MOLECULAR ECOLOGY

Molecular Ecology (2017) **26**, 163–177 doi: 10.1111/mec.13881

SPECIAL ISSUE: THE MOLECULAR MECHANISMS OF ADAPTATION AND SPECIATION: INTEGRATING GENOMIC AND MOLECULAR APPROACHES

Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation

BILLIE A. GOULD,* YANI CHEN* and DAVID B. LOWRY*†

*Department of Plant Biology, Michigan State University, Plant Biology Laboratories, 612 Wilson Road, Room 166, East Lansing, MI 48824, USA, †Program in Ecology, Evolutionary Biology and Behavior, Michigan State University, Giltner Hall, 293 Farm Ln Rm 103, East Lansing, MI 48824, USA

Abstract

The early stages of speciation are often characterized by the formation of partially reproductively isolated ecotypes, which evolve as a by-product of divergent selective forces that are endemic to different habitats. Identifying the genomic regions, genes and ultimately functional polymorphisms that are involved in the processes of ecotype formation is inherently challenging, as there are likely to be many different loci involved in the process. To localize candidate regions of the genome contributing to ecotype formation, we conducted whole-genome pooled sequencing (pool-seq) with 47 coastal perennial and 50 inland annual populations of the yellow monkeyflower, *Mimulus guttatus*. Coastal perennial and inland annual ecotypes of *M. guttatus* have previously been shown to be ecologically reproductively isolated and highly locally adapted to their respective habitats. Our pool-seq results found allelic differentiation between the ecotypes for two chromosomal inversions, suggesting that frequencies of inversion heterokaryotypes are strongly differentiated between the ecotypes. Further, there were elevated levels of nonsynonymous change across chromosomal inversions. Across the genome, we identified multiple strong candidate genes potentially driving the morphological, life history and salt tolerance differences between the ecotypes. Several candidate genes coincide with previously identified quantitative trait locus regions and also show a signature of recent natural selection. Overall, the results of our study add to growing support for a major role of chromosomal inversions in adaptation and speciation and provide new insights into the genetic mechanisms underlying classic plant ecotype adaptations to wet and dry habitats.

Keywords: adaptation, ecotype, genomics, inversion, *Mimulus*, pool-seq, salt tolerance, speciation

Received 28 April 2016; revision received 22 August 2016; accepted 29 August 2016

Introduction

The formation of new species occurs as a continuum over time as reproductive isolating barriers accumulate due to fundamental population genetic processes: mutation, selection and genetic drift (Coyne & Orr 2004; Lowry *et al*. 2008a; Sobel & Chen 2014). Both theoretical and empirical studies have established that ecological natural selection plays a key role in the origin of new species, especially for the early stages of the process (Rundle & Nosil 2005; Schluter 2009; Sobel *et al*. 2010). These early stages are often characterized by the formation of partially reproductively isolated ecotypes, which form as a by-product of divergent selective forces that are endemic to different habitats (Clausen 1951; Hendry *et al*. 2002; Lowry 2012; Nosil 2012; Lamichhaney *et al*. 2015). While ecotype formation and ecological

Correspondence: Billie A. Gould, Fax: (517) 353 1926; E-mail: bgould@msu.edu

reproductive isolation are very important for speciation, we have a very limited understanding of the genetic mechanisms by which ecotypes evolve. The process of ecotype formation is by its nature complex because there are many different environmental factors that impose natural selection on organisms among habitats (Nosil *et al.* 2009; Schluter 2009). That set of selective pressures in turn leads to a complex array of phenotypic changes, which can involve many different loci across the genome (Coop *et al.* 2009; Pritchard *et al.* 2010; Rockman 2012; Gompert *et al.* 2013). Thus, the formation of ecotypes is expected to involve changes in allele frequencies at multiple loci that underlie multiple different adaptations. Those adaptations in turn can lead directly or indirectly to reproductive isolating barriers (Coyne & Orr 2004; Schluter 2009; Nosil 2012).

Identifying the genomic regions, genes and ultimately SNPs that are involved in the processes of ecotype formation is inherently challenging, as there are likely to be many different loci involved in the process. Quantitative trait locus (QTL) mapping is effective for identifying regions of the genome that are involved in adaptive trait divergence between ecotypes (Colosimo *et al.* 2005; McKay *et al.* 2008; Lowry & Willis 2010; Savolainen *et al.* 2013). However, going from QTL to underlying gene for traits with complex genetic architectures is an arduous process often involving years or even a decade of fine mapping (Frary *et al.* 2000; Des Marais *et al.* 2014; Sweigart & Flagel 2015). Further, QTL identified in a single cross are often not representative of the causes of divergence between ecotypes. Population genomic analyses can help to speed up the process of understanding the underlying genetic mechanisms by which ecotypes evolve because they take advantage of numerous genetic recombination events that have occurred in natural populations (Korte & Farlow 2013; Lu *et al.* 2013; Rellstab *et al.* 2015). To be effective, these studies require whole-genome sequencing of many individual lines (Tiffin & Ross-Ibarra 2014), which can be cost-prohibitive for modern laboratories running on a tight budget.

Whole-genome pooled sequencing (pool-seq) of populations is an economical population genomic alternative for identifying polymorphisms associated with adaptive divergence (Schlötterer *et al.* 2014). While not ideal for estimating population genetic parameters of a single population (Lynch *et al.* 2014), pool-seq has proven to be a very effective way to identify loci with high allelic divergence between populations (Turner *et al.* 2010; Magwene *et al.* 2011; Cheng *et al.* 2012; Fabian *et al.* 2012; Kofler *et al.* 2012; Raineri *et al.* 2012; Flagel *et al.* 2014a,b; Kapun *et al.* 2016). Pool-seq, like any allele frequency outlier analysis, may have low power to identify outliers if population structure is high between the pools being evaluated (Lotterhos & Whitlock 2014; Tiffin & Ross-Ibarra 2014; Rellstab *et al.* 2015). Further, it is often hard to tell whether pool-seq analyses have succeeded, as there are often no a priori expectations of which genomic regions should be differentiated in allele frequencies between pools. Our study addresses these two major challenges for pool-seq and identifies candidate genes for differentiation between a pair of divergent plant ecotypes.

Soil water availability is one of the most important factors that drive physiological and developmental differences between plant ecotypes (Clausen 1951; Porter 1966; Lowry 2012; Andrew *et al.* 2013). Coastal and inland ecotypes of the yellow monkeyflower *Mimulus guttatus* represent an excellent model system for studying the genetic basis of divergence between plant ecotypes driven by soil water availability (Hall & Willis 2006; Hall *et al.* 2006, 2010; Lowry *et al.* 2008b; Lowry & Willis 2010; Baker & Diggle 2011; Baker *et al.* 2012; Oneal *et al.* 2014; Twyford & Friedman 2015). Populations of *M. guttatus* found in dunes, sea cliffs and coastal prairies of Oregon and California collectively comprise a distinct coastal ecotype (Lowry *et al.* 2008b; Twyford & Friedman 2015). Previous studies suggest that populations have spread from southern glacial refuges towards the north and that the coastal ecotype is derived from inland populations (Lowry *et al.* 2008b; Twyford & Friedman 2015). Coastal ecotype plants are always perennial and have large leaves, wide stems and large flowers, produce many prostrate vegetative branches and are late-flowering (Hall *et al.* 2006, 2010; Lowry *et al.* 2008b). In contrast, inland populations are often annual, produce small leaves, thin stems and small flowers, produce mostly upright flowering branches and flower much earlier than coastal perennials (Hall & Willis 2006; Lowry *et al.* 2008a,b; Hall *et al.* 2010; Baker & Diggle 2011; Baker *et al.* 2012). Using field experiments, we have shown that inland populations of *M. guttatus* have evolved a genetically based, locally adaptive, early-flowering annual life history as a mechanism of escape from hot summer drought conditions (Lowry *et al.* 2008b; Lowry & Willis 2010). Individuals of coastal ecotype maintain a perennial life history, adapted to year-round soil water availability maintained by coastal summer fog and cool temperatures (Lowry *et al.* 2008b). In addition, coastal populations of *M. guttatus* are more tolerant to salt spray and soil salinity than inland populations (Lowry *et al.* 2008b; Lowry *et al.* 2009; Selby *et al.* 2014). Both ecotypes accumulate Na^+ in their leaves, but coastal plants have far less leaf necrosis in soil salinity and salt spray treatments (Selby *et al.* 2014). Coastal perennial and inland annual ecotypes are also highly reproductively isolated as a result of local adaptation to their respective

environments (Lowry *et al.* 2008b). Much of that reproductive isolation is due to ecological barriers including immigrant inviability and asynchrony in flowering time of the two ecotypes (Lowry *et al.* 2008b; Lowry & Willis 2010).

The genetic basis of ecotypic differences between inland annual and coastal perennial populations of *M. guttatus* has previously been explored using both QTL mapping and statistical genetic approaches (Hall *et al.* 2006, 2010; Lowry *et al.* 2009; Holeski *et al.* 2014; Twyford & Friedman 2015). Vegetative and reproductive differences between the ecotypes can be traced to a small number of QTL regions, many of which have pleiotropic effects. One of the key adaptive QTL maps to a chromosomal inversion on chromosome 8. One orientation of the inversion is found in coastal perennials, while the other orientation is found in inland annuals (Lowry & Willis 2010). Field reciprocal transplant experiments demonstrated that the inversion contributes to reproductive isolation between the ecotypes through immigrant inviability, flowering asynchrony and extrinsic postzygotic isolation (Lowry & Willis 2010).

Chromosomal inversions have long been thought to be involved in evolutionary adaptations (Dobzhansky 1970; Kirkpatrick & Barton 2006; Ayala *et al.* 2014; Adrion *et al.* 2015). Inversions strongly suppress genetic recombination in heterokaryotic individuals (inversion heterozygotes) either because they directly prevent crossing-over or because recombinant gametes are unbalanced (Dobzhansky 1970; Coyne *et al.* 1993; Rieseberg 2001). Researchers have thus hypothesized that inversions could act as adaptation supergenes by holding together long haplotypes containing multiple adaptive polymorphisms through suppressed recombination (Dobzhansky 1970; Joron *et al.* 2011; Schwander *et al.* 2014; Thompson & Jiggins 2014; Küpper *et al.* 2016; Lamichhaney *et al.* 2016). Recently, Twyford & Friedman (2015) found that SNPs within the chromosome 8 inversion region in *M. guttatus* were more divergent between annual and perennial plants on average than the rest of the genome. However, that study only surveyed 276 SNPs across the 6.7-Mb inverted region and so had little resolution to detect patterns of evolution within the inversion. Further, Holeski *et al.* (2014) identified two more inversions, through linkage mapping in a single cross, on chromosomes 5 and 10, but the extent to which these two regions contribute to adaptation is unknown. If chromosomal inversions are involved in adaptive divergence between the ecotypes, then we would predict that SNPs unique to a specific orientation of the inversion are unevenly distributed between the ecotypes and thus show a pattern of significantly higher differentiation than the rest of the genome. It should also be noted that, to some extent, reduced recombination in rare heterokaryotypic individuals may reduce diversity in inversions, slightly driving up relative measures of divergence (Nachman & Payseur 2012).

Here, we use whole-genome pool-seq to identify candidate genomic regions, including chromosomal inversions, in adaptation and reproductive isolation. Pool-seq in plants such as *M. guttatus* presents major challenges because of high levels of structure between populations (mean pairwise F_{ST} = 0.46–0.48; Lowry *et al.* 2008a; Twyford & Friedman 2015). However, there is much lower population structure between regional groups, including comparisons between the coastal perennials and inland annuals at the ecotype level (F_{CT} = 0.085; Lowry *et al.* 2008b). Therefore, we sequenced pools of individuals at the ecotype level instead of at the population level. Overall, we evaluated the following four questions: (i) What is the genomewide pattern of allelic differentiation between two geographically widespread plant ecotypes? (ii) Do chromosomal inversions have higher levels of differentiation in allele frequencies than colinear regions of the genome? (iii) What are the patterns of differentiation within those inversions and can we identify candidate genes for the phenotypic effects of inversions? (iv) Which genes are most differentiated in allele frequencies between coastal and inland ecotypes and, thus, candidates for adaptation? Our study demonstrates the benefits of a new approach for pool-seq, focusing on the ecotype level, and in the process provides valuable new insights into the evolution of adaptation and speciation.

Methods

Plant material and pooled ecotype sequencing

To understand patterns of genomewide divergence between coastal perennial and inland annual ecotypes of *Mimulus guttatus*, we created two ecotype pools for whole-genome sequencing. From the seed stocks available to us at the time of the study, we were able to successfully germinate seeds from 47 coastal perennial and 50 inland annual populations (Fig. 1; Table S1, Supporting information). All plants were grown in a Biochambers (Winnipeg, MB, Canada) FXC-19 flex growth chamber at Michigan State University. To minimize the influence of any particular population on final ecotype pools, we allowed at most three individuals per population to be included in each ecotype pool. Our final pools consisted of 101 coastal perennial accessions combined into the *coastal pool* and 92 inland annual accessions combined into the *inland pool* (Table S1, Supporting information). While obligate perennial populations of *M. guttatus* do occur in inland habitat

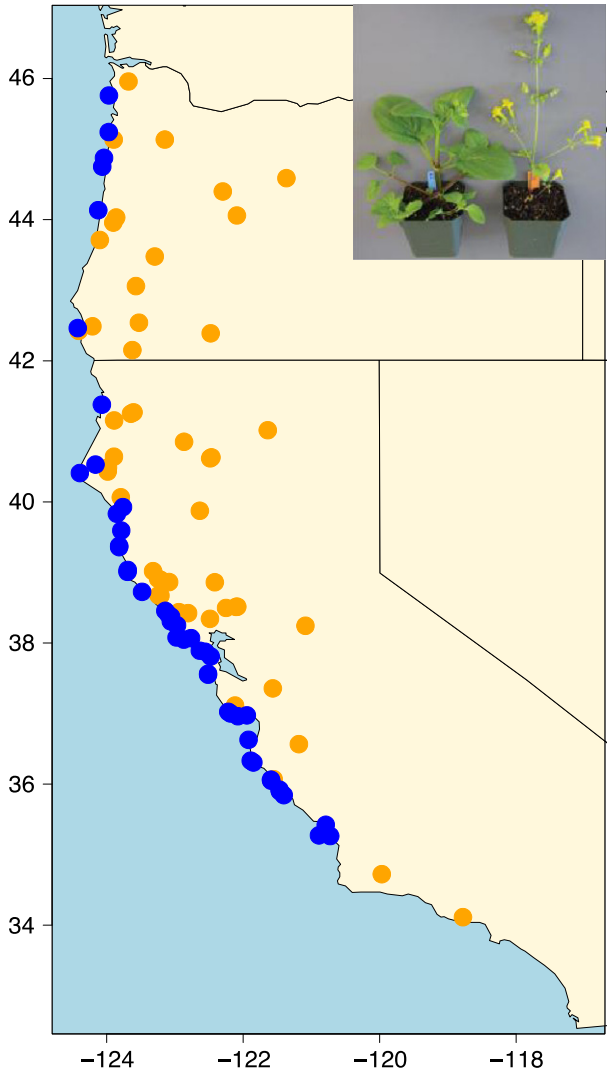

Fig. 1 The locations of coastal (blue) and inland (orange) populations. Inset: a coastal (left) and an inland (right) plant of similar age. [Colour figure can be viewed at wileyonlinelibrary.com]

(Twyford & Friedman 2015), we only selected plants with morphologies consistent with an annual life history for the inland pool. Floral bud tissue was collected from all individuals and frozen in a −80 °C freezer. Tissue samples for each pool were equalized across all individuals by only combining buds of similar size and developmental stage. The combined tissue for each pool was ground to a fine powder with liquid nitrogen using a mortar and pestle. Genomic DNA was extracted from homogenized powdered tissue using a Qiagen DNeasy Plant Mini Kit (Hilden, Germany). The Michigan State University Research Technology Support Facility prepared TruSeq Illumina (San Diego, CA, USA) libraries from extracted genomic DNA. The coastal and inland pools were then barcoded and subsequently combined together into one final pool. We conducted 2 × 250 bp paired-end sequencing on the final combined pool split across two v2 Rapid lanes on the Illumina High-Seq 2500 platform.

Read processing, alignment and variant calling

Sequencing produced approximately 125.9 and 109.6 million raw read pairs for the coastal and inland pools, respectively. We summarized raw read attributes using FastQC (Andrews *et al.* 2014) and then used Cutadapt (Martin 2011) to trim and filter reads for quality. Forward and reverse reads were quality-trimmed at a score threshold of 20, maximum proportion of Ns 50% and minimum length of 50 bp. We aligned filtered reads with BWA v0.7.12 (Li & Durbin 2009) to the *Mimulus guttatus* version 2.0 genome (PHYTOZOME v10.3; http://phytozome.jgi.doe.gov), retaining only uniquely mapped reads. We removed optical and PCR duplicates using PICARDTOOLS v1.113 (http://broadinstitute.github.io/picard/). The final alignments retained 57.8% and 56.7% of raw reads from the coastal and inland sequencing pools, respectively. For subsequent analyses and SNP calling, all unpaired reads were removed with SAMTOOLS v1.2 (Li *et al.* 2009).

A pileup file was created using SAMTOOLS, and SNPs were called using the program SNAPE v201303 (Raineri *et al.* 2012). SNAPE uses a Bayesian approach to calculate the posterior probability that a SNP is segregating in the sequence pool and calculates an expected alternate allele frequency for each base in the genome. We used an informative prior on the allele frequency spectrum. Prior values for theta and divergence were both set to 0.05 according to estimates generated by Brandvain *et al.* (2014).

Population genetic analyses

We used allele frequencies generated by SNAPE to calculate windowed population genetic summary statistics across the genome. We defined 'called bases' as having a probability of the presence of a SNP ≥ 95% (SNP sites) or ≤5% (invariant reference sites). Sites with probabilities between these values were treated as missing data. We also excluded bases with less than 50× coverage in either sequencing pool. We split the genome into nonoverlapping windows of 1000 called bases each. Window size was chosen based on estimates of linkage disequilibrium (LD) from previous studies which show that LD falls to background levels at about 1 kb (Brandvain *et al.* 2014). To avoid using SNPs introduced due to inaccurate mapping in repetitive parts of the genome, we excluded windows with average depth of coverage greater than two standard deviations from

the mean of all windows (mean depth = 262.2, SD = 147.0).

For each genomic window, we calculated three statistics: F_{ST}, G and the ratio of inland annual (IA) to coastal perennial (CP) nucleotide diversity (π_{IA}/π_{CP}). Each statistic was calculated on a per-site basis and averaged across each window. Outlier windows with high allelic differentiation (G, F_{ST}) between coastal and inland population pools can indicate genomic regions that are involved in local adaptation. We calculated F_{ST} (Wright 1951) based on expected heterozygosity within and across pools with a lower bound of zero. The results of F_{ST} analyses are shown in Supplemental Fig. 1 but not discussed at length in the text, as F_{ST} does not control for differences in sequencing coverage across the genome and other biases with pooled data. G is a measure of differentiation traditionally used in bulk segregant analysis that takes into account read count variation caused by population sampling and variability inherent in short-read sequencing of pooled samples (Magwene *et al.* 2011; Wright *et al.* 2015). We also calculated genomewide values of absolute divergence, D_{xy} (Nei & Li 1979), between the ecotypes.

Lastly, nucleotide diversity (π) was calculated for each sequencing pool (Tajima 1989), and the π ratio of the inland pool to the coastal pool was calculated for each window. Population genetic analyses were conducted using R (v3.0.2) and custom PYTHON (v2.7.2) scripts. Scripts used for population genomic analyses can be found at https://bitbucket.org/billiegould/bioinfo-tools.

The v2.0 *Mimulus* genome is a pseudomolecule assembly constructed from scaffolds from the v1.1 assembly using linkage from a recombinant inbred line mapping population (Hellsten *et al.* 2013; phytozome.org). Many of the scaffolds that make up chromosome pseudomolecules are likely to be incorrectly ordered, especially in the vicinity of chromosomal rearrangements (Hellsten *et al.* 2013; Holeski *et al.* 2014). When examining windows in putatively inverted regions of the genome, we included v1.1 scaffolds that are inferred to be in inverted regions from a previous linkage map study (Holeski *et al.* 2014), even though some of them may not be continuously ordered in the v2.0 genome assembly. For the inversion on chromosome 5, we included v1.1 scaffolds 36, 281, 149, 158, 170, 197, 266, 288, 327 and 368. For the inversion on chromosome 8, we included v1.1 scaffolds 11, 59, 76, 155 233, 604 and 1093. For the inversion on chromosome 10, we included v1.1 scaffolds 4, 48, 90, 206, 210, 223 and 445.

For the G-statistic, we isolated the top 1% of genomic windows and looked to see which genes and promoter regions overlapped those windows. For the π ratio statistic (π_{IA}/π_{CP}), we examined both the top and bottom 1% of windows which are potentially indicative of selective sweeps in the coastal and inland populations, respectively. Putative promoters were defined as the region 1000 bp upstream of each gene's transcriptional start site. Transcribed regions and promoters overlapping the G outlier windows were considered candidate genes for adaptive ecotypic differentiation. *Mimulus* candidate gene annotations were downloaded from Phytozome. We also found the best *Arabidopsis thaliana* protein match for each candidate gene (E-value cut-off of 10^{-3}) using stand-alone BLAST (blastp). Functional term enrichments for *Arabidopsis* matches to the candidate gene lists were conducted against the set of top matches to all genes in the *Mimulus* genome using the R BIOCONDUCTOR package topGO. We used Bioconductor database 'org.At.tair.db' and the algorithm 'classic' (statistic = 'fisher') (Alexa & Rahnenfuhrer 2016). GO term significance values were corrected for false discovery rate using the Benjamini and Hochberg method. We annotated SNPs genomewide using SnpEff (Cingolani *et al.* 2012) and the current genome annotation available at Phytozome. Finally, we calculated Tajima's D statistic (Tajima 1989) for each gene in the genome. Genes with greater than 50% missing data were excluded.

Results

Overall patterns of ecotype divergence

Between two sequencing pools of 47 coastal perennial (CP) and 50 inland annual (IA) populations, we identified 29 693 578 SNPs, which is consistent with the results of previous studies that show *Mimulus guttatus* has extremely high, genomewide nucleotide diversity (Hellsten *et al.* 2013; Brandvain *et al.* 2014). Interestingly, only four SNPs were completely fixed between coastal and inland sequencing pools (two were in the chromosome 8 inversion region, and only one was in a gene region, within an intron of Migut.H00680). Nucleotide diversity was higher in inland populations than in coastal populations (Fig. 2), and in general, coding regions were less diverse (CP $\pi = 0.039$; IA $\pi = 0.046$) than noncoding regions (CP $\pi = 0.056$; IA $\pi = 0.065$). Values for the absolute divergence (D_{xy}) between the ecotypes were approximately equivalent to the values of π, which suggests a high level of shared ancestral polymorphism between the ecotypes (Cruickshank & Hahn 2014). Given that absolute divergence between the ecotypes was not appreciably greater than within-ecotype diversity, we only present plots of D_{xy} values across the genome in Fig. S5 (Supporting

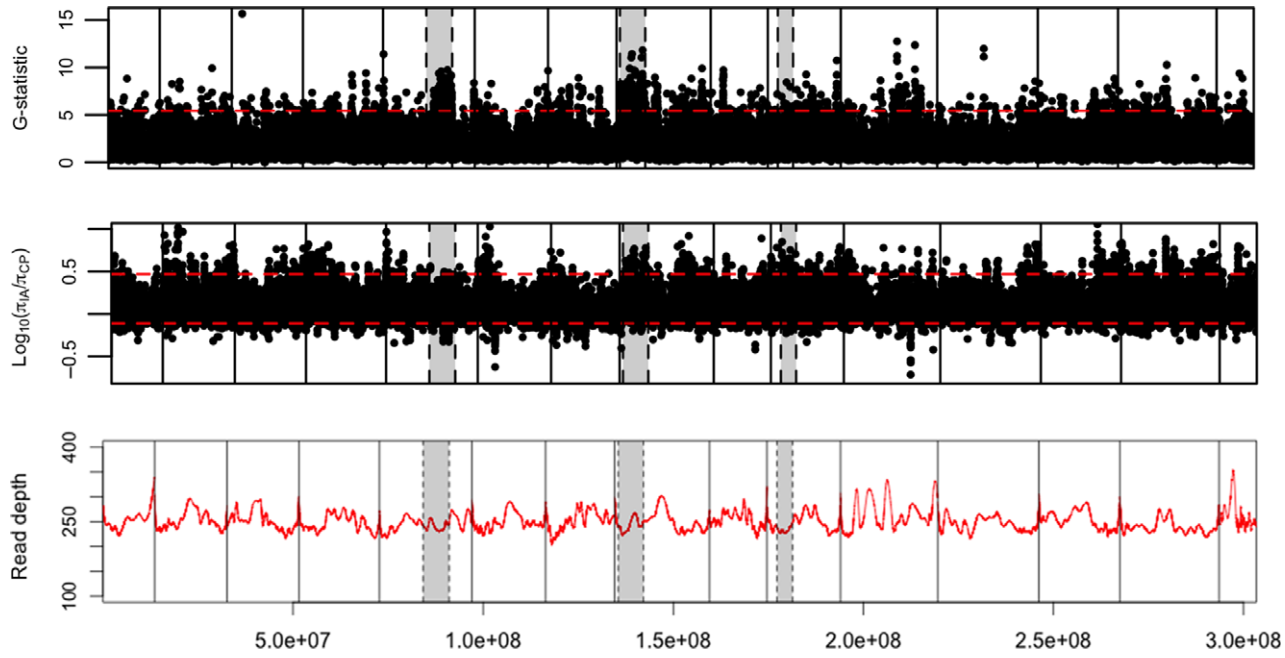

Fig. 2 G, log10 diversity ratio, and read depth values in 1000-bp windows across the genome. Inversion regions are shaded in grey. The top 1% of windows are above (or below) the red dashed line for each statistic. U, unordered scaffolds. Depth is represented as a Loess smoothed average with span = 0.1. [Colour figure can be viewed at wileyonlinelibrary.com]

information) and do not discus results further in the text. Allelic differentiation, G, was slightly higher in noncoding regions than in coding regions (3.15 vs. 2.42). Genewise Tajima's D was on average significantly lower (t-test, $P < 0.0001$) in the inland populations (average $D = 0.46$) than in the coastal populations (average $D = 0.78$), possibly indicating more recent population bottlenecks or higher population structure between inland groups.

Patterns of divergence for chromosomal inversions

All three chromosomal inversion regions of the genome identified in previous QTL analyses were more differentiated between ecotypes than the rest of the genome (Figs 3 and 4). While heterokaryotypes of the inversion on chromosome 8 are known to be widespread and to be at very different frequencies in coastal vs. inland populations, the geographic distribution of inversion heterokaryotypes for the chromosome 5 and 10 inversions are unknown. If alternative heterokaryotypes of an inversion were involved in adaptive divergence between coastal perennial and inland annual ecotypes, then we would expect elevated values for G in the region. We found that the average G-statistic across the inversion on chromosome 8 was 3.03, significantly higher than for the rest of genome ($G = 1.66$, d.f. = 2892.2, $P < 0.0001$). The average G across the inversion on chromosome 5 was also significantly elevated ($G = 3.01$ vs. 1.67, d.f. = 2151.5, $P < 0.0001$). The inverted region on chromosome 10 had only marginally higher G values than the rest of the genome ($G = 1.74$ vs. 1.70, d.f. = 1973.3, $P = 0.03$).

Coding regions in all three inversions were enriched for nonsynonymous SNPs compared with coding regions across the rest of the genome (all inversions, Fisher's exact test, $P < 0.0001$). Inversion regions on chromosomes 5, 8 and 10 contain 2.3%, 3.3% and 2.0% of the coding bases in the genome, respectively, but contain 2.6%, 4.2% and 7.9% of the nonsynonymous mutations in the genome, respectively. We expect a subset of the nonsynonymous mutations to be deleterious to protein function. Despite an overabundance of nonsynonymous SNPs in all the inverted regions, only the inversion region on chromosome 10 was also strongly enriched for SNPs predicted to have deleterious effects (i.e. create premature stops and starts, disrupt splice junctions) ($P < 0.0001$). The inversion on chromosome 8 was only moderately enriched ($P = 0.004$), and the inversion on chromosome 5 was not enriched ($P = 0.12$). This may indicate that nonsynonymous mutations in the chromosome 8 and 5 inversions are less deleterious on average than those found in inversion 10 and in the rest of the genome.

We also calculated Tajima's D for genes inside and outside the inverted regions (Fig. S3, Supporting information). Calculated within each ecotype pool, low or

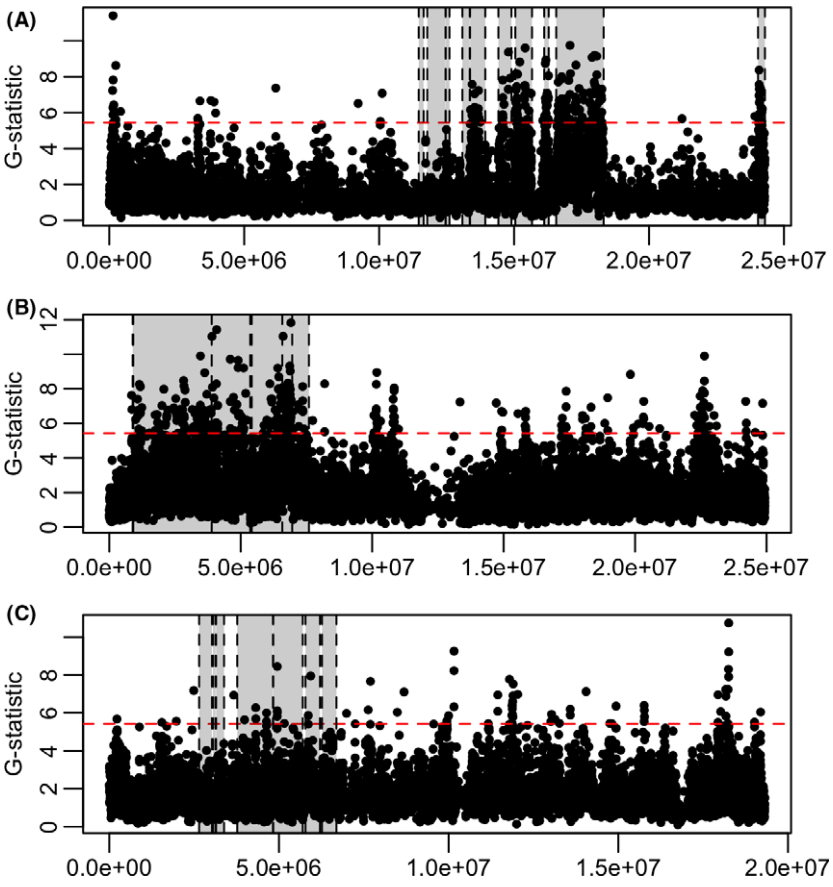

Fig. 3 G-statistic values across chromosomes 5 (A), 8 (B) and 10 (C). Shaded regions denote genomic scaffolds that are thought to be in inverted regions based on QTL mapping studies. Some scaffolds may be misordered in the current genome assembly. [Colour figure can be viewed at wileyonlinelibrary.com]

negative values of D may indicate an excess of rare SNPs within a gene, caused either by recent population expansion or by recent directional or purifying selection. Exceptionally low D values co-occurring with high differentiation or extreme diversity ratio values in the genome support the role of natural selection in those regions. D values were somewhat higher than estimates from previous studies (Flagel *et al.* 2014a,b; Puzey *et al.* 2016), which may result from pooling differentiated populations. While D values were positive on average, within both ecotypes the average D value for genes in inversions 5 and 8 (but not 10) was significantly lower (Wilcoxon tests, $P < 0.05$) inside the inversion than outside, potentially indicating recent selection on one or more inversion genes.

Lastly, we examined the ratio of nucleotide diversity in the inland annual ecotype to diversity in the coastal perennial ecotype (π ratio). Throughout the genome, on average there was greater nucleotide diversity within the inland annual ecotype than within the coastal perennial ecotype (average π ratio = 1.26). A local π ratio significantly larger or smaller than found across the rest of the genome may be caused by recent selection that leads to reduced nucleotide diversity in one group and not the other. Within all three inverted genome regions, we found the π ratio of inland vs. coastal diversity was significantly elevated above background levels, more so for chromosomes 5 and 8 than for chromosome 10 (Fig. S4, Supporting information). While the background diversity ratio was 1.26 genomewide, the diversity ratio was 1.47 across the chromosome 8 inversion, 1.31 in the chromosome 5 inversion and 1.36 in the chromosome 10 inversion (*t*-tests, $P < 0.0001$).

Identification of candidate genes for adaptive divergence

Overall, the technique of sequencing DNA pools by ecotype successfully reduced background levels of differentiation allowing for identification of candidate genomic regions associated with adaptive divergence. For *M. guttatus*, average population pairwise F_{ST} has been estimated to be ~0.48 (Lowry *et al.* 2008a), while average genomewide F_{ST} between ecotype pools in this study was 0.016 (for variable sites only). Not only was genomewide differentiation very low between sequencing pools, there was little genomewide relationship between allele frequency and differentiation (Fig. S2, Supporting information; $R^2_{CP} = 0.17$ and $R^2_{IA} = 0.16$).

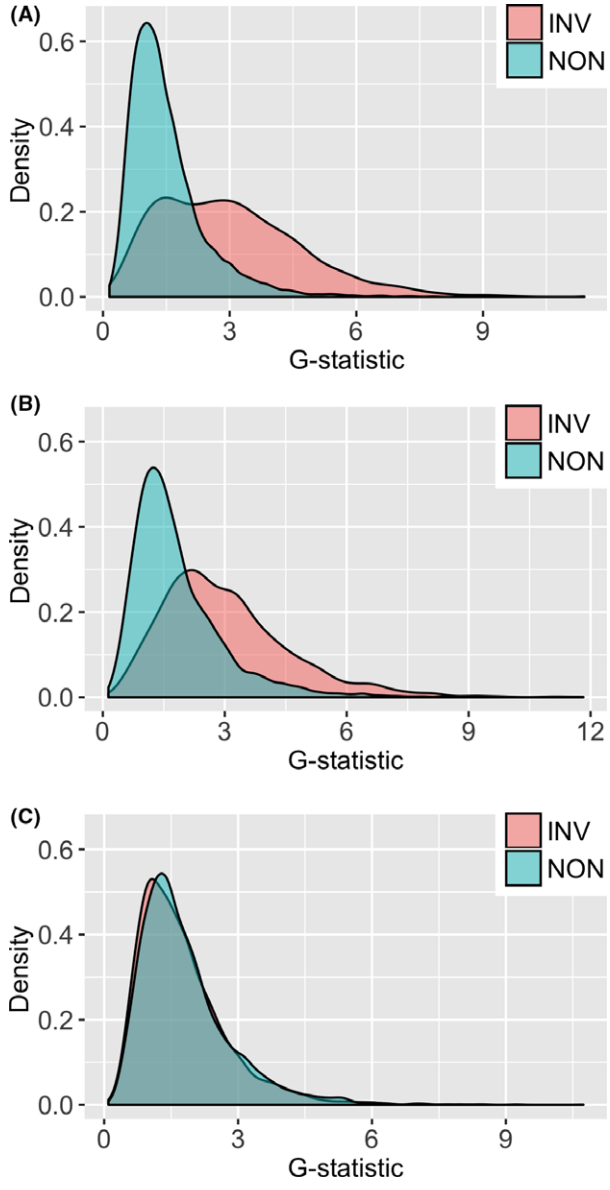

Fig. 4 The distribution of *G*-statistic window values inside and outside the chromosomal inversion on chromosomes 5 (A), 8 (B) and 10 (C). [Colour figure can be viewed at wileyonlinelibrary.com]

This implies that the presence of differentiated regions of the genome is not due to low-frequency variants that are present in only a few geographically isolated populations within one ecotype.

We identified candidate genes for adaptive divergence between the ecotypes, as those with either genic or promoter regions that overlapped with the 1% most differentiated windows for the *G*-statistic across the genome (Fig. 5). The top windows for *G* across the genome (N = 1131 windows) contained 667 distinct candidate genes and/or their promoter regions. Twenty-four

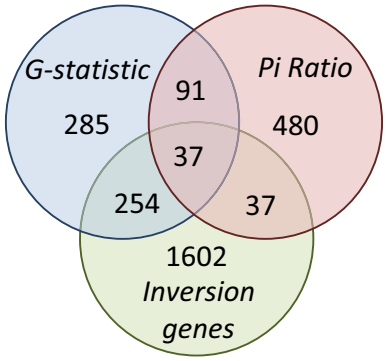

Fig. 5 Overlap of genes or their promoters identified in the top 1% of statistical outlier windows and in inversion regions on chromosomes 5, 8 and 10. [Colour figure can be viewed at wileyonlinelibrary.com]

per cent of these (N = 163) had both highly differentiated genic and promoter regions. Twenty per cent of the transcribed regions and 8% of the promoters were also in the top or bottom 1% of π ratio outlier windows, indicating they may have been affected by recent selective sweeps (Fig. 5). Forty-four per cent of the candidate genes (N = 291) fell in the chromosomal inversion regions on chromosome 5, 8 or 10 even though those inverted regions contain only 7% of all genes in the genome.

We determined the closest *Arabidopsis* sequence match for each candidate gene and examined their functional annotations (Table S2, Supporting information). Transcribed regions and their promoters in outlier G windows were not enriched for any biological functions. However, many genes with putatively adaptive functions were among the candidates (Fig. 6; Table S2, Supporting information), including at least 16 salt stress response genes, four gibberellic acid pathway genes, and other genes involved in flowering, growth, development and drought stress response (see Discussion).

Discussion

Overall, our study identified patterns of genomic divergence between plant ecotypes through pool sequencing. Our analysis found that coding regions are less diverse and less differentiated between the ecotypes than noncoding regions. We found that three previously identified chromosomal inversions were significantly more differentiated between the ecotypes than the rest of the genome, which may indicate they have been under selection during their evolution. Further, inversions were enriched for nonsynonymous mutations and contained candidate genes that appear to have undergone recent selective sweeps within ecotypes. We also identified a set of candidate genes for adaptive divergence in

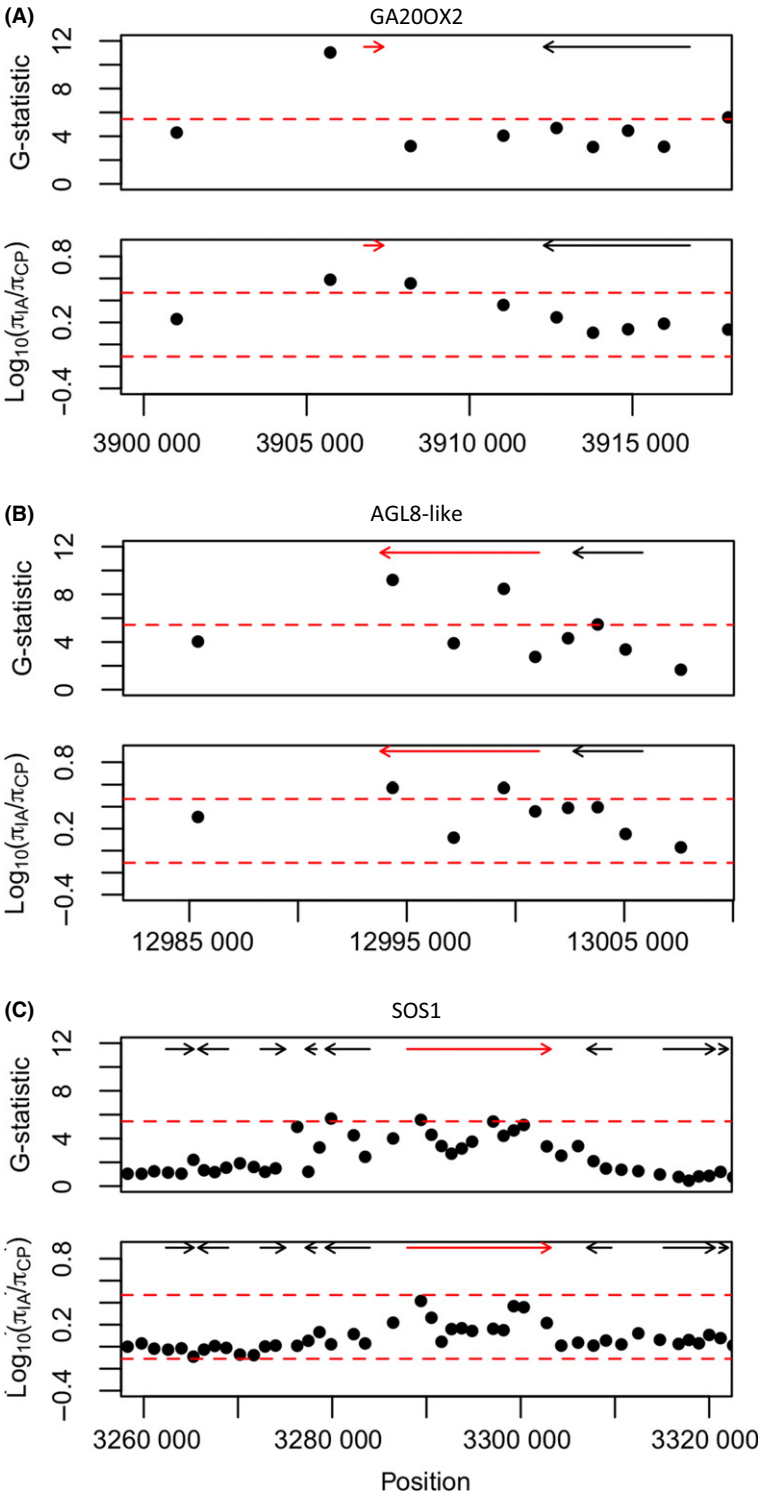

Fig. 6 Divergence (G) and diversity ratio window values near three candidate genes. (A) The GA20OX2 gene region, (B) the AGL8 gene region and (C) the SOS1 gene region. Arrows indicate genic regions, candidate gene in red. The dotted lines represent the 1% outlier window cut-off value for each statistic. [Colour figure can be viewed at wileyonlinelibrary.com]

development, flowering time, branching architecture and salt tolerance divergence between the ecotypes. Our results suggest that pooled ecotype sequencing has great potential for adaptation outlier analyses and may be especially beneficial in species where among population structure is high.

Differentiation and divergence for chromosomal inversions

As predicted, we found elevated allelic differentiation (*G*) across the inverted region of chromosome 8, which is known to play a role in adaptive divergence between

the coastal and inland ecotypes. The elevated allelic differentiation for the *G*-statistic in the chromosome 8 inverted regions is a nice validation of the pooled ecotype sequencing approach. We also found elevated differentiation for chromosome 5, but only marginally so for the inversion on chromosome 10. The frequency of the chromosome 5 and 10 inversions across *Mimulus guttatus* populations is unknown. However, there is evidence for the chromosome 5 inversion segregating in at least two other mapping populations (Friedman & Willis 2013; Garner *et al.* 2016). Friedman & Willis (2013) mapped a QTL for vernalization sensitivity to the chromosome 5 inversion within a perennial × perennial cross, indicating this polymorphism may contain adaptive variation. While elevated differentiation in inverted regions might result simply from reduced recombination alone (Nachman & Payseur 2012), there is other evidence that recent selection events have contributed to differentiation. Within ecotypes, Tajima's *D* on average was lower in genes in inverted regions 5 and 8, but not 10, supporting a role for positive selection in the former but not the latter. Genes in all inversions were enriched for nonsynonymous SNPs, but only the chromosome 10 inversion was enriched for strongly deleterious mutations. Together, these results suggest that the chromosome 5 inversion polymorphism (along with the chromosome 8 inversion) might be important for coast/inland adaptive divergence on a broad scale, while the chromosome 10 inversion might be rare and/or not consistently involved in adaptive divergence between coastal perennial and inland annual ecotypes across their geographic ranges. The chromosome 10 inversion may also be younger than the other two, not having had time to spread to high frequency.

Genetic mechanisms of adaptive ecotype divergence

One of the major factors driving ecotype formation in plants is differences in soil water availability across habitats (Clausen 1951; Porter 1966; Andrew *et al.* 2013; Lowry *et al.* 2014, 2015). Water availability directly impacts plants because it is the primary limiting factor for photosynthesis on Earth. Plants adapted to wetter habitats typically have larger leaves, more branches and greater overall size and flower later than plants from drier habitats (Stebbins 1952; Clausen & Hiesey 1958; Roux *et al.* 2006; Lowry 2012; Juenger 2013). These common differences in morphology between wet- and dry-adapted plant ecotypes are likely the result of a trade-off between growth and the consequent demand for water that results from that growth (Geber & Dawson 1990; Dudley 1996; Ackerly *et al.* 2000). While these patterns are well established, very little is known about the genetic basis of trade-offs in growth and reproduction that underlie plant ecotype formation. The divergence between coastal perennial and inland annual ecotypes of *M. guttatus* clearly reflects the classic plant trade-off in growth and reproduction (Hall & Willis 2006; Hall *et al.* 2006; Lowry *et al.* 2008b).

Our study identified multiple candidate genes that could be involved in the divergence in development between the ecotypes. Among overdifferentiated candidate genes and their promoters, 19% also had significantly reduced nucleotide diversity in one ecotype (Fig. 5), potentially indicating recent ecotype-specific selective sweeps. There is particularly strong evidence that some candidates are involved in adaptive divergence because they colocalize (occur within the 2-LOD drop confidence interval) with QTL identified in previous studies. Hall *et al.* (2006) originally identified QTL for flower size and developmental traits on chromosomes 3, 4 and 8 in the first annual (IM62) x perennial (DUN10) mapping experiment. The chromosome 3 QTL colocalizes with the outlier candidate gene *ent*-kaurene oxidase (*KO*; Migut.C00648), which is involved in gibberellic acid biosynthesis. The chromosome 4 QTL of Hall *et al.* (2006) colocalizes with developmental candidate gene AGAMOUS-like 8 (*AGL8*, Migut.D01622; Fig. 6B). The pleiotropic chromosome 11 QTL for flower and leaf size colocalized with candidate gene auxin response factor 8 (*ARF8*, Migut.K01158).

Chromosome 8 has two well-studied QTL (*DIV1* and *DIV2*) that both contribute to divergence in flower size, vegetative traits (internode length/width, leaf size, etc.), branching (stolon production), flowering time and critical photoperiod (Hall *et al.* 2006, 2010; Lowry & Willis 2010; Friedman & Willis 2013; Friedman 2014; Friedman *et al.* 2015; Twyford & Friedman 2015). The *DIV1* QTL is the chromosomal inversion, which represents 2.1% of the genome, but contains 21.9% of all candidate genes identified in this study. Many of these outliers are likely a by-product of reduced recombination in the inverted region (Noor & Bennett 2009; Nachman & Payseur 2012; Cruickshank & Hahn 2014). However, allelic differentiation (*G*) was far from uniformly elevated across the chromosome 8 inversion (Fig. 3B). The inversion contained many outliers that could be involved in developmental differences between the ecotypes, including genes involved in gibberellic acid biosynthesis and signalling (*GA20OX2*, Migut.H00683, Fig. 5A; *GASA4*, Migut.H00923), multiple auxin-related genes (*IAA8*, Migut.H00298; *VFB2*, Migut.H00412; *PGP19*, Migut.H00909; *AAR3*, Migut.H00929) and an *R2R3*-Myb transcription factor involved in the brassinosteroid signalling pathway (*MYB56*, Migut.H00933). *IAA8* had significantly reduced Tajima's *D* value in the coastal populations ($D = 0.43$ inland vs. $D = -1.39$ coastal) indicating it may have recently been under selection. A

critical gene involved in abscisic acid biosynthesis (*ABA1*; Migut.H00431) was also highly differentiated in the chromosome 8 inversion, which could potentially be involved in adaptation to wet vs. dry habitats. As the inversion has been established to contribute to ecological reproductive isolation in nature (Lowry & Willis 2010), these genes are candidates for both adaptation and speciation.

The coastal perennial ecotype, when compared to the inland annual ecotype, in many ways resembles the dwarfing phenotype that results from gibberellin acid (GA) gene mutations in cultivated plants (Peng *et al.* 1999; Sasaki *et al.* 2002). Further, addition of GA to plants 'rescues' some of the phenotypic differences between coastal perennial and inland annual ecotypes of *M. guttatus*, especially differences in internode elongation and growth architecture (D.B. Lowry, unpublished). The GA biosynthetic gene gibberellin 20-oxidase (*GA20OX2*, Migut.H00683; Fig. 3A), found within the chromosome 8 inversion (*DIV1*), is a compelling candidate gene that causes dwarfing in rice (Sasaki *et al.* 2002; Spielmeyer *et al.* 2002). This candidate also had highly reduced Tajima's *D* in the coastal ecotype (−1.22) vs. the inland ecotype (0.46), a potential indicator of recent selective sweep at this locus. The other chromosome 8 QTL (*DIV2*) had a *G* outlier peak in the promoter of another important green revolution dwarf gene, *gibberellic acid insensitive* (*GAI*, Migut.H02266), which is involved in GA signalling (Peng *et al.* 1999). The *DIV1* and *DIV2* QTL are responsible for much of the morphological and life history divergence between coastal and inland ecotypes and have recently been shown to interact epistatically (Friedman 2014). *GA20OX* and the DELLA genes, including *GAI*, are known to regulate each other in a feedback loop (Fleet & Sun 2005; Achard & Genschik 2009) and thus represent a hypothesis to explain the epistatic interaction of *DIV1* and *DIV2*.

The coastal headlands, cliffs and dunes of the world are populated by salt-tolerant plants which have adapted to oceanic salt spray and saline soils (Boyce 1954; Ahmad & Wainwright 1976; Griffiths & Orians 2003; Busoms *et al.* 2015). While salt tolerance is well established in coastal plants, the physiological and genetic mechanisms underlying the salt tolerance of coastal plants are poorly understood. Both coastal and inland ecotypes of *M. guttatus* take up soil Na^+ ions into the shoot (which could be a mechanism of osmotic adjustment) and experience nutrient deficiencies in shoot K^+ under high-salt conditions (Lowry *et al.* 2009). However, the coastal ecotype has profoundly greater leaf tissue tolerance to salt than the inland ecotype in both soil salinity and salt spray experiments (Lowry *et al.* 2008a,b, 2009; Selby *et al.* 2014). These observations collectively suggest that ion homoeostasis plays a key role in the evolution of salt tolerance of the coastal ecotype. Thus, we hypothesized that genes involved in salt tolerance through ion homoeostasis were likely to be outliers in our analysis. Indeed, many of the genes that are well established to be involved in ion homoeostasis in other plants (or their promoters) were in the top 1% most differentiated regions of the genome, including *SOS1* (Migut.E00570; Fig. 5C), *SOS3* (Migut.E01258), *SnRK2* (Migut.H00468), two orthologs of *CBL10* (Migut.A00138, Migut.E01561) and four orthologs of *HKT1* (Migut.J00828, Migut.L01905, Migut.L01906, Migut.L01907). *SOS3*, *CBL10* and *SnRK2* are all involved in regulation of responses to salt stress (Qiu *et al.* 2002; Kim *et al.* 2007; Fujita *et al.* 2009). *SOS1* and *HKT1* are plant plasma membrane Na^+ transporters that play a critical role in removal of sodium from plant cells and organs to prevent damage and maintain homoeostasis (Deinlein *et al.* 2014). *SOS1* had locally reduced diversity in coastal pool compared with inland pool with the highest π ratio value for the gene in the top 1.7% of ratio values genomewide (Fig. 4C). *HKT1* has been shown to contribute to salt tolerance variation in *Arabidopsis thaliana* (Rus *et al.* 2006), and allelic variation in *A. thaliana HKT1* is correlated with proximity to the ocean (Baxter *et al.* 2010). Tajima's *D* was strongly reduced in the *M. guttatus* coastal ecotype (−0.87) vs. the inland ecotype (0.78) for one of the *HKT1* homologs (Migut.J00828), a potential indicator of recent selection at the locus.

The future of pooled ecotype sequencing

Pooled ecotype sequencing has multiple benefits for identification of the loci involved in local adaptation. Whole-genome sequencing of hundreds of lines is still too expensive for most research programmes. Further, genotyping using reduced representation libraries, such as RAD, MSG and GBS, is likely to miss the vast majority of outlier loci (Tiffin & Ross-Ibarra 2014; Hoban *et al.* 2016), especially in systems such as ours where linkage disequilibrium extends for only short distances (Brandvain *et al.* 2014). Pool-seq allows for economical and efficient whole-genome population genetics. However, pool-seq of a small number of populations can be challenged by high levels of population structure. By pooling at the ecotype level, studies such as ours can identify outliers between groups, where structure can be much lower than among populations. Pooled ecotype sequencing can be aided if there are a priori expectations of divergence for particular chromosomal regions (e.g. inversions and QTL), which can be used as a validation of the method.

Despite the great potential of pooled ecotype sequencing, there are also major caveats that should be taken into

account. As with all genomic outlier studies, there are many issues that can contribute to both false positive and false negative results (Tiffin & Ross-Ibarra 2014; Hoban *et al*. 2016). Further, pooled ecotype sequencing contains hierarchical population structure, which could lead to false positives if an island model is assumed for the two ecotypes (Excoffier *et al*. 2009). While we did not assume any model for establishing significance thresholds for summary statistics (instead opting to use a top 1% threshold), future research should attempt to develop better methods for establishing significant thresholds for pooled ecotype sequencing studies.

Conclusions

Overall, our study contributes to the mounting evidence in *Mimulus guttatus* and other species (Dobzhansky 1970; Joron *et al*. 2011; Schwander *et al*. 2014; Thompson & Jiggins 2014; Küpper *et al*. 2016; Lamichhaney *et al*. 2016) that chromosomal inversions play a key role in the processes of adaptation and speciation. Ecotype pool-seq allowed us to identify candidate genes for classic wet and dry habitat adaptations within inversions as well as across the rest of the genome. In the near future, we plan to combine these outlier results with gene expression studies to identify a set of top candidate genes for ecotype divergence. Ultimately, functional molecular genetics will be required to understand the underlying genetic mechanisms in the evolution of ecotypes. With the recent discovery of new genome editing methods (Doudna & Charpentier 2014), there are now excellent ways forward to test gene function.

Acknowledgements

We thank John Kelly for supplying us with detailed information regarding the locations of the chromosome 5 and 10 inversions. Kevin Wright and members of the NSF NIMBioS Computational Landscape Genomics working group contributed ideas about experimental design and analyses. Findley Finseth, Chris Oakley, Emily Ditmar, Damian Popovic, Caitlyn Byron, Sam Perez and three anonymous reviewers provided helpful comments on the manuscript. Funding for this study was provided by Michigan State University.

References

Achard P, Genschik P (2009) Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. *Journal of Experimental Botany*, **60**, 1085–1092.

Ackerly DD, Dudley SA, Sultan SE *et al*. (2000) The evolution of plant ecophysiological traits: recent advances and future directions. *BioScience*, **50**, 979–995.

Adrion JR, Hahn MW, Cooper BS (2015) Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends in Genetics*, **31**, 434–444.

Ahmad I, Wainwright SI (1976) Ecotype differences in leaf surface properties of *Agrostis stolonifera* from salt marsh, spray zone and inland habitats. *New Phytologist*, **76**, 361–366.

Alexa A, Rahnenfuhrer J (2016). *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.24.0.

Andrew RL, Kane NC, Baute GJ, Grassa CJ, Rieseberg LH (2013) Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Molecular Ecology*, **22**, 799–813.

Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S (2014) FastQC. A quality control tool for high throughput sequence data. *Babraham Bioinformatics* http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ayala D, Ullastres A, González J (2014) Adaptation through chromosomal inversions in *Anopheles*. *Frontiers in Genetics*, **5**, 129.

Baker RL, Diggle PK (2011) Node-specific branching and heterochronic changes underlie population-level differences in *Mimulus guttatus* (Phrymaceae) shoot architecture. *American Journal of Botany*, **98**, 1924–1934.

Baker RL, Hileman LC, Diggle PK (2012) Patterns of shoot architecture in locally adapted populations are linked to intraspecific differences in gene regulation. *New Phytologist*, **196**, 271–281.

Baxter I, Brazelton JN, Yu D *et al*. (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *AtHKT1;1*. *PLoS Genetics*, **6**, e1001193.

Boyce SG (1954) The salt spray community. *Ecological Monographs*, **24**, 29–67.

Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL (2014) Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genetics*, **10**, e1004410.

Busoms S, Teres J, Huang XY *et al*. (2015) Salinity is an agent of divergent selection driving local adaptation of *Arabidopsis* to coastal habitats. *Plant Physiology*, **168**, 915–929.

Cheng C, White BJ, Kamdem C *et al*. (2012) Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics*, **190**, 1417–1432.

Cingolani P, Platts A, Wang LL *et al*. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.

Clausen J (1951) *Stages in the Evolution of Plant Species*. Cornell University Press, Ithaca, New York, USA.

Clausen J, Hiesey W (1958) *Experimental Studies on the Nature of Species*. Carnegie Institution of Washington, Washington, DC, USA.

Colosimo PF, Hosemann KE, Balabhadra S *et al*. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.

Coop G, Pickrell JK, Novembre J *et al*. (2009) The role of geography in human adaptation. *PLoS Genetics*, **5**, e1000500.

Coyne J, Orr H (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts, USA.

Coyne JA, Meyers W, Crittenden AP, Sniegowski P (1993) The fertility effects of pericentric inversions in *Drosophila melanogaster*. *Genetics*, **134**, 487–496.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Deinlein U, Stephan AB, Horie T, Luo W, Xu G, Schroeder JI (2014) Plant salt-tolerance mechanisms. *Trends in Plant Science*, **19**, 371–379.

Des Marais DL, Auchincloss LC, Sukamtoh E *et al.* (2014) Variation in MPK12 affects water use efficiency in *Arabidopsis* and reveals a pleiotropic link between guard cell size and ABA response. *Proceedings of the National Academy of Sciences of the USA*, **111**, 2836–2841.

Dobzhansky T (1970) *Genetics of the Evolutionary Process*. Columbia University Press, New York, New York, USA.

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.

Dudley SA (1996) The response to differing selection on plant physiological traits: evidence for local adaptation. *Evolution*, **50**, 103–110.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Fabian DK, Kapun M, Nolte V *et al.* (2012) Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*, **21**, 4748–4769.

Flagel LE, Bansal R, Kerstetter RA *et al.* (2014a) Western corn rootworm (*Diabrotica virgifera virgifera*) transcriptome assembly and genomic analysis of population structure. *BMC Genomics*, **15**, 195.

Flagel LE, Willis J, Vision TJ (2014b) The standing pool of genomic structural variation in a natural population of *Mimulus guttatus*. *Genome Biology and Evolution*, **1**, 53–64.

Fleet CM, Sun TP (2005) A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Current Opinion in Plant Biology*, **8**, 77–85.

Frary A, Nesbitt TC, Frary A *et al.* (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85–88.

Friedman J (2014) Genetic determinants and epistasis for life history trait differences in the common monkeyflower, *Mimulus guttatus*. *Journal of Heredity*, **105**, 910–921.

Friedman J, Willis JH (2013) Major QTLs for critical photoperiod and vernalization underlie extensive variation in flowering in the *Mimulus guttatus* species complex. *New Phytologist*, **199**, 571–583.

Friedman J, Twyford AD, Willis JH, Blackman BK (2015) The extent and genetic basis of phenotypic divergence in life history traits in *Mimulus guttatus*. *Molecular Ecology*, **24**, 111–122.

Fujita Y, Nakashima K, Yoshida T *et al.* (2009) Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in *Arabidopsis*. *Plant and Cell Physiology*, **50**, 2123–2132.

Garner AG, Kenney AM, Fishman L, Sweigart AL (2016) Genetic loci with parent-of-origin effects cause hybrid seed lethality in crosses between *Mimulus* species. *New Phytologist*, **211**, 319–331.

Geber MA, Dawson TE (1990) Genetic variation in and covariation between leaf gas exchange, morphology, and development in *Polygonum arenastrum*, an annual plant. *Oecologia*, **85**, 153–158.

Gompert Z, Lucas LK, Nice CC, Fordyce JA, Alex Buerkle C, Forister ML (2013) Geographically multifarious phenotypic divergence during speciation. *Ecology and Evolution*, **3**, 595–613.

Griffiths ME, Orians CM (2003) Salt spray differentially affects water status, necrosis, and growth in coastal sandplain heathland species. *American Journal of Botany*, **90**, 1188–1196.

Hall MC, Willis JH (2006) Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. *Evolution*, **60**, 2466–2477.

Hall MC, Basten C, Willis JH (2006) Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in *Mimulus guttatus*. *Genetics*, **172**, 1829–1844.

Hall MC, Lowry DB, Willis JH (2010) Is local adaptation in *Mimulus guttatus* caused by tradeoffs at individual loci? *Molecular Ecology*, **19**, 2739–2753.

Hellsten U, Wright KM, Jenkins J *et al.* (2013) Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Science of the USA*, **110**, 19478–19482.

Hendry AP, Taylor EB, McPhail JD (2002) Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system. *Evolution*, **56**, 1199–1216.

Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the genomic basis of local adaptation in non-model organisms: pitfalls, practical solutions, and future directions. *The American Naturalist*. In press.

Holeski LM, Monnahan P, Koseva B, McCool N, Lindroth RL, Kelly JK (2014) A high-resolution genetic map of yellow monkeyflower identifies chemical defense QTLs and recombination rate variation. *G3: Genes, Genomes, Genetics*, **4**, 813–821.

Joron M, Frezal L, Jones RT *et al.* (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–206.

Juenger TE (2013) Natural variation and genetic constraints on drought tolerance. *Current Opinions in Plant Biology*, **16**, 274–281.

Kapun M, Fabian DK, Goudet J, Flatt T (2016) Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **5**, 1317–1336.

Kim BG, Waadt R, Cheong YH *et al.* (2007) The calcium sensor CBL10 mediates salt tolerance by regulating ion homeostasis in *Arabidopsis*. *The Plant Journal*, **52**, 473–484.

Kirkpatrick M, Barton NH (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, **173**, 419–434.

Kofler R, Betancourt AJ, Schlötterer C (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1002487.

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, **9**, 29.

Küpper C, Stocks M, Risse JE *et al.* (2016) A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, **48**, 79–83.

Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371–375.

Lamichhaney S, Berglund J, Almén MS *et al.* (2016) Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, **48**, 84–88.

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 (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Lowry DB (2012) Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society*, **106**, 241–257.

Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology*, **8**, e1000500.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH (2008a) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B*, **363**, 3009–3021.

Lowry DB, Rockwood RC, Willis JH (2008b) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution*, **62**, 2196–2214.

Lowry DB, Hall MC, Salt DE, Willis JH (2009) Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytologist*, **3**, 776–788.

Lowry DB, Behrman KD, Grabowski P (2014) Adaptation between ecotypes and along environmental gradients in *Panicum virgatum*. *The American Naturalist*, **183**, 682–692.

Lowry DB, Hernandez K, Taylor SH *et al.* (2015) The genetics of divergence and reproductive isolation between ecotypes of *Panicum hallii*. *New Phytologist*, **205**, 402–414.

Lu F, Lipka AE, Glaubitz J *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, **9**, e1003215.

Lynch M, Bost D, Wilson S (2014) Population-genetic inference from pooled-sequencing data. *Genome Biology and Evolution*, **6**, 1210–1218.

Magwene PM, Willis JH, Kelly JK (2011) The statistics of bulk segregant analysis using next generation sequencing. *PLoS Computational Biology*, **7**, e1002255.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, **17**, 10–12.

McKay JK, Richards JH, Nemali KS (2008) Genetics of drought adaptation in *Arabidopsis thaliana II*. QTL analysis of a new mapping population, KAS-1 x TSU-1. *Evolution*, **62**, 3014–3026.

Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 409–421.

Nei M, Li W-H (1979) Mathematical model for studying genic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **79**, 5269–5273.

Noor MA, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**, 439–444.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford, UK.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.

Oneal E, Lowry DB, Wright KM, Zhu Z, Willis JH (2014) Divergent population structure and climate associations of a chromosomal inversion polymorphism across the *Mimulus guttatus* species complex. *Molecular Ecology*, **23**, 2844–2860.

Peng J, Richards DE, Hartley NM *et al.* (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, **400**, 256–261.

Porter CL (1966) An analysis of variation between upland and lowland Switchgrass *Panicum virgatum* L in Central Oklahoma. *Ecology*, **47**, 980–992.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, R208–R215.

Puzey J, Willis J, Kelly J (2016) Whole genome sequencing of 56 Mimulus individuals illustrates population structure and local selection. *Molecular Ecology*, In review.

Qiu QS, Guo Y, Dietrich MA, Schumaker KS, Zhu JK (2002) Regulation of SOS1, a plasma membrane Na+/H+ exchanger in Arabidopsis thaliana, by SOS2 and SOS3. *Proceedings of the National Academy of Sciences*, **99**, 8436–8441.

Raineri E, Ferretti L, Esteve-Codina A, Nevado B, Heath S, Pérez-Enciso M (2012) SNP calling by sequencing pooled samples. *BMC Bioinformatics*, **13**, 239.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–4370.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.

Roux F, Touzet P, Cuguen J, Le Corre V (2006) How to be early flowering: an evolutionary perspective. *Trends in Plant Science*, **11**, 375–381.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Rus A, Baxter I, Muthukumar B *et al.* (2006) Natural variants of AtHKT1 enhance Na+ accumulation in two wild populations of *Arabidopsis*. *PLoS Genetics*, **2**, e210.

Sasaki A, Ashikari M, Ueguchi-Tanaka M *et al.* (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature*, **416**, 701–702.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Schwander T, Libbrecht R, Keller L (2014) Supergenes and complex phenotypes. *Current Biology*, **24**, R288–R294.

Selby JP, Jeong AL, Toll K, Wright KM, Lowry DB (2014). Methods and discoveries in the pursuit of understanding the genetic basis of adaptation to harsh environments in *Mimulus*. In: *Plant Ecology and Evolution in Harsh Environments* (eds Rajakaruna N, Boyd RS, Harris TB), pp. 243–265. NOVA Science Publishers, Hauppauge, New York, USA.

Sobel JM, Chen GF (2014) Unification of methods for estimating the strength of reproductive isolation. *Evolution*, **68**, 1511–1522.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution*, **64**, 295–315.

Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (*sd-1*), "green revolution" rice, contains a defective gibberellin

20-oxidase gene. *Proceedings of the National Academy of Sciences*, **99**, 9043–9048.
Stebbins GL (1952) Aridity as a stimulus to plant evolution. *The American Naturalist*, **86**, 33–44.
Sweigart AL, Flagel LE (2015) Evidence of natural selection acting on a polymorphic hybrid incompatibility locus in *Mimulus*. *Genetics*, **199**, 543–554.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. *Heredity*, **113**, 1–8.
Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.
Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, **42**, 260–263.
Twyford AD, Friedman J (2015) Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, **69**, 1476–1486.
Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.
Wright KM, Arnold B, Xue K, Šurinová M, O'Connell J, Bomblies K (2015) Selection on meiosis genes in diploid and tetraploid Arabidopsis arenosa. *Molecular Biology and Evolution*, **32**, 944–955.

D.L. and B.G. conceived the study and wrote the manuscript. Y.C. conducted laboratory experiments. B.G. conducted data analysis.

Data accessibility

Raw pooled sequencing files are available through the NCBI Sequence Read Archive under project identifier SRP074310 – accessions SRX1740606 (inland pool) and SRX1740605 (coastal pool). SNAPE genotype call files for each pool are available upon request.

Supporting information

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

Fig. S1 F_{ST} for 1000 bp windows across the genome.

Fig. S2 Differentiation vs. average alternate allele frequency in the coastal perennial sequencing pool for 1 kb windows across the *Mimulus* genome.

Fig. S3 Average Tajima's *D* values for genes inside and outside the inversion regions.

Fig. S4 Pi and pi ratio values for inversion regions and non-inversion regions in the two ecotypes (IA and CP).

Fig. S5 D_{xy} for 1000 bp windows across the genome.

Table S1 Populations used for pooled sequencing.

Table S2 Candidate ecotypic adaptation genes.