New Phytologist

Research

Recent mating-system evolution in *Eichhornia* is accompanied by *cis*-regulatory divergence

Ramesh Arunkumar, Teresa I. Maddison, Spencer C. H. Barrett and Stephen I. Wright

Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada

Author for correspondence:
Ramesh Arunkumar
Tel: +1 416 978 5603
Email: ram.arunkumar@utoronto.ca

Received: *23 September 2015*
Accepted: *30 January 2016*

New Phytologist (2016) **211:** 697–707
doi: 10.1111/nph.13918

Key words: allele-specific expression (ASE), deleterious mutations, *Eichhornia paniculata*, evolution of selfing, outcrossing bias, transcriptomes.

Summary

- The evolution of predominant self-fertilization from cross-fertilization in plants is accompanied by diverse changes to morphology, ecology and genetics, some of which likely result from regulatory changes in gene expression. We examined changes in gene expression during early stages in the transition to selfing in populations of animal-pollinated *Eichhornia paniculata* with contrasting mating patterns.
- We crossed plants from outcrossing and selfing populations and tested for the presence of allele-specific expression (ASE) in floral buds and leaf tissue of F_1 offspring, indicative of *cis*-regulatory changes.
- We identified 1365 genes exhibiting ASE in floral buds and leaf tissue. These genes preferentially expressed alleles from outcrossing parents. Moreover, we found evidence that genes exhibiting ASE had a greater nonsynonymous diversity compared to synonymous diversity in the selfing parents.
- Our results suggest that the transition from outcrossing to high rates of self-fertilization may have the potential to shape the *cis*-regulatory genomic landscape of angiosperm species, but that the changes in ASE may be moderate, particularly during the early stages of this transition.

Introduction

The evolution of self-fertilization (selfing) from cross-fertilization (outcrossing) is the most frequent reproductive transition in flowering plants. This change in mating system has a major influence on the morphology, ecology and genetics of selfing lineages. Populations with high selfing rates generally have smaller and less showy flowers, often lacking physical separation between stigmas and anthers, and usually exhibit reduced pollen-to-ovule ratios and nectar production (Darwin, 1876; Lloyd, 1965; Cruden, 1977). Due to the independent origin among angiosperm families of these convergent floral changes they are jointly referred to as the 'selfing syndrome' (Stebbins, 1957; Ornduff, 1969; Sicard & Lenhard, 2011). Predominant selfing can also have significant demographic and biogeographical consequences as this mating strategy allows individuals to reproduce under pollinator- and mate-limited conditions, and to found colonies following dispersal events (Lloyd, 1980). Theoretical studies also indicate that changes in mating system from outcrossing to selfing can have important genomic consequences (reviewed in Charlesworth & Wright, 2001; Wright *et al.*, 2008; Barrett *et al.*, 2014). The reduction in effective population size following the transition to selfing can be associated with the accumulation of weakly deleterious mutations, higher mutation and recombination rates, and altered representation of selfish genetic elements in selfing genomes, when compared to outcrossing genomes. Genetic analyses of populations have detected both minor- and major-effect mutations governing morphological traits promoting self-pollination (e.g. *Turnera* – Shore & Barrett, 1985, 1990; *Mimulus* – Fishman *et al.*, 2002, 2015; *Capsella* – Sicard *et al.*, 2011; Slotte *et al.*, 2012). However, the underlying molecular genetic changes involved in transitions from outcrossing to selfing remain largely uncharacterized.

The evolution of selfing may often be associated with changes in gene expression, but little is known about this potential association. Gene expression is controlled by the interaction between linked elements proximal to the gene (*cis*-regulatory elements) and unlinked elements that occur elsewhere in the genome (*trans*-acting elements) (Carroll, 2005). *Cis*-regulatory elements are of particular interest as mutations in these regions have a lower chance of resulting in deleterious pleiotropic effects compared to those in coding regions, or at *trans*-acting elements (Emerson & Li, 2010; Wittkopp & Kalay, 2012). A common approach for isolating the effects of *cis*-regulatory divergence is to generate hybrids between related species and test for allele-specific expression (ASE) (Pastinen, 2010). For example, using RNA-seq data, Bell *et al.* (2013) detected ASE in 51% of genes (*c.* 35 000) in hybrids generated from invasive and noninvasive populations of *Cirsium arvense*. In a study of hybrids between wild teosinte and domesticated maize lines, Lemmon *et al.* (2014) reported that 70% of genes (*c.* 17 000) exhibited ASE; however, only 1079 showed consistent differences among all lines

indicating standing genetic variation in both the wild and domesticated samples. Recently, significant genetic variation associated with ASE was detected at *c.* 6000 genes in a single population of the outcrossing *Capsella grandiflora* (Josephs *et al.*, 2015). These studies detected widespread ASE in populations of the same species, or between closely related taxa.

Shifts in gene expression may accompany mating-system evolution for several reasons. First, adaptive phenotypic changes accompanying the transition in mating system may be mediated by shifts in expression of genes involved in modifications to floral morphology and/or life history (Sicard & Lenhard, 2011). Second, species and population differences in the genomic distribution and/or regulation of transposable elements (TEs) can drive shifts in gene expression of adjacent genes through the effects of TE silencing on nearby genes (Lister *et al.*, 2008; Hollister *et al.*, 2011). Third, genome-wide reductions in the efficacy of selection on regulatory sequences in selfing populations (Charlesworth & Wright, 2001), and/or relaxed selective pressures due to loss of pollinators, and life history evolution (Lloyd, 1980) could each lead to changes in gene expression. Finally, recent theory also predicts that a proliferation of DNA regions that enhance gene expression can accumulate in outcrossing genomes, but inbreeding and selfing have the potential to reduce the rate of accumulation of such elements (Fyon *et al.*, 2015).

The influence of mating system evolution on gene expression is largely unexplored in plants. Using SNP arrays, He *et al.* (2012) found that *c.* 2205 out of the 14 462 genes that were assayed showed evidence for ASE in hybrids generated from outcrossing *Arabidopsis lyrata* and selfing *A. thaliana*. Strikingly, the outcrossing *A. lyrata* allele was preferentially expressed over the selfing allele for 90% of the ASE genes, and these genes accumulated more nonsynonymous mutations than synonymous mutations when compared to non-ASE genes. He *et al.* (2012) attributed the observed outcrossing bias in allelic expression to transposable element-induced gene silencing in *A. thaliana*. Generally, increased expression intensity is associated with slower protein evolution due to stronger selection against protein misfolding (Drummond & Wilke, 2008), protein–protein misinteractions (Yang *et al.*, 2012) and mRNA misfolding (Park *et al.*, 2013). Therefore, gene silencing could be associated with the accumulation of nonsynonymous mutations in the downregulated ASE genes in *A. thaliana* (He *et al.*, 2012). As the proportion of genes displaying ASE in floral bud and leaf tissues in *A. thaliana* were similar, there was no direct evidence indicating that *cis*-regulatory divergence is associated specifically with adaptive differences in floral morphology or function. Steige *et al.* (2015) reported that 44% of 18 452 genes showed evidence for ASE in hybrids generated from more closely related outcrossing and selfing *Capsella* species. In contrast to the *Arabidopsis* results, they observed preferential expression of selfing alleles for ASE genes. Analysis of small RNA expression indicated that the observed bias may have resulted from transposable element-induced silencing of outcrossing alleles due to their greater representation near ASE genes. Although they did not observe an enrichment of ASE genes in floral tissues of *Capsella*, a number of genes showing ASE patterns were found in previously identified regions containing quantitative trait loci (QTL) for floral and reproductive traits, consistent with the hypothesis that some of the genes exhibiting ASE may have been subject to adaptive regulatory evolution in association with the evolution of the selfing syndrome. Currently, the strength and extent of the association of *cis*-regulation with mating-system variation has been assessed in a single plant family only, and the general nature of the association during the early stages of mating-system divergence remains unknown.

Here, we investigate whether divergence in *cis*-regulatory elements has accompanied the recent evolution of selfing from outcrossing in the diploid, annual *Eichhornia paniculata* (Pontederiaceae). Outcrossing in *E. paniculata* is promoted by the floral polymorphism tristyly, in which populations are composed of three floral morphs with a reciprocal arrangement of stigma and anther heights (Darwin, 1877; Barrett, 1992). By contrast, selfing populations exhibit stigmas and anthers at the same position within a flower and self-pollinate autonomously. The change in spatial separation of stigmas and anthers in selfing variants is governed by recessive modifiers that cause elongation of stamens to the same height as the stigma (Fenster & Barrett, 1994). Whereas tristylous populations possess large, showy, blue-purple flowers, those in selfing populations are much smaller and possess traits characteristic of the selfing syndrome (Morgan & Barrett, 1989). In *E. paniculata* the transition to selfing has occurred on multiple occasions and is associated with long-distance dispersal from Brazil to the Caribbean and Central America (Husband & Barrett, 1993; Barrett *et al.*, 2009). Ness *et al.* (2010) estimated that the colonization of the Caribbean occurred *c.* 120 000 yr ago, suggesting that the transition to selfing occurred relatively recently. Bottlenecks and the shift to selfing are associated with a reduction in genetic diversity in populations (Barrett & Husband, 1990; Ness *et al.*, 2010). Recent studies report reduced codon usage bias and a small increase in the proportion of potentially deleterious and effectively neutral mutations in selfing populations, findings consistent with a genome-wide reduction in selection efficacy (Ness *et al.*, 2012; Arunkumar *et al.*, 2015). *Eichhornia paniculata* therefore provides a valuable system to explore the associations among selection efficacy, *cis*-regulatory divergence and mating-system transitions.

We tested for the occurrence of ASE in F_1 offspring from four crosses between different outcrossing genotypes from Brazil and independently derived selfing genotypes from the Caribbean and Central America. We investigated ASE in two contrasting plant tissues: floral bud tissue with gametophytically and sporophytically expressed genes, and leaf tissue that only expresses sporophytic genes, and analyzed the biological function and patterns of selection acting on ASE genes. We addressed the following specific questions. (1) What proportion of genes sampled from F_1 plants exhibit ASE, and is the observed ASE associated specifically with modifications to floral architecture associated with the evolution of selfing? If the evolution of the selfing syndrome in flowers was the major factor influencing *cis*-regulatory variation, we would expect to see a strong signal of ASE in floral tissue but only minimal levels in leaf tissue, and genes showing ASE to be

involved in outcrossing floral function. However, a variety of other factors unrelated to floral evolution could also influence patterns of *cis*-regulation, including changes in vegetative and life-history traits associated with the demographic history and ecology of selfing populations, genome-wide changes in TE distributions, or alterations in the efficacy of selection. If these factors play an important role in influencing ASE expression we might expect to see minimal differences in levels of ASE between floral and leaf tissue. (2) Do genes displaying ASE preferentially express outcrossing or selfing alleles, and are there any differences in the proportion of nonsynonymous and synonymous mutations between ASE and non-ASE genes in outcrossing and selfing populations? If transitions to selfing are accompanied by widespread gene silencing, we might expect to see higher expression of outcrossing alleles and an accumulation of nonsynonymous mutations in selfing compared to outcrossing populations. Conversely, outcrossing alleles might be underexpressed if rare mutations resulting in gene silencing in outcrossers were absent in selfers, due to stochastic loss or the reduced activity of transposable elements during the transition to selfing (see Ågren *et al.*, 2014; Steige *et al.*, 2015).

Materials and Methods

Sampling parents and generating F_1 crosses

We obtained open-pollinated seeds from four outcrossing *Eichhornia paniculata* (Spreng.) Solms. populations from NE Brazil and four selfing populations from Jamaica ($n = 2$), Cuba and Mexico. We used the morph structure of populations as a determinant of their mating systems, as trimorphic and monomorphic populations are predominantly outcrossing and selfing, respectively (Barrett & Husband, 1990). We grew plants under uniform glasshouse conditions at the University of Toronto and chose one individual from each population and crossed each outcrosser to a unique selfer (Supporting Information Table S1). For each cross, the outcrosser and selfer were the paternal and maternal parents, respectively. We emasculated maternal parents 4 h before anther dehiscence and applied pollen from the outcrossing paternal parent using forceps. Approximately 2 wk after each cross, we collected mature seeds and stored them in dry conditions. Following germination of F_1 crosses 8 months later, we chose one plant from each of the four crosses for further study.

RNA extraction and sequencing

We extracted RNA from floral buds and leaf tissue. We note that what we refer to as 'leaf tissue' is technically tissue obtained from elongated internodes or petioles subtending 'leaves' (see Richards & Barrett (1984) for details of the organography of *E. paniculata*). The reproductive and vegetative tissue was obtained from each of the eight parents and four F_1 offspring for sequencing. We extracted RNA using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich). We used the extracted RNA samples to make Illumina TruSeq RNA libraries that were sequenced using the 100 bp paired end protocol on three lanes of the Illumina HiSeq 2000 at the McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada). The parental bud transcriptomes were sequenced as part of Arunkumar *et al.* (2015) on lanes where 12 samples were multiplexed in each lane. The raw sequence data for the parental bud transcriptomes are available under accession number SRP049636 at the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra, accessed 2 February 2016) and the associated BioProject alias is PRJNA266681 (http://www.ncbi.nlm.nih.gov/bioproject/, accessed 2 February 2016). For this study, eight libraries were prepared from the leaf tissue of parents: four were leaf tissue of F_1 progeny and four were floral buds of F_1 progeny. The 12 libraries from leaf tissue were multiplexed on each lane. We sequenced each of these libraries twice, with a total of two lanes dedicated to these samples. However, we used only one of the two sequenced libraries for parental leaf tissue in our analyses to keep the sequencing depth among the floral bud and leaf parental transcriptomes the same. By contrast, we used both of the sequenced libraries prepared from the leaf tissue of the F_1 progeny in our analyses. The four libraries prepared from floral buds of F_1 progeny were sequenced on an individual lane. The raw sequence data for the parental leaf transcriptomes and the F_1 floral bud and leaf transcriptomes are available under accession number SRP060405 at the Sequence Read Archive and the associated BioProject alias is PRJNA288861.

Read mapping

We previously generated a 65.53-Mbp *de novo* transcriptome assembly ($N_{50} = 2.2$ Kbp) from selfing transcriptomes from six populations in the Caribbean using the programs Velvet 1.2.08 (Zerbino & Birney, 2008) and Oases 0.2.08 (Schulz *et al.*, 2012) and predicted coding regions from BLAST searches to plant databases (Arunkumar *et al.*, 2015). Further, we filtered paralogous loci by removing contigs containing sites heterozygous across multiple selfing populations, and those for which coverage of mapped genomic reads were greater than the genomic sequencing depth. We mapped the RNA-seq short reads and genomic reads with Burrows-Wheeler Aligner v0.7.8-r455 (Li & Durbin, 2009) using default parameters and used the Sampe command to combine the paired end read mapping results. Then, we used the Stampy 1.0.23 software (Lunter & Goodson, 2011) with default parameters to map more divergent reads and identify insertions and deletions (indels). We processed the read mapping output using the SamFormatConverter, ReorderSam, AddOrReplaceReadGroups and BuildBamIndex programs from of the Picard tools package 1.124 using default settings (http://picard.sourceforge.net, accessed 2 February 2016). We assigned the libraries prepared from leaf tissue of the F_1 progeny that were sequenced on separate lanes to the same read group. We also used the MarkDuplicates program, part of the Picard tools package 1.124, to identify and tag duplicate reads. As erroneous mismatches might have occurred near indels, we used the RealignerTargetCreator and IndelRealigner programs, part of the Genome Analysis Toolkit (GATK)

3.3-0-g37228af (DePristo *et al.*, 2011), with default parameters to realign sequences within *c.* 3 kb of an indel. We subsampled the reads from the resulting BAM files of F_1 floral bud transcriptomes to 67% of their original value using Samtools view command (SAMTOOLS v.1.2; Li *et al.*, 2009) to control for the higher sequencing depth of these libraries compared to those prepared from the leaf tissue of the F_1 progeny.

Identifying genes showing ASE

We ran the UNIFIEDGENOTYPER program, also part of GATK 3.3-0-g37228af (DePristo *et al.*, 2011), with the BadCigar read filter to identify invariant sites, SNPs and indels from all samples and we retained sites with Phred-scaled quality scores > 60 and sequencing depth > 20. We retained SNPs if their Phred-scaled genotype quality for all parental and progeny samples from the floral bud and leaf tissue was > 60 and if they were at least 5 bp away from either side of an indel. We used custom Perl scripts to identify sites that were homozygous for alternative SNPs in both parents and heterozygous in F_1 plants. For these sites, we extracted the depth of the outcrossing and selfing alleles in the F_1 transcriptomes from the UNIFIEDGENOTYPER output. We used the depth information to estimate the posterior probability that a gene shows ASE using hierarchical Bayesian inference developed by Skelly *et al.* (2011). We compared the level of ASE detected using this Bayesian inference to that detected using a binomial exact test, a widely applied method for estimating the level of differential expression (Robinson *et al.*, 2010; Skelly *et al.*, 2011). Although the binomial test has a higher false discovery rate (FDR) when compared to Bayesian methods (Skelly *et al.*, 2011; León-Novelo *et al.*, 2014), it is a powerful approach when low quality sites are excluded (Castel *et al.*, 2015). As we lacked genomic reads from the parents, we adopted the genomic model inferred for humans and yeast by Skelly *et al.* (2011). In their paper, they summarized the genomic model using the *a hat* and *d hat* parameters. For both species, the *a hat* ranged from 3000 to 6500 indicating that both parental copies were expressed at approximately equal levels in the genomic data and *d hat* of *c.* 550 indicating that noise around this 50 : 50 expectation for each gene was small. Following their model, we used large values for both parameters (*a hat* = 5000, *d hat* = 500). Even so, we investigated the effect of increasing or decreasing noise in the genomic model by repeating the analysis using 10-fold larger and smaller values of *a hat* and *d hat* parameters. We ran 500 000 MCMC iterations, sampling every 100th iteration to check for convergence. We ran 2000 scaling iterations allowing for a maximum of eight rounds of scaling. After convergence, we calculated the posterior probability that each gene showed ASE. We also investigated the presence of ASE by performing a two-sided negative binomial exact test on each gene. To do this, we averaged the reads counts across all SNPs for each gene using EDGER v.3.1 (Robinson *et al.*, 2010). We only used genes that had at least one site heterozygous in all progeny. We used a FDR of 5% with a null hypothesis that paternal and maternal alleles were expressed equally to call ASE genes using this approach.

Filtering ASE calls

Although the lack of availability of parental genomic sequence required us to explore a range of parameters in the Bayesian analysis, availability of transcriptome data from one outcrossing genotype and genomic data from a cross between selfing populations allowed us to identify and filter some potential SNPs showing spurious ASE, or ASE reflecting standing variation within outcrossing populations. We inferred ASE in the outcrossing parents by estimating allelic expression for sites heterozygous in each outcrosser. We excluded from our analyses genes that had posterior probabilities of ASE > 0.7 in transcriptomes generated for outcrossing parents, as such genes likely reflect ASE present in natural populations unrelated to differences between the outcrossing and selfing genotypes. To exclude spurious ASE sites caused by possible mapping biases, we also inferred the extent of ASE when genomic reads from two *E. paniculata* individuals were mapped to the transcriptome reference. One individual was a selfer from Oaxaca, Mexico and the other was a genotype derived from a cross between a selfing individual from Slipe, Jamaica and the aforementioned selfer from Mexico. We calculated allelic depth for sites heterozygous in each individual for identifying ASE and used the DNA model scripts generated by Skelly *et al.* (2011) to estimate the *a hat* and *b hat* parameters. We excluded from our analyses of the genomic read data genes that had posterior probabilities of ASE > 0.7, as genomes would be uninformative for assessing differential expression, and any observed ASE may be spurious and associated with technical variability in read mapping.

Molecular functional analyses

For genes displaying evidence for ASE, we performed custom translated nucleotide BLAST searches (Altschul *et al.*, 1990), with *E. paniculata* genes as the query against the *Arabidopsis* protein database obtained from The Arabidopsis Information Resource (TAIR) (Lamesch *et al.*, 2012). We performed a singular enrichment analysis, using the genes expressed in each tissue as the query and the TAIR10 *Arabidopsis* gene model as the reference, using the agriGO toolkit and database (Du *et al.*, 2010). We found that significant gene ontology terms were found after performing hypergeometric tests with the Yekutieli multiple test correction at a significance level of 0.05.

Calculating expression intensity

We calculated the number of reads that mapped to the transcriptome reference assembly from the floral bud and leaf transcriptomes of the parents and F_1 offspring using HTSeq-count 0.6.1p1 (Anders *et al.*, 2015) under the union mode. For genes with significant evidence for ASE, we averaged the allelic depth across all sites to calculate the mean expression intensity for the outcrossing and selfing alleles in the F_1 progeny. Further, we estimated the expression levels for these ASE genes in the parental floral bud and leaf transcriptomes by using HTSeq-count 0.6.1p1. We assessed if comparisons of expression levels between

outcrossing and selfing alleles in the F_1 progeny, and expression levels of ASE genes in the outcrossing and selfing parental transcriptomes, were statistically significant using R (R Development Core Team, 2011) by first generating bootstrap replicates, after resampling randomly across the genes. We performed a two-tailed permutation test comparing the overall sums for bootstrap datasets to generate a *P*-value for each comparison.

Comparing nonsynonymous and synonymous diversity

We performed a 2×2 contingency test to investigate if there was interaction between nonsynonymous and synonymous polymorphisms and floral bud and leaf ASE genes. For this test, we summed the number of polymorphisms observed across all ASE genes in each tissue. Further, we compared genetic diversity (π) for nonsynonymous ($\pi_{nonsynonymous}$) and synonymous sites ($\pi_{synonymous}$) for genes with ASE based on posterior probabilities > 0.7 against genes without evidence for ASE. All genes with ASE would have had heterozygous site(s). To reduce potential directional biases due to invariant genes from the set that did not show ASE, we only analyzed genes with one or more heterozygous sites. We excluded the genotype from Mexico when calculating π for selfing populations. This genotype was a result of an independent transition to selfing (see Arunkumar *et al.*, 2015), and therefore substitutions between genotypes from separate transitions may have been identified as SNPs and thus distorted estimates of π. Our ASE gene set only included those with posterior probability for ASE > 0.7 across all of the remaining three progeny. All other genes were defined as non-ASE genes. We randomly sampled three outcrossing genotypes from the four sequenced parents to keep the number of chromosomes being compared the same. We estimated π using the Polymorphorama script (Andolfatto, 2007; Haddrill *et al.*, 2008) and generated mean values and 95% confidence intervals (CI) across 1000 bootstrap replicates, after resampling randomly across the genes using R (R Development Core Team, 2011). To obtain the 95% CI, we excluded bootstrap replicates with mean values below the 2.5 percentile and above the 97.5 percentile. Then, we used the smallest and largest values from the rest to represent the lower and upper limits of the CIs, respectively. We performed permutation tests, as in the previous section, to assess if $\pi_{nonsynonymous}/\pi_{synonymous}$ comparisons between ASE and non-ASE genes were statistically significant. This resampling approach would make it less likely that a few loci with extremely high or low levels of genetic diversity would lead to the erroneous detection of significant differences in the comparisons of ASE and non-ASE genes.

Results

Allele-specific expression

The distribution of coverage of genes from floral bud and leaf transcriptomes from parents and offspring were similar, but there were many genes that were not expressed in leaf transcriptomes (Fig. S1). We identified 60 000–70 000 sites homozygous for alternative SNPs in outcrossing and selfing parents and heterozygous in F_1 progeny, which we used to test for the presence of ASE in floral bud and leaf tissues (Table S2). Of these sites, only 3118 and 2836 sites were heterozygous in floral bud and leaf tissues, respectively, across all four F_1 progeny, and 1798 sites were heterozygous in both tissue types in all four progeny. 450–800 genes sampled from the floral bud tissue had probability values > 0.7 (Fig. 1a), but only 53 genes showed ASE across all four F_1 genotypes. Similarly, 350–650 genes sampled from leaf tissue had probability values > 0.7 (Fig. 1b), but only 46 genes showed ASE across all four F_1 genotypes. Only five genes showed ASE from both floral bud and leaf tissues across all four progeny at this posterior probability threshold. Note that a 2×2 contingency test did not reveal an interaction between tissue type – that is, floral bud or leaf – and the number of nonsynonymous and synonymous polymorphisms summed across all ASE genes for each tissue at the 5% significance threshold. Genes displaying ASE in floral buds and leaves were involved in a range of functions including growth, binding, regulation, maintaining cell structure and defense response (Table S3), but there was no clear association to floral function for any of the genes.

We tested the effects of varying the parameters and approaches for detecting ASE. First, we varied the *a hat* and *d hat* parameters, which represent parameters quantifying technical variability, using the Skelly *et al.* (2011) approach. When we used larger *a hat* (50 000) and *d hat* (5000) parameters to identify genes showing ASE in floral buds of one of the F_1s, the mode of the posterior probability distribution shifted to the right (Fig. S2). In this case, many more genes had larger posterior probabilities for ASE. By contrast, the use of smaller *a hat* (500) and *d hat* (50) parameters (allowing for greater overdispersion due to technical variability) resulted in a distribution similar to the original, using an *a hat* of 5000 and *d hat* of 500 (Fig. 1a), with a similar number of genes with probabilities > 0.7. Further, a negative binomial exact test indicated that of the 6360 genes that had at least one site heterozygous across the floral buds of all four F_1 progeny, 764 showed significant evidence for ASE at a 5% FDR threshold.

Expression fold difference

We calculated differences in expression intensity of outcrossing and selfing alleles in F_1 plants for genes with posterior probability of ASE > 0.7. We chose this probability threshold because there were few ASE genes with larger posterior probability values (Fig. 1). Strikingly, almost all genes with ASE in floral bud (Fig. 2a) and leaf tissues (Fig. 2b) had a 1.5–2-fold increase in expression of outcrossing compared to selfing alleles, across all comparisons. Permutation tests indicated that the differences in expression intensity of outcrossing and selfing alleles were significant at the 0.1% threshold for all comparisons. This result does not appear to be due to mapping biases; although we observed a general outcrossing bias in allelic expression, it was often the selfing allele that matched the reference allele in the three progeny with Caribbean

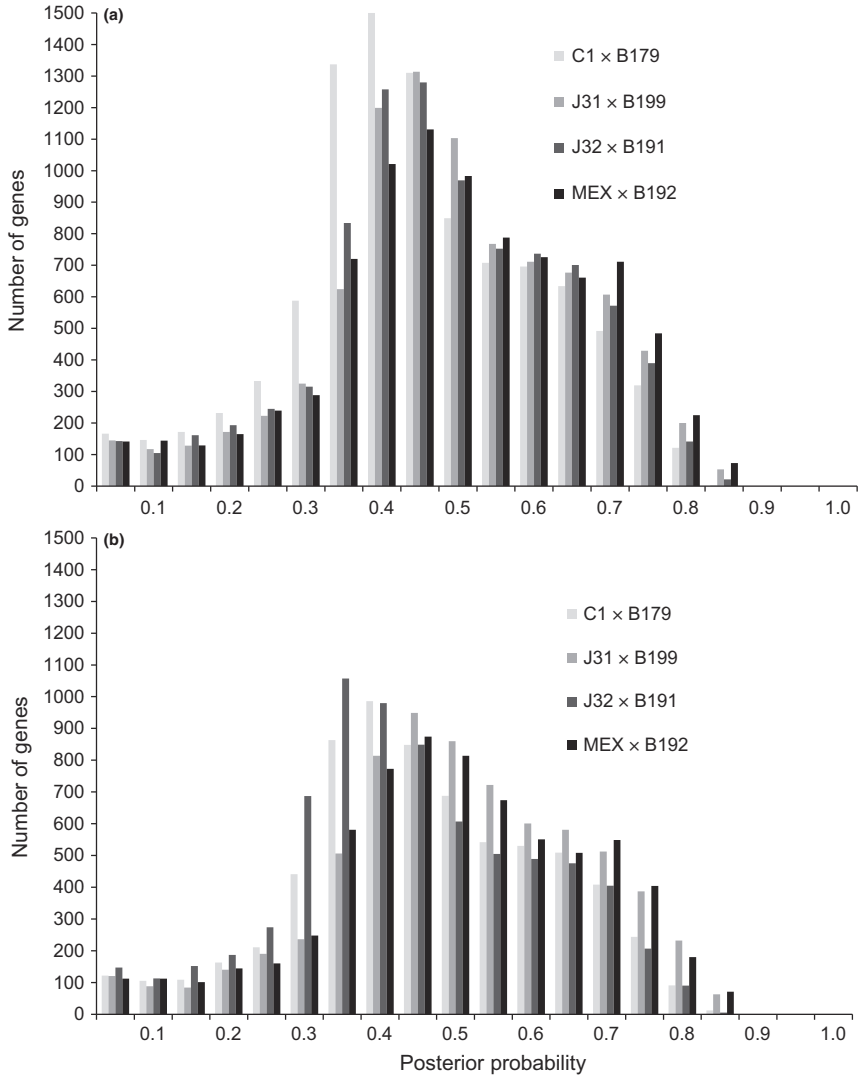

Fig. 1 Distribution of posterior probabilities for genes showing allele-specific expression (ASE) in *Eichhornia paniculata*. Probability distributions are illustrated for genes sampled from (a) floral buds and (b) leaf tissue of F_1 plants generated from crosses between selfing (maternal) and outcrossing (paternal) parents. We used a Bayesian binomial test as implemented in Skelly *et al.* (2011) to determine the probability of ASE against the null expectation of equal expression intensity of both alleles.

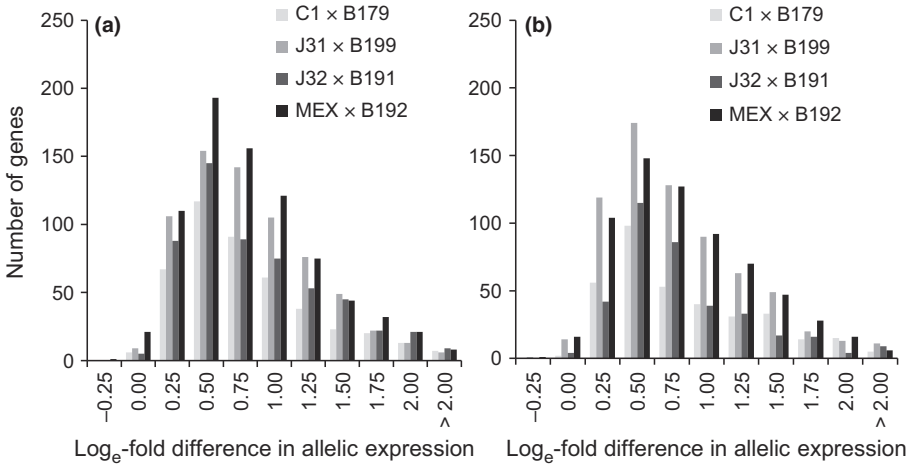

Fig. 2 Distribution of log$_e$-fold expression difference between outcrossing and selfing alleles of genes with posterior probability for allele-specific expression > 0.7 in F_1 plants of *Eichhornia paniculata*. Illustrated are the distributions for genes sampled from (a) floral buds and (b) leaf tissue of F_1 plants. Large (> 0) and small (< 0) values indicate higher expression of outcrossing and selfing alleles, respectively.

selfing maternal parents, because of the source of our reference transcriptome (Table S2). For the progeny of the selfing parent from Mexico, the outcrossing allele matched the reference allele for *c.* 40% of sites and the selfing allele matched the reference allele for the remaining 60%. For this progeny, we consistently observed an outcrossing bias in allelic expression intensities for ASE genes, when repeating the calculations using only sites where the outcrossing allele (Fig. S3a), or selfing allele (Fig. S3b), matched the reference allele. By contrast, for genes without significant evidence for ASE, both alleles were expressed at approximately equal levels regardless of which site types were analyzed (Fig. S3a–c). Genes showing ASE in floral bud and leaf tissues of the F_1 plants also had a general trend of increased expression levels in outcrossing compared to selfing parents (Fig. S4a,b), although this difference was not significant at the 5% level across all comparisons (Table S4).

Relative proportions of nonsynonymous and synonymous mutations

We compared $\pi_{nonsynonymous}/\pi_{synonymous}$ for 106 floral bud and 86 leaf tissue genes that had posterior probability > 0.7 for ASE across the three progeny with Caribbean selfing maternal parents against all other genes (posterior probability < 0.7). We randomly subsampled 106 floral bud and 86 leaf tissue genes from the set without significant evidence for ASE to keep the number of genes being compared the same. As expected, there was a *c.* 80% and 50% reduction in diversity for ASE and non-ASE genes, respectively, in floral bud and leaf tissues of selfing parents when compared to outcrossing parents (Table S5). The percentage diversity reductions at nonsynonymous and synonymous sites were similar. ASE genes for floral bud and leaf tissues of outcrossing parental genotypes, $\pi_{nonsynonymous}/\pi_{synonymous}$ was not significantly different from the estimate based on all other genes (Fig. 3a,b). By contrast, $\pi_{nonsynonymous}/\pi_{synonymous}$ was significantly higher for ASE genes compared to all other genes in floral buds of parental selfing genotypes (Fig. 3c). Although a similar trend of higher $\pi_{nonsynonymous}/\pi_{synonymous}$ was also observed for ASE in leaf tissue compared to non-ASE genes, this difference was not significant at the 5% threshold (Fig. 3d). Note that a non-normal distribution of $\pi_{nonsynonymous}$ and $\pi_{synonymous}$ values resulted in lower $\pi_{nonsynonymous}/\pi_{synonymous}$ when we resampled individual estimates for each gene to generate bootstrap replicate datasets (Fig. 3) compared to a single global estimate (Table S5). The former approach is less susceptible to deviations from assumptions of normality. The $\pi_{nonsynonymous}/\pi_{synonymous}$ estimates remained significantly different between ASE and non-ASE genes when we repeated the analyses, after restricting the ASE gene set to those with posterior probabilities for ASE > 0.75 across the three progeny. By contrast, when the posterior probability threshold for ASE genes was reduced below 0.65, there were no longer significant differences in $\pi_{nonsynonymous}/\pi_{synonymous}$ between ASE and non-ASE genes.

Discussion

We detected allele-specific expression for genes sampled from floral bud and leaf tissue of F_1 plants from crosses between outcrossing and selfing genotypes of the annual plant *E. paniculata*.

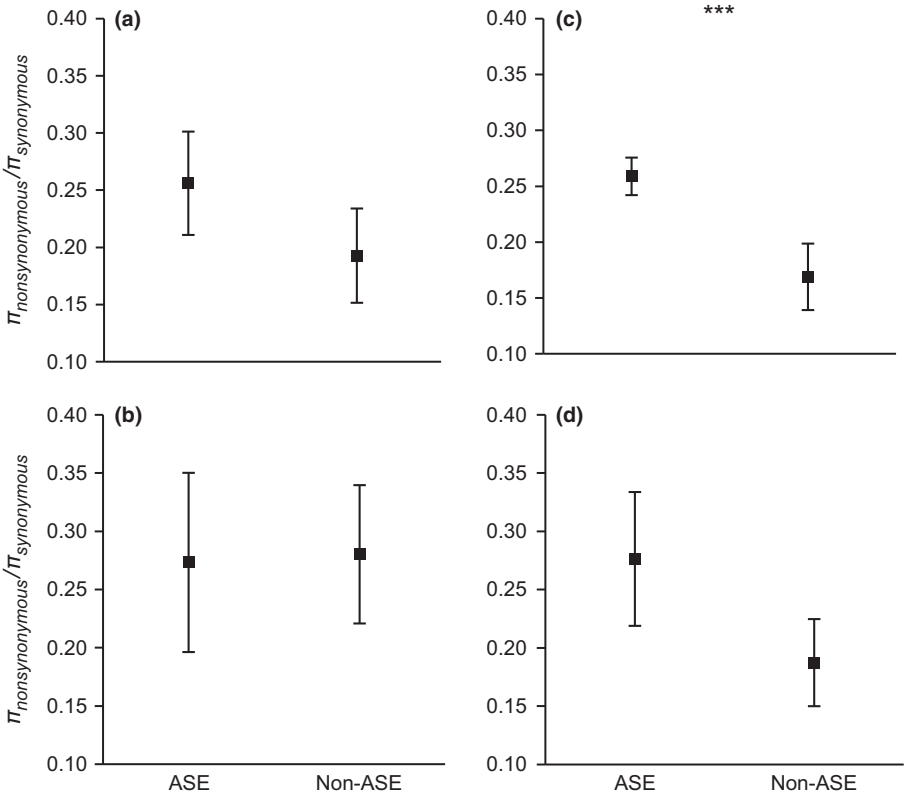

Fig. 3 Diversity at nonsynonymous ($\pi_{nonsynonymous}$) and synonymous sites ($\pi_{synonymous}$) for genes with posterior probability > 0.7 for allele-specific expression (ASE), and for genes without evidence of ASE (posterior probability < 0.7) in *Eichhornia paniculata*. Illustrated are the mean $\pi_{nonsynonymous}/\pi_{synonymous}$ estimates for ASE and non-ASE genes in: (a) floral buds of outcrossing genotypes, (b) leaf tissue of outcrossing genotypes, (c) floral buds of selfing genotypes and (d) leaf tissue of selfing genotypes. Three outcrossing genotypes from NE Brazil and three selfing genotypes from the Caribbean were used to estimate π. All analyzed genes had one or more heterozygous sites. There were 106 and 86 genes showing ASE in floral buds and leaf tissues, respectively. We randomly subsampled from the set of non-ASE genes to keep the number of genes compared the same. 95% confidence intervals were generated from 1000 bootstrap replicates, resampling across genes. ***, Comparison significant at the 0.1% level based on permutation tests.

However, only moderate levels of *cis*-regulatory differentiation accompanied the transition to selfing, probably because our study involved an intraspecific comparison of mating system differentiation in which selfing is of relatively recent origin, and we factored out genes with evidence of allele-specific expression (ASE) segregating within outcrossing populations. The majority of genes that we investigated showed biased expression of the outcrossing allele in both floral bud and leaf tissue. When compared against genes without evidence for ASE in F_1 plants, genes with ASE had more nonsynonymous mutations than synonymous mutations in selfing parents but not outcrossing parents. Below, we discuss the possible reasons for the associations between *cis*-regulatory variation, nonsynonymous polymorphism and mating system divergence in *E. paniculata*.

The evolution of selfing is accompanied by differences in *cis*-regulation

Our results are consistent with the hypothesis that there is an association between mating system divergence and patterns of ASE. We observed preferential expression of the outcrossing allele in genes showing ASE in floral and leaf tissues, similar to patterns reported from *Arabidopsis* (Chang *et al.*, 2010; He *et al.*, 2012), and in contrast to those from *Capsella* (Steige *et al.*, 2015). We would have expected preferential mapping of selfing alleles if read mapping bias significantly influenced the detection of ASE (Degner *et al.*, 2009; Stevenson *et al.*, 2013), because the transcriptome reference we used was from Caribbean selfing populations (Arunkumar *et al.*, 2015). However, this was not the case as we observed a strong bias in the expression of outcrossing alleles. Genes showing ASE in progeny also showed generally greater expression intensities in outcrossing compared to selfing parents. The difference in level of gene expression between the two parents was marginal in some of the parental pairs across both tissue types. This difference may be associated with the reduced power, owing to the lower sequencing depth of the parents when compared to the F_1 offspring. Alternatively, it is possible in some cases that compensatory *trans* effects are present, whereby *cis*-regulatory changes revealed in F_1 hybrids are compensated for by *trans* effects in the parental populations (Bell *et al.*, 2013).

Given that we observed as much ASE in leaves as flower buds, there is no clear evidence for specific regulatory changes associated with the evolution of the selfing syndrome in *E. paniculata*. Patterns of ASE in this species also may be associated with adaptation to new biogeographical zones with potentially novel environments, as a result of the colonization of the Caribbean and Central America following long-distance dispersal from Brazil (Ness *et al.*, 2010). However, it is not clear why such changes would consistently be associated with reduced expression in selfing populations. Instead, the overall patterns we observed may be more likely driven by general shifts in gene regulation caused by enhancer evolution (Fyon *et al.*, 2015), transposable element (TE)-mediated gene silencing (Lister *et al.*, 2008; Hollister *et al.*, 2011) and/or changes in the strength or efficacy of selection on ASE genes following the transition to selfing (Charlesworth & Wright, 2001). Elevated nonsynonymous polymorphism at downregulated genes is consistent with a reduced strength and/or efficacy of selection, because it implies that this set of genes in particular has experienced less constraint. The finding of ASE in both floral and leaf tissue suggests that a variety of evolutionary forces have probably shaped *cis*-regulatory variation accompanying the change in mating system in *E. paniculata*.

Another factor that can contribute to allele-specific expression is genomic imprinting. In flowering plants imprinting has been found to primarily affect endosperm (Messing & Grossniklaus, 1999; Köhler & Weinhofer-Molisch, 2010; Nodine & Bartel, 2012), although more recent evidence from *A. thaliana* suggests that imprinting may also affect the contribution of alleles from the maternal and paternal parents to the embryo (Raissig *et al.*, 2013). Brandvain & Haig (2005) predicted that in crosses between an outcrossing paternal parent and a selfing maternal parent, maternally inherited genes might exhibit imprinting. Under this scenario, the outcrossing allele might be expressed preferentially in progeny, similar to the allelic bias we observed. However, we collected floral buds from F_1 progeny before any possibility of autonomous selfing, and thus it seems unlikely that the ASE we observed resulted from early endosperm development. Although our experimental design minimized the influence of genetic imprinting, it cannot be ruled out completely as a possible contributor to the observed patterns of ASE. A recent study suggested that parent-of-origin effects on gene expression may exist in later stages of development (Videvall *et al.*, 2015). It is also worth noting that in studies of allele-specific expression involving mating system differences, (*Capsella* – Steige *et al.*, 2015; *Arabidopsis* – He *et al.*, 2012; this study) the paternal allele showed ASE bias in all cases. Reciprocal crosses between outcrossing and selfing genotypes might provide further insights on the potential contribution of genetic imprinting and/or inheritance patterns of small RNAs to ASE.

Patterns of ASE and association to the evolution of the selfing syndrome

The observed *cis*-regulatory differences in selfers and outcrossers could be due to a variety of causes including changes to floral function and life history traits associated with the evolution of selfing, genomic changes such as enhancer evolution (Fyon *et al.*, 2015) or genome-wide reductions in the efficacy of selection (Charlesworth & Wright, 2001). As there was no significant interaction between tissue type and ASE shaping nonsynonymous and synonymous diversity, or evidence for an enrichment of genes involved in floral or reproductive function, selection on floral function is clearly not the only factor associated with the downregulation of selfing alleles in F_1 plants of *E. paniculata*.

The evolution of the selfing syndrome is unlikely to explain the overall patterns of regulatory evolution that we observed. However, this does not rule out a role for regulatory evolution in morphological changes associated with the shift to selfing. Some of the genes showing ASE in floral tissue may have important roles in traits associated with the selfing syndrome. We detected 100 genes showing ASE in floral buds and not leaves; similarly,

85 genes of the 1305 genes displaying ASE in floral buds of hybrid *Capsella* were found within genetic regions associated with floral trait differentiation between selfers and outcrossers (Steige *et al.*, 2015). In comparison to the number of ASE genes identified, genetic mapping studies have found only 1–6 QTLs governing changes in 6–10 traits associated with floral morphological differences in *Mimulus* (Fishman *et al.*, 2002, 2015) and *Capsella* (Sicard *et al.*, 2011; Slotte *et al.*, 2012). Although the number of genes showing ASE is higher than the number of QTLs associated floral phenotypic differentiation in these earlier studies, there may be multiple genes occurring within the QTL regions identified. Also, some QTLs underlying floral traits may comprise 'master regulators' that co-ordinate the expression of a suite of genes. Our comparisons of patterns of ASE in floral bud and leaf tissues suggest that overall *cis*-regulatory differentiation is more likely to be a general consequence of mating system divergence. Nevertheless, the floral-specific ASE genes that we have identified comprise an important set of candidate genes for investigating the role of morphological evolution in mating system transitions.

Acknowledgements

We would like to thank Wei Wang for assistance with the bioinformatics analyses. This work was supported by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to S.C.H.B. and S.I.W. R.A. was supported by student fellowships from the Ministry of Training, Colleges and Universities Ontario Graduate Scholarship (OGS), University of Toronto and an NSERC graduate fellowship. T.I. was supported by an undergraduate student summer bursary from NSERC.

Author contributions

R.A., S.C.H.B. and S.I.W. planned and designed the research and wrote the manuscript. R.A. and T.I.M. generated the F_1 plants and analysed data.

References

Ågren JA, Wang W, Koenig D, Neuffer B, Weigel D, Wright SI. 2014. Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genomics* **15**: 602.

Altschul S, Gish W, Miller W, Myers E, Lipman D. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Anders S, Pyl PT, Huber W. 2015. HTSeq A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Andolfatto P. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Research* **17**: 1755–1762.

Arunkumar R, Ness RW, Wright SI, Barrett SCH. 2015. The evolution of selfing is accompanied by reduced efficacy of selection and purging of deleterious mutations. *Genetics* **199**: 817–829.

Barrett SCH. 1992. *Evolution and function of heterostyly*. Berlin: Springer.

Barrett SCH, Arunkumar R, Wright SI. 2014. The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philosophical Transactions of the Royal Society B* **369**: 20130344.

Barrett SCH, Husband BC. 1990. Variation in outcrossing rate in *Eichhornia paniculata*: the role of demographic and reproductive factors. *Plant Species Biology* **5**: 41–56.

Barrett SCH, Ness RW, Vallejo-Marín M. 2009. Evolutionary pathways to self-fertilization in a tristylous plant species. *New Phytologist* **183**: 546–556.

Bell GDM, Kane NC, Rieseberg LH, Adams KL. 2013. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biology and Evolution* **5**: 1309–1323.

Brandvain Y, Haig D. 2005. Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *American Naturalist* **166**: 330–338.

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biology* **3**: e245.

Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. 2015. Tools and best practices for allelic expression analysis. *Genome Biology* **16**: 195.

Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV. 2010. Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. *Genome Biology* **11**: R125.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics & Development* **11**: 685–690.

Cruden RW. 1977. Pollen–ovule ratios: a conservative indicator of breeding systems in flowering plants. *Evolution* **31**: 32–46.

Darwin C. 1876. *The effects of cross and self-fertilisation in the vegetable kingdom*. London, UK: John Murray.

Darwin C. 1877. *The different forms of flowers on plants of the same species*. London, UK: John Murray.

Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* **25**: 3207–3212.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**: 341–352.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research* **38**: W64–W70.

Emerson JJ, Li W-H. 2010. The genetic basis of evolutionary change in gene expression levels. *Philosophical Transactions of the Royal Society B* **365**: 2581–2590.

Fenster CB, Barrett SCH. 1994. Inheritance of mating-system modifier genes in *Eichhornia paniculata* (Pontederiaceae). *Heredity* **72**: 433–445.

Fishman L, Beardsley PM, Stathos A, Williams CF, Hill JP. 2015. The genetic architecture of traits associated with the evolution of self-pollination in Mimulus. *New Phytologist* **205**: 907–917.

Fishman L, Kelly AJ, Willis JH. 2002. Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution* **56**: 2138–2155.

Fyon F, Cailleau A, Lenormand T. 2015. Enhancer runaway and the evolution of diploid gene expression. *PLoS Genetics* **11**: e1005665.

Haddrill PR, Bachtrog D, Andolfatto P. 2008. Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Molecular Biology and Evolution* **25**: 1825–1834.

He F, Zhang X, Hu J, Turck F, Dong X, Goebel U, Borevitz J, de Meaux J. 2012. Genome-wide analysis of *cis*-regulatory divergence between species in the *Arabidopsis* genus. *Molecular Biology and Evolution* **29**: 3385–3395.

Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences, USA* **108**: 2322–2327.

Husband BC, Barrett SCH. 1993. Multiple origins of self-fertilization in tristylous *Eichhornia paniculata* (Pontederiaceae): inferences from style morph and isozyme variation. *Journal of Evolutionary Biology* **6**: 591–608.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences, USA* **12**: 15 390–15 395.

Köhler C, Weinhofer-Molisch I. 2010. Mechanisms and evolution of genomic imprinting in plants. *Heredity* **105**: 57–63.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M *et al.* 2012. The Arabidopsis information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**: D1202–D1210.

Lemmon ZH, Bukowski R, Sun Q, Doebley JF. 2014. The role of *cis* regulatory evolution in maize domestication. *PLoS Genetics* **10**: e1004745.

León-Novelo LG, McIntyre LM, Fear JM, Graze RM. 2014. A flexible Bayesian method for detecting allelic imbalance in RNA-seq data. *BMC Genomics* **15**: 920.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536.

Lloyd DG. 1965. Evolution of self-compatibility and racial differentiation in Leavenworthia (Cruciferae). *Contributions from the Gray Herbarium of Harvard University* **195**: 3–134.

Lloyd DG. 1980. Demographic factors and mating patterns in angiosperms. In: Solbrig OT, ed. *Demography and evolution in plant populations*. Oxford, UK: Blackwell, 67–88.

Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* **21**: 936–939.

Messing J, Grossniklaus U. 1999. Genomic imprinting in plants. In: Ohlsson R, ed. *Genomic imprinting*. Berlin: Springer, 23–40.

Morgan MT, Barrett SCH. 1989. Reproductive correlates of mating system evolution in *Eichhornia paniculata* (Spreng.) Solms (Pontederiaceae). *Journal of Evolutionary Biology* **2**: 183–203.

Ness RW, Siol M, Barrett SCH. 2012. Genomic consequences of transitions from cross-to self-fertilization on the efficacy of selection in three independently derived selfing plants. *BMC Genomics* **13**: 611.

Ness RW, Wright SI, Barrett SCH. 2010. Mating-system variation, demographic history and patterns of nucleotide diversity in the tristylous plant *Eichhornia paniculata*. *Genetics* **184**: 381–392.

Nodine MD, Bartel DP. 2012. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**: 94–97.

Ornduff R. 1969. Reproductive biology in relation to systematics. *Taxon* **18**: 121–133.

Park C, Chen X, Yang JR, Zhang J. 2013. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences, USA* **110**: E678–E686.

Pastinen T. 2010. Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews Genetics* **11**: 533–538.

R Development Core Team. 2011. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raissig MT, Bemer M, Baroux C, Grossniklaus U. 2013. Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2. *PLoS Genetics* **9**: 1003862.

Richards JH, Barrett SCH. 1984. The developmental basis of tristyly in *Eichhornia paniculata* (Pontederiaceae). *American Journal of Botany* **71**: 1347–1363.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**: 1086–1092.

Shore JS, Barrett SCH. 1985. Genetics of distyly and homostyly in the *Turnera ulmifolia* complex (Turneraceae). *Heredity* **55**: 167–174.

Shore JS, Barrett SCH. 1990. Quantitative genetics of floral characters in homostylous *Turnera ulmifolia* var. *angustifolia* (Turneraceae). *Heredity* **64**: 105–112.

Sicard A, Lenhard M. 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany* **107**: 1433–1443.

Sicard A, Stacey N, Hermann K, Dessoly J, Neuffer B, Bäurle I, Lenhard M. 2011. Genetics, evolution, and adaptive significance of the selfing syndrome in the genus *Capsella*. *Plant Cell* **23**: 3156–3171.

Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM. 2011. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Research* **21**: 1728–1737.

Slotte T, Hazzouri KM, Stern D, Andolfatto P, Wright SI. 2012. Genetic architecture and adaptive significance of the selfing syndrome in *Capsella*. *Evolution* **66**: 1360–1374.

Stebbins GL. 1957. Self-fertilization and population variability in the higher plants. *American Naturalist* **91**: 337–354.

Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T. 2015. *Cis*-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella*. *Molecular Biology and Evolution* **32**: 2501–2514.

Stevenson KR, Coolon JD, Wittkopp PJ. 2013. Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. *BMC Genomics* **14**: 536.

Videvall E, Sletvold N, Hagenblad J, Ågren J, Hansson B. 2015. Strong maternal effects on gene expression in *Arabidopsis lyrata* hybrids. *Molecular Biology and Evolution*, in press. doi: 10.1093/molbev/msv342.

Wittkopp PJ, Kalay G. 2012. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**: 59–69.

Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences* **169**: 105–118.

Yang JR, Liao BY, Zhuang SM, Zhang J. 2012. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proceedings of the National Academy of Sciences, USA* **109**: E831–E840.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de bruijn graphs. *Genome Research* **18**: 821–829.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Number of reads that mapped to reference assembly from floral bud and leaf parental transcriptomes, and floral bud and leaf F_1 offspring transcriptomes of *Eichhornia paniculata*.

Fig. S2 Distribution of posterior probabilities that genes show allele-specific expression (ASE) under various *a hat* and *d hat* parameters in *Eichhornia paniculata*.

Fig. S3 Distribution of $\log_e$-fold expression difference between outcrossing and selfing alleles for ASE and non-ASE genes calculated using only sites where the outcrossing allele matched the reference allele, only sites where the selfing allele matched the reference allele, or all sites.

Fig. S4 Distribution of $\log_e$-fold gene expression difference between outcrossing and selfing *Eichhornia paniculata* parents for the set of genes with posterior probability for allele-specific expression > 0.7 in F_1 offspring.

Table S1 Population codes and localities for parental genotypes of *Eichhornia paniculata* used in the study

Table S2 Number of sites homozygous for alternative SNPs in outcrossing and selfing parents and heterozygous in F_1 progeny

Table S3 Functional classification for genes with posterior probability for allele-specific expression > 0.7 sampled from floral buds and leaf tissue of F_1 plants from crosses between selfing (maternal) and outcrossing (paternal) parents of *Eichhornia paniculata*

Table S4 *P*-values from comparisons of expression level differences between outcrossing and selfing *Eichhornia paniculata* parents for the set of genes showing allele-specific expression in the F_1 offspring

Table S5 Diversity at nonsynonymous and synonymous sites for *Eichhornia paniculata* floral bud and leaf genes in outcrossing parents from Brazil and selfing parents from the Caribbean

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.