MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 4472–4485 doi: 10.1111/mec.12875

Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*)

JOSHUA PUZEY*† and MARIO VALLEJO-MARÍN‡

*Department of Biology, Duke University, Durham, NC 27705, USA, †Department of Biology, College of William and Mary, Williamsburg, VA 23185, USA, ‡Biological and Environmental Sciences, University of Stirling, Stirling FK9 4LA, UK

Abstract

Global trade and travel is irreversibly changing the distribution of species around the world. Because introduced species experience drastic demographic events during colonization and often face novel environmental challenges from their native range, introduced populations may undergo rapid evolutionary change. Genomic studies provide the opportunity to investigate the extent to which demographic, historical and selective processes shape the genomic structure of introduced populations by analysing the signature that these processes leave on genomic variation. Here, we use next-genera tion sequencing to compare genome-wide relationships and patterns of diversity in native and introduced populations of the yellow monkeyflower (*Mimulus guttatus*). Genome resequencing data from 10 introduced populations from the United Kingdom (UK) and 12 native *M. guttatus* populations in North America (NA) demonstrated reduced neutral genetic diversity in the introduced range and showed that UK populations are derived from a geographic region around the North Pacific. A selective-sweep analysis revealed site frequency changes consistent with selection on five of 14 chromosomes, with genes in these regions showing reduced silent site diversity. While the target of selection is unknown, genes associated with flowering time and biotic and abiotic stresses were located within the swept regions. The future identification of the specific source of origin of introduced UK populations will help determining whether the observed selective sweeps can be traced to unsampled native populations or occurred since dispersal across the Atlantic. Our study demonstrates the general potential of genome-wide analyses to uncover a range of evolutionary processes affecting invasive populations.

Keywords: anthropogenic dispersal, genome scan, introduced species, long-distance colonization, next-generation sequencing, selective sweeps

Received 30 April 2014; revision received 2 July 2014; accepted 22 July 2014

Introduction

The introduction of species beyond their native ranges can affect ecological and evolutionary interactions in the new habitat (Cox 2004; Phillips & Shine 2006; Liu & Pemberton 2009; Ricciardi *et al.* 2013) and can negatively impact levels of local biodiversity and result in high economic costs (Pimentel 2002; Williams *et al.* 2010; Vila *et al.* 2011). Introduced populations are often used as models to investigate rapid genetic changes and adaptation to novel environments, thus providing valuable insights into basic biological processes including local adaptation (Sax *et al.* 2007; Prentis *et al.* 2008). In particular, genetic analyses continue to play a central role in studies of the origin and establishment of introduced populations, as well as of the mechanisms that permit the colonization and drive the spread of popula-

Correspondence: Mario Vallejo-Marín,
Fax: +44 (0) 1786 467843; E-mail: mario.vallejo@stir.ac.uk

tions beyond their native range (Baker & Stebbins 1965; Lee 2002).

The genomic structure of non-native populations is influenced by a variety of processes, including population bottlenecks, multiple introductions, population expansion, gene flow between populations and selection, among others (Lee 2002). For instance, in populations established after limited long-distance dispersal events, the level of genetic diversity can be significantly lower than in the native range, reflecting population bottlenecks (Lachmuth *et al.* 2010; Ness *et al.* 2012). However, introduction of multiple individuals from the same population, or multiple introductions from genetically diverse source populations, can counteract the loss of diversity or even result in higher levels of genetic variation within introduced populations compared with native ones (Dlugosch & Parker 2008). The level of standing variation in introduced populations is relevant to the colonization process, as severe bottlenecks and reduced diversity could indicate potential limitations for the rapid evolution of adaptive traits in novel environments (Barrett & Schluter 2008; Lachmuth *et al.* 2010; Siol *et al.* 2010; Messer & Petrov 2013). Severe bottlenecks resulting in globally reduced diversity may indicate that natural selection is mutation limited. However, genetic variation resulting from introduction of multiple individuals can provide ample standing variation for natural selection.

Genome-wide studies have been employed to investigate genetic patterns in natural populations, including the relationship between native and introduced populations as well as invasion pathways of exotic plants and animals (Jahodová *et al.* 2007; Dlugosch *et al.* 2013; Tarnowska *et al.* 2013). Genome scans allow detecting selection acting on specific locations in the genome (Nielsen *et al.* 2005), and by comparing the sites under selection in the genomes of different populations, it is possible to identify candidates for genetic regions associated with local adaptation (Savolainen *et al.* 2013). A prerequisite to any genome-wide study is identifying a large number of genetic markers, such as restriction site polymorphisms (e.g. AFLPs, Vos *et al.* 1995), or single-nucleotide polymorphisms (SNPs). The growing access to high-throughput sequencing technologies at low costs opens the opportunity to conduct genome-wide studies at an unprecedented depth, even in nonmodel organisms (Prentis *et al.* 2010; Twyford & Ennos 2012; Ellegren 2014).

The generation of genome-wide markers by high-throughput sequencing can employ methods for genome complexity reduction, such as transcriptome sequencing (Dlugosch *et al.* 2013) or RAD sequencing (Davey *et al.* 2011; Roda *et al.* 2013). However, for the increasing number of species in which a reference genome is available, whole-genome resequencing allows genotyping markers, such as SNPs, which may occur at high densities across the genome (Davey *et al.* 2011; Twyford & Ennos 2012; Savolainen *et al.* 2013). Importantly, whole-genome resequencing removes many of the ascertainment biases associated with SNP chips or other genome reduction technologies. The dense marker saturation achieved through genome resequencing is particularly useful for detecting the footprint of selection acting on specific locations across the genome. For instance, selective sweeps, in which selection drives previously rare alleles to fixation, also reduce diversity at neighbouring regions around the selected site (Messer & Petrov 2013). The signal left behind by selective sweeps can be detected by comparing patterns of variation along the genome with the level expected under a null model. Hard selective sweeps, where a single variant is driven to fixation, leave a characteristic footprint in the genome, which can be identified using summary statistics such as Tajima's *D* or the composite likelihood ratio (CLR) (Nielsen *et al.* 2005; Messer & Petrov 2013). These statistics may be particularly powerful to detect recent selective sweeps as linkage disequilibrium (LD) between the selected site and the surrounding variation is expected to be highest immediately following the fixation of the adaptive allele.

Genomic studies of native and introduced populations can uncover demographic, historical and selective processes by analysing the signature that these processes leave on genomic variation. Here, we use whole-genome resequencing to assess the relationship between native and introduced populations and to uncover selective episodes in specific regions of the genome of introduced populations. We study the yellow monkeyflower (*Mimulus guttatus*, Phrymaceae), a species that has long been used as a model for ecological and evolutionary studies in its native range (Vickery 1959; Wu *et al.* 2008), and which has become naturalized in eastern North American, New Zealand, Iceland, the Faroe Islands and Western Europe (van Kleunen & Fischer 2008; Murren *et al.* 2009; Tokarska-Guzik & Dajdok 2010), becoming particularly widespread in the UK (Preston *et al.* 2002; Vallejo-Marín & Lye 2013). *Mimulus guttatus* is ideally suited for studying the ecological genomics of non-native populations due to its recent introduction and spread (<200 years), abundant information on the ecology and evolution of native populations, and the availability of a full-genome sequence, which provides a backbone for analysing and interpreting patterns of genetic variation in introduced populations. The relatively small genome of *M. guttatus* (1 N = 430 MB) makes this species a good candidate for population genomic studies through resequencing, as multiple individuals can be analysed with a relatively small budget. We analysed previously available and

newly generated whole-genome sequence data for 12 native and 10 introduced British populations of *M. guttatus*, as well as five additional related taxa (n = 35 *Mimulus* genomes in total). Our data set allowed us to address three specific aims: (i) to determine the level of genome-wide diversity present in introduced populations of *M. guttatus* in the UK; (ii) to investigate the genetic relationships between native and introduced populations; and (iii) to search for evidence of hard selective sweeps in introduced populations.

Methods

Study system

The *Mimulus guttatus* species complex includes a set of phenotypically variable, interfertile taxa with a native range of distribution in western North America from northern Mexico to Alaska (Grant 1924; Wu *et al.* 2008). Within this complex, populations of *M. guttatus* Fischer ex DC. (Grant 1924) show marked variation in characteristics including life history (annual/perennial) (Hall & Willis 2006; Lowry & Willis 2010), mating system (Ritland 1990; Dole 1992), phenology (Hall & Willis 2006; Friedman & Willis 2013), floral morphology (Fishman *et al.* 2002), edaphic adaptations (e.g. tolerance to elevated concentrations of heavy metals or salt, Macnair & Watkins 1983; Lowry *et al.* 2008, 2009), habitat preferences (Wu *et al.* 2008), chromosome number (most populations are diploid: $2n = 2x = 28$, but tetraploids also occur in the native range, Sweigart *et al.* 2008) and clonal growth (Dole 1992; van Kleunen 2007), among others. This incredible diversity has led some taxonomists to subdivide *M. guttatus* into numerous morphological species (e.g. Pennell 1951; Nesom 2012). Here, we adhere to the broader circumscription of *M. guttatus* Fischer ex DC. (Grant 1924; Wu *et al.* 2008).

Mimulus guttatus was introduced into the British Isles in 1812, and the first naturalized populations were reported in England around 1830 (Roberts 1964; Parker 1975). In the UK, *M. guttatus* is currently widespread and occurs in wet habitats along the banks of rivers and streams, in ditches, marshy areas and other wet places (Stace 2010; Truscott et al. 2006; Vallejo-Marín & Lye 2013). It propagates via both seeds and clonally through lateral stems that root freely at the nodes. The source of the first naturalized populations of *M. guttatus* in the UK is unknown, but one of the earliest specimens of this taxon to reach Europe was derived from material collected by Langsdorff between 1806 and 1810 in the Aleutian Islands in Alaska and transmitted to the Botanic Gardens at Cambridge (Sims 1812; Pennell 1935, p. 116). The use of *Mimulus* spp. as a horticultural species in Victorian England, as reflected by being readily available in seed catalogues of the time (e.g. Gardeners' Chronicle 1852), raises the distinct possibility that *M. guttatus* was introduced into the UK on repeated occasions and from multiple sources.

Population sampling

Analysing genomes across a wide geographic scale represents a trade-off between the numbers of individuals vs. populations sampled. The goal of this study was to determine the introduction history of *Mimulus* into the UK and the effects of the introduction on nucleotide diversity and to identify signals of selective sweeps that are common across the UK. To do this, we sought to obtain samples from geographically disparate regions from across the UK. Obtaining geographically distant samples increases the likelihood of identifying introductions from multiple different donor populations. This sampling strategy also facilitated our goal of identifying selective sweeps shared across the UK *M. guttatus* populations. Population-specific selective sweeps caused by local adaptation to narrow geographic and ecological niches in the UK are not detected in our analyses and would require multiple individuals from the specific population of interest. Previous molecular analyses of *M. guttatus* have demonstrated that a scattered sampling design, with one individual per population, is sufficient to capture regional differentiation and can avoid clustering biases resulting from sampling multiple individuals from fewer populations (Oneal *et al.* 2014). In total, we analysed genome data from 27 populations: 12 *M. guttatus* populations in the native range, 10 UK *M. guttatus* and five out-groups. From one of the native populations (Iron Mountain, IM), we sampled an additional eight individuals, which allowed us to explore the sensitivity of our findings to the particular individual sampled within a population.

Introduced populations. We sampled 10 populations of *M. guttatus* spanning the range of distribution of this species in the British Isles (Table 1; Fig. 1). The northernmost population came from the Shetland Islands (QUA, N 60.105° W 1.227°) and the southernmost from Cornwall, England (CRO, N 50.163°, W 5.293°). A population from Northern Ireland was also included (VIC, N 54.763°, W 7.454°). Non-native populations were collected from banks of canals, streams or rivers (HOU, CER, VIC, AYR, DBL, TOM and PAC), on roadside ditches (QUA), on waterlogged ground in an abandoned field (CRO) or in a bog near a small stream (TRE). A single wild-collected individual per population was randomly selected from each population for sequencing.

Native-range populations and out-groups. We obtained sequence data from the Sequence Read Archive (SRA)

Table 1 Populations sampled for genome resequencing of *Mimulus guttatus* and related taxa. A single individual was sequenced per population, except for IM, where sequence data were available for nine individuals. Life history and, for native populations, their classification as coastal or inland were provided for each population when available (Lowry *et al.* 2008; Lowry & Willis 2010)

Population	Taxon	Latitude	Longitude	Life history	Coastal/ Inland	Location
M. guttatus (native)						
PED		32.711	−110.628	Perennial	Inland	San Pedro River, Pinal Co., AZ
MED		37.829	−120.345	Annual	Inland	Moccasin, Tuolumne Co., CA
REM		38.859	−122.410	Annual	Inland	Rumsey, Yolo Co., CA
LMC		38.864	−123.084	Annual	Inland	Yorkville, Mendocino Co., CA
SWB		39.036	−123.691	Perennial	Coastal	Sperm Whale Beach, Mendocino Co., CA
BOG		41.924	−118.804	Perennial	Inland	Bog Hot Springs, Humboldt, Co., NV
MAR		43.479	−123.294	Annual	Inland	Marshanne Landing, Douglas Co., OR
DUN		43.893	−124.130	Perennial	Coastal	Dunes, Lane Co., OR
IM		44.401	−122.151	Annual	Inland	Iron Mountain, Linn Co., OR
AHQ		44.431	−110.813	Perennial	Inland	Lonestar Basin Thermal Spring, Teton Co., WY
YJS		44.951	−114.585	Perennial	Inland	Yellowjacket creek, Lemhi Co., ID
TSG		53.419	−131.916	Perennial	Coastal	Graham Island, Haida Gwaii (Queen Charlotte Islands), British Columbia, Canada
M. guttatus (introduced)						
CRO		50.163	−5.293	Perennial	—	Crowan, Cornwall
TRE		50.498	−4.465	Perennial	—	Tremar Coombe, Cornwall, England
HOU		51.097	−1.508	Perennial	—	Houghton Lodge, Hampshire, England
CER		53.006	−3.549	Perennial	—	Cerrigydrudion, Denbigshire, Wales
VIC		54.763	−7.454	Perennial	—	Victoria Bridge, Northern Ireland
AYR		55.461	−4.625	Perennial	—	Ayr, Ayrshire, Scotland
DBL		56.197	−3.965	Perennial	—	Dunblane, Perthshire, Scotland
TOM		57.255	−3.368	Perennial	—	Tomintoul, Moray, Scotland
PAC		57.355	−3.336	Perennial	—	Packhorse Bridge, Speyside, Scotland
QUA		60.105	−1.227	Perennial	—	Quarff, Shetland Islands
Out-groups						
SF	*Mimulus nasutus*	45.635	−120.914	Annual	Inland	Sherars Falls, Wasco Co., OR
MCN	*Mimulus cupriphilus*	37.912	−120.724	Annual	Inland	McNulty Mine, Calaveras, Co., CA
CVP	*Mimulus platycalyx*	38.372	−123.055	Annual	Inland/ Coastal	Coleman Valley Road, Sonoma Co., CA
EBR	*Mimulus micranthus*	39.631	−123.532	Annual	Inland	Branscomb, Mendocino Co., CA
DENT	*Mimulus dentilobus*	NA	NA	—	—	NA

(http://www.ncbi.nlm.nih.gov/sra) from 12 native populations of *M. guttatus* and five out-groups within section Simiolus: *Mimulus nasutus* (SF), *Mimulus cupriphilus* (MCN), *Mimulus platycalyx* (CVP), *Mimulus micranthus* (EBR) and *Mimulus dentilobus* (DENT). The 12 native populations of *M. guttatus* covered a linear transect of ~2800 km from Haida Gwaii (Queen Charlotte Islands), British Columbia (TSG, N 53.419°, W 131.916°) to Arizona (PED, N 32.711°, W 110.628°; Table 1). Populations in the native range occurred in a diversity of wet habitats including river and stream banks, seeps, beach dunes, bogs and springs. A single individual represented all but one of the native populations. In the case of the Iron Mountain (IM) population, we were able to obtain data for nine separate individuals. Mean coverage per genotyped base per individual ranged between 4 and 29× with an average of 10×.

DNA isolation and sequencing

We collected leaf tissue of British *M. guttatus* individuals (one per population) in the field and preserved it in resealable plastic bags with self-indicating silica gel (Fisher Scientific, Loughborough, UK) at room temperature. This dry tissue was used for DNA extraction using the Leaf MasterPure total DNA extraction kit (Cambio Ltd, Cambridge, UK). DNA libraries were created and barcoded using the Nextera DNA sample preparation kit (Illumina, San Diego, CA, USA), which uses a transposon-based method to randomly tag DNA for multiplexed sequencing. After library construction, an Agilent Bioanalyzer (Santa Clara, CA, USA) was used to measure length distribution of library, and a fluorometer (Qubit 2.0; Life Technologies, Paisley, UK) was used to measure concentration. Equimolar quantities of

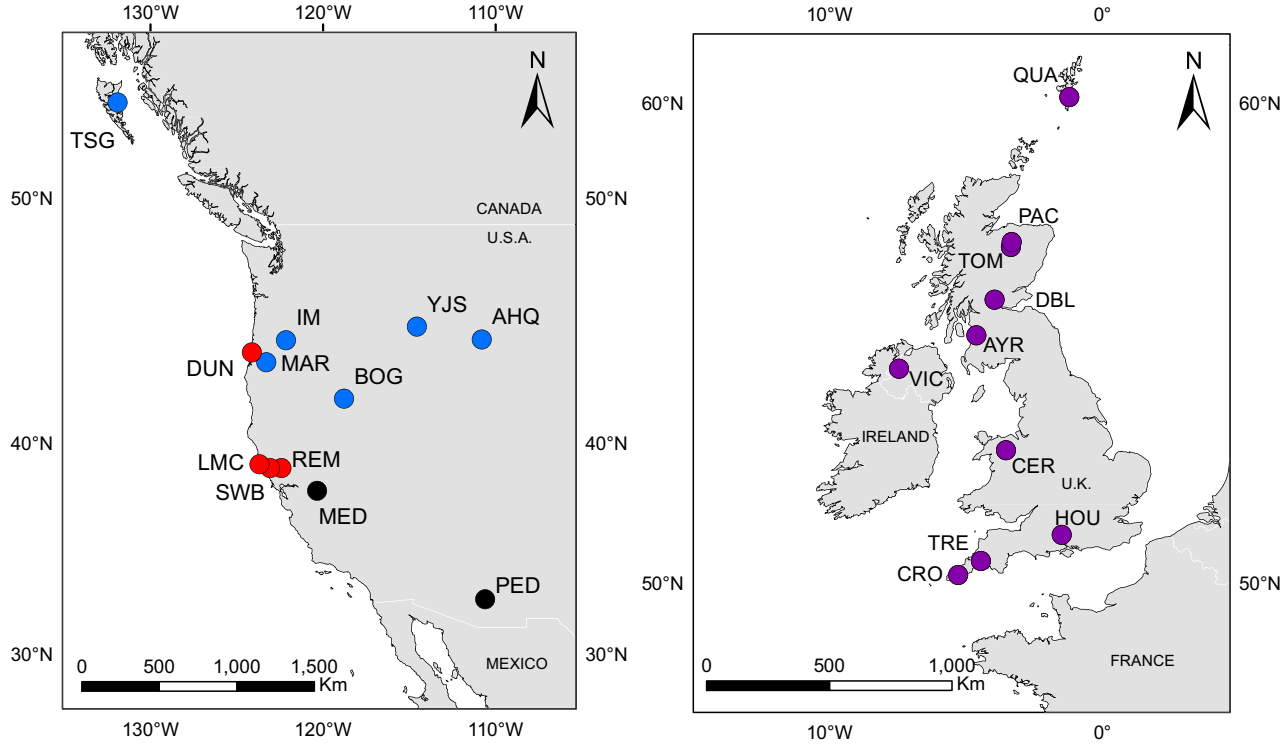

Fig. 1 Location of the 22 *Mimulus guttatus* populations sampled in the native range in North America (left-hand side panel), and in the introduced range in the UK (right-hand-side panel). Notice the different scales in the two maps. The colour of the symbols corresponds to the clades shown in the neighbour-joining tree in Fig. 2.

each library were pooled and sequenced in an Illumina HiSeq 2500 rapid-run producing 150-base pair paired-end reads. Overall, we obtained raw coverage of 1.5–11× per individual with an average of 5.7×. Raw sequence data for UK *Mimulus* samples are deposited in the JGI SRA (SRA accession numbers are given in the Data Accessibility Section).

Sequence data analysis

Genome alignment and SNP genotyping. Raw reads were aligned to the *M. guttatus* v2.0 genome available from PHYTOZOME (http://www.phytozome.net) using BOWTIE2 (Langmead & Salzberg 2014) using fast-local searches, allowing soft-clipping of poorly mapped read ends. After alignment, *Picard tools* (http://picard.sourceforge.net) was used to remove duplicates, add read groups, and verify that all mate information was accurate. After processing in *Picard tools*, the *Genome Analysis Toolkit* (GATK, DePristo *et al.* 2011) was used to call genotypes using the 'Unified Genotyper'. Minimum alignment quality was 25, and base quality was 25. Called genotypes were filtered to include genotypes with a call quality threshold of Q30 or greater. Insertions, deletions and heterozygous sites were not included in subsequent analyses. Detailed command-line methods can be found in the Supporting information. After all filtering, mean coverage per genotyped base per individual for the 10 UK samples ranged from 1.7 to 5.8× with an average of 3.5×. Of the 293 Mb located on the main 14 genomic scaffolds (representing 14 linkage groups), after all filtering, 71 Mb were genotyped in at least one of the UK individuals. A total of 18.3, 18.5 and 8.9 Mb were genotyped in 8, 9 and 10 of the UK samples, respectively (Fig. S7, Supporting information).

Measures of genetic diversity. Nucleotide diversity at silent and nonsilent sites was calculated using software described in Zhang *et al.* (2006). Briefly, genomes for all sequenced lines were recalled using the genotype data. Missing data were not imputed. Measurements of pairwise synonymous (π_{syn}) and nonsynonymous nucleotide diversity ($\pi_{non\text{-}syn}$) were calculated through pairwise comparison of coding sequences. Coding sequences were extracted from the recalled genomes using the gff3 gene annotation file available on phytozome.net. A Fisher *P*-value associated with each diversity value and indicates the confidence of that particular value. We only considered π_{syn} and $\pi_{non\text{-}syn}$ values for genes with a Fisher *P*-value ≤ 0.001 and alignment length >200 bases. In addition to calculating nucleotide diversity at synonymous and

nonsynonymous sites, whole-genome alignments were used to calculate genome-wide nucleotide diversity (π) in sliding windows using VariScan (Hutter *et al.* 2006). Windows of 50 000 genotyped bases and overlapping steps of 1000 bases were used.

Genetic relationships between introduced and native Mimulus guttatus. To determine the genetic relationships between introduced and native populations, we conducted an analysis of genetic similarity using a random subset of 1 400 000 SNPs. To create this data set, we randomly selected 100 000 SNPs for each of the 14 major linkage groups (chromosomes) that were genotyped in at least 30 individuals (of 35). Our SNP data set is therefore not subject to ascertainment bias arising from selecting, for example, only coding or noncoding SNPs (Garvin *et al.* 2010). Instead, the SNP data set analysed here should represent a snapshot of the total genetic diversity of each sample and be shaped by both neutral and non-neutral processes (Helyar *et al.* 2011). Within each linkage group, neighbouring SNPs were separated by 209 bp on average (209 ± 3.34; mean ± SE). Each SNP was coded as '0' if it matched the reference allele, and '1' for the alternative allele. In this analysis, we included all native and introduced individuals and the five out-groups (n = 35 individuals). Multiple individuals from IM were included as a reference of the variation seen within a single population.

We constructed a genetic distance matrix using *p*-distance (the proportion of nucleotide sites that differ between a pair of sequences) from the binary SNP data using the package *ape* (Paradis *et al.* 2004) in R ver. 3.0.3 (R Development Core Team 2014). The combined distance matrix was then used to estimate the relationships between all samples using a neighbour-joining (NJ) analysis in *ape*. Support for nodes in the NJ tree was calculated using 100 bootstrap replicates. Trees were drawn using FigTree v. 1.4.0 (Rambaut 2014). The NJ distance-based approach used here is appropriate for genome-wide analyses (e.g. Brandvain *et al.* 2014), as maximum-likelihood and Bayesian phylogenetic methods depend on specifying a mutational model, which is not practical for genome-wide data. We also conducted a principal component analysis (PCA) using the function *glPca* in adegenet (Jombart & Ahmed 2011). This analysis provides an independent estimate of the relationships between native and introduced populations and can be used to compare with the results of the NJ analysis. For the PCA, we selected only one individual for each of the 12 native and 10 introduced populations of *M. guttatus*. The identity of the particular individual chosen from the IM population had no qualitative effect on the relationships inferred from the PCA (Fig. S8, Supporting information), and similarly, randomly choosing one IM individual instead of nine for the NJ analysis did not change the tree topology (data not shown).

Selective-sweep analysis. Regions in the genome showing the signature of selective sweeps were detected using the parametric approach described in Nielsen *et al.* (2005) and implemented in the program SweepFinder. This method compares the observed site frequency spectrum within local regions in the genome (windows) against the background site frequency spectrum seen across the entire genome (or linkage group) and calculates statistical departures from this background expectation using a CLR. Importantly, the null hypothesis employed by this method is derived from the background data itself and does not depend on specific population genetic models or assumptions about demographic equilibrium (Nielsen *et al.* 2005), which are unlikely to hold in recently introduced populations. SweepFinder is robust to models that include population growth with recombination (Nielsen *et al.* 2005).

One potential issue with SweepFinder is that it is sensitive to SNP density (Nielsen *et al.* 2005). To account for both shared ancestral sweeps and artefacts due to genotype density, we independently analysed the North American (NA) and UK data using the exact same criteria. Ten samples from the North American populations (AHQ, BOG, DUN, IM, LMC, MAR, PED, SWB, TSG and YJS) were chosen based on the results of the genome-wide relationship analysis. Next, independently, for both the NA and UK data sets, we determined whether a given site was polymorphic and asked how many individuals were genotyped at that particular site. Using this information, we choose sites that were genotyped in at least eight individuals (of the 10 total) in both the UK and NA samples and were polymorphic in at least one of these populations. Thus, we ended up with a data set that included the exact same number of sites from the exact same genomic locations. Next, we calculated genome-wide site frequency spectrum (SFS) for each data set (UK and NA). Each chromosome was divided into 5000 bins, and the SFS within each bin was compared to the genome-wide SFS to look for signals of a selective sweep using the parametric approach described (Nielsen *et al.* 2005) and implemented in SweepFinder.

After running SweepFinder, we independently plotted the genome-wide CLR distribution for UK and NA samples. Within the NA samples, all genomic locations with CLR scores above the median value were marked for masking. Liberal masking based on the NA analyses removes hard sweeps that occurred in the last common ancestor and removes artefacts due to variable genotype density. The NA SweepFinder results were used to mask

the UK genome SweepFinder results. Only genomic positions that survived masking were considered in further analyses. Within the UK, the top 1% CLR outliers were identified as regions that have possibly experienced a hard sweep and subjected to further investigation. Gene coordinates are available in the gff3 gene annotation on phytozome, and genes with positions at least partially overlapping the swept regions were extracted for further analyses.

Results

Nucleotide diversity in Mimulus guttatus

Overall genome-wide nucleotide diversity in the UK was $\pi = 0.015$ (Fig. S1, Supporting information). Figures S2 and S3 (Supporting information) show patterns of nucleotide diversity across the genome for both native and introduced populations. For genes, within UK samples, diversity at silent sites was $\pi_{syn} = 0.0325$ while nonsynonymous diversity was $\pi_{non\text{-}syn} = 0.0035$ (Fig. S4, Supporting information). Within the native North American populations, nucleotide diversity was calculated through comparisons of 10 individuals (same 10 individuals used as the NA samples in sweep analyses). Genome-wide diversity within the NA samples is $\pi = 0.031$ (Fig. S1, Supporting information). Synonymous diversity within NA is $\pi_{syn} = 0.0610$, while nonsynonymous diversity is $\pi_{non\text{-}syn} = 0.0075$ (Fig. S4, Supporting information). Comparing NA and UK nucleotide diversity indicates an overall reduction in the introduced populations of ~50%.

Genetic relationships between native and introduced populations

The relationships between 22 native and introduced populations of *Mimulus guttatus*, and five out-groups based on the genetic distance of 1 400 000 SNPs distributed across the genome are shown in Fig. 2. All the population-level nodes in this NJ tree had a bootstrap support of 100%. The NJ tree shows that all 10 UK populations form a single well-supported clade (Fig. 2). The UK clade is most closely related to the native TSG population, a coastal perennial *M. guttatus* from Graham Island (Queen Charlotte Islands) in British Columbia (Lowry & Willis 2010). The UK and TSG samples form part of a clade of populations located north of the N 40° parallel, and which includes other inland perennial (BOG and YJS), and annual populations (AHQ, IM and MAR) (Figs 1 and 2; Table 1). The NJ tree shows a second clade composed mainly of more southern populations, and which includes inland plants (LMC, REM), coastal perennials (SWB, DUN) and two annual out-groups (CVP, *Mimulus platycalyx*; EBR, *Mimulus micranthus*). The DUN population is the only one in this group located north of the N 40° parallel (Fig. 1). Finally, the NJ tree shows a third clade, including three out-groups (MCN, *Mimulus cupriphilus*; SF, *Mimulus nasutus*; and DENT, *Mimulus dentilobus*) and the two most southern populations of *M. guttatus*, MED, an annual inland population, and PED, an inland perennial from Arizona. Our results indicate that native *M. guttatus* is separated into two main clades corresponding mostly to geographic location (north and south groups), and not to different life histories (annual/perennial) or habitat types (coastal/inland), as has been recently described by Brandvain *et al.* (2014). Finally, our NJ analysis also indicates that *M. cupriphilus* and *M. platycalyx* are nested within broadly circumscribed *M. guttatus* (Fig. 2).

The results of the PCA show clear support for a close relationship between all UK samples and also indicate that the most genetically similar native population sampled here is the coastal perennial TSG (Fig. 3). The first principal component separates the north and south groups of native *M. guttatus*, with DUN partly overlapping with the north group. The second principal component separates the UK samples from most of the other northern accessions (Fig. 3). Together, the NJ and PCA show a common ancestry of British *M. guttatus* and its association with northern, native populations.

Evidence of selection in introduced populations

Our analysis of selective sweeps in native and introduced populations identified several genomic regions displaying changes in the site frequency spectrum (measured using the CLR statistic), consistent with the signatures of positive selection acting in these regions (Fig. 4; results for all linkage groups are shown in Fig. S5, Supporting information). The comparison of high CLR regions in the separate analyses conducted in NA and UK samples allowed us to detect selective sweeps shared by both native and introduced populations. By masking these high CLR sites in NA, we located genomic regions that are candidates for selective sweeps occurring after the separation of the clade leading to the UK populations, including potentially unsampled North American donor populations. Moreover, the masking also accounted for possible regions with high CLR scores that were simply due to low genotyped SNP density. Our analysis revealed selective sweeps on five of the 14 linkage groups in the UK, which are not shared with NA samples (Fig. 4, Table 1). Regions with high CLR scores showed significantly reduced overall diversity: diversity within sweeps was $\pi = 0.0076$, half that of nucleotide diversity outside sweeps $\pi = 0.0152$ ($P \leq 0.0001$; Fig. S6, Supporting information).

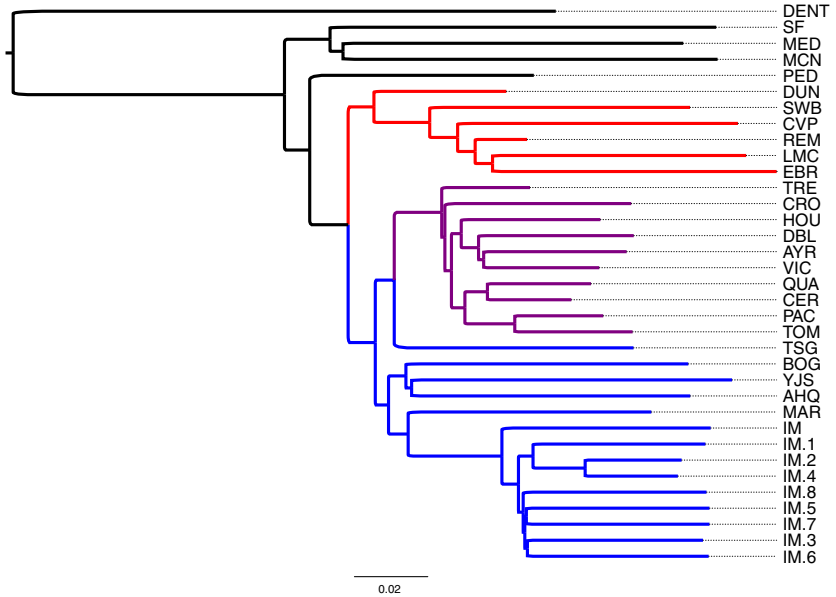

Fig. 2 Relationships between native (North America) and introduced (UK) populations of *Mimulus guttatus* inferred from 1 400 000 SNPs, sampled at a density of 100 000 SNPs per chromosome. The neighbour-joining tree was built from a matrix of pairwise genetic distance (*p*-distance) of 30 individuals of *M. guttatus* and four related taxa and rooted with *Mimulus dentilobus*. Branches leading to introduced populations are shown in purple. The two clades containing native *M. guttatus* in the north and south groups (see Results) are shown in blue and red, respectively. All nodes at the population level have a 100/100 bootstrap support. Population names as in Table 1. SNP, single-nucleotide polymorphism.

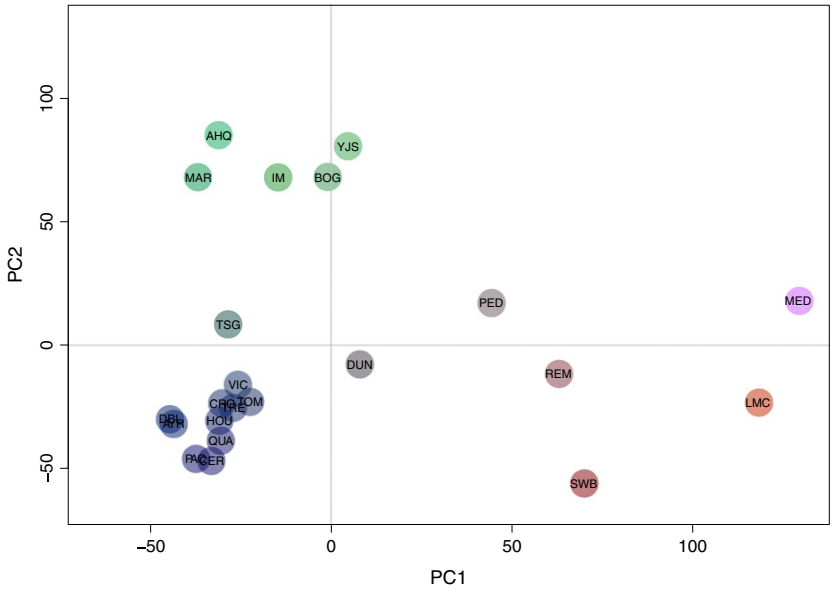

Fig. 3 Principal component analysis of 12 native (North America) and 10 introduced (UK) populations of *Mimulus guttatus*. Introduced populations cluster in the lower-left quadrat. Only one individual of the Iron Mountain (IM) population was included in this analysis. Colours denote values at first two principal component (PC) axes. Population names as in Table 1.

We identified a total of 299 genes located under the candidate region for selective sweeps within the UK clade (Table S1, Supporting information). Synonymous diversity for genes within sweeps was significantly lower than silent site diversity outside sweeps (π_{syn} = 0.0147 vs. π_{syn} = 0.0323; $P \leq 0.0001$) (Fig. S6, Supporting information). Within sweeps, 28 genes had $\pi_{syn} < 0.01$ (Table S2, Supporting information) and only

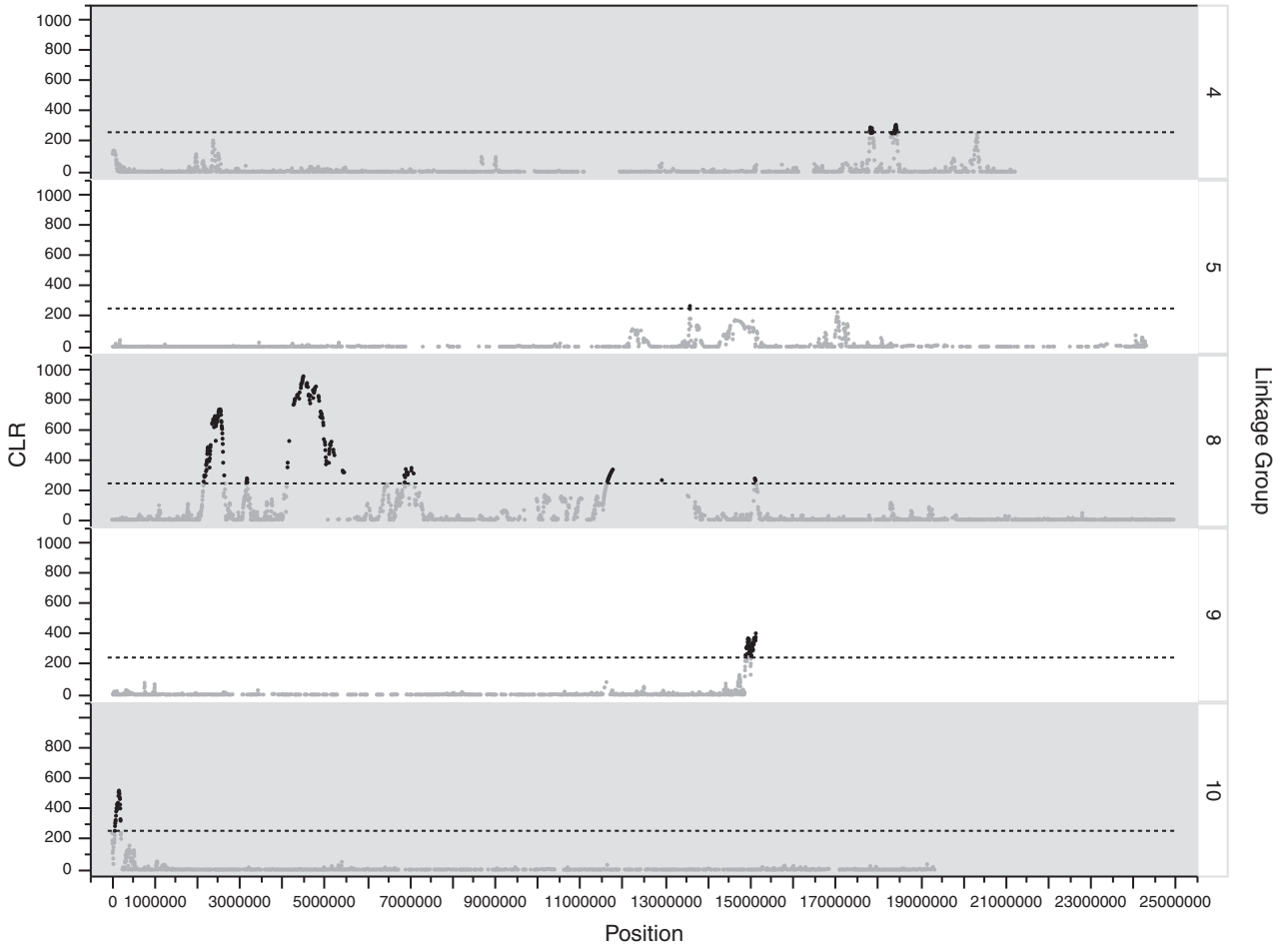

Fig. 4 Evidence of hard selective sweeps within UK *Mimulus guttatus* was found on five of the 14 linkage groups using SweepFinder (Nielsen *et al*. 2005). The figure shows selective sweeps after masking for shared sweeps between native and introduced populations (see Methods). Dashed line indicates genome-wide 1% outlier cut-off based on the composite likelihood ratio (CLR) statistic. Only linkage groups with sweeps exceeding 1% outlier cut-off are shown here; CLR profiles for all linkage groups are shown for both native and introduced populations in the Supporting information.

two genes under sweeps had silent site diversity (π_{syn}) above the genome-wide mean. Taking advantage of the annotated genome of *M. guttatus*, we recorded genes located within the swept regions, which included genes involved in flowering time, abiotic stress response including nutrient transport and tolerance to freezing, and biotic stress responses (Table S1, Supporting information).

Discussion

Our study represents the first genome resequencing study of native and introduced populations of *Mimulus guttatus*. By analysing whole-genome sequences of 35 individuals from 22 populations, we demonstrated that introduced plants in the UK are characterized by a ~50% reduction in synonymous (π_{syn}) genetic diversity and that UK populations form a single clade, relative to the North American samples included here, suggesting a common origin for non-native populations. Our analysis revealed changes in the site frequency spectrum at multiple locations across the genome, consistent with selective sweeps, some of which were restricted to UK samples. The reduction in both synonymous and nonsynonymous nucleotide diversity in genes located within sweep regions, compared to genes outside of these sweeps, was consistent with the expected loss of genetic diversity in regions linked to selected loci. Future studies are required to assess whether selection is indeed responsible for the observed patterns of variation in the sweep candidate regions, as well as to determine whether such selection acted before or after the introduction of *M. guttatus* into the UK. This study illustrates the potential of whole-genome sequencing studies to provide the initial steps for understanding the genomic consequences of invasion and rapid adaptation to new environments.

Origin and diversity of Mimulus guttatus populations in the UK

Determining the geographic origin of introduced populations provides a reference point for studies of the potential ecological and evolutionary changes occurring during the colonization and establishment phases of biological invasions (Milne & Abbott 2000). Our sample of North American populations includes a large part of the native range of *M. guttatus* (Fig. 1), but admittedly still represents a small fraction of populations from this widely distributed taxon (Grant 1924). However, we were able to include populations with different life histories, morphologies and habitat preferences (Table 1), including different ecological and morphological groups (Lowry *et al.* 2008; Lowry & Willis 2010).

Our analysis of genome-wide polymorphism revealed two main clades of *M. guttatus* in the native range, which broadly correspond to their geographic origin (north and south groups; Brandvain *et al.* 2014). An exception to the geographic arrangement of these two groups of *M. guttatus* is the DUN population. Although geographically located in the northern range (Fig. 1), DUN is nested within the southern group (Fig. 2) and may represent a secondary dispersal or reflect an introgression event between these two groups (Brandvain *et al.* 2014). Interestingly, the North and South clades of *M. guttatus* include populations with contrasting habitats, morphologies and life histories (Table 1), suggesting that taxonomic groupings based on general morphological and ecological attributes are unlikely to correspond to monophyletic clades (Nesom 2012). At the same time, the polyphyletic nature of the *M. guttatus* species complex is reflected in our results by the fact that a *M. guttatus* population (MED) is nested within out-group taxa (SF: *Mimulus nasutus*, and MCN: *Mimulus cupriphilus*; Fig. 2), while *Mimulus platycalyx* (CVP) and *Mimulus micranthus* (EBR) fall within *M. guttatus* populations. Elucidating the phylogenetic relationships within the *M. guttatus* species complex has proved challenging (Beardsley *et al.* 2003, 2004), but the use of genome-wide sequences in a phylogenetic context (Wagner *et al.* 2013) could provide a tool to establish the genetic relationships between populations of this interfertile group.

The genetic structure observed in indigenous populations allowed us to determine with confidence that introduced populations in the UK are most genetically similar to populations from the northern end of the native distribution. In particular, our results indicate that UK samples are most genetically similar to the coastal perennial population of TSG (Figs 2 and 3). However, without additional sampling, the exact source for introduced populations in the UK remains unknown. The genetic similarity within UK populations, which fall within a single clade (Fig. 2), suggests that this part of the introduced range has been established either via a single introduction event or, perhaps more likely, via multiple introductions from closely related native populations. To determine the origin and number of introductions of UK *M. guttatus* with more accuracy, it will be necessary to conduct further sampling in the Alaskan end of the distribution, as historical records point to this region as a potential source for the first specimens of this taxon in Europe (Sims 1812; Pennell 1935, p. 116).

Despite a 50% reduction in nucleotide diversity (π) in non-native *M. guttatus*, UK populations still harbour a significant amount of diversity among populations. In particular, the estimate of nucleotide diversity at synonymous sites in UK populations ($\pi_{syn} = 0.0325$) is above the median estimate of neutral nucleotide diversity for outcrossing flowering plants ($\pi = 0.0148$) and much higher than the estimate for selfing taxa ($\pi = 0.0035$) (Leffler *et al.* 2012). The diversity observed in UK populations suggests that the bottleneck experienced by *M. guttatus* during invasion was weak enough to permit the maintenance of a significant amount of nucleotide diversity. In general, introduced populations tend to show reduced genetic diversity compared with native ones, but phenomena including large initial propagule numbers, multiple introductions and admixture can result in equal or higher levels of genetic variation (Dlugosch & Parker 2008). The existence of nonsynonymous variation in UK *M. guttatus* ($\pi_{non\text{-}syn} = 0.0035$) is potentially important for naturalization and spread in the new range, as introduced populations may be able to respond to new selective pressures from standing genetic variation and are not necessarily limited by mutation rate (Prentis *et al.* 2008). Adaptation from standing genetic variation is particularly important for invasive species as it may allow responding more rapidly to novel selective pressures (Barrett & Schluter 2008).

Detecting selection in nonequilibrium populations

Our study uncovered several areas of the genome that bear the signature of positive selection. The signals left by selective sweeps include increased frequency of previously rarely derived alleles, reduction in variation in linked sites and increase in local LD (Smith & Haigh 1974; Barton 1998). However, past demographic changes can confound the pattern of a selective sweep (Barton 1998; Barton & Etheridge 2004). This is particularly the case in populations that have experienced a bottleneck and subsequent population growth with recombination (Barton & Etheridge 2004; Nielsen *et al.* 2005).

Furthermore, recombination between lineages that either escape or not the population bottleneck will result in a combination of short and long coalescent branches creating a sweep-like pattern in the SFS (Pavlidis *et al.* 2010). Thus, inferences of selection based on genome scans in nonequilibrium populations must be made with caution.

The implementation that we used to detect the signature of positive selection (SweepFinder) is relatively robust to the effects of bottlenecks and population expansion and is particularly powerful to detect recent selective sweeps (Nielsen *et al.* 2005; Pavlidis *et al.* 2010). Recent studies have shown that SweepFinder performs better than other tests, such as the ω-statistic, in nonequilibrium populations (Pavlidis *et al.* 2010). We consider that the very recent introduction of *M. guttatus* into the UK, relatively weak population bottleneck and lack of within-UK population structure make SweepFinder an appropriate method for detecting recent and strong sweeps in the introduced range. Moreover, the independently analysed NA data using the exact same SNP sampling strategy allowed us to mask shared sweeps and eliminate false positives due solely to SNP sampling and variable site density. Therefore, we consider that the genomic regions identified here are strong candidates for selective sweeps in the lineage leading to UK populations. Nevertheless, it is important to recognize that we cannot currently determine whether the selection events happened in an ancestral (unsampled) native population or after the dispersal of *M. guttatus* to the UK. Establishing the timing of the potential selective event is essential to determine whether introduced populations can exploit novel environments through (pre)adaptations brought in from the native range, or whether adaptive evolution occurs subsequent to dispersal during the establishment and spread phases of a biological invasion (Maron *et al.* 2004; Colautti *et al.* 2009).

Genes within swept regions

The potential selective sweeps we detected in five *M. guttatus* chromosomes include ~300 genes. As predicted for selective sweeps, we found reduced diversity in both coding and noncoding genic regions under these sweeps (Fig. S6, Supporting information). While it is possible to identify candidate regions for selective sweeps and to determine the genes located under these sweeps, it is not possible to know without directly testing which genes were the actual targets of selection. However, the selective sweeps identified here contain genes involved in flowering time, nutrient stress and biotic stress (Table S2, Supporting information) and could be involved in adaptation to different day lengths, soil types, novel pathogens and general responses to stress (e.g. Hodgins *et al.* 2012).

Of particular interest is the identification of selective sweeps in linkage group 8 (LG8). The selective sweeps at positions 2 Mbp (2 million base pairs) and 5 Mbp of LG8 are exceptionally wide, consistent with very strong selection pulling a large haplotype block to high frequency very quickly giving very little time for recombination to break up linked sites. These regions of LG8 are gene rich, suggesting that the width of the peak is not an artefact of reduced recombination associated with repetitive regions (Hellsten *et al.* 2013). Instead, these selective sweeps are located near or at a known inversion region of ~6 Mb in length (Oneal *et al.* 2014). This inversion (*DIV1*) is polymorphic in the native range and is associated with a number of morphological and life history differences between annual and perennial ecotypes (Lowry & Willis 2010). Therefore, this region is a good candidate for bearing adaptive variation that could be selected for in the lineage leading to the introduced populations. Mapping experiments in *M. guttatus* have demonstrated that a large region on LG8 is involved in critical photoperiod to flower (Friedman & Willis 2013). The selective sweep we identified on LG8 is located near (within 200 000 base pairs) one of the major QTLs for critical photoperiod identified in Friedman and Willis' study. Flowering time is a crucial component of fitness in seasonal environments, and therefore, it is expected to be under selection in the introduced range. Studies of invasive species have often demonstrated the potential for rapid evolution of phenology in the introduced range (Maron *et al.* 2004; Colautti *et al.* 2009). The geographic distribution of introduced populations at high latitudes (approximately between 50° N and 60° N in the UK) suggests that these populations experience day length conditions similar to the northern end of the distribution of *M. guttatus* in North America. It would be of interest to determine whether genes involved in the control of flowering under long days are under selection in northern indigenous or in introduced populations. One such candidate is *Migut.D02071*, a phytochrome-associated protein phosphatase gene located in one of the selective sweep regions in LG4 (Table S2, Supporting information). In *Arabidopsis thaliana*, a similar gene (*ATFYPP3*) participates in the regulation of flowering time in long days (www.string-db.org). Preliminary data suggest that UK populations require long days (16 h) to flower (M. Vallejo-Marín, unpublished data). Future studies could address whether flowering time has indeed played a role in the establishment of introduced populations in high latitudes.

To conclude, our results demonstrate the enormous potential for whole-genome sequencing studies to

contribute to the study of non-native populations. With a modest sequencing effort, we were able to quickly obtain previously unavailable information on the origin and diversity of introduced populations of *M. guttatus*, as well as on the genomic consequences of biological introductions, including identifying potential regions under selection. As whole-genome reference data become available for other nonmodel organisms, resequencing studies are likely to be increasingly used to study the history and consequences of biological invasions and to establish the contribution of adaptive processes to shaping the genomes of rapidly evolving populations.

Acknowledgements

We thank J. H. Willis for his support and advice, the *Mimulus* community for providing information and sharing unpublished data, and A. Twyford, J. Friedman, V. Sork and three anonymous reviewers for comments on a previous version of the manuscript. We thank the Joint Genome Institute, supported by the U.S. Department of Energy under contract no. DE-AC02-05CH11231, for providing sequence data of native accessions of *M. guttatus*. This work was supported by a US National Science Foundation—National Plant Genome Initiative Postdoctoral Fellowship to JP (NSF, 1202778), and a UK Natural Environment Research Council grant (NERC, NE/J012645/1) and a Carnegie Trust Travel grant to MVM.

References

Baker HG, Stebbins GL (eds) (1965) *The Genetics of Colonizing Species*. Academic Press, New York, NY.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Barton NH (1998) The effect of hitch-hiking on neutral genealogies. *Genetical Research*, **72**, 123–133.

Barton NH, Etheridge AM (2004) The effect of selection on genealogies. *Genetics*, **166**, 1115–1131.

Beardsley PM, Yen A, Olmstead RG (2003) AFLP phylogeny of *Mimulus* section *Erythranthe* and the evolution of hummingbird pollination. *Evolution*, **57**, 1397–1410.

Beardsley PM, Schoenig SE, Whittall JB, Olmstead RG (2004) Patterns of evolution in Western North American *Mimulus* (Phrymaceae). *American Journal of Botany*, **91**, 474–489.

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

Colautti RI, Maron JL, Barrett SCH (2009) Common garden comparisons of native and introduced plant populations: latitudinal clines can obscure evolutionary inferences. *Evolutionary Applications*, **2**, 187–199.

Cox GW (2004) *Alien Species and Evolution*. Island Press, Washington, District of Columbia.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Dlugosch KM, Lai Z, Bonin A, Hierro J, Rieseberg LH (2013) Allele identification for transcriptome-based population genomics in the invasive plant *Centaurea solstitialis*. *G3: Genes Genomes Genetics*, **3**, 359–367.

Dole JA (1992) Reproductive assurance mechanisms in 3 taxa of the *Mimulus guttatus* complex (Schrophulariaceae). *American Journal of Botany*, **79**, 650–659.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.

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

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.

Gardeners' Chronicle (1852) *The Gardeners' Chronicle and Agricultural Gazette*. Bradbury and Evans, London.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources*, **10**, 915–934.

Grant AL (1924) A monograph of the genus *Mimulus*. *Annals of the Missouri Botanical Garden*, **11**, 99–380.

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

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 Sciences of the United States of America*, **110**, 19478–19482.

Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, **11**, 123–136.

Hodgins KA, Lai Z, Kurkowski K, Huang J, Rieseberg LH (2012) The molecular basis of invasiveness: differences in gene expression of native and introduced common ragweed (*Ambrosia artemisiifolia*) in stressful and benign environments. *Molecular Ecology*, **22**, 2496–2510.

Hutter S, Vilella AJ, Rozas J (2006) Genome-wide DNA polymorphism analyses using *VariScan*. *BMC Bioinformatics*, **7**, 409.

Jahodová Š, Trybush S, Pyšek P, Wade M, Karp A (2007) Invasive species of *Heracleum* in Europe: an insight into genetic relationships and invasion history. *Diversity and Distributions*, **13**, 99–114.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

van Kleunen M (2007) Adaptive genetic differentiation in life-history traits between populations of *Mimulus guttatus* with annual and perennial life-cycles. *Evolutionary Ecology*, **21**, 185–199.

van Kleunen M, Fischer M (2008) Adaptive rather than non-adaptive evolution of *Mimulus guttatus* in its invasive range. *Basic and Applied Ecology*, **9**, 213–223.

Lachmuth S, Durka W, Schurr FM (2010) The making of a rapid plant invader: genetic diversity and differentiation in

the native and invaded range of *Senecio inaequidens*. *Molecular Ecology*, **19**, 3952–3967.
Langmead B, Salzberg SL (2014) Bowtie 2.2.1. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Nature Methods*, **9**, 357–359.
Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386–391.
Leffler EM, Bullaughey K, Matute DR *et al.* (2012) Revisiting and old riddle: what determines genetic diversity levels within species? *PLOS Biology*, **10**, e1001388.
Liu H, Pemberton RW (2009) Solitary invasive orchid bee outperforms co-occurring native bees to promote fruit set of an invasive *Solanum*. *Oecologia*, **159**, 515–525.
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, Rockwood RC, Willis JH (2008) 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*, **183**, 776–788.
Macnair MR, Watkins AD (1983) The fitness of the copper tolerance gene of *Mimulus guttatus* in uncontaminated soil. *New Phytologist*, **95**, 133–137.
Maron JL, Vila M, Bommarco R, Elmendorf S, Beardsley P (2004) Rapid evolution of an invasive plant. *Ecological Monographs*, **74**, 261–280.
Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, **28**, 659–669.
Milne RI, Abbott RJ (2000) Origin and evolution of invasive naturalized material of *Rhododendron ponticum* L. in the British Isles. *Molecular Ecology*, **9**, 541–556.
Murren CJ, Chang CC, Dudash MR (2009) Patterns of selection of two North American native and nonnative populations of monkeyflower (Phrymaceae). *New Phytologist*, **183**, 691–701.
Nesom G (2012) Taxonomy of *Erythranthe* Sect. *Simiola* (Phrymaceae) in the USA and Mexico. *Phytoneuron*, **40**, 1–123.
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.
Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. *Genome Research*, **15**, 1566–1575.
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.
Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
Parker PF (1975) *Mimulus* in Great Britain: a cytotaxonomic note. *New Phytologist*, **74**, 155–160.
Pavlidis P, Jensen JD, Stephan W (2010) Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*, **185**, 907–922.
Pennell FW (1935) The Scrophulariaceae of eastern temperate North America. *The Academy of Natural Sciences of Philadelphia*, **1**, 1–650.
Pennell F (1951) *Mimulus*. In: *Illustrated Flora of the Pacific States* (eds Abrams LR, Ferris RS), vol. 3, pp. 688–731. Stanford University Press, Stanford, CA.
Phillips BL, Shine R (2006) An invasive species induces rapid adaptive change in a native predator: cane toads and black snakes in Australia. *Proceedings of the Royal Society of London B: Biological Sciences*, **273**, 1545–1550.
Pimentel D (ed.) (2002) *Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal, and Microbe Species*. CRC Press, Boca Raton, Florida.
Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, **13**, 288–294.
Prentis PJ, Woolfit M, Thomas-Hall SR *et al.* (2010) Massively parallel sequencing and analysis of expressed sequence tags in a successful invasive plant. *Annals of Botany*, **106**, 1009–1017.
Preston CD, Pearman DA, Dines TD (eds) (2002) *New Atlas of the British and Irish Flora*. Oxford University Press, Oxford.
R Development Core Team (2014) *R. A Language and Environment for Statistical Computing, Ver. 3.0.3*. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org/.
Rambaut A (2014) *FigTree version 1.4.0*. Available from http://tree.bio.ed.ac.uk/software/figtree/ (accessed 12 March 2014).
Ricciardi A, Hoopes MF, Marchetti MP, Lockwood JL (2013) Progress toward understanding the ecological impacts of nonnative species. *Ecological Monographs*, **83**, 263–282.
Ritland K (1990) Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution*, **44**, 1230–1241.
Roberts RH (1964) *Mimulus* hybrids in Britain. *Watsonia*, **6**, 70–75.
Roda F, Ambrose L, Walter GM *et al.* (2013) Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology*, **22**, 2941–2952.
Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.
Sax DF, Stachowicz JJ, Brown JH *et al.* (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology & Evolution*, **22**, 465–471.
Sims J (1812) *Curtis's Botanical Magazine. Flower-Garden Displayed*, vol. 35. Sherwood, Neeley & Jones, London.
Siol M, Wright SI, Barrett SCH (2010) The population genomics of plant adaptation. *New Phytologist*, **188**, 313–332.
Smith JM, Haigh J (1974) The hitchhiking effect of a favorable gene. *Genetical Research*, **23**, 23–35.
Stace C (2010) *New Flora of the British Isles*, 3rd edn. Cambridge University Press, Cambridge.
Sweigart AL, Martin NH, Willis JH (2008) Patterns of nucleotide variation and reproductive isolation between a *Mimulus* allotetraploid and its progenitor species. *Molecular Ecology*, **17**, 2089–2100.
Tarnowska K, Daguin-Thiebaut C, Pain-Devin S, Viard F (2013) Nuclear and mitochondrial genetic variability of an old invader, *Dreissena polymorpha* (Bivalvia), in French river basins. *Biological Invasions*, **15**, 2547–2561.

Tokarska-Guzik B, Dajdok Z (2010) *NOBANIS. Invasive alien species fact sheet:* Mimulus guttatus. Online Database of the European Network on Invasive Alien Species—NOBANIS. Available from www.nobanis.org (accessed 1 March 2014).

Truscott AM, Soulsby C, Palmer SCF, Newell L, Hulme PE (2006) The dispersal characteristics of the invasive plant *Mimulus guttatus* and the ecological significance of increased occurrence of high-flow events. *Journal of Ecology*, **94**, 1080–1091.

Twyford AD, Ennos RA (2012) Next-generation sequencing as a tool for plant ecology and evolution. *Plant Ecology & Diversity*, **5**, 411–413.

Vallejo-Marín M, Lye GC (2013) Hybridisation and genetic diversity in introduced *Mimulus* (Phrymaceae). *Heredity*, **110**, 111–122.

Vickery RK (1959) Barriers to gene exchange within *Mimulus guttatus* (Scrophulariaceae). *Evolution*, **13**, 300–310.

Vila M, Espinar JL, Hejda M *et al.* (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters*, **14**, 702–708.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Williams F, Eschen R, Harris A *et al.* (2010) *The economic cost of invasive non-native species on Great Britain*. CABI report, 198 pp.

Wu CA, Lowry DB, Cooley AM, Wright KM, Lee YW, Willis JH (2008) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity*, **100**, 220–230.

Zhang Z, Li J, Zhao X, Wang J, Wong GK, Yu J (2006) KaKs_calculator: calculating *Ka* and *Ks* through model selection and model averaging. *Genomics, Proteomics & Bioinformatics*, **4**, 259–263.

M.V.M. and J.P. designed the research. M.V.M. made the field surveys and collected the samples. J.P. and M.V.M. organized laboratory work. J.P. generated the genomic data and conducted the bioinformatic analyses. J.P. and M.V.M. analysed the genomic data. M.V.M. and J.P. wrote the manuscript.

Data accessibility

DNA sequences: NCBI SRA (individual population codes in parentheses): SRR1462346 (10-AYR-10), SRR1475232 (10-CER-10), SRR1475385 (10-DBL-20), SRR1481643 (10-HOU-17), SRR1481644 (10-QUA-47), SRR1482404 (10-TOM-23), SRR1482405 (12-CRO-5), SRR1482406 (12-PAC-39), SRR1482407 (12-VIC-18), SRR1482409 (12-TRE-17).

SNP Genotype data: Dryad doi:10.5061/dryad.3gp32.

Supporting information

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

Appendix S1 Bioinformatics. Commands used for sequence analysis, quality control, and genotyping.

Fig. S1 Distribution of genome-wide pairwise nucleotide diversity (i.e. the average number of nucleotide differences per site between two sequences, π) in native and introduced populations of the yellow monkeyflower, *Mimulus guttatus*.

Fig. S2 Genome-wide pattern of pairwise nucleotide diversity (π) in 10 introduced populations of *Mimulus guttatus* in the United Kingdom.

Fig. S3 Genome-wide of pairwise nucleotide diversity (π) in 10 native populations of *Mimulus guttatus* in North America (AHQ, BOG, DUN, IM, LMC, MAR, PED, SWB, TSG, YJS).

Fig. S4 Distribution of nonsynonymous ($\pi_{non\text{-}syn}$) and synonymous (π_{syn}) pairwise nucleotide diversity in 10 introduced (left-hand side panels; United Kingdom), and in 10 native populations (right-hand side panels; North America) of *Mimulus guttatus*.

Fig. S5 Selective sweeps identified in introduced populations (United Kingdom) of the yellow monkeyflower, *Mimulus guttatus*, across the 14 major linkage groups (scaffolds) in this species.

Fig. S6 Evidence for reduced pairwise nucleotide diversity (π and π_{syn}) in genes located within (sweep) vs. outside selective sweeps (no sweep) of introduced populations of *Mimulus guttatus* in the United Kingdom.

Fig. S7 Genome-wide distribution of genotyped base pairs per individual across 10 individuals from 10 introduced populations of *Mimulus guttatus* in the United Kingdom (UK).

Fig. S8 Principal component analysis (PCA) of 30 individuals of *Mimulus guttatus* from both native and introduced ranges, including multiple individuals of the Iron Mountain (IM) population.

Table S1 Annotated list of 299 genes located within the selective sweeps identified in introduced populations of the yellow monkeyflower, *Mimulus guttatus*, in the United Kingdom.

Table S2 Annotated list of a subset of 28 genes within the selective sweeps described in Table S1 (Supporting information) that display reduced synonymous diversity ($\pi_{syn} < 0.01$).