Received: 2 November 2020 | Revised: 18 January 2021 | Accepted: 3 March 2021

DOI: 10.1111/jeb.13784

RESEARCH PAPER

JOURNAL OF Evolutionary Biology eseb WILEY

Hybridization, ecogeographical displacement and the emergence of new lineages – A genotyping-by-sequencing and ecological niche and species distribution modelling study of *Sempervivum tectorum* L. (Houseleek)

Armin G. Fabritzek | **Eva Maria Griebeler** | **Joachim W. Kadereit**

Institut für Organismische und Molekulare Evolutionsbiologie, Johannes Gutenberg-Universität Mainz, Mainz, Germany

Correspondence
Joachim W. Kadereit, Institut für Organismische und Molekulare Evolutionsbiologie, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany.
Email: kadereit@uni-mainz.de

Funding information
Deutsche Forschungsgemeinschaft (DFG), Grant/Award Number: KA 635/17-1&2

Abstract

Ecogeographical displacement of homoploid hybrid lineages from their parents is well documented and considered an important mechanism to achieve reproductive isolation. In this study, we investigated the origin of the flowering plant species *Sempervivum tectorum* in the Massif Central (France) through homoploid hybridization between lineages of the species from the Rhine Gorge area (Germany) and the Pyrenees (France). We used genotyping-by-sequencing genetic data as evidence for the hybrid origin of the Massif Central lineage, and WorldClim climatic data and soil pH and soil temperature data collected by us for ecological niche and species distribution modelling. We could show that the Massif Central lineage shows hybrid admixture and that the niche of this lineage is significantly different from those of the parental lineages. In comparison with the parental niches, different variables of the niche of the hybrid lineage are intermediate, parental-combined or extreme. The different niche of the Massif Central populations thus can plausibly be interpreted as hybridization-derived. Our species distribution modelling for the Last Glacial Maximum and Mid-Holocene showed that the potential distribution of the hybrid lineage at the likely time of its origin in the Quaternary possibly was parapatric in relation to the largely sympatric distributions of the parental lineages. We hypothesize that reproductive isolation of the hybrid lineage from the parental lineages resulted from the segregation of distribution ranges by a differential response of the three lineages to a warming climate.

KEYWORDS

homoploid hybrid speciation, niche evolution, reproductive isolation, soil pH, soil temperature

1 | INTRODUCTION

Interspecific hybridization has great evolutionary potential and can lead to homoploid hybrid speciation, including speciation through introgressive hybridization (Abbott et al., 2010), or to polyploid hybrid speciation (Abbott et al., 2013; Abbott & Rieseberg, 2012; Arnold, 1997; Mallet, 2007). Whereas polyploid hybrid speciation may account for 2%–15% of speciation events in flowering

plants (Otto & Whitton, 2000; Wood et al., 2009), homoploid hybrid speciation appears to be rarer (Kadereit, 2015; Yakimowski & Rieseberg, 2014).

Ecogeographical displacement of hybrid lineages from their parents is well documented for homoploid hybrid species. Compilations of homoploid hybrid plant species (Abbott et al., 2010; Gross & Rieseberg, 2005; Kadereit, 2015; Rieseberg, 1997; Yakimowski & Rieseberg, 2014) clearly showed that ecogeographical differentiation between hybrid and parent species can be observed in the large majority of cases. This ecogeographical differentiation results from intermediate trait expression, combination of parental traits, or, probably most commonly, transgressive trait expression in hybrid lineages (Gross & Rieseberg, 2005). Also, modelling studies have shown that ecological and spatial isolation are required to achieve substantial reproductive isolation of incipient hybrid species (Buerkle et al. 2000), that the unavailability of a habitat different from parental habitats reduces the frequency of homoploid hybrid speciation (Buerkle et al., 2003), and that the evolution of reproductive isolation by intrinsic mechanisms alone is unlikely to lead to speciation (McCarthy et al., 1995; Buerkle et al., 2000). However, it has also been shown that reproductive isolation of hybrid populations from their parents by genetic incompatibilities can originate through hybridization (Schumer et al., 2015). The effect of interspecific hybridization on ecology and geographical distribution has also been illustrated in studies demonstrating range expansion of species through hybridization (Ma et al., 2019; Pfennig et al., 2016; Suarez-Gonzalez et al., 2016; Sun et al., 2020; Whitney et al., 2015). It, thus, seems that a possibly important result of interspecific hybridization is ecogeographical isolation of the hybridization product from its parents. As this ecogeographical isolation generally is assumed to be a result of hybridization, it would also satisfy the demand by Schumer et al. (2014) that demonstration of homoploid hybrid speciation requires demonstration of hybridization-derived reproductive isolation (for further discussion see Nieto Feliner et al., 2017 and Schumer et al., 2018).

Sempervivum tectorum L. (Houseleek), one of 46 species (Klein & Kadereit, 2015) of *Sempervivum* L. (Crassulaceae), is widespread in open rocky habitats at mostly high altitudes across the European high mountains (mainly Pyrenees, Alps and Apennine, more rarely Balkans). The species is also widely cultivated and has become naturalized outside its natural range (Parnell & Favarger, 1993). The age of the clade *S. tectorum* belongs to has been dated to within the last 0.5 million years ago (ma; Klein & Kadereit, 2015). In *Sempervivum*, interspecific hybridization is very common ('t Hart et al., 2003; Klein & Kadereit, 2016), and allopatric distributions of hybrid individuals and one or both parental taxa have been documented in several instances. Such distributions imply large-scale geographical range shifts of either hybrid or parent(s) (Klein & Kadereit, 2016). In the western part of its range, *S. tectorum* is highly disjunctly distributed in the Rhine Gorge area in Germany (comprising the Upper Middle Rhine, Mosel and Ahr river valleys), the Massif Central in France and the Pyrenees in France and Spain (Figure 1). In a recent phylogenetic study of *S. tectorum* in these areas using genotyping-by-sequencing (GBS) data by Fabritzek and Kadereit (2018), which aimed at demonstrating that the Rhine Gorge material of the species represents an independent phylogenetic lineage native to that area, it was found that populations from the Massif Central might be composed of the gene pools of Rhine Gorge area and Pyrenean populations. Although

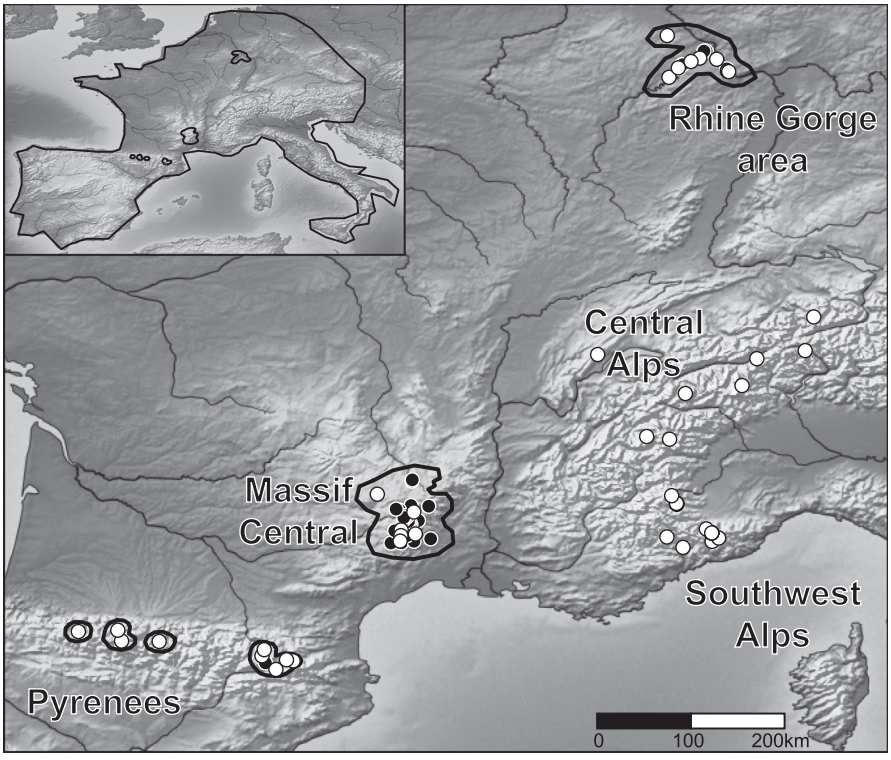

FIGURE 1 Location of study areas. White circles: sampling localities of the 52 samples included in the GBS analysis by Fabritzek and Kadereit (2018); black circles: additional occurrences recorded during field work or provided by the Parc National des Cévennes; black polygons around occurrences: background area used for Maxent modelling; inset: black line marks the area used for spatial projection of Maxent models. Samples were taken at distances of at least 1.5 km

currently no generally accepted intraspecific classification of *S. tectorum* is available (Parnell & Favarger, 1993), material from the Rhine Gorge area, the Massif Central and the Pyrenees has been recognized at specific (and also other) rank in the past: *S. rhenanum* (Hegi & Schmid ex Hayek) Lawalrée for the Rhine Gorge material, *S. avernense* Lec. and Lamotte for the Massif Central material, and *S. boutignyarum* Gren. & Billot for the Pyrenean populations. These three geographical lineages were clearly distinct in the phylogenetic analysis of the species (Fabritzek & Kadereit, 2018) and are reproductively isolated from each other by their widely allopatric distribution. Accordingly, we here do not investigate the hybrid origin of a generally accepted species but are dealing with a case of incipient speciation. The analysis of young hybrid taxa has been considered ideal for elucidating early steps in hybrid species evolution (Abbott et al., 2010, 2013; Nolte & Tautz, 2009).

The overall aim of the present study of *S. tectorum* is to explore the possibility that novel trait combinations in hybrids resulted in ecological differentiation from their parents, which in turn, in response to Quaternary climate changes, led to geographical displacement. In particular, we will (a) further analyse the GBS genetic data obtained by Fabritzek and Kadereit (2018) in order to establish that the Massif Central lineage shows hybrid admixture. (b) Using climatic data obtained from the WorldClim database (Hijmans et al., 2005) as well as soil pH and soil temperature data collected by us in the field, we investigate, using different approaches including niche modelling, whether the ecological niche of the Massif Central lineage is different from that of the Rhine Gorge and Pyrenees lineages, and whether niche differences can plausibly be interpreted as the result of hybridization by being intermediate, parental-combined or extreme as observed in hybrids and hybrid progeny by other authors (Rieseberg et al., 1993; Gross & Rieseberg, 2005; Nolte & Tautz, 2009; Abbott et al., 2010, 2013). (c) Using the ecological data in a species distribution modelling approach, we then investigate whether ecological differentiation of the Massif Central lineage resulted in its geographical displacement from the Rhine Gorge and Pyrenees lineages at the likely time of its origin in the Quaternary.

2 | MATERIAL AND METHODS

2.1 | Genetic characterization of lineages

2.1.1 | Samples and genotyping-by-sequencing analysis

Fabritzek and Kadereit (2018) conducted a GBS analysis using 52 accessions of *Sempervivum tectorum* sampled in the western part of its overall geographical range (Figure 1). Of these, eight samples were from the Rhine Gorge (RG) area in Germany (MRT_1353, MRT_1354, MRT_1361, MRT_1391, MRT_1394, MRT_1396, MRT_1424, MRT_1425), six from the Massif Central (MC) in France (MC_1501, MC_1502, MC_1503, MC_1504, MC_1505, MC_1506) and nine from the French Pyrenees (PYR; PY_1508, PY_1509, PY_1510, PY_1511, PY_1512, PY_1513, PY_1514, PY_1515, PY_1516). The remaining 29 samples were collected in the Central Alps (C Alps, 12 samples) and Southwest Alps (SW Alps, 17 samples).

DNA extraction, library preparation for GBS and the ipyrad – GBS pipeline have been described in Fabritzek and Kadereit (2018).

2.1.2 | Phylogenetic network inference

The GBS data of the above 52 samples were used to calculate a phylogenetic network with NeighborNet as implemented in Splitstree version 4.14.2 (Huson & Bryant, 2006). The network was calculated using uncorrected P distances and assuming equal angle splits. Bootstrap support of splits was obtained from 1,000 bootstrap replicates.

2.1.3 | Inference of genetic groups and admixture among groups

Genetic groups and admixture among groups were inferred using STRUCTURE version 2.3.4 (Pritchard et al., 2000). All 23 GBS samples from RG, MC and PYR were included in this analysis without a priori assignment of genotypes to geographical regions. For more details, refer Appendix S1.

2.1.4 | Inference of admixture and introgression

As a further test of admixture and to detect potential introgression among the RG, MC and PYR lineages, we performed D-Statistics (i.e., ABBA-BABA tests; Durand et al., 2011) as implemented in ipyrad version 0.7.28 (Eaton & Overcast, 2020). D-Statistics was computed for two tree topologies. These were ((P1: MC, P2: RG), P3: PYR), outgroup: Alps) and (((P1: MC, P2: PYR), P3: RG), outgroup: Alps). As outgroup, all 29 samples from the C Alps and SW Alps were used (Fabritzek & Kadereit, 2018). For both topologies, the significance of D was tested by 1,000 bootstrap replicates in which loci were resampled with replacement. Significance was assessed for each replicate by transforming the Z-score into a two-tailed *p*-value, and using 0.01 as a conservative cut-off for significance after using Holm-Bonferroni correction for multiple comparisons (Eaton, & Ree, 2013). D-Statistics was run on the MOGON cluster at Johannes Gutenberg-Universität, Mainz.

2.1.5 | Genetic distinctness of GBS genotypes

In order to explore the genetic differentiation among the three geographical groups, principal component analysis (PCA) and discriminant analysis of principal components (DAPC) were performed using the software R for all 23 GBS samples from RG, MC and PYR. For details of methods, refer Appendix S2.

2.1.6 | Assessment of ploidy level

The R package GBS2PLOIDY version 1.0 (Gompert & Mock, 2017) was used to infer cytotypes of RG, MC and PYR from the observed heterozygosity and the allelic ratio of heterozygous SNPs. For this analysis, MC ($N = 6$) and PYR ($N = 9$) were assumed to be tetraploid ($2n = 72$) as reported by Favarger and Scherbatoff (1973). As the cytotype of RG ($N = 8$) is not known, we assumed an unknown ploidy level for all eight samples. Samples of *Sempervivum calcareum* ($N = 14$; $2n = 38$, Favarger & Scherbatoff, 1973) and of *S. marmoreum* ($N = 1$; $2n = 34$, Favarger & Zésiger, 1964) from Fabritzek and Kadereit (2018) were our diploid references. In the ipyrad GBS pipeline, we increased the minimal depth on base calling per locus from six (Fabritzek & Kadereit, 2018) to 15 (Gompert & Mock, 2017) for all samples and allowed four alleles per locus. For converting the vcf files produced to the format used by *gbs2ploidy*, the python script VCFCONVERTER2.PY version 2.2 (https://github.com/dandewaters/VCF-File-Converter) was used.

The *gbs2ploidy* pipeline was run using the settings from Gompert and Mock (2017) on aspen (*Populus tremuloides*), that is, two independent Markov chain Monte Carlo (MCMC) runs, 10,000 post-burn-in iterations for each chain, and 1,000 iterations as burn-in with a thinning interval of three. To infer probabilities of being tetraploid for RG, MC, PYR, *S. calcareum* and *S. marmoreum*, we conducted five independent MCMC runs on posterior estimates of allelic ratios. After each MCMC run, the dataset was randomly split into 50% training and 50% test samples. For each sample, 100 independent runs of ploidy estimation based on posterior estimates of allelic ratios were then carried out and the average probability of assigning the sample to a tetraploid level was calculated from runs. Finally, average assignment probabilities of samples from RG, MC, PYR, *S. calcareum* and *S. marmoreum* were averaged for each group over the five MCMC runs.

2.2 | Ecological niches and geographical distribution

2.2.1 | Ecological fieldwork

To assess microclimatic thermal conditions experienced by specimens sampled from RG, MC and PYR, we recorded the temperature of the soil in which specimens grew (hereafter soil temperature time series) between 1 October 2015 and 31 July 2017. We further took soil samples close to the roots to determine soil pH. For details of methods, refer Appendix S3.

2.2.2 | Modelling bioclimatic niches of lineages in Maxent

In order to characterize the ecological niches of the RG, MC and PYR lineages at the macroclimatic level, we established ecological niche models (ENMs) using the Maxent algorithm (Phillips et al., 2006). The basis for our three datasets on occurrences of the RG, PYR and MC lineages used for ENM development was the georeferenced locations of specimens sampled for the GBS analysis. These were eight records for the RG lineage, six for the MC lineage and nine for the PYR lineage (hereafter GBS occurrences). To each of these datasets, we added further occurrences seen by us in the field or obtained from the Parc National des Cévennes. For the RG lineage, only records from natural populations were added (refer Fabritzek & Kadereit, 2018 for details). Altogether, 14 occurrences from RG, 29 from MC and 10 from PYR were used to establish ENMs (hereafter extended occurrences, Figure 1). As potential predictors of lineage occurrences, we used the 19 bioclimatic variables (BIO1 through BIO19) for the period 1950–2000 provided by the WorldClim database (Hijmans et al., 2005). Variables capture annual trends, seasonality and extremes in temperature and precipitation. Bioclimatic datasets are available not only for current but also for past and future climates. Details on ENM development and their geographical projections for different periods (current, Mid-Holocene, Last Glacial Maximum (LGM)) are provided in Appendix S4.

We finally derived two well-constrained ENMs for each of the lineages RG, MC and PYR (Table S1, Appendix S4). Together, these made use of 10 bioclimatic variables (hereafter BIOVars), six representing the thermal niche (BIO3, BIO4, BIO6, BIO7, BIO8, BIO11; hereafter temperatureBIOVars) and four the precipitational niche (BIO12, BIO13, BIO16, BIO17; hereafter precipitationBIOVars). All subsequent analyses on the ecological distinctness of lineages were restricted to these variables.

2.3 | Tests on niche dissimilarity

To compare the microclimatic thermal niche (soil temperature series recorded close to samples, Appendix S3) and the macroclimatic thermal niche (values of bioclimatic variables close to species occurrences, WorldClim dataset) of lineages, we first calculated five climatic variables comparable to those from the WorldClim dataset (Hijmans et al., 2005) for each of the 19 soil temperature time series. Hereafter, these variables are referred to as iBIOVars and the specific variables are iBIO3, iBIO4, iBIO6, iBIO7 and iBIO11. For the calculation of iBIOVars, we applied the function *biovars* from the r package dismo version 1.1–4 (Hijmans et al., 2017). A transformed temperature series derived from the respective soil temperature series by averaging per day was passed to this function.

For the macroclimatic niche, values for each of the 10 BIOVars (those used by ENMs of lineages, Table S1) were extracted from the WorldClim database for each occurrence of the three lineages used in Maxent modelling (using GPS coordinates of the extended occurrence dataset; $N_{RG} = 14$, $N_{MC} = 29$, $N_{PYR} = 10$) using the function *getData* from the r package raster version 2.6–7 (Hijmans & van Etten, 2011). We then built three different datasets for each lineage. In the first dataset, we included bioclimatic conditions only for all 19 occurrences at which we sampled plant material for GBS analysis and had soil temperature time series (iButton dataset; $N_{RG} = 6$, $N_{MC} = 5$, $N_{PYR} = 8$). In the second dataset, we included conditions for

all 23 occurrences sampled for GBS analysis (GBS dataset; $N_{RG} = 8$, $N_{MC} = 6$, $N_{PYR} = 9$). In the third dataset, conditions of other validated occurrences were included (extended dataset, the 53 occurrences comprising the extended occurrences; $N_{RG} = 14$, $N_{MC} = 29$, $N_{PYR} = 10$).

We then used LDA to assess (multivariate) niche differences between lineages at the micro-scale and the macro-scale. For the micro-scale, we conducted one LDA with the combinations of iBIOVars values (iBIO3, iBIO4, iBIO6, iBIO7 and iBIO11) from the iButton dataset ($N_{RG} = 6$, $N_{MC} = 5$, $N_{PYR} = 8$) and another in which we added soil pH values to these combinations ($N_{RG} = 5$, $N_{MC} = 5$, $N_{PYR} = 8$). With this analysis, we aimed to investigate whether lineages are distinct with respect to their microclimatic thermal niche and whether differences in their edaphic niche alone lead or add to their potential ecological distinctness.

For differences at the macro-scale three LDAs were performed. The first used bioclimatic variables describing macroclimatic thermal conditions (temperatureBIOVars; BIO3, BIO4, BIO6, BIO7, BIO8 and BIO11) of the GBS dataset ($N_{RG} = 8$, $N_{MC} = 6$, $N_{PYR} = 9$). With this LDA, we assessed whether thermal differences among lineages were consistent between the micro-scale and the macro-scale. Two other LDAs were conducted with combinations of all BIOVars. One was done with the GBS dataset ($N_{RG} = 8$, $N_{MC} = 6$, $N_{PYR} = 9$) and the other with the extended dataset ($N_{RG} = 14$, $N_{MC} = 29$, $N_{PYR} = 10$). By comparing the LDA for the GBS dataset using only temperatureBIOVars with that using both temperatureBIOVars and precipitationBIOVars (BIO12, BIO13, BIO16 and BIO17), we explored the contribution of the thermal and precipitational niches to the overall ecological distinctness of lineages. Comparison of LDAs using all BIOVars for the GBS dataset and for the extended dataset allowed us to explore the influence of sample sizes on results.

Prior to all LDAs, we standardized values of variables to a zero mean and a variance of unity. We used Wilks' λ and one-way MANOVA (Bartlett χ^2) to test for differences between lineages. We further used a repeated k-fold cross validation for estimating the probability of assigning a combination of environmental values to the correct lineage. We chose 10 subsets (k) and calculated the average accuracy of 10 repetitions with a training/test split of 0.75/0.25 for the data. R packages *MASS* (version 7.3–51.5, Venables & Ripley, 2002), *caret* (version 6.0–84), FactoMineR (version 2.3, Lê et al., 2008), DiscriMiner (version 0.1–29, Sanchez, 2013), klaR (version 0.6–15, Weihs et al., 2005) and rrcov (version 1.5–2) were used to calculate LDAs, Wilks' λ, one-way MANOVA, and k-fold cross-validation accuracies.

2.4 | Tests of niche similarity in ENMTools

All preceding analyses ignored differences in suitability of climatic conditions for lineages (which ENMs assess by logistic values). To determine the extent of ecological similarity between RG, MC and PYR, we conducted tests of niche similarity in ENMTools (Warren et al., 2010). ENMTools provides two quantitative tests of niche similarity, that is, the identity (equivalency) test and the background (similarity) test (Warren et al., 2008). Both tests quantify niche similarity with Schoener's (1968) D and Hellinger's distance I (Warren et al., 2010). As our final ENMs of lineages shared no bioclimatic variables, except for PYR and MC, which shared BIO13 (Precipitation of the Wettest Month) and BIO16 (Precipitation of the Wettest Quarter, Table S2)), we passed all 10 BIOVars to ENMTools (Table S2). To assess pairwise potential differences between lineages in their thermal niche (temperatureBIOVars), their precipitational niche (precipitationBIOVars) and in both (BIOVars), we carried out the identity test and the symmetric background test (lineage X background vs. lineage Y background, X,Y ϵ {RG, MC, PYR}) for all three sets of variables. To build lineage models in ENMTools, we used the Maxent algorithm with the *species files* (occurrences of lineages) and *background files* (background raster of lineages) that we used for ENM development (Appendix S4). To test the significance of D and I values, we ran the identity test and the background test with 500 replications in order to create pseudoreplicate null distributions. ENMTools was run on the MOGON cluster at Johannes Gutenberg-Universität, Mainz.

2.5 | Potential present and past distribution

The final ENMs derived from Maxent modelling of lineages (Table S1) were used to visualize the potential current and past (Mid-Holocene and LGM) distribution of lineages (i.e., areas of climatic suitability for the lineages) in Central and Southwest Europe and the Alps. Therefore, the two ENMs obtained for each lineage were evaluated for the respective WorldClim datasets under different Global Circulation Models (GCM). For details, refer Appendix S5.

3 | RESULTS

3.1 | Genetic characterization of lineages

3.1.1 | Phylogenetic network inference

The NeighborNet network of *S. tectorum* revealed five distinct clusters of GBS genotypes that corresponded to the geographical regions RG, MC, PYR, C Alps and SW Alps. The network indicated strong genetic differentiation between and weak differentiation within geographic clusters. Bootstrap support values for the RG, MC and PYR clusters were 100%. The cluster with all MC genotypes was located between the RG and PYR clusters, and was closer to RG than to PYR (Figure 2).

3.1.2 | Inference of genetic groups and admixture among groups

Structure analysis of GBS genotypes from RG, MC and PYR yielded an optimal number of two ($K = 2$, "delta K" criterion, Evanno et al.,

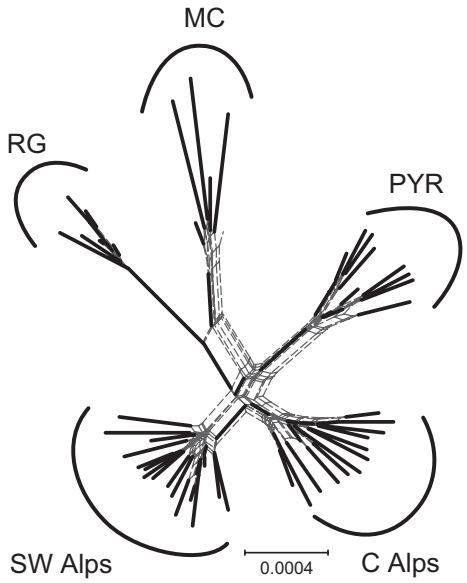

FIGURE 2 NeighborNet network of 52 GBS genotypes of *Sempervivum tectorum* from the Rhine Gorge area (RG), Massif Central (MC), Pyrenees (PYR) and Central (C) and Southwest (SW) Alps. Bootstrap support values ≥ 70% are in bold

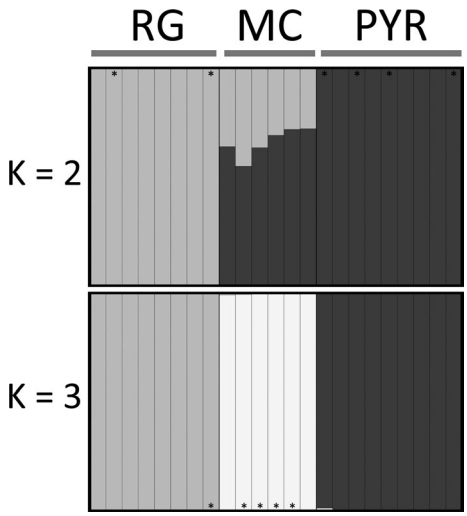

FIGURE 3 Structure analysis of 23 GBS genotypes from the Rhine Gorge area (RG), Massif Central (MC) and Pyrenees (PYR). The optimal number of clusters was $K = 2$ (Evanno et al. 2005) or $K = 3$ (Falush et al., 2003). Asterisks indicate small proportions (≤ 0.1%) of other clusters

2005) or three genetic groups ($K = 3$, "probability of K" criterion, Falush et al., 2003). Admixture of MC was strong under the "delta K" criterion where MC was admixed between RG and PYR with a higher amount of PYR. For $K = 3$, only a very small amount of admixture was seen between RG and MC, and between MC and PYR. RG, MC and PYR formed three genetically essential distinct groups (Figure 3).

3.1.3 | Inference of admixture and introgression

D-statistics of GBS genotypes yielded significant D values for both tree topologies tested. The D value derived for the ((((MC, RG), PYR), Alps) topology indicated low introgression for PYR and MC (9,608 significant tests out of 143,055 individual tests, 7%). For the (((MC, PYR), RG), Alps)) topology, D indicated high introgression for RG and MC (102,179 significant tests out of 143,055 individual tests, 71%).

3.1.4 | Genetic distinctness of GBS genotypes

Both the PCA and DAPC of the 23 GBS genotypes indicated a strong genetic distinctness of RG, MC and PYR (Figure 4). Wilks' λ was small and highly significant for the PCA (Wilks' $\lambda = 0.011$, $\chi^2 = 81.422$, $p < 10^{-12}$, for the first five principal components suggested by the Kaiser-Guttman criterion, the broken stick criterion suggested only one component). For the DAPC Wilks' λ was somewhat larger and again highly significant (Wilks' $\lambda = 0.05$, $\chi^2 = 58.611$, $p < 10^{-11}$, for two linear discriminants, LDA step done with the first five principal components recommended by cross-validation). All genotypes were correctly classified with respect to lineages (ratio within group to between group variability = 0.305, *k*-fold cross validation accuracy = 1, *SD* (*k*-fold cross validation accuracy) = 0). PCA and DAPC placed GBS genotypes of MC between those of RG and PYR on PC1 and LD1, respectively (Figure 4).

3.1.5 | Assessment of ploidy level

The high assignment probability estimate of 0.959 (standard deviation, $SD = 0.090$) indicated a tetraploid level for RG. Consistent with reported chromosome numbers (Favarger & Scherbatoff, 1973), assignment probabilities to a tetraploid level were high for MC (0.977, $SD = 0.038$) and PYR (0.997, $SD = 0.009$) and were low for our diploid references *S. calcareum* (< 0.001, $SD < 0.001$; Favarger & Scherbatoff, 1973) and *S. marmoreum* (0.191, $SD = 0.011$; Favarger & Zésiger, 1964).

3.2 | Ecological distinctness of the RG, MC and PYR lineages

3.2.1 | Ecological fieldwork

Soil temperature was successfully recorded for six specimens from RG, five from MC and eight from PYR. These soil temperature series were used to establish iBIOVars for the micro-climatic niche of lineages (Figure 5a). Soil pH determination yielded seven values for RG, six for MC and nine for PYR. Values were used to describe the edaphic niche of lineages (Figure 5b).

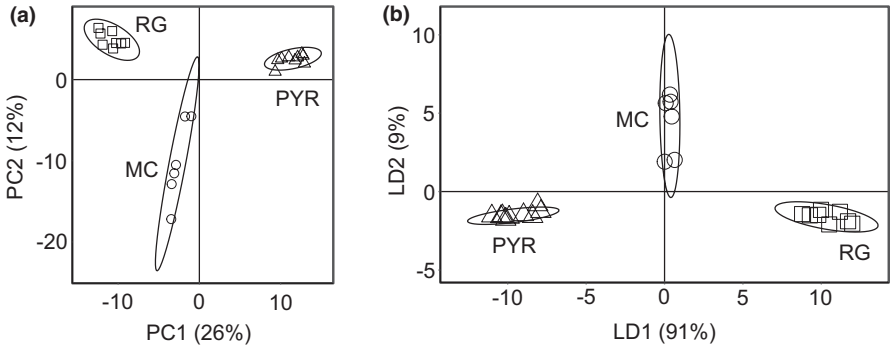

FIGURE 4 Genetic distinctness of 23 GBS genotypes from the Rhine Gorge area (RG), Massif Central (MC) and Pyrenees (PYR) as shown by a) PCA and b) DAPC. 95% confidence level eclipses were plotted around genotypes

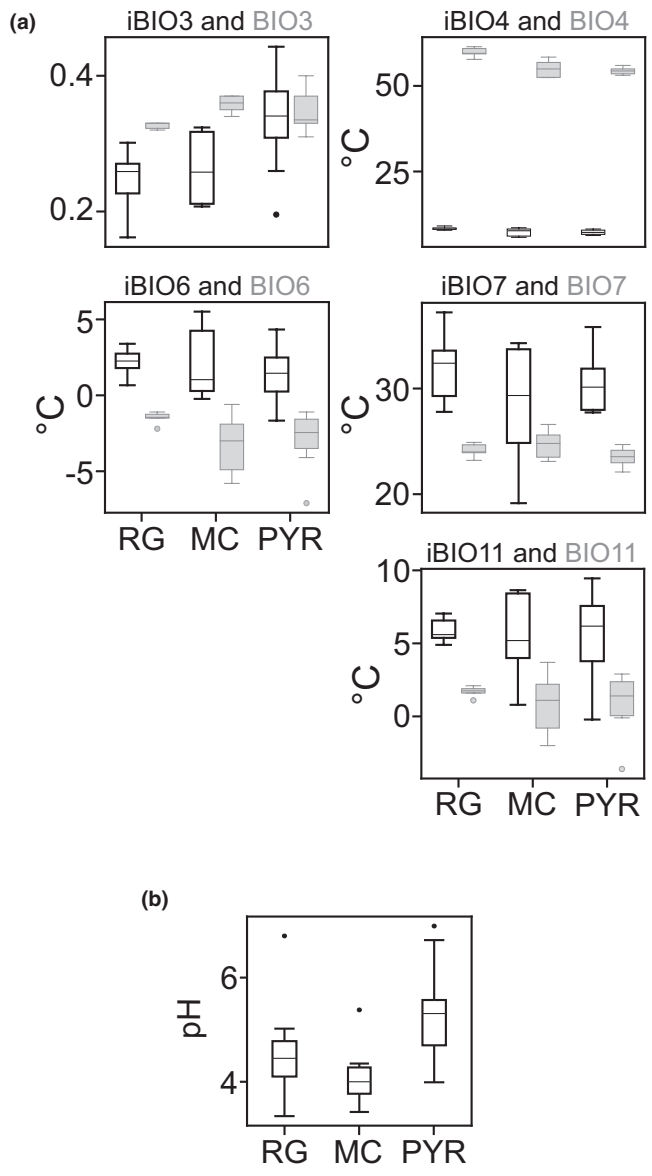

FIGURE 5 Environmental preferences of the Rhine Gorge area (RG), Massif Central (MC) and Pyrenees (PYR) lineages. Shown are medians, quartiles, 1.5-fold interquartile range and extreme values. a) Thermal preferences of lineages at the micro-scale (white, iBIOVars, soil temperature) and macro-scale (grey, BIOVars) for the iButton dataset. Kruskal-Wallis H-statistics on lineage differences in medians of iBIOVars and of BIOVars are given in Table S2. (b) Edaphic preferences of lineages based on rhizosphere soil pH

3.2.2 | Modelling bioclimatic niches of lineages in Maxent

Ecological niche modelling yielded two final ENMs for each of the three lineages (Table S1). Their average bootstrapping AUCs ranged between 0.793 and 0.85. High AUCs indicated that all models were good. Average correct classification rates of all ENMs were large (79.31% to 92.86% of occurrences were correctly predicted), except for that for the second MC model (65.52%). Transferability of models to "novel" environments was good (ratios of Test AUC and Training AUC values ranged from 1.04 to 1.12, internal validation of ENMs).

The ENM of RG with the highest AUC and the highest correct classification rate used BIO6 (Min Temperature Coldest Month) and BIO12 (Annual Precipitation) and the second best used BIO17 (Precipitation of Driest Quarter) and BIO11 (Mean Temperature Coldest Quarter), with variables of both models ordered by their descending contribution to the training gain. The best ENM of MC used BIO8 (Mean Temperature of Wettest Quarter), BIO3 (Isothermality), BIO7 (Temperature Annual Range) and BIO16, and the second best used BIO13, followed by BIO3 and BIO8. The ENMs of PYR used BIO13 (Precipitation of Wettest Month) followed by BIO4 (Temperature Seasonality) in the best and BIO16 (Precipitation Wettest Quarter) followed by BIO4 in the second-best model. In summary, the six final ENMs made use of 10 BIOVars of which six were temperature-related (BIO3, BIO4, BIO6, BIO7, BIO8, BIO11; temperatureBIOVars) and four precipitation-related (BIO12, BIO13, BIO16, BIO17; precipitationBIOVars). Only PYR and MC shared BIO13 and BIO16 (Table S1).

3.3 | Tests on niche dissimilarity

3.3.1 | Microclimatic and edaphic niche

With respect to distinctness of lineages in their thermal (Figure 5a) and edaphic preferences (micro-scale), RG, MC and PYR were similar with respect to each of the iBIOVars (except for RG and PYR which differed significantly for iBIO4, Table S2, Appendix S5). A Kruskal-Wallis H-test on differences in the edaphic niche of lineages indicated that their soil pH preferences differed significantly among each other ($H = 5.984$, $p = 0.05$; $N_{RG} = 7$, $N_{MC} = 6$, $N_{PYR} = 9$; posthoc

pairwise tests: $p < 10^{-16}$ for RG vs. MC, for MC vs. PYR and for RG vs. PYR, Figure 5b). MC soil pH was extreme in relation to RG and PYR (Figure 5b).

Linear discriminant analysis revealed marginally significant distinctness of RG, MC and PYR for microclimatic thermal conditions (iBIOVars; Table S3, Figure 7a). Adding soil pH increased lineage distinctness which, however, was not significant (Table S3, Figure 7b). In both LDAs on the micro-scale (Figure 7a,b), specimens from MC showed the largest overall variability in environmental conditions, which considerably overlapped with those of RG and PYR, whereas conditions used by RG and PYR differed much more with only a small overlap of 95% confidence interval eclipses. This result was consistent with the very poor correct classification rates derived from k-fold cross-validation analysis of both LDAs (around 0.54, Table S3).

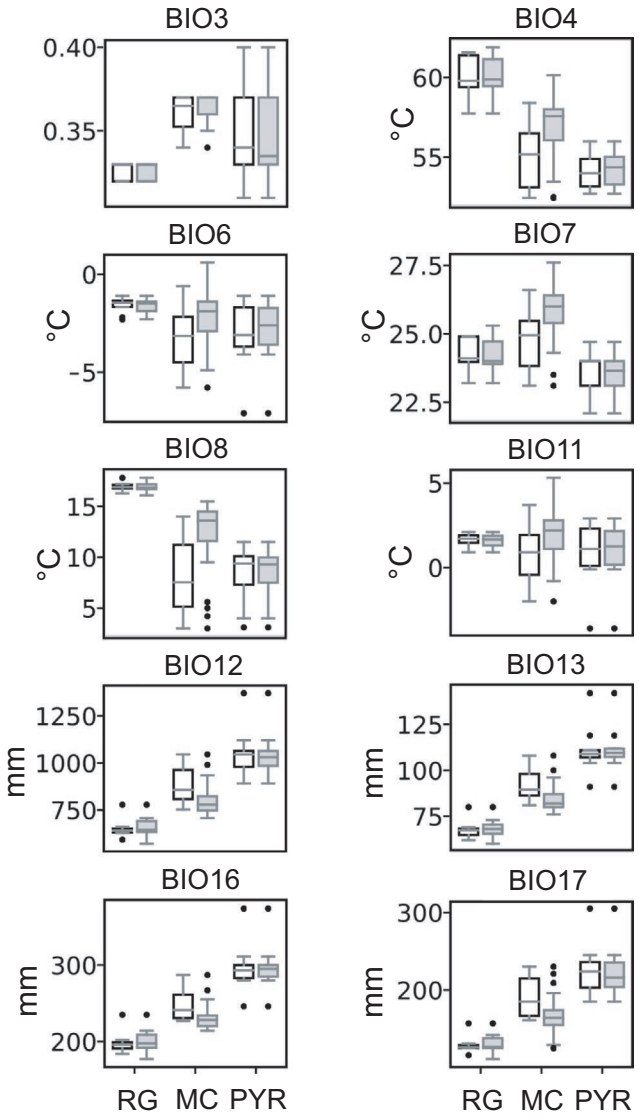

FIGURE 6 Thermal and precipitational preferences of the Rhine Gorge area (RG), Massif Central (MC) and Pyrenees (PYR) lineages at the macro-scale for the GBS (white) and extended (grey) datasets. Shown are medians, quartiles, 1.5-fold interquartile ranges and extreme values. Kruskal-Wallis H-statistics on differences in medians of BIOVars of lineages are found in Table S2

Although not statistically significant, both LDAs on the micro-scale showed differences in the microclimatic thermal niches of RG and PYR, whereas the thermal niche of MC combined that of RG and PYR (Figure 7a,b). Adding the edaphic niches increased distinctness of the three lineages and reflected the preference for more acid soils of MC (Figures 5b and 7b).

3.3.2 | Macroclimatic niche

At the macro-scale and when using the extended dataset, MC was intermediate between RG and PYR for BIO4 and BIO8. RG, MC and PYR did not differ for BIO6 and BIO11, and MC was extreme in BIO3 and BIO7 (Figure 6, Table S2). For all precipitationBIOVars (Figure 6), MC was intermediate between RG and PYR in both the GBS and the extended dataset.

All three LDA plots for macroclimatic conditions for the GBS dataset clearly showed distinctness of the three lineages (Table 3, Figure 7c–e), irrespective of which dataset and whether only all temperatureBIOVars or BIOVars were analysed. In all LDA plots, the 95% confidence interval eclipses of the three lineages showed no pairwise overlap (Figure 7c–e). All Wilks' λ values were small and highly significant, and all correct classification rates were high (around 0.9, Table S3). For the GBS dataset, combining all temperatureBIOVars and precipitationBIOVars resulted in a Wilks' λ nearly two magnitudes smaller than that resulting from temperatureBIOVars only (Table S3). For this dataset, the LDA plot of all temperatureBIOVars indicated an intermediate niche for PYR, which was more similar to that of MC than to that of RG (Figure 7c). Consistent with the substantial decrease in Wilks' λ, addition of all precipitationBIOVars led to a stronger separation of PYR and MC without changing the position of MC closer to PYR than to RG (Figure 7d). The distinctness of lineages was similar when all temperatureBIOVars and precipitationBIOVars were analysed together for the extended dataset (Figure 7e), although Wilks' λ here was almost a magnitude higher than that obtained for the GBS dataset. All this suggested that RG is clearly distinct from both MC and PYR with respect to the six temperature-related bioclimatic variables, while differences in thermal and precipitational conditions increase distinctness of the MC and PYR lineages (Figure 7c–e).

3.4 | Tests of niche similarity in ENMTools

All results obtained from the identity and (symmetric) background tests indicated significant differences between the niches of the three lineage pairs. D values were always smaller than I values and thus indicated a smaller niche overlap irrespective of which bioclimatic variable set (all BIOVars, only all temperatureBIOVars, only all precipitationBIOVars) was analysed (Table S4). When using all BIOVars both D and I indicated small niche overlap between RG and MC, intermediate overlap between RG and PYR, and large overlap between MC and PYR (Table S4). This pattern was also found when

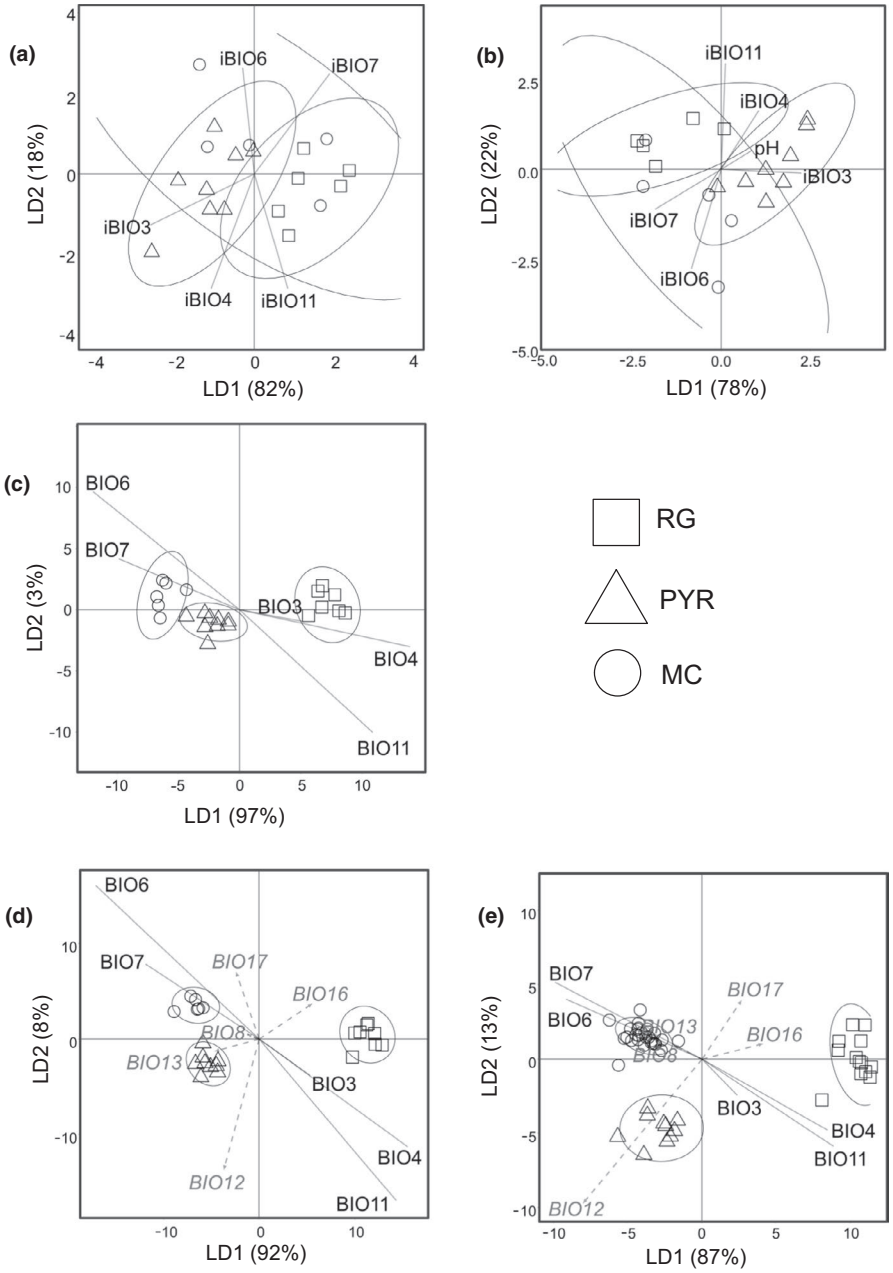

FIGURE 7 LDA analysis on ecological distinctness of the Rhine Gorge area (RG), Massif Central (MC) and Pyrenees (PYR) lineages. (a) Thermal niche derived from the iButton dataset for iBIOvars. (b) Thermal and edaphic niche (soil pH) derived from the iButton dataset. (c) Thermal niche derived from the GBS dataset. (d) Thermal and precipitational niche derived from the GBS dataset and (e) from the extended dataset. LDA statistics on lineage distinctness and *k*-fold cross validation are shown in Table S3. LDAs in (c) through (e) were done with BIOVars. BIOVars in grey and with dashed lines refer to precipitation, all others to temperature. 95% confidence level eclipses were plotted around lineages

using only all temperatureBIOVars, but the amount of niche overlap between RG and PYR and between MC and PYR was larger than for all BIOVars (Table S4). When using only all precipitationBIOVars, niche overlap between RG and MC and between RG and PYR was more similar than for all BIOVars (Table S4). For all three lineage pairs, niche overlap indicated by D and I was always larger for all precipitationBIOVars than for all temperatureBIOVars and for all these variables combined (Table S4). This suggests that temperature-related variables contributed more than precipitation-related variables to the ecological distinctness of the RG, MC and PYR lineages. Overall, the thermal and precipitational niche of MC was much more similar to that of PYR than to that of RG, which corroborated the LDA results (Figure 7e). Differences in the thermal and precipitational niches of RG and PYR were moderate, and these niches were similar as shown by the identity and background test, but the LDA suggested similarly strong differences between RG and PYR and between RG and MC.

3.5 | Potential present and past distribution

The potential present and past (LGM, Mid-Holocene) spatial distributions (i.e., areas with climatic conditions suitable for occurrence) of the RG, MC and PYR lineages are shown in Figure 8a. The Mid-Holocene and LGM maps show areas in which both final ENMs of lineages (Table S1) predicted a presence for each of the three GCMs considered by us (refer Appendix S4). The present map shows presences predicted by both of its two ENMs under present climatic conditions. Figure 8b shows the potential past spatial distribution of the three lineages in the Iberian Peninsula.

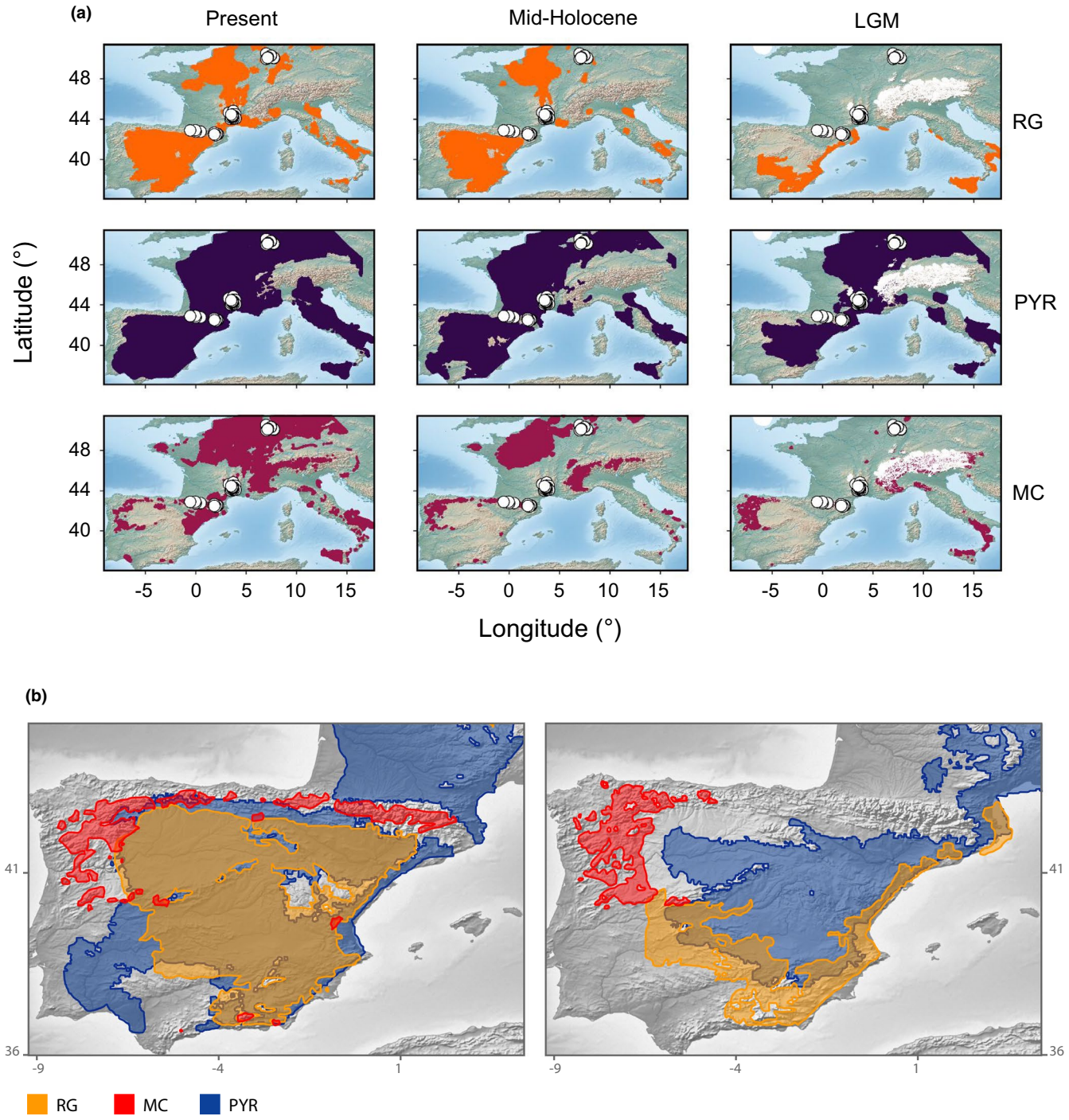

FIGURE 8 (a) Modelled potential present and past (LGM, Mid-Holocene) spatial distribution of areas with climatic conditions suitable for the Rhine Gorge area (RG), Massif Central (MC) and Pyrenees (PYR) lineages. The Mid-Holocene and LGM maps show areas in which both final ENMs of lineages (Table S1) predicted suitable conditions under each of the three GCMs used (refer Appendix S5). The present map shows suitable areas predicted by ENMs for lineages under present climatic conditions. For sampling localities see Figure 1. (b) As 8a for the Iberian Peninsula only. Left: Mid-Holocene; right: LGM. Orange: RG, red: MC, blue: PYR

4 | DISCUSSION

4.1 | A hybrid origin of the MC lineage of *S. tectorum*

All analyses of our GBS data show that the MC lineage of *S. tectorum* owes its genetic identity to hybridization between the RG and PYR lineages. D-statistics as the most direct test of hybridization revealed 71% significant tests between RG and MC and 7% significant tests between MC and PYR. This result is also reflected in the Splitstree analysis (Figure 2). Here, MC is linked to both RG and PYR, but more strongly to the former. This asymmetrical relationship of MC is also reflected in the finding by Fabritzek and Kadereit

(2018) that in an SVDquartets analysis of a larger GBS dataset (also containing samples of *S. tectorum* from the Alps and samples of the distantly related *S. calcareum* and *S. marmoreum*), considered to represent a species tree, MC is more closely related to RG than to PYR. In the Structure analysis using samples from only RG, MC and PYR (Figure 3), MC contains a somewhat larger proportion of PYR than RG genetic material at $K = 2$, the most likely K using Evanno et al. (2005). At $K = 3$, the most likely K using Falush et al. (2003), MC is essentially identified as a separate lineage with a very small amount of admixture with RG, and PYR showing a small amount of admixture with MC. In an earlier structure analysis including samples from the Alps, Fabritzek and Kadereit (2018) found a somewhat larger proportion of RG than of PYR in MC at $K = 4$, the most likely K using Evanno et al. (2005), and identified MC as a separate lineage with only very small amounts of admixture from other clusters at $K = 7$, the most likely K using Falush et al. (2003). Identification of admixed lineages as separate lineages at higher values of K has also been observed by Ma et al. (2019) in their study of hybrid populations in a species of cypress. Finally, in both the PCA and DAPC analyses of the GBS data (Figure 4), MC is clearly intermediate between RG and PYR on PC1/LD1 but not on PC2/LD2. All these results imply that the MC lineage most likely is the result of hybridization between RG and PYR. As all three lineages most likely have the same ploidy level as estimated from allelic ratios, confirming chromosome number reports for MC and PYR by Favarger and Scherbatoff (1973), this has been homoploid hybridization. However, as D-statistics implies high levels of introgression between RG and MC and low levels between MC and PYR, a result also reflected in Splitstree, and results from Structure, PCA and DAPC imply a near-equal contribution of RG and PYR to MC, we cannot finally decide whether the MC lineage is the result of hybridization with equal contributions from both parents or of introgressive hybridization with a minor contribution from only one parent. Either way MC will have undergone independent evolution after its origin through hybridization.

4.2 | The ecological niches of RG, MC and PYR

As evident from the LDAs of both the GBS and extended datasets, the niches of RG, MC and PYR are different (Figure 7). Our tests of niche similarity in ENMTools (Table S4, Warren et al., 2010) also revealed significant differences between the niches of the three lineages although small niche overlap between RG and MC, intermediate overlap between RG and PYR, and large overlap between MC and PYR were found when considering, for example, all BIOVars. When considering individual bioclimatic and edaphic variables at the micro-scale, that is, using only those samples for which soil temperature data and soil samples have been collected, MC is more variable than RG and PYR, and its variation more or less encompasses the variation of these two lineages (Figures 4 and 5). MC is extreme for soil pH (Figure 5) and in comparison with RG and PYR grows on more acid soils. At the macroclimatic scale (Figure 6), MC is intermediate between RG and PYR for BIO4 (Temperature Seasonality), BIO8 (Mean Temperature of Wettest Quarter), BIO12 (Annual Precipitation), BIO13 (Precipitation of Wettest Month), BIO16 (Precipitation of Wettest Quarter) and BIO17 (Precipitation of Driest Quarter), and extreme in comparison with RG and PYR in BIO3 (Isothermality) and BIO7 (Temperature Annual Range). When considering overall niche similarity as assessed with ENMTools, with small niche overlap between RG and MC (Table S4), this clearly is not congruent with some analyses of our genetic data, where D-statistics and Splitstree had implied a close relationship between RG and MC. On the other hand, the often intermediate niche at the macroclimatic scale of MC probably reflects the Structure and PCA/DAPC results from our genetic data. In summary, our ecological data illustrate that the MC lineage, by being partly intermediate between the parental lineages, partly more variable than the parental lineages and partly extreme in relation to the parental lineages, shows a pattern which is fully in line with what has been observed in hybrids and hybrid progeny by other authors (Rieseberg et al., 1993; Gross & Rieseberg, 2005; Nolte & Tautz, 2009; Abbott et al., 2010, 2013). This pattern, on the background of our genetic data, can thus plausibly be interpreted as hybridization-derived.

4.3 | The geographical setting of hybridization between RG and PYR

After having established that the MC lineage of *S. tectorum* most likely originated through homoploid hybridization between RG and PYR, and considering the highly disjunct distributions of RG, MC and PYR today (Figure 1), the question arises where this hybridization might have taken place.

Following Birks and Willis (2008) and Tzedakis et al. (2013), many of the open LGM environments in Europe would have been suitable for the growth and widespread distribution of Alpine plants. The potential widespread LGM distribution modelled by us particularly for PYR (Figure 8) clearly fits this hypothesis. Also, the possibly widespread LGM distribution of PYR in the Iberian Peninsula is supported by the finding that a sample of *Sempervivum* from the Sierra Nevada (southern Spain), classified as *S. minutum* (Kunze ex Willk.) Nyman ex Pau, was found to group within PYR (Figure S1). More generally, Kropf et al. (2006, 2008, 2012) postulated continuous distribution ranges in Quaternary glacials of high mountain plant taxa between the Massif Central, the Pyrenees and the Sierra Nevada, implying widespread distributions of such taxa in the Iberian Peninsula. However, our modelling of ecological niches and geographical distributions is based on climatic data only and thus ignores, for example, solar radiation, topography and edaphic conditions which all will be relevant considering that *S. tectorum* is a species of open rocky places at mostly high altitudes.

The modelled potential distribution of RG and PYR in the LGM showed substantial overlap in the Iberian Peninsula, the southern Apennine and on Sicily (Figure 8) where hybridization might have taken place. The modelled potential distribution of MC populations in the LGM (Figure 8) shows that this lineage could have been

distributed largely parapatrically (sensu Mallet et al., 2009) with both parental lineages with a small area of sympatry of all three lineages at the western end of the Sierra de Gredos of the Iberian Central System (ICS) in Spain and some overlap in the southernmost Apennine and on Sicily.

Considering the extant geographical distribution of the three lineages (RG, MC, PYR) in western Europe (Figure 1) and the fact that Apennine populations of *S. tectorum* were never found to be closest relative to any of them by Fabritzek and Kadereit (2018; see Figure S1), we assume that an origin of the MC lineage in Iberia is more likely than its origin in the southern Apennine or on Sicily. Also, close relationships between populations from the Iberian Peninsula, the Massif Central and various parts of Central Europe have been reported for several plant and animal taxa (Hewitt, 2000; Taberlet et al., 1998), and Iberian-Pyrenean-Massif Central (Zetzsche, 2004), Pyrenean-Massif Central (Valtueña et al., 2012) and Massif Central-Central European (Huck et al., 2009) relationships have been reported for other plant species.

The modelled potential LGM distribution of RG, MC and PYR may then imply that the small area of overlap with climatic conditions suitable for the three lineages at the western end of the Sierra de Gredos (Figure 8b) may represent the area in which hybridization took place should the MC lineage have originated in the last Quaternary glacial (or in an earlier Quaternary glacial assuming that range shifts were similar in different glacials).

Should hybridization between RG and PYR have taken place in a warmer period similar to the mid-Holocene for which we modelled potential distributions (Figure 8), these two lineages could have overlapped in the Iberian Peninsula, the Apennine and on Sicily, and in large areas north of the Pyrenees, particularly large lowland areas mainly in northern France. All these areas have modelled climatic conditions suitable for the three lineages. Following our argumentation above, the Apennine and Sicily can be excluded from further consideration. A mid-Holocene distribution of RG and PYR in lowland areas of northern France (and further east in case of PYR) seems unlikely because this area was also modelled as a potential extant distribution area of the two lineages (Figure 8), but today *S. tectorum* cannot be found there. This possibly can be explained either with the absence in this area of suitable habitat, that is, open rocky places, for this species, or with its inability to migrate into such habitats in warmer periods with higher levels of competition in intervening nonrocky habitats. Also, as stated above, our modelling is based on climatic data only. Although an origin of MC through hybridization between RG and PYR under warmer climatic conditions north of the Pyrenees cannot be entirely excluded, the Iberian Peninsula as area of origin again seems more likely. Quite intriguingly, the modelled potential mid-Holocene distribution in the Iberian Peninsula of MC on the one hand and RG and PYR on the other hand again, as found in our LGM models, is largely parapatric with a small area of sympatry or close proximity of all three lineages in the Galician-Portuguese Mountains (Macizo Galaico-Leonés) in northwest Spain (Figure 8b), which may represent the area in which hybridization took place. Both the Sierra de Gredos and the Galician-Portugies Mountains, modelled as potential areas of origin of MC, provide open rocky habitats as required by the species.

Considering, however, that our distribution models are based on climatic data only, the observation of a parapatric distribution of the hybrid lineage with its parental lineages in both the LGM and the mid-Holocene, although intriguing, should not be overinterpreted.

4.4 | Hybridization, niche evolution, ecogeographical displacement and the origin of MC

The finding of a hybrid origin of MC, its possibly parapatric distribution in relation to RG and PYR in colder times of the past, and its ecological distinctness from its parents leads us to postulate the following scenario for the origin of MC. (a) The large areas of potential overlap of RG and PYR, which today are widely allopatric in distribution, in the Iberian Peninsula in the LGM and mid-Holocene provided the opportunity for between-lineage hybridization. This confirms the importance of between-lineage contact as a consequence of distributional change in response to climate change (Anderson, 1948; Anderson & Stebbins, 1954; Hewitt, 2011). (b) The hybrid lineage, MC, is ecologically different from its parents. It is intermediate in some characters, encompasses parental properties in others, and is extreme (novel) in yet others. Such set of character expression – intermediate, parental-combined, extreme (novel) – of hybrids and hybrid progeny in comparison with their parents has been observed frequently (Rieseberg et al., 1993) and is highly likely to be the immediate result of hybridization and of genetic changes in later hybrid generations (Abbott et al., 2010, 2013; Nolte & Tautz, 2009). All these observations for MC underline that evolutionary novelty is an emergent property of hybrids (Abbott & Brennan, 2014; Anderson & Stebbins, 1954; Arnold, 1997; Ehrendorfer, 1980; Soltis et al., 2014). (c) The divergent ecology of the hybrid lineage resulted in geographical displacement. While the lineage must have originated in sympatry with the parental lineages, it either dispersed into more suitable habitats under climatically stable conditions or obtained its parapatry and later allopatry in relation to the parental lineages as a consequence of climatic changes in the Quaternary to which the hybrid and parental lineages responded differently (Kadereit, 2015). Comparison of the modelled LGM and extant distributions of the three lineages (Figure 8) clearly illustrates segregation of sympatric/parapatric distribution areas due to differential response of the three lineages to climate change, that is, the transition from a cold to a warmer climate in the Quaternary. Extant MC also has the lowest soil pH requirements of all three lineages (Figure 5b). As either the Sierra de Gredos (LGM) or the Galician-Portuguese Mountains (mid-Holocene) have been identified as possible areas of origin of MC, and both these mountain ranges are largely granitic (IGME, 2015), it seems possible that differing soil pH requirements of MC (Figure 5b) played a large role in its ecogeographical displacement. Intriguingly, ecogeographical displacement here was inferred from SDMs building on climatic data, but the potential displacement areas very well fit the extreme

edaphic niche of the emerging hybrid lineage. The origin of extreme soil requirements appears to have been of outstanding importance also in the evolution of homoploid hybrid species in *Helianthus* (Rieseberg et al., 2003). However, as pointed out above, identification of the Sierra de Gredos or the Galician-Portuguese Mountains as possible areas of origin of the MC lineage is far from certain. (d) Ecogeographical displacement of the hybrid lineage, possibly catalysed by past climatic change, resulted in reproductive isolation from the parental lineages and thus initiated the independent evolution of MC, which had been identified as a distinct phylogenetic lineage by Fabritzek and Kadereit (2018) and as a distinct genetic cluster at some values of K (refer above).

ACKNOWLEDGMENTS

We are grateful to the following agencies for granting permission to collect plant material. Germany: Regierungspräsidium Darmstadt, Struktur- und Genehmigungsdirektion Nord - Koblenz, Struktur- und Genehmigungsdirektion Süd - Neustadt an der Weinstraße, Regierungspräsidium Karlsruhe; Switzerland: Parc Jura Vaudois, L'ufficio della natura e del paesaggio - Cantone Ticino; Italy: Ministero dell'Ambiente e della tutela del Territorio e del Mare, Giunta Regionale Regione Liguria, Parco Naturale Alpi Marittime, Parco dell'Aveto, Parco Corno alle Scale, Parco Nazionale Gran Paradiso, Parco Nazionale dello Stelvio; France: Parc national des Cévennes. Very helpful comments on an earlier version of this paper by Richard J. Abbott (St. Andrews/Scotland), the editor and two anonymous reviewers are gratefully acknowledged. We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support (grant KA 635/17-1&2 to JWK). Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare we have no conflict of interest.

AUTHOR CONTRIBUTIONS

JWK designed the study; AGF collected the data; AGF and EMG analysed the data; and AGF, EMG and JWK wrote the manuscript; AFG and EMG contributed equally to this paper.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jeb.13784.

DATA AVAILABILITY STATEMENT

All genetic data are available as demultiplexed raw reads in an NCBI sequence read archive as accession SRP135257. Soil pH- and raw-field temperature time series data are available on Dryad (https://doi.org/10.5061/dryad.c866t1g66).

ORCID

Armin G. Fabritzek https://orcid.org/0000-0002-0199-6802
Eva Maria Griebeler https://orcid.org/0000-0003-2166-487X
Joachim W. Kadereit https://orcid.org/0000-0002-9274-3359

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott, R. J., & Brennan, A. C. (2014). Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130346. https://doi.org/10.1098/rstb.2013.0346

Abbott, R. J., Hegarty, M. J., Hiscock, S. J., & Brennan, A. C. (2010). Homoploid hybrid speciation in action. *Taxon*, *59*, 1375–1386. https://doi.org/10.1002/tax.595005

Abbott, R. J., & Rieseberg, L. H. (2012). Hybrid speciation. In *Encyclopedia of life sciences (eLS)*. John Wiley & Sons, Ltd.

Anderson, E. (1948). Hybridization of the habitat. *Evolution*, *2*, 1–9. https://doi.org/10.2307/2405610

Anderson, E., & Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution*, *8*, 378–388. https://doi.org/10.2307/2405784

Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press.

Birks, H. J. B., & Willis, K. J. (2008). Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, *1*, 147–160. https://doi.org/10.1080/17550870802349146

Buerkle, C. A., Morris, R. J., Asmussen, M. A., & Rieseberg, L. H. (2000). The likelihood of homoploid hybrid speciation. *Heredity*, *84*, 441–451. https://doi.org/10.1046/j.1365-2540.2000.00680.x

Buerkle, C. A., Wolf, D. E., & Rieseberg, L. H. (2003). The origin and extinction of species through hybridization. In C. A. Brigham, & M. W. Schwartz (Eds.), *Population viability in plants: Conservation, management, and modelling of rare plants* (pp. 117–141). Springer. https://doi.org/10.1007/978-3-662-09389-4_5

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*, 2239–2252. https://doi.org/10.1093/molbev/msr048

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*, 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, *62*, 689–706. https://doi.org/10.1093/sysbio/syt032

Ehrendorfer, F. (1980). Polyploidy and distribution. In W. Lewis (Ed.), *Polyploidy. Basic life sciences* (Vol. *13*, pp. 45–66). Springer. https://doi.org/10.1007/978-1-4613-3069-1_3

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. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fabritzek, A. G., & Kadereit, J. W. (2018). Identity and relationships of *Sempervivum tectorum* (Crassulaceae) in the Rhine Gorge area. *Willdenowia*, *48*, 405–414. https://doi.org/10.3372/wi.48.48310

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–1587.

Favarger, C., & Zesiger, F. (1964). *Sempervivum*. In T. G. Tutin (Ed.). *Flora Europaea* (pp. 351–355). Cambridge University Press.

Favarger, C., & Scherbatoff, M. (1973). Sur une espèce intéressante des Alpes austro-occidentales: *Sempervivum calcareum* Jordan. *Candollea*, *28*, 219–235.

Gompert, Z., & Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Molecular Ecology Resources*, *17*, 1156–1167. https://doi.org/10.1111/1755-0998.12657

Gross, B. L., & Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, *96*, 241–252. https://doi.org/10.1093/jhered/esi026

't Hart, H., Bleij, B., & Zonneveld, B. (2003). *Sempervivum*. In U. Eggli (Ed.), *Illustrated handbook of succulent plants: Crassulaceae*. Springer-VerlagBerlin Heidelberg GmbH.

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907–913. https://doi.org/10.1038/35016000

Hewitt, G. M. (2011). Quaternary phylogeography: The roots of hybrid zones. *Genetica*, *139*, 617–638. https://doi.org/10.1007/s10709-011-9547-3

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *dismo: species distribution modeling*. R package version 1.1-4. Retrieved from https://CRAN.R-project.org/package=dismo

Hijmans, R. J., & van Etten, J. (2011). *Raster: Geographic analysis and modeling with raster data*. R package version 2.6-7. Retrieved from http://CRAN.R-project.org/package=raster

Huck, S., Büdel, B., Kadereit, J. W., & Printzen, C. (2009). Range-wide phylogeography of the European temperate-montane herbaceous plant *Meum athamanticum* Jacq.: Evidence for periglacial persistence. *Journal of Biogeography*, *36*, 1588–1599. https://doi.org/10.1111/j.1365-2699.2009.02096.x

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

Instituto Geológico y Minero de España (IGME), C/ RÍOS ROSAS, 23 28003 MADRID. (2015). *Mapa Geológico de la Península Ibérica, Baleares y Canarias*. Retrieved from http://info.igme.es/cartografiadigital/datos/geologicos1M/Geologico1000_(2015)/pdfs/EditadoG1000_(2015).pdf

Kadereit, J. W. (2015). The geography of hybrid speciation in plants. *Taxon*, *64*, 673–687. https://doi.org/10.12705/644.1.

Klein, J. T., & Kadereit, J. W. (2015). Phylogeny, biogeography, and evolution of edaphic association in the European oreophytes *Sempervivum* and *Jovibarba* (Crassulaceae). *International Journal of Plant Sciences*, *176*, 44–71. https://doi.org/10.1086/677948

Klein, J. T., & Kadereit, J. W. (2016). Allopatric hybrids as evidence for past range dynamics in *Sempervivum* (Crassulaceae), a western Eurasian high mountain oreophyte. *Alpine Botany*, *126*, 119–133. https://doi.org/10.1007/s00035-016-0164-8

Kropf, M., Comes, H. P., & Kadereit, J. W. (2006). Long-distance dispersal vs. vicariance: The origin and genetic diversity of alpine plants in the Spanish Sierra Nevada. *New Phytologist*, *172*, 169–184. https://doi.org/10.1111/j.1469-8137.2006.01795.x

Kropf, M., Comes, H. P., & Kadereit, J. W. (2008). Causes of the genetic architecture of south-west European high mountain disjuncts. *Plant Ecology & Diversity*, *1*, 217–228. https://doi.org/10.1080/17550870802331938

Kropf, M., Comes, H. P., & Kadereit, J. W. (2012). Past, present and future of mountain species of the French Massif Central - the case of *Soldanella alpina* L. subsp. *alpina* (Primulaceae) and a review of other plant and animal studies. *Journal of Biogeography*, *39*, 799–812. https://doi.org/10.1111/j.1365-2699.2011.02668.x

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR : An R package for multivariate analysis. *Journal of Statistical Software*, *25*, 1–18. https://doi.org/10.18637/jss.v025.i01

Ma, Y., Wang, J., Hu, Q., Li, J., Sun, Y., Zhang, L., Abbott, R. J., Liu, J., & Mao, K. (2019). Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Nature Communications Biology*, *2*, 213. https://doi.org/10.1038/s42003-019-0445-z

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*, 279–283. https://doi.org/10.1038/nature05706

Mallet, J., Meyer, A., Nosil, P., & Feder, J. L. (2009). Space, sympatry and speciation. *Journal of Evolutionary Biology*, *22*, 2332–2341. https://doi.org/10.1111/j.1420-9101.2009.01816.x

McCarthy, E. M., Asmussen, M. A., & Anderson, W. W. (1995). A theoretical assessment of recombinational speciation. *Heredity*, *74*, 502–509. https://doi.org/10.1038/hdy.1995.71

Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., Piñeiro, R., Riina, R., Rosselló, J. A., Soltis, P. S., & Villa-Machío, I. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity*, *118*, 513–516. https://doi.org/10.1038/hdy.2017.7

Nolte, A. W., & Tautz, D. (2009). Understanding the onset of hybrid speciation. *Trends in Genetics*, *26*, 54–58. https://doi.org/10.1016/j.tig.2009.12.001

Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*, *34*, 401–437. https://doi.org/10.1146/annurev.genet.34.1.401

Parnell, J., & Favarger, C. (1993). *Sempervivum* L., *Jovibarba* Opiz. In T. G. Tutin, N. A. Burges, A. O. Chater, J. R. Edmondson, V. H. Heywood, D. M. Moore, D. H. Valentine, S. M. Walters, & D. A. Webb (Eds.), *Flora Europaea* (2nd ed., Vol. *1*, pp. 425–428). Cambridge University Press.

Pfennig, K. S., Kelly, A. L., & Pierce, A. A. (2016). Hybridization as a facilitator of species range expansion. *Proceedings of the Royal Society B*, *283*, 20161329. https://doi.org/10.1098/rspb.2016.1329

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, *28*, 359–389. https://doi.org/10.1146/annurev.ecolsys.28.1.359

Rieseberg, L. H., Ellstrand, N. C., & Arnold, M. (1993). What can molecular and morphological markers tell us about plant hybridization? *Critical Reviews in Plant Sciences*, *12*, 213–241. https://doi.org/10.1080/07352689309701902

Rieseberg, L. H., Raymond, O., Rosendahl, D. M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J. L., Schwarzbach, A. E., Donovan, L. A., & Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, *301*, 1211–1216. https://doi.org/10.1126/science.1086949

Sanchez, G. (2013). *DiscriMiner: Tools of the trade for discriminant analysis. R package version 0.1-29*. Retrieved from https://CRAN.R-project.org/package=DiscriMiner

Schoener, T. W. (1968). The *Anolis* lizards of Bimini: Resource partitioning in a complex fauna. *Ecology*, *49*, 704–726. https://doi.org/10.2307/1935534

Schumer, M., Cui, R., Rosenthal, G. G., & Andolfatto, P. (2015). Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genetics*, *11*(3), e1005041. https://doi.org/10.1371/journal.pgen.1005041

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution*, *68*, 1553–1560. https://doi.org/10.1111/evo.12399

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2018). What do we mean when we talk about hybrid speciation? *Heredity*, *120*, 379–382. https://doi.org/10.1038/s41437-017-0036-z

Soltis, P. S., Liu, X., Marchant, D. B., Visger, C. J., & Soltis, D. E. (2014). Polyploidy and novelty: Gottlieb's legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130351. https://doi.org/10.1098/rstb.2013.0351

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cron, Q. C. B., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, *25*, 2427–2442. https://doi.org/10.1111/mec.13539

Sun, Y., Lu, Z., Zhu, X., & Ma, H. (2020). Genomic basis of homoploid hybrid speciation within chestnut trees. *Nature Communications*, *11*, 3375. https://doi.org/10.1038/s41467-020-17111-w

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*, 453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, *28*, 696–704. https://doi.org/10.1016/j.tree.2013.09.001

Valtueña, F. J., Preston, C. D., & Kadereit, J. W. (2012). Phylogeography of a Tertiary relict plant, *Meconopsis cambrica* (Papaveraceae), implies the existence of northern refugia for a temperate herb. *Molecular Ecology*, *21*, 1423–1437. https://doi.org/10.1111/j.1365-294X.2012.05473.x

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer. https://doi.org/10.1007/978-0-387-21706-2

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, *62*, 2868–2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography*, *33*, 607–611. https://doi.org/10.1111/j.1600-0587.2009.06142.x

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR analyzing German business cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data analysis and decision support* (pp. 335–343). Springer-Verlag. https://doi.org/10.1007/3-540-28397-8_36

Whitney, K. D., Broman, K. W., Kane, N. C., Hovick, S. M., Randell, R. A., & Rieseberg, L. H. (2015). Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Molecular Ecology*, *24*, 2194–2211. https://doi.org/10.1111/mec.13044

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences*, *106*, 13875–13879. https://doi.org/10.1073/pnas.0811575106

Yakimowski, S. B., & Rieseberg, L. H. (2014). The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany*, *101*, 1247–1258. https://doi.org/10.3732/ajb.1400201

Zetzsche, H. (2004). *Die Phylogeographie des Artkomplexes Pulsatilla alpina*. (Ranunculaceae). Dissertation, Martin-Luther-Universität Halle-Wittenberg. https://doi.org/10.25673/3638

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Fabritzek AG, Griebeler EM, Kadereit JW. Hybridization, ecogeographical displacement and the emergence of new lineages – A genotyping-by-sequencing and ecological niche and species distribution modelling study of *Sempervivum tectorum* L. (Houseleek). *J Evol Biol*. 2021;34:830–844. https://doi.org/10.1111/jeb.13784