

Genomic analyses elucidate S-locus evolution in response to intra-specific losses of distyly *Primula vulgaris*

Emiliano Mora-Carrera¹, Rebecca Lynn Stubbs¹, Giacomo Potente¹, Narjes Yousefi¹, Barbara Keller², Juriaan de Vos³, Peter Szövényi¹, and Elena Conti¹

¹University of Zurich

²University of Zürich

³University of Basel

July 21, 2023

Abstract

Distyly, a floral dimorphism that promotes outcrossing, is controlled by a hemizygous genomic region known as the S-locus. Disruptions of genes within the S-locus are responsible for the loss of distyly and the emergence of homostyly, a floral monomorphism that favors selfing. Using whole genome resequencing data of distylous and homostylous individuals from populations of *Primula vulgaris* and leveraging high-quality reference genomes of *Primula* we tested, for the first time, predictions about the evolutionary consequences of transitions to selfing on S-locus genes. Our results confirm the presence of previously reported homostyle-specific, loss-of-function mutations in the exons of the S-locus gene *CYP?*, while also revealing a previously undetected structural rearrangement in *CYP?* associated with the shift to homostyly. Additionally, we discovered that the promoter region of *CYP?* in distylous and homostylous individuals is identical, suggesting that down-regulation of *CYP?* via mutations in its promoter region is not a cause of shift to homostyly. Furthermore, we found that hemizygoty leads to reduced genetic diversity and less efficient purifying selection in S-locus genes compared to genes outside the S-locus, and that the shift to homostyly further lowers genetic diversity, as expected for mating-system shifts. Finally, we tested, for the first time, long-standing theoretical models of changes in S-locus genotypes during early stages of the transition to homostyly, supporting the assumption that two (diploid) copies of the S-locus might reduce homostyle viability.

1 INTRODUCTION

The repeated shift from outcrossing to selfing is a central topic in plant evolution (Stebbins, 1957; Cutter, 2019). Previous studies used phenotypic traits typically associated with selfing to estimate, for example, the number and tempo of transitions to selfing in phylogenies of ancestrally outcrossing taxa (Goldberg & Iqic, 2012; de Vos *et al.* , 2014). However, missing knowledge of the genes that control mating systems has hindered the study of molecular processes associated with transitions to selfing until recently, especially in non-model organisms. Current advances in genomics now facilitate the identification of the genes and mutations associated with mating-system shifts.

A prime model to investigate the transition from outcrossing to selfing has been the shift from distyly to homostyly in *Primula* (Barrett, 2019). Distyly is characterized by the co-occurrence in populations of two types of self-incompatible individuals, called pins and thrums, distinguished by the reciprocal arrangement of male (anthers) and female (stigma) sexual organs in their flowers (Figure 1A; Ganders, 1979; Lloyd & Webb, 1992; Keller *et al.* , 2014). This floral heteromorphism represents an adaptation for outcrossing reported in at least 26 angiosperm families (Naiki, 2012). Conversely, homostyly is a floral homomorphism that enables selfing. It is characterized by self-compatible individuals bearing flowers with both stigma and anthers at the same level in the corolla tube (Figure 1A; Barrett, 2019). Evidence supporting higher selfing in homostylous

than distylous plants has been reported in diverse taxa (Belaoussoff & Shore, 1995; Schoen *et al.* , 1997; Mora-Carrera *et al.* , 2021). Independent shifts from distyly to homostyly have been documented both within and among species (Zhou *et al.* , 2012; Kissling & Barrett, 2013; de Vos *et al.* , 2014; Ruiz-Martín *et al.* , 2018; Costa *et al.* , 2019).

It has long been known that the S-locus supergene controls distyly and the shift to homostyly (Lewis & Jones, 1992). However, the molecular and functional characterization of the S-locus has been performed only recently. The breakthrough occurred in the *Primula* system, where the S-locus comprises five genes (*CCM^T* , *CYP^T* , *GLO^T* , *KFB^T* , and *PUM^T*) and is hemizygous in thrums (S/0) but absent in pins (0/0; Figure 1A; Li *et al.* , 2016; Potente *et al.* , 2022). Two S-locus genes were recently shown to control key traits in thrum flowers: *GLO^T* determines high anthers, while *CYP^T* determines short stigma and self-incompatibility (Huu *et al.* , 2016; 2020; 2022). Specifically, experimental silencing of *GLO^T* in *Primula forbesii* thrums lowered anther position, producing flowers with both anthers and stigma in the middle of the corolla tube (i.e., short-homostyly). However, self-incompatibility was retained, preventing self-fertilization in short-homostyles (Huu *et al.* , 2020). Conversely, silencing of *CYP^T* in *Primula veris* was associated with both style elongation and loss of self-incompatibility, thus turning self-incompatible thrum flowers into self-compatible, homostylous flowers with both stigma and anthers at the mouth of the corolla tube (i.e., long-homostyly; Huu *et al.* , 2016; 2022). Although both short- and long-homostyly have been reported, the latter type is most common in *Primula* (Charlesworth & Charlesworth, 1979; Lewis & Jones, 1992), likely because self-compatibility in long-homostylous flowers enables self-fertilization and reproductive assurance (Mora-Carrera *et al.* , 2021). Therefore, we hereafter refer to long-homostyly simply as homostyly (Figure 1A).

The shift from distyly to homostyly has been intensely studied in populations of *P. vulgaris* from Somerset, England, that display variation of thrums, pins, and homostyles (Crosby 1940; 1949). Targeted Sanger sequencing of the five individual *CYP^T* exons of homostyles from the mentioned populations revealed that all tested thrums shared the same functional *CYP^T* allele (*CYP^T* -1; Figure 1B). Contrariwise, 21 homostyles harbored six different *CYP^T* alleles, each with a unique, potentially disruptive mutation (*CYP^T* -2 to *CYP^T* -7; Figure 1B; Li *et al.* , 2016; Mora-Carrera *et al.* , 2021). One possible explanation for the lack of shared *CYP^T* mutations among the homostyles is that homostyly evolved independently multiple times. However, the same study also found that six homostyles from two different populations had the same *CYP^T* allele as that of thrums (i.e., *CYP^T* -1). This result raised the possibility that homostyly initially arose via *CYP^T* silencing caused by either a structural rearrangement (such as an inversion or a translocation) involving any of the *CYP^T* exons or an inactivating mutation in the *CYP^T* promoter, followed by multiple, unique mutations in *CYP^T* exons, as those found in *CYP^T* -2 to *CYP^T* -7 (Mora-Carrera *et al.* , 2021; Charlesworth, 2022). Both types of mutations (structural rearrangements in *CYP^T* or silencing of *CYP^T* promoter) cannot be detected using Sanger sequencing of individual *CYP^T* exons. Determining whether homostyly in *P. vulgaris* arose multiple times via independent mutations in *CYP^T* exons or once through a shared structural rearrangement involving *CYP^T* exons or a mutation in the *CYP^T* promoter requires the mapping against a genomic reference of extensive genomic sequences covering both the S-locus and its upstream region. Both types of resources are now available from whole genome resequencing data (WGR) and published genomes for *P. vulgaris* (Cocker *et al.* , 2018) and its close relative *P. veris* (Potente *et al.* , 2022)

Furthermore, the availability of WGR data and reference genomes in the selected study group facilitates the testing of population genetic predictions concerning the evolution of the entire S-locus and S-locus gene-paralogs in thrums, pins and homostyles. First, the thrum-specific segregation of the hemizygous S-locus should cause a 3/4th reduction of effective population size (N_e) (Huu *et al.* , 2016), hence a decrease of genetic diversity in S-locus genes compared to the rest of the genome (Gutiérrez-Valencia *et al.* , 2021). Secondly, hemizygosity could have contrasting effects on the efficacy of purifying selection on S-locus genes. On the one hand, the reduction of N_e in the S-locus should make purifying selection less efficient (Huu *et al.* , 2016). On the other hand, similarly to what happens in the Y sex-chromosome (Gossmann *et al.* , 2011), selection to maintain function of S-locus genes and the exposure of recessive deleterious mutations under hemizygosity should enhance the efficacy of purifying selection. However, the extent to which the efficacy of purifying selection differs between genes within and outside the S-locus remains poorly understood (Potente

et al., 2022). Finally, the transition to homostyly could also reduce genetic diversity in S-locus genes due to increased homozygosity in homostyles (Mora-Carrera *et al.*, 2021). The high-quality annotation of the five S-locus genes and their four paralogs (*CCM1*, *CYP734A51*, *GLO1*, and *KFB1*) in *Primula* (Li *et al.*, 2016; Potente *et al.*, 2022), combined with sequences of these nine genes extracted from WGR data, provides an ideal opportunity to test the above predictions for the molecular evolution of the S-locus.

Additionally, the ability to assess the ploidy level of specific genomic regions extracted from WGR data now enables the testing of predictions about the changing frequencies of haploid and diploid S-locus genotypes during the transition from distyly to homostyly. In a pioneering study, Crosby (1949) proposed a model for how the frequencies of thrum, pin, and homostylous phenotypes change over time (Figure 1C and D). This model rested on then accepted genetic model for the S-locus, which assumed that thrums were typically heterozygous dominant at the S-locus, pins homozygous recessive, and homostyles stemmed from thrums via recombination at the S-locus (Bateson & Gregory, 1905). Crosby assumed that the viability of homozygous homostyles was either 35% lower than or equal to the viability of pins, thrums, and heterozygous homostyles. The assumption of lower viability for homozygous homostyles rested on previous studies by Mather and Winton (1941) proposing that homozygous dominant thrums (S/S) had lower viability than heterozygous thrums. Crosby’s model is applicable also under the recently demonstrated hemizyosity of the S-locus in *Primula*, by assuming that homostyles with a diploid S-locus (S*/S*-genotypes; where S* indicates an S locus with disrupted *CYP^T*) can have either lower or equal viability as that of homostyles with a haploid S-locus (S*/0-genotype), thrums (S/0-genotype), and pins (0/0-genotype; see Figures 1C and D).

In *P. vulgaris*, repeated phenotypic surveys conducted in Somerset, England, have shown that, when homostyles are present at high frequency, thrums tend to be less frequent and, in some cases, absent, compared to pins (Crosby, 1949; Curtis & Curtis, 1985; Mora-Carrera *et al.*, 2021). These findings align with the predictions of Crosby’s model under lower viability of S*/S*-homostyles (Figure 1C). However, one Somerset population consisted exclusively of homostyles (Curtis & Curtis, 1985; Mora-Carrera *et al.*, 2021), suggesting that the fixation of homostyly is possible, as expected under the model with equal viability for S*/S*- and S*/0-homostyles. However, previous phenotypic surveys could not discriminate between S*/S*- and S*/0-genotypes for homostyles. Recently developed sequencing technologies enable the estimation of sequencing depth at the S-locus (Gutiérrez-Valencia *et al.*, 2022), allowing us to determine whether the S-locus is haploid or diploid in both homostylous and thrum individuals. Therefore, it is now possible to estimate whether the observed frequencies of S*/S*- and S*/0-homostyles in natural populations support the model assuming lower or equal viability for S*/S*- genotypes in relation to the other genotypes.

Here, we analyze WGR data from nine populations of *P. vulgaris* with varying frequencies of pins, thrums, and homostyles, to answer the following questions: 1) Do all homostyles carrying different disrupted *CYP^T* alleles share either a mutation in the promoter region and/or a structural rearrangement involving *CYP^T* exons that might disrupt *CYP^T* function, allowing for the possibility of a single origin of homostyly? 2) Do S-locus genes have lower genetic diversity and efficacy of purifying selection than their paralogs? 3) Do homostyles have lower genetic diversity in S-locus genes than thrums? 4) Do observed frequencies of S*/0- and S*/S*-homostyles in natural populations better match genotypic frequencies predicted under the assumptions of lower or equal viability for S*/S*- homostyles? Our study illustrates how knowledge of the genes controlling mating systems combined with high-quality genomic resources generates novel insights into the genotypic changes and evolutionary consequences associated with phenotypic transitions from outcrossing to selfing.

2 MATERIAL AND METHODS

2.1 Study species

Primula vulgaris (primrose) is an ancestrally heterostylous, diploid (2n=22) perennial, rosette-forming plant blooming from February to April. Distylous populations of *P. vulgaris* occur across Eurasia, including in Turkey, most of Western Europe, larger Mediterranean Islands, and all British Isles (Jacquemyn *et al.*, 2009). In contrast, populations with varying frequencies of homostyles have been discovered only in Somerset and Chiltern Hills, England (Crosby, 1940, 1949), and one population in the Netherlands (Barmantlo *et al.*, 2017).

Habitat fragmentation due to intensive pastoral and agricultural activities in these areas were suggested as the potential selective pressure favoring the shift to self-fertilizing homostyles in these populations (Mora-Carrera *et al.*, 2021).

2.2 Sampling and sequencing of plant material

In spring 2019, we collected leaf tissue from 105 individuals of *P. vulgaris*, comprising 74 heterostyles (37 pins and 37 thrums) and 31 homostyles. The samples were obtained from nine populations, including: six dimorphic populations (i.e., pins and thrums) from Turkey (TR-D), Slovakia (SK-D), Switzerland (CH-D) and England (EN1-D, EN2-D, and EN3-D); two trimorphic (i.e., pins, thrums, and homostyles) populations from England (EN4-T and EN5-T); and one monomorphic (i.e., only homostyles) population from England (EN6-M) (Table 1). Populations EN4-T, EN5-T and EN6-M corresponded to populations T04, T07, and M01, respectively, from our previous study (Mora-Carrera *et al.*, 2021). Moreover, the 11 homostyles from EN6-M analyzed here included three of the homostyles carrying *CYP^T*-1 from population M01 of Mora-Carrera *et al.* (2021; Figure 1B). DNA extractions were performed using the Maxwell extraction method (Promega, USA) at the Functional Genomics Center Zurich (Zurich, Switzerland). Library preparation and paired-end sequencing (150bp) were conducted by RAPiD GENOMICS (Gainesville, Florida, USA) using Illumina Novaseq6000 platform, generating 7,077,866,510 paired-end sequencing reads with an average sequencing depth of 18.9 (\pm SD = 3.08).

2.3 Mapping and variant calling of WGR data to *CYP^T* from the *P. vulgaris* genome

To determine the sequences of all five exons and four introns of *CYP^T*, along with its up- and down-stream intergenic regions, we mapped WGR reads from all 105 individuals to a previously published genome of *P. vulgaris* (Cocker *et al.*, 2018). We opted to use the less contiguous genome of *P. vulgaris* (Cocker *et al.*, 2018) rather than the highly contiguous, chromosome-scale genome of *P. veris* (Potente *et al.*, 2022) to enhance the likelihood of accurate mapping of WGR reads to intronic as well as up- and downstream intergenic regions of *CYP^T*. Prior to mapping, we also produced a reference S-locus assembly for *P. vulgaris* by replacing all S-locus contigs (LH_v2.0002458, LH_v2.0067593, LH_v2.0003915, and LH_v2.0000241) in the genome of *P. vulgaris* with a published 450 kb sequence of the *P. vulgaris* S-locus (Li *et al.*, 2016). To determine the position of *CYP^T* in the reference S-locus assembly, we aligned the coding sequence of *CYP^T* against the reference using the *query* function of *blastn* with default parameters, which is part of the NCBI BLAST+ toolkit v2.6.0 (Camacho *et al.*, 2009).

Prior to mapping, Illumina adapters were clipped from raw reads with Trimmomatic v0.38 (Bolger *et al.*, 2014) using default parameters. Mapping was performed using BWA-mem v7.17 (Li & Durbin, 2009) with default parameters. As negative control, pin individuals (0/0) were included in the analysis and, as expected, none of the sequencing reads from the 37 pins mapped to the S-locus. Duplicated reads were marked with the MarkDuplicates tool included in Picard v2.18.4 (<http://broadinstitute.github.io/picard/>). Variant calling of SNPs and indels for the S-locus was conducted using HaplotypeCaller, implemented in the Genome Analysis Toolkit (GATK) v4.1.2.0 (McKenna *et al.*, 2010) pipeline. Finally, SNP variants were filtered from the Variant Call Format (VCF) file using the SelectVariants with following filters: quality-by-depth (QD) > 2.0; mapping quality (MQ) > 40.0; strand bias (FS) < 60.0; mapping quality rank-sum test (MQRankSum) > -12.5; a rank-sum test (ReadPositionRankSum) > -8.0; site read depth (DP < $\frac{1}{2}$ X) || (DP > 3X). Additionally, sites with fixed heterozygosity (i.e., InbreedingCoeff < -0.99) likely representing incorrect SNP calling (O’Leary *et al.*, 2018; Pavanet *et al.*, 2020) were filtered out.

2.4 Identification of mutations putatively disrupting *CYP^T* function in *P. vulgaris*

*Disruptive mutations in *CYP^T* coding regions* - To identify potential homostyle-specific, loss-of-function *CYP^T* mutations, including non-synonymous mutations, insertions, and deletions, we compared the sequences of the five *CYP^T* exons in 31 homostyles with the functional *CYP^T* allele of the 37 thrums. For this, we extracted the respective sequences of the five *CYP^T* exons from the S-locus VCF file using the *intersect* function included in BEDtools v2.29.2 (Quinlan & Hall, 2010) and converted them into a single FASTA file using *vcf2phylyp.py* (<https://github.com/edgardomortiz/vcf2phylyp>). The sequence alignment of all exons and

the detection of putatively disruptive mutations in CYP^T of homostyles and thrums were performed with MEGA X (Kumar *et al.*, 2018). Finally, we compared the resulting sequence alignment with an alignment of previously detected mutations in CYP^T exons reported by Mora-Carrera *et al.* (2021).

Structural rearrangements in CYP^T - To determine whether the shift to homostyly is associated with structural rearrangements involving CYP^T exons, we examined the paired-end sequencing reads mapped to the introns and exons of CYP^T in both thrums (as a reference) and homostyles using the Interactive Genomic Viewer (IGV) v2.8.6 (Robinson *et al.*, 2011). Translocations can be identified by analyzing the mapped paired-end reads, where one read is mapped to one position in the genome (e.g., CYP^T in the S-locus), while its mate-pair is mapped to a different position, either in the same or different chromosome. Inversions can be detected by comparing the orientation of the mapped read-pairs to CYP^T in the reference genome. If there is a small inversion, both mapped paired-end reads should be oriented in the same direction (-- or --), whereas in the absence of a structural change normal mapped paired-end reads should be oriented towards each other (--). Finally, deletions are characterized by drops in sequencing read coverage at specific positions in the genome, in this case, within CYP^T .

Disruptive mutations in CYP^T promoter region - To determine whether mutations in the promoter region are responsible for CYP^T loss of function in homostyles, we conducted an analysis to identify the putative CYP^T transcription-factor binding-site and searched it for homostyle-specific mutations. It is expected that the 3kb region upstream of a gene of interest contains its promoter, including the transcription-factor binding-site (Yu *et al.*, 2016), thus we first extracted and aligned the 3kb sequence upstream of CYP^T exon 1 from 20 thrums in dimorphic populations TR-D, SK-D, CH-D, and EN1-D of *P. vulgaris*. We then added the 3kb sequence upstream of CYP^T from the *P. veris* high-quality reference genome (Potente *et al.*, 2022) to the aligned dataset. Secondly, we inspected the aligned upstream sequence of CYP^T to look for any single nucleotide polymorphism (SNP) that was fixed in homostyles but absent in thrums. Thirdly, to identify enriched motifs of transcription binding-sites in the aligned dataset, we employed the Simple Enrichment Analysis (SEA) implemented in the Multiple Em for Motif Elicitation (MEME) suite program v5.5.0 (McLeay & Bailey, 2010). We used the *Arabidopsis thaliana* DAP motifs database (O'Malley *et al.*, 2016) and default parameters. The identified enriched motif sequences were compared between the 37 thrums and 31 homostyles by aligning these regions with MEGA X (Kumar *et al.*, 2018). To determine the potential function of the identified enriched motifs, we consulted The *Arabidopsis* Information Resource (TAIR) (<http://www.arabidopsis.org>).

2.5 Genetic variation in S-locus genes and their paralogs

To quantify genetic variation in the five S-locus genes (CCM^T , CYP^T , GLO^T , KFB^T , PUM^T) and their four paralogs ($CCM1$, $CYP734A51$, $GLO1$, $KFB1$), we calculated nucleotide diversity (π) at synonymous (π_S) and non-synonymous (π_N) sites. For this analysis, we mapped the WGR reads of *P. vulgaris* to the chromosome-scale reference genome of *P. veris* (Potente *et al.*, 2022) because it provides a better annotation of all S-locus genes and their paralogs outside the S-locus than the *P. vulgaris* reference genome. Prior to mapping, we annotated sites at 4- and 0-fold degenerate sites as synonymous and non-synonymous sites, respectively, for all nine genes using the script NewAnnotateRef.py (https://github.com/fabbyrob/science/tree/master/pileup_analyzers [last accessed on March 25, 2022]) (Williamson *et al.*, 2014). Mapping of sequencing reads, variant calling, and filtering were performed with BWA-mem v7.17 (Li & Durbin, 2009) and GATK v4.1.2.0 (McKenna *et al.*, 2010) as described in section 2.3. Subsequently, we split the VCF files of each of the nine genes into synonymous and non-synonymous sites using BEDtools v2.29.2 (Quilan & Hall, 2010). Using these VCF files, we estimated π_S and π_N with pixy v.1.0.0 (Korunes & Samuk, 2021). Specifically, we estimated π_S and π_N of S-locus paralogs in heterostyles (i.e., pins and thrums combined) and homostyles, while the estimates of π_S and π_N for S-locus genes included thrums (since pins lack the S-locus) and homostyles. Finally, we estimated the strength of purifying selection in all S-locus genes and their paralogs by calculating the ratio of nucleotide diversity at non-synonymous vs. synonymous sites (π_N/π_S).

2.6 S- locus genotypes in natural populations of *P. vulgaris*

To determine whether homostyles and thrums have either a haploid (i.e., S*/0 or S/0, respectively) or diploid (i.e., S*/S* or S/S, respectively) S-locus, we calculated the relative sequencing depth of the S-locus ($Rel_{S-locus\ depth}$), as follows. We first estimated the average site depth of the S-locus coding regions ($depth_{S-locus}$) and of the genome-wide coding regions ($depth_{Genome-wide}$) from filtered VCF files using BCFtools v1.9, as indicated in section 2.5, then we calculated $Rel_{S-locus\ depth}$ as $depth_{S-locus}/depth_{Genome-wide}$. This normalization allowed us to account for differences of sequencing depth among individuals. A $Rel_{S-locus\ depth}$ value of approximately 0.5 ± 0.25 indicates a haploid S-locus (S*/0 or S/0 for homostyles and thrums, respectively), while a value close to 1 ± 0.25 indicates a diploid S-locus (S*/S* or S/S for homostyles and thrums, respectively).

To determine whether the proportions of individuals carrying 0/0-, S/0-, S*/S*-, and S*/0-genotypes in natural populations support the model with lower or the one with equal viability of S*/S*-homostyles (Figure 1C and D), we compared the observed frequencies of the four genotypes in EN4-T, EN5-T, and EN6-M with the expected genotype frequencies predicted by Crosby’s model (1949; Figure 1C and D). First, using Crosby’s equations (1949), we calculated the expected frequencies of all four genotypes (0/0-, S/0-, S*/S*-, and S*/0) at generations 10, 20, 30, and 40 after the onset of homostyly, specifying either lower ($v = 0.65$) or equal ($v = 1$) viability of S*/S*-homostyles. Since natural frequencies of the four genotypes might be compatible with levels of viability not included in Crosby’s (1949) model, we additionally estimated expected genotype frequencies under $v = 0.9, 0.8, 0.7, 0.6,$ and 0.5 . Viability values below 0.5 were not used because phenotypic frequencies reflecting these conditions (roughly equal frequencies of pins and homostyles and absence of thrums) have never been reported in natural populations.

Secondly, we estimated the observed frequencies of all four genotypes as follows. We tallied the raw counts of pins (0/0-genotypes) and thrums (S/0-genotypes) in EN4-T and EN5-T based on previous population surveys (T04 and T07, respectively, in Mora-Carrera *et al.*, 2021; Material and Methods section 2.2, here). Furthermore, we calculated the number of homostyles with S*/0- and S*/S*-genotypes in EN4-T, EN5-T, and EN6-M by multiplying the proportion of S*/0- and S*/S*-genotypes reported here (see Results section 3.3) by the raw number of homostyles in each of the three populations. Finally, to determine if observed and expected genotype frequencies are compatible with Crosby’s model under lower or equal viability of S*/S*-homostyles at 10, 20, 30, and 40 generations, we used chi-squared tests with Bonferroni corrections. A significant difference between observed and expected frequencies indicates that observed genotype frequencies are not compatible with Crosby’s model, while non-significant results indicate that they are compatible with it.

3 RESULTS

3.1 Mutations putatively disrupting CYP^T function in *P. vulgaris*

Disruptive mutations in CYP^T coding region - Consistent with expectations (Huu *et al.*, 2016; Gilmartin *et al.*, 2016; Potente *et al.*, 2022) and previous findings by Mora-Carrera *et al.* (2021), we detected putatively disruptive CYP^T mutations exclusively in homostyles (Figure 2). Specifically, we identified a non-synonymous mutation [Serine to Stop codon] in exon 2 (referred to as allele CYP^T-2 in Mora-Carrera *et al.*, 2021) and an 8bp deletion in exon 1 that shifts the open reading frame of CYP^T (referred to as allele CYP^T-6 in Mora-Carrera *et al.*, 2021; Figure 1B). Both mutations in CYP^T-2 and CYP^T-6 introduce a premature stop codon, likely causing incomplete CYP^T translation. In our dataset, CYP^T-2 occurred in two homostyles of population EN4-T and all ten homostyles of EN5-T, while CYP^T-6 occurred in eight homostyles of EN4-T.

Additionally, we discovered four non-synonymous CYP^T mutations in thrums never reported before. These mutations were observed in different exons and individuals. Specifically, in the 37 thrums analyzed, two non-synonymous mutations were detected in exon 1 of four out of five individuals from SK-D; one non-synonymous mutation was found in exon 3 of two thrums from EN5-T; and one non-synonymous mutation was identified in exon 5 of one individual from CH-D. In contrast to the non-synonymous mutations discovered in homostyles (see above), none of the above mentioned mutations found in thrums introduced a premature stop codon.

Structural rearrangements in CYP^T - A drop in sequencing coverage compared to the rest of the genome

revealed for the first time a large deletion (ca. 2150 bps) encompassing *CYP^T* exon 1 and its upstream and downstream regions (Figure 3), which was exclusively detected in the 11 homostyles from population EN6-M (Figures 2 and 3). Additional analyses of WGR data in IGV showed that a small number of reads within this region had paired-mates mapping to different chromosomes (not shown), suggesting a local deletion of exon 1 and its potential translocation to another location in the genome. Notably, the remaining four *CYP^T* exons in these homostyles showed sequencing read coverage values comparable to average sequence read coverage of the genome and did not exhibit any additional disruptive mutations. Apart from the local deletion of exon 1, we did not identify additional inversions or translocations involving the remaining *CYP^T* exons. We designated this previously unreported allele as *CYP^T* -8 (Figure 1B).

Mutations in putative promoter region of CYP^T- We identified 20 conserved motifs in the region upstream of *CYP^T* that are potential binding sites for seven different transcription-factor families, as summarized in Table 2. Among these motifs, ten were associated with various subfamilies of the DREB (Dehydration-Responsive Element-Binding protein) group within the ERF/AP2 transcription-factor family, four were related to the MYB (Myeloblastosis viral oncogene) transcription factor, two were associated with the basic leucine-zipper transcription factor, and one each were linked to the telomere-repeat binding factor, basic pentacysteine, CHC (Clathrin Heavy Chain) protein, and a FAR1 (Fatty Acyl-CoA Reductase 1) related protein. All these conserved motifs were found within a range of 1361 to 15 base pairs upstream of *CYP^T* exon 1. Since *CYP^T* is a brassinosteroid-degrading enzyme expressed exclusively in the style in *P. vulgaris* (Huu *et al.*, 2016), we focused on eight conserved motifs involved in brassinosteroid regulation and associated with genes expressed in the carpel of *Arabidopsis thaliana*, based on the TAIR database (Table 2). Inspection of the aligned sequences upstream of *CYP^T* indicated that the 3kb sequence upstream of *CYP^T* in 20 out of the 31 homostyles were identical to those of the 37 thrums with a functional S-locus (Supplementary Data). However, in the 11 homostyles with the 2150 bp deletion affecting *CYP^T* exon 1 and surrounding upstream and downstream regions (i.e., *CYP^T* -8 allele: see above), all eight conserved motifs were absent from the promoter region. In summary, *CYP^T* promoter regions are conserved and likely functional in most analyzed homostyles, while the 11 homostyles with a deleted *CYP^T* exon 1 (*CYP^T* -8 allele) also appear to lack the promoter region.

3.2 Genetic variation in S-locus genes and their paralogs

Our results showed that, as expected due to S-locus hemizyosity, π_S was lower in S-locus genes of thrums (0.0012 ± 0.0006 [mean \pm SE]) than in their paralogs in pins and thrums (0.0034 ± 0.0008 ; Table 3A). Moreover, π_S in homostyles was zero for both S-locus genes and their paralogs, except for *GLO^T* and *CYP734A51*, where π_S was extremely low (0.0012 and 0.0008, respectively; Table 3B), thus supporting the prediction that the shift to predominant selfing should be associated with lower π_S in homostyles than in heterostyles.

Furthermore, our results indicated that, on average, π_N/π_S values were higher for S-locus genes of thrums than for their paralogs in pins and thrums (1.01 ± 0.37 vs. 0.53 ± 0.29 , respectively; Table 3A), implying lower purifying selection in S-locus genes. Additionally, π_N/π_S was lower in *KFB1* than in *KFB^T* ($\pi_N/\pi_S = 0.23$ and 1.83, respectively), but higher in *CCM1* than in *CCM^T* ($\pi_N/\pi_S = 1.5$ and 0.91, respectively; Table 3A), implying stronger and weaker purifying selection on the two paralogs than on their respective S-locus genes, respectively. Finally, we found that, within the S-locus, π_N/π_S was higher for *CCM^T*, *KFB^T*, and *PUM^T* ($\pi_N/\pi_S = 0.91$, 1.83, and 10.36, respectively; Table 3A) than for *CYP^T* ($\pi_N/\pi_S = 0.28$), whereas π_N/π_S in *GLO^T* was not calculated due to the lack of variation at synonymous sites in this gene. In homostyles, π_N/π_S was zero for most S-locus genes and their paralogs, except for *CYP734A51*, due to the lack variation at synonymous and non-synonymous sites (Table 3B).

3.3 S-locus genotypes in natural populations of *P. vulgaris*

Relative S-locus sequencing depth (*Rel_{S-locus depth}*) allowed us to determine S-locus ploidy in the analyzed thrums and homostyles. Of the 37 thrums collected from six dimorphic and two trimorphic populations across our sampling range, 34 had a haploid S-locus (i.e., S/0) and three had a diploid S-locus (S/S; Figure 4). Of these three thrums, two were homozygous for the functional copy of *CYP^T* (*CYP^T*-1/*CYP^T*-1; i.e.,

S/S) and belonged to one trimorphic and one dimorphic population each, respectively (Table 1), while one thrum was heterozygous and carried one functional and one disrupted copy of $CYP^T(CYP^T-1/CYP^T-2$; i.e., S/S*). Of the 31 homostyles collected from two trimorphic (EN4-T and EN5-T) and one monomorphic population (EN6-M) in England, 10 (32%) had a haploid S-locus (S*/0) and 21 (68%) had a diploid S-locus (S*/S*) (Figure 4). Specifically, S*/0- homostyles represented 40% and 60% of homostyles in the trimorphic populations EN4-T and EN5-T, respectively, while all tested homostyles of the monomorphic population EN6-M had the S*/S*- genotype (Table 1; purple triangles in Figure 4).

We also calculated expected frequencies and observed frequencies of 0/0, S/0-, S*/0-, and S*/S*-genotypes under different assumptions for viability of S*/S*-genotypes compared to the other three genotypes and after different numbers of generations following the origin of homostyles (Table 4). The results of chi-squared tests showed that non-significant differences were found only in seven cases, of which four occurred in the monomorphic, homostylous population EN6-M and three in the two trimorphic populations EN5-T and EN4-T. Specifically, in EN6-M, observed frequencies matched expected frequencies at generations 30 and 40 under the assumption of equal and slightly lower viability ($v = 1$ and 0.9 , respectively) for the S*/S*-homostyles. In EN5-T, observed frequencies matched expected frequencies at 20 generations under lower viability ($v = 0.8$) of S*/S*-homostyles. Finally, in EN4-T, observed frequencies matched expected genotypic frequencies at generations 20 and 30 under levels of viability for S*/S*-homostyles close or equal to those of the model in Figure 1C (i.e., $v = 0.70$ and 0.65 , respectively).

4 DISCUSSION

We integrated whole genome resequencing data from a comprehensive sampling of long-monitored populations of *P. vulgaris* with knowledge of the recently assembled genomes of *P. vulgaris* and *P. veris* to explain the causes and consequences of the transition from distyly to homostyly. We identified a novel loss-of-function structural mutation in CYP^T associated with the transition to homostyly that had remained undetected using exonic Sanger sequencing (Mora-Carrera *et al.*, 2021). Importantly, we found no evidence for a potential single origin of homostyly in *P. vulgaris* via mutations in CYP^T promoter region or structural mutations involving CYP^T exons, thus the previously supported hypothesis of multiple transitions to homostyly via independent loss-of-function mutations in CYP^T exons stands (Mora-Carrera *et al.*, 2021). Furthermore, population genetic analyses validated theoretical expectations for the evolutionary consequences of hemizyosity on S-locus genes and revealed differences of selective constraints among S-locus genes. Finally, the genomic resources newly available in the *Primula* system enabled, for the first time, the testing of long-standing predictions on changing frequencies of S-locus genotypes during intraspecific transitions from distyly to homostyly, partially supporting the possible role of viability differences between homostyles with haploid vs. diploid S-locus genotypes in preventing the fixation of homostyly. Jointly, our study provides a detailed overview of the early molecular- and population genetic causes and consequences of mating-system transitions.

Genetic basis of transitions from distyly to homostyly in Primula vulgaris

Shifts from outcrossing to selfing are common in flowering plants and can be caused by loss-of-function mutations in the genes of interest, structural rearrangements of their exons, or mutations in their promoters (Shimizu & Tsuchimatsu, 2015). One question concerns whether disruptive mutations in the alleles that determine outcrossing act dominantly or recessively. In Brassicaceae, loss of self-incompatibility often stems from mutations in dominant alleles of genes controlling this trait (Busch *et al.* , 2011; Nasrallah *et al.* , 2017; Tsuchimatsu *et al.*, 2012; Bachmann *et al.*, 2019), although in *Arabidopsis lyrata* loss of self-incompatibility caused by mutations in recessive alleles has also been discovered (Mable *et al.* , 2017). In *Primula* , the S-locus controlling distyly is hemizygous in both *P. vulgaris* and *P. veris* (Huu *et al.*, 2016; 2020), but previous models assumed S-locus heterozygosity, with thrum phenotype associated with the dominant S-locus allele (Bateson & Gregory, 1905). Based on this model and greenhouse crossing experiments, Crosby (1949) assumed that S-locus alleles associated with homostyly should be recessive. Previously, we documented seven CYP^T haplotypes (CYP^T-2 to CYP^T-7) with putative loss-of-function mutations that are occurring exclusively in *P. vulgaris* homostyles (Mora-Carrera *et al.*, 2021). Two of these haplotypes (CYP^T-2 and CYP^T-6) have an early stop codon causing premature termination of translation. Our results indicate

that, when hemizygous or homozygous, these disrupted CYP^T alleles lead to homostyly. However, CYP^T -2 behaves recessively in the single heterozygous individual carrying one functional and one disrupted copy of CYP^T (i.e., CYP^T -1/ CYP^T -2, represented as S/S* in Figure 4), determining a thrum phenotype. This finding aligns with results of crossing experiments between thrums and homostyles of *Primula oreodoxa* showing that S* is recessive to S when the two alleles co-occur (Yuan *et al.*, 2018), corroborating Crosby’s prediction (1949). Therefore, our results indicate that the presence of a disrupted CYP^T allele does not alter the thrum morph when paired with a functional CYP^T allele, hence the disrupted allele acts recessively.

In addition to loss-of-function mutations in coding regions, mating-system shifts can also stem from transcription silencing or exonic rearrangements in pertinent genes (Chakraborty *et al.*, 2023). For instance, down-regulation caused by transposon-like insertions in the promoter regions of the male-determining self-incompatibility genes *MGST* and *BnSP11-1* trigger the shift from self-incompatibility to self-compatibility in *Prunus avium* and *Brassica napus*, respectively (Gao *et al.*, 2016; Ono *et al.*, 2020). In *Primula vulgaris*, previous Sanger sequencing of individual CYP^T exons identified homostyles with an apparently functional CYP^T allele (CYP^T -1; Figure 1B), suggesting that homostyles might also arise through CYP^T silencing caused by a disruptive mutation in its promoter region or exonic rearrangements in CYP^T not detectable via Sanger sequencing (Mora-Carrera *et al.*, 2021). However, 20 of the 31 homostyles analyzed in the present study had a promoter region identical to that of 37 thrums, implying that the shift to homostyly in these plants was likely caused by loss-of-function mutations in CYP^T exons rather than in its promoter. The remaining 11 homostyles were characterized by a large 2150 bp deletion that eliminated both CYP^T exon 1 and its promoter region (Figure 3). Thus, our current results do not support the conclusion that mutations in the promoter region or exonic rearrangements in CYP^T can alone cause the shift to homostyly in *P. vulgaris*.

The evidence above also has implications for determining whether homostyly arose once or multiple times in *P. vulgaris*. The single origin of homostyly, followed by independent mutations in CYP^T , would have been supported if all studied homostyles had shared the same promoter mutation or rearrangement in CYP^T . However, this is not the case, favoring the hypothesis of multiple origins of homostyly via independent mutations in CYP^T exons, as previously proposed (Mora-Carrera *et al.*, 2021). Nevertheless, a study of a single homostyle from Chiltern Hills, England (population not included in our analyses), found reduced expression of CYP^T when compared to a thrum, suggesting epigenetic silencing might play a role in the shift to homostyly (Huu *et al.*, 2016). The mentioned study however did not provide sequences of CYP^T exons, thus it remains unknown whether they contained any potentially disruptive mutations in the coding region. Therefore, transcriptome analyses of homostylous flowers are necessary to conclusively discard the possibility that disruptive promoter mutations causing reduced CYP^T expression might also cause the shift to homostyly.

Finally, it remains to be explained why the three homostyles previously thought to have the functional CYP^T -1 allele based on Sanger sequencing of the five CYP^T exons (Mora-Carrera *et al.*, 2021) were here found to contain the 2150 bp deletion including exon 1 (i.e., CYP^T -8 haplotype: see Figures 1B, 2, and 3). A possible explanation is that exon 1 was deleted from the S-locus (causing CYP^T loss of function, hence homostyly) and translocated to a highly repetitive genomic region. The translocation could have allowed targeted amplification and subsequent Sanger sequencing using exon-1-specific PCR primers, while preventing exon-1 detection via next generation sequencing due to biases arising, for example, during genomic DNA sonication used to produce short DNA fragments prior to short-read library preparation (Poptsova *et al.*, 2014; Garafutdinov *et al.*, 2016; Jennings *et al.*, 2017). Notably, a few low-quality sequencing reads did map to CYP^T exon 1, suggesting this exon is indeed present in the genome of these homostyles but was not successfully sequenced using short-read sequencing methodology. Long-read sequencing technologies capable of sequencing through repetitive regions would be necessary to definitively resolve whether CYP^T exon 1 was translocated to a highly repetitive genomic region in these homostyles. To summarize, our findings indicate that the homostyles previously identified as having a functional CYP^T allele in fact possess a disrupted CYP^T allele due to exon 1 deletion (designated CYP^T -8 allele: Figure 3). Overall, these results emphasize that not only non-synonymous mutations or small deletions, but also structural rearrangements such as large deletions and translocations can cause mating-system transitions.

Population genetic consequences of hemizygoty and the transition to homosity on S-locus genes

One of the most notable, recent discoveries on the S-locus is that it is hemizygous and present only in thrums in all systems where its genetic architecture has been investigated, including in *Primula* (Huu *et al.*, 2016; Li *et al.*, 2016), *Turnera* (Shore *et al.*, 2019), possibly *Fagopyrum* (Matsui & Yasui, 2020), *Linum* (Gutiérrez-Valencia *et al.*, 2022), and *Gelsemium* (Zhao *et al.*, 2023), representing different families and orders of flowering plants. The hemizygoty of the S-locus should affect patterns of molecular diversity. Specifically, tight genetic linkage provided by recombination suppression and thrum-specific occurrence of S-locus genes are expected to cause a reduction of genetic diversity inside the S-locus compared to other genomic regions (Gutiérrez-Valencia *et al.*, 2021). Our results demonstrate that the mean π_S of S-locus genes (CCM^T , CYP^T , GLO^T , and KFB^T ; π_S : 0.0012 ± 0.0006) is lower than that of their paralogs ($CCM1$, $CYP734A51$, $GLO1$, and $KFB1$; π_S : 0.0034 ± 0.0007) located elsewhere in the genome (Table 3). This result corroborates previous studies that found an overall decrease in genetic diversity between the S-locus and its flanking regions in *Primula* (Potente *et al.*, 2022) and *Linum* (Gutiérrez-Valencia *et al.*, 2022). Thus, our work confirms predictions that S-locus genomic architecture influences patterns of molecular evolution in S-locus genes.

The non-recombining nature of the S-locus also affects its response to natural selection. Specifically, absence of recombination is expected to reduce S-locus Ne , decreasing the efficacy of purifying selection (Gossmann *et al.*, 2011) on S-locus genes compared to genes outside the S-locus. Thus, increased degeneration due to accumulation of deleterious mutations is expected in these genes (Charlesworth & Charlesworth, 2000; Huu *et al.*, 2016). Conversely, if selection to maintain function were strong, purifying selection should be more efficient on S-locus genes than on their paralogs due to the dominant nature of the hemizygous S-locus (Gutiérrez-Valencia *et al.*, 2021; Potente *et al.*, 2022). Regarding the former hypothesis, a greater accumulation of transposable elements in S-locus non-coding regions compared to the rest of the genome was detected, supporting the conclusion that purifying selection on the S-locus might be relaxed (Potente *et al.*, 2022). However, whether the efficacy of purifying selection differs between coding regions of S-locus genes and their paralogs remains poorly understood (Potente *et al.*, 2022). Our results indicate that, on average, S-locus genes exhibit higher accumulation of non-synonymous mutations than their paralogs, implying purifying selection is less effective on the former ($\pi_N/\pi_S = 1.01 \pm 0.37$ and 0.53 ± 0.25 , respectively; Table 3), conformant with predicted effects of reduced S-locus Ne . However, patterns of selective constraints within and outside the S-locus vary among gene duplicates. For example, the strength of purifying selection is similar between CYP^T and $CYP734A51$, albeit slightly stronger in the former ($\pi_N/\pi_S = 0.28$ and 0.38 , respectively). Conversely, purifying selection is less efficient in the S locus for KFB ($\pi_N/\pi_S = 1.83$ [KFB^T] and 0.23 [$KFB1$]), whereas CCM shows the opposite pattern ($\pi_N/\pi_S = 0.91$ [CCM^T] and 1.50 [$CCM1$]; Table 3). Taken together, the results imply that the effects of hemizygoty on purifying selection vary among *P. vulgaris* S-locus genes, corroborating previous results in *P. veris* (Potente *et al.*, 2022).

A key question for the genetics of distyly concerns whether the strength and nature of selection differ between S-locus genes with and without a demonstrated function in distyly. Among the three, nine, and five protein-coding genes identified in the S-locus of *Gelsemium*, *Linum*, and *Primula*, respectively, (Li *et al.*, 2016; Gutiérrez-Valencia *et al.*, 2022; Potente *et al.*, 2022; Zhao *et al.*, 2023) only two, namely CYP^T and GLO^T of *Primula*, have been functionally characterized, showing that CYP^T determines short styles and female self-incompatibility (Huu *et al.*, 2016, 2022), while GLO^T determines high anthers in thrums (Huu *et al.*, 2020). However, it remains unclear whether CCM^T , PUM^T , and KFB^T play a role in *Primula* distyly. The markedly reduced and non-floral specific expression of CCM^T , PUM^T , and KFB^T compared to CYP^T and GLO^T in both *P. vulgaris* and *P. veris* (Cocker *et al.*, 2018; Potente *et al.*, 2022) cast doubt on whether the former three genes are essential for distyly. In the distylous *Gelsemium elegans* (Gentianales), the homolog of *Primula* CCM^T was absent from the genome, while homologs of PUM^T and KFB^T were present but did not localize to the putative S-locus and were expressed in both pin and thrum flowers (Zhao *et al.*, 2023). Taken together, previous evidence suggests that CCM^T , PUM^T , and KFB^T may not be essential for the core traits of distyly (i.e., reciprocal placement of sexual organs and self-incompatibility), hence they might be under relaxed purifying selection. If this is true, one might expect thrums to exhibit higher accumulation

of non-synonymous mutations in CCM^T , KFB^T , and PUM^T than in CYP^T and GLO^T . Indeed, our results support this prediction, for we found weaker purifying selection on CCM^T , KFB^T , and PUM^T ($\pi_N/\pi_S = 0.91$, 1.83, and 10.36, respectively) compared to CYP^T ($\pi_N/\pi_S = 0.28$; Table 3A). It is unlikely that the results are explained by positive directional selection on advantageous non-synonymous mutations of the three genes above in thrums, because positive selection should cause rapid fixation of advantageous mutations, hence absence of polymorphism at non-synonymous sites (Hahn, 2020), which is not what we found (Table 3A). To summarize, in *P. vulgaris* purifying selection seems stronger on the only two S-locus genes for which a key function in distyly has been demonstrated (namely, CYP^T and GLO^T) than on CCM^T , KFB^T , and PUM^T , which were not found in the S-locus of other species and for which no differential expression between pin and thrum flowers was detected. Discovering whether the three genes above may play a role in controlling ancillary traits of distyly (e.g., pollen size and number, male incompatibility) requires additional functional studies in *Primula* and other distylous taxa.

Comprehensive population genetic analyses of variability in S-locus genes and their paralogs had never been performed until now, due to missing knowledge of relevant genes, unavailability of sequences from said genes, and inadequate population sampling. Here, we expanded on previous Sanger sequencing analyses of CYP^T in Somerset (England) populations (Mora-Carrera *et al.*, 2021) by analyzing also sequences of S-locus genes and their paralogs extracted from WGR data of Slovakian, Swiss, and Turkish populations of *P. vulgaris*. First, homostyles, found exclusively in three Somerset populations, exhibited lower genetic diversity than thrums for both S-locus genes and their paralogs (Table 3), corroborating previous reports of reduced genetic diversity in homostyles (Husband & Barrett, 1993; Ness *et al.*, 2010; Yuan *et al.*, 2017; Zhou *et al.*, 2017; Zhong *et al.*, 2019). Second, both S-locus genes and their paralogs have markedly lower genetic variation in English populations than in other Eurasian populations of *P. vulgaris* (Table S2). This finding suggests a recent genetic bottleneck in English populations. This bottleneck could be associated with colonization of England following glacial retreat during the Last Glacial Maximum (ca. 10,000-12,000 years ago), as suggested for other plant species (Birks, 1989). Future genomic and demographic investigations will determine whether the signatures of genetic bottlenecks detected in S-locus genes and their paralogs apply to the entire genome, thus helping to infer the timing and mode of *P. vulgaris* colonization of the British Isles.

Does lower viability of S/S*-homostyles prevent the fixation of homostyly in P. vulgaris?*

Theoretical and experimental work suggests that, all else being equal, once selfing originates, the selfing phenotype should increase in frequency and eventually become fixed over time (Fisher, 1941; Lande & Schemske, 1985; Charlesworth *et al.*, 1990). In the transition from distyly to homostyly, Crosby's model (1949) predicted that the rate of increase and ultimate fixation of homostyles in a population depends on whether homostyles with diploid S-locus have lower or equal viability as the other genotypes in the population (Figure 1C and D). The assumption of lower viability for S*/S*-homostyles of *P. vulgaris* expanded upon evidence from crossing experiments in *P. sinensis* suggesting that homozygous dominant thrums had 30% lower viability than heterozygous thrums (de Winston & Mather, 1941). More recently, results of crossing experiments in a *Primula* hybrid (*Primula x tommasinii*) were interpreted as evidence of inviability for S/S-thrums (Kurian & Richards, 1997). Furthermore, population surveys of pin-to-thrum ratios in *P. oreodoxa* indicated that thrums were overrepresented at the seed (~1:3) but not adult stage (~1:1), implying that differences in viability could occur during the life cycle (Yuan *et al.*, 2018). However, genotyping of thrums was not carried out, thus preventing the determination of whether the decrease of thrums from seed to adult stage was caused by lower viability of S/S-thrums. Our observed frequencies of S*/0- and S*/S*-homostyles from the two trimorphic, English populations EN4-T and EN5-T of *P. vulgaris* are consistent with Crosby's prediction of a recent transition to homostyly (20-30 generations) under 30-40% lower viability of S*/S*-homostyles (Table 4), supporting the model that assumes lower fitness for S*/S*-homostyles than S*/0-homostyles (Figure 1C).

Conversely, the occurrence of a monomorphic, homostylous population of *P. vulgaris* in England, first reported by Curtis and Curtis (1985) 38 years ago and recently sampled by Mora Carrera *et al.* (2021 and present

study) is congruent with the assumption of equal viability for S*/S* homostyles. All 11 genotyped homostyles in this population (here named EN6-M) carry the S*/S*- genotype (Table 1 and Figure 3), thus EN6-M could represent a case in which homostyly increased in frequency over time and became fixed in the population by displacing pins and thrums, as predicted under the assumption of equal viability for S*/S* homostyles (Figure 1D). Alternatively, EN6-M could have been established by an S*/S*- homostyle stemming from a nearby population, thus it might have been monomorphic homostylous from the beginning. Indeed, Curtis and Curtis (1985) reported that this monomorphic population was located only about 240 m away from a trimorphic population which might have served as a source for the initial homostyle that gave origin to EN6-M. Finally, EN6-M had a very low population size (n = 19; Mora-Carrera *et al.*, 2021) suggesting that stochasticity could have played a role in the fixation of S*/S*-homostyles in this population and that homozygosity of an S-locus with disrupted *CYP^T* might have detrimental effects on population growth.

To summarize, our results suggest that a diploid S-locus with inactivated *CYP^{T*}* may not *per se* be incompatible with homostyle viability. However, the occurrence of two copies of the remaining S-locus genes [i.e., *CCM^T*, *GLO^T*, *KFB^T*, and *PUM^T*] in the genome of a homostyle could have detrimental effects on viability at different stages of the life cycle, possibly stemming from gene-dosage effects (Rice & McLysaght, 2017; Ascencio *et al.*, 2021; Liet *et al.*, 2015; Tasdighian *et al.*, 2017). Future research combining S-locus genotyping and characterization of function and dosage effects of S-locus genes at different life-cycle stages with fitness measurements in the field and in greenhouse experiments is essential to address whether differences in viability prevent the widespread fixation of homostyly in *P. vulgaris*.

DATA ACCESSIBILITY STATEMENT

Sequence alignment of all exons of *CYP^T* and alignment of *CYP^T* promoter region has been submitted to the Wiley as supporting material and will be uploaded to Dryad (upon acceptance). BAM files of S-locus assembly has been uploaded temporarily to the following repository (https://github.com/EmilianoMora/temp-S-locus-BAM_files-Mora-Carrera.etal/tree/main). These files will be uploaded to Dryad (upon acceptance). Original sequencing reads will be uploaded upon acceptance to the NCBI repository (<https://www.ncbi.nlm.nih.gov/>) under the BioProject IDs PRXXXXXX. Code for the analyses and figures will be available in github.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS SECTION

EM-C, JdV, PS and EC designed the research. EM-C, RS, and BK collected the data. EM-C and RS performed laboratory work. EM-C analyzed the data with the help of RS, GP, NY, and BK. NY provided computational infrastructure and bioinformatics support. EM-C and EC wrote the manuscript. All coauthors provided feedback.

ACKNOWLEDGMENTS

We thank Natural England for permits and landowners for access to populations; Ferhat Celep and Judita Kochjarová for assistance during fieldwork collection. This research was supported by the Graduate Campus office at the University of Zurich through a GRC-Travel Grant and the Forschungskredit Candoc (grant no. K-74202-05-01) to EM-C, by the Swiss Government Excellence Scholarship (grant no. 2018.0475) to EM-C, by the Swiss National Science Foundation (grant no. 310030_185251) to JdV, and (grant nos. 3100-061674.00/1 and 31003A_175556) to EC.

LITERATURE CITED

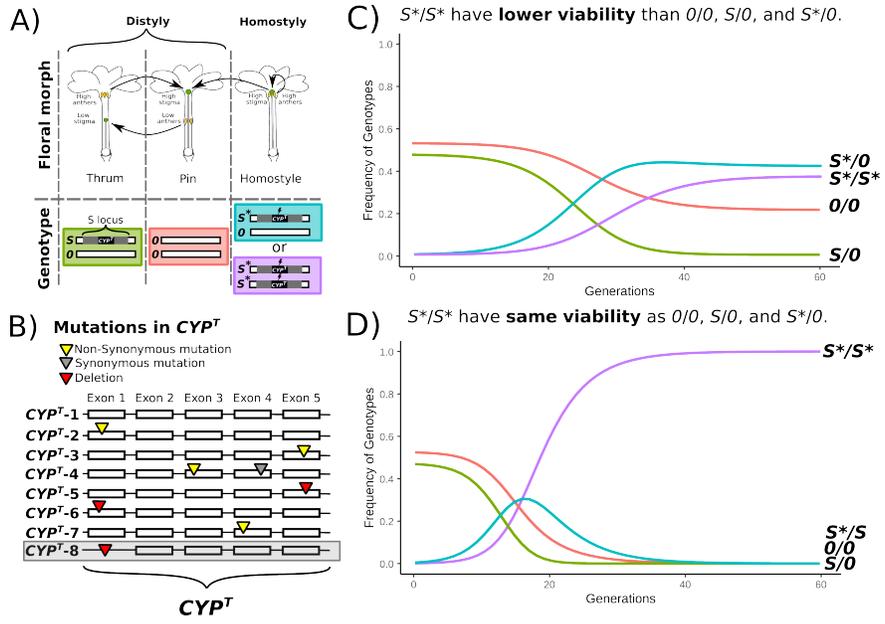
Ascencio D, Diss G, Gagnon-Arsenault I, Dubé AK, DeLuna A, Landry CR. 2021. Expression attenuation as a mechanism of robustness against gene duplication. *Proceedings of the National Academy of Sciences of the United States of America* 118: 1-10.

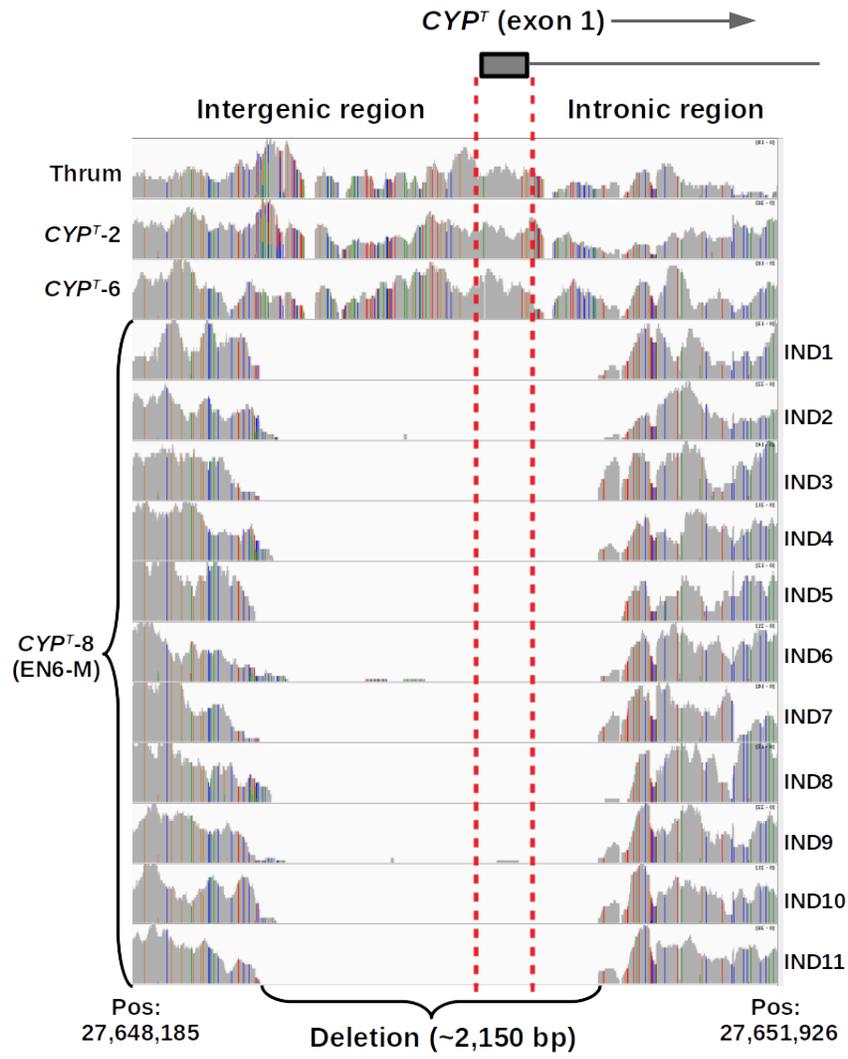
- Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoré A, Lafon-Placette C, Steige KA, Callot C, Marande W, Neuffer B, *et al.* 2019. Genetic basis and timing of a major mating system shift in *Capsella*. *New Phytologist* 224: 505–517.
- Barmentlo SH, Meirmans PG, Luijten SH, Triest L, Oostermeijer JGB. 2017. Outbreeding depression and breeding system evolution in small, remnant populations of *Primula vulgaris*: consequences for genetic rescue. *Conservation Genetics* 19: 545–554.
- Barrett SCH. 2019. ‘A most complex marriage arrangement’: Recent advances on heterostyly and unresolved questions. *New Phytologist* 224: 1051–1067.
- Bateson W, Gregory RP. 1905. On the inheritance of heterostylism in *Primula*. *Proceedings of the Royal Society of London. Series B* 76: 581–586.
- Belaoussoff S, Shore JS. 1995. Floral correlates and fitness consequences of mating-system variation in *Turnera ulmifolia*. *Evolution* 49: 545–556.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Busch JW, Joly S, Schoen DJ. 2011. Demographic signatures accompanying the evolution of selfing in *Leavenworthia alabamica*. *Molecular Biology and Evolution* 28: 1717–1729.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10: 1–9.
- Charlesworth D. 2022. Primrose homostyles: A classic case of possible balancing selection revisited. *Molecular Ecology* 32: 1–3.
- Charlesworth B, Charlesworth D. 1979. The maintenance and breakdown of distyly. *The American Naturalist* 114: 499–513.
- Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 355: 1563–1572.
- Charlesworth D, Morgan MT, Charlesworth B. 1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* 44: 1469–1489.
- Cocker JM, Wright J, Li J, Swarbreck D, Dyer S, Caccamo M, Gilmartin PM. 2018. *Primula vulgaris* (primrose) genome assembly, annotation and gene expression, with comparative genomics on the heterostyly supergene. *Scientific Reports* 8: 1–13.
- Costa J, Torices R, Barrett SCH. 2019. Evolutionary history of the buildup and breakdown of the heterostyly syndrome in Plumbaginaceae. *New Phytologist* 224: 1278–1289.
- Crosby JL. 1940. High proportions of homostyle plants in populations of *Primula vulgaris*. *Nature* 145: 672–673.
- Crosby JL. 1949. Selection of an unfavourable gene-complex. *Evolution* 3: 212–230.
- Curtis J, Curtis CF. 1985. Homostyle primroses re-visited. I. Variation in time and space. *Heredity* 54: 227–234.
- Cutter AD. 2019. Reproductive transitions in plants and animals: selfing syndrome, sexual selection and speciation. *New Phytologist* 224: 1080–1094.
- Fisher R. 1941. Average excess and average effect of a gene substitution. *Annals of Eugenics* 11: 53–63.
- Ganders FR. 1979. The biology of heterostyly. *New Zealand Journal of Botany* 17: 607–635.

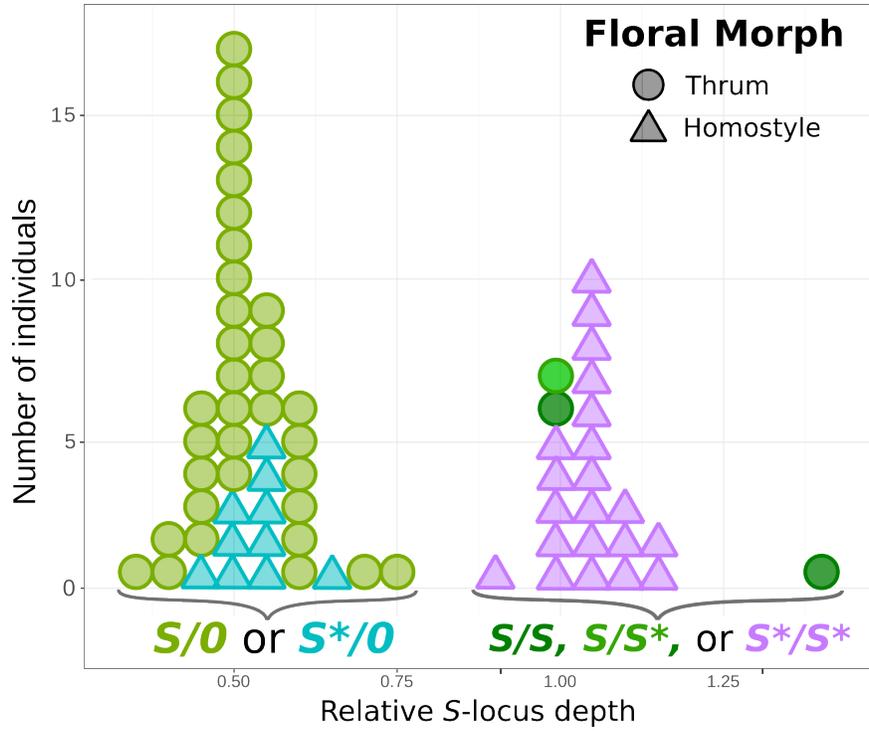
- Gao C, Zhou G, Ma C, Zhai W, Zhang T, Liu Z, Yang Y, Wu M, Yue Y, Duan Z, *et al.* 2016. Helitron-like transposons contributed to the mating system transition from out-crossing to self-fertilizing in polyploid *Brassica napus* L. *Scientific Reports* 6: 1–10.
- Garafutdinov RR, Galimova AA, Sakhabutdinova AR, Chemeris A V. 2016. PCR-based evaluation of sequence specificity of DNA fragmentation by ultrasound. *Molecular Biology* 50: 236–241.
- Goldberg EE, Iqic B. 2012. Tempo and mode in plant breeding system evolution. *Evolution* 66: 3701–3709.
- Gossmann TI, Woolfit M, Eyre-Walker A. 2011. Quantifying the variation in the effective population size within a genome. *Genetics* 189: 1389–1402.
- Gutiérrez-Valencia J, Fracassetti M, Berdan EL, Bunikis I, Soler L, Dainat J, Kutschera VE, Losvik A, Désamoré A, Hughes PW, *et al.* 2022. Genomic analyses of the *Linum* distyly supergene reveal convergent evolution at the molecular level. *Current Biology* 32: 4360–4371.e6.
- Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T. 2021. The genomic architecture and evolutionary fates of supergenes. *Genome Biology and Evolution* 13: 1–19.
- Husband BC, Barrett SCH. 1993. Multiple origins of self-fertilization in tristylous *Eichhornia paniculata* (Pontederiaceae): Inferences from style morph and isozyme variation. *Journal of Evolutionary Biology* 6: 591–608.
- Huu CN, Kappel C, Keller B, Sicard A, Takebayashi Y, Breuninger H, Nowak MD, Bäurle I, Himmelbach A, Burkart M, *et al.* 2016. Presence versus absence of *CYP73A50* underlies the style-length dimorphism in primroses. *eLife* 5: 1–15.
- Huu CN, Keller B, Conti E, Kappel C, Lenhard M. 2020. Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proceedings of the National Academy of Sciences of the United States of America* 117: 23148–23157.
- Huu CN, Plaschil S, Himmelbach A, Kappel C, Lenhard M. 2022. Female self-incompatibility type in heterostylous *Primula* is determined by the brassinosteroid-inactivating cytochrome P450 *CYP73A50*. *Current Biology* 32: 671–676.
- Jacquemyn H, Endels P, Brys R, Hermy M, Woodell SRJ. 2009. Biological flora of the British Isles: *Primula vulgaris* Huds. (*P. acaulis* (L.) Hill). *Journal of Ecology* 97: 812–833.
- Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding K V., Nikiforova MN. 2017. Guidelines for validation of next-generation sequencing–based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *Journal of Molecular Diagnostics* 19: 341–365.
- Keller B, Thomson JD, Conti E. 2014. Heterostyly promotes disassortative pollination and reduces sexual interference in Darwin’s primroses: Evidence from experimental studies. *Functional Ecology* 28: 1413–1425.
- Kissling J, Barrett SCHH. 2013. Variation and evolution of herkogamy in *Exochaenium* (Gentianaceae): Implications for the evolution of distyly. *Annals of Botany* 112: 95–102.
- Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources* 21: 1359–1368.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549.
- Kurian V, Richards AJ. 1997. A new recombinant in the heteromorphy “S” supergene in *Primula*. *Heredity* 78: 383–390.
- Lande R, Schemske DW. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic Models. *Evolution* 39: 24–40.

- Lewis D, Jones DA. 1992. The genetics of heterostyly. In: Barrett S, ed. *The Evolution and Function of Heterostyly*. Berlin: Springer-Verlag, 129–151.
- Li J, Cocker JM, Wright J, Webster MA, McMullan M, Dyer S, Swarbreck D, Caccamo M, Van Oosterhout C, Gilmartin PM. 2016. Genetic architecture and evolution of the S-locus supergene in *Primula vulgaris*. *Nature Plants* 2: 1–7.
- Li Z, Defoort J, Tasdighian S, Maere S, Van De Peer Y, De Smet R. 2015. Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* 28: 326–344.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Lloyd DG, Webb CJ. 1992. The selection of heterostyly. In: Barrett S, ed. *The Evolution and Function of Heterostyly*. Heidelberg: Springer-Verlag Berlin, 179–207.
- Mable BK, Hagmann J, Kim ST, Adam A, Kilbride E, Weigel D, Stift M. 2017. What causes mating system shifts in plants? *Arabidopsis lyrata* as a case study. *Heredity* 118: 52–63.
- Mather K, Winton D DE. 1941. Adaptation and counter-adaptation of the breeding system in *Primula*. *Annals of Botany* 5: 297–311.
- Matsui K, Yasui Y. 2020. Genetic and genomic research for the development of an efficient breeding system in heterostylous self-incompatible common buckwheat (*Fagopyrum esculentum*). *Theoretical and Applied Genetics* 133: 1641–1653.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* : 1297–1303.
- McLeay RC, Bailey TL. 2010. Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. *BMC Bioinformatics* 11: 1–11.
- Mora-Carrera E, Stubbs RL, Keller B, Léveillé-Bourret É, de Vos JM, Szövényi P, Conti E. 2021. Different molecular changes underlie the same phenotypic transition: Origins and consequences of independent shifts to homostyly within species. *Molecular Ecology* 32: 1–18.
- Naiki A. 2012. Heterostyly and the possibility of its breakdown by polyploidization. *Plant Species Biology* 27: 3–29.
- Nasrallah JB. 2017. Plant mating systems: self-incompatibility and evolutionary transitions to self-fertility in the mustard family. *Current Opinion in Genetics and Development* 47: 54–60.
- Ness RW, Wright SI, Barrett SCH. 2010. Mating-system variation, demographic history and patterns of nucleotide diversity in the tristylous plant *Eichhornia paniculata*. *Genetics* 184: 381–392.
- O'Malley RC, Huang SSC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. 2016. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 165: 1280–1292.
- Ono K, Akagi T, Morimoto T, Wónsch A, Tao R. 2018. Genome re-sequencing of diverse sweet cherry (*Prunus avium*) individuals reveals a modifier gene mutation conferring pollen-part self-compatibility. *Plant and Cell Physiology* 59: 1265–1275.
- Poptsova MS, Il'icheva IA, Nechipurenko DY, Panchenko LA, Khodikov M V., Oparina NY, Polozov R V., Nechipurenko YD, Grokhovsky SL. 2014. Non-random DNA fragmentation in next-generation sequencing. *Scientific Reports* 4: 1–6.
- Potente G, Léveillé-Bourret É, Yousefi N, Choudhury RR, Keller B, Diop SI, Duijsings D, Pirovano W, Lenhard M, Szövényi P, *et al.* 2022. Comparative genomics elucidates the origin of a supergene controlling floral heteromorphism. *Molecular Biology and Evolution* 39: 1–16.

- Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- Rice A, McLysaght A. 2017. Dosage-sensitive genes in evolution and disease. *BMC Biology* 15: 1-10.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* 29: 24–26.
- Ruiz-Martín J, Santos-Gally R, Escudero M, Midgley JJ, Pérez-Barrales R, Arroyo J. 2018. Style polymorphism in *Linum* (Linaceae): a case of Mediterranean parallel evolution? *Plant Biology* 20: 100–111.
- Schoen DJ., Johnston MO., L’Heureux A-M, Marsolais JV., Heureux A-ML’, Marsolais JV. 1997. Evolutionary history of the mating system in *Amsinckia* (Boraginaceae). *Evolution* 51: 1090–1099.
- Shimizu KK, Tsuchimatsu T. 2015. Evolution of selfing: Recurrent patterns in molecular adaptation. *Annual Review of Ecology, Evolution, and Systematics* 46: 593–622.
- Shore JS, Hamam HJ, Chafe PDJ, Labonne JDJ, Henning PM, McCubbin AG. 2019. The long and short of the S-locus in *Turnera*(Passifloraceae). *New Phytologist* 224: 1316–1329.
- Tsuchimatsu T, Kaiser P, Yew CL, Bachelier JB, Shimizu KK. 2012. Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica* . *PLoS Genetics* 8: 16–19.
- Stebbins GL. 1957. Self fertilization and population variability in the higher plants. *The American Naturalist* 91: 337–354.
- Tasdighian S, Van Bel M, Li Z, Van De Peer Y, Carretero-Paulet L, Maere S. 2017. Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *Plant Cell* 29: 2766-2785.
- de Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E. 2014. Heterostyly accelerates diversification via reduced extinction in Darwin’s primroses. *Proceedings of the Royal Society B: Biological Sciences* 281: 20140075.
- Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI. 2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora* . *PLoS Genetics* 10: 1-12.
- Yu CP, Lin JJ, Li WH. 2016. Positional distribution of transcription factor binding sites in *Arabidopsis thaliana* . *Scientific Reports* 6: 1–7.
- Yuan S, Barrett SCH, Duan T, Qian X, Shi M, Zhang D. 2017. Ecological correlates and genetic consequences of evolutionary transitions from distyly to homostyly. *Annals of Botany* 120: 775–789.
- Yuan S, Barrett SCH, Li C, Li X, Xie K, Zhang D. 2018. Genetics of distyly and homostyly in a self-compatible *Primula* . *Heredity* : 1–10.
- Zhao Z, Zhang Y, Shi M, Liu Z, Xu Y, Luo Z, Yuan S, Tu T, Sun Z, Zhang D, et al. 2023. Genomic evidence supports the genetic convergence of a supergene controlling the distylous floral syndrome. *New Phytologist* 237: 601–614.
- Zhong L, Barrett SCH, Wang X, Wu Z, Sun H, Li D, Wang H, Zhou W. 2019. Phylogenomic analysis reveals multiple evolutionary origins of selfing from outcrossing in a lineage of heterostylous plants. *New Phytologist* 224: 1290–1303.
- Zhou W, Barrett SCH, Li HD, Wu ZK, Wang XJ, Wang H, Li DZ. 2017. Phylogeographic insights on the evolutionary breakdown of heterostyly. *New Phytologist* 214: 1368–1380.
- Zhou W, Barrett SCH, Wang H, Li DZ. 2012. Loss of floral polymorphism in heterostylous *Luculia pinceana* (Rubiaceae): A molecular phylogeographic perspective. *Molecular Ecology* 21: 4631–4645.







| Population | Latitude | Longitude | Population type | Collected individuals (thrum:pin:homostyle) | Thrum | | Homostyle | |
|------------|----------|-----------|-----------------|--|-------------|------|-----------|-------|
| | | | | | S/O : S/S | O/O | S*/O | S*/S* |
| TR-D | 40.74 | 39.55 | D | 5:5:-- | 0.5 | 0.5 | - | - |
| SK-D | 48.77 | 19.05 | D | 5:5:-- | 0.5 | 0.5 | - | - |
| CH-D | 46.44 | 6.91 | D | 5:5:-- | 0.5 | 0.5 | - | - |
| EN1-D | 52.14 | -0.13 | D | 5:4:-- | 0.55 | 0.45 | - | - |
| EN2-D | 51.35 | -2.33 | D | 4:5:-- | 0.22 : 0.22 | 0.55 | - | - |
| EN3-D | 51.09 | -2.71 | D | 5:5:-- | 0.5 | 0.5 | - | - |
| EN4-T | 51.08 | -2.30 | T | 3:4:10 | 0.17 | 0.24 | 0.24 | 0.35 |
| EN5-T | 51.13 | -2.42 | T | 4:5:10 | 0.16 : 0.04 | 0.27 | 0.32 | 0.21 |
| EN6-M | 51.03 | -2.63 | M | --:--:11 | - | - | - | 1 |

| TAIR ID CODE | TF FAMILY | Expressed in carpel | Expressed with BRs treatment | Distance from exon 1 CYP1 (in bp) | Consensus sequence |
|--------------|--------------|---------------------|------------------------------|--------------------------------------|---------------------|
| AT1G01250 | DREB | Yes | No | 1351 | TGTCGGTGRKDNKD |
| AT1G19210 | DREB | Yes | Yes | 917 | DTGGWCGGTGRHGR |
| AT1G36060 | DREB | No | No | 460 | DTKGKCGGTGGHGR |
| AT4G28140 | DREB | No | Yes | 177 | KGTCGGTGGHGRND |
| AT5G65130 | DREB | No | Yes | 177 | HYNH-NHYDCCRCGCMW |
| AT3G50260 | DREB | No | No | 174 | CCDYDCCACCGMCA |
| AT5G67190 | DREB | Yes | No | 173 | YCDYDCCACCGACA |
| AT2G23340 | DREB | Yes | Yes | 173 | HDWGTGCGGTGRHND |
| AT4G06746 | DREB | No | - | 173 | HNNNHNDYCACCGCAWH |
| AT1G21910 | DREB | Yes | Yes | 171 | YCNCCDYDCCACCGMCA |
| AT5G42520 | BPC | Yes* | No | 171 | CTCTCTCTCTCTCTCTC |
| AT3G22780 | CHC | Yes* | Yes | 171 | WWTTWAAAATTTAAA |
| AT3G23250 | MYB | No | No | 171 | YHHRH-HWHYCACCAACCH |
| AT4G01680 | MYB | No | Yes | 168 | WGGTGGTTRRRNDD |
| AT1G09540 | MYB | No | - | 160 | NYYACCWACWH |
| AT1G34670 | MYB | No | No | 80 | NNWDBYACCCWAMC |
| AT5G67580 | TRBF2 | Yes | Yes | 64 | WWWHTWRRCCCAAWTH |
| AT4G38170 | FAR1-related | Yes* | Yes | 64 | CTCTCTCTCTCTCTCTC |
| AT2G40620 | bZip | Yes | Yes | 16 | NMCAGCTGKCA |
| AT1G06850 | bZip | Yes | Yes | 16 | DTGMCAGCTGKHW |

Note: - DREB=Dehydration-Responsive Element-Binding; BPC=Basic Pentacysteine; CHC=Chlathrin Heavy Chain; MYB=Myeloblastosis viral oncogene; FAR1=Fatty Acyl-CoA Reductase 1; TRBF2=Telomere Repeat Binding Factor 2; bZip=Basic leucine zipper transcription factor; *highly overexpressed.

A)

| Gene | n | π_s | π_N | π_N/π_s |
|------------------------|-----------------|---|---|---|
| <i>CYP^T</i> | 37 [†] | 0.0026 | 0.0007 | 0.28 |
| <i>GLO^T</i> | 37 [†] | 0 | 0.0015 | NA |
| <i>CCM^T</i> | 37 [†] | 0.0018 | 0.0016 | 0.91 |
| <i>KFB^T</i> | 37 [†] | 0.0004 | 0.0007 | 1.83 |
| <i>PUM^T</i> | 37 [†] | 0.0001 | 0.0011 | 10.36 |
| Average (± SE) | | 0.0010 (± 0.0005) 0.0012 (± 0.0006) ^a | 0.0011 (± 0.0002) 0.0011 (± 0.0003) ^a | 3.35 (± 2.04) 1.01 (± 0.37) ^a |
| <i>CYP734A51</i> | 74 [‡] | 0.0040 | 0.0015 | 0.38 |
| <i>GLO1</i> | 74 [‡] | 0.0019 | 0 | NA |
| <i>CCM1</i> | 74 [‡] | 0.0019 | 0.0029 | 1.50 |
| <i>KFB1</i> | 74 [‡] | 0.0056 | 0.0013 | 0.23 |
| Average (± SE) | | 0.0034 (± 0.0008) | 0.0014 (± 0.005) | 0.53 (± 0.29) |

B)

| Gene | n | π_s | π_N | π_N/π_s |
|------------------------|----|-----------------|-------------------|---------------|
| <i>CYP^T</i> | 31 | 0 | 0.0005 | NA |
| <i>GLO^T</i> | 31 | 0.0012 | 0 | 0 |
| <i>CCM^T</i> | 31 | 0 | 0 | 0 |
| <i>KFB^T</i> | 31 | 0 | 0 | 0 |
| <i>PUM^T</i> | 31 | 0 | 0 | 0 |
| Average (± SE) | | 0.0002 (± 0.00) | 0.0001 (± 0.00) | 0 |
| <i>CYP734A51</i> | 31 | 0.0008 | 0.0002 | 0.28 |
| <i>GLO1</i> | 31 | 0 | 0 | 0 |
| <i>CCM1</i> | 31 | 0 | 0.0003 | NA |
| <i>KFB1</i> | 31 | 0 | 0.0008 | NA |
| Average (± SE) | | 0.0002 (± 0.00) | 0.0003 (± 0.0001) | 0.14 (± 0.14) |

Note.- n=sample size; †=thrums; ‡=pins and thrums; NA=not available; ^a=average without *PUM^T*.

| | | observed frequencies (0/0 : S /0 : S' /0 : S' /S*) | | | |
|------------|---|--|-----------------------------|-----------------------------|--|
| | | v = 1 | | | |
| Generation | expected freq (0/0 : S /0 : S' /0 : S' /S*) | EN4-T | EN5-T | EN6-M | |
| 10 | 0.50 : 0.42 : 0.05 : 0.03 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 20 | 0.17 : 0.04 : 0.29 : 0.50 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 30 | 0.02 : 0.00 : 0.06 : 0.92 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1 ^{NS} | |
| 40 | 0.00 : 0.00 : 0.01 : 0.99 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1 ^{NS} | |

| | | observed frequencies (0/0 : S /0 : S' /0 : S' /S*) | | | |
|------------|---|--|-----------------------------|-----------------------------|--|
| | | v = 0.9 | | | |
| Generation | expected freq (0/0 : S /0 : S' /0 : S' /S*) | EN4-T | EN5-T | EN6-M | |
| 10 | 0.46 : 0.34 : 0.14 : 0.07 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 20 | 0.15 : 0.02 : 0.29 : 0.54 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 30 | 0.04 : 0.00 : 0.11 : 0.85 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1 ^{NS} | |
| 40 | 0.01 : 0.00 : 0.04 : 0.95 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1 ^{NS} | |

| | | observed frequencies (0/0 : S /0 : S' /0 : S' /S*) | | | |
|------------|---|--|---|-----------------|--|
| | | v = 0.8 | | | |
| Generation | expected freq (0/0 : S /0 : S' /0 : S' /S*) | EN4-T | EN5-T | EN6-M | |
| 10 | 0.48 : 0.48 : 0.10 : 0.04 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 20 | 0.25 : 0.05 : 0.37 : 0.33 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36^{NS} | 0 : 0 : 0 : 1** | |
| 30 | 0.12 : 0.00 : 0.27 : 0.62 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 40 | 0.08 : 0.00 : 0.19 : 0.73 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |

| | | observed frequencies (0/0 : S /0 : S' /0 : S' /S*) | | | |
|------------|---|--|-----------------------------|-----------------|--|
| | | v = 0.7 | | | |
| Generation | expected freq (0/0 : S /0 : S' /0 : S' /S*) | EN4-T | EN5-T | EN6-M | |
| 10 | 0.50 : 0.40 : 0.07 : 0.03 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 20 | 0.35 : 0.13 : 0.35 : 0.17 | 0.36 : 0.16 : 0.29 : 0.19^{NS} | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 30 | 0.21 : 0.01 : 0.40 : 0.38 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 40 | 0.17 : 0.00 : 0.37 : 0.46 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |

| | | observed frequencies (0/0 : S /0 : S' /0 : S' /S*) | | | |
|------------|---|--|-----------------------------|-----------------|--|
| | | v = 0.65 | | | |
| Generation | expected freq (0/0 : S /0 : S' /0 : S' /S*) | EN4-T | EN5-T | EN6-M | |
| 10 | 0.52 : 0.46 : 0.01 : 0.01 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 20 | 0.48 : 0.37 : 0.12 : 0.03 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 30 | 0.34 : 0.01 : 0.38 : 0.18 | 0.36 : 0.16 : 0.29 : 0.19^{NS} | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |
| 40 | 0.24 : 0.01 : 0.43 : 0.32 | 0.36 : 0.16 : 0.29 : 0.19** | 0.29 : 0.09 : 0.25 : 0.36** | 0 : 0 : 0 : 1** | |

Note. ** = P<0.01; ^{NS}=Non-Significant differences.