


# The Genomic Selfing Syndrome Accompanies the Evolutionary Breakdown of Heterostyly

Xin-Jia Wang <sup>1,2,3</sup> Spencer C.H. Barrett,<sup>4</sup> Li Zhong,<sup>1,3</sup> Zhi-Kun Wu,<sup>5</sup> De-Zhu Li,<sup>1,2</sup> Hong Wang,<sup>\*,1</sup> and Wei Zhou<sup>\*,1,2</sup>

<sup>1</sup>CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

<sup>2</sup>Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

<sup>3</sup>University of Chinese Academy of Sciences, Beijing, China

<sup>4</sup>Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

<sup>5</sup>Department of Pharmacy, Guizhou University of Traditional Chinese Medicine, Guiyang, Guizhou, China

\*Corresponding authors: E-mails: zhouwei@mail.kib.ac.cn; wanghong@mail.kib.ac.cn.

Associate editor: Yuseob Kim

## Abstract

The evolutionary transition from outcrossing to selfing can have important genomic consequences. Decreased effective population size and the reduced efficacy of selection are predicted to play an important role in the molecular evolution of the genomes of selfing species. We investigated evidence for molecular signatures of the genomic selfing syndrome using 66 species of *Primula* including distylous (outcrossing) and derived homostylous (selfing) taxa. We complemented our comparative analysis with a microevolutionary study of *P. chungensis*, which is polymorphic for mating system and consists of both distylous and homostylous populations. We generated chloroplast and nuclear genomic data sets for distylous, homostylous, and distylous–homostylous species and identified patterns of nonsynonymous to synonymous divergence ( $d_N/d_S$ ) and polymorphism ( $\pi_N/\pi_S$ ) in species or lineages with contrasting mating systems. Our analysis of coding sequence divergence and polymorphism detected strongly reduced genetic diversity and heterozygosity, decreased efficacy of purifying selection, purging of large-effect deleterious mutations, and lower rates of adaptive evolution in samples from homostylous compared with distylous populations, consistent with theoretical expectations of the genomic selfing syndrome. Our results demonstrate that self-fertilization is a major driver of molecular evolutionary processes with genomic signatures of selfing evident in both old and relatively young homostylous populations.

**Key words:** deleterious mutations, effective population size, mating system, *Primula*, selection efficacy.

## Introduction

The evolution of predominant self-fertilization (selfing) from obligate cross-fertilization (outcrossing) is recognized as the most frequent reproductive transition in flowering plants (Stebbins 1957). Although only ~10–15% of angiosperm species are highly autogamous, there is extensive comparative evidence of multiple independent origins of selfing from outcrossing, particularly in herbaceous taxa (Stebbins 1974; Jain 1976; Barrett et al. 1996; Igic et al. 2006). The frequency in which predominant selfing originates in hermaphrodite organisms (Jarne and Charlesworth 1993) provides opportunities to investigate the evolution of convergent morphological (Sicard and Lenhard 2011) and genomic (Cutter 2019) syndromes. Efforts to understand how and why selfing evolves has attracted considerable interest since the early work of Charles Darwin, George Henslow, and Hermann Muller (reviewed in Lloyd 1980), and continues today employing diverse comparative, experimental, and genetic

approaches (Busch and Delph 2012; Igic and Busch 2013; Wright et al. 2013; Barrett et al. 2014). Numerous theoretical models have explored the selective factors influencing variation in plant mating systems (reviewed in Goodwillie et al. 2005), and the recent development of molecular tools enables investigation of the consequences of mating-system transitions on genomic variation and molecular evolution.

Compared with obligate outcrossing, predominant selfing is expected to reduce the effective population size,  $N_e$ , because of nonindependent gamete sampling and the prevalence of genome-wide homozygosity (Pollak 1987; Nordborg 2000). When all else is equal,  $N_e = N / (1 + F_{IS})$ , where  $N$  is the effective population size under panmixia and  $F_{IS}$  is Wright's inbreeding coefficient,  $N_e$  should be diminished up to 2-fold in the case of complete self-fertilization with  $F_{IS} = 1$ . Moreover,  $N_e$  is expected to be reduced further by the action of linked selection (Charlesworth et al. 1993; Kamran-Disfani and Agrawal 2014; Roze 2016), including background selection and selection sweeps, because of the restricted levels of

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

effective recombination and higher linkage disequilibrium (LD) in selfers than outcrossers (Flint-Garcia et al. 2003). In addition, the colonizing ability of many selfers, owing to their ability to benefit from reproductive assurance, may further contribute to decreasing  $N_e$  as a result of recurrent colonization–extinction cycles and severe bottlenecks (Baker 1955; Lloyd 1980; Charlesworth and Wright 2001; Pannell and Fields 2014).

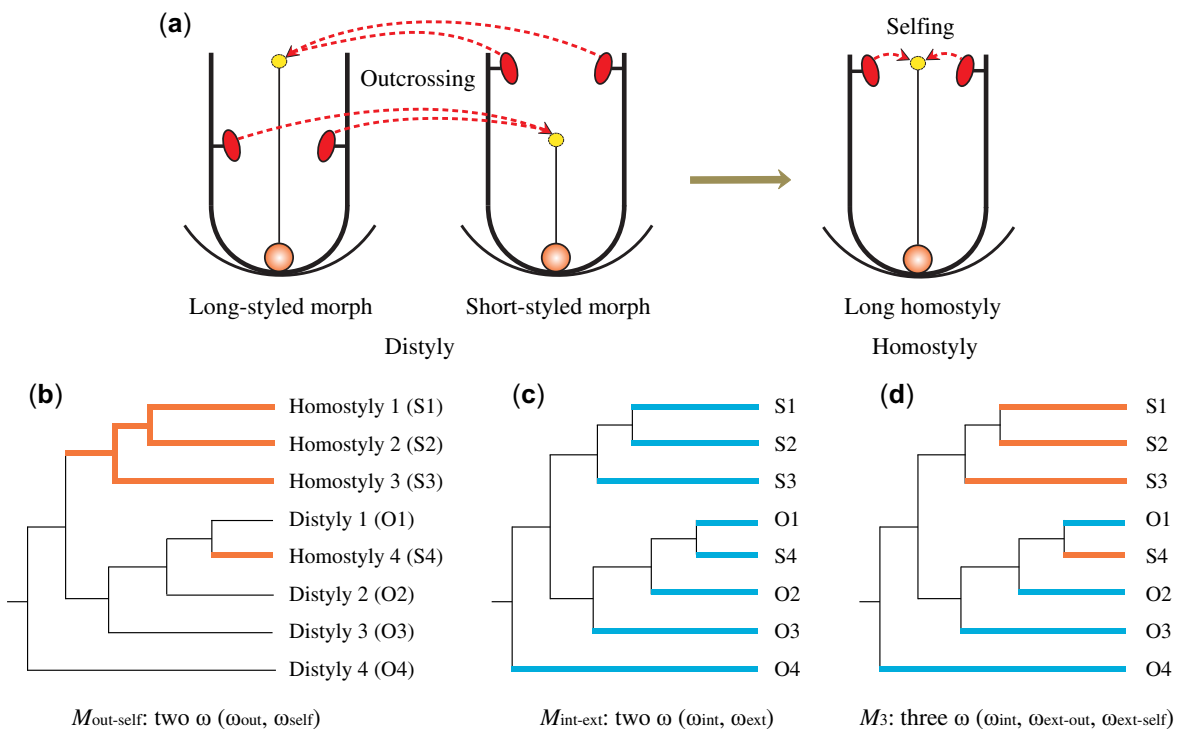
The rate of molecular evolution of mutations subject to selection is a function of  $N_e s$ , the product of the effective population size and the selection coefficient ( $s$ ) (Kimura 1983). As proposed by the nearly neutral theory, there should be a large fraction of slightly deleterious mutations with selection coefficients in the order of  $1/N_e$ , thus differences in the effective population size can strongly affect the outcome of purifying selection against these mutations (Ohta 1992; Akashi et al. 2012). The reduced  $N_e$  in selfers should increase the role of drift relative to selection resulting in a higher proportion of mutations that behave almost neutrally. This process can thus lead to the accumulation of deleterious mutations and stochastic loss of beneficial mutations (“Muller’s ratchet”; Heller and Smith 1978). Assuming similar mutation rates ( $\mu$ ), both the ratio of nonsynonymous to synonymous polymorphism ( $\pi_N/\pi_S$ ) (equivalently, the ratio of 0- and 4-fold degenerate positions in protein-coding sequences,  $\pi_0/\pi_4$ ) and the degree of divergence ( $d_N/d_S$ ) are expected to be higher in selfing than outcrossing lineages, as a consequence of the accumulation and fixation of weakly deleterious mutations. However, strongly deleterious and highly recessive mutations are more likely to be purged by selection because the homozygous genetic background of selfers limits opportunities for the masking of deleterious mutations (Wang et al. 1999; Arunkumar et al. 2015). Despite the theoretical expectation of genomic degradation under recurrent selfing, empirical evidence for selfing as an “evolutionary dead end” is mixed (Takebayashi and Morrell 2001; Igic and Busch 2013). Therefore, further comparisons of the genetics of related outcrossing and selfing plant populations to determine whether “genomic selfing syndromes” (Cutter 2019) are evident should provide further insight into why some selfing lineages persist whereas others go extinct.

Why the efficacy of purifying selection varies among species is a key question in evolutionary genetics. A recent survey of plant and animal species investigated the relation between effective population size and the efficacy of selection and found that among plants, mating system and longevity were significant correlates (Chen et al. 2017). For genes under strong purifying selection, different patterns of selection are expected between selfers and outcrossers, as indicated by a higher ratio of nonsynonymous to synonymous polymorphisms ( $\pi_N/\pi_S$ ) and/or substitutions ( $d_N/d_S$ ). However, these predictions are not always supported by empirical data from comparisons of species pairs with contrasting mating systems. Polymorphism-based measures have mostly supported the prediction of a reduced selection efficacy in selfers (Cutter et al. 2008; Ness et al. 2012; Arunkumar et al. 2015; Burgarella et al. 2015), and this pattern was further confirmed by comparative analysis of genome-wide polymorphism data from

both plants and animals (Chen et al. 2017). However, for comparisons involving  $d_N/d_S$  the evidence for relaxed selection is mixed with some studies supporting weak purifying selection in selfing plants (Slotte et al. 2010, 2013; Qiu et al. 2011; Hazzouri et al. 2012) whereas others have failed to detect the reduced efficacy of selection (e.g., *Arabidopsis*, Wright et al. 2002; Triticeae, Haudry et al. 2008; Escobar et al. 2010; Glémin and Muyle 2014) and also in animals (e.g., *Caenorhabditis*, Cutter et al. 2008).

The recent origin of selfing is often invoked to explain discrepancies between theoretical predictions and empirical data (Wright et al. 2008; Glémin and Galtier 2012; Ness et al. 2012; Cutter 2019). Divergence patterns revealed in relatively young selfers may largely reflect the influence of a long history of outcrossing in the immediate ancestor. This effect is likely to be magnified when the evolution of selfing is associated with speciation, with a change in mating system evolving gradually after reproductive isolation, as may have occurred in the highly self-fertilizing *Arabidopsis thaliana* (Tang et al. 2007). Phylogenetically informed comparative analyses of the efficacy of purifying selection involving related selfers of contrasting age should provide opportunities to investigate changes in the genomic selfing syndrome during the evolutionary history of taxa.

The evolutionary breakdown of the sexual polymorphism heterostyly to homostyly (fig. 1a) provides rich opportunities for investigating transitions from outcrossing to selfing and their genomic consequences (reviewed in Barrett 2019). Populations with this mating polymorphism are usually composed of two (distyly) or three (tristyly) floral morphs that differ reciprocally in stigma and anther height and possess a suite of ancillary polymorphisms involving stigmas and pollen (Darwin 1877; Ganders 1979; Barrett 1992). Heterostylous populations commonly exhibit a heteromorphic self-incompatibility system that limits or entirely prevents self- and intramorph mating, thus enforcing phenotypic disassortative mating. Heterostyly is reliably reported from 28 angiosperm families and in the majority, the polymorphism has broken down giving rise to a monomorphic homostylous condition, with populations usually composed of a single, fully self-compatible floral form with anthers and stigmas close together within a flower (Ganders 1979; Weller 1992; Barrett 2019). Homostylous plants most frequently have the capacity for autonomous self-pollination and, as a result, possess high selfing rates (Piper et al. 1984; Ganders et al. 1985). In distylous taxa, homostyles can either have long styles and long-level stamens (long homostyles) or short styles and short-level stamens (short homostyles), with the former more commonly encountered in natural populations, either as intraspecific variants in distylous populations (Crosby 1949; Charlesworth and Charlesworth 1979; Yuan, Barrett, et al. 2017; Zhou et al. 2017; Shao et al. 2019), or as derived monomorphic varieties or species in otherwise distylous lineages (Schoen et al. 1997; Truyens et al. 2005; de Vos, Wüest, et al. 2014). In contrast to the relatively gradual evolution of high selfing rates via mixed mating, characteristic of most mating-system transitions in angiosperms, predominant selfing arises very rapidly in distylous taxa, by genetic changes at the



**Fig. 1.** (a) Floral phenotypes associated with the breakdown of distyly to long homostyly and the mating-system transition from outcrossing to selfing that is involved. Solid arrow indicates the evolutionary transition; dashed arrows indicate predominant matings. Branch models tested with *codeml* in PAML analyses: (b)  $M_{out-self}$ : two  $\omega$  for outcrossing (thin black) and selfing (thick orange) branches; (c)  $M_{int-ext}$ : two  $\omega$  for internal (thin black) and external (thick blue) branches; (d)  $M_3$ : three  $\omega$  for internal branches (thin black), outcrossing (thick blue), and selfing (thick orange) external branches. Outcrossing and selfing branches on models are indicated by O and S.

S-locus linkage group governing the heterostylous syndrome (reviewed in Kappel et al. 2017).

*Primula* (Primulaceae) provides a valuable comparative system for investigating the evolutionary and genomic consequences of mating-system transitions. Beginning with Darwin's classic work on heterostyly in *Primula* (Darwin 1877), the genus has received over a century of sustained interest and today is the most well-studied heterostylous taxon (Mast and Conti 2006; Gilmartin 2015). The vast majority (92%) of ~400–500 *Primula* species are distylous, with the remaining species monomorphic for styler condition and homostylous. Homostyles are distributed among 19 of the 38 sections in the genus and in all but one they co-occur with distylous species (Mast et al. 2006; de Vos, Wüest, et al. 2014). Phylogenetic analyses and ancestral state reconstructions of *Primula* species demonstrate a single origin of distyly but numerous independent breakdown events to homostyly (Mast et al. 2006; de Vos, Wüest, et al. 2014; Zhong et al. 2019). Significantly, intraspecific investigations of several *Primula* species have provided additional microevolutionary evidence that homostyles are derived from distylous morphs (Crosby 1949; Charlesworth and Charlesworth 1979; Yuan, Barrett, et al. 2017; Zhou et al. 2017; Shao et al. 2019). Thus, in *Primula*, the transition from outcrossing to selfing has occurred repeatedly in different lineages over a range of contrasting timescales, which should allow an evaluation of the

time-dependent effects of selfing among multiple independent transitions.

Here, we employ phylogenetic and comparative genomic approaches of both nuclear and chloroplast protein-coding genes to investigate the genetic and genomic consequences of mating-system transitions in a sample of 66 species of distylous and homostylous *Primula*. We also conducted a detailed microevolutionary investigation of *P. chungensis*, because we previously demonstrated that this species is comprised both distylous, homostylous, and mixed populations containing both floral forms (Zhou et al. 2017). Contrary to most other homostylous species of *Primula*, the homostylous lineages of *P. chungensis* do not exhibit well-developed morphological selfing syndromes and probably represent more recent transitions to selfing from outcrossing than occur in fully homostylous species (Zhou et al. 2017). Thus, our sampling of homostyles provides a novel opportunity to evaluate the genomic consequences of different time scales of the history of selfing in a system in which the initial transition from outcrossing to selfing is rapid.

Our investigation specifically addressed the following questions: 1) Do homostylous species of *Primula* exhibit higher  $d_N/d_S$  ratios than distylous outcrossing species? And if so, is this pattern of divergence consistent for both chloroplast and nuclear genomes given their contrasting patterns of inheritance? 2) At the species level, does the more recently derived selfing lineage of *P. chungensis* exhibit higher  $\pi_N/\pi_S$  ratios

than the ancestral outcrossing lineage? 3) Is the elevated proportion of nonsynonymous mutations observed in *P. chungensis* (see Results) associated with a reduced efficacy of purifying selection on nearly neutral mutations? And, in addition to relaxed selection, is there evidence for the purging effect of selfing on large deleterious mutations?

## Results

### Molecular Evolution of the Chloroplast Genome

The reconstructed phylogeny of 60 *Primula* species using the whole chloroplast genome and ML analyses is presented in [supplementary figure S1, Supplementary Material](#) online. Detailed statistics for the complete plastid genomes are provided in [supplementary table S1, Supplementary Material](#) online. Seventy-two chloroplast coding genes were evident in all 78 accessions.

For protein-coding sequences, we used the *codeml* implementation in PAML to apply different branch models both on individual gene alignments and the concatenated data sets ([fig. 1](#)). In the concatenated data set, the  $\omega$  value was 0.220 under the null one-ratio model (hereafter  $M_0$ ; [supplementary table S2a, Supplementary Material](#) online), indicating that most *Primula* chloroplast genes evolve under purifying selection.  $M_{\text{out-self}}$  provided a significantly better fit than  $M_0$  ( $\chi^2=44.384$ ,  $df=1$ ,  $P < 0.001$ ; [supplementary table S2a, Supplementary Material](#) online). Increased values of  $\omega$  in the plastid coding region of homostylous relative to distylous branches (0.278 and 0.209, respectively) indicated that selfers experience an accelerated accumulation of slightly deleterious mutations compared with their outcrossing counterparts. Concatenated alignment data sets exhibited a statistically significant better fit to  $M_{\text{int-ext}}$  compared with  $M_0$  ( $\chi^2=95.200$ ,  $df=1$ ,  $P < 0.001$ ; [supplementary table S2a, Supplementary Material](#) online). Higher  $\omega$  values in external relative to internal branches (0.251 and 0.177, respectively) were consistent with the hypothesis that most mildly deleterious mutations are removed by purifying selection in the long term, whereas slightly deleterious mutations showed relatively more persistence over the short term.  $M_3$  provided a significantly better fit than  $M_{\text{int-ext}}$  ( $\chi^2=9.398$ ,  $df=1$ ,  $P < 0.01$ ; [supplementary table S2a, Supplementary Material](#) online). In  $M_3$ , internal branches had lower  $\omega$  values (0.177) relative to the two external branches (0.240 and 0.278 for outcrossing and selfing branches, respectively) concordant with the results for  $M_{\text{int-ext}}$  and  $M_{\text{out-self}}$ .

When these genes were analyzed separately, we identified 11 (15.3%) as “evolving differentially” and 10 of 11 (90.9%) had higher  $\omega$  values for selfing branches ([fig. 2a](#)). Furthermore, the results from the branch-site model identified that in selfers none of the ten fast-evolving genes showed signs of positive selection ([supplementary table S3, Supplementary Material](#) online). We identified 23 genes (31.9%) with a statistically significant better fit of  $M_{\text{int-ext}}$  compared with  $M_0$ , based on different  $\omega$  values between internal and external branches, although in most cases (22 of 23) there were higher  $\omega$  values for external branches ([fig. 2b](#)). Overall,  $\omega$  values for selfing and external branches were significantly higher (Wilcoxon-signed

rank test,  $P < 0.001$ ) resulting in a positive difference in ( $\omega_{\text{self}} - \omega_{\text{out}}$ ) and ( $\omega_{\text{ext}} - \omega_{\text{int}}$ ) ([fig. 3a](#)).

### Molecular Evolution of the Nuclear Genome

We provide details of the de novo transcriptome assemblies and coding region identification and annotation in the supplementary results S1, [Supplementary Material](#) online. Values for genome-wide homozygosity of the homostylous species were  $\sim 6$ -fold higher than for distylous species.

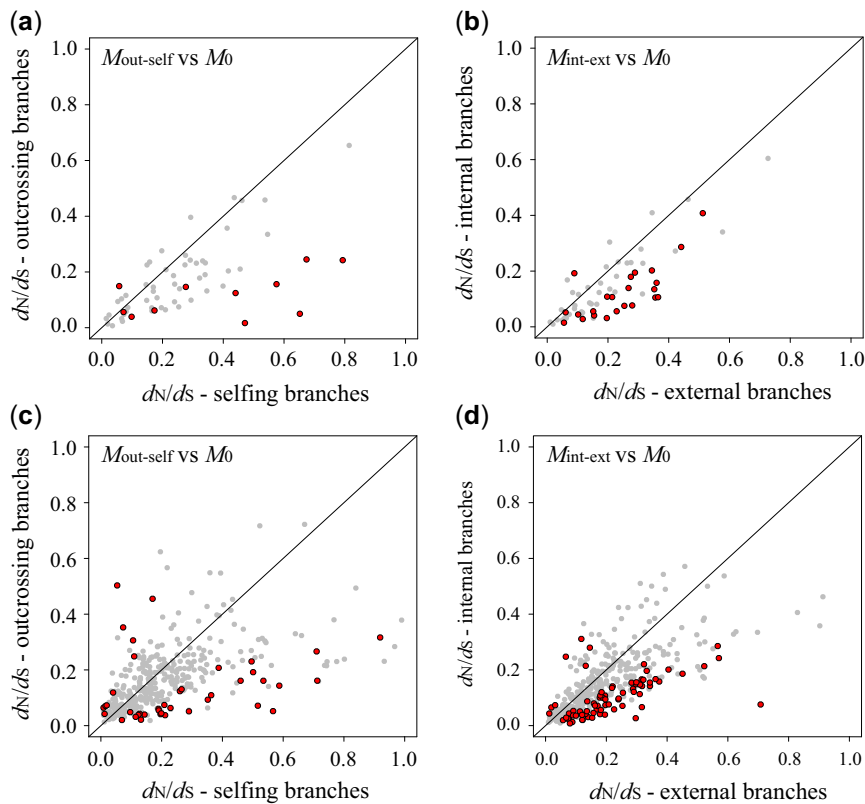
For the concatenated data set, both the two-ratio models  $M_{\text{out-self}}$  and  $M_{\text{int-ext}}$  provided a significantly better fit than either  $M_0$  ( $\chi^2=27.686$ ,  $df=1$ ,  $P < 0.001$ ; [supplementary table S2b, Supplementary Material](#) online) or  $M_3$  ( $\chi^2=3.576$ ,  $df=1$ ,  $P = 0.059$ ; [supplementary table S2b, Supplementary Material](#) online). The  $\omega$  value was higher in selfing than outcrossing species ( $\omega_{\text{self}}=0.175$  vs.  $\omega_{\text{out}}=0.157$ ) in the model  $M_{\text{out-self}}$  and  $\omega_{\text{ext}}$  (0.179) was higher than  $\omega_{\text{int}}$  (0.146) in the model  $M_{\text{int-ext}}$  ([supplementary table S2b, Supplementary Material](#) online).

When analyses were performed on each gene independently, 41 orthologs (9.3% of the investigated genes) were identified as differentially evolving between distylous and homostylous species. Thirty (71%) showed a significantly higher  $\omega_{\text{self}}$  than  $\omega_{\text{out}}$  ([fig. 2c](#)). Few positively selected sites contributed to the elevated  $\omega_{\text{self}}$  in nuclear genes tested by the branch-site model ([supplementary table S3, Supplementary Material](#) online). Similarly, 78 orthologs (17.8%) revealed a significant difference between internal and external branches ([fig. 2d](#)):  $\omega_{\text{ext}}$  was higher than  $\omega_{\text{int}}$  in 71 genes (92.3%). Similar to our results for chloroplast genes, we observed significantly higher values of  $d_N/d_S$  in selfing and external branches when these genes were analyzed independently (Wilcoxon-signed rank test,  $P < 0.001$ ; [fig. 3b](#)).

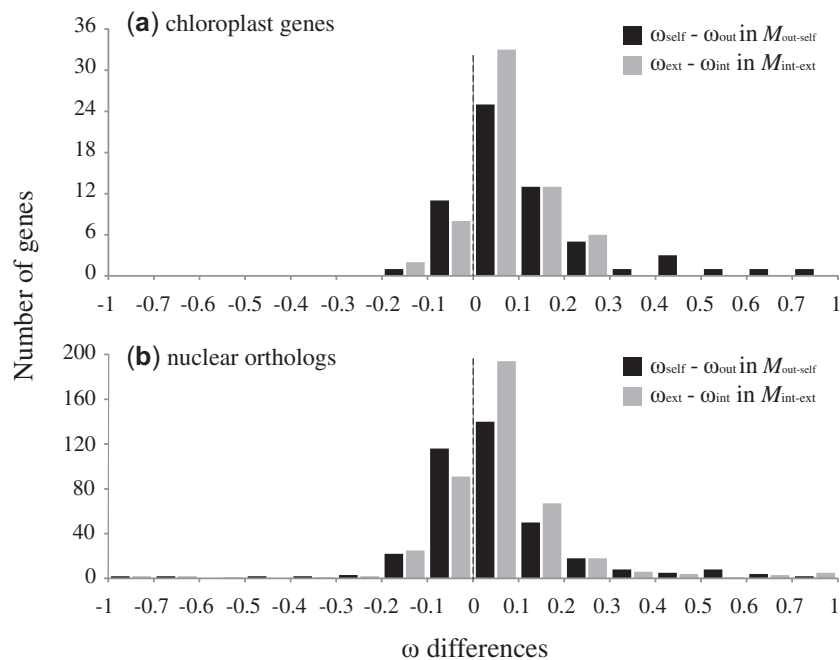
### Patterns of Polymorphism in Lineages of *Primula chungensis* with Contrasting Mating Systems

The chloroplast data set of the *P. chungensis* coding region (72 genes) contained a total of 62,766 sites, of which 16 and 10 were polymorphic in the outcrossing and selfing lineage, respectively. The two lineages shared no polymorphic sites in common and showed divergence at 20 sites. General patterns of nucleotide diversity variation between the two lineages were consistent with that of nuclear sequences described below (selfing vs. outcrossing for  $\pi$ : 0.007% vs. 0.010%;  $\theta_W$ : 0.006% vs. 0.012%; [table 1a](#)). The  $\pi_N/\pi_S$  ratio was also higher in the selfing than outcrossing lineage (selfing vs. outcrossing: 0.385 vs. 0.368). The mean Tajima's  $D$  was negative in the outcrossing lineage, but positive in the selfing lineage ([table 1a](#)).

The nuclear genome data set of *P. chungensis* contained a total of 2,954 gene alignments and 1,858,710 sites. Compared with the outcrossing lineage (mean  $\theta_W=0.074\%$ ,  $\pi = 0.074\%$ ; [table 1b](#)), the level of polymorphism of the selfing lineage was significantly reduced (mean  $\theta_W=0.052\%$ ,  $\pi = 0.053\%$ ; [table 1b](#)). The ratio of nonsynonymous to synonymous nucleotide diversity ( $\pi_N/\pi_S$ ) was slightly higher in the selfing than in the outcrossing lineage (0.284 and 0.280, respectively). The mean Tajima's  $D$  was slightly negative for polymorphic



**FIG. 2.** Protein evolution of 72 chloroplast genes among 60 *Primula* species and 445 nuclear orthologs among 12 *Primula* species. Plot of  $d_N/d_S$  values in selfing ( $x$ -axis) and outcrossing ( $y$ -axis) branches from the model  $M_{out-self}$  for (a) chloroplast genes and (c) nuclear orthologs, respectively. Plot of  $d_N/d_S$  values in external ( $x$ -axis) and internal ( $y$ -axis) branches from the model  $M_{int-ext}$  for (b) chloroplast genes and (d) nuclear orthologs, respectively. Genes with statistically significant differences between model  $M_{out-self}/M_{int-ext}$  and null model ( $M_0$ ) are in red.



**FIG. 3.** Distributions of the differences of  $\omega$  ( $d_N/d_S$ ) estimated separately in  $M_{out-self}$  and  $M_{int-ext}$  for (a) 72 chloroplast genes and for (b) 445 nuclear orthologs, respectively. Two ratios ( $\omega_{out}$  and  $\omega_{self}$ ) in  $M_{out-self}$  are given for outcrossing and selfing branches. Two ratios ( $\omega_{int}$  and  $\omega_{ext}$ ) in  $M_{int-ext}$  are estimated for internal and external branches.

**Table 1.** Summary of Chloroplast and Nuclear Genome Data for the Two Lineages (outcrossing vs. selfing) of *Primula chungensis*.

Lineage	<i>n</i>	Site Class	<i>N</i> <sub>sites</sub>	<i>S</i>	$\pi$ ( $10^{-3}$ )	$\theta_w$ ( $10^{-3}$ )	$\pi_N/\pi_S$	$\theta_N/\theta_S$	<i>D</i>
<b>(a) Chloroplast data</b>									
Outcrossing lineage (distylous)	5	Synonymous	14,549	7	0.19	0.23	0.368	0.391	−1.222
		Nonsynonymous	48,216	9	0.07	0.09			
Selfing lineage (homostylous)	8	Synonymous	14,547	4	0.13	0.11	0.385	0.455	0.826
		Nonsynonymous	48,218	6	0.05	0.05			
<b>(b) Nuclear data</b>									
Outcrossing lineage (distylous)	4	Synonymous	435,550	1,321	1.65	1.64	0.280	0.287	−0.050
		Nonsynonymous	1,432,160	1,222	0.47	0.47			
Selfing lineage (homostylous)	9	Synonymous	435,534	1,331	1.17	1.12	0.284	0.295	0.131
		Nonsynonymous	1,423,176	1,258	0.34	0.33			

NOTE.—Nucleotide diversity ( $\pi$ ), Watterson's theta ( $\theta_w$ ), and test for neutrality (*D*; Tajima's *D* based on  $\pi$  and  $\theta_w$ ) for each lineage are given. *n*, sample size; *N*<sub>sites</sub>, number of sites; *S*, polymorphisms.

sites in the outcrossing lineage but positive in the selfing lineage (table 1b), likely because of the higher subdivision among selfing populations, a result consistent with previous work on the phylogeography of the species (Zhou et al. 2017).

For the concatenated nuclear genomic data set of *P. chungensis*, we also conducted protein evolution analyses by adding three- and four-ratio models ( $M_{3-1}$  and  $M_{4i}$ ; supplementary fig. S2, Supplementary Material online). Here, the  $M_{3-1}$  model, which allows different  $\omega$  values for internal selfing, internal outcrossing, and external branches, was a better fit than model  $M_{int-ext}$  ( $\chi^2=7.788$ ,  $df=1$ ,  $P < 0.01$ ; supplementary table S2c, Supplementary Material online). Significant divergence of the  $\omega$  ratio was obtained between the two internal branches with the selfing lineage higher than outcrossing lineage ( $\omega_{int-out}=0.284$ ;  $\omega_{int-self}=0.328$ ; supplementary table S2c, Supplementary Material online).

### McDonald–Kreitman Tests for Selection in Lineages of *P. chungensis* with Contrasting Mating Systems

We conducted McDonald–Kreitman (MK) tests on each nuclear locus in selfing and outcrossing lineages of *P. chungensis* and summarized our results using the neutrality index (NI). We found a significant mating system effect on the NI. The NI value in the outcrossing lineage was close to 1 (NI=1.079; Fisher's exact test,  $P = 0.088$ ; fig. 4a) which fit the null expectation of equal ratios of divergence and polymorphism. In contrast, there was a significant deviation of neutrality (NI=1) for the selfing lineage with higher NI=1.162 (Fisher's exact test,  $P < 0.001$ ; fig. 4a), and thus evidence for the accumulation of slightly harmful mutations.

We also conducted MK tests on the chloroplast data set of *P. chungensis* (fig. 4b). Both selfing (NI=1.906; Fisher's exact test,  $P = 0.317$ ) and outcrossing lineages (NI=1.653; Fisher's exact test,  $P = 0.442$ ) had higher NI values than 1, but the differences were not significant (fig. 4b).

### Distribution of Fitness Effects of the Selfing and Outcrossing Lineage of *P. chungensis*

For the distylous and homostylous lineage of *P. chungensis*, we estimated the genome-wide distribution of fitness effects (DFE) of nonsynonymous mutations by using the folded synonymous and nonsynonymous site frequency spectra (SFS) (Eyre-Walker and Keightley 2009). We found differences

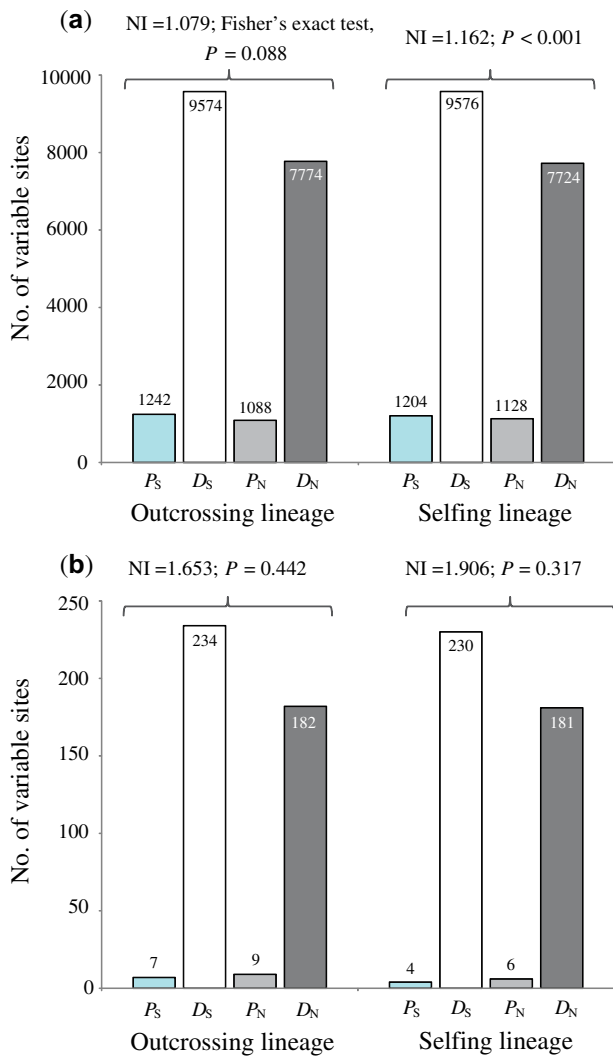
between the outcrossing and selfing lineage (fig. 5). In the outcrossing lineage, there were relatively fewer effectively neutral nonsynonymous mutations, with  $\sim 19\%$  (SE=0.023) falling into this category ( $0 < N_e s < 1$ ; fig. 5). In contrast, in the selfing lineage,  $\sim 23\%$  (SE=0.022) of new nonsynonymous mutations were classified as effectively neutral, and the difference between these proportions is significant (fig. 5). Furthermore, there was a significantly greater proportion of sites that were strongly selected against and very rarely fixed ( $N_e s > 100$ ) in the selfing compared with the outcrossing lineage (45% vs. 24%; fig. 5).

### Adaptive Evolution of Selfing and Outcrossing Lineage of *P. chungensis*

We used the method of Eyre-Walker and Keightley (2009) to estimate the proportion of adaptive nonsynonymous substitutions ( $\alpha$ ) during the divergence of the selfing and outcrossing lineage of *P. chungensis*, with *P. mallophylla* as the outgroup. Our estimate of the proportion of adaptive substitutions differed significantly between the selfing and outcrossing lineage of *P. chungensis* (table 2). The estimate of  $\alpha$  fixed by positive selection between the outcrossing lineage of *P. chungensis* and *P. mallophylla* was 25%, with the confidence interval (CI) not encompassing zero (95% CI: 5–44%). In contrast, no adaptive substitutions were found for the selfing lineage of *P. chungensis* since  $\alpha$  was not significantly different from zero (95% CI: −16% to 32%) (table 2). Splitting nonsynonymous substitutions into adaptive ( $\omega_a$ ) and nonadaptive proportions ( $\omega_d$ ), we obtained a lower  $\omega_a$  and a higher  $\omega_d$  in the selfing compared with the outcrossing lineage of *P. chungensis* (table 2).

### Discussion

Using both chloroplast (72 genes) and nuclear coding sequences (445 orthologs), we found significantly higher nonsynonymous versus synonymous divergence ratios ( $d_N/d_S$ ) in homostylous than distylous species (supplementary table S2, Supplementary Material online and fig. 3). This result likely reflects the fixation of deleterious mutations due to weaker selection over relatively long-time scales. Over shorter time scales, as evident in distylous and homostylous samples of *P. chungensis* (2,954 orthologs), weaker natural selection was also evident with a higher  $\pi_N/\pi_S$  ratio (table 1), a significantly

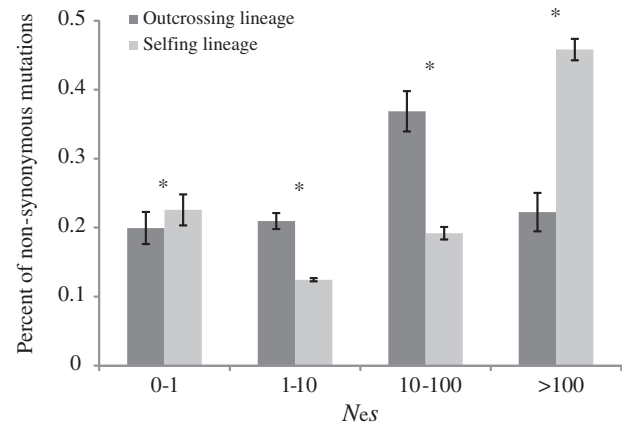


**Fig. 4.** The MK test separating the outcrossing and selfing lineages of *Primula chungensis* from outgroup (*P. mallophylla* and *P. secundiflora*) based on (a) nuclear and (b) chloroplast genome data. Synonymous polymorphism ( $P_S$ ) and divergence ( $D_S$ ) are indicated by blue and white bars, respectively; nonsynonymous polymorphism ( $P_N$ ) and divergence ( $D_N$ ) are indicated by gray and black bars, respectively; numbers above each bar indicate the total numbers of each type of site.

different DFE (elevated proportion in:  $0 < N_e s < 1$ ; fig. 5), and a lower rate of adaptive evolution ( $\alpha$ ,  $\omega_a$ ) in a selfing lineage than in an outcrossing lineage (table 2). These results are consistent with theoretical predictions of mating-system transitions from outcrossing to selfing at both macro- and microevolutionary time scales and are predicted features of the genomic selfing syndrome (Cutter 2019). Below, we compare our results with other recent genomic studies of selfing and outcrossing taxa and discuss their implications for the evolutionary fate of homostyles in *Primula*.

### Mating Systems and Selection Efficacy

Contrasting patterns of selection are expected between related selfing and outcrossing populations for genes under strong purifying selection as indicated by a higher ratio of



**Fig. 5.** Distribution of fitness effects (DFE) of new nonsynonymous mutations falling in different  $N_e s$  categories for outcrossing and selfing lineages of *Primula chungensis*.  $N_e s$  is the product of  $N_e$  and the selection coefficient ( $s$ ). Nine selfing and four outcrossing individuals were used to generate the DFEs. Error bars on top of each  $N_e s$  category are 95% confidence intervals from 200 bootstrap replicates generated by resampling over loci. Comparisons between the two lineages at  $P < 0.05$  significance level are indicated by \*.

**Table 2.** Parameters of the Fitness Distribution of Nonsynonymous Mutations ( $\beta$ , mean  $S$ ), Rate of Adaptive Evolution ( $\alpha$ ), Adaptive ( $\omega_a$ ), and Nonadaptive Proportion ( $\omega_d$ ) of  $d_N/d_S$  for the Selfing and Outcrossing Lineages of *Primula chungensis* Estimated with the Method of DFE (see Materials and Methods for details).

	Outcrossing Lineage	Selfing Lineage
$\beta$ (CI)	0.322 (0.243, 0.375)	0.195 (0.166, 0.226)
Mean $S$ (CI)	69.048 (56.698, 97.609)	732.82 (603.530, 915.048)
$\alpha$ (CI)	0.250 (0.058, 0.445)	0.163 (-0.167, 0.322)
$\omega_a$ (CI)	0.057 (0.012, 0.112)	0.038 (-0.035, 0.079)
$\omega_d$ (CI)	0.170 (0.139, 0.231)	0.198 (0.167, 0.247)

$\beta$ , the shape parameter of the gamma distribution assumed for the estimations of  $S$ .

nonsynonymous to synonymous polymorphisms ( $\pi_N/\pi_S$ ) and/or substitutions ( $d_N/d_S$ ). In *P. chungensis*, we detected a modest signature of elevated  $\pi_N/\pi_S$  at both nuclear (0.284 vs. 0.280) and chloroplast genes (0.385 vs. 0.368) in our comparison of a selfing versus outcrossing lineage, respectively, a pattern consistent with the results of our MK tests (fig. 4). The reduced efficacy of selection in selfing samples was also evident by a significant increase in the proportion of effectively neutral nonsynonymous mutations (24% vs. 20% in the category:  $0 < N_e s < 1$ ; fig. 5). This result is in accord with DFE data from another heterostylous species, *Eichhornia paniculata*, in which the floral polymorphism has broken down and the transition to selfing is accompanied by a larger fraction of effective neutral sites, although in this case, the difference was not significant (Arunkumar et al. 2015).

One important caveat of our implementation of the DFE analysis is that this approach is likely to be biased when used in highly selfing populations because LD will result in per-locus estimates that are nonindependent of one another. Such an effect will depend critically on the amount of selfing in populations. Our previous SSR marker-based estimates (Zhong et al. 2019) indicated that the selfing rate of

homostylous *Primula* populations averaged 0.8 ( $n = 15$  populations) but was strongly associated with the degree of herkogamy (and see de Vos et al. 2018). Although we do not currently have selfing rate estimates for *P. chungensis*, the focal species in our DFE study, homostylous populations have relatively large flowers and a variable degree of stigma-anther separation and thus are unlikely to be highly selfing (Zhou et al. 2017). Therefore, our comparative use of the DFE in comparing distylous and homostylous population in *P. chungensis* is probably qualitatively valid, even if the point estimates we obtained are unlikely to be accurate.

In accord with the patterns we obtained for polymorphism data, the  $d_N/d_S$  ratio was also higher in homostylous than distylous species (supplementary table S2, Supplementary Material online). This result contrasts with several earlier studies using divergence data that have found no evidence of relaxed selection in selfing species (Wright et al. 2002; Cutter et al. 2008; Haudry et al. 2008; Escobar et al. 2010). In the case of homostylous *Primula* species, the transition to high selfing rates is usually rapid through an abrupt rearrangement of sexual organs in one generation (fig. 1a) and in some cases is likely associated with reproductive isolation and speciation, as appears to also be the case in *Capsella* (Mattila et al. 2018). This may help to explain why a signature of divergence between selfing and outcrossing lineages was also evident in *P. chungensis* (supplementary table S2c, Supplementary Material online).

Another issue that needs to be considered in detecting divergence differences between selfing and outcrossing lineages is the number of loci that are investigated. Insufficient genomic sampling probably accounts for the earlier finding of no signal of elevated  $d_N/d_S$  in a comparison homostylous versus distylous *Primula* species as only sequences from the plastid genes *matK* and *rbcl* were used (Glémin and Muyle 2014). In our study, genomic sampling involved the whole chloroplast genome, with 72 genes included in the analysis of protein evolution, and  $\omega$  was significantly higher in homostylous than distylous branches ( $\omega_{\text{self}} > \omega_{\text{out}}$  in  $M_{\text{out-self}}$   $P < 0.001$ ; fig. 3a). When we analyzed chloroplast genes independently, we identified 11 genes that evolved differentially between outcrossing and selfing species (fig. 2a). Although the chloroplast genes we analyzed are unlikely to evolve independently because of the absence of recombination, they are likely to mutate at different rates and the evolutionary fate of mutations at different genes will exhibit some degree of independence. Interestingly, *matK* did not show a significantly higher  $\omega_{\text{self}}$  than  $\omega_{\text{out}}$  (0.462 vs. 0.457,  $\chi^2 = 0.003$ ,  $P = 0.959$ ). And *rbcl* exhibited the opposite pattern to that predicted in  $M_{\text{self-out}}$  with  $\omega_{\text{self}}$  lower than  $\omega_{\text{out}}$  (0.057 vs. 0.149) and this model was rejected when compared with  $M_0$  ( $\chi^2 = 4.848$ ,  $P = 0.028$ ). Sample size issues were also apparent in studies of *A. thaliana*, with Slotte et al. (2010) finding evidence for weak purifying selection using genome-wide data in contrast to an earlier study by Wright et al. (2002) which failed to detect a signature using a more limited number of loci. Given the stochasticity and slow rate of mutation accumulation in chloroplast genes, large numbers of loci are

probably required to consistently detect the influence of predominant selfing on plant genomes.

Although the genomic signatures we detected from polymorphism and divergence data are consistent with the expected effects of contrasting mating systems, the magnitude of the differences in  $\omega$  values between homostylous and distylous species were relatively small (supplementary table S2, Supplementary Material online). A plausible explanation for this result is that we detected few positively selected sites in homostylous species (supplementary table S3, Supplementary Material online) but more efficient positive selection was evident in outcrossing lineages. As a result, adaptive substitutions should strongly increase the  $d_N/d_S$  ratio, and this in turn may lower the difference in  $d_N/d_S$  between outcrossing and selfing lineages. This explanation is supported by our analysis of genome-wide adaptive evolution which found a higher proportion of adaptive nonsynonymous substitutions ( $\alpha$ ) and adaptive rates ( $\omega_{\alpha}$ ) in distylous than homostylous lineage (table 2). Slower rates of adaptive evolution accompanying shifts in reproductive systems have also been reported in *A. thaliana* (Slotte et al. 2010), the selfing snail *Galba truncatula* (Burgarella et al. 2015), and asexual species of *Oenothera* (Hersch-Green et al. 2012).

### Potential Implications for the Evolution of Homostyly in *Primula*

Our results provide insights on the evolutionary fate of homostylous *Primula* species, with potential implications for the evolution of selfing species in general. According to the “dead-end hypothesis” initially proposed by Stebbins (1957), selfing lineages are expected to arise frequently from outcrossing ancestors due to the transmission advantage of mating system modifiers and reproductive assurance under pollinator and/or mate limited conditions. But once selfers have originated they are more prone to extinction than outcrossers as a result of a variety of genetic and ecological factors (Busch and Delph 2012; Igic and Busch 2013; Wright et al. 2013; Barrett et al. 2014). Rapid transitions from outcrossing to selfing in *Primula* by single-gene (CYP734A50) mutations (Huu et al. 2016; Kappel et al. 2017) appear to be most commonly associated with speciation (cladogenic shifts) rather than species replacement (anagenic shifts) (see Goldberg and Igic 2012). However, over long-time scales, the frequent origin of homostyly in *Primula* (de Vos, Wüest, et al. 2014; Zhong et al. 2019) does not appear to contribute significantly to the overall species richness of the genus. Indeed, the “live fast and die young” nature of selfing homostylous lineages described for *Primula* helps to explain their transient nature over long-time scales (de Vos, Hughes, et al. 2014).

In most *Primula* species, enforced outcrossing, governed by heteromorphic self-incompatibility, may result in long-term genetic advantages by increasing opportunities for survival and adaptation. This may be of particular importance in the topographically complex mountains of southwest China where most species occur. In contrast, the negative genomic consequences of persistent selfing in homostyles, including reduced genetic diversity and the accumulation of deleterious mutations, may potentially limit long-term adaptive potential



(although see [Grant and Kalisz 2019](#)). Thus, the evolution of the genomic selfing syndrome could be one of the primary drivers of the extinction of selfing lineages ([Glémin and Ronfort 2013](#); [Cutter 2019](#)), despite the short-term advantages of selfing ([Busch and Delph 2012](#)) and ephemeral bursts of speciation ([de Vos, Hughes, et al. 2014](#)). Our empirical evidence of genomic degradation in homostyles of *Primula* is therefore generally consistent with Stebbin's dead-end theory for selfing lineages.

Not all homostyles in heterostylous species necessarily have a doomed long-term fate. Several features of the reproductive and genetic systems of some taxa may ameliorate the negative effects of selfing and delay extinction risk. The simplest mechanism involves the evolution of some degree of herkogamy in homostyles resulting in the evolution of mixed mating rather than predominant selfing ([Barrett and Shore 1987](#); [de Vos et al. 2018](#)). Investigations of species with mixed mating indicate that moderate levels of selfing are associated with more efficient selection and the absence of characteristic features of the genomic selfing syndrome ([Salcedo et al. 2014](#); [Laenen et al. 2018](#)). Although we observed some evidence for the genomic syndrome in homostylous populations of *P. chungensis*, this species is unusual in displaying considerable variation in herkogamy thus providing opportunities for mixed mating ([Zhou et al. 2017](#)). Significantly, *P. chungensis* has a much larger geographical range than other homostyles from SW China that were investigated in this study. Similarly, three other genera (*Dodecatheon*, *Cortusa*, and *Sredinskya*), inferred to be descended from the most recent common ancestor of *Primula* ([Mast et al. 2006](#)) which was distylous, are each monomorphic with stigmas positioned above anthers ([Mast and Reveal 2007](#)). Their possession of approach herkogamy may represent a floral strategy that limits the most harmful genomic consequences of selfing when it arises through the breakdown of distyly.

A second mechanism that may reduce the extinction risk for some homostylous lineages of *Primula* is polyploidy. The ability to self may facilitate the establishments of polyploid and genome duplication can ameliorate the negative effects of selfing and inbreeding depression, particularly in autopolyploids ([Rausch and Morgan 2005](#)). A close association between polyploidy and homostyly is evident in *Primula* sect. *Aleuritia* ([Guggisberg et al. 2006, 2009](#); [Theodoridis et al. 2013](#)). In this section, polyploid homostyles are more commonly found in previously glaciated areas of high latitude and appear to be more ecologically successful than their distylous ancestors ([Guggisberg et al. 2006](#)). However, there is no evidence that among the sections of *Primula* that are the basis of our study that polyploidy plays a significant role in the evolutionary fate of homostyles. Indeed, surveys of chromosome numbers indicate that the vast majority of the species we investigated are diploid, and the only known exception is *P. oreodoxa* which has both diploid and tetraploid homostyly populations ([Yuan, Barrett, et al. 2017](#)).

## Conclusions

Our study represents the most comprehensive analysis of sequence divergence and polymorphism data based on

both plastid and nuclear genes for related outcrossing and selfing plants. We found both macro- and microevolutionary evidence for components of the genomic selfing syndrome. The transition to selfing via homostyle evolution has resulted in significant effects on genetic diversity, the efficacy of selection, and on rates of adaptive evolution. Most studies on the genomic consequences of mating-system transition have focused on model systems by comparing related species pairs (e.g., *Arabidopsis* and *Capsella*, reviewed in [Mattila et al. 2018](#)) or populations (e.g., *Eichhornia paniculata*, [Arunkumar et al. 2015](#)) with contrasting mating systems. *Primula* provides a valuable opportunity to extend these approaches as it is composed of numerous independent transitions from outcrossing to selfing as a result of the repeated breakdown of distyly to homostyly. Future comparative investigations of transitions of different ages among *Primula* species should provide novel insights on diverse features of genome evolution and the extent to which the evolution of selfing increases extinction risk.

## Materials and Methods

### Plant Materials

To evaluate the potential influence of the evolutionary history of selfing on chloroplast sequence divergence, we obtained 56 samples each from 44 distylous and 12 homostylous species, and 22 samples from four distylous–homostylous species. We analyzed nuclear genome divergence by using the RNA-seq data of 12 samples each from nine distylous and three homostylous species. To investigate molecular divergence over shorter time scales, we focused on *P. chungensis*, a species with at least two independent origins of homostyly in the Tibet and Hengduan mountains region ([Zhou et al. 2017](#)). We sampled 26 individuals from 13 populations from the Hengduan mountains (Sichuan lineage); for population information see [Zhou et al. \(2017\)](#). Thirteen individuals were used for plastome assembly, with five and eight individuals each from distylous and homostylous populations, respectively. The remaining 13 individuals from four distylous and nine homostylous populations were investigated using RNA-seq. In total, we generated 51 new chloroplast genome assemblies and 18 transcriptomes with the chloroplast and RNA-seq data obtained from our previous phylogenomic study ([Zhong et al. 2019](#)) and Sequence Read Archives (see [Zhang et al. 2013](#); [Cocker et al. 2015](#); [Nowak et al. 2015](#); [Yuan, Zeng, et al. 2017](#)), respectively. Information on all samples are presented in [supplementary table S4, Supplementary Material](#) online.

### Chloroplast Genome Data Preparation for Molecular Evolutionary Analysis

We obtained chloroplast genomes of our samples ([supplementary table S1, Supplementary Material](#) online) from de novo assemblies following the procedures detailed below. First, we isolated total genomic DNA with a modified cetyl trimethyl ammonium bromide method ([Doyle 1991](#)) from silica-dried leaf tissue. Then we constructed libraries using a NEBNext Ultra II DNA Library Prep Kit for Illumina

(New England Biolabs, Ipswich, MA). Paired-end sequencing was carried out on a HiSeq X-Ten sequencer (Illumina, San Diego, CA) with  $2 \times 150$  bp read length. We obtained  $\sim 2$  Gb of sequence data for each sample. After quality control using NGS QC TOOL Kit (Patel and Jain 2012) with default parameters, we performed de novo assemblies of chloroplast genomes employing the GetOrganelle pipeline (Jin et al. 2018) for each accession. We subsequently annotated the circular plastomes following the PGA program (Qu et al. 2019) and all the annotated transfer RNA (tRNA) genes were further verified using the corresponding structures predicted by tRNAscan-Se v.1.21 (Schattner et al. 2005). We retrieved five annotated *Primula* plastid genomes from NCBI database to complement our data set. Finally, we extracted 72 chloroplast coding genes from each annotated plastome for further molecular evolutionary analysis.

### Transcriptome Data Preparation for Molecular Evolutionary Analysis

We isolated total RNA of each sample using an RNAPrep Pure Plant Kit (TIANGEN Biotechnologies Corporation, Beijing, China). We prepared the cDNA library for transcriptome sequencing using a cDNA Synthesis Kit (Illumina, San Diego, CA) following the manufacturer's recommendations, and the 18 cDNA libraries were then sequenced on the Illumina HiSeq 4000 platform (Illumina, San Diego, CA) to obtain short sequences of 150 bp from both ends of each cDNA.

We performed preliminary quality check of raw reads within each data set with FastQC v.0.11.2. The clean reads were filtered from the raw reads with Trimmomatic v.0.32 (Bolger et al. 2014) after trimming adapters and removing both ambiguous ( $N > 10\%$ ) and low-quality reads (Phred score  $< 30$ ). We generated de novo assemblies of transcriptomes of individuals for each species using the Trinity platform with default parameters (Haas et al. 2013). We discarded contigs with length  $< 300$  bp due to their low annotation rate. To remove the redundant sets of sequences, similar sequence stretches were filtered and merged with the TGI Clustering tool (Perteza et al. 2003). We selected the longest transcript as representative for each cluster. We characterized the quality of assemblies, by examining the RNA-seq read representation by aligning reads to the assemblies using bowtie2 (Langmead and Salzberg 2012) and counted the overall alignment rate and number of proper pairs. We performed an assessment of the representation of single-copy conserved plant orthologs using the program BUSCO (Simão et al. 2015) to evaluate whether the assembly was full-length or nearly full-length. We calculated the ExN50 statistic against a fraction of the most highly expressed transcripts (Ex), and retained transcripts with an expression level higher than the saturation point for subsequent analysis.

We predicted open reading frames (ORFs) from the assembled transcripts using the program TransDecoder and scanned the predicted ORFs for homology to known proteins in two ways: a BLAST search (Camacho et al. 2009) against Swiss-Prot database, and by identifying hidden Markov model protein domains with HMMER v.3.1b (Eddy 2011) against the HMMER/Pfam protein database. Those coding peptides with

BLAST hits ( $e\text{-value} < 1e-5$ ) or domain hits (FullSeqScore and FullDomainScore  $> 20$ ) were retained in the set of reported likely coding regions. We integrated the final coding region predictions BLAST and Pfam search results using TransDecoder. We selected the top scores for predicted ORFs for each transcript in situations where there was more than one prediction within a transcript.

Finally, we used OrthoFinder v.2.3 (Emms and Kelly 2019) to identify 445 single-copy orthologs among annotated ORFs data set in the 12 *Primula* species for interspecific divergence analysis, and 2,954 orthologs from the 13 individuals of *P. chungensis* used for intraspecific divergence and polymorphism analysis. We aligned the protein-coding sequences corresponding to orthologous genes with ParaAT (Zhang et al. 2012). We excluded misalignments manually based on the sequence alignment for each ortholog set.

### Molecular Divergence between Distylous and Homostylous Species

To investigate the association between mating-system differentiation and the accumulation of deleterious mutations, we estimated the ratio of nonsynonymous to synonymous divergence ( $d_N/d_S = \omega$ ) among distylous and homostylous species. We used the *codeml* program, included in PAML package v.4.4 (Yang 2007), to test the fit of several hierarchical models using both plastid and nuclear sequences. We evaluated the fit of the following maximum likelihood branch models to the data: 1) the null one-ratio model ( $M_0$ ), which assumes the same  $\omega$  value for all branches; 2) a two-ratio model, in which two separate  $\omega$  values were estimated for distylous and homostylous species ( $M_{\text{out-self}}$ ; fig. 1b); 3) a second two-ratio model contrasting internal versus external branches ( $M_{\text{int-ext}}$ ; fig. 1c); 4) a three-ratio model ( $M_3$ ; fig. 1d), where internal and external distylous and external homostylous branches were allowed to have different  $\omega$ . We applied different models to both individual gene alignments and to all genes from concatenated data sets. We used the branch-site model to test if there was any evidence for signatures of positive selection of loci with significantly elevated  $\omega$  values in homostyles. We then compared the fit of models by using likelihood ratio tests with the appropriate degrees of freedom (df).

For the chloroplast data set, we estimated codon evolution based on the phylogenetic tree reconstructed by RAxML v.8 (Stamatakis 2006) with 1,000 bootstrap under the GTRGAMMA model (supplementary fig. S1, Supplementary Material online), and the outgroup *Androsace bulleyana* (accession number: KU513438) was excluded from the tree in the PAML analyses. For the transcriptome data set, we examined the inter- and intraspecific analyses separately using the inferred phylogenetic tree specific to the corresponding inter- and intraspecific orthologs.

### Within-Species Polymorphism in Distylous and Homostylous Lineages of *P. chungensis*

To test the hypothesis that homostylous populations of *P. chungensis* have lower genetic diversity and a higher ratio of nonsynonymous versus synonymous nucleotide diversity ( $\pi_N/\pi_S$ ) compared with distylous populations, we estimated

the population statistics of nucleotide diversity ( $\pi$ ), Watterson's theta ( $\theta_w$ ), and Tajima's  $D$  for both chloroplast and nuclear sequences. We then used a contingency table-based variation of the MK test (McDonald and Kreitman 1991) and Fisher's exact test to evaluate whether the ratio of nonsynonymous to synonymous polymorphisms ( $P_N/P_S$ ) of *P. chungensis* was differentiated from the ratio of nonsynonymous to synonymous divergence ( $D_N/D_S$ ) of the outgroup species (*P. secundiflora* and *P. mallophylla*, both distylous). We summarized the MK results using the NI, defined as  $(P_N/D_N)/(P_S/D_S)$  (Rand and Kann 1996).

We also performed tests of selection on distylous and homostylous population samples of *P. chungensis* by estimating the genome-wide DFE of new nonsynonymous mutations, the rate of adaptive evolution ( $\alpha$ ), and the rate of adaptive [ $\omega_a = \alpha (d_N/d_S)$ ] and nonadaptive [ $\omega_d = (1-\alpha) (d_N/d_S)$ ] nonsynonymous mutations using DFE- $\alpha$  v.2.15 (Eyre-Walker and Keightley 2009). For these analyses, synonymous sites were assumed to evolve neutrally and we obtained the estimates of DFE and  $\alpha$  for both selfing and outcrossing lineages. We generated the locus-specific nonsynonymous- and synonymous-folded SFS and number of invariant sites using a Perl script, Polymorphorama (Bachtrog and Andolfatto 2006) and we report the proportion of mutations falling into a given  $N_e s$  range ( $<1$ ,  $1-10$ ,  $10-100$ , and  $>100$ ) of the DFE. We estimated the rate of adaptive evolution ( $\alpha$ ) by comparing the selfing and outcrossing lineages of *P. chungensis* with the outgroup *P. mallophylla*. We derived CIs and SEs from 200 bootstrap replicates by randomly resampling across loci using R (R Development Core Team 2014).

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

We thank Asher Cutter, Ramesh Arunkumar, and Stephen Wright for helpful comments on the article. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000), the National Natural Science Foundation of China (31971394, 31770417, and 31570384), the Yunling Scholarship of Yunnan Province (YLXL20170001), Light of West China Program of the Chinese Academy of Sciences, and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

## Author Contributions

W.Z., S.C.H.B., H.W., and D.-Z.L. planned and designed the research. X.-J.W., W.Z., L.Z., and Z.-K.W. performed experiments and conducted field work. X.-J.W. and W.Z. analyzed the data. W.Z., X.-J.W., and S.C.H.B. wrote the article.

## References

Akashi H, Osada N, Ohta T. 2012. Weak selection and protein evolution. *Genetics* 192(1):15–31.

- Arunkumar R, Ness RW, Wright SI, Barrett SCH. 2015. The evolution of selfing is accompanied by reduced efficacy of selection and purging of deleterious mutations. *Genetics* 199(3):817–829.
- Bachtrog D, Andolfatto P. 2006. Selection, recombination and demographic history in *Drosophila miranda*. *Genetics* 174(4):2045–2059.
- Baker HG. 1955. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* 9(3):347–349.
- Barrett SCH. 1992. Evolution and function of heterostyly. Berlin: Springer-Verlag.
- Barrett SCH. 2019. 'A most complex marriage arrangement': recent advances on heterostyly and unresolved questions. *New Phytol.* 224(3):1051–1067.
- Barrett SCH, Arunkumar R, Wright SI. 2014. The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philos Trans R Soc B.* 369(1648):20130344.
- Barrett SCH, Harder LD, Worley AC. 1996. The comparative biology of pollination and mating in flowering plants. *Philos Trans R Soc Lond B Biol Sci.* 351(1345):1271–1280.
- Barrett SCH, Shore JS. 1987. Variation and evolution of breeding systems in the *Turnera ulmifolia* L. complex (Turneraceae). *Evolution* 41(2):340–354.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Burgarella C, Gayral P, Ballenghien M, Bernard A, David P, Correa A, Hurtrez-Boussés S, Escobar J, Galtier N, et al. 2015. Molecular evolution of freshwater snails with contrasting mating systems. *Mol Biol Evol.* 32(9):2403–2416.
- Busch JW, Delph LF. 2012. The relative importance of reproductive assurance and automatic selection as hypotheses for the evolution of self-fertilization. *Ann Bot.* 109(3):553–562.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421.
- Charlesworth B, Charlesworth D. 1979. The maintenance and breakdown of distyly. *Am Nat.* 114(4):499–513.
- Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–1303.
- Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. *Curr Opin Genet Dev.* 11(6):685–690.
- Chen J, Glémin S, Lascoux M. 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol Biol Evol.* 34(6):1417–1428.
- Cocker JM, Webster MA, Li JH, Wright J, Kaithakottil G, Swarbreck D, Gilmartin PM. 2015. *Oakleaf*: an S locus-linked mutation of *Primula vulgaris* that affects leaf and flower development. *New Phytol.* 208(1):149–161.
- Crosby JL. 1949. Selection of an unfavourable gene-complex. *Evolution* 3(3):212–230.
- Cutter AD. 2019. Reproductive transitions in plants and animals: selfing syndrome, sexual selection, and speciation. *New Phytol.* 224(3):1080–1094.
- Cutter AD, Wasmuth JD, Washington NL. 2008. Patterns of molecular evolution in *Caenorhabditis* preclude ancient origins of selfing. *Genetics* 178(4):2093–2104.
- Darwin C. 1877. The different forms of flowers on plants of the same species. London: John Murray.
- de Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E. 2014. Heterostyly accelerates diversification via reduced extinction in primroses. *Proc R Soc B.* 281(1784):20140075.
- de Vos JM, Keller B, Zhang LR, Nowak MD, Conti E. 2018. Mixed mating in homostylous species: genetic and experimental evidence from an alpine plant with variable herkogamy, *Primula halleri*. *Int J Plant Sci.* 179(2):87–99.
- de Vos JM, Wüest RO, Conti E. 2014. Small and ugly? Phylogenetic analyses of the "selfing syndrome" reveal complex evolutionary fates of monomorphic primrose flowers. *Evolution* 68(4):1042–1057.

- Doyle J. 1991. DNA protocols for plants: CTAB total DNA isolation. In: Hewitt GM, Johnston A, editors. *Molecular techniques in taxonomy*. Berlin: Springer-Verlag. p. 283–293.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. 7(10):e1002195.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20(1):238.
- Escobar JS, Cenci A, Bolognini J, Haudry A, Laurent S, David J, Glémin S. 2010. An integrative test of the dead-end hypothesis of selfing evolution in *Triticeae* (Poaceae). *Evolution* 64(10):2855–2872.
- Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol*. 26(9):2097–2108.
- Flint-Garcia SA, Thornsberry JM, Buckler IVES. 2003. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol*. 54(1):357–374.
- Ganders FR. 1979. The biology of heterostyly. *N Z J Bot*. 17(4):607–635.
- Ganders FR, Denny SK, Tsai D. 1985. Breeding systems and genetic variation in *Amsinckia spectabilis* (Boraginaceae). *Can J Bot*. 63(3):533–538.
- Gilmartin PM. 2015. On the origins of observations of heterostyly in *Primula*. *New Phytol*. 208(1):39–51.
- Glémin J, Galtier N. 2012. Genomic evolution in outcrossing versus selfing versus asexual species. In: Anisimova M, editor. *Evolutionary genomics. Methods in molecular biology*. New York: Humana. p. 311–335.
- Glémin J, Muyle A. 2014. Mating systems and selection efficacy: a test using chloroplastic sequence data in Angiosperms. *J Evol Biol*. 27(7):1386–1399.
- Glémin J, Ronfort J. 2013. Adaptation and maladaptation in selfing and outcrossing species: new mutations versus standing variation. *Evolution* 67(1):225–240.
- Goldberg EE, Igic B. 2012. Tempo and mode in plant breeding system evolution. *Evolution* 66(12):3701–3709.
- Goodwillie C, Kalisz S, Eckert CG. 2005. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical expectations, and empirical evidence. *Annu Rev Ecol Syst*. 36(1):47–79.
- Grant AG, Kalisz S. 2019. Do selfing species have greater niche breadth? Support from ecological niche modeling. *Evolution* 74(1):73–88.
- Guggisberg A, Mansion G, Conti E. 2009. Disentangling reticulate evolution in an arctic-alpine polyploid complex. *Syst Biol*. 58(1):55–73.
- Guggisberg A, Mansion G, Kelso S, Conti E. 2006. Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid-polyploid species complex of *Primula*. *New Phytol*. 171(3):617–632.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8(8):1494–1512.
- Haudry A, Cenci A, Guilhaumon C, Paux E, Poirier S, Santoni S, David D, Glémin S. 2008. Mating system and recombination affect molecular evolution in four *Triticeae* species. *Genet Res*. 90(1):97–109.
- Hazzouri KM, Escobar J, Ness RW, Newman LK, Randle A, Kalisz S, Wright SI. 2012. Comparative population genomics in *Collinsia* sister species reveals evidence for reduced effective population size, relaxed selection and evolution of biased gene conversion with an ongoing mating system shift. *Evolution* 67(5):1263–1278.
- Heller R, Smith JM. 1978. Does Muller's ratchet work with selfing? *Genet Res*. 32(3):289–293.
- Hersch-Green EI, Myburg H, Johnson MT. 2012. Adaptive molecular evolution of a defence gene in sexual but not functionally asexual evening primroses. *J Evol Biol*. 25(8):1576–1586.
- 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 *CYP734A50* underlies the style length dimorphism in primroses. *eLife* 5:e17956.
- Igic B, Bohs L, Kohn JR. 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proc Natl Acad Sci U S A*. 103(5):1359–1363.
- Igic B, Busch JW. 2013. Is self-fertilization an evolutionary dead end? *New Phytol*. 198(2):386–397.
- Jain SK. 1976. The evolution of inbreeding in plants. *Annu Rev Ecol Syst*. 7(1):469–495.
- Jarne P, Charlesworth D. 1993. The evolution of the selfing rate in functionally hermaphroditic plants and animals. *Annu Rev Ecol Syst*. 24(1):441–466.
- Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, Li DZ. 2018. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *BioRxiv*: 256479. doi: <https://doi.org/10.1101/256479>.
- Kamran-Disfani A, Agrawal AF. 2014. Selfing adaptation and background selection in finite populations. *J Evol Biol*. 27(7):1360–1371.
- Kappel C, Huu CN, Lenhard M. 2017. A short story gets longer: recent insights into the molecular basis of heterostyly. *J Exp Bot*. 68(21-22):5719–5730.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.
- Laenen B, Tedder A, Nowak MD, Toräng P, Wunder J, Wötzel S, Steige KA, Kourmpetis Y, Odong T, Drouzas AD, et al. 2018. Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*. *Proc Natl Acad Sci U S A*. 115(4):816–821.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359.
- Lloyd DG. 1980. Demographic factors and mating patterns in angiosperms. In: Solbrig OT, editor. *Demography and evolution in plant populations*. Oxford: Blackwell. p. 67–88.
- Mast AR, Conti E. 2006. The primrose path to heterostyly. *New Phytol*. 171(3):439–442.
- Mast AR, Kelso S, Conti E. 2006. Are any primroses (*Primula*) primitively monomorphic? *New Phytol*. 171(3):605–616.
- Mast AR, Reveal JL. 2007. Transfer of *Dodecatheon* to *Primula* (Primulaceae). *Brittonia*. 59(1):79–82.
- Mattila TM, Laenen B, Slotte T. 2018. Population genomics of transitions to selfing in Brassicaceae model systems. In: Dutheil JY, editor. *Statistical population genomics*. New York: Springer. p. 269–288.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351(6328):652–654.
- Ness RW, Siol M, Barrett SCH. 2012. Genomic consequences of transitions from cross- to self-fertilization on the efficacy of selection in three independently derived selfing plants. *BMC Genomics* 13(1):611.
- Nordborg M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154(2):923–929.
- Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E. 2015. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biol*. 16(1):12.
- Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst*. 23(1):263–286.
- Pannell JR, Fields PD. 2014. Evolution in subdivided plant populations: concepts, recent advances and future directions. *New Phytol*. 201(2):417–432.
- Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7(2):e30619.
- Pertege G, Huang XQ, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee YD, White J, Cheung F, Parvizi B, et al. 2003. TIGR Gene Indices Clustering Tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19(5):651–652.
- Piper JG, Charlesworth B, Charlesworth D. 1984. A high rate of self-fertilization and increased seed fertility of homostyle primroses. *Nature* 310(5972):50–51.
- Pollak E. 1987. On the theory of partially inbreeding finite population. I. Partial selfing. *Genetics* 117(2):353–360.
- Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D. 2011. Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. *Genome Biol Evol*. 3(1):868–880.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 15:50.
- R Development Core Team. 2014. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: <http://www.R-project.org/>.

- Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol.* 13(6):735–748.
- Rausch JH, Morgan MT. 2005. The effect of self-fertilization, inbreeding depression, and population size on autopolyploid establishment. *Evolution* 59(9):1867–1875.
- Roze D. 2016. Background selection in partially selfing populations. *Genetics* 203(2):937–957.
- Salcedo A, Kalisz S, Wright SI. 2014. Limited genomic consequences of mixed mating in the recently derived sister species pair, *Collinsia concolor* and *Collinsia parryi*. *J Evol Biol.* 27(7):1400–1412.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33(Web Server):W686–W689.
- Schoen DJ, Johnston MO, L'Heureux A-M, Marsolais JV. 1997. Evolutionary history of the mating system in *Amsinckia* (Boraginaceae). *Evolution* 51(4):1090–1099.
- Shao JW, Wang HF, Fang SP, Conti E, Chen YJ, Zhu HM. 2019. Intraspecific variation of self-incompatibility in the distylous plant *Primula merrilliana*. *AoB Plants* 11(3):plz030.
- Sicard A, Lenhard M. 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Ann Bot.* 107(9):1433–1443.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Mol Biol Evol.* 27(8):1813–1821.
- Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet.* 45(7):831–835.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
- Stebbins GL. 1957. Self-fertilization and population variability in the higher plants. *Am Nat.* 91(861):337–354.
- Stebbins GL. 1974. Flowering plants: evolution above the species level. Cambridge: Belknap Press.
- Takebayashi N, Morrell PL. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *Am J Bot.* 88(7):1143–1150.
- Tang C, Toomajian C, Sherman-Broyles S, Plagnol V, Guo YL, Hu TT, Clark RM, Nasrallah JB, Weigel D, Nordborg M. 2007. The evolution of selfing in *Arabidopsis thaliana*. *Science* 317(5841):1070–1072.
- Theodoridis S, Randin C, Broennimann O, Patsiou T, Conti E. 2013. Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia*. *J Biogeogr.* 40(7):1278–1289.
- Truyens S, Arbo MM, Shore JS. 2005. Phylogenetic relationships, chromosome and breeding system evolution in *Turnera* (Turneraceae): inferences from ITS sequence data. *Am J Bot.* 92(10):1749–1758.
- Wang J, Hill WG, Charlesworth D, Charlesworth B. 1999. Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genet Res.* 74(2):165–178.
- Weller SG. 1992. Evolutionary modifications of tristylous breeding systems. In: Barrett SCH, editor. Evolution and function of heterostyly. Berlin: Springer-Verlag. p. 247–272.
- Wright SI, Kalisz S, Slotte T. 2013. Evolutionary consequences of self-fertilization in plants. *Proc Biol Sci.* 280(1760):20130133.
- Wright SI, Lauga B, Charlesworth D. 2002. Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol Biol Evol.* 19(9):1407–1420.
- Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. *Int J Plant Sci.* 169(1):105–118.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yuan S, Barrett SCH, Duan TT, Qian X, Shi MM, Zhang DX. 2017. Ecological correlates and genetic consequences of evolutionary transitions from distyly to homostyly. *Ann Bot.* 120(5):775–789.
- Yuan S, Zeng G, Shi MM, Zhang DX. 2017. Development of EST-SSR markers for *Primula ovalifolia* (Primulaceae) by transcriptome sequencing. *Appl Plant Sci.* 5(12):1700100.
- Zhang L, Yan HF, Wu W, Yu H, Ge XJ. 2013. Comparative transcriptome analysis and marker development of two closely related Primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genomics* 14(1):329.
- Zhang Z, Xiao JF, Wu JY, Zhang HY, Liu GM, Wang XM, Dai L. 2012. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem Biophys Res Commun.* 419(4):779–781.
- Zhong L, Barrett SCH, Wang XJ, Wu ZK, Sun HY, Li DZ, Wang H, Zhou W. 2019. Phylogenomic analysis reveals multiple evolutionary origins of selfing from outcrossing in a lineage of heterostylous plants. *New Phytol.* 224(3):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 Phytol.* 214(3):1368–1380.