RNA-Seq Analysis of Allele-Specific Expression, Hybrid Effects, and Regulatory Divergence in Hybrids Compared with Their Parents from Natural Populations

Graeme D.M. Bell, Nolan C. Kane, Loren H. Rieseberg, and Keith L. Adams*
Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada
*Corresponding author: E-mail: keitha@mail.ubc.ca.
Accepted: May 3, 2013

Abstract
Hybridization is a prominent process among natural plant populations that can result in phenotypic novelty, heterosis, and changes in gene expression. The effects of intraspecific hybridization on F₁ hybrid gene expression were investigated using parents from divergent, natural populations of Cirsium arvense, an invasive Compositae weed. Using an RNA-seq approach, the expression of 68,746 unigenes was quantified in parents and hybrids. The expression levels of 51% of transcripts differed between parents, a majority of which had less than 1.25-fold changes. More unigenes had higher expression in the invasive parent (P₁) than the noninvasive parent (P₂). Of those that were divergently expressed between parents, 10% showed additive and 81% showed nonadditive (transgressive or dominant) modes of gene action in the hybrids. A majority of the dominant cases had P₂-like expression patterns in the hybrids. Comparisons of allele-specific expression also enabled a survey of cis- and trans-regulatory effects. Cis- and trans-regulatory divergence was found at 70% and 68% of 62,281 informative single-nucleotide polymorphism sites, respectively. Of the 17% of sites exhibiting both cis- and trans-effects, a majority (70%) had antagonistic regulatory interactions (cis x trans); trans-divergence tended to drive higher expression of the P₁ allele, whereas cis-divergence tended to increase P₂ transcript abundance. Trans-effects correlated more highly than cis with parental expression divergence and accounted for a greater proportion of the regulatory divergence at sites with additive compared with nonadditive inheritance patterns. This study explores the nature of, and types of mechanisms underlying, expression changes that occur in upon intraspecific hybridization in natural populations.

Key words: intraspecific hybridization, hybrids, gene expression, gene regulation, Cirsium, Compositae.

Introduction
Hybridization events are widespread in plants and are thought to be an important evolutionary stimulus (Rieseberg 1997; Arnold 2004). Both inter- and intraspecific hybridization are capable of creating genetic variation and novelty that can facilitate rapid adaptive evolution, most notably in crop plants and invasive weeds (Ellstrand and Schierenbecks 2000; Abbott et al. 2003; Ainouche et al. 2003; Prentis et al. 2008). The related genetic phenomenon, heterosis, in which heterozygous hybrid lines achieve faster growth rates than either parental line, is widely exploited by commercial plant breeders (Duvick 1999; Hochholdinger and Hoecker 2007; Birchler et al. 2010; Xing and Zhang 2010).

The ability of hybrids to exploit novel or extreme habitats is often attributed to the altered regulatory environment in hybrids, where novel interactions between parental alleles can result in nonmidparental (i.e., transgressive) hybrid phenotypes (Rieseberg et al. 1999, 2007; Arnold and Martin 2010). Transgressive hybrid effects have been shown to account for the appearance of certain adaptive characteristics that may even drive hybrid speciation (Rieseberg, Widmer, et al. 2003). Evidence is emerging that modified gene expression levels in hybrids contribute to transgressive hybrid phenotypes (Song and Messing 2003; Springer and Stupar 2007). The molecular mechanisms underlying these “hybrid effects” concern the additive and dominant nature of regulatory interactions between parental alleles. Other ways hybridization can facilitate adaptive evolution are by generally increasing genetic variation and through the introduction of new and possibly beneficial alleles. Studying the extent of additive and nonadditive gene expression in hybrids can provide insight into...
the genetic mechanisms by which hybridization leads to the transgressive segregation implicated in the adaptive success of hybrids (Gibson et al. 2004; Birchler et al. 2010; Chen 2010).

Genes that exhibit dominant or transgressive (over- and underdominant) hybrid expression are thought to be important in conferring novel or nonmidparental hybrid traits because even subtle changes can have drastic phenotypic effects (Coors and Pandey 1999; Crow 1999; reviewed in Springer and Stupar 2007; Chen 2010). For example, studies in Arabidopsis and Drosophila have linked specific point mutations in transcriptional regulators of key developmental genes to variation in flowering time and morphological evolution, respectively (e.g., Gompel et al. 2005; Prud’homme et al. 2006; Schwartz et al. 2009). Other plant studies of gene expression inheritance using inbred maize, Senecio, and rice lines found that additive and dominant modes of gene action predominated, whereas fewer genes exhibited transgressive expression (Swanson Wagner et al. 2006; Hegarty et al. 2006; He et al. 2010, respectively).

Gene expression is governed at the level of transcription by interactions between cis- and trans-acting regulatory elements (Tautz 2000; Williams et al. 2007). A test comparing parental and hybrid allele-specific expression (ASE) has been devised to discern between the cis- and trans-effects underlying gene expression divergence (Wittkopp et al. 2004). It is based on the fundamental difference between the two types of regulators: cis-regulatory changes have allele-specific effects on expression of nearby genes, whereas changes in trans-regulators can affect both alleles in the diploid hybrid nucleus. In the hybrid, allelic imbalance (i.e., unequal expression of parental alleles) is a signature of cis-regulatory divergence, because both parental alleles are exposed to a common set of trans-acting regulators.

Studying the nature of the regulatory changes (cis- or trans-acting) that underlie parental divergence and transgressive hybrid segregation can provide insight into the evolutionary forces that influence gene expression variation within and between populations (Lemos et al. 2008; Emerson and Li 2010). There is ongoing debate over the relative importance of cis- and trans-regulatory mutations in terms of their contributions toward adaptive evolution (Hoekstra and Coyne 2007; Prud’homme et al. 2007). Cis-regulatory mutations have been shown to account for at least some instances of evolutionarily significant phenotypic change (Wray 2007), but trans-regulatory evolution can also affect adaptive morphological change, as demonstrated by the trans-mediation of flowering-time genes FRI and FLC in Arabidopsis allopolyploids (Wang et al. 2006).

Cis/trans-ASE tests using hybrids of inbred maize lines found that cis-acting regulatory variation accounted for the majority of the observed parental expression divergence and that cis-variation correlated with additive expression patterns in the F1 hybrid (Guo et al. 2004; Stupar and Springer 2006).

A similar trend has emerged in the animal literature (Cowles et al. 2002; Pastinen and Hudson 2004; Wittkopp et al. 2004). However, there is also evidence for significant nonadditive gene expression in maize, Arabidopsis, Drosophila, and rice, including many cases of transgressive hybrid expression (Gibson et al. 2004; Ranz et al. 2004; Auger et al. 2005; Vuylsteke et al. 2005; Huang et al. 2006). These and other studies identify a general trend of cis-regulatory changes accounting for more of the divergent expression between more genetically divergent (i.e., interspecific) parents (Stern and Orgogozo 2008; Wittkopp et al. 2008a), whereas trans-effects account for a higher proportion of the variation in gene expression between less divergent (i.e., intraspecific) parents (Wittkopp et al. 2004, 2008b; Stern and Orgogozo 2008; Zhang and Borevitz 2009). Exceptions to this trend have been noted in cases of suspected population bottlenecks in one or both parental population, where a higher-than-expected contribution of trans-variation might be the result of drift, rather than selection (McManus et al. 2010).

Although there have been many studies of gene and ASE changes in hybrids using interspecific and intraspecific crosses in both plants and animals, most of those studies have used inbred parental lines (e.g., Stupar and Springer 2006; Springer and Stupar 2007; Gruber and Long 2008; Guo et al. 2008; Wittkopp et al. 2004, 2008a; Chang et al. 2008; Zhang and Borevitz 2009; Sung et al. 2009). Thus, the effects of hybridization on gene expression in natural populations remain largely unknown.

High-throughput sequencing approaches have recently been used to investigate the role of hybridization in invasive plant evolution, particularly in weeds in the Compositae (Asteraceae) family (Basu et al. 2004; Chao et al. 2005; Stewart et al. 2009; Lai et al. 2012). Several Compositae weeds emerging as the focus of genomic weed research include sunflower, ragweed, diffuse and common knapweeds, yellow starthistle, Canada thistle, and horseweed (Chao et al. 2005; Stewart et al. 2009; Lai et al. 2012). Evidence for the role of interspecific hybridization in adaptive evolution comes from wild sunflowers, where multiple natural hybridization events facilitated range expansion and the colonization of novel environments (Rieseberg et al. 2003, 2007). Intraspecific hybridization can also contribute to adaptive success in cases where invasive weeds experienced multiple introductions from divergent source populations (Bosshop et al. 2005; Dlugosch and Parker 2008). There is genetic evidence for multiple introductions having occurred with numerous invasive weeds including Ambrosia artemisiifolia (Common ragweed), Cytisus scoparius (Scotch broom), Verbascum thapsus (Common Mullein), and Hypericum canariense (Canary Island St. John’s wort; Gento et al. 2005; Kang et al. 2007; Dlugosch and Parker 2008). In the Compositae species, there is evidence for interspecific hybridization between diffuse and spotted knapweeds, as well as for intraspecific hybridization having
occurred between wild and cultivated sunflower populations (Lai et al. 2012).

In this study, we examined the nature and extent of changes in gene expression that occur in intraspecific hybrids from natural, divergent populations of outcrossing Cirsium arvense (L.) Scop. (Canada thistle) by using an RNA-Seq approach to profile and compare gene expression levels in parents and F1 hybrids. High-throughput Illumina sequencing of total cDNA was used to quantify overall gene, as well as allele-specific, transcript abundance (Cloonan and Grimmond 2008; Wang et al. 2008; Fontanillas et al. 2010). The extent of parental divergence was assessed and genes exhibiting additive, dominant, and transgressive hybrid expression patterns were identified. A cis\(\text{cis}\)/trans ASE test was also used to infer the putative regulatory mechanisms responsible for altered hybrid expression patterns (Wittkopp et al. 2004). The magnitude and direction of cis- and/or trans-regulatory changes were identified for each contig and their net effect on allelic expression (cis + trans, cis \(\times\) trans, and compensatory) was determined. Finally, this analysis examined a putative link between the mode of gene action (additive/nonadditive) and the mechanism of regulatory divergence (cis\(\text{cis}\)/trans), with a focus on cases of transgressive hybrid expression. Compared with previous intraspecific hybrid studies that used inbred parental lines, the use of Cirsium parents from natural outcrossing populations more accurately reflects the ecological context in which natural hybridization occurs.

Materials and Methods

Cirsium arvense Reference Transcriptome

Given the challenges of de novo short read assembly, the Illumina paired-end reads were mapped to a reference transcriptome previously assembled using reads generated by Roche 454 pyrosequencing of cDNA libraries derived from flower, leaf, and root tissues from an invasive genotype of C. arvense (Lai et al. 2012). This accession derives from a North Dakotan population with collection ID: NW-22-1-M. The total number of unfiltered raw reads in the leaf, flower, and root libraries were 609,458, 565,129, and 636,795, respectively.

Before assembly, the 454-generated sequence data were cleaned to remove uninformative and contaminating sequences from the cDNA libraries using SnoWhite v1.1.2 (Barker et al. 2010). SnoWhite is a pipeline of custom scripts and existing programs, Seqclean and TagDust (Chen et al. 2007; Lassmann et al. 2009), designed to trim off the poly-(A/T) tails and filter out adaptor and primer sequences. The proportion of reads eliminated during cleaning steps for the three libraries ranged from 0.9% to 3.2%.

MIRA3 short read aligner was used to assemble the cleaned 454 reads into contiguous sequences (Chevreux et al. 2004). The output contigs from MIRA3 were fed into another assembler, CAP3, to correct for MIRA’s tendency to break apart highly expressed contigs due to assumptions of normalized input (Huang and Madan 1999). There were 88,374 contigs on the reference transcriptome when we required more than 94% similarity for final assembly of contigs with CAP3. The quality of this assembly was validated using the DupPipe tool to model the age (\(K\)) distribution of duplicated genes (contigs). The observation of a peak corresponding to the ancient Compositae duplication at \(K\) approximately 0.6 is consistent with expectations, indicating no problems with over- or underassembly (Barker et al. 2010).

Sample Preparation

The following Cirsium populations were selected to achieve parentage representative of distinct geographic populations. Dr. Alessia Guggisberg collected seeds from the natural native and invasive Cirsium populations used in this study, 2800808-2 (Romania) and KN-ON (Ontario), respectively, in the summer of 2008. In this experiment, the female parent seed derives from the Romanian population and the male parent is from Ontario. Vouchers are available in the Rieseberg laboratory at the UBC Biodiversity Research Centre in Vancouver, BC.

Seeds were scarified with sandpaper and germinated in Petri dishes incubated at 28 °C (22 °C when dark) in a germination chamber running a 16-h photoperiod. After 1–2 cm of root growth, newly germinating seedlings were potted in soil, comprising 75% regular peat moss and 25% perlite (pH of 5.5–6.5), and moved to a growth chamber set for 23 °C and a 16-h photoperiod. Fertilization was controlled by enclosing all unopened flower heads in 3” \(\times\) 4” mesh bags and perforated plastic bags. As an imperfectly dioecious species, male and female individuals were crossed pairwise by brushing pollen directly onto the elongated stigmas of receptive female flowers. Fruits fully matured approximately 2 weeks after pollination, evident by the characteristic drying out and release of the achenes from the capitulum. Collected \(F_1\) seeds were germinated and raised as described earlier. Once 4–6 true leaves emerged, seedling leaf tissue was collected and stored at \(-80^\circ\)C. Parental tissue from the same developmental stage was acquired via clonal propagation from 1 to 2 cm long root fragments. All tissues were harvested during the same time of day to control for photoperiod effects on differential gene expression.

RNA Extraction

Total RNA was extracted from parental and hybrid seedling leaf tissues using a phenol-based Trizol protocol followed by treatment with RNase-free DNase I (Qiagen). Biological replicates for the parents were not possible given that our primary goal was to assess expression changes in hybrids that result from intraspecific hybridization. For this purpose, exact parents were required when comparing with hybrids, so that alleles can be tracked accurately. It is important to point out that
we are not trying to comment directly on differences between native versus invasive populations. Separate extractions were performed for each of 60 \( F_1 \) hybrids, which were pooled before sequencing forming two biological replicates (each \( n = 30 \)). A spectrophotometer was used to calculate final RNA concentrations for the male, female, and hybrid pools 1 and 2. Agarose gel electrophoresis confirmed RNA integrity and the absence of genomic contamination.

**Illumina Sequence Data Acquisition and Read Mapping**

Non-normalized cDNA libraries were prepared from each of the parental and pooled hybrid RNA using the standard mRNA-seq protocol (Illumina, San Diego, CA) at the Indiana University Center for Genomics and Bioinformatics (http://cgb.indiana.edu/), where they were subsequently sequenced on Illumina’s GA-II platform. Four sequencing lanes were used on an Illumina GA II flowcell. Both parental libraries were sequenced in their own lanes, and hybrid pools 1 and 2 were each sequenced separately. The millions of resulting raw paired-end reads (2 \( \times \) 76 bp) were trimmed using a custom Python script to remove the low-quality stretches typically found at the 3'-end of GA-II Illumina sequence data. All scripts are available upon request.

MOSAIK (v1.1.0018) short-read aligner was used to map the Illumina-generated reads to the contigs of the reference transcriptome, such that the depth coverage of reads mapped to a particular locus provided a direct readout of mRNA transcript abundance (Hillier et al. 2008). Parameters were optimized for identifying true single-nucleotide polymorphisms (SNPs) by including only uniquely mapped reads and allowing for 8 mismatches per read (~10% read length). MosaikAligner parameters used: -hs 10 -m unique -mm 8 -bw 29 -act 35 -mhp 100.

**Normalization and Filtering**

A total-count scaling method of normalization was applied to standardize the total number of reads between lanes, so that the average expression level across all genes would be the same in each of the four libraries (Robinson and Smyth 2007; Marioni et al. 2008; McManus et al. 2010). This step prevented against making misleading inferences of expression divergence due solely to technical differences in sequencing depth between libraries. Normalized count data were rounded to the nearest integer to satisfy the requirements of downstream statistical (e.g., binomial) tests. For brevity, the normalized count data for the four libraries is hereafter referred to as \( P_1 \) (invasive male parent), \( P_2 \) (native female parent), Hyb\(_1\), and Hyb\(_2\). Expression of each allele \((P_1 \text{ and } P_2)\) in the hybrids is given as Hyb\(_1\)\(_P_1\), Hyb\(_2\)\(_P_1\), Hyb\(_1\)\(_P_2\), and Hyb\(_2\)\(_P_2\). The average total expression of the hybrid pools was used for some analyses, where \( \text{Hyb}_{\text{avg}} = (\text{Hyb}_1 + \text{Hyb}_2)/2 \).

Before assessing parental expression divergence, a minimum read-count cutoff was imposed to filter out low-coverage sites where there is generally a poor signal-to-noise ratio. A cutoff of \( P_1 + P_2 \geq 20 \) was chosen, and all contigs satisfying this threshold were retained for further analysis. This same cutoff was used with a similar RNA-Seq data set (McManus et al. 2010).

**Patterns of Gene and ASE Divergence**

For establishing patterns of overall gene expression, transcript abundance was observed as the normalized total reads mapped per contig, with the exception of instances where one parent had a read count of zero; in these cases, a 0 \( \rightarrow \) 1 adjustment was made to satisfy the cis and trans tests’ requirement of positive integers (provided that the minimum coverage threshold of \( P_1 + P_2 \geq 20 \) was met before adjustment). In contrast, the depth of reads mapped at individual SNP sites was used to quantify ASE. In both analyses, binomial exact tests were used to identify statistically significant differences in expression between any two genotypes. For all statistical tests performed, the resulting \( P \) values were adjusted to correct for multiple testing using Q-value software (Storey and Tibshirani 2003; Storey et al. 2004). A conservative false discovery rate (FDR) of 0.5% was imposed as the threshold for significance.

**Inheritance Classifications**

Statistical software, R (v 2.9.0 CRAN), was used to sort genes by the mode of expression inheritance based on the results of the pairwise binomial tests between parents and also between each parent and the hybrid. The percentage of the total reads per library that mapped to a given contig was log transformed to simplify interpretation of comparisons between genotypes and graphical representation of the data.

Log-transformed percent expression was subtracted between parents and hybrids to establish the direction and magnitude of change, where positive values for \( \log(P_1) - \log(P_2) \) indicate that \( P_1 > P_2 \), and the value itself gives a direct readout of the magnitude fold change; this is the same as taking the log-transformed ratio, \( \log(P_1/P_2) \). Parents were compared with each other and with the hybrids to assign a significant pattern of inheritance for each contig (e.g., \( P_1 \sim F_1 > P_2 \)). Based on these tests, each contig was categorized according to a commonly used system of describing additive and dominant modes of gene action (Swanson-Wagner et al. 2006; McManus et al. 2010).

Determining conserved expression between two genotypes was based on the results of significance (binomial) tests, where more than 0.5% FDR indicated failure to reject the null hypothesis of similar expression. Contigs where hybrid expression was not similar to either parent but fell within a midparental range were classified as additive, and contigs for which expression in the hybrid was either above high-parent
or below low-parent were considered transgressive and classified as overdominant and underdominant, respectively. Contigs for which expression in the hybrid was similar (FDR < 0.5%) to one of the two parents were classified as dominant for that parent.

Overdominant and underdominant genes were further investigated for possible enrichment in genes showing cis x trans mechanism of regulatory divergence or an over-representation of certain cellular processes. A conservative list of transgressive genes, requiring a minimum 1.25 fold-change in expression between the F1 hybrid and near-parent (NP), was used for gene ontology (GO) term enrichment analysis. For these contigs, gene functions were predicted based on sequence homology to annotated Arabidopsis thaliana gene models in the TAIR9 database (blastx using an e-value cutoff of e^-10).

ASE Patterns Reveal Cis- and Trans-Regulatory Effects

SNPs were identified as single-base differences between the Illumina-generated reads in the hybrid alignments and the 454-generated reference transcriptome using the SAMtools (v 0.1.13) sort, index, and pileup functions (Li et al. 2009). SAMtools varFilter script was applied to allow a maximum variant coverage of 10,000 to a minimum quality threshold (Phred-score) of 20 for the final SNP calls. Once SNPs were identified in the individual libraries, a custom Perl script was applied to produce a list of SNP sites at which both parents were homozygous for differences. Fixed polymorphisms between parents were required to assign a parental origin unambiguously to hybrid reads, an important consideration in studies of ASE. All scripts are available upon request.

As described in the introduction, gene expression can be altered due to changes in either cis- or trans-acting regulators. To determine their relative contributions, we applied a test that compares the ratio of expression of the two parental alleles in F1 hybrids with the relative expression of the same alleles in the homozygous parents. In the genetic regulatory background of the hybrid, both parental alleles are exposed to a common set of trans-acting factors, and, therefore, any observed hybrid allelic imbalance is characteristic of cis-regulatory divergence. Trans-acting variation can then be detected by comparing hybrid allelic imbalance back to the parental ratio.

A binomial exact test (FDR 0.5%) was used to determine significant cis-effects, based on allelic imbalance in the hybrids, and the extent of cis-effects was quantified as: $\text{cis} = \log_2(P_1/P_2)$. Statistically significant trans-effects were identified using a Fisher’s exact test (FDR 0.5%) to compare the ratios of allelic expression between hybrids and parents. By comparing the hybrid allelic expression ratios to the allelic ratio observed between parents, the magnitude and direction of trans-acting regulatory change are revealed, such that an equal allele frequency (0.5 A1, 0.5 A2) in the hybrids means that only trans-acting changes underlie the expression divergence between parents: $\text{trans} = \log_2(P_1/P_2) - \log_2(P_1/P_1)$.

Custom R (v 2.9.0 CRAN) scripts were designed to sort all contigs into their mechanism of regulatory divergence based on the results of binomial and Fisher’s tests as well as the direction of change (as in Landry et al. 2005; McManus et al. 2010). Contigs were categorized as “cis-only” or “trans-only” if parental expression divergence was observed as well as significant cis-effects or trans-effects, but not both. “Cis + trans” refers to cases where cis- and trans-effects were both significant and promoted expression of the same allele, whereas “cis x trans” indicates counteracting effects of cis- and trans-regulators. Compensatory interactions were identified as the subset of “cis x trans” interactions that resulted in no parental divergence, where counteracting cis- and trans-effects effectively cancelled each other out. A “conserved” pattern of regulation was assigned to cases where there was no significant cis-effect, trans-effect, or parental expression divergence. A low frequency of contigs examined was scored as “ambiguous” based on patterns of significance tests that could not be categorized according to these criteria.

GO Prediction and Enrichment Analysis for Misexpressed Contigs

For predicting gene function, we used the best-hit Arabidopsis annotation resulting from a Blastx search, which used misexpressed contig sequences as a query against a database of all Arabidopsis peptide representative gene models (TAIR9). The result with the lowest E value was inferred to be orthologous to the given queried gene. Enrichment analysis for transgressive contigs was done using AmiGO (v 1.7) to investigate possible over-representation of certain gene functions (Carbon et al. 2009). A background set of Arabidopsis best-hit gene products for all contigs in our data set was used for comparison.

Results

Sequencing and Mapping

Native and invasive parental cDNA libraries were sequenced, as were two pooled hybrid cDNA libraries, using an RNA-seq approach. A total of 82,713,256 paired-end read sequences were obtained from the Illumina runs, corresponding to 20–21 million raw read-pairs for each of the parental and the hybrid libraries. Reads were then trimmed for quality and mapped using Mosaik short-read aligner to the 454-generated reference C. arvense transcriptome reported in Lai et al. (2012). Bases were trimmed to remove terminal bases below Q of 20, with trimmed reads shorter than 40 bp and sequences having internal Ns removed entirely. An average of 3.6 bp were trimmed off of each read, resulting in an average final read length of 72.4 bp. Mapping results are presented in table 1. An average of 82.4% of the reads mapped across all
Expression Divergence between Native and Invasive Parents

Following paired-end read mapping and normalization, mRNA abundance was summarized for each contig in the parental and hybrid alignment files using SAMTools pileup function (Li et al. 2009). Significance of parental expression divergence was scored using binomial exact tests with an FDR correction of 0.5% (as in McManus et al. 2010). Using this significance threshold, 35,120 contigs (51%) of the 68,746 that passed the coverage filter were scored as divergently expressed between parents, regardless of magnitude of change (supplementary fig. S1A, Supplementary Material online). Of the 35,120 differentially expressed contigs, significantly more (71%, n = 24,863) had higher expression in the invasive parent, P1 (fig. 1; log2(P1/P2) > 0), compared with just 10,257 contigs that had higher expression in the native parent, P2 (binomial exact test \( P = 4.9e^{-324} \)). After imposing a required minimum fold difference of 1.25\( \times \), the number of differentially expressed contigs between parents was substantially reduced to 12,446 (18%), indicating many subtle differences in transcript abundance between the parents (fig. 1A and B). A 1.25-fold minimum has been used in similar studies (McManus et al. 2010). Using this cutoff, there were still significantly more genes with higher expression in the invasive parent, with 9,597 (77%) contigs exhibiting higher expression in P1 and only 2,849 (23%) contigs with higher expression in P2 (Binomial P = 4.9e\(-324\)). Upon imposing a more stringent 1.5-fold minimum expression difference cutoff, 9,412 contigs (13.7% of total) remained differentially expressed between parents, with the same majority (7,251 contigs; 77%) still showing higher expression in P1 than P2, compared with just 2,161 (23%) with higher expression in P2.

Modes of Expression Inheritance Include Extensive “Hybrid Effects”

Once parental divergence was assessed, genes were sorted into their modes of expression inheritance (additive, dominant, or transgressive) in the hybrids based on pairwise comparisons of total expression levels, using normalized total mapped reads per contig, between parents and hybrids (supplementary fig. S1A–C, Supplementary Material online). Categorization of contigs based on their expression inheritance pattern was achieved by considering significant binomial test results (0.5% FDR) and the direction of divergence, regardless of magnitude fold-change (see Materials and Methods). Significant patterns of expression inheritance were reported for 454-generated EST data, comprising 88,374 unigene sequences.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>Hybrid Pool 1</th>
<th>Hybrid Pool 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total read pairs</td>
<td>19,705,882</td>
<td>19,658,330</td>
<td>19,086,096</td>
<td>19,351,444</td>
</tr>
<tr>
<td>Filtered out</td>
<td>5,231,547</td>
<td>7,778,752</td>
<td>6,810,056</td>
<td>7,490,454</td>
</tr>
<tr>
<td>Total aligned</td>
<td>34,180,217</td>
<td>31,537,908</td>
<td>31,362,136</td>
<td>31,212,434</td>
</tr>
<tr>
<td>Percent aligned</td>
<td>86.70</td>
<td>80.20</td>
<td>82.20</td>
<td>80.60</td>
</tr>
</tbody>
</table>

Note.—Illumina GAII-generated sequence reads were mapped to a Cirsium reference transcriptome, assembled from publically available 454-generated EST data, comprising 88,374 unigene sequences.
a majority (62%) had \(P_2\)-like expression patterns in the hybrid. However, those contigs exhibiting \(P_1\)-like hybrid expression were more often more highly expressed than \(P_2\). Of the differentially expressed contigs that showed \(P_1\)-like dominance patterns, 31% had lower expression in hybrids than in \(P_2\) (\(P_2 > P_1 \sim F_1\); binomial exact \(P = 3.2\, 10^{-223}\)), whereas a majority (74%) of those that had \(P_2\)-dominance patterns had lower expression in the hybrids than in \(P_1\) (\(P_1 > P_2 \sim F_1\); \(P = 4.9\, 10^{-324}\); fig. 2D); these included 9,986 cases of high-parent dominance and 13,720 cases of low-parent dominance (fig. 2D). Of the contigs showing high-parent dominance in the hybrids, 4,663 (66%) showed \(P_1\)-like expression (\(P = 7.3\, 10^{-166}\)) compared with only 2,077 contigs (19%) of the low-parent dominance cases exhibiting \(P_1\)-like expression (\(P = 4.9\, 10^{-324}\); fig. 2E).

For some of the most transgressive contigs, the magnitude of expression change between hybrid and NP is summarized in supplementary table S1A, Supplementary Material online, along with the observed pattern of expression inheritance and predicted gene type. The top 50 overdominant contigs that exhibit the largest fold changes between hybrid and the NP are shown in supplementary table S1A, Supplementary Material online, whereas the top 50 underdominant contigs are shown in supplementary table S1B, Supplementary Material online. Transgressive contigs with above high-parent and below low-parent expression were found to have NP fold changes of 0.02–6.41 and 0.02–4.45, respectively. Sequence similarity to functionally annotated \textit{A. thaliana} genes enabled prediction of biological processes. Of the overdominant cases, 12,083 (87%) could be matched to a best-hit orthologous gene (representing 4,815 unique gene models) along with 6,328 (92%) of the underdominant cases (representing 2,652 unique gene models). Transgressively expressed contigs are involved in numerous functional roles (supplementary tables S1A and S1B, Supplementary Material online).

**ASE Analysis**

Next, ASE patterns in the hybrids were analyzed. To infer hybrid ASE levels, parent-specific SNPs were identified and used to discern between alleles in hybrids. Custom Perl scripts were used to generate a list of SNP sites in the hybrids at which the parents were fixed for different alleles (i.e., using only homozygous loci in each parent). This analysis required expression in both parents at each locus, so that genotypes could be determined unambiguously and parental alleles tracked in the hybrids; this is especially important because these parents are descendants of natural, outcrossing (heterozygous) populations. Of the 94,181 resulting SNP sites, 62,281 (66.1%) passed the filter for minimum coverage in the parents (\(P_1 + P_2 \geq 20\times\)). The distribution of the informative SNP sites among contigs is explored in supplementary figure S2, Supplementary Material online. The total number of contigs represented by these 20×-filtered SNPs was 24,377 contigs, representing 27.6% of the possible 88,374 assembled reference sequence contigs.

Normalized mapped read-depth coverage at SNP sites in the hybrid and parental alignments was used to quantify the expression of alleles. The expression of both parental
Fig. 2.—Modes of hybrid expression inheritance are investigated for the 68,746 contigs that passed the filter for minimum expression. (A) A scatterplot comparing total expression between Hyb_Avg and P_1 (x axis), and between Hyb_Avg and P_2 (y axis), enabled sorting of contigs into their modes of expression inheritance as described in text. (B) and (C) Total expression (log_2 normalized read counts) is compared between hybrid and invasive and native parents, left to right, respectively. Barplots (D) and (E) further explore the cases of parental dominance in hybrid expression patterns.
alleles ($P_1$ and $P_2$) was compared with their corresponding expression in hybrids ($F_1P_1$ or $F_1P_2$, respectively) for each informative SNP site. The expression of allele $P_2$ in hybrids more closely resembled parental $P_2$ expression ($\tau = 0.51$; fig. 3E) compared with allele $P_1$, whose expression showed more differences between hybrids and parents ($\tau = 0.13$; fig. 3D). The expression of allele $P_1$ differed between parent and hybrids at 43,843 (70.4%) of SNP sites based on binomial tests with FDR = 0.5%. By comparison, there were fewer differences in expression of allele $P_2$ between parents and hybrids, with 32,099 sites (51.2%) showing significant differences. Of the SNP sites supporting differential $P_1$ allele expression between generations, a majority (87.9%) had higher expression in parents than hybrids (binomial exact test $P = 4.94e^{-324}$). In contrast, a majority (88%) of the contigs showing $P_2$ allele expression differences between parents and hybrid had higher expression in hybrids compared with parents (binomial exact test $P = 4.9e^{-324}$).

In the hybrids, $P_2$ alleles were significantly more often expressed in greater abundance than $P_1$ alleles (binomial exact test $P = 4.9e^{-324}$), as indicated by the excess (84%) of negative values for $\log_2(F_1P_1/F_1P_2)$ (fig. 3C). Consistent with this observation, this analysis also found far more cases of probable monoallelic expression of the $P_2$ allele in the hybrids, supported by 22,703 (36.5%) of informative SNP sites, compared with only 165 SNPs (0.3%) at which only allele $P_1$ is expressed in hybrids.

ASE Tests Reveal Cis- and Trans-Regulatory Divergence

Next, all 62,281 SNP sites that passed the minimum coverage filter were scored based on the nature of their regulatory divergence by comparing parental and hybrid allelic expression ratios. Cis- and trans-regulatory effects were revealed for all informative SNP sites, including those exhibiting parental expression divergence. It should be noted that inferred parental divergence used in this allele-specific analysis was based on relative coverage at loci corresponding to hybrid SNP sites rather than on total reads mapped per contig, as was reported earlier. This was required because the hybrid allelic data are reported at the SNP level and so the parental data must be as well to make meaningful comparisons. Cases where either parent has an expression of 0 were dropped from this analysis because parental expression is a prerequisite for tracing the origin of hybrid alleles.

Significant trans-regulatory divergence was found at 43,571 (70%) informative SNP sites, and 42,223 (68%) showed evidence of significant cis-regulatory divergence (FDR 0.5%; fig. 3C; supplementary fig. S1F and S1E, Supplementary Material online, respectively). Evidence for both cis- and trans-effects were exhibited by 32,598 (52%) sites and were further divided into three categories commonly used in studies of regulatory evolution (Landry et al. 2005; McManus et al. 2010): “cis + trans,” “cis x trans,” and “compensatory,” based on the direction and magnitude of regulatory change as described in Materials and Methods section. Of these, 3,176 (9.8%) were categorized as cis + trans, 7,342 (22.5%) as cis x trans, and 22,080 (67.7%) as compensatory (fig. 3C). A total of 7,817 SNP sites that did not show evidence of either parental expression divergence or significant cis- or trans-effects were classified as “conserved.” An “ambiguous” pattern was found for 18,624 SNP sites.

The median significant trans-regulatory difference between parents (3.5-fold) was significantly different than the median cis-regulatory difference between parents in direction but not in magnitude (−3.6-fold). In other words, cis-regulatory divergence tended to cause higher expression of the native $P_2$ allele, whereas trans-regulatory divergence tended to drive higher expression of the invasive $P_1$ allele (Wilcoxon’s $P < 2.2 \times 10^{-16}$, fig. 3A). Trans-regulatory divergence correlated more highly with expression differences between parents than cis-regulatory divergence ($P < 2.2e^{-16}$, $\tau = 0.33$; $P < 2.2e^{-16}$, $\tau = -0.13$, respectively). Furthermore, the proportion of total expression divergence between parents explained by trans-effects (% trans) increased with the magnitude of divergence (fig. 3B).

Types of Regulatory Divergence Influence Patterns of Overall Expression Inheritance

The next analysis tested the hypothesis that cis-regulatory mutations tend to have additive effects on hybrid expression, whereas mutations acting in trans are more often subject to dominance interactions. The proportion of regulatory change due to cis-regulatory divergence (median % cis) was significantly lower for sites categorized as additive (32%) than those categorized as nonadditive (46%; Wilcoxon’s rank-sum test, $P < 2.2e-16$; fig. 4A), indicating that trans-regulatory divergence contributes more to expression changes between parents for sites that showed additive inheritance as opposed to those that show nonadditive inheritance (fig. 4A). The magnitude of the cis-regulatory divergence between parents was not significantly different for genes with additive (median = 3.32-fold) than those with nonadditive (median = 3.08-fold) expression. The percent cis-regulatory divergence underlying transgressive and nontransgressive modes of gene action was also compared, and no significant difference was detected between transgressive and nontransgressive gene sets (Wilcoxon’s $P < 2.2e-16$; fig. 4B).

Genes with Transgressive Expression Are Enriched for Cis + Trans Interactions

Using the SNP-level data, we also tested the hypothesis that cases of transgressive expression are enriched for cis x trans mechanisms of regulatory divergence, based on a study of interspecific Drosophila hybrids, which found an enrichment of cis x trans in the set of genes exhibiting transgressive hybrid expression (Landry et al. 2005). Of the 14,196 sites exhibiting...
FIG. 3.—(A) Absolute magnitude (fold-change) of parental divergence resulting from cis- and trans-effects. (B) A box-and-whisker plot showing the percent of cis-effects for genes binned based on the magnitude of expression divergence between parents. (C) Scatterplot showing relative allelic expression levels in parents and hybrids; results used to sort genes into categories based on their mechanism of regulatory evolution as described in Materials and Methods. Scatterplots (D) and (E) compare parental and hybrid allelic expression for alleles $P_1$ and $P_2$, respectively.
nontransgressive (i.e., additive, dominant, or conserved) patterns of inheritance, 50% had cis x trans mechanisms of regulatory divergence compared with just 34% (n = 914) cases of hybrid over- or underexpression (Fisher’s exact test P value = 2.8e−09). In contrast, 34% of transgressive SNP sites showed cis + trans regulatory divergence, compared with just 20% of nontransgressive sites (Fisher’s exact test P value = 6.95e−14). In this intraspecific analysis, nontransgressively expressed sites showed a significantly higher frequency of cis x trans allelic interactions, whereas transgressive sites were enriched for cis + trans allelic interactions.

**Discussion**

**Parental Expression Divergence**

By sequencing total mRNA from two C. arvense parents and their F1 hybrid, significant expression divergence was found to affect 51% of contigs. That a majority (65%) of differentially expressed contigs had fold-changes of less than 1.25 indicate many subtle differences between populations. Significantly more contigs had higher expression in the invasive parent (P1) than the noninvasive parent (P2). One caveat is that these expression data represent only a single time point in development and, in this case, are specific to seedling leaf tissue. In addition, no biological replicates from the population were used for the outcrossing parents because comparisons with hybrids required using the exact heterozygous parents to track alleles. Because the parents from a single cross and their hybrids were used, the observed parental divergence may not be relevant directly to the context of native versus invasive populations. Instead, these results provide a more general estimate of the level of expression divergence between parents from natural, divergent populations.

**Hybrid Expression Patterns Reveal All Modes of Gene Action**

This study found evidence for all modes of gene action, including dominant and transgressive (especially overdominant) allelic interactions contributing to parental expression divergence. A majority (81%) of differentially expressed contigs showed nonadditive (i.e., dominant, overdominant, and underdominant) patterns of expression inheritance in hybrids, whereas only 10% showed additive modes of gene action. Interestingly, most (62%) of the dominant cases had P1-like expression patterns in the hybrid, although those with P2-like expression were more often more highly expressed. Extensive nonadditive inheritance of gene expression has also been found within *Drosophila melanogaster* (Gibson et al. 2004), and 84% of differentially expressed genes showed nonadditive inheritance in an interspecific hybrid of *D. melanogaster* and *D. sechellia* (McManus et al. 2010). Numerous studies have found evidence that parental divergence is often caused by dominant and transgressive (especially overdominant) modes of gene action (Rieseberg et al. 2003; Landry et al. 2005; Swanson-Wagner et al. 2006).

The 30% of differentially expressed contigs classified as transgressive in this study is lower than that has been reported in interspecific *Drosophila* hybrids (35–69%; Ranz et al. 2004; McManus et al. 2010). The frequency of overdominance was significantly higher than the frequency of underdominance. Some studies of interspecific hybrids observed that underexpression is more common than overexpression (Ranz et al. 2004; McManus et al. 2010), although others found that overexpression contributed more to parental divergence (Swanson-Wagner et al. 2006). Although transgressive segregation seems to be especially common in interspecific systems (Rieseberg et al. 2003; Landry et al. 2005; Ranz et al. 2004),
nonadditivity has also been found in F1 hybrids from inbred D. melanogaster lines and in hybrids of Arabidopsis ecotypes (Gibson et al. 2004; Zhang and Borevitz 2009, respectively). Transgressive hybrid effects are generally less common within than between species (Landry et al. 2007; Wittkopp et al. 2008b), although exceptions exist (McManus et al. 2010).

ASE Analysis

In the hybrid, the P2-derived (maternal) allele was significantly more often more highly expressed than the P1-derived (paternal) allele in the hybrid, suggesting that an excess of regulatory changes have increased expression of the P2 alleles in the hybrids. This analysis also identified potential cases of allele-specific silencing supported by 36.8% of SNPs and revealed that monoallelic expression of P2 is much more common in hybrids than monoallelic expression of P1. Allele-specific silencing has been found to occur upon hybridization in F1 hybrids (Adams 2007; Springer and Stupar 2007) and may even do so selectively in only specific organ types (reviewed in Adams 2007).

Cis- and Trans-Regulatory Divergence

By incorporating ASE data, the nature of regulatory evolution underlying the various modes of gene action was also revealed. Cis-acting changes affected a similar number of contigs as trans-acting changes (68% and 70%, respectively), and over half of all contigs (52%) exhibited both types of regulatory divergence. This is consistent with results from other intraspecific ASE studies including one using hybrids between Arabidopsis ecotypes, which found similar frequencies of cis- and trans-effects but slightly more cases of cis-variation (Zhang and Borevitz 2009). Results from ASE studies within the yeast Saccharomyces cerevisiae found a significantly higher number of trans-regulated genes (Sung et al. 2009; Emerson et al. 2010). In contrast, interspecific comparisons have generally found higher relative contributions of cis-effects to parental expression divergence, including between closely related fruit fly species D. melanogaster and D. simulans (Wittkopp et al. 2004, 2008a; Graze et al. 2009). Other interspecific reports of a high proportion of cis-effects come from studies of poplar, maize, and yeast (Zhuang and Adams 2007; Springer and Stupar 2007; Tiross et al. 2009, respectively). To our knowledge, this study is novel among plants in its application of the cis/trans test to obligate outcrossers, although a similar approach was used to score cis regulatory divergence in humans (Serre et al. 2008) using only loci at which parents were fixed for different alleles. The cis-regulatory differences detected in this study are more likely to reflect intraspecific polymorphisms rather than being the consequence of divergent natural selection; the latter is unlikely given the relatively small genetic distance between native and invasive parents (i.e., an intraspecific system). Instead, these cis-differences likely reflect differences in the standing genetic variation between populations. One caveat relating to the inferred proportion of trans-effects in this study is that our test will have captured some novel allelic interactions that only arise in the hybrid and which do not actually contribute to the expression divergence observed between parents (Graze et al. 2009; Zhang and Borevitz 2009).

Of the 10,518 sites that showed evidence for both cis- and trans-effects, a majority (70%) had antagonistic cis- and trans-regulatory interactions (cis x trans). This is consistent with studies from both flies and yeast finding an excess of these cases in which cis- and trans-variation have opposite effects on allelic expression (Landry et al. 2005; Tiross et al. 2009, respectively). Such "compensatory" interactions have been interpreted as evidence that stabilizing selection acts to maintain gene expression levels (Birchler and Veitia 2010). The median magnitude of cis-divergence was similar to the median magnitude of trans-divergence, but the direction of their effects differed: cis-variation tended to drive higher expression of the native P1 allele, whereas trans-varation tended to cause higher expression of the invasive P1 allele.

Compared with cis-effects, trans-regulatory variation correlated more highly with expression differences between parents and played a larger role with increasing magnitude of parental divergence. Generally, most of the trans-acting changes detected in interspecific studies of ASE have had only minor overall effects on gene expression divergence, whereas the magnitude of expression divergence caused by cis effects tended to be greater (Wittkopp et al. 2004, 2008a; Tiross et al. 2009; Birchler and Veitia 2010). However, in one study using closely related Drosophila species, expression differences between parents have been shown to generally correlate more strongly with trans-divergence than with cis-divergence (McManus et al. 2010).

Correlations between Cis- or Trans-Regulatory Divergence and Mode of Gene Action

Finally, we looked for correlations between mode of gene action and the mechanism of regulatory divergence. Trans-effects contributed more to expression divergence between parents for sites that showed additive (68% trans) as opposed to nonadditive (54% trans) inheritance patterns. This is in contrast to evidence suggesting the typical effect of trans-regulatory variation is to cause a departure from additivity, due to a susceptibility to dominance interactions, whereas cis-regulatory variation tends to have additive effects on gene expression divergence (Lemos et al. 2008; Stern and Orgogozo 2008).

Relative to contigs exhibiting additive and dominant modes of inheritance, transgressive contigs were enriched for a "cis + trans" mechanism of regulatory divergence. These data fit a model in which cis and trans divergence with similar effects on allelic expression contribute to transgressive expression in the hybrids. Indeed, genes with "cis + trans" regulatory
evolution have been found to be more likely to exhibit divergent parental expression as a result of directional selection (Tirosh et al. 2009). However, the opposite trend has also been documented in interspecific Drosophila hybrids, in which “cis x trans” regulatory interactions were found to be more common among transgressively expressed genes (Landry et al. 2005; McManus et al. 2010).

Concluding Remarks
This work contributes to the body of genomic resources being developed for an assortment of weedy species in the Compositae family (Stewart et al. 2009; Lai et al. 2012). As demonstrated here, transcriptome sequencing using high-throughput methods enables studies of the effects of hybridization on gene expression in even nonmodel genetic organisms. A greater degree of nonadditivity (i.e., proportion of dominant and transgressive cases) was observed in this study, which used parents from natural outcrossing populations, compared with earlier work which relied on crosses between inbred parental lines. This is a departure from the typical view that transgressive hybrid effects are generally less common within than between species and may reflect the more complex genetic contributions of the heterozygous parents, compared with homozygous parents, to the hybrid’s regulatory environment. Functional characterization of transgressively expressed contigs will further help to resolve their importance as potential contributors to the adaptive success of hybrids in an ecological context. Future comparisons of gene expression among a wider sampling of genotypes, tissues, and developmental stages may help predict genes that contribute to transgressive hybrid traits. Unique insight was also gleaned into the mechanisms of regulatory evolution that prevail in intraspecific systems. This work confirms important contributions of both cis- and trans-regulatory evolution underlying parental expression divergence yet reveals a slightly greater contribution of trans-effects than has been typically observed in crosses between selfers. These findings fit a proposed model that adaptation proceeds through a combination of cis- and trans-regulatory mutations, with a significant contribution of trans-effects.

Supplementary Material
Supplementary figures S1 and S2 and tables S1 and S2 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments
The authors thank Sally Otto for advice on statistics. Also they acknowledge Alessia Guggisberg for collecting the C. arvense parental populations. This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC), including a Special Research Opportunities grant to L.H.R. and K.L.A. and by a Discovery Grant to K.L.A.

Literature Cited


Associate editor: Michael Purugganan