Genetic architecture of spatially complex color patterning in hybrid *Mimulus*

Arielle M. Cooley[1,*], Caroline Schlutius[2], Melia Matthews[1], Xingyu Zheng[2,3], Daniel Thomas[1], Patrick Edger[4], Adrian Platts[4], Logan George[2], Aaron Williams[1], Amy LaFountain[5], Douglas Hundley[6], Yao-Wu Yuan[5], Alex Twyford[7] and Joshua R. Puzey[2]

[1]Biology Dept, Whitman College, Walla Walla WA, USA, [2]Biology Dept, College of William & Mary, Williamsburg VA, USA, [3]School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA, [4]Department of Horticulture, Michigan State University, East Lansing, MI, USA, [5]Dept of Ecology & Evolutionary Biology, University of Connecticut, Storrs CT, USA, [6]Math Dept, Whitman College, Walla Walla WA, USA, [7]School of Biological Sciences, University of Edinburgh, Edinburgh, Scotland, UK, *Corresponding author: cooleya@whitman.edu, 345 Boyer Ave., Walla Walla WA 99362 U.S.A.

ABSTRACT

Interspecies hybridization generates inter-genomic interactions, which may result in unique traits not seen in either parent species. Here we explore the genetic basis of two types of floral pigment phenotypes, in hybrids between monkeyflower species *Mimulus cupreus* and *M. luteus* var. variegatus. *Mimulus cupreus* has abundant yellow carotenoid pigmentation in its petal lobes, while *M. l. variegatus* has a derived reduction in carotenoid intensity. Thus, as expected, carotenoid intensity segregates in an F2 hybrid population. More surprisingly, both species appear to have petal lobes solidly and identically covered in magenta anthocyanin pigment (leading to an orange color in the high-carotenoid species *M. cupreus*), yet F1 and F2 hybrids exhibit novel and complex spatial patterns of anthocyanin spotting. A rare yellow morph of *M. cupreus*, which lacks petal anthocyanins, also generates spatially patterned offspring when hybridized with *M. l. variegatus*. We use this cross, together with newly developed genomic and image analysis tools, to investigate the genetic architecture of color and pattern variation in an F2 hybrid population. We report that the non-patterned carotenoid reduction in *M. l. variegatus* is genetically simple, and is explained by a single QTL which contains the *Beta-carotene hydroxylase-1* (*BCH1*) gene. HPLC results show that beta-carotene content differs between dark yellow and light yellow petals, which supports a causal role for *BCH1*. The hybrid-specific anthocyanin patterning phenotypes are more complex, with one QTL of large effect and four detectable QTLs of small effect. These results illustrate how different types of traits may have predictably distinct genetic architectures, and provide candidate genomic regions for investigating the molecular mechanisms of both simple and complex floral color patterning.

KEYWORDS Keyword; Keyword2; Keyword3; ...

1 Introduction

Hybridization between species forces two divergent genomes to co-exist within the same organism. These genomic interactions can yield surprising results at both molecular and phenotypic levels, with evolutionary consequences ranging from species extinction (Levin *et al.* 1996; Epifanio and Philipp 2000) to adaptation (Anderson and Stebbins Jr 1954; Rieseberg *et al.* 2003; Suarez-Gonzalez *et al.* 2018), adaptive radiation (Seehausen 2004; Marques *et al.* 2019; Grant and Grant 2019), and hybrid speciation (Grant 1971; Rieseberg 1997; Mallet 2007). Finding the genetic architecture of hybrid-specific phenotypes is an important step towards understanding the origins and evolutionary consequences of inter-genome interactions.

Plant pigmentation has long served as a vehicle for investigating genomic, developmental, and evolutionary mechanisms (McClintock 1950; Davies *et al.* 2012; Sobel and Streisfeld 2013). A change in plant pigmentation can have profound ecological and evolutionary consequences, due to the diverse roles of pigment including pollinator attraction, herbivore deterrence, and protection against abiotic stressors such as light and extreme

Manuscript compiled: Friday 29th April, 2022
[2]Corresponding author: Whitman College Biology Dept., 345 Boyer Ave, Walla Walla WA, USA, cooleya@whitman.edu.

temperatures (Chalker-Scott 1999; Gould 2004; Demmig-Adams *et al.* 1996). These factors have the potential to impose strong purifying selection, but may also promote the maintenance of phenotypic variation (Takahashi *et al.* 2015; Sapir *et al.* 2021; Kellenberger *et al.* 2019) as well as the evolution of new pigment traits as environmental conditions change (Trunschke *et al.* 2021; Niu *et al.* 2017). Indeed, color patterning has evolved frequently and has generated spectacular amounts of phenotypic diversity, particularly in the petals of flowering plants.

In flowers, the two major classes of pigments are the yellow-to-orange carotenoids and the red-to-purple anthocyanins (Grotewold 2006). The anthocyanin biosynthetic pathway, and its regulation by R2R3 MYB, bHLH, and WDR transcription factors, is among the best understood of any pigment pathway in any kingdom (Holton and Cornish 1995; Davies *et al.* 2012; Xu *et al.* 2015; Chen *et al.* 2019).

Much of floral pigmentation research to date has focused on binary (presence-absence) traits (e.g. Zufall and Rausher (2003); Cooley *et al.* (2011)), or overall pigment intensity (e.g. Bradshaw Jr and Schemske (2003)), with the bulk of studies addressing anthocyanin rather than carotenoid pigmentation. Complex patterning traits such as speckling and spotting offer an opportunity to explore fundamental principles of evolution and development, as has been done with eyespot formation in butterflies (Beldade and Brakefield (2002)), yet such traits in plants are only beginning to be addressed (Martins *et al.* (2013, 2017); Neher and Hallatschek (2013); Yamagishi *et al.* (2014); Ding *et al.* (2020)). Unanswered questions include: To what extent is this type of variation genetic, rather than environmental or stochastic? Do complex spatial patterns reflect an underlying genetic complexity (many loci of individually small effect), or is a simple one- or two-locus system sufficient? And, in the case of hybrid-specific color patterning, what is the role of genetic divergence between the parent species in creating the overall phenotype?

Genetic mapping is an important first step in answering many of these questions. A necessary prerequisite to genetic mapping is a quantifiable phenotype. The difficulty of scoring and analyzing spatially complex variation has been a major barrier to genetic and developmental analyses of many evolutionarily fascinating traits (Houle *et al.* (2010); Minervini *et al.* (2015); Gehan and Kellogg (2017)), and a variety of algorithms are now being developed to help lessen this barrier. These include machine-learning approaches as in Boogaard *et al.* (2020) and computer vision based techniques as in Galkovskyi *et al.* (2012). In plants, phenotypic image analysis has been developed for traits including color intensity (De Keyser *et al.* 2013; Trivellini *et al.* 2014; Li *et al.* 2020), leaf shape (de Souza *et al.* 2016; Weight *et al.* 2008), root characteristics (Kimura *et al.* 1999; Nakano *et al.* 2012), and overall plant architecture (Knecht *et al.* 2016). The analysis of complex color patterning is still relatively uncommon, perhaps because of the sheer abundance of flower color traits that are readily categorized using "by-eye" methods. As interest in complex color patterning increases, the development of new tools will be essential, because techniques developed for other types of biological traits are not easily adapted to capturing variation across a heterogeneously pigmented petal surface. Examples of digital image processes developed specifically for complex color patterning in plants include an older principal-components based approach by Yoshioka *et al.* (2004) and a not-yet-published pre-print (Li *et al.* 2019).

An attractive biological system for studying the evolution of plant pigmentation and color patterning is the monkeyflower genus *Mimulus* (synonym *Erythranthe* (Barker *et al.* 2012; Lowry *et al.* 2019)). The genus features an abundance of floral diversity and an array of ecological, molecular, and genomic resources (Wu *et al.* 2008; Sobel and Streisfeld 2013; Yuan 2019). Anthocyanin and carotenoid pigmentation vary across *Mimulus*. Unfortunately, most of the complex color patterns are found in species that are rare, unstudied, or both. One exception is the *luteus* species group from Chile (Grant 1924; Watson and Von Bohlen 2000; Cooley *et al.* 2008), which combines ease of growth in the greenhouse, a solid foundation of prior pigmentation research (Medel *et al.* 2003, 2007; Cooley *et al.* 2011; Zheng *et al.* 2021), ease of hybridization (Stanton *et al.* 2016), and an intriguing hybrid-specific petal pattern phenotype. The *luteus* group consists of the broadly distributed *M. luteus* var. *luteus*, as well as *M. l. variegatus*, *M. naiandinus*, *M. cupreus*, and *M. depressus*, all of which have more limited distributions in mid- to high-elevation regions of the Andes mountains (Von Bohlen 1995).

The *luteus* group is nested within the *Simiolus* section of *Mimulus*, which is characterized almost uniformly by yellow-flowered plants displaying red dots of anthocyanin pigment in the nectar guide region of the flower (Grant 1924). Within the *luteus* group, however, flower color has evolved dramatically (Fig. 1A). Petal carotenoid pigmentation has been lost or greatly downregulated in petals of the white-and-pink flowered *M. naiandinus* and the magenta-flowered *M. l. variegatus* (Cooley and Willis 2009). The two carotenoid reductions are genetically distinct: recessive in *M. naiandinus* versus semi-dominant in *M. l. variegatus* (Fig. 1B-C: compare yellow intensity of *naiandinus* x *luteus* F1s to *variegatus* x *luteus* F1s).

Petal lobe anthocyanin pigmentation, meanwhile, has been gained in *M. naiandinus*, *M. l. variegatus*, and *M. cupreus* (Cooley and Willis 2009). In each case, petal lobe anthocyanin is dominant and segregates in a 3:1 ratio in an F2 hybrid population, but maps to either of two genomic regions - *pla1* for *M. naiandinus* and the common orange morph of *M. cupreus*, and *pla2* for *M. l. variegatus* (Cooley and Willis 2009; Cooley *et al.* 2011). A rare yellow morph of *M. cupreus*, found at only a single locality, likely represents a secondary loss of petal lobe anthocyanins, and a complementation test indicates that the causal locus is in the *pla1* region (Cooley and Willis 2009).

Hybridization reveals an unexpected property of the anthocyanin pigmentation in *M. l. variegatus* versus *M. cupreus*: hybrids are characterized by a spatially complex distribution of petal anthocyanin pigmentation not seen in either parent (Fig. 1B-D). This patchy color patterning is seen regardless of whether the cross involves the common orange morph of *M. cupreus* (high carotenoids and high anthocyanins), or a rare yellow morph (high carotenoids but no petal lobe anthocyanins) (Fig. 1B-D).

Genetic mapping (Cooley *et al.* 2011) and functional genetic experiments ((Zheng *et al.* 2021)) indicate that the R2R3 MYB transcription factor gene *MYB5a/NEGAN* is responsible for the gain of petal anthocyanin in *M. l. variegatus*, while an unlinked genomic region containing candidate genes *MYB2b* and *MYB3a* is responsible for the gain of petal anthocyanin in the orange-flowered morph of *M. cupreus* and its subsequent loss in the yellow morph. Interestingly, crosses to a third taxon (*M. l. luteus*) that lacks petal anthocyanins reveal underlying pattern differences between *M. l. variegatus* and *M. cupreus*. Using *M. l. variegatus* as the anthocyanin donor yields F1 hybrids with a "globular" phenotype of large patches of petal anthocyanin. In contrasting, using the orange morph of *M. cupreus* as the anthocyanin donor

yields F1 hybrids with a fine spray of anthocyanin, referred to here as the "blush" phenotype (Fig. 1B-C).

The variation in spot size between *variegatus* x *luteus* hybrids versus *cupreus* x *luteus* hybrids suggests that the apparently solid-colored *M. l. variegatus* and *M. cupreus* have in fact colored their petals by means of two evolutionarily distinct spot formation systems, functioning at different spatial scales. We hypothesize that the two species' divergent spot formation systems are interacting in the hybrids to generate the observed diversity of color patterns. This hypothesis gives rise to the hypothesis that, at QTLs associated with spot size variation, *M. l. variegatus* alleles will confer larger spots than *M. cupreus* alleles.

To help investigate these ideas, we have developed digital tools for quantifying color patterning on a petal surface, as well as a new and highly improved *M. l. luteus* reference genome. Using these resources, we demonstrate that the spatially complex red (anthocyanin) color patterns found in an F2 hybrid mapping population have a correspondingly complex genetic basis, compared to the spatially simple trait of carotenoid intensity. We identify a candidate gene and propose a possible mechanism for the evolutionarily recent loss of yellow carotenoid pigmentation in *M. l. variegatus*. Finally, we explore the effects of *cupreus* compared to *variegatus* alleles on spot-size related traits, as a step towards understanding the evolutionary and developmental mechanisms that generate hybrid-specific phenotypic complexity.

Materials and Methods

Plant Materials and Growth Conditions

Two highly inbred lines, *Mimulus luteus* var. *variegatus* RC6 and *M. cupreus* LM43 (Table 6), were grown in the Whitman College greenhouse under 14-hour days with daily misting from an automatic watering system. Temperatures ranged from approximately 10-15 °C overnight and 15-30 °C during the day. Pollen from *M. cupreus* LM43 was applied to *M. l. variegatus* stigmas to generate F1 hybrid seeds. F1 plants were raised in the Whitman College greenhouse and manually self-fertilized to generate F2 hybrid seeds. The cross was performed in only one direction, because previous work indicated that cross direction does not influence color patterning traits in either an F1 or F2 hybrid population of *variegatus* x *cupreus* (Cooley and Willis 2009). F2 plants were grown in two locations: the Whitman College greenhouse, and the College of William & Mary greenhouse. At the latter location, plants were grown under 16 hour light regiment at 18-25 °C. *Mimulus l. luteus* line EY7 was grown at the College of William & Mary and was used for preparation of a reference genome.

Petal Photography and Pattern Quantification

Each *Mimulus* flower consists of five petals: two dorsal, two lateral, and one ventral (Fig. 1). In many *Mimulus* species, including those studied here, the ventral petal differs from the other five in that it has a series of anthocyanin spots leading from the petal lobe into the throat. These spots presumably act as nectar guides.

We decided to examine both dorsal petals of each flower, considering them to be approximate (though mirror image) replicates of the same pattern, and thus potentially able to provide some insight into the degree to which pattern variation is genetically determined versus stochastic. We also examined the ventral petal, hypothesizing that the genes which pattern the nectar guide region might alter how color patterns are produced throughout this unique petal.

From each of 2-4 flowers per F2 plant, the two dorsal petals and the ventral petal were cut at the end of the nectar guide and placed face-up on a strip of white tape. Petals were photographed in a darkroom. A single 60-watt bulb was used to provide illumination. Each photo included a color standard and a ruler. Photographs were taken using an Olympus VG-120 digital camera (Whitman College), and a Nikon D3200 (William and Mary). Each photograph was cropped and rotated in Gimp v. 2.10.8 (https://www.gimp.org/) so that only the tape background and petals were visible. The resulting .jpg image was processed using the digital image analysis pipeline described below. Additional traits were scored by eye (Table 2).

Digital Image Analysis Pipeline

To enhance downstream analysis of flower photos, full-color photos were first transformed to a 3-color space (L*a*b*), using k-means clustering with 3 centroids: red for anthocyanin-pigmented petal tissue, yellow for non-anthocyanin-pigmented petal tissue, and white for the background (Fig. 2). Using a custom script in combinaton with the Matlab plotter, clustering was supervised with manual repositioning of color centroids to ensure optimum retention of detail of spot shapes. Color categorization was implemented in Matlab version 2017b (Matlab 2017).

Following color-categorization, image analysis was undertaken using a custom pipeline (Fig. 2), implemented within the SciPy Ecosystem (Virtanen et al. 2020), with tools from the Scikit Image (Van der Walt et al. 2014) and Shapely (Gillies 2007) packages. In brief, spots and petals were vectorized to polygons from rasters using the Marching Squares (Cubes) algorithm (Lorenson and Cline 1987), holes in spots were detected with a custom algorithm, and all polygons were checked for validity and repaired, if possible. Repairs to polygons were implemented by both automated methods based on Shapely validation tools, and a series of manual correction programs based on the Matplotlib pyplot framework. These manual curations also acted as a final quality control step for vector images. All petal polygons were assigned a unitless area of one, and their respective spots were scaled accordingly, to allow generalized comparisons among flower petals and images of different sizes. Several programs for hand-curation of spot and petal vectors were created and used to ensure final image quality.

The final set of vectors representing the petal, spots, and holes-within-spots, were saved by the pipeline as geojson files that are readable with standard GIS software. Following the conversion of photos to vectors, spots were counted and measured in relation to the size of their respective petals and their position within the petals (Table 3). These measurements were then used as quantitative phenotypic traits in the QTL survey.

Code for the python image processing package is available on github at:<github.com/danchurch/mimulusSpeckling>. It is also available at pypi.org: <pypi.org/project/makeFlowerPolygons-dcthom/> A guided tour to the use of the pipeline is available as a jupyter notebook at: <nbviewer.jupyter.org/github/danchurch/mimulusSpeckling/blob/master/make_polygons/notebooks/petals_to_polygons.ipynb>.

Variance Components Analysis

To determine the proportion variation in our measured petal spot phenotypes that is due to genetic variation between individuals, we conducted a variance component analysis. We performed a PCA using Python to reduce all 15 quantitative phenotypes (Table 3) to four PC axes. A total of 353 individual plants were included in this analysis. For variance component analysis, the 'VCA' packing in R was used. The PCA dataframe was split by petal type. Variance components were determined separately for the dorsal petals and the ventral petal.

Dimensionality reduction of traits for QTL analysis

A total of 94 petal spot traits were computationally measured, to assess the features listed in Table 3 for both upper petals and lower petals (47 traits each). Due to high degrees of correlation between traits, these were reduced to the main axes of variation using PCA. Within a plant, for a given petal type (upper petals or lower center petal), each phenotype was separately averaged to give a single value per plant. A total of 353 unique plants were used. PCA dimension reduction was performed separately for the upper and lower petals. The first and second PC axes explained 39% and 22% of the variation for the upper petals and 36% and 24% for the lower petals. For the final QTL mapping, a dataset of the manually measured traits and the first four PC axes for the upper and lower petal were used.

Discrete phenotypic traits

In addition to the quantitative traits described above, eleven traits were scored "by eye" based on visual assessment of the petal photographs (Table 2). These traits were not included in the dimensionality reduction described above.

Assembly of the M. l. luteus reference genome

High molecular weight DNA was extracted from dark treated (to reduce levels of secondary compounds) *Mimulus luteus* var. *luteus* line EY7, and sequenced using the PacBio HiFi system. Approximately 35x genome-wide PacBio HiFi coverage was generated. In addition, a HiC library was constructed and sequenced to approximately 100x to aid in scaffolding. The primary assembly was constructed using hifiasm (https://github.com/chhylp123/hifiasm) with primarily default parameters (except -D 10) to assemble PacBio highly accurate long reads (HiFi reads) into a set of contigs with a low probability of chimerism between subgenomes. Next, HiC reads were aligned to the primary assembly with bwa (mem) and processed with SALSA 2.2 (https://github.com/marbl/SALSA) to generate a scaffolded assembly. Analysis of synteny (COGE synmap) with the diploid *M. guttatus* suggested that read alignment included some mis-alignments between subgenomes and this had been incorporated into the scaffolding. We therefore moved away from bwa to novoalign for HiC alignments with parameters (-r None -t 30) selected to reject any alignments with even a low chance of alignment ambiguity. To aid in annotation, RNA was extracted from young floral and vegetative buds and sequenced using PacBio IsoSeq. The resulting assembly was annotated using two rounds of Maker with a transcriptome input generated using stringtie (https://ccb.jhu.edu/software/stringtie/) in reference guided long-read mode based on a set of *M. l. luteus* IsoSeq reads, together with the *M. guttatus* v5.0 proteome from Phytozome 12 (https://phytozome.jgi.doe.gov/pz/portal.html) and a repeat library generated on the assembly using repeatModeler (https://www.repeatmasker.org/RepeatModeler/). HMMs for Augustus (http://bioinf.uni-greifswald.de/augustus/) and SNAP (https://github.com/KorfLab/SNAP) were initially trained on *M. guttatus* gene models and then retrained on the first round *M. l. luteus* maker gene models, while geneMark used the standard eukaryotic gene models. Genome statistics were generated using the assemblathon stats (https://github.com/KorfLab/Assemblathon/) code and assembly/annotation completeness was assessed using BUSCO v4 (https://busco.ezlab.org/) with Eudicot odb10 models in genome and proteome modes respectively. Completeness assessment with the recently released BUSCO v5 code did not significantly change these results.

QTL Mapping

Fresh leaf tissue (0.09 - 0.10 g per leaf) was collected from each of 373 *M. l. variegatus* RC6 x yellow *M. cupreus* LM43 F2 plants and was snap frozen in liquid nitrogen. DNA was extracted a Qiagen DNeasy Plant Mini Kit (Germantown, MD, USA), double eluted in 30-35 uL of warm dH20, and checked for purity and concentration using either a Nanodrop Lite (Thermo Fisher Scientific, Waltham, MA, U.S.A.) or a Qubit 4 Fluorometer (Invitrogen, Carlsbad, CA, U.S.A.). Genotyping by Sequencing (GBS) libraries were subsequently prepared using 100ng DNA per sample. The GBS protocol followed Elshire et al. 2011 with the enzyme ApeKI, with up to 95 samples and a water control pooled per lane. Three lanes of Illumina sequencing were performed by the Duke University Center for Genomic and Computational Biology using 100bp single end sequencing. Sequences were demultiplexed using Stacks version 2.1 (Catchen *et al.* 2011) and aligned to the *M. l. luteus* genome using bowtie2 and SNPs were called using GATK HaplotypeCaller. The resultant VCF file was filtered to remove sites with greater than 50% missing data, converted into a HapMap using Tassel, and than processed with the Tassel 'run_pipeline.pl' to convert HapMap to csv. The 'run_pipeline.pl' script filters the SNP dataset to include locations that meet the following cutoffs: (1) the parents were genotyped and (2) the parental genotype is not heterozygous. (https://bitbucket.org/tasseladmin/tassel-5-source/wiki/UserManual/GenosToABH/GenosToABHPlugin). A total of 7767 sites survived this filtering.

The resultant CSV file was used in R/QTL to create linkage groups and conduct QTL mapping (Broman *et al.* 2003). To create linkage groups, the data was further filtered to remove individuals with more than 250 missing marker calls and remove markers with more than 100 missing individuals. Next, duplicate individuals, defined as those who shared greater than 80% identity at markers, were removed from the dataset. Finally, markers with distorted segregation patterns (as determined by cutoff of p<1e-5) were filtered. Using this filtered dataset, we estimated recombination fractions between alleles, formed linkage groups, and reordered markers using this newly formed linkage map. Finally, we carried out QTL mapping for 19 petal spot patterning traits using R/qtl. The R/QTL 'scanone' function, using the 'em' method and 'np' model and 500 permutations was used to identify QTL. After QTL mapping, genes falling under the peak or nearby wings (within X cM) of the peak were extracted.

Carotenoid extraction and analysis

Carotenoids were extracted from ventral petals of yellow-flowered *M. cupreus* and *M. cupreus* x *M. l. variegatus* F1 hybrids, as well as from a series of four F2 hybrids that display pigmentation ranging from pale to dark yellow (Fig. 5). The

petal area used for each extraction was standardized by closing a 1.5 mL Eppendorf tube over the petals to create a "punch". Two punches per flower were placed in 200 ul of methanol and ground with a nylon pestle until the tissue appeared colorless.

The ground tissue was centrifuged at 16,000 g for 1 min to pellet debris, after which the supernatant was transferred to a clean Eppendorf tube. Carotenoids and flavonoids were then partitioned by adding 150 ul of methylene chloride and 150 ul of distilled water was added to each tube. The carotenoid-containing layer (i.e., lower phase) of selected samples was collected and dried under a gentle stream of nitrogen gas. The dried extracts were treated with ethanolic potassium hydroxide to remove fatty acid esters according to the methods of Schiedt (1995). Saponified carotenoids were dried under nitrogen gas, redissolved in 9:1 (v/v) hexanes/acetone and chromatographed according to the method outlined in LaFountain *et al.* (2015).

Carotenoids were identified based on their absorption spectra and relative order of elution as compared to a previously published analysis (LaFountain *et al.* 2015). The relative percentages of each carotenoid per sample were determined by integrating the area under each carotenoid peak (Waters Empower 2.0 Software), manually correcting for differences in their extinction coefficients (Britton 1995) and comparing each value to the total carotenoid content.

Data Availability

Seeds are available upon request. File S1 contains trait data used for dimensionality reduction via Principal Components Analysis. File S2 lists the 94 traits used for the PCA. File S3 contains genes identified within each mapped QTL. Sequence data will become available on GenBank upon publication. The Python image processing package is available on github at <github.com/danchurch/mimulusSpeckling>.

Results

Chromosome-level assembly of the *M. l. luteus* reference genome

The *Mimulus luteus* assembly was generated from PacBio HiFi reads assembled (see methods) with Hifiasm Cheng *et al.* (2021) and scaffolded with Illumina HiC reads using SALSA2 (https://github.com/marbl/SALSA) after a careful alignment of scaffolding reads to the largely homozygous tetraploid contigs (round of selfing = Z) with the novoalign (http://www.novocraft.com/products/novoalign/) aligner. The genome size of 599MBase with a scaffold L50 of 32 and N50 of 6.4MBase compares favorably with estimates for the genome size of 640-680MBase made by flow cytometry and a previous assembly of 410MBase that had a scaffold N50 of 0.28MBase Vallejo-Marin (2012); Edger *et al.* (2017). Notably this revised assembly appears to incorporate both more duplicated polyploid gene space and pericentromeric sequence space missing from the previous luteus assembly and the assembly of the diploid relative *M. guttatus ssp. guttatus* (JGI version 5.0, Phytozome-13: MguttatusTOL551v5.0) respectively (Fig. 3). The distribution of Ks scores from a self-self synmap alignment validates the rich history of genome amplifications previously described in this species.

Consistent with a more complete polyploid gene space, the maker annotation (methods) gene count increased to 53,411 genes from 46,855 in the previous assembly and approaches twice the *M. guttatus* diploid protein-coding gene count (2x26,718(guttatus): 53,436). The similarity of these two values may however be somewhat coincidental since the short read guttatus assembly was likely depleted in gene space relative to the long-read *M. luteus* assembly whilst diploidization will have begun to reduce the gene count in *M. luteus*, bringing these two values together. Missing BUSCO genes remained around 2.5%, similar to the level in the previous assembly (2.6%).The luteus assembly contains significant blocks of sequence without evident CDS homology in the assembled *M. guttatus* genome (Fig. 3), albeit similar sequence can be found in unassembled *M. guttatus* BAC and clone sequence available in genbank. The size of these gene-poor blocks missing from the *M. guttatus* assembly ranges up to c. 10MBase, and is consistent with the incorporation of more complete centromeric and pericentromeric sequence in the long-read *M. luteus* assembly. However tandem duplicate arrays make scaffolding of these regions with HiC reads more prone to error than HiC scaffolding of chromosome arm regions.

Pattern variation is genetically influenced

A trait that is completely genetically determined would be expected to show no variation among the flowers on a single plant; all variation would be found among the genetically diverse F2 hybrid plants. Conversely, a trait that is completely determined by non-genetic factors should similar variation, on average, between any two flowers, regardless of whether they originated from the same or different plants. In other words, within-plant variation would be similar to among-plant variation.

The first two PCs for both the upper and lower petal clearly show that the majority of variation in the petal spot phenotypes is found among plants. For the PC1 trait, 81% and 72% of the variation was among plants for the dorsal and ventral petals, respectively, consistent with a substantial genetic component to trait variation.

Five anthocyanin QTLs and one carotenoid QTL contribute to hybrid flower color variation

QTL mapping revealed six distinct QTL peaks – five corresponding to patterning of anthocyanin and one to the level of carotenoid pigment (Fig. 4). Of the 19 traits mapped (11 "by-eye" traits; 4 principal components (PC) traits extracted from upper-petal phenotypes; and 4 PC traits extracted from ventral-petal phenotypes), 15 had significant peaks as determined by a permutation test.

The peak on LG3, which contains known anthocyanin-activating gene *MYB5*, was shared by the largest number of traits: AnthocyaninPresence, Lower-petal Proportion Red, Upper-petal Proportion Red, PC1-LowerPetals, PC1-UpperPetals, PC2-LowerPetals, PC2-UpperPetals, PC3-LowerPetals, PC4-UpperPetals, PC4-LowerPetals, and RimSpots. The peak on LG17 is linked to two traits that are specific to the upper petals: GlobularSpray-UpperPetals and PC3-UpperPetals. The peaks on LG4, 6, 14, and 15 each corresponded to only one trait (Blush, PC2-lower, AnthocyaninPresence, and CarotenoidIntensity, respectively). Four traits (FineSpray, HugeSpot, TipSpotsOnly, and Column) did not map to any QTL.

Overall, these QTLs explained relatively low proportions of total trait variation. The two traits that appear to segregate in a single-locus Mendelian fashion – AnthocyaninPresence and CarotenoidIntensity – have 16.6% and 17.2%, respectively, of trait variation explained by the identified QTLs. Other traits have about one third as much trait variation explained (Table X), or less (not shown).

Carotenoid composition changes with pigment intensity

Liquid-liquid phase partitioning of flavonoids and carotenoids, depicted in Fig. 5, reveals that the yellow pigmentation of these flowers is due to carotenoids. Flavonoid-based pigments (e.g., anthocyanins or chalcones) would be expected to migrate to the aqueous (upper) layer but were not observed in these samples. Carotenoids are expected to migrate to the methylene chloride (lower) layer. The range of yellow pigmentation observed in the lower layer of these samples corresponds to their respective floral colors, providing strong support that this pigmentation is carotenoid-based.

To further interrogate the identity of these pigments, the extracts from the *M. cupreus*, *M. cupreus* x *M. l. variegatus* F1 hybrids, and darkest and lightest F2 hybrids were separated by high-performance liquid chromatography (HPLC). These data reveal that all flowers sampled can synthesize the xanthophylls produced by the late carotenoid biosynthetic pathway enzymes (e.g., neoxanthin, deepoxyneoxanthin, and mimulaxanthin; Fig. 6). However, the F1 and light F2 individual show an increased relative concentration of beta-carotene, as would be expected if an early carotenoid biosynthesis pathway structure gene such as *BCH1* were disrupted.

Discussion

Interspecies hybrids expand the range of phenotypic variation seen in nature, presenting opportunities to investigate traits that are not typically seen in existing species. Here we compare the genetic architecture of a classic, spatially simple interspecies difference in yellow carotenoid pigmentation with that of a spatially complex anthocyanin pigmentation phenotype that emerges only in interspecies hybrids.

The two traits show distinct genetic architectures. The evolutionarily recent carotenoid reduction in *M. l. variegatus* maps to a single QTL, which encompasses the carotenoid biosynthetic gene *Beta-carotene hydroxylase-1*. In contrast, hybrid-specific anthocyanin pattern variation is influenced by a large-effect locus and four detectable smaller-effect loci, approximately consistent with the idea of a distribution of effect sizes proposed by Orr (1998) for adaptive evolution. The adaptive significance is unknown for either of the traits studied here. However, within the very small *luteus* group of *Mimulus*, carotenoid loss appears to have evolved twice (as a recessive trait in *M. naiandinus* and as a semi-dominant trait in *M. l. variegatus*; Fig. 1) and the gain of petal lobe anthocyanin pigmentation has evolved at least twice (Cooley and Willis 2009; Cooley *et al.* 2011). The repeated evolution of similar traits under similar ecological conditions is suggestive of adaptation (Endler 1986), but regardless of adaptive benefit, the argument that genetic architecture will vary by system and trait type (Dittmar *et al.* 2016) appears to be illustrated by our results.

In the course of investigating these questions, new resources were developed for the plant research community. The substantially improved, high-quality *M. l. luteus* genome will contribute to genetic research within the *luteus* group as well as comparative studies across *Mimulus*. The digital image analysis pipeline adds to the repertoire of tools available for studying complex spatial patterns in plants, an area that has been less accessible to geneticists compared to more qualitative traits.

BCH1 is a strong candidate for the evolution of reduced carotenoid pigmentation in M. l. variegatus

The carotenoid-based yellow intensity variation was mapped to a major QTL that contains the *Beta-Carotene Hydroxylase 1* gene (*BCH1*). *BCH1* is a particularly promising candidate gene for two reasons. (i) The dark yellow genotypes, including the *M. cupreus* parent and dark yellow F2s, accumulate no beta-carotene but a mixture of xanthophylls that are downstream of beta-carotene in the flower petals, whereas the light yellow genotypes, including the *M. cupreus* x *M. l. variegatus* F1 hybrid and pale yellow F2 individuals, accumulate beta-carotene as a major carotenoid component (Fig. 6). This is consistent with low level of *BCH* activity in the light yellow genotypes, as *BCH* is the enzyme converting beta-carotene to downstream xanthophylls. (ii) *BCH* is not only a key enzyme determining the relative composition between beta-carotene and downstream xanthophylls, but also underlies major QTLs explaining total carotenoid content variation in other plant systems. For example, *BCH* was implicated as the causal gene underlying the Y-locus determining potato tuber flesh color (yellow vs. white); elevated expression level of *BCH* is tightly associated with yellow flesh color (Kloosterman *et al.* 2010). Similarly, *BCH* expression level was found to be a major determinant of petal carotenoid content variation among *Ipomoea* species (Yamamizo *et al.* 2010).

How *BCH1* activity affects total carotenoid accumulation is unknown. One possibility is that higher *BCH1* activity increases metabolic flux into the carotenoid biosynthesis pathway, and hence leads to higher carotenoid production (Zhou *et al.* 2011). Another possibility is that chromoplasts in some plants have better storage capacity for esterified carotenoids than beta-carotene (Ariizumi *et al.* 2014). Since *BCH* catalyzes the addition of hydroxyl residues to beta-carotene and these hydroxyl groups are required for carotenoid esterification, *BCH* activity could affect total carotenoid accumulation by enhancing carotenoid storage in chromoplasts.

Transcriptional activator NEGAN/MYB5a underlies the largest-effect QTL for anthocyanin patterning

In flowering plants, the anthocyanin biosynthetic pathway that produces red to purple pigments is regulated by an "MBW" complex consisting of a subgroup-6 R2R3 MYB transcriptional activator and bHLH and WDR cofactors, possibly characterized by homodimers of the MYB and bHLH components (Petroni and Tonelli 2011). In dicots, the MBW complex specifically activates the "late" biosynthetic genes in the pathway, typically *DFR* and *ANS/LDOX* (Lepiniec *et al.* 2006; Dubos *et al.* 2010). In contrast to this relatively constant mechanism of anthocyanin activation, anthocyanin repression has been ascribed to at least 14 different protein and/or small RNA families to date, most of which act upon the MBW complex (LaFountain and Yuan 2021).

In principle, an anthocyanin pigmentation feature could be gained either through activation (evolutionary change in a member of the MBW anthocyanin-activating complex or its downstream target genes), or through de-repression (evolutionary change in an anthocyanin repressor). In practice, color patterning changes have frequently been tracked to the anthocyanin-activating subgroup 6 R2R3 MYB genes (Yuan *et al.* 2014; Streisfeld *et al.* 2013; Lowry *et al.* 2012; Yamagishi 2013; Schwinn *et al.* 2006), possibly because their high copy number and correspondingly tissue-specific expression make them less pleiotropically constrained than other regulators (Streisfeld and Rausher 2011).

The evolutionarily recent and genetically dominant gains of

petal lobe anthocyanin pigmentation, in both the the common orange morph of *M. cupreus* and the magenta-flowered *M. l. variegatus*, follow this pattern. Pigment gain has been mapped in each case to a single genomic region (*pla1* and *pla2* respectively). Each genomic region spans a tandem array of subgroup 6 R2R3 MYB genes, and the "activating" allele at each region is associated with higher expression of the late biosynthetic genes that are the targets of the MBW complex (Cooley *et al.* 2011).

The *pla1* genomic region, responsible for petal lobe anthocyanin in the orange morph of *M. cupreus*, harbors seven candidate MYB genes. In the present study, however, we used a rare yellow-flowered morph of *M. cupreus* that lacks the anthocyanin-activating allele at *pla1* (Cooley *et al.* 2011). Thus, all anthocyanin activation in our genetic mapping population is expected to be caused by the *M. l. variegatus* allele of the *pla2* genomic region. Consistent with this expectation, approximately a quarter of all F2 plants (82/310) lacked petal lobe anthocyanin, having inherited two copies of the recessive *M. cupreus pla2* allele.

Within *pla2*, a combination of transgenic overexpression, RNAi, and transcriptomics has identified *MYB5a* as the causal gene for petal lobe anthocyanin activation in *M. l. variegatus* (Zheng *et al.* 2021). Thus, it is not surprising that the QTL of largest effect discovered in this study, on LG3, is centered directly over the *MYB5a* gene. What is more suprising is that, in fact, a weak second QTL (on LG14) was also associated with the trait of anthocyanin presence versus absence, even though the trait segregated in a 3:1 ratio in the F2 population. Out of 310 F2 plants, 228 (73.5%) had petal lobe anthocyanin and 82 (26.5%) did not. It is possible that these two QTLs are genetically (though not physically) linked, or that the weaker LG14 peak is a false positive.

The major-effect LG3 QTL impacts variation in ten other traits in addition to anthocyanin presence versus absence, including many of the summary traits drawn from the principal components dimensionality reduction. Variation in these 11 traits is likely dominated by the complete lack of petal lobe anthocyanin pigment in one quarter of our F2 plants. A QTL analysis restricted to the 228 plants that possess petal lobe anthocyanins did not identify any new QTLs (unupbl. data), possibly due to the modest sample size of this subset of our mapping population. In future work, focusing research effort on plants with the dominant anthocyanin-present phenotype would maximize statistical power to detect non-LG3 QTLs that contribute to quantitative pattern variation, rather than to the qualitative presence versus absence of petal lobe anthocyanins.

While the *variegatus* allele of *MYB5a*, at the LG3 locus, is required to activate anthocyanin production (Cooley and Willis 2009; Cooley *et al.* 2011; Zheng *et al.* 2021), we expect that the QTLs outside of the LG3 region are largely responsible for creating the spatial patterns of anthocyanin distribution. Each of these loci on its own contributes to one or two specific aspects of color patterning, according to the main effects in our QTL analysis, and interactions amongst loci may also be important in explaining the genetic component of pattern variation.

Inter-genomic interactions

In the *luteus* group of *Mimulus*, patterned petal anthocyanin pigmentation is seen only in hybrids. This observation alone indicates an important role for inter-genomic interactions. Perhaps the simplest interaction that could create a hybrid-specific trait is between two alleles at a single locus, with heterozygotes displaying a different phenotype from either homozygote. We tested this hypothesis by more closely examining the genotype-phenotype relationship for two QTLs: LG4, which associates with the "blush" or fine-spray phenotype, and LG17, which associates with the "globular" or large-spot phenotype. We asked whether either of these traits might arise only in a heterozygous genotype. Tables 5 and 6 show that this is not the case: the "blush" phenotype is commonly observed in both c/v heterozygotes and c/c homozygotes (where c indicates a *cupreus* allele at LG4, and v indicates a *variegatus* allele). Similarly, the "globular" phenotype is commonly observed for both c/v heterozygotes and v/v homozygotes at LG17.

Although this pattern does not support a simple within-locus explanation for the cause of hybrid-specific complexity, it does shed light on the question of whether the apparently solid anthocyanin pigmentation of orange-flowered *M. cupreus* is in fact the product of a dense spray of tiny spots, while the lobes of *M. l. variegatus* petals are each pigmented by one large spot with a single developmental origin. If so, we predict that *variegatus* alleles will, on average, be correlated with larger spot sizes than *cupreus* alleles, and the overall spot size will vary linearly with the total number of *variegatus* versus *cupreus* alleles at spot-size QTLs. At the LG4 (blush) QTL, nearly all plants with the blush phenotype had at least one cupreus allele. At the LG17 (globular spray-upper petals and PC3-upper petals) QTL, nearly all plants with the globular spray phenotype had at least one variegatus allele. Taken together, these two patterns suggests that spot size is a trait that differs genetically between the two apparently solid-colored parents, with the small-spot alleles tending to originate from M. cupreus.

Although plants with blush tended to have a *cupreus* allele (c) at the LG4 QTL, plants without blush showed all three possible genotypes (c/c, c/v, and v/v). Plants without blush included many c/c and c/v individuals, and in fact did not differ significantly from the 1:2:1 ratio expected for an F2 population (X-squared = 2.6175, df = 2, p-value = 0.2702). Thus, having a cupreus allele at the LG4 QTL appears to be nearly always necessary, but not at all sufficient, to generate the blush trait (Table 5).

The converse pattern was observed for the trait of globular spray. Plants with globular spray tended to have a *variegatus* allele at the LG17 QTL. Plants that lacked globular spray had fewer heterozygotes at the LG17 QTL than predicted by the 1:2:1 null hypothesis (X-squared = 15.833, df = 2, p-value = 0.0003646), but showed similar numbers of the two homozygous genotypes (c/c and v/v). Having a *variegatus* allele at the LG17 QTL, therefore, appears to be nearly necessary but not sufficient to generate the globular spray trait (Table 6).

If a dominant genotype is necessary but not sufficient to generate either blush or globular spots, then other factors must contribute - possibly including inter-locus epistasis. In favor of this hypothesis is the discovery that nectar-guide spots in *M. guttatus* are consistent with a two-component reaction-diffusion system, in which an autocatalytic R2R3 MYB anthocyanin activator orthologous to *MYB5a* eventually activates its own repressor (Ding *et al.* 2020). Interactions between the activator and the repressor determine spot size through dynamic mechanisms that are likely to be non-additive, a finding that is supported by preliminary mathematical modeling results in our lab (unpubl. data). A larger mapping population will be required to provide enough power to search for the epistatic interactions predicted by this mechanism.

Conclusions

Hybrids can be intermediate in phenotype to their two parents; similar to only one parent; or, in some cases, entirely different from what would be predicted based on the parents' appearance. Here we present an intriguing example of the latter case. We show that the genetic architecture of a hybrid-specific anthocyanin color patterning trait is complex, especially as compared to a more 'typical' pigmentation trait (carotenoid intensity) that differs between the two parent *Mimulus* species. Interspecific allele interactions at a single locus do not appear to explain hybrid patterning. Rather, two- or multi-locus interactions may be responsible for these emergent traits, perhaps due to an underlying developmental mechanism such as a reaction-diffusion system.

Acknowledgements

We thank former William & Mary and Whitman undergraduates Matthias Leu, Taylor Wilke, and Aaron Williams, for earlier work that helped pave the way for ideas explored in this paper. Funding was provided by NSF-DEB 1655311 (to AMC), NSF-DEB 1754075 (to AMC and JRP), NSF-IOS 2031272 (to AMC and JRP), Whitman College, and the College of William & Mary.

Literature Cited

Anderson, E. and G. L. Stebbins Jr, 1954 Hybridization as an evolutionary stimulus. Evolution **8**: 378–388.

Ariizumi, T., S. Kishimoto, R. Kakami, T. Maoka, H. Hirakawa, *et al.*, 2014 Identification of the carotenoid modifying gene pale yellow petal 1 as an essential factor in xanthophyll esterification and yellow flower pigmentation in tomato (s olanum lycopersicum). The Plant Journal **79**: 453–465.

Barker, W., G. Nesom, P. M. Beardsley, and N. S. Fraga, 2012 A taxonomic conspectus of phrymaceae: a narrowed circumscription for mimulus, new and resurrected genera, and new names and combinations. Phytoneuron **39**: 1–60.

Beldade, P. and P. M. Brakefield, 2002 The genetics and evo–devo of butterfly wing patterns. Nature Reviews Genetics **3**: 442.

Boogaard, F. P., K. S. Rongen, and G. W. Kootstra, 2020 Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging. Biosystems Engineering **192**: 117–132.

Bradshaw Jr, H. and D. W. Schemske, 2003 Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature **426**: 176.

Britton, G., 1995 Carotenoids today and challenges for the future. Carotenoids **1**: 13–26.

Broman, K. W., H. Wu, Ś. Sen, and G. A. Churchill, 2003 R/qtl: Qtl mapping in experimental crosses. Bioinformatics **19**: 889–890.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait, 2011 Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, genomes, genetics **1**: 171–182.

Chalker-Scott, L., 1999 Environmental significance of anthocyanins in plant stress responses. Photochemistry and photobiology **70**: 1–9.

Chen, L., B. Hu, Y. Qin, G. Hu, and J. Zhao, 2019 Advance of the negative regulation of anthocyanin biosynthesis by myb transcription factors. Plant physiology and biochemistry .

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li, 2021 Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature methods **18**: 170–175.

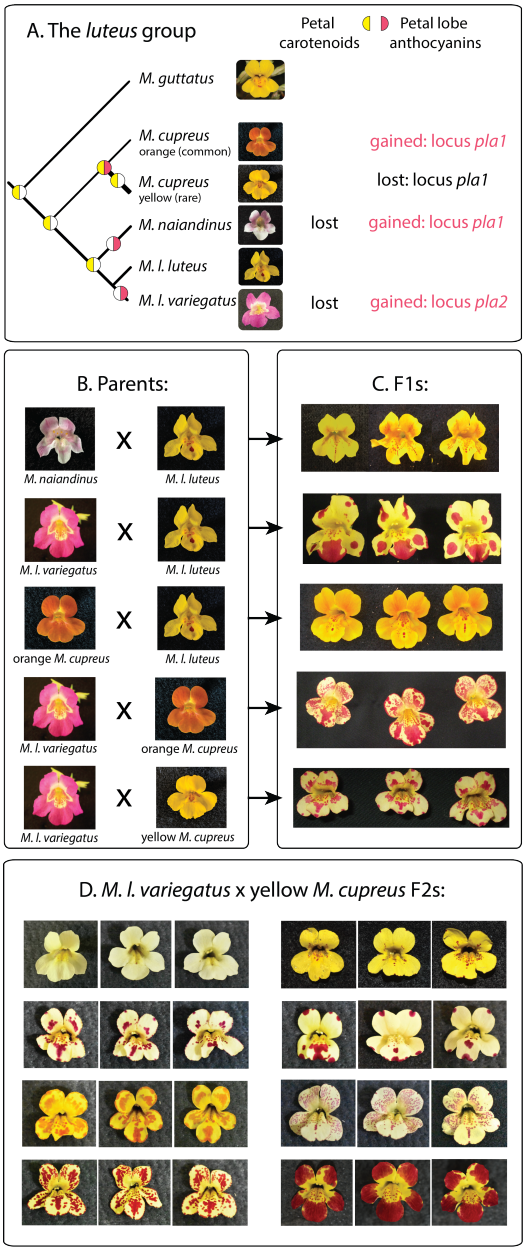

Figure 1 Flower color divergence yields complex spatial patterning in *Mimulus* hybrids. A. Because the *luteus* group is nested within a large monophyletic grouping of yellow-flowered species, the ancestral state is inferred to consist of yellow, carotenoid-pigmented petals (indicated by yellow color in the left side of the circle) with no petal lobe anthocyanins (indicated by a lack of magenta color in the right side of the circle). Carotenoids are hypothesized to have been lost twice independently, while anthocyanins are hypothesized to have been gained three times and lost once. B. Crosses are shown with maternal parent first; no differences were observed between reciprocal crosses. C. Three flowers per plant are shown for each F1 hybrid to illustrate the degree of consistency across flowers of the same genotype. D. Three flowers per plant are shown for each of X different F2 hybrids derived from the *M. l. variegatus* x yellow-flowered *M. cupreus* cross. Photos courtesy Bella Rivera, Joshua Shin, Leah Samuels.

Table 1 Seed sources

Taxon	Line ID	Inbreeding	Population	Location[a]
M. luteus var. *luteus*	Mll-EY7	13 generations	El Yeso	33.4°S, 70.0°W (2600 m)
M. luteus var. *variegatus*	Mlv-RC6	13 generations	Río Cipreses	34.2°S, 70.3°W (1200 m)
M. cupreus yellow morph	Mcu-LM43	10 generations	Laguna del Maule	36.0°S, 70.3°W (2300 m)

[a] Locations from which seeds were collected is given as latitude, longitude (meters above sea level.

Table 2 Qualitative traits utilized in genetic mapping

Trait name	Description
Carotenoid intensity	Darkness of yellow petal pigment, scored as low, medium, or high.
Anthocyanin presence	Binary indicator of whether any anthocyanin pigment is present on the petal lobes outside of the nectar guide region.
Blush	Binary indicator of whether any of the petals had a very thin layer of diffuse anthocyanin pigment.
Tip spot	Binary indicator of whether the two upper (dorsal) petals have a single spot at the tip of the petal.
Rim	Amount of rim covered by spots, scored as low, medium, or high.
Huge Spot	Binary indicator of whether there is a spot that covers nearly all of the petal.
Column	Binary indicator of whether majority of the central column of the lower (ventral) petal is anthocyanin pigmented.
Spray	Binary indicator of whether there is a spray of spots coming up from the throat on the two upper (dorsal) petals.

Table 3 Quantitative traits utilized in dimensionality reduction, for genetic mapping of principal components

Trait category	Traits assessed
General polygon info	Average Spot Size
	Number of spots
	Size of largest spot
Centeredness	Number of spots in the central zone
	Percentage of center covered by spots
	Average distance to center from all spots
Edgeness	Number of spots in the edge zones
	Percentage of total spots located in the edge zones
	Percent of edge zone covered by spots
	Average distance from all spots to edge
Throat region	Number of spots in the throat zone
	Percentage of total spots located in the throat zone
	Percentage of throat zone covered by spots
Quadrants	Number of spots in each of the four quadrants (proximal, distal, lower, upper)
	Percentage of spots in each of the four quadrants (proximal, distal, lower, upper)

Table 4 Phenotypic Variation Explained (PVE), by mapped QTLs. The five traits shown in Figure 3 are listed below; these are the mapped traits with the highest PVE.

Trait	PVE
Carotenoid intensity	0.172
Anthocyanin presence	0.166
Blush	0.062
Globular spray - upper petals	0.059
PC2 - lower petals	0.057

Table 5 Genotype at the LG4 QTL is predictive of blush phenotype. Genotypes and phenotypes are non-randomly associated (Chi-square contingency table: X-squared = 17.823, df = 2, p-value = 0.0001348). Plants with the Blush trait tend to have at least one *cupreus* allele at this locus. c, a cupreus allele at the LG4 QTL. v, a variegatus allele at the LG4 QTL.

Genotype	Blush present	Blush absent
c/c	24	53
c/v	25	127
v/v	4	71

Table 6 Genotype at the LG17 QTL is predictive of globular spray phenotype. Genotypes and phenotypes are non-randomly associated (Chi-square contingency table: X-squared = 17.333, df = 2, p-value = 0.0001723). Plants with the Globular Spray trait tend to have at least one variegatus allele at this locus. c, a cupreus allele at the LG17 QTL. v, a variegatus allele at the LG17 QTL.

Genotype	Blush present	Blush absent
c/c	4	80
c/v	32	90
v/v	25	70

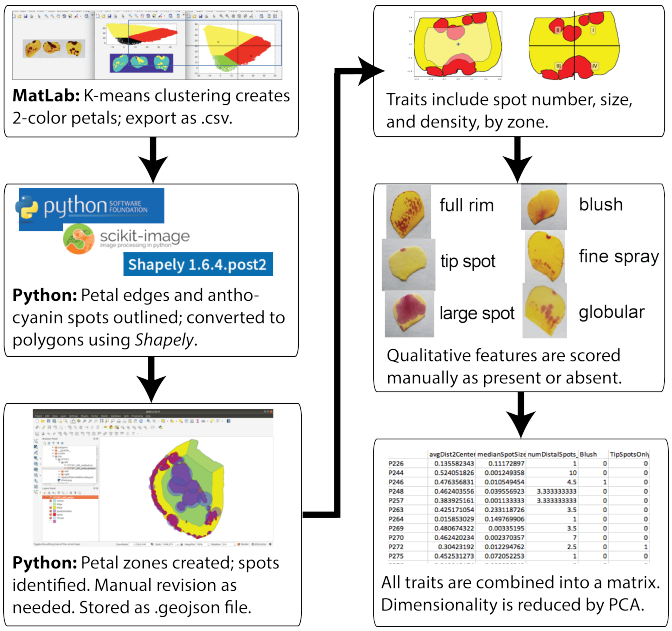

Figure 2 Digital image analysis workflow.

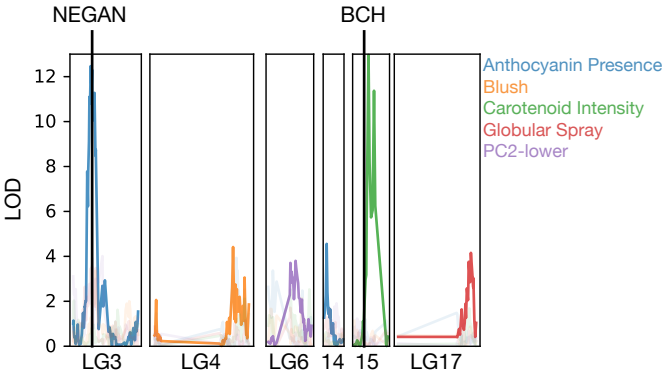

Figure 4 QTL mapping revealed six statistically significant peaks. The locations of two genes known to be important in anthocyanin spotting and carotenoid levels, NEGAN and BCH, are noted. The peak on LG3 was shared by eleven traits, only one of which is shown in full color here. The peak on LG17 is shared by two traits: GlobularSpray-UpperPetals (shown in color) and PC3-UpperPetals (in grey).

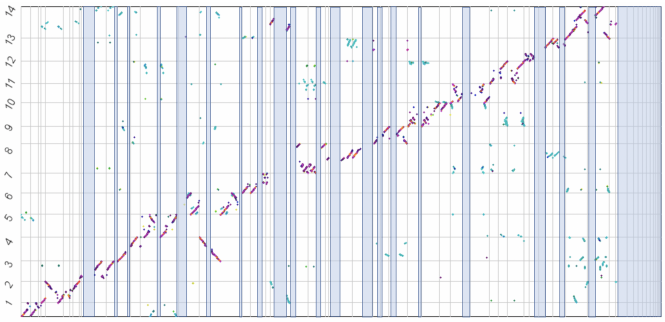

Figure 3 Syteny mapping between newly assembled *M. luteus* and publicaly available *M. guttatus* genome reveals many missing portions in the *M. guttatus* assembly. The *M. luteus* genome is on the x-axis and *M. guttatus* on the y-axis. Blue vertical boxes indicate portions of the *M. luteus* assembly that are missing in *M. guttatus*. The *M. luteus* whole genome duplication is also clearly evident in this plot as each *M. guttatus* genomic portion is typically represented by two *M. luteus* genomic segments.

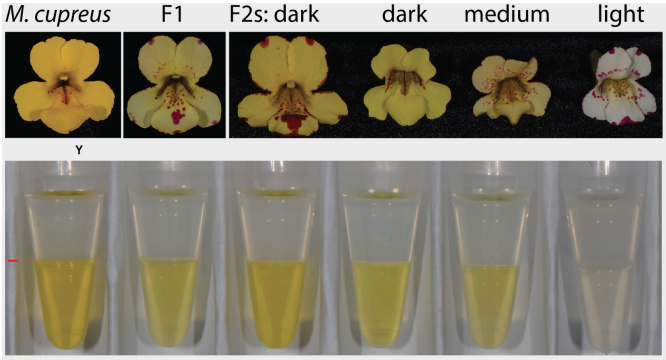

Figure 5 Flowers chosen for pigment extraction (top panel) and corresponding liquid phase partitioning of flavonoids and carotenoids (bottom panel). Red arrow denotes the interface between phases. The *M. cupreus* parent, *M. cupreus* x *M. l. variegatus* F1 hybrid, darkest F2 individual (third from left), and a pale-yellow individual (fifth from left) were further analyzed by high-performance liquid chromatography (HPLC).

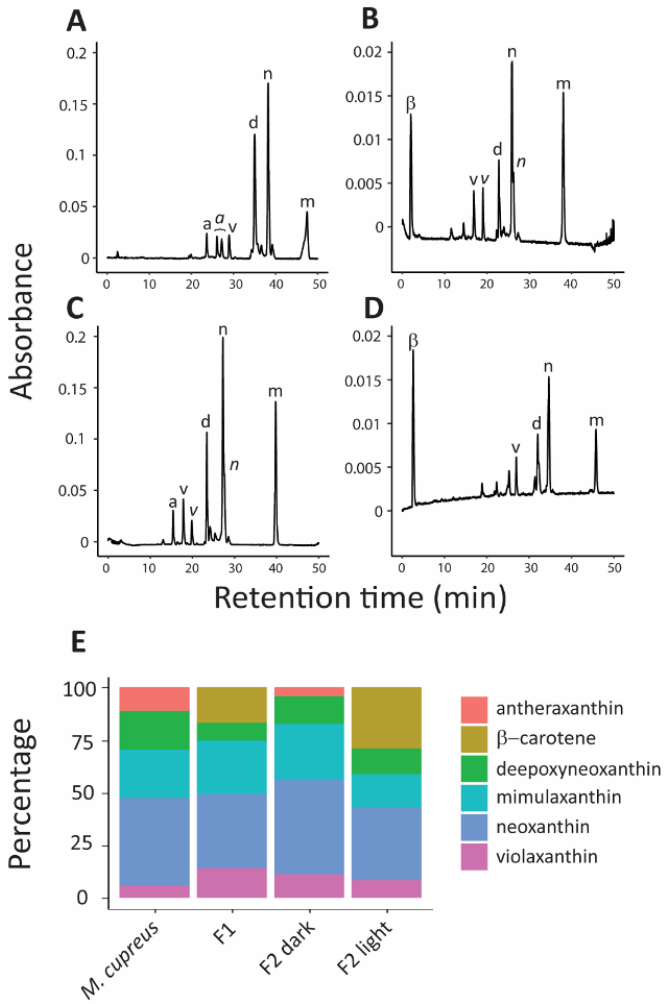

Figure 6 (A-D) HPLC chromatograms of floral extracts from *M. cupreus* (A), F1 hybrid (B), dark yellow F2 (C), and light yellow F2 (D). Detection wavelength for all chromatograms is 450nm. Absorbance values in panels A and C have been corrected for differences in injection volume, which were 4x and 2x, respectively. Abbreviations are as follows: b, beta-carotene; a, antheraxanthin; v, violaxanthin; d, deepoxyneoxanthin; n, neoxanthin; m, mimulaxanthin. Labels in italics denote cis-isomers. (E) Relative percentages of carotenoids as determined by integration of HPLC peak areas. **Note: I think we could place the photo of each plant within each graph, for easier reference. I can do that once I get my Adobe Illustrator license renewed - Arielle.**

Cooley, A., G. Carvallo, and J. Willis, 2008 Is floral diversification associated with pollinator divergence? flower shape, flower colour and pollinator preference in chilean mimulus. Annals of Botany **101**: 641–650.

Cooley, A. M., J. L. Modliszewski, M. L. Rommel, and J. H. Willis, 2011 Gene duplication in mimulus underlies parallel floral evolution via independent trans-regulatory changes. Current Biology **21**: 700–704.

Cooley, A. M. and J. H. Willis, 2009 Genetic divergence causes parallel evolution of flower color in chilean mimulus. New Phytologist pp. 729–739.

Davies, K. M., N. W. Albert, and K. E. Schwinn, 2012 From landing lights to mimicry: the molecular regulation of flower colouration and mechanisms for pigmentation patterning. Functional Plant Biology **39**: 619–638.

De Keyser, E., P. Lootens, E. Van Bockstaele, and J. De Riek, 2013 Image analysis for qtl mapping of flower colour and leaf characteristics in pot azalea (rhododendron simsii hybrids). Euphytica **189**: 445–460.

de Souza, M. M., F. N. Medeiros, G. L. Ramalho, I. C. de Paula Jr, and I. N. Oliveira, 2016 Evolutionary optimization of a multi-scale descriptor for leaf shape analysis. Expert Systems with Applications **63**: 375–385.

Demmig-Adams, B., A. M. Gilmore, and W. Adams 3rd, 1996 Carotenoids 3: in vivo function of carotenoids in higher plants. The FASEB Journal **10**: 403–412.

Ding, B., E. L. Patterson, S. V. Holalu, J. Li, G. A. Johnson, *et al.*, 2020 Two myb proteins in a self-organizing activator-inhibitor system produce spotted pigmentation patterns. Current Biology .

Dittmar, E. L., C. G. Oakley, J. K. Conner, B. A. Gould, and D. W. Schemske, 2016 Factors influencing the effect size distribution of adaptive substitutions. Proceedings of the Royal Society B: Biological Sciences **283**: 20153065.

Dubos, C., R. Stracke, E. Grotewold, B. Weisshaar, C. Martin, *et al.*, 2010 Myb transcription factors in arabidopsis. Trends in plant science **15**: 573–581.

Edger, P. P., R. Smith, M. R. McKain, A. M. Cooley, M. Vallejo-Marin, *et al.*, 2017 Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. The Plant Cell **29**: 2150–2167.

Endler, J., 1986 Natural selection in the wild. Monogr. Popul. Biol. **21**: 1–33.

Epifanio, J. and D. Philipp, 2000 Simulating the extinction of parental lineages from introgressive hybridization: the effects of fitness, initial proportions of parental taxa, and mate choice. Reviews in Fish Biology and Fisheries **10**: 339–354.

Galkovskyi, T., Y. Mileyko, A. Bucksch, B. Moore, O. Symonova, *et al.*, 2012 Gia roots: software for the high throughput analysis of plant root system architecture. BMC plant biology **12**: 116.

Gehan, M. A. and E. A. Kellogg, 2017 High-throughput phenotyping. American Journal of Botany **104**.

Gould, K. S., 2004 Nature's swiss army knife: the diverse protective roles of anthocyanins in leaves. BioMed Research International **2004**: 314–320.

Grant, A. L., 1924 A monograph of the genus mimulus. Annals of the Missouri Botanical Garden **11**: 99–388.

Grant, P. R. and B. R. Grant, 2019 Hybridization increases population variation during adaptive radiation. Proceedings of the National Academy of Sciences **116**: 23216–23224.

Grant, V., 1971 *Plant Speciation*. Columbia University Press.

Grotewold, E., 2006 The genetics and biochemistry of floral pigments. Annu. Rev. Plant Biol. **57**: 761–780.

Holton, T. A. and E. C. Cornish, 1995 Genetics and biochemistry of anthocyanin biosynthesis. The Plant Cell **7**: 1071.

Houle, D., D. R. Govindaraju, and S. Omholt, 2010 Phenomics: the next challenge. Nature reviews genetics **11**: 855–866.

Kellenberger, R. T., K. J. Byers, R. M. D. B. Francisco, Y. M. Staedler, A. M. LaFountain, *et al.*, 2019 Emergence of a floral colour polymorphism by pollinator-mediated overdominance. Nature communications **10**: 1–11.

Kimura, K., S. Kikuchi, and S.-i. Yamasaki, 1999 Accurate root length measurement by image analysis. Plant and Soil **216**: 117–127.

Kloosterman, B., M. Oortwijn, J. Uitdewilligen, T. America, R. de Vos, *et al.*, 2010 From qtl to candidate gene: genetical genomics of simple and complex traits in potato using a pooling strategy. BMC genomics **11**: 1–16.

Knecht, A. C., M. T. Campbell, A. Caprez, D. R. Swanson, and H. Walia, 2016 Image harvest: an open-source platform for high-throughput plant image processing and analysis. Journal of experimental botany **67**: 3587–3599.

LaFountain, A. M., H. A. Frank, and Y.-W. Yuan, 2015 Carotenoid composition of the flowers of mimulus lewisii and related species: implications regarding the prevalence and origin of two unique, allenic pigments. Archives of biochemistry and biophysics **573**: 32–39.

LaFountain, A. M. and Y.-W. Yuan, 2021 Repressors of anthocyanin biosynthesis. New Phytologist .

Lepiniec, L., I. Debeaujon, J.-M. Routaboul, A. Baudry, L. Pourcel, *et al.*, 2006 Genetics and biochemistry of seed flavonoids. Annu. Rev. Plant Biol. **57**: 405–430.

Levin, D. A., J. Francisco-Ortega, and R. K. Jansen, 1996 Hybridization and the extinction of rare plant species. Conservation biology **10**: 10–16.

Li, D., C. Li, Y. Yao, M. Li, and L. Liu, 2020 Modern imaging techniques in plant nutrition analysis: A review. Computers and Electronics in Agriculture **174**: 105459.

Li, M., M. H. Frank, and Z. Migicovsky, 2019 Colourquant: a high-throughput technique to extract and quantify colour phenotypes from plant images. arXiv preprint arXiv:1903.01652 .

Lowry, D. B., C. C. Sheng, J. R. Lasky, and J. H. Willis, 2012 Five anthocyanin polymorphisms are associated with an r2r3-myb cluster in mimulus guttatus (phrymaceae). American journal of botany **99**: 82–91.

Lowry, D. B., J. M. Sobel, A. L. Angert, T.-L. Ashman, R. L. Baker, *et al.*, 2019 The case for the continued use of the genus name mimulus for all monkeyflowers. Taxon **68**: 617–623.

Mallet, J., 2007 Hybrid speciation. Nature **446**: 279–283.

Marques, D. A., J. I. Meier, and O. Seehausen, 2019 A combinatorial view on speciation and adaptive radiation. Trends in ecology & evolution .

Martins, T. R., J. J. Berg, S. Blinka, M. D. Rausher, and D. A. Baum, 2013 Precise spatio-temporal regulation of the anthocyanin biosynthetic pathway leads to petal spot formation in clarkia gracilis (onagraceae). New Phytologist **197**: 958–969.

Martins, T. R., P. Jiang, and M. D. Rausher, 2017 How petals change their spots: cis-regulatory re-wiring in clarkia (onagraceae). New Phytologist **216**: 510–518.

McClintock, B., 1950 The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences **36**: 344–355.

Medel, R., C. Botto-Mahan, and M. Kalin-Arroyo, 2003 Pollinator-mediated selection on the nectar guide phenotype in the andean monkey flower, mimulus luteus. Ecology **84**: 1721–1732.

Medel, R., A. Valiente, C. Botto-Mahan, G. Carvallo, F. Pérez, *et al.*, 2007 The influence of insects and hummingbirds on the geographical variation of the flower phenotype in mimulus luteus. Ecography **30**: 812–818.

Minervini, M., H. Scharr, and S. A. Tsaftaris, 2015 Image analysis: the new bottleneck in plant phenotyping [applications corner]. IEEE signal processing magazine **32**: 126–131.

Nakano, A., H. Ikeno, T. Kimura, H. Sakamoto, M. Dannoura, *et al.*, 2012 Automated analysis of fine-root dynamics using a series of digital images. Journal of Plant Nutrition and Soil Science **175**: 775–783.

Neher, R. A. and O. Hallatschek, 2013 Genealogies of rapidly adapting populations. Proceedings of the National Academy of Sciences **110**: 437–442.

Niu, Y., Z. Chen, M. Stevens, and H. Sun, 2017 Divergence in cryptic leaf colour provides local camouflage in an alpine plant. Proceedings of the Royal Society B: Biological Sciences **284**: 20171654.

Orr, H. A., 1998 The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution **52**: 935–949.

Petroni, K. and C. Tonelli, 2011 Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant science **181**: 219–229.

Rieseberg, L. H., 1997 Hybrid origins of plant species. Annual review of Ecology and Systematics **28**: 359–389.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, *et al.*, 2003 Major ecological transitions in wild sunflowers facilitated by hybridization. Science **301**: 1211–1216.

Sapir, Y., M. K. Gallagher, and E. Senden, 2021 What maintains flower colour variation within populations? Trends in Ecology & Evolution .

Schiedt, K., 1995 Isolation and analysis. Carotenoids **1**: 81–108.

Schwinn, K., J. Venail, Y. Shang, S. Mackay, V. Alm, *et al.*, 2006 A small family of myb-regulatory genes controls floral pigmentation intensity and patterning in the genus antirrhinum. The Plant Cell **18**: 831–851.

Seehausen, O., 2004 Hybridization and adaptive radiation. Trends in ecology & evolution **19**: 198–207.

Sobel, J. M. and M. A. Streisfeld, 2013 Flower color as a model system for studies of plant evo-devo. Frontiers in plant science **4**: 321.

Stanton, K., C. M. Valentin, M. E. Wijnen, S. Stutstman, J. J. Palacios, *et al.*, 2016 Absence of postmating barriers between a selfing vs. outcrossing chilean mimulus species pair. American journal of botany **103**: 1030–1040.

Streisfeld, M. A. and M. D. Rausher, 2011 Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. Evolution: International Journal of Organic Evolution **65**: 629–642.

Streisfeld, M. A., W. N. Young, and J. M. Sobel, 2013 Divergent selection drives genetic differentiation in an r2r3-myb transcription factor that contributes to incipient speciation in mimulus aurantiacus. PLoS genetics **9**: e1003385.

Suarez-Gonzalez, A., C. Lexer, and Q. C. Cronk, 2018 Adaptive introgression: a plant perspective. Biology letters **14**: 20170688.

Takahashi, Y., K.-i. Takakura, and M. Kawata, 2015 Flower color

polymorphism maintained by overdominant selection in sisyrinchium sp. Journal of plant research **128**: 933–939.

Trivellini, A., B. Gordillo, F. J. Rodriguez-Pulido, E. Borghesi, A. Ferrante, *et al.*, 2014 Effect of salt stress in the regulation of anthocyanins and color of hibiscus flowers by digital image analysis. Journal of agricultural and food chemistry **62**: 6966–6974.

Trunschke, J., K. Lunau, G. H. Pyke, Z.-X. Ren, and H. Wang, 2021 Flower color evolution and the evidence of pollinator-mediated selection. Frontiers in plant science **12**.

Vallejo-Marin, M., 2012 Mimulus peregrinus (phrymaceae): A new british allopolyploid species. PhytoKeys p. 1.

Von Bohlen, C., 1995 El género mimulus l.(scrophulariaceae) en chile. Gayana Botanica **52**: 7–28.

Watson, J. M. and C. Von Bohlen, 2000 Plate 400. mimulus naiandinus. Curtis's Botanical Magazine **17**: 195–201.

Weight, C., D. Parnham, and R. Waites, 2008 Technical advance: Leafanalyser: a computational method for rapid and large-scale analyses of leaf shape variation. The Plant Journal **53**: 578–586.

Wu, C., D. Lowry, A. Cooley, K. Wright, Y. Lee, *et al.*, 2008 Mimulus is an emerging model system for the integration of ecological and genomic studies. Heredity **100**: 220.

Xu, W., C. Dubos, and L. Lepiniec, 2015 Transcriptional control of flavonoid biosynthesis by myb–bhlh–wdr complexes. Trends in plant science **20**: 176–185.

Yamagishi, M., 2013 How genes paint lily flowers: regulation of colouration and pigmentation patterning. Scientia Horticulturae **163**: 27–36.

Yamagishi, M., S. Toda, and K. Tasaki, 2014 The novel allele of the lh myb 12 gene is involved in splatter-type spot formation on the flower tepals of asiatic hybrid lilies (lilium spp.). New Phytologist **201**: 1009–1020.

Yamamizo, C., S. Kishimoto, and A. Ohmiya, 2010 Carotenoid composition and carotenogenic gene expression during ipomoea petal development. Journal of experimental botany **61**: 709–719.

Yoshioka, Y., H. Iwata, R. Ohsawa, and S. Ninomiya, 2004 Quantitative evaluation of flower colour pattern by image analysis and principal component analysis of primula sieboldii e. morren. Euphytica **139**: 179–186.

Yuan, Y.-W., 2019 Monkeyflowers (mimulus): new model for plant developmental genetics and evo-devo. New Phytologist **222**: 694–700.

Yuan, Y.-W., J. M. Sagawa, L. Frost, J. P. Vela, and H. D. Bradshaw Jr, 2014 Transcriptional control of floral anthocyanin pigmentation in monkeyflowers (*Mimulus*). New Phytologist **204**: 1013–1027.

Zheng, X., K. Om, K. A. Stanton, D. Thomas, P. A. Cheng, *et al.*, 2021 The regulatory network for petal anthocyanin pigmentation is shaped by the myb5a/negan transcription factor in mimulus. Genetics **217**: iyaa036.

Zhou, X., R. McQuinn, Z. Fei, A.-M. A. WOLTERS, J. Van Eck, *et al.*, 2011 Regulatory control of high levels of carotenoid accumulation in potato tubers. Plant, Cell & Environment **34**: 1020–1030.

Zufall, R. A. and M. Rausher, 2003 The genetic basis of a flower color polymorphism in the common morning glory (ipomoea purpurea). Journal of Heredity **94**: 442–448.