Genomic analyses elucidate the causes and consequences of breakdown of distyly in *Linum trigynum*

Juanita Gutiérrez-Valencia[1], Panagiotis-Ioannis Zervakis[1], Zoé Postel[1], Marco Fracassetti[1], Aleksandra Losvik[1], Sara Mehrabi[1], Ignas Bunikis[2], Lucile Soler[3], P. William Hughes[1], Aurélie Désamoré[1], Benjamin Laenen[1], Mohamed Abdelaziz[4], Olga Vinnere Pettersson[2], Juan Arroyo[5], Tanja Slotte[1,*]

[1]Department of Ecology, Environment and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden
[2]Uppsala Genome Center, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
[3]Department of Medical Biochemistry and Microbiology, Uppsala University, National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[4]Department of Genetics, University of Granada, Granada, Spain
[5]Department of Plant Biology and Ecology, University of Seville, Seville, Spain

*Corresponding author:
Tanja Slotte, Dept. of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm SE-106 91 Stockholm, Sweden. Email: tanja.slotte@su.se

Running Title: Genomic analyses of breakdown of distyly in *Linum*

Key words: homostyly, self-fertilization, distribution of fitness effects, genome assembly, plant mating system.

Abstract

Distyly is an iconic floral polymorphism governed by a supergene, which promotes efficient pollen transfer and outcrossing through reciprocal differences in the position of sexual organs in flowers, often coupled with heteromorphic self-incompatibility (SI). Distyly has evolved convergently in multiple flowering plant lineages, but has also broken down repeatedly, often resulting in homostylous, self-compatible populations with elevated rates of self-fertilization relative to their distylous ancestors. Here, we aimed to study the causes and consequences of the shift to homostyly in *Linum trigynum*, which is closely related to distylous *Linum tenue.* Building on a high-quality genome assembly, we show that *L. trigynum* harbors a genomic region homologous to the dominant haplotype of the distyly supergene in *L. tenue*, suggesting that loss of distyly first occurred in a short-styled individual. In contrast to homostylous *Primula* and *Fagopyrum*, *L. trigynum* harbors no fixed loss-of-function mutations in coding sequences of *S*-linked distyly candidate genes. Instead, floral gene expression analyses and controlled crosses suggest that mutations downregulating the *S*-linked *LtWDR-44* candidate gene for male SI and/or anther height could underlie homostyly in *L. trigynum*. Population genomic analyses of 224 whole-genome sequences further demonstrate that *L. trigynum* is highly self-fertilizing, exhibits significantly lower genetic diversity genome-wide and is experiencing relaxed purifying selection on nonsynonymous mutations relative to *L. tenue*, despite the relatively recent split of *L. trigynum* and *L. tenue*. Our analyses elucidate the tempo and mode of loss of distyly in *L. trigynum*, and advances our understanding of a common evolutionary transition in flowering plants.

Introduction

Floral adaptations to different pollination modes have long fascinated evolutionary biologists (Darwin, 1876; 1877). One such adaptation is distyly, a floral polymorphism that promotes efficient pollen transfer and outcrossing via insect pollinators (Darwin, 1877; Lloyd & Webb 1992; reviewed by Barrett, 2019). Natural populations of distylous species are polymorphic for floral morph, such that individuals have one of two types of flowers that differ reciprocally in the positions of male and female sexual organs (anthers and stigmas, respectively) within the flower. Pin individuals have long styles (L-morph) and short stamens with stigmas at a high position and anthers at a low position in the flower, whereas thrum individuals have the reciprocal arrangement with short styles (S-morph) and long stamens (Fig. 1). In most distylous species, these morphological differences are coupled with a heteromorphic self-incompatibility (SI) system which allows inbreeding avoidance (Charlesworth and Charlesworth, 1979) and reinforces disassortative mating by preventing self- and intra-morph fertilization.

Despite its multi-trait nature, distyly is inherited as a single Mendelian locus, called the distyly supergene or the *S*-locus. The distyly supergene typically harbors two alleles: one dominant exclusive to thrums and one recessive for which pins are homozygous (Bateson and Gregory, 1905; Laibach, 1923) (Fig. 1). It is only recently that distyly supergenes have begun to be sequenced and characterized in detail (reviewed in Gutiérrez-Valencia et al., 2021). So far, characterization of independently evolved distyly supergenes in six systems (*Primula*: Li et al., 2016; *Turnera*: Shore et al., 2019; *Linum*: Gutiérrez-Valencia et al., 2022a; *Fagopyrum*: Yasui et al., 2012; Fawcett et al., 2023; *Gelsemium*: Zhao et al., 2023; *Nymphoides indica*: Yang et al., 2023) suggest a remarkable degree of convergence in the genetic architecture of distyly supergenes. Common features of independently evolved distyly supergenes include presence of indel variation and hemizygosity in thrum individuals, and similarities in patterns of molecular evolution (Gutiérrez-Valencia et al., 2022a).

While distyly has evolved independently at least 13 times in flowering plants (Naiki, 2012), it has broken down far more frequently, often resulting in homostylous species that are monomorphic and self-compatible (SC), with flowers harboring anthers and stigmas at the same height (reviewed by Ganders, 1979). Given their capacity for self-fertilization, homostylous individuals (homostyles) can be favored whenever insect-mediated pollination becomes unreliable, due to selection for reproductive assurance (e.g., Piper et al., 1986; Yuan et al., 2017). Genomic characterization of transitions from distyly to homostyly may therefore allow the identification of genetic changes underlying this mating system shift, as well as its population genomic consequences.

Because early models of the distyly supergene posited that thrum plants were heterozygous at the distyly supergene (Ernst, 1936), it was frequently hypothesized that rare recombination events between dominant and recessive *S*-haplotypes caused homostyly (Dowrick, 1956; Charlesworth and Charlesworth, 1979; reviewed by Ganders, 1979). This idea has been revisited after the realization that thrums are predominantly hemizygous rather than heterozygous at the distyly *S*-locus in several distylous systems (Li et al., 2016; Shore et al., 2019; Gutiérrez-Valencia et al., 2022a; Fawcett et al., 2023). Since hemizygosity precludes the possibility of recombination between the dominant and recessive haplotypes, other genetic

causes of distyly breakdown should be considered. First, unlinked genetic modifiers could act by reducing floral herkogamy (Ganders, 1979; Mather & De Winton, 1941), precipitating distyly breakdown. Second, distyly breakdown could be a consequence of loss-of-function mutations at *S*-linked genes (as suggested by Ernst, 1936). Theory predicts that individuals that exhibit functional dominant male SI in combination with recessive style length and female SI function should be favored during the establishment of homostyly (Dowrick, 1956; Charlesworth and Charlesworth, 1979). In systems with hemizygous *S*-loci, it seems likely that such combinations would arise by mutation and not recombination. So far, results on the breakdown of distyly in *Primula* and *Fagopyrum* are in line with this prediction. In both systems, mutations affecting *S*-linked genes responsible for female SI function and short styles (i.e., thrum-exclusive *CYP734A50* in *Primula* and *S-ELF3* in *Fagopyrum*) can readily lead to the formation of long-homostylous SC plants, because these *S*-linked genes jointly govern style length and female SI (Huu et al., 2016; Huu et al., 2022; Fawcett et al., 2023). Moreover, independently evolved homostyles in natural populations of *Primula vulgaris* harbor putative loss-of-function mutations in *CYP734A50* (Mora-Carrera et al., 2021). Mutations at *S*-linked genes, particularly in genes affecting style length and female SI, thus constitute a feasible pathway to loss of distyly, but it remains unknown if similar events have unfolded in other lineages of distylous plants.

Genomic studies hold the promise to quantify the impact of homostyly on outcrossing rates, as well as to characterize the consequences for patterns of polymorphism and the efficacy of selection. If the evolution of homostyly is associated with shifts to high selfing rates, we expect the transition to result in marked reductions in the effective population size (N_e), exacerbated by linked selection due to reduced effective recombination rates in selfers, and potentially by founder events and bottlenecks associated with selfing (reviewed by Wright et al., 2013; Slotte, 2014; Hartfield et al., 2017; Cutter, 2019). In combination, these processes should result in reduced genetic diversity genome-wide, more marked population structure and a decreased efficacy of selection, especially against weakly deleterious mutations, in selfing homostylous species compared to their distylous relatives. Although transitions from distyly to homostyly have occurred repeatedly in the history of flowering plants, the genomic consequences of this transition have so far primarily been studied in one system, *Primula* (e.g., Wang et al., 2021, Zhong et al., 2019).

Linum is a promising system for studying the evolution and breakdown of distyly (e.g. Gutiérrez-Valencia et al., 2022a), because it shows a remarkable diversity of stylar conditions, including several independent losses of distyly (Ruiz-Martín et al., 2018; Maguilla et al., 2021). In agreement with the expectation that selfers have improved colonization ability (Baker, 1955; Stebbins, 1957), phylogenetic analyses suggest that the evolution of homostyly is associated with the expansion of *Linum* outside its center of origin (Maguilla et al., 2021). It remains largely unknown how homostyly has emerged in *Linum*, but we have recently shown that, similar to other distylous lineages, the distyly *S*-locus in *L. tenue* harbors a hemizygous region exclusively inherited by thrums (Gutiérrez-Valencia et al., 2022a).

Here, we investigated the genetic causes and evolutionary consequences of loss of distyly in *Linum trigynum*, an annual self-compatible species distributed across Spain and North Africa, in which homostyly has been inferred as a derived state (Ruiz-Martín et al., 2018). For this purpose, we assembled and annotated

a high-quality genome sequence of the homostylous *L. trigynum* using PacBio high-fidelity (HiFi) long reads and chromatin conformation capture (Hi-C) data. To investigate the genetic basis of loss of distyly, we compared our *L. trigynum* genome assembly as well as two additional linked-read draft assemblies of *L. trigynum* to eight genome assemblies of the closely related distylous species *L. tenue* (Gutiérrez-Valencia et al., 2022a), with a focus on mutations and regulatory changes at *S*-linked genes. To characterize the timing and mode of origin of homostyly, we analyzed polymorphism data based on 224 whole genome sequences from eight natural populations each of *L. trigynum* and *L. tenue*. Finally, we conducted comparative population genomic analyses to investigate the genome-wide effects of the shift to homostyly, in terms of levels of diversity, inbreeding levels, population structure and the efficacy of purifying selection.

Results

A Genomic Framework for Studying Breakdown of Distyly in Linum trigynum

To provide a genomic framework for studies of the evolution of homostyly in *L. trigynum*, we generated a high-quality *L. trigynum* genome assembly based on HiFi PacBio long reads (~41x coverage) and Hi-C data from a single *L. trigynum* individual (see Materials and Methods and Table S1, Supplementary Information for sample information). The resulting assembly was highly complete (complete BUSCOs=96.6%; Table S2, Supplementary Information) and spanned 498 Mb (N50=47.03 Mb), divided in 10 chromosomes and five scaffolds (less than 500 kb) with large-scale synteny to the *L. tenue* genome assembly (Gutiérrez-Valencia et al., 2022a) (Fig. 2a, Fig. S1, Supplementary Information). Repetitive sequences constituted 56.49% of the assembly, and we annotated a total of 54,692 coding genes, similar to the number in *L. tenue* (Gutiérrez-Valencia et al., 2022a).

We supplemented our PacBio genome assemblies of *L. trigynum* and *L. tenue* with two additional draft assemblies of *L. trigynum* based on 10x Chromium linked-read sequencing data (contig N50=4.82 Mb vs 4.47 Mb; assembly size=461.69 Mb vs 463.46 Mb), and seven additional *L. tenue* linked-read assemblies for thrum individuals which carry both the dominant and recessive haplotypes of the *S*-locus (Table S3, Supplementary Information).

Analyses of the S*-locus Region Elucidate the Origin of Homostyly*

Whole-genome alignments of our high-quality *L. trigynum* and *L. tenue* assemblies showed that *L. trigynum* carries a region on chromosome 10 that is homologous to the dominant haplotype of the *L. tenue* distyly *S*-locus (Fig. 2b). Analyses of linked-read assemblies and genome coverage at this region based on short-read data from 104 *L. trigynum* individuals further suggest that *L. trigynum* is fixed for the longer *S*-haplotype that is dominant in *L. tenue* (Fig. 2c). Haplotype network analyses indicate that variation at the *L. trigynum* *S*-locus stems from a single *S*-haplotype from its distylous ancestor (Fig. S2, Supplementary Information). Together, these results suggest that events affecting the functioning of the supergene in thrum individuals led to the breakdown of distyly, and that the haplotype identified in *L. trigynum* is likely derived from the dominant *S*-haplotype.

We first investigated the gene content in the *L. trigynum* genome region homologous to the dominant *S*-haplotype of *L. tenue*. After curation of *S*-locus annotation (see Materials and Methods for details), we retained nine *S*-linked genes (Fig. 2d) in *L. tenue*. Out of these, five homologous *S*-linked genes were present in our high-quality long-read *L. trigynum* assembly (Fig. 2d). These included the thrum-specific gene *LtTSS1* which has pistil-specific expression and is likely to reduce cell length in thrum styles, and *LtWDR-44* potentially involved in controlling anther position and/or pollen incompatibility in *L. tenue* (discussed in Gutiérrez-Valencia et al., 2022a). In addition, *L. trigynum* harbored homologs of genes at the 5' end and 3' ends of the *S*-locus region (Fig. 2d). The remaining genes, which are of unknown function (Gutiérrez-Valencia et al. 2022a), were not present in the *L. trigynum* genome assembly (Fig. 2).

Next, we compared sequences of *L. tenue* and *L. trigynum* to identify candidate loss-of-function changes at the *S*-locus region affecting gene function in *L. trigynum*, and to quantify synonymous and nonsynonymous divergence between *L trigynum* and the dominant *S*-haplotype of *L. tenue.* We only identified putative loss-of-function changes in *L. trigynum* in the gene *LtTSS1* (Table 1). This gene only harbors one exon, as demonstrated by our updated annotation, validated by PCR-assays of cDNA structure (see Materials and Methods for details; Fig. S3, Supplementary Information). In *LtTSS1*, we identified a C-to-A mutation resulting in a premature stop codon in *L. trigynum* that was not fixed but segregating at a frequency of 0.26 in our population data set, suggesting that it is unlikely to have initally caused loss of distyly. Although all genes also harbored nonsynonymous mutations, there were no frameshift or other major-effect mutations in the coding sequence of the other analyzed genes, nor markedly accelerated rates of nonsynonymous substitution (Table 1).

The S*-linked Gene* LtWDR-44 *is Downregulated in Floral Buds of* L. trigynum

To determine if distyly breakdown in *L. trigynum* was associated with changes in transcript abundance at candidate genes, we contrasted gene expression at *S*-linked genes in *L. tenue* thrums and *L. trigynum* homostyles. We focused on detecting altered expression of any of the genes shared between the dominant *S* allele of *L. tenue* and its derived allele fixed in *L. trigynum* (i.e., *liteg52183*, *indelg2*, *LtTSS1*, *LtWDR44* and *LtTPR/liteg52188*). Out of these genes, only the stamen length and/or male SI candidate gene *LtWDR-44* was significantly differentially expressed in floral buds, being downregulated in floral buds of *L. trigynum* compared to *L. tenue* (Log_2-fold change = -1.58, $p < 0.01$, Fig. 3). *LtWDR-44* was not differentially expressed between *L. tenue* thrums and *L. trigynum* homostyles in leaves (Supplementary Fig. S4; Supplementary Information). None of the remaining *S*-linked genes showed significantly different levels of expression in floral buds or leaves between *L. tenue* thrum and homostylous *L. trigynum* (Supplementary Fig. S4; Supplementary Information).

Pollination Assays Indicate Breakdown of Male SI in L. trigynum

Distyly breakdown often evolves in association with loss of heteromorphic SI (reviewed by Barrett, 2019). If the *S*-linked gene *LtWDR-44* is a determinant of male SI, as we previously hypothesized (Gutiérrez-Valencia et al., 2022a), then we would expect downregulation of this gene to affect the male (pollen) SI expressed by

L. trigynum. Specifically, we would expect *L. trigynum* pollen to behave similarly to *L. tenue* pin pollen and elongate successfully in thrum but not pin *L. tenue* pistils (Supplementary Fig. S5, Supplementary Information). Crossing assays followed by pollen tube staining showed that *L. trigynum* grew long pollen tubes in pistils of *L. tenue* thrum plants, but not in pistils of *L. tenue* pin plants (Fig. 3e, Supplementary Fig. S5, Supplementary Information; note that for incompatible crosses, pollen tube rejection occurs in the style in *Linum*; Murray 1986). These results are consistent with expectations if downregulation of *LtWDR-44* in *L. trigynum* impairs male SI function.

High Levels of Inbreeding in the Homostylous L. trigynum

The fact that *L. trigynum* exhibits symptoms of the selfing syndrome in floral traits (Fig. 1) and has an approximately fivefold lower pollen:ovule ratio compared to *L. tenue* (Repplinger, 2009) suggests that self-pollination is likely to be its main mating strategy. If so, we would expect *L. trigynum* to be highly inbred. To assess whether this was the case, we estimated the inbreeding coefficient using genome-wide polymorphism data from 224 individuals representing eight populations per species (average n=14 individuals per population; Table S1, Supplementary Information). We found that estimates of the inbreeding coefficient (F_{IS}) for *L. trigynum* were significantly higher than those for *L. tenue* populations (Kruskal–Wallis test followed by Dunn's test with Benjamini-Hochberg adjustment of P-values, $P < 0.05$ for all except two *L. trigynum – L. tenue* population comparisons; summary of median F_{IS} values across populations: *L. tenue* mean=0.09, S.E.=0.03, n=8, *L. trigynum* mean=0.87, S.E.=0.04, n=8) (Fig. 4a). Assuming equilibrium (Wright, 1969), the mean effective self-fertilization rate in *L. trigynum* is 0.93 ($2 F_{IS} /[1+ F_{IS}]$). These results show that *L. trigynum* is more inbred than *L. tenue*, consistent with the shift to homostyly and breakdown of SI resulting in elevated selfing rates.

Stronger Population Structure and Reduced Polymorphism in L. trigynum *Compared to* L. tenue

Transitions to self-fertilization are expected to result in reductions of the effective population size (N_e), and thus reduced genetic diversity as well as more pronounced population structure in selfers relative to outcrossers (Wright et al., 2013). To assess if the homostylous *L. trigynum* had reduced polymorphism levels genome-wide compared to the distylous *L. tenue,* we obtained windowed estimates of nucleotide diversity (π, 100 kb windows) for eight populations per species. In agreement with expectation, *L. trigynum* showed markedly lower genome-wide π than *L. tenue* (Fig. 4b) (Kruskal–Wallis test followed by Dunn's test with Benjamini-Hochberg adjustment of P-values, $P < 0.001$ for all *L. trigynum – L. tenue* populations comparisons average π values across populations: *L. tenue*: mean=1.3×10^{-3}, S.E.=1.4×10^{-4}, n=8, *L. trigynum*: mean=5.3×10^{-4}, S.E.=1.2×10^{-4}, n=8).

Next, we investigated population structure based on our genome-wide polymorphism (see Fig. 5a for geographical origin of the sampled individuals). Structure analyses with ADMIXTURE, principal component analyses (PCA), and TreeMix-based inference based on 76,934 SNPs in non-coding regions (see Methods for details) showed that *L. tenue* and *L. trigynum* form clearly differentiated groups (Fig. 5b, c; Fig. S6, Supplementary Information). We found the highest degree of differentiation between *L. tenue* and *L.*

trigynum populations (F_{ST} median=0.96, 1st and 3st quartile=0.95 and 0.97), followed by *L. trigynum* (F_{ST} median=0.37, 1st and 3st quartile=0.28 and 0.50) and finally *L. tenue* (F_{ST} median=0.05, 1st and 3st quartile=0.03 and 0.05) (Kruskal–Wallis test followed by Dunn's test with Bonferroni corrected *P* values, $P < 0.001$) (Fig. S7, Supplementary Information). TreeMix analyses (Fig. 5c) further suggested the absence of gene flow between *L. tenue* and *L. trigynum*, and stronger population structure within *L. trigynum* than in *L. tenue*. Overall, these results show low population structure within *L. tenue*, and more marked population structure within *L. trigynum* (Fig. 5b, c).

Finally, divergence population genomic analyses jointly estimating demographic parameters and inbreeding levels suggested that the split between *L. tenue* and *L. trigynum* was relatively recent (about 340 kya) and associated with a marked effective population size reduction in *L. trigynum* (Fig. 5d) as well as elevated inbreeding in *L. trigynum* (F_{IS} = 0.88). Together, these results suggest that the transition to homostyly was associated with higher rates of self-fertilization and reduced N_e, and occurred within a relatively recent evolutionary timeframe.

Relaxed Purifying Selection Against Weakly Deleterious Mutations in L. trigynum

Comparisons of highly selfing species and outcrossing relatives have repeatedly identified genome-wide signatures of relaxed purifying selection in selfers (reviewed by Cutter, 2019). Here, we investigated if similar patterns are observed in *L. trigynum*. First, we annotated repetitive regions in the genome sequences of *L. trigynum* and *L. tenue*, and compared the proportion of transposable elements (TEs) in 50 kb windows. Contrary to the expectation that selfing would allow the accumulation of harmful TEs insertions in the *L. trigynum* genome after weakened selection due to reduced N_e (Charlesworth and Wright, 2001), we found no significant differences in TE content of the genomes of these species (Fig. 4c) (t-test: t=1.49, df=20561, *N.S.*).

We next investigated if the shift to selfing in *L. trigynum* has resulted in reduced efficacy of purifying selection on nonsynonymous mutations, which could yield an elevated ratio of nonsynonymous to synonymous polymorphism (π_N/π_S). In line with this expectation, we found that π_N/π_S estimates were slightly higher in *L. trigynum* than in *L. tenue* (Fig. S8, Supplementary Information) (*L. tenue*: mean= 0.40, S.E.= 0.02, *n*=8, *L. trigynum*: mean= 0.50, S.E.= 0.15, *n*=8). To test if elevated π_N/π_S in *L. trigynum* was due to weaker purifying selection, we estimated the distribution of negative fitness effects (DFE) of new nonsynonymous mutations in each population of *L. trigynum* and *L. tenue* using fastDFE v1.0.0 (Fig. 4d). In line with the expectation that reduced N_e in selfers should lead to relaxed selection against weakly deleterious mutations (reviewed by Wright et al., 2013), we found that the proportion of new nonsynonymous mutations that are effectively neutral was significantly higher in *L. trigynum* than in *L. tenue* populations ($0 > N_es > -1$: median=0.442) compared to *L. tenue* ($0 > N_es > -1$: median=0.24) ($0 > N_es > -1$: Wilcoxon rank-sum test, $P < 0.001$, *n*=8 populations per species), suggesting that selection against weakly deleterious mutations is relaxed in *L. trigynum* relative to *L. tenue*. Moreover, we found a lower proportion of new nonsynonymous mutations with moderate and strongly deleterious effects in *L. trigynum* ($-1 > N_es > -10$: median=0.064, $-10 > N_es > -100$: median=0.074, $N_es < -100$: median=0.074) compared to *L.*

tenue (-1 > $N_e s$ > -10: median=0.096, and -10 > $N_e s$ > -100: median=0.132, $N_e s$ < -100: median=0.524) (Wilcoxon rank-sum test, $P < 0.001$, n=8 populations per species). Together, these results show that the efficacy of purifying selection on new nonsynonymous mutations is lower in *L. trigynum* than in *L. tenue*.

Discussion

Understanding the causes and consequences of major evolutionary transitions is an important aim in evolutionary biology. The remarkable diversity of mating system strategies that plants exhibit provides an opportunity to elucidate the causes of reproductive transitions, and how they impact the capacity of populations to respond to selection. Here, we first generated a genomic framework to study genetic causes of loss of distyly in *Linum*, and then used this framework to characterize the timing and mode of the transition as well as its genomic consequences in terms of inbreeding levels, patterns of polymorphism, and the efficacy of natural selection genome-wide.

While earlier models posited that rare recombination events between the recessive and dominant alleles at the *S*-locus were the cause for distyly breakdown (Dowrick, 1956; Charlesworth and Charlesworth, 1979), recent findings have indicated that distyly *S*-loci harbor presence-absence variation and that thrums are predominantly hemizygous at the *S*-locus rather than heterozygous (Li et al., 2016; Shore et al., 2019; Gutiérrez-Valencia et al., 2022a; Fawcett et al., 2023; Yang et al., 2023; Zhao et al., 2023). The recent characterization of the genetic architecture of the *S*-locus makes previous models of breakdown less likely and rather indicates that mutations either at the *S*-locus or at unlinked modifier loci are more plausible causes for the evolutionary shift from distyly to homostyly.

The presence of a genomic region in *L. trigynum* homologous to the thrum-specific dominant *S*-haplotype of *L. tenue* strongly suggests that breakdown of distyly first occurred in a thrum individual. This finding allowed us to identify candidate mutations for distyly breakdown, by comparing sequences of *S*-linked genes in *L. tenue* and their homologues in *L. trigynum*. We also conducted differential expression analyses between *L. trigynum* and *L. tenue* to determine if changes in the expression of *S*-linked genes might have contributed to the evolution of homostyly in *L. trigynum.* Work in *Primula* (Huu et al., 2016; Huu et al., 2022; Mora-Carrera et al., 2021) and *Fagopyrum* (Fawcett et al., 2023) has shown that shifts to homostyly in both lineages were most likely precipitated by mutations at *S*-linked genes that simultaneously govern pistil elongation and female SI. Although we found a premature stop codon in *LtTSS1,* which is a strong candidate gene for style length and possibly female SI in *Linum* (Gutiérrez-Valencia et al., 2022a), this major-effect mutation was not fixed in *L. trigynum* populations, suggesting that loss of distyly has different genetic basis in this species. Our differential expression analyses showed that *LtWDR-44,* which likely governs anther position and/or pollen functioning, was the only *S*-linked gene downregulated in floral buds of *L. trigynum* compared to *L. tenue*. Pollination assays further showed that *L. trigynum* pollen tubes can grow in styles of *L. tenue* thrums but not in those of pins, consistent with expectations if downregulation of *LtWDR-44* in *L. trigynum* disrupted the male component of the SI system. These results suggest that mutations resulting in downregulation of the expression of *LtWDR-44* are a plausible cause of the breakdown of distyly in *L. trigynum* via disruption of pollen SI, and should be subject to further functional study. Future studies could

expand our understanding of the breakdown of distyly in *L. trigynum* by investigating whether the alteration of *LtWDR-44* might simultaneously disrupt male SI and contribute to decreased herkogamy, or if more complex evolutionary pathways could have led to the evolution of homostylous flowers. Studies in *Turnera* support the plausibility of this final scenario by showing that the *S*-linked *YUCG* gene determines the functioning and size of pollen, but does not affect stamen length (Henning et al., 2022). Our finding of downregulation of *LtWDR-44* associated with SC and a switch from thrum- to pin-type male SI in *L. trigynum* constitutes an interesting contrast to the mode of loss of distyly in *Primula* and *Fagopyrum*, which involved mutations at *S*-linked candidate genes for style length and female SI function (Huu et al., 2016; Huu et al., 2022; Mora-Carrera et al., 2021, Fawcett et al., 2023).

Results from *Primula* and *Fagopyrum* are in line with theoretical predictions that long homostyles harboring functional dominant male SI function in combination with recessive style length and female SI function should be favored during establishment of homostyly (Dowrick, 1956; Charlesworth and Charlesworth, 1979). Our findings so far do not conform to these theoretical expectations. It is possible that this is the result of historical contingency. The early work by Dowrick (1956) showed that short homostyles could increase in frequency in a distylous population, but they would do so at a slower rate than long homostyles. However, short homostyles would still be expected to have an advantage over distylous individuals under conditions that favor selfing. It is also possible that the function of *S*-locus genes in *Linum* render a transition to homostyly via long homostyles more difficult than in *Primula*. Importantly, in the absence of functional genetic studies of *LtTSS1* it remains uncertain whether loss-of-function mutations at this gene could simultaneously alter both style length and female SI, as is the case in *Primula* (but see results referring to the absence of *TSS1* in homostylous *L. lewisii* by Innes et al., 2023).

On the other hand, both theoretical work (Uyenoyama et al., 2001; Tsuchimatsu and Shimizu, 2013; Shimizu and Tsuchimatsu, 2015) and empirical evidence (e.g., Tsukamoto et al., 2003; Tsuchimatsu et al., 2010; Tsuchimatsu et al., 2012; Bachmann et al., 2019) have shown that mutations impacting male components of the *S*-locus are a likely and common mechanism for transitions to SC in homomorphic SI plant lineages. However, the loss of thrum-specific male SI does not instantly lead to compatibility with all individuals in a distylous population, unlike the case of loss-of-function mutations in dominant *SCR* alleles governing male SI in crucifers (e.g., Bachmann et al., 2019), so these two situations are not entirely analogous. Studies in other clades will be required to determine how gene content at the *S*-locus and the expression patterns of *S*-linked genes varies across distylous and homostylous species, and whether similar genetic mechanisms are generally associated with distyly breakdown in *Linum*.

A caveat regarding the potential causes of loss of distyly in *L. trigynum* that we have identified is that we cannot completely rule out a contribution of mutations at non-*S*-linked loci to reduced herkogamy and SC in *L. trigynum* (Ganders, 1979; Mather and Winton, 1941). Unfortunately, such studies are precluded here by the marked genome-wide differentiation between *L. trigynum* and *L. tenue* which prevents identification of genetic variants associated with floral morph using association mapping. Additionally, we were unable to obtain viable offspring from crosses of *L. trigynum* and *L. tenue*, preventing genetic mapping

in interspecific F2 mapping populations. Functional assays should instead be used to validate candidate genetic changes that triggered loss of distyly in *L. trigynum.*

Transitions from distyly to homostyly are expected to frequently result in elevated self-fertilization rates, with major consequences for genetic variation and the efficacy of selection. To elucidate the timeframe within which the shift to homostyly occurred, as well as its genomic consequences, we analyzed 224 whole-genome sequences of individuals from eight populations each of *L. trigynum* and *L. tenue*. We found that homostyly was associated with predominant self-fertilization, and that the shift to homostyly occurred relatively recently, within the approximately 340 ky since the split between *L. trigynum* and *L. tenue*. Despite this relatively recent shift, there are already marked genomic consequences of self-fertilization. For instance, our finding that *L. trigynum* is less genetically diverse and shows stronger population structure than *L. tenue* are in line with some of the most frequently reported consequences of shifts to selfing (reviewed in Cutter, 2019).

Selfing is often associated with genome-wide relaxed purifying selection (Cutter, 2019). Here, we assessed whether the genome of *L. trigynum* has accumulated TEs and deleterious mutations as a result of reductions in N_e associated with selfing. Repetitive elements like TEs can duplicate and insert across the genome with either neutral or deleterious effects (Quesneville, 2020), and thus they could increase in the genome under weakened purifying selection (Charlesworth and Wright, 2001). On the other hand, selfers could experience a reduction in TE content due to the efficient purging of recessive TE insertions with deleterious fitness effects in selfers (Wright et al., 2008, Roessler et al., 2020). Our results suggest that selfing in *L. trigynum* has not precipitated the accumulation of TEs compared to the outcrossing distylous *L. tenue*. Studies conducted in other plant selfing lineages showed higher copy numbers of TEs per individual, but accumulation patterns differed across TE families (Bonchev and Willi, 2018). Furthermore, the strength of selection against TEs is expected to be especially strong if they interfere with genes or their regulatory regions (reviewed by Quesneville, 2020; Horvath et al., 2017, Roessler et al., 2020). In consequence, forthcoming analyses would benefit from exploring TE accumulation at a more refined scale by comparing different TE families, and variation with respect to gene proximity.

Reductions in N_e resulting from bottlenecks and an increased impact of linked selection due to elevated inbreeding are expected to result in reduced efficacy of selection against weakly deleterious mutations in selfers (Wright et al., 2013; Burgarella & Glémin, 2017; Cutter, 2019). In line with this prediction, we found evidence for weaker purifying selection in *L. trigynum* than in *L. tenue* based on analyses of the distribution of fitness effects of new nonsynonymous (0-fold degenerate) mutations. With the data at hand, we were unable to assess the full distribution of fitness effects, including positive selection, but using improved outgroup genomes for reliable polarization future studies should address this question. In addition, detailed analyses separately assessing selection on specific classes of genes, such as those involved in the evolution of the selfing syndrome or pollen competition (Gutiérrez-Valencia et al., 2022b) would be warranted for a more complete understanding of the impact of shifts to selfing on selection across the genome. With these caveats, our analyses demonstrate genomic signatures consistent with relaxed selection on nonsynonymous mutations in association with loss of distyly. These results are in line with previous

studies on the genomic impact of loss of distyly in *Primula* (Wang et al., 2021), and with a wealth of studies on the genomic impact of shifts to selfing in other plant lineages (e.g., Slotte et al., 2010; Slotte et al., 2013; Laenen et al., 2018; Mattila et al., 2020; Yi et al., 2022). Finally, comparative population genomic analyses of multiple shifts from distyly to homostyly in *Linum* will be valuable to better understand general genomic effects of this evolutionary transition.

Taken together, this study harnesses new genome assemblies, gene expression analyses, and population genomic analyses to extend our understanding of the breakdown of distyly to a new system, provide a basis for future research on the pathways associated with loss of distyly, and broaden our knowledge on the evolutionary genomic consequences of shifts to selfing.

Methods

Plant Material and Sequencing

For genome sequencing and population genomic analyses, seeds and leaves of *L. trigynum* and *L. tenue* were sampled in 16 different localities in southern Spain, representing 8 populations per species (Table S1, Supplementary Information). Plants were grown in a controlled climate chamber at Stockholm University (Stockholm, Sweden) set to 16 h light at 20°C: 8 h dark at 18°C, 60% maximum humidity, 122 µE light intensity. We extracted DNA from 224 individuals for whole-genome short-read sequencing on an Illumina NovaSeq 6000 on a S4 flowcell (150-bp paired-end reads, v1.5 sequencing chemistry). For long-read and linked-read sequencing we extracted high molecular weight (HMW) DNA from young leaves using the CTAB method (as in Fulton et al., 1995), followed by two purification steps with Genomic tip 500/G (QIAGEN, Germany). To generate a high-quality genome assembly, HMW DNA of individual *Ltri*. 6-30-2 was sequenced on 2 Sequel II SMRT cells in HiFi mode, which led to the production of 27 Gb of HiFi data. This was supplemented with Dovetail Hi-C sequencing data (OmniC) for the same individual. For linked-read sequencing of two samples of *L. trigynum* (*Ltri.* 1-1-1 and *Ltri*. 1-42-2) and four samples of *L. tenue* (*Lten*. CL-3-1, *Lten*. CL-75-1, *Lten*. STM-5-2, and *Lten.* STM-30-1) we used the Chromium Genome Library preparation kit with sequencing on an Illumina HiSeqX system (paired-end 150bp read length, v2.5 sequencing chemistry). For genome annotation, we extracted total RNA from stems, leaves, floral buds, and mature flowers of individual *Ltri*. 6-30-2, and for differential expression analyses, we extracted total RNA from floral buds and leaves of *L. tenue* and *L. trigynum* (Sample size: *L. tenue* thrum=6, pin=4, *L. trigynum* homostyle=5) using the RNeasy Plant Mini Kit (QIAGEN, Germany). Libraries were sequenced on an Illumina NovaSeq S1 Sequencing System to produce paired-end 150bp read length reads. Full details on plant growth, sampling, extraction and sequencing are given in Supplementary Methods, Supplementary Information.

L. trigynum de novo *genome assemblies*

HiFi PacBio subreads were assembled with IPA (v1.3.2) (https://github.com/PacificBiosciences/pbipa) using reads with QV20 or higher, which lead to a preliminary assembly of 103 primary contigs (total length = 498.60 Mb) and 2,280 associated contigs (total length = 100.57 Mb). To check for primary contigs that

should be identified as haplotigs we used Purge Haplotigs (Roach et al., 2018). Illumina short-reads obtained from individual *Ltri*. 6-30-2 were mapped to the primary IPA assembly using minimap2 (Li, 2018), and the alignments were processed with the function purge_haplotigs to generate a coverage histogram for contigs. The resulting histogram showed a unimodal distribution, suggesting that purging was not required. The assembly was scaffolded using Dovetail Hi-C data using 3D-DNA scaffolding (Dudchenko et al., 2017), which was pre-filtered to only keep contacts with a mapping quality higher than 30. The expected haploid number of chromosomes in *L. trigynum* is 10 (Pastor et al., 1990), and the resulting scaffolded assembly consisted of 13 pseudochromosomes (N50 = 47.03 Mb, Length = 498.10 Mb). This assembly was further edited to correct three misassemblies that were not supported by mapping of our PacBio reads to the *L. trigynum* assembly. The edited *L. trigynum* genome assembly was polished two times with Pilon (Walker et al., 2014) using Illumina short reads of the same individual. We screened the genome assemblly for regions with high coverage and with similarity with the NCBI Plastid database (https://www.ncbi.nlm.nih.gov/genome/organelle/). We detected and hard-masked one region with plastid contamination on chromosome 6 (positions 59960001-60370000). We visualized broad-scale genome synteny between our *L. tabrigynum* and *L. tenue* (Gutiérrez-Valencia et al., 2022a) assemblies by aligning genome assemblies using minimap2 v2.4 (Li, 2018) and plotting alignments larger than 100 kb using the R package circlize (Gu et al., 2014) (Fig. 2a).

Genomic linked-read sequences (10x Genomics) of two *L. trigynum* and four *L. tenue* samples were assembled with the Supernova pipeline (https://github.com/10XGenomics/supernova) using default parameters to obtain the output type pseudohap2. Furthermore we included three additional *L. tenue* linked-read assemblies from Gutiérrez-Valencia et al. (2022a).

Genome Annotation

The annotation of *L. trigynum* was made with TSEBRA (Gabriel et al., 2021) which combines the gene prediction of BRAKER1 (Hoff et al., 2015) and BRAKER2 (Brůna et al., 2021). BRAKER1 predicts genes using RNAseq data, therefore we trimmed the raw reads from four different tissues (leaves, stems, floral buds and mature flowers) with fastp (Chen et al., 2018), after we aligned them using STAR (Dobin and Gingeras, 2015). BRAKER2 predict genes using protein databases, therefore we use protein data from *L. tenue* (Gutiérrez-Valencia et al., 2022a) and four additional Malpighiales species (*L. usitatissimum*, *M. esculenta*, *P. trichocarpa*, *S. purpurea*) download from phytozome (https://phytozome-next.jgi.doe.gov/) (Goodstein et al., 2012). We used AGAT (https://github.com/NBISweden/AGAT) to extract the coding DNA sequences needed to assess the genome completeness with BUSCO v5 (Manni et al., 2021). Repetitive elements were identified using RepeatMasker (Smit et al., 2013) using a custom repeat library modelled using RepeatModeler (Smit & Hubley, 2008).

We reannotated the *S*-locus genes in *L. tenue* and in *L. trigynum*. First, we identified the transcripts with StringTie v2.1.4 (Pertea et al., 2015) using the RNA-seq data from (Gutiérrez-Valencia et al., 2022a). Second, we predicted the proteins using TransDecoder v5.7.0 (https://github.com/TransDecoder/). Third, we visually inspected the annotated genes using IGV v2.12.3 (Thorvaldsdóttir et al., 2012) retaining those well-

supported by RNA-seq evidence. In comparison to the previous *L. tenue* annotation (Gutiérrez-Valencia et al., 2022a), two genes (*indelg1* and *LtTSS1*) were modified, leaving a total of nine genes at the dominant allele of the *L. tenue S*-locus. Six genes with high similarity to repeats (*indelg3, indelg5, indelg6, indelg7, indelg8, indelg9*) that were not analyzed in Gutiérrez-Valencia et al. (2022a) were not retained in our updated annotation. In *L. trigynum* S-locus region we removed 22 genes annotated by TSEBRA because they are not well-supported by RNA-seq evidence and we manually reannotated two genes (*indelg2* and *LtTSS1*) in order to have the same intron-exon structure of the orthologs genes in *L. tenue*. Finally, we used NUCmer (Kurtz et al., 2004) to identify orthologous regions between *L. tenue* and *L. trigynum* in the *S*-locus region (Fig. 2d).

Identification of Candidate Loss of Function S-*Linked Mutations*

Using minimap2 (Li, 2018), we mapped the sequence of the *L. tenue* distyly *S*-locus (Gutiérrez-Valencia et al., 2022a) against the genome assembly of *L. trigynum* to identify and extract the sequence homologous to this region (Fig. 2b). The same approach was used to retrieve contigs containing the *S*-locus from 10X genomics supernova assemblies (two *L. trigynum* and six *L. tenue.* We conducted BLAST analyses (Camacho et al., 2009) to identify the sequences of *S*-locus genes in each assembly. Sequences of each gene were independently aligned using MUSCLE v3.8.31 (Edgar, 2004), and only coding sequences were kept for further analyses. Coding sequences were aligned using codon-aware alignment in webPRANK (Löytynoja and Goldman, 2010). We inspected each alignment using AliView (Larsson, 2014) to identify major effect mutations (non-consensus splice sites and premature stop codons). Estimates of mean synonymous (d_S) and nonsynonymous divergence (d_N) between *L. tenue* and *L. trigynum* were obtained in MEGA X (Kumar et al., 2018; Stecher et al., 2020) using the Nei-Gojobori model (Nei and Gojobori, 1986), with standard error estimates obtained using 1000 bootstrap replicates. Ambiguous codon positions were removed for each sequence pair.

Sequence Processing, Mapping, Variant Calling and Filtering

Illumina short reads from 224 individuals representing populations of *L. tenue* and *L. trigynum* were quality and adaptor trimmed with *bbduk* from BBMap/BBTools (Bushnell, 2015). Trimmed paired-end reads were mapped to the *L. tenue* genome assembly using BWA-MEM (v0.7.17) (Li, 2013). Alignments of *L. tenue* and *L. trigynum* short reads to the *L. tenue* reference genome were used for coverage analyses focusing on the *S*-locus region. For joint inference of demographic history and analyses of population structure, short-read sequences of both species were mapped to the *L. tenue* genome assembly (Gutiérrez-Valencia et al., 2022a). For all remaining analyses, sequences of each species were mapped to their corresponding reference genome, and processed independently in downstream analyses leading to variant calling.

Alignments with mapping quality lower than 20 were discarded, and we used *MarkDuplicates* from Picard tools v2.0.1 (Broad Institute, 2019) to remove duplicated reads from the alignment. The resulting alignments were used to obtain genotype likelihoods with *mpileup*, and variants (SNPs/INDELs) and invariant sites were identified by samtools/bcftools, using the model for multiallelic and rare-variant calling

(Danecek and McCarthy, 2017). The VCF file was processed to keep only biallelic SNPs and invariant sites, and then filtered based on the maximum proportion of missing data (pm = 0.1) and read depth (5 < dp < 200). To avoid false heterozygous calls based on a low number of alternate alleles, we used a combination of allele balance and coverage filtering, which has previously been successful for highly repetitive plant genomes (see Laenen et al., 2018 and Gutiérrez-Valencia et al., 2022b for a detailed description).

Coverage Analyses

To investigate differences in depth of coverage between pin (n=25), thrum (n=26) (*L. tenue*) and homostyle (n=104) (*L. trigynum*) at the *S*-locus, sequences mapped to both the *L. tenue* and *L. trigynum* genome assemblies were processed to remove repetitive regions identified with RepeatMasker (Smit et al., 2013) using the *L. usitatissimum* repeat library. We used BEDTools (Quinlan et al., 2010) to estimate coverage for 50 kb windows across the genome. Estimates were further processed using in R to estimate normalized mean coverage across windows, and differences between morphs were tested using a Kruskal-Wallis test, followed by a post-hoc Dunn's test with Bonferroni correction for multiple testing.

Haplotype Network Analyses

To investigate whether loss of distyly in *L. trigynum* might have occurred repeatedly, we conducted a haplotype network analysis of two genes at the *S*-locus, *LtTSS1* and *LtWDR-44*. We used the genome annotations of *L. tenue and L. trigynum* species to extract coding sequences for these two genes. We then extracted the corresponding sequences from our short-read data using bam2consensus, a tool from the package tool bambam v.1.4 (Page et al., 2014). In total, 67 *L. tenue* and 100 *L. trigynum* individuals had sufficient coverage and were included in this analysis. We aligned all sequences using codon-aware alignment in PRANK (Löytynoja and Goldman, 2010). Haplotypes and haplotype network were assessed using the R package pegas v.1.2 (Paradis, 2010).

Differential Expression Analyses

RNASeq raw reads from floral buds and leaves were processed with the function bbduk from BBMap/BBTools (Bushnell, 2015) for quality and adapter trimming (parameters k=2, mink=11, ktrim=r, minlength=50, qtrim=rl, trimq=20, hdist=1, tbo, tpe). Reads were mapped and quantified with STAR (Dobin and Gingeras, 2015), using the genome reference and the longest isoform per transcript from our updated annotation of *L. tenue*. Multimapping reads were discarded by using the flag outFilterMultimapNmax=1. Files listing counts mapped to each feature (ReadsPerGene.out.tab) were further processed in R to conduct differential expression analyses using the package DESeq2 (Love, Huber and Anders, 2014). Two independent contrasts were conducted for floral buds and leaves separately: homostyles were first contrasted to thrums, and then to pins. We corrected for multiple testing with the Benjamini-Hochberg method, and genes with an adjusted Log2-fold change > |1.5| and $p < 0.01$ were considered significantly differentially expressed.

Pollination Assays

To test the functionality of male and female SI in *L. trigynum*, we conducted controlled reciprocal crosses of *L. trigynum* (*n*=2 individuals) to both pin (*n*=2-5) and thrum (*n*=2-4) morphs of *L. tenue.* For comparison we also conducted self-pollination of *L. trigynum*, compatible (pin x thrum, thrum x pin) and incompatible pollinations (thrum x thrum, pin x pin) in *L. tenue* and negative controls without pollination. We did three technical replicates of each type of cross of two individuals. For pollination, whole pistils were removed from mature flower buds or recently opened flowers, placed on agar plates and hand pollinated. Pollen tube growth was observed after 4 h. For pollen tube staining, we adapted a protocol by (Mori et al. 2006). Specifically, hand-pollinated pistils were fixed with 9:1 ethanol: acetic acid solution 4 hours after pollination. Pistils were then hydrated with an ethanol series (70%, 50% and 30% ethanol) and softened overnight in 1 M NaOH. Pollen tube growth was observed under an UV fluorescence microscope (Olympus BX60) after staining with 0.1% (w/v) aniline blue solution in 100 mM K_3PO_4 and 2% glycerol.

Population Structure and Timing of Split

The filtered VCF was pruned based on linkage disequilibrium (r^2) prior to conducting structure and demographic analyses with PLINK (Chang et al., 2015) (parameters: window size in kilobase=50, variant count to shift the window at the end of each step=5, pairwise r^2 threshold=0.5). We used the function *--pca* also implemented in PLINK (Chang et al., 2015) to conduct a Principal Component Analysis (PCA) on SNPs in the pruned VCF. The same data set was used for the structure analysis in ADMIXTURE (Alexander et al., 2009) using values of *K* ranging from 2 to 16. The most likely number of subpopulations in the population was determined after identifying the *K* value with the lowest cross-validation error. The results of both the PCA and structure analyses were plotted in R (R Core Team, 2021). Weighted F_{ST} values were calculated using the Weir and Cockerham estimator (Weir & Cockerham, 1984) implemented in VCFTools using the function *--weir-fst-pop* (Danecek et al., 2011). Pairwise F_{ST} values were then compared within and between species using a Kruskal–Wallis test followed by Dunn's test , and *P*-values were corrected using the Bonferroni method in R (R Core Team, 2021). We investigated historical relationships among populations using TreeMix (Pickrell and Pritchard, 2012) with 0 to 5 migration edges, running 100 iterations to get the optimal number of migration edges. Using the evanno method implemented in the R package OptM (Evanno et al., 2005; Fitak, 2021), we got an optimum of two migration edges, which were between populations of *L. trigynum*. We further ran 60 iterations with two migration edges to ensure we had the maximum likelihood tree, as well as 1000 bootstrap replicates to obtain confidence intervals. No admixture was detected between the species. Finally, we used an extension of *dadi* (Gutenkunst et al., 2009) by Blischak et al. (2020) to coestimate inbreeding and demographic parameters of a simple split model for *L. tenue* and *L. trigynum*. For simplicity, we only included one population each of *L. tenue* and *L. trigynum* (populations 32 and 11, respectively) in this analysis.

Estimates of Inbreeding and Polymorphism

We estimated the inbreeding coefficient (F_{IS}) (Wright, 1951) to assist our understanding of the prevalence of selfing in *L. trigynum* using the option *--het* in VCFTools (Danecek et al., 2011). Nucleotide polymorphism (π) was estimated in 100 kb windows per population using pixy (Korunes and Samuk, 2021). Statistical testing for significant differences in F_{IS} and windowed π across populations was conducted in R (R Core Team, 2021).

Estimates of Purifying Selection

To compare the content of TEs between the genomes of *L. tenue* and *L. trigynum*, we created custom libraries of repeats using RepeatModeler (Smit and Hubley, 2008). The resulting libraries were then used to identify loci harboring these repeats using RepeatMasker (Smit et al., 2013). We divided genome sequences in 50 kb windows using BEDTools (Quinlan and Hall, 2010), and estimated the proportion of TEs per window in R (R Core Team, 2021).

We investigated purifying selection at coding sequences using the annotation of both *L. tenue* (Gutiérrez-Valencia et al., 2022a) and *L. trigynum*. We calculated π_N/π_S using pixy with the options *--bed_file* and *--sites_file* to estimate π on a per-gene basis and by restricting the analyses to 0-fold and 4-fold sites. π_N and π_S were computed in R to obtain and compare values of π_N/π_S. These sites were identified using the python script NewAnnotateRef.py (https://github.com/fabbyrob/science/tree/master/pileup_analyzers) (Williamson et al., 2014) ran separately on our annotated high-quality long-read *L. tenue* and *L. trigynum* assemblies, considering only the longest transcript per gene. Finally, estimates of DFE for each population were obtained using fastdfe (v1.0.0) (https://github.com/Sendrowski/fastDFE), a python implementation of polyDFE (Tataru et al., 2017) which supports deleterious DFE inference from folded frequency spectra. We used the model "GammaExpParametrization" which models the DFE under a Γ distribution. In the model, we parametrized the folded SFS using the nuisance parameters with the option "get_demography" to account for demographic history effects. Since we did not have divergence information, we could only obtain the deleterious DFE by setting "p_b=0", while other models provided are equivalent for that purpose. We ran 200 iterations to get the highest maximum likelihood fit, while confidence intervals were obtained with 300 bootstrap replications. As before, 0-fold degenerate sites were considered to be under stronger purifying selection than 4-fold degenerate sites which were assumed to evolve neutrally. Site frequency spectra for 0- and 4-fold sites were obtained with easySFS (https://github.com/isaacovercast/easySFS) which uses a modified implementation of *dadi* (Gutenkunst et al., 2009) with the option "-a" to keep all SNPs. DFE results were summarized in four bins depicting the proportion of new mutations evolving as effectively neutral ($0 > N_e s > -1$), moderately ($-1 > N_e s > -10$), and strongly deleterious ($-10, N_e s > -100$) and ($-100 > N_e s$). The proportion of mutations in each category was compared between species using R (R Core Team, 2021).

Data Availability

609 All sequencing data, genome assemblies and their annotation produced in this study has been uploaded to the
610 European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/) under study accession number
611 PRJEB67577.
612

613 **Acknowledgments**

614 We thank José Ruíz-Martin for assistance with field work, Jerker Eriksson for technical assistance, and
615 Björn Nystedt for discussion of the project. This project has received funding from the European Research
616 Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant
617 agreement No 757451), from the Swedish Research Council (grant no. 2019-04452) and the Erik Philip-
618 Sörensen foundation to T.S., and from the Bergströms foundation to J.G.V., and from the Nilsson-Ehle
619 foundation to P.W.H. Z.P. is financially supported by a Carl Tryggers fellowship (CTS 21:1471). The
620 authors acknowledge support from the National Genomics Infrastructure (NGI) in Stockholm and Uppsala
621 (Uppsala Genome Center, SNP&SEQ), funded by the Knut and Alice Wallenberg foundation, the Swedish
622 Research Council and Science for Life Laboratory. We acknowledge support of the Swedish National
623 Infrastructure for Computing (SNIC) and Uppsala Multidisciplinary Center for Advanced Computational
624 Science (UPPMAX) for assistance with massively parallel sequencing and access to computational
625 infrastructure. Support by NBIS (National Bioinformatics Infrastructure Sweden) is gratefully
626 acknowledged.
627

628 **References**

629 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated
630 individuals. *Genome Research*, *19*, 1655–1664.
631 Bachmann, J. A., Tedder, A., Laenen, B., Fracassetti, M., Désamoré, A., Lafon-Placette, C., Steige, K. A.,
632 Callot, C., Marande, W., Neuffer, B., Bergès, H., Köhler, C., Castric, V. and Slotte, T. (2019). Genetic
633 basis and timing of a major mating system shift in *Capsella*. *New Phytologist*, *224*, 505-517.
634 Baker, H. G. (1955). Self-compatibility and establishment after "long-distance" dispersal. *Evolution*, *9*, 347–
635 349.
636 Barrett, S. C. H. (2019). 'A most complex marriage arrangement': Recent advances on heterostyly and
637 unresolved questions. *New Phytologist*, *224*, 1051–1067.
638 Bateson, W., & Gregory, R. P. (1905). On the inheritance of heterostylism in *Primula*. *Proceedings of the*
639 *Royal Society B: Biological Sciences*, *76*, 581–586.
640 Blischak, P. D., Barker, M. S., Gutenkunst, R. N. (2020). Inferring the demographic history of inbred species
641 from genome-wide SNP frequency data. *Molecular Biology and Evolution*, *37*, 2124–2136.
642 Bonchev, G., Willi, Y. (2018). Accumulation of transposable elements in selfing populations of *Arabidopsis*
643 *lyrata* supports the ectopic recombination model of transposon evolution. *New Phytologist*, *219*, 767–
644 778.
645 Broad Institute. (2019). Picard Tools. *Broad Institute, GitHub Repository*.
646 http://broadinstitute.github.io/picard/

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*, lqaa108.

Burgarella, C., & Glémin, S. (2017). Population genetics and genome evolution of selfing species. In *ELS* (pp. 1–8). John Wiley & Sons, Ltd.

Bushnell, B. (2015). *BBMap/BBTools*. sourceforge.net/projects/bbmap/

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, s13742–015–0047–8.

Charlesworth, D., Charlesworth, B. (1979). A model for the evolution of distyly. *The American Naturalist*, 114(4), 467–498.

Charlesworth, D., Wright, S. I. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics & Development*, *11*, 685–690.

Chen, S., Zhou, Y., Chen, Y., Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*, i884-i890.

Cutter, A. D. (2019). Reproductive transitions in plants and animals: Selfing syndrome, sexual selection and speciation. *New Phytologist*, *224*, 1080–1094.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics, 33*, 2037–2039.

Darwin, C. (1876). *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom*. Cambridge University Press; Cambridge Core.

Darwin, C. (1877). The effects of cross and self fertilisation in the vegetable kingdom. Cambridge University Press; Cambridge Core.

Dobin, A., Gingeras, T.R. (2015). Mapping RNA-Seq reads with STAR. Current Protocols in Bioinformatics, 51, 11.14.1-11.14.19.

Dowrick, V. P. J. (1956). Heterostyly and homostyly in *Primula obconica*. *Heredity*, *10*, 219–236.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, *356*, 92–95.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797.

Ernst, A. (1936). Heterostylie-Forschung. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre*, *71*, 156–230.

Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*, 2611–20.

Fawcett, J.A., Takeshima, R., Kikuchi, S. *et al.* (2023). Genome sequencing reveals the genetic architecture of heterostyly and domestication history of common buckwheat. *Nature Plants*, 9, 1236–1251.

Fitak, R. R. (2021). *OptM*: estimating the optimal number of migration edges on population trees using Treemix. Biology Methods and Protocols, 6, bpab017.

Fulton, T. M., Chunwongse, J., & Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Molecular Biology Reporter*, *13*, 207–209.

Gabriel, L., Hoff, K.J., Brůna, T. et al. (2021). TSEBRA: transcript selector for BRAKER. BMC *Bioinformatics*, *22*, 566

Ganders, F. R. (1979). The biology of heterostyly. *New Zealand Journal of Botany*, *17*, 607–635.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, *40*, D1178-1186.

Gu, Z., Gu, L., Eils, R., Schlesner, M., Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, 19, 2811–2812.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, 5, e1000695.

Gutiérrez-Valencia, J., Hughes, P.W., Berdan, E.L., &Slotte, T. (2021). The genomic architecture and evolutionary fates of supergenes. *Genome Biology and Evolution*, *13*, evab057.

Gutiérrez-Valencia, J., Fracassetti, M., Berdan, E. L., Bunikis, I., Soler, L., Dainat, J., Kutschera, V. E., Losvik, A., Désamoré, A., Hughes, P. W., Foroozani, A., Laenen, B., Pesquet, E., Abdelaziz, M., Pettersson, O. V., Nystedt, B., Brennan, A. C., Arroyo, J., & Slotte, T. (2022a). Genomic analyses of the *Linum* distyly supergene reveal convergent evolution at the molecular level. *Current Biology*, *32*, 4360-4371.

Gutiérrez-Valencia, J., Fracassetti, M., Horvath, R., Laenen, B., Désamore, A., Drouzas, A.D., Friberg, M., Kolár, F., & Slotte, T. (2022b). Genomic signatures of sexual selection on pollen-expressed genes in *Arabis alpina*. *Molecular Biology and Evolution*, 39, msab349.

Henning, P. M., Shore, J. S., McCubbin AG. (2022). The S-gene YUC6 pleiotropically determines male mating type and pollen size in heterostylous Turnera (Passifloraceae): a novel neofunctionalization of the yucca gene family. *Plants*, 11, 2640.

Hartfield, M., Bataillon, T., & Glémin, S. (2017). The evolutionary interplay between adaptation and self-fertilization. *Trends in Genetics*, *33*, 420–431.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., Stanke, M. (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*, 767–769.

Horvath, R., Slotte, T. 2017. The role of small RNA-based epigenetic silencing for purifying selection on transposable elements in *Capsella grandiflora*. *Genome Biology and Evolution*, *9*, 2911-2920.

Huu, C. N., Kappel, C., Keller, B., Sicard, A., Takebayashi, Y., Breuninger, H., Nowak, M. D., Bäurle, I., Himmelbach, A., Burkart, M., Ebbing-Lohaus, T., Sakakibara, H., Altschmied, L., Conti, E., & Lenhard, M. (2016). Presence versus absence of *CYP734A50* underlies the style-length dimorphism in primroses. *eLife*, *5*, e17956.

Huu, C. N., Plaschil, S., Himmelbach, A., Kappel, C., & Lenhard, M. (2022). Female self-incompatibility type in heterostylous *Primula* is determined by the brassinosteroid-inactivating cytochrome P450 CYP734A50. *Current Biology*, *32*, 671-676.e5.

Innes, P. A., Smart, B. C., Barham, J. A. M., Hulke, B. S., Kane, N. C. (2023). Chromosome-scale genome assembly of lewis flax (*Linum lewisii* Pursh.) bioRxiv 2023.10.10.561607; doi: https://doi.org/10.1101/2023.10.10.561607

Korunes, K. L., 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.

Kurtz, S., Phillippy, A., Delcher, A.L. et al. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*, R12.

Laenen, B., Tedder, A., Nowak, M.D., Toräng, P., Wunder, J., Wötzel, S., Steige, K.A., Kourmpetis, Y., Odong, T., Drouzas, A.D., Bink, M., Ågren, J., Coupland, G. & Slotte, T. 2018. Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*. *Proceedings of the National Academy of Sciences of the USA* 115: 816-821.

Laibach, F. (1923). Die Abweichungen vom ‘‘mechanischen’’ Zahlenverhältnis der Long- under Kurz-griffel bei heterostylen Pflanzen. *Biologisches Zentralblatt*, 43, 148–157.

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*, 3276–3278.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint, arXiv:1303.3997.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*, 3094–3100.

Li, J., Cocker, J. M., Wright, J., Webster, M. A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., Oosterhout, C. van, & Gilmartin, P. M. (2016). Genetic architecture and evolution of the *S* locus supergene in *Primula vulgaris*. *Nature Plants*, *2*, 16188.

Lloyd, D. G., & Webb, C. J. (1992). The evolution of heterostyly. In S. C. H. Barrett (Ed.), *Evolution and Function of Heterostyly* (pp. 151–178). Springer.

Love, M. I., Huber, W. Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. (2014). *Genome Biology*, *15*, 550.

Löytynoja, A., Goldman, N. (2010). webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics*, *11*, 579.

Maguilla, E., Escudero, M., Ruíz-Martín, J., & Arroyo, J. (2021). Origin and diversification of flax and their relationship with heterostyly across the range. *Journal of Biogeography*, *48*, 1994–2007.

760 Manni, M., Berkeley, M. R., Seppey, M., Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality
761 and beyond. *Current Protocols*, *1*, e323.
762 Mattila, T. M., Laenen, B., Slotte, T. (2020). Population genomics of transitions to selfing in Brassicaceae
763 model systems. In J. Y. Dutheil (Ed.), Statistical population genomics (pp. 269–287). Springer US.
764 Mather, K., Winton, D. D. (1941). Adaptation and counter-adaptation of the breeding system in *Primula*.
765 *Annals of Botany*, *5*, 297–311.
766 Mora-Carrera, E., Stubbs, R. L., Keller, B., Léveillé-Bourret, É., de Vos, J. M., Szövényi, P., & Conti, E.
767 (2021). Different molecular changes underlie the same phenotypic transition: Origins and consequences
768 of independent shifts to homostyly within species. *Molecular Ecology*, *00*, 1– 18.
769 Mori, T., Kuroiwa, H., Higashiyama, T., & Kuroiwa T. (2006). Generative Cell Specific 1 is essential for
770 angiosperm fertilization. *Nature Cell Biology*, *8(1)*, 64-71.
771 Murray BG. 1986. Floral biology and self-incompatibility in *Linum*. *Bot Gazette* 147: 327-333.
772 Naiki, A. (2012). Heterostyly and the possibility of its breakdown by polyploidization. *Plant Species*
773 *Biology*, *27*, 3–29.
774 Nei, M., Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and
775 nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, *3*, 418–426.
776 Page, J. T., Liechty, Z. S., Huynh, M. D., Udall, J. A. (2014). BamBam: Genome sequence analysis tools for
777 biologists. *BMC Research Notes*, *7*, 829.
778 Paradis, E. 2010. pegas: An R package for population genetics with an integrated-modular approach.
779 *Bioinformatics*, *26*, 419–420.
780 Pastor, J., Diosdado, J. C., Santa Bárbara, C., Vioque, J., & Pérez, E. (1990). Números cromosómicos para la
781 flora Española: 556-619. *Lagascalia*, *15*, 271–274.
782 Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., Salzberg, S. L. (2015). StringTie
783 enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnology*, 3, 290–
784 295.
785 Piper, J. G., Charlesworth, B., Charlesworth, D. 1986. Breeding system evolution in *Primula vulgaris* and
786 the role of reproductive assurance. *Heredity*, *56*, 207-217.
787 Pickrell, J. K., Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele
788 frequency data. *PLOS Genetics*, *8*, e1002967.
789 Quesneville, H. (2020). Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome.
790 *Mobile DNA*, *11*, 28.
791 Quinlan, A. R., Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features.
792 *Bioinformatics*, *26*, 841–842.
793 R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical
794 Computing. https://www.R-project.org/
795 Repplinger, M. (2009). Phylogenie und Biogeographie der Linoideae (Linaceae) – Evolution von
796 Heterostylie / Homostylie in *Linum*. PhD thesis. Johannes Gutenberg-Universität, Mainz, Germany.

797 Roessler, K., Muyle, A., Diez, C. M., Gaut, G. R. J., Bousios, A., Stitzer, M. C., Seymour, D. K., Doebley, J.
798 F., Liu, Q., Gaut, B. S. (2019). The genome-wide dynamics of purging during selfing in maize. *Nature*
799 *Plants*, 5, 980–990.
800 Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for
801 third-gen diploid genome assemblies. *BMC Bioinformatics*, *19*, 460.
802 Ruiz-Martín, J., Santos-Gally, R., Escudero, M., Midgley, J. J., Pérez-Barrales, R., & Arroyo, J. (2018).
803 Style polymorphism in *Linum* (Linaceae): A case of Mediterranean parallel evolution? *Plant Biology*, *20*
804 *Suppl 1*, 100–111.
805 Shimizu, K. K., Tsuchimatsu, T. 2015. Evolution of selfing: recurrent patterns in molecular adaptation.
806 *Annual Review of Ecology, Evolution, and Systematics*, 46, 593–622.
807 Shore, J. S., Hamam, H. J., Chafe, P. D. J., Labonne, J. D. J., Henning, P. M., & McCubbin, A. G. (2019).
808 The long and short of the *S*-locus in *Turnera* (Passifloraceae). *New Phytologist*, *224*, 1316–1329.
809 Slotte, T., Foxe, J. P., Hazzouri, K. M., Wright, S. I. (2010). Genome-wide evidence for efficient positive
810 and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size.
811 *Molecular Biology and Evolution*, *27*, 1813–1821.
812 Slotte, T., Hazzouri, K., Ågren, J. et al. (2013). The *Capsella rubella* genome and the genomic consequences
813 of rapid mating system evolution. *Nature Genetics*, *45*, 831–835.
814 Slotte, T. (2014). The impact of linked selection on plant genomic variation. *Briefings in Functional*
815 *Genomics*, *13*, 268–75.
816 Smit, A. F. A., & Hubley, R. (2008). *RepeatModeler*. https://www.repeatmasker.org/RepeatModeler/
817 Smit, A. F. A., Hubley, R., & Green, P. (2013). *RepeatMasker* (v.4.1.0). https://www.repeatmasker.org/
818 Stebbins, G. L. (1957). Self-fertilization and population variability in the higher plants. *The American*
819 *Naturalist*, *91*, 337–354.
820 Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for
821 macOS. *Molecular Biology and Evolution*, *37*, 1237–1239.
822 Tataru, P., Mollion, M., Glémin, S., Bataillon, T. (2017). Inference of distribution of fitness effects and
823 proportion of adaptive substitutions from polymorphism data. *Genetics*, *207*, 1103–1119.
824 Thorvaldsdóttir, H., Robinson, J. T., Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-
825 performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*, 178-192.
826 Tsuchimatsu, T., Suwabe, K., Shimizu-Inatsugi, R., Isokawa, S., Pavlidis, P., Städler, T., Suzuki, G.,
827 Takayama, S., Watanabe, M., Shimizu, K. K. (2010). Evolution of self-compatibility in Arabidopsis by a
828 mutation in the male specificity gene. *Nature*, 464, 1342-1346.
829 Tsuchimatsu, T., Kaiser, P., Yew, C. L., Bachelier, J. B., Shimizu, K. K. Recent loss of self-incompatibility
830 by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. (2012). *PLoS Genetics*,
831 *8*, e1002838.
832 Tsuchimatsu, T., Shimizu, K. K. (2013). Effects of pollen availability and the mutation bias on the fixation
833 of mutations disabling the male specificity of self-incompatibility. *Journal of Evolutionary Biology*, *26*,
834 2221–2232.

Tsukamoto, T., Ando, T., Takahashi, K., Omori, T., Watanabe, H., Kokubun, H., Marchesi, E., Kao, T. H. (2003). Breakdown of self-incompatibility in a natural population of *Petunia axillaris* caused by loss of pollen function. *Plant Physiology*, *131*, 1903–1912.

Uyenoyama, M. K., Zhang, Y., Newbigin, E. (2001). On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. *Genetics*, *157*, 1805–1817.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, *9*, e112963.

Wang, X.-J., Barrett, S. C. H., Zhong, L., Wu, Z.-K., Li, D.-Z., Wang, H., & Zhou, W. (2021). The genomic selfing syndrome accompanies the evolutionary breakdown of heterostyly. *Molecular Biology and Evolution*, *38*, 168–180.

Weir, B. S., Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Williamson, R. J., Josephs, E. B., Platts, A. E., Hazzouri, K. M., Haudry, A., et al. (2014) Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. *PLOS Genetics*, *10*, e1004622.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354.

Wright, S. (1969). Evolution and the genetics of populations. In *The theory of gene frequencies. Vol 2.* (University of Chicago Press.).

Wright, S. I., Kalisz, S., & Slotte, T. (2013). Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 1760.

Wright, S. I., Ness, R. W., Foxe, J. P., Barrett, S. C. H. (2008). Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences*, *169*, 105–118.

Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O., Ota, T. (2012). *S-LOCUS EARLY FLOWERING 3* is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. PLoS One, 7, e31264.

Yang, J., Xue, H., Li, Z., Zhang, Y., Shi, T., He, X., Barrett, S. C. H., Wang, Q., Chen, J. (2023). Haplotype-resolved genome assembly provides insights into the evolution of *S*-locus supergene in distylous *Nymphoides indica*. *New Phytologist*, In press.

Yi, H., Wang, J., Wang, J., Rausher, M., Kang, M. (2022). Genomic insights into inter- and intraspecific mating system shifts in *Primulina*. *Molecular Ecology*. 31, 5699–5713.

Yuan, S., Barrett, S. C. H., 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.

Zhao, Z., Zhang, Y., Shi, M., Liu, Z., Xu, Y., Luo, Z., Yuan, S., Tu, T., Sun, Z., Zhang, D. and Barrett, S.C.H. (2023). Genomic evidence supports the genetic convergence of a supergene controlling the distylous floral syndrome. *New Phytologist*, *237*, 601–614.

Zhong, L., Barrett, S. C. H., Wang, X.-J., Wu, Z.-K., Sun, H.-Y., Li, D.-Z., 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.

873 **Tables**

874

875 **Table 1.** Sequence divergence between *L. trigynum* and *L. tenue* genome assemblies at the five homologous
876 genes they share in the *S*-locus region. Synonymous and nonsynonymous substitution rates are indicated by
877 d_S and d_N, respectively. Standard errors (SE) of both are given as well as the number of major-effect
878 mutations.

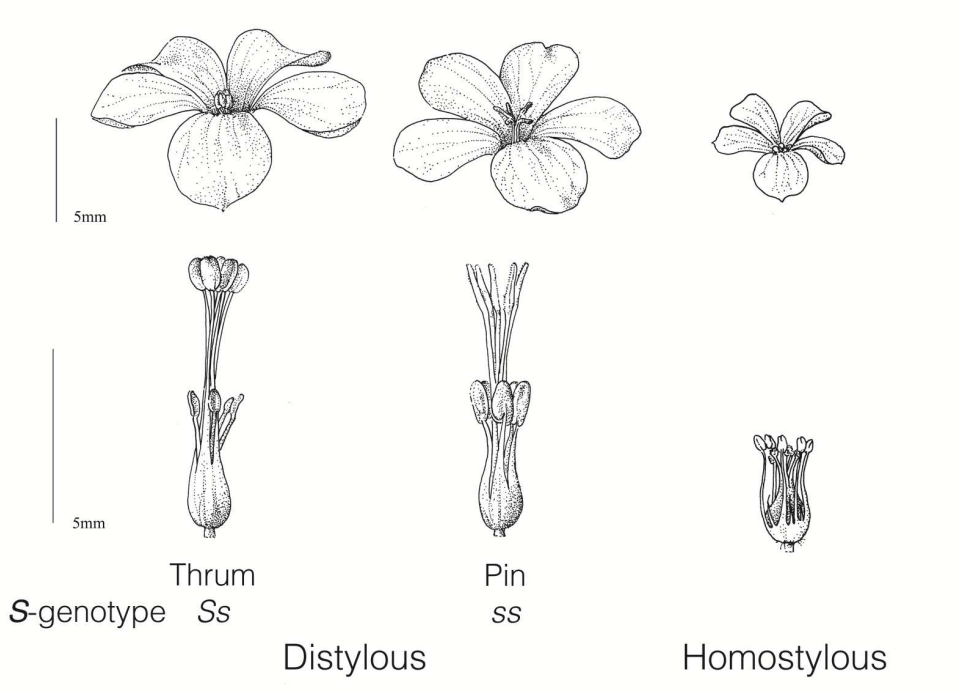
Gene	n_{Ltri}[1]	n_{Lten}[2]	d_S	SE_{dS}	d_N	SE_{dN}	Coding sites	Major-effect
LITEG00000052183	3	4	0.10	0.034	0.018	0.007	176	0
indelg2	3	6	0.036	0.023	0.073	0.020	86	0
LtTSS1	3	8	0.021	0.015	0.006	0.004	144	Segregating premature stop
LtWDR-44	3	8	0.017	0.005	0.012	0.002	925	0
LtTPR LITEG00000052188	3	5	0.018	0.010	0.009	0.004	267	0

879 [1]Sample size, *L. trigynum,* [2]Sample size, *L. tenue*

880

881 **Figures**

882



883

884

885 **Fig. 1. Flower morphs of the distylous *L. tenue* and the homostylous *L. trigynum*.** The expected
886 genotypes of thrum, pin and homostylous flowers are indicated. *S*= dominant haplotype, *s*=recessive
887 haplotype, and *S**= haplotype homologous to the *L. tenue S* allele. Illustrations by Alison Cutts.

888

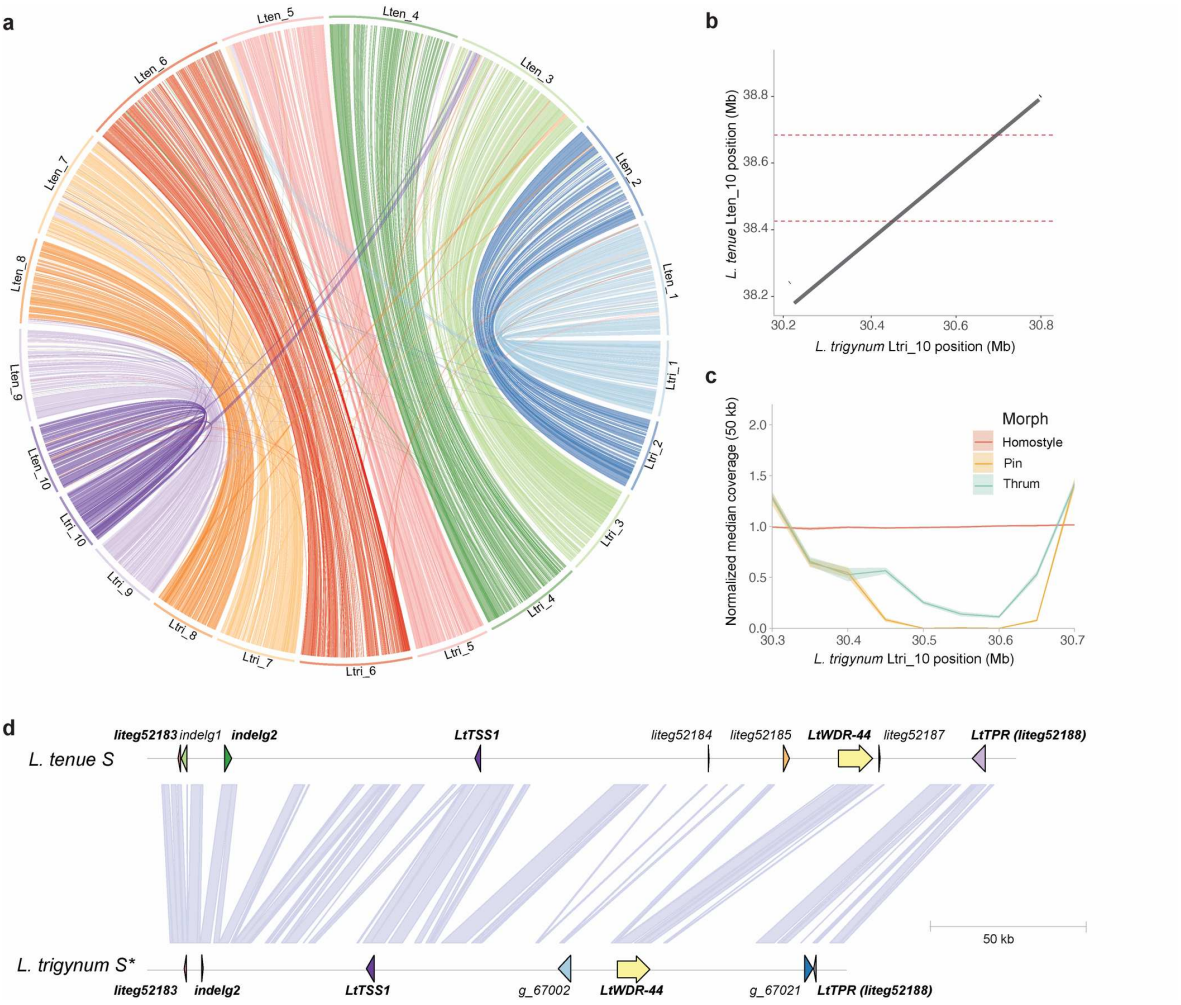


Fig. 2. Broad-scale synteny between *L. trigynum* and *L. tenue* and comparative analyses of the *S*-locus region. (a) Circos plot showing broad-scale synteny between *L. tenue* (top) and *L. trigynum* (bottom). **(b)** Dotplot showing significant minimap2 alignments between the genomic region harboring the dominant *S*-haplotype of *L. tenue* and a region of *L. trigynum* chromosome 10. The limits of the *L. tenue S*-locus are indicated by dotted lines. **(c)** Plot of median and 95% confidence intervals of median coverage in 50 kb windows for *L. trigynum* (homostyle), *L. tenue* thrum and pin individuals, all mapped to our *L. trigynum* assembly. **(d)** Comparison of gene content at the *S*-haplotype of *L. tenue* and corresponding region (here named *S**) in *L. trigynum*. Names of genes shared between the *L. tenue S*-haplotype and the *L. trigynum S**-haplotype are written in bold text.

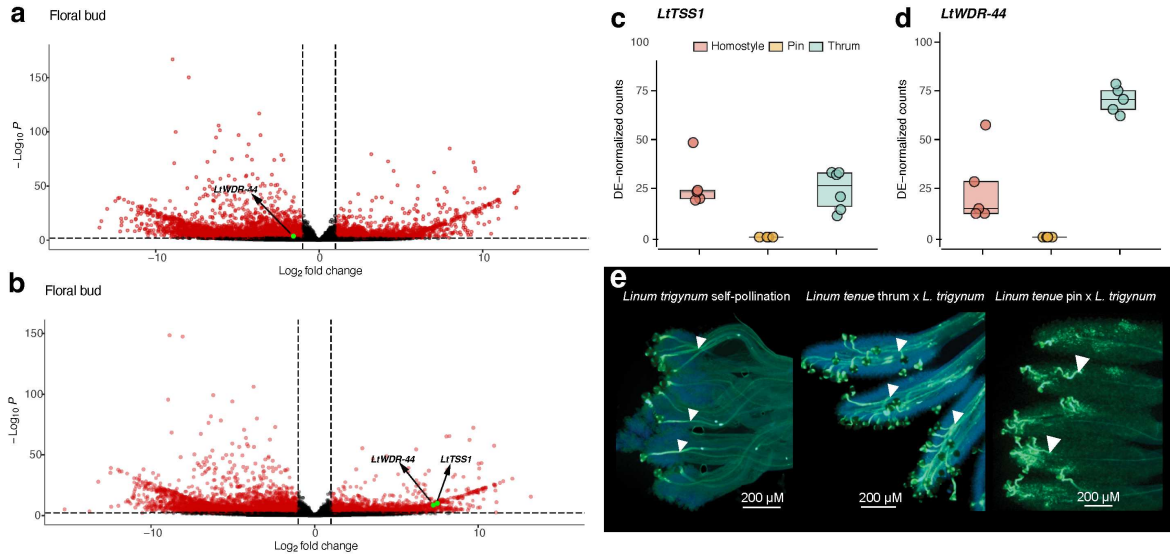


Fig. 3. Differential expression of *S*-linked candidate genes in floral buds and pollination assays suggest a role for downregulation of *LtWDR-44* in transition to homostyly. (a) Volcano plot depicting fold change vs significance of differential expression between *L. trigynum* homostyles and *L. tenue* thrums. Significant *S*-linked genes (only *LtWDR-44* in this analysis) are indicated. (b) Volcano plot depicting fold change vs significance of differential expression between *L. trigynum* homostyles and *L. tenue* pins. Significant *S*-linked genes (*LtWDR-44* and *LtTSS1*) are indicated. (c) Normalized counts for *LtTSS1* in *L. trigynum* homostyles, *L. tenue* pins and *L. tenue* thrums. (d) Normalized counts for *LtWDR-44* in *L. trigynum* homostyles, *L. tenue* pins and *L. tenue* thrums. (e) Representative epifluorescence micrographs of pollination assays demonstrating self-compatibility of *L. trigynum* (left), a compatible reaction when pollinating *L. tenue* thrum (middle) but not pin (right) with *L. trigynum* pollen. Pollen tubes are indicated by white arrows. Note that the site of pollen tube rejection is in the style in *Linum*, such that an incompatible cross yields shorter, aborted pollen tubes. Scale bars indicate the degree of magnification.

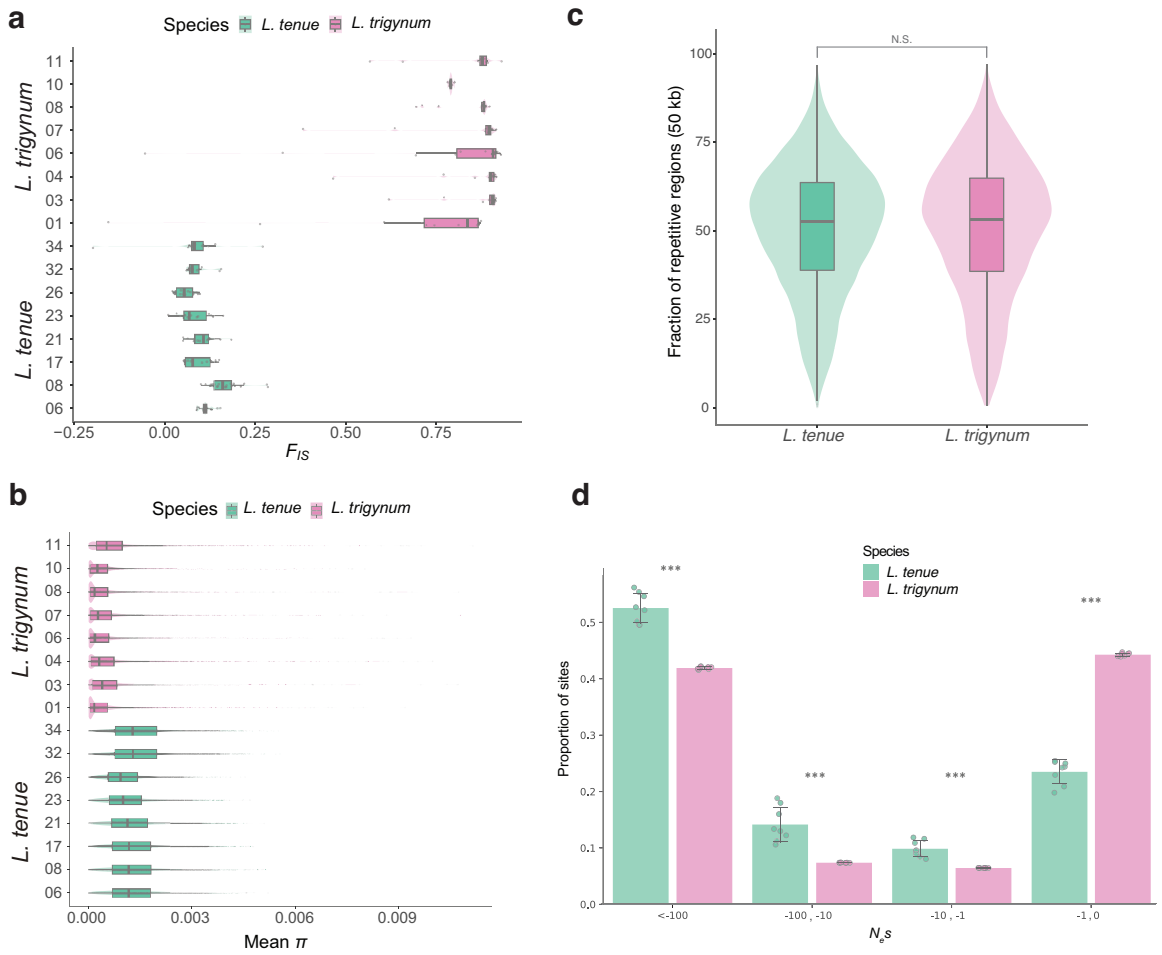


Fig. 4. The shift to homostyly in *L. trigynum* is associated with higher levels of inbreeding and marked effects on population genomic variation. (a) Inbreeding coefficient (F_{IS}) estimates for *L. trigynum* and *L. tenue* indicate higher inbreeding in *L. trigynum.* **(b)** Comparisons of π estimates in windows of 100 kb across the genome show drastically lower nucleotide diversity in *L. trigynum* compared to *L. tenue* populations. (**c)** Estimates of the percentage of TE-derived sequences in 50 kb windows across the genome (t-test: t=1.49, df=20561, $P > 0.05$, N.S.) shows that *L. trigynum* and *L. tenue* have similar TE content. **(d)** Comparison of genome-wide distribution of negative fitness effect estimates for 0-fold degenerate nonsynonymous mutations between *L. trigynum* (n=8) and *L. tenue* (n=8). Error bars represent the standard deviation derived from populations of each species. The fraction of mutations between species was significantly different for each category of N_es ($0 > N_es > -1$ = effectively neutral, $-1 > N_es > -10$ = moderately deleterious, $N_es > 10$ -10 $> N_es > -100$ and $N_es < -100$ = strongly deleterious) (***$P < 0.001$, Wilcoxon rank sum test).

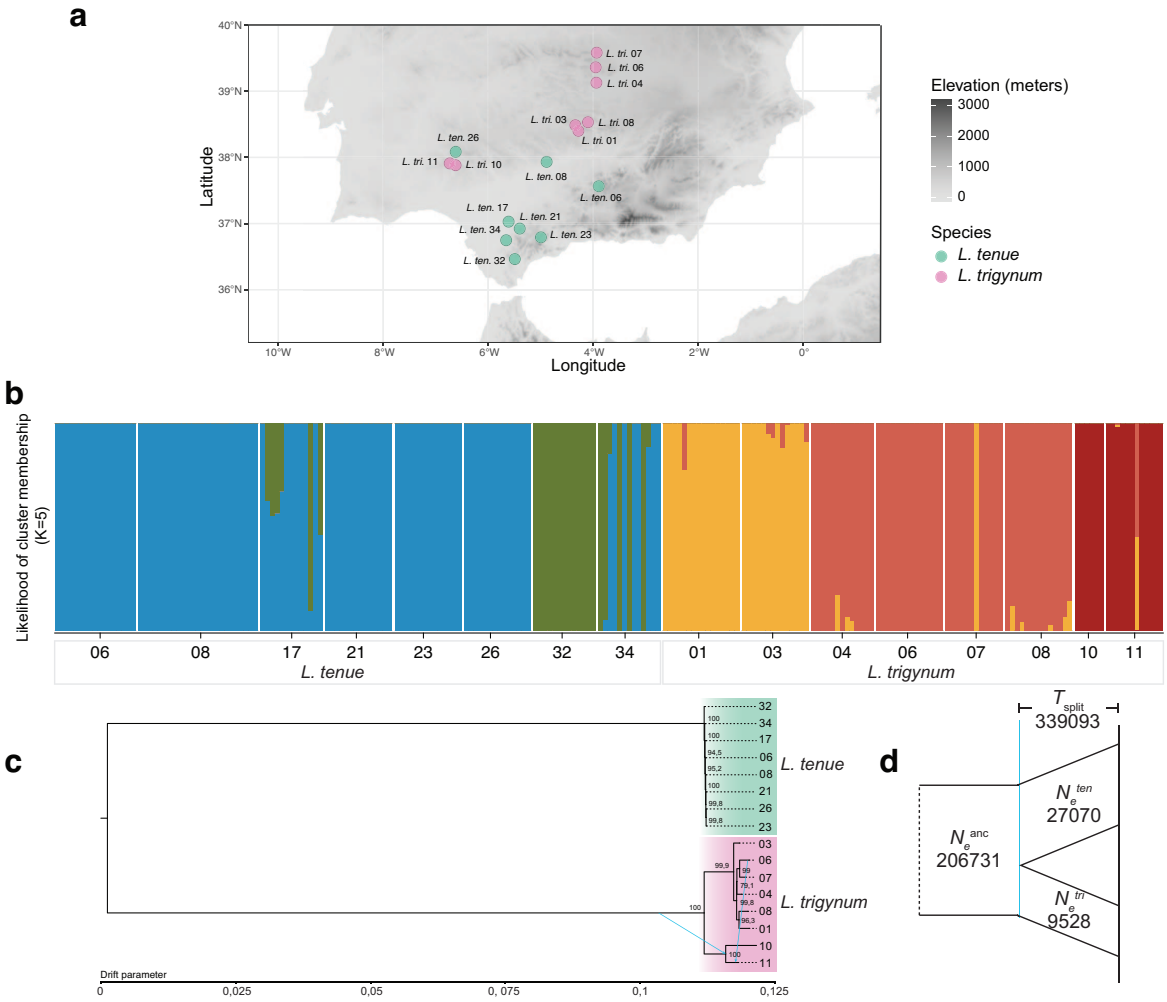


Fig. 5. Population structure and demographic history of *L. trigynum* and *L. tenue*. **(a)** Geographic origin of the Iberian populations of *L. trigynum* (*n*=8) and *L. tenue* (*n*=8) included in this study. **(b)** Assignment of individuals to each of the five inferred ancestral clusters (*K*=5) based on admixture models to describe population structure in ADMIXTURE. **(c)** Maximum likelihood tree inferred by TreeMix. Bootstrap values for each bifurcation and two migration edges within *L. trigynum* are shown. **(d)** The demographic model inferred by dadi. Estimates of ancestral and current effective population sizes (in numbers of individuals) and the time of the split (in number of generations) are shown.