The genomes of Darwin's primroses reveal chromosome-scale adaptive introgression and differential permeability of species boundaries

Stubbs, Rebecca L ; Theodoridis, Spyros ; Mora-Carrera, Emiliano ; Keller, Barbara ; Potente, Giacomo ; Yousefi, Narjes ; Jay, Paul ; Léveillé-Bourret, Étienne ; Choudhury, Rimjhim Roy ; Celep, Ferhat ; Kochjarová, Judita ; Conti, Elena

Abstract: Introgression is an important source of genetic variation that can determine species adaptation to environmental conditions. Yet, definitive evidence of the genomic and adaptive implications of introgression in nature remains scarce. The widespread hybrid zones of Darwin's primroses (Primula elatior, Primula veris, and Primula vulgaris) provide a unique natural laboratory for studying introgression in flowering plants and the varying permeability of species boundaries. Through analysis of 650 genomes, we provide evidence of an introgressed genomic region likely to confer adaptive advantage in conditions of soil toxicity. We also document unequivocal evidence of chloroplast introgression, an important precursor to species-wide chloroplast capture. Finally, we provide the first evidence that the S-locus supergene, which controls heterostyly in primroses, does not introgress in this clade. Our results contribute novel insights into the adaptive role of introgression and demonstrate the importance of extensive genomic and geographical sampling for illuminating the complex nature of species boundaries.

DOI: https://doi.org/10.1111/nph.19361

New Phytologist

Research

The genomes of Darwin's primroses reveal chromosome-scale adaptive introgression and differential permeability of species boundaries

Rebecca L. Stubbs[1]*, Spyros Theodoridis[2]*, Emiliano Mora-Carrera[1], Barbara Keller[1], Giacomo Potente[1], Narjes Yousefi[1], Paul Jay[3], Étienne Léveillé-Bourret[4], Rimjhim Roy Choudhury[5], Ferhat Celep[6], Judita Kochjarová[7] and Elena Conti[1]

[1]Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, 8008, Switzerland; [2]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, 60325, Germany; [3]Center for GeoGenetics, University of Copenhagen, Copenhagen, 1350, Denmark; [4]Département de Sciences Biologiques, Institut de Recherche en Biologie Végétale (IRBV), Université de Montréal, Montreal, QC, H1X 2B2, Canada; [5]Department of Biology, University of Fribourg, Fribourg, 1700, Switzerland; [6]Department of Biology, Faculty of Arts and Sciences, Kırıkkale University, Kırıkkale, 71450, Turkey; [7]Department of Phytology, Faculty of Forestry, Technical University in Zvolen, Zvolen, 96001, Slovak Republic

Authors for correspondence:
Rebecca L. Stubbs
Email: stubbsrl@gmail.com

Spyros Theodoridis
Email: spyrostheodoridis@gmail.com

Elena Conti
Email: elena.conti@systbot.uzh.ch

Received: *13 June 2023*
Accepted: *5 October 2023*

New Phytologist (2024) **241:** 911–925
doi: 10.1111/nph.19361

Key words: adaptive introgression, genomics, hybridization, Primula, whole-genome resequencing.

Summary

- Introgression is an important source of genetic variation that can determine species adaptation to environmental conditions. Yet, definitive evidence of the genomic and adaptive implications of introgression in nature remains scarce.
- The widespread hybrid zones of Darwin's primroses (*Primula elatior*, *Primula veris*, and *Primula vulgaris*) provide a unique natural laboratory for studying introgression in flowering plants and the varying permeability of species boundaries.
- Through analysis of 650 genomes, we provide evidence of an introgressed genomic region likely to confer adaptive advantage in conditions of soil toxicity. We also document unequivocal evidence of chloroplast introgression, an important precursor to species-wide chloroplast capture. Finally, we provide the first evidence that the S-locus supergene, which controls heterostyly in primroses, does not introgress in this clade.
- Our results contribute novel insights into the adaptive role of introgression and demonstrate the importance of extensive genomic and geographical sampling for illuminating the complex nature of species boundaries.

Introduction

Hybridization is a prevalent phenomenon in nature affecting various evolutionary processes, such as adaptation, speciation, and extinction, contributing to the emergence and persistence of biodiversity on Earth (Abbott *et al.*, 2013; Taylor & Larson, 2019). There are several potential outcomes when species hybridize. In most cases, hybrids are nonviable or sterile, and the movement of genetic material between hybridizing species is prevented (Runemark *et al.*, 2019). Alternatively, if viable hybrids backcross with either parental species, genetic introgression can occur, indicating permeable species boundaries. The degree of permeability reflects the structure and function of genomic regions underlying reproductive barriers and responses to natural selection (Mallet, 2007; Harrison & Larson, 2014; Runemark *et al.*, 2019). For instance, if alleles are subject to divergent selection, linked to deleterious genes, or contribute to the reproductive isolation of a species, their introgression will be prevented (Mallet, 2007; Runemark *et al.*, 2019). Conversely, if novel alleles in donor species provide a fitness advantage to the recipient species, either directly or by being linked to advantageous alleles, their introgression will be favored (Martin & Jiggins, 2017; Fu *et al.*, 2022). Indeed, introgression, alongside newly arisen mutations and standing variation, is now considered an important source of genetic variation contributing to adaptation, in a process known as adaptive introgression (Suarez-Gonzalez *et al.*, 2018a; Stolle *et al.*, 2022). In the last decade, the availability of whole-genome sequences from natural populations has provided significant insights into the genomic and evolutionary consequences of introgression (Meier *et al.*, 2017; Suarez-Gonzalez *et al.*, 2018a; Stolle *et al.*, 2022), yet definitive evidence from across the tree of life is still scarce. Efficiently studying introgression and the permeability of species

*These authors contributed equally to this work.

boundaries in natural systems requires study groups that have multiple occurrences of hybrid zones, high-quality reference genomes, and extensive geographic sampling, yet, to date, very few taxa meet these requirements (e.g. cichlids, Meier *et al.*, 2018; *Heliconius* butterflies, Martin *et al.*, 2013; Jay *et al.*, 2018; sticklebacks, Ravinet *et al.*, 2018b; Dixon *et al.*, 2019; Ravinet *et al.*, 2021). Study groups that do meet these requirements can be used to identify and compare varied patterns of introgression across the entire genome and can deepen our understanding of genomic regions that are key to the evolution of species (Harrison & Larson, 2014; Payseur & Rieseberg, 2016).

Primula sect. *Primula*, a clade of flowering plants in the family Primulaceae, represents an ideal model to study the evolutionary consequences of hybridization and introgression because of its accumulated knowledge about the patterns of crossability among species and availability of genomic resources (Schmidt-Lebuhn *et al.*, 2012; Keller *et al.*, 2016, 2021; Mora-Carrera *et al.*, 2021; Potente *et al.*, 2022a; Stubbs *et al.*, 2023). Seminal studies of hybridization and interspecific crossability between three species in this section, namely *Primula veris*, *Primula elatior*, and *Primula vulgaris*, including work done by Darwin (1868), established that crosses between *P. vulgaris* and *P. elatior* are the most likely to produce viable seeds, while crosses between *P. veris* and *P. elatior* are the least likely to do so, with crosses between *P. vulgaris* and *P. veris* falling in an intermediate position (Fig. 1a). All three species are widespread and are found in both allopatry and sympatry across Europe, and multiple hybrid zones have been documented for each species pair. In particular, hybrid zones in England have been well-documented and studied for over 150 yr (Darwin, 1868; Valentine, 1947, 1952, 1955; Gurney *et al.*, 2007). These studies have established that within England, *P. veris* is widespread and occurs on a range of soils while *P. elatior* is confined to areas where the soils are seasonally waterlogged and have high concentrations of ferrous ions (Martin, 1968; Taylor & Woodell, 2008; Brys & Jacquemyn, 2009). By contrast, *P. vulgaris* thrives on drier soils and the levels of ferrous ions tolerated by *P. elatior* are toxic to *P. vulgaris* plants (Jacquemyn *et al.*, 2009). Yet, both *P. vulgaris* and *P. elatior* occur in a hybrid zone in Buff Wood, near Hatley, England (Fig. 1b, EN4), where soil saturation and iron levels resemble the habitat of *P. elatior* (Martin, 1968). The numerous observations in this region of the distribution and ecology of these two species generated open questions on the adaptive mechanisms that allow *P. vulgaris* to occupy the otherwise unsuitable edaphic habitats of *P. elatior* in Buff Wood (Valentine, 1947; Martin, 1968; Taylor & Woodell, 2008).

Most *Primula* species produce two floral morphs at equal frequencies, that is S-morphs (thrums) that have tall anthers and short stigmas, and L-morphs (pins) that have short anthers and tall stigmas. This floral polymorphism, called heterostyly, promotes outcrossing and is controlled by the S-locus supergene, a 278-kb nonrecombining region containing five genes (*CCM*^{T}, *GLO*^{T}, *CYP*^{T}, *PUM*^{T}, and *KFB*^{T}) that is hemizygous in S-morphs and absent in L-morphs (Huu *et al.*, 2016; Li *et al.*, 2016). Acting as a prezygotic barrier to self- and intramorph pollination, the S-locus supergene is expected to be under balancing selection resulting from disassortative mating (Thompson & Jiggins, 2014; Barrett, 2019; Gutiérrez-Valencia *et al.*, 2021). Previous studies supported the hypothesis that supergenes controlling mating strategies exhibit low levels of introgression, thus contributing to reproductive isolation and species divergence in butterflies (Martin *et al.*, 2013), *Drosophila* (Garrigan *et al.*, 2012), flycatchers (Ellegren *et al.*, 2012), and hominins (Sankararaman *et al.*, 2014). In contrast to these examples, Stölting *et al.* (2013) detected high levels of introgression across incipient sex chromosomes between *Populus* species. Previous analyses of reproductive barriers in *Primula* sect. *Primula* suggest that heterostyly may not strongly contribute to reproductive isolation and that the strength of reproductive barriers may be influenced by which species and morphs serve as male or female parents (Keller *et al.*, 2016, 2021). Specifically, intramorph incompatibilities are weakened when S-morph *P. elatior* plants provide the pollen in *P. vulgaris* × *P. elatior* crosses (Keller *et al.*, 2016). Moreover, in a *P. vulgaris* × *P. veris* hybrid zone, 11 and 8 F1 hybrids had L- and S-morph *P. vulgaris* fathers, respectively, when self- and intramorph incompatibility mechanisms are maintained across species boundaries (Keller *et al.*, 2021). Thus, an open question remains as to whether the heterostyly supergene moves across species boundaries or not.

Similarly to supergenes, the plastid genome is ideal for investigating introgression of nonrecombining genomic regions (Schwander *et al.*, 2014). Introgression of the chloroplast between species, or 'chloroplast capture', has been proposed as a primary cause of cytonuclear discordance (i.e. inconsistencies between nuclear and cytoplasmic phylogenies) and can occur in the absence of nuclear introgression between the hybridizing species (Rieseberg & Soltis, 1991; Tsitrone *et al.*, 2003). Chloroplast capture has been documented in several plant systems and is proposed to have occurred in *Primula* (Guggisberg *et al.*, 2009; Stubbs *et al.*, 2023), yet has never been documented on a genomic scale.

Here, we used six allopatric populations and nine hybrid zones of three *Primula* species to generate whole-genome resequencing (WGR) data from 650 individuals and elucidate the complex evolutionary dynamics of hybridization and introgression. In this study, we ask the following: (1) Do genomic data support previous observations on the patterns of crossability and the porosity of species boundaries? (2) Are there shared introgression patterns, potentially adaptive, across the hybrid zones of the three species? (3) Do nonrecombining genomic regions such as the S-locus supergene and the plastid genome introgress? Overall, we provide the first extensive genomic evidence that unequivocally identifies patterns of hybridization and introgression in Darwin's primroses, and our results illuminate the complex nature of permeable species boundaries across space and taxa.

Materials and Methods

DNA extraction and whole-genome resequencing

For *Primula veris* L., *P. vulgaris* Huds., and *P. elatior* (L.) Hill and their hybrids (Supporting Information Methods S1; Tables S1, S2), DNA was extracted using the Maxwell Rapid

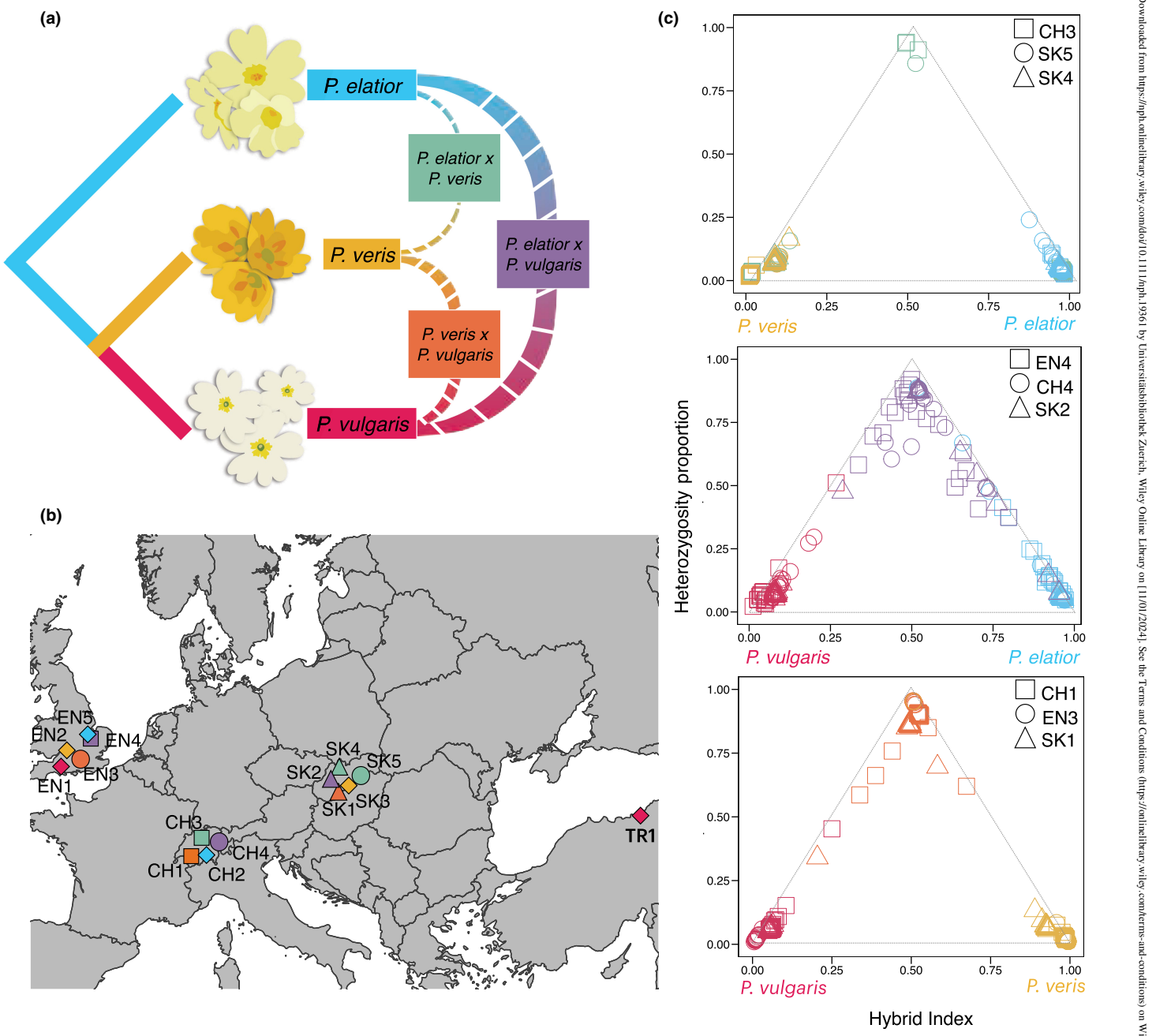

Fig. 1 Overview of crossability, sampling locations, and genomic admixture of *Primula* species. Pure individuals are indicated with primary colors, that is *P. elatior* (blue), *P. veris* (yellow), *P. vulgaris* (red), hybrids are indicated with secondary colors, that is *P. elatior* × *P. veris* hybrids (green), *P. veris* × *P. vulgaris* hybrids (orange), and *P. elatior* × *P. vulgaris* hybrids (purple). (a) Schematic phylogeny (based on Stubbs *et al*., 2023) showing relationships between the three species (solid lines) and described previously degrees of interspecific crossability (dashed lines). Thicker lines between species pairs represent higher crossability. (b) Sampling locations used in this study. Each hybrid zone, designated with the associated hybrid's color, includes two species and their hybrid. Shapes for hybrid zones correspond to (c), while allopatric populations are diamonds. (c) Triangle plots show hybrid index (*x*-axis) and heterozygosity proportion (*y*-axis) for all individuals from three hybrid zones of each species pair. Color corresponds to phenotypic designation as either parental species or hybrid. Parental genotypes have hybrid indices close to 0 or 1 and low heterozygosity, F1 hybrid genotypes have high heterozygosity and intermediate hybrid indices (top-center of each plot), and post-F1 hybrids (F_{1+} and backcrosses) have lower heterozygosity and intermediate hybrid indices.

Sample Concentrator 48 and Maxwell RSC PureFood GMO and Authentication Kit. The manufacturer's instructions were followed with the exclusion of Proteinase K. Library preparation and WGR were conducted at Rapid Genomics (Gainesville, FL, USA; rapid-genomics.com). Samples were sequenced on the Illumina NovaSeq 6000 S4 (paired-end 150-bp reads) to an average depth of 20×. Reading mapping and variant calling followed Stubbs *et al.* (2023), and details can be found in Methods S2.

Phylogenomic and population structure analyses

To confirm our phenotypic assignments, investigate cytonuclear discordance, and test for S-locus and chloroplast introgression, we constructed maximum likelihood phylogenetic trees for the nuclear and chloroplast genomes and the S-locus (Methods S3). For each alignment, ModelFinder (Kalyaanamoorthy *et al.*, 2017) was used to select the best nucleotide substitution model (adjusted for ascertainment bias using -m MFP + ASC) and otherwise default parameters were used (Table S3). IQ-tree v.2.1.2 (Nguyen *et al.*, 2015) was chosen to estimate each phylogeny. We generated statistical support for branches with 1000 ultrafast bootstraps (Minh *et al.*, 2013). Identical methods were followed for the plastome and S-locus, with the exception that we used the entirety of those regions in each analysis (Methods S3).

Population genetic structure was assessed to identify shared genetic polymorphisms among individuals. These analyses were performed nine times for all nine hybrid zones. A principal component analysis (PCA) was performed for allele frequencies following Ravinet *et al.* (2018a) using Plink and visualized in R v.3.6.3 (R Core Team, 2020) using ggplot2 (Wickham, 2016). We used Admixture v.1.3 (Alexander & Lange, 2011) as a model-based analysis of population structure. Scripts to visualize the Admixture results are available at https://github.com/speciationgenomics/scripts (Meier *et al.*, 2017). Additional details for both the phylogenomic and population structure analyses are available in Methods S3.

Classification of parental and hybrid individuals and detection of introgression

We calculated the proportion of ancestry and the interspecific heterozygosity for all individuals in hybrid zones to further determine the extent of hybridization in each sample (Table S4). We followed the methods of Liu *et al.* (2022) and used python scripts available from https://github.com/jingwanglab/Populus-Genomic-Consequences-of-Hybridization. Further details can be found in Methods S4.

To identify genomic regions subjected to introgression and quantify the average extent of introgression, we calculated *D*-statistics, admixture ratios, and a combination of population genomic summary statistics (F_{ST}, d_{XY}, π) for each parental taxa in hybrid zones. Dsuite (Malinsky *et al.*, 2021) was used to calculate f_d, a modified *D*-statistic that is more sensitive to introgression in smaller windows and better able to detect introgression compared with traditional ABBA–BABA tests (Martin *et al.*, 2015). For this analysis, we used sliding windows consisting of 1000 sites with an overlap of 100 sites. The f_4 admixture ratio, or f_4-ratio, was also calculated for the entire genome and for every chromosome in Dsuite. For both introgression analyses, we set up the tests to focus on introgression of one parental taxa into the other (and vice versa) in sympatry. For example, to investigate introgression in *P. vulgaris* from *P. elatior* in the hybrid zone EN4 we tested the following: (((allopatric *P. vulgaris* EN1, sympatric *P. vulgaris* EN4), sympatric *P. elatior* EN4), outgroup). All tested combinations are available in Fig. S1. The F_{ST}, d_{XY}, and π analyses were performed using the python script 'popgenWindows.py' (available from https://github.com/simonhmartin/genomics_general) (Martin *et al.*, 2015) in nonoverlapping 50-kb windows across the genome. For the F_{ST} and d_{XY} analyses, we compared the target species from the hybrid zone with the same species from an allopatric population. For example, allopatric *P. vulgaris* EN1 and sympatric *P. vulgaris* EN4 (all combinations in Fig. S2). Additionally, we merged both allopatric populations for each species and calculated d_{XY} between species, resulting in three pairwise species comparisons in total. To provide further insight into the patterns of inheritance in the S-locus, we assessed d_{XY} and π using the same methods as above, but we used nonoverlapping 10 kb windows across the region.

Additionally, we visualized patterns of ancestry in hybrid species using 'ancestry painting'. This tool aids in creating initial hypotheses and visualizing patterns of hybridization and introgression by coloring chromosomes based solely on the genotypes carried at sites that are fixed ($F_{ST} = 1.0$) between the presumed parental species. This was performed with scripts available at https://github.com/mmatschiner/tutorials/tree/master/analysis_of_introgression_with_snp_data. The genotypes of all individuals were visualized for each of these fixed sites. We ran this analysis for both phenotypic hybrids and parental taxa from hybrid zones with the expectation that F1 hybrids should be heterozygous at most sites and backcrossed hybrids should show a mix of homozygosity and heterozygosity. Due to the possibility of ancestral hybridization and introgression in the parental taxa in hybrid zones, we used samples of each species from the 'pure' allopatric populations to determine fixed sites.

Demographic inference

To determine the direction of gene flow, estimate of migration rate, and time of hybridization for introgressing populations (i.e. *P. elatior*–*P. vulgaris* and *P. veris*–*P. vulgaris*), we performed coalescent simulation using fastsimcoal2 v.2.7 (Excoffier *et al.*, 2021). The joint two-dimensional site frequency spectrum (2D-SFS) of both parental species in each hybrid zone used as input for fastsimcoal2 was estimated with easySFS (https://github.com/isaacovercast/easySFS). As our primary interest was direction and relative age of hybridization, we tested four models: (1) no gene flow, (2) equal gene flow, (3) stronger gene flow in the direction of Species A, and (4) stronger gene flow in the direction of Species B. Additional information is available in Methods S5.

Structure and composition of the introgressed region in EN4

Genotype at the introgressed region was assessed using PCA and computed using scripts available from https://github.com/PaulYannJay/SuperInfer and based on Jay *et al.* (2021). To validate these results, putative inversions were also identified in this region using the program Smoove v.0.2.8 (https://github.com/

brentp/smoove) which integrates LUMPY (Layer *et al.*, 2014) and other associated tools into one package to identify structural variants (Methods S6).

We estimated the strength of positive and purifying selection through calculating the ratio of nonsynonymous to synonymous polymorphisms (pN/pS) and the ratio of nonsynonymous to synonymous substitutions (dN/dS). SNPEFF v.5.0c (Cingolani *et al.*, 2012) was used with default parameters to annotate *P. vulgaris* and *P. elatior* samples homozygous for the introgressed. We used the script CountMutSubSlidingGeneV2.pl available from https://github.com/PaulYannJay/Mutation-load-analysis/ to count the number of nonsynonymous and synonymous mutations and to obtain pN/pS and dN/dS ratios. We compared the dN/dS of genes between the introgressed region and (1) the rest of chromosome 4 and (2) all other chromosomes using the Welch two-sample *t*-test. Furthermore, we calculated pN/pS and dN/dS for the entire introgressed region and compared these values using the McDonald–Kreitman test (MKT; McDonald & Kreitman, 1991). Additional details are in Methods S6.

The R package CLUSTERPROFILER v.4.4.4 (Wu *et al.*, 2021) was used to identify Gene Ontology (GO) terms enriched in the introgressed region in *P. vulgaris* EN4 compared with the rest of the genome, using the functional annotation of *P. veris* from Potente *et al.* (2022b). *P*-value correction was performed using the Benjamini–Hochberg method, and the GO terms with $P \leq 0.05$ were considered significantly enriched. Fold enrichment was obtained through comparing the background frequency of total genes annotated to that term in the designated species to the sample frequency representing the number of genes inputted that fall under the same term. Significant terms associated with the gene set were obtained along with the degree of fold enrichment, *P*-value, and gene counts linked with that term. Fold enrichment was used to indicate how genes of a certain GO category are over-represented, while the *P*-value reflects the chance of observing that number of genes for a specific term (Dalmer & Clugston, 2019). The heatmap generated by CLUSERPROFILER was used to group terms into categories based on genes shared by each term (Fig. S3).

Results

Nuclear phylogenomic and population structure analyses confirm phenotypic assignments and previous hypotheses of permeability of species boundaries

We collected and performed WGR for 530 individuals sampled from three hybrid zones of each species pair (i.e. *P. elatior*–*P. vulgaris*, *P. elatior*–*P. veris*, and *P. veris*–*P. vulgaris*), and 120 individuals from two allopatric populations for each of the three species (Fig. 1a,b; Tables S1, S2). Overall, the results of the nuclear phylogenetic analyses suggest the vast majority of samples were correctly identified in the field as parental or hybrid taxa based on their phenotypic characters (Fig. S4). The nuclear phylogenetic results were supported by the population structure analyses (Figs S5, S6). The population clustering analysis also recovered population structure within all three widespread species, with populations of each species being composed of one to three clusters (Fig. S6).

We estimated the hybrid indices and heterozygosity proportion for all individuals in each hybrid zone using > 1.3 million SNPs per species pair (Fig. 1c; Table S5). Aligning with expectations, the hybrid zones of *P. elatior*–*P. veris* had few hybrids, as shown by the few individuals with both high heterozygous proportions and intermediate hybrid indices (Fig. 1c). Moreover, *P. elatior*–*P. vulgaris* hybrid zones had the most hybrids, while *P. veris*–*P. vulgaris* hybrid zones had the second most hybrid samples. Many individuals in *P. elatior*–*P. vulgaris* hybrid zones had intermediate proportions of heterozygous sites and hybrid indices closer to 1 or 0, indicative of F_{1+} hybrids and backcrosses (Fig. 1c).

Our demographic inference using coalescent simulations with FASTSIMCOAL2 with introgressing populations showed that in both *P. elatior*–*P. vulgaris* and *P. veris*–*P. vulgaris* hybrid zones, *P. vulgaris* was always the recipient of higher levels of gene flow compared with its sympatric species (Fig. S7; Table S6). With the exception of one population (CH1), migration rates and timing for the initiation of gene flow were older in *P. elatior*–*P. vulgaris* compared with *P. veris*–*P. vulgaris* (Table S6), indicating a potential for more frequent gene flow in *P. elatior*–*P. vulgaris* hybrid zones.

Population genomics and signatures of hybridization and introgression

The mean genetic diversity (π) was highest in *P. elatior* (0.011 ∓ 0.007), lowest in *P. veris* (0.006 ∓ 0.004), and intermediate in *P. vulgaris* (0.010 ∓ 0.007; Table S7). The hybrid samples had the highest nucleotide diversity (*P. elatior* × *P. vulgaris*: 0.018 ∓ 0.007; *P. veris* × *P. vulgaris*: 0.016 ∓ 0.006). The genome-wide average of absolute divergence (d_{XY}) followed an identical pattern, with *P. elatior* having the highest d_{XY} (0.012 ∓ 0.007), *P. veris* having the lowest d_{XY} (0.007 ∓ 0.004), and *P. vulgaris* being intermediate (0.011 ∓ 0.007; Table S8). The d_{XY} was greater between species compared with within species. The d_{XY} of *P. elatior* vs *P. vulgaris* was 0.024 ∓ 0.008, the d_{XY} of *P. veris* vs *P. vulgaris* was 0.024 ∓ 0.007, and the d_{XY} of *P. elatior* vs *P. veris* was 0.023 ∓ 0.007 (Table S8).

To compare genomic signatures of introgression within and among species pairs and to assess whether potentially adaptive alleles were introgressed, we analyzed parental taxa in hybrid zones. We employed the f_4-ratio statistic, which estimates the proportion of admixture in a population. Overall, the average f_4-ratio was lowest in *P. veris*–*P. elatior* hybrid zones and highest in *P. elatior*–*P. vulgaris* hybrid zones, with *P. veris*–*P. vulgaris* being intermediate (Fig. 2a; Table S9). *Primula vulgaris* had higher proportions of foreign ancestry (i.e. higher f_4-ratio) when compared to either *P. elatior* or *P. veris* in all hybrid zones ($P < 0.05$, $z > 3$; Fig. 2a). In other words, *P. vulgaris* was more likely to be the recipient of introgression compared with the other two species. Notably, Chromosome 4 in the hybrid zone EN4 population of *P. vulgaris* was found to have an f_4-ratio almost fourfold

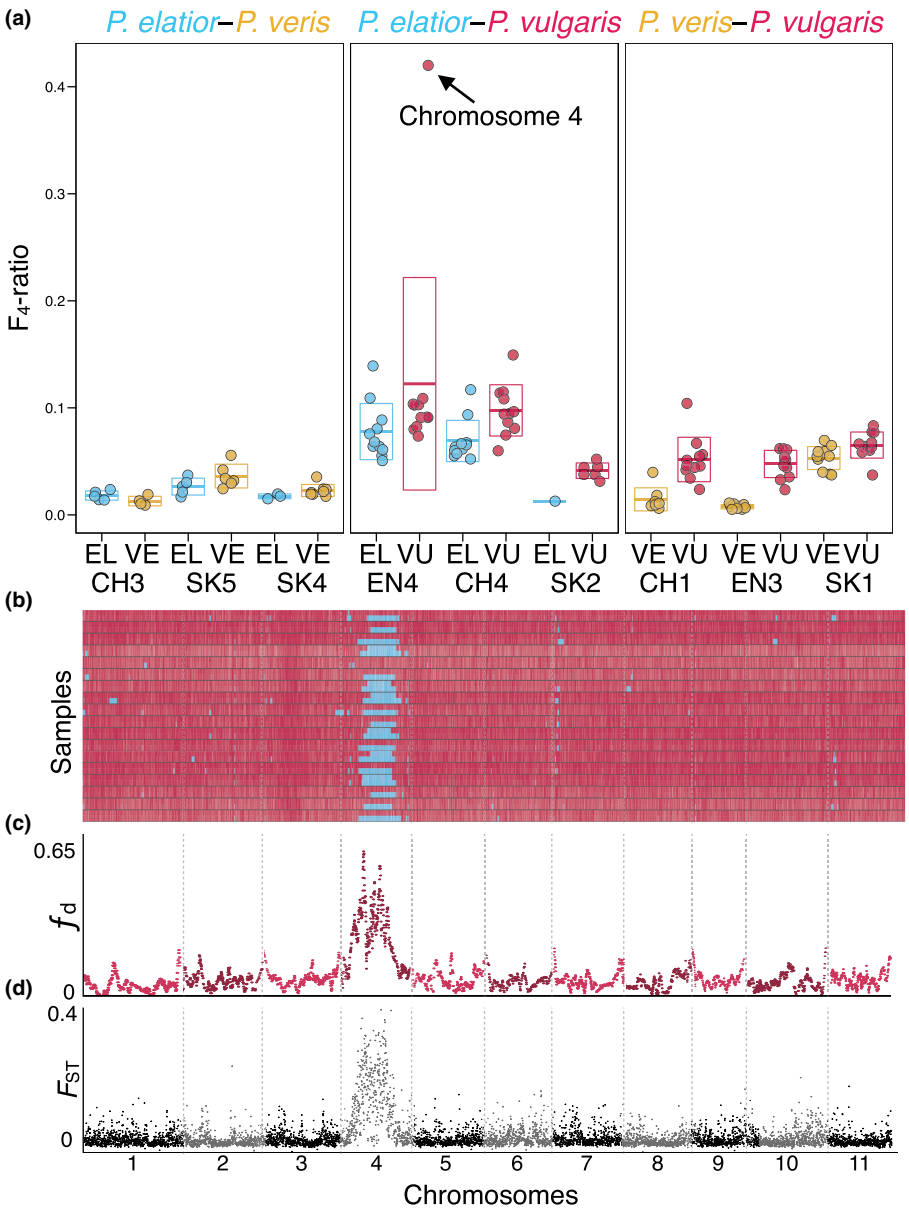

Fig. 2 Introgression and hybridization are asymmetrical and unequal between and within species pairs in hybrid zones. (a) Each panel represents a species pair collected in a hybrid zone and includes three populations. Circles represent a chromosome with a significant ($P < 0.05$ and $z > 3$) proportion of foreign ancestry, as measured by the f_4-ratio for each parental species. The arrow denotes Chromosome 4 from the *Primula vulgaris* samples in EN4. (b) Ancestry painting of alleles fixed at different sites across all chromosomes in *Primula vulgaris* individuals from *P. elatior*–*P. vulgaris* hybrid zone EN4. Blue denotes alleles shared with allopatric *P. elatior*, red denotes alleles shared with allopatric *P. vulgaris*. (c) f_d across the genome based on the relationships (((allopatric *P. vulgaris* TR1, sympatric *P. vulgaris* EN4), sympatric *P. elatior* EN4), outgroup). (d) F_{ST} across the genome between allopatric *P. vulgaris* EN1 and sympatric *P. vulgaris* EN4. EL, *P. elatior*; VE, *P. veris*, VU, *P. vulgaris*.

that of any other chromosome in any other population investigated (Fig. 2a, denoted with arrow).

We used ancestry painting to visualize heterozygosity levels and fixation of parental allele frequencies for both phenotypic parental species and hybrid individuals from each hybrid zone (Fig. S8; Runemark *et al.*, 2018). For hybrids, we confirmed them as being F_1 hybrids if they were almost entirely heterozygous at sites fixed in the parental species, while post-F_1 hybrids (F_{1+} hybrids and backcrosses) had heterozygous and homozygous regions interspersed. Of the 18 populations of parental species collected in nine hybrid zones, six populations included individuals with chromosome painting patterns indicative of multiple generations of backcrossing (Fig. S8). Generally, the introgressed regions were randomly dispersed throughout the genomes (Fig. S8). However, in the *P. elatior*–*P. vulgaris* EN4 hybrid zone there was a genomic region introgressed from *P. elatior* to *P. vulgaris* in Chromosome 4 and present in 19 of 20 phenotypic *P. vulgaris* plants (Fig. 2a,b).

The f_d analysis, which quantifies the average extent of introgression across a genome, was used to identify statistically supported genomic regions with higher levels of introgression in all populations (Fig. S1). The f_d analysis strongly supported an introgressed region in *P. vulgaris* EN4 that contains the pericentromeric region of Chromosome 4 (Fig. 2c). We also estimated F_{ST}, a relative measure of genetic differentiation that can identify genomic regions putatively under selection, between the same species in hybrid zones and allopatric populations. The same region in Chromosome 4 from *P. vulgaris* EN4 that was identified as being introgressed was strongly differentiated in comparison with allopatric *P. vulgaris* EN1 (Fig. 2d; genome-wide mean $F_{ST} = 0.022 \pm 0.04$, Chromosome 4 mean $F_{ST} = 0.168 \pm 0.088$). Genome-wide average d_{XY} showed the

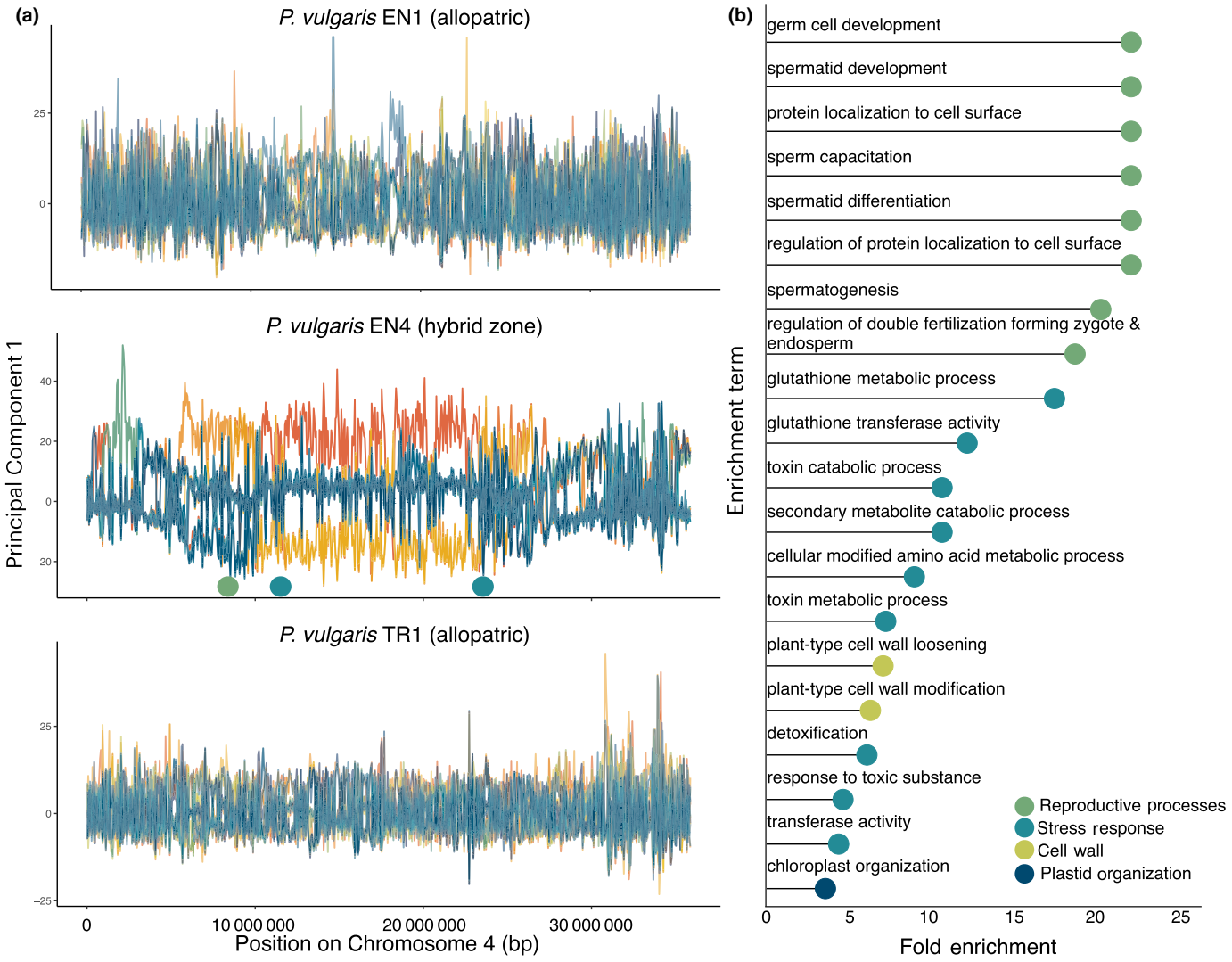

Fig. 3 Characterization of Chromosome 4 from *Primula vulgaris*. (a) Sliding-window principal component analysis (PCA) comparing Chromosome 4 of three populations of *P. vulgaris*. Each colored line represents the variation in the position of a specimen on the first PCA axis along chromosome 4. Within EN4 (hybrid zone of *P. elatior*–*P. vulgaris*, middle plot), individual genomes are characterized by one of three genotypes for the introgressed region spanning *c.* 14–23 Mb: homozygous for the introgression (bottom, yellow lines), heterozygous (middle, blue lines), and homozygous for the ancestral arrangement (top, red lines). Circles along the *x*-axis in the middle plot show approximate location of the genes corresponding to reproductive processes (green circle) and stress responses (teal circles). (b) Gene Ontologies for enriched genes in the introgressed region of chromosome 4 of *P. vulgaris* EN4 ordered by fold enrichment. Circle color indicates the category of each term.

same pattern as F_{ST} with Chromosome 4 having an elevated d_{XY} (0.019 ± 01) compared with the rest of the genome (0.009 ± 0.008; Table S8).

Inferring the structure and function of the introgressed region in *P. vulgaris* in Buff Wood

The 50-kb sliding-window PCA of Chromosome 4 of *P. vulgaris* EN4 samples clustered individuals into three genotypic groups (Fig. 3a), indicating a large, *c.* 17 Mb, nonrecombining region, encompassing 1066 genes. Within *P. vulgaris* samples homozygous for the introgressed region, we detected a significantly higher (P-value < 0.05) average dN/dS and pN/pS ratio in genes in the introgressed region (Tables S10, S11). Furthermore, when comparing the entire introgressed region in both *P. vulgaris* and *P. elatior*, all three tests for selection were greater than one (*P. vulgaris*: dN/dS = 1.512, pN/pS = 1.940, and MKT = 1.282; *P. elatior*: dN/dS = 1.450, pN/pS = 2.006, and MKT = 1.383; Table S12). The structural variant analysis characterized this nonrecombining region as an inversion (Table S13). We investigated whether the genes in the putative inversion were enriched for specific biological or molecular functions. The GO terms with the highest fold enrichment corresponded to sperm and processes involved in reproduction (Figs 3b, S3). This included GO terms such as, 'spermatid development', 'spermatogenesis', and 'germ cell development'. These genes are found close together at the beginning of the introgressed region (Fig. 3a; Table S14). Notably, we also found many GO terms related to processing toxins

and reacting to environmental stressors, supporting the hypothesis of adaptive introgression of genes related to waterlogged-induced soil toxicity (Martin, 1968; Kim *et al.*, 2018). The GO terms included the following: 'glutathione metabolic process', 'glutathione transferase activity', 'toxin catabolic process', 'toxin metabolic process', and 'detoxification'. The genes supporting these biological functions are found after the reproductive processes genes and are in two clusters separated by 11 Mb (Fig. 3a; Table S14). Additionally, these genes all belong to the same orthogroup and are classified as belonging to the glutathione s-transferases (GST) superfamily (Potente *et al.*, 2022a,b).

Introgression of nonrecombining genomic regions

We investigated our data for evidence of chloroplast introgression by building phylogenies from the complete chloroplast genome and comparing them with the nuclear genome phylogenies (Fig. 4). If chloroplast introgression had occurred between the three species, we would expect the phenotype and nuclear genotype of an individual to match one species, while the chloroplast genome would be most closely related to a different species. Within *P. elatior*–*P. vulgaris* hybrid zones, of 45 *P. elatior* individuals, eight had a chloroplast aligning with *P. vulgaris* (Fig. 4, blue asterisks). There were no instances of *P. vulgaris* plants with *P. elatior* chloroplast genomes in these hybrid zones. Notably, seven *P. vulgaris* specimens from a *P veris*–*P. vulgaris* hybrid zone in Switzerland (CH1) had chloroplast genomes most similar to *P. elatior* (Fig. 4, red asterisks). Finally, within *P. elatior*–*P. veris* hybrid zones, there were no individuals that showed any evidence of chloroplast introgression.

Presuming maternal inheritance of the chloroplast genome, we can ask whether there is bias in the maternal contribution to hybrids. This was determined by using the rooted plastome phylogeny to first establish clades corresponding to each pure species, and then identifying in which clade the hybrid samples occurred (Fig. 4; Table 1). Due to the limited number of *P. elatior* × *P. veris* hybrids, no major patterns can be deduced. For *P. veris* × *P. vulgaris* hybrids, the majority (98%) of hybrids were recovered in a clade with *P. veris*, suggesting that *P. veris* provided the chloroplast for most hybrid samples in these populations and is more likely to be the maternal parent (Table 1). By contrast, in *P. elatior* × *P. vulgaris* hybrid populations, more hybrids samples (76% of hybrids) were recovered in a clade with *P. vulgaris* compared with *P. elatior* (24% of hybrids), suggesting that *P. vulgaris* is more likely to be the maternal parent for hybrid samples in these populations.

Applying a similar methodology to that used in investigating chloroplast introgression, we can ask whether there are S-morph specimens, which contain the S-locus, with a phenotype and nuclear genotype aligned with one species, but an S-locus that is genetically similar to a different species. Unlike for the plastome, we do not see any evidence of introgression of the S-locus (Fig. 4). We further investigated the population genomics of the S-locus. Overall, π in the S-locus showed similar values for all three species (*P. elatior*: 0.005 ∓ 0.003, *P. vulgaris*: 0.005 ∓ 0.004, *P. veris*: 0.002 ∓ 0.001; Table S15). However, as expected due to hemizygosity, π was smaller in the S-locus compared with the rest of the genome (Tables S8, S15). The d_{XY} of the S-locus for each species were also very similar, but the averages showed a different pattern, with *P. vulgaris* having the highest d_{XY} (0.007 ∓ 0.005), followed by *P. elatior* (0.006 ∓ 0.003), and then *P. veris* (0.002 ∓ 0.001; Table S16).

We were also interested in asking whether there were morph-dependent asymmetries of reproductive isolation. This was determined using the S-locus and chloroplast phylogenies (Fig. 4). For *P. veris* × *P. vulgaris* hybrids, 79% of S-morph hybrids received the S-locus from *P. veris* while 21% of S-morph hybrids received the S-locus from *P. vulgaris* (Fig. 4; Table 1). Aligning with the majority of our wild collected samples having *P. veris* as the maternal parent, our results suggest that *P. veris* is more likely to be the S-morph, seed-producing parent when it is fertilized by *P. vulgaris* pollen (Table 1). For *P. elatior* × *P. vulgaris* hybrids, which species served as the S-morph was almost equally split, with 52% and 48% of S-morph hybrids receiving the S-locus from *P. elatior* and *P. vulgaris*, respectively. However, S-morph *P. elatior* parents were more likely to be the pollen-contributing parent, while S-morph *P. vulgaris* plants were more likely to be the seed-producing parent (Table 1).

Discussion

The *P. veris*–*P. elatior*–*P. vulgaris* species complex provides a unique system for studying hybridization and genetic introgression. These three species, which have intrigued biologists for centuries, have varied levels of crossability, reproductive isolation, and natural hybridization. We utilized this system as a natural laboratory and provided the first genomic-scale study of hybridization within *Primula* through analysis of 650 complete genomes. We discovered a large, inverted, introgressed region in one hybrid zone that apparently provides an adaptive advantage in conditions of soil toxicity. We additionally teased apart the complexities of species- and morph-dependent asymmetries in interspecific gene flow, presented unequivocal evidence of chloroplast introgression, and provided the first evidence that the S-locus supergene does not introgress.

Genomics of hybridization in *P. veris*, *P. elatior*, and *P. vulgaris*

Hybridization in *P. veris*, *P. elatior*, and *P. vulgaris* has been extensively studied for over a century (Darwin, 1868; Valentine, 1947, 1952, 1955; Keller *et al.*, 2016, 2021). Nevertheless, a genome-wide study of these species focusing on the consequences of differential genome permeability to introgression has not been undertaken. Here, we provide the first extensive genomic evidence confirming and enriching the predictions of crossability first made by Darwin and corroborated by over 150 yr of research (Darwin, 1868; Valentine, 1947, 1952, 1955; Gurney *et al.*, 2007; Keller *et al.*, 2016, 2021; Fig. 1a). Specifically, demographic analysis (Fig. S7), *D*-statistics (Figs 2a, S1), and analyses classifying parental and hybrid individuals (Figs 1c, S8) confirmed the following: *P. vulgaris* has the most porous species

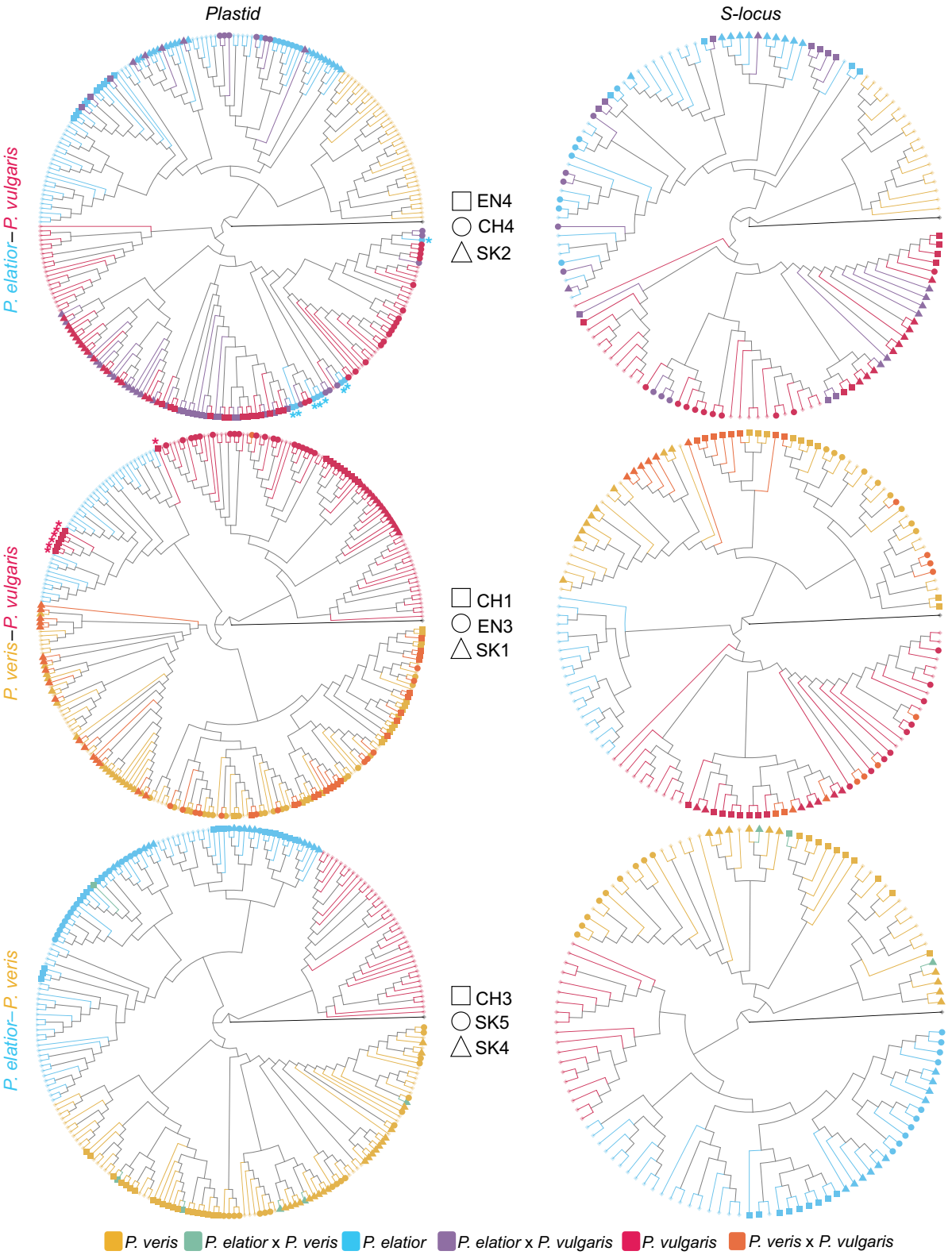

Fig. 4 Plastid and S-locus phylogenies for *Primula veris*, *Primula elatior*, *Primula vulgaris* and their hybrid species. Rooted cladograms with samples from allopatric populations (translucent, small squares) and hybrid zones (opaque, shapes correspond to Fig. 1). The same allopatric samples are used in all phylogenies. Individuals are colored according to nuclear and phenotypic identity and correspond to Fig. 1. Individuals with mismatched chloroplast and nuclear identity are signaled with asterisks. S-locus phylogenies contain only S-morph plants.

New Phytologist

Table 1 Origin of the chloroplast genome and S-locus in hybrid samples of three species of *Primula* sect. *Primula*.

Chloroplast

Hybrid	F_1/BC	*P. elatior*	*P. veris*	*P. vulgaris*	Total
P. elatior × *P. veris*	F_1	1 (25.0%)	3 (75.0%)	–	4
	BC	0	0	–	0
					4
P. veris × *P. vulgaris*	F_1	–	42 (84.0%)	1 (2.0%)	43
	BC	–	7 (14.0%)	0	7
					50
P. elatior × *P. vulgaris*	F_1	0	–	6 (11.1%)	6
	BC	13 (24.1%)	–	35 (64.8%)	48
					54

S-locus

		P. elatior		*P. veris*		*P. vulgaris*		
Hybrid	F_1/BC	♀	♂	♀	♂	♀	♂	Total
P. elatior × *P. veris*	F_1	0	0	1 (100%)	0	–	–	1
	BC	0	0	0	0	–	–	0
								1
P. veris × *P. vulgaris*	F_1	-	-	14 (54.8%)	0	1 (3.8%)	7 (26.9%)	22
	BC	-	-	4 (15.4%)	0	0	0	4
								26
P. elatior × *P. vulgaris*	F_1	0	1 (3.4%)	–	–	1 (3.4%)	0	2
	BC	4 (13.8%)	9 (31.0%)	–	–	10 (34.5%)	4 (13.8%)	27
								29

(Upper panel) Origin of parental chloroplast in phenotypic F_1 (first generation) or BC (backcrossed/post-F_1) hybrids: *P. elatior* × *P. veris* ($n = 4$), *P. veris* × *P. vulgaris* ($n = 50$), and *P. elatior* × *P. veris* ($n = 54$). (Lower panel) Origin of parental S-locus in phenotypic S-morph hybrids (F_1 or BC): *P. elatior* × *P. veris* ($n = 1$), *P. veris* × *P. vulgaris* ($n = 26$), and *P. elatior* × *P. veris* ($n = 29$). In both cases, there were a limited number of *P. elatior* × *P. veris* hybrids. Classification of F_1 and BC hybrids was performed with the hybrid index analysis (see Materials and Methods and Classification of Parental and Hybrid Individuals and Detection of Introgression).

boundaries of the three species, hybridization between *P. elatior* and *P. vulgaris* produces the most backcrossed hybrids, hybridization between *P. veris* and *P. vulgaris* produces mostly F1 hybrids with *P. veris* more likely to be the maternal parent, and hybridization between *P. elatior* and *P. veris* produces very few hybrids. Furthermore, we recovered similar π and d_{XY} for all species. Nucleotide diversity is a reflection of effective population size (Ne; Hague & Routman, 2016), which might suggest a similar effective population size for all three *Primula* species in this study, and the low estimates of d_{XY} may reflect that these three species have experienced limited divergence (Dutoit *et al.*, 2017).

We did recover discrepancies between our analyses and previous work in both species- and morph asymmetries in *P. elatior* × *P. vulgaris* hybrids. Prior work recovered *P. elatior* as more likely to be the maternal parent for *P. elatior* × *P. vulgaris* hybrids (Keller *et al.*, 2016), while our research found that *P. vulgaris* was more likely to be the maternal parent. A likely reason for this discrepancy is that previous research was conducted with nursery-grown, F_1 hybrids, while in our study, the majority of the wild-sampled *P. elatior* × *P. vulgaris* hybrids were post-F_1 hybrids (Fig. 1c; Table 1). It has been shown, albeit rarely, that reproductive isolation in backcrosses via fertile hybrids can be asymmetrical in the opposite direction of the pure parental crosses due to breakdown in reproductive barriers. Thus, fertile F_1 hybrid crossing equally well with both parental species (Moreira-Hernández & Muchhala, 2019). Therefore, it is possible that hybrids in *P. elatior*–*P. vulgaris* hybrid zones backcross equally well with either parental species, effectively erasing the asymmetry found in crosses between pure parental individuals. We argue that the increase in the interspecific crossability between parental species and backcrosses compared with pure parental crosses represents an exciting and underexplored topic that could greatly elucidate the mechanisms involved in reproductive isolation.

Adaptive introgression into *P. vulgaris*

A major finding of our study is that in the hybrid zone located in Buff Wood in England (EN4) we discovered a region of Chromosome 4 in *P. vulgaris* plants that had been introgressed from *P. elatior* (Figs 2b–d, 3). Our genomic analyses suggest that the introgressed region is 25–40% of the chromosome length and contains the pericentromeric region. We found that this region is highly differentiated from the equivalent region of other *P. vulgaris* populations (Figs 2d, S9), which is characteristic of genomic regions that have been the target of selection (Ravinet *et al.*, 2018a). Furthermore, we established that the introgressed region is nonrecombining and is likely an inversion (Fig. 3a;

Table S13). The introgression and inheritance of coadapted alleles can be strengthened by genetic architectures, such as inversions, that maintain desirable combinations of alleles and reduce the production of recombinant genotypes (Thompson & Jiggins, 2014). Thus, it is likely that the inversion recovered in this population preserved a favorable combination of coadapted alleles through suppressed recombination and resulted in the entire genomic region being inherited together (Jay *et al.*, 2021; Stolle *et al.*, 2022). Moreover, the adaptive role of an introgressed region needs to be supported by strong signatures of selection acting on it (Burgarella *et al.*, 2019). In support of this prerequisite, we detected a higher dN/dS ratio in the introgressed region than compared with genes in the rest of the genome (Tables S11, S12), indicating that some genes in the introgressed region may be under positive selection. However, the introgressed region had MKT > 1 in both *P. elatior* and *P. vulgaris*, suggesting that nonsynonymous mutations in this region were segregating as polymorphism rather than fixing in both species (Table S12).

This introgressed region was enriched for genes involved in biological functions that pertained mostly to the categories of reproductive processes and stress responses (Figs 3b, S3). The first category, reproductive processes, is important because hybrid sterility and inviability are common forms of intrinsic barriers of postzygotic isolation (Orr, 1995). Specifically, one of the genetic incompatibilities that cause hybrid dysfunction is abnormal expression of genes involved in spermatogenesis (Mack & Nachman, 2017), and studies have suggested that spermatogenesis is more sensitive to disruption in hybrids compared with oogenesis (Li *et al.*, 2009). Indeed, Keller *et al.* (2016) found that F_1 male sterility was a significant component of reproductive isolation in *P. vulgaris* plants in *P. elatior* × *P. vulgaris* crosses. Therefore, the enrichment of genes in the introgressed region in *P. vulgaris* that control 'spermatid development', 'regulation of double fertilization forming zygote and endosperm', and 'germ cell development' most likely reflect the flow of genetic information from *P. elatior* as a means for backcrossed hybrids to overcome sterility, thus aiding the adaptive introgression of the genomic region.

The second category of GO terms enriched in the introgressed region of *P. vulgaris* related to responses to toxins and environmental stress. Notably, two of the populations we sampled in England, Buff Wood (EN4) and nearby Hayley Wood (EN5), have been the subject of increased interest to botanists and ecologists since the early 20th century (Valentine, 1948; Rackham, 1975; Gurney *et al.*, 2007; Taylor & Woodell, 2008). These two sites are separated by *c.* 2.85 km and comprise almost identical habitats with seasonal waterlogging and high concentrations of ferrous ions. However, both *P. vulgaris* and *P. elatior* grow in Buff Wood, while only *P. elatior* grows in Hayley Wood. In fact, when *P. vulgaris* was experimentally grown in Hayley Wood, the root systems died, an indication of ferrous ion toxicity (Martin, 1968). Therefore, introgression of genes enriched for biological properties such as 'detoxification' and 'toxin metabolic process' is most likely enabling *P. vulgaris* to grow in the otherwise toxic soils of Buff Wood. Furthermore, two GO terms relating to toxicity in the introgressed region are 'glutathione metabolic process' and 'glutathione transferase activity' (Figs 3b, S3). Kim *et al.* (2018) found evidence that increased expression of glutathione transferases mitigates waterlogging stress in soybeans, and multiple studies have shown that genes in the GST superfamily can metabolize an array of toxic exogenous compounds and are responsive to many environmental stressors (Hasanuzzaman *et al.*, 2017; Kumar & Trivedi, 2018). Taken together, we propose that the introgressed region of Chromosome 4 from *P. elatior* to *P. vulgaris* provides a fitness advantage allowing *P. vulgaris* to grow in the iron-rich, waterlogged soils of Buff Wood. We present this population as an ideal case study of adaptive introgression with regard to soil toxicity resulting from waterlogging, a phenomenon that is becoming even more relevant as the frequency and duration of flooding is increasing due to land use and climate change (Ponting *et al.*, 2021). More research is needed to dissect the mechanisms providing the inferred adaptive benefits in this unique natural laboratory.

Introgression of nonrecombining regions

Recent studies have shown that supergene alleles may or may not cross species boundaries (Jay *et al.*, 2018; Dixon *et al.*, 2019; Jeong *et al.*, 2022; Stolle *et al.*, 2022). However, no study has examined whether hybridization and introgression affect the heterostyly supergene. We found no evidence of introgression for the S-locus in any of the species studied (Fig. 4). One possible reason for this is that for alleles to be introgressed they would need to provide either an adaptive advantage or be linked to loci that do confer an advantage in response to environmental conditions (Martin *et al.*, 2013; Suarez-Gonzalez *et al.*, 2018b; Fu *et al.*, 2022). It is possible that the genes residing in the S-locus in primroses are not involved in adaption to environment. Indeed, the origin of the S-locus in Primulaceae predated the origin of *Primula* sect. *Primula* (Potente *et al.*, 2022a) and the five genes forming the S-locus are highly conserved within this section (Potente *et al.*, 2022a; Stubbs *et al.*, 2023). Thus, the introgression of the S-locus from one species into another would not provide any selective advantage. Additionally, it has been observed across multiple systems that introgression is reduced in functionally important regions, implying that selection against minor parent ancestry generates barriers to introgression (Moran *et al.*, 2021). Because our research is limited to *Primula* sect. *Primula*, we cannot rule out the possibility that at deeper phylogenetic scales S-locus introgression may have occurred. Future comparative genomic and population genetic research that includes the S-locus and sampling across *Primula* would be needed to address this question.

Considering our finding with the S-locus, it is of interest to compare the S-locus in *Primula* to other systems where polymorphism is maintained by frequency-dependent selection. One such system is the self-incompatibility S-locus in Brassicaceae. Sporophytic self-incompatibility (SI) in the Brassicaceae plant family is controlled by at least two genes (SCR and SRK) in the nonrecombining S-locus and occurs in multiple genera throughout the family (Edh *et al.*, 2009). Despite both being nonrecombining, the SI S-locus in Brassicaceae differs from the heterostyly S-locus in *Primula* in a few key ways. In Brassicaceae, the SI system is homomorphic and the S-locus has been reported to have

50–60 alleles with complex dominance and recessive relationships among them. It is suggested that negative frequency-dependent selection for the SI S-locus in Brassicaceae may result in a selective advantage for migrants with rare alleles, promoting gene flow and facilitating introgression, thus increasing genetic diversity at the S-locus (Castric *et al.*, 2008; Schierup & Vekemans, 2008; Pickup *et al.*, 2019). By contrast, the SI system in *Primula* is heteromorphic where the number of mating types is kept to two (or three, in tristyly) by floral heteromorphism, thus lacking the selective advantage of the rare allele as in Brassicaceae that would promote introgression. Moreover, research from Potente *et al.* (2022a) documented that there is a higher ratio of nonsynonymous to synonymous substitutions in the S-locus compared with its paralogs indicating relaxed selection in the hemizygous region. The low levels of genetic diversity (π) for the five genes composing the S-locus (Table S15) compared with the rest of the genome (Table S7) is likely a consequence of the S-locus being hemizygous and having a reduced effective population size (Kappel *et al.*, 2017) and the lack of introgression for this region. Altogether, although the S-locus in both Brassicaceae and *Primula* are maintained by balancing selection, the key differences in their evolution and maintenance through natural selection likely affect their likelihood to be introgressed.

In contrast to the S-locus, we did recover evidence of introgression for the plastid genome. Every instance of chloroplast introgression we recovered involved either *P. vulgaris* or *P. elatior*. The most astonishing illustration of this phenomenon was found in a hybrid zone of *P. veris*–*P. vulgaris* (CH1), where seven *P. vulgaris* samples had a chloroplast genome inherited from *P. elatior* (Fig. 4). To our knowledge, there are no extant *P. elatior* plants in this hybrid zone. The phenomenon of a three-species hybrid has been recorded in other *Primula* studies (Taylor & Woodell, 2008), so it is possible that this was once a tri-species hybrid zone. Previous work has proposed chloroplast capture as a cause of cytonuclear discordance in *Primula*, particularly with regard to the contrasting phylogenetic position of different populations of *P. vulgaris* and *P. elatior* between the nuclear and chloroplast phylogenies (Stubbs *et al.*, 2023). Here, we have provided evidence of unilateral introgression of the chloroplast across species boundaries driving the cytonuclear discordance in this section.

Acknowledgements

The authors are very thankful to Carolina Potente for designing the flower graphics in Fig. 1. We are grateful to Marie-Eve Garon-Labrecque, Monika Janisova, and Bryan Drew for fieldwork assistance. We thank The Wildlife Trust for Bedfordshire, Cambridgeshire & Northamptonshire, and specifically Mark Ricketts, for permitting access to Buff Wood. We thank Mark Ravinet and Joana Meier for the guidance with analyses. This work was supported by the Swiss National Science Foundation (grant no: 175556).

Competing interests

None declared.

Author contributions

ST, BK and EC designed the research. RLS, BK, EM-C, NY, EL-B, FC and JK carried out the fieldwork. NY set up the computational infrastructure and provided bioinformatics support. EM-C performed the demographic analysis. GP and RLS analyzed the gene ontology data, and PJ and RLS analyzed the structure of the introgressed region. RRC performed additional analyses. RLS performed the remaining analyses. RLS, ST and EC conceived of the manuscript. All authors interpreted the data and contributed to the writing of the manuscript. RLS and ST contributed equally to this work.

ORCID

Ferhat Celep https://orcid.org/0000-0003-3280-8373
Rimjhim Roy Choudhury https://orcid.org/0000-0002-0499-4124
Elena Conti https://orcid.org/0000-0003-1880-2071
Paul Jay https://orcid.org/0000-0001-5979-1263
Barbara Keller https://orcid.org/0000-0002-7903-8938
Judita Kochjarová https://orcid.org/0000-0002-0554-7879
Étienne Léveillé-Bourret https://orcid.org/0000-0002-0069-0430
Emiliano Mora-Carrera https://orcid.org/0000-0001-8237-4265
Giacomo Potente https://orcid.org/0000-0002-4343-3952
Rebecca L. Stubbs https://orcid.org/0000-0001-7386-2830
Spyros Theodoridis https://orcid.org/0000-0001-5188-7033
Narjes Yousefi https://orcid.org/0000-0001-6292-1516

Data availability

The 650 whole-genome sequences used in this study are deposited in SRA (BioProject PRJNA1023265). Sequence alignments and phylogenies are deposited in the Dryad Digital Repository doi: 10.5061/dryad.wh70rxwtz.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R *et al.* 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229–246.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**: 246.

Barrett SCH. 2019. 'A most complex marriage arrangement': recent advances on heterostyly and unresolved questions. *New Phytologist* **224**: 1051–1067.

Brys R, Jacquemyn H. 2009. Biological flora of the British Isles: *Primula veris* L. *Journal of Ecology* **97**: 581–600.

Burgarella C, Barnaud A, Kane NA, Jankowski F, Scarcelli N, Billot C, Vigouroux Y, Berthouly-Salazar C. 2019. Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. *Frontiers in Plant Science* **10**: 4.

Castric V, Bechsgaard J, Schierup MH, Vekemans X. 2008. Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genetics* **4**: e1000168.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SNPEFF: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**: 80–92.

Dalmer TRA, Clugston RD. 2019. Gene ontology enrichment analysis of congenital diaphragmatic hernia-associated genes. *Pediatric Research* **85**: 13–19.

Darwin C. 1868. On the specific difference between *Primula veris*, Brit. Fl. (var. officinalis of Linn.), *P. vulgaris*, Brit. Fl. (var. acaulis, Linn.), and *P. elatior*, Jacq.; and on the hybrid nature of the common Oxlip. With supplementary remarks on naturally-produced hybrids in the genus verbascum. *Botanical Journal of the Linnean Society* **10**: 437–454.

Dixon G, Kitano J, Kirkpatrick M. 2019. The origin of a new sex chromosome by introgression between two stickleback fishes. *Molecular Biology and Evolution* **36**: 28–38.

Dutoit L, Burri R, Nater A, Mugal CF, Ellegren H. 2017. Genomic distribution and estimation of nucleotide diversity in natural populations: perspectives from the collared flycatcher (*Ficedula albicollis*) genome. *Molecular Ecology Resources* **17**: 586–597.

Edh K, Widén B, Ceplitis A. 2009. The evolution and diversification of S-locus haplotypes in the Brassicaceae family. *Genetics* **181**: 977–984.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A. 2012. The genomic landscape of species divergence in *Ficedula flycatchers*. *Nature* **491**: 756–760.

Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. FASTSIMCOAL2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**: 4882–4885.

Fu R, Zhu Y, Liu Y, Feng Y, Lu R-S, Li Y, Li P, Kremer A, Lascoux M, Chen J. 2022. Genome-wide analyses of introgression between two sympatric Asian oak species. *Nature Ecology & Evolution* **6**: 924–935.

Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton KR, Presgraves DC. 2012. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Research* **22**: 1499–1511.

Guggisberg A, Mansion G, Conti E. 2009. Disentangling reticulate evolution in an arctic-alpine polyploid complex. *Systematic Biology* **58**: 55–73.

Gurney M, Preston CD, Barrett J. 2007. Hybridisation between Oxlip *Primula elatior* (L.) Hill and primrose *P. vulgaris* Hudson, and the identification of their variable hybrid *P.* × *digenea* A. Kerner. *Watsonia* **26**: 239–251.

Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T. 2021. The genomic architecture and evolutionary fates of supergenes. *Genome Biology and Evolution* **13**: evab057.

Hague MTJ, Routman EJ. 2016. Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity* **116**: 92–98.

Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* **105**: 795–809.

Hasanuzzaman M, Nahar K, Anee TI, Fujita M. 2017. Glutathione in plants: biosynthesis and physiological role in environmental stress tolerance. *Physiology and Molecular Biology of Plants* **23**: 249–268.

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.

Jacquemyn H, Endels P, Brys R, Hermy M, Woodell SRJ. 2009. Biological flora of the British Isles: *Primula vulgaris* Huds. (*P. acaulis* (L.) Hill). *Journal of Ecology* **97**: 812–833.

Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M. 2021. Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics* **53**: 288–293.

Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J, Dasmahapatra KK, Joron M. 2018. Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology* **28**: 1839–1845.

Jeong H, Baran NM, Sun D, Chatterjee P, Layman TS, Balakrishnan CN, Maney DL, Yi SV. 2022. Dynamic molecular evolution of a supergene with suppressed recombination in white-throated sparrows. *eLife* **11**: e79387.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. MODELFINDER: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Kappel C, Huu CN, Lenhard M. 2017. A short story gets longer: recent insights into the molecular basis of heterostyly. *Journal of Experimental Botany* **68**: 5719–5730.

Keller B, de Vos JM, Schmidt-Lebuhn AN, Thomson JD, Conti E. 2016. Both morph- and species-dependent asymmetries affect reproductive barriers between heterostylous species. *Ecology and Evolution* **6**: 6223–6244.

Keller B, Ganz R, Mora-Carrera E, Nowak MD, Koutroumpa K, Conti E. 2021. Asymmetries of reproductive isolation are reflected in directionalities of hybridization: integrative evidence on the complexity of species boundaries. *New Phytologist* **229**: 1795–1809.

Kim Y, Seo C-W, Khan AL, Mun B-G, Shahzad R, Ko J-W, Yun B-W, Park S-K, Lee I-J. 2018. Exo-ethylene application mitigates waterlogging stress in soybean (*Glycine max* L.). *BMC Plant Biology* **18**: 254.

Kumar S, Trivedi PK. 2018. Glutathione S-transferases: role in combating abiotic stresses including arsenic detoxification in plants. *Frontiers in Plant Science* **9**: 751.

Layer RM, Chiang C, Quinlan AR, Hall IM. 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology* **15**: R84.

Li J, Cocker JM, Wright J, Webster MA, McMullan M, Dyer S, Swarbreck D, Caccamo M, van Oosterhout C, Gilmartin PM. 2016. Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nature Plants* **2**: 1–7.

Li XC, Barringer BC, Barbash DA. 2009. The pachytene checkpoint and its relationship to evolutionary patterns of polyploidization and hybrid sterility. *Heredity* **102**: 24–30.

Liu S, Zhang L, Sang Y, Lai Q, Zhang X, Jia C, Long Z, Wu J, Ma T, Mao K *et al.* 2022. Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome. *Molecular Biology and Evolution* **39**: msac008.

Mack KL, Nachman MW. 2017. Gene regulation and speciation. *Trends in Genetics* **33**: 68–80.

Malinsky M, Matschiner M, Svardal H. 2021. DSUITE-FAST D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* **21**: 584–595.

Mallet J. 2007. Hybrid speciation. *Nature* **446**: 279–283.

Martin MH. 1968. Conditions affecting the distribution of *Mercurialis Perennis* L. in certain Cambridgeshire Woodlands. *Journal of Ecology* **56**: 777.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**: 1817–1828.

Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular Biology and Evolution* **32**: 244–257.

Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development* **47**: 69–74.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**: 652–654.

Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* **8**: 14363.

Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018. Genomics of parallel ecological speciation in lake Victoria Cichlids. *Molecular Biology and Evolution* **35**: 1489–1506.

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**: 1188–1195.

Mora-Carrera E, Stubbs RL, Keller B, Léveillé-Bourret É, de Vos JM, Szövényi P, Conti E. 2021. Different molecular changes underlie the same phenotypic transition: origins and consequences of independent shifts to homostyly within species. *Molecular Ecology* **00**: 1–18.

Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. 2021. The genomic consequences of hybridization. *eLife* **10**: e69016.

Moreira-Hernández JI, Muchhala N. 2019. Importance of pollinator-mediated interspecific pollen transfer for angiosperm evolution. *Annual Review of Ecology, Evolution, and Systematics* **50**: 191–217.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**: 1805–1813.

Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology* **25**: 2337–2360.

Pickup M, Brandvain Y, Fraïsse C, Yakimowski S, Barton NH, Dixit T, Lexer C, Cereghetti E, Field DL. 2019. Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. *New Phytologist* **224**: 1035–1047.

Ponting J, Kelly TJ, Verhoef A, Watts MJ, Sizmur T. 2021. The impact of increased flooding occurrence on the mobility of potentially toxic elements in floodplain soil – a review. *Science of the Total Environment* **754**: 142040.

Potente G, Léveillé-Bourret É, Yousefi N, Choudhury RR, Keller B, Diop SI, Duijsings D, Pirovano W, Lenhard M, Szövényi P *et al.* 2022a. Comparative genomics elucidates the origin of a supergene controlling floral heteromorphism. *Molecular Biology and Evolution* **39**: msac035.

Potente G, Stubbs RL, Yousefi N, Pirovano W, Szövényi P, Conti E. 2022b. Comparative transcriptomics reveals commonalities and differences in the genetic underpinnings of a floral dimorphism. *Scientific Reports* **12**: 20771.

R Core Team. 2020. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rackham O. 1975. *Hayley wood: its history and ecology*. Cambridge, UK: Cambridgeshire and Isle of Ely Naturalists Trust.

Ravinet M, Elgvin TO, Trier C, Aliabadian M, Gavrilov A, Sætre G-P. 2018a. Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B: Biological Sciences* **285**: 20181246.

Ravinet M, Kume M, Ishikawa A, Kitano J. 2021. Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats. *Journal of Evolutionary Biology* **34**: 114–127.

Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. 2018b. The genomic landscape at a late stage of stickleback speciation: high genomic divergence interspersed by small localized regions of introgression. *PLoS Genetics* **14**: e1007358.

Rieseberg LH, Soltis DE. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**: 65–84.

Runemark A, Trier CN, Eroukhmanoff F, Hermansen JS, Matschiner M, Ravinet M, Elgvin TO, Sætre G-P. 2018. Variation and constraints in hybrid genome formation. *Nature Ecology & Evolution* **2**: 549–556.

Runemark A, Vallejo-Marin M, Meier JI. 2019. Eukaryote hybrid genomes. *PLoS Genetics* **15**: e1008404.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**: 354–357.

Schierup MH, Vekemans X. 2008. Genomic consequences of selection on self-incompatibility genes. *Current Opinion in Plant Biology* **11**: 116–122.

Schmidt-Lebuhn AN, de Vos JM, Keller B, Conti E. 2012. Phylogenetic analysis of *Primula* section *Primula* reveals rampant non-monophyly among morphologically distinct species. *Molecular Phylogenetics and Evolution* **65**: 23–34.

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

Stolle E, Pracana R, López-Osorio F, Priebe MK, Hernández GL, Castillo-Carrillo C, Arias MC, Paris CI, Bollazzi M, Priyam A *et al.* 2022. Recurring adaptive introgression of a supergene variant that determines social organization. *Nature Communications* **13**: 1180.

Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, Lexer C. 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* **22**: 842–855.

Stubbs RL, Theodoridis S, Mora-Carrera E, Keller B, Yousefi N, Potente G, Léveillé-Bourret É, Celep F, Kochjarová J, Tedoradze G *et al.* 2023. Whole-genome analyses disentangle reticulate evolution of primroses in a biodiversity hotspot. *New Phytologist* **237**: 656–671.

Suarez-Gonzalez A, Hefer CA, Lexer C, Douglas CJ, Cronk QCB. 2018a. Introgression from *Populus balsamifera* underlies adaptively significant variation and range boundaries in *P. trichocarpa*. *New Phytologist* **217**: 416–427.

Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018b. Adaptive introgression: a plant perspective. *Biology Letters* **14**: 20170688.

Taylor K, Woodell SRJ. 2008. Biological flora of the British Isles: *Primula elatior* (L.) Hill. *Journal of Ecology* **96**: 1098–1116.

Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution* **3**: 170–177.

Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity* **113**: 1–8.

Tsitrone A, Kirkpatrick M, Levin DA. 2003. A model for chloroplast capture. *Evolution* **57**: 1776–1782.

Valentine D. 1952. Studies in British Primulas. III. Hybridization between *Primula elatior* (L.) Hill and *P. veris* L. *New Phytologist* **50**: 383–399.

Valentine DH. 1947. Studies in British Primulas. I. Hybridization between primrose and Oxlip (*Primula Vulgaris* Huds. and *P. Elatior* Schreb.). *New Phytologist* **46**: 229–253.

Valentine DH. 1948. Studies in British Primulas. II. Ecology and taxonomy of primrose and Oxlip (*Primula vulgaris* Huds. and *P. elatior* Schreb.). *New Phytologist* **47**: 111–130.

Valentine DH. 1955. Studies in British primulas. IV. Hybridization between *Primula vulgaris* Huds. and *P. veris* L. *New Phytologist* **54**: 70–80.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. New York, NY, USA: Springer-Verlag.

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovations* **2**: 100141.

Supporting Information

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

Fig. S1 f_d statistics from sliding windows across the genome for each species in each hybrid zone.

Fig. S2 F_{ST} statistic presented in a Manhattan plot for each hybrid zone.

Fig. S3 Heatmap plot of enriched terms.

Fig. S4 Rooted nuclear phylogeny with allopatric and sympatric samples for every population used in this study.

Fig. S5 Principal component analysis of sympatric and allopatric samples for every population used in this study.

Fig. S6 Population structure analyses conducted in Admixture for each sympatric population.

Fig. S7 Best-supported demographic scenario estimated by the coalescent simulations for hybridizing species pairs and hybrid zones.

Fig. S8 Ancestry painting of alleles for each hybrid zone including phenotypic parent species 1, phenotypic hybrids, and phenotypic parent species 2.

Methods S1 Geographic sampling.

Methods S2 Read mapping and variant calling.

Methods S3 Phylogenomic and population structure analyses.

Methods S4 Classification of parental and hybrid individuals.

Methods S5 Demographic inference.

Methods S6 Structure and composition of the introgressed region in EN4.

Table S1 Samples and species used in this study.

Table S2 Collecting locations.

Table S3 Samples, number of sites, and model selected by ModelFinder for each nuclear phylogenetic analysis.

Table S4 Total number of fixed sites for each population for the Hybrid Index analyses.

Table S5 Heterozygosity proportion and hybrid index of every sample.

Table S6 Output from demographic analysis including migration rate and generations since initiation of hybridization.

Table S7 Mean and standard deviation of nucleotide diversity (π) by species and population.

Table S8 Mean and standard deviation of genome-wide average of absolute divergence (d_{XY}) by species and population.

Table S9 *D*-statistic and *f*4-ratio for every population by chromosome.

Table S10 Ratio of nonsynonymous to synonymous polymorphisms (pN/pS) and substitutions (dN/dS) for genes in Chromosome 4 for *Primula vulgaris* sample homozygous for introgressed region in EN4.

Table S11 Average nonsynonymous to synonymous polymorphisms (pN/pS) and nonsynonymous to synonymous substitutions (dN/dS) for every scaffold in the *Primula vulgaris* genome.

Table S12 Total nonsynonymous to synonymous polymorphisms (pN/pS), nonsynonymous to synonymous substitutions (dN/dS), and McDonald–Kreitman test (MKT) for the entire introgressed region for both *Primula vulgaris* and *Primula elatior* from population EN4.

Table S13 Inversions identified by structural variant caller.

Table S14 Gene ontology information for the genes relating to stress response and reproductive processes.

Table S15 Mean and standard deviation of nucleotide diversity (π) for the S-locus in S-morph samples by species and population.

Table S16 Mean and standard deviation of genome-wide average of absolute divergence (d_{XY}) for the S-locus in S-morph samples by species and population.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.