Strongly asymmetric hybridization barriers shape the origin of a new polyploid species and its hybrid ancestor[1]

Mario Vallejo-Marín[2]
Arielle M. Cooley[3]
Michelle Lee Yue Qi[2]
Madison Folmer[5]
Michael R. McKain[4]
Joshua R. Puzey[5,6]

[1] Manuscript received _____; revision accepted _____.

[2] Biological and Environmental Science, School of Natural Sciences. University of Stirling. FK9 4LA. Stirling, Scotland, United Kingdom
[3] Biology Department, Whitman College, Walla Walla, Washington 99362
[4] Donald Danforth Plant Science Center, St. Louis, MO 63132
[5] Department of Biology, College of William and Mary, Williamsburg, Virginia 23185
[6] Author for correspondence. Email: jrpuzey@wm.edu

ABSTRACT

- *Premise of the study:* Hybridization between diploids and tetraploids can lead to new allopolyploid species, often via a triploid intermediate. Viable triploids are often produced asymmetrically, with greater success observed for "maternal-excess" crosses where the mother has a higher ploidy than the father. Here we investigate the evolutionary origins of *Mimulus peregrinus*, an allopolyploid recently derived from the triploid *M. x robertsii*, to determine whether reproductive asymmetry has shaped the formation of this new species.
- *Methods:* We used reciprocal crosses between the diploid (*M. guttatus*) and tetraploid (*M. luteus*) progenitors to determine the viability of triploid hybrids resulting from paternal- versus maternal-excess crosses. To investigate whether experimental results predict patterns seen in the field, we performed parentage analyses comparing natural populations of *M. peregrinus* to its diploid, tetraploid, and triploid progenitors. Organellar sequences obtained from pre-existing genomic data, supplemented with additional genotyping was used to establish the maternal ancestry of multiple *M. peregrinus* and *M. x robertsii* populations.
- *Key results:* We find strong evidence for asymmetric origins of *M. peregrinus*, but opposite to the common pattern, with paternal-excess crosses significantly more successful than maternal-excess crosses. These results successfully predicted hybrid formation in nature: 111 of 114 *M. x robertsii* individuals, and 27 of 27 *M. peregrinus*, had an *M. guttatus* maternal haplotype.
- *Conclusion:* This study, which includes assembly of the first *Mimulus* chloroplast genome, demonstrates the utility of parentage analysis through genome skimming. We highlight the benefits of complementing genomic analyses with experimental approaches to understand asymmetry in allopolyploid speciation.

KEYWORDS: Allopolyploidy; asexual reproduction, chloroplast genome; genome skimming, hybridization; introduced species; *Mimulus peregrinus*; mitochondrial genome; sterile hybrid; triploid block.

INTRODUCTION

The extraordinary abundance of polyploid lineages among flowering plants (Otto and Whitton, 2000; Soltis et al., 2014; Barker et al., 2015) is surprising, given that nascent polyploids have to overcome significant hurdles in their formation and establishment. For example, genome duplication may cause deleterious changes in nuclear and cellular architecture, alter mitosis and meiosis, and modify gene expression (Comai, 2005). In addition, the initial rarity of polyploids originating in diploid populations imposes ecological disadvantages that result from reproductive costs in inter-ploidy matings (minority cytotype exclusion; Levin, 1975). The widespread association of polyploidy and inter-specific hybridization (allopolyploidy) means that in many cases, new polyploid lineages also have to overcome other genetic and ecological reproductive barriers that otherwise keep species apart (Ramsey and Schemske, 1998; Mallet, 2007; Rieseberg and Willis, 2007). Studies of recently formed allopolyploid lineages provide unique opportunities to investigate the early stages of polyploid formation and establishment.

Polyploids (individuals carrying more than two sets of chromosomes) can originate through a variety of pathways, including the somatic doubling of chromosomes in meristematic tissue during mitosis and, perhaps more commonly, through the fusion of one or more gametes with the somatic chromosome number (unreduced gametes) (Ramsey and Schemske, 1998; Mallet, 2007). When the formation of polyploids involves mating between gametes or individuals of different ploidy levels (e.g., reduced and unreduced gametes, or between a diploid and a tetraploid), an individual with an unbalanced ploidy level (e.g., a triploid) will be formed (Levin, 2002). In most cases, these odd-ploidy individuals are a transient phase in polyploid formation, leading to the eventual production of higher ploidy, stable lineages (e.g., tetraploids, hexaploids; Levin, 2002). The occurrence of an intermediate triploid stage in polyploid formation is known as the triploid bridge (Mallet, 2007).

Odd-ploidy levels (e.g., triploidy) such as triploids represent a significant challenge to overcome in the formation of polyploids, as inter-ploidy crosses often display strong post-zygotic reproductive barriers, severely reducing hybrid viability and fertility (Ramsey and Schemske, 1998; Levin, 2002; Comai, 2005; Köhler, Scheid, and Erilova, 2010). The term “triploid block” has been coined to describe the regular observation that inter-ploidy crosses, specifically between diploids and tetraploids, often result in early hybrid inviability and sterility (Köhler, Scheid, and

Erilova, 2010). However, evidence from both natural and experimental systems has shown that interploidy crosses can sometimes produce odd-ploidy offspring (e.g., triploids) that are nevertheless viable (Burton and Husband, 2000; Husband, 2004; Stace, Preston, and Pearman, 2015).

The ability to overcome the triploid block can be asymmetric; in other words, whether the maternal or paternal parent has the higher ploidy level, affects the viability of the resulting offspring (Köhler, Scheid, and Erilova, 2010). Asymmetric hybridization can be explained by a variety of mechanisms including cyto-nuclear incompatibilities (Tiffin, Olson, and Moyle, 2001), or imbalances in either the ploidy ratio of endosperm:embryo (Watkins, 1932; Burton and Husband, 2000) or between maternal:paternal genomes (Haig and Westoby, 1989; Köhler, Scheid, and Erilova, 2010). Barriers to inter-ploidy mating within and between species, including triploid block, have been studied using controlled crosses in a variety of natural and artificially produced diploids and polyploids (Ramsey and Schemske, 1998; Husband, 2004; Köhler, Scheid, and Erilova, 2010; Scott, Tratt, and Bolbol, 2013). Earlier surveys of interploidy crossing data have determined that maternal-excess crosses—those in which the maternal parent has a higher ploidy level than the paternal one—are consistently more successful than paternal-excess crosses at producing viable triploid seeds (Stebbins, 1957; Ramsey and Schemske, 1998). However, a number of counter-examples, in which paternal-excess crosses are more successful, have been recently discovered (Table 1; Supplemental Table 1).

Asymmetric reproductive barriers during interploidy hybridization can be identified experimentally by performing reciprocal crosses (e.g. Husband, 2004), but discovering whether asymmetries observed in the lab translate to patterns in nature, requires tracing the maternal *versus* paternal ancestry of wild populations. Ancestry analysis of this kind requires uniparentally inherited genetic markers, such as mitochondria or chloroplast that contain enough genetic differences between parental taxa. The rapid growth of next generation sequencing projects offers an extraordinary opportunity to generate (sometimes inadvertently) exactly the type of data necessary to elucidate the maternal ancestry of wild populations at a genomic scale. Because cytoplasmic genomes (mitochondria and chloroplast) are present at a much higher number per cell than the nuclear genome, targeted and non-targeted sequencing often results in incidental high coverage of cytoplasmic genomes, as well as of repetitive elements in the nuclear genome (Dodsworth, 2015). Thus mitochondrial and chloroplast information can be "skimmed" from the

larger collection of sequences obtained in whole-genome sequencing projects (Straub et al., 2012), or from targeted sequencing approaches (Weitemier et al., 2014), and used for parentage analyses. These resource-efficient approaches not only rescue information that would otherwise be discarded, but also provide effective means for addressing longstanding questions about the ancestries of naturally-occurring polyploid lineages.

A new system for investigating allopolyploid speciation, using the complementary approaches of experimental manipulation and next-generation sequencing, is the complex of naturalized *Mimulus* L. (Phrymaceae) species in Great Britain and Ireland. Hybridization between introduced populations of the North American diploid *M. guttatus* (2n = 2x = 28) and the South American tetraploid *M. luteus* (2n = 4x = 60-62) have given rise to *M. x robertsii,* a triploid (2n = 3x = 44-46), sexually sterile, but widely naturalized taxon distributed across the British Isles (Roberts, 1964; Parker, 1975; Silverside, 1990; Vallejo-Marín et al., 2015). By means of rapid vegetative propagation, *M. x. robertsii* has become common and is now found in about 40% of British *Mimulus* populations (Vallejo-Marín and Lye, 2013). The reproductive asymmetry typical of interploidy matings has been reported for *M. guttatus* x *M. luteus* hybrids, but in the less common direction: triploids have been successfully produced only by paternal-excess crosses, with tetraploid *M. luteus* as the pollen donor (Roberts, 1964; Parker, 1975). These early studies suggest that *M. x robertsii* is likely to have arisen in the same way, from a *M. guttatus* mother and *M. luteus* father. *Mimulus x robertsii* has given rise at least twice to a sexually fertile allohexaploid species, *M. peregrinus* Vallejo-Marín ($2n = 6x = 92$), through genome doubling (Vallejo-Marín, 2012; Vallejo-Marín et al., 2015). This speciation event is quite recent, as both progenitors were introduced to the United Kingdom (UK) in the last 200 years, and the earliest records of hybrid populations date back approximately 140 years (Preston, Pearman, and Dines, 2002; Vallejo-Marín et al., 2015). Given that *M. peregrinus* evolved from *M. x robertsii*, one would expect that the maternal ancestor of *M. peregrinus* is also *M. guttatus*.

Here we investigate the occurrence and consequences of inter-ploidy reproductive barriers in polyploid *Mimulus* (*M. x robertsii* and *M. peregrinus*) produced by hybridization between diploid (*M. guttatus*) and tetraploid taxa (*M. luteus*). We complement classic crossing experiments with analyses of the parentage of natural hybrid populations using genome-skimming of chloroplast and mitochondrial genomes and single-marker genotyping. Inter-ploidy hybrids in *Mimulus* are ideally suited for this purpose due to their experimental tractability (Wu et al., 2008) as well as

the growing genomic resources available for this genus, including whole genome sequences for both nuclear and mitochondrial genomes for *M. guttatus* (Mower et al., 2012; Hellsten et al., 2013), ongoing work on the *M. luteus* genome (Edger et al., in prep), and recent whole-genome and targeted sequence data for native and introduced *Mimulus* populations (Puzey and Vallejo-Marín, 2014; Vallejo-Marín et al., 2015). We experimentally evaluate the success of reciprocal crosses between *M. guttatus* and *M. luteus* to confirm previous observations of triploid block asymmetry, and compare for the first time the phenotype of hybrids produced in reciprocal directions. We use a combination of genome skimming and genotyping of natural populations to evaluate the hypothesis that reproductive asymmetry observed between *M. guttatus* and *M. luteus* violates the trend (Ramsey and Schemske, 1998) of triploid formation via maternal-excess crosses. Our study addresses the following specific questions: (1) What is the effect of cross direction on the phenotype and viability of hybrids? (2) Are naturalized triploid hybrids between *M. guttatus* and *M. luteus* (i.e., *M. x robertsii*) asymmetrically produced, and does the asymmetry match the pattern observed in experimental hybridizations? (3) Who is the maternal parent of hexaploid *M. peregrinus*, and does this fit the pattern seen for experimental crosses and naturalized *M. x robertsii*?

MATERIAL AND METHODS

Study system—The introduction of *Mimulus* into Europe in the 1800's resulted in hybridization between closely related, but previously geographically isolated taxa. Hybridization between taxa in the North American *M. guttatus* DC. and South American *M. luteus* L. species aggregates has yielded a taxonomically complex array of hybrids that establish naturalized populations in the British Isles (Roberts, 1964; Silverside, 1998; Stace, Preston, and Pearman, 2015). The *M. luteus* group includes a number of interfertile taxa: *M. luteus* var. *luteus, M. luteus* var. *variegatus, M. naiandinus, M. cupreus,* and *M. depressus* (Grant, 1924; Watson, 1989; von Bohlen, 1995; Cooley and Willis, 2009), some of which have become naturalized in the British Isles (Stace, 2010). Extant populations of *M. luteus* in the British Isles include highly polymorphic individuals, which may have resulted from hybridization events between different native varieties. Here we focus on interspecific hybrids between *M. guttatus* and different varieties of *M. luteus s.l.,* excluding hybrids with another closely related South American taxon, *M. cupreus*, which are rarer and easily distinguished (Stace, 2010). Hybrids between *M. guttatus* and *M. luteus s.l.* receive different taxonomic names, mostly based on color patterns of the corolla (Stace,

2010). For simplicity, hereafter we refer to these hybrids as *M. x robertsii*, and the interested reader is referred to Silverside (1998) for a more detailed morphological classification of hybrid types.

Hybridization between diploid *M. guttatus* (2n= 2x = 28) and tetraploid *M. luteus* (2n = 4x = 60-62) yields highly sterile, triploid individuals (2n = 3x = 44 - 46, *M. x robertsii*). Despite their sterility (Roberts, 1964; Parker, 1975), these triploid hybrids can be vegetatively vigorous and display a strong capacity to reproduce by clonal propagation from plant fragments that root at the nodes. *M. x robertsii* has been very successful at establishing naturalized populations throughout the British Isles (Stace, Preston, and Pearman, 2015), and it can become locally abundant, forming populations of thousands of individuals. A recent genetic survey of *M. x robertsii* showed that populations of this sterile hybrid are genetically variable and composed of multiple highly heterozygous genotypes (Vallejo-Marín and Lye, 2013). Approximately 40-50% of extant *Mimulus* populations in the United Kingdom contain *M. x robertsii* (Vallejo-Marín and Lye, 2013).

***Crossing design*—**To generate the plants used in this study, seeds were collected from natural populations of *M. guttatus* from Dunblane, Perthshire, Scotland (56.19°N, 3.98°W), and *M. luteus s.l.* from Coldstream, Scottish Borders, Scotland (55.65°N, 2.24°W) on 8 Aug 2009 and 10 July 2010, respectively. Individuals of *M. luteus s.l.* from Coldstream are polymorphic for the number of purple blotches present in the corolla lobes. Phenotypically, these *M. luteus s.l.* plants resemble intraspecific hybrids between *M. luteus* var. *luteus* and *M. luteus* var. *variegatus* (sometimes called *M. smithii*). From each population, we randomly chose six individuals collected as seeds from separate maternal plants and crossed them to generate 12 outbred lines for each species using an equal contribution design in which each individual contributed to the genome of four lines, two as the paternal parent and two as the maternal parent.

To generate interspecific hybrids, each of the 12 *M. guttatus* outbred lines were randomly paired with one of 12 *M. luteus s.l.* outbred lines. Two of these interspecific pairs were then assigned to each of six mating groups, and each of these groups was subject to a diallel cross, except selfs. To avoid half-sib matings, conspecifics in the same mating group were chosen to be unrelated. Therefore, all matings within each group are outbred and have the same inbreeding coefficient relative to the parental generation. Flowers were emasculated before conducting hand pollinations. Only 33 out of 72 possible hybrid crosses were performed due to flower availability.

The crossing design yielded individuals of four different types: (**a**) *M. guttatus* x *M. guttatus* (MG x MG); (**b**) *M. luteus s.l.* x *M. luteus s.l.* (ML x ML); and reciprocal F1 hybrid crosses of (**c**) *M. guttatus* x *M. luteus s.l.* (MG x ML), and (**d**) *M. luteus s.l.* x *M. guttatus* (ML x MG). Hereafter, crosses are represented as maternal x paternal parent. In total we obtained seeds of 33 lines of the following crossing types: MG x MG (4 lines), ML x ML (10 lines), MG x ML (10 lines), and ML x MG (9 lines). For all subsequent experiments, we randomly chose four lines of each crossing type (16 lines total; Appendix S2; see Supplemental Data with the online version of this article).

***Seed production*—**To determine the number of seeds produced per fruit, between 14 to 20 days after hand pollination, 9 to 11 randomly selected fruits of each parental and F1 cross types were collected into glassine paper bags. The fruits were dissected in a stereo microscope (MZ6, Leica Microsystems, Milton Keynes, UK) at 0.63x – 4.0x magnification to remove the fruit casing and count all seeds.

***Germination rate*—**To determine seed viability and germination, for the F1 crosses, 100 randomly selected seeds from each line were surface-sown in 3.5”, 370 mL, round pots (9F; Desch PlantPak, Netherlands) using a low nutrient compost (Modular Seed Growing Medium, William-Sinclair, Lincoln, UK). The pots were arranged in plastic trays in standing water (approximately 3-4cm water depth). Trays were placed in controlled environment chambers (Microclima; Snijders Scientific B.V, Tilberg, Netherlands) at the University of Stirling, and maintained at cycles of 16 light-hours (24°C) and 8 dark-hours (16°C) at 70% constant humidity. Seedling emergence was scored daily by counting the cumulative number of individuals in which the cotyledons could be observed.

***Floral and vegetative phenotyping*—**To compare the phenotypes of parental and hybrid lines, a second set of seeds from the same lines were planted and germinated as described above. Individual seedlings of each line were transplanted 10 days after germination into 3.5” pots with low nutrient compost and arranged in flooded, plastic trays. Trays were placed at the University of Stirling glasshouses under natural light conditions, with 16 hours of supplemental light provided by compact-fluorescent lamps. Heating was only provided if the temperature inside the glasshouse dropped below 15°C during the day and 10°C at night. Plants were sprayed once to control for an aphid infestation, with 1.5L of diluted Provado Ultimate Bug Killer (20 mL/L) (Bayer Garden, Cambridge, UK). This experiment was carried out between 3 July and 2 October 2012. The date and plant height at the time when the first flower opened was recorded, as well as

the following measurements from the first two flowers of each plant, which were taken using digital calipers: (a) corolla width (maximum distance between lateral petal lobes in frontal view); (b) corolla height (maximum distance between upper and lower petal lobes in frontal view); (c) corolla throat (size of the opening of the corolla tube taken near the center of the flower); (d) calyx length (measured at the upper sepal); (e) tube length (measured in side view); (f) pedicel length; (g) bract width (width of the subtending leaf at the node of the flower being measured) ; and (h) bract length (including petiole). Nectar production was measured on a subset of flowers using a 50µL calibrated glass pipette (Drummond Scientific Company, Broomall, PA, USA).

***Confirming the success of artificial crosses*—**To further check that the individuals produced through artificial crosses were indeed derived from interspecific hybridization, we genotyped all plants used in the morphological analyses. Leaf tissue samples were collected in August-October 2012 and stored in silica gel. DNA was extracted using a modification of the CTAB protocol described in Doyle and Doyle (1990), and quantified with Nanodrop 2000 (Thermo Scientific, Wilmington, DE, USA). We genotyped putative hybrids and parents at six microsatellite loci (ATT240, AAT217, ATT267, AAT230, AAT225, AAT300; (Kelly and Willis, 1998) following the protocol described in Vallejo-Marín and Lye (2013). We used STRand 2.4.59 (Toonen and Hughes, 2001) to analyze fluorescence profiles. Based on the genotyping results, four ML x MG individuals were identified as self-pollinations of the ML maternal plant based on lack of expected heterozygosity, and excluded from subsequent analyses.

***Data analysis*—**The number of seeds produced by the four cross types (MG x MG, MG x ML, ML x MG, and ML x ML) was statistically compared with a general linear model (with normal distribution) with cross type as explanatory variable using the *glm* package in *R* ver. 3.1.2 (R Core Team, 2015). Pairwise posthoc comparisons among cross types was done using Tukey tests on the fitted model using *multcomp* (Hothorn, Bretz, and Westfall, 2008). Germination curves for each cross type were calculated as the mean of the four studied lines. The proportion of plants that flowered was analyzed using a binomial distribution (logit link) in *glm*, and posthoc pairwise comparisons among cross types done with Tukey tests in *multcomp*. Nectar content and plant height were also analyzed using *glm* with cross type as an explanatory variable, and Tukey tests performed in *multcomp.* Floral and leaf traits are presented as mean and standard error for illustration only, and no further general linear model or pairwise posthoc comparisons for cross types were done on these traits.

***Chloroplast assembly*—**Despite the availability of both nuclear (Hellsten et al., 2013) and mitochondrial (Mower et al., 2012) genomes for *M. guttatus,* no chloroplast genome for any *Mimulus* species had been previously published. Thus we sought to generate an assembly of the entire chloroplast genome using newly generated whole-genome sequences of an advanced inbred line of *M. luteus var. luteus* (MLll/MLl2, Table 2). Whole genome shotgun reads were trimmed using Trimmomatic v.0.32 (Bolger, Lohse, and Usadel, 2014) with the SLIDINGWINDOW:10:20 and MINLEN:40 parameters. Trimmed reads were mapped against the *Salvia miltiorrhiza* (NC_020431.1; Qian et al., 2013) using bowtie2 v.2.1.0 (Langmead and Salzberg, 2012) with the default "very-sensitive-local parameter" set. Mapped reads were assembled using SPAdes v.3.1.0 (Bankevich et al., 2012) with k-mer sizes of 55 and 87 under the "only-assembler" option. The resulting contigs were assembled with the full trimmed data set using afin (bitbucket.org/benine/afin) with the following parameters: -l 50 -f .1 -d 100 -x 100. afin spans contigs by trimming contig ends, identifying matching reads to the contig ends, and extending the contigs iteratively while attempting to fuse contigs at each iteration. Ultimately, afin was able to assemble the plastome from ~10 contigs to one. The assembled plastome was put into Sequencher 5.0.1 (Genecodes), and the boundaries of the quadripartite regions (large single copy-LSC, inverted repeat B-IRB, small single copy-SSC, and inverted repeat A-IRA) were identified. The final plastome is presented in the LSC-IRB-SSC-IRA format. The final assembly of the plastome was verified through a coverage analysis. Jellyfish v.2.1.3 (Marçais and Kingsford, 2011) was used to estimate 25-mer abundance from the cleaned reads. These abundances were used to map a 25 base pair sliding window of coverage across the assembled plastome. The coverage was found to be uniform and equal for the single copy regions (LSC=373X; SSC=328X) with the inverted repeat having ~2x the coverage of the single copy regions (IR=756X), as expected. DOGMA (Wyman, Jansen, and Boore, 2004) was used to annotate the *Mimulus luteus* plastome. A graph of the annotated plastome structure was created using Circos v.0.66 (Krzywinski et al., 2009), as implemented in Verdant (verdant.iplantcollaborative.org).

***Samples for genomic parentage analysis*—**In order to determine the relationships between the cytoplasmic genomes of *M. guttatus* and *M. luteus*, we analyzed previously published and newly generated sequencing data of both parental taxa (Table 2). We included 24 individuals of *M. guttatus* from 10 introduced (UK) and 9 native populations. Within the introduced range we included replicate individuals for four populations (AYR, CER, DBL and HOU), as well as a

controlled cross of one of these populations (DBL). For *M. luteus*, we included an inbred line of *M. luteus var. luteus* and one inbred line of *M. luteus var. variegatus* both originally collected in the native range (Chile). The *M. luteus var. luteus* inbred line was represented by two individuals (MLl1 and MLl2, from the 7th and 9th inbred generations). In addition, we analyzed *M. luteus s.l.* from two populations in the introduced range in the UK (COL and EVI), including two individuals originating from the same population (COL): a wild collected plant, and an individual from an experimental cross (Table 2). For *M. x robertsii,* we included samples from four natural populations (GIO, LED, NEN and TOR). For *M. peregrinus* we analyzed samples from three populations (GON, STR, and LED), including three individuals from the type populations for the allopolyploid species (LED). As a control, we also analyzed an experimentally-produced "*M. x robertsii*" hybrid obtained by crossing *M. guttatus* x *M. luteus s.l.* (DBL x COL, maternal x paternal parent; CG-1-1 x CS-4-3 = H003b), as well as an S1 synthetic "*M. peregrinus*" allohexaploid produced by colchicine treatment of the H003b "*M. x robertsii*" line (13-SYN-1-JP1). In addition, we also included three samples from taxa closely related to *M. guttatus* in North America (*M. nasutus, M. micranthus,* and *M. platycalyx*), plus one outgroup (*M, dentilobus*) (Table 2). In terms of sequencing approach, 16 samples (four of each *M. guttatus, M. luteus, M. x robertsii,* and *M. peregrinus*) were derived using sequence capture (Vallejo-Marín et al. 2015), and the remaining through whole-genome sequencing obtained from previously published data (10 samples; Puzey and Vallejo-Marín 2014, 13 samples from the NIH sequence read archive (deposited by JGI; http://www.ncbi.nlm.nih.gov/sra)), or data newly generated here (six samples). In total we analyzed 45 genomic samples (Table 2).

***Genotyping of plastid genomes*—**We used the publicly available *M. guttatus* mitochondrial genome (Mower et al., 2012) (JN098455) or the newly assembled *M. luteus var. luteus* chloroplast genome (this study) as references for base calling and genotyping. All reads were aligned to the references using *bowtie2* (Langmead and Salzberg, 2012). Alignments were sorted, duplicates marked and discarded, and read groups added using *Picard* tools (http://broadinstitute.github.io/picard). Following alignment filtration, genotypes were called used GATK UnifiedGenotyper (McKenna et al., 2010; DePristo et al., 2011; Auwera et al., 2013) (Appendix S3; see Supplemental Methods with the online version of this article). Following genotyping, indels and heterozygous sites were excluded.

***Filtering of genomic data for phylogenetic analysis*—**The 17 chloroplast-derived regions of the *Mimulus guttatus* mitochondrial genome (NC_018041.1; Mower et al., 2012) were extracted and aligned against the *Mimulus luteus* chloroplast genome using Sequencher v.5.0.1 (http://www.genecodes.com). The *Mimulus luteus* chloroplast sequence was trimmed prior to alignment so that only the LSC, IRB, and SSC were present. Alignments were inspected for sequence similarity. Multiple instances where portions of the mitochondrial chloroplast-derived sequences did not have an ordinal alignment with the majority of the sequence were identified. In these cases, the region that did not align was removed and realigned to the full plastome. These chimeric chloroplast-derived sequences may reflect multiple recombination events, and subsequence recombination hotspots of the mitochondrial genome, or they are representative of multiple chloroplast-genome insertion events in the history of the mitochondrial genome. Once aligned, the coordinates of alignment to the chloroplast genome were recorded. Data corresponding to these positions were eventually removed from the data matrix (see below).

Chloroplast sequences were reconstructed from VCF files representing mapping to the *Mimulus luteus* plastome (see above). For each position of the *M. luteus* plastome, the corresponding base in the mapped data set was recorded. If an indel was identified, that position was ignored. Indels are difficult to distinguish from missing data using this method, so we chose to completely ignore them to reduce potentially conflicting signal. Each reconstructed plastome was split into features (protein-coding genes, tRNAs, and rRNAs) and inter-feature regions, which were used for individual alignments and concatenation (see below). Features were identified based on their alignment to known coordinates of the annotated the *Mimulus luteus* chloroplast genomes. Chloroplast regions identified in mitochondrial genome were excluded in phylogenetic analyses. This process was done using a Perl script (https://github.com/mrmckain/AJB_2015_Mimulus_Allopolyploidy).

Mitochondrial sequences were reconstructed in a similar manner to that of the plastid sequences. In this case, the annotation of the *M. guttatus* mitochondrial genomes was used to recognize features in the reconstructed sequences. Regions mapping to the chloroplast-derived regions of the mitochondrial were ignored. As before, reconstructed features and inter-feature regions were used in alignments and then concatenated (see below).

***Phylogenetic analyses*—**Individual features and inter-feature regions were aligned using MAFFT v7.029b (Katoh et al., 2005) on auto settings. Individual alignments were concatenated into a

single alignment that was assessed for presence/absence at each site. For the chloroplast alignment, if a gap was identified at a site, that site was removed. As with indels, we chose to do this to alleviate the effects of missing data on the phylogeny estimation. The final, un-gapped alignment was 86,843 bp for all 45 samples. The chloroplast phylogeny was estimated using RAxML v8.0.22 (Stamatakis, 2006) under the GTR+Γ model of evolution with 500 bootstrap replicates.

Mitochondrial sequences were processed in a similar fashion to the plastid sequences. The individual feature and inter-feature regions were aligned using MAFFT and concatenated. The mitochondrial sequence data was missing a larger percentage of the total mitochondrial genome, on average, across the 45 sample set. We noted that the sequence-captured samples were extremely low in their total representation of the mitochondrial genome. Filtering out all gapped sites (as above) resulted in a very short alignment of 4,875 bp. After removal of the sequence-captured samples and the WGS sample CG, which demonstrated a low coverage of the mitochondrial genome, the un-gapped alignment of the remaining 28 accessions totaled 402,717 bp. The mitochondrial phylogeny was estimated as above using the reduced 28 sample set.

We included CS-4-3 (*M. luteus*), CG-1-1 (*M. guttatus*), first generation lab generated triploid hybrid HOO3b (CG-1-1 x CS-4-3), and lab generated allohexaploid SYN (13-SYN-1-JP1) to verify that this approach could determine maternal ancestry in a controlled hybridization event. In addition, we also included two *M. luteus* individuals (ML1l and ML1l2) derived from the same inbred line but genotyped using different methods (WGS and SC, respectively).

Haplotype networks— We estimated matrilineal haplotype networks for both chloroplast and mitochondria using statistical parsimony (TCS) implemented in PopArt (http://popart.otago.ac.nz). For this analysis we used a panel of 929 SNPs for the chloroplast, and 1,454 for the mitochondria. Only variable sites were included in the SNP matrix. Indels and sites filtered as described in above section (*Filtering of genomic data for phylogenetic analysis*) were excluded. For chloroplast, we included both whole-genome sequence (WGS) and sequence capture (SC) samples, and for mitochondria we excluded the SC samples as above. For both data sets we excluded sites with more than 5% missing data. This allowed us to analyze 493 sites of which 161 were parsimony-informative for chloroplast, and 784 sites of which 300 were parsimony-informative for the mitochondria.

Principal component analysis—We also conducted a principal component analysis (PCA) using the SNP data of the chloroplast and mitochondrial genomes using the *glPca* function in the *R* package *adegenet* v. 2.0 (Jombart and Ahmed, 2011). For this analysis we removed genomic regions with low genotyping success, and included only those SNP sites that were successfully genotyped in 90% (41/45) or more individuals. This filtering step removed poorly genotyped genomic regions, and reduced the total number of SNPs from 1454 to 434 for the mitochondrial genome and from 929 to 694 in the chloroplast. For the chloroplast analysis, we also removed a single WGS individual which had poor genotyping success across the genome even after filtering (CG). As expected the amount of missing data across genotyped SNP loci was higher for the SC than for the WGS data set after filtering. For the mitochondria data set, the amount of missing data per site per individuals was 0.0013 for WGS and 0.1080 for SC. For the chloroplast, the same threshold yielded an average amount of missing data per site per individuals of 0.0054 for WGS and 0.0176 for SC.

Mitochondrial genotyping in natural hybrids—We used mitochondrial genome sequences to identify potential markers that could help us distinguish *M. guttatus* and *M. luteu*s in natural hybrid populations. We sought loci segregating for alternative alleles in *M. luteus* (ML11 and CS) and *M.* guttatus (CG and AYR1) in the sequenced panel. Our goal was to identify fragment length polymorphic sites (indels) that could be scored using microcapillary fragment analysis. To identify segregating indels, we sorted BAM alignments in IGV genome browser, and manually screened for indels that were alternatively fixed between *M. luteus* and *M. guttatus*. We identified three indels (3-4 bp) that distinguished *M. guttatus* and *M. luteus*. Genotyping individuals at these indels produced a distinct haplotype with which we could potentially identify the maternal (mitochondrial) parent of hybrid individuals. To facilitate large-scale genotyping, we designed primers flanking each of the indel regions using Primer3Plus (Untergasser et al., 2012), and confirmed their amplification, and sequence in a separate test panel of parental and hybrid individuals using Sanger sequencing. We checked the compatibility of the three primer pairs for pooling them in a single multiplex reaction using Multiplex Manager v 1.2 (Holleley and Geerts, 2009). We then used fluorescently-labelled forward primers (6-FAM, Eurofins Genomics, Germany), and unlabeled reverse primers to generate PCR products in a single multiplex reaction (Type-It, QIAGEN, UK). PCR cycles consisted of a denaturing step of 5 min at 95 ° C, followed by 30 cycles of 95 ° C for 30 s, 55 ° C for 180 s, and 72 ° C for 30 s, and a final elongation step of 30 min at 60 ° C. Fragment length of the PCR products was measured in an ABI3730

sequencer using LIZ500 as a size standard (DNA Sequencing and Services, Dundee, UK). Primer sequences for these mitochondrial markers are given in Appendix S4 (see Supplemental Data with the online version of this article). We genotyped 163 individuals of *M. guttatus* (3 populations), *M. luteus s.l.* (4 populations), *M. x robertsii* (13 populations), and *M. peregrinus* (2 populations), as well as a synthetic allohexaploid line product of an *M. guttatus* (DBL) x *M. luteus* (COL) cross treated with colchicine and selfed for one generation (SYN; Table 2). Between 1 and 15 individuals were genotyped per taxon per population (6.8 ± 3.5, mean ± SD) (Table 3).

RESULTS

***Seed set and germination of artificial hybrids*—**Reciprocal crosses between *M. guttatus* and *M. luteus* resulted in a relatively high number of seeds per fruit in both directions. The average number of seeds per fruit was 723 ± 73 (mean ± SE) for the MG x ML cross, and 635 ± 100 for ML x MG. This difference was not statistically significant (Tukey test, $P = 0.72$; Fig. 1A). Within species crossing yielded 1,237 ± 95, and 467 ± 57 seeds per fruit for *M. guttatus* and *M. luteus*, respectively (Fig. 1A). In comparison, we found a marked effect of the identity of the maternal parent on seed germination. While approximately half of the seeds of the MG x ML cross germinated (0.525 ± 0.062), only 4% of the ML x MG were able to do so (0.042 ± 0.018) (Fig. 1B). Both parental taxa had high germination proportions (0.835 ± 0.037 and 0.825 ± 0.020, for MG x MG, and ML x ML, respectively), although *M. guttatus* seeds achieved maximum germination at a much faster rate (Fig. 1B).

***Floral and vegetative phenotype of F1 hybrids*—**The proportion of hybrid plants that flowered was significantly different depending on crossing direction. The MG x ML cross had a higher proportion of flowering individuals (0.983 ± 0.017, $n = 60$) than the reciprocal cross (0.706 ± 0.114, $n = 17$) (Tukey test $P = 0.021$). The proportion of flowering plants for the MG x ML cross also exceeded both parental crosses (0.633 ± 0.044, $n = 120$; and 0.741 ± 0.040, $n = 120$; for MG x MG and ML x ML, respectively; Tukey tests $P < 0.05$). Moreover, MG x ML hybrids began flowering on average nine days earlier than the reciprocal cross (54.91 ± 0.96 days, 63.25 ± 4.08). Flowering occurred slightly earlier for *M. luteus* (53.08 ± 0.98) than for *M. guttatus* (57.77 ± 1.14).

Among plants that flowered, the MG x ML hybrid achieved the largest stature of all cross types (267 ± 9mm, *vs.* 139.1 ± 12.3, 188.7 ± 5.0, and 138.9 ± 5.4; for ML x MG, MG x MG, and ML x ML, respectively; Fig. 2; Tukey tests $P < 0.05$). Flowers and bracts of both types of hybrids were intermediate or larger than the parental values (Table 4). Both MG x ML and ML x MG hybrids had similar-sized flowers, although hybrids with *M. guttatus* as the maternal parent had longer pedicels and larger bracts (Table 4). Finally, parental taxa differed significantly in nectar production (Tukey test $P < 0.001$), with MG x ML hybrids producing significantly more nectar than MG x MG ($P < 0.001$), while ML x MG produced larger amounts, but not significantly so ($P = 0.092$), than the *M. guttatus* parent (Fig. 2).

Analysis of Mimulus chloroplast—The complete *Mimulus luteus* chloroplast genome was assembled, providing the first published sequence for both the genus *Mimulus* and the family Phrymaceae. The plastome follows the typical quadripartite structure found in most of the photosynthetic angiosperm lineages. The total length of the chloroplast genome is 153,150 bp and is made up of a 84,293 bp LSC, a 17,851 bp SSC, and two 25,503 bp IR regions (Fig. 3). The complete chloroplast sequence is available on Genbank (accession will be available after review). The plastome contains 80 unique protein-coding genes (including *ycf1*, *ycf15*, and *ycf2*), 30 unique tRNAs, and 4 unique rRNAs. No major gene losses or rearrangements were identified relative other published plastomes from Lamiales (Qian et al., 2013; Zhang et al., 2013; Nazereno, Carlsen, and Lohmann 2015.

Genomic analysis of hybrid parentage based on cytoplasmic genomes— On average, sequencing coverage of the cytoplasmic genomes was very high, although also very skewed. For WGS datasets, per individual average read coverage of the mitochondrial and chloroplast genome ranged from 48-940x and 347-5556x with global means of 318 and 2303, respectively. As expected, the sequence capture datasets had high read coverage but totals were considerably lower than the WGS datasets (Supp. Table 2). Per individual average read coverage for SC datasets mapped against the mitochondrial and chloroplasts genomes from 6-13x and 55-112x with means of 9x and 85x, respectively. The chloroplast has considerably greater read depth than the mitochondria. The fraction of genotyped bases varied considerably between the WGS and SC dataset (Appendix S5; see Supplemental Data with the online version of this article). On average, 68% and 80% of chloroplast sites were genotyped in the SC and WGS datasets, respectively (fraction of sites genotyped calculated from entire mitochondrial and chloroplast datasets before

filtering). The mitochondria exhibited larger span between the average percentage of called bases between the SC and WGS datasets with 57% and 85% called.

The chloroplast tree had relatively high resolution and strong support at the taxon level (i.e., *M. guttatus vs. M. luteus*), but was less useful to resolve relationships between populations within species (Fig. 4), suggesting recent divergence of populations. A single monophyletic *M. luteus* clade (minus MLv) is well supported in the chloroplast tree. The mitochondrial dataset does not support *M. luteus* monophyly. Within the chloroplast tree, *M. peregrinus* (allopolyploid) forms two distinct groups and are more closely related to geographically proximate sterile triploid hybrids. *M. x robertsii* sample (LED1), and *M. peregrinus*, LED2, LED3, and LED4, are closely related and all collected form Leadhills, Scotland. *M. peregrinus* sample GON, also falls within the above group and is from nearby Glengonnar, Scotland. Based on the chloroplast data, the other naturally *M. peregrinus* sample included in this analysis, STR, also falls within a clade with geographically proximate *M. x robertsii* (TOR). The relationship of *M. x robertsii* and *M. peregrinus* supports the finding of Vallejo-Marín *et al.* (2015) using a different aspect of the genomic data, i.e., the nuclear component of the genome. Naturally occurring allopolyploids (LED2, LED3 LED4, GON, and STR) as well as naturally occurring hybrids (LED1, TOR, NEN, GIO) are more closely related to *M. guttatus* than they are to *M. luteus* based on chloroplast analysis. The mitochondrial tree topology indicates that naturally occurring allopolyploid LED is nested within a *M. guttatus* clade. The chloroplast tree places MLl1 and MLl2 (samples derived from the same inbred lines but genotyped using different methods) sister to each other. SYN and HOO, triploid hybrid and allohexaploid samples, respectively, derived from an artificial cross between CG x CS, fall within the same clade in both the chloroplast and mitochondrial trees. The maternal parent of these lines (CG) is not sister to the SYN and HOO samples as expected. This is most likely due to homoplasy affecting tree reconstruction due to relatively low sequence divergence between accessions.

The haplotype network of the chloroplast showed that both *M. x robertsii* and *M. peregrinus* samples have identical or nearly identical haplotypes to *M. guttatus* (Fig. 5A). Samples of *M. peregrinus* from Leadhills (LED2, LED3, LED4) and nearby Glengonnar (GON) share a haplotype with *M. x robertsii* from Leadhills (LED1). Samples of *M. peregrinus* from Orkney (STR) belong to a closely related haplotype that also includes *M. x robertsii* from Orkney (TOR), northern England (rob3), and the Outer Hebrides (rob4). This haplotype also includes most of the

M. guttatus samples from the British Isles, the synthetic triploid and hexaploid (HOO and SYN), and one native population (AHQT). Samples of *M. luteus s.l.* from the British Isles (CS and COL, Coldstream; EVI, Evie) fall in a distinct group of haplotypes separated from other *Mimulus* by several mutational steps, and which also includes the two replicates from the *M. luteus var. luteus* inbred line (MLl1 and MLl2; Chile). As in the phylogenetic reconstruction, the sample from a Chilean *M. luteus var. variegatus* (MLv) is situated near *M. guttatus* haplotypes. The mitochondrial network, which excludes all SC samples due to low genotyping success, shows the single individual of wild *M. peregrinus* (LED2) as part of a haplotype that includes most of the British *M. guttatus* as well as the synthetic triploid and hexaploid accessions (HOO and SYN). Both *M. luteus var. luteus* (MLl1) and *M. luteus s.l.* (CS) are separated from all other *Mimulus* by a large number of mutational steps (Fig. 5B).

The PCA of both chloroplast and mitochondria data sets also showed that *M. x robertsii* and *M. peregrinus* cluster together with *M. guttatus* samples (Fig. 6). In the chloroplast PCA, the two known origins of *M. peregrinus* (Leadhills and Orkney) fall in separate clusters with *M. guttatus*, each of them associated with local *M. x robertsii* samples (Fig 6A, Appendix S6). In the mitochondrial PCA, there is not enough resolution to differentiate these two origins, and all *M. peregrinus* and *M. x robertsii* samples fall in the same cluster, along with other native and introduced *M. guttatus* (Fig. 6B, Appendix S7). As above, the *M. luteus var. variegatus* sample from Chile (MLv) is closer to *M. guttatus* samples than to other *M. luteus*.

***Natural frequency of* M. guttatus *as maternal parent*—**We successfully genotyped 163 *Mimulus spp.* individuals from 21 populations and one synthetic line at three mitochondrial loci (Table 3). We recovered only two haplotypes: *237/296/313* and *240/292/317* (fragment size in base pairs for markers mit-259, mit-262, and mit-357, respectively). For *M. luteus*, we genotyped 10 individuals, all of which had haplotype *240/292/317*. For *M. guttatus,* we genotyped 14 individuals from three populations of which all but one had haplotype *237/296/313*. The exceptional individual belonged to population DBL (Table 3). We genotyped 110 individuals of *M. x robertsii* from 12 populations across the U.K. From these, 107 had haplotype *240/292/317* (*"M. guttatus* haplotype"), while three individuals from population NEN had the other haplotype. Finally, the 23 individuals of *M. peregrinus* sampled in populations STR and LED also had the haplotype *240/292/317*. The synthetic allohexaploid line created with *M. guttatus* as the maternal parent, had haplotype *240/292/317* as expected.

DISCUSSION

Our study combines experimental crosses, genomic analyses of chloroplast and mitochondrial genomes, and surveys of natural hybrid populations to demonstrate how a strongly asymmetric post-zygotic hybridization barrier has shaped the origin of the recently formed allopolyploid species *Mimulus peregrinus*, and its triploid ancestor *M. x robertsii* (= *M. guttatus* x *M. luteus*). Thus our study confirms early crossing experiments (Roberts, 1964; Parker, 1975), and adds to a growing list of exceptions to the "rule" of poorly-performing paternal-excess crosses (Table 1; Appendix S1 (see Supplemental Data with the online version of this article)). Specifically, we showed that viable hybrids are considerably more likely to be produced when diploid *M. guttatus* is used as the maternal parent and tetraploid *M. luteus* as the paternal parent (MG x ML), than in the opposite direction (ML x MG). We showed that MG x ML hybrids are larger in size than the parental taxa and the reciprocal hybrid ML x MG. The MG x ML hybrids also flowered earlier and produced more nectar than their *M. guttatus* parents. Motivated by this observed asymmetry of hybridization, we used previously available genomic sequences, and developed novel genomic resources, including the first assembled chloroplast in the genus *Mimulus* and in the family Phrymaceae, to show that most individuals in natural populations of the triploid hybrid *M. x robertsii* have *M. guttatus* as their maternal parent. As expected, *M. peregrinus*, which arose by genome duplication from *M. x robertsii,* also has *M. guttatus* as its maternal parent. Beyond revealing the asymmetric origin of *M. x robertsii* and *M. peregrinus*, our study shows the general potential of exploiting cytoplasmic genome data from both targeted and whole-genome sequence projects (Straub et al., 2012; Dodsworth, 2015), to investigate the evolution and speciation of polyploid and hybrid taxa.

Pattern of hybridization asymmetry: seed viability and germination—Cytonuclear incompatibilities and differences in ploidy level, or a combination of both, can result in post-zygotic asymmetric hybridization (Tiffin, Olson, and Moyle, 2001; Köhler, Scheid, and Erilova, 2010; Scott, Tratt, and Bolbol, 2013). We currently cannot distinguish between these two classes of mechanisms to explain the observed pattern of asymmetric hybridization leading to *M. x robertsii* and *M. peregrinus*. However, analysis of the morphology of hybrid seeds obtained in reciprocal crosses may offer some insight into the mechanisms of such asymmetry. For example, intra-specific hybridization between diploid and tetraploid *Arabidopsis thaliana* has shown that crosses with maternal excess (4x – 2x) yield smaller seed sizes than the reciprocal paternal excess

cross (2x – 4x) (Scott, Tratt, and Bolbol, 2013). The seed size differences are thought to result from misregulation of imprinted genes that control growth and cell division in the endosperm (Köhler and Kradolfer, 2011; Schatlowski and Köhler, 2012; Lafon-Placette and Köhler, 2015). Paternal excess may cause endosperm over-proliferation and failed endosperm cellularization (resulting in large seeds), while maternal excess results in endosperm proliferation and small seeds (Haig and Westoby, 1991; Scott et al., 1998; Bushell, Spielman, and Scott, 2003; Scott, Tratt, and Bolbol, 2013). As a first step to understand the mechanism of reproductive asymmetry in *Mimulus*, we compared seed size in the 16 controlled crosses used in the phenotypic experiment. These crosses included four accessions of each parental taxa and the reciprocal hybrid crosses (Appendix S2; see Supplemental Data with the online version of this article). We estimated seed size as seed area obtained from digital images of 300-600 seeds per cross type (Appendix S8; see Supplemental Data with the online version of this article). Our results indicate that although both types of hybrids have smaller seeds than the parental taxa, the ML x MG cross produces larger (if mostly inviable) seeds than the reciprocal (and more viable) MG x ML cross (Appendix S8; see Supplemental Data with the online version of this article). This suggests that interploidy crosses in *Mimulus* do not mimic the seed phenotypes of interploidy crosses in *Arabidopsis* (Scott, Tratt, and Bolbol, 2013). Further analyses of the morphology and histology of interploidy reciprocal crosses of *M. guttatus* x *M. luteus* are needed to elucidate the mechanistic and genetic basis of post-zygotic hybridization asymmetry in this group.

***Pattern of hybridization asymmetry: adult plant phenotype*—**Most previous experimental studies comparing the viability of asymmetrically produced hybrids have focused on the immediate consequences of intrinsic post-zygotic incompatibilities at the seed and seedling stage (seed number and seed germination). However, as we have shown, surviving hybrids may also differ in their adult phenotypes depending on cross direction. Although our sample size is small, our results show that paternal-excess *Mimulus* triploids possessed characteristics potentially associated with high extrinsic fitness under ecological conditions, including larger plant size. Paternal excess hybrids also flowered earlier, and produced more nectar than maternal-excess triploids, but the direct effects of these traits on hybrid fitness are expected to be negligible given that *M. x robertsii* is thought to be sexually sterile (Roberts, 1964; Stace, 2010). A major gap in our knowledge is whether phenotypic differences among reciprocal hybrids observed in greenhouse conditions, translate to performance differences in the field, and further studies on the ecology of reciprocal hybrids are needed. The combination of *Mimulus* genomic and transgenic

resources (Wu et al., 2008; Hellsten et al., 2013; Yuan et al., 2013; Twyford et al., 2015) with a rich ecological knowledge base, makes the naturalized hybrids and neopolyploids of the British Isles an exciting model for connecting molecular mechanisms of interploidy asymmetry to observed patterns of speciation in the wild.

***Genetic ancestry of* M. x robertsii *and* M. peregrinus—**Regardless of mechanism, our study seems to be an exception to the general pattern in which viable inter-ploidy hybrids are more likely to be produced when the maternal parent is of higher ploidy level than the paternal one (maternal excess; Stebbins, 1957; Ramsey and Schemske, 1998). Both chloroplast and mitochondrial genomic analyses clearly show that the triploid and hexaploid individuals cluster more closely with *M. guttatus* than with any native or introduced *M. luteus* (Figs. 4, 5, 6). The genomic data also shows that even whole-genome sequences are unable to completely separate *M. guttatus* and *M. luteus*, which may reflect incomplete lineage sorting between these relatively closely related taxa (Madison and Knowles, 2006) Moreover, the origin of *M. luteus* itself is unknown, but nuclear genomic data suggests that this taxon has a hybrid origin (Vallejo-Marín et al., 2015), perhaps with a *M. guttatus*-like ancestor (Mukherjee and Vickery, 1962). If the putative allopolyploid *M. luteus* has had multiple origins, then the simple expectation of reciprocal monophyly for the plastid genomes of *M. luteus* and *M. guttatus* may not be met. Solving the mystery of the origin of *M. luteus* and its relationship with *M. guttatus* and related taxa will require detailed analyses of natural populations throughout the native South American range.

Our survey of hybrid *Mimulus* populations across the British Isles using three mitochondrial loci also supports *M. guttatus* as the maternal parent of *M. x robertsii* and *M. peregrinus*. Of 114 triploid *M. x robertsii* from 13 populations, and 27 individuals from three populations of the hexaploid *M. peregrinus* that were genotyped at these three mitochondrial loci, all but three *M. x robertsii,* individuals from a single population had maternal haplotypes that could be confidently assigned to *M. guttatus*. The three individuals of *M. x robertsii* (all from population NEN) that had *M. luteus* mitochondrial haplotypes (Table 3) pose an exception to the general pattern detected here. An *M. luteus*-like maternal haplotype in hybrids could be explained in a number of ways. First, it could simply be due to contamination during genotyping. However, we re-extracted and re-genotyped this samples and confirmed the mitochondrial haplotype. Second, the three samples could have been mistakenly identified as *M. x robertsii* but instead belong to *M. luteus.*

We think this is also unlikely as plants were morphologically identified on site, and we have also conducted surveys of this population (NEN) in 2010, 2011, and 2013, and *M. luteus* has never been observed in this or nearby populations. Third, the haplotypes in these three hybrids could arise via introgression from *M. luteus*. Introgression across ploidy barriers has been shown for nuclear markers in other species (Chapman and Abbott, 2010), and could possibly occur at the mitochondrial level as well. Although, we cannot rule out this possibility, the lack of local *M. luteus* (the nearest known *M. luteus* natural population is ~100km north), and principally the fact that hybrids are not known to set seed (i.e., are almost completely sterile; Roberts, 1964; Vallejo-Marín, 2012), makes the introgression scenario unlikely. Fourth, the "*M. luteus*" haplotype in hybrids could be explained by incomplete lineage sorting of ancestral polymorphisms between *M. guttatus* and *M. luteus* (Twyford and Ennos, 2012). We cannot currently rule out ancestral polymorphism, and this remains a distinct possibility given the observations that a single individual of *M. guttatus* (DBL) bears a mitochondrial haplotype otherwise characteristic of *M. luteus* (Table 2). Similarly the mitochondrial haplotype for the British individual of *M. luteus s.l.* (CS) nested within the *M. guttatus* clade also suggests incomplete lineage sorting. Fifth, it is possible that homoplasy resulting from the sequences converging on a similar haplotype (especially if the sequences are not that distinct) could explain this pattern. Finally, a simpler explanation may be that although the majority of hybrids are produced when *M. guttatus* is the maternal parent, occasionally, the opposite cross can yield viable offspring. Indeed, our experimental results show that viable hybrids in the ML x MG direction are produced albeit with low probability (Figure 1).

Skimming of cytoplasmic genomes — Acquisition of complete organellar genomes is increasingly feasible with the advent of "genome skimming" techniques associated with next-generation sequencing projects. In fact, mitochondrial sequences usually represent 1-5% of the reads obtained in such projects (Steele, et al., 2012; Smith, 2015) and are often discarded as contamination during nuclear genomic analyses. Our study showed that data generated by whole-genome sequencing contain enough cytoplasmic 'contamination' to yield between 1,515 and 208-fold mean coverage of the chloroplast and mitochondria, respectively. Even the approach of sequence capture, in which specific (nuclear) regions were targeted (Vallejo-Marín et al., 2015) contained sufficient reads to build large contigs of both chloroplast and mitochondrial genes (86,843 bp and 402,717 bp, respectively). As expected, however, the sequence capture data did not allow us to genotype all the same sites in all individuals. The limited coverage of specific

sites was particularly acute for the mitochondrial genome, which reduced its utility in some of the analyses conducted here. Our results suggest that using off-targeted reads from sequence capture approaches may be sufficient to generate moderate numbers of genetic markers (e.g., Fig. 6), but this approach is of limited utility when coverage of large numbers of sites across multiple individuals is required (e.g., Figs. 4, 5).

The* Mimulus *chloroplast genome —Shallow whole-genome sequencing (also referred to as genome skimming) allowed us to reconstruct the chloroplast genome of *M. luteus*. The structure and gene content of this genome follows what has been seen for other taxa in Lamiales (Nazareno et al., 2015), including that of *Utricularia gibba* (Ibarra-Laclette et al., 2013), a species known for a diminutive nuclear genome. The relatively constant architecture and gene content identified across multiple lineages of the order Lamiales suggests that there may be selection for maintenance of gene content and structure relative to other eudicots lineages like saguaro (Sanderson et al., 2015) and legumes (Saski et al., 2005). A more comprehensive taxon sampling of chloroplast genomes is needed throughout the Lamiales and all angiosperm to truly understand the dynamics and evolution of this vital organelle.

The use of genome skimming has repeatedly proven to be a useful technique for acquiring either complete gene sets (e.g. Washburn et al., 2015) or full plastomes (e.g. Bock et al., 2013). Here we demonstrate the utility of a single plastome and mitochondrial genome as anchors to identify organellar genome sequences rapidly from genome skimming data. Though full plastid and mitochondrial genome sequences were not recovered, a large proportion of each genome was and those regions were identified to annotated features and alignable providing resolution of relationships at the population level. Future work in the group will include *de novo* assembly of whole chloroplast genomes across the genus providing valuable resources for population-level studies of hybridization and polyploidy.

Conclusions —The value of genome skimming for general phylogenetic analyses is well recognized (e.g. Hahn, Bachmann, and Chevreux, 2013). However, the utility of discarded reads for analyses of recent hybridization events has only recently been exploited (Bock et al., 2013). We demonstrate that the analysis of the entire mitochondrial and chloroplast genomes can yield information to infer the maternal ancestry of a hybrid and neo-allopolyploid derived from recently diverged taxa (Beardsley and Olmstead, 2002; Nie et al., 2006). However, our study also shows that even whole genome sequences might only imperfectly resolve phylogenetic

relationships between closely related species. Therefore, genomic analysis of hybrid ancestry can be significantly strengthened by experimental crossing data such as the one conducted here. The ongoing proliferation of genomic sequencing opens the door for low-cost genome-skimming based analyses across hybrid species, including many ecologically fascinating yet non-model organisms. Exploration of reproductive isolation asymmetry across the plant kingdom will be invaluable in gaining a better understanding in the origin of hybrid and allopolyploid species. Connecting the genomic and molecular underpinnings of reproductive asymmetry to the morphological and ecological consequences of asymmetric barriers will ultimately generate a better solution to the puzzle of allopolyploids' evolutionary success.

ACKNOWLEDGEMENTS

We thank H.E. O'Brien for comments and suggestions on haplotype analysis. This work was supported in part by a UK Natural Environment Research Council grant (NERC, NE/J012645/1) and a Carnegie Trust Travel grant to MVM and a Suzann Wilson Matthews Research Award to JRP.

REFERENCES

AUWERA, G. A., M. O. CARNEIRO, C. HARTL, R. POPLIN, G. DEL ANGEL, A. LEVY-MOONSHINE, T. JORDAN, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*: 11.10. 11-11.10. 33.

BANKEVICH, A., S. NURK, D. ANTIPOV, A. A. GUREVICH, M. DVORKIN, A. S. KULIKOV, V. M. LESIN, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455-477.

BARKER, M. S., N. ARRIGO, A. E. BANIAGA, Z. LI, AND D. A. LEVIN. 2015. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*: In press.

BEARDSLEY, P. M., AND R. G. OLMSTEAD. 2002. Redefining Phrymaceae: The placement of *Mimulus*, tribe Mimuleae and Phryma. *American Journal of Botany* 89: 1093-1102.

BEHREND, A., A. GLUSCHAK, A. PRZYBYLA, AND A. HOHE. 2015. Interploid crosses in heather (*Calluna vulgaris*). *Scientia Horticulturae* 181: 162-167.

BOCK, D. G., N. C. KANE, D. P. EBERT, AND L. H. RIESEBERG. 2013. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. *New Phytologist* 201: 1021-1030.

BOLGER, A. M., M. LOHSE, AND B. USADEL. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*: btu170.

BURTON, T. L., AND B. C. HUSBAND. 2000. Fitness differences among diploids, tetraploids, and their triploid progeny in *Chamerion angustifolium*: mechanisms of inviability and implications for polyploid evolution. *Evolution* 54: 1182-1191.

BUSHELL, C., M. SPIELMAN, AND R. J. SCOTT. 2003. The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *The Plant Cell* 15: 1430-1442.

CASTRO, S., Z. MÜNZBERGOVÁ, J. RAABOVÁ, AND J. LOUREIRO. 2011. Breeding barriers at a diploid–hexaploid contact zone in *Aster amellus*. *Evolutionary Ecology* 25: 795-814.

CHAPMAN, M. A., AND R. J. ABBOTT. 2010. Introgression of fitness genes across a ploidy barrier. *New Phytologist* 186: 63-71.

CHEUNG, K. W., F. M. RAZEQ, C. A. SAUDER, T. JAMES, AND S. L. MARTIN. 2015. Bidirectional but asymmetrical sexual hybridization between *Brassica carinata* and *Sinapis arvensis* (Brassicaceae). *Journal of plant research* 128: 469-480.

COMAI, L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* 6: 836-846.

COOLEY, A. M., AND J. H. WILLIS. 2009. Genetic divergence causes parallel evolution of flower color in Chilean *Mimulus*. *New Phytologist* 183: 729-739.

DEPRISTO, M. A., E. BANKS, R. POPLIN, K. V. GARIMELLA, J. R. MAGUIRE, C. HARTL, A. A. PHILIPPAKIS, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491-498.

DINU, I., R. HAYES, R. KYNAST, R. PHILLIPS, AND C. THILL. 2005. Novel inter-series hybrids in *Solanum*, section Petota. *Theoretical and applied genetics* 110: 403-415.

DODSWORTH, S. 2015. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science* 20: 525-527.

DOYLE, J. J., AND J. L. DOYLE. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

EDGER, P., K. CHILDS, A. PLATTS, C. PIRES, A. M. COOLEY, AND J. PUZEY. in prep. Transposon density contributes to homeolog specific gene expression in allopolyploid monkeyflower (*Mimulus luteus*)

GRANT, A. L. 1924. A monograph of the genus *Mimulus*. *Annals of the Missouri Botanical Garden* 11: 99-389.

GREINER, R., AND C. OBERPRIELER. 2012. The role of inter-ploidy block for reproductive isolation of the diploid *Leucanthemum pluriflorum* Pau (Compositae, Anthemideae) and its tetra-and hexaploid relatives. *Flora-Morphology, Distribution, Functional Ecology of Plants* 207: 629-635.

HAHN, C., L. BACHMANN, AND B. CHEVREUX. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* 41: e129-e129.

HAIDER, N. 2011. Chloroplast-specific universal primers and their uses in plant studies. *Biologia Plantarum* 55: 225-236.

HAIG, D., AND M. WESTOBY. 1989. Parent-specific gene expression and the triploid endosperm. *American Naturalist*: 147-155.

HAIG, D., AND M. WESTOBY. 1991. Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philosophical Transactions of the Royal Society of London Series B* 333: 1-13.

HELLSTEN, U., K. WRIGHT, J. JENKINS, S. SHU, Y. YUAN, S. WESSLER, J. SCHMUTZ, et al. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 110: 19478-19482.

HERSCH-GREEN, E. I. 2012. Polyploidy in Indian paintbrush (*Castilleja*; Orobanchaceae) species shapes but does not prevent gene flow across species boundaries. *American Journal of Botany* 99: 1680-1690.

HOLLELEY, C. E., AND P. G. GEERTS. 2009. Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques* 46: 511-517.

HOTHORN, T., F. BRETZ, AND P. WESTFALL. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50: 346-363.

http://www.genecodes.com. Sequencher sequence analysis software, version 5.0.1. website: http://www.genecodes.com.

HUSBAND, B. C. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* 82: 537-546.

IBARRA-LACLETTE, E., E. LYONS, G. HERNÁNDES-GUZMÁN, ET AL. 2013. Architecture and evolution of a minute plant genome. *Nature* 498: 94-98.

JOMBART, T., AND I. AHMED. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070-3071.

KATOH, K., K.-I. KUMA, H. TOH, AND T. MIYATA. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research* 33: 511-518.

KELLY, A. J., AND J. H. WILLIS. 1998. Polymorphic microsatellite loci in *Mimulus guttatus* and related species. *Molecular Ecology* 7: 769-774.

KÖHLER, C., AND D. KRADOLFER. 2011. Epigenetic mechanisms in the endosperm and their consequences for the evolution of flowering plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1809: 438-443.

KÖHLER, C., O. M. SCHEID, AND A. ERILOVA. 2010. The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* 26: 142-148.

KRZYWINSKI, M., J. SCHEIN, I. BIROL, J. CONNORS, R. GASCOYNE, D. HORSMAN, S. J. JONES, AND M. A. MARRA. 2009. Circos: an information aesthetic for comparative genomics. *Genome research* 19: 1639-1645.

LAFON-PLACETTE, C., AND C. KÖHLER. 2015. Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Current Opinion in Plant Biology* 23: 39-44.

LANGMEAD, B., AND S. L. SALZBERG. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9: 357-359.

LEVIN, D. A. 1975. Minority cytotype exclusion in local plant populations. *Taxon* 24: 35-43.

LEVIN, D. A. 2002. The role of chromosomal change in plant evolution. Oxford University Press on Demand.

MADDISON, W. P., AND L. L. KNOWLES. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21-30.

MALLET, J. 2007. Hybrid speciation. *Nature* 446: 279-283.

MARÇAIS, G., AND C. KINGSFORD. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764-770.

MCKENNA, A., M. HANNA, E. BANKS, A. SIVACHENKO, K. CIBULSKIS, A. KERNYTSKY, K. GARIMELLA, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297-1303.

MIYASHITA, T., AND Y. HOSHINO. 2015. Interploid and intraploid hybridizations to produce polyploid Haskap (*Lonicera caerulea* var. *emphyllocalyx*) plants. *Euphytica* 201: 15-27.

MOWER, J. P., A. L. CASE, E. R. FLORO, AND J. H. WILLIS. 2012. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome biology and evolution* 4: 670-686.

MUKHERJEE, B. B., AND R. K. VICKERY. 1962. Chromosome counts in the section *Simiolus* of the genus *Mimulus* (Scrophulariaceae). V. The chromosomal homologies of *M. guttatus* and its allied species and varieties. *Madroño* 16: 141-172.

NAZARENO, A. G., M. CARLSEN, AND L. G. LOHMANN. 2015. Complete chloroplast genome of *Tanaecium tetragonolobum*: The first Bignoniaceae plastome. *PloS ONE* 10: e0129930.

NGHIEM, Q., J. HARBARD, C. HARWOOD, A. GRIFFIN, T. HA, AND A. KOUTOULIS. 2013. Pollen-pistil interactions between autotetraploid and diploid *Acacia mangium* and diploid *A. auriculiformis*. *Journal of Tropical Forest Science* 25: 96-110.

NIE, Z., H. SUN, P. BEARDSLEY, R. OLMSTEAD, AND J. WEN. 2006. Evolution of biogeographic disjunction between eastern Asia and eastern North America in *Phryma* (Phrymaceae). *American Journal of Botany* 93: 1343-1356.

OTTO, S. P., AND J. WHITTON. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401-437.

PARKER, P. F. 1975. *Mimulus* in Great Britain - A cytotaxonomic note. *New Phytologist* 74: 155-160.

PRESTON, C. D., D. A. PEARMAN, AND T. D. DINES. 2002. New atlas of the British and Irish flora. An atlas of the vascular plants of Britain, Ireland, the Isle of Man and the Channel Islands. Oxford University Press.

PUZEY, J., AND M. VALLEJO-MARÍN. 2014. Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Molecular ecology* 23: 4472-4485.

QIAN, J., J. SONG, H. GAO, Y. ZHU, J. XU, X. PANG, H. YAO, et al. 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One* 8: e57607.

R CORE TEAM. 2015. R: A language and environment for statistical computing. website: http://www.R-project.org/.

RAMSEY, J., AND D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29: 467-501.

RIESEBERG, L. H. 1997. Hybrid origins of plant species. *Annual review of ecology and systematics*: 359-389.

RIESEBERG, L. H., AND J. H. WILLIS. 2007. Plant speciation. *Science* 317: 910-914.

ROBERTS, R. 1964. *Mimulus* hybrids in Britain. *Watsonia* 6: 70-75.

ROCCAFORTE, K., S. E. RUSSO, AND D. PILSON. 2015. Hybridization and reproductive isolation between diploid *Erythronium mesochoreum* and its tetraploid congener *E. albidum* (Liliaceae). *Evolution*.

SABARA, H. A., P. KRON, AND B. C. HUSBAND. 2013. Cytotype coexistence leads to triploid hybrid production in a diploid–tetraploid contact zone of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* 100: 962-970.

SANDERSON, M. J., D. COPETTI, A. BÚRQUEZ ET AL. 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. *American Journal of Botany* 102: 1-13.

SASKI, C. S-B. LEE, H. DANIELL ET AL. 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Molecular Biology* 59: 309-322.

SCHATLOWSKI, N., AND C. KÖHLER. 2012. Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. *Journal of experimental botany* 63: 6059-6067.

SCOTT, R. J., J. L. TRATT, AND A. BOLBOL. 2013. Seed development in interploidy hybrids. *Polyploid and Hybrid Genomics*: 271-290.

SCOTT, R. J., M. SPIELMAN, J. BAILEY, AND H. G. DICKINSON. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125: 3329-3341.

SEKINE, D., T. OHNISHI, H. FURUUMI, A. ONO, T. YAMADA, N. KURATA, AND T. KINOSHITA. 2013. Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *The Plant Journal* 76: 792-799.

SILVERSIDE, A. J. 1990. A new binomial in *Mimulus*. *Watsonia* 18: 210-212.

------. 1998. *Mimulus* Section *Simiolus*. *In* T. Rich AND A. Jermy [eds.], Plant Crib, 259-261. Botanical Society of the British Isles, London.

SMITH, D. R. 2015. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Briefings in functional genomics*: elv027.

SOLTIS, D. E., M. C. SEGOVIA- SALCEDO, I. JORDON-THADEN, L.MAJURE, N. M. MILES, E. V. MAVRODIEV, W. MEI, et al. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytologist* 202: 1105-1117.

STACE, C.A. 2010. New flora of the British Isles. Cambridge University Press.

STACE, C. A., C. D. PRESTON, AND D.A. PEARMAN. 2015. Hybrid Flora of the British Isles. BSBI Publications.

STAMATAKIS, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.

STEBBINS, G. L. 1957. Self fertilization and population variability in higher plants. *American Naturalist* 91: 337-354.

STEELE, P. R., K. L. HERTWECK, D. MAYFIELD, M. R. MCKAIN, J. LEEBENS-MACK, AND J. C. PIRES. 2012. Quality and quantity of data recovered from massively parallel sequencing: examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330-348.

STOUTE, A. I., V. VARENKO, G. J. KING, R. J. SCOTT, AND S. KURUP. 2012. Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. *The Plant Journal* 71: 503-516.

STRAUB, S. C., M. PARKS, K. WEITEMIER, M. FISHBEIN, R. C. CRONN, AND A. LISTON. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349-364.

TIFFIN, P., S. OLSON, AND L. C. MOYLE. 2001. Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society of London B: Biological Sciences* 268: 861-867.

TOONEN, R. J., AND S. HUGHES. 2001. Increased throughput for fragment analysis on an ABI Prism® 377 automated sequencer using a membrane comb and STRand software. *BioTechniques* 31: 1320-1325.

TWYFORD, A., AND R. ENNOS. 2012. Next-generation hybridization and introgression. *Heredity* 108: 179-189.

TWYFORD, A. D., M. A. STREISFELD, D. B. LOWRY, AND J. FRIEDMAN. 2015. Genomic studies on the nature of species: adaptation and speciation in *Mimulus*. *Molecular Ecology* 24: 2601-2609.

UNTERGASSER, A., I. CUTCUTACHE, T. KORESSAAR, J. YE, B. C. FAIRCLOTH, M. REMM, AND S. G. ROZEN. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40: e115.

VALLEJO-MARÍN, M. 2012. *Mimulus peregrinus* (Phrymaceae): A new British allopolyploid species. *Phytokeys* 14: 1-14.

VALLEJO-MARÍN, M., AND G. C. LYE. 2013. Hybridisation and genetic diversity in introduced *Mimulus* (Phrymaceae). *Heredity* 110: 111-122.

VALLEJO-MARÍN, M., R. J. BUGGS, A. M. COOLEY, AND J. R. PUZEY. 2015. Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. *Evolution* 69: 1487-1500.

VAN LAERE, K., D. HERMANS, L. LEUS, AND J. VAN HUYLENBROECK. 2015. Interspecific hybridisation within *Buxus* spp. *Scientia Horticulturae* 185: 139-144.

VON BOHLEN, C. 1995. El género *Mimulus* L. (Scrophulariaceae) en Chile. *Gayana Botanica* 52: 7-28.

WASHBURN, J. D., J. C. SCHNABLE, G. DAVIDSE, AND J. C. PIRES. 2015. Phylogeny and photosynthesis of the grass tribe Paniceae. *American Journal of Botany* 102: 1493-1505.

WATKINS, A. 1932. Hybrid sterility and incompatibility. *J. Genet* 25: 125-162.

WATSON, J. M. 1989. 'Andean Nymph,' a pink *Mimulus*. Thompson & Morgan, Ipswich.

WEITEMIER, K., S. C. STRAUB, R. C. CRONN, M. FISHBEIN, R. SCHMICKL, A. MCDONNELL, AND A. LISTON. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2.

WU, C. A., D. B. LOWRY, A. M. COOLEY, K. M. WRIGHT, Y.-W. LEE, AND J. H. WILLIS. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 101: 220-230.

WYMAN, S. K., R. K. JANSEN, AND J. L. BOORE. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252-3255.

YUAN, Y.-W., J. M. SAGAWA, R. C. YOUNG, B. J. CHRISTENSEN, AND H. D. BRADSHAW. 2013. Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*. *Genetics* 194: 255-263.

ZHANG, H, C. LI, H. MIAO, AND S. XIONG. 2013. Insights form the complete chloroplast genome in the evolution of *Sesamum indicum* L. *PloS ONE* 8: e80508.

FIGURE LEGENDS

Figure 1. (A) Mean seed number and 95% confidence intervals (CI) obtained in crosses within species (MG x MG, and ML x ML), and between species (MG x ML and ML x MG). Sample sizes: 9, 11, 10, 10 fruits. **(B)** Mean germination of *M. guttatus* x *M. guttatus* (black circle), *M. luteus* x *M. luteus* (black square), *M. guttatus* x *M. luteus* (gray circle), *M. luteus* x *M. guttatus* (gray square). Crosses are indicated as maternal x paternal parent. Four lines for each cross type were analyzed, planting 100 seeds per line (1600 seeds total). Each data point represents the average (+/- SE) proportion of germinated seeds across the four lines for a given cross type across the germination period.

Figure 2. Plant height at first flower **(A),** and nectar content per flower (**B**) of *Mimulus* crosses within- (MG x MG and ML x ML) and between-species (MG x ML and ML x MG). Sample size (number of individuals) for plant height: n = 72, 54, 11, 78; sample size (number of flowers): n = 67, 73, 14, 73. Dotted line shows the mid-parent value. Bars are 95% confidence intervals.

Figure 3. Circular plot of *Mimulus luteus* var. *luteus* chloroplast genome. The plastome is 153,150 base pairs with a 84,293 base pair large single copy (LSC) region, a 17,851 base pair small single copy (SSC) region, and two 25,503 base pair inverted repeat regions (IRB and IRA). Photosystem genes are in green, RNA polymerase subunits are in orange, ribosomal genes are in black, other protein encoding genes are in yellow, rRNAs are in blue, and tRNAs are in black.

Figure 4. Maximum likelihood tree of *Mimulus* species and populations based on alignments of 86,843 base pairs from chloroplast genomes (A) and 402,717 base pairs from mitochondrial genomes (B). Bootstrap values are indicated. Nodes with less than 50% bootstrap support were collapsed. Samples genotyped using sequence-capture (N=16) are excluded from the mitochondrial analysis due to low genotyping success. Trees are rooted with *M. dentilobus*, Color-coding is based on species identity.

Figure 5. Matrilineal haplotype network (statistical parsimony; TCS) of (A) chloroplast and (B) mitochondrial SNPs. The chloroplast analysis is based on 493 sites, and the mitochondrial on 784 sites. Sites with more than 5% missing data were excluded. The mitochondrial dataset excludes 16 samples obtain through sequence capture due to low genotyping success (see Results). Names for individual samples are only shown here for *M. x robertsii* and *M. peregrinus*.

Figure 6. Principal component analysis of (A) 929 SNPs from the chloroplast and (B) 1454 SNPs from the mitochondria. Only sites successfully genotyped in 41/45 samples were included in this analysis. The chloroplast dataset excludes one individual (CG) with low genotyping success for the analyzed SNPs. The mitochondrial dataset excludes 16 samples obtain through sequence capture due to low genotyping success (see Results). A fully annotated version of this Figure is available as Appendix S6 and Appendix S7.

Table 1. Recent (1998-2015) reports of asymmetric crossing success following interploidy crosses. Some of the studies listed below verified that the seeds or seedlings produced were triploid; others did not and could potentially include diploid, tetraploid, or aneuploid offspring in their measure of success.

Genus	Reference	Ploidy Type	Superior Direction	Measure of Success
Brassica	Stoute et al., 2012	auto (a)	m	% seed germination; all seedlings tested were triploid
Brassica x *Sinapsis*	Cheung et al., 2015	allo	m	triploid seedlings per pollination, relative to intraploidy crosses
Buxus	Van Laere et al., 2015	unknown	m	seedlings per pollination
Calluna	Behrend et al., 2015	auto (n)	m	triploid seedlings per pollination
Erythronium	Roccaforte, Russo, and Pilson, 2015	auto (n)	m	seeds per pollination relative to intraploidy crosses
Lonicera	Miyashita and Hoshino, 2015	unknown	m*	triploid seedlings per pollination
Oryza (rice)	Sekine et al., 2013	auto (u)	m	% seed germination
Acacia	Nghiem et al., 2013	auto (c)	p	filled fruits per pollination and filled seeds per fruit, relative to intraploidy crosses

				Asymmetric hybridization during allopolyploid speciation
Aster	Castro et al., 2011	autohexaploid (n)	p	seeds per pollination, % seed germination, triploid seedlings per pollination, relative to intraploidy crosses
Castilleja	Hersch-Green, 2012	unknown naturally ocurring 2x, 4x, and 8x	p	filled fruits per pollination and seeds per fruit, relative to intraploidy crosses
Chamerion	Sabara, Kron, and Husband, 2013	unknown naturally occurring 2x and 4x	p	observed vs expected triploid production from wild-collected maternal seed families
Leucanthemum	Greiner and Oberprieler, 2012	allo	p	seeds per pollination relative to intraploidy crosses
Solanum	Dinu et al., 2005	auto (u)	p*	seed development and % seed germination

Notes: Ploidy type: a, derived from anther culture; c, colchicine-induced; n, naturally-occurring; u, type of autotetraploid unknown. *Superior Direction:* m, crosses more successful when maternal parent has higher ploidy; p, crosses more successful when paternal parent has higher ploidy.

*with embryo rescue

Table 2. Sample names and sampling locations for *Mimulus* accessions used in this study. Data type: WGS = whole-genome sequencing; SC = sequence capture.

Taxon	Population	Sample ID	Short Name	Location	Latitude	Longitude	Data Type*	Accession Number
***M. guttatus* (2x)**	AYR	10-AYR-10	AYR1	Ayr, Ayrshire, Scotland	55.461	−4.625	WGS	SRR1462346
	AYR	10-AYR-18-1	AYR2	Ayr, Ayrshire, Scotland	55.461	−4.625	SC	dryad.gg273
	CER	10-CER-10	CER1	Cerrigydrudion, Wales	53.006	−3.549	WGS	SRR1475232
	CER	10-CER-25-1	CER2	Cerrigydrudion, Wales	53.006	−3.549	SC	dryad.gg273
	CRO	12-CRO-5	CRO	Crowan, Cornwall	50.163	−5.293	WGS	SRR1482405
	DBL	10-DBL-20	DBL1	Dunblane, Scotland	56.197	−3.965	WGS	SRR1475385
	DBL	CG-1-1-JP1	CG	Dunblane, Scotland	56.197	−3.965	WGS	this study
	DBL	11-DBL-1-1	DBL2	Dunblane, Scotland	56.187	−3.965	SC	dryad.gg273
	HOU	10-HOU-17	HOU1	Houghton Lodge, England	51.097	−1.508	WGS	SRR1481643
	HOU	10-HOU-23-1	HOU2	Houghton Lodge, England	51.097	−1.508	SC	dryad.gg273
	PAC	12-PAC-39	PAC	Packhorse Bridge, Scotland	57.355	−3.336	WGS	SRR1482406
	QUA	10-QUA-47	QUA	Quarff, Shetland Islands	60.105	−1.227	WGS	SRR1481644
	TOM	10-TOM-23	TOM	Tomintoul, Scotland	57.255	−3.368	WGS	SRR1482404
	TRE	12-TRE-17	TRE	Tremar Coombe, England	50.498	−4.465	WGS	SRR1482409
	TSG	TSG	TSG	Haida Gwaii, Canada	53.419	−131.916	WGS	SRS120816
	VIC	12-VIC-18	VIC	Victoria Bridge, Northern Ireland	54.763	−7.454	WGS	SRR1482407
	AHQ	AHQT	AHQT	Lonestar Basin Thermal Spring, Teton Co., WY	44.431	−110.813	WGS	SRR486613
	BOG	BOG	BOG	Bog Hot Springs, Humboldt, Co., NV	41.924	−118.804	WGS	SRR071996
	LMC	LMC	LMC	Yorkville, Mendocino Co., CA	38.864	−123.084	WGS	SRR072031
	MAR	MAR	MAR	Marshanne Landing, Douglas Co., OR	43.479	−123.294	WGS	SRR071967
	PED	PED	PED	San Pedro River, Pinal Co., AZ	32.711	−110.628	WGS	SRR071969
	REM	REM	REM	Rumsey, Yolo Co., CA	38.859	−122.410	WGS	SRR071971
	SWB	SWB	SWB	Sperm Whale Beach, Mendocino Co., CA	39.036	−123.691	WGS	SRR072030

	YJS	YJS	YJS	Yellowjacket creek, Lemhi Co., ID	44.951	−114.585	WGS	SRR071970
***M. guttatus* complex**								
M. platycalyx	CVP	CVP	CVP	Coleman Valley Road, Sonoma Co., CA	38.372	−123.055	WGS	SRR486607
M. micranthus	EBR	EBR	EBR	Branscomb, Mendocino Co., CA	39.631	−123.532	WGS	SRR486609
M. nasutus	SF	SF	SF	Sherars Falls, Wasco Co., OR	45.635	−120.914	WGS	SRR400478
***M. luteus s.l.* (4x)**	COL	CS-4-JP1	CS	Coldstream, Scotland	55.655	−2.240	WGS	this study
	COL	10-COL-20e-1	COL	Coldstream, Scotland	55.655	−2.240	SC	dryad.gg273
	EVI	13-EVI-5	EVI	Evie, Scotland	59.112	−3.108	SC	dryad.gg273
M. l. luteus	EY	MLl-EY	MLl1	El Yeso, Chile	−33.4	−70.0	WGS	this study
M. l. luteus	EY	MLl-EY-7x4-1-3	MLl2	El Yeso, Chile	−33.4	−70.0	SC	dryad.gg273
M. l. variegatus	RC	MLv-RC-5x4-4-1	MLv	Río Cipreses, Chile	−34.2	−70.3	SC	dryad.gg273
***M. robertsii* (3x)**	DBL x COL (3x)	H003b-JP1	HOO	Artificial triploid (CG-1-1 x CS-4-3)			WGS	this study
	GIO	13-GIO-13	GIO	Giordail Beach, Scotland	58.337	-6.202	SC	dryad.gg273
	LED	12-LED-3-11	LED1	Leadhills, Scotland	55.424	−3.735	SC	dryad.gg273
	NEN	12-NEN-41	NEN	Nenthall, England	54.806	−2.376	SC	dryad.gg273
	TOR	13-TOR-13	TOR	Tormiston, Scotland	58.996	−3.183	SC	dryad.gg273
***M. peregrinus* (6x)**	DBL x COL (6x)	13-SYN-1-JP1	SYN	Artificial hexaploid (CG-1-1 x CS-4-3)			WGS	this study
	GON	12-GON-106	GON	Glen Gonnar, Scotland	55.467	3.738	SC	dryad.gg273
	LED	11-LED-seed-1-JP1	LED2	Leadhills, Scotland	55.423	−3.735	WGS	this study
	LED	11-LED-seed-30	LED3	Leadhills, Scotland	55.423	−3.735	SC	dryad.gg273
	LED	11-LED-seed-2-11	LED4	Leadhills, Scotland	55.424	−3.735	SC	dryad.gg273
	STR	13-STR-32	STR	Stromness, Scotland	58.969	−3.283	SC	dryad.gg273
***M. dentilobus* (outgroup)**	DENT	DENT	DENT	NA	NA	NA	WGS	SRX030541

*Note: *Data type*: WGS, Whole Genome Sequencing; SC, Sequence Capture.

Table 3. Mitochondrial haplotypes of 163 individuals of *M. guttatus, M. luteus,* and their hybrid derivatives, the triploid *M. x robertsii* and allopolyploid *M. peregrinus*. Haplotypes are given as the size in base pairs of three mitochondrial loci with fragment-length polymorphism (indels). Between 1 and 15 individuals of each of 21 populations, and one artificially created allopolyploid, were genotyped at all three loci. The synthetic allopolyploid was the product of a *M. guttatus* (DBL) x *M. luteus s.l.* (COL) cross, with *M. guttatus* as the maternal parent. A single individual phenotyped as *M. guttatus* (DBL), but with a "*M. luteus"* maternal haplotype is presented separately.

Taxon	Population	# Samples	Haplotype	Inferred maternal parent
M. guttatus	COL	5	237/296/313	*M. guttatus*
	DBL	5	237/296/313	*M. guttatus*
	TOM	3	237/296/313	*M. guttatus*
M. luteus s.l.	COL	5	**240/292/317**	***M. luteus***
	EVI	3	**240/292/317**	***M. luteus***
M. luteus var. luteus	EY, Chile	1	**240/292/317**	***M. luteus***
M. luteus var. variegatus	RC, Chile	1	**240/292/317**	***M. luteus***
M. x robertsii	BCH	6	237/296/313	*M. guttatus*
	EOR	9	237/296/313	*M. guttatus*
	EVI	1	237/296/313	*M. guttatus*
	GIO	8	237/296/313	*M. guttatus*
	GON	9	237/296/313	*M. guttatus*
	GOO	10	237/296/313	*M. guttatus*
	GRO	10	237/296/313	*M. guttatus*
	NEN	9	237/296/313	*M. guttatus*
	NEN	3	**240/292/317**	***M. luteus***
	NOR	9	237/296/313	*M. guttatus*
	PLY	7	237/296/313	*M. guttatus*
	SCA	9	237/296/313	*M. guttatus*
	TAR	10	237/296/313	*M. guttatus*
	TOR	10	237/296/313	*M. guttatus*
M. peregrinus	LED	15	237/296/313	*M. guttatus*

	STR	8	237/296/313	*M. guttatus*
	Synthetic (DBL x COL)	6	237/296/313	*M. guttatus*
Incongruent sample				
M. guttatus	DBL	1	**240/292/317**	***M. luteus***
Total	22	163		

Table 4. Phenotypic measurements of *M. guttatus* (MG x MG), *M. luteus* (ML x ML), and their reciprocal hybrids (MG x ML, and ML x MG). Crosses are shown as maternal x paternal parent. Mean +/- standard error (SE). All values in millimeters.

	MG x MG (*n*=146)	**MG x ML** (*n*=117)	**ML x MG** (*n*=23)	**ML x ML** (*n*=155)
Floral traits				
Corolla height	23.54 ± 0.47	26.85 ± 0.44	27.43 ± 1.47	21.18 ± 0.42
Corolla width	27.26 ± 0.4	30.53 ± 0.39	29.79 ± 1.44	24.21 ± 0.46
Corolla throat	2.56 ± 0.07	3.63 ± 0.07	3.98 ± 0.26	4.43 ± 0.08
Corolla tube length	20.74 ± 0.22	24.38 ± 0.24	24.45 ± 0.5	23.51 ± 0.27
Flower depth	36.04 ± 0.47	36.54 ± 0.38	35.34 ± 1.15	31.53 ± 0.4
Calyx length	13.12 ± 0.14	14.81 ± 0.15	15.03 ± 0.32	13.87 ± 0.12
Pedicel length	18.91 ± 0.45	42.16 ± 1.31	35.31 ± 3.12	46.53 ± 1.01
Vegetative traits				
Bract length	16.58 ± 0.48	56.59 ± 2.34	43.42 ± 5.93	52.62 ± 1.78
Bract width	16.21 ± 0.39	42.95 ± 1.46	31.2 ± 2.86	34.62 ± 0.93

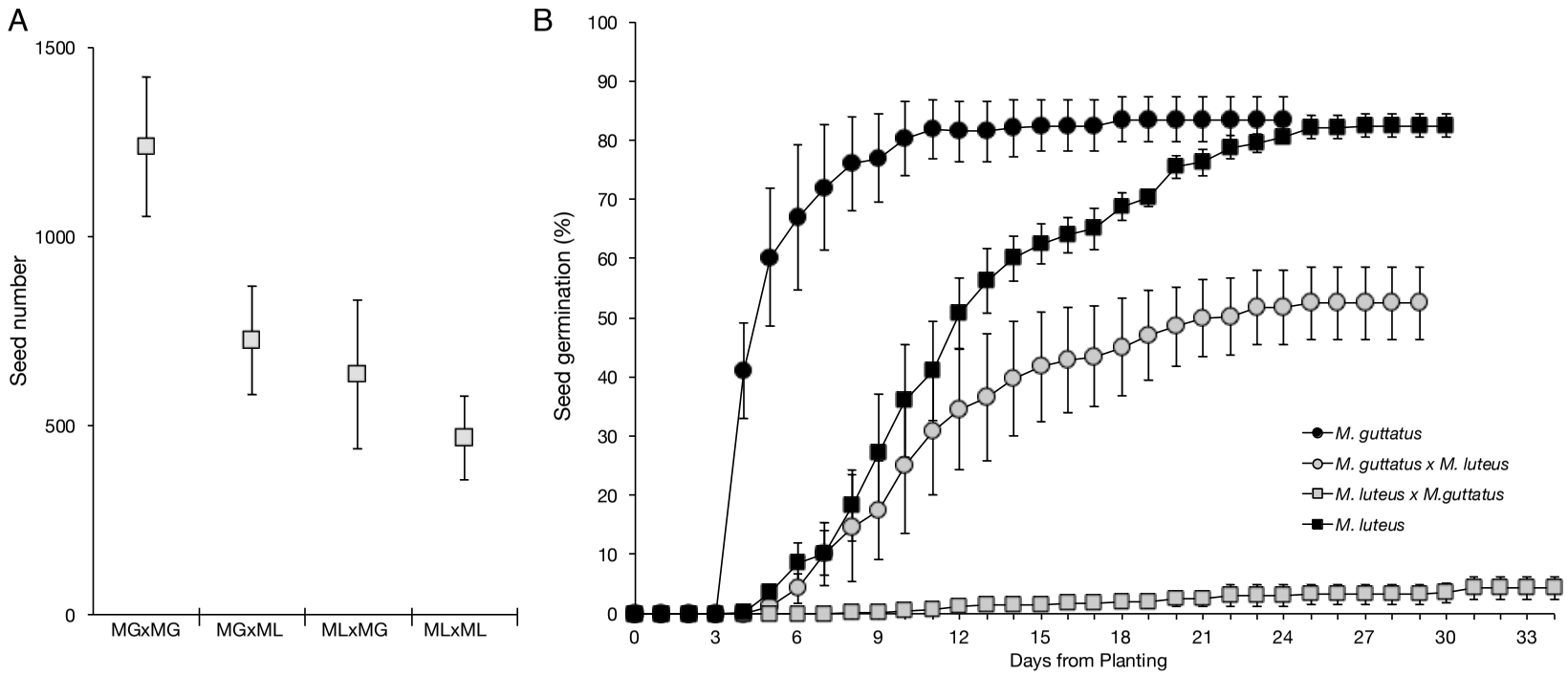

A
1500
1000
500
0
Seed number
MGxMG
MGxML
MLxMG
MLxML
B
100
90
80
70
60
50
40
30
20
10
0
Seed germination (%)
0
3
6
9
12
15
18
21
24
27
30
33
Days from Planting
M. guttatus
M. guttatus x M. luteus
M. luteus x M.guttatus
M. luteus

Figure 1.

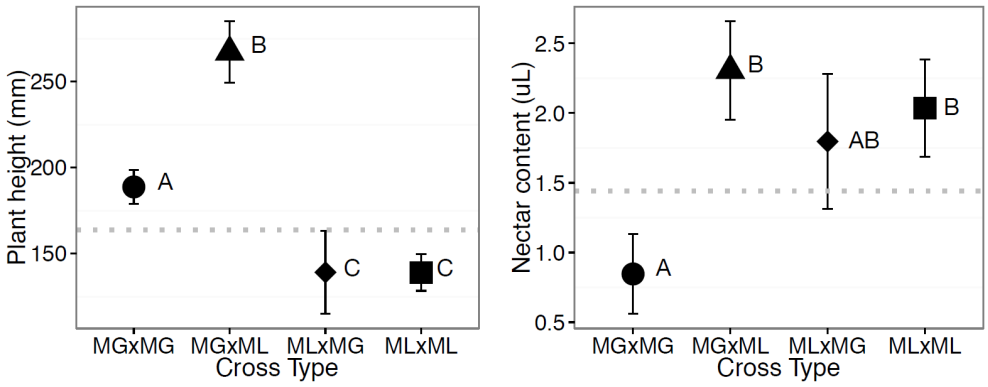

Plant height (mm)
250
200
150
A
B
C
C
MGxMG
MGxML
MLxMG
MLxML
Cross Type
Nectar content (uL)
2.5
2.0
1.5
1.0
0.5
A
B
AB
B
MGxMG
MGxML
MLxMG
MLxML
Cross Type

Figure 2.

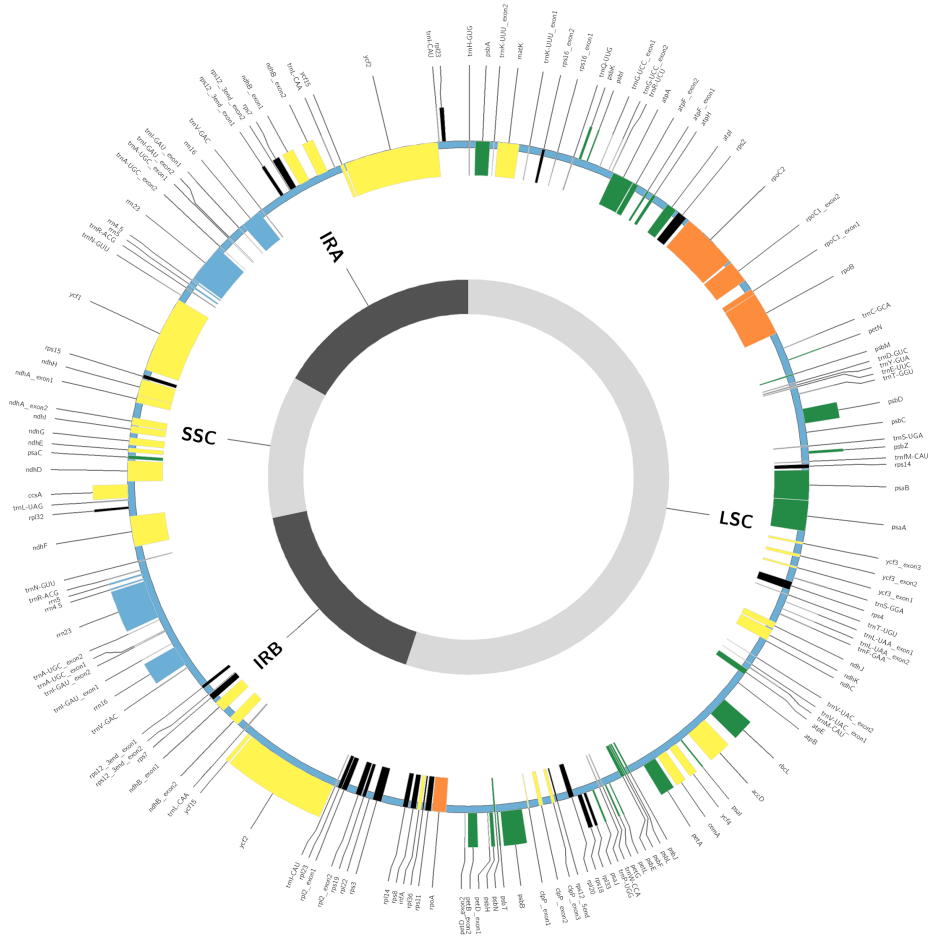


Figure 3.

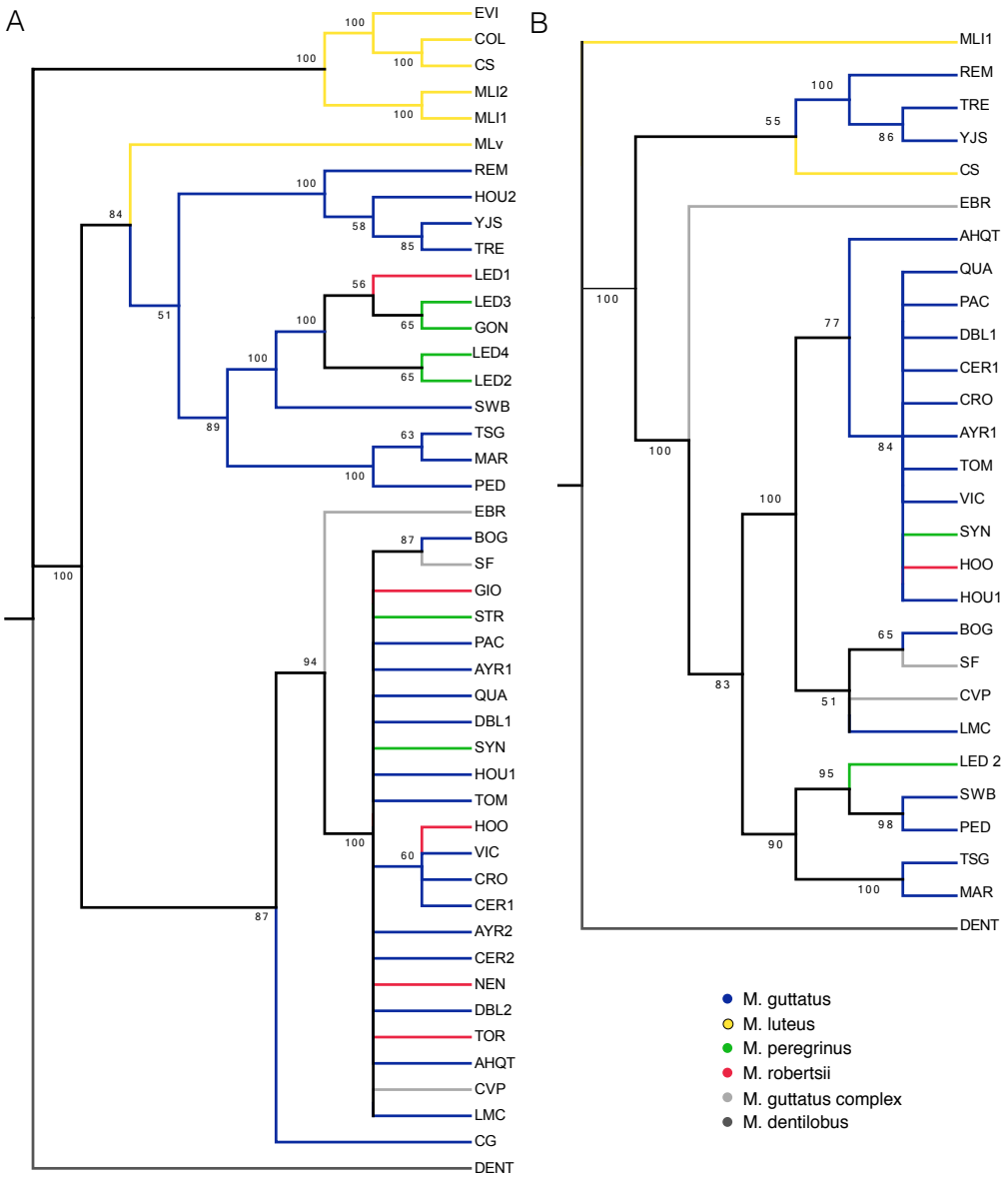

A
EVI
COL
CS
MLI2
MLI1
MLv
REM
HOU2
YJS
TRE
LED1
LED3
GON
LED4
LED2
SWB
TSG
MAR
PED
EBR
BOG
SF
GIO
STR
PAC
AYR1
QUA
DBL1
SYN
HOU1
TOM
HOO
VIC
CRO
CER1
AYR2
CER2
NEN
DBL2
TOR
AHQT
CVP
LMC
CG
DENT
B
MLI1
REM
TRE
YJS
CS
EBR
AHQT
QUA
PAC
DBL1
CER1
CRO
AYR1
TOM
VIC
SYN
HOO
HOU1
BOG
SF
CVP
LMC
LED 2
SWB
PED
TSG
MAR
DENT
M. guttatus
M. luteus
M. peregrinus
M. robertsii
M. guttatus complex
M. dentilobus

Figure 4.

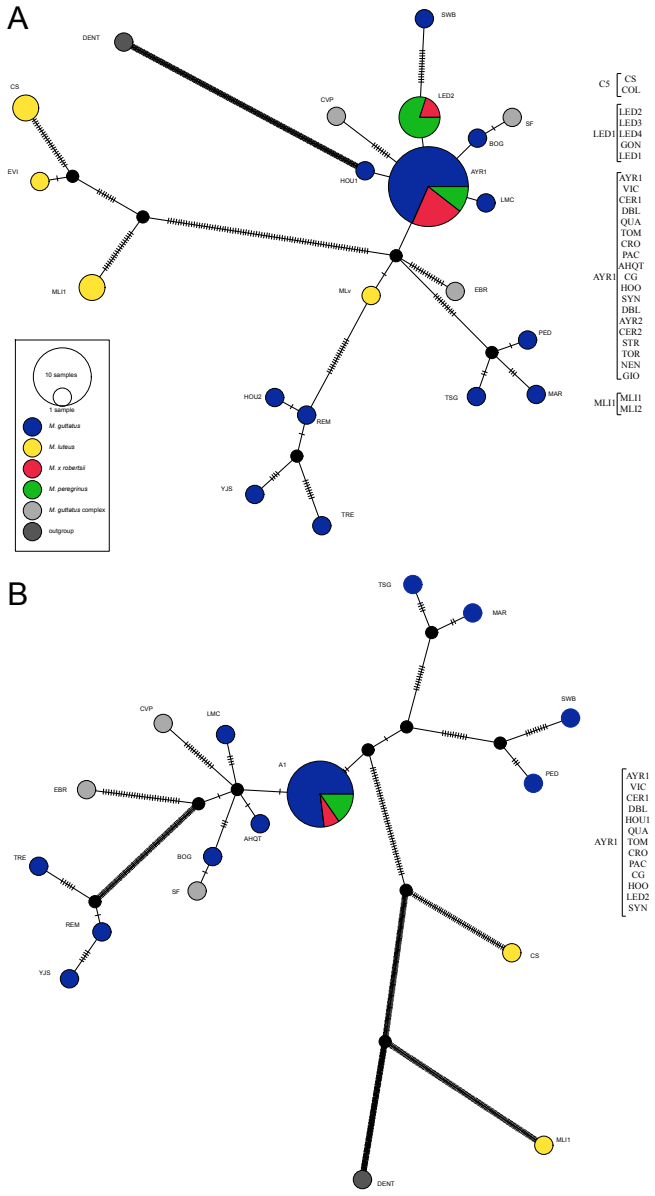

A
DENT
SWB
CS
CVP
LED2
SF
BOG
EVI
HOU1
AYR1
LMC
MLI1
MLv
EBR
PED
HOU2
REM
TSG
MAR
YJS
TRE
10 samples
1 sample
M. guttatus
M. luteus
M. x robertsii
M. peregrinus
M. guttatus complex
outgroup
CS
CS
COL
LED1
LED2
LED3
LED4
GON
LED1
AYR1
AYR1
VIC
CER1
DBL
QUA
TOM
CRO
PAC
AHQT
CG
HOO
SYN
DBL
AYR2
CER2
STR
TOR
NEN
GIO
MLI1
MLI1
MLI2
B
TSG
MAR
SWB
CVP
LMC
A1
PED
EBR
AHQT
BOG
SF
TRE
REM
YJS
CS
MLI1
DENT
AYR1
AYR1
VIC
CER1
DBL
HOU1
QUA
TOM
CRO
PAC
CG
HOO
LED2
SYN

Figure 5.

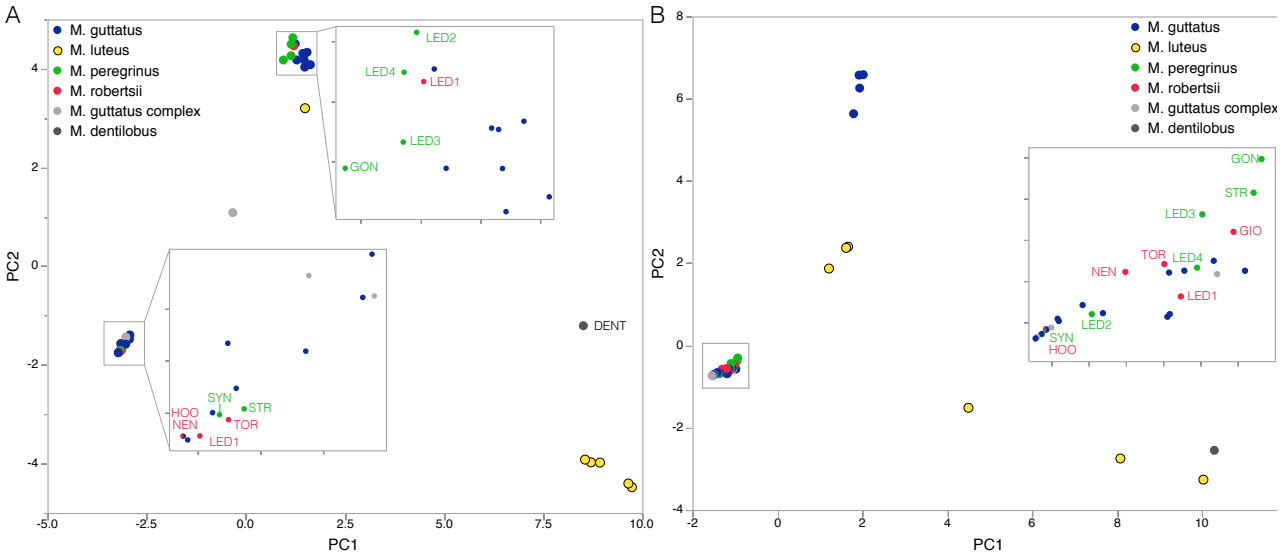

A
M. guttatus
M. luteus
M. peregrinus
M. robertsii
M. guttatus complex
M. dentilobus
LED2
LED4
LED1
LED3
GON
SYN
STR
HOO
NEN
TOR
LED1
DENT
PC2
PC1
B
M. guttatus
M. luteus
M. peregrinus
M. robertsii
M. guttatus complex
M. dentilobus
GON
STR
LED3
GIO
TOR
LED4
NEN
LED1
LED2
SYN
HOO
PC2
PC1

Figure 6

SUPPLEMENTAL: Analysis of cytoplasmic genomes demonstrates the asymmetric origin of a new allopolyploid species[1]

Mario Vallejo-Marin[2]
Arielle M. Cooley[3]
Michelle Lee Yue Qi[2]
Michael McKain[4]
Joshua R. Puzey[5,6]

[1] Manuscript received _____; revision accepted _____.
[2] Biological and Environmental Science, School of Natural Sciences. University of Stirling.
[3] Biology Department, Whitman College, Walla Walla, Washington 99362
[4] Donald Danforth Plant Science Center, St. Louis, MO 63132
[5] Department of Biology, College of William and Mary, Williamsburg, Virginia 23185
[6] Author for correspondence. Email: jrpuzey@wm.edu

Appendix S1. Reports of asymmetric crossing success following interploidy crosses, taken from Ramsey and Schemske (1998)

Genus	Reference	Ploidy Type	Superior Direction
Beta	(Woodell and Valentine, 1961)	unknown	m
Brassica	(Howard, 1939)	auto (c)	m
Campanula	(Gairdner and Darlington, 1931)	unknown (n)	m
Cucumis	(Shifriss, 1942)	auto (c)	m
Datura	(Sansome, Satina, and Blakeslee, 1942)	unknown (n)	m
Lycopersicon	(Cooper and Brink, 1945)	auto (c)	m
Salpiglossis	(Needham and Erickson, 1992)	unknown	m
Secale	(Håkansson and Ellerström, 1950)	unknown	m

Sinapsis	(Olsson and Rufelt, 1948)	auto (c)	m
Spinacia	(Janick and Stevenson, 1955)	auto (c)	m
Zea	(Randolph, 1935)	auto (e)	m
Euphorbia	(Milbocker and Sink, 1968, 1969)	unknown (n)	p
Galium	(Woodell and Valentine, 1961)	unknown (n)	p
Hordeum	(Tsuchiya, 1958, 1960)	auto (c)	p
Luffa	(Roy and Dutt, 1972)	auto (c)	p
Solanum	(Marks, 1966)	auto (c)	p
Vitis	(Olmo, 1952)	unknown	p

Notes: Ploidy type: c, colchicine-induced autotetraploid; n, naturally-occurring autotetraploid; u, type of autotetraploid unknown; e, experimentally produced by means other than colchicine. *Superior Direction:* m, crosses more successful when maternal parent has higher ploidy; p, crosses more successful when paternal parent has higher ploidy.

Appendix S2. Experimental lines used for the phenotypic analyses.

Cross Type (Maternal x Paternal)	Accession number	Parentage (Maternal x Paternal)	Number of plants in flowering experiment
M. guttatus* x *M. guttatus	H065	CG-6.2 x CG-3.3	30
(MG x MG)	H063	CG-3.3 x CG-6.2	30
	H082	CG-1.1 x CG-8.1	30
	H103	CG-8.1 x CG-1.1	30
M. luteus* x *M. luteus	H115	CS-11.3 x CS-7.1	30
(ML x ML)	H027	CS-7.1 x CS-11.3	30
	H014	CS-12.3 x CS-4.1	30
	H108	CS-4.1 x CS-12.3	30
M. guttatus* x *M. luteus	H112	CG-6.2 x CS-11.3	15
(MG x ML)	H022	CG-3.3 x CS-7.1	15
	H011	CG-1.1 x CS-12.3	15
	H104	CG-8.1 x CS-4.1	15
M. luteus* x *M. guttatus	H110	CS-11.3 x CG-6.2	4
(ML x MG)	H001	CS-7.1 x CG-3.3	9
	H012	CS-12.3 x CG-1.1	4
	H107	CS-4.1 x CG-8.1	0

Appendix S3. Supplemental methods.

Genotyping methods: All reads were aligned to the references using *bowtie2* (Langmead and Salzberg, 2012) (bowtie2 -I 150 -X 2000 --local --fast-local -x REF -1 fastq1 -2 fastq2 | samtools view -bS - > name.bam). Alignments were sorted, duplicates marked and discarded, and read groups added using picard tools (http://broadinstitute.github.io/picard). Following alignment filtration, genotypes were called used GATK UnifiedGenotyper (McKenna et al., 2010; DePristo et al., 2011; Auwera et al., 2013). with the following command: java -Xmx5g -jar ../GenomeAnalysisTK.jar -R Mimulus_mitochondria_JN098455.fa -T UnifiedGenotyper -rf MaxInsertSize --maxInsertSize 10000 -rf DuplicateRead -rf BadMate -rf BadCigar --min_base_quality_score 25 -rf MappingQuality --min_mapping_quality_score 25 -ploidy 2 --genotype_likelihoods_model BOTH --output_mode EMIT_ALL_SITES --max_alternate_alleles 2 --standard_min_confidence_threshold_for_emitting 30 -I BAM.list -o OUT.vcf.

Appendix S4. Primers used to genotype length polymorphisms in mitochondria in order to distinguish between *Mimulus guttatus* and *M. luteus* maternal parentage. Marker location refers to the position of the indel in the *M. guttatus* mitochondrial genome (Mower et al., 2012; GenBank accession number: JN098455.1).

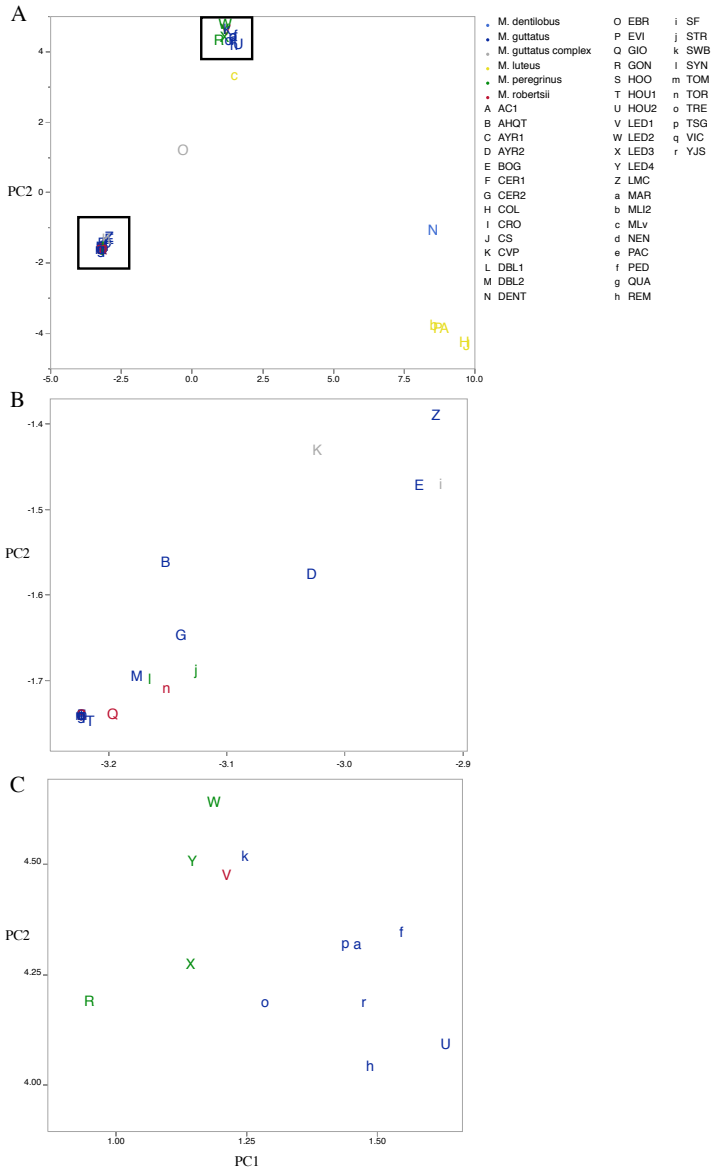
Locus	**Marker type**	**Marker location (bp)**		**Primer sequence (5'-3')**	**Dye**	**Expected PCR Product**	
						M. guttatus	*M. luteus*
mit-357	4bp indel	37,145	F:	TCAGCTTCGCCATTGCTACA	FAM	321	325
			R:	AGTACTGCAGTGAATGGGCC			
mit-259	3bp indel	261,294	F:	TGAGGAAAAGCCCGTGTTCT	FAM	239	242
			R:	CTATGAGGTGGCCCCTTTCG			
mit-262	4bp indel	263,336	F:	TGTCAGGTTGCATGGGGTAC	FAM	296	292
			R:	ACACGACCTGAAATGGGGAC			

Appendix S5. Sequencing depth for samples.

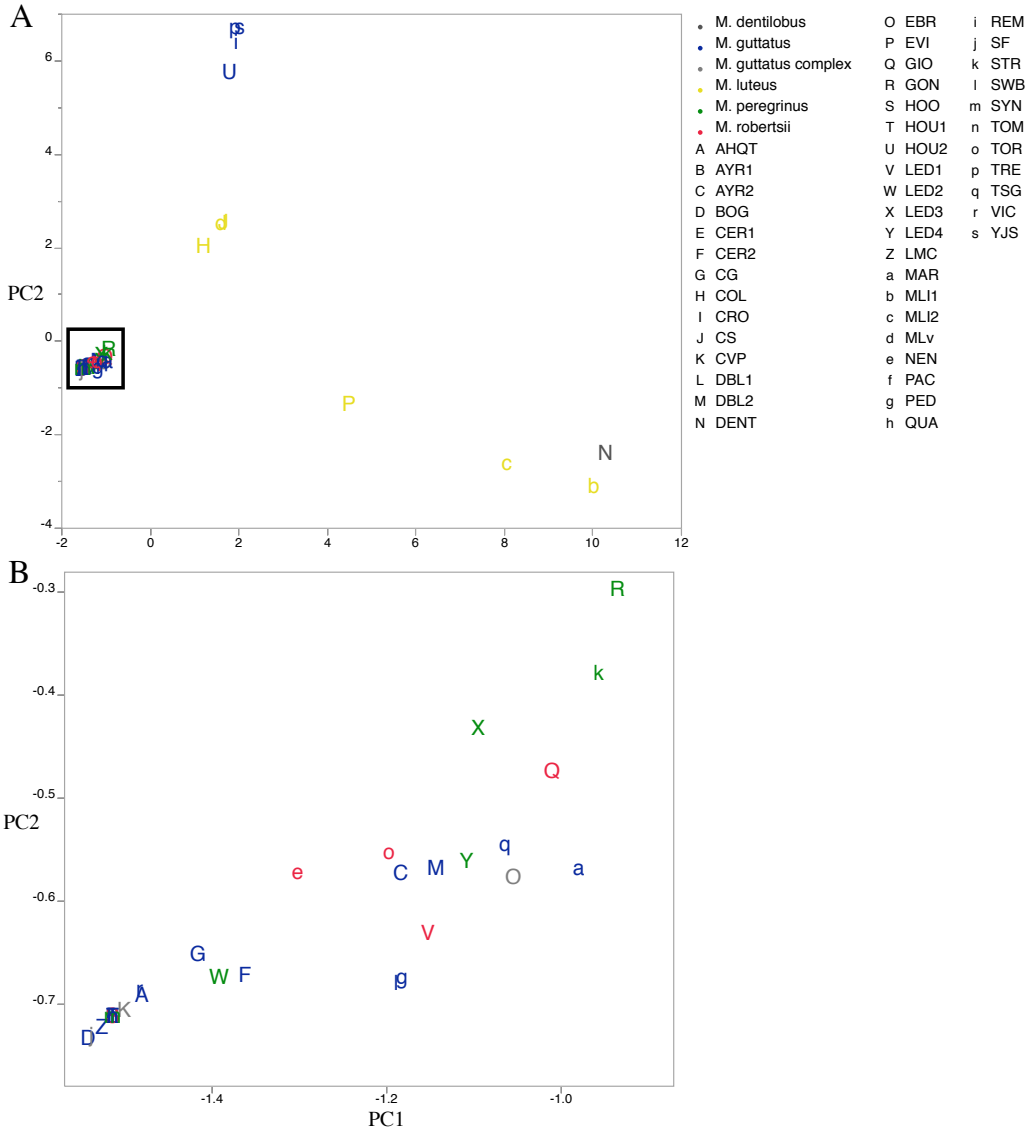
Data Type	Sample ID	Sequence ID	Average Read Depth	Number of genotyped bases	Fraction genotyped bases	Average depth of genotyped bases	Average Read Depth	Number of genotyped bases	Fraction genotyped bases	Average depth of genotyped bases
WGS	10-AYR-10	A1	1456.1	112775	0.74	228.5	171.4	441010	0.84	120.9
WGS	12-VIC-18	A10	346.6	105537	0.69	235.9	48.4	440724	0.84	42.9
WGS	10-CER-10	A2	2585.6	113715	0.74	227.3	246.7	441201	0.84	161.0
WGS	10-DBL-20	A3	1455.7	112367	0.73	229.1	173.2	441169	0.84	121.4
WGS	10-HOU-17	A4	426.5	106836	0.70	237.0	119.9	442976	0.84	112.6
WGS	10-QUA-47	A5	1718.6	107548	0.70	240.8	184.7	442413	0.84	127.5
WGS	10-TOM-23	A6	399.5	106496	0.70	236.3	50.0	442999	0.84	41.3
WGS	12-CRO-5	A7	917.8	106843	0.70	242.1	133.7	442430	0.84	110.4
WGS	12-PAC-39	A8	2909.2	115889	0.76	224.1	337.2	445981	0.85	217.1
WGS	12-TRE-17	A9	3815.4	111936	0.73	232.3	573.7	448311	0.85	244.9
WGS	MLl-EY	AC1	2486.4	107876	0.70	238.3	240.8	432469	0.82	174.3
WGS	AHQT	AHQT	4928.8	108772	0.71	236.5	844.6	438173	0.83	248.6
WGS	BOG	BOG	1708.9	137338	0.90	187.7	197.6	442700	0.84	162.0
WGS	CS-4-JP1	C5	2232.3	107077	0.70	242.1	206.5	448375	0.85	150.8
WGS	CG-1-1-JP1	CG	3262.1	109797	0.72	235.6	320.2	443677	0.84	210.1
WGS	CVP	CVP	4548.2	111857	0.73	229.8	940.1	436690	0.83	240.4
WGS	DENT	DENT	1630.7	153003	1.00	169.0	365.3	461374	0.88	224.3
WGS	EBR	EBR	5556.0	110374	0.72	233.0	713.3	438791	0.83	248.1
WGS	H003b-JP1	HOO	4384.4	111664	0.73	231.0	439.8	443851	0.84	239.3
WGS	11-LED-seed-1-	LED	3535.7	111650	0.73	232.2	319.9	447990	0.85	206.6

	JP1									
WGS	LMC	LMC	1920.8	149613	0.98	173.1	255.3	450802	0.86	207.0
WGS	MAR	MAR	1923.9	151430	0.99	171.0	349.2	453537	0.86	236.5
WGS	PED	PED	1999.1	146656	0.96	176.0	244.6	445281	0.85	195.6
WGS	REM	REM	2349.7	152132	0.99	171.0	305.3	461747	0.88	224.1
WGS	SF	SF	1515.3	145172	0.95	177.5	225.9	451390	0.86	184.4
WGS	SWB	SWB	1790.8	145926	0.95	176.9	580.2	458817	0.87	237.4
WGS	13-SYN-1-JP1	SYN	2476.9	109378	0.71	221.8	311.4	443903	0.84	194.7
WGS	TSG	TSG	889.4	152345	0.99	170.1	169.3	459080	0.87	138.0
WGS	YJS	YJS	1622.6	146730	0.96	176.1	167.7	447788	0.85	123.0
SC	11-DBL-1-1	gut1	93.9	105310	0.69	22.5	8.7	275679	0.52	5.6
SC	10-AYR-18-1	gut2	90.0	104474	0.68	19.4	8.9	285412	0.54	5.4
SC	10-CER-25-1	gut3	97.9	105822	0.69	25.4	10.1	313618	0.60	5.7
SC	10-HOU-23-1	gut4	112.0	105588	0.69	28.8	12.9	350977	0.67	6.2
SC	10-COL-20e-1	lut1	79.4	103726	0.68	15.7	8.7	300567	0.57	4.7
SC	MLl-EY-7x4-1-3	lut2	76.6	103911	0.68	15.1	8.4	292091	0.56	4.8
SC	13-EVI-5	lut3	87.3	103696	0.68	16.1	8.2	202412	0.39	5.5
SC	MLv-RC-5x4-4-1	lut4	90.5	104626	0.68	20.7	10.5	355082	0.68	5.2
SC	11-LED-seed-30	per1	54.8	100751	0.66	11.3	6.0	309412	0.59	4.2
SC	11-LED-seed-2-11	per2	84.9	103887	0.68	18.9	9.7	352591	0.67	5.4
SC	13-STR-32	per3	74.9	103142	0.67	12.5	7.8	233876	0.44	4.7
SC	12-GON-106	per4	59.5	101556	0.66	11.4	6.6	311248	0.59	4.3
SC	13-TOR-13	rob1	86.3	103984	0.68	16.2	9.5	283396	0.54	4.9
SC	12-LED-3-11	rob2	80.9	104376	0.68	22.2	8.0	329161	0.63	5.4
SC	12-NEN-41	rob3	99.3	104621	0.68	24.6	10.9	334022	0.64	5.8
SC	13-GIO-13	rob4	91.7	104848	0.68	20.6	8.6	255706	0.49	5.6

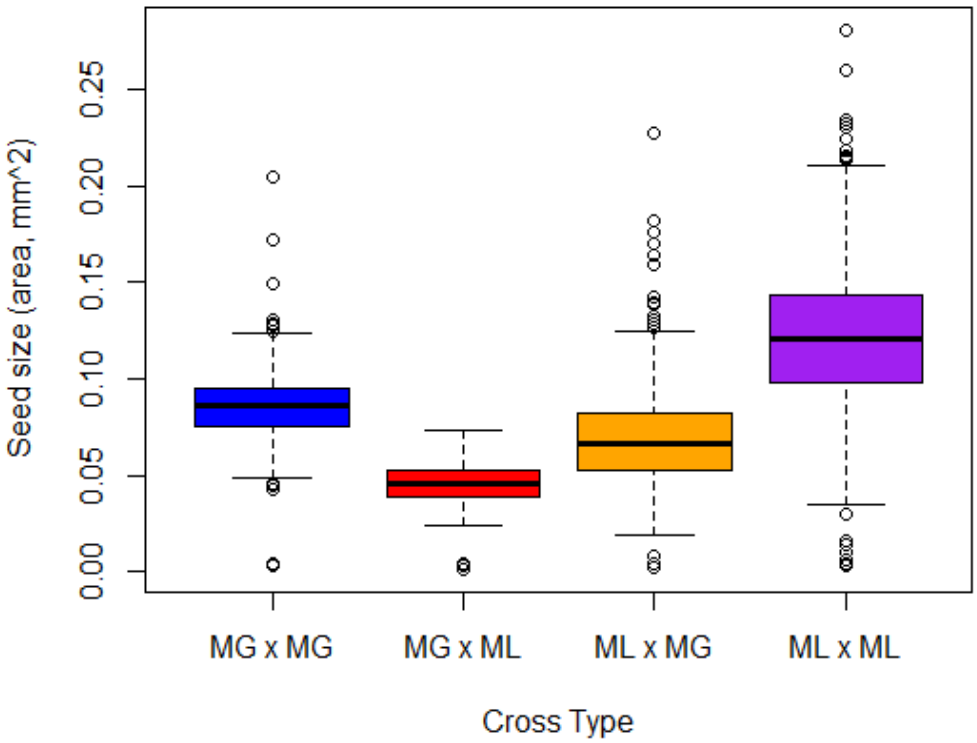
		Global average	1515	116290	0.76	144	208	393798	0.75	116
		SC average	85	104020	0.68	19	9	299078	0.57	5
		WGS average	2303	123060	0.80	213	318	446057	0.85	177



Appendix S6. Fully labeled PCA of samples based on chloroplast data. (A) Full PCA showin all samples. (B) Zoomed in view of bottom left cluster. (C) Zoomed in view of top cluster.



Appendix S7. Fully labeled PCA of samples based on mitochondria data. (A) Full PCA showing all samples. (B) Zoomed in view of bottom left cluster.



Appendix S8. Seed size of parental and hybrid crosses.

REFERENCES

COOPER, D., AND R. BRINK. 1945. Seed collapse following matings between diploid and tetraploid races of Lycopersicon pimpinellifolium. *Genetics* 30: 376.

GAIRDNER, A., AND C. DARLINGTON. 1931. Ring-formation in diploid and polyploid Campanula persicifolia. *Genetica* 13: 113-150.

HÅKANSSON, A., AND S. ELLERSTRÖM. 1950. Seed development after reciprocal crosses between diploid and tetraploid rye. *Hereditas* 36: 256-296.

HOWARD, H. 1939. The size of seeds in diploid and autotetraploidBrassica oleracea L. *Journal of Genetics* 38: 325-340.

JANICK, J., AND E. STEVENSON. 1955. The effects of polyploidy on sex expression in spinach. *Journal of Heredity* 46: 151-156.

MARKS, G. 1966. The enigma of triploid potatoes. *Euphytica* 15: 285-290.

MILBOCKER, D., AND K. SINK. 1968. The diploid-tetraploid cross in poinsettia. *American Society of Horticultural Science* 93: 663-666.

------. 1969. Embryology of diploid× diploid and diploid× tetraploid crosses in poinsettia. *Canadian Journal of Genetics and Cytology* 11: 598-601.

MOWER, J. P., A. L. CASE, E. R. FLORO, AND J. H. WILLIS. 2012. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (Mimulus guttatus) lineage with cryptic CMS. *Genome biology and evolution* 4: 670-686.

NEEDHAM, D. C., AND H. T. ERICKSON. 1992. Fecundity of tetraploid× diploid crosses and fertility of the resultant triploids in Salpiglossis sinuata. *HortScience* 27: 835-837.

OLMO, H. 1952. Breeding tetraploid grapes. Proc. Amer. Soc. Hort. Sci, 59: 285-290.

OLSSON, G., AND B. RUFELT. 1948. Spontaneous crossing between diploid and tetraploid Sinapis alba. *Hereditas* 34: 351-365.

RANDOLPH, L. 1935. Cytogenetics of tetraploid maize. *J. agric. Res* 50: 1-605.

ROY, R., AND B. DUTT. 1972. Cytomorphological studies in induced polyploids of Luffa acutangula Roxb. *Nucleus* 15: 171-180.

SANSOME, E. R., S. SATINA, AND A. BLAKESLEE. 1942. Disintegration of ovules in tetraploid-diploid and in incompatible species crosses in Datura. *Bulletin of the Torrey Botanical Club*: 405-420.

SHIFRISS, O. 1942. Polyploids in the genus *Cucumis*. *Journal of Heredity* 33: 144-152.

TSUCHIYA, T. 1958. Studies on cross compatibility of diploid, triploid, and tetraploid barleys. I. Seed development in reciprocal crosses between diploids and induced autotetraploids. *Bot. Mag.(Tokyo)* 71: U8-52.

------. 1960. Studies on cross compatibility of diploid, triploid and tetraploid barleys II. Results of crosses between triploids, diploids and induced autotetraploids. *Idengaku Zasshi= Jap. J. Genet.* 35: 337-343.

WOODELL, S., AND D. VALENTINE. 1961. Studies in British Primulas IX. Seed Incompatibility in Diploid-Autotetraploid Crosses. *New Phytologist*: 282-294.