Identity and relationships of Sempervivum tectorum (Crassulaceae) in the Rhine Gorge area

Authors: Fabritzek, Armin G., and Kadereit, Joachim W.

Source: Willdenowia, 48(3) : 405-414

Published By: Botanic Garden and Botanical Museum Berlin (BGBM)

URL: https://doi.org/10.3372/wi.48.48310

Willdenowia
Annals of the Botanic Garden and Botanical Museum Berlin

ARMIN G. FABRITZEK[1] & JOACHIM W. KADEREIT[1*]

Identity and relationships of *Sempervivum tectorum* (*Crassulaceae*) in the Rhine Gorge area

Version of record first published online on 16 November 2018 ahead of inclusion in December 2018 issue.

Abstract: *Sempervivum tectorum* (*Crassulaceae*), an orophyte widespread in the European high mountains, also grows in rocky habitats of the Rhine Gorge area (Upper Middle Rhine, Mosel and Ahr river valleys). On the background of its long history of cultivation, it is unclear whether *S. tectorum* is native or naturalized in the Rhine Gorge area. Using 52 accessions of *S. tectorum* from across its geographical range (except SE Europe) as well as 15 samples of *S. calcareum* and *S. marmoreum* in our final sample, we conducted a genotyping-by-sequencing analysis. The genetic data were used for the identification of genetic groups and for the reconstruction of phylogenetic relationships. We found that the material from the Rhine Gorge area forms a homogeneous genetic group that is clearly distinct from populations from the Massif Central, Alps and Pyrenees. The Rhine Gorge material appears to be most closely related to material from the Massif Central. We hypothesize that the disjunct distribution of the Rhine Gorge/Massif Central clade is best interpreted as the remnant of a formerly wider distribution area obtained in a Quaternary glacial. The possibility of recognizing Rhine Gorge material taxonomically is discussed and rejected for morphological and nomenclatural reasons.

Key words: Alps, *Crassulaceae*, genotyping-by-sequencing, Massif Central, Pyrenees, Rhine Gorge area, Quaternary migration, *Sempervivum*

Article history: Received 21 March 2018; peer-review completed 25 July 2018; received in revised form 26 September 2018; accepted for publication 1 October 2018.

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

Introduction

The various rocky habitats of the Rhine Gorge area, here understood to comprise the Upper Middle Rhine, Mosel and Ahr river valleys, harbour a number of species that outside this area are distributed disjunctly mainly in the Alps and other European high-mountain ranges. These include, e.g. *Amelanchier ovalis* Med., *Biscutella laevigata* L., *Cotoneaster integerrimus* Med., *Rumex scutatus* L. and *Sempervivum tectorum* L. Of these, *S. tectorum* (*Crassulaceae*) has a long history of cultivation going back at least to the Middle Ages (Lippert 1995) when Charlemagne, in the early 9th century *Capitulare de villis vel curtis imperii*, decreed: *"Et ille hortulanus habeat super domum suam Iovis barbam*" (And the gardener shall have growing on his house Jove's beard). Following Lippert (1995), this decree refers to *S. tectorum* and most likely is based on the widespread belief that the species protects against lightning and fire. On this background, the native status of *S. tectorum* in the Rhine Gorge area has been discussed (Hayek 1922) or doubted (Jäger 2011). In the area, *S. tectorum* is rare, and of all known populations only 12 grow at sites considered natural by us (see below).

In the context of a phylogenetic study of *Sempervivum* L. and *Jovibarba* Opiz (Klein & Kadereit 2015), an analysis of the geographical distribution of interspecific hybrids and their parental species (Klein & Kadereit 2016), and ongoing studies on the evolution of *S. tecto-*

1 Institut für Organismische und Molekulare Evolutionsbiologie, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany; *e-mail: kadereit@uni-mainz.de (author for correspondence); armin.fabritzek@gmail.com

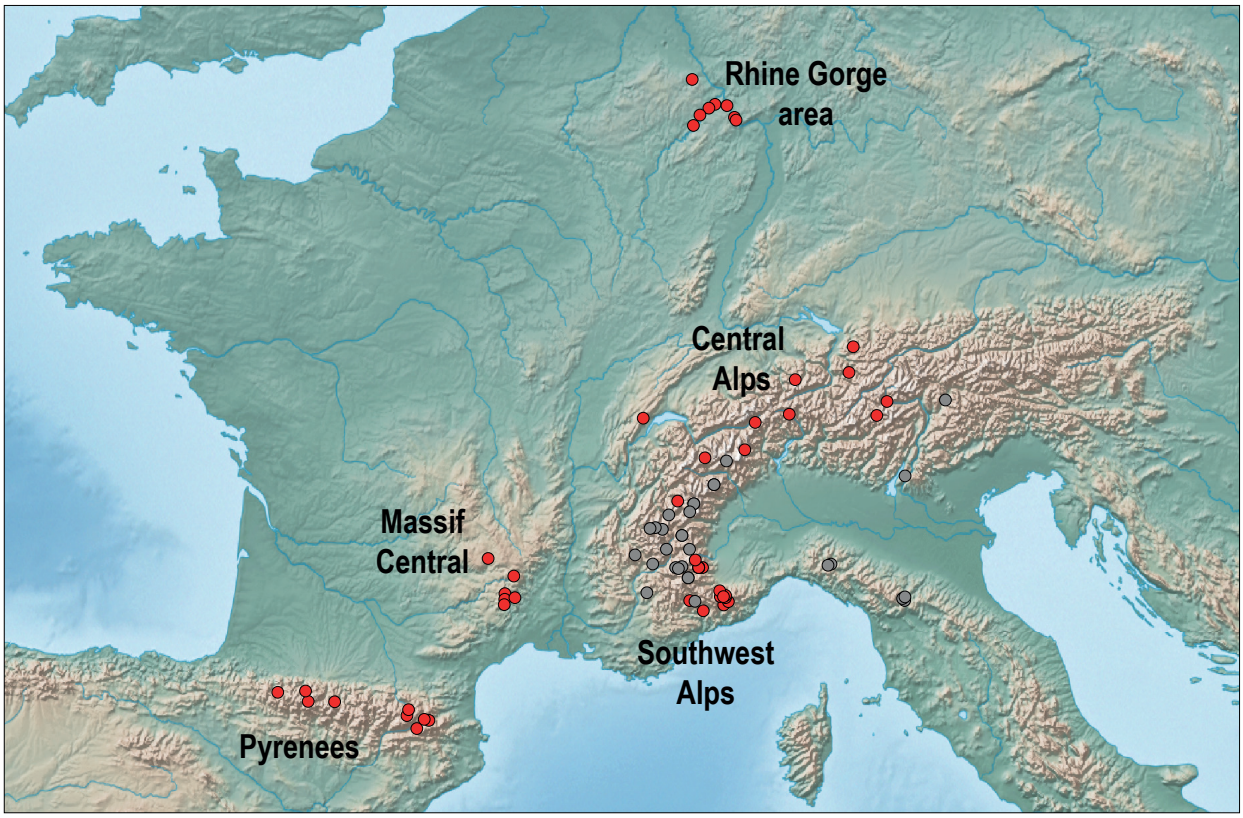

Fig. 1. Sampling sites for *Sempervivum tectorum* included in this study. Grey dots indicate samples that were excluded from the final analyses because of admixture (for explanation see text).

rum, we obtained a broad sample of the species from most parts of its range except SE Europe, which will be used here to clarify the status of the species as either native or naturalized in the Rhine Gorge area and to further clarify the phylogenetic relationships of the Rhine Gorge area populations.

Material and methods

Plant material

For preliminary analyses, a total of 99 samples were used. These were 83 samples of *Sempervivum tectorum* covering its entire distribution range except for the Balkans, one sample each of *S. minutum* (Kunze ex Willk.) Nyman ex Pau and *S. marmoreum* Griseb. as two of six species most closely related to *S. tectorum* (for details see Klein & Kadereit 2015) and 14 samples of *S. calcareum* Jord. as outgroup. In contrast to the findings by Klein & Kadereit (2015), resulting from a misidentification, *S. calcareum* has since been identified as a distant relative of *S. tectorum* (Fabritzek & Kadereit unpubl. results). To avoid sampling at sites strongly influenced by man, especially in the Rhine Gorge area, we assessed all 20 known sites. Of these, eight were either clearly man-made structures such as stone walls or monuments harbouring very small populations of fewer than five (presumably clonal) individuals or sites where other and non-native species of *Sempervivum* were found. All populations considered to grow at natural sites were found on steep rocky slopes along the river valleys across an area of around 80 km linear distance from the northernmost population in the Ahr valley to the southernmost population in the Rhine valley. Estimated population sizes ranged between 15–1500 individuals. Of the 12 populations considered to grow at natural sites, eight were sampled for this study. In order not to sample individuals of clonal origin, minimal distance between sampling sites was 1.5 km in the Rhine Gorge area. Minimal distance between sampling sites was c. 7 km in the Pyrenees, c. 8 km in the Massif Central, c. 23 km in the Central Alps and c. 1.4 km in the Southwest Alps (except for one site where five plants from different patches of leaf rosettes were sampled within a radius of c. 100 m).

All sampling sites for *Sempervivum tectorum* are shown in Fig. 1. The preliminary analyses were used to identify samples that showed a significant amount of admixture (see below) or for which too little sequence information was available. After exclusion of these samples, the final data set consisted of 67 samples comprising 52 samples of *S. tectorum*, one sample of *S. marmoreum* and 14 samples of *S. calcareum* as outgroup. Of the total sample, 95 samples were collected in the field in 2015 for this study. One garden specimen of *S. minutum* and frozen DNA of three samples from a previous study (Klein

& Kadereit 2015) were also used. Origin and identity of all samples are listed in Supplementary Table 1, which also contains information about voucher material deposited in MJG (herbarium code according to Thiers 2018+).

DNA extraction

For total genomic DNA extraction fresh leaves were chopped and immediately dried on silica gel for two days. Longer drying and storage of dried material resulted in decreasing DNA quality and content. DNA was isolated either with the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol with some modifications as described below, or a modified CTAB protocol. For the DNeasy Plant Mini Kit, 20 mg of dried leaves were ground for 90 seconds using a Retsch MM 301 mill (Retsch, Haan, Germany), immediately incubated in 600 µL lysis buffer for 50 min and inverted every 10 minutes. For the last 10 minutes, 4 µL RNase were added. P3-Buffer volume was adjusted to 195 µL. DNA was eluted once using 40 µL elution buffer. For DNA extraction using CTAB, the protocol of Schulte & al. (2010) was used, based on the original CTAB protocol by Doyle & Doyle (1987) with modifications described by Weising & al. (2005) including an additional purification step to precipitate remaining polysaccharides as described in Michaels & al. (1994). When the DNA concentration was too low, water content was reduced by evaporating excessive water at 40 °C.

Library preparation for genotyping-by-sequencing

Samples were genotyped using the genotyping-by-sequencing protocol described by Elshire & al. (2011) with some modifications. Briefly, DNA was digested with ApeK1 (New England Biolabs, Ipswich, MA, U.S.A.) and adapters including a custom barcode for each sample were ligated to the fragments. After ligation, up to 25 samples were pooled and fragments were amplified using 18 cycles of polymerase chain reaction (PCR) in five separate PCR reactions before pooling the PCR products of one library for clean-up. The 25 barcode adapters used were a subsample of the barcode adapters from Elshire & al. (2011). The common adapters with the following nucleotide sequence were designed to fit the Illumina TrueSeq® sequencing primers: 5′- CWG AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA -3′ and 5′- TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT -3′. Both adapters were prepared following Elshire & al. (2011) and stored at -21 °C. Digestion of sample DNA was carried out in a total volume of 20 µL containing ~50 ng (library F18) or ~100 ng DNA (libraries L01–L05), 2 µL NEBuffer 2.1, water and 2.5 (F18) or 3.6 Units (L01–L05) ApeK1 for 2 hours at 37 °C (F18) or 75 °C (L01–L05). Ligation was performed in 50 µL reactions containing the whole restriction reaction, 0.96 ng adapter per 10 ng restricted DNA, 1× T4 DNA Ligase Buffer and 400 Units T4 Ligase (New England Biolabs, Ipswich, MA, U.S.A.). Because of different amounts of starting DNA, different adapter concentrations were used for libraries F18 and L01–L05: working solution with 0.6 ng/µL for F18 and the concentrated stock with 3 ng/µL for L01–L05. Ligation was carried out at 22 °C for 60 min (F18) or 70 min (L01–L05) and denaturation of the enzyme at 65 °C for 30 min. The ligation products were cleaned with the Macherey-Nagel PCR purification kit (Macherey-Nagel, Düren, Germany). Up to 25 ligation products of one library were combined in one column, bound with 2:1 NTI:H_2O, washed three times with NT3 and eluted in 2× 25 mL elution buffer. Selective PCRs were carried out in 50 µL (F18) or 25 µL (L01–L05) reactions with 10 µL (F18) or 5 µL (L01–L05) NEB Taq 5× Master Mix (New England Biolabs, Ipswich, MA, U.S.A.), 2.5 µL of 10 µM concentration of each primer and 3 µL (F18) or 1 µL (L01–L05) (~44 ng DNA) of the combined and cleaned ligation products of up to 25 samples. Selective PCR was performed with the following program: 5 min at 72 °C, 30 s at 98 °C, followed by 35 cycles of 98 °C for 10 s, 65 °C for 30 s and 72 °C for 90 s followed by a final elongation step at 72 °C for 5 min. The first PCR primer was the TruSeq PE PCR Primer 1.0 (Illumina, Inc., San Diego, CA, U.S.A.) which is complementary to the 3′ strand of the barcode adapters. The second primer used in this study was one of five NEBNext Index Primer for Illumina (TS01, TS04, TS05, TS06 and TS07; New England, Biolabs, Ipswich, MA, U.S.A.) that bind to the 3′ strand of the common adapter. Libraries were cleaned on one column each with the Macherey-Nagel PCR purification kit (Macherey-Nagel, Düren, Germany) using 1:1 NTI:H_2O for binding, NT3 for three washing steps and eluted in 2× 15 mL elution buffer. The libraries were quantified using a Qubit dsDNA HS Assay Kit (Life Technologies, MA, U.S.A.), QIAxcel (QIAGEN, Hilden, Germany) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, U.S.A.). Libraries L01–L05 were standardized for DNA concentration and pooled before sequencing.

Library F18 included 25 samples and was paired-end sequenced (100nt) on a half lane on an Illumina HiSeq 2500 system (Illumina, Inc., San Diego, CA, U.S.A.) at the Institut für Organismische und Molekulare Evolutionsbiologie, Johannes Gutenberg-Universität Mainz (7 of 25 samples had too few raw reads to be analysed). The five libraries L01–L05 were paired-end sequenced (150nt) on an Illumina NextSeq 500 (StarSEQ, Mainz, Germany) on two separate half lanes as technical replicates. The five libraries L01–L05 included altogether 90 samples (L01–L05: 16, 19, 20, 20, 15, respectively). Ten individuals were technical replicates of samples from F18 that had the lowest sequencing depth in F18 (Sample 1508, 1515, 1516, 1517, 1518, 1526, 1529, 1536, 1549, 1598). Raw data of samples 1517, 1526, 1598, 1515 and 1536 were merged from F18 and the later libraries. Of all other technical replicates only raw data from one sequencing run were used. Trueseq primers used were TS01 (L01), TS4 (F18 and L02), TS05 (L03), TS06 (L04) and TS07 (L05).

ipyrad – genotyping-by-sequencing pipeline

Raw reads were processed with ipyrad (Eaton & Overcast in prep.) versions 0.5.15 (demultiplexing) and 0.7.15 (rest of the pipeline) using the *de novo* method for paired-end GBS sequencing reads. First, all six libraries were demultiplexed. No mismatch within the barcode adapters was allowed. Demultiplexed raw reads of technical replicates were manually merged. Restriction overhang was CWGC. Maximum amount of low-quality base calls (Phred Q score < 20) in a read was set to 5. Phred Q score offset was set to 33. Reads shorter than 35 base pairs were discarded after trimming of the common Illumina adapter, the reverse complement of the second restriction site and the barcode. Minimum depth for statistical base calling and majority rule base calling was set to 6. The maximum number of uncalled bases and heterozygous sites in consensus sequences in forward and reverse reads were set to 5 and 8, respectively. For maximum number of heterozygous sites per locus we used the standard setting of 50 %. Several authors (Fischer & al. 2015; Derkarabetian & al. 2016; Johannesen & al. 2017) used and tested different levels (5–55 %) of maximum number of heterozygous sites and reported a strong effect on number of loci recovered but a small effect on derived tree topologies or on structure analyses. The maximum number of single nucleotide polymorphisms (SNPs) in forward and reverse reads per locus was set to 20 each. Final loci had to have a maximum of two alleles in the final output files. One base at the 5′ and two bases at the 3′ ends were deleted from the finally aligned loci in order to reduce error during alignment. In order to obtain a moderate amount of missing data with a reasonable amount of shared loci between all samples, the minimum number of samples containing information for a specific locus (minCov) was set to 60 % (excluding the outgroup). Samples with a sequencing depth lower than mean minus 2 standard deviations were excluded (a total of two samples, Supplementary Table 1) because large amounts of missing data can have strong effects in different analyses. To assess suitable settings for the clustering threshold, we generated data sets with three different clustering thresholds, i.e. 0.86, 0.9 and 0.94. The total number of recovered loci at the end of the ipyrad pipeline increased with higher clustering thresholds (5247/6063/7644). The overall topologies of maximum likelihood trees based on the different data sets were almost identical, but the highly supported sister relationship of *S. marmoreum* to all samples of *Sempervivum tectorum* (incl. *S. minutum*) was only resolved with clustering threshold 0.9. As preliminary structure plots or Splits Trees 4.14.2 (Huson & Bryant 2006) resulted in almost identical results, all subsequent data sets were assembled using the clustering threshold of 0.9.

Data analysis

For preliminary analyses Maximum Likelihood (ML) and Singular Value Decomposition Scores for Species Quartets (SVDQ), trees were calculated using 97 samples. In order to identify admixture between genetic lineages identified by ML and SVDQ, a Bayesian structure analysis was conducted after exclusion of *Sempervivum marmoreum*, *S. minutum* and *S. calcareum*. Using the results of this analysis for the most probable smallest K estimated by delta K (K = 5), samples that were assigned to two different clusters by at least 30 % or had assignment probabilities to a cluster lower than 85 % were excluded. As a result, 30 samples that showed a significant amount of admixture were excluded: three samples from the more eastern sampling location in the Apennines that were sister to *S. tectorum* and *S. minutum*; one sample from *S. minutum* that was sister to the *S. tectorum* samples from the Pyrenees; two samples from the western sampling location in the Apennines; and 24 samples from the Alps. This resulted in a data set consisting of 67 (out- and sister-group included) and 52 (out- and sister-group excluded) samples. The data set including the out- and sister-group was used to calculate trees using ML and SVDQ. The dataset without the out- and sister-group was used to conduct a second Bayesian Structure analysis and a Principal Component Analysis. The initial data set was constructed using ipyrad, and data pruning for obtaining the final data set was conducted manually.

Tree inference

The Maximum Likelihood analyses were performed using RAxML-HPC2 (ver. 8.2.10) on XSEDE (Stamatakis 2014) in the CIPRES Science Gateway (Miller & al. 2010). The GTR-GAMMA model was used as nucleotide substitution model and the number of necessary bootstrap replicates was calculated automatically using the autoMRE Majority Rule Criterion Selection. To obtain a tree based on a coalescent method, SVDquartets v.1.0 (Chifman & Kubatko 2014) implemented in PAUP*v.4.0a157 and v.4.0a159 (Swofford 2002) was used. SVDquartets uses a multispecies coalescent method that combines quartets of samples into a single combined tree including all samples. All possible quartets were sampled and at least 500 bootstrap replicates were carried out. For the tree search the QFM quartets assembly algorithm was used with no parameter for local searches. Samples of *Sempervivum calcareum* were used as outgroup. Environment for Tree Exploration (ETE3; Huerta-Cepas & al. 2016) was used for tree visualization. The tree was rooted with *S. calcareum* accessions and nodes with bootstrap support values < 80 % were collapsed.

Structure

We used Structure v.2.3.4 (Pritchard & al. 2000) for the analysis of the data sets in order to infer genetic groups and possible admixture. We ran 500000 generations with an initial burn-in of 100000 generations for values of K rang-

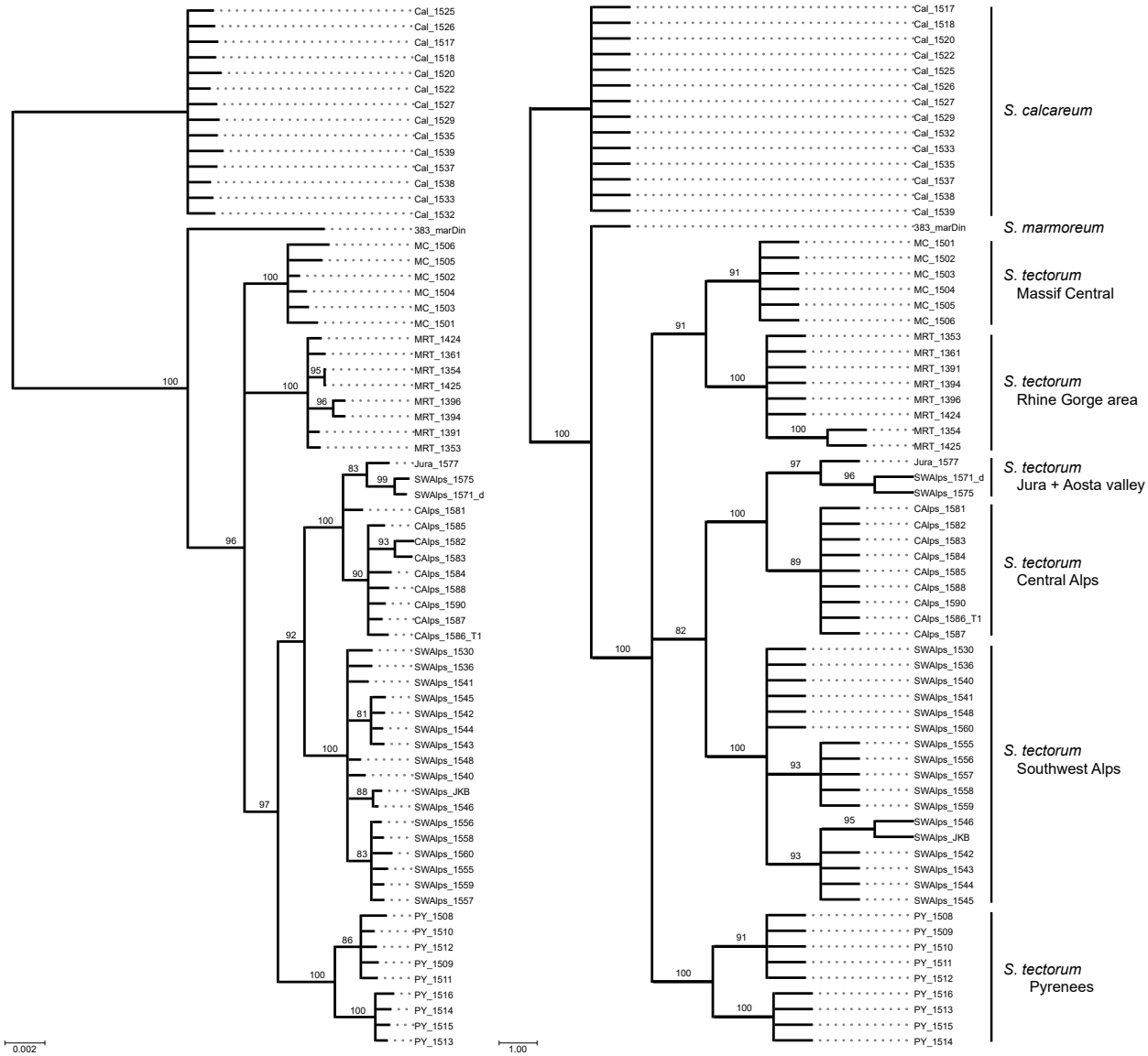

Fig. 2. Best-scoring Maximum Likelihood tree calculated with RAxML (left) and majority rule consensus tree calculated with SVDquartets (right). Branches with bootstrap support < 80 % were collapsed.

ing from 1 to 10 with 10 iterations each, with one sample retained every 100 iterations. Convergence of alpha was checked manually. The CLUMPAK web platform was used to determine the optimal number of K using either delta K following Evanno & al. (2005) or the probability of K following Falush & al. (2003), and to combine 10 iterations for the optimal output (Kopelman & al. 2015). CLUMPP was run with 2000 repeats and the greedy algorithm and random input order. Graphical output of the CLUMPP results was obtained using Distruct (Rosenberg 2004). Structure was run on the MOGON cluster at Johannes Gutenberg-Universität Mainz for all data sets.

Principal Component Analysis

For the Principal Component Analysis (PCA) we used the R packages adegenet version 2.1.0 (Jombart, 2008) and ade4 version 1.7.10–1.7.13 (Dray & Dufour 2007; Bougeard & Dray 2018). To read and transform the ipyrad vcf file (dataset without excluded samples) into a genlight object we used the vcfR package version 1.6.0 (Knaus & Grünwald 2017). Ploidy level was set to 2 (using ploidy level of 4 led to identical results; data not shown). To plot the data ggplot2 version 2.2.1 (Wickham 2016) was used.

Results

A total of 119140324 raw reads were retrieved from the Illumina sequencing platform. Ipyrad could build a total of 208149 loci with a mean depth of coverage of 5.36 and a standard deviation of 65.86. A total of 166948 loci were excluded from the final data set because of low coverage.

After filtering for minimum coverage of > 5 of raw reads within loci, and for a minimum of 60 % of samples containing a locus (excluding the outgroup *Sempervivum calcareum*), a total of 6063 loci were recovered. The mean number of loci in the final data set shared among all samples was 4239 with a standard deviation of 1102. These loci contained a total of 28387 single nucleotide polymorphism (SNP). For phylogenetic tree construction, the final data set consisted of concatenated loci that matched the filtering criteria described above. The total length of the concatenated alignment of the data set was 474479 base pairs with 6386 parsimony informative characters. The structure files contained 5763 informative loci with one random SNP sampled per locus. Demultiplexed raw reads can be downloaded in NCBI sequence read archive accession SRP135257.

Tree inference

For the description of tree topologies, support values were defined as: > 95 highly supported, 95–90 well-supported, 90–80 supported. In the Maximum Likelihood (ML) and Singular Value Decomposition Scores for Species Quartets (SVDQ) trees (Fig. 2), all samples of the outgroup *Sempervivum calcareum* (Cal) formed a highly supported group, *S. marmoreum* was identified as sister to all samples of *S. tectorum*, and geographical groups within *S. tectorum* were well resolved and highly supported. In the ML tree, the samples from the Massif Central (MC) and Rhine Gorge area (MRT) form a trichotomy with a highly supported clade in which the samples from the Pyrenees (PY) are highly supported sister to a well-supported clade of the C Alps and SW Alps. In the SVDQ tree, *S. tectorum* consists of a trichotomy of (1) MC and MRT as well-supported sister to each other, (2) SW Alps and C Alps as supported sister to each other, and (3) Pyrenees. In both analyses the C Alps clade contains two samples from south of the Aosta valley as sister to the one sample from the Jura Mts included in the analysis.

Except for the Alps, the same geographical groups were obtained in the ML and SVDQ analyses of the dataset before exclusion of samples. In the SVDQ analysis, samples from the SW Alps and C Alps formed one supported clade (BS 87) without resolution of the two geographical groups. In the ML analysis, samples from the Alps are part of a large polytomy containing the other geographical groups as one supported clade.

Structure

The structure analysis of our final data set for K = 2–7 is shown in Fig. 3. The optimal number of clusters is K = 4 (identified by the probability of K; Falush & al. 2003) or K = 7 (identified by delta K; Evanno & al. 2005). MC always shows signs of admixture. Either MC is admixed between MRT and the Alps and Pyrenees (K = 2) or between MRT and the Pyrenees with a small amount of admixture from the Alps (K = 3). At K = 4, when a subdivision of the Alps samples can be seen for the first time, MC shows admixture with all other groups with clear predominance of MRT and PY. At K = 5, MC forms its own cluster, which is admixed with mainly MRT but also SW Alps in some samples. At K = 6 and K = 7, the proportion of the MC cluster increases and the proportion of the MRT cluster decreases. In the structure analysis before exclusion of samples, material from the MRT did not show any admixture at the most likely K values (data not shown).

Principal Component Analysis (PCA)

The first two principal components account for 22 % of variation, and the samples fall into five groups as indicated by the 95 % confidence level eclipse drawn around samples (Fig. 4). These five groups correspond to the five regions of the study area, i.e. Rhine Gorge area, Massif Central, Pyrenees, Central Alps and Southwest Alps. The PCA plot clearly illustrates that genetic variation of samples within regions is similar across regions. Distances between genotypes within regions most likely imply that no clonal samples were included in the analysis.

Discussion

Identity and relationships of Rhine Gorge *Sempervivum tectorum*

As evident from our genetic clustering and PCA analyses (Fig. 3, 4), *Sempervivum tectorum* from the Rhine Gorge area forms a very homogeneous genetic group clearly distinct from populations from the Massif Central, Alps and Pyrenees. In our phylogenetic analyses (Fig. 2), the Rhine Gorge material forms a highly supported clade. These results in our opinion clearly show that *S. tectorum* from the Rhine Gorge area is a genetically unique and clearly distinct lineage and make it very likely that *S. tectorum* is native to the Rhine Gorge area. While our data allow us to exclude the possibility that the species here is naturalized material of Alpine or Pyrenean origin, it is conceivable that it is naturalized material originating from the Massif Central, where the populations appear to be most closely related to those in the Rhine Gorge area. However, this seems unlikely to us mainly because of the genetic differences between material from these two areas. As all individuals from the Massif Central analysed showed admixture at all values of K (Fig. 3), this admixture should be recognizable in Rhine Gorge material should this have originated from the Massif Central. Although the absence of admixture in Rhine Gorge material may be the result of a genetic bottleneck after migration or synanthropic introduction from the Massif Central, the genetic variation as seen in Rhine Gorge material in our PCA analysis (Fig. 4), which is as high as in other areas, provides no evidence for such a bottleneck. Those

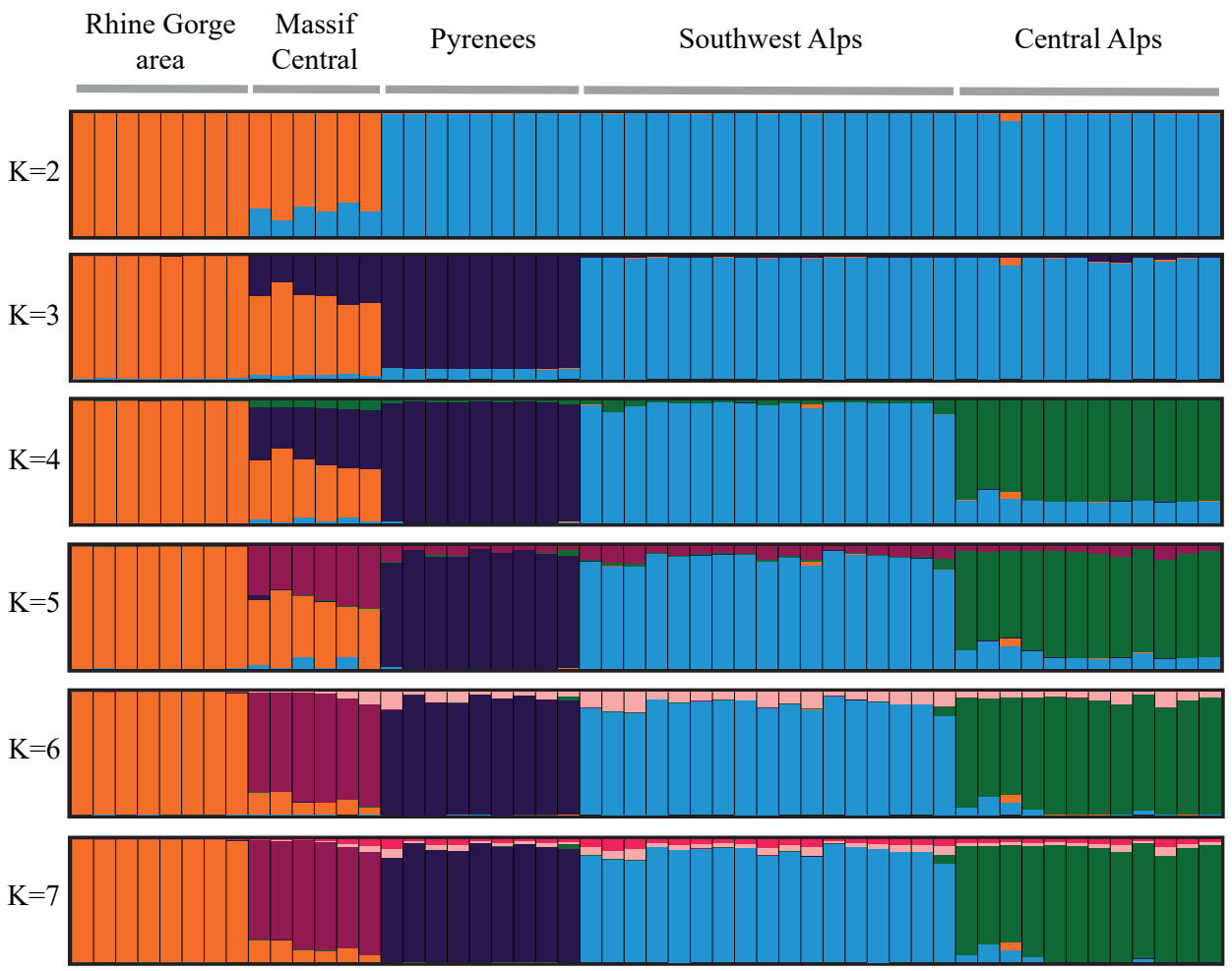

Fig. 3. Structure analysis of final dataset for K = 2 – 7. The optimal number of clusters is K = 4 (identified by the probability of K; Falush & al. 2003) or K = 7 (identified by delta K; Evanno & al. 2005).

individuals from the Rhine Gorge area included in our analyses were collected only from sites that we considered natural, and plants from clearly anthropogenic sites were not sampled. In consequence, we cannot exclude the possibility that the Rhine Gorge area harbours both native and naturalized material of the species.

As evident from our phylogenetic analyses, Rhine Gorge *Sempervivum tectorum* is most closely related to material from the Massif Central. While such relationship was resolved using SVDQ (Fig. 2), and in our opinion is supported by genetic similarity (Fig. 3, 4), it is not contradicted by the results of our ML analysis where material from the Rhine Gorge area, the Massif Central and the Alps plus Pyrenees forms three clades of a trichotomy (Fig. 2). Quite remarkably, a possible relationship of the Rhine Gorge populations to populations from C and E France had already been suspected by Hayek (1922).

Southwestern European relationships of Central European plant and animal populations had already been recognized in early reviews of phylogeographic studies of European plants and animals (Taberlet & al. 1998; Hewitt 1999), and had been interpreted as reflecting postglacial immigration via, e.g., the Burgundy Gate or the Moselle Valley (e.g. Linn & Griebeler 2015). Considering that *Sempervivum tectorum* essentially is a mountain plant which most likely could not migrate between mountain ranges in postglacial times, the geographical distribution of the Rhine Gorge plus Massif Central clade of our SVDQ phylogeny (Fig. 2) is unlikely to be the result of postglacial migration. Instead, this disjunct distribution much more likely originated in a glacial period in which the species likely occupied a wide geographical range outside the European high mountains. The extant occurrences of the species in the Massif Central and Rhine Gorge area would then be the remnants of a formerly continuous and larger distribution area. Based on the non-sympatric distribution of hybrid and parent individuals in *Sempervivum*, long-distance migration of the genus in glacial times has indeed been postulated by Klein & Kadereit (2016). The existence and persistence through at least the last glacial of such "northern" lineages including populations from the Massif Central (but also other populations from elsewhere in Europe) has been postulated before for other mountain species such as *Pulsatilla alpina* (L.) Delarbre (Zetzsche 2004), *Meum athamanticum* Jacq. (Huck & al. 2009) and *Papaver cambricum* L.

(= *Meconopsis cambrica* (L.) Vig.; Valtueña & al. 2012), and also for animal species (see Schmitt & Varga 2012 for review).

Taxonomic considerations

Considering the clear geographic structure of our data, with distinct and non-overlapping lineages in the Pyrenees, the Massif Central, the Rhine Gorge area, the Central Alps, the Southwest Alps and the Apennine (results not shown for the last), it would be tempting to recognize these lineages at subspecific rank. Material from these areas has been named at various ranks before (see e.g. Hayek 1922; Huber 1961; Lippert 1995; Conti & al. 2005; Tison & de Foucault 2014). As regards the Rhine Gorge material, Hayek (1922) interpreted these populations of *Sempervivum tectorum* as native and described them as var. *rhenanum* Hegi & Schmid ex Hayek, a name later elevated to species rank by Lawalreé (1956). Such action, however, would be premature for three reasons. First, our sample is not complete and particularly material from the S Apennines and the Balkans was not included. Second, the variability of the species across its broad distribution range (Hayek 1922; Huber 1961; Lippert 1995) as well as its high plasticity, particularly in rosette and leaf size and coloration, do not allow us to name morphological characters by which the different lineages, including the Rhine Gorge material, can be distinguished. Therefore, a revision based on broad material from across the entire range is necessary, and such a revision would ideally also require the examination of material in the field because observation of characters is often very difficult in herbarium specimens. Third, the identity of the lectotype of *S. tectorum* is highly problematic. Following Gallo & Jarvis (2016), the lectotype chosen by Parnell (1993) best matches hybrid material between *S. arachnoideum* L. and *S. tectorum*. In consequence, Gallo & Jarvis (2016) proposed to conserve the name *S. tectorum* with a conserved type, and this proposal has since been recommended (W. Applequist, pers. comm.). The type proposed by Gallo & Jarvis (2016) is a specimen collected in Sweden, where the species is not native but naturalized. As the geographical origin of this naturalized material is unknown, it remains unclear into which of the geographical lineages of *S. tectorum* the type specimen falls, which prevents stable naming of all other lineages. Clarity about the geographical origin of Swedish *S. tectorum* probably could be achieved by genotyping the Linnaean type material, as has been done in other instances (e.g. Andreasen & al. 2014).

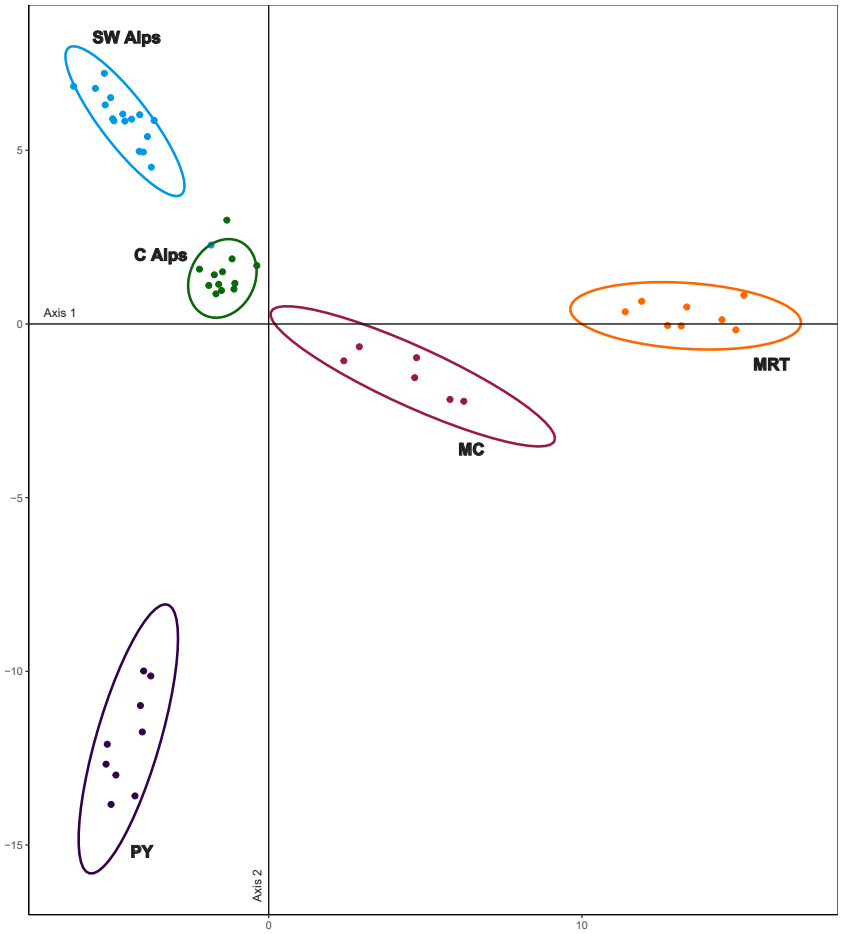

Fig. 4. Principal Component Analysis (PCA) of genotyping-by-sequencing genotypes. The first two axes explain 11.6 % and 10.4 % of total variance, respectively. MRT: Rhine Gorge area, MC: Massif Central, C Alps: Central Alps, SW Alps: Southwest Alps, PY: Pyrenees.

Acknowledgements

We would like to acknowledge the Deutsche Forschungsgemeinschaft (DFG) for financial support (KA 635/17-2). 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. Markus Dillenberger, Wolfgang Ehmke, Michael Gerschwitz-Eidt, Katrin Jahn, Martha Kandziora, Johannes Klein,

Maximilian Lauterbach, Michael Pirie, Natalie Schmalz and Christian Uhink provided various support, Nicholas Turland and Wendy Applequist are acknowledged for nomenclatural advice, and thanks are due to Katja Reichel (Freie Universität Berlin) and two anonymous reviewers for helpful comments on the manuscript.

References

Andreasen K., Manktelow M., Sehic J. & Garkava-Gustavsson L. 2014: Genetic identity of putative Linnaean plants: successful DNA amplification of Linnaeus's crab apple *Malus baccata*. – Taxon **63:** 408–416.

Bougeard S. & Dray S. 2018: Supervised multiblock analysis in R with the ade4 package. – J. Stat. Softw. **86:** 1–17.

Chifman J. & Kubatko L. 2014: Quartet inference from SNP data under the coalescent model. – Bioinformatics **30:** 3317–3324.

Conti F., Abbate G., Alessandrini A. & Blasi C. (ed.) 2005: An annotated checklist of the Italian vascular flora. – Roma: Palombi e Partner S.r.l.

Derkarabetian S., Burns M., Starrett J. & Hedin M. 2016: Population genomic evidence for multiple Pliocene refugia in a montane-restricted harvestman (Arachnida, Opiliones, *Sclerobunus robustus*) from the southwestern United States. – Molec. Ecol. **25:** 4611–4631.

Doyle J. J. & Doyle J. L. 1987: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. – Phytochem. Bull. **19:** 11–15.

Dray S. & Dufour A. B. 2007: The ade4 package: implementing the duality diagram for ecologists. – J. Stat. Softw. **22:** 1–20.

Eaton D. A. R. & Overcast I. [in prep]: ipyrad: interactive assembly and analysis of RAD-seq data sets. – Available at https://github.com/dereneaton/ipyrad

Elshire R. J., Glaubitz J. C., Sun Q., Poland J., Kawamoto K., Buckler E. S. & Mitchell S. E. 2011: A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. – PLoS ONE **6:** e19379.

Evanno G., Regnaut S. & Goudet J. 2005: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. – Molec. Ecol. **14:** 2611–2620.

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.

Fischer G., Azorsa F., Hita Garcia F., Mikheyev A. S. & Economo E. P. 2015: Two new phragmotic ant species from Africa: morphology and next-generation sequencing solve a caste association problem in the genus *Carebara* Westwood. – ZooKeys **2015:** 77–105.

Gallo L. & Jarvis C. E. 2016: (2486) Proposal to conserve the name *Sempervivum tectorum* (*Crassulaceae*) with a conserved type. – Taxon **65:** 1431–1432.

Hayek A. von 1922: *Sempervivum* L. – Pp. 543–562 in: Hegi G., Illustrierte Flora von Mittel-Europa. Bd. IV/2. – München: J. F. Lehmann.

Hewitt G. M. 1999: Post-glacial re-colonization of European biota. – Biol. J. Linn. Soc. **68:** 87–112.

Huber H. 1961: *Crassulaceae*. – Pp. 62–125 in: Huber H. (ed.), Gustav Hegi – Illustrierte Flora von Mitteleuropa. Ed. 2, Bd. IV/2A. – Berlin & Hamburg: Paul Parey.

Huck S., Büdel B., Haase P., Kadereit J. W. & Printzen C. 2009: Range wide phylogeography of the European temperate-montane *Meum athamanticum*: evidence for periglacial persistence and multiple northern refugia. – J. Biogeogr. **36:** 1588–1599.

Huerta-Cepas J., Serra F. & Bork P. 2016: ETE 3: reconstruction, analysis, and visualization of phylogenomic data. – Molec. Biol. Evol. **33:** 1635–1638.

Huson D. H. & Bryant D. 2006: Application of phylogenetic networks in evolutionary studies. – Molec. Biol. Evol. **23:** 254–267.

Jäger E. J. 2011: Rothmaler – Exkursionsflora von Deutschland. Gefäßpflanzen: Grundband. Ed. 20. – Heidelberg: Spektrum Akademischer Verlag.

Johannesen J., Fabritzek A. G., Ebner B. & Bikar S.-E. 2017: Characterisation of microsatellite and SNP markers from Miseq and genotyping-by-sequencing data among parapatric *Urophora cardui* (*Tephritidae*) populations. – PeerJ **5:** e3582.

Klein J. T. & Kadereit J. W. 2015: Phylogeny, biogeography and evolution of edaphic association in the European oreophytes *Sempervivum* and *Jovibarba* (*Crassulaceae*). – Int. J. Pl. Sci. **176:** 44–71.

Klein J. T. & Kadereit J. W. 2016: Allopatric hybrids as evidence for past range dynamics in *Sempervivum* (*Crassulaceae*). – Alpine Bot. **126:** 119–133.

Knaus B. J. & Grünwald N. J. 2017: vcfr: a package to manipulate and visualize variant call format data in R. – Molec. Ecol. Res. **17:** 44–53.

Kopelman N. M., Mayzel J., Jakobsson M., Rosenberg N. A. & Mayrose I. 2015: Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. – Molec. Ecol. Res. **15:** 1179–1191.

Lawalrée A. 1956: Plantes Médio-Mosano-Rhénanes. – Biol. Jaarb. **23:** 228–253.

Linn C. A. & Griebeler E. M. 2015: Reconstruction of two colonization pathways of *Mantis religiosa* (Mantodea) in Germany using four mitochondrial markers. – Genetica **143:** 11–20.

Lippert W. 1995: *Crassulaceae*. – Pp. 69–129 in: Weber H. E. (ed.), Gustav Hegi – Illustrierte Flora von Mitteleuropa. Ed. 3, Bd. IV/2A. – Berlin: Blackwell.

Michaels S. D., John M. C. & Amasino R. M. 1994: Removal of polysaccharides from plant DNA by ethanol precipitation. – BioTechniques **17:** 274–276.

Miller M. A., Pfeiffer W. & Schwartz T. 2010: Creating the CIPRES Science Gateway for inference of large

phylogenetic trees. – New Orleans: 2010 Gateway Computing Environments Workshop (GCE).

Parnell J. 1993: *Sempervivum*. – P. 87 in: Jarvis C. E., Barrie F. R., Allan D. M. & Reveal J. L. (ed.), A list of Linnaean generic names and their types. – Regnum Veg. **127.**

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

Rosenberg N. A. 2004: DISTRUCT: a program for the graphical display of population structure. – Molec. Ecol. Notes **4:** 137–138.

Schmitt T. & Varga Z. 2012: Extra-Mediterranean refugia: the rule and not the exception? – Frontiers Zool. **9:** 22.

Schulte K., Silvestro D., Kiehlmann E., Vesely S., Novoa P. & Zizka G. 2010: Detection of recent hybridization between sympatric Chilean *Puya* species (*Bromeliaceae*) using AFLP markers and reconstruction of complex relationships. – Molec. Phylogen. Evol. **57:** 1105–1119.

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

Swofford D. L. 2002: PAUP*. Phylogenetic analysis using parsimony (*and other methods). Ed. 4. – Sunderland: Sinauer Associates.

Taberlet P., Fumagalli L., Wust-Saucy A.-G. & Cosson J.-F. 1998: Comparative phylogeography and postglacial colonization routes in Europe. – Molec. Ecol. **7:** 453–464.

Thiers B. 2018+ [continuously updated]: Index herbariorum. A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. – Published at http://sweetgum.nybg.org/science/ih/ [accessed 9 Nov 2018].

Tison J.-M. & de Foucault B. 2014: Flora gallica. Flore de France. – Mèze: Biotope.

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. – Molec. Ecol. **21:** 1423–1437.

Weising K., Nybom H., Pfenninger M., Wolff K. & Kahl G. 2005: DNA fingerprinting in plants: principles, methods, and applications. Ed. 2. – Boca Raton: CRC Press.

Wickham H. 2016: ggplot2: elegant graphics for data analysis. – New York: Springer.

Zetzsche H. 2004: Die Phylogeographie des Artkomplexes *Pulsatilla alpina* (*Ranunculaceae*). – Halle: Dissertation, Martin-Luther-Universität Halle-Wittenberg.

Willdenowia
Open-access online edition www.bioone.org/loi/will
Online ISSN 1868-6397 · Print ISSN 0511-9618 · Impact factor 1.500
Published by the Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin