Original Article, published in Annals of Botany (doi.org/10.1093/aob/mcaa145)

Recent hybrid speciation at the origin of the narrow endemic *Pulmonaria helvetica*

Sandra Grünig, Markus Fischer, Christian Parisod*

Institute of Plant Sciences, University of Bern, Switzerland

*Corresponding author: Institute of Plant Sciences, University of Bern, Altenbergrain 21, CH-3013 Bern; e-mail: christian.parisod@ips.unibe.ch; phone: +41 (0)31 631 4949

Abstract

Background and Aims: Hybridization is known to drive plant speciation through the establishment of homoploid or allopolyploid hybrid species. Here we investigate the origin of *Pulmonaria helvetica*, a narrow endemic species described across a restricted area of Switzerland that was entirely covered by ice during the last glacial maximum. This species presents an original number of chromosomes (2n = 24) and morphological traits suggestive of a hybrid origin.

Methods: We sequenced a plastid locus and 1077 double-digest restriction site associated DNA (ddRAD) loci in 67 individuals from across the distribution range of *P. helvetica* and candidate progenitor species growing in the same area. Assignment of genotypes to main genetic clusters within and among taxa using STRUCTURE tested whether *P. helvetica* represents a genetically differentiated lineage and addressed the hypothesis of its hybrid origin. Comparative ecological modelling further addressed possible niche differentiation among taxa.

Key Results: P. helvetica was highlighted as a genetically homogeneous species distinct from co-occurring taxa. Consistent with a scenario of hybrid speciation, it presented clear evidence of balanced admixture between *P. officinalis* (2n = 16) and *P. mollis* s.l. (2n = 18, 22) that was also highlighted as maternal progenitor based on plastid sequences. Limited genetic structure within the maternal progenitor is consistent with an origin of *P. helvetica* through either homoploid hybridization with considerable karyotype changes or via complex scenarios of allopolyploidy involving a dysploid taxon of *P. mollis* s.l. Comparative niche modelling indicated non-significant ecological differences between *P. helvetica* and its progenitors, supporting intrinsic factors resulting from hybridization as main drivers of speciation.

Conclusions: Hybridization appears as a major process having promoted the postglacial origin of the narrow endemic *P. helvetica*, suggesting hybrid speciation as an effective process that rapidly produces new species under climate changes.

Keywords: ddRAD, dysploidy, ecological modelling, endemic species, homoploid hybridization, speciation, postglacial expansion, *Pulmonaria*

Introduction

Hybridization, considered as crossing between species or genetically divergent populations, may impact the processes of speciation in various ways (Abbott *et al.*, 2013). It can slow or reverse differentiation through gene flow and recombination or it can accelerate speciation via adaptive introgression or the formation of new hybrid taxa (Arnold, 1997; Rieseberg, 1997; Soltis and Soltis, 2009). Hybridization has been involved as a significant driver of speciation in plants, most often in association with duplication of chromosome sets (i.e. allopolyploidy; (Wood *et al.*, 2009). Homoploid hybrid speciation, whereby new species of mixed ancestry originate without change in ploidy, is supposedly more challenging than allopolyploidy (Gross and Rieseberg, 2005). Early generation hybrids usually show reduced fitness due to genic incompatibilities, but partial infertility resulting from meiosis of divergent chromosomes is indeed alleviated by genome duplication (Coyne and Orr, 2004; Rieseberg and Willis, 2007). Given the few documented cases of homoploid hybrid speciation, evolution towards complete reproductive isolation remains poorly understood (Yakimowski and Rieseberg, 2014).

In contrast to nascent allopolyploid species, homoploid hybrids lack any immediate intrinsic postzygotic isolation through ploidy and have to maintain genetic integrity in face of back-crossing with parental species. Accordingly, well-documented homoploid hybrid species in plants are known to also show karyotypic differences with their parents (Lai *et al.*, 2005), supporting the role of chromosomal rearrangements in promoting intrinsic reproductive isolation (Ortiz-Barrientos *et al.*, 2016). Noticeably, both homoploid and polyploid hybrid species usually present ecological divergence to their progenitor species and thus some degree of ecogeographic isolation (Gross and Rieseberg, 2005). Although empirical studies suggest that extrinsic reproductive isolation may not be necessary in allopolyploids (e.g. Huynh *et al*., 2020), simulations of homoploid hybrids suggest that strong ecological selection considerably promotes speciation (Buerkle *et al.*, 2000). Detailed studies in sunflowers accordingly indicated adaptive QTLs promoting habitat isolation in the presence of gene flow (Lexer *et al.*, 2003) and extrinsic reproductive isolation appear critical to the establishment of homoploid hybrid species (Rieseberg, 1997); (Levin, 2002). However, underpinnings of homoploid vs polyploid hybrid speciation remain elusive and further investigation based on different biological models is necessary.

The genus *Pulmonaria* (Boraginaceae) is a taxonomically difficult group that includes ca. 18 species distributed across Eurasia and that shows considerable morphological and karyotypic variation (Sauer, 1975; Bolliger, 1982). Following events of polyploidy and aneuploidy, chromosome numbers vary widely among taxa and have been extensively used to distinguish species among otherwise morphologically similar lineages (Sauer, 1975; Bolliger, 1982; Meeus *et al.*, 2016). Meeus *et al.* (2016) recently provided a phylogeny based on plastid loci and highlighted the presence of two major clades among European *Pulmonaria* species. Considering species included in our study area, the first clade encompasses the *P. officinalis* agg. (i.e. *P. officinalis* and *P. obscura*), whereas the second clade encloses *P. mollis* s.l. (incl. *P. montana*) and its multiple chromosomal lineages. Interspecific hybrids were regularly reported in sympatry (Hess, H.E., Landolt, E., Hirzel, R., 1972; Sauer, 1975; Bolliger, 1982) and inconsistencies between phylogenies based on nuclear and chloroplastic loci (Kirchner, 2004; Meeus *et al.*, 2016) further suggest that hybridization likely contributed to the evolutionary history of *Pulmonaria*. Switzerland being a major contact zone between *Pulmonaria* lineages that recolonized Europe after the last glacial maximum, the area appears particularly suitable to study the impact of hybridization on recent speciation.

Pulmonaria helvetica was described as a species of the aggregate *P. officinalis* agg. distinct from the previously reported *P. officinalis* (2n = 16) based on the discovery of populations presenting 2n = 24 chromosomes in western Switzerland (Bolliger, 1978, 1982). Its current distribution range is restricted to an area smaller than 1000 km^2 of temperate forests at low elevation in between the Jura mountains and the Alps, and was entirely covered by ice during the last glacial maximum, which excludes the possibility of a relictual lineage. Its morphology, distinct from other *Pulmonaria* species, combines phenotypic characters from both clades within the genus. Accordingly, *P. helvetica* was postulated to be of recent hybrid origin between *P. officinalis* and a representative of the dysploid *P. mollis* s.l. (2n = 18-24) (Bolliger, 1978, 1982). Although these species form a mosaic across western Switzerland and are mostly found in allopatric populations, hybrids between *P. officinalis* and populations of *P. mollis* with 18 chromosomes were reported with 17 chromosomes. Accordingly, the origin and evolution of populations described as *P. helvetica* remain elusive.

Multiple scenarios could explain the formation of *P. helvetica*: (i) The taxon may constitute a hybrid swarm due to continual production of early generation hybrids between species in secondary contact. Such a hybrid zone is expected to encompass a genetic cline from one to the other lineage (Barton and Hewitt, 1985) and would also refute *P. helvetica* as a distinct species. (ii) *P. helvetica* may have initially differentiated from the gene pool of a single species and was then introgressed with loci from another species. Limited evidence of admixture would accordingly be expected. (iii) *P. helvetica* results from hybrid speciation and accordingly presents a largely admixed but homogeneous gene pool. Here, we use plastid and nuclear sequences to characterize genetic variation and investigate the phylogenetic origin and population structure of *P. helvetica*. Particularly, we investigate whether *P. helvetica* is a distinctive gene pool to be considered as an independent species and infer its recent origin by addressing the hypothesis of hybrid speciation between distinct parental lineages. Finally, we investigate the ecological niche occupied by *Pulmonaria* species to assess the role of extrinsic factors and provide an initial account of main drivers of such recent speciation.

Material and methods

Sampling and DNA extraction

The study area is located in western Switzerland and encompasses the complete distribution range of *P. helvetica* as well as neighbouring populations of related species from the aggregate *P. officinalis* agg. (*P. helvetica* Bolliger, *P. officinalis* L, and *P. obscura* Dumort), and the *P. mollis* s.l. aggregate. As it matches phylogenetic insights (Meeus, *et al*., 2016), we followed the taxonomic treatment of Hess, H.E., Landolt, E., Hirzel, R. (1972) who nested species from aggregates of *P. montana* Lej.s.l. (i.e. *P. montana* Lej. s.str., 2n=22/24, and *P. montana* subsp. *jurana* (Graber) W. Sauer, 2n=22) and *P. mollis* agg. (*P. mollis* subsp. *alpigena* W. Sauer, 2n=18, and *P. collina* W. Sauer, 2n=18) into a *P. mollis* s.l. aggregate (Supplementary Table S1). Our sampling therefore comprehensively accounts for existing variation described as species and subspecies by (Sauer, 1975). Whenever possible, we revisited and included populations where (Sauer, 1975) or (Bolliger, 1982) reported chromosome counts, but individuals described as *P. montana* Lej. s.str with 2n=24 were not recovered (Sauer, 1975).

Within the distribution range of *P. helvetica*, three main regions were distinguished: a northern area hereafter referred to as region A, a south-western area (region B) and a south-eastern (region C). Three populations were sampled within each region by collecting

individuals whose coordinates were recorded using a GPS Juno SB (Trimble), with a precision of 2-3m (Table 1, Supplementary Table S2). Individuals from other species as well as contact zones between *P. helvetica* and either *P. obscura* or *P. officinalis* were similarly sampled and georeferenced. For each individual, a summer leaf is available as herbarium voucher deposited at the herbarium of the University of Bern (SG001-SG067).

Table 1: Individuals within species of *Pulmonaria* genotyped using ddRADseq (N) georeferences (Coord) in populations (Pop) of western Switzerland, with prior assignment to a taxon based on morphology (Table S1).

Species (chromosomes)	Taxon	Pop	Coord	N
***P. officinalis* agg. (phylogenetic clade I)**				
P. helvetica (2n = 24)	Region A (North)	H-A1[a]	46°48'43 - 6°47'59	4
		H-A2	46°47'49 - 6°49'45	4
		H-A3	46°47'58 - 6°46'22	1
	Region B (South west)	H-B1	46°36'38 - 6°37'47	3
		H-B2	46°37'11 - 6°38'59	2
		H-B3[a]	46°37'20 - 6°36'29	4
	Region C (South east)	H-C1[a]	46°36'52 - 6°48'17	3
		H-C2[a]	46°36'45 6°48'55	2
		H-C3[a]	46°36'45 - 6°47'02	1
P. obscura (2n = 14)	-	OB-A[a]	46°45'49 - 6°52'13	4
		OB-B	46°34'18 - 6°37'10	3
		OB-BE[a]	46°55'19 - 7°14'39	4
P. officinalis (2n = 16)	-	OFF1[a]	46°19'00 - 6°59'25	3
		OFF2[a]	46°39'36 - 6°33'34	3
***P. mollis* s.l. (phylogenetic clade II)**				
P. mollis agg. (2n = 18)	*P. collina*	COL1[a]	46°32'38 - 6°49'19	2
		COL2[a]	46°54'26 - 7°31'11	2
	P. mollis subsp. *alpigena*	MOL1[a]	47°02'20 - 6°54'59	2
		MOL2[a]	47°16'06 - 7°22'59	4
P. montana (2n = 22)	*P. montana* subsp. *jurana*	MON1[a]	46°33'44 - 6°14'22	3
	P. *montana* s.str.	MON2[a]	46°43'42 - 6°27'45	1
Contact zones between *P. helvetica* and another *Pulmonaria* species				
P. helvetica – *P. obscura*	*P. helvetica* (2) *P. obscura* (3)	S-FRO	46°35'24 - 6°40'54	5
P. helvetica- *P. officinalis*	*	S-GOU	46°35'42 - 6°48'29	7

[a] Populations with prior chromosome counts (Sauer, 1975; Bolliger, 1982)
*Individuals from S-GOU presented intermediate phenotypes and were not a priori attributed to species

A representative subset of 67 individuals was selected among the different species for genotyping at plastid and nuclear (ddRAD seq) loci. Namely, 24 *P. helvetica*, 4 *P. montana* Lej. s.l. (thereafter *P. montana*), 11 *P. obscura*, 6 *P. officinalis*, 4 *P. collina*, 6 *P. mollis* subsp. *alpigena*, 5 individuals from the contact zone between *P. helvetica* and

P. obscura, and 7 individuals from the contact zone between *P. helvetica* and *P. officinalis* (Table 1, Supplementary Table S2). For each individual, DNA was extracted from silica-dried leaves using the DNeasy Plant Mini Kit (Qiagen) following the manufacturer's protocol. Extracts have been purified a second time with SPRI beads and checked by electrophoresis on agarose. DNA concentration has been quantified using a Spark 10M Multimode Microplate Reader (TECAN).

Plastid sequences

The plastid *rps16* locus was PCR amplified following (Oxelman *et al.*, 1997) in 22 individuals from *P. helvetica* (12), *P. mollis* subsp. *alpigena* (1), *P. collina* (2), *P. montana* (1), *P. officinalis* (2) and *P. obscura* (4). PCR products were Sanger sequenced (service by Microsynth AG, Balgach Switzerland) and sequences were submitted to Genbank (accessions MK962107-MK962128). For each species investigated here as well as *Borago officinalis* as an outgroup, the *rps16* accessions (KT737605–KT737662; AJ431019.1) from (Meeus *et al.*, 2016) and here-produced sequences were aligned using MAFFT v7.308 (Katoh and Standley, 2013) under the E-INS-i algorithm, a 100PAM/k = 2 scoring matrix, a gap open penalty of 1.3 and an offset value of 0.123. After manual curation of the alignment, gaps were removed and a haplotype network based on the Median Joining method was performed following (Bandelt *et al.*, 1999) in SplitsTree4 (Huson and Bryant, 2006). Maximum Likelihood analyses were carried out with RAxML (Stamatakis, 2014) under the GTRGAMMA model, with 1000 bootstrap replicates.

ddRAD seq library preparation

ddRAD libraries were prepared on 67 individual samples following (Peterson *et al.*, 2012). Briefly, 120 nanograms of DNA was digested with NEB EcoRI-HF (rare-cutting) and NEB MseI (frequent-cutting) restriction enzymes, and adapters including NGS flow-cell annealing regions, PCR priming sequences and 5bp indexes were ligated to genomic DNA at restriction sites. The EcoR1 adapters included 48 barcodes that, in combination with three MseI indexes adjoined a biotin tag and four degenerated bases to identify PCR duplicates, individually tagged all samples. After ligation of adapters, samples were pooled and fragments around a mean length of ca. 550bp were selected by AMPure XP beads (Agencourt). Fragment retained after a selection with Dynabeads M-270 Streptavidin (Invitrogen) were used to prepare the library.

The concentration and distribution of fragment sizes were measured on the 2200 Tapestation (Agilent) using the High sensitivity D1000 ScreenTape (Agilent) before individual libraries were pooled. This final library was quality-controlled again and sequenced as single-end 100 bp reads on a lane of Illumina HiSeq3000 at the Next Generation Sequencing Platform of the University of Bern. Raw reads from each sample are available in the NCBI SRA repository SUB5539358.

Reference assembly and SNP calling

The quality of the sequencing reads was checked with FASTQC (Andrews, 2010) and data were processed following the dDocent pipeline (Puritz *et al.*, 2014). The process_radtag program (Stacks; (Catchen *et al.*, 2013)) was used to demultiplex raw fastq reads and remove

low quality bases. Only reads present at least ten times and then in at least seven individuals were retained and clustered using CD-HIT (Fu *et al.*, 2012), allowing for 20% mismatch. Corresponding loci generated a reference catalogue fitting best practice (LaCava *et al.*, 2019) against which all reads were remapped using BWA (Li and Durbin, 2009). SNPs calling was performed using Freebayes (Garrison and Marth, 2012).

Resulting SNPs were filtered with VCFtools (Danecek *et al.*, 2011) and only SNPs with a quality of more than Q20, minimum depth of 3, mean depth of 10, minor allele count of 3, minor allele frequency of 0.05 and less than 50% missing data were retained. Furthermore, the following procedure was applied to generate a final dataset of robust SNPs: (a) SNPs with more than 20% NAs per population were removed, (b) loci with extreme high coverage, i.e. depth above 100, were excluded, (c) complex SNPs were decomposed into phased SNP and indels, and indels were later removed, (d) loci with more than 10% missing data among all individuals were removed and only biallelic sites were conserved, and (e) sites out of (global) Hardy-Weinberg were removed. Finally, SNP haplotypes were inferred using rad_haplotyper (Willis *et al.*, 2017) to eliminate non-independent SNP loci at specific RAD loci and maximize the information content of each locus.

Genetic structure within and among species

We used the R package StAMPP (Pembleton *et al.*, 2013) to calculate pairwise Nei's genetic distances (Nei, 1972). The resulting matrix was exported and visualized as a network in SPLITSTREE v.4 (Huson and Bryant, 2006).

Genetic structure was inferred without a priori population grouping using Bayesian assignment in STRUCTURE v2.3.4 (Pritchard *et al.*, 2000). A first analysis was applied to the complete dataset using the admixture model for ancestry and the correlated model for allele frequencies, without any prior information about the geographical or population origin of individuals. It was run five times for K between K=1 and K=12 with 50,000 iterations of burn-in, and 1,000,000 iterations. Similar STRUCTURE analyses were run on 31 individuals belonging to *P. helvetica* to address genetic variation within that species. The 14 individuals belonging to *P. mollis* s.l. (i.e. *P. montana*, *P. mollis* subsp. *alpigena* and *P. collina*) were similarly analysed but did not show any genetic differentiation.

Structure Harvester (Earl and vonHoldt, 2012) was used to process results of STRUCTURE runs, and produce plots of the mean likelihood values per K, execute the Evanno method to select the optimal K value (Evanno *et al.*, 2005) and generate input files for CLUMPP (Jakobsson and Rosenberg, 2007). The presumed optimal K value was assessed as the maximum value before the second-order derivative rate of change in Ln probabilities L(K) reaches a plateau. A hierarchical presentation of different levels of clustering was however privileged as biologically more meaningful (Janes *et al.*, 2017). Cluster assignment across replicate analyses was summarised in CLUMPP and plotted in R. Genotypic data were also visualised by performing a PCA with the adegenet R package (Jombart, 2008).

Analysis of molecular variance (AMOVA) was performed using Arlequin ver. 3.5 (Excoffier and Lischer, 2010) and partitioned genetic variance across species, populations and individuals with significance estimated through 1000 permutations. A first analysis compared the six species defined in our study area (i.e. *P. helvetica*, *P. obscura*, *P. mollis* subsp. *alpigena*, *P. collina*, *P. montana* and *P. officinalis*). Populations within species corresponded to field populations except for *P. helvetica* where samples were pooled into regions A, B, and C. A subsequent analysis partitioned genetic variance within *P. helvetica* by comparing the three populations nested within each sampling regions (A, B and C).

Patterns of isolation by distance were tested within *P. helvetica* and within *P. mollis* s.l. Pairwise genetic distances among individuals were estimated using Rousset's â (Rousset, 2000) in SPAGeDi v1.3 (Hardy and Vekemans, 2002), and associated with Euclidian spatial distances among pairs of individuals using Mantel tests in the R package vegan (Oksanen *et al*., 2019) . Significance was evaluated through 10000 permutations.

Modelling of occupied environmental niches

Environmental analyses were based on available occurrences retrieved from the Info Flora database based on field observations reported within the last two decades with an accuracy of at least 1km (www.infoflora.ch). All 3242 georeferenced occurrences were filtered from duplicated observations within 100x100m and only occurrences included within a defined geographical space (i.e. sites below 2000m elevation across the study area) were kept. Due to possible misidentification, *P. mollis*, *P. collina* and *P. mollis* subsp. *alpigena* were treated as a single taxon (i.e. *P. mollis* agg.). Likewise, *P. montana* and *P. montana* subsp *jurana* were similarly treated (i.e. *P. montana*). This yield a dataset of 1314 occurrences including 281 *P. helvetica*, 125 *P. officinalis*, 610 *P. obscura*, 188 *P. mollis* agg. and 110 *P. montana*.

Values from 18 environmental variables (Supplementary Table S3) were extracted from the Swiss Eco-Climatic GIS data available at 25m-resolution (Zimmermann and Kienast, 1999) for each occurrence (Supplementary File 1), using the spatial analysts tools in ArcMap 10.6 (ESRI, Redlands, CA, USA). When the Pearson correlation coefficient between pairs of variables equalled or exceeded 0.9, a variable was selected based on the presumed ecophysiology and biology of *Pulmonaria* species. Accordingly, proximal predictors were privileged over indirect ones such as elevation and 12 variables were kept for ecological modelling: annual degree days, global radiation, topographic aspect, topographic position, topographic wetness index, average evapotranspiration, slope, number of precipitation days per growing season, mean precipitation, soil-water balance, continentality index of Gams, number of frost days during the growing season.

Comparative ecological modelling among species followed (Broennimann *et al.*, 2012). Accordingly, principal component (PC) analyses were performed within the defined geographical space out of 1 Mio points randomly distributed across this background using a gridded environmental space with 100x100 cells corresponding to unique combinations of environmental conditions from the 12 environmental variables mentioned above. For each species, the density of occurrences in each cell is estimated through a kernel density function. Realized niches between all pairs of species were compared by testing for niche similarity (Warren *et al.*, 2008) using the Ecospat R package (Di Cola *et al.*, 2017). As each species was modelled in a common background environmental space, the inferred ecological niches could be straightforwardly compared and their overlap calculated using the metric D (Warren *et al.*, 2008). A niche similarity test was performed by randomly shifting observed density of occurrences from one species into the environment with the second species and inversely. The resulting overlap is measured across 100 repetitions building a null distribution to reject niche overlap between species due to regional availability of habitats when falling outside of the 95% confidence interval of the null hypothesis. Accordingly, an observed niche overlap higher than expected indicates niche conservatism, whereas a lower overlap indicates significant niche differentiation.

Results

SNP calling

Sequencing of ddRADseq libraries produced ca. 360 10^6 reads, among which 252 10^6 were retained after demultiplexing. Our *de novo* reference catalogue presented 29,595 contigs that mapped 78% reads after parameter optimization and called 167,782 SNPs. After filtering (Supplementary Table S4), the final dataset included 4,286 biallelic SNPs matching overall Hardy-Weinberg expectation (Supplementary File 2). Haplotyping of SNPs further reduced the number of SNPs to 1,077 independent loci (Supplementary File 3).

Genetic distance at plastid and nuclear loci

The median joining network (Fig. 1a) and the Maximum Likelihood phylogenetic tree (Supplementary Fig. S1) inferred from the plastid *rps16* locus indicated three main haplotypes. The first plastid haplotype was shared among samples from *P. mollis* s.l. without species distinction as well as all *P. helvetica* individuals. As expected, *P. officinalis* individuals presented different haplotypes, with one private to *P. officinalis* and the other shared with *P. obscura*. This third haplotype includes a subgroup including only sequences from *P. obscura*.

The splits network based on Nei's genetic distance between the 67 individuals based on nuclear loci of the ddRAD dataset (Fig. 1b) showed two main sets of splits. Individuals from *P. montana*, *P. mollis*, and *P. collina* formed a coherent group corresponding to *P. mollis* s.l., whereas individuals from *P. obscura* were grouped together, close to a group including samples from *P. officinalis*. All individuals of *P. helvetica* were clustered at the intersection of splits between *P. mollis* s.l. and *P. officinalis*.

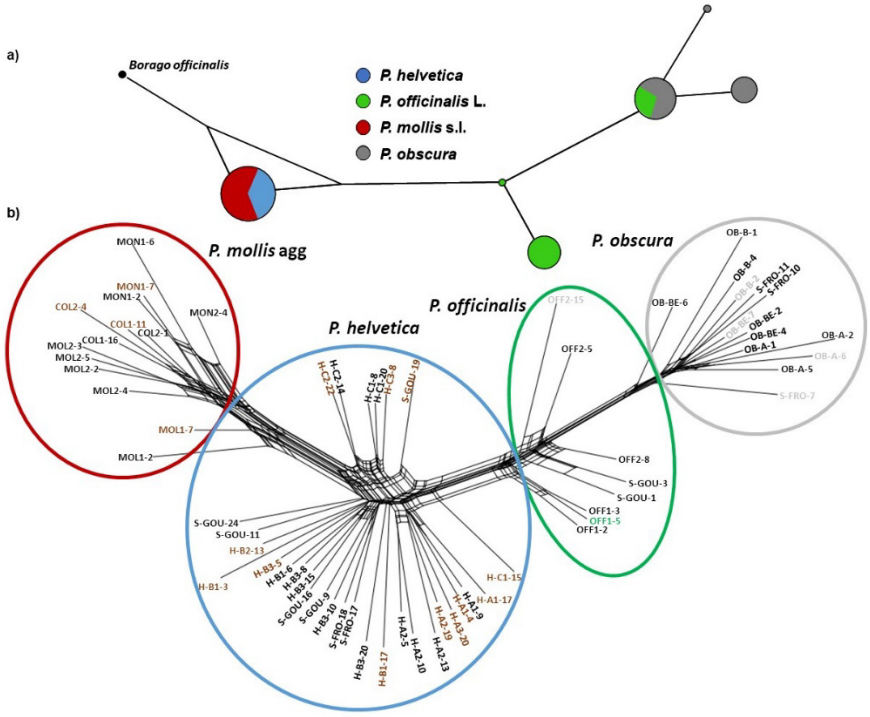

Fig. 1: Distance-based median joining network of *Pulmonaria* samples assigned to species according to the panel. (a) Network based on plastid *rps16* sequences from 22 samples from western Switzerland and sequences from European samples (Meeus *et al.*, 2016), including the outgroup *Borago officinalis*. (b) Network based on Nei's distance between 67 samples from western Switzerland genotyped at 1077 nuclear loci (ddRAD dataset). Individuals in bold are coloured according to their plastid haplotype.

Genetic structure analysis

Analysis of genetic variation in the complete dataset using STRUCTURE supported K=2 as the optimal number of clusters following the Evanno ΔK method, whereas the mean Ln probabilities increased from K=1 to K=4 before a plateau is reached (Supplementary Fig. S2). As the PCA on genotypic data was consistent with four clusters (Supplementary Fig. S3), different levels of partitioning ranging from K=2 to K=4 were considered to grasp the hierarchical clustering (Fig. 2).

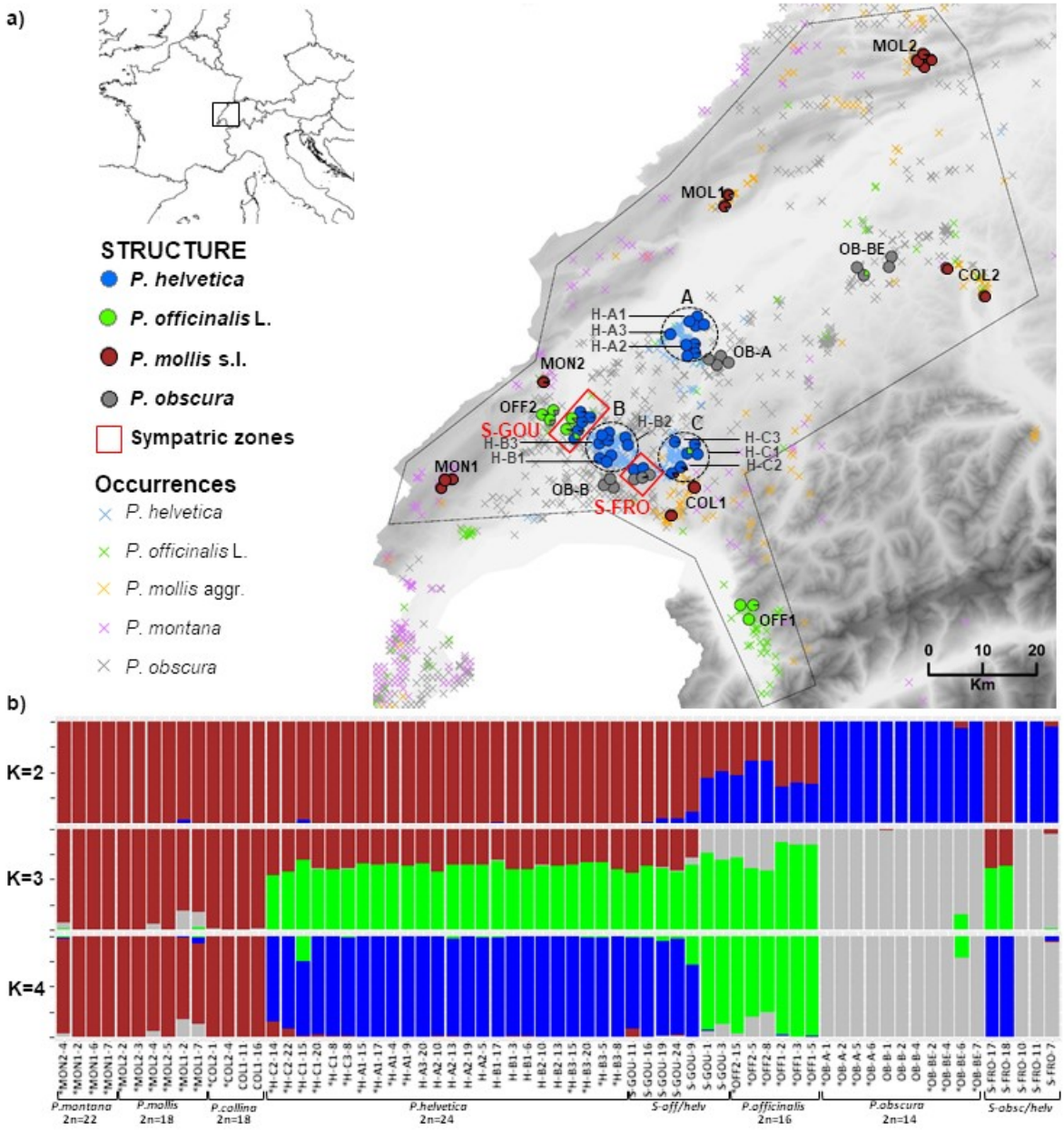

Fig. 2: Genetic clusters based on 1077 ddRAD loci within and among *Pulmonaria* species and their distribution across the study area. (a) Study area with genotyped individuals as circles coloured according to their genetic clusters out of STRUCTURE at K=4. Occurrences are shown as crosses to illustrate the range of the different taxa according to the panel. (b) STRUCTURE plots showing the partitioning from K=2 to K=4, with species and populations according to Table 1. * is marking populations with prior chromosome counts and populations beginning with "S-" are in zones of sympatry between species.

At K=3, genetic clusters distinguished *P. mollis* s.l. from *P. officinalis*, and from *P. obscura*. Individuals of *P. officinalis* were however not fully assigned to this species and appeared to share genetic variation with *P. obscura*. At this level, *P. helvetica* appeared as a coherent genepool presenting admixture between *P. officinalis* and the cluster representative of *P. mollis* s.l. and *P. montana*, matching expectations of the hypothesis that *P. helvetica* originated through their hybridization.

At K=4, *P. helvetica* and *P. officinalis* became clearly assigned to different genetic clusters, supporting the hypothesis that these taxa are reproductively isolated. Genotypes across zones where species are in sympatry were consistent with limited admixture. The contact zone between *P. helvetica* and *P. officinalis* (S-GOU population) showed individuals strongly assigned to either *P. officinalis* or the genetic cluster encompassing *P. helvetica*, with only limited evidence of admixture from *P. officinalis* (i.e. 1% to 5% assignment to that cluster). Out of seven samples, only one individual (S-GOU9) presented appreciable admixture with ca. 70% *P. helvetica* and 30% *P. officinalis.* In the area where *P. helvetica* is in parapatry with *P. mollis* s.l. (H-C2), sampled individuals were unambiguously assigned to *P. helvetica*, but a sample revealed noticeable admixture with the corresponding genetic cluster (8% and 15%). Samples collected in the contact zone between *P. helvetica* and *P. obscura* (S-FRO population) presented 99% pure assignment to their respective cluster, indicating no gene flow between those two species of the *P. officinalis* agg. Further partition of genetic variation at K=5 indicated substructure within *P. helvetica*, whereas putative taxa such as *P. mollis* subsp. *alpigena, P. collina* and *P. montana* remained genetically homogeneous across levels of partitioning. Accordingly, we considered the four genetic clusters as main species under study and estimated their genetic diversity (Table S5). Although the comparison of five simple scenarios with an approximate Bayesian computation approach did not allow to fully capture the complex evolution of *P. helvetica,* scenarios involving admixture yielded data close to the observed dataset (Supplementary Text T1).

Within *P. helvetica,* STRUCTURE analyses between K=1 to K=4 appeared equally probable and highlighted substructure largely matching the geographic distribution of populations (Supplementary Fig. S4). At K=2, the clustering reflected sampling regions with individuals from regions A and B appearing distinct and region C appearing admixed. At K=3, individuals from each region formed distinct clusters (Fig. 3a). At K=4, supported as optimal partitioning based on ΔK, all individuals appeared as admixture between four clusters.

Within *P. mollis* s.l., STRUCTURE analyses from K=1 to K=3 were also equally probable, with a ΔK suggesting an optimal clustering at K=3 (Supplementary Fig. S5). Although inconsistent with taxonomy, individuals described as *P. montana* (MON1 and MON2 populations) were partially distinguished from *P. mollis* agg. at K=2, whereas individuals of *P. mollis* subsp. *alpigena* (MOL1 and MOL2 populations) were assigned to two distinct clusters from K=3 and showed admixture within *P. collina* (COL1 and COL2 populations; Fig. 3b).

Consistent with the STRUCTURE analyses, partitioning of genetic variation using the AMOVA framework showed that the species level mostly explained the genetic structure in the dataset (ca. 21%). F_{ST} estimated between described taxa (i.e. *P. helvetica, P. mollis* subsp. *alpigena*, *P. collina*, *P. montana*, *P. obscura*, and *P. officinalis*) ranged from 0.06 to 0.46 and were all significant (Supplementary Table S5). Within *P. helvetica*, sampling regions A, B,

and C included most of the variation (ca. 4%), whereas pairwise F_{ST} between regions were occasionally significant but always low (<0.015; Supplementary Table S5).

Mantel tests between genetic distances (â) and geographical distances among individuals showed a significant correlation of 0.33 (p<0.01) that was consistent with isolation by distance within *P. helvetica*. The correlation was also significant (0.23; p<0.01) among individuals of *P. mollis* s.l. (Fig. 3; Supplementary Fig. S6).

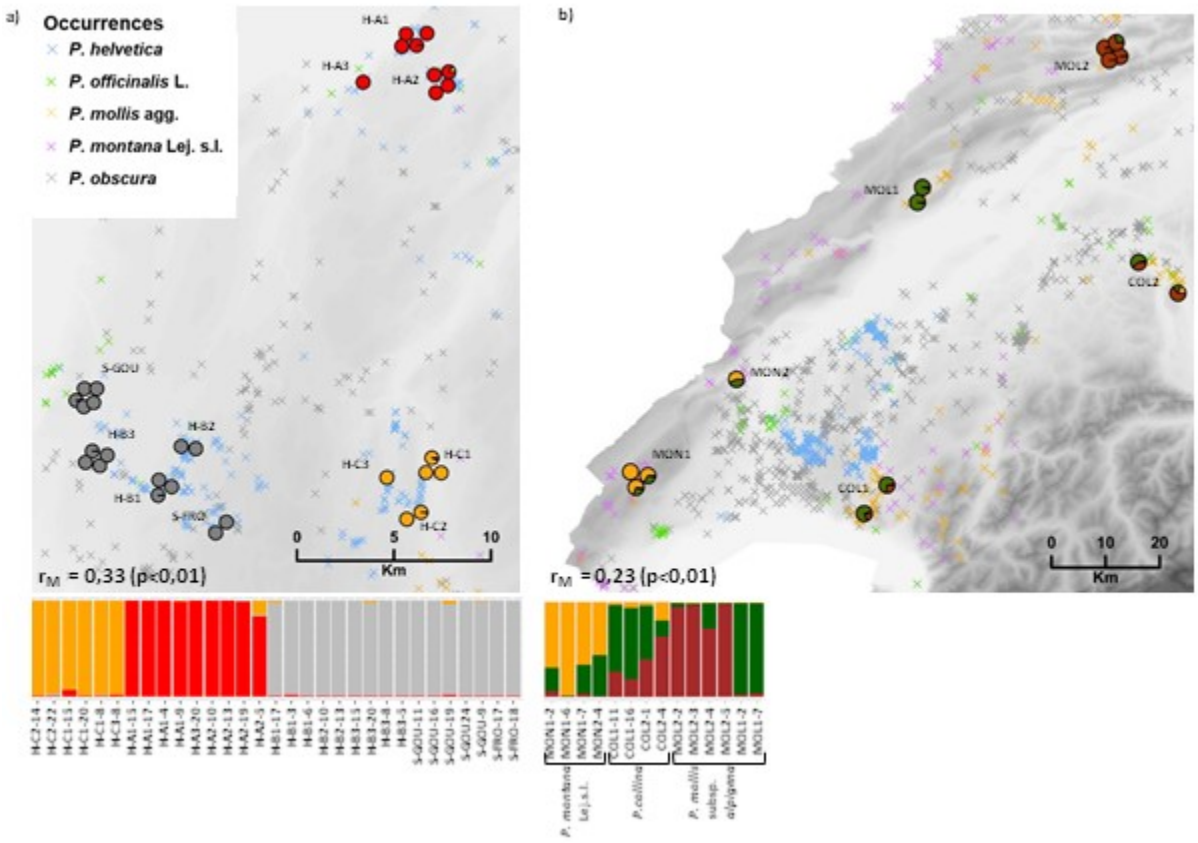

Fig. 3: Distribution of genotyped individuals (a) within *Pulmonaria helvetica* and (b) within the *P. mollis* s.l. among K=3 genetic clusters. The slope of the correlation between genetic and geographic distances is indicated with r_M and its corresponding significance is given by the p-value.

Realized ecological niches

The first PC-axis of the environmental space defined within the study area by ordination of 12 variables explained 43.5% of the total inertia and was positively associated to precipitation variables, soil-water-balance and slope (Fig. 4). The second axis (11.8%) was negatively correlated to global radiation and average number of frost days during the growing season.

The different *Pulmonaria* species occupied similar fractions of this space and presented high overlap of their realized niches. *P. helvetica* presented a noticeably narrow ecological niche as compared to all other species. Accordingly, tests between pairs of species showed non-significantly lower similarity than expected, consistent with a lack of niche differentiation among species (Table 2).

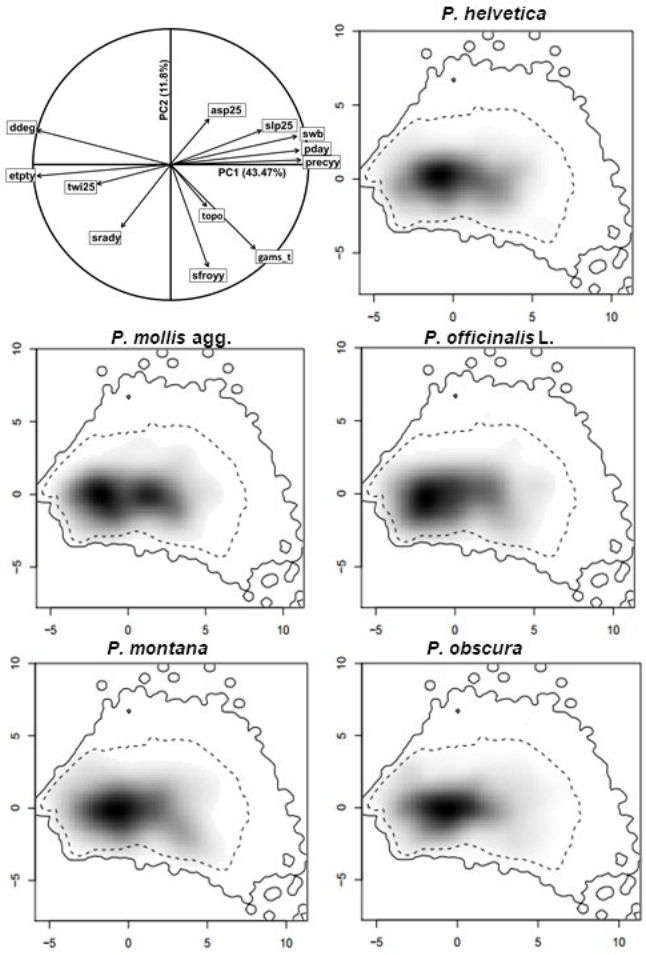

Fig. 4: Realized ecological niches of *Pulmonaria* species modelled by ordination in the common environmental space of the study area and defined in the correlation circle (top left). Shading represents the density of occurrences for each species across the common environmental space with solid and dashed lines representing 100% and 50% of its distribution.

ddeg: annual degreedays (3°C threshold), slp: slope, swb: soil water balance, pday: number of precipitation days during growing season, precyy: yearly mean precipitation sum, sfroyy: number of frost days during growing season, srady: global radiation, topo: relative topographic exposure, asp: direction that a slope faces, gams_t: continentality indices of Gams, twi: topographic wetness index, etpty: average evapotranspiration per month (Supplementary Table S2).

Table 2: Overlap of modeled environmental niches between all pairs of species, with tests for greater and lower similarity than expected by chance based on 100 randomization (p-values). A significant greater similarity indicates niche conservationism, whereas significantly lower similarity would be coherent with niche differentiation

Comparison between		Niche overlap	Greater similarity, p-value	Lower similarity, p-value
P. helvetica	*P. officinalis*	0.618	0.118	0.921
P. helvetica	*P. obscura*	0.641	0.118	0.911
P. helvetica	*P. montana*	0.613	0.129	0.842
P. helvetica	*P. mollis* agg.	0.544	0.139	0.921
P. mollis agg.	*P. montana*	0.350	0.178	0.832
P. mollis agg.	*P. officinalis*	0.510	0.129	0.891
P. montana	*P. obscura*	0.596	0.109	0.811
P. mollis agg.	*P. obscura*	0.400	0.208	0.802
P. montana	*P. officinalis*	0.584	0.089	0.901
P. obscura	*P. officinalis*	0.638	0.248	0.960

Discussion

Pulmonaria helvetica *as a reproductively isolated species*

Genotyping of representative samples of *Pulmonaria* species growing across western Switzerland efficiently delineated taxa in this morphologically and taxonomically confuse group of plants. Limited substructure was highlighted within *P. mollis* s.l. based on 1077 ddRAD seq loci and the distinction of nested taxa such as *P. mollis* subsp. *alpigena* (2n=18), *P. collina* (2n=18) or *P. montana* (2n=22) was not supported here. In the *P. officinalis* agg., nuclear loci contrastingly distinguished *P. obscura* (2n=14), *P. officinalis* (2n=16) and *P. helvetica* (2n=24), suggesting restricted gene flow among those taxa with different karyotypes. In particular, throughout partitioning levels of the STRUCTURE analysis, samples from *P. helvetica* formed a differentiated gene pool appearing genetically homogeneous and distinct from neighbouring species.

Sharp delimitation of *P. helvetica* based on multilocus genotypes at K=4 and evidence of balanced admixture between the *P. mollis* s.l. and *P. officinalis* at K=3 contrasts with the genetic cline expected under a scenario of a hybrid swarm after secondary contact. Genotyping across contact zones where *P. helvetica* is in sympatry with other *Pulmonaria* species further revealed evidence of limited introgression and therefore supports this narrow endemic as an independent species. Together with the few experimental crosses of (Bolliger, 1982) between *P. helvetica* and *P. officinalis* that produced progenies with reduced pollen fertility, this study indicates that *P. helvetica* is reproductively isolated from neighbouring taxa across the study area.

Hybrid origin of Pulmonaria helvetica

The distinctive gene pool of *P. helvetica* shows considerable admixture that is thus coherent with hybrid allelic combinations stabilized in a lineage that can be recognized as a species (i.e. hybrid speciation as defined by (Mallet, 2007). Plastid and nuclear loci here supported such hybrid origin involving *P. mollis* s.l. as maternal progenitor and of *P. officinalis* as paternal progenitor. As the current distribution range of *P. helvetica* was entirely covered by ice until 18 - 16 kyr BP, hybridization likely took place after recolonization of suitable forests on the Swiss plateau (ca. 13 kyr BP) (Burga, 1998) that promoted secondary contact between populations of those species.

Natural hybrids between *P. mollis* s.l. and *P. officinalis* have been documented with 2n = 17 and a morphology similar to *P. helvetica* (Bolliger, 1982). Available data thus appear coherent with a scenario of recombinational speciation having generated *P. helvetica* through elusive chromosomal rearrangements towards a stabilized lineage with 2n = 24. As the maternal taxon presents variable karyotypes following hypothetical dysploidy, Bolliger (1982) also pointed out the possible role of specific populations of *P. montana* s. str. with 2n = 24. Although populations recovered here with highest number of chromosomes were only 2n = 22, hybridization involving a population of *P. mollis* s.l. with 24 chromosomes and unreduced triploid pollen from *P. officinalis* (n = 24) cannot be ruled out. The role of hybridization with dysploid taxa at the origin of biological species is poorly documented and, although corresponding hybrids could benefit from immediate reproductive isolation, adaptive changes required for a balanced meiosis in such admixed species remain particularly elusive. Although a more comprehensive survey would be desirable, genome-wide loci here revealed genetically indistinct putative maternal taxa despite different chromosome numbers and

therefore hardly favour a specific scenario. Additional work using e.g. herbarium samples or molecular cytogenetics may fully characterize events at the origin of *P. helvetica* and shed light on the evolution of its highly heterochromatic bivalents (Bolliger, (1982).

Establishment of the narrow endemic species

Hybrid speciation is necessarily initiated in the vicinity of the parental species and the fitness of hybrids is thus critical for the establishment of hybrid species (Abbott *et al.*, 2013). Accordingly, new hybrid species investigated so far appeared typically fitter than either parental species in divergent habitats (Rieseberg, 1997). In contrast to such expectations, comparative niche modelling here did not support significant differentiation between the realized niche of *P. helvetica* and either parental species (*P. officinalis*, *P. mollis* agg., or *P. montana*), despite their current parapatry or allopatry. It should be noted that such lack of differentiation between species holds for environmental abiotic variables considered here and thus remains blind to other factors such as edaphic or biotic factors that may be of particular relevance. It however appears consistent with the hybrid *P. helvetica* having an intrinsic fitness advantage, as also reported in some polyploid hybrids species (Huynh *et al.*, 2020).

The narrow endemic *P. helvetica* did not only establish but spread towards the three main regions of its current distribution range (< 1000 km^2) in which it became abundant. Limited genetic differentiation and the pattern of isolation by distance within the species is consistent with such expansion. The effective expansion of such a recent species may have been promoted by hybridization that is known to mask the load of deleterious recessive alleles fixed in parental populations and yield heterosis (Lynch, 1991; Wang *et al.*, 2015). This has been shown to affect populations of perennial plants subjected to genetic drift (Oakley and Winn, 2012) and may have promoted the establishment and expansion of *P. helvetica*. It should however be noticed that the origin of this species also coincided with postglacial recolonization of multiple species and it may have benefitted from “open habitats” promoting hybrid speciation in face of the parental populations (Buerkle *et al.*, 2000; Gross and Rieseberg, 2005). The fitness of *Pulmonaria* species and their hybrids should be evaluated to conclude on the underpinnings of such a speciation event.

Presenting genetic evidence of admixture and strong reproductive isolation with parental lineages, *P. helvetica* satisfies criteria of a hybrid species (Mallet, 2007). Its origin however matches typical scenarios of neither homoploid nor polyploid speciation, but suggests that hybridization supports such postglacial speciation (Arnold, 1997; Levin, 2002). Our data could not demonstrate that reproductive isolation is derived from hybridization itself (Schumer *et al.*, (2014). However, our results indicate a limited impact of niche shift and a central role of chromosomal changes in reducing gene flow. As also discussed by Nieto Feliner *et al.* (2017), admixed genetic variation is likely to support rapid speciation. Accordingly, range shifts bringing taxa in secondary contact may promote the origin of new species under climate changes (Levin, 2019). Further investigation should shed light on the consequences of reticulate evolution for diversification and expansion of hybrid species.

Supplementary Information

The supplementary file contains Table S1 to S5 and Figure S1 to S6

The supplementary Text 1 presents approximate Bayesian computations

Supplementary Data 1 presents spatial coordinates and environmental values of all samples

Supplementary Data 2 presents genetic data in vcf format

Supplementary Data 3 presents genetic data in genepop format

Acknowledgement

Data analysed in this paper were generated at the Genetic Diversity Centre, ETH Zurich, and sequenced at the Next Generation Sequencing Platform of the University of Bern. We thank Markus Bolliger, Christophe Bornand and two anonymous reviewers for their valuable comments.

References

Abbott R, Albach D, Ansell S, *et al.* 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229–246. doi:10.1111/j.1420-9101.2012.02599.x.

Andrews S. 2010. *FastQC: A quality control tool for high throughput sequence data.* Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arnold ML. 1997. *Natural Hybridization and Evolution*. New York: Oxford University Press.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**: 37–48. doi:10.1093/oxfordjournals.molbev.a026036.

Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**: 113–148. doi:10.1146/annurev.es.16.110185.000553.

Bolliger M. 1978. *Die Pulmonaria obscura-officinalis-Gruppe in der Schweiz*: Büchler.

Bolliger M. 1982. *Die Gattung Pulmonaria in Westeuropa*. Vaduz: J. Cramer.

Broennimann O, Fitzpatrick MC, Pearman PB, *et al.* 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* **21**: 481–497. doi:10.1111/j.1466-8238.2011.00698.x.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH. 2000. The likelihood of homoploid hybrid speciation. *Heredity* **84**: 441–451. doi:10.1046/j.1365-2540.2000.00680.x.

Burga CA. 1998. Swiss Vegetation History during the Last 18 000 Years. *New Phytologist* **110**: 581-602. doi:10.1111/j.1469-8137.1988.tb00298.x.

Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140. doi:10.1111/mec.12354.

Coyne JA, Orr AH. 2004. *Speciation*. Sunderland, Massachusetts: Sinauer Associates Inc. Publ.

Danecek P, Auton A, Abecasis G, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158. doi:10.1093/bioinformatics/btr330.

Di Cola V, Broennimann O, Petitpierre B, *et al.* 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* **40**: 774–787. doi:10.1111/ecog.02671.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361. doi:10.1007/s12686-011-9548-7.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**: 564–567. doi:10.1111/j.1755-0998.2010.02847.x.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152. doi:10.1093/bioinformatics/bts565.

Garrison E, Marth G. 2012. *Haplotype-based variant detection from short-read sequencing*.

Gross BL, Rieseberg LH. 2005. The ecological genetics of homoploid hybrid speciation. *Journal of Heredity* **96**: 241–252. doi:10.1093/jhered/esi026.

Hardy OJ, Vekemans X. 2002. spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620. doi:10.1046/j.1471-8286.2002.00305.x.

Hess, H.E., Landolt, E., Hirzel, R. 1972. *Flora der Schweiz und angrenzender Gebiete*. Basel und Stuttgart: Birkhäuser Verlag.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267. doi:10.1093/molbev/msj030.

Huynh S, Broennimann O, Guisan A, Felber F, Parisod C. 2020. Eco-genetic additivity of diploids in allopolyploid wild wheats. *Ecology Letters* **23**: 663–673. doi:10.1111/ele.13466.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806. doi:10.1093/bioinformatics/btm233.

Janes JK, Miller JM, Dupuis JR, *et al.* 2017. The K = 2 conundrum. *Molecular Ecology* **26**: 3594–3602. doi:10.1111/mec.14187.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405. doi:10.1093/bioinformatics/btn129.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780. doi:10.1093/molbev/mst010.

Kirchner DE. 2004. *Molekulare Phylogenie und Biogeographie der Gattung Pulmonaria L. (Boraginaceae),* 1. Aufl. Aachen: Mainz.

LaCava MEF, Aikens EO, Megna LC, Randolph G, Hubbard C, Buerkle CA. 2019. Accuracy of de novo assembly of DNA sequences from double-digest libraries varies substantially among software. *Molecular ecology resources* **20**: 360–370. doi:10.1111/1755-0998.13108.

Lai Z, Nakazato T, Salmaso M, *et al.* 2005. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* **171**: 291–303. doi:10.1534/genetics.105.042242.

Levin DA. 2002. *The Role of Chromosomal Change in Plant Evolution*. New York: Oxford University Press.

Levin DA. 2019. Plant speciation in the age of climate change. *Annals of botany* **124**: 769–775. doi:10.1093/aob/mcz108.

Lexer C, Welch ME, Durphy JL, Rieseberg LH. 2003. Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: Implications for the origin of Helianthus paradoxus, a diploid hybrid species. *Molecular Ecology* **12**: 1225–1235. doi:10.1046/j.1365-294X.2003.01803.x.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760. doi:10.1093/bioinformatics/btp324.

Lynch M. 1991. The Genetic Interpretation of inbreeding depression and outbreeding depression. *Evolution* **45**: 622–629. doi:10.1111/j.1558-5646.1991.tb04333.x.

Mallet J. 2007. Hybrid speciation. *Nature* **446**: 279–283. doi:10.1038/nature05706.

Meeus S, Janssens S, Helsen K, Jacquemyn H. 2016. Evolutionary trends in the distylous genus Pulmonaria (Boraginaceae): Evidence of ancient hybridization and current interspecific gene flow. *Molecular Phylogenetics and Evolution* **98**: 63–73. doi:10.1016/j.ympev.2015.11.022.

Nei M. 1972. Genetic Distance between Populations. *American Naturalist* **106**: 283–292.

Nieto Feliner G, Álvarez I, Fuertes-Aguilar J, *et al.* 2017. Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* **118**: 513–516. doi:10.1038/hdy.2017.7.

Oakley CG, Winn AA. 2012. Effects of population size and isolation on heterosis, mean fitness, and inbreeding depression in a perennial plant. *New Phytologist* **196**: 261–270. doi:10.1111/j.1469-8137.2012.04240.x.

Oksanen J, Blanchet FG, Friendly M., Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H 2019. *Vegan: Community Ecology Package*. https://cran.r-project.org, https://github.com/vegandevs/vegan.

Ortiz-Barrientos D, Engelstädter J, Rieseberg LH. 2016. Recombination Rate Evolution and the Origin of Species. *Trends in Ecology & Evolution* **31**: 226–236. doi:10.1016/j.tree.2015.12.016.

Oxelman B, Lidn M, Berglund D. 1997. Chloroplastrps16 intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* **206**: 393–410. doi:10.1007/BF00987959.

Pembleton LW, Cogan NOI, Forster JW. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular ecology resources* **13**: 946–952. doi:10.1111/1755-0998.12129.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one* **7**: e37135. doi:10.1371/journal.pone.0037135.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Puritz JB, Hollenbeck CM, Gold JR. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**: e431. doi:10.7717/peerj.431.

Rieseberg LH. 1997. Hybrid origins of plant species. *Annual Review of Ecology, Evolution and Systematics* **28**: 359–389. doi:10.1146/annurev.ecolsys.28.1.359.

Rieseberg LH, Willis JH. 2007. Plant speciation. *Science (New York, N.Y.)* **317**: 910–914. doi:10.1126/science.1137729.

Rousset F. 2000. Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**: 58–62. doi:10.1046/j.1420-9101.2000.00137.x.

Sauer W. 1975. *Karyo-systematische Untersuchungen an der Gattung Pulmonaria (Boraginaceae): Chromosomen-Zahlen, Karyotyp-Analysen und allgemeine Hinweise auf die Entwicklungsgeschichte*. Stuttgart: S. Schweizerbart'sche Verlagsbuchhandlung.

Schumer M, Rosenthal GG, Andolfatto P. 2014. How common is homoploid hybrid speciation? *Evolution* **68**: 1553–1560. doi:10.1111/evo.12399.

Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* **60**: 561–588. doi:10.1146/annurev.arplant.043008.092039.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033.

Wang L, Greaves IK, Groszmann M, Wu LM, Dennis ES, Peacock WJ. 2015. Hybrid mimics and hybrid vigor in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **112**: E4959-67. doi:10.1073/pnas.1514190112.

Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**: 2868–2883. doi:10.1111/j.1558-5646.2008.00482.x.

Willis SC, Hollenbeck CM, Puritz JB, Gold JR, Portnoy DS. 2017. Haplotyping RAD loci: an efficient method to filter paralogs and account for physical linkage. *Molecular ecology resources* **17**: 955–965. doi:10.1111/1755-0998.12647.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 13875–13879. doi:10.1073/pnas.0811575106.

Yakimowski SB, Rieseberg LH. 2014. The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *American Journal of Botany* **101**: 1247–1258. doi:10.3732/ajb.1400201.

Zimmermann NE, Kienast F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *Journal of Vegetation Science* **10**: 469–482. doi:10.2307/3237182.

Data Accessibility

The datasets supporting the conclusions of this article are available in the NCBI SRA repository SUB5539358 and hyperlink to datasets in Bioproject 544114. Plastid *rps16* sequences have been submitted to NCBI with the accessions number MK962107-MK962128.