MOLECULAR ECOLOGY

Molecular Ecology (2017) **26**, 752–765 doi: 10.1111/mec.13946

The genetic architecture of tristyly and its breakdown to self-fertilization

RAMESH ARUNKUMAR, WEI WANG, STEPHEN I. WRIGHT and SPENCER C. H. BARRETT
Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, Ontario Canada, M5S 3B2

Abstract

The floral polymorphism tristyly involves three style morphs with a reciprocal arrangement of stigma and anther heights governed by two diallelic loci (*S* and *M*). Tristyly functions to promote cross-pollination, but modifications to stamen position commonly cause transitions to selfing. Here, we integrate whole-genome sequencing and genetic mapping to investigate the genetic architecture of the *M* locus and the genetic basis of independent transitions to selfing in tristylous *Eichhornia paniculata*. We crossed independently derived semi-homostylous selfing variants of the long- and mid-styled morph fixed for alternate alleles at the *M* locus (*ssmm* and *ssMM*, respectively), and backcrossed the F_1 to the parental *ssmm* genotype. We phenotyped and genotyped 462 backcross progeny using 1450 genotyping-by-sequencing (GBS) markers and performed composite interval mapping to identify quantitative trait loci (QTL) governing style-length and anther-height variation. A QTL associated with the primary stylemorph differences (style length and anther height) mapped to linkage group 5 and spanned ~13–27.5 Mbp of assembled sequence. Bulk segregant analysis identified 334 genes containing SNPs potentially linked to the *M* locus. The stamen modifications characterizing each selfing variant were governed by loci on different linkage groups. Our results provide an important step towards identifying the *M* locus and demonstrate that transitions to selfing have originated by independent sets of mating-system modifier genes unlinked to the *M* locus, a pattern inconsistent with a recombinational origin of selfing variants at a putative supergene.

Keywords: bulk segregant analysis, *M* locus, modifier genes, QTL mapping, self-fertilization, tristyly

Received 10 April 2016; revision received 2 November 2016; accepted 7 November 2016

Introduction

Heterostyly is a floral polymorphism characterized by two (distyly) or three (tristyly) style morphs with a reciprocal arrangement of stigmas and anthers (Darwin 1877; Ganders 1979; Barrett 1992). The polymorphism has evolved in at least 28 angiosperm families and promotes animal-mediated cross-pollination. Most mating in heterostylous populations occurs between plants with anthers and stigmas of equivalent height. Although there has been considerable work on the function and adaptive significance of heterostyly (reviewed in Lloyd & Webb 1992a, b; Barrett & Shore 2008), little is known about the underlying genetic architecture of the suite of polymorphisms comprising the heterostylous syndrome. Most heterostylous species are distylous with populations composed of long- and short-styled morphs (hereafter L- and S-morphs), and this polymorphism is governed by the *S* locus (Bateson & Gregory 1905; reviewed in Lewis & Jones 1992). Tristylous species are restricted to six flowering plant families, and populations are composed of long-, mid- and short-styled morphs (hereafter L-, M- and S-morphs), each with two stamen levels within a flower corresponding in height to stigmas in the remaining style morphs. Despite the polyphyletic origins of tristyly, a similar two-locus diallelic model (*S* and *M* loci) with dominance and epistasis

Correspondence: Ramesh Arunkumar, Fax: (416) 978-5878;
E-mail: ram.arunkumar@utoronto.ca

between the two loci (reviewed in Lewis & Jones 1992; Barrett 1993) governs the inheritance of the floral polymorphism in Lythraceae (Barlow 1923; Fisher & Mather 1943; Eckert & Barrett 1993), Oxalidaceae (Von Ubisch 1926; Fisher & Martin 1948; Fyfe 1950, 1956; Weller 1976; Bennett *et al.* 1986; Weller *et al.* 2015) and Pontederiaceae (Gettys & Wofford 2008; S.C.H. Barrett & R. Arunkumar, unpublished data).

Investigations of the molecular genetic architecture of heterostyly have focused solely on distyly. The *S* locus in distylous species is a candidate supergene region composed of a cluster of linked genes with two distinct haplotypes maintained by balancing selection (Ernst 1928; Charlesworth & Charlesworth 1979a, b; Lewis & Jones 1992; Yoshida *et al.* 2011; Charlesworth 2016). Nowak *et al.* (2015) assembled the draft genome of distylous *Primula veris* and performed bulk segregant analysis on L- and S-morph pools. They identified 13 variants in the candidate *S*-linked region, 113 genes with morph-biased expression and one that was completely silenced in flowers of the L-morph. More recently, Huu *et al.* (2016) identified a gene (*CYP734A50*) associated with the *S* locus governing style-length dimorphism in *P. veris* from bulk segregant analysis. The gene is only expressed in styles of the S-morph, and its loss or inactivation results in long styles. Following the identification of a paralog in a closely related species, the authors suggested that it arose via a gene duplication event. A major goal of this study was to apply molecular approaches to a tristylous species to investigate the genetic architecture of this more complex heterostylous polymorphism.

The long-term frequency-dependent selection that maintains tristyly is expected to result in a strong signal of balancing selection in regions of the genome in which the loci responsible for the polymorphism reside. Indeed, strong signatures of balancing selection have been detected for homomorphic self-incompatibility systems in the Brassicaceae (Takahata 1990; Vekemans & Slatkin 1994) and Solanaceae (Clark & Kao 1991; Richman & Kohn 1999). A key requirement to detect signals of balancing selection in outcrossing populations is suppressed recombination (Charlesworth 2006). It is not yet clear whether the evolution of the tristylous polymorphism involves recombination suppression. This would be expected if tristyly were governed by supergene(s) and several workers have visualized the *S* and *M* loci as supergenes, each containing several tightly linked loci responsible for the morphological and physiological components of the tristylous syndrome (Ornduff 1972; Lewis & Jones 1992). However, there is no empirical evidence for supergene control and an alternative model has been proposed involving a smaller number of genes with pleiotropy playing an important role in governing several tristylous traits (Charlesworth 1979). One motivation for the present study was to obtain evidence that may shed light on these alternative hypotheses for the genetic architecture of tristyly.

A common feature of heterostylous groups is the breakdown of the polymorphism and the evolution of derived mating systems. The most frequent transition involves the evolution of selfing variants with stigmas and anthers at a similar height within a flower (Charlesworth & Charlesworth 1979b; Barrett 1989; Weller 1992). In distylous and tristylous taxa, these are referred to as homostyles or semi-homostyles, respectively. In the latter case, only one of the two stamen levels within a flower is modified, hence the term semi-homostyly (Ornduff 1972; Barrett 1988). The breakdown of tristyly is especially well documented in *Eichhornia* (Pontederiaceae), where in each of the three tristylous species semi-homostyles have established in populations (Barrett 1988). Studies of *E. paniculata* indicate multiple independent transitions to selfing via semi-homostyle evolution (Husband & Barrett 1993; Barrett *et al.* 2009) with the loss of herkogamy (stigma–anther separation) governed by recessive modifiers apparently nonallelic to the *S* or *M* loci (Fenster & Barrett 1994; Vallejo-Marín & Barrett 2009). However, the location, genetic basis and effect size of the loci underlying the breakdown of tristyly in *E. paniculata* remains unknown.

Eichhornia paniculata (Pontederiaceae) is a diploid ($n = 8$) self-fertile annual with a cryptic trimorphic incompatibility system (Cruzan & Barrett 1993, 2016). Fertilizations from intermorph cross-pollination are favoured over intramorph cross-pollination and self-pollination, thus promoting phenotypic disassortative mating (Barrett *et al.* 1987). The flowers of each style morph have upper and lower level stamens each composed of three stamens (Richards & Barrett 1984, 1992). The inheritance of tristyly is controlled by two diallelic loci with the *S* locus epistatic to *M*. The S-morph is governed by a dominant *S* allele (genotypes: *SsMM*, *SsMm*, *Ssmm*), the M-morph by a dominant *M* allele (genotypes: *ssMM* or *ssMm*) and the L-morph is of genotype *ssmm*; controlled crosses indicate that the *S* and *M* loci are linked and separated by a map distance of 2.7 cM (S.C.H. Barrett & R. Arunkumar, unpublished data).

The largest concentration of *E. paniculata* populations occur in N.E. Brazil, with a smaller number of populations in Cuba and Jamaica, and a few isolated populations in Nicaragua and Mexico. Brazilian populations are largely tristylous and outcrossing whereas populations in the Caribbean and Central America are predominantly selfing (Barrett *et al.* 2009). It has been estimated that colonization of the Caribbean from Brazil occurred ~120 000 years ago (Ness *et al.* 2010). Semi-homostylous variants of the M-morph (hereafter M′)

occur sporadically in N.E. Brazil, but predominate in Jamaica and Cuba (Barrett *et al.* 1989, 2009). In the M′ semi-homostyle, one to three short-level stamens have elongated to the mid-level of the flower causing autonomous self-pollination (Richards & Barrett 1992; Barrett *et al.* 2009). In Nicaragua and Mexico, a different semi-homostylous phenotype occurs which is absent from the Caribbean and Brazil. This variant is a semi-homostylous L-morph (hereafter L′) resulting from a separate breakdown of tristyly (Barrett *et al.* 2009) and exhibits three mid-level stamens in close proximity to long-level stigmas, also due to stamen elongation.

A recombinational origin for homostyles appears to have occurred in several distylous species (Dowrick 1956; Charlesworth & Charlesworth 1979b; Shore & Barrett 1985). However, although semi-homostyles have often been reported in tristylous taxa (e.g. *Lythrum salicaria* – Stout 1925; *Oxalis* spp. – Ornduff 1972; *Eichhornia* spp. – Barrett 1988), there is no evidence that these have originated by recombination or through mutation at the tristyly loci (see Charlesworth 1979). Morphological, geographic and genetic evidence indicate that the two semi-homostyles in *E. paniculata* have independent origins. Molecular studies indicate that the semi-homostylous L′ variant from Nicaragua shares more SNPs with outcrossing populations from Brazil than with M′ semi-homostyles from Jamaica (Ness *et al.* 2011, 2012), a pattern consistent with their separate origins. Also, controlled crosses suggest that transitions to selfing in the M′ semi-homostyle involve one or two recessive modifier genes, whereas both major and minor mutations appear to control stamen modification in the L′ semi-homostyle (Fenster & Barrett 1994; Barrett *et al.* 2009). *Eichhornia paniculata* thus provides an excellent opportunity to study the genetic architecture of morphological adaptations associated with higher selfing rates, and also for determining whether the same or different genes are involved in independent transitions to predominant self-fertilization.

Here, we use the highly inbred semi-homostyles to investigate the genetic architecture of the *M* locus and the QTLs associated with stamen modifications causing self-pollination. To identify the genomic location of the *M* locus and the loci associated with the evolutionary breakdown of tristyly in *E. paniculata*, we assembled a draft genome and conducted quantitative trait loci (QTL) mapping of floral traits distinguishing the semi-homostyles. We also performed bulked segregant analysis on L′- and M′-morph pools with the goal of identifying genes with SNP differentiation and differential expression to complement our genetic mapping results and to validate candidate genes. To accomplish this, we performed a backcross between an F_1, obtained from crossing the semi-homostylous L′ (*ssmm*) and M′ (*ssMM*) genotypes, and the semi-homostylous L′ parent. This cross should result in segregation at the *M* locus for style length and unmodified anther height, as well as for any unlinked modifier loci controlling the anther heights of the modified stamens that distinguish the two semi-homostyles.

Our study addressed the following specific questions: (i) What is the location of the *M* locus controlling style length and stamen height, and does the region containing the *M* locus show signatures of long-term balancing selection? We used genetic mapping to identify the genetic region governing the traits distinguishing the L′- and M′-morphs, and using additional samples from outcrossing tristylous populations from Brazil, we scanned for evidence of balancing selection in the region containing the *M* locus. (ii) Are there genes that segregate for alternate SNPs, and/or that show differential gene expression between the L′ and M′ semi-homostyles in backcross progeny and, if so, what is their molecular function? Such genes represent candidates for being associated with the *M* locus. (iii) Are genetic regions separate from the *M* locus responsible for the stamen modifications causing independent transitions to selfing in the two semi-homostylous variants? Given their independent origins and contrasting patterns of stamen modification, we predicted that the regions of the genome governing selfing in the two semi-homostyles would likely differ.

Materials and methods

F_1 and backcross populations

We performed crosses to produce individuals segregating for the *M*/*m* allele to investigate the genetic architecture of the *M* locus. We crossed a semi-homostylous M′ (*ssMM*) plant from Slipe, St. Elizabeth, Jamaica, with a semi-homostylous L′ (*ssmm*) plant from Oaxaca, San Mateo del Mar, Nr. Tehuantepec, Mexico. The cross involved the M′ plant as the female parent and the L′ plant as the male parent. The female parent was emasculated prior to cross-pollination. Seeds from this cross were germinated to produce an F_1 of M′ plants (all *ssMm*), and a single plant was chosen to backcross to the L′ (*ssmm*) parent with the *ssmm* plant as the female parent. Further details of the glasshouse culture of the backcross population are available in the Supplementary methods. We phenotyped all flowering progeny, but were only able to obtain near complete genotype data for 462 progeny (see Methods, Supporting information, for genotype quality controls).

Floral measurements

We performed floral measurements to examine phenotypic variation in the backcross progeny. At the start of flowering for each plant, we recorded the date, and 2 weeks after flowering of each plant commenced, we measured flower (perianth) breadth, width and length, style length and the heights of the six stamens within a flower (Fig. S1, Supporting information). Flower breadth and width were the longest distances along the vertical and horizontal axes, respectively, of the top view of the flower using the nectar guide to orient the flower. Flower length, style length and anther height were measured as the distances from the base of the floral tube to the tip of the flower or the floral organ, as viewed from the side view of the flower. We sequentially labelled anthers from 1 to 6 in each plant, corresponding to the position relative to the base of the flower, with 6 being furthest away from the base. We measured two flowers per plant and used the average value in our analyses. We also measured the height of the plant from the surface of the soil to the tip of the tallest leaf. We estimated the mean and 95% confidence intervals for plant height and all floral traits. We compared measurements of plant height, flower width, breadth and length and style length between progeny of the L′ and M′ semi-homostyles using two-sample *t*-tests assuming unequal variances and analysed the correlations among all measured traits. All analyses were conducted using R (R Development Core Team 2011).

Genotyping by sequencing

We used genotyping by sequencing (GBS) to identify and generate genetic markers for F_1 and backcross progeny. For all backcross progeny, F_1 genotypes (*ssMm*), and the *ssMM* and *ssmm* parental genotypes, we collected ~1.5-cm floral buds for DNA extractions. We extracted DNA using the Qiagen DNeasy Plant Mini Kit. The DNA material was shipped to the Biotechnology Resource Center Genomic Diversity Facility at Cornell University to create genotyping-by-sequencing libraries made with the *PstI* enzyme (Elshire *et al.* 2011). Libraries were placed randomly across five 96-well plates and were sequenced using the 100-bp single-end approach on the individual lanes of the Illumina HiSeq2500. After sequencing, we used the denovo_map.pl pipeline version 1.24 part of the STACKS software to identify informative SNPs (Catchen *et al.* 2013). We set 20 as the minimum coverage to create a stack and removed highly repetitive tags. We used the default values for the remaining parameters for SNP calling using $m = 3$, $M = 3$ and $n = 2$. At the termination of SNP calling, we identified a total of ~20 000 markers. We removed markers that were not typed in >10% of genotypes, had unexpected segregation patterns or showed evidence for segregation distortion, resulting in 1450 markers being used for genetic map construction (see Methods, Supporting information).

Composite interval mapping

For each measured trait, we performed composite interval mapping (CIM) to identify the interval regions for QTLs using the R/QTL software (Broman *et al*. 2003; Broman & Sen 2009). First, we constructed a genetic map for *E. paniculata* (detailed in supplementary methods) and calculated conditional genotype probabilities for each cM in the genetic map assuming a genotyping error rate of 0.03. Further, we performed CIM using the imputation method. In some cases, we compared the interval scans using the expectation–maximization algorithm (EM) and extended Haley–Knott regression (EHK). For each LOD peak, we performed 1000 permutations to attain its genome-scan-adjusted *P*-value. We estimated the per cent of variation in phenotype explained under a single QTL model using the formula $1–10^{-2\mathrm{LOD}/n}$, where n was the sample size. For a multiple QTL model, we used the 'makeqtl' and 'fitqtl' functions, using the imputation method for the latter function, in the R/QTL software to estimate the proportion of phenotypic variances explained by each QTL. We used *P*-values from the 'addint' function to test whether there was any evidence for interactions among QTLs in a multiple QTL model. For the markers with the highest LOD scores underneath each LOD peak, we identified their corresponding effect sizes.

Identifying genes within QTLs

We sequenced the genomes of the parental L′ and M′ semi-homostyles to examine the genes found within QTLs identified for measured traits. We then extracted DNA from floral bud tissue using the QIAGEN DNEASY PLANT MINI KIT. Illumina TruSeq DNA libraries with 400-bp insert sizes were prepared from both samples, and an additional library with a 5-kbp insert size was prepared from DNA extracted from the L′ parent. Each library was sequenced on individual lanes using the 100-bp paired-end protocol on Illumina HiSeq 2000 at the McGill University and Génome Québec Innovation Centre. The genomic assembly approach is detailed in the supplementary methods. We performed a reciprocal nucleotide BLAST search (Altschul *et al.* 1990) for the assembled genome against the genetic map to identify the scaffolds containing markers with LOD scores above the significance threshold in each QTL. We used the markers from the genetic map as the query, the

genomic assembly as the database and extracted the top hit. Further, we conducted a BLAST search with the scaffold as the query and a 65.53 Mbp *E. paniculata* transcriptome assembly (16 416 contigs, N_{50} = 2.2 kbp) generated in Arunkumar *et al.* (2015) as the database. This procedure was used to identify genes occurring within scaffolds from the genomic assembly.

Bulk segregant analysis

To identify candidate transcripts putatively linked to the *M* locus, we performed bulk segregant analysis (BSA) on the L′ and M′ semi-homostylous backcross progeny to identify differentially expressed genes. We collected a single floral bud from each backcross progeny of sizes ranging from 0.5 to 1 cm. We pooled floral buds of 10 backcross progeny of the same morph and extracted RNA using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich). We then pooled equal amounts of RNA from four extractions to generate six L′ and six M′ pools each containing RNA from ~40 backcross progeny. The extracted RNA samples were used to make Illumina TruSeq RNA libraries that were sequenced using the 100-bp paired-end protocol on two lanes of the Illumina HiSeq 2000 at the Génome Québec Innovation Centre at McGill University. Three L′ and three M′ pooled libraries were sequenced on each lane. We mapped the RNA-Seq short reads from the backcross progeny, and the L′ and M′ grandparents and genomic reads from the F_1s to the previously generated transcriptome assembly (Arunkumar *et al.* 2015) and the genomic assembly generated in this study. The read mapping and variant calling approaches are detailed in the supplementary methods. We identified candidate genes associated with the *M* locus by assessing the SNP segregation in the bulked backcross progeny transcriptomes, and our approach is detailed in the supplementary methods. We also inferred the molecular function and expression intensity of candidate genes, and these approaches are also detailed in the supplementary methods.

Population genetic summary statistics

We calculated population genetic summary statistics for candidate genes underlying each QTL and compared them to genomewide patterns to investigate evidence for balancing selection in the outcrossing tristylous populations. For the candidate genes, we used a population genetic data set previously reported in Table S1 (Supporting information) in Arunkumar *et al.* (2015). We sequenced the transcriptomes, mapped the reads and called SNPs in each plant using the aforementioned transcriptome reference and using a similar approach to this study. Here, using the SNP data set, we generated estimates of pairwise nucleotide diversity at synonymous sites (Π_{syn}) and Tajima's *D* at synonymous sites using the program POLYMORPHORAMA (Andolfatto 2007; Haddrill *et al.* 2008). We estimated the population recombination rate (ρ) between adjacent sites in a scaffold using the program DNASP version 5 (Librado & Rozas 2009). To do this, we concatenated sequences of genes that occurred within a scaffold and specified their positions in the scaffold in DNASP version 5.

Results

Phenotypic variation of L′ and M′ backcross progeny

There were 227 L′ and 235 M′ progeny in the backcross generation, which is not significantly different from the expected 1:1 ratio for a cross between *ssmm* × *ssMm* genotypes (Deviation from 1:1; G = 0.201, P = 0.647). There was no significant difference in flowering time between the L′ and M′ progeny (mean 72.5 days, confidence interval [CI] ±2 days). L′ progeny were significantly taller than M′ progeny but had smaller flowers (Figs S1–S2, Supporting information). Two-sample *t*-tests using unequal variances indicated that plant height (P = 1.11E−08) and flower size (P = 4.46E−04) were significantly different between the L′ and M′ progeny at the 1% threshold.

As expected, there were significant differences in style length and anther position between the L′ and M′ phenotypes. Style lengths for the L′ and M′ backcross progeny were 15.66 ± 0.10 mm and 11.65 ± 0.12 mm (mean ± CI), respectively (Fig. 1), and this difference was significant (two-sample *t*-test: P = 5.93E−213). The upper three anthers of L′ progeny (anthers 4–6) were close to each other and were 0.9–3 mm below the stigma. This range in stigma–anther separation reflected the occurrence of two distinguishable phenotypes underlying segregation of the anther modifications associated with the transition to selfing; one in which anther 6 was very close or touching the stigma and the other in which the anther was separated from the stigma. Thus, the average value for anther 6 indicates a small separation. The two phenotypes could not be distinguished in the mapping study. The lower three anthers of L′ progeny (anthers 1–3) were 5–8 mm below the stigma. In M′ progeny, the upper three anthers (anthers 4–6) were close to each other and were 4–5.5 mm above the stigma whereas the lower three anthers were ~3 mm below and above the stigma, respectively, with another two very close to or touching the stigma. Style length and anther heights were weakly associated with plant height and with flower width, breadth and length within each morph (Table S1, Supporting information).

Even so, the comparisons of style-length and anther-height differences within and between L′ and M′ progeny were not affected when each measure was first divided by flower size, calculated as flower width × breadth × length or by plant height.

Genomic reference and genetic map

We generated a genomic assembly and genetic map for *E. paniculata* to assist with identifying the genetic regions associated with the *M* locus and to identify regions associated with the breakdown of tristyly. The genomic assembly of parental genotypes, assisted by genome size estimation, indicated a large section of the genome contained repetitive content. The haploid genome size of *E. paniculata* based on flow cytometry was 0.61 Gbp. Assemblies with K-MER values >40 had a large number of <100-bp contigs resulting in a fragmented assembly. When using a K-MER of 31, there were 40 320 scaffolds totalling 573 Mbp of assembled sequences with 4.6% of 'N' bases (Table S2, Supporting information); 40.3% and 0.9% of scaffolds were greater than 10 and >100 kbp, respectively, and this assembly had an N_{50} of 31.7 kbp. We used the assembly generated using a K-MER = 31 for further analyses.

We generated genetic markers in regions with low repeat content to construct the genetic map of *E. paniculata*. Using 1450 genetic markers, the genetic map had the expected eight linkage groups (hereafter LGs), consistent with the number of chromosomes in *E. paniculata*, with 90–250 markers in each LG (Fig. 2); 410 Mbp of the assembled genome mapped to these markers. The sizes of the LGs ranged from 10 to 50 cM with a total length of 255 cM. Note that the small genetic maps of some LGs and the relatively low overall rate of recombination using the estimated genome size (0.4 cM/Mbp on average) may reflect the presence of chromosomal rearrangements such as inversions given our use of an interpopulation cross. Similarly, larger

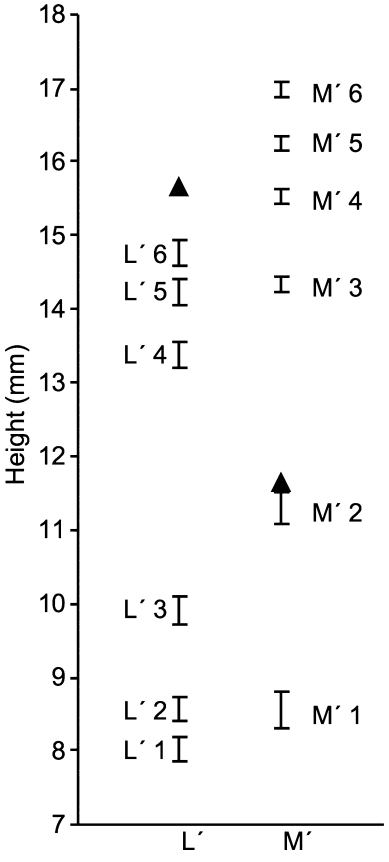

Fig. 1 Mean anther heights of 227 L′ and 235 M′ backcross *Eichhornia paniculata* progeny. We crossed a semi-homostylous M′ plant (*ssMM*) with a semi-homostylous L′ plant (*ssmm*) and backcrossed the resulting F_1 M′ plant (*ssMm*) to the L′ parent to generate the backcross population. We sequentially labelled anthers from 1 to 6 within each flower, corresponding to their positions relative to the base of the flower, with six being furthest away from the base. The bars represent 95% confidence intervals. The black triangles indicate the position of the stigma in each semi-homostyle. [Colour figure can be viewed at wileyonlinelibrary.com].

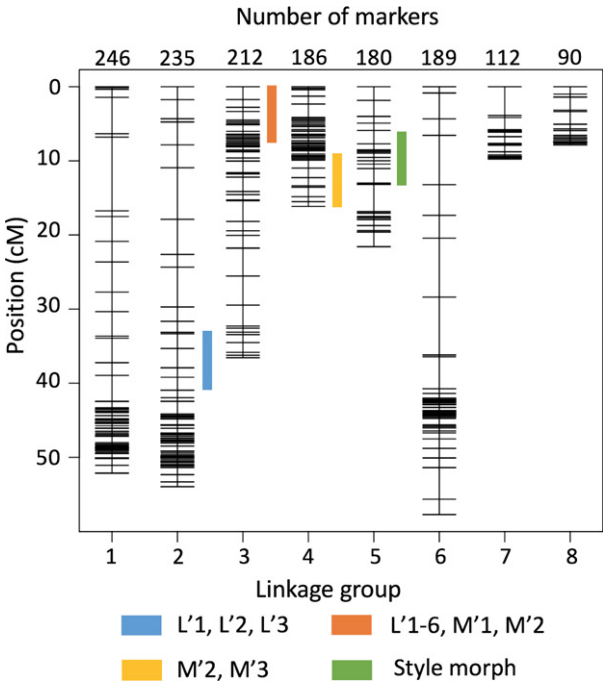

Fig. 2 Genetic map of *Eichhornia paniculata*; the number of markers in each linkage group is indicated on the top of the diagram. The coloured bars on the right of the linkage groups indicate the QTLs identified for style length and anther heights in the L′ and M′ backcross progeny. We sequentially labelled anthers from 1 to 6 within each flower corresponding to their positions relative to the base of the flower with 6 being furthest away from the base. [Colour figure can be viewed at wileyonlinelibrary.com].

LGs, particularly groups 1, 2, 3 and 6, had long stretches containing few markers, likely representing repetitive regions or regions with chromosomal rearrangements.

Composite interval mapping of the M locus

Composite interval mapping (hereafter CIM) using both L′ and M′ backcross progeny with the imputation method indicated that there was a single peak on LG5 with LOD scores of > 35 shared by style length (Fig. 3a) and the upper three anthers, which are the stamen levels that distinguish the L′ and M′ semi-homostyles (L′ and M′ anthers 4–6; Fig. 3b). There were no other genomic regions shared between style length and anther heights, and this shared peak on linkage group 5 was the only one evident for style length. Analysing the style-length peak indicated that there were 47 markers located between 5.5 and 11.5 cM (Fig. 3a, Table S3, Supporting information). Using estimates of the total map size and genome size (and assuming the region experiences an average rate of recombination), the interval could be 13.1–27.5 Mbp. BLAST searches using the markers as the query and the genomic scaffolds as the database indicated that there was a total of 35 nonoverlapping scaffolds of various sizes totalling 2.39 Mbp of genomic sequence in the region. The scaffolds found within this interval contained 204 genes, and the position near 8.9 cM contained the greatest density of genes (Table S3, Supporting information). Among the 35 scaffolds, the longest one was 200 kbp, and it mapped to 10 markers from the linkage map containing 41 genes. In contrast, other scaffolds of sizes 180–220 kbp in the genomic assembly contained fewer than 24 genes (Table S4, Supporting information). Finally, CIM using EM and EHK also yielded similar results (Fig. S3, Supporting information). Thus, this genetic region likely contains the *M* locus.

Population genomic signatures associated with the M locus

We investigated whether there were population genomic footprints suggesting that tristyly is subject to

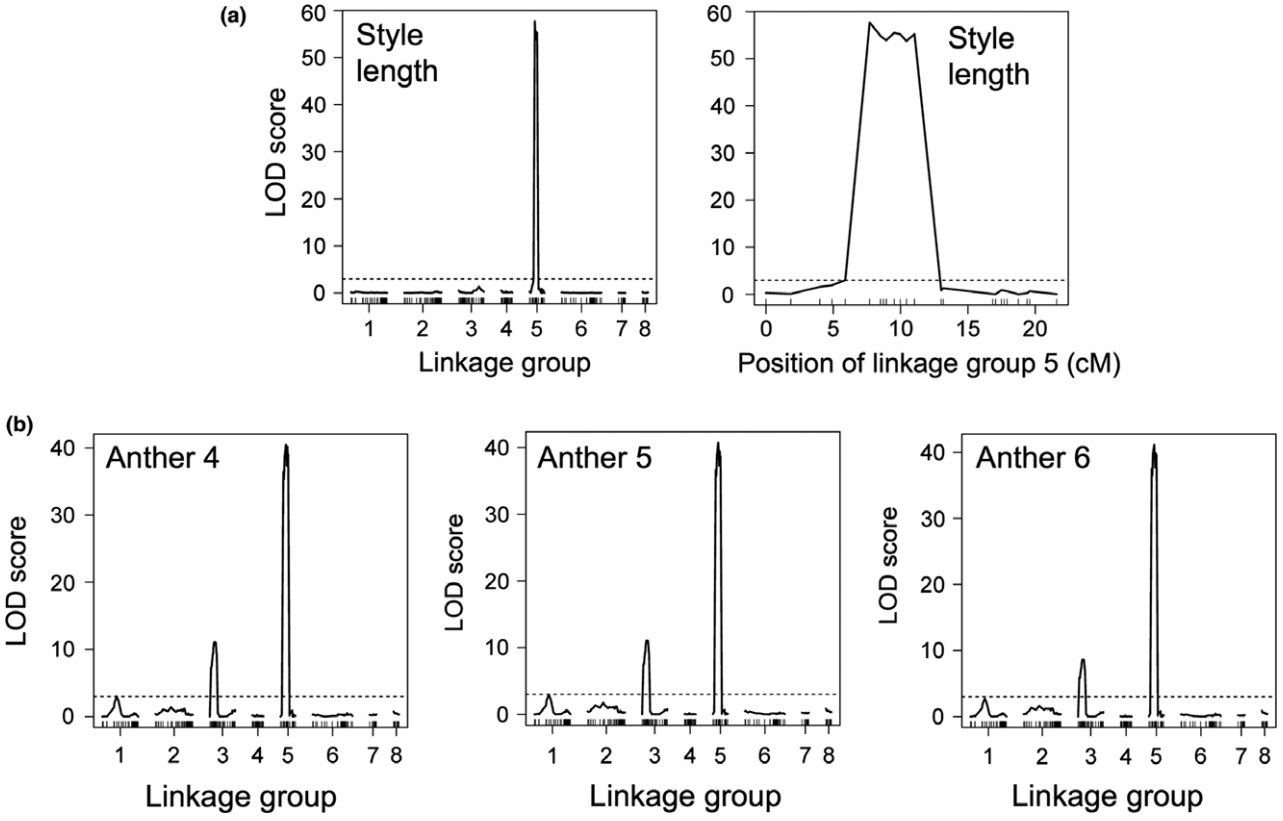

Fig. 3 Composite interval mapping of the backcross progeny of *Eichhornia paniculata*. Shown are LOD scores for: (a) style-length variation and (b) variation in the anthers of the upper level stamens in 227 L′ and 235 M′ backcross progeny. Shown are the map distances and LOD scores across all linkage groups. The upper level anthers in the L′ and M′semi-homostyles correspond to the mid- and long-level stamens, respectively. The horizontal dashed lines indicate the threshold for significance (LOD > 3) at the 0.1% level based on 1000 permutations. [Colour figure can be viewed at wileyonlinelibrary.com].

long-term balancing selection. However, for the 204 genes in the interval on LG5, neither the average Π_{syn} nor Tajima's D was significantly different from the genome-wide averages, when assessed using a two-sample t-test at the 5% significance threshold. Three genes showed Tajima's D values greater than the 2.5% genomewide distribution. However, no genes showed both elevated Π_{syn} and Tajima's D, and the proportion of genes showing elevated diversity was not above the number observed genomewide (Fig. S4, Supporting information). Further, a chi-squared test indicated that the ratio of shared to unique polymorphisms at the interval on LG5 (15:178) was not significantly different compared to the remainder of the genome (1672:16 857) at the 1% significance threshold.

To investigate whether the QTL region on LG5 showed evidence of recombination suppression, we estimated the population recombination rate $\rho = 4N_e r$, where r is the recombination rate and N_e is the effective population size. The ρ between adjacent sites in the 200-kbp scaffold mapping to 8.9 cM was 0.0004. This value was lower than the range of 0.0010–0.0018 observed in 20 scaffolds of comparable sizes 180–220 kbp. SNPs were separated by a distance of 2.5–5 kbp along the 180–220-kbp scaffolds, and there is no clear evidence to suggest that average SNP distances differed between the scaffolds. Finally, while there was no evidence that any of the *Arabidopsis* orthologs for the genes identified within the interval on LG 5 were enriched for a particular molecular function, six genes were characterized as being involved in reproductive development (Table S5, Supporting information).

Transcriptomes of L′ and M′ pools

Although the QTL mapping revealed the broad interval containing the *M* locus, the incomplete genome assembly could mean that the genome assembly may not contain the causal gene(s). To find additional genic regions within the *M* locus region, we performed bulk segregant analysis (BSA) using RNA from pooled backcross progeny to identify candidate genes associated with differentiation between the L′ and M′ semi-homostyles. A total of 334 of 16 147 contigs in the transcriptome assembly had one or more sites that were homozygous across all L′ pools and heterozygous across all M′ pools, as expected if such SNPs were linked to the stylemorph locus. There were no contigs that contained sites homozygous in all M′ pools and heterozygous in L′ pools, suggesting a low rate of false positives. To compare estimates of Π_{syn} and Tajima's D for BSA genes against the genomewide background, we identified 3984 genes with one or more heterozygous sites in the genome, as all BSA genes would have had at least one such site. Two and six of the BSA genes were in the 2.5% of genes with the largest Π_{syn} or most positive Tajima's D in the genome, respectively (Fig. S5, Supporting information). Although this was lower than expected by chance, we still identified nine BSA genes showing $\Pi_{syn} > 0.03$, including a locus with both high polymorphism (0.049) and high Tajima's D (1.14) (Table S5, Supporting information). As expected with a loss of balancing selection in the selfing semi-homostylous populations, Π_{syn} was 0 for all nine of these genes. However, neither average Π_{syn} nor Tajima's D was significantly different from the genomewide averages, when assessed using a two-sample t-test at the 5% significance threshold. There was evidence to indicate significant functional enrichment in postembryonic development among the *Arabidopsis* orthologs in the 334 genes identified from BSA, and twenty-one genes were involved in this function (Fig. S6, Supporting information). We were unable to identify any genes showing significant differential expression. Thus, while we did not observe a general signature of balancing selection on genes in the *M* locus region, we have identified a number of candidate genes with high diversity.

Our results from BSA were generally congruent with those from linkage mapping. Twenty-four of the 334 BSA genes mapped to markers in the 8.7–11.3 cM interval on LG5 and another eight mapped to other parts of LG5 (Table S3, Supporting information). Custom BLAST searches indicated that the other LGs contained far fewer candidate genes ($n = 6$, Table S6, Supporting information). Whereas all genes identified by BSA mapped to one or more scaffolds in the genomic assembly, the majority of these candidate genes ($n = 296$) were not localized using linkage mapping, indicating that our BSA approach has allowed for additional gene localization within the *M* locus region.

Regions governing anther-height variation in L′ and M′ semi-homostyles

Our analysis of backcross progeny revealed that anther-height variation in L′ and M′ semi-homostyles was governed by QTLs separated from the region governing style length, reflecting anther-height modifier loci associated with the transition to selfing. When we performed QTL mapping of L′ and M′ progeny separately for each semi-homostyle, we found QTLs that were unique to anthers with the smallest degree of physical separation from the stigma, corresponding to the modified anthers causing autonomous self-pollination. One QTL in the L′ semi-homostyle was specific to the upper three anther heights (anthers 4–6) on LG2 and spanned ~8 cM containing 54 markers (Fig. 4a, b). The markers with the highest LOD score underneath this peak were

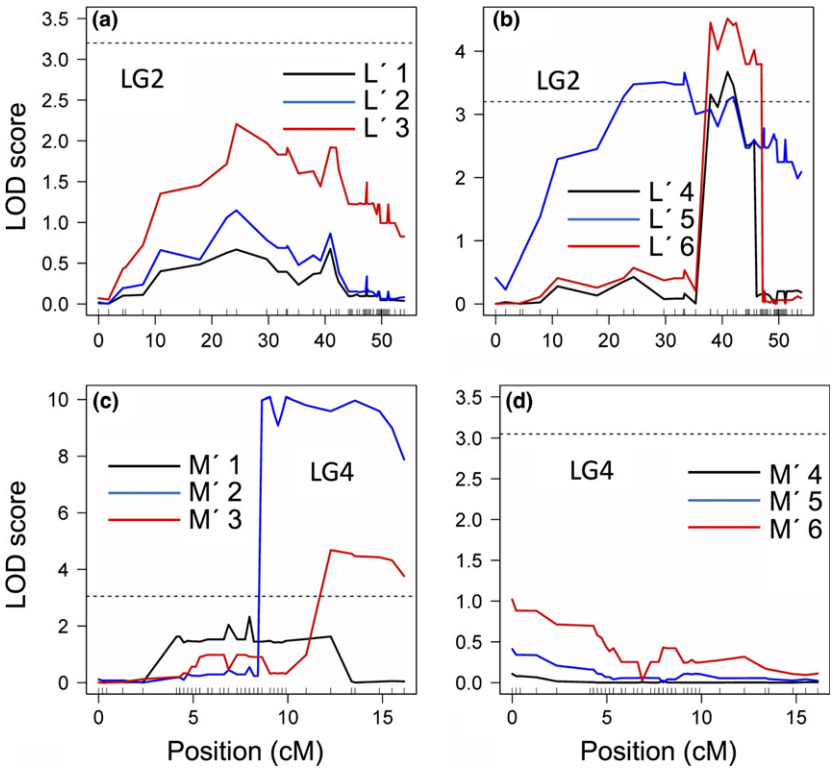

Fig. 4 LOD scores on linkage groups 2 and 4 following composite interval mapping of anther heights in *Eichhornia paniculata*. Shown are LOD scores for L′ and M′ semi-homostyles in backcross progeny: (a) the lowest three L′ anthers, (b) the upper three L′ anthers in linkage group 2, (c) the lowest three M′ anthers and (d) the upper three M′ anthers in linkage group 4. We sequentially labelled anthers from 1 to 6 within each flower, corresponding to their positions relative to the base of the flower, with 6 being furthest away from the base. The horizontal dashed lines indicate the threshold for significance at the 0.1% level based on 1000 permutations. [Colour figure can be viewed at wileyonlinelibrary.com].

of genotypes AA, BB and AB in the L′ parent, M′ parent and F_1 M′, respectively. An AA genotype at this marker was associated with ~0.6 mm increase in anther height when compared with AB genotypes (Fig. S7, Supporting information). This QTL explained 6–8.5% of the variation in the height of upper level anthers (Table S7, Supporting information) and likely arose as a selfing modifier during the evolution of the L′ semi-homostyle. Very few genes were identified in the scaffolds containing these markers. This interval contained 39 nonoverlapping scaffolds totalling a size of 1.68 Mbp of genomic sequence (Table S8, Supporting information). The genomic scaffolds contained a total of 195 genes. No corresponding QTL on LG2 was found in the M′ semi-homostylous progeny (Fig. S8, Supporting information).

Similarly, the M′ semi-homostyle had a QTL on LG4 shared by the second and third lowest anthers (M′ anthers 2–3). The interval for this QTL was at least 8 cM, and the 21 markers underneath this peak were widely distributed (Fig. 4c, d). The marker with the highest LOD score underneath this peak was of genotypes BB, AA and AB in the L′ parent, M′ parent and F_1 M′, respectively. An AA genotype at this marker was associated with ~1.5 mm (Fig. S9A, Supporting information) and ~0.5 mm (Fig. S9B, Supporting information) increase in the height of M′ anthers 2 and 3, respectively, when compared with AB genotypes. These results are consistent with this QTL being an anther-height selfing modifier that originated during the evolution of the M′ semi-homostyle. Although AA and AB genotypes associated with this marker also segregated in L′ semi-homostylous progeny, genotypic differentiation was not accompanied by differences in the heights of L′ anthers 2 and 3 (Fig. S9C, D, Supporting information); similarly, we did not detect a significant QTL for these anther heights in this region following CIM of the L′ progeny. The identified QTL explained 18.1% and 8.6% of the variation in the heights of M′ anthers 2 and 3, respectively. This interval matched 31 nonoverlapping scaffolds totalling a size of 1.05 Mbp of genomic sequence (Table S9, Supporting information). Different markers in 7–11 cM of the interval matched the same genomic scaffolds and matched a total of 90 genes. CIM using EHK and EM approaches also indicated a QTL on LG4 for the uppermost anther height in L′ progeny (anther 6, Fig. S10, Supporting information) and a QTL on LG2 for the second lowest anther height in M′ progeny (anther 2, Fig. S11, Supporting information). Therefore, different genetic regions are associated with stamen elongation causing self-pollination in the L′ and M′ semi-homostyles.

We also identified a QTL affecting anther height of both semi-homostyles on LG 3 (Fig. 5). The QTL was not solely associated with modified anther levels in L′ and M′ semi-homostyles. The interval shared between all anther heights in L′ progeny and the lowest two anthers in M′ progeny (M′ anthers 1–2) spanned 8.3 cM,

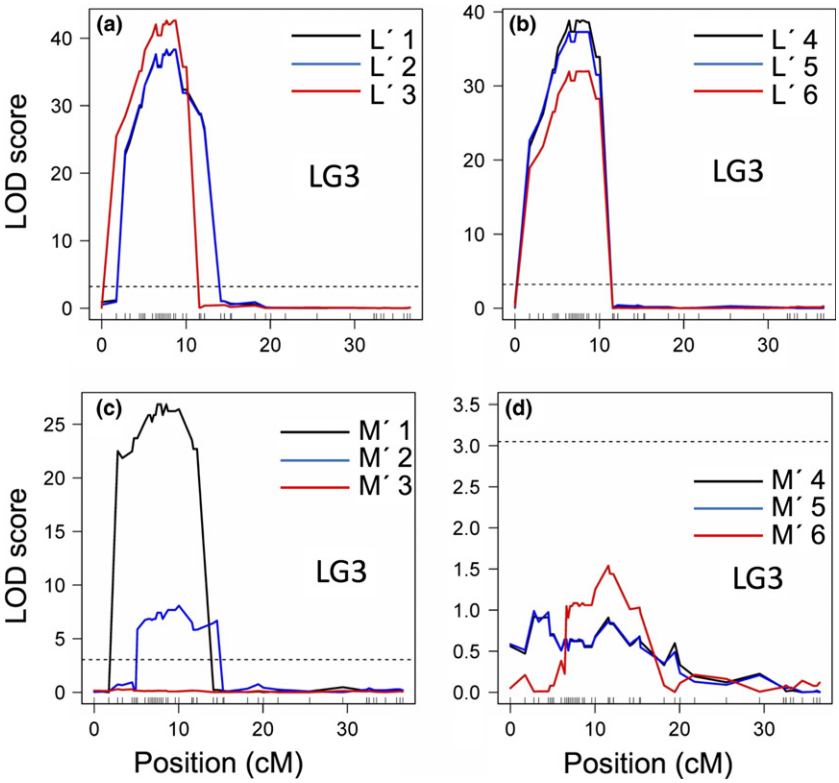

Fig. 5 LOD scores on linkage group 3 following composite interval mapping of anther heights in *Eichhornia paniculata*. Shown are LOD scores for of L′ and M′ semi-homostyles in backcross progeny: (a) the lowest three L′ anthers, (b) the upper three L′ anthers, (c) the lowest three M′ anthers and (d) the upper three M′ anthers. We sequentially labelled anthers from 1 to 6 within each flower, corresponding to their positions relative to the base of the flower, with 6 being furthest away from the base. The horizontal dashed lines indicate the threshold for significance at the 0.1% level based on 1000 permutations. [Colour figure can be viewed at wileyonlinelibrary.com].

and there were 178 markers underneath this peak (Fig. 5, Table S10, Supporting information). A completely additive multiple QTL model was assumed as *P*-values for testing interactions among the identified QTLs in each morph were >0.05. The QTL on LG3 explained 42–55% of the variation in height of L′ anthers (Table S7, Supporting information) and the lowest M′ anther (anther 1). The QTL explained 14.5% of variation in the second lowest M′ anther (anther 2). This interval matched 685 kbp of nonoverlapping genomic sequence containing 256 genes.

All anther heights in L′ progeny also had peaks in LGs 1 and 6, but the LOD scores for L′ anthers 1 and 4 were below the significance threshold for the QTL on LG6 (Fig. S12, Supporting information). Each QTL explained 6–8% of the variation in the height of anthers (Table S7, Supporting information). The interval on LG1 spanned ~7 cM with two markers underneath the peak, and the interval on LG6 spanned ~4 cM with a single marker underneath the peak. The upper three anthers in the M′ progeny (M′ anthers 4–6) did not show any significant peaks across the LGs.

Discussion

Our study investigated the genetic architecture of variation in style length and anther height characterizing the tristylous syndrome of *E. paniculata* and its evolutionary breakdown to selfing. We found that a single large genetic interval near the centre of linkage group 5 governed style length and the stamen levels that distinguish the L- and M-morphs. We also found 334 genes with contrasting patterns of SNPs in the L′ and M′ semi-homostylous backcross progeny, although we found no evidence for morph-specific gene expression. In both the L′ and M′ semi-homostyles, modified stamens contained unique QTLs that were not associated with the *M* locus and these mapped elsewhere in the genome. Significantly, different QTLs were involved with stamen modifications in the two semi-homostyles consistent with their independent origins. Below we discuss insights gained from mapping style length and anther height on the genetic architecture of tristyly and its breakdown to selfing.

Broad mapping of the M locus

Our mapping study confirmed that the *M* locus responsible for style-length dimorphism localized to a single linkage group. One of the goals of this investigation was to use population genomic signatures to distinguish whether supergenes, or a polymorphism involving pleiotropy, control tristyly. If the *M* locus represents a multigene region exhibiting suppressed recombination, we would expect to be able to identify a large block of genes with elevated polymorphism and high

linkage disequilibrium. In contrast, if the polymorphism was governed by a small number of genes with pleiotropy in a region of normal recombination, the signal of balancing selection may be highly localized, with little evidence of elevated linkage disequilibrium (Charlesworth 2016). Although we did not detect evidence that genes in the *M* locus interval generally show elevated polymorphism, our mapped interval was large and many scaffolds remain unmapped to the region. As we obtained no clear evidence for recombination suppression associated with the *M* locus, it remains unclear which of the two models for the genetic control of tristyly is correct and in reality, some combination of both may be involved.

Our genetic mapping approach identified 204 candidate markers within LG5. Although we did not identify a region of elevated diversity, a small number of genes from the BSA showed signs of both elevated polymorphism and Tajima's *D* in the region containing the *M* locus. Some of these genes were enriched for embryonic development and represent candidates for further study. Some BSA genes mapped to linkage groups other than LG5 and this could be associated with their occurrence in repetitive regions. Given the stringent selection criteria for candidate SNPs and the lack of false positives, the list of candidate genes generated from BSA should be helpful for future fine-mapping studies. An improved genome assembly and more comprehensive polymorphism analyses combined with fine mapping of the *M* locus should enable a more direct test of whether the loci controlling tristyly involve supergenes and/or pleiotropic polymorphisms. In addition, our study did not involve the S-morph of *E. paniculata* and future work should obviously also include this morph in efforts to determine the location of the tristyly loci and refine our understanding of the genetic architecture of the polymorphism.

Genetic architecture of stamen modifications promoting self-pollination

The stamens promoting self-pollination in each semi-homostyle had QTLs unlinked to the *M* locus and some were not shared between them. This finding supports other genetic evidence indicating that the evolution of the semi-homostyles occurred independently in *E. paniculata* (Fenster & Barrett 1994; Barrett *et al.* 2009). Our results confirm that at least some genes involved in anther-height modification are unlinked to the tristyly loci. The modified upper three anthers in the L′ semi-homostyle and the second and third lowest anthers in the M′ semi-homostyle had QTLs not found in the remaining stamen levels of each variant. These QTLs had the largest influence on anther position, and the anthers exhibited the lowest degree of physical separation from the stigma in each semi-homostyle. The genotype associated with stamen elongation of modified anthers in L′ backcross progeny was the same as that of the L′ parent; similarly, the allele found in the homozygous genotype associated with stamen elongation of modified anthers in M′ progeny was fixed in the M′ parent. Both patterns therefore confirm that the modifier alleles associated with stamen elongation in each morph also occurred in the parents. Interestingly, the allele in the QTL occurring within LG4 associated with modification of the short-level stamens in M′ progeny (M′ anthers 2 and 3, see Fig. 4) did not have a similar effect in modifying the short-level stamens of L′ progeny. This indicates that the effect of this allele is morph-limited in expression, as proposed by Lloyd & Webb (1992b) for other genes associated with heterostyly. To our knowledge, this is the first mapping study that has demonstrated this phenomenon. Similar effects of morph-limited expression of modifiers could not be established for L′ plants, as all M′ progeny were heterozygous for the QTL on LG2 causing stamen modification given the backcross design used for genetic mapping.

In addition to the one QTL shared with style length, anther heights in the L′ and M′ semi-homostyles were governed by different genetic regions, some of which were shared between the two selfing variants and others that were unique to each semi-homostyle. These QTLs are significant because they provide insight into the genetic architecture of evolutionary transitions from outcrossing to selfing in *E. paniculata*. Anthers that were positioned below the stigma in each selfing variant (L′ anthers 1–6 and M′ anthers 1 and 2) shared the same QTL on LG3. This QTL exhibited the most influence on anther-height variation compared to the other identified QTLs. It is probable that this region contains the modifier genes that are nonallelic to the *S* and *M* loci governing stamen elongation and causing the loss of herkogamy, as predicted from crossing studies (Fenster & Barrett 1994; Vallejo-Marín & Barrett 2009). The lack of evidence that the QTL on LG3 also governs the upper three anthers (long-level stamens) found only in M′ plants may be because further elongation of this stamen level has no functional significance as any modification would be unlikely to promote self-pollination. All stamen levels in the L′ semi-homostyle had additional QTLs modifying anther position on LGs 1 and 6. In contrast, there were only one or two significant QTLs governing most of the anther-height variation in the M′ semi-homostyle. Our results therefore conform to expectations from crossing studies indicating that multiple genes are associated with the evolution of the L′ semi-homostyle in Central America whereas only a few

genes govern the transition to selfing in the M′ semi-homostyle from the Caribbean (Fenster & Barrett 1994; Barrett *et al.* 2009; Vallejo-Marín & Barrett 2009). Our mapping results therefore indicate that the independent evolution of selfing in the L- and M-morphs of *E. paniculata* through semi-homostyle formation is governed by modifiers located elsewhere in the genome to the tristyly loci and are associated with gradual changes to stamen position.

Acknowledgements

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. and student fellowships from Ministry of Training, Colleges and Universities Ontario Graduate Scholarship (OGS) and University of Toronto and an NSERC graduate fellowship to R.A. We thank Minkyu Kim for assistance with generating and phenotyping the backcross population and for help with DNA and RNA extraction, and Adrian Platts for assistance and advice on the genomic assembly.

References

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

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.

Barlow N (1923) Inheritance of the three forms in trimorphic species. *Journal of Genetics*, **13**, 133–146.

Barrett SCH (1988) Evolution of breeding systems in *Eichhornia*, a review. *Annals of the Missouri Botanical Garden*, **75**, 741–760.

Barrett SCH (1989) The evolutionary breakdown of heterostyly. In: *The Evolutionary Ecology of Plants* (eds Linhart Y, Bock J), pp. 151–169. Westview Press, Boulder.

Barrett SCH (ed.) (1992) *Evolution and Function of Heterostyly*. Springer, New York.

Barrett SCH (1993) The evolutionary biology of tristyly. In: *Oxford Surveys in Evolutionary Biology*, vol. **9** (eds Futuyma D, Antonovics J), pp. 283–326. Oxford University Press, Oxford.

Barrett SCH, Shore JS (2008) New insights on heterostyly: comparative biology, ecology and genetics. In: *Self-Incompatibility in Flowering Plants: Evolution, Diversity and Mechanisms* (ed. Franklin-Tong V), pp. 3–32. Springer-Verlag, Berlin.

Barrett SCH, Brown AHD, Shore JS (1987) Disassortative mating in tristylous *Eichhornia paniculata* (Pontederiaceae). *Heredity*, **58**, 49–55.

Barrett SCH, Morgan MT, Husband BC (1989) The dissolution of a complex genetic polymorphism: the evolution of self-fertilization in tristylous *Eichhornia paniculata* (Pontederiaceae). *Evolution*, **43**, 1398–1416.

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

Bateson W, Gregory RP (1905) On the inheritance of heterostylism in *Primula*. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **76**, 581–586.

Bennett JH, Leach CR, Goodwins IR (1986) The inheritance of style length in *Oxalis rosea*. *Heredity*, **56**, 393–396.

Broman KW, Sen Ś (2009) *A Guide to QTL Mapping with R/qtl*, **46**. Springer, New York.

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**, 889–890.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Charlesworth D (1979) The evolution and breakdown of tristyly. *Evolution*, **33**, 489–498.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2**, e64.

Charlesworth D (2016) The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications*, **9**, 74–90.

Charlesworth D, Charlesworth B (1979a) A model for the evolution of distyly. *The American Naturalist*, **114**, 467–498.

Charlesworth B, Charlesworth D (1979b) The maintenance and breakdown of distyly. *The American Naturalist*, **114**, 499–513.

Clark AG, Kao TH (1991) Excess nonsynonymous substitution of shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proceedings of the National Academy of Sciences United States of America*, **88**, 9823–9827.

Cruzan MB, Barrett SCH (1993) Contribution of cryptic incompatibility to the mating system of *Eichhornia paniculata* (Pontederiaceae). *Evolution*, **47**, 925–934.

Cruzan MB, Barrett SCH (2016) Post-pollination discrimination between self- and outcross-pollen covaries with the mating system of a self-compatible flowering plant. *American Journal of Botany*, **103**, 568–576.

Darwin C (1877) *The Different Forms of Flowers on Plants of the Same Species*. John Murray, London.

Dowrick VPJ (1956) Heterostyly and homostyly in *Primula obconica*. *Heredity*, **10**, 219–236.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Ernst A (1928) Zur verebung der morphologischen heterostylie-merkmale. *Berichte der Deutschen Botanischen Gesellschaft*, **46**, 573–588.

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

Fisher RA, Martin VC (1948) Genetics of style length in *Oxalis*. *Nature*, **162**, 533.

Fisher RA, Mather K (1943) The inheritance of style length in *Lythrum salicaria*. *Ann Eugen*, **12**, 1–23.

Fyfe VC (1950) The genetics of tristyly in *Oxalis valdiviensis*. *Heredity*, **4**, 365–371.

Fyfe VC (1956) Two modes of inheritance of the short-styled form in the genus *Oxalis*. *Nature*, **177**, 942–943.

Ganders FR (1979) The biology of heterostyly. *New Zealand Journal of Botony*, **17**, 607–635.
Gettys LA, Wofford DS (2008) Genetic control of floral morph in tristylous pickerelweed (*Pontederia cordata* L.). *Journal of Heredity*, **99**, 558–563.
Haddrill PR, Bachtrog D, Andolfatto P (2008) Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Molecular Biology and Evolution*, **25**, 1825–1834.
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.
Huu CN, Kappel C, Keller B *et al.* (2016) Presence versus absence of *CYP734A50* underlies the style-length dimorphism in primroses. *eLife*, **5**, e17956.
Lewis D, Jones DA (1992) The genetics of heterostyly. In: *Evolution and Function of Heterostyly* (ed. Barrett SCH), pp. 129–150. Springer, New York.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
Lloyd DG, Webb CJ (1992a) The evolution of heterostyly. In: *Evolution and Function of Heterostyly* (ed. Barrett SCH), pp. 151–178. Springer, New York.
Lloyd DG, Webb CJ (1992b) The selection of heterostyly. In: *Evolution and Function of Heterostyly* (ed. Barrett SCH), pp. 179–208. Springer, New York.
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.
Ness RW, Siol M, Barrett SCH (2011) De novo sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants. *BMC Genomics*, **12**, 298.
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.
Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E (2015) The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biology*, **16**, 12.
Ornduff R (1972) The breakdown of trimorphic incompatibility in *Oxalis* section Corniculatae. *Evolution*, **26**, 52–65.
R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
Richards JH, Barrett SCH (1984) The developmental basis of tristyly in *Eichhornia paniculata* (Pontederiaceae). *American Journal of Botany*, **71**, 1347–1363.
Richards JH, Barrett SCH (1992) The development of heterostyly. In: *Evolution and Function of Heterostyly* (ed. Barrett SCH), pp. 85–127. Springer, New York.
Richman AD, Kohn JR (1999) Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. *Proceedings of the National Academy of Sciences United States of America*, **96**, 168–172.
Shore JS, Barrett SCH (1985) Genetics of distyly and homostyly in the *Turnera ulmifolia* complex (Turneraceae). *Heredity*, **55**, 167–174.
Stout AB (1925) Studies in *Lythrum salicaria*. 2. A new form of flowers in this species. *Bulletin of the Torrey Botanical Club*, **52**, 81–85.
Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species polymorphism. *Proceedings of the National Academy of Sciences United States of America*, **87**, 2419–2423.
Vallejo-Marín M, Barrett SCH (2009) Modification of flower architecture during early stages in the evolution of self-fertilization. *Annals of Botany*, **103**, 951–962.
Vekemans X, Slatkin M (1994) Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics*, **137**, 1157–1165.
Von Ubisch G (1926) Koppelung von farbe und heterostylie bei *Oxalis rosea*. *Biologisches Zentralblatt*, **46**, 633–645.
Weller SG (1976) The genetic control of tristyly in *Oxalis* section Ionoxalis. *Heredity*, **37**, 387–393.
Weller SG (1992) Evolutionary modifications of tristylous breeding systems. In: *Evolution and Function of Heterostyly* (ed. Barrett SCH), pp. 247–272. Springer, New York.
Weller SG, Sakai AK, Lucas CA, Weber JJ, Domínguez CA, Molina-Freaner FE (2015) Genetic basis of tristyly in tetraploid *Oxalis alpina* (Oxalidaceae). *Botanical Journal of the Linnean Society*, **179**, 308–318.
Yoshida Y, Ueno S, Honjo M *et al.* (2011) QTL analysis of heterostyly in *Primula sieboldii* and its application for morph identification in wild populations. *Annals of Botany*, **108**, 133–142.

R.A., S.I.W. and S.C.H.B. conceived and designed the study and wrote the manuscript. R.A. produced the F_1 and backcross population and collected DNA and RNA from samples. R.A. and W.W. generated the genome assembly, and R.A. performed the genetic mapping, population genomic comparisons, bulk segregant analysis and all remaining analyses. All authors read and approved the final manuscript.

Data accessibility

GBS sequence data for the backcross plants are available under Accession no. SRP069136 at the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra, accessed 13 March 2016), and the associated BioProject alias is PRJNA310331.

Parental genomic sequence data are available under Accession no. SRP069167 at the SRA (http://www.ncbi.nlm.nih.gov/sra, accessed 13 March 2016) and the associated BioProject alias is PRJNA310303.

The *Eichhornia paniculata* genomic assembly has been deposited at DDBJ/ENA/GenBank under the accession LTAE00000000. The version described in this study is

version LTAE01000000 and the associated BioProject alias is PRJNA310303.

Transcriptomes from the bulk segregant analyses are available under Accession no. SRP069122 at the SRA (http://www.ncbi.nlm.nih.gov/sra, accessed 13 March 2016) and the associated BioProject alias is PRJNA310302.

Supporting information

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

Fig. S1 Phenotypic measurements of the flowers of the L′ and M′ semi-homostyles of *Eichhornia paniculata*.

Fig. S2 Plant height and flower size of semi-homostyles in the backcross F_2 of *Eichhornia paniculata* progeny.

Fig. S3 Comparison of composite interval mapping of style length of L′ and M′ semi-homostyles in F_2 progeny of *Eichhornia paniculata* using the imputation method (IMP), expectation-maximization algorithm (EM) and extended Haley-Knott regression (EHK).

Fig. S4 Comparison of π_{syn} and Tajima's *D* for all genes compared to those found within the interval identified for style length on linkage group LG 5.

Fig. S5 Comparison of π_{syn} and Tajima's *D* for all genes with one or more heterozygous sites in the bulk segregant population (BSA) and for genes with candidate SNPs suggesting linkage to the style-morph locus.

Fig. S6 Functional enrichment of *Eichhornia paniculata* genes with candidate SNPs suggesting linkage to the style-morph locus.

Fig. S7 Effect plots for the marker with the highest LOD score on linkage group 2 identified during interval mapping of anther height for the upper three anther levels in the L′ semi-homostyle of *Eichhornia paniculata* in backcross progeny: A) anther 4, B) anther 5, and C) anther 6.

Fig. S8 LOD scores on linkage group 2 following composite interval mapping of upper level anthers in M′semi-homostyle of *Eichhornia paniculata*.

Fig. S9 Effect plots for the marker with the highest LOD score on linkage group 4 identified during interval mapping of anther height for modified anther levels in the L′ semi-homostyle of *Eichhornia paniculata* in backcross progeny: A) anther 2, B) anther 3 in the M′morph, and the anther height: C) anther 2, D) anther 3 in the L′ semi-homostyle.

Fig. S10 Comparison of composite interval mapping of anther 6 of the L′semi-homostyle of *Eichhornia paniculata* in backcross progeny using the imputation method (IMP), expectation-maximization algorithm (EM) and extended Haley-Knott regression (EHK).

Fig. S11 Comparison of composite interval mapping of the second lowest anther of the M′ semi-homostyle of *Eichhornia paniculata* in backcross progeny using the imputation method (IMP), expectation-maximization algorithm (EM) and extended Haley-Knott regression (EHK).

Fig. S12 Composite interval mapping of anther heights of the L′ semi-homostyle of *Eichhornia paniculata* in backcross progeny.

Table S1 Pearson's correlation between the measurements of traits in *Eichhornia paniculata*.

Table S2 Summary metrics of genomic assembly for *Eichhornia paniculata*.

Table S3 Number of markers and genes within the interval on linkage group 5 that had high LOD scores following composite interval mapping of style length, and number of genes with candidate SNPs suggesting linkage to the style-morph locus in that interval in backcross *Eichhornia paniculata* progeny.

Table S4 Number of genes in genomic scaffolds of sizes 180-220 kbp in *Eichhornia paniculata* backcross progeny.

Table S6 Number of genes with candidate SNPs suggesting linkage to the style morph locus in each linkage group in *Eichhornia paniculata* backcross progeny.

Table S7 Percent of variation in height of anthers in the L′ semi-homostyle explained by each identified QTL in *Eichhornia paniculata* backcross progeny.

Table S8 Number of markers and genes within the interval on linkage group 2 that had high LOD scores following composite interval mapping of anthers of the L′ semi-homostyle in the backcross progeny of *Eichhornia paniculata*.

Table S9 Number of markers and genes found within the interval on linkage group 4 that had high LOD scores following composite interval mapping of anther height of the M′ semi-homostyle in the backcross progeny of *Eichhornia paniculata*.

Table S10 Number of markers within the interval on linkage group 3 that had high LOD scores following composite interval mapping of anther height in L′ and M′ semi-homostyles in the backcross progeny of *Eichhornia paniculata*.

Table S5 List of candidate genes identified from bulk segregant analysis of *Eichhornia paniculata* F_2 progeny.