormalities. Based on our informal experimentation, we are doubtful that this additional work would yield sufficient benefits to make it worthwhile.

CHOICE OF DONOR AND RECIPIENT SPECIES

Whichever transgenomic strategy is used, the choice of donor and recipient species is important. As the preeminent plant genetic model system, blessed with efficient transformation methods, *A. thaliana* is the ideal recipient species with which to initially implement and assess a transgenomics approach. While other readily transformed plant species such as rice, tobacco, and petunia are worth considering, we believe that the efficiency of dip-transformation makes Brassicaceae the clade of choice in which to first try transgenomic research.

The strengths of transgenomics would be greatest when working with a donor species that is closely related to the recipient so that many of the core developmental process and genes are shared, but distant enough that many visible phenotypes differ. Alternatively, if one is committed to a particular trait (e.g., salt tolerance, metal hyperaccumulation, leaf shape, etc.), one can pick a donor species that differs from *A. thaliana* in at least that phenotype. What we do not know is how far from the phylogenetic neighborhood of *A. thaliana* you can go before most phenotypic effects are hard to make sense of: *Brassica* L., *Cleome* L., papaya, cotton, tobacco, rice, moss?

An additional consideration is the availability of genetic tools in the donor species. If the donor genome has been sequenced it will be that much easier to home in on causal regions once an interesting phenotype has been found. Also, if the donor species is one that can be transformed then it will be possible to introduce *A. thaliana* genes, which may be useful for exploring gene function. Furthermore, if both species can be transformed with high-throughput methods, it becomes possible to undertake a full, bidirectional transgenomics screen, which should allow one to identify almost all genes of large effect that have contributed to the phenotypic divergence of the two species (**Figure 3**).

CLONING STRATEGY

Once the donor and recipient species have been identified, a number of detailed practical issues will need to be addressed, many of which could significantly influence the efficacy of the screen. Among these is the strategy used to assemble a genomic library, most notably insert size, choice of vector, selectable marker, and possible target-enrichment strategy.

Large inserts will presumably allow for more rapid screening of a genome to completion. A further advantage of long inserts is they are less likely to include truncated genes, which can yield false positive results if, for example, a truncated donor species protein binds to and prevents a partner protein from interacting normally with the homologous recipient species protein (Villagarcia et al., 2012) or if the transgene region is missing a negative regulatory protein domain or a *cis*-repressor element. On the other hand, longer clone inserts are more difficult to work with in the lab and typically show lower transformation efficiencies. Furthermore, inserts that are too large to allow amplification from T-DNA primers will make it markedly more difficult to isolate the insert sequence from transformant plants, as might be necessary for a shotgun transgenomic screen.

Vectors should be ones that can achieve sufficient transformation efficiency for inserts of the target size, yet should tend to yield only one insert per transgenic line so as to minimize confounding dosage effects. Additionally, all things being equal, it might be helpful to engineer a vector that results in inserts being flanked by insulator sequences (She et al., 2010; Singer and Cox, 2013). By reducing the extent to which gene expression is affected by where in the genome an insert lands, insulator sequences might reduce line-to-line variability and, thus, lower the false negative rate.

The choice of selectable marker will be guided by speed of screening as well as the desirability of being able to identify transgenic lines at a very early developmental stage so that genes causing early lethality can be found. For example, it might be possible to use a selectable marker that causes embryos to fluoresce (Ali et al., 2012), allowing transformants to be visually identified as seeds and then grown up, non-selectively, on soil.

The final factor to consider in designing a transgenomic screen is whether it might be possible to manipulate the genomic library to increase the proportion of clones that includes potentially causal genes. While transgenomics is premised on the idea that we want to focus on genes with their native cisregulatory machinery, meaning that we must include abundant non-coding content in our inserts, there is certainly much of the genome that is, a priori, less likely to cause informative transdominant phenotypes in a foreign genome. Our initial focus might be to enrich for the gene-space or, if we are focusing on a particular phenotype, we might be specifically interested in enriching for genes that are expressed in a particular organ. There is a diversity of methods available for enriching genomic libraries (Cronn et al., 2012), though most are optimized for short rather than long inserts. Possibilities include enriching for low-copy number genes based on melting kinetics (Yuan et al., 2003) or methylation (Palmer et al., 2003) or using hybridization against expressed genes (Lovett et al., 1991; Fu et al., 2010). While further method development would be needed, there is abundant scope for generating transgenomic libraries where a far higher proportion of clones contain causally important genes, thereby making screens dramatically more efficient.

PROSPECTS FOR TRANSGENOMICS

Transgenomics has great potential for contributing to the study of gene function in genetic model species and for identifying the molecular changes that underlie phenotypic evolution. While transgenomic screens will require a significant investment of effort and money, evolutionary theory and data from candidate gene transformation experiments show that, at least in Brassicaceae, the costs may well be outweighed by the fundamental data that will be obtained.

The initial ventures in transgenomics should aim to answer some important outstanding questions. In particular it is critical that an effort be made to quantify the proportion of a genome that causes different kinds of dominant phenotypes (morphological changes, sterility, lethality, etc.) as a function of phylogenetic distance between donor and recipient species. Further, there would be great value in studying two closely related species and conducting a complete, reciprocal transgenomics screen in order to determine the total number of major genes responsible for their different phenotypes. Lastly, we believe it would be beneficial for at least the Arabidopsis community to invest in an ordered transgenomics resource deposited in stock centers. If sets of sequenced clones from different donor species were each associated with transgenic seed, a researcher interested in a particular gene family could order up seed containing exogenous versions of that gene to look for phenotypes that may reveal aspects of gene function. Similarly, a scientist seeking new genes involved in the development of a phenotype could identify a donor and recipient species pair that differ in the phenotype and then could screen the corresponding transgenomic lines.

Looking to the future, it seems likely that transgenomics will emerge as an important tool for basic plant research. It is also possible that transgenomics could have applied importance. As a complement to traditional mass selection and candidate gene genetic modification, breeders may find it useful to introduce genetic variation in bulk from foreign species followed by screening and selection based on desirable traits that emerge. For all these reasons, we hope that this article will stimulate efforts to develop transgenomics as a tool for plant genetic research.

ACKNOWLEDGMENTS

We are very grateful to the following for thoughtful discussion and helpful suggestions over the years we have been exploring transgenomics: S. Cutler, B. Deakin, B. Dilkes, J. Doebley, B. Glover, E. Kramer, P. Krysan, N. Liu, P. Masson, T. C. Pires, S. D. Smith, K. Shimizu, M. K. Sliwinski, J. Stanga, and D. Weigel. DB gratefully acknowledges funding from the NSF (IOS-0641428), the UW-Madison Vilas fund, and the Guggenheim Foundation. We thank Melody Wu for her work on *L. alabamica SWA2*. Artwork was prepared with assistance from Kandis Elliot.

REFERENCES

- Ali, M. A., Shah, K. H., and Bohlmann, H. (2012). pMAA-Red: a new pPZP-derived vector for fast visual screening of transgenic *Arabidopsis* plants at the seed stage. *BMC Biotechnol.* 12:37. doi: 10.1186/1472-6750-12-37
- Arnaud, N., Lawrenson, T., Østergaard, L., and Sablowski, R. (2011). The same regulatory point mutation changed seed-dispersal structures in evolution and domestication. *Curr. Biol.* 21, 5–5. doi: 10.1016/j.cub.2011.06.008
- Arthur, W. (2002). The emerging conceptual framework of evolutionary developmental biology. *Nature* 415, 757–764. doi: 10.1038/415757a
- Baum, D. A. (2002). "Identifying the genetic causes of phenotypic evolution: a review of experimental strategies," in *Developmental Genetics and Plant Evolution*, eds Q. C. B. Cronk, R. M. Bateman, and J. A. Hawkins (London: Taylor & Francis), 493–507.
- Beilstein, M. A., Al-Shehbaz, I. A., Mathews, S., and Kellogg, E. A. (2008). Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. *Am. J. Bot.* 95, 1307–1327. doi: 10.3732/aib.0800065
- Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc. Natl. Acad. Sci. U.S.A. 107, 18724–18728. doi: 10.1073/pnas.0909766107
- Birchler, J. A., and Veitia, R. A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19, 395–402. doi: 10.1105/tpc.106.049338
- Blazquez, M. A., Soowal, L. N., Lee, I., and Weigel, D. (1997). LEAFY expression and flower initiation in *Arabidopsis*. *Development* 124, 3835–3844.
- Bomblies, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangl, J. L., et al. (2007). Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* 5:e236. doi: 10.1371/journal.pbio.0050236
- Burger, A. L., Watts, L., and Botha, F. C. (2006). Grapevine promoter directs gene expression in the nectaries of transgenic tobacco. *Physiol. Plant.* 126, 418–434. doi: 10.1111/j.1399-3054.2006.00598.x
- Busch, A., and Zachgo, S. (2007). Control of corolla monosymmetry in the Brassicaceae Iberis amara. Proc. Natl. Acad. Sci. U.S.A. 104, 16714–16719. doi: 10.1073/pnas.0705338104
- Carlson, A., Skadsen, R., and Kaeppler, H. F. (2006). Barley hordothionin accumulates in transgenic oat seeds and purified protein retains antifungal properties in vitro. *In Vitro Cell. Dev. Biol. Plant* 42, 318–323. doi: 10.1079/IVP2006780
- Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. (2004). From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design. Oxford: Wiley-Blackwell.
- Correa, R., Stanga, J., Larget, B., Roznowski, A., Shu, G., Dilkes, B., et al. (2012). An assessment of transgenomics as a tool for identifying genes involved in the evolutionary differentiation of closely related plant species. *New Phytol.* 193, 494–503. doi: 10.1111/j.1469-8137.2011.03 949.x
- Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., et al. (2012). Targeted enrichment strategies for next-generation plant biology. *Am. J. Bot.* 99, 291–311. doi: 10.3732/ajb.1100356
- de Majnik, J., Weinman, J. J., Djordjevic, M. A., Rolfe, B. G., Tanner, G. J., Joseph, R. G., et al. (2000). Anthocyanin regulatory gene expression in transgenic white clover can result in an altered pattern of pigmentation. *Austral. J. Plant Physiol.* 27, 659–667.
- de Martino, G., Pan, I., Emmanuel, E., Levy, A., and Irish, V. F. (2006). Functional analyses of two tomato APETALA3 genes demonstrate diversification in their roles in regulating floral development. *Plant Cell* 18, 1833–1845. doi: 10.1105/tpc.106.042978
- Doczi, R., Kondrak, M., Kovacs, G., Beczner, F., and Banfalvi, Z. (2005). Conservation of the drought-inducible DS2 genes and divergences from their ASR paralogues in solanaceous species. *Plant Physiol. Biochem.* 43, 269–276.
- Doebley, J., and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. *Plant Cell* 10, 1075–1082. doi: 10.2307/3870712
- Dornelas, M. C., and Rodriguez, A. P. M. (2006). The tropical cedar tree (*Cedrela fissilis Vell.*, Meliaceae) homolog of the *Arabidopsis LEAFY* gene is expressed in reproductive tissues and can complement

Arabidopsis leafy mutants. Planta 223, 306-314. doi: 10.1007/s00425-005-0086-y

- Fourquin, C., Del Cerro, C., Victoria, F. C., Vialette-Guiraud, A., De Oliveira, A. C., and Ferrandiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. *Plant Physiol.* 162, 907–917. doi: 10.1104/pp.113.21 7570
- Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* 16, 9–9. doi: 10.1016/j.tplants.2010.11.005
- Fu, Y., Springer, N. M., Gerhardt, D. J., Ying, K., Yeh, C. T., Wu, W., et al. (2010). Repeat subtraction-mediated sequence capture from a complex genome. *Plant J.* 62, 898–909. doi: 10.1111/j.1365-313X.2010.04196.x
- Gottlieb, L. D. (1984). Genetics and morphological evolution in plants. *Am. Nat.* 123, 681–709. doi: 10.1086/284231
- Haag, E. S. (2007). Compensatory vs. pseudocompensatory evolution in molecular and developmental interactions. *Genetica* 129, 45–55. doi: 10.1007/s10709-006-0032-3
- Hay, A., and Tsiantis, M. (2006). The genetic basis for differences in leaf form between Arabidopsis thaliana and its wild relative Cardamine hirsuta. Nat. Genet. 38, 942–947. doi: 10.1038/ng1835
- He, C. Y., and Saedler, H. (2005). Heterotopic expression of MPF2 is the key to the evolution of the Chinese lantern of Physalis, a morphological novelty in Solanaceae. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5779–5784. doi: 10.1073/pnas.0501877102
- Hempel, F. D., and Feldman, L. J. (1995). Specification of chimeric flowering shoots in wild-type Arabidopsis. Plant J. 8, 725–731. doi: 10.1046/j.1365-313X.1995.08050725.x
- Hoekstra, H. E., and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016. doi: 10.1111/j.1558-5646.2007.00105.x
- Hovav, R., Chehanovsky, N., Moy, M., Jetter, R., and Schaffer, A. A. (2007). The identification of a gene (Cwp1), silenced during Solanum evolution, which causes cuticle microfissuring and dehydration when expressed in tomato fruit. *Plant J.* 52, 627–639. doi: 10.1111/j.1365-313X.2007. 03265.x
- Hsieh, T. H., Lee, J. T., Charng, Y. Y., and Chan, M. T. (2002). Tomato plants ectopically expressing *Arabidopsis* CBF1 show enhanced resistance to water deficit stress. *Plant Physiol.* 130, 618–626. doi: 10.1104/pp.0 06783
- Kondrashov, A. S., Sunyaev, S., and Kondrashov, F. A. (2002). Dobzhansky– Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14878–14883. doi: 10.1073/pnas.232565499
- Landry, C. R., Hartl, D. L., and Ranz, J. M. (2007). Genome clashes in hybrids: insights from gene expression. *Heredity* 99, 483–493. doi: 10.1038/sj.hdy.6801045
- Lee, J.-H., Cho, Y.-S., Yoon, H.-S., Suh, M. C., Moon, J., Lee, I., et al. (2005). Conservation and divergence of FCA function between *Arabidopsis* and rice. *Plant Mol. Biol.* 58, 823–838. doi: 10.1007/s11103-005-8105-8
- Lee, S. Y., and Lee, D. H. (2005). Expression of MbR4, a TIR-NBS type of apple R gene, confers resistance to bacterial spot disease in *Arabidopsis. J. Plant Biol.* 48, 220–228. doi: 10.1007/BF03030411
- Lee, Y. W., Gould, B. A., and Stinchcombe, J. R. (2014). Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB Plants* 6:plu004. doi: 10.1093/aobpla/plu004
- Li, N., Yuan, L., Liu, N., Shi, D., Li, X., Tang, Z., et al. (2009). SLOW WALKER2, a NOC1/MAK21 homologue, is essential for coordinated cell cycle progression during female gametophyte development in *Arabidopsis. Plant Physiol.* 151, 1486–1497. doi: 10.1104/pp.109.142414
- Lin, E., Burns, D. J. W., and Gardner, R. C. (1993). Fruit developmental regulation of the kiwifruit actinidin promoter is conserved in transgenic petunia plants. *Plant Mol. Biol.* 23, 489–499. doi: 10.1007/BF00019297
- Liu, N., Chen, A. P., Zhong, N. Q., Wang, F., Wang, H. Y., and Xia, G. X. (2007). Functional screening of salt stress-related genes from *Thellungiella halophila* using fission yeast system. *Physiol. Plant.* 129, 671–678. doi: 10.1111/j.1399-3054.2007.00857.x
- Liu, N., Sliwinski, M. K., Correa, R., and Baum, D. A. (2011). Possible contributions of TERMINAL FLOWER 1 to the evolution of rosette flowering

in Leavenworthia (Brassicaceae). New Phytol. 189, 616–628. doi: 10.1111/j.1469-8137.2010.03511.x

- Lovett, M., Kere, J., and Hinton, L. M. (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci.* U.S.A. 88, 9628–9632. doi: 10.1073/pnas.88.21.9628
- Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., et al. (2005). The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308, 260–263. doi: 10.1126/science.1108229
- Mallory, A. C., Reinhart, B. J., Jones-Rhoades, M. W., Tang, G., Zamore, P. D., Barton, M. K., et al. (2004). MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. *EMBO J.* 23, 3356–3364. doi: 10.1038/sj.emboj.7600340
- Mietkiewska, E., Giblin, E. M., Wang, S., Barton, D. L., Dirpaul, J., Brost, J. M., et al. (2004). Seed-specific heterologous expression of a nasturtium FAE gene in *Arabidopsis* results in a dramatic increase in the proportion of erucic acid. *Plant Physiol.* 136, 2665–2675. doi: 10.1104/pp.104.046839
- Morel, J. B., Mourrain, P., Beclin, C., and Vaucheret, H. (2000). DNA methylation and chromatin structure affect transcriptional and posttranscriptional transgene silencing in *Arabidopsis. Curr. Biol.* 10, 1591–1594. doi: 10.1016/S0960-9822(00)00862-9
- Mori, S., Kobayashi, H., Hoshi, Y., Kondo, M., and Nakano, M. (2004). Heterologous expression of the flavonoid 3 '5 '-hydroxylase gene of Vinca major alters flower color in transgenic *Petunia hybrida*. *Plant Cell Rep.* 22, 415–421. doi: 10.1007/s00299-003-0709-3
- Nasrallah, M. E., Liu, P., and Nasrallah, J. B. (2002). Generation of selfincompatible Arabidopsis thaliana by transfer of two S locus genes from A. lyrata. Science 297, 247–249. doi: 10.1126/science.1072205
- Nasrallah, M. E., Liu, P., Sherman-Broyles, S., Boggs, N. A., and Nasrallah, J. B. (2004). Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: Implications for the evolution of selfing. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16070–16074. doi: 10.1073/pnas.0406970101
- Palmer, L. E., Rabinowicz, P. D., O'shaughnessy, A. L., Balija, V. S., Nascimento, L. U., Dike, S., et al. (2003). Maize genome sequencing by methylation filtration. *Science* 302, 2115–2117. doi: 10.1126/science.1091265
- Papoyan, A., and Kochian, L. V. (2004). Identification of Thlaspi caerulescens genes that may be involved in heavy metal hyperaccumulation and tolerance. Characterization of a novel heavy metal transporting ATPase. *Plant Physiol.* 136, 3814–3823.
- Rieseberg, L. H., Archer, M. A., and Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity* 83(Pt 4), 363–372. doi: 10.1038/sj.hdy.6886170
- Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., and Schmidt, R. (2004). Silencing in *Arabidopsis* T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell* 16, 2561–2572. doi: 10.1105/tpc.104.024547
- She, W., Lin, W., Zhu, Y., Chen, Y., Jin, W., Yang, Y., et al. (2010). The gypsy insulator of Drosophila melanogaster, together with its binding protein suppressor of Hairy-wing, facilitate high and precise expression of transgenes in *Arabidopsis thaliana*. *Genetics* 185, 1141–1150. doi: 10.1534/genetics.110.117960
- Singer, S., and Cox, K. (2013). A gypsy-like sequence from Arabidopsis thaliana exhibits enhancer-blocking activity in transgenic plants. J. Plant Biochem. Biotechnol. 22, 35–42. doi: 10.1007/s13562-012-0108-3
- Sliwinski, M. K., Bosch, J. A., Yoon, H. S., Von Balthazar, M., and Baum, D. A. (2007). The role of two LEAFY paralogs from *Idahoa scapigera* (Brassicaceae)

in the evolution of a derived plant architecture. *Plant J.* 51, 211–219. doi: 10.1111/j.1365-313X.2007.03148.x

- Sliwinski, M. K., White, M. A., Maizel, A., Weigel, D., and Baum, D. A. (2006). Evolutionary divergence of LFY function in the mustards *Arabidopsis thaliana* and *Leavenworthia crassa*. *Plant Mol. Biol.* 62, 279–289. doi: 10.1007/s11103-006-9020-3
- Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evolution* 54, 1079–1091. doi: 10.1111/j.0014-3820.2000.tb00544.x
- True, J. R., and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3, 109–119. doi: 10.1046/j.1525-142x.2001.003002109.x
- Tzeng, T. Y., and Yang, C. H. (2001). A MADS box gene from lily (*Lilium longiflorum*) is sufficient to generate dominant negative mutation by interacting with PISTILLATA (PI) in *Arabidopsis thaliana*. *Plant Cell Physiol.* 42, 1156–1168. doi: 10.1093/pcp/pce151
- Villagarcia, H., Morin, A. C., Shpak, E. D., and Khodakovskaya, M. V. (2012). Modification of tomato growth by expression of truncated ERECTA protein from *Arabidopsis thaliana*. J. Exp. Bot. 63, 6493–6504. doi: 10.1093/jxb/ers305
- Vlad, D., Kierzkowski, D., Rast, M. I., Vuolo, F., Dello Ioio, R., Galinha, C., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343, 780–783. doi: 10.1126/science.1248384
- Wagner, G. P., Chiu, C.-H., and Laubichler, M. (2000). Developmental evolution as a mechanistic science: the inference from developmental mechanisms to evolutionary processes. *Am. Zool.* 40, 819–831. doi: 10.1093/icb/ 40.5.819
- Wang, M. Y., Gu, D., Liu, T. S., Wang, Z. Q., Guo, X. Y., Hou, W., et al. (2007). Overexpression of a putative maize calcineurin B-like protein in *Arabidopsis* confers salt tolerance. *Plant Mol. Biol.* 65, 733–746. doi: 10.1007/s11103-007-9238-8
- Whipple, C. J., Ciceri, P., Padilla, C. M., Ambrose, B. A., Bandong, S. L., and Schmidt, R. J. (2004). Conservation of B-class floral homeotic gene function between maize and *Arabidopsis. Development* 131, 6083–6091. doi: 10.1242/dev.01523
- Yoon, H. S., and Baum, D. A. (2004). Transgenic study of parallelism in plant morphological evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6524–6529. doi: 10.1073/pnas.0401824101
- Yuan, Y., Sanmiguel, P. J., and Bennetzen, J. L. (2003). High-Cot sequence analysis of the maize genome. *Plant J.* 34, 249–255. doi: 10.1046/j.1365-313X.2003.01716.x
- Zahn, L. M., Leebens-Mack, J., DePamphilis, C. W., Ma, H., and Theissen, G. (2005). To B or not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J. Heredity* 96, 225–240. doi: 10.1093/jhered/esi033

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Correa and Baum. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.