Journal of Experimental Botany, Vol. 68, No. 21–22 pp. 5719–5730, 2017
doi:10.1093/jxb/erx387 Advance Access publication 1 November 2017

FLOWERING NEWSLETTER REVIEW

A short story gets longer: recent insights into the molecular basis of heterostyly

Christian Kappel, Cuong Nguyen Huu and Michael Lenhard*

Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24–25, House 26, D-14476 Potsdam-Golm, Germany

* Correspondence: michael.lenhard@uni-potsdam.de

Received 1 September 2017; Editorial decision 4 October 2017; Accepted 9 October 2017

Editor: Frank Wellmer, Trinity College Dublin, Ireland

Abstract

Heterostyly is a fascinating adaptation to promote outbreeding and a classical paradigm of botany. In the most common type of heterostyly, plants either form flowers with long styles and short stamens, or short styles and long stamens. This reciprocal organ positioning reduces pollen wastage and promotes cross-pollination, thus increasing male fitness. In addition, in many heterostylous species selfing and the generation of unfit progeny due to inbreeding depression is limited by a self-incompatibility system, thus promoting female fitness. The two floral forms are genetically determined by the *S* locus as a complex supergene, namely a chromosomal region containing several individual genes that control the different traits, such as style or stamen length, and are held together by very tight linkage due to suppressed recombination. Recent molecular-genetic studies in several systems, including *Turnera*, *Fagopyrum*, *Linum*, and *Primula* have begun to identify and characterize the causal heterostyly genes residing at the *S* locus. An emerging theme from several families is that the dominant *S* haplotype represents a hemizygous region not present on the recessive *s* haplotype. This provides an explanation for the suppressed recombination and suggests a scenario for the chromosomal evolution of the *S* locus. In this review, we discuss the results from recent molecular-genetic analyses in light of the classical models on the genetics and evolution of heterostyly.

Key words: *CYP734A50*, distyly, *GLOBOSA2*, hemizygosity, heterostyly, *Primula*, *S* locus, supergene, tristyly.

Introduction

Heterostyly is a classical paradigm in plant genetics, evolution, and ecology (Barrett, 1992). It has fascinated plant scientists for centuries (Gilmartin, 2015), with Charles Darwin laying the basis for its functional dissection with his book on 'The different forms of flowers on plants of the same species' in 1877 (Darwin, 1877). In heterostylous species, the individuals fall into two (distyly) or three (tristyly) distinct, genetically determined categories or morphs based on their flower morphology. In distylous species, L-morph plants, also called pins, form flowers with long styles and short stamens, while S-morph individuals, also called thrums, form flowers with short styles and long stamens (Fig. 1). As heterostylous species often possess tubular flowers with stamen filaments fused to the floral tube (Ganders, 1979), this leads to anthers attached to the petal tube in a low position in L-morph flowers and towards the opening of the tube in S-morph flowers. The two morphs thus show reciprocal herkogamy, that is the parts of the sexual organs involved in pollination, anthers and stigma, are spatially separated within a flower, yet the position of male and female organs matches up between flowers of the two morphs. In tristylous species, stigma and anthers can be located at three positions in the flower – low,

Abbreviations: BRs, brassinosteroids; CCM, C-terminal domain; ELF3, EARLY FLOWERING 3; QTL, quantitative trait loci; RAD, restriction-associated DNA.

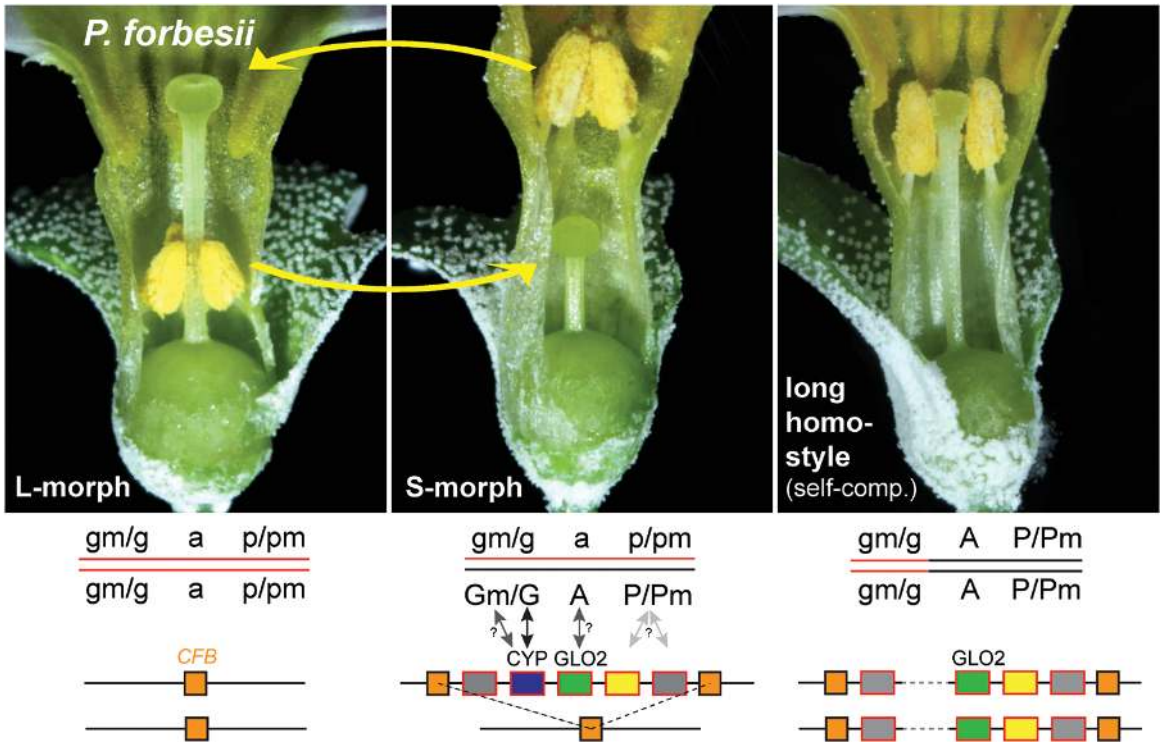

Fig. 1. Heterostyly in *Primula* and its molecular and genetic basis. (Top) Flowers of L-morph plants (left) and S-morph plants (middle) of *Primula forbesii* are shown, demonstrating the reciprocal herkogamy and spatial matching between the anthers of L-morph flowers and stigma of S-morph flowers and vice versa. Compatible crosses are indicated by the yellow arrows. Right image shows the flower of a self-compatible long homostyle, with anthers and stigma at the same high position. (Middle) The classical diallelic supergene model for the *S* locus is indicated, with the five proposed individual causal genes for style length (*G*), female incompatibility (*Gm*), anther position (*A*), pollen size (*P*) and male incompatibility (*Pm*). Uppercase letters indicate dominant alleles responsible for S-morph trait expression; lowercase letters indicate recessive alleles. (Bottom) Schematic representation of the molecular structure of the *S* locus. Genes are indicated by boxes. The recessive haplotype (bottom) only carries a single copy of the *CFB* gene, while the dominant haplotype (top) carries an approximately 280 kb additional segment containing five predicted genes (red frames). The correspondence between these genes and the genetically defined factors is indicated by double-headed arrows, with the darkness of the arrows indicating the strength of support for the equivalence.

middle and high – and flowers have two sets of stamens of different heights from each other and from the stigma position. In L-morph plants, a long style is combined with anthers at low and medium heights; in M-morph flowers, the stigma is at the middle position, with both high and low anthers; and in S-morph flowers, the style is short, while the anthers are at the medium and high positions (Barrett, 1992). In addition to this reciprocal herkogamy, the morphs are often also distinguished by a suite of ancillary features, such as differences in pollen size and exine structure and different lengths of stigmatic papillae (Dulberger, 1992). Besides these anatomical and morphological features, many, but by no means all heterostylous species exhibit a physiological incompatibility system. This heteromorphic self-incompatibility, also called intra-morph incompatibility, is manifested in low fertilization success and seed set after intra-individual and intra-morph pollinations, effectively limiting self-fertilization (Wedderburn and Richards, 1990).

Heterostyly is evolutionarily very widespread amongst flowering plants and is found in at least 28 families (Ganders, 1979; Barrett, 2002). It is thought to have arisen more than 20 times independently (Lloyd and Webb, 1992*a*), and even within families, such as the Boraginaceae, multiple independent origins have been inferred (Cohen, 2011). By contrast, in other families, such as the Primulaceae, a single origin of heterostyly appears more likely (Mast *et al.*, 2006; de Vos *et al.*, 2014).

Regarding its function, heterostyly is a complex adaptation to simultaneously promote both male and female fitness (Barrett and Shore, 2008). Self-incompatibility prevents or at least strongly reduces self-fertilization; it thus limits the number of less fit offspring due to inbreeding depression and increases female fitness of the plants. By contrast, the morphological aspect of heterostyly, namely reciprocal herkogamy, increases male fitness by promoting inter-morph pollen transfer over selfing (Ganders, 1979; Lloyd and Webb, 1992*b*; Stone, 1995; Keller *et al.*, 2014). This reduces sexual interference and pollen wastage, that is the unproductive deposition of pollen on incompatible self stigmas (Barrett and Shore, 2008), or in cases without self-incompatibility, on self stigmas leading to the production of less fit inbred offspring. It was particularly this latter point of promoting cross-pollination that Darwin had proposed as the selective advantage of heterostyly (Darwin, 1877), and a number of studies in both field and garden settings has confirmed this so-called 'cross promotion hypothesis' (Ganders, 1979; Lloyd and Webb, 1992*b*; Stone, 1995; Keller *et al.*, 2014).

The developmental, ecological and phylogenetic aspects of heterostyly have been excellently and comprehensively reviewed elsewhere (Barrett, 1992; Barrett and Shore, 2008; Cohen, 2010). Given recent progress in understanding the molecular-genetic basis of heterostyly, the following discussion will largely focus on its molecular and genetic control.

Classical genetics of heterostyly

The genetic basis of heterostyly has been studied for over a century (Bateson and Gregory, 1905), both because of the position of heterostyly as a prime example of convergent evolution in plants and because of the supergene architecture of the responsible locus (Schwander *et al.*, 2014) that became apparent early on. Many of these studies have used *Primula* species, but examples from several other families have also been investigated.

Soon after the rediscovery of Mendel's laws, Bateson and Gregory established that heterostyly in *Primula* was controlled by a single locus, termed *S*, of which there is a dominant *S* and a recessive *s* allele (Bateson and Gregory, 1905). S-morph individuals are *S*/*s* heterozygotes, while L-morph plants are homozygous for the recessive allele *s*/*s*. Plants homozygous for the dominant *S* allele are virtually never found in *Primula*, such that each compatible cross represents an *S*/*s* x *s*/*s* backcross, resulting in the expectation of equal proportions of L- and S-morph individuals in the progeny, that is isoplethy. A major breakthrough in our understanding of the genetics of heterostyly came from the work of Alfred Ernst who described the existence of plants with unusual trait combinations, such as long styles and high anthers in long homostyles, or short styles and low anthers in short homostyles (Ernst, 1928; Ernst, 1936; Ernst, 1955). In addition, pollen size was found to be separable from the anther positions, with unusual combinations of high anthers, but small pollen grains and vice versa. The pollen (male) and style/stigma (female) incompatibility phenotypes always followed pollen size and style length, respectively. Based on the inheritance data, the novel morphs were found to be caused by novel alleles at the *S* locus, leading to the suggestion that the *S* locus in fact represents at least three distinct, yet very tightly linked genes controlling the different features of heterostyly (Ernst, 1936). According to this notion, the *G* locus determines style length and female incompatibility behaviour, with the dominant allele causing short styles; the *A* locus determines anther position, with the dominant allele leading to high anthers; and the *P* locus controls pollen size and male compatibility. The *S* locus would thus be a supergene, namely a chromosomal segment containing very tightly linked individual genes that together control an integrated phenotype (Schwander *et al.*, 2014; Charlesworth, 2016). The S-morph then has the genotype *GPA*/*gpa*, and the L-morph *gpa*/*gpa* (Fig. 1). The novel morphs were interpreted by Ernst to result from mutations of one or two of the three genes, such that long homostyles would be *gPA*/*gpa* for example. Pamela Dowrick expanded on these findings by confirming that homostyly in tetraploid *Primula obconica* is governed by a novel allele at the *S* locus (Dowrick, 1956). In addition, her own results and a re-examination of data from Ernst and others led her to conclude that homostyles were more likely to have arisen by recombination within the *S* locus supergene than by mutation as envisaged by Ernst and that the order of the loci was likely *GPA* (Dowrick, 1956). The well-documented observation that long homostyles were observed much more frequently than short homostyles, even though the relevant haplotypes *gPA* and *Gpa* should arise in equal proportions from recombination within the *S* locus, was ascribed to differential fitness of the two forms (Charlesworth and Charlesworth, 1979).

These views about the *S* locus supergene were reinforced by Lewis and Jones, based also on the occurrence of homostyles at high frequencies that appeared incompatible with their origin by novel mutations (Lewis and Jones, 1992). Their study summarized the widely accepted model of the heterostyly *S* locus (Lewis and Jones, 1992; Richards, 2003). The main features of this model were that the *S* locus is a diallelic supergene consisting of at least three separable loci, *G*, *P*, and *A*; all dominant alleles are linked on one *S* haplotype, all recessive alleles on the *s* haplotype; recombination within the supergene is strongly suppressed; homostyles and other abnormal trait combinations result from rare events of recombination in the *S* locus; and the *S* locus is located close to a centromere based on the frequency of double reduction in tetraploids. Additional genes responsible for female and male incompatibility behaviour separate from *G* and *P*, respectively, have been proposed on the basis of the physiological arguments that a single locus was unlikely to control both the morphological and incompatibility phenotypes (Lewis and Jones, 1992; Richards, 2003). For the male side, this conclusion was supported by an unusual novel type observed by Kurian and Richards, in which pollen size was separated from male incompatibility behaviour (Kurian and Richards, 1997), yet on the female side, to our knowledge, style length and incompatibility have never been separated due to mutation/recombination at the *S* locus. This supergene model raised several questions regarding, for example, the mechanism of recombination suppression, the differences between the dominant and recessive haplotypes, and of course the identity and function of the component individual loci.

As outlined above, the supergene model of the *S* locus was mainly established on the basis of work in *Primula* species. How does this translate to independently evolved examples of heterostyly? In the additional systems that have been studied, heterostyly, or at least the reciprocal herkogamy, also seems to be controlled by a single diallelic locus (Lewis and Jones, 1992). As in *Primula*, the S-morph is determined by the dominant haplotype in six out of eight studied genera with distyly. Evidence for this locus being a supergene is available in several systems. For example, in *Turnera subulata* a long homostyle was generated by an x-ray induced deletion at the *S* locus, indicating the presence of at least two separate genes for style and stamen length, respectively (Labonne *et al.*, 2010). Similarly, long homostyly in *Turnera ulmifolia* is due to a novel allele at the *S* locus that causes both long styles and long stamens (Shore and Barrett, 1985). Also in the genus *Fagopyrum* (buckwheat), homostyly in naturally occurring species appears to be due to an altered *S* locus haplotype,

supporting its status as a supergene (Matsui *et al.*, 2003). As for the more complex case of tristyly, the three different morphs result from different alleles at two epistatically interacting loci. For Lythraceae, Oxalidaceae, and Pontederiaceae the dominant allele at the *S* locus leads to the formation of S-morph flowers, while homozygous recessive *s*/*s* individuals are either L-morphs or M-morphs depending on the genotype at the *M* locus, such that *s*/*s M*/*m* and *s*/*s M*/*M* leads to M-morph flowers and *s*/*s m*/*m* to L-morph flowers (Lewis and Jones, 1992; Barrett, 1993; Barrett and Shore, 2008).

Population-genetic theory makes a number of predictions about the molecular features of the *S* locus and causal heterostyly genes. First, as the polymorphism in the floral morphs is under negative frequency-dependent balancing selection, the *S* and *s* alleles are expected to be maintained as stable polymorphisms for much longer periods of time than alleles at other background loci in the genome (Uyenoyama, 2005; Charlesworth, 2006). At the same time, recombination is suppressed within the supergene and the dominant *S* allele is only ever present in a heterozygous state. This results in at least a 4-fold reduction in the effective population size for the dominant *S* allele, which in turn should lead to a greater effect of genetic drift relative to selection on the *S* allele as compared with non-*S*-locus associated diploid regions of the genome (Uyenoyama, 2005). The three effects combined, namely long-term maintenance, suppressed recombination, and stronger genetic drift predict high levels of sequence divergence of *S* locus genes and the accumulation of repetitive elements and transposons (Uyenoyama, 2005; Charlesworth, 2006).

Models for the evolution of heterostyly

Two main models have been elaborated to explain the evolution of heterostyly (Charlesworth and Charlesworth, 1979; Lloyd and Webb, 1992*a*; Lloyd and Webb, 1992*b*). They differ markedly in their presumed ancestral state and in the inferred sequence of trait acquisition. The Charlesworth and Charlesworth model assumes a self-compatible homostylous ancestor, in which a mutation to a novel incompatible pollen type spreads and establishes a polymorphism, if the product of selfing rate and inbreeding depression is above 0.5 (Charlesworth and Charlesworth, 1979). A mutation to a novel female type, compatible with the novel pollen type could then spread to establish a diallelic incompatibility system as seen in the classical examples of heterostyly. Reciprocal herkogamy would evolve subsequently, most easily first by a dominant mutation to a shorter style and then a recessive mutation to a higher anther position; spreading of these mutations would be promoted or could only occur if they arose in tightly linked genes. A special case would be that the same mutation is responsible for a novel female incompatibility type and a short style and this would be essential for the evolution of heterostyly, if the recessive morph was more self-compatible than the dominant one, as seen in many *Primula* species (Charlesworth and Charlesworth, 1979; Wedderburn and Richards, 1990).

By contrast, the model by Lloyd and Webb is based on the taxonomic observation that heterostyly has only evolved in families with relatively simple, 'depth-probed' flowers with an intermediate level of specialization (Lloyd and Webb, 1992*a*). Such 'depth-probed' flowers often show approach herkogamy, with the stigma protruding above the anthers, leading to the suggestion that heterostyly evolved from approach-herkogamous ancestors with at least partial outbreeding. The first step in this model would be the invasion of such a population by a dominant mutation shortening the style, followed by a second mutation to elevate the anthers in the short-styled form to the level of the stigma in the original form. Thus, reciprocal herkogamy would evolve before self-incompatibility. The latter is proposed to evolve secondary due to reciprocal physiological adaptations of the pollen tubes to growth in the different stylar environments via either morph-linked or functionally morph-limited modifier mutations. The authors of both models concur that the molecular mechanism of heteromorphic self-incompatibility is likely to be fundamentally different from the self/non-self recognition mechanisms in homomorphic self-incompatibility systems (Charlesworth and Charlesworth, 1979; Lloyd and Webb, 1992*a*; Shimizu and Tsuchimatsu, 2015). This is based on variability in inhibition sites of incompatible pollen (tubes) and the fact that in self-compatible homostyles both female and male incompatibility responses are still intact but have been switched around, while self-compatible plants in homomorphic self-incompatibility systems arise by inactivating mutations of either the male or female components.

Some assumptions of the Charlesworth and Charlesworth model have been questioned early on and arguments have been put forward in support of the Lloyd and Webb model. Conceptually, a homostylous ancestral state with the high levels of inbreeding depression required by the Charlesworth and Charlesworth model appears unlikely. This is because non-herkogamous species and derived homostyles tend to be highly selfing (Piper *et al.*, 1984), which in turn should purge the genetic load causing inbreeding depression (Ganders, 1979). A phylogenetic argument for the Lloyd and Webb model is the wide distribution of some form of herkogamy, such as approach herkogamy amongst self-compatible monomorphic species of flowering plants (Ganders, 1979). Indeed, while this has not been achieved in many cases, ancestral-character reconstruction in *Exochaenium* has shown that heterostyly evolved once from an approach herkogamous ancestral state (Kissling and Barrett, 2013). Lastly, as mentioned above, not all heterostylous species show self-incompatibility, consistent with only having evolved the morphological aspect of heterostyly as the first step in the Lloyd and Webb model (Barrett and Cruzan, 1994).

With special reference to *Primula*, Al-Wadi and Richards suggested a scenario where distyly evolved from a long homostylous condition represented in some species in the section Sphondylia, with acquisition of reciprocal herkogamy first, followed by that of intra-morph incompatibility (Alwadi and Richards, 1993). However, this appears very unlikely in the light of phylogenetic reconstructions demonstrating that the most recent common ancestor of the entire genus *Primula* was distylous and that the species in

section Sphondylia are not primitively monomorphic (Mast *et al.*, 2006).

Molecular genetics of heterostyly

Identifying the molecular basis of heterostyly has been an important aim in order to shed light on its development, physiology, and evolution. Three main approaches have been employed. Differential expression or abundance of proteins or mRNAs in different morphs has been used to identify candidate factors; deletion mutagenesis and the search for homostyles have helped to delimit the chromosomal location of the *S* locus; genetic mapping and positional cloning have been used to define the genomic sequence of the *S* locus.

Turnera

All three of the above approaches have been combined in the neotropical shrubs of the genus *Turnera*. Species in this genus show typical distyly, with the S-morph controlled by the dominant *S* allele and the L-morph homozygous for the recessive *s* allele (Shore and Barrett, 1985). A difference is the relatively open, bowl-like structure of the *Turnera* flowers that contrasts with the more frequent tubular flowers found in many other heterostylous species (Ganders, 1979). A polygalacturonase and α-dioxygenase were identified as differentially abundant proteins between the styles of S- and L-morph flowers; in particular, both proteins are only detected in S-morph styles (Athanasiou and Shore, 1997; Athanasiou *et al.*, 2003; Khosravi *et al.*, 2004). The gene encoding the polygalacturonase is linked to, but separate from the *S* locus, while the α-dioxygenase gene is unlinked, suggesting that expression of both is secondarily controlled by the causal distyly factors. The *S* locus was genetically mapped in a large population of backcross individuals and three markers in complete linkage with the *S* locus were discovered (Labonne *et al.*, 2009). One of these markers identified a retrotransposon, and another one a sterol sulfotransferase.

The identified markers were used in two respects: to characterize induced deletion mutants (Labonne *et al.*, 2010) and to initiate a chromosome walk (Labonne and Shore, 2011). X-ray mutagenesis had identified a number of deletions affecting either the entire *S* locus or parts thereof; in particular, as mentioned above, one deletion resulted in a long homostyle phenotype and genotyping showed that the deletion was at the *S* locus, suggesting that the equivalent to the *Primula G* locus had been deleted, yet the equivalent to *A* was still present and functional (Labonne *et al.*, 2010). In parallel, bacterial artificial chromosome (BAC) contigs were constructed starting from the BACs harbouring the *S* locus associated markers. This identified a single BAC clone carrying the recessive *s* allele; also two overlapping BACs derived from the dominant *S* haplotype were isolated, yet their location relative to the causal *S* locus genes remains unclear (Labonne and Shore, 2011). Unfortunately, the sequences of these BAC contigs have not been reported to our knowledge; also, in light of the recent results from *Primula* it is unclear whether the BAC harbouring the *s* allele indeed contains recessive alleles of the causal heterostyly genes, or whether it lacks a fragment of *S*-allele specific sequence harbouring the causal heterostyly genes. Together with the establishment of a transformation protocol for *Turnera joelii* (Chafe *et al.*, 2015), these studies have brought us tantalizingly close to the molecular identities of the *Turnera* heterostyly genes, yet the last step remains to be taken.

Linum

Heterostyly is also observed in some species of flax (*Linum*), including *Linum tenuifolium*, *Linum usatissimum* and *Linum grandiflorum* (Dulberger, 1992). As in other systems, S-morph plants are assumed to be *S*/*s* heterozygous, while L-morph individuals are *s*/*s* homozygous recessive (Ushijima *et al.*, 2012). Supergene control similar to *Primula* has been suggested (Ushijima *et al.*, 2012) as self-compatible monomorphic populations have been identified, yet these morphologically resemble the L-morph (Nicholls, 1985). Compared with other systems, the difference in anther positions is not very pronounced in distylous *Linum* species, yet many of the typical ancillary polymorphisms are found (Dulberger, 1992) and the distylous form is strongly self-incompatible. Using a combination of proteomics and transcriptomics, differentially abundant proteins and genes with morph-specific expression were identified in *L. grandiflorum* (Ushijima *et al.*, 2012). These included in particular the *THRUM STYLE-SPECIFIC GENE 1* (*TSS1*) that is exclusively expressed in S-morph styles. This morph-restricted expression appears to be based on hemizygosity of the gene, being present only in S-morph plants, but not in L-morph plants, thus likely residing on the dominant *S* haplotype. Two additional genes with expression enriched in styles of S-morph plants encode an aspartyl protease and a MYB transcription factor, whose overexpression limits pistil growth in *Arabidopsis thaliana* (Ushijima *et al.*, 2012). The finding of likely hemizygosity for *TSS1* is particularly relevant as it foreshadowed recent findings in *Primula*.

Fagopyrum

The buckwheat *Fagopyrum esculentum* shows typical morphological distyly with intra-morph incompatibility (Darwin, 1877). Also genetically, the system conforms to the common pattern, with the S-morph determined by the dominant *S* haplotype; S-morph plants are heterozygous *S*/*s* and L-morph plants homozygous *s*/*s* (Garber and Quisenberry, 1927; Lewis and Jones, 1992). The naturally occurring, self-compatible long homostylous species *Fagopyrum homotropicum* appears to carry an altered *S* locus haplotype that is dominant to the *s* haplotype yet recessive to *S*, with female and male incompatibility behaviours consistent with organ positions (Matsui *et al.*, 2003). This is again consistent with this homostyle having arisen by recombination or mutation in an *S* locus supergene. Both molecular markers linked to the *S* locus and differentially expressed proteins between the morphs have been identified (Ali *et al.*, 1998; Miljus-Dukic *et al.*, 2004; Yasui *et al.*,

2004) and a BAC library was constructed to aid in the identification of the *S* locus (Yasui *et al.*, 2008).

A breakthrough came from the use of next generation sequencing to identify S-morph restricted transcripts on the hypothesis that the dominant *S* haplotype may contain alleles not present on the *s* haplotype or at least not expressed from it. This indeed led to the identification of four short-style specific transcripts, named *SSG1* to *SSG4* (Yasui *et al.*, 2012). A chimaeric plant sporting a side branch with long-styled flowers on an otherwise S-morph plant carried a deletion of the *SSG2* and *SSG3* genes, and *SSG3* was found to be present in each S-morph individual in a large mapping population but not in any L-morph plants, indicating complete linkage between the presence of *SSG3* and the dominant *S* haplotype. *SSG3* encodes a protein related to the *A. thaliana* EARLY FLOWERING 3 (ELF3) protein; a second homologue to *A. thaliana* ELF3 is also expressed in flowers and is indeed more closely related to *At*ELF3 than the *S* locus specific S-ELF3, indicating that the latter arose by a duplication. The complete correspondence between the presence of a functional *S-ELF3* gene and S-morph individuals extends even to different *Fagopyrum* species, indicating that balancing selection has maintained it as a trans-specific presence/absence polymorphism. Furthermore, two independently arisen self-compatible long homostylous genotypes of *Fagopyrum* carry inactivating mutations in *S-ELF3*, providing very strong evidence that *S-ELF3* is a causal heterostyly locus that determines style length and possibly also female incompatibility behaviour in *Fagopyrum*.

Exploiting the availability of a BAC library, a 610 kb region surrounding *S-ELF3* was defined that contained *SSG2* and another functionally uncharacterized gene, as well as a large number of repetitive sequences and transposable elements. This work was recently expanded on by a draft sequence of the *F. esculentum* genome combined with genotyping-by-sequencing of L-morph and S-morph landrace individuals (Yasui *et al.*, 2016); searching for further S-morph specific contigs identified a total of at least 5.4 Mb of sequence with evidence for absence in all L-morph individuals but presence in most or all S-morph individuals. The identified contigs contained a high fraction, almost 75%, of transposable element-derived sequences, yet only 32 predicted genes, suggesting a large, non-recombining hemizygous region that accumulates transposable elements due to the population-genetic effects outlined above. Functional analysis of these additional genes in the candidate *S* locus should identify further causal heterostyly genes in buckwheat.

Lithospermum

Heterostyly has evolved multiple times independently within the Boraginaceae and even within certain genera heterostyly appears to have evolved several times. One such example is *Lithospermum* (Cohen, 2010). To gain insights into the molecular control and ultimately the repeated evolution of heterostyly, a transcriptomic study in *Lithospermum multiflorum* identified differentially expressed genes between styles and corolla tubes of S- and L-morph plants at three different developmental stages (Cohen, 2016). This showed a dynamic pattern of changes in differentially expressed genes. In particular, identified genes in early stages included many candidate regulators of organ growth and development, while those in late stages were enriched in factors affecting physiology and stress response.

Eichhornia

The molecular basis of tristyly has recently been studied in *Eichhornia paniculata*, an annual relative of the clonal aquatic invader *Eichhornia crassipes* (Arunkumar *et al.*, 2017). *E. paniculata* shows the classical two-locus, two-allele control of tristyly, with the *S* locus epistatic over the *M* locus. Tristyly has been modified several times independently in *E. paniculata*, giving rise to largely selfing semi-homostylous mid- and long-morphs (M' and L'), where some of the stamens are located at the same height as the stigma. Using a backcross of an M' (genotype *s*/*s M*/*M*) x L' (genotype *s*/*s m*/*m*) hybrid to the parental L', a large population segregating for the *M* locus was generated (Arunkumar *et al.*, 2017). This was used to identify genes linked to the *M* locus and to map quantitative trait loci (QTL) for the floral-organ traits. A large chromosomal region of more than 10 Mb was found to co-segregate with the *M* locus, likely containing more than 300 genes. There was some evidence for reduced recombination in this region, yet little population-genetic signals of balancing selection. Thus, the question whether the *M* locus represents a supergene in the classical sense or merely contains one or very few pleiotropically acting genes remains open. Consistent with earlier studies (Fenster and Barrett, 1994), the modified stamen lengths leading to the semi-homostylous arrangement were due to modifier loci, at least some of which segregate independently of the *M* locus; in particular, some of the modifiers in the M' and L' variants were different, confirming their independent origins (Arunkumar *et al.*, 2017).

Primula

Early molecular-genetic studies in the *Primula* system identified differentially expressed genes between flowers of the two morphs and determined their location relative to the *S* locus. In one study, none of 11 differentially expressed genes was found linked to the *S* locus, indicating that these genes are potential downstream components in the elaboration or function of the two morphs (McCubbin *et al.*, 2006). A second study identified two genes, *SLL1* and *SLL2*, tightly linked to the *S* locus (Li *et al.*, 2007). Also, a restriction-fragment length polymorphism, *PvSLP1*, was identified in a highly repetitive region closely linked to the *S* locus (Manfield *et al.*, 2005).

In addition to these molecular markers, several genes identified by their mutant phenotypes were localized in close linkage to the *S* locus. This includes the dominant *Hose in hose* (*Hih*) mutant that causes the homeotic conversion of sepals to petals in *Primula vulgaris* due to the ectopic expression of the *P. vulgaris* orthologue to the B-class homoeotic gene *GLOBOSA* (*PvGLO1*) (Li *et al.*, 2010). Conversely, the

sepaloid mutation leads to flowers containing only sepals and carpels, reminiscent of B-function mutants, yet no mutation of *PvGLO1* was detected (Li *et al.*, 2008). The dominant *Oakleaf* mutation is also linked to the *S* locus and affects the shape of *P. vulgaris* leaves, rendering them more deeply lobed than wild-type leaves (Cocker *et al.*, 2015). Based on these various linked markers and genes, a high resolution genetic map of the *P. vulgaris S* locus region was constructed, placing *GLO1* and *Oakleaf* on opposite sides of the *S* locus (Li *et al.*, 2015). Three molecular markers were placed between *Oakleaf* and *S*, and between *PvGLO1* and *S*, respectively, and constructing and sequencing of BAC contigs confirmed this marker order. Fluorescence *in situ* hybridization using *S*-locus linked BACs as probes also confirmed that the *S* locus is located next to the centromere, as predicted based on classical genetics, and is found on the largest *Primula* chromosome.

Parallel work in *Primula veris* had led to a reference genome sequence from a pool of S- and L-morph DNA (Nowak *et al.*, 2015). This was combined with restriction-associated DNA (RAD) tag genotyping of large S- and L-morph pools from a semi-natural population with very short linkage disequilibrium and consequently a high mapping resolution. RAD tagging identified two markers as completely linked to the *S* locus, while for another four markers only a single crossover event was found. These six linked markers were located on four scaffolds of a combined size of more than 640 kb, suggesting that meiotic recombination is suppressed in a larger region at or surrounding the *S* locus. Genes on these contigs showed higher heterozygosity in S-morph than in L-morph individuals, consistent with sequence divergence due to suppressed recombination. Transcriptome analysis of flowers from S- and L-morphs identified a *GLOBOSA*-like gene as the most highly differentially expressed, being restricted to S-morph flowers (Nowak *et al.*, 2015). This gene, termed *GLO2*, represents a duplicated paralogue to *GLO1* identified based on the *Hih* mutant in *P. vulgaris*.

These studies in *P. vulgaris* and *P. veris* set the stage for two major recent breakthroughs in our understanding of heterostyly in *Primula*, the identification and functional characterization of the *G* locus in the heterostyly supergene (Huu *et al.*, 2016), and the assembly of sequence models of the dominant and recessive *S* locus alleles (Li *et al.*, 2016; Burrows and McCubbin, 2017). The latter came from the assembly and sequencing of BAC contigs starting from the above markers. Intriguingly, the assemblies of the *S* and the *s* haplotypes differed by a region of approximately 280 kb that was exclusively found in S-morph DNA (Li *et al.*, 2016). Presence of this 280 kb insertion co-segregated perfectly with the S-morph phenotype in different laboratory and natural populations. This region contains five predicted genes, including *GLO2* and *CYP734A50*, a gene encoding a kelch-repeat F-box protein (*KFB*), a gene for a Pumilio-like RNA-binding protein (*PUM*), and a gene encoding a protein with a highly conserved C-terminal domain (*CCM*). To avoid confusion, the nomenclature based on *P. veris* will be used for *GLO2* and *CYP734A50*, as the presence of *GLO2* as a duplicated paralogue and of *CYP734A50* was first described in this species (Nowak *et al.*, 2015; Huu *et al.*, 2016). The inserted region is flanked by a duplicated *cyclin-like F Box* gene (*CFB*). The regions just flanking the insertion contain stretches without elevated rates of polymorphism between S- and L-morph haplotypes, arguing that recombination has occurred recently in the vicinity of the *S* locus. This observation is in contrast to the conclusion from *P. veris* (Nowak *et al.*, 2015); further studies will be needed to resolve whether this is a difference between the two species analyzed or depends on the analysis method. Analysis of naturally occurring long homostyles in *P. vulgaris* identified a one-base pair insertion in *CYP734A50* in one line and an amino-acid mutation of unknown significance in another line. A spontaneous short homostyle mutant carried a transposon insertion in *GLO2*, suggesting that *CYP734A50* and *GLO2* represent the *G* and *A* loci in the original supergene model, respectively. Phylogenetic dating of the *GLO2*/*GLO1* duplication suggests an age of more than 50 million years ago, considerably older than the divergence of *Primula* and the most closely related non-heterostylous genus *Androsace* (Li *et al.*, 2016).

In a parallel approach, tissue-resolved transcriptomics in *P. veris* identified the *CYP734A50* gene as being exclusively expressed in the styles of S-morph plants (Huu *et al.*, 2016). This morph-restricted expression is due to the gene only being present in S-morph plants and this presence/absence polymorphism of *CYP734A50* perfectly co-segregates with the morphs, not just in a large *P. veris* population but also in all other heterostylous *Primula* species tested, including *Primula forbesii*. The latter diverged from *P. veris* more than 20 million years ago, indicating that this trans-specific *CYP734A50* presence/absence polymorphism has been maintained by balancing selection, as predicted for a causal heterostyly locus. That *CYP734A50* was indeed the *G* locus controlling style length was supported by several lines of evidence. First, the gene was absent from several naturally occurring long homostylous species but present in all S-morph individuals of closely related heterostylous species. Second, naturally occurring *P. vulgaris* long homostyles showed a strongly reduced expression; incidentally, this was the population that only carried the amino acid exchange mutation (see above). Third, two independent spontaneous mutants of cultivated primroses carried deletions of the fourth exon. Fourth, virus-induced gene silencing (VIGS) of *CYP734A50* in *P. forbesii* resulted in long homostyles. The encoded cytochrome P450 enzyme is homologous to several functionally characterized family members from tobacco, rice, and *A. thaliana* that all inactivate brassinosteroids (BRs) (Neff *et al.*, 1999; Ohnishi *et al.*, 2006; Thornton *et al.*, 2011) and indeed styles of S-morph *P. vulgaris* plants contain strongly reduced levels of bioactive BRs compared with those of L-morph plants (Huu *et al.*, 2016). This reduced level appears to be the reason for the reduced cell expansion in S-morph styles that underlies their short length (Webster and Gilmartin, 2006), as it can be rescued by exogenous BR treatment. The *CYP734A50* gene also arose from a duplication, like *GLO2*, and as expected from population genetic theory it shows a higher rate of sequence divergence between *Primula* species compared with the paralogue outside of the *S* locus, most likely reflecting the

reduced efficiency of purifying selection and greater importance of drift at the non-recombining *S* allele. Also, comparing the genomic sequences of the *S* locus genes *CYP734A50* and *GLO2* with those of their paralogues identifies much longer introns in the *S* locus genes that contain many transposable element-derived sequences and repeats.

A third independent study also established and sequenced two BAC contigs from *P. vulgaris* containing the two *GLO* paralogues (Burrows and McCubbin, 2017). One of these contigs contains part of the dominant *S*-haplotype, confirming the presence of *CYP734A50* and *GLO2* as duplicated genes and of *CCM* in the hemizygous region of the *S* supergene, even though *CCM* is not annotated in this contig (Li *et al.*, 2016; Burrows and McCubbin, 2017). Interestingly, the 285 kb contig containing *GLO1* did not identify any additional genes with homology to those located at the *S*-locus. Also, comparing the *P. vulgaris S* locus contig with the *P. veris* contigs from the genome study suggested that the corresponding *P. veris* contig is from the *s* haplotype lacking the hemizygous region; this *P. veris* contig contains one of the two completely linked RAD tag markers (Nowak *et al.*, 2015).

Thus, the following major conclusions can be drawn from these three studies. First, in contrast to the classical model, the *S* locus is not a diallelic supergene but rather a hemizygous region containing at least five genes. Second, *CYP734A50* is the *G* gene determining style length and possibly female incompatibility behaviour, while *GLO2* represents a strong candidate for the *A* gene controlling anther position. Third, the hemizygosity of the dominant *S* haplotype provides a straightforward explanation for suppressed recombination as there is no homologous sequence on the *s* haplotype for it to pair with and undergo crossing-over. Fourth, long and short homostyles, as well as other abnormal trait combinations have not arisen by crossing-over within the *S* locus between homologous sequences, but rather by mutations. Fifth, at least the two studied genes at the *S* locus show the molecular-evolutionary signatures expected for a non-recombining, permanently hemizygous region.

Implications of recent findings for the evolution, function, and breakdown of heterostyly

How do the above findings on the *Primula S* locus integrate with previous work and thinking on the topic, and what are their implications for the function and the evolution of heterostyly in *Primula* and possibly beyond?

An important question is how the conclusion from the present work that homostyles and other abnormal types cannot have resulted from homologous recombination between the *S* and *s* alleles can be reconciled with the previous interpretation that they arose by crossing-over within the *S* locus. Examining flanking-marker genotypes given for previously studied *P. vulgaris* short and long homostyles indicates that for the long homostyle the *PvSLL1* genotype is homozygous S-morph like, consistent with this plant being homozygous for a mutant version of the dominant *S* haplotype; the *PvSLL2* marker was homozygous for a novel allele, very similar in size to the described S-morph allele, which may reflect intraspecific polymorphism in *P. vulgaris* (Li *et al.*, 2007). By contrast, the *PvSLL1* genotype of the short homostyle was homozygous L-morph like and heterozygous at *PvSLL2*, inconsistent with a mutational origin from the dominant *S* allele; however, this short homostyle was suggested to be due to mutations outside of the *S* locus (Li *et al.*, 2007), similar to the situation in a short homostyle line in *F. esculentum* (Matsui *et al.*, 2004). Thus, overall the available results from genotyping *Primula* homostyles appear consistent with their origin by mutation, rather than recombination in the *S* locus.

A major argument against mutation was the frequent occurrence of such abnormal types at frequencies above 10^{-3}, which appeared incompatible with known *per locus* mutation rates of 10^{-4} to 10^{-7} (Lewis and Jones, 1992). Concerning the different types of mutations, the rate of point mutations does not appear to be higher at the *S* locus than elsewhere in the genome. This conclusion is based on the comparison of sequence divergence between the *CYP734A50* orthologues from different *Primula* species with that between their *CYP734A51* paralogue sequences (Huu *et al.*, 2016). The rate of synonymous changes was essentially the same between the *CYP734A50* orthologues and the non-*S* locus paralogues, indicating that the rate of point mutations was comparable; it was only the rate of non-synonymous changes that was higher at the *S* locus, consistent with less efficient purifying selection. However, a major difference between the *S* locus and general euchromatic regions of the *Primula* genome is the much higher density of transposable element-derived and repetitive sequences (Huu *et al.*, 2016; Li *et al.*, 2016; Burrows and McCubbin, 2017). It is conceivable that some of these transposable elements are still potentially mobile and not entirely silenced and that their activity underlies the high frequency at which abnormal types arise. Alternatively, these transposable element-derived sequences may provide a substrate for illegitimate, non-homologous recombination with other copies outside of the *S* locus, which would lead to disruption of the *S* locus. In support of these possibilities, several of the *S* haplotypes from long homostyles, either from naturally occurring homostylous species or from cultivated strains, harbour deletions of all or part of the *CYP734A50* sequence (Huu *et al.*, 2016). That said, at least one natural long homostyle in *P. vulgaris* is due to a single-base insertion (Li *et al.*, 2016). Also, several of the novel abnormal types reported by Ernst, such as a *GPA* haplotype arising from a *GP*/*A* transheterozygous individual remain difficult to explain (Ernst, 1936; Dowrick, 1956; Lewis and Jones, 1992), except with trivial explanations like rare pollen cross-contamination. Nevertheless, it appears that a higher mutation rate at the *S* locus resulting from residual transposable element activity or non-homologous recombination between repeated sequences offers a plausible explanation for the relatively frequent occurrence of abnormal types.

Another important implication of these recent findings is that they suggest a model for the chromosomal evolution of the *S* locus supergene. For both *CYP734A50* and *GLO2* there is strong evidence that they have arisen by duplication, with

paralogues found outside of the *S* locus (Huu *et al.*, 2016; Li *et al.*, 2016; Burrows and McCubbin, 2017). However, a scenario of step-wise gene-by-gene duplication from unlinked ancestral genes, with the duplicates fortuitously ending up next to each other, appears highly unlikely. Rather, segmental duplication with subsequent gene loss and neofunctionalization of retained genes appears to be a more plausible scenario (Fig. 2). The suggestion of gene loss is based on the observation that the *GLO1* contigs in *P. vulgaris* and *P. veris* do not contain the *CYP734A51* paralogue (Nowak *et al.*, 2015; Burrows and McCubbin, 2017), that is the two ancestral genes do not appear to be neighbours as are their duplicates at the *S* locus. How this segmental duplication arose remains unclear at present; one possibility suggested by the duplicated *CFB* genes at the boundaries of the *S* locus region is that a small duplication of an ancestral *CFB* gene generated the substrate for illegitimate crossing-over and the origin of a segmental duplication. Importantly, the results from *Fagopyrum* and *L. usatissimum* described above also indicate that the dominant *S* locus allele is a hemizygous region potentially derived from duplication, suggesting that the above scenario may apply more widely for the evolution of heterostyly.

A third implication of the *Primula* work is that style length and the female incompatibility behaviour may both be controlled by the same gene, *CYP734A50*. This is suggested by the findings concerning the naturally occurring long homostyles in Southern English *P. vulgaris* populations (Crosby, 1940; Crosby, 1949). These long homostyles are self-compatible, with the long style showing the incompatibility phenotype of normal long L-morph styles. Sequencing identified only a single base-pair insertion in *CYP734A50* in the long homostyles from Somerset and an amino acid exchange and strongly reduced gene expression was reported for the long homostyle from the Chiltern Hills (Huu *et al.*, 2016; Li *et al.*, 2016). Thus, unless these lines harbour undetected linked mutations, it appears that loss of *CYP734A50* function alters both style length and the female incompatibility response. In turn, this would suggest that self-incompatibility is not based on a dedicated self/non-self recognition system as in the well-studied examples of homomorphic self-incompatibility in plants, but rather depends on matching physiological adaptations of pollen tubes and stylar environment. These would only permit successful pollen tube growth in compatible combinations, as suggested in the Lloyd and Webb model for the evolution of heterostyly (Lloyd and Webb, 1992*a*). It is well-established that the strength of intra-morph incompatibility varies greatly between species and even between the two morphs of a given species in *Primula* (Wedderburn and Richards, 1990; Richards, 2003); the proposed physiological scenario would seem more compatible with this variability, based on the effects of modifier mutations on pollen tube growth in the style, than would a more classical self/non-self recognition mechanism. Evidence for polygenic modification of the intra-morph incompatibility has been obtained in several examples in *Turnera* (Shore and Barrett, 1986). Also, there may not be a common intra-morph incompatibility mechanism in both morphs of one species, based on differential cellular phenotypes of incompatible pollen of L- versus S-morphs, for example in *Turnera* (Safavian and Shore, 2010) and in distylous species of *Rubiaceae* (Klein *et al.*, 2009). Thus, on balance these observations seem to be more easily explained by intra-morph incompatibility as a secondary physiological adaptation as proposed by Lloyd and Webb, even though more elaborate mechanisms cannot be excluded at present, for example the BR-dependent expression of cognate receptor-ligand pairs in styles and pollen tubes.

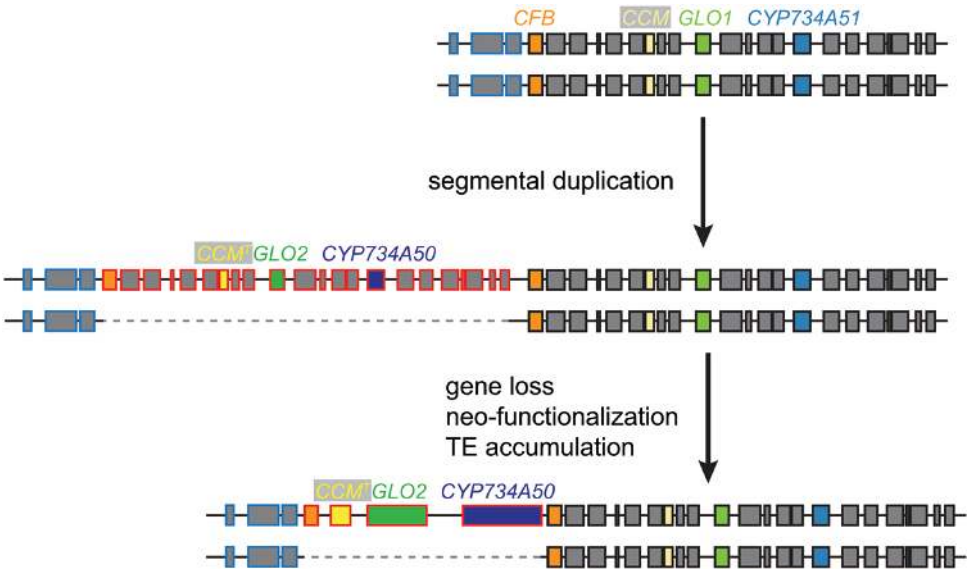

Fig. 2. Model for the chromosomal evolution of the *Primula S* locus supergene. A segmental duplication arises and is maintained as a polymorphism in the population as a *proto-S* locus. Duplicated genes in this region (red boxes) can be lost or acquire new activities by neofunctionalization, establishing the dominant *S* locus haplotype. Once this dominant haplotype is only ever present in heterozygotes and thus does not undergo recombination, transposable elements (TEs) will accumulate there, increasing the sizes of introns and intergenic regions. Coloured boxes indicate genes labelled in the same coloured font. The two additional predicted genes in the dominant *S* locus haplotype (see Fig. 1) are not shown, as no paralogues outside of the *S* locus have been reported for them, making their origin unclear.

In summary, the above considerations add several plausible pieces to the model for the evolution of heterostyly in *Primula* (Fig. 2). Segmental duplication in an approach herkogamous ancestor gave rise to the *proto-S* locus and this duplication was maintained as a polymorphism in the population. The *proto-S* locus underwent gene loss and several of the duplicates evolved novel functions related to flower morphology and function. In particular, *CYP734A50* gained style-specific expression, reducing both style length and altering the permissiveness of the style for pollen tube growth; *GLO2* evolved a novel function in promoting cell proliferation in the lower part of the corolla tube beneath the anther insertion point (Webster and Gilmartin, 2006), leading to the high anthers of S-morph flowers. Male incompatibility may then have evolved by reciprocal adaptation of the pollen tubes to the different stylar environments. In summary, this scenario is more in line with the Lloyd and Webb model than with the Charlesworth and Charlesworth model; in particular, all mutations for acquiring novel functions by *S* locus genes would be dominant, a scenario not covered by the latter model.

Several obvious experiments suggest themselves to test the proposed scenario. First, a better, more contiguous genome assembly should be able to answer whether the *S* locus indeed arose from a segmental duplication via gene loss and neofunctionalization. Second, studying many more molecularly defined, preferably small-scale mutations in *CYP734A50* will answer whether alterations in style length and female incompatibility type always go together; establishing a transformation protocol should allow for testing this hypothesis further (Hayta *et al.*, 2016). Third, functional characterization of the remaining genes in the dominant *S* locus haplotype will identify the gene(s) required for male compatibility behaviour; this in turn may shed further light on the evolutionary sequence of events.

Comparison with supergenes from animals

Supergenes like the *S* locus also underlie complex adaptive polymorphisms in animals. How do the recent insights from heterostyly compare with findings in these animal systems?

Arguably the best understood examples at the molecular level are the mimicry supergenes in *Heliconius* and *Papilio* butterflies. *Heliconius numata* butterflies show Mullerian mimicry, with different morphs resembling one of seven different species of *Melinea* butterflies (Joron *et al.*, 2006; Joron *et al.*, 2011). The seven distinct phenotypes are controlled by a supergene with haplotypes in a strict dominance hierarchy (Le Poul *et al.*, 2014). The supergene is localized in a region containing three closely linked wing-patterning genes in other *Heliconius* species (Joron *et al.*, 2006), one of which has recently been identified (Nadeau *et al.*, 2016). In *H. numata*, beneficial allele combinations at these three patterning genes have been locked into supergene haplotypes by two nested inversions that suppress recombination between mimetic and non-mimetic haplotypes, but also between different mimetic haplotypes (Joron *et al.*, 2011).

Harmless butterflies in the genus *Papilio* show Batesian mimicry, imitating unpalatable species, for example from Danainae. Mimicry is polymorphic, with different forms resembling different model species. A single Mendelian locus determines the different phenotypes, each involving variation in several pattern elements of the wings. Supergene architecture had therefore long been assumed (Charlesworth, 2016). However, recent molecular studies have indicated that allelic variants of only single genes – *doublesex* (*dsx*) in *Papilio polyetes* and *engrailed* in *Papilio dardanus* – determine the different mimicry morphs (Kunte *et al.*, 2014; Timmermans *et al.*, 2014; Nishikawa *et al.*, 2015). In *P. polyetes*, the mimetic *dsx* haplotype is inverted in a 130 kb region relative to the non-mimetic one, explaining the suppressed recombination between sequence polymorphisms within *dsx*. By contrast, no structural rearrangement has been found in *P. dardanus*.

Another supergene based on a chromosomal inversion has recently been identified in the ruff, a bird with three distinct male morphs. Two morphs are associated with different dominant haplotypes of a large derived inversion, while the third morph is homozygous for the recessive ancestral haplotype (Kupper *et al.*, 2016; Lamichhaney *et al.*, 2016). An even larger inversion affecting a region >100 Mb controls two alternative morphs in the white-throated sparrow (Tuttle *et al.*, 2016). One of them is always heterozygous for the inversion, while the other morph is homozygous for the non-inverted chromosome. Similarly, in fire ants a large chromosomal inversion suppresses recombination between two haplotypes affecting social organization (Wang *et al.*, 2013); as in the sparrows, one of the chromosomes is never homozygous. As expected for non-recombining regions that are only ever present in heterozygotes, the respective chromosomes in sparrows and ants shows signs of degeneration, reflecting the accumulation of mutations due to Muller's ratchet.

Thus, virtually all molecularly defined supergenes in animals involve inversions that suppress recombination and maintain co-adapted alleles segregating as a single unit, yet no case of a segregating hemizygous region has been described. By contrast, no inversions have been found in examples of the heterostyly supergene. It will be fascinating to see whether this difference between the type of structural chromosomal variation that suppresses recombination in animal versus plant supergenes holds up as more examples are molecularly dissected.

Outlook

In summary, research into the molecular basis of heterostyly has come a long way during the last decade. Several examples are beginning to be understood in considerable detail and hemizygosity of the dominant *S* haplotype is emerging as a potentially widespread feature of heterostylous systems. This offers both an elegant solution to the problem of suppressing recombination and suggests a possible evolutionary path based on segmental duplication, gene loss, and neofunctionalization. At the same time, in contrast to the intensively studied and molecularly characterized homomorphic self-incompatibility systems (Fujii *et al.*, 2016), the molecular and physiological basis of heteromorphic self-incompatiblity remains poorly understood. We argue that more work should be directed towards this problem as it has critical implications

for the evolutionary sequence of the different component traits of heterostyly. In any case, if the last decade is anything to go by, the future looks bright for understanding the molecular and genetic basis of heterostyly as a fascinating floral adaptation that functions to promote cross pollination and limit the harmful effects of inbreeding.

Acknowledgements

We thank Adrien Sicard, Isabel Bäurle and members of the Lenhard and Bäurle groups for stimulating discussion and comments.

References

Ali J, Nagano M, Penner GA, Campbell CG, Adachi T. 1998. Identification of RAPD markers linked to the homostylar (Ho) gene in buckwheat. Breeding Science **48,** 59–62.

Alwadi H, Richards AJ. 1993. Primary homostyly in *Primula L* Subgenus *Sphondylia* (Duby) rupr and the evolution of distyly in primula. New Phytologist **124,** 329–338.

Arunkumar R, Wang W, Wright SI, Barrett SC. 2017. The genetic architecture of tristyly and its breakdown to self-fertilization. Molecular Ecology **26,** 752–765.

Athanasiou A, Shore JS. 1997. Morph-specific proteins in pollen and styles of *Distylous turnera* (Turneraceae). Genetics **146,** 669–679.

Athanasiou A, Khosravi D, Tamari F, Shore JS. 2003. Characterization and localization of short-specific polygalacturonase in distylous *Turnera subulata* (Turneraceae). American Journal of Botany **90,** 675–682.

Barrett SC. 2002. The evolution of plant sexual diversity. Nature Reviews Genetics **3,** 274–284.

Barrett SCH. 1992. Evolution and function of heterostyly. Berlin Heidelberg: Springer.

Barrett SCH. 1993. The evolutionary biology of tristyly. In: Futuyma D, and Antonovics J, eds. Oxford surveys in evolutionary biology, Oxford: Oxford University Press, 283–326.

Barrett SCH, Cruzan MB. 1994. Incompatibility in heterostylous plants. In: Williams EG, Knox RB, Clarke AE, eds. Genetic control of self-incompatibility and reproductive development in flowering plants, The Netherlands: Kluwer Academic Publ, 189–219.

Barrett SCH, Shore JS. 2008. New insights on heterostyly: Comparative biology, ecology and genetics. In: Franklin-Tong VE, ed. Self-incompatibility in flowering plants: evolution, diversity and mechanisms, Berlin: Springer, 3–32.

Bateson W, Gregory RP. 1905. On the inheritance of heterostylism in Primula. Proceedings of the Royal Society of London B Series **76,** 581–586.

Burrows BA, McCubbin AG. 2017. Sequencing the genomic regions flanking S-linked PvGLO sequences confirms the presence of two GLO loci, one of which lies adjacent to the style-length determinant gene CYP734A50. Plant Reproduction **30,** 53–67.

Chafe PDJ, Lee T, Shore JS. 2015. Development of a genetic transformation system for distylous *Turnera joelii* (Passifloraceae) and characterization of a self-compatible mutant. Plant Cell, Tissue and Organ Culture **120,** 507–517.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLOS genetics **2,** e64.

Charlesworth D. 2016. The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. Evolutionary Applications **9,** 74–90.

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

Cocker JM, Webster MA, Li J, 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 Phytologist **208,** 149–161.

Cohen JI. 2010. "A case to which no parallel exists": The influence of Darwin's different forms of flowers. American Journal of Botany **97,** 701–716.

Cohen JI. 2011. A phylogenetic analysis of morphological and molecular characters of *Lithospermum L*. (Boraginaceae) and related taxa: evolutionary relationships and character evolution. Cladistics **27,** 559–580.

Cohen JI. 2016. De novo sequencing and comparative transcriptomics of floral development of the distylous species *Lithospermum multiflorum*. Frontiers in Plant Science **7,** 1934.

Crosby JL. 1940. High proportions of homostyle plants in populations of *Primula vulgaris*. Nature **145,** 672.

Crosby JL. 1949. Selection of an unfavourable gene-complex. Evolution **3,** 212–230.

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. Proceedings of The Royal Society of London B: Biological Sciences **281,** 20140075.

Dowrick VPJ. 1956. Heterostyly and homostyly in *Primula obconica*. Heredity **10,** 219–236.

Dulberger R. 1992. Floral polymorphisms and their functional significance in the heterostylous syndrome. In: Barrett SCH, ed. Evolution and function of heterostyly, Berlin Heidelberg: Springer, 41–84.

Ernst A. 1928. Zur Vererbung der morphologischen Heterostyliemerkmale. Berichte der Deutschen Botanischen Gesellschaft **46,** 573–588.

Ernst A. 1936. Heterostylie-Forschung: Versuche zur genetischen Analyse eines Organisations-und „Anpassungs"merkmales. Zeitschrift für Induktive Abstammungs- und Vererbungslehre **71,** 156–230.

Ernst A. 1955. Self-fertility in monomorphic Primulas. Genetica **27,** 391–448.

Fenster CB, Barrett SCH. 1994. Inheritance of Mating-System Modifier Genes in *Eichhornia-Paniculata* (Pontederiaceae). Heredity **72,** 433–445.

Fujii S, Kubo K, Takayama S. 2016. Non-self- and self-recognition models in plant self-incompatibility. Nature Plants **2,** 16130.

Ganders FR. 1979. The biology of heterostyly. New Zealand Journal of Botany **17,** 607–635.

Garber RJ, Quisenberry KS. 1927. The inheritance of length of style in buckwheat. Journal of Agricultural Research **34,** 181–183.

Gilmartin PM. 2015. On the origins of observations of heterostyly in Primula. New Phytologist **208,** 39–51.

Hayta S, Smedley MA, Li J, Harwood WA, Gilmartin PM. 2016. Plant Regeneration from Leaf-derived Callus Cultures of Primrose (*Primula vulgaris*). HortScience **51,** 558–562.

Huu CN, Kappel C, Keller B, *et al*. 2016. Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. eLife **5,** e17956.

Joron M, Papa R, Beltrán M, *et al*. 2006. A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. PLOS Biology **4,** e303.

Joron M, Frezal L, Jones RT, *et al*. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature **477,** 203–206.

Keller B, Thomson JD, Conti E. 2014. Heterostyly promotes disassortative pollination and reduces sexual interference in Darwin's primroses: evidence from experimental studies. Functional Ecology **28,** 1413–1425.

Khosravi D, Yang ECC, Siu KWM, Shore JS. 2004. High level of alpha-dioxygenase in short styles of distylous Turnera species. International Journal of Plant Sciences **165,** 995–1006.

Kissling J, Barrett SC. 2013. Variation and evolution of herkogamy in *Exochaenium* (Gentianaceae): implications for the evolution of distyly. Annals of Botany **112,** 95–102.

Klein DE, Freitas L, Da Cunha M. 2009. Self-incompatibility in a distylous species of Rubiaceae: is there a single incompatibility response of the morphs? Sexual Plant Reproduction **22,** 121–131.

Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, Mullen SP, Kronforst MR. 2014. Doublesex is a mimicry supergene. Nature **507,** 229–232.

Küpper C, Stocks M, Risse JE, *et al*. 2016. A supergene determines highly divergent male reproductive morphs in the ruff. Nature Genetics **48,** 79–83.

Kurian V, Richards AJ. 1997. A new recombinant in the heteromorphy 'S' supergene in Primula. Heredity **78,** 383–390.

Labonne JD, Shore JS. 2011. Positional cloning of the s haplotype determining the floral and incompatibility phenotype of the long-styled morph of distylous Turnera subulata. Molecular Genetics and Genomics **285,** 101–111.

Labonne JD, Tamari F, Shore JS. 2010. Characterization of X-ray-generated floral mutants carrying deletions at the S-locus of distylous *Turnera subulata*. Heredity **105,** 235–243.

Labonne JJ, Goultiaeva A, Shore JS. 2009. High-resolution mapping of the S-locus in Turnera leads to the discovery of three genes tightly associated with the S-alleles. Molecular Genetics and Genomics **281,** 673–685.

Lamichhaney S, Fan G, Widemo F, *et al*. 2016. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). Nature Genetics **48,** 84–88.

Le Poul Y, Whibley A, Chouteau M, Prunier F, Llaurens V, Joron M. 2014. Evolution of dominance mechanisms at a butterfly mimicry supergene. Nature Communications **5,** 5644.

Lewis D, Jones DA. 1992. The genetics of heterostyly. In: Barrett SCH, ed. Evolution and function of heterostyly, Berlin Heidelberg: Springer, 129–150.

Li J, Webster M, Furuya M, Gilmartin PM. 2007. Identification and characterization of pin and thrum alleles of two genes that co-segregate with the *Primula* S locus. The Plant Journal **51,** 18–31.

Li J, Webster M, Dudas B, Cook H, Manfield I, Davies B, Gilmartin PM. 2008. The S locus-linked *Primula* homeotic mutant sepaloid shows characteristics of a B-function mutant but does not result from mutation in a B-function gene. The Plant Journal **56,** 1–12.

Li J, Webster MA, Wright J, Cocker JM, Smith MC, Badakshi F, Heslop-Harrison P, Gilmartin PM. 2015. Integration of genetic and physical maps of the *Primula vulgaris* S locus and localization by chromosome in situ hybridization. Phytologist **208,** 137–148.

Li J, Cocker JM, Wright J, *et al*. 2016. Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. Nature Plants **2,** 16188.

Li JH, Dudas B, Webster MA, Cook HE, Davies BH, Gilmartin PM. 2010. Hose in Hose, an S locus-linked mutant of *Primula vulgaris*, is caused by an unstable mutation at the Globosa locus. Proceedings of the National Academy of Sciences, USA **107,** 5664–5668.

Lloyd DG, Webb CJ. 1992a. The evolution of heterostyly. In: Barrett SCH, ed. Evolution and function of heterostyly, Berlin Heidelberg: Springer, 151–178.

Lloyd DG, Webb CJ. 1992b. The selection of heterostyly. In: Barrett SCH, ed. Evolution and function of heterostyly, Berlin Heidelberg: Springer, 179–208.

Manfield IW, Pavlov VK, Li J, Cook HE, Hummel F, Gilmartin PM. 2005. Molecular characterization of DNA sequences from the *Primula vulgaris* S-locus. Journal of Experimental Botany **56,** 1177–1188.

Mast AR, Kelso S, Conti E. 2006. Are any primroses (*Primula*) primitively monomorphic? New Phytologist **171,** 605–616.

Matsui K, Nishio T, Tetsuka T. 2004. Genes outside the S supergene suppress S functions in buckwheat (*Fagopyrum esculentum*). Annals of Botany **94,** 805–809.

Matsui K, Tetsuka T, Nishio T, Hara T. 2003. Heteromorphic incompatibility retained in self-compatible plants produced by a cross between common and wild buckwheat. New Phytologist **159,** 701–708.

McCubbin AG, Lee C, Hetrick A. 2006. Identification of genes showing differential expression between morphs in developing flowers of *Primula vulgaris*. Sexual Plant Reproduction **19,** 63–72.

Miljus-Dukic J, Ninkovic S, Radovic S, Maksimovic V, Brkljacic J, Neskovic M. 2004. Detection of proteins possibly involved in self-incompatibility response in distylous buckwheat. Biologia Plantarum **48,** 293–296.

Nadeau NJ, Pardo-Diaz C, Whibley A, *et al*. 2016. The gene cortex controls mimicry and crypsis in butterflies and moths. Nature **534,** 106–110.

Neff MM, Nguyen SM, Malancharuvil EJ, *et al*. 1999. BAS1: A gene regulating brassinosteroid levels and light responsiveness in Arabidopsis. Proceedings of the National Academy of Sciences, USA **96,** 15316–15323.

Nicholls MS. 1985. The Evolutionary Breakdown of Distyly in *Linum tenuifolium* (Linaceae). Plant Systematics and Evolution **150,** 291–301.

Nishikawa H, Iijima T, Kajitani R, *et al*. 2015. A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. Nature Genetics **47,** 405–409.

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 Biology **16,** 12.

Ohnishi T, Nomura T, Watanabe B, Ohta D, Yokota T, Miyagawa H, Sakata K, Mizutani M. 2006. Tomato cytochrome P450 CYP734A7 functions in brassinosteroid catabolism. Phytochemistry **67,** 1895–1906.

Piper JG, Charlesworth B, Charlesworth D. 1984. A high rate of self-fertilization and increased seed fertility in homostyle primroses. Nature **310,** 50–51.

Richards J. 2003. Primula. Portland, Oregon: Timber Press.

Safavian D, Shore JS. 2010. Structure of styles and pollen tubes of distylous *Turnera joelii* and *T. scabra* (Turneraceae): are there different mechanisms of incompatibility between the morphs? Sexual Plant Reproduction **23,** 225–237.

Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. Current Biology **24,** R288–R294.

Shimizu KK, Tsuchimatsu T. 2015. Evolution of Selfing: recurrent patterns in molecular adaptation. Annual Review of Ecology, Evolution, and Systematics **46,** 593–622.

Shore JS, Barrett SCH. 1985. The genetics of distyly and homostyly in *Turnera ulmifolia* L (Turneraceae). Heredity **55,** 167–174.

Shore JS, Barrett SCH. 1986. Genetic modifications of dimorphic incompatibility in the *Turnera ulmifolia* L complex (Turneraceae). Canadian Journal of Genetics and Cytology **28,** 796–807.

Stone JL. 1995. Pollen donation patterns in a tropical distylous shrub (*Psychotria suerrensis* Rubiaceae). American Journal of Botany **82,** 1390–1398.

Thornton LE, Peng H, Neff MM. 2011. Rice CYP734A cytochrome P450s inactivate brassinosteroids in Arabidopsis. Planta **234,** 1151–1162.

Timmermans MJTN, Baxter SW, Clark R, *et al*. 2014. Comparative genomics of the mimicry switch in *Papilio dardanus*. Proceedings of The Royal Society B: Biological Sciences **281,** 20140465.

Tuttle EM, Bergland AO, Korody ML, *et al*. 2016. Divergence and functional degradation of a sex chromosome-like supergene. Current Biology **26,** 344–350.

Ushijima K, Nakano R, Bando M, *et al*. 2012. Isolation of the floral morph-related genes in heterostylous flax (*Linum grandiflorum*): the genetic polymorphism and the transcriptional and post-transcriptional regulations of the S locus. The Plant Journal **69,** 317–331.

Uyenoyama MK. 2005. Evolution under tight linkage to mating type. New Phytologist **165,** 63–70.

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang YC, Shoemaker D, Keller L. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. Nature **493,** 664–668.

Webster MA, Gilmartin PM. 2006. Analysis of late stage flower development in *Primula vulgaris* reveals novel differences in cell morphology and temporal aspects of floral heteromorphy. New Phytologist **171,** 591–603.

Wedderburn F, Richards AJ. 1990. Variation in within-morph incompatibility inhibition sites in heteromorphic *Primula L*. New Phytologist **116,** 149–162.

Yasui Y, Wang Y, Ohnishi O, Campbell CG. 2004. Amplified fragment length polymorphism linkage analysis of common buckwheat (*Fagopyrum esculentum*) and its wild self-pollinated relative *Fagopyrum homotropicum*. Genome **47,** 345–351.

Yasui Y, Mori M, Matsumoto D, Ohnishi O, Campbell CG, Ota T. 2008. Construction of a BAC library for buckwheat genome research - an application to positional cloning of agriculturally valuable traits. Genes & Genetic Systems **83,** 393–401.

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.

Yasui Y, Hirakawa H, Ueno M, Matsui K, Katsube-Tanaka T, Yang SJ, Aii J, Sato S, Mori M. 2016. Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. DNA Research **23,** 215–224.