

# The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone *Mimulus guttatus*

MARGARET F. HENDRICK,\*† FINDLEY R. FINSETH,\* MINNA E. MATHIASSEN,‡  
KRISTEN A. PALMER,§ EMMA M. BRODER,¶ PETER BREIGENZER\* and LILA FISHMAN\*

\*Division of Biological Sciences, University of Montana, 32 Campus Dr., Missoula, MT 59812, USA, †Department of Earth and Environment, Boston University, 685 Commonwealth Ave., Boston, MA 02215, USA, ‡School of Biology and Ecology, University of Maine, 5751 Murray Hall, Orono, ME 04469, USA, §Department of Biology, Wheaton College, 26 E. Main St., Norton, MA 02766, USA, ¶Biology Department, Wesleyan University, 45 Wyllys Ave., Middletown, CT 06259, USA

## Abstract

Microgeographic adaptation provides a particularly interesting context for understanding the genetic basis of phenotypic divergence and may also present unique empirical challenges. In particular, plant adaptation to extreme soil mosaics may generate barriers to gene flow or shifts in mating system that confound simple genomic scans for adaptive loci. Here, we combine three approaches – quantitative trait locus (QTL) mapping of candidate intervals in controlled crosses, population resequencing (PoolSeq) and analyses of wild recombinant individuals – to investigate one trait associated with *Mimulus guttatus* (yellow monkeyflower) adaptation to geothermal soils in Yellowstone National Park. We mapped a major QTL causing dense leaf trichomes in thermally adapted plants to a <50-kb region of linkage Group 14 (Tr14) previously implicated in trichome divergence between independent *M. guttatus* populations. A PoolSeq scan of Tr14 region revealed a cluster of six genes, coincident with the inferred QTL peak, with high allele frequency differences sufficient to explain observed phenotypic differentiation. One of these, the R2R3 MYB transcription factor Migut.N02661, is a plausible functional candidate and was also strongly associated ( $r^2 = 0.27$ ) with trichome phenotype in analyses of wild-collected admixed individuals. Although functional analyses will be necessary to definitively link molecular variants in Tr14 with trichome divergence, our analyses are a major step in that direction. They point to a simple, and parallel, genetic basis for one axis of *Mimulus guttatus* adaptation to an extreme habitat, suggest a broadly conserved genetic basis for trichome variation across flowering plants and pave the way for further investigations of this challenging case of microgeographic incipient speciation.

**Keywords:** adaptation, admixture, population genomics, quantitative trait locus mapping

Received 12 January 2016; revision received 15 June 2016; accepted 22 June 2016

## Introduction

Diverse taxa exhibit microgeographic adaptation (Richardson *et al.* 2014), but flowering plants may be

particularly prone to differentiate adaptively over spatial scales within their normal dispersal distance. Because plants are rooted, their ecology and evolution are tightly tied to local soil conditions, which may exert local selection on a wide range of traits. In some cases, divergent selection across soil mosaics occurs in the absence of barriers to gene flow and maintains intrapopulation polymorphism (as in the classic case of

Correspondence: Lila Fishman, Fax: 406 243 4184;

E-mail: lila.fishman@mso.umt.edu

The first two authors contributed equally to the manuscript.

*Linanthus parryae*; Epling & Dobzhansky 1942; Schemske & Bierzychudek 2007). However, microgeographic adaptation may sometimes generate barriers to gene flow that further accentuate opportunities for both neutral and adaptive divergence (Levin 2009). For example, shifts in both mating system and flowering phenology generate partial reproductive isolation between sweet grass (*Anthoxanthum odoratum*) populations adapted to mine tailings and adjacent nonmine populations (Antonovics 1967, 2006; Caisse & Antonovics 1978). Microgeographic adaptation and associated phenological shifts may even lead to sympatric speciation as has been proposed for *Howea* palms and other island taxa (Savolainen *et al.* 2006; Papadopulos *et al.* 2011). Closely related plant species are often defined by distinct edaphic requirements/tolerances, and adaptation and speciation across microgeographic soil mosaics may be an important generator of flowering plant diversity (e.g. Macnair 1989; Levin 1993; Sambatti & Rice 2006; Anacker & Strauss 2014; Ferris *et al.* 2014). However, while microgeographic differentiation is increasingly appreciated as a contributor to patterns of trait and fitness variation (Anderson *et al.* 2015; Kubota *et al.* 2015), it has been relatively unexplored from an evolutionary genetic perspective.

In particular, it is not yet clear whether microgeographic adaptation differs in its genetic architecture from adaptive divergence in other spatial contexts. For example, are alleles of major effect, which may (arguably) have larger selection coefficients, commonly involved in microgeographic divergence in the face of gene flow? A first step in addressing this question, and the host of other interesting questions about the mechanistic basis of adaptation in any spatial context, is the identification of loci underlying traits diverging across microgeographic ecological gradients (Renaut *et al.* 2012; Arnegard *et al.* 2014; Erickson *et al.* 2016). Recently, the wide accessibility of genomewide scans of sequence variation has made it possible to identify genetic regions associated with traits of interest (genomewide association mapping [GWAS]) or under putative divergent selection (Beaumont 2005) directly in wild populations (reviewed in Hunter *et al.* 2013; Savolainen *et al.* 2013; Zuellig *et al.* 2014). These population genomic approaches, which may be particularly powerful in cases of microgeographic adaptation in the face of gene flow, have led to a remarkable expansion of adaptation genomics research. However, both GWAS approaches and genome scans for adaptive variants require low levels of neutral differentiation between populations (in the case of outlier approaches) and/or repeated parallel adaptation in order to distinguish signal from background differentiation (Tiffin & Ross-Ibarra 2014). These conditions may be met in some cases of microgeographic adaptation,

allowing identification of individual loci and genomic regions associated with habitat-specific phenotypes (Turner *et al.* 2010; Andrew *et al.* 2012; Andrew & Rieseberg 2013). However, especially in self-compatible plant species, shifts in mating system associated with local adaptation may result in rapid genomewide differentiation by drift (De Mita *et al.* 2013; Hough *et al.* 2013; Wright *et al.* 2013). Similarly, shifts in phenology leading to reduced pollen flow (Levin 2006), combined with local seed dispersal in many plant taxa, may generate reproductive isolation and high levels of linkage disequilibrium among traits and loci. In such cases, distinguishing signals of adaptation or phenotypic association from background noise may be very difficult (Le Corre & Kremer 2012; Crawford & Nielsen 2013).

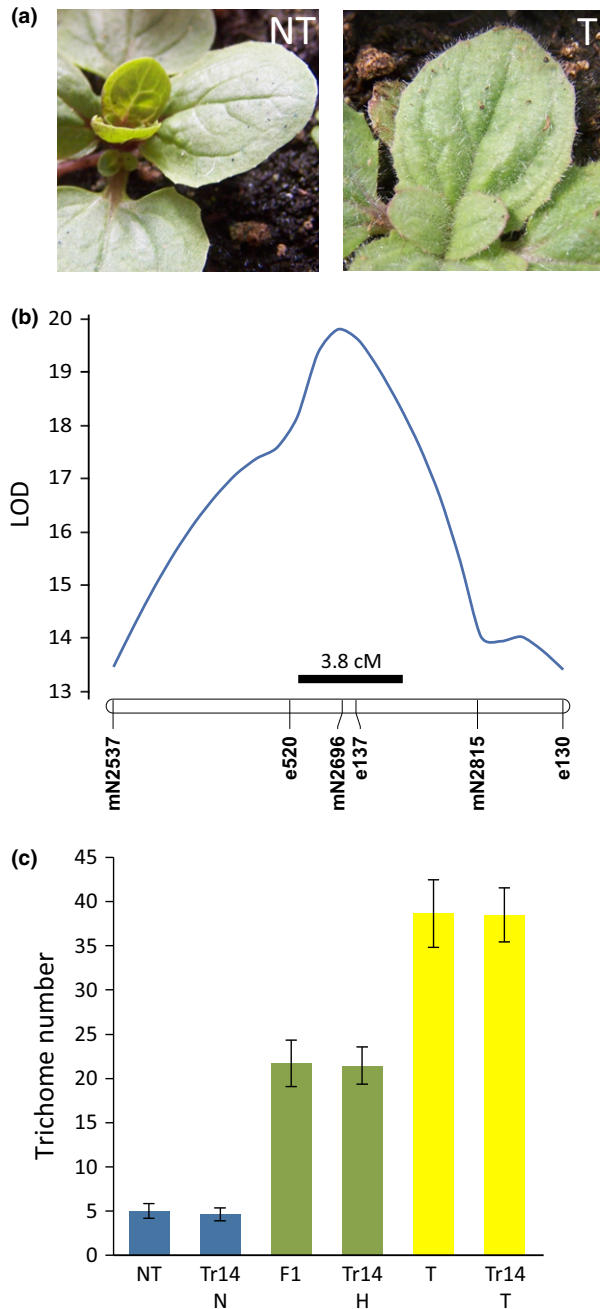
When adaptation is accompanied by strong nonrandom mating, a multipronged approach that capitalizes on both natural and experimental recombination may be necessary to determine the genetic basis of divergent traits and assess how phenotypically associated alleles sort on the landscape. Historically, quantitative trait locus (QTL) mapping in experimental hybrids between divergent populations or species has been the primary approach to characterizing loci underlying adaptive divergence, followed by fine mapping and positional cloning to identify causal sequence polymorphisms for major QTLs. Unlike genome scan approaches based only on natural variation, controlled mapping crosses also generate recombinant genotypes that may be selected against in the wild, thus allowing characterization of complex genetic architectures (Arnegard *et al.* 2014). However, such mapping approaches are limited in resolution by the amount of recombination achievable in a few experimental generations and also run the risk of characterizing major mutations present in parental lines not actually relevant to trait divergence in the wild (Rockman 2011). Thus, ground-truthing such experimental approaches with genome scans for differentiation and surveys of genotype–phenotype association in the wild may help refine QTL positions and can also test the relevance of the QTL region to patterns of natural variation. The combination of QTL and population genomic approaches has illuminated classic cases of adaptation in animals (e.g., Hoekstra *et al.* 2006; Rogers & Bernatchez 2007; Jones *et al.* 2012; Bradic *et al.* 2013), and integrated approaches to the genetic of adaptation are generally considered advantageous in any system (Stinchcombe & Hoekstra 2008; Pardo-Diaz *et al.* 2015). However, outside of crop systems, few studies of plant adaptation (especially over microgeographic scales) have combined controlled mapping approaches with population genomics to address questions about the genetic basis of adaptation (reviewed in Strasburg *et al.* 2011; but see Chapman *et al.* 2016).

Here, we investigate the genetic basis of trichome density in *Mimulus guttatus* (common yellow monkeyflower) populations adapted to geothermally heated soil mosaics in Yellowstone National Park (Lekberg *et al.* 2012). The *M. guttatus* species complex is a diverse but interfertile group of wildflowers found in moist-soil habitats throughout western North America. In Yellowstone, *M. guttatus* is common in cool bogs and is also one of a handful of flowering plant species to thrive in the extremely hot and nutrient-poor soils around geothermal vents. *Mimulus guttatus* in nonthermal bog sites belong to a widespread perennial ecotype defined by a chromosomal inversion on linkage Group 8 (LG8) (Lowry & Willis 2010). Yellow monkeyflowers from thermal vent sites are genetically and ecologically annual (Lekberg *et al.* 2012), but multiple lines of evidence (Lekberg *et al.* 2012; Puzey & Vallejo Marín 2014) suggest that thermally adapted *M. guttatus* are locally derived from nearby (perennial) nonthermal populations. Thermal *M. guttatus* thrive in their extreme habitat by taking advantage of winter–spring snowmelt on geothermally heated soils, which creates a narrow window of temperate and moist conditions for plant growth. In greenhouse common gardens, plants from thermal (T) populations remain differentiated from nearby nonthermal (NT) populations for a large suite of putatively adaptive traits (Lekberg *et al.* 2012). These include dwarfism, flowering under short days (12 vs. >14 h days), loss of stolons/increased allocation to flowers, small flowers and efficient self-fertilization, and dense leaf hairs and anthocyanin spot on early leaves. Thus, they provide an excellent opportunity to investigate the genetics of novel and parallel trait evolution across an extreme microgeographic gradient.

Our investigations of the genetic basis of thermal *M. guttatus* adaptation focus on the high elevation Agrostis Headquarters site, where extreme thermal (AHQT) and nonthermal (AHQNT) populations are highly phenotypically and phenologically distinct (Lekberg *et al.* 2012). Although only ~500 m apart, the AHQT and AHQNT populations exhibit high  $F_{ST}$  at polymorphic markers (0.34; Lekberg *et al.* 2012) and also in genomewide comparisons using PoolSeq data (genomewide average = 0.19; see Results). High differentiation probably reflects high selfing rates in the AHQT population, strong (genetically based) divergence in flowering phenology and strong selection against migrants. Genomewide differentiation is useful for marker-based QTL mapping, but confounds population genomic scans for adaptive outliers. However, population scans, along with admixture mapping where highly divergent populations come into contact, may provide complementary information to evaluate potential candidate loci within well-defined QTLs. Therefore, we

utilize a combination of QTL and fine mapping, population PoolSeq, and admixture analyses to infer the genetic basis of one dramatic trait difference between thermal and nonthermal populations of *M. guttatus*, the production of dense glandular trichomes on the cotyledons and young leaves of AHQT plants (Fig. 1; Lekberg *et al.* 2012).

Trichomes are multifunctional, important for many aspects of plant biology, and highly variable within and among plant species (Hauser 2014). Trichomes play numerous adaptive roles, including defence against biotic attack (e.g. insect herbivores and fungal pathogens) and mitigation of abiotic stress (e.g. UV radiation, cold, drought). Because trichomes (especially glandular trichomes) appear to be costly in the absence of herbivores or other selective agents (Hare *et al.* 2003; Kärkkäinen *et al.* 2004; Sletvold *et al.* 2010; Steets *et al.* 2010), they often vary genetically among populations. In *Mimulus guttatus*, both biotic and abiotic factors are likely to be important in structuring trichome variation, as population mean trichome density is correlated with latitude (Kooyers *et al.* 2015) and trichome production increases with simulated herbivory in some populations (Holeski 2007). To date, the only investigation of the genetics of trichome variation in *Mimulus guttatus* have focused on leaf hairs as an antiherbivore defence. Holeski *et al.* (2010) mapped two major QTLs (on LG10 and LG14) underlying divergence in constitutive trichome density between a glabrous annual line (IM, Oregon Cascades) and a relatively hairy perennial line (PR; Northern California). Because PR is from a mild and humid coastal site, whereas the annual IM is from a drier and colder high elevation site, Holeski (2007) inferred that the dense trichomes of PR plants were likely adaptations to herbivory. In contrast, dense glandular trichomes on the cotyledons and early (<5th pair) leaves of AHQT (Lekberg *et al.* 2012) are unlikely to reflect adaptation to herbivory, as we have seen no evidence of insect damage to leaves in numerous observations of AHQT plots (M. Hendrick and L. Fishman, personal observation). AHQT plants germinate in October/November and senesce by mid-June when snowfall ends; subzero ambient temperatures during their growing season probably preclude insect herbivores as well as pollinators. Dense AHQT trichomes are thus a plausible adaptation to abiotic factors associated with their unusual phenology. In particular, glandular trichomes have been associated with increased cold tolerance (Agrawal *et al.* 2004) and decreased leaf wettability (Brewer *et al.* 1991). Hairy AHQT *M. guttatus* leaves exhibit lower leaf wettability than glabrous AHQNT leaves, as indicated by relatively low angles of tangent between standardized water droplets and the leaf surface (E. M. Broder and L. Fishman, unpublished). Low



**Fig. 1** A major QTL on linkage Group 14 (Tr14) is sufficient to explain divergence in rosette leaf trichome density between geothermally adapted (AHQT) and nearby nonthermal (AHQNT) *Mimulus guttatus*. (a) Parental phenotypes. (b) Log-odds score (LOD) trace indicating strength of genotype–phenotype association across the QTL region of LG14. All values are highly significant. The bar indicates the location of Tr14 (1.5 LOD drop around peak). (c) Phenotypic values (means  $\pm 1$  SE) for AHQNT parent (1 generation inbred), F<sub>1</sub> hybrids and AHQT parent (1 generation inbred) control plants (shown as NT, F<sub>1</sub> and T, respectively) and for each F<sub>2</sub> genotype at Tr14 peak marker (e137).

leaf wettability is thought to be important for maintaining photosynthetic gas exchange under high precipitation conditions, such as the >5 m of snow characteristic of the AHQT winter–spring growing season. Thus, the evolutionary transition from glabrous (AHQNT) to hairy (AHQT) in Yellowstone *M. guttatus* is very likely to be ecologically (as well as evolutionarily) independent from divergence in trichome density between widespread annuals and perennials elsewhere in the species range. This provides an ideal framework both for investigating the genetics of microgeographic adaptation and for understanding the role of genetic constraint vs. environmental context in shaping the genetic architecture of trait differentiation.

Specifically, we use mapping approaches to ask if the genetic basis of trichome differentiation between *M. guttatus* adapted to geothermally heated soils and their nonthermal putative ancestors is (i) under the control of a major locus and (ii) parallel at the QTL level with other shifts in leaf hairiness within *Mimulus*. Although no one study can answer the very big question of what structures the genetic architecture of adaptation, this work is an important puzzle piece in the context of other studies of trichome variation and ongoing studies of other traits in the same system. In addition, by integrating fine mapping in a controlled cross with scans of genomic differentiation, we are able to identify a small number of positional candidate genes, including one functional candidate that also exhibits a strong genotype–phenotype association with the trait in wild admixed individuals.

## Materials and methods

### System

The yellow monkeyflower, *Mimulus guttatus* (Phrymaceae), is widespread across western North America, occurring commonly in habitats with at least ephemerally soggy soil (seeps, creeks, bogs, snowfields, etc.). The primarily outcrossing *M. guttatus* is the central member of a species complex of interfertile but ecologically and evolutionarily distinct taxa (Brandvain *et al.* 2014), and is also highly diverse within (Mojica *et al.* 2012; Flagel *et al.* 2014) and among (Oneal *et al.* 2014; Kooyers *et al.* 2015; Twyford & Friedman 2015) populations. This diversity, as well as experimental tractability, has made *M. guttatus* a model system for understanding the genetics of adaptation and speciation (Wu *et al.* 2007; Twyford *et al.* 2015). A reference genome ([www.Phytozome.jgi.doe.gov](http://www.Phytozome.jgi.doe.gov); Hellsten *et al.* 2013), as well as linkage maps anchored to the reference by

gene-based markers (Fishman *et al.* 2014b), facilitates the identification and characterization of genes and genomic regions underlying adaptation.

In Yellowstone, *M. guttatus* occurs in a range of habitats, from cool (nonthermal) bogs along riverbanks to the fringes of hot springs to dry thermal vents made hospitable only by winter snowmelt. At the Agrostis Headquarters site (Lekberg *et al.* 2012), the focal extreme thermal population (AHQT; census size >10 000) occurs on a large open hillside covered with dry thermal vents. Soils in this area reach >50 °C during the growing season and plants at the site flower from March through May–June (with the timing of senescence depending on the snowfall regime in a given year). The nonthermal population nearby (AHQNT, census size in 100s) is ~500 m downhill, in a bog with little thermal influence. AHQNT plants emerge from rhizomes after snow melts (May at earliest) and flower in July–September. Additional *M. guttatus* populations occur in smaller thermal soil patches in the canyon separating AHQT and AHQNT.

#### Marker genotyping

DNA for genetic mapping and population genetics was extracted using a standard CTAB-chloroform protocol (Doyle & Doyle 1987) modified for use in 96-well format. All markers (both pre-existing MgSTS markers and newly designed ones; Table S1, Supporting Information) were exon-primed intron-containing length polymorphisms amplified using a standard touchdown PCR protocol). They were sized on ABI 3100 or ABI 3700 automated sequencers with an in-lane size standard. Markers were visualized with 5' fluorescent (6-FAM or HEX) labels, either by direct labelling the forward (F) primer or by M13 tailing of the F-primer plus addition of a labelled M13 primer in the amplification reaction. Genotypes were scored in GENEMAPPER software (Applied Biosystems, Waltham, MA) with visual verification and double-checking of all putative crossovers in mapping populations.

#### Targeted QTL mapping

As an initial screen for major QTLs underlying trichome differences, we created an outbred F<sub>2</sub> mapping population by intercrossing two F<sub>1</sub> hybrids between wild-derived AHQT and AHQNT individuals. The F<sub>2</sub>s ( $N = 198$ ) were grown in a greenhouse common garden (University of Montana, 12-h light–12-h dark, January–April 2011) along with AHQNT and AHQT parental individuals ( $N = 40$  and  $29$ , respectively) and F<sub>1</sub> hybrids ( $N = 50$ ). All plants were phenotyped for multiple traits, including trichome number (counted along the

leaf edge from petiole to tip) and density (number standardized by the length of the leaf) on one of the second pair of rosette leaves. We initially screened three genomic regions containing trichome QTLs in an independent *M. guttatus* QTL study (Holeski *et al.* 2010) using the single MgSTS markers (www.mimulusevolution.org) at the QTL peaks in that study and then additional flanking markers for a major QTL region on LG14. We created a linkage map of the region using the regression mapping algorithm and Kosambi map function in Joinmap 4.0 (Kyazma, Wageningen, Netherlands) and localized the QTL using interval mapping in WINDOWS QTL CARTOGRAPHER 2.5 (Statistical Genetics, University of North Carolina, Raleigh, NC).

#### Fine mapping

For finer mapping, we created a new F<sub>2</sub> ( $N = 222$ ) by intercrossing 3<sup>rd</sup>-generation inbred lines from each population (AHQT 1.1 and AHQNT 1.6, respectively) and selfing a single F<sub>1</sub> hybrid. These F<sub>2</sub> hybrids were grown in a common garden in the UM greenhouse under 16-h days with supplemental lighting and genotyped at QTL-flanking markers m18080 and mN2537. Recombinant individuals ( $N = 48$ ) were then phenotyped by counting trichomes (as above) on one leaf from the second pair and genotyped at additional markers designed across the interval (Table S1, Supporting Information). Because only a targeted subset of this fine-mapping population was phenotyped, we could not estimate the proportion of the F<sub>2</sub> variance explained by our peak markers in this experiment. However, we also present the full phenotypic distribution and Tr14 association from a 3<sup>rd</sup> (not fine mappable) F<sub>2</sub> population ( $N = 336$ ) grown under 16-h days and similar greenhouse conditions and phenotyped using the same methods.

#### Pooled resequencing (PoolSeq) of AHQT and AHQNT populations

To identify nucleotide sites and genes highly differentiated between AHQT and AHQNT populations, we generated population pools consisting of many independent plants sampled from each population. For each population, we extracted genomic DNA from AHQT and AHQNT seedlings (one per wild-collected maternal family;  $N = 200$  and  $100$ , respectively) in pools of 20 plants, quantified each genomic subpool using a fluorometer (Hoescht 33258 dye, Turner Biosystems) and then combined those subpools in equimolar amounts. A genomic DNA library (Illumina TruSeq, following kit protocols) was constructed from each population pool, and each PoolSeq library was sequenced on one lane of Illumina HiSeq 2500 (paired end, 100-bp reads).

To trim adaptors and remove low-quality sequences, we applied TRIMMOMATIC version 0.30 to the raw sequence data (Bolger *et al.* 2014). The sequences were mapped to the hardmasked *M. guttatus* v2.0 genome (<http://phytozome.jgi.doe.gov/>) with the aln algorithm of BWA version 0.6.2 with default parameters (Li & Durbin 2009). We fixed mate information and removed duplicates using PICARD tools version 1.104 (<http://picard.sourceforge.net>). From the resultant bam files, we removed sequences with mapping quality lower than 29 and filtered out improperly paired, unpaired and unmapped sequences with SAMTOOLS version 0.1.19 (Li *et al.* 2009). Variant sites (SNPs) were called using the UNIFIEDGENOTYPYER tool from the Genome Analysis Toolkit version 2.8 with default parameters (McKenna *et al.* 2010; DePristo *et al.* 2011), and SNP coverages per population were tabulated with VCFTOOLS version 0.1.11 (Danecek *et al.* 2011).  $F_{ST}$  per gene and Fisher's exact test to assess differentiation per SNP were calculated genomewide with POPOOLATION2 version 1.201 (Kofler *et al.* 2011) using SNPs with coverage of 30 $\times$ –500 $\times$  and diploid pool sizes set to 400 and 200 (AHQT and AHQNT, respectively). We then calculated the difference in allele frequencies between population pools for SNPs with coverage >40 $\times$  and <200 $\times$ , focusing on the Tr14 interval.

#### Analyses of natural recombinants

To characterize the phenotypic/genotypic variation at AHQT and AHQNT extremes as well as spatially intermediate sites, we established 16 1  $\times$  1 m plots in areas across AHQ where *M. guttatus* was abundant. Five of these plots were within the highly thermal AHQT site, four in the AHQNT bog and seven in the moderately thermal Canyon area that separates the extremes. The Canyon plots are also in geothermally heated soils (although generally less hot and later to dry completely than AHQT; M. F. Hendrick and L. Fishman, unpublished), and *M. guttatus* there have a life history generally like that of AHQT plants (over-winter growth, spring reproduction, and obligate annuality). However, the lower Canyon sites are within 150 m of AHQNT and plants at these more moderately thermal sites overlap in flowering phenology with AHQNT plants in at least some years (M. F. Hendrick and L. Fishman, personal observation), generating opportunities for ongoing pollen flow.

To determine the opportunities for admixture in intermediate sites, we collected leaf tissue from up to five wild plants within each plot (spring–summer 2010), extracted genomic DNA as above and genotyped plants at seven unlinked PCR-based markers known to be polymorphic in Yellowstone *M. guttatus* (Lekberg *et al.*

2012). We used STRUCTURE (Pritchard *et al.* 2000) to assign individuals to clusters (proportionally) based on their multilocus genotypes, following Lekberg *et al.* (2012). Because our primary goal was to assess whether there was evidence of admixture in Canyon thermal individuals collected between the highly distinct AHQT and AHQNT extremes (Lekberg *et al.* 2012 and Results), we present data here from runs of STRUCTURE at  $k = 2$  clusters.

To identify and characterize wild-derived individuals with informative recombination events for Tr14, we grew plants from seeds collected in each plot (1–2 per maternal family, up to 10 maternal families per plot) in a randomized greenhouse common garden at the University of Montana and measured their trichome numbers as above. All plants were then genotyped at five markers across the Tr14 QTL region, each of which had a single allele at high frequency (82.5–100%) in AHQT plants ( $n = 63$ ) but at low frequency at AHQNT (5–33%). Genotypes were recoded as N (no AHQT allele), H (heterozygous for AHQT allele) or T (homozygous for AHQT allele). We then examined the relationship between genotype and trichome number in the subset of recombinant individuals from the variable AHQNT and Canyon populations ( $N = 53$ ) using both ANOVA (treating genotype as a categorical variable) and regression (treating genotype as a continuous variable) using JMP 12 (SAS Institute, Cary, NC).

## Results

#### Targeted QTL mapping

Consistent with previous population surveys, the outbred AHQNT and AHQT parents were strongly differentiated for trichome count and density (Fig. 1a), with  $F_1$  and  $F_2$  hybrids intermediate. In the  $F_2$  hybrids, our targeted marker on LG14, e137, was strongly associated with trichome density ( $r^2 = 0.35$ ,  $N = 162$ ,  $P < 0.0001$  for trichome density). Genotypes in the other targeted region, on LG10, exhibited no association with trichome traits (MgSTS528: ANOVA  $r^2 = 0.0007$ ,  $P = 0.92$ ,  $N = 220$ ). To flank and localize the major LG14 QTL, we genotyped four additional markers spanning ~14 cM (1.5 Mb) centred about e137. The QTL remained centred at e137 and adjacent marker mN2696 (Fig. 1b), and its action appeared essentially additive (Fig. 1c). This major QTL (Tr14) explains 100% of the parental difference in trichome number and >83% of the parental difference in trichome density (because number and density show similar patterns, and number is more intuitive scale, we use the former for further analyses). Thus, Tr14 is an additive Mendelian locus controlling a

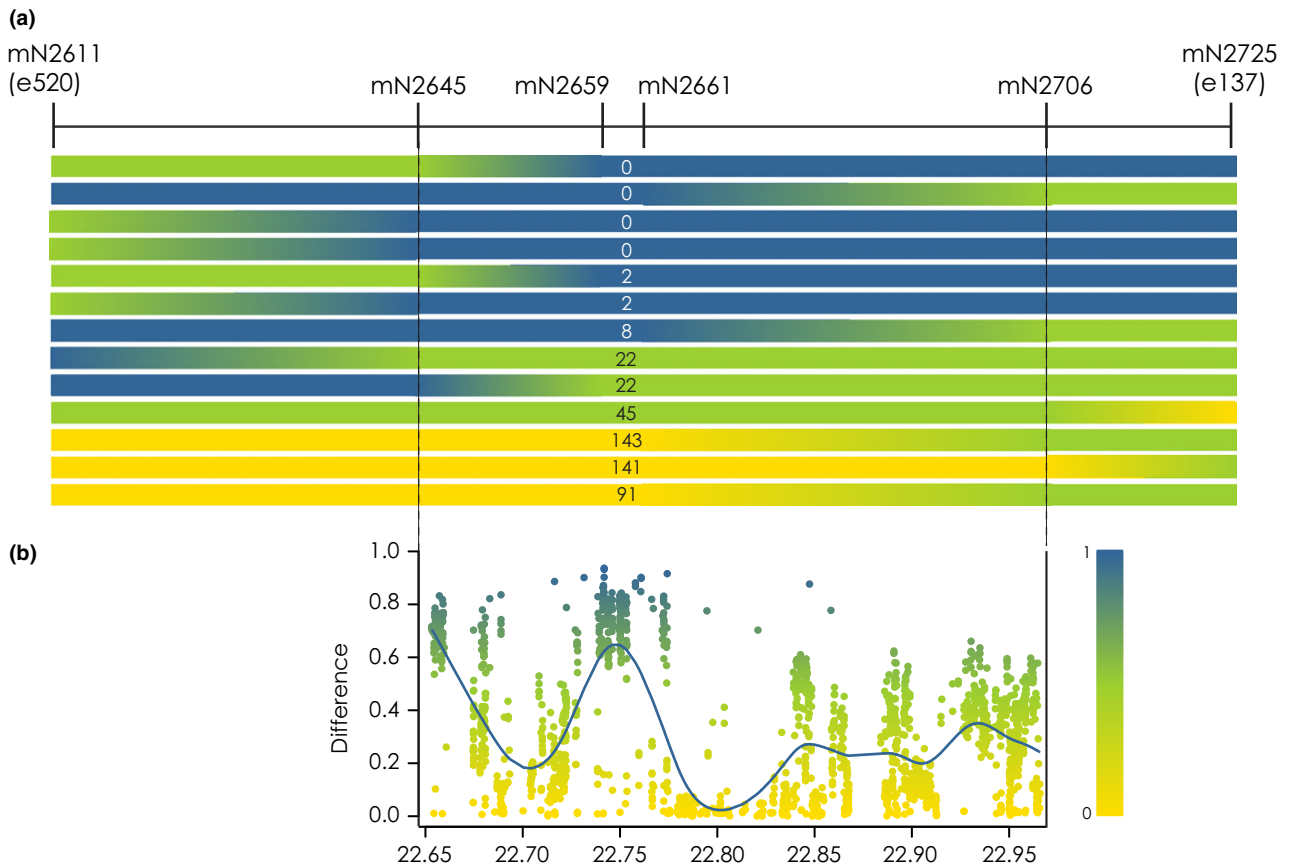
large proportion of the divergence in rosette leaf trichome production between thermal and nonthermal *M. guttatus* populations in Yellowstone.

*Fine mapping*

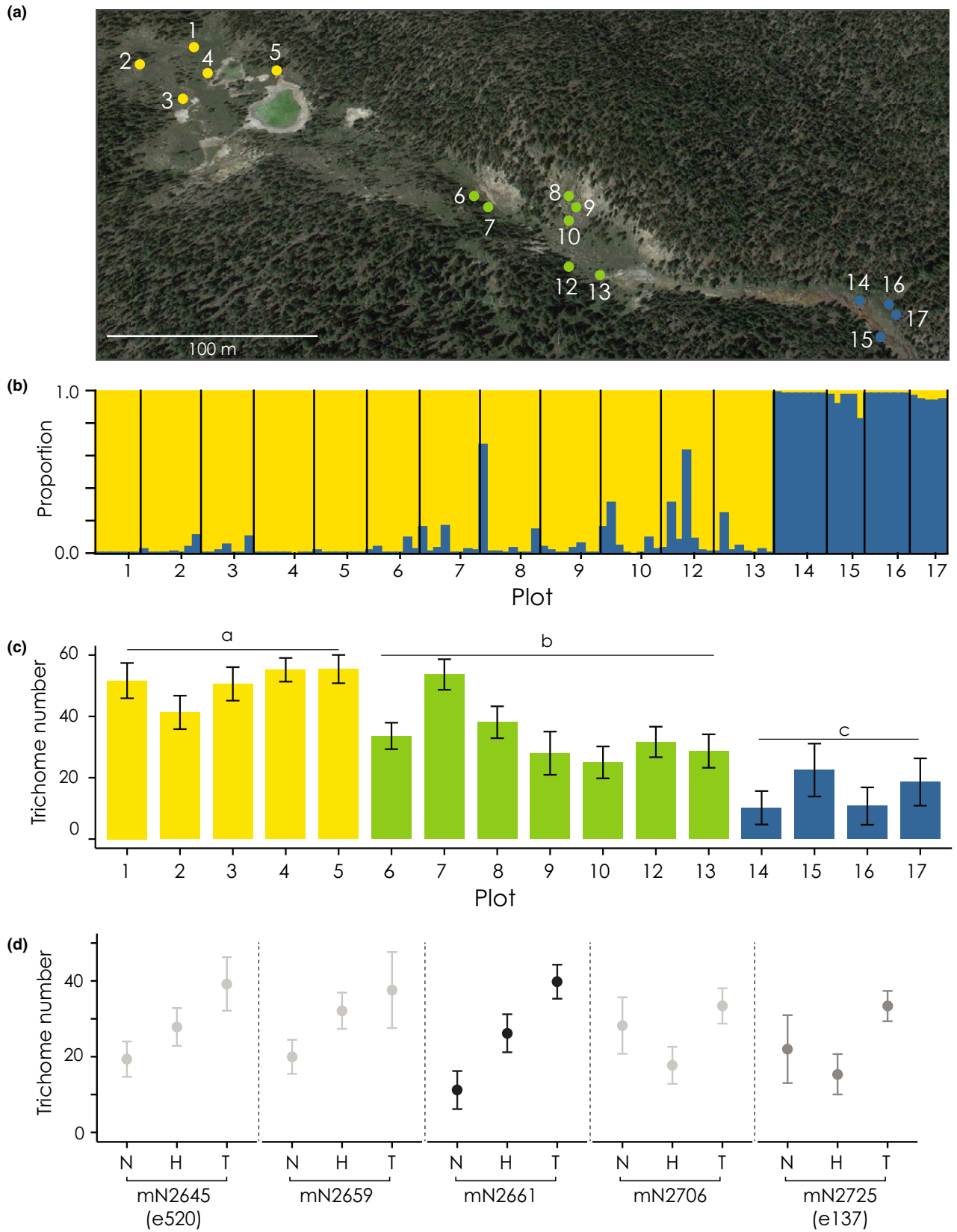
To refine the location of the Tr14 QTL, we constructed a second F<sub>2</sub> mapping population from inbred AHQT and AHQNT lines (*n* = 222). From this set, we identified 46 individuals recombinant between flanking markers mN2537 and mN2815, phenotyped them and genotyped them at new markers designed across the region. Break-point mapping narrowed the location of Tr14 to the 314-kb interval between mN2645 and mN2706, with mN2661 and mN2659 exhibiting a perfect association with trichome phenotype (Fig. 2a). A handful of recombinants remain between the flanking markers but we were not successful in designing new informative markers to further shave down the interval on either side of mN2659/mN2661. Nonetheless, the categorical

association between trichome phenotype and genotype at this pair of markers suggests that Tr14 is located near the centre of the fine-map interval rather than near either boundary.

Because we selectively phenotyped the fine-mapping population, we cannot provide an estimate of the F<sub>2</sub> variance explained by these markers in this mapping population. However, analyses of a 3rd line-cross F<sub>2</sub> data set (*N* = 336) were also consistent with Tr14 as a major QTL for trichome number (Fig. S1, Supporting Information). In this F<sub>2</sub> population (grown under similar conditions), 18% of hybrids made no trichomes at all, producing a strongly bimodal distribution, and mN2661 genotype explained more than half of the F<sub>2</sub> phenotypic variance in trichome number (*r*<sup>2</sup> = 0.52, ANOVA with 2 d.f., *F* = 158). This pattern is consistent with Tr14 acting as a major locus with the T allele *dominant* for trichome presence (Mendelian expectation: 25% of F<sub>2</sub>S = N and glabrous), plus an additional *recessive* T allele at a second unlinked locus that makes 1/4 of the



**Fig. 2** Fine mapping of Tr14 trichome QTL. (a) Bars (scaled to genetic distance) show genotypes of recombinant F<sub>2</sub> hybrids (blue/dark = NT, green/medium = H, yellow/pale = T). Numbers on bars are trichome counts. The dotted lines indicate the maximum bounds of Tr14, but the pattern of recombination indicates that it is near the centre of interval. (b) Differences in PoolSeq allele frequency across the Tr14 fine-map interval (scaled to Mb position on *M. guttatus* V2 scaffold\_14 = LG14) showing area of elevated differentiation at mN2657-mN2662. Line shows a smoothing spline with  $\lambda = 0.0014$ .





Tr14\_N plants hairy rather than glabrous (25–6.25% = 18.75%). However, additional mapping will be necessary to test this hypothesis.

#### *Population differentiation across the Tr14 QTL region and identification of candidate genes*

Our high-coverage Pool-Seq data set exhibited the high genomewide differentiation expected from previous marker based estimates; across 543 675 snps from 24 258 genes on 14 chromosomes, mean  $F_{ST}$  per gene is  $0.19 \pm 0.001$  (Fig. S2, Supporting Information) between the AHQT and AHQNT extreme populations. This is due in large part to low variation within AHQT ( $H_e = 0.17$ ) relative to AHQNT ( $H_e = 0.33$ ), consistent with the evolution of routine self-pollination (Lekberg *et al.* 2012) and perhaps bottlenecks associated with adaptation to the thermal soil environment. Although it does contain some highly differentiated sites, the Tr14 region is only one of hundreds of regions of similarly (or greater) elevated divergence across chromosome 14 (Fig. S2, S3, Supporting Information). Thus, while we can exclude sites or genes as candidate causes of trichome divergence due to low differentiation, we cannot use high differentiation alone as a guide to the loci of adaptation in this system.

To explain the strong differentiation between AHQT (all plants hairy) and AHQNT (70–100% of AHQNT completely glabrous, depending on survey year), the underlying genetic variant must also be near fixation at AHQT and at <20% at AHQNT. Both sliding window and raw differences in allele frequency identify a <36-kb region (containing the six genes Migut.N02657 – Migut.N02662) of elevated AHQT–AHQNT differentiation coincident with the inferred fine-map location of Tr14 (Fig. 2b). This region contains 45 of the 53 snps with frequency differences  $\geq 80\%$  found across the ~314-kb fine-mapped region. The six genes in this highly differentiated cluster are annotated as 60S ribosomal protein L35 (Migut.N02657), PPR repeat//Dirigent-like protein//Rad51 (Migut.N02658), PAPS transporter 1 (Migut.N02659), DUF1635 (Migut.N02660), R2R3 MYB

transcription factor (Migut.N02661) and DUF966 (Migut.N02662).

The R2R3 MYB transcription factor in the highly differentiated region, Migut.N02661, is a strong candidate on the basis of both position and known roles for R3 and R2R3 MYBs in controlling trichome density and other epidermal traits in plants. Compared to other genes in the candidate interval, Migut.N02661 has relatively few polymorphisms and low genewise  $F_{ST}$  between AHQT and AHQNT (Fig. S2, Supporting Information). However, both an indel in the first intron (the length polymorphism used for mapping above and analyses below) and a nonsynonymous substitution (glutamine  $\Rightarrow$  lysine) in the 3rd exon exhibit high frequency differences between the AHQT and AHQNT pools (~80%). The nonreference allele at the latter is near-fixed at AHQT (88%; 37 of 42 reads) and very rare at AHQNT (10%; 6 of 62 reads) in the PoolSeq data. The nonsynonymous (NS) variant in Migut.N02661 is one of 12 highly differentiated (frequency difference in PoolSeq data  $>0.7$ ) NS variants across the six-gene focal region (Table S2, Supporting Information). The others are in Migut.N02660 ( $n = 2$ ) and Migut.N02658 ( $n = 9$ ); five nonsynonymous SNPs in the latter gene are, like the Migut.N02661 variant, also differentiated between AHQT and the glabrous line IM767 (Holeski *et al.* 2010). However, Migut.N02658 appears to be two separate genes (a RAD51 homologue and a pentatricopeptide repeat protein) mis-annotated as one, which makes evaluating its variants difficult (Table S2, Supporting Information). Neither of the annotated Migut.N02658 gene families or the DUF1665 group of Migut.N02660 has any known functional connection to epidermal traits in other taxa.

#### *Analyses of wild recombinant individuals*

Although the AHQT and AHQNT populations are highly differentiated (Lekberg *et al.* 2012), individuals from spatially and ecologically intermediate Canyon populations (Fig. 3a) show evidence of admixture between the extremes. At  $k = 2$ , STRUCTURE assigns most Canyon individuals largely to the same cluster as

**Fig. 3** Phenotypic and genetic data from naturally admixed individuals at AHQ supports the association of Tr14 with Migut.N02661. (a) Location of sample plots at AHQT (yellow; 1–5), Canyon (green; 6–13), and AHQNT (blue; 14–17). (b) Proportional assignment to AHQT and AHQNT populations (STRUCTURE analysis of seven unlinked markers;  $k = 2$ ) of individuals from plots across AHQ site (total  $N = 113$ ), illustrating admixture in Canyon population between AHQT and AHQNT extremes. (c) Trichome counts (means  $\pm 1$  SE) for each plot, from plants grown in a greenhouse common garden ( $N = 168$  total,  $n = 5$ –20 per plot). Letters indicate least-squared means (LSMs) were significantly different among the three populations (LSMs =  $50.8 \pm 2.26$ ,  $34.1 \pm 2.03$ , and  $15.5 \pm 3.5$  for AHQT, Canyon, and AHQNT, respectively; Tukey's HSD,  $P < 0.05$ ). (d) Associations of marker genotype (coded as T, H and N on the basis of near-fixed T alleles) and trichome phenotype for five markers across the Tr14 region, in admixed individuals from Canyon and AHQNT plots. mN2661 genotype was strongly associated with trichome number (black;  $r^2 = 0.27$ ,  $P < 0.0005$ ), whereas markers in flanking and similarly differentiated genes were only weakly associated (dark grey; e137/mN2725;  $r^2 = 0.13$ ,  $P = 0.03$ ; outside the fine-map interval) or not statistically associated (grey;  $P > 0.05$ ) with phenotype. Least-squared means ( $\pm 1$  SE) from ANOVA are shown.

AHQT (consistent with their thermally influenced habitat and phenology), but many individuals are partially assigned to the AHQNT cluster (Fig. 3b). Similarly, in the greenhouse common garden of wild-derived individuals, plants from Canyon plots had significantly lower mean trichome numbers and significantly higher coefficients of variation than AHQT (Fig. 3c), suggesting segregating variation for this trait. In addition, while most AHQNT individuals were entirely glabrous, a fraction of AHQNT plants in this sample were hairy, suggesting either gene flow from AHQT/Canyon or segregating ancestral variation. This is in contrast to the initial survey of phenotypic variation at AHQ, which found that all (20 of 20) AHQNT plants sampled were hairless (Lekberg *et al.* 2012).

In the set of recombinant Canyon and AHQNT individuals ( $N = 53$ ), we observed a strong association between the number of AHQT-associated alleles at mN2661 and trichome number ( $r^2 = 0.27$ ,  $P < 0.001$ , ANOVA) (Fig. 3c). Only one of the four flanking markers was significantly associated with trichome number at  $\alpha = 0.05$  (mN2725/e137:  $r^2 = 0.13$ ,  $P = 0.03$ ). Thus, although levels of admixture and gene flow are generally low between the AHQT and AHQNT extremes, there is enough recombination that variation in the strength of genotype–phenotype association can be detected. Natural recombinants point to Migut.N02661, rather than highly differentiated neighbouring markers in the fine-map interval, as most predictive of trichome number or density. Although this association does not rule out polymorphisms in other highly differentiated genes (e.g. Migut.N02658) or nongenic sequences in the region as causal, it strengthens the case for coding or regulatory variation at or very near the MYB Migut.N02661 as a major source of the striking microgeographic divergence in trichome production at AHQ.

## Discussion

### *Triangulation of a major locus underlying local adaptation*

Using a mixture of QTL, divergence scan and admixture approaches, we have identified a major locus and strong candidate gene associated with morphological adaptation to an extreme habitat. Yellow monkeyflowers adapted to life in geothermal soils in Yellowstone National Park have evolved fuzzy (trichome dense) cotyledons and rosette leaves, likely for cold tolerance or snow shedding during their unique winter growing season, whereas nearby nonthermal is generally glabrous. Genetic mapping in controlled crosses, accelerated by targeting of trichome QTLs previously identified in a distinct pair of *Mimulus* populations (Holeski *et al.* 2010)

identified a major locus sufficient to completely explain this striking divergence and fine mapping narrowed our range of inference to ~50 genes. On its own, the finding of a major parallel QTL argues for a partially shared genetic basis for ecologically and evolutionary independent transitions in leaf trichomes in *Mimulus* and also supports the emerging consensus that rapid adaptation often occurs via mutations of major effect. However, in order to fully address questions about the genetics of parallel evolution, as well as to understand the molecular mechanism and evolutionary history of trichome variation, we must work towards connecting QTL-level variation to the underlying genes.

Towards this goal (and to understand trichome variation in its natural context), we used both PoolSeq and individual genotyping to examine patterns of population genetic divergence across the Tr14 QTL region. In PoolSeq data from the AHQT and AHQNT extremes, fewer than 10% of polymorphic sites in the region (1749 SNPs with coverage  $>40\times$ ) were differentiated enough to possibly explain the absence of glabrous individuals at AHQT and the low frequency of hairy plants at AHQNT; these clustered in a set of six contiguous genes coincident with the inferred fine-map position of the Tr14 QTL (Fig. 2). These highly differentiated genes included one strong functional candidate, Migut.N02661, an R2R3 MYB with homology to the MYB48/MYB59 subfamily in *Arabidopsis thaliana*. We then took advantage of natural recombinants (Fig. 3) from AHQNT and an intermediate Canyon thermal site to further refine genotype–phenotype relationships. Consistent with a causal role for Migut.N02661, marker genotype at this locus was the most strongly predictive of phenotype, even in comparison with nearby markers mN2645 and mN2659, which are in genes more highly differentiated in the PoolSeq data. This suggests that local selection on trichome number at AHQT may have elevated differentiation across a multigene region, but that subsequent admixture at the lower end of the soil temperature gradient has been sufficient to partially break down local linkage disequilibrium. Alternatively, it is possible that the presence of hairy individuals (and associated mN2661 genotypes) at AHQNT reflects ancestral standing variation rather than secondary genetic exchange. Further mapping and functional analyses will be necessary to definitively link sequence variants at our candidate locus to differentiation in trichome number; however, our success in identifying a shared positional and functional candidate illustrates the utility of a multipronged approach to the genomics of adaptation. Such a combined strategy may be particularly useful in cases of incipient microgeographic speciation in plants, where shifts in life history and mating system can rapidly elevate differentiation genomewide.

*A discrete genetic basis for a fuzzy trait*

Trichome count, both in our hybrids and in plants in general, is a quantitative trait, with both environmental and genetic components generating near-continuous trait variation. Nonetheless, we find that the dramatic difference in leaf pubescence between AHQT geothermal *M. guttatus* (all very hairy) and nearby nonthermal AHQNT (most completely glabrous) can be largely accounted for by allelic substitution at a single genetic locus, Tr14. This suggests a simple genetic switch that regulates the presence/absence of trichomes on young leaves in Yellowstone *Mimulus guttatus*, although the intermediacy of F<sub>1</sub> hybrids and Tr14 heterozygotes also argues for an additive effect of Tr14 dosage on trichome production. Similar switches underlying trichome variation between ecotypes in *Arabidopsis*, and subsequent mapping/validation has characterized several major effect genes (reviewed in Hauser 2014). For example, a single amino acid change in the R3 MYB ETC2 was identified as the cause of a major QTL accounting for a >15-fold difference in trichome density between divergent accessions (Hilscher *et al.* 2009). Similarly, multiple loss-of-function mutations in the R2R3 MYB GL1 cause glabrous leaves in some populations, and more subtle haplotypic variation contributes to natural trichome variation rangewide (Bloomer *et al.* 2012).

In *Mimulus*, trichome variation appears to have a primarily oligogenic basis, with a few major QTLs explaining divergence in both constitutive and induced trichome phenotypes (this study, Holeski *et al.* 2010). A similarly simple genetic basis also often characterizes divergence in pigment traits in plants and animals (Hoekstra 2006; Smith & Rausher 2011), the presence/absence traits in plants (Gottlieb 1984) and flowering time in *Mimulus* (Fishman *et al.* 2014a, 2015). In contrast, divergence in traits such as flower size (Fishman *et al.* 2002, 2015) or height (Dobzhansky 1970; reviewed in Olson-Manning *et al.* 2012) often occurs via accumulation of many small adaptive changes, likely built from the quantitative standing variation within populations. It is not yet clear the relative roles of ecology (which shapes the strength of selection), development (which may impose pleiotropic constraints), and history (which influences the amount and nature of standing variation available for adaptation to new environments) play in generating these patterns. The diverse suite of characters divergent between thermal and nonthermal *M. guttatus* populations, many of which are also independently differentiated among other taxa in the complex, provides an interesting opportunity for addressing these issues.

In addition to explaining AHQT and AHQNT divergence in trichome number, our characterization of Tr14 suggests a parallel genetic basis for trichome patterning

at two scales. Within *Mimulus*, Tr14 is coincident with the largest effect QTL for 2nd-leaf trichome incidence and constitutive trichome density mapped in a distinct annual–perennial cross with opposite parental phenotypes (Holeski *et al.* 2010). Although that QTL has not been finely mapped, our successful targeting of marker e137 and the localization of our fine-map interval within the 2-Mb QTL region of Holeski *et al.* (2010) argues for a shared genetic basis. Notably, the LG14 QTL identified in that study specifically affected early trichome variation, with no effect on 5th leaf trichome incidence and effects on postdamage trichome induction opposite to the parental trait difference (presumably because dense early trichomes reduce the scope for increases in density postinduction). This suggests that any shared molecular basis for Tr14 is specific to trichome production at the early rosette stage. Our finding of substantial parallelism at the QTL level, coupled with the inference that different kinds of selection (biotic vs. abiotic) are involved in trichome divergence, suggests an important role for developmental constraint. Similarly, convergent genetic bases have been implicated many cases of parallel phenotypic divergence in diverse taxa (reviewed in Nadeau & Jiggins 2010; Pavey *et al.* 2011), although most (like ours) are coloured by the bias inherent in using QTL or molecular candidates.

At a larger scale, our identification of a MYB transcription factor as the most likely gene underlying Tr14 suggests substantial conservatism in the genetic basis of natural trichome variation. MYB transcription factors are a very large gene family in flowering plants, with important regulatory roles in plant development and responses to biotic and abiotic stress. MYB genes have up to three tandemly repeated, DNA-binding MYB domains; the two-domain R2R3 MYBs are the most diverse and often exhibit tissue-specific patterns of expression. In *Arabidopsis*, the R2R3 MYBs GLABROUS1 (GL1) and MYB66 (WEREWOLF) are central determinants of epidermal cell patterning in leaves and roots, respectively (reviewed in Ishida *et al.* 2008). GL1 is a positive regulator of trichome development, but several single-repeat R3 MYBs are negative regulators that competitively bind to GL1's targets and thus inhibit trichome formation. MYBs have been implicated in trichome development in Rosids beyond *Arabidopsis*, including cotton (where a functional homologue of GL1 contributes to seed fibre production; Wang *et al.* 2004), peach (Vendramin *et al.* 2014) and poplar (Plett *et al.* 2010). *Mimulus* makes single-celled trichomes like those of *Arabidopsis* (Scoville *et al.* 2011), but Asterids (such as *Mimulus*) have been argued to have a fundamentally different trichome development pathway not involving homologues of GL1 (Serna & Martin 2006; Yang *et al.* 2011; Yang & Ye 2013). Nonetheless, MYBs have been

implicated in Asterid epidermal cell patterning; in particular, the MIXTA-LIKE group of R2R3 MYBs is involved in conical cell and trichome formation in *Antirrhinum* (Glover *et al.* 1998). In *Mimulus*, MIXTA homologues have been linked to trans-generation responses to leaf damage including trichome induction (Scoville *et al.* 2011) and also to variation in pigmentation traits (Cooley *et al.* 2011; Lowry *et al.* 2012). The main cluster of MIXTA-LIKE loci in *M. guttatus* is on LG8 and was not associated with trichome variation in our cross (although that region does contain a major QTL for a novel anthocyanin leaf spot phenotype in AHQT plants; M. F. Hendrick and L. Fishman, unpublished).

Our candidate Migut.N02661 belongs to a relatively poorly characterized subgroup of MYBs containing *Arabidopsis* MYB48 and MYB59 (Dubos *et al.* 2010). Conserved alternative splicing (spanning monocots and dicots) in this subgroup produces up to four distinct transcripts per gene (Li *et al.* 2006), including both full R2R3 MYBs and single-repeat MYB-LIKE transcripts; these appear to be expressed in different organs and in response to different environmental stresses. At least three transcripts (two including the first intron, one not) of Migut.N02661 were annotated in the *M. guttatus* V2 build (<http://phytozome.jgi.doe.gov/>), and we have preliminarily confirmed expression of both full R2R3 and expected shorter MYB-like transcripts of Migut.N02661 in Yellowstone *M. guttatus* leaves (K. Anderson and L. Fishman, unpublished). Identification of Migut.N02661 as a strong positional candidate suggests a novel, although not entirely unexpected, role for this subgroup of MYBs in trichome formation.

#### *An integrated approach to genotype–phenotype associations*

– By combining three different approaches to the problem of linking genotype and phenotype (QTL mapping, genome scans for population differentiation and admixture analysis), we have made substantial progress towards identifying a major locus underlying morphological adaptation to an extreme habitat. None of these approaches is individually novel, and many reviews call for precisely such integration (e.g. Stinchcombe & Hoekstra 2008; Pardo-Diaz *et al.* 2015). However, it is still rare for a single study to combine these approaches as part of the identification of major genes, in part because different systems may lend themselves more to one or the other (e.g. QTL mapping is simplest in crosses between parents with many fixed differences, whereas population genomics scans for adaptive/trait-associated loci work best when genomewide differentiation is low). Nonetheless, there is much to be gained from a multipronged approach, especially when (as in our system and other incipient species) population genomics alone is unlikely to provide single-gene resolution. In addition to providing necessary

recombination, QTL studies link phenotype and genotype even when populations/species are highly differentiated genomewide and thus reduce the biases inherent in only studying the genetics of adaptation and speciation in the context of abundant gene flow. Not surprisingly, in several of the classic examples of adaptation by major genes in animals, loci were first identified using QTL mapping and then further refined (and placed in an ecological/geographical context) with population genomics of parallel populations or admixture zones (e.g. Colosimo *et al.* 2005; Reed *et al.* 2011).

In addition, even when single-gene resolution is possible with population genomics, experimental approaches can yield complementary insight. Compared to reduced representation sequencing of individuals, high-coverage PoolSeq has the advantage (at least in taxa with reference genomes and where large numbers of individuals can be pooled) of efficiently identifying the maximal set of site- or trait-associated polymorphisms (Schlötterer *et al.* 2014); this may be particularly important when low linkage disequilibrium limits signal to sites very close to a causal polymorphism (Kardos *et al.* 2016). However, especially when samples are pooled by site rather than phenotype, the connections between outlier loci and any given trait may be diffuse (Cutler & Jensen 2010; and also prone to false positives; Anderson *et al.* 2014). For example, to our knowledge, none of the loci identified as strongly (and plausibly from a functional perspective) associated with serpentine adaptation in a seminal PoolSeq genome scan (Turner *et al.* 2010) has been directly connected to physiological or fitness traits by mapping, and thus, it is not yet clear how the outliers identified in that study contribute individually or in aggregate to tolerance of serpentine soils. At least in nonwoody plants and other tractable organisms, QTL and individual-based admixture mapping in focal regions are useful complements to genome scans of populations or phenotypic pools; every method has weaknesses, and cross-validation is ideal. As next-generation sequencing approaches continue to make both QTL mapping and population genomic scans even more accessible, multiple lines of inference are likely to be the gold standard for studying the genetics of adaptation, along with functional confirmation of individual candidate mutations (Pardo-Diaz *et al.* 2015).

## Conclusions

In addition to identifying a major locus underlying phenotypic differentiation during microgeographic adaptation to a novel soil habitat, this work provides a strong foundation for further investigations. First, we have identified a positional candidate gene for trichome variation that is also a plausible functional candidate.

Because our QTL is shared with another transition in the same trait elsewhere in the species, this provides the necessary foundation for comparative work to investigate the molecular basis of parallel phenotypic divergence. Second, by combining PoolSeq and admixture analyses of wild-collected plants with QTL mapping, we strengthen the inference that divergence in early-leaf trichome density is adaptive. Although inbreeding and low levels of variation at AHQT make drift a broad-scale null hypothesis for its high differentiation from AHQNT (genomewide  $F_{ST} = 0.19$ ; and see Fig. 3b), near fixation of alternative SNPs across a six-gene region coincident with the fine-map position of Tr14 argues for divergent phenotypic selection with genomically local effects. Additional genomewide analyses and field tests of fitness in 'mismatched' individuals will be necessary to fully pin down patterns of gene flow at loci underlying phenotypic differentiation in trichome density and other traits across the AHQ site. Finally, this work illustrates how complementary approaches can illuminate microgeographic adaptation accompanied by shifts in mating system and phenology, which may be common in plants. Such cases of rapid differentiation may better represent incipient speciation than cases of recent adaptation where a few major loci define distinct phenotypes against a freely segregating background, and thus are particularly interesting to understand genetically.

### Acknowledgements

We thank Dan Crowser, Andrew Demaree, Leah Grunzke, Joey Latsha, John Mason, Angela Stathos and Jeff Van Noppen for assistance with plant care and Mike Weltzer, Natalie McLenaghan, Angela Stathos and Dan Crowser for assistance with field sample collection, Colin Callahan for laboratory assistance and University of Montana Murdock Sequencing Core and Hudson Alpha Genome Services Laboratory staff for marker genotyping and genome resequencing, respectively. We are grateful to Christie Hendrix and other Yellowstone National Park staff for research permits (YELL-2010/2011-SCI-5834) and generous assistance with field logistics. Funding support was provided by National Science Foundation (NSF) grants DEB-0846089 and DEB-1457763 to L. F. M. M., K. A. P. and E. M. B. were supported by NSF-Research Experience for Undergraduates Grants DBI-0755560 & DBI-1157101.

### References

Agrawal AA, Conner JK, Stinchcombe JR (2004) Evolution of plant resistance and tolerance to frost damage. *Ecology Letters*, **7**, 1199–1208.

Anacker BL, Strauss SY (2014) The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proceedings of the Royal Society London B*, **281**, 20132980.

Anderson EC, Skaug HJ, Barshis DJ (2014) Next-generation sequencing for molecular ecology: a caveat regarding pooled samples. *Molecular Ecology*, **23**, 502–512.

Anderson JT, Perera N, Chowdhury B, Mitchell-Olds T (2015) Microgeographic patterns of genetic divergence and adaptation across environmental gradients in *Boechera stricta* (Brassicaceae). *American Naturalist*, **186**, S60–S73.

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–2482.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, **21**, 2078–2091.

Antonovics J (1967) Evolution in closely adjacent plant populations V. Evolution of self-fertility. *Heredity*, **23**, 219–238.

Antonovics J (2006) Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity*, **97**, 33–37.

Arnegard ME, McGee MD, Matthews B *et al.* (2014) Genetics of ecological divergence during speciation. *Nature*, **511**, 307–311.

Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Bloomer RH, Juenger TE, Symonds VV (2012) Natural variation in GL1 and its effects on trichome density in *Arabidopsis thaliana*. *Molecular Ecology*, **21**, 3501–3515.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bradic M, Teotónio H, Borowsky RL (2013) The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus*. *Molecular Biology and Evolution*, **30**, mst136–2400.

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

Brewer CA, Smith WK, Vogelmann TC (1991) Functional interaction between leaf trichomes, leaf wettability and the optical properties of water droplets. *Plant, Cell and Environment*, **14**, 955–962.

Caisse M, Antonovics J (1978) Evolution in closely adjacent plant populations. IX. Evolution of reproductive isolation in clinal populations. *Heredity*, **40**, 371–384.

Chapman MA, Hiscock SJ, Filatov DA (2016) The genomic bases of morphological divergence and reproductive isolation driven by ecological speciation in *Senecio* (Asteraceae). *Journal of Evolutionary Biology*, **29**, 98–113.

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

Cooley AM, Modliszewski JL, Rommel ML, Willis JH (2011) Gene duplication in *Mimulus* underlies parallel floral evolution via independent trans-regulatory changes. *Current Biology*, **21**, 700–704.

Crawford JE, Nielsen R (2013) Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? *Molecular Ecology*, **22**, 6131–6148.

Cutler DJ, Jensen JD (2010) To pool, or not to pool? *Genetics*, **186**, 41–43.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

- De Mita S, Thuillet A-C, Gay L *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**, 1383–1399.
- 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.
- Dobzhansky T (1970) *Genetics of the Evolutionary Process*. Columbia University Press, New York, NY.
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.
- Dubos C, Stracke R, Grotewold E *et al.* (2010) MYB transcription factors in *Arabidopsis*. *Trends in Plant Science*, **15**, 573–581.
- Epling C, Dobzhansky T (1942) Genetics of natural populations. VI. Microgeographic races in *Linanthus parryae*. *Genetics*, **27**, 317–332.
- Erickson PA, Glazer AM, Killingbeck EE *et al.* (2016) Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks. *Evolution*, **70**, 887–902.
- Ferris KG, Sexton JP, Willis JH (2014) Speciation on a local geographic scale: the evolution of a rare rock outcrop specialist in *Mimulus*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **369**, 20140001.
- 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.
- Fishman L, Sweigart AL, Kenney AM, Campbell S (2014a) Major quantitative trait loci control divergence in critical photoperiod for flowering between selfing and outcrossing species of monkeyflower (*Mimulus*). *New Phytologist*, **201**, 1498–1507.
- Fishman L, Willis JH, Wu CA, Lee YW (2014b) Comparative linkage maps suggest that fission, not polyploidy, underlies near-doubling of chromosome number within monkeyflowers (*Mimulus*; Phrymaceae). *Heredity*, **112**, 562–568.
- Fishman L, Beardsley P, Stathos A, Williams CF, Hill JP (2015) The genetic architecture of traits associated with the evolution of self-pollination in *Mimulus*. *New Phytologist*, **205**, 907–917.
- Flagel LE, Willis JH, Vision TJ (2014) The standing pool of genomic structural variation in a natural population of *Mimulus guttatus*. *Genome Biology and Evolution*, **6**, 53–64.
- Glover BJ, Perez-Rodriguez M, Martin C (1998) Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. *Development*, **125**, 3497–3508.
- Gottlieb LD (1984) Genetics and morphological evolution in plants. *American Naturalist*, **123**, 681–709.
- Hare JD, Elle E, Dam NM (2003) Costs of glandular trichomes in *Datura wrightii*: a three-year study. *Evolution*, **57**, 793–805.
- Hauser M-T (2014) Molecular basis of natural variation and environmental control of trichome patterning. *Frontiers in Plant Science*, **5**, 320.
- 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 USA*, **110**, 19478–19482.
- Hilscher J, Schlötterer C, Hauser M-T (2009) A single amino acid replacement in ETC2 shapes trichome patterning in natural *Arabidopsis* populations. *Current Biology*, **19**, 1747–1751.
- Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, **97**, 222–234.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101–104.
- Holeski LM (2007) Within and between generation phenotypic plasticity in trichome density of *Mimulus guttatus*. *Journal of Evolutionary Biology*, **20**, 2092–2100.
- Holeski LM, Chase-Alone R, Kelly JK (2010) The genetics of phenotypic plasticity in plant defense: trichome production in *Mimulus guttatus*. *American Naturalist*, **175**, 391–400.
- Hough J, Williamson RJ, Wright SI (2013) Patterns of selection in plant genomes. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 31–49.
- Hunter B, Wright KM, Bombles K (2013) Short read sequencing in studies of natural variation and adaptation. *Current Opinion in Plant Biology*, **16**, 85–91.
- Ishida T, Kurata T, Okada K, Wada T (2008) A genetic regulatory network in the development of trichomes and root hairs. *Annual Review of Plant Biology*, **59**, 365–386.
- Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.
- Kardos M, Husby A, McFarlane SE, Qvarnström A, Ellegren H (2016) Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources*, **16**, 727–741.
- Kärkkäinen K, Løe G, Agren J (2004) Population structure in *Arabidopsis lyrata*: evidence for divergent selection on trichome production. *Evolution*, **58**, 2831–2836.
- Kofler R, Pandey RV, Schlotter C (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, **27**, 3435–3436.
- Kooyers NJ, Greenlee AB, Colicchio JM, Oh M, Blackman BK (2015) Replicate altitudinal clines reveal that evolutionary flexibility underlies adaptation to drought stress in annual *Mimulus guttatus*. *New Phytologist*, **206**, 152–165.
- Kubota S, Iwasaki T, Hanada K *et al.* (2015) A genome scan for genes underlying microgeographic-scale local adaptation in a wild *Arabidopsis* species. *PLoS Genetics*, **11**, e1005361.
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.
- Lekberg Y, Roskilly B, Hendrick MF *et al.* (2012) Phenotypic and genetic differentiation among yellow monkeyflower populations from thermal and non-thermal soils in Yellowstone National Park. *Oecologia*, **170**, 111–122.
- Levin DA (1993) Local speciation in plants: the rule not the exception. *Systematic Botany*, **18**, 197–208.
- Levin DA (2006) Flowering phenology in relation to adaptive radiation. *Systematic Botany*, **31**, 239–246.
- Levin DA (2009) Flowering-time plasticity facilitates niche shifts in adjacent populations. *New Phytologist*, **183**, 661–666.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Li J, Li X, Guo L *et al.* (2006) A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in *Arabidopsis* and rice. *Journal of Experimental Botany*, **57**, 1263–1273.

- Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- 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, Sheng CC, Lasky JR, Willis JH (2012) Five anthocyanin polymorphisms are associated with an R2R3-MYB cluster in *Mimulus guttatus* (Phrymaceae). *American Journal of Botany*, **99**, 82–91.
- Macnair MR (1989) The potential for rapid speciation in plants. *Genome*, **31**, 203–210.
- McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.
- Mojica JP, Lee YW, Willis JH, Kelly JK (2012) Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. *Molecular Ecology*, **21**, 3718–3728.
- Nadeau NJ, Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics*, **26**, 484–492.
- Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, **13**, 867–877.
- 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.
- Papadopulos AST, Baker WJ, Crayn D *et al.* (2011) Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences USA*, **108**, 13188–13193.
- Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, **6**, 445–464.
- Pavey SA, Bernatchez L, Aubin-Horth N, Landry CR (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution*, **26**, 298–306.
- Plett JM, Wilkins O, Campbell MM, Ralph SG, Regan S (2010) Endogenous overexpression of *Populus* MYB186 increases trichome density, improves insect pest resistance, and impacts plant growth. *The Plant Journal*, **64**, 419–432.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Puzey J, Vallejo Marín M (2014) Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Molecular Ecology*, **23**, 4472–4485.
- Reed RD, Papa R, Martin A *et al.* (2011) *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, **333**, 1137–1141.
- Renaut S, Mailet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**, 354–363.
- Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution*, **29**, 165–176.
- Rockman MV (2011) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.
- Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. *Molecular Biology and Evolution*, **24**, 1423–1438.
- Sambatti JBM, Rice KJ (2006) Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution*, **60**, 696–710.
- Savolainen V, Anstett M-C, Lexer C *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature*, **441**, 210–213.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.
- Schemske DW, Bierzychudek P (2007) Spatial differentiation for flower color in the desert annual *Linanthus parryae*: was Wright right? *Evolution*, **61**, 2528–2543.
- 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.
- Scoville AG, Barnett LL, Bodbyl Roels SA, Kelly JK, Hileman LC (2011) Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in *Mimulus guttatus*. *New Phytologist*, **191**, 251–263.
- Serna L, Martin C (2006) Trichomes: different regulatory networks lead to convergent structures. *Trends in Plant Science*, **11**, 274–280.
- Sletvold N, Huttunen P, Handley R, Kärkkäinen K, Agren J (2010) Cost of trichome production and resistance to a specialist insect herbivore in *Arabidopsis lyrata*. *Evolutionary Ecology*, **24**, 1307–1319.
- Smith SD, Rausher MD (2011) Gene loss and parallel evolution contribute to species difference in flower color. *Molecular Biology and Evolution*, **28**, 2799–2810.
- Steets JA, Takebayashi N, Byrnes JM, Wolf DE (2010) Heterogeneous selection on trichome production in Alaskan *Arabidopsis kamchatica* (Brassicaceae). *Amer. J. Bot.*, **97**, 1098–1108.
- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.
- Strasburg JL, Sherman NA, Wright KM *et al.* (2011) What can patterns of differentiation across plant genomes tell us about adaptation and speciation?. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 364–373.
- 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.
- Twyford AD, Streisfeld MA, Lowry DB, Friedman J (2015) Genomic studies on the nature of species: adaptation and speciation in *Mimulus*. *Molecular Ecology*, **24**, 2601–2609.

- Vendramin E, Pea G, Dondini L *et al.* (2014) A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. (C Peace, Ed.). *PLoS One*, **9**, e90574.
- Wang S, Wang J-W, Yu N *et al.* (2004) Control of plant trichome development by a cotton fiber MYB gene. *The Plant Cell*, **16**, 2323–2334.
- Wright SI, Kalisz S, Slotte T (2013) Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society London B*, **280**, 20130133–20130133.
- Wu CA, Lowry DB, Cooley AM *et al.* (2007) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity*, **100**, 220–230.
- Yang C, Ye Z (2013) Trichomes as models for studying plant cell differentiation. *Cellular and Molecular Life Sciences*, **70**, 1937–1948.
- Yang C, Li H, Zhang J *et al.* (2011) A regulatory gene induces trichome formation and embryo lethality in tomato. *Proceedings of the National Academy of Sciences, USA*, **108**, 11836–11841.
- Zuellig MP, Kenney AM, Sweigart AL (2014) Evolutionary genetics of plant adaptation: insights from new model systems. *Current Opinion in Plant Biology*, **18**, 44–50.

---

L.F. and M.F.H. designed and oversaw the research; L.F., M.F.H., M.E.M., K.A.P., E.M.B. and P.B. collected and analysed the data; L.F., M.F.H. and F.R.F. analysed and interpreted the data and wrote the manuscript.

---

## Data accessibility

Genetic mapping, marker population genetics and admixture phenotypic and genotypic data sets: doi:10.5061/dryad.42517. PoolSeq genomic data for AHQT and AHQNT: NCBI SRP077752.

## Supporting information

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

**Fig. S1** Distribution of rosette leaf trichome number in AHQT × AHQNT *Mimulus guttatus* AHQT × AHQNT F<sub>2</sub> hybrids greenhouse-grown under long (16 h) daylengths and spring temperature (13C/21C) conditions (N = 336).

**Fig. S2** Genome-wide distributions of (a) F<sub>ST</sub> and (b) allele frequency differences between bulk pools of individuals from AHQT and AHQNT.

**Fig. S3** A scan of SNP frequency differences across Chromosome 14 (top) and Tr14 trichome QTL region (bottom) reveals high differentiation.

**Table S1** Markers used for QTL mapping and admixture analyses.

**Table S2** Highly-differentiated (frequency difference >0.7; >× coverage in PoolSeq dataset) genic SNPs across six-gene focal region of TR14.