

RESEARCH ARTICLE

Extensive phenotypic diversity in the cultivated Florist's Gloxinia, *Sinningia speciosa* (Lodd.) Hiern, is derived from the domestication of a single founder population

Tomas Hasing¹ | Elijah Rinaldi¹ | Silvia Manrique² | Lucia Colombo² | David C. Haak¹ | David Zaitlin³ | Aureliano Bombarely^{1,2} 

¹School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA

²Department of Biosciences, Università degli Studi di Milano, Milan, Italy

³Kentucky Tobacco Research and Development Center (KTRDC), University of Kentucky, Lexington, KY, USA

Correspondence

Aureliano Bombarely, School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA.

Email: aurebg@vt.edu

David Zaitlin, Kentucky Tobacco Research and Development Center (KTRDC), University of Kentucky, Lexington, KY, USA.
Email: dzait2@uky.edu

Funding information

Gesneriad Society, Grant/Award Number: Non applicable

Societal Impact Statement

Domesticated plants are essential for agriculture and human societies. Hence, understanding the processes of domestication will be crucial as we strive for more efficient crops and improvements to plants that benefit humankind in other ways. Here, we study the ornamental plant *Sinningia speciosa*, and reveal that despite the incredible variety found in domesticated varieties (e.g., in flower colour and form), they are all derived from a single founder population near Rio de Janeiro, Brazil. Knowledge of the domestication of horticultural plants is scarce and given its small, low-complexity genome, and ease of cultivation, we suggest that *S. speciosa* is a good model for studying genomic variation during domestication.

Summary

- The process of domestication often involves a complex genetic structure with contributions from multiple founder populations, interspecific hybridization, chromosomal introgressions, and polyploidization events that occurred hundreds to thousands of years earlier. These complex origins complicate the systematic study of the sources of phenotypic variation. The Florist's Gloxinia, *Sinningia speciosa* (Lodd.) Hiern, was introduced into cultivation in England two hundred years ago from botanical expeditions that began in the 18th century. Since that time, amateur plant breeders and small horticultural companies have developed hundreds of cultivars with a wide range of flower colors and shapes.
- In our genetic study of *S. speciosa*, we examined an extensive diversity panel consisting of 115 individuals that included different species, wild representatives, and cultivated accessions.
- Our analysis revealed that all of the domesticated varieties are derived from a single founder population that originated in or near the city of Rio de Janeiro in Brazil. We did not detect any major hybridization or polyploidization events that could have contributed to the rapid increase in phenotypic diversity.

- Our findings, in conjunction with other features such as a small, low-complexity genome, ease of cultivation, and rapid generation time, makes this species an attractive model for the study of genomic variation under domestication.

KEYWORDS

genetic bottleneck, ornamental crop, phenotypic diversity, plant domestication, population structure

1 | INTRODUCTION

Darwin's treatise (Darwin, 1868) on variation and domestication sparked over a century of research into the patterns of variation that are found in domesticated plants and animals. The evolutionary histories of major domesticated food crops such as wheat, soybean, rice, and maize have been well-studied by geneticists, evolutionary biologists, and anthropologists interested in multiple aspects of their divergence from wild progenitors (Kantar, Nashoba, Anderson, Blackman, & Rieseberg, 2017). Such crops have been essential to our understanding of plant domestication as a process that began ca. 10,000 years ago with the harvest of wild material for the purpose of propagation. The initial and subsequent cycles of harvesting and planting allowed humans to (consciously and/or unconsciously) select phenotypes that satisfied human needs and benefitted cultivation across different geographical regions. As would be expected, the details underlying this simplified description of plant domestication are complex and vary across species. For example, unintentional gene flow between crops and wild plants can significantly slow the process of domestication (Baute, Kane, Grassa, Lai, & Rieseberg, 2015), and when geographically differentiated relatives are involved, estimating the time and location of domestication of major crops can become a complex task (Larson et al., 2014). This situation often leads researchers to erroneously infer multiple domestication events from what was most likely a single process (Huang et al., 2012). Clearly, human-plant interactions have changed drastically since the time when landraces were selected. Factors such as increased levels of selection, the ability to efficiently transport plant material around the world, and the development of specialized agricultural systems and breeding programs have all had a major impact on the dynamics of plant domestication. However, studies on recently domesticated horticultural crops are limited, and major food crops continue to contribute disproportionately to the literature on plant domestication (Meyer, DuVal, & Jensen, 2012). As a group, ornamental crops have been overlooked, and genomic comparisons with their wild relatives are uncommon. The study of nontraditional, nonmodel species, such as *Sinningia speciosa*, is essential to expand the knowledge that we have on modern patterns of plant domestication and to broaden our understanding of general trends.

S. speciosa (Gesneriaceae), also known as the "florist's gloxinias", is a perennial, herbaceous, long-lived tuberous flowering plant native to the Atlantic Coastal Forests of southeastern Brazil. This ornamental houseplant has a well-documented domestication history covering the

brief 200 years since it was first collected from natural populations. *S. speciosa* belongs to the monophyletic tribe Sinningieae, a relatively small clade encompassing three genera and 85 species. Despite this paucity of species, the group spans a remarkable range of flower forms and colors that have diverged into four pollinator syndromes (Perret, Chautems, Spichiger, Kite, & Savolainen, 2003). The species range from minute herbaceous plants with perennial tubers (such as *S. pusilla*) to large, woody shrubs that do not have tubers (*Paliavana prasinata*, for example) (Perret, Chautems, & Spichiger, 2006). Artificial interspecific hybrids are not uncommon across the tribe, so many species are potential sources of phenotypic diversity within the cultivated forms.

S. speciosa is an excellent model for studying plant domestication for several reasons. First, it is of relatively recent introduction (ca. 200 years), so founder population alleles may remain well-represented among extant natural populations. Moreover, given that most cultivated forms of *S. speciosa* have been maintained and bred far from their native habitat in Brazil, the chance of unintentional gene flow between wild and cultivated plants is minimal. Like other model domestication systems (e.g. maize) cultivated forms of *S. speciosa* show striking phenotypic divergence from natural populations (Figure 1, Table 1). While the diversity of some ornamental traits such as flower color and corolla pattern has increased over time, flower symmetry has shifted from the natural zygomorphic type to the mutant actinomorphic form, which is practically fixed in the commercial cultivars. Similar patterns of trait fixation and diversification have previously been associated with the initial stages of crop domestication and later phases of improvement (Meyer & Purugganan, 2013). Importantly, because of the short domestication history, there are numerous written and pictorial records in horticultural magazines and catalogues from nurseries and botanical gardens that provide information about the introduction, cultivation, breeding, and commercialization of particular phenotypes. This information reveals the chronological order in which most of the major mutations appeared and were incorporated into the cultivated material. Together, these resources significantly enhance the application of genomic tools to accurately reconstruct the domestication of the species, identify founder populations, and search for hybridization and polyploidization events that could explain the generation of so much phenotypic diversity over such a relatively short period of time.

Genotyping-by-sequencing (GBS) (Elshire et al., 2011) combines high-throughput next-generation DNA sequencing with the construction of highly multiplexed pools of reduced-representation genomic libraries to provide genome-wide genotyping at an affordable

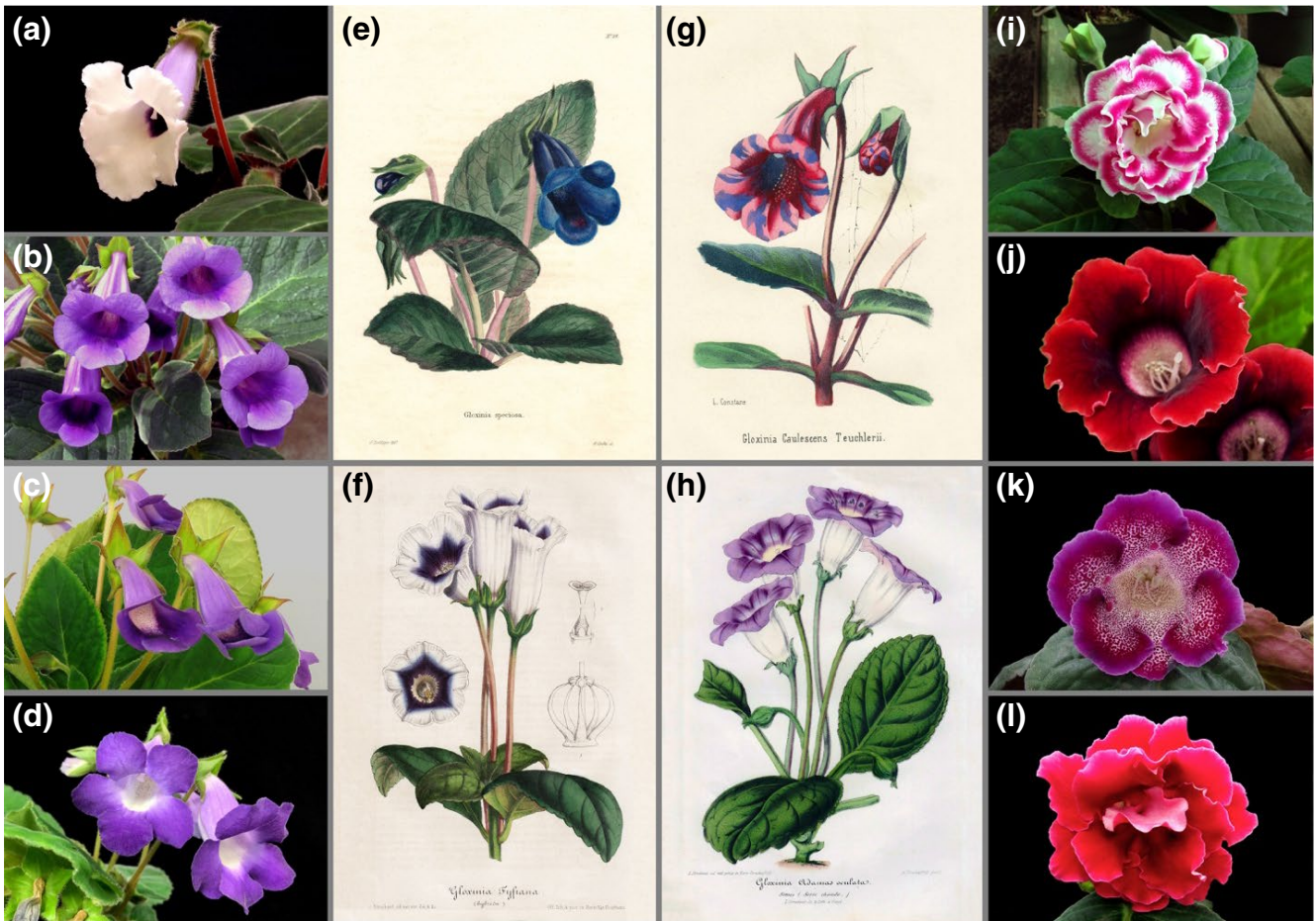


FIGURE 1 Floral diversity in wild and cultivated accessions of *Sinningia speciosa*. Flowers of four wild collections from Brazil (a–d): *S. speciosa* “Pedra Lisa” (a); “São Conrado” (b); “Avenida Niemeyer” (c); and “Chiltern Seeds” (d). Hand-colored botanical prints from the 19th century (e–h): the first known image of *Gloxinia speciosa* (Loddiges 1817) (e); the first *S. speciosa* cultivar with peloric flowers, *G. Fyfiana*, (Lemaire and Van Houtte, 1848) (f); *G. caulescens* “Teuschlerii” (Neumann 1846) (g); and *G. “Adamas Oculata”* (Lemaire and Verschaffelt, 1855) (h). Flowers of modern-day peloric cultivars (i–l): unknown double-flowered cultivar from a nursery in Costa Rica (i); the commercial cultivar “Empress” with single red and burgundy corollas (j); “Peridot’s Darth Vader” (k); and double corolla “Bristol’s Love Potion” (l)

cost. This level of efficiency is essential when studying nonmodel organisms of low economic importance for which whole genome resequencing may be cost prohibitive. Using single nucleotide polymorphisms (SNPs) obtained through GBS, we characterized a collection of wild and cultivated *S. speciosa* individuals to measure the effects of domestication on the level of genetic diversity, genetic structure, and demography. We also estimated genome sizes using flow cytometry to determine the impact of polyploidization events or significant genome expansions that could be potential sources of phenotypic diversity. Finally, we genotyped several species across the three major clades in the tribe Sinningieae to explore the potential for hybridization among members of the tribe.

2 | MATERIALS AND METHODS

2.1 | Plant material

We analyzed a collection of 115 individuals classified in the tribe Sinningieae. Living plant material or extracted DNA was obtained

from different sources. Most cultivated types were purchased from two commercial growers; The Violet Barn in the USA (Naples, NY) and Koeman Flowerbulbs in the Netherlands (Hem, North Holland). Most of the wild accessions were obtained through The Gesneriad Society Seed Fund and/or from the private collections of members of The Gesneriad Society. The wild accessions have been maintained and propagated either clonally or by seed in the USA for several years, and the names given to the different populations represent their collection sites in Brazil (Figure 2). To capture as much genetic diversity as possible among the domesticated material, we purchased a number of cultivars sold exclusively in Europe. These plants were sent to the Colombo laboratory in the University of Milan (Italy) where the DNA was extracted and then shipped to the USA. A number of other samples were also received as purified DNA. Of the 115 individuals, 58 were *S. speciosa*: 21 wild representatives, 30 domesticated, four semidomesticated, and three F_1 individuals produced in the Virginia Tech School of Plant and Environmental Sciences (VA, USA) greenhouse (Table S1). The semidomesticated group consisted of old cultivated material of generally wild appearance, but with larger

flowers. The F_1 individuals were included in the study as controls for the population structure analysis; one was the result of a cross between a cultivated and a wild form (Búzios × “Empress”), while the other two were full-sibs generated from a cross between two commercial cultivars (“Love Potion” × “Good Morning”). The remaining 57 individuals represented 32 other species and nine hybrids sampled from across the three major clades of the tribe Sinningieae;

TABLE 1 Comparison of phenotypic diversity across wild and cultivated forms of *S. speciosa*

Trait	Wild	Domesticated
Overall flower size	Small	Small
	Medium	Medium
		Large
Corolla color	Lavender	Lavender
	Purple	Purple
	White	White
	Pink	Pink
		Blue
		Red
Overall color pattern	Dotted-low	Dotted-low
		Dotted-medium
		Dotted-full
		Solid color
Color pattern on edge	Soft edge	Soft edge
		Defined edge
Orientation	Zygomorphic	Zygomorphic
		Actinomorphic
Corolla	Single	Single
		Double
		Multiple
Number of petals	Five	Five
		Six to nine

Sinningia, Corytholoma, and Dircaea. Four other *Sinningia* individuals of unknown species were also included in this group (Table S2).

2.2 | DNA extraction, library preparation, and Illumina DNA sequencing

Genomic DNA was extracted using a modified CTAB protocol (Doyle & Doyle, 1987) and further purified with the Monarch PCR & DNA Cleanup Kit (New England Biolabs (NEB), Ipswich). Each GBS library was prepared from 100 ng of genomic DNA digested with the restriction enzyme ApeKI (NEB) for 2 hr at 75°C. The reactions contained 10 µl gDNA solution (10 ng/µl), 1 µl ApeKI (5 U/µl), 2 µl 10X NEBuffer 3.1 (NEB) and 7 µl ddH₂O. Illumina adapters were then attached with T4 DNA ligase (NEB) by incubation for 2 hr at 22°C followed by 30 min at 65°C to inactivate the ligase. Each ligation reaction contained 20 µl digested gDNA, 6 µl adapter stock solution (0.6 ng/µl, included both common and barcode adapters), 5 µl ligase buffer (10X, NEB), 1.6 µl T4 DNA ligase (400 U/µl), and 17.4 µl ddH₂O. Subsequently, 5 µl aliquots from each library were pooled and purified using the Monarch PCR & DNA Cleanup Kit. Libraries were amplified by PCR in reactions containing 2 of the pooled libraries (15 ng/µl), 25 µl Taq Master Mix (2×, NEB), 1 µl forward and 1 µl reverse primers (10 µM), and 21 µl ddH₂O. The thermocycler program was 1 min at 95°C, followed by 18 cycles of 30 s at 95°C, 20 s at 62°C, and 30 s at 68°C, with a 5 min extension at 68°C and final cooling to 4°C. The amplified DNA was purified using the NEB Monarch PCR & DNA Cleanup Kit, and DNA fragment sizes between 250 and 550 bp were selected using the BluePippin instrument (Sage Science) with a 2% Agarose Dye-Free cassette and external marker V1. Finally, the size distribution of DNA fragments in the libraries was confirmed using the Agilent 2100 Bioanalyzer system (Agilent Genomics). The GBS libraries were sequenced with two Illumina single-end runs at the Duke Center for Genomic and Computational Biology. The first run (33 samples) was sequenced with 100 cycles on the Illumina HiSeq 2500 (Rapid Run), while the second run (82 samples) was sequenced with 75 cycles on the Illumina NextSeq 500

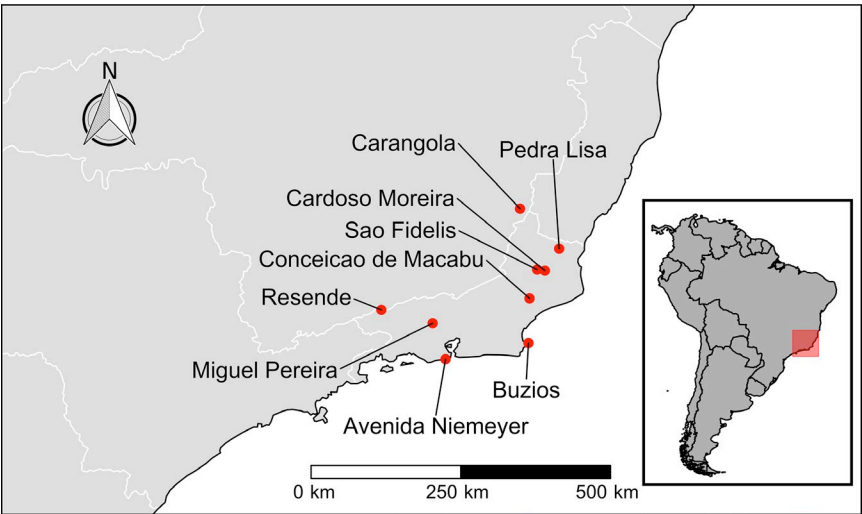


FIGURE 2 Approximate collection sites in Brazil for nine wild populations of *S. speciosa* for which geographic data were available. “Buzios” refers to the collection site in the seaside town of Armação dos Búzios in southeastern Rio de Janeiro state

instrument (both at the Duke University Center for Genomic and Computational Biology).

2.3 | Read processing and SNP calling

The fastq files were demultiplexed, and adapter sequences were removed using CUTADAPT V1.13 (Martin, 2011). The reads were then processed with FASTQ-MCF V1.04.807 (Aronesty, 2013) to perform base trimming, removing bases with quality scores <30 (phred-scaled quality score) from both ends and discarding reads shorter than 50 bases. Reads were then aligned against a draft genome of *S. speciosa* currently assembled onto 8,027 scaffolds (assembly available upon request) using BOWTIE2 V2.2.4 (Langmead & Salzberg, 2012). We used FREEBAYES V0.9.20 (Garrison & Marth, 2012) with the default parameters to call variants; we retained only bi-allelic SNPs using BCFTOOLS V1.3.1 (Li et al., 2009) and discarded individual variants with quality below 30 (phred-scaled quality score for the assertion made on the alternative allele) and depth less than 10 using VCFILTER V1.0.0 (Garrison, 2016). Finally, we removed SNPs with missing data using VCFILTER V0.1.12 (Danecek et al., 2011).

2.4 | Population structure and genetic diversity

We analyzed population structure using three different clustering approaches: (a) principal component analysis (PCA) based on genetic distances; (b) maximum likelihood estimation of population membership using ADMIXTURE V1.3.0 (Alexander, Novembre, & Lange, 2009); and (c) the Bayesian-based clustering method as implemented in FINESTRUCTURE V2.0.7 (Lawson, Hellenthal, Myers, & Falush, 2012). Pairwise genetic distances for PCA were estimated using TASSEL V5.2.48 (Bradbury et al., 2007), which defines the distance at a given locus between two individuals as 1–pIBS (probability of identity by state—the probability that two alleles drawn at random at a given locus are the same). ADMIXTURE and FINESTRUCTURE differ with respect to their model-based approaches. ADMIXTURE models the probability of each genotype using ancestry proportions and population allele frequencies, but requires a predefined number of populations (K). The optimal K was chosen by running ADMIXTURE's standard cross-validation procedure for values from 1 to 10, with each one tested 10 times using different random seeds. FINESTRUCTURE uses a “chromosome painting” approach that takes into account the relative positions of SNPs to define the haplotypes that are donated/received across individuals. This information leads to a “co-ancestry matrix” that is used to partition the dataset into groups with indistinguishable genetic ancestry (populations). Finally, estimates of genetic diversity were calculated using the number of segregating sites (S), nucleotide diversity (π), and Watterson's theta (θ_w) estimators. These values were generated using the POPGENOME R package (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014). Population differentiation was measured using Weir and Cockerham's weighted fixation index (F_{ST}), calculated with VCFILTER.

2.5 | Genome size measurement

We used flow cytometry to estimate the genome size of a subset of individuals from which fresh tissue was available. Each specimen was replicated three times using *Nicotiana benthamiana* as an internal standard (1C value = 3.2 pg). Nuclei were extracted by chopping young tissue from both the test sample and the reference (50 mg fresh tissue from each) together in 1 ml De Laat's buffer (de Laat & Blaas, 1984). We also analyzed each sample and the reference separately to confirm the individual peaks. After chopping, the sample slurries were passed through a 30 μ m CellTrics filter (Sysmex, Kobe, Japan) and centrifuged at 4 RCF (200 rpm) at 4°C for 5 min to pellet the nuclei. The volume was reduced to 200 μ l by removing the upper solution, and 200 μ l of 2X staining solution (100 μ g/ml propidium iodide, 100 μ g/ml RNase A, 2.2 μ l/ml 2-mercapthoethanol, and De Laat's buffer to final volume) was added. After gentle mixing, the tubes were incubated at room temperature in the dark for 20 min. Samples were kept at 4°C until measurements were taken, usually within 2 hr. The stained nuclei samples were analyzed with a BD FACSCalibur flow cytometer (BD Biosciences) using a 488 nm laser and a 585/42 bandpass filter to measure the fluorescence of propidium iodide, and the data was analyzed with FlowJo VX software (Treestar Inc). Aggregated nuclei and debris were excluded from the analysis using a PI-A versus PI-W plot, and the median fluorescent intensity was calculated for each population of nuclei.

3 | RESULTS

3.1 | Alignment and SNP calling

The tribe Sinningieae had previously been divided into five clades based on a phylogenetic analysis of several plastid DNA sequences and one nuclear gene (Perret et al., 2006): the clades are Dircaea, Corytholoma, Sinningia, Vanhouttea, and Thamnoligeria, the last two of which contain only seven species. Our collection initially included 128 individuals distributed across the larger clades Dircaea, Corytholoma, and Sinningia. Thirteen samples were excluded from the analysis because of insufficient data (less than 750,000 reads). The remaining 115 individuals averaged >4 million aligned reads each. The genome of *S. speciosa* that we used as reference for read alignment belongs to the wild accession “Avenida Niemeyer”. At present, the assembly contains 395.6 Mbp fragmented into 8,078 scaffolds (N90 = 1,776 bp) with an average length of 49.0 Kb. The total size of the assembly is very close to the 389.9 Mb (± 4.9) genome size estimated for “Avenida Niemeyer” using flow cytometry. After calling and filtering variants, we obtained 4,636,365 bi-allelic SNPs with a minimum depth of 10 and quality of 30 (phred-scaled quality score for the assertion made on the alternative allele). Finally, after removing all SNPs with any missing information, we retained 9,913 high-quality SNPs among the 115 individuals across the tribe, and 25,083 SNPs among the 58 *S. speciosa* individuals.

3.2 | Population structure

We examined the population structure by using PCA on genetic distances, as well as the model-based clustering methods from ADMIXTURE and FINESTRUCTURE. The first two components of the PCA explained 92.4% of the variance measured across the entire tribe, and provided a clear separation between the three major clades (Figure 3a). The first component (79.3%) mostly separated the Sinningia clade from Corytholoma and Dircaea, while the second component (13.1%) separated the latter two. As expected, the artificial hybrids fell between their parental species. For example, the hybrid “Yma”, was located between its progenitors *S. bullata* and *S. muscicola*. Similarly, “Apricot Bouquet” was positioned between the three species reported to be in its background - *S. aggregata*, *S. warmingii*, and *S. tubiflora*. XSinvana ‘Mount Magazine’, a hybrid between *Paliavana tenuiflora* and *S. conspicua* (Becker, 2008), was positioned between *P. prasinata* (a close relative of *P. tenuiflora*) and *S. conspicua*. Other unnamed F_1 hybrids included the crosses *S. bullata* x *S. conspicua* and *S. speciosa* x *S. helleri*, and they followed the same trend. Parents of other hybrids in the analysis are either unknown or

were not included in our dataset. Even though the 58 *S. speciosa* individuals were spread over a relatively large area formed by the first two principal components, both the cultivated and wild forms clustered together with no obvious signs of hybridization events pulling domesticated forms toward any other species.

The higher resolution PCA based on the 25,083 SNPs identified within *S. speciosa* revealed more details about the relationships between the domesticated cultivars and wild types (Figure 3b). We included *S. macrophylla* in this group because it clustered well within *S. speciosa*, in agreement with previous studies (Perret et al., 2003; Zaitlin, 2012). The first two components explained 67.8% of the variance. The commercial cultivars were clearly separated from most wild accessions across the first component (62.3%), while the semi-domesticated types grouped close to the transition area. The second component (5.5%) mainly isolated two wild populations, ‘Imbé’ and ‘Poço Parado’, from the rest. The F_1 individual from an intraspecific *S. speciosa* test cross was situated between its parents, the wild-type ‘Búzios’ and the red-flowered form of the cultivar “Empress”, following the same trend as observed for the interspecific hybrids. A visual comparison of the area covered by the cultivars in relation to the area covered by wild accessions suggests a general reduction in genetic diversity resulting from domestication. “Avenida Niemeyer” and “WT01” are wild-type accessions that clustered with the domesticated types.

The population structure inferred with ADMIXTURE (Figure 4) produced similar results to those from the PCA. The cross-validation errors were smallest at values of K ranging from 2 to 4 (Figure S1). At K=2, ADMIXTURE separated the domesticated types from the wild accessions almost perfectly. In agreement with the PCA, “Avenida Niemeyer” and “WT01” were associated with the domesticated group. In fact, “Diego Pink” and “NT-Milye Vesnushki” were the only cultivars that showed any level of admixture with any other wild types. “Diego Pink” is likely the result of a recent backcross of cultivated material to a wild form, aimed to produce zygomorphic flowers of large size and distinctive color. “NT-Milye Vesnushki” is a Russian cultivar with an apparently complex background that is exposed at higher levels of K. At K=3 the wild types were subdivided into two groups, while all the commercial cultivars remained together and maintained the connection with “Avenida Niemeyer” and “WT01”. At K = 4, the domesticated group split into two subgroups. The first subgroup included the semidomesticated types and older cultivars which remained directly associated with “Avenida Niemeyer” and “WT01”. Newer cultivars clustered in the second subgroup, possibly due to significant differentiation that occurred after additional cycles of breeding.

In general, the results from FINESTRUCTURE (Figure 5) agreed with the results from the PCA and ADMIXTURE. However, FINESTRUCTURE identified 33 small clusters that better fit the traditional definition of populations established in freely interbreeding groups of individuals. For instance, it perfectly separated all the wild accessions from different locations into individual populations. It also clustered together all four individuals from the cultivar “Empress” into a single population, as well as most “Bristol” type

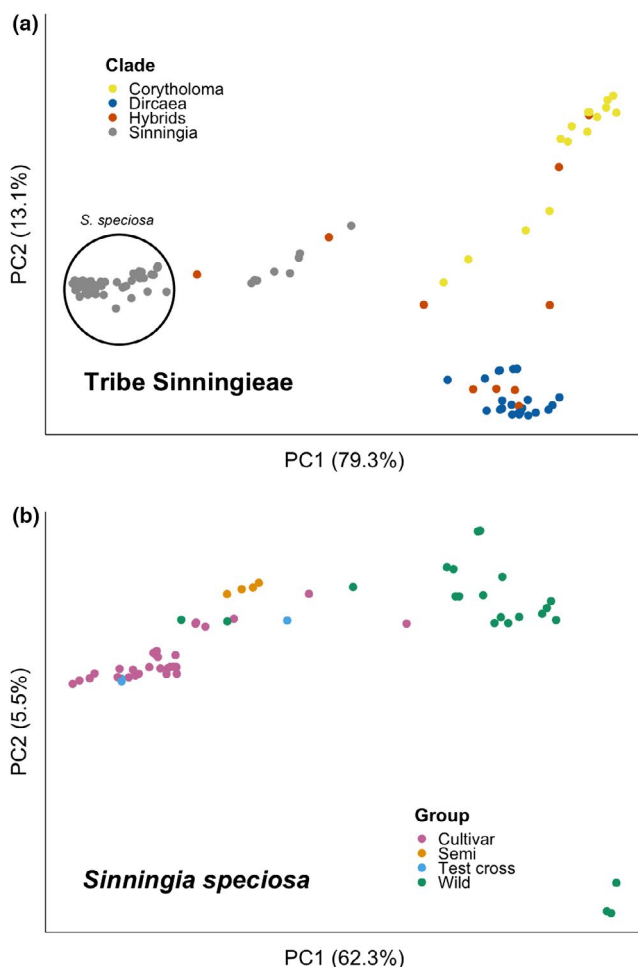
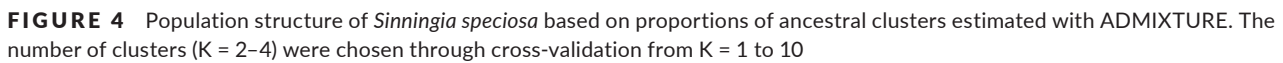


FIGURE 3 Principal component analysis based on genetic distances. (a) Members of tribe Sinningieae are color coded based on the three main phylogenetic clades; (b) *S. speciosa* accessions are color coded based on their relative level of domestication



One of the main weaknesses of GBS, especially when conducted at low coverage, is the amount of missing data it generates (Glaubitz et al., 2014), which forces researchers to either utilize sophisticated imputation methods or to include variants with some degree of missing information in their analyses. This is often problematic, because population genetic estimates of commonly used statistics can deviate considerably from true values (Arnold, Corbett-Detig, Hartl, & Bomblies, 2013). However, we were able to identify several thousand nonmissing SNPs at the intra- and interspecific levels. Working exclusively with nonmissing SNPs across species restricts our comparisons to conserved genomic regions that were originally present in the tribe's common ancestor. Any bias in our multispecies analyses seems marginal or nonexistent, because we were able to cleanly separate the three main clades in the tribe and make sense of reported interspecific hybrids. Nonetheless, to reduce potential bias and increase resolution within *S. speciosa*, we re-filtered the original SNP data based on nonmissing observations within the species, effectively creating two groups of SNPs. The larger number of nonmissing SNPs identified in *S. speciosa* is attributed to the reduced number of samples and the closer genetic relationships among them and to the reference genome, factors that improve the consistency of the alignments and the number of common sites respectively. Although the fragmented

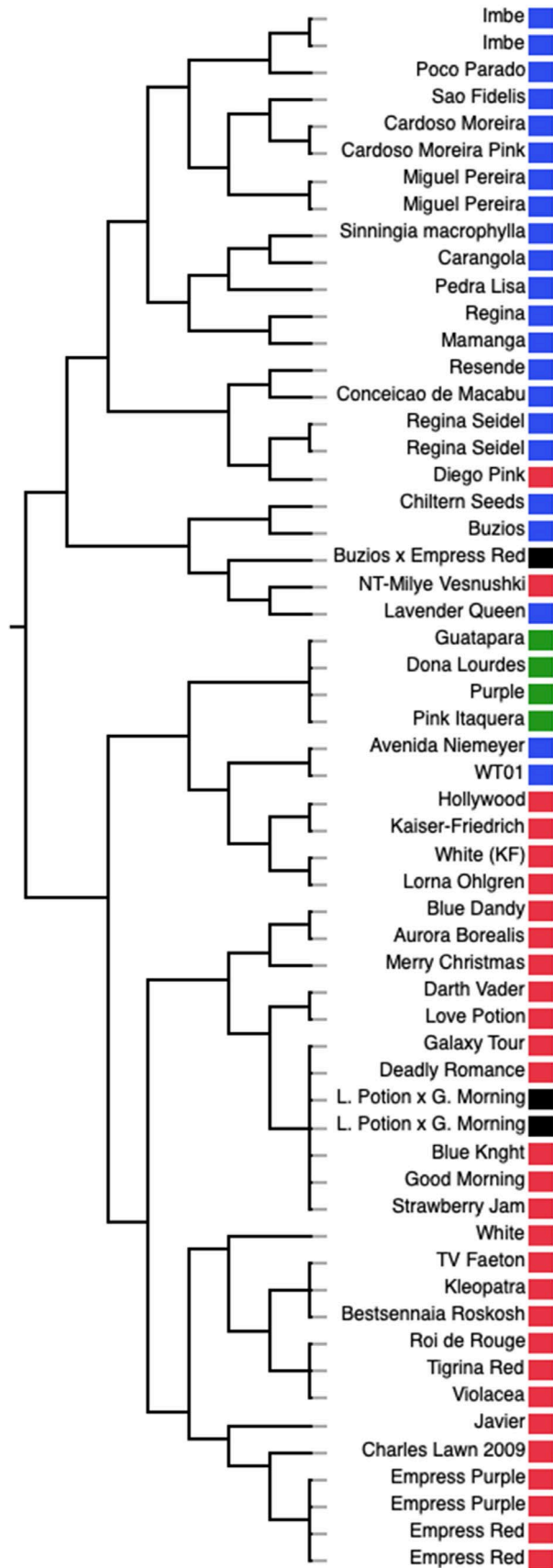


FIGURE 5 Relationships between *Sinningia speciosa* individuals based on FINESTRUCTURE. The gray tips that are directly connected with a vertical line represent a single population. Samples are color coded based on the level of domestication. Blue: wild, Green: semidomesticated, Red: domesticated, Black: test-crosses

condition of our draft reference genome did not allow us to identify specific regions showing hallmarks of domestication, we were able to generate genome-wide estimates of parameters associated with genetic diversity and population structure.

4.2 | Cultivated *S. speciosa*; single or multiple species?

The relatively short period of time in which *S. speciosa* has been subjected to cultivation and breeding has produced numerous phenotypic variations that are unknown in natural populations. In less than two hundred years, the flowers have diverged with respect to size, shape, color, and corolla pattern, as well as in the number of stamens, petals, and petal whorls. Furthermore, most of these changes were introduced within the first four decades of cultivation (Zaitlin, 2011). Comparable levels of human-induced phenotypic variation have been introduced in other ornamental flowering crops. However, such crops have often gone through longer periods of cultivation and/or complex interspecies hybridization schemes. Roses, for example, have been cultivated for ~5,000 years, during which time seven or more species with different levels of ploidy were cross-pollinated to create the genomes of contemporary roses (Bombarely, 2018; Martin, Piola, Chessel, Jay, & Heizmann, 2001). Lilies, tulips, and amaryllis (Christenhusz et al., 2013; Meerow, 2009; van Tuyl & Arens, 2011) are three other examples of highly hybridized ornamental plants. Despite the short time frame in which major aesthetic changes were introduced into *S. speciosa*, we found no evidence of inter-specific hybridization events that could conceivably accelerate the process. Early reports of multiple hybridizations include crosses between *Gloxinia speciosa*, *G. candida*, *G. maxima*, and *G. caulescens* (Burbidge, 1877; Harrison, 1847; Paxton, 1838), all of which are today considered to be synonyms of *S. speciosa*. Allegedly, some or all of these formerly distinct (and illegitimate) species feature in the genetic background of *Gloxinia* “Fyfiana”, the first reported plant with actinomorphic flowers that was bred in Scotland in 1844–45 (Fyfe, 1879; Harrison, 1847). This particular cultivar occupies an important place in the history of *S. speciosa*, because most modern cultivars have inherited their distinctive flower shape from a single recessive mutation that was disseminated extensively during the early stages of domestication. As this mutation has probably occurred only once, all cultivated forms with actinomorphic flowers are likely descendants of *Gloxinia* “Fyfiana” (Citerne & Cronk, 1999; Dong et al., 2018). The identical mutation was confirmed in multiple cultivars by Dong et al. (2018) who identified a small deletion in the single *CYCLOIDEA*-like gene found in

S. speciosa (SsCYC). The wild-type allele of this and related genes are responsible for suppressing growth in the dorsal organs of wild zygomorphic flowers such as *S. speciosa* and *Antirrhinum majus* (snapdragon). Dong et al. (2018) hypothesized that the mutation first occurred in *G. caulescens* sometime between 1820 and 1832, and remained hidden in the heterozygous genotype during a series of hybridization events between the individuals mentioned above. The homozygous *Gloxinia* “Fyfiana” was then postulated to have arisen in 1844 after a backcross to the original heterozygous form of *G. caulescens*. According to this reconstruction, the mutation remained undetected for 12 to 24 years in the heterozygous form, a situation that we consider to be extremely unlikely for a self-compatible species that was in high demand and under constant sexual and vegetative propagation. Instead, we speculate that the mutation originally arose in *G. maxima*, a generally accepted parent of *G. “Fyfiana”* (Harrison, 1847), and subsequent self-pollination could have then generated homozygous progeny expressing the actinomorphic phenotype in the first generation. Unfortunately, the precise origin of the mutation in SsCYC may be lost to history. The originator of *G. “Fyfiana”*, John Fyfe, published a short article

where he stated: “... the parent plant of *Gloxinia Fyfiana* was profusely dusted with the pollen of *Digitalis purpurea* (foxglove), *Lophospermum scandens* [possibly *L. erubescens*], *Datura wrightii*, [and] *Brugmansia sanguinea*” (Fyfe, 1879). However, because none of these four species are classified in the Gesneriaceae, successful hybridizations with *S. speciosa* are unlikely. Fyfe also failed to mention the number of plants he initially grew from such crosses, nor did he describe any of the siblings of *G. “Fyfiana”*. Thus, neither scenario about the introduction of the mutation in SsCYC can be confirmed through historical records.

4.3 | Domestication founders

The initial movement of *S. speciosa* from Brazil to England in 1815 was followed by 30 years of additional introductions of wild-collected plants (Zaitlin, 2011). Although records detailing the specific collection sites in Brazil are far from exhaustive, a number of articles point towards areas close to or within the city of Rio de Janeiro, such as the Serra dos Órgãos and Corcovado Mountain (Brackenridge, 1886; Hooker, 1842; Paxton, 1846). Our results indicate that this is, in fact, the area from where most of the founder collections could have originated. All three computational approaches, PCA, ADMIXTURE, and FINESTRUCTURE indicate that the wild form “Avenida Niemeyer”, collected in 1975 from a population along the coastal road of that name in the southern part of Rio de Janeiro, is the closest known wild relative of the modern and semidomesticated cultivars. Our findings are in agreement with the results of Zaitlin (2012) who used amplified fragment length polymorphisms (AFLPs) in combination

TABLE 2 Estimates of genetic diversity

Group	N	S	π	θ_w
Domesticated	29	9,303	0.00102	0.00126
Semidomesticated	4	2,874	0.00087	0.00070
Wild	22	21,430	0.00162	0.00309
Test crosses	3	3,999	0.00104	0.00110

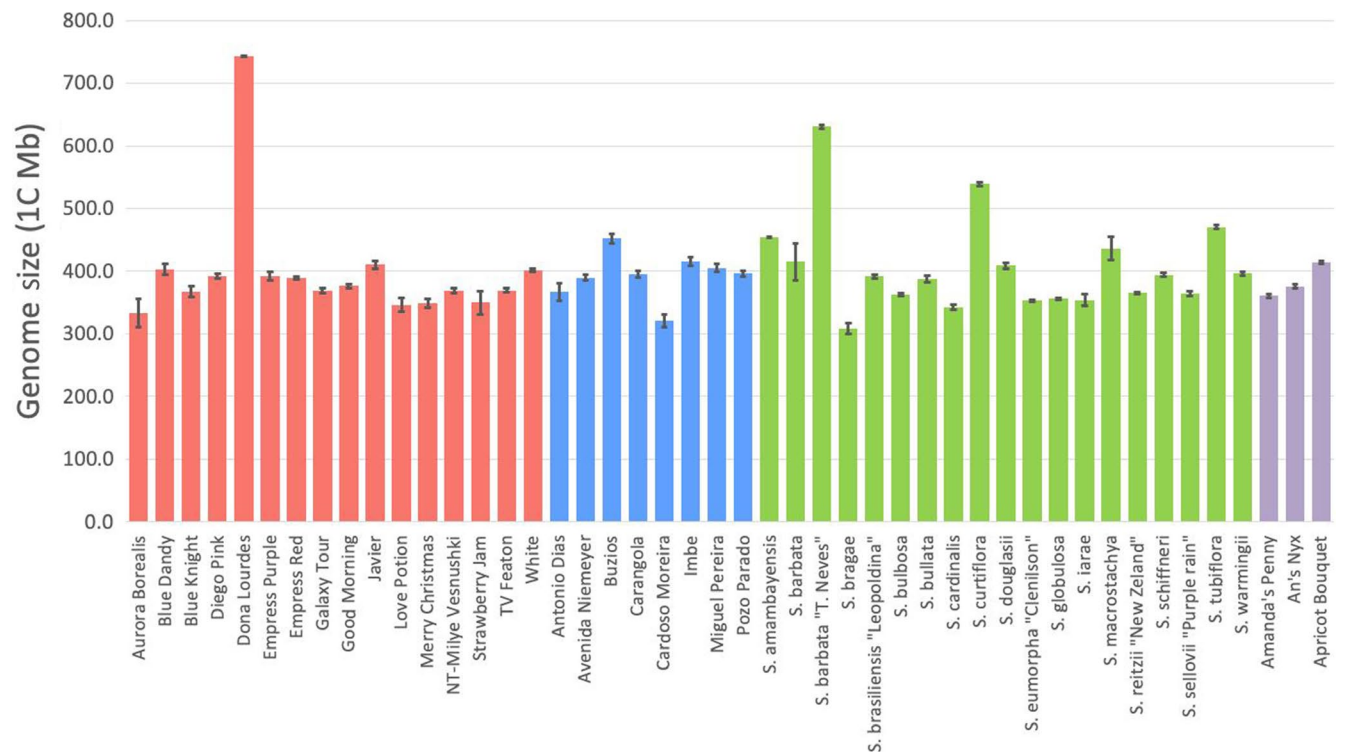


FIGURE 6 Genome sizes in individual plants across the tribe Sinningieae from which fresh leaf tissue was available. Red: domesticated *S. speciosa*, Blue: wild *S. speciosa*, Green: other species in Sinningia, Purple: hybrids

with DNA sequence data from the nuclear ribosomal internal transcribed spacer (nrITS) region to determine the relationships within a much smaller group of wild and cultivated *S. speciosa* plants. He associated seven domesticated cultivars with the wild forms “Avenida Niemeyer” and ‘São Conrado’, also collected along the coast in the city of Rio de Janeiro. Dong et al. (2018) however, obtained contrasting results. Their phylogenetic analysis was based on the *SsCYC* gene sequence and suggests that the actinomorphic allele, which is homozygous in most cultivated material, originated from the wild-type “Cardoso Moreira”. The plant material used in the Dong et al. (2018) study traces to a population located 2–3 km from the town of the same name, located in the northern part of Rio de Janeiro state (more than 300 km distant from Rio de Janeiro city) ~50 km from the southern border of the state of Espírito Santo (Zaitlin, 2012). Both of our “Cardoso Moreira” samples (purple- and pink-flowered forms) clustered together with “São Fidelis”, an accession that was collected 30–40 km from the “Cardoso Moreira” population, validating its geographical origin. The results from this and other studies, combined with the absence of records for early collections in the northern part of Rio de Janeiro state and the fact that plants from the Cardoso Moreira population grow quite tall (rather than as rosettes), steer us away from considering “Cardoso Moreira” as a potential founder. Several studies have previously shown that single genes contain insufficient phylogenetic information, often leading to poor resolution and extensive incongruence among phylogenies (Rokas & Chatzimanolis, 2008).

4.4 | Genetic diversity

Our study does not use the traditional multi-individual sampling approach for each population that is usually employed in “classical” population genetics studies. Instead we sampled a single or few representatives of several wild and breeding populations to perform comparisons across these two groups. A general concern is whether we can recover meaningful information from them, especially because we have already shown that they are genetically structured. Using both empirical and simulated data, St. Onge, Palmé, Wright, and Lascoux (2012) demonstrated that the scattered sampling technique, which is analogous to the one employed here, yields genetic population estimates that are descriptive of true values and tend to outperform those from alternative nonexhaustive sampling strategies such as local or pooled approaches.

Similar to most crop species, the genetic diversity of *S. speciosa* has been reduced significantly during domestication, as suggested by the strong negative changes in our estimates of genetic diversity (S , π , and d_{w}). Perhaps the most striking indicator of the magnitude of the contraction is the fact that 72% of the 21,430 wild SNPs have been fixed in the tested cultivars. This contrasts with rice, for example, where approximately 82% of SNPs ($\text{MAF} > 0.05$) that segregate in the wild ancestor *Oryza rufipogon* also segregate in cultivated *Oryza sativa* (Huang et al., 2012). As previously discussed, the substantial loss in genetic diversity can be attributed mainly to two reasons. First, there is a strong founder effect driven by the small number

of individuals that were collected from an apparently restricted geographical region with limited genetic diversity. These results are supported by our population structure analyses in conjunction with previous phenetic and phylogenetic analyses (Zaitlin, 2012) as well as by written records that disclose collection sites (Brackenridge, 1886; Hooker, 1842; Paxton, 1846). Second, the already narrow pool of alleles in cultivated material could have shrunk even further following the substantial selective sweep around the *SsCYC* gene (Dong et al., 2018) that has practically fixed the mutant actinomorphic allele in the modern cultivars.

We observed additional patterns of genetic diversity that are characteristic of genetic bottlenecks. The number of SNPs with rare alleles ($\text{MAF} \leq 0.05$) dropped from 66% among the wild types to 47% among the cultivars. Such a disproportionate decline in rare alleles is expected in population contractions (i.e. small founder populations), in which minor alleles have less chance of being brought into cultivation and a large proportion are lost immediately. This reduction is usually followed by subsequent losses due to the stronger genetic drift of minor alleles that is inherent in discrete populations of small size. In our dataset, a staggering 86% of rare alleles present in the wild populations were lost during domestication.

There are several examples of changes in genome size driving domestication through whole-genome duplications (Salman-Minkov, Sabath, & Mayrose, 2016) and transposon expansions (Chia et al., 2012). Although we did not count chromosomes to confirm polyploidy in our samples, we identified a semidomesticated cultivar, “Dona Lourdes”, in which the genome size is approximately twice that of the wild-type genomes and is thus likely to be a tetraploid (Figure 5). This cultivar and “Guatapara” are the only two reported cases of potential polyploidy in *S. speciosa* (Zaitlin & Pierce, 2010). Such a low frequency of genome duplications among cultivars provides no evidence of positive selection during domestication. The effects of subtle changes in genome size resulting from increases in transposable element content are much more difficult to detect without full genome resequencing information. However, our preliminary analysis comparing genome size variation across domesticated and wild types strongly suggests that there are no significant expansions driving the process of domestication in *S. speciosa*.

5 | CONCLUSIONS

The genetic analysis of a biodiversity panel of *S. speciosa* has led to the identification of a single extant founder population as the origin of most of the domesticated “gloxinia” cultivars. This founder event, along with a selective sweep, drove a strong genetic bottleneck among commercial cultivars. Despite the loss in genetic diversity, phenotypic diversity has increased as a result of selection for mutations that occurred during domestication, and is not driven by hybridization or polyploidization. These results establish an attractive foundation for the use of *S. speciosa* as a model to study the genetic mechanisms involved in the production of new phenotypes during the plant domestication process.

ACKNOWLEDGMENTS

We thank the Gesneriad Society (www.gesneriadsociety.org) and Virginia Tech for their financial support of the research described in this manuscript. We also thank Jim Roberts (Marriottsville, MD), Charles Marden Fitch (Mamaroneck, NY), Dale Martens (Sherrard, IL), Julia Mavity-Hudson (Nashville, TN), Deborah Ohlgren (Moose Lake, MN), Irina Nicholson (Lakewood, CO), Jon Dixon (Woodside, CA), and Mauro Peixoto (Mogi das Cruzes, SP, Brazil) for the contribution of some of the plant material used in this study. We also acknowledge Qian Zhang and the Virginia Tech Flow Cytometry Lab for their assistance with the genome size estimations. Special thanks go to Jeff Burr, the Virginia Tech greenhouse manager, and to our summer student Holly Waterman for the expert care they provided to keep our plants healthy. We also acknowledge Simona Masiero and Francesca Lopez for providing greenhouse space and technical support with DNA extractions in Milan. Finally, we would like to thank to the Translational Plant Science community and the Department of Horticulture at Virginia Tech for their scientific support.

AUTHORS CONTRIBUTIONS

T.H. designed the experiments, collected the accessions for the biodiversity panel, grew the plants, performed the flow cytometry measurements, extracted the DNA, prepared the GBS libraries, and performed the analyses. E.R. grew the plants and performed the flow cytometry measurements. S.M. grew the plants and extracted the DNA. L.C. and D.H. collaborated in the experimental design. D.Z. collaborated in collecting accessions for the biodiversity panel, growing plants, isolating DNA, and the experimental design. A.B. designed the experiments, collected accessions for the biodiversity panel, and performed the computational analyses. All of the authors participated in the discussion of the results and in the writing and editing of the manuscript.

DATA AVAILABILITY STATEMENT

DNA sequences as raw reads can be found at NCBI under the BioProject PRJNA552788.

ORCID

Aureliano Bombarely  <https://orcid.org/0000-0001-6257-8914>

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22(11), 3179–3190. <https://doi.org/10.1111/mec.12276>
- Aronesty, E. (2013). Comparison of sequencing utility programs. *The Open Bioinformatics Journal*. <https://doi.org/10.2174/1875036201307010001>
- Baute, G. J., Kane, N. C., Grassa, C. J., Lai, Z., & Rieseberg, L. H. (2015). Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytologist*, 206(2), 830–838. <https://doi.org/10.1111/nph.13255>
- Becker, J. (2008). Gesneriad register. *Gesneriads*, 58(3), 44–51.
- Bombarely, A. (2018). Roses for darwin. *Nature Plants*, 4(7), 406–407. <https://doi.org/10.1038/s41477-018-0195-9>
- Brackenridge, W. D. (1886). A chapter in the secret history of the gloxinia. *The Gardeners' Monthly and Horticulturist*, 28, 59–60.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Burbidge, F. (1877). The gesnera family (Gesneraceae). *Cultivated plants: Their propagation and improvement* (pp. 331–339). Edinburgh and London: William Blackwood and Sons.
- Chia, J.-M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., ...Ware, D. (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics*, 44(7), 803–807. <https://doi.org/10.1038/ng.2313>
- Christenhusz, M. J. M., Govaerts, R., David, J. C., Hall, T., Borland, K., Roberts, P. S., ...Fay, M. F. (2013). Tiptoe through the tulips - cultural history, molecular phylogenetics and classification of Tulipa (Liliaceae). *Botanical Journal of the Linnean Society*, 172(3), 280–328. <https://doi.org/10.1111/boj.12061>
- Citerne, H., & Cronk, Q. C. B. (1999). The origin of Peloric Sinningia. *The New Plantsman*, 6(4), 219–222.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ...Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Darwin, C. (1868). *The variation of animals and plants under domestication*. London: John Murray.
- deLaat, A. M. M., & Blaas, J. (1984). Flow-cytometric characterization and sorting of plant chromosomes. *Theoretical and Applied Genetics*, 67(5), 463–467. <https://doi.org/10.1007/BF00263414>
- Dong, Y., Liu, J., Li, P., Li, C., Lu, T., Yang, X., & Wang, Y. (2018). Evolution of Darwin's peloric gloxinia (*Sinningia speciosa*) is caused by a null mutation in a pleiotropic TCP gene. *Molecular Biology and Evolution*, 38(8), 1901–1915. <https://doi.org/10.1093/molbev/msy090>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Fyfe, J. (1879). Hints on the cultivation of choice gloxinias. *The Gardeners' Monthly and Horticulturist*, 21, 359–360.
- Garrison, E. (2016). Vcfliib, a simple C++ library for parsing and manipulating VCF files. <https://github.com/vcfliib/vcfliib>
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv: 1207.3907. <https://arxiv.org/abs/1207.3907>
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, 9(2), e90346. <https://doi.org/10.1371/journal.pone.0090346>
- Harrison, J. (1847). Gloxinia Fyfiana - Mr. Fyfe's Gloxinia. *The Floricultural Cabinet and Florist's Magazine*, 15, 49–51.
- Hooker, W. J. (1842). Gloxinia speciosa; var. macrophylla, variegata. Rough Gloxinia; large variegated-leaved var. *Curtis's Botanical Magazine*, 15, Plate 3934.

- Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., ...Han, B. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 490(7421), 497–501. <https://doi.org/10.1038/nature11532>
- Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K., & Rieseberg, L. H. (2017). The genetics and genomics of plant domestication. *BioScience*, 67(11), 971–982. <https://doi.org/10.1093/biosci/bix114>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., ...Fuller, D. Q. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17), 6139–6146. <https://doi.org/10.1073/pnas.1323964111>
- Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, 8(1), 11–17. <https://doi.org/10.1371/journal.pgen.1002453>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ...Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1), 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martin, M., Piola, F., Chessel, D., Jay, M., & Heizmann, P. (2001). The domestication process of the modern rose: Genetic structure and allelic composition of the rose complex. *Theoretical and Applied Genetics*, 102(2–3), 398–404. <https://doi.org/10.1007/s00120051660>
- Meerow, A. W. (2009). Tilting at windmills: 20 years of *Hippeastrum* breeding. *Israel Journal of Plant Sciences*, 57(4), 303–313. <https://doi.org/10.1560/IJPS.57.4.303>
- Meyer, R. S., DuVal, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. *The New Phytologist*, 196(1), 29–48. <https://doi.org/10.1111/j.1469-8137.2012.04253.x>
- Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics*, 14(12), 840–852. <https://doi.org/10.1038/nrg3605>
- Paxton, J. (1838). *Gloxinia maxima*. *Paxton's Magazine of Botany, and Register of Flowering Plants*, 5, 219–220.
- Paxton, J. (1846). *Gloxinia passinghamii* (Mr. Passingham's *Gloxinia*). *Paxton's Magazine of Botany, and Register of Flowering Plants*, 12, 267–268.
- Perret, M., Chautems, A., & Spichiger, R. (2006). Dispersal-vicariance analysis in the tribe Sinningieae (Gesneriaceae): A clue to understanding the biogeographical history of the Brazilian atlantic forest. *Annals of the Missouri Botanical Garden*, 93(2), 340–358. [https://doi.org/10.3417/0026-6493\(2006\)93\[340:DAITTS\]2.0.CO;2](https://doi.org/10.3417/0026-6493(2006)93[340:DAITTS]2.0.CO;2)
- Perret, M., Chautems, A., Spichiger, R., Kite, G., & Savolainen, V. (2003). Systematics and evolution of tribe Sinningieae (Gesneriaceae): Evidence from phylogenetic analyses of six plastid DNA regions and nuclear ncpGS. *American Journal of Botany*, 90(3), 445–460. <https://doi.org/10.3732/ajb.90.3.445>
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. <https://doi.org/10.1093/molbev/msu136>
- Rokas, A., & Chatzimanolis, S. (2008). From Gene-Scale to Genome-Scale Phylogenetics: the Data Flood in, but the Challenges Remain. In W. J. Murphy (Ed.), *Phylogenomics. Methods in Molecular Biology* (pp. 1–12). Totowa, NJ: Humana Press.
- Salman-Minkov, A., Sabath, N., & Mayrose, I. (2016). Whole-genome duplication as a key factor in crop domestication. *Nature Plants*, 2(8), 1–4. <https://doi.org/10.1038/NPLANTS.2016.115>
- St. Onge, K. R., Palmé, A. E., Wright, S. I., & Lascoux, M. (2012). Impact of sampling schemes on demographic inference: An empirical study in two species with different mating systems and demographic histories. *G3 (Bethesda, MD)*, 2(7), 803–814. <https://doi.org/10.1534/g3.112.002410>
- vanTuyl, J. M., & Arens, P. (2011). *Lilium: Breeding History of the Modern Cultivar Assortment*. In *Acta Horticulturae* (pp. 223–230). Leuven, Belgium: International Society for Horticultural Science (ISHS). <https://doi.org/10.17660/ActaHortic.2011.900.27>
- Zaitlin, D. (2011). A history of the Florist *Gloxinia* in pictures and words. *Gesneriads*, 61(3), 14–27.
- Zaitlin, D. (2012). Intraspecific diversity in *Sinningia speciosa* (Gesneriaceae: Sinningieae), and possible origins of the cultivated Florist's *Gloxinia*. *AoB PLANTS*, 2012, <https://doi.org/10.1093/aobpl/a/pls039>
- Zaitlin, D., & Pierce, A. J. (2010). Nuclear DNA content in *Sinningia* (Gesneriaceae); intraspecific genome size variation and genome characterization in *S. speciosa*. *Genome*, 53(12), 1066–1082. <https://doi.org/10.1139/G10-077>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Hasing T, Rinaldi E, Manrique S, et al. Extensive phenotypic diversity in the cultivated Florist's *Gloxinia*, *Sinningia speciosa* (Lodd.) Hiern, is derived from the domestication of a single founder population. *Plants, People, Planet*. 2019;00:1–12. <https://doi.org/10.1002/ppp3.10065>