ZENTRUM
FÜR BIODIVERSITÄT UND NACHHALTIGE LANDNUTZUNG

- CENTRE OF BIODIVERSITY AND SUSTAINABLE LAND USE -

Evolution of a hybrid zone of willows (*Salix* L.) in the Alps analysed by RAD-seq and morphometrics

Master's Thesis for achievement of the academic degree
"Master of Science" (M.Sc.) in the study program
"Biodiversity, Ecology and Evolution"
at the Georg-August University of Göttingen
prepared at the

Albrecht-von-Haller Institute for Plant Sciences

Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium)

submitted by

B.Sc. Pia Marinček

born in
Ljubljana

Göttingen, May 2022

First reviewer: Prof. Dr. Elvira Hörandl

Second reviewer: Prof. Dr. Maria Teresa Aguado Molina

Day of announcement of the Master's thesis: 29.11.2021

Date of delivery of the Master's thesis: 20.05.2022

Abstract

Natural hybridization of plants can result in many outcomes with several evolutionary consequences, such as hybrid speciation and introgression. Natural hybrid zones can arise as a result of fluctuating climate during the exchange of glacial and interglacial periods, where species retract and expand their territories, resulting in secondary contacts. On mountain ranges, such as the European Alps, hybrid zones of alpine species are frequently formed where different lineages meet. Willows are a large genus of woody plants with an immense capability of interspecific crossing. About 33 species of this genus inhabit the Alps, where intermediate phenotypes were frequently observed and proclaimed as hybrids, but rarely examined in detail. One of the reasons was a lack of informative molecular markers and suitable analytical tools to analyse this highly diverse and convergent genus. With the development of the RAD-seq technology, several questions regarding willow phylogeny and hybridization could be answered. In this study, a putative hybrid zone of two sister species, *S. foetida* and *S. waldsteiniana,* was investigated to study the genomic structure of populations within and outside their contact zone, to find evidence for hybrid speciation or introgression, and to analyse if morphological phenotypes are reflected by their genotypes. Individuals of the two species were sampled across their distribution range in the Alps and examined with the use of RAD-seq data and morphometric analyses. The results showed that a hybrid zone between the two species was established within the range of their contact zone. Patterns of genetic admixture in homoploid hybrids indicated introgression with asymmetric backcrossing to one of the parental species. Morphometric characteristics of hybrids supported the molecular data and showed intermediacy with a bias towards *S. waldsteiniana*. Only one potentially divergent hybrid lineage was observed, hinting to a possibility of localised hybrid speciation events.

Structure

1. Introduction

1.1. Hybridization

What is hybridization?

Hybridization is described as a (successful) interbreeding of individuals from two (or more) populations, which are distinguishable on the basis of at least one heritable character (Harrison, 1990, Arnold, 1997). Commonly this term is used to describe the crossing of two distinct species. Scientists have long tried to determine the role of hybridization in the evolutionary processes of living beings. In most animal groups, hybridization seemed to have been a rare natural event, often considered a reproductive mistake (Mayr, 1963). On the contrary, it was appraised as a widespread phenomenon among flowering plants (Stebbins, 1959). Several studies from the late 20th and early 21st centuries have revealed that angiosperms tend to hybridize frequently with an estimated 3 - 25% of observed hybridization events spread unevenly across 40% of families and 16% of genera (Mallet, 2005, Whitney et al., 2010). Studying hybridization is important for several different reasons: to understand the systematics of a particular taxon, to investigate mechanisms that restrict gene flow and promote speciation, and because it can act as a potential source of new variation and evolutionary creation (Arnold, 1997). Since traditional species concepts are formulated in a way that the effects of hybridization are either ignored, contradicting, or cannot be entirely included in a given concept's definition (Arnold 1997, Hörandl, 2022), studying and understanding hybridization processes under natural circumstances has extreme importance in shaping the very fundamental biological concepts and their definitions.

Outcomes of hybridization

Interbreeding of distinct populations can have various different outcomes and is hard to predict. Areas where hybridization naturally occurs are described as hybrid zones (Harrison, 1993) where different mixed ancestry situations may be observed. Production of a viable first generation of hybrids (F1) is usually the most difficult on the account of pre- and postzygotic incompatibilities of divergent taxa (Mallet, 2005). A rise in ploidy levels of hybrids – allopolyploidization – often results in chromosomal or genetic incompatibilities between hybrids and parents, but homoploid hybrids that have the same ploidy levels as their parents are less likely to follow this path (Mallet, 2007, Rieseberg, 1997, Soltis & Soltis, 2009, Soltis et al., 2014). Ecological selection might become the driver to reproductive isolation of genetically compatible parental and admixed lineages. Hybridization can boost genetic variance and, consequently, hybrids might be capable of inhabiting different ecological niches than their parental species (Mallet 2007). Homoploid offspring that are not reproductively isolated allow for backcrossing with one or both of their parents (Mallet, 2007, Yakimowski & Rieseberg, 2014). Such an outcome of hybridization produces a wide array of recombinant types, described as a hybrid swarm (Harrison, 1993). This reticulate exchange of intraspecific genetic material through hybrid offspring is described with the term introgression (Anderson, 1949) and can result in genetic, ecological, phenotypical, or other changes in hybridizing populations (Arnold, 1997, Mallet, 2007). Introgression is a well-documented process in flowering plants (Anderson, 1949, Stebbins, 1950, Stebbins, 1959, reviewed in Yakimowski & Rieseberg, 2014). High levels of introgressive hybridization in nature are often associated with disturbances that open new ecological niches (reviewed in Yakimowski & Rieseberg, 2014). Several studies on

adaptive trait introgression of plants suggested that homoploid hybrid variants could directly be involved in important ecological adaptations (reviewed in Yakimowski & Rieseberg, 2014, Kadereit, 2015).

Hybrid phenotypes

Studies of plant hybrids revealed, that F1 hybrids showed a mosaic of intermediate, as well as parental morphological traits, which could be explained by the dominant inheritance patterns of certain characters (Rieseberg & Ellstrand, 1993). Moreover, first, as well as later-generation hybrids, frequently exhibit transgressive (extreme, compared to parental) morphological characteristics (Rieseberg et al., 1999). Patterns of the dominance of morphological traits were found to be hard to predict even in F1 hybrids, exhibiting a mismatch of characters with dominance in conflicting directions (Thompson et al., 2021). Conclusively, hybrid morphology can be diverse, exceeding the simple assumption of intermediacy, and should be studied case-specific.

Hybridization in time and space

In spatio-temporal context, many hybrid zones are a result of secondary contact of diverging populations that were separated during the last glacial maximum (LGM), between 30 to 18 thousand years ago (Ivy-Ochs, 2008). In Europe during this time, the unfavourable conditions forced the species to retreat to southern refugia on the Iberian, Apennine, and Balkan peninsula, as well as the Caucasus region and near the Caspian Sea, from where they rapidly colonised northwards, once the conditions were favourable (Hewitt, 1999). Mountain regions, such as the European Alps, act as a range expansion barrier, resulting in a saturation of many hybrid zones between different populations (Hewitt, 2004). Most hybrid zones have remained relatively spatially stable since their establishment hundreds or thousands of years ago (Harrison, 1990). Recent movements and range expansions are frequently linked with environmental changes, many of which are a result of anthropogenic factors, such as habitat disturbance, assisted invasion of species, and climate change (Hewitt, 2011). Many hybrid zone studies in the Alps were conducted on single contact zone instances and only a few studies focused on metapopulational studies over larger scopes of species' contact zones. Such studies give a more conclusive picture of the extent of hybrid zones and general patterns of admixture and processes underlining them.

Willows and hybridization

Genus *Salix* L., with about 450 species distributed worldwide, is the largest genus in the willow family (Salicaceae) (Argus, 1997). Willows come in many different growth forms, ranging from large, mighty trees to tiny dwarf shrubs and inhabit a wide range of habitats, including the unfavourable mountain and arctic regions, where they are one of the dominating arborescent plant groups (Skvortsov, 1999). In Europe, *Salix* is the largest woody plant genus with about 65 described species (Rechinger, 1993), 33 of which were described in the mountain systems of the European Alps (Aeschimann & Lauber, 2004). However, for systematic biologists, willows have always posed a difficulty. One of the troubling taxonomic and phylogenetic aspects of this genus is frequent hybridization. High crossing ability across the genus and even among distantly related species has been examined with experimental crossings (Wichura, 1865, Argus, 1974) and observed in nature (Gramlich et al., 2018). Flexible pollination by insects, as well as facultatively by wind, and long-distance dispersal of airborne seeds, enable

high gene flow between individuals, which, on the account of the absence of reproductive barriers, can result in hybridization (Hörandl et al., 2012). Favouring of permeable interspecific gene flow to enhance genetic variability in combination to long-distance dispersal is linked to willows' ecological role to colonise open niches as pioneer species (Hörandl et al., 2012). Particularly closely related taxa of the same ploidy level are likely to naturally form homoploid hybrids (Mosseler, 1990). Homoploid hybridization of willows is frequently linked to species coming into secondary contact after glacier retreat. Several studies have examined hybrid zones of different *Salix* species pairs. While some pairs, like *S. alba* and *S. fragilis,* were observed to only form distinct homoploid F1 hybrids (Oberprieler et al., 2013), most other studies detected patterns of introgression (Hardig et al., 2000, Fogelqvist et al., 2015, Gramlich & Hörandl, 2016, Gramlich et al., 2016, 2018). Hybrid individuals were also observed to evolve adaptive growth form traits and exhibited higher tolerances to certain abiotic conditions (Gramlich et al., 2016, 2018). However, even though this species-rich genus is well known for many hybridizing taxa with frequently reported hybrid phenotypes in nature, they were rarely examined in detail and over a broader geographical extent (Hörandl, 1992) and therefore studies on willow hybrid zones remain relatively scarce.

Study case: Salix foetida and Salix waldsteiniana

Two closely related European *Salix* species of section *Arbuscella* (as classified by Skvortsov, 1999), *S. foetida* Schleich. ex. DC. and *S. waldsteiniana* Willd., have been observed to form intermediate phenotypes in regions in the Alps where their distributions overlap (Hörandl, 1992). Traditionally, the two species were recognised as a vicariant pair with a west-east distribution pattern (Hörandl et al., 2012). In the Alps, the species show an overlap in the central region (Figure 1). Outside the Alps, *S. foetida* has a western distribution in the Pyrenees and Apennines and *S. waldsteiniana* an eastern distribution in the Dinarides and the Balkan Peninsula. While it is unclear, whether the species separated as a result of allopatric speciation during glacier retreat or whether their origins lie outside the Alps and the species have reached the area via long-distance dispersal (Wagner et al., 2021), their sister position has so far been genetically confirmed (Wagner et al., 2020). Moreover, the species tend to show differences in habitat preferences, with *S. waldsteiniana* primarily occurring on carbonate bedrock soils and *S. foetida* on silicate bedrock soils and more moist habitats, which can also aid in species identification in the field (Hörandl, 1992, Hörandl et al., 2012). Whether this has played a role in ecological niche differentiation remains unclear (Wagner et al., 2021). Both species are diploid with 19 chromosome pairs (2n = 38) (Neumann & Polatschek, 1972, Büchler, 1985, Dobes et al., 1997). The key morphological characteristics for identification are based on the anatomy of leaves and catkins. As all the members of *S.* sect *Arbuscella,* both species, as described by Skvortsov (1999), can be recognised by their typical reddish shoots, contrasting bicolouration of leaves (adaxial site dark green and lustrous, abaxial whiteish with straight and slender veins), and denticulate leaf margin. According to Hörandl et al. (2012) the two species can be identified by their leaf dentation patterns, with *S. foetida* having dense, pronounced, well perceptible teeth with well visible, brightly coloured glands, and *S. waldsteiniana* having widely separated, less pronounced teeth with inconspicuous, darker glands (Figure 2). The leaves of *S. waldsteiniana* are also often larger, elliptic to obovate with a larger width-to-length ratio, than those of *S. foetida*, which are sharper elliptic to lanceolate (Figure 3). Catkins, both

male and female, are larger in *S. waldsteiniana*, but otherwise very similar in both species. The observed putative hybrid phenotypes, as described in Hörandl (1992), show intermediate dentation patterns with bright glands, resembling those of *S. foetida*, or they are densely dentated, but the glands are small and more similar in appearance to those of *S. waldsteiniana*. The size of the leaves was reported to be within the range of variability of both species. Observation of deformed and fruitless female catkins led to a hypothesis, that the intermediate individuals are most likely majorly represented by a reproductively unsuccessful first generation of hybrids.

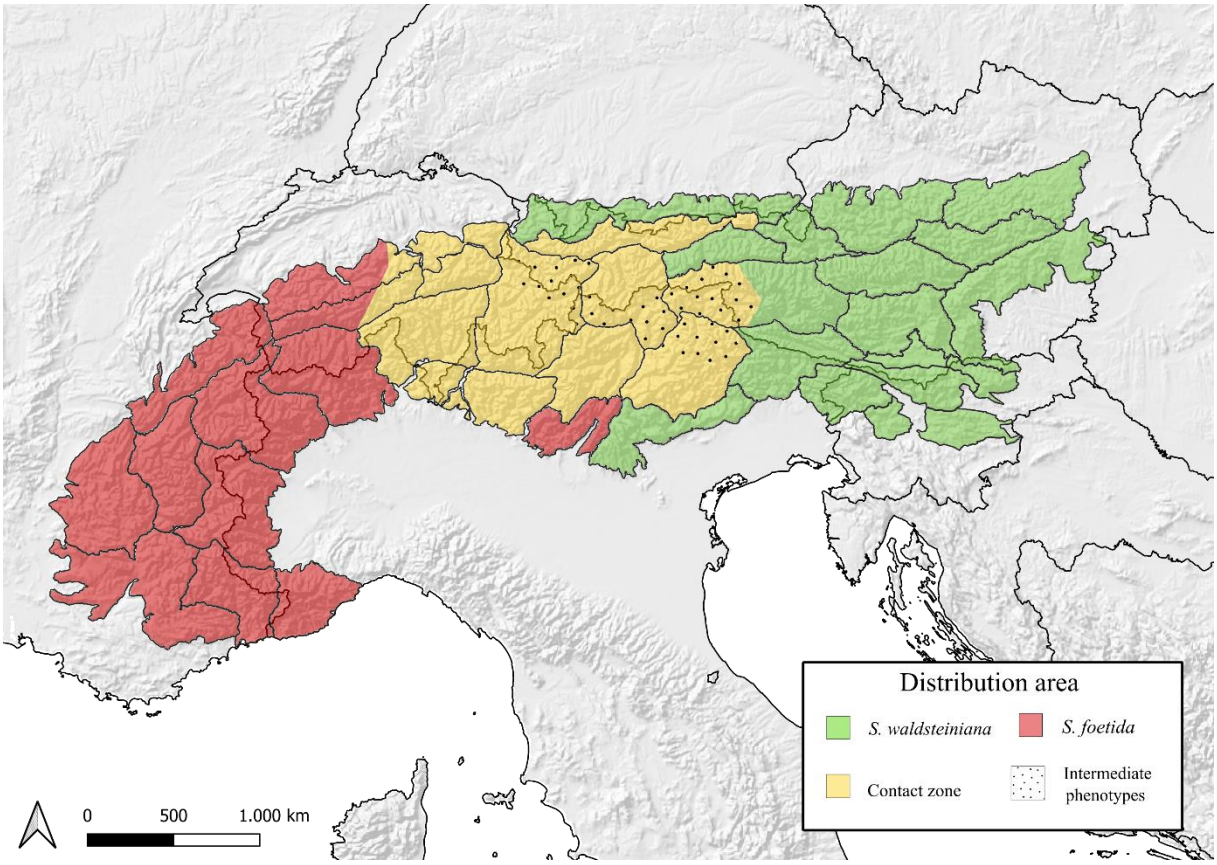

Figure 1: Distribution map of *S. foetida* and *S. waldsteiniana* in the Alps. The distribution is mapped according to distribution maps in Flora alpina (Aeschimann & Lauber, 2004). The areas of intermediate phenotypes are mapped according to observations recorded in Hörandl (1992). The map was created with *QGIS* (v.3.16.3).

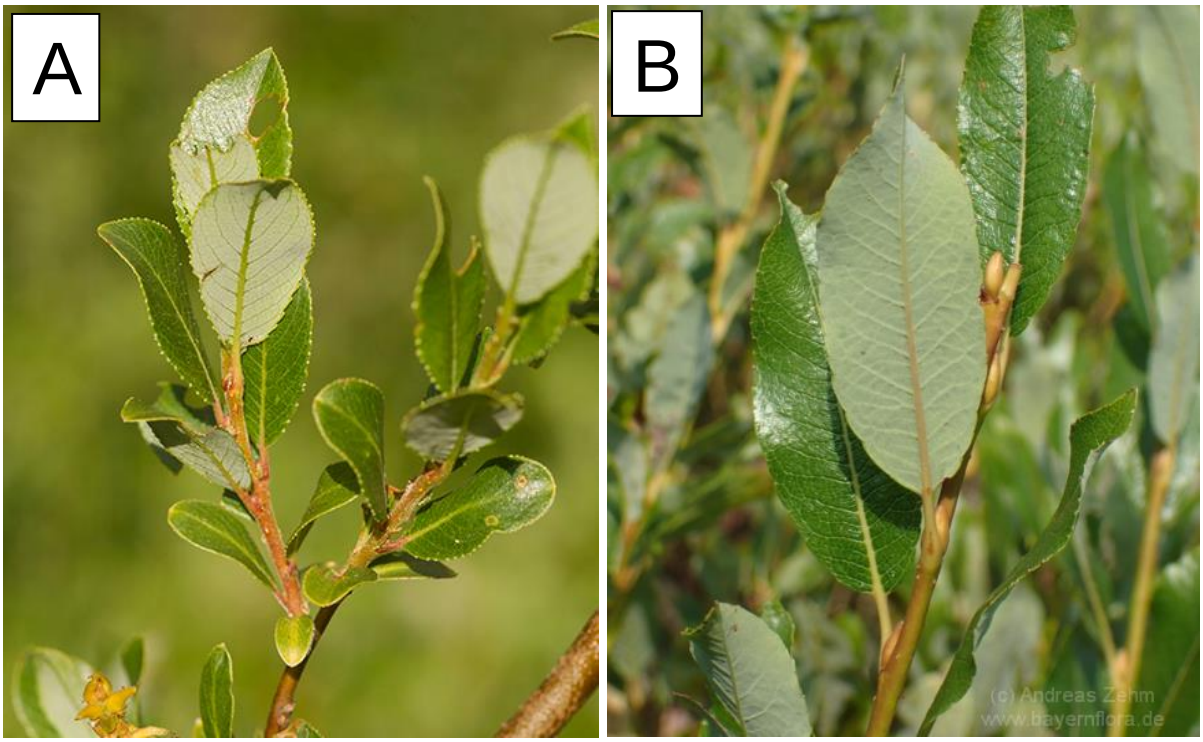

Figure 2: Comparison of photographed vegetative shoots of **A) *S. foetida*** (© orchid-nord.com, author unknown) and **B) *S. waldsteiniana*** (© Andreas Zehm)

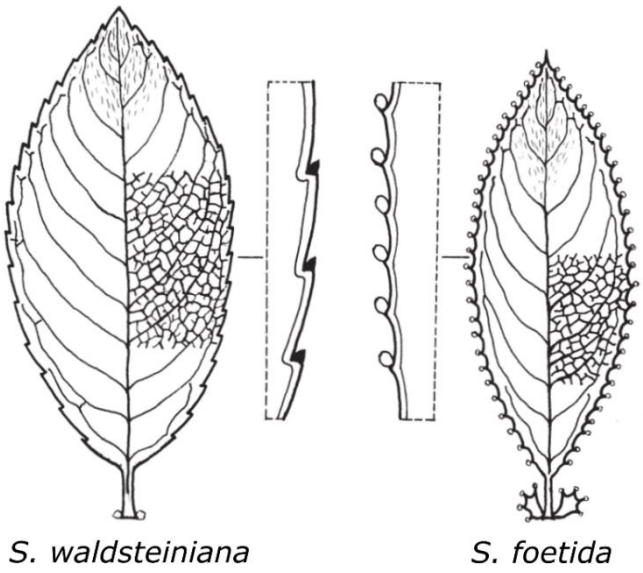

Figure 3: Illustration of leaves of *S. waldsteiniana* and *S. foetida*. Adapted from Hörandl (1992).

1.2. Methodological approaches to investigate hybridization

Restriction site-associated DNA sequencing (RAD-seq)

RAD-seq technique, described by Baird et al. (2008), is a cost-effective method for reduced-representation, genome-wide SNP genotyping. The genomic DNA of the individual is cut using a restriction enzyme that recognises specific target sequences and breaks the DNA strain at these sites. This creates fewer target sequences in comparison to whole genomic DNA, resulting in deeper sequencing coverage and highly confident genotype calls. With the use of widely applicable restriction site enzymes, an individual's genome can be studied without any prior genomic background, making this technique widely popular in ecological and evolutionary genomics of non-model organisms (reviewed in Andrews et al., 2016). In *Salix*, several studies using chloroplast, ribosomal and nuclear markers (Azuma et al., 2000, Hardig et al., 2010, Percy et al., 2014, Wu et al., 2015) could not resolve infrageneric phylogeny with high confidence and showed a high degree of polyphyly as a result of low levels of variability and reticulate processes, such as hybridization, introgression, and incomplete lineage sorting. But in the last decade, RAD-seq approaches were proven to be a powerful tool to disentangle phylogenetic relationships of *Salix* clades and species (Wagner et al., 2018, Wagner et al., 2020), radiation patterns (He et al., 2021) as well as genetic structure of natural hybrid populations (Gramlich et al., 2018).

Morphometrics

The shapes of leaves, petals, and whole plants are greatly significant in plant science, as they are crucial in the identification of species and can provide information on the individual's ecology, health and evolutionary biology (Cope et al., 2012). To study them, accurate and objective measurements and their statistical analyses ("morphometrics") are required. Informative quantitative traits of floral or vegetative tissues can be measured, calculated, and compared as simple functions, usually as distance metrics (Lexter et al., 2009). Digitalised images of selected tissues are easily stored and can be used in subsequent analyses with a new set of methodologies to come. Such examples are the analytical methodologies used in the digital analysis of shapes, coined under the term "geometric morphometrics" (Mitteroecker & Gunz, 2009). Leaves offer several characters that are extensively used in traditional taxonomic keys (Cope et al., 2012). Genus *Salix* is no exception. The taxa are most easily distinguished during the growing season by observing leaf anatomy (Hörandl et al., 2012). Several leaf traits, such as leaf area, dentation pattern, length, width, and distance to the widest point of the leaf, can be quantitatively measured and analysed.

1.3. Aims of the study

The scope of this Master's thesis was to investigate the hybrid zone and hybridization of *S. foetida* and *S. waldsteiniana* in the European Alps using SNP molecular markers obtained via RAD-seq protocol and to conduct morphometric analysis of selected phenotypic leaf traits. With this, I aimed to answer the following questions: i) Do the two sister species hybridize in the overlapping range of the distribution area in the Alps resulting in local homoploid hybridization events or in a broader hybrid zone? ii) What is the genetic structure of those hybrid zone populations and what conclusions can be made regarding the processes involved (e.g. hybrid speciation or permeable gene flow and introgression? iii) Can analyses of

morphological traits unravel hybrid phenotypes, distinguish them from the parental ones, and are they in concordance with the genetic structure of individuals?

2. Material and Methods

2.1. Collection of samples

In the years between 2014 and 2020, over 190 individuals of *S. foetida, S. waldsteiniana,* and putative hybrids were sampled over the whole distribution area in the European Alps. Sampling included purebred populations outside the contact zone and populations from within the two species' distribution overlap, focusing on localities with intermediate phenotypes, as summarised in Hörandl (1992) (Figure 1). Populations outside the contact zone included: Two *S. foetida* populations from Switzerland (population CH3 from Great St. Bernard and population CH4 from Julier pass) and two *S. waldsteiniana* populations from Slovenia (population SI1 from Srednji Golak and population SI2 from Ratitovec) (Figure 4A). Further samples of single individuals from previous projects (Wagner et al., 2018, Wagner et al., 2020) were added to the purebred populations: three *S. waldsteiniana* individuals sampled in Styria and Lower Austria, Austria, marked as population AU10, three *S. foetida* individuals from Rhône Glacier, Switzerland, marked as population CH11 and a single individual from Vinschgau, Italy, marked as population IT12. Populations within the contact zone included: Populations IT5 and IT8 from South Tirol, Seiser Alm, Italy, populations IT7 and IT9 from Grödnerjoch, Italy, and population IT6 from Zirog Alps, Brennerbad, Italy (Figure 4B). Per population, between 16 to 23 individuals were sampled. Several undamaged leaves were removed from each selected individual, isolated in paper tea bags, and preserved in silica gel for DNA extraction. Herbarium vouchers for each sample were deposited in the herbarium of the University of Göttingen (GOET). Each individual was identified by the respective collectors and successful identification was later evaluated on herbarium vouchers by E. Hörandl. Individuals with intermediate phenotypical characters were classified as hybrids (population IT9 from Zirog Alps, Brennerbad, Italy, and population IT6 from Grödnerjoch (direction Corvara), Italy.

The selection of 95 appropriate samples for molecular analysis was done based on the amount of extracted DNA (for more information see chapter "*2.2. DNA extraction and RAD-seq"*). Together with additional, already sequenced individuals used in Wagner et al. (2018) and Wagner et al. (2020), this added up to 102 individuals from twelve different sampling localities included in the molecular analyses (Table 1).

Table 1: Table of all sampling sites ordered geographically from west to east. Populations labelled with an asterisk (*) mark populations of single samples used in previous studies (Wagner et al., 2018, Wagner et al., 2020). Populations sampled in the contact zone are coloured grey.

Population ID	Country	Species (phenotype)	Locality	Latitude	Longitude	Sample IDs	Altitude	No of individuals per site	No of DNA samples used	No of individuals used for morphometrics	Date of collection	Collector
CH3	Switzerland	*S. foetida*	Great St. Bernard	45°53'45.1" N	7°11'11.9" E	EH10958 - EH10974	2040	16	10	10	4.08.2020	Elvira Hörandl
CH11*	Switzerland	*S. foetida*	Rhône Glacier	46°33'59.2" N	8°22'23.7" E	SG35, 36, 41	1780	3	3	0	18.7 - 19. 7. 2014	Susanne Gramlich
CH4	Switzerland	*S. foetida*	Julier pass, direction Tiefencastel	46°27'40.0" N	9°40'58.1"E	EH10832 - EH10847	1905	18	10	10	26.07.2020	Elvira Hörandl
IT12*	Italy	*S. foetida*	Vinschgau	46°31'33.8" N	10°35'10.9" N	EH10273	1850	1	1	0	4.08.2015	Elvira Hörandl
IT6	Italy	hybrid	Zirog Alps, Brennerbad	46°57'57.0" N	11°29'53.4" E	EH10542 - EH10564	1855	23	11	11	2.08.2019	Elvira Hörandl & Franz Hadacek
IT8	Italy	*S. waldsteiniana*	Seiseralm, Alpenrosenhütte	46° 30' 55.3" N	11°38'41.2" E	EH10684 - EH10702	2035	19	11	11	8.08.2019	Elvira Hörandl & Franz Hadacek
IT5	Italy	*S. foetida*	Seiseralm, Sonne	46°33'01.5" N	11°39'56.0" E	EH10703 - EH10722	1827	18	11	11	8.08.2019	Elvira Hörandl & Franz Hadacek
IT7	Italy	*S. waldsteiniana*	Grödnerjoch	46° 32' 46.8" N	11° 48'21.9" E	EH10644 - EH10663	2050	20	11	11	6.08.2019	Elvira Hörandl & Franz Hadacek
IT9	Italy	hybrid	Grödnerjoch, direction to Corvara	46° 32' 54.4" N	11°48'48.3" E	EH10664 - EH10681	2056	18	11	11	6.08.2019	Elvira Hörandl & Franz Hadacek
SI1	Slovenia	*S. waldsteiniana*	Srednji Golak	45°58'33" N	13°52'32" E	PM20_006 - PM20_026	1383	21	10	10	13.06.2020	Pia Marinček
SI2	Slovenia	*S. waldsteiniana*	Ratitovec	46°14'17" N	14°5'56" E	PM20_031 - PM20_055	1350	21	10	10	17.6. - 18.6. 2020	Pia Marinček
AU10*	Austria	*S. waldsteiniana*	Styria & Lower Austria	47°39'54.1" N - 47°43'01.3" N	13°46'55.8" E - 15°46'33.7" E	NW17_018, 063, 069	1550 - 1780	3	3	0	3.8. - 8.8. 2017	Natascha Wagner & Katrin Scheufler

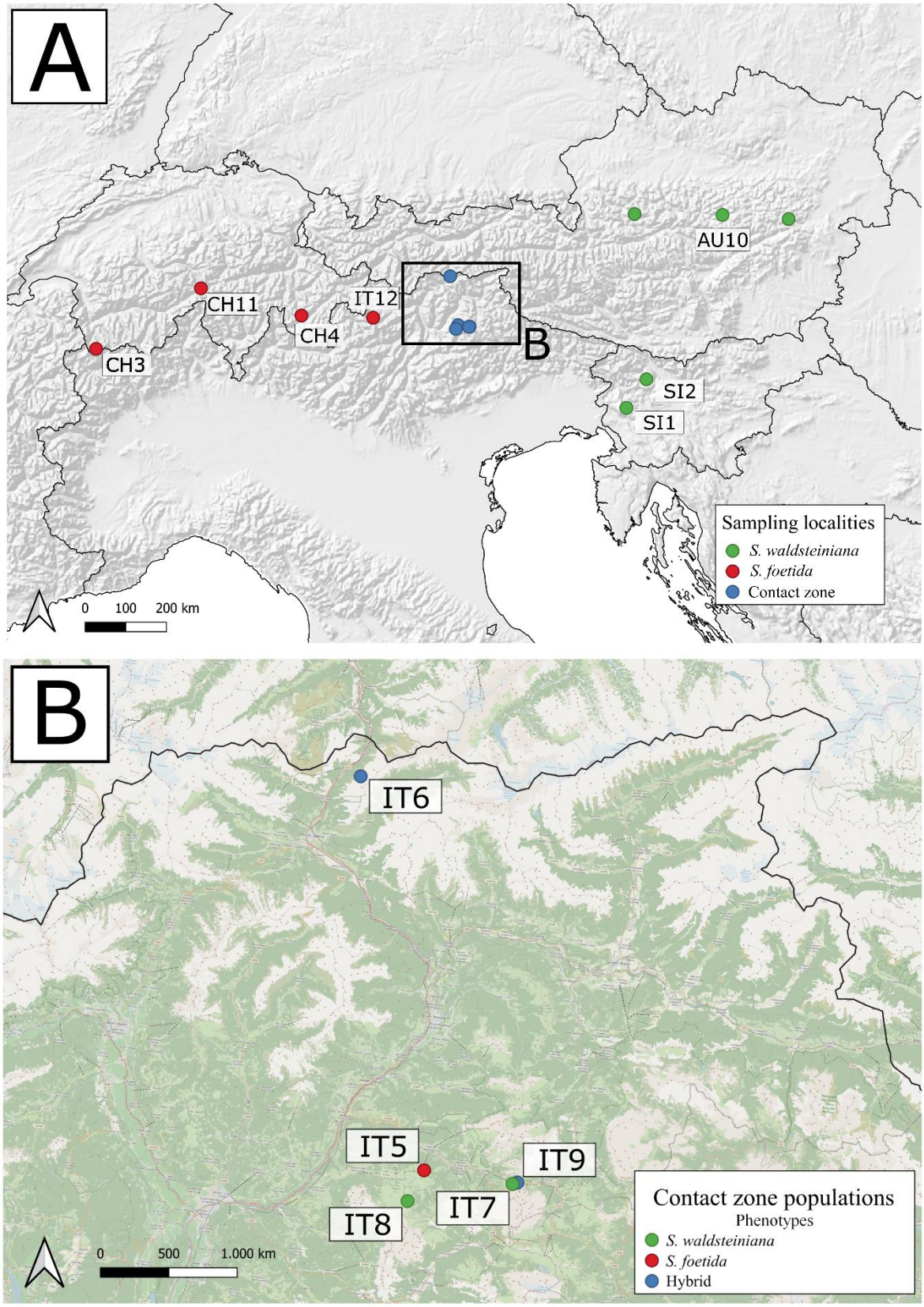

Figure 4: **A) Sampling localities used in the study**. The black square marks the closed-up region of populations of the contact zone. **B) Close-up of the populations from the contact zone.** Maps were created with *QGIS* (v.3.16.3).

2.2. DNA extraction and RAD-seq

Between 10 – 20 mg of silica-dried leaf material per individual extraction were collected and crushed using TissueLyser II (Qiagen, Hilden, Germany) at 30 Hz for one minute with the use of 5 ⌀ mm steel beads. DNA was extracted using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following a modified manufacturer's protocol to increase the DNA yield. In step 2, the volume of AP1 buffer was adjusted between 400 to 450 µL, depending on the quantity of leaf material used in the extraction, the volume of RNase A was decreased to 2.5 µL, and incubation at 65°C was 30 minutes long. In step 3, incubation on ice was increased to a minimum of 30 minutes. In step 11, 60 µL of AE buffer was added and the sample was incubated for 30 minutes at room temperature. In step 12, step 11 was repeated, but with 30-40 µL of AE buffer, adding to a final DNA extract volume of 90-100 µL. Each individual was extracted (at least) twice due to previous experience with low DNA yields of the genus *Salix*. In total, the DNA of 133 individuals was extracted in 285 extractions and re-extractions (Appendix 7.3). DNA quantity was measured with a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA). The samples were required to contain at least 20 ng/µL of DNA, or average concentrations of two extracts of the same sample had to be at least 15 ng/µL for successful pooling. Based on these criteria, more than 180 extracts were eligible for sequencing. Individuals with highest DNA yields were selected for the implementation of RAD-seq technology. This included ten individuals from each population outside the contact zone and eleven individuals from each of the remaining five populations from the contact zone, adding up to a total of 95 samples. They were sent to Floragenex Inc. (Beaverton, USA), where the company performed library preparation for single-end RAD-seq using the *Pst*I restriction enzyme, and sequenced the samples on Illumina HiSeq 2500 platform (Illumina Inc., San Diego, USA). All DNA extracts had to be normalised to 20 ng/µL and shipped frozen, as required by the company. After sequencing, raw sequences were available in the fastq file format.

2.3. Bioinformatic analysis

2.3.1. Assembly of RAD-loci

Software for the RAD-loci assembly

The reads produced by the RAD-seq workflow are assembled into so-called RAD-loci. Several open access software tools were developed for the construction of the RAD-loci assembly. The selection of an appropriate tool must be study specific, as it determines the outcome of all subsequent analyses. In this study, two frequently used RAD-loci assembly tools were used: *ipyrad* (v.0.9.52) (Eaton and Overcast, 2020) and *STACKS* (v.2.2.) (Catchen et al., 2013). Due to different assembly algorithms, the use of *ipyrad* is intended for phylogenetic research and *STACKS* for population genetics. In both pipelines, the *de-novo* assembly approach was used.

Before the assembly, raw reads must be assigned to each sample -"demultiplexed"-, based on the designated 10 bp long barcodes. In *ipyrad*, demultiplexing is performed in step 1 of the pipeline, in *STACKS* in the *process_radtags* program. Because this is a uniform step, it must only be performed once. The *ipyrad* assembly approach was chosen for this. On average, 6.94 million reads were assembled per individual, the minimum was 3.86 million, and the maximum was 40.44 million. No population significantly deviated from the mean. After demultiplexing, the quality of sequenced bases was checked using *FastQC* (v.0.11.9) (Andrews, 2010). In this

software tool, the information of base call quality stored in the fastq file format is summarised and visualised, along with some other statistics of the sequenced reads. The accuracy of base calls in the fastq file is represented with a Phred score (Q) (Ewing et al., 1998, Ewing & Green, 1998), which is a measure of the quality of the identification of nucleobases. Q30 and higher represents a nearly 100 % (99.9%) accurately called base and is considered as a threshold for a successful call. All individuals showed high Q of all bases that dropped slightly after the first 20 bp, but in no cases below the threshold of Q30, and no sequences were flagged for poor quality. Thus, no additional quality filtering was needed

In *ipyrad*, RAD-loci are assembled in seven consecutive steps, the first one being the demultiplexing. Steps 2 – 5 process the data for each sample through filtering, clustering, and identifying a consensus sequence. In step 6, orthologues across samples are identified and clustered and finally, in step 7 the orthologues are filtered as specified by the user and written in selected output formats. The information about applied parameters and filters for the final assembly is stored in the so-called "params file", which the user can edit according to their own specific requirements.

The *STACKS* pipeline consists of three programs. The first is the *process_radtags* program, which demultiplexes the data. The second is the assembly pipeline, wrapped in the *denovo_map.pl* wrapper for *de-novo* assembly. It consists of three programs: *ustacks* merges RAD-tags into loci for each individual, *vstacks* creates a catalogue of merged loci of all individuals used in the pipeline, and *sstacks* matches the loci from each individual against the assembled catalogue. Then the *populations* pipeline is run where loci and SNPs can be filtered and several different outputs can be created for further downstream analyses. In *STACKS*, parameters and filters are applied as flags when calling the pipeline command. The *populations* pipeline also requires the use of a population map; a simple text file, where each individual is assigned a population (in *ipyrad* this is optional).

Similarity threshold for clustering of loci

In the de-novo approach, putative RAD-loci are assembled without a reference. Thus, the assembly of homologous loci is not straightforward and depends on the choice of appropriate parameters that define the similarity at which reads assemble into a homologous locus. In *ipyrad,* this is defined with the *clustering_threshold* parameter. In *STACKS*, parameters *M* (number of nucleotide mismatches allowed between stacks to merge them into a putative locus of an individual) and *n* (number of nucleotide mismatches allowed between stacks during construction of the loci catalogue for all analysed individuals) define the similarity threshold. The default settings of both assembly software tools are rather different, and there is no other rule in selecting appropriate thresholds, but to empirically examine a range of them and decide on the most appropriate parameter value(s) by investigating certain metrics.

In *ipyrad*, a range of clustering values from 0.75 to 0.97 was tested, with a step increase of 0.05 between 0.75 to 0.9, and a step increase of 0.01 between 0.9 and 0.95 for a finer scale resolution. In *STACKS*, parameter *m* (minimum number of raw reads required to form a stack) was set to 10, due to observed high coverages of stacks. The use of the same setting of *m* in Gramlich et al. (2018) backed up this selection. Then, a range of *M* and *n* from 2 to 8 was tested for the selected *m*, where *M* and *n* were kept at the same value, as recommended by Rochette & Catchen, (2017).

Frequently, metrics such as maximised number of loci or SNPs, are suggested to be used as a criterion of optimality (Mastretta-Yanes et al., 2015, Paris et al., 2017). However, the number of loci or SNPs is not a sole indicator of an appropriate similarity threshold selection. To investigate and identify the most appropriate similarity parameters, six metrics from a recently published empirical pipeline by McCartney-Melstad et al. (2019), and an additional seventh metric for the *STACKS* assembly from Paris et al. (2017), were used to help identify the optimal threshold of the *ipyrad* and *STACKS* pipeline. A detailed description of the investigated metrics can be seen in Appendix 7.4.

Optimal assembly of RAD-loci

Empirical runs for clustering similarity thresholds were run with all individuals and tested with metrics described in Appendix 7.4. For *ipyrad* assembly, a clustering threshold of 0.93 (93% similarity) was selected as the most appropriate, and for the *STACKS* assembly, the combination of *m* 10, *M* 3, and *n* 3 (translating to roughly 93% similarity) was identified as the most fitting (Table 2). The assembly of both software tools was compared, with a minimum sample per locus filter in *ipyrad* set to 81 and filter *r* in *STACKS* set to 80 (Figure 5).

Table 2: Optimal similarity settings for clustering of putative loci based on six metrics from McCartney-Melstadt et al. (2019) and an additional metric for maximisation of polymorphic loci at *r 80* from Paris et al. (2017) that was only investigated for the empirical *STACKS* assemblies.

	Optimal similarity threshold settings	
	ipyrad *(clustering threshold)*	***STACKS*** *(M/n)*
Fraction of inferred paralogues	0.9 - 0.93	3 - 5
Genetic similarity: heterozygosity	0.9 - 0.95	2 - 7
Genetic similarity: number of SNPs	0.9 - 0.93	4 - 6
Variance explained by first 8 PCs	0.93	3
PCC between pairwise missingness and genetic divergence	0.93	4
Percentage of increased SNP divergence per 100 km	0.9 - 0.93	3
Maximisation of polymorphic loci at *r 80*	/	3
Selected optimal similarity clustering threshold	**0.93**	**3**

.

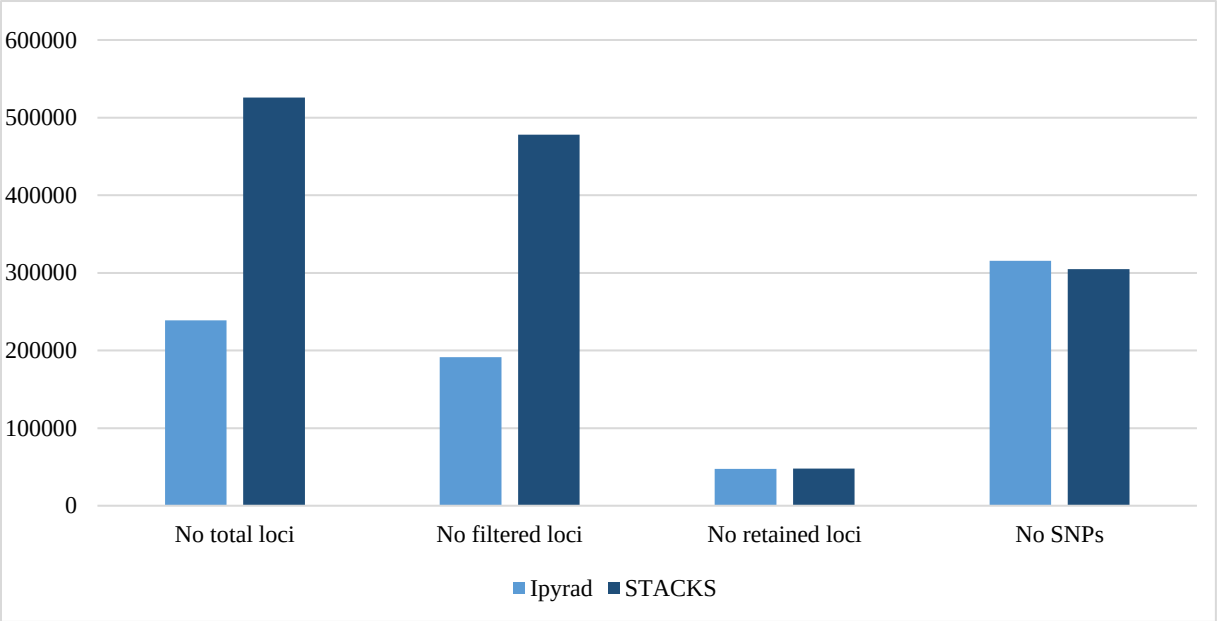

Figure 5: Comparison of RAD-loci assembly for *STACKS* and *ipyrad* pipeline at optimal similarity thresholds. Both final assemblies consist of loci shared between at least 80% of all individuals in *STACKS* and at least 81 samples in *ipyrad* assembly (79.4% of all individuals).

To explore the content of loci belonging to the more conserved plastid regions, which would not be informative in this analysis, two *ipyrad* assemblies were run with the two species' plastid genome references. The complete chloroplast genome sequences were obtained from the GenBank database (accession numbers MW435429.1 and MW435453.1.). Only 40 and 42 of the total loci mapped to the *S. waldsteiniana* and *S. foetida* reference plastome, respectively, indicating that the fraction of more conserved plastid loci was low. No filtering for plastid regions was needed.

Preparing the RAD-loci assembly for population genetics analysis

For the analysis of population genetics and population structure, further filters were applied to the *STACKS* assembly in order to remove problematic aspects of the full RAD-loci assembly (missing data, non-informative loci, linkage disequilibrium, etc.) The selection of loci based on their frequency among individuals and populations is defined with flags *r* (minimal percentage of samples within a defined population where a locus must be present to be retained) and *p* (minimal number of populations where a filtered locus must be present to be retained). Even when these flags are kept at the same value, the information in the population map affects the assembly with the definition of groups (Graham et al., 2020). In this analysis, individuals were divided into three larger groups. Two groups consisted of parental individuals collected outside the contact zone, and in the third group, all populations from within the contact zone were grouped together, regardless of their identified phenotype. The parameters *r* and *p* were set to 80 and 1, respectively. To minimise the amount of missing data, loci below the minimum allele frequency of 5% were discarded using a *min-maf* flag in *STACKS*. Potentially paralogous loci, putatively consequential to a recent salicoid duplication event (Tuskan et al., 2006), that are characterised by an excess of heterozygosity (Hohenlohe et al., 2011) were filtered based on

their levels of heterozygosity and F_{IS} - *value*. Loci above the observed heterozygosity of 60 % were discarded with the *max-obs-het* flag, and loci with a F_{IS} - *value* (calculated with the *fstats* flag in *STACKS*) below zero via manual extraction from the *sumstats* file to a blacklist using *Microsoft Office Excel* (v.2016) (Microsoft Corporation, Redmond, USA). Parental loci (referring to the loci present in the populations of parental species outside the contact zone) that were out of Hardy-Weinberg equilibrium (HWE) at *p-value* < 0.05, were extracted to a blacklist in the same manner. Because several analysis tools require unlinked markers, the *write-single-snp* flag was used where only the first SNP (that passed the minimum allele frequency and maximum observed heterozygosity constraints) per locus was kept. *STACKS* also offers an option to keep a random SNP from a locus, but to ensure the reproducibility of the results, the option of keeping the first SNP per locus was made. Next, the appropriate outputs (generated by the *plink* flag) were used in *PLINK* (v.1.90) (Purcell et al., 2007) to remove the blacklisted loci with F_{IS} - *value* < 0 and parental loci out of HWE with the use of the *exclude* flag. To decrease missing data, SNPs with a genotyping rate below 10% were removed with the flag *geno* in *PLINK*. Throughout these steps, several thousand of loci had to be discarded from the initial assembly, but this ensured that the remaining dataset was appropriate for all the subsequent population genetics analyses.

Final RAD-loci assemblies

The *ipyrad* assembly, run on the optimal clustering threshold (0.93) with a filter for minimum sample per locus set to 81, generated 47,297 RAD-loci with 315,576 SNPs with 6.3 % of missing SNP data.

An additional *ipyrad* assembly on the same settings was constructed, with an inclusion of two outgroup *S. viminalis* samples (from Wagner et al., 2018), adding to 104 individuals. The final assembly consisted of 47,899 RAD-loci with 336,649 SNPs and 6.6% missing SNP data.

The *STACKS* assembly, run with optimal parameters (*m* 10, *M* 3, *n* 3) and filtered for loci present in 80% of individuals per population, generated 50,924 RAD-loci with 329,581 SNPs. After additional filtering, the assembly consisted of 6,173 unlinked SNPs with 4.4% missing SNP data.

2.3.2. Downstream analysis

Phylogenetic analysis

To investigate the relationships between the individuals in a phylogenetic tree format, two tree-building methods were applied. A maximum likelihood (ML) building approach with a GTR+Γ model of nucleotide substitution was used in a software *RAxML* (v.8.2.4) (Stamatakis, 2014) with rapid bootstrapping analysis of 100 replicates. The ML tree was generated using the complete RAD-loci alignment of the *ipyrad* assembly with outgroups. It was visualised with *FigTree* (v.1.4.4) (Rambaut, 2012). Then, the branch support values of the ML tree were quantified with the Quartet Sampling (QS) method (Pease et al., 2018), implemented in *quartetsampling* (v.1.3.1). This approach takes an already existing phylogenetic topology and evaluates it by generating replicates for each internal (i.e., non-terminal) branch, calculating the likelihoods of the three possible topological arrangements of the four taxa linked to the branch (the quartets) for all internal branches. The pipeline generates three branch scores that give a further evaluation of the tested topology. The first score is quartet concordance (*QC*), which

describes whether the frequency of the concordant quartet is inferred over both discordant quartets (*QC* = 1), whether the frequencies of concordant and discordant quartets are equivocal (*QC* = 0), or whether discordant quartets are represented more frequently than concordant ones (*QC* < 0*)*. The second score is quartet differential (*QD*) which explains whether the frequencies of the two discordant quartets are equal *(QD* = 1), skewed (*QD* = 0.3), or whether all discordant topologies belong to one of the discordant quartets (*QD* = 0). The last score is quartet informativeness (*QI*), which tells what proportion of replicates were informative.

Phylogenetic networks extend phylogenetic trees and are able to more explicitly represent the reticulate evolutionary history of taxa in processes like hybridization (Huson & Bryant, 2006). Neighbour-Net (NN) (Bryant & Moulton, 2004) is a variant of a widely used Neighbour-Joining (Saitou & Nei, 1987) distance-based method for the construction of phylogenetic networks. The NN algorithm is implemented in the *SplitsTree* (v.4.17.1) (Huson & Bryant, 2006) software to calculate and visualise split phylogenetic networks that represent incompatibilities in the dataset. This results in a reticulate representation of the phylogenetic network. A NN network-building algorithm with 100 bootstrap replicates was applied to the complete SNP alignment of the *ipyrad* assembly with outgroups.

Detection of reticulate evolution

Though we can get some impression of the presence of hybridization in a dataset by creating phylogenetic trees and networks, an algorithm implemented in the *HyDe* software package (Blischak et al., 2018) uses a phylogenetic inference approach to test the dataset for hybridization patterns. The algorithm of this software was designed to detect admixture with a set of functions called phylogenetic invariants, that arise under a coalescent model of hybridization when applied to quartet networks. One group of the quartet is an outgroup, two are treated as parents, and one is a hybrid group. Based on *D* - statistics (Green et al., 2010), the probability of the prediction of a constructed model to fit the observed pattern of the putative hybrid taxon, is evaluated. Significant predictions have a *p–value* of 0.05 or less. Estimated *γ-values* explain detected hybridization patterns: values between 0.4 and 0.6 hint to (recent) hybridization events and values between 0.2 - 0.4 and 0.6 - 0 .8 hint to introgression or ancient hybridization events. The data can be tested for populations or individuals as specified in a population map. For this analysis, the complete RAD-loci alignment of the *ipyrad* assembly with outgroups was used. The population map defined all parental individuals from populations outside the contact zone as either *S. waldsteiniana* or *S. foetida* population, and each hybrid individual was assigned its own group. A triplet file of 55 triplets was created to test each individual within the contact zone as a hybrid of both parental species, where *S. foetida* population was assigned as the first parent and *S. waldsteiniana* as the second parent. Triplets with significant hybridization (*p-value* < 0.05) were tested with 100 bootstrap replicates, using *bootstrap_hyde.py* script. The average *γ-values* of the replicates were visualised with *ggplot2* (v.3.3.5) package (Wickham, 2016) in *R* (v. 4.1.2) (R Core Team, 2021).

Population structure with STRUCTURE

To explore the presence of distinct genetic groups within the dataset, *STRUCTURE* (v.2.3.4) (Pritchard et al., 2000) was used. In this software, the genetic structure is analysed with a calculation of ancestry coefficients that represent the proportions of an individual's genome that originate from multiple ancestral gene pools. This is performed with a Bayesian likelihood

model, which calculates probabilities that the observed genetic patterns of each individual match ancestral patterns of a predefined number of genetic populations. The number of populations is defined with a parameter *K*. Five runs per *K* ranging from 1 to 13 were calculated with 50,000 (Monte Carlo) MCMC iterations and a 10,000 burn-in period. The most likely *K* was determined with *STRUCTURE HARVESTER* (v.0.6.94) (Earl & vonHoldt, 2012) using Evanno's method (Evanno et al., 2005) for detecting the most likely number of *K* based on calculated mean likelihood values per K. The separate runs per *K* were summarised in a single output using *CLUMPP* (v.1.1.2) (Jakobsson & Rosenberg, 2007), with a *Greedy* method for 10,000 repeats. The results were visualised using *Microsoft Office Excel.* A pre-analysis of the filtered *STACKS* unlinked SNP assembly revealed, that population IT5 presented a strong bias and consistently formed a separate partition (Appendix 7.5). A reduction of this population to three individuals removed the bias and therefore the dataset was reduced to 94 individuals. The reduced dataset was also used in subsequent analyses where population IT5 posed the same problem.

Population structure with sNMF

The likelihood model analysis in *STRUCTURE* requires a lot of computational effort and time. In addition, it is sensitive to systematically appearing missing data (Pritchard et al., 2010, Arnold et al., 2013), because the model recognises hierarchical missingness as an ancestral pattern. *STRUCTURE* also assumes HWE and linkage equilibrium among loci, which can bias the analysis. A much faster, yet reliable alternative is the Sparse Non-Negative Matrix Factorization algorithm (sNMF) (Frichot et al., 2014) implemented in the function *snmf* in *R* package *LEA* (v.1.4.0) (Frichot & François, 2015). In this approach, ancestry proportions of *K* ancestral populations are estimated with least-square estimates. The analysis is robust for deviations from HWE. To evaluate the right number of *K*, an entropy criterion is estimated and visualised on a plot. The *sNMF* analysis of population structure was performed on unlinked SNP filtered dataset from *STACKS* assembly for the reduced dataset of 94 samples, due to population IT5 bias (Appendix 7.6). The *K* was tested in a range from 1 to 20. In addition, the advantageous speed of this calculation approach also allowed for application on 46,695 unfiltered, unlinked SNPs of the *ipyrad* assembly for comparison. Results were visualised with *Microsoft Office Excel.*

Population structure with RADpainter and fineRADstructure

An approach offering a high resolution of genetic structure based on haplotype linkage information, focusing on the most recent common ancestry, is implemented in the *RADpainter* and *fineRADstructure* pipeline (Malinsky et al., 2018). With the use of a coancestry matrix (a summary of nearest neighbour haplotype relationships in the dataset) and MCMC clustering algorithm, a fine-scale structural resolution of individuals can be calculated relatively fast even from a large SNP dataset. The outputs are visualised with a heatmap of coancestry coefficients with the use of an *R* script provided by the authors. Filtered *STACKS* assembly was used for this analysis.

Assignment to parental and hybrid classes

Even though some conclusions on the level of admixture can be done based on the results of the population structure analyses in *STRUCTURE* and *sNMF*, the software *NewHybrids* (v.1.1 beta) (Anderson & Thompson, 2002) was developed specifically to identify and assign parental

and hybrid classes of examined individuals. The classes are calculated as the posterior distribution of individuals falling into different parental or hybrid categories, based on the match of their SNP patterns to assigned allele frequency classes. The probability model can assign six classes: two parental classes, F1 hybrids, F2 hybrids, and backcrosses to either of the parental classes. Default allele frequency information, as provided by the software developer, was used. The analysis was run for 100,000 MCMC iterations, with a 25,000 burn-in period. The filtered, unlinked *STACKS* assembly, reduced to 94 individuals, had to be restricted to 300 loci, because of the software tool's computational restrictions. The first 300 loci of the filtered assembly were used to ensure reproducibility of results. The calculated posterior probabilities for the assignment of parental and hybrid classes were visualised with *Microsoft Office Excel.*

Multivariate analysis

Statistical multivariate analyses are powerful and fast approaches to investigate the grouping of individuals within genetic datasets. They are practical to use when exploring high-dimensional data because they reduce its dimensionality. This allows for good graphical visualisation and interpretation. One of the most commonly used multivariate methods associated with genetic data is Principal Coordinate Analysis (PCoA) (Gower, 1966). PCoA can be applied to qualitative or discrete variables, represented in a distance matrix. From this matrix, the dimensions are reduced and represented with Principal Coordinates (PCos), which can be plotted in pairs on simple, two-dimensional plots. PCos are ordered by the proportion of inertia (dissimilarity) of the data they describe, and therefore usually the first two PCos, that explain the most inertia, are graphically represented and explored. PCoA analysis was calculated with *R* package *adegenet* (v.2.0.1) (Jombart, 2008), an extension of the *ade4* (v.1.7.18) package (Dray & Dufour, 2007). *R* package *ggplot2* was used to visualise the plots of the first two PCos. Filtered *STACKS* assembly was investigated based on the grouping of individuals for sampling localities. Additionally, the groups were also defined and analysed based on the results of the *NewHyrbids* analysis (parents, hybrids or, backcrosses).

F-statistics

The divergence of populations can be measured via statistical evaluation of genetic variance. The F_{ST}*-value* gives the variance among subpopulations relative to the total variance. Pairwise calculations of F_{ST}*-value* between two populations can be interpreted as levels of genetic divergence: When nearing zero, two populations show no convergence, and when nearing one, maximum divergence. Calculation and evaluation of F_{ST}*-value* are used to infer the population structure of molecular datasets. Several different methods and equations are used to calculate F_{ST}*-value.* Here, *R* package *genepop* (v. 1.1.7) (Raymond, 1995) was used on the unlinked filtered *STACKS* dataset. In this package, the F_{ST} *– value* is calculated using the following formula:

$$F_{ST} = \frac{Q_2 - Q_3}{1 - Q_3}$$

Where Q is the probability of identity in state; Q_2 among genes in different individuals within defined populations and Q_3 among defined populations. F_{ST} *-value* calculation was performed on unlinked, filtered *STACKS* assembly, where all individuals of *S. foetida* and *S. waldsteiniana* outside contact zones were assigned to two parental populations, and each population within

the contact zone was assigned to its own group to evaluate pairwise divergence between each parental species, as well other populations of the contact zone.

Isolation by distance

The term Isolation By Distance (IBD) was first used by Wright (1938, 1940) to describe the observed pattern of genetic variation resulting from spatial limitations of gene flow. As the geographic distance increases, the genetic similarity of individuals and populations tends to decrease. When individuals are randomly distributed in space, this change should follow a linear regression. In a hybrid swarm, individuals of the same and nearby populations would show a versatile array of genetic distances, disrupting the assumptions of the IBD linear model. Modelling of IBD was done by performing Mantel test statistics (Mantel, 1967) with 1000 permutations. This approach statistically tests the correlation between two matrices, the first one, in this case, being the Euclidean distance matrix of the genetic data, based on calculated F_{ST}*-values,* and the second one being a geographic distance matrix of pairwise distances calculated from geographic coordinates of the sampled individuals' localities. IBD was calculated for the filtered *STACKS* dataset, using the *adegenet* package in *R*, and visualised with the *R* base *plot* function.

2.4. Morphometrics

Morphometric measurements

Morphometric measurements were applied to the 95 individuals with RAD-sequenced data. In an attempt to represent an individual's average leaf, up to 10 mature leaves per herbarium specimen were selected. Each individual leaf was positioned on a blank paper sheet with a 10 x 10 mm black square as a scale, and scanned with a Canon CanoScan LiDE 220 (Canon, Tokyo, Japan) scanner, producing a high-quality image. If possible, leaves were not removed from specimens to prevent the destruction of herbarium vouchers. The number of teeth on the perimeter of each leaf was counted manually. Parameters length, width, widest point (distance from the base of the leaf to the widest part), perimeter, and area were measured on the scanned images with *Digimizer* (v.5.7.2) image analysis software (MedCalc Software Ltd, Ostend, Belgium). The ratios between length and width, as well as length and widest point, were calculated to quantitatively represent the shape of each leaf. A higher length-to-width ratio describes a narrower, lanceolate leaf and a lower ratio describes a wider, elliptic leaf. A higher length-to-widest point ratio describes an ovate or lanceolate (in combination with a high length-to-width ratio) leaf shape and a lower ratio describe an obovate leaf shape.

In total, it was possible to obtain 781 leaves of sufficient quality for morphometric analysis, adding up to approximately 8 leaves per individual. A table of all measured traits averaged per individual, can be seen in Appendix 7.7.

Morphometric analysis

A multivariate analysis approach was used to investigate several morphometric variables in one analysis. Prior to that, data were explored for possible correlation among measured variables, as well as external factors that might bias the analysis. All measured leaf traits, as well as altitude, were taken into consideration. An increase in altitude is associated with higher solar radiation (Blumthaler, 2012) and lower air pressure, air density, and temperature (Körner, 1999). These factors might affect leaf physiology and anatomy (Körner, 1989, Milla & Reich,

2011). Pearson's correlation coefficients of all the measured morphometric variables and altitude were calculated using the *R* base function (Appendix 7.8). Based on the correlation inspection, four independent metrics (leaf area, no of teeth per cm, length-to-width ratio, length-to-widest point ratio) were selected for multivariate analysis.

Principal Component Analysis (PCA) (Hotelling, 1933) is used to calculate the variance of multidimensional quantitative variables using a covariance matrix and reducing the dimensions with principal components (PCs). To calculate PCA, the base *R* function *prcomp* was used and the first two PCs were visualised with *ggbiplot* (v.0.55) package (Vu, 2011). With a Discriminant Analysis of Principal Components (DAPC) (Jombart et al, 2010), the grouping of individuals was evaluated on calculated PCs with an estimation of Posterior Group Memberships Probabilities (PGMPs). DAPC, implemented in the package *adegenet*, was designed to identify and evaluate genetic clusters in molecular markers datasets. However, it can be applied to any quantitative data (Jombart & Collins, 2015). So far, it has been successfully used to analyse datasets of morphometric traits (e.g. Gotzek et al., 2012, Ferreyra et al., 2013, Ruiz et al., 2021). DAPC was performed for two datasets. First, only 20 individuals outside the contact zone were grouped to either *S. foetida* or *S. waldsteiniana*, according to phenotypic identification. They were analysed to investigate the differences in morphometric traits of true parental phenotypes. Then, all 95 individuals were used. They were grouped based on phenotype identification in the field (*S. foetida, S. waldsteiniana,* or hybrids). Finally, an analysis based on grouping inferred from the results of the *NewHybrids* analysis based on 87 individuals was performed to identify differences between parents and identified genetically admixed individuals of different hybrid classes.

3. Results

3.1. Bioinformatic analysis

Phylogenetic analysis

The ML phylogenetic tree showed a topology that was generally well supported with high bootstrap values (Figure 6). All locally sampled populations, except one, formed well-supported groups (Figure 7). The exception was a contact zone population IT8, identified as *S. waldsteiniana,* which was split into two clades. *S. foetida* and *S. waldsteiniana* populations sampled outside the contact zone formed the two most divergent monophyletic clades. Contact zone populations formed paraphyletic groups positioned intermediate to both parental clades. Individuals from population IT5 showed very little divergence among each other with generally low bootstrap supports. Six individuals from population IT7 formed a well-supported group within the population and five of those showed little divergence with low bootstrap support. In other populations, only a few (or no) pairs of individuals showed such low divergence patterns.

Quartet sampling analysis revealed that despite high bootstrap supports, some topologies showed discordant topologies (Figure 7). All (or close to all) replicates were informative across the whole tree. No conflicting topologies were found for the branch separating the populations into two large clades (1/NA/1). Both clades of population IT8 (-0.0077/0.96/0.98 and (-0.42/0.55/0.99), population IT7 (-0.027/0.74/0.98), and population AU10 (-0.029/0.77/0.99) showed negative *QC* scores and *QD* between 0.3 and 1. This indicates that discordant topologies were represented more often than the initial concordant topology of the ML tree. Population

IT6 (0.021/0.93/0.99) and the clade of neighbouring populations IT7 and IT9 (0.07/0.56/1) showed *QC* values near 0 with *QD* scores between 0.3 and 1. This indicates an equivocal frequency of concordant and discordant topologies. The remaining branches showed positive *QC* scores, indicating the prevalence of concordant topologies.

The NN phylogenetic network showed a topology, corresponsive to the ML tree (Figure 8). All individuals of the parental species outside the putative hybrid zone were positioned closely together and on the opposite sites of the network. Populations from the contact zone were situated between them, following the same pattern as that observed in the ML tree. The three single individuals from population AU10 had visibly longer branches than the rest of the examined individuals.

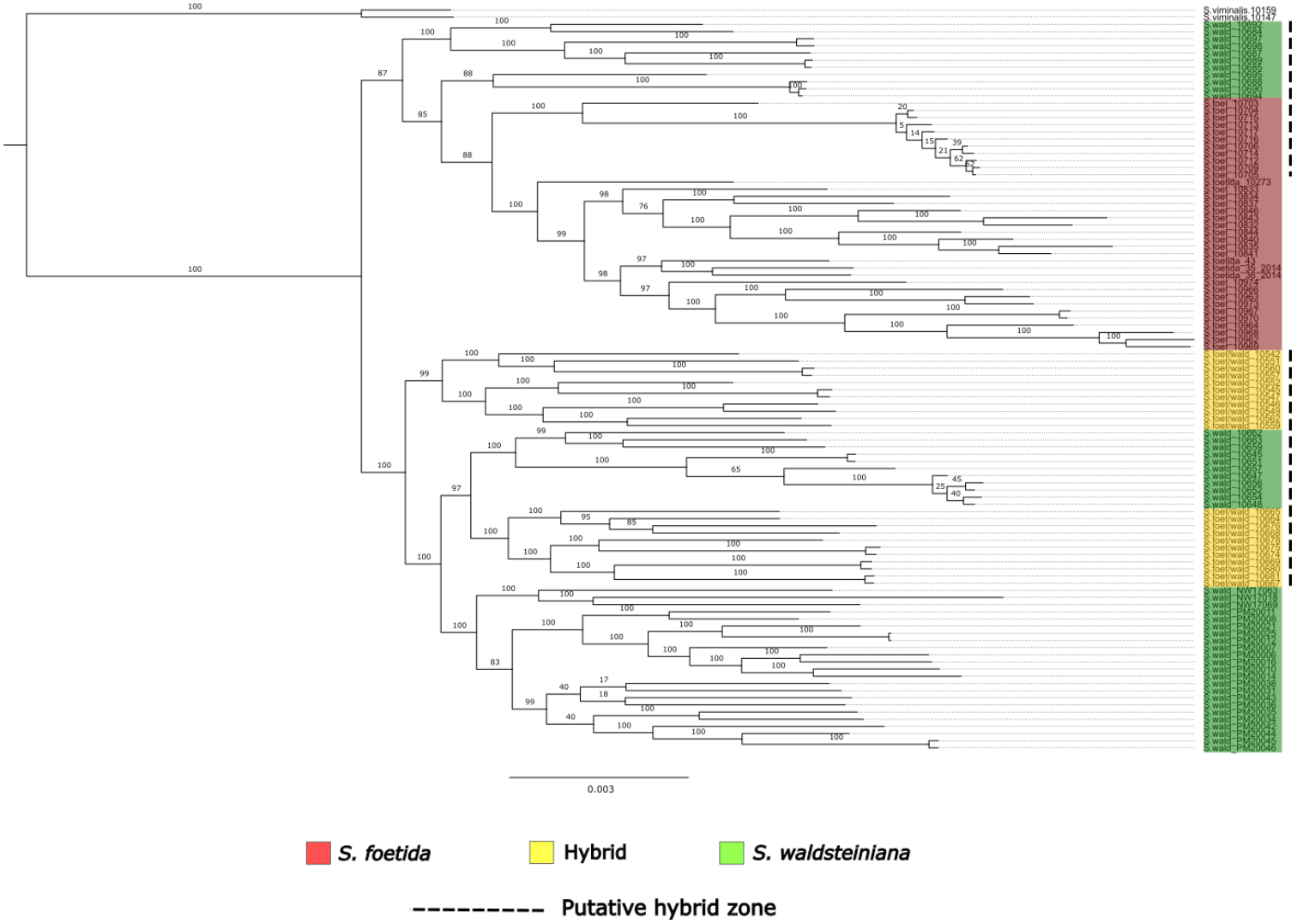

Figure 6: Maximum likelihood tree with 104 individuals, including the two outgroup *S.viminalis* individuals to root the tree. Colours indicate phenotype as described in the legend. Dotted lines represent populations collected in the putative hybrid zone (contact zone). The numbers displayed above branches are calculated bootstrap support values of 100 bootstrap replicates.

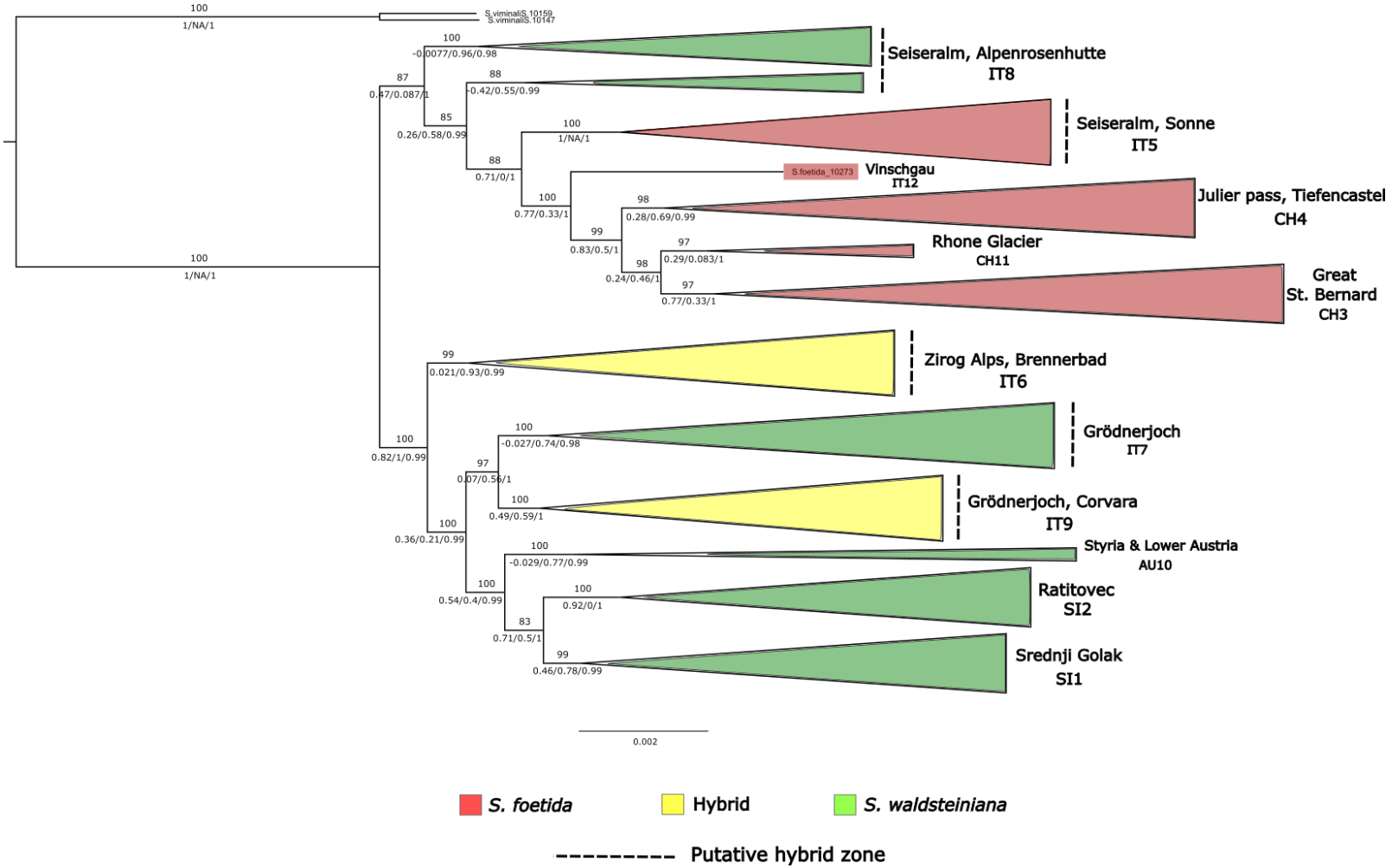

Figure 7: Maximum likelihood tree collapsed into populations with calculated *QS* scores The numbers displayed above branches are calculated bootstrap support values of 100 bootstrap replicates and the numbers displayed below branches are the calculated *QS* scores. Individuals are collapsed into populations based on sampling locality and labelled by their respective location and identifier, as presented in Table 1. Colours indicate phenotype as described in the legend. Dotted lines represent populations collected in the putative hybrid zone (contact zone).

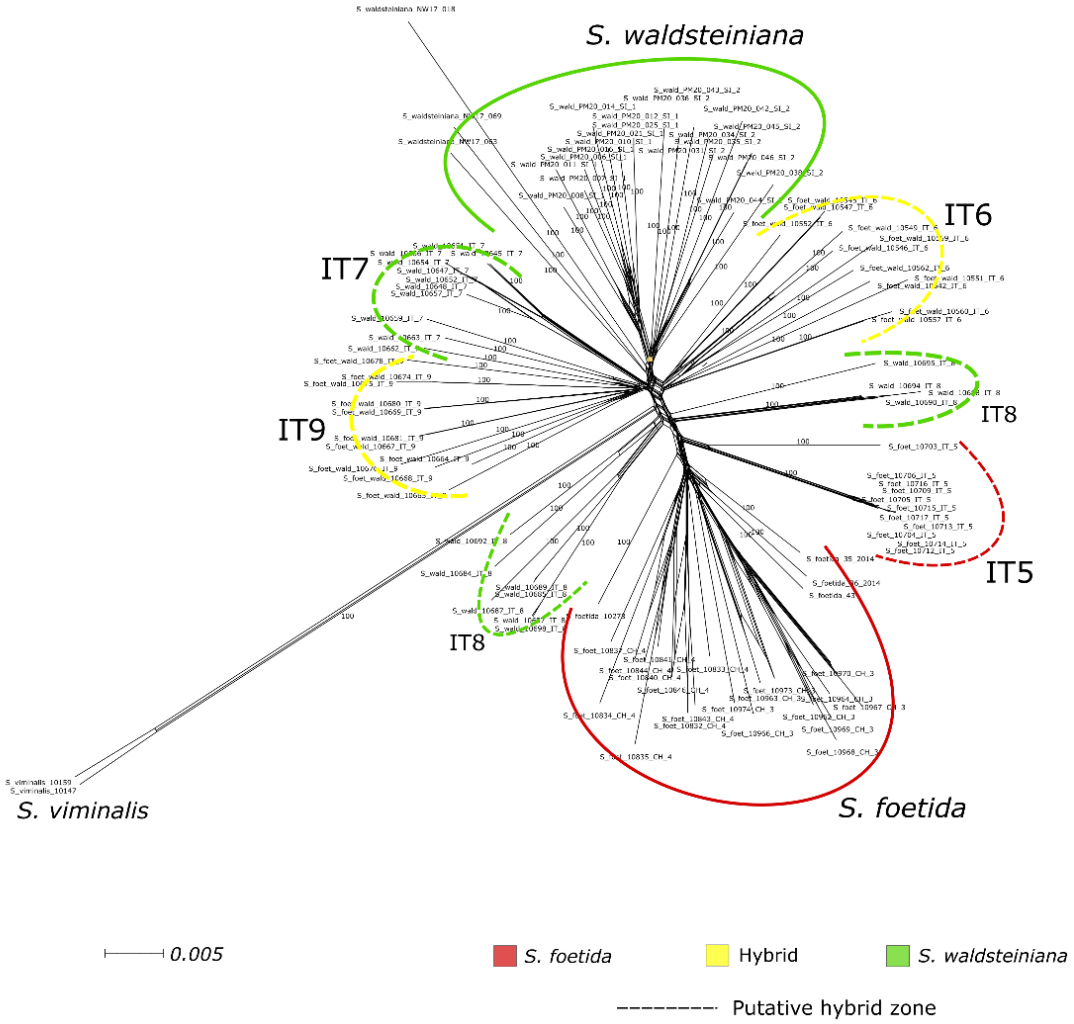

Figure 8: Neighbour-Net calculated phylogenetic network of 104 individuals, including the two outgroup *S. viminalis* individuals with displayed bootstrap values for longest branches. The scale for branch length is indicated on the left side below. Individuals are coloured by populations and marked with corresponding population identifiers (Table 1). Colours indicate phenotype as described in the legend. Dotted lines represent populations collected in the putative hybrid zone (contact zone).

Detection of reticulate evolution

Significant hybridization between the parental *S. foetida* and *S. waldsteiniana* populations was detected for all 55 individuals of the putative hybrid zone (Figure 9). Population IT5, identified as *S. foetida*, showed the highest portions of the *S. foetida* genetic patterns (γ-*values* 0.62 - 0.72. Population IT8, identified as *S. waldsteiniana*, also showed high portions of the *S. foetida* genetic patterns (γ-*values* 0.51 – 0.69), however several individuals showed a nearly equal admixture of both parental patterns. Population IT7, identified as *S. waldsteiniana*, displays the most equally admixed patterns of both parents (γ-*values* 0.37 – 0.52). Populations IT6 and IT9, both identified as hybrids, showed the highest variance (γ-*values* 0.24 – 0.52 and 0.29 – 0.48, respectively), with admixture patterns slightly shifted towards *S. waldsteiniana*.

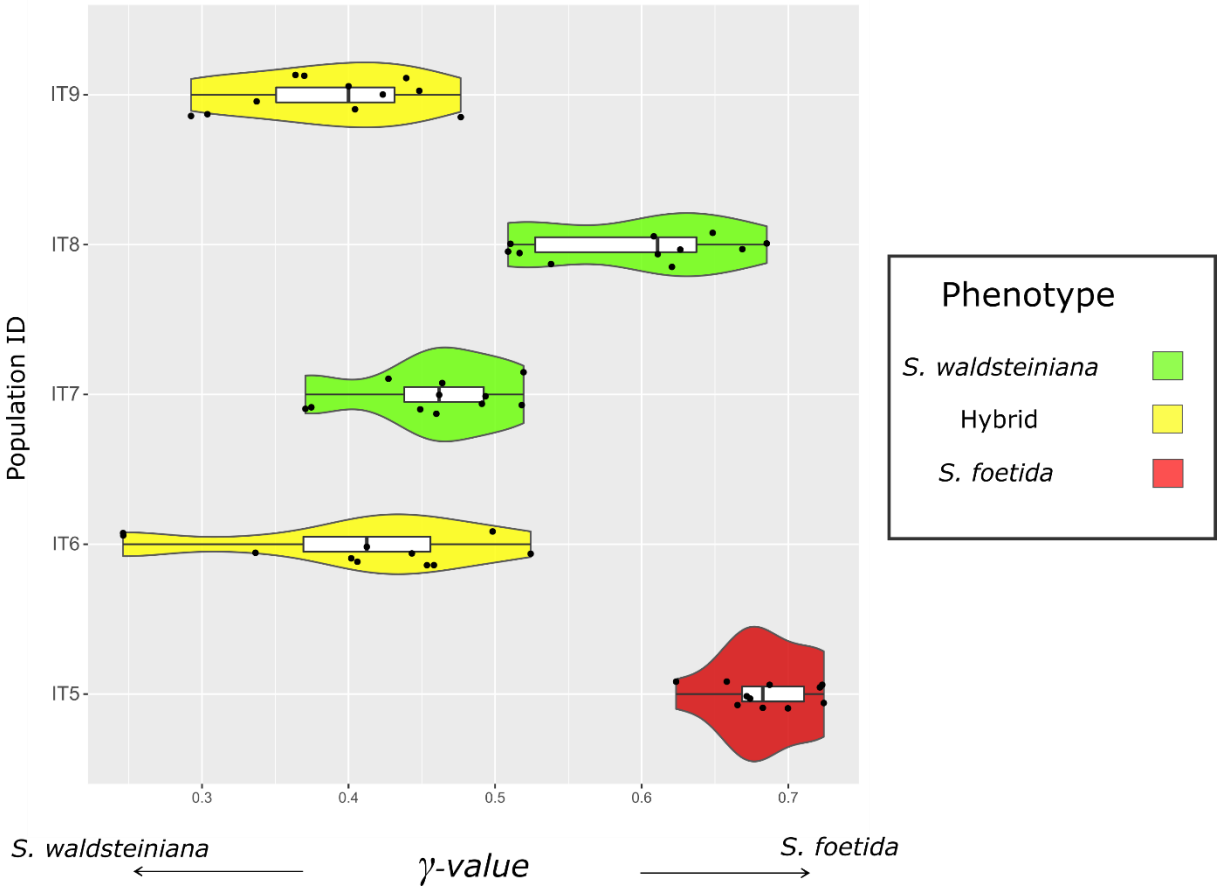

Figure 9: Calculated *γ-values* for each of the 55 individuals of the five populations from the contact zone, represented with boxplots. Higher values indicate higher similarity to *S. foetida*, and lower values higher similarity to *S. waldsteinana* genetic patterns. Populations are labelled with corresponding population identifiers (Table 1). Colours represent observed phenotype, as described in the legend on the right.

Population structure with STRUCTURE

For the reduced dataset of 94 individuals, the model-based clustering approach implemented in *STRUCTURE* assumed that a model with $K = 4$ was best fitting with the highest ΔK, followed by $K = 2$ with a slightly lower ΔK (Appendix 7.5). All remaining values of K showed uniformly low ΔK. For $K = 2$, the model identified two genetic clusters that showed different levels of admixture in populations within the contact zone, whereas the populations outside the contact zone mostly exhibited genetic patterns of a single cluster (Figure 10). $K = 4$ revealed two additional clusters, one in the remaining individuals of IT5 and a subset of individuals of IT8, and the other one in the majority of individuals of IT7 (Figure 10).

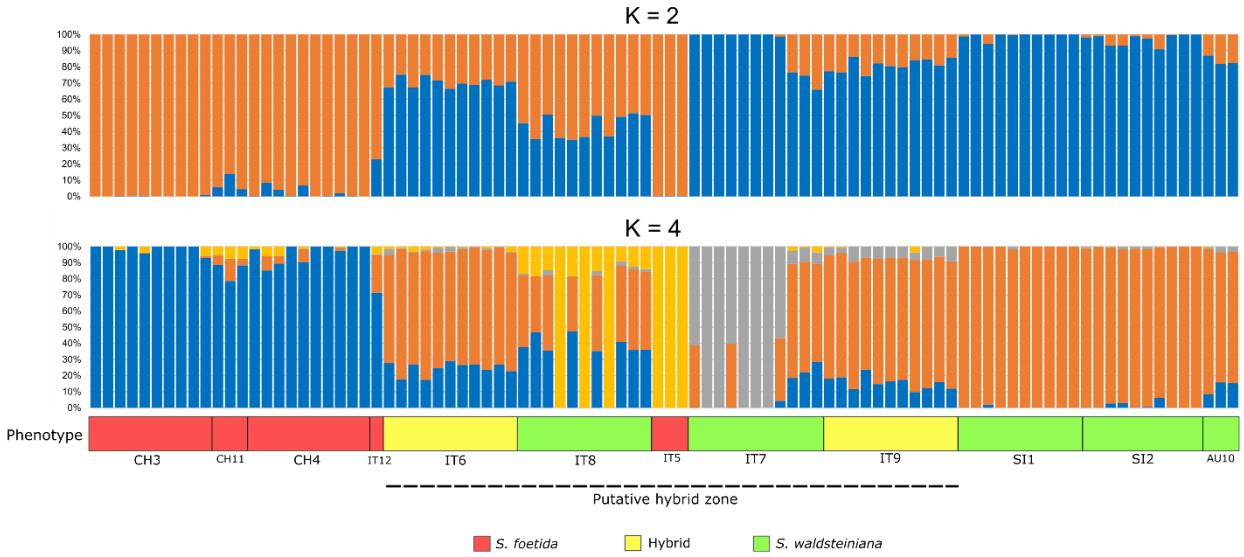

Figure 10: Results of the *STRUCTURE* analysis for the two *K* with highest Δ*K* values for the reduced dataset of 94 individuals**.** Results are represented with stacked bar plots. Different colours of the plots represent posterior probability for different genetic groups as calculated in the analysis. Populations are ordered by their geographical location from west to east, labelled by their identifiers (Table 1) and coloured by phenotype as described in the legend below. The dotted line marks populations of the putative hybrid zone (contact zone).

Population structure with sNMF

The plot of calculated entropy (Appendix 7.6) suggested that the plateau was reached somewhere around *K = 6.* Therefore, the range of *K* from 2 to 6 was taken into further investigation (Figure 11). For *K = 2,* the observed genetic patterns of admixture matched those calculated in *STRUCTURE,* however, both parental populations seemingly showed higher portions of the other species' genetic patterns. This is more pronounced in *S. waldsteiniana,* where individuals from populations SI1 and SI2 showed between 11.5% to 25.0% posterior probability of the *S. foetida* genetic patterns (mean = 17.8%). The three individuals collapsed into the AU10 population, showed even higher admixture, with posterior probabilites for *S. foetida* genetic cluster ranging between 34,5% to 43.6% (mean = 37.8%). Comparably, several *S. foetida* individuals from outside the contact zone showed no *S. waldsteiniana* genetic patterns, and the highest observed posterior probability was 25.1% (mean = 6.2%). This seemed to change with an increase in *K*. With higher values of *K,* internal structures of local populations were recognised as genetic patterns. Decreased variation of posterior probabilities of different genetic groups could be observed in *S. waldsteiniana* populations outside the contact zone when *K* was set between 3 and 6 (the exception was population AU10). On the contrary, the posterior probabilities of other genetic groups increased in *S. foetida* individuals outside the contact zone with an increase of *K.*

Matching admixture patterns with slight differences in posterior probabilities of the two genetic clusters could be observed for *K = 2* of the analysis ran on the unlinked 46,695 SNPs from the *ipyrad* assembly (Figure 12).

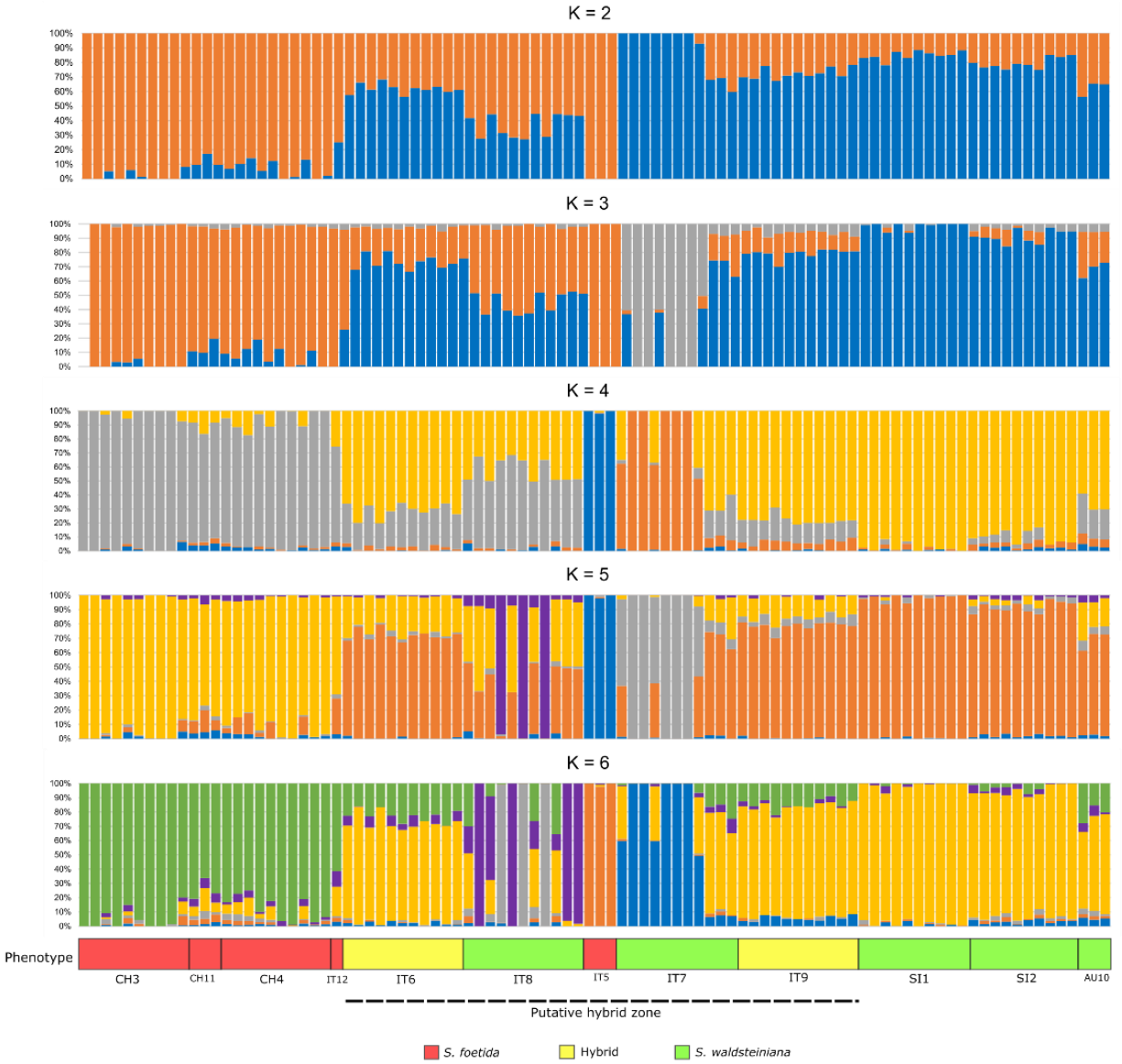

Figure 11: results of the *sMNF* analysis for *K* between 2 and 6 for the reduced dataset of 94 individuals**.** Results are represented with stacked bar plots. Different colours of the plots represent posterior probabilities for different genetic groups as calculated in the analysis. Populations are ordered by their geographical location from west to east, labelled by their identifiers (Table 1) and coloured by phenotype as described in the legend below. The dotted line marks populations from the putative hybrid zone (contact zone).

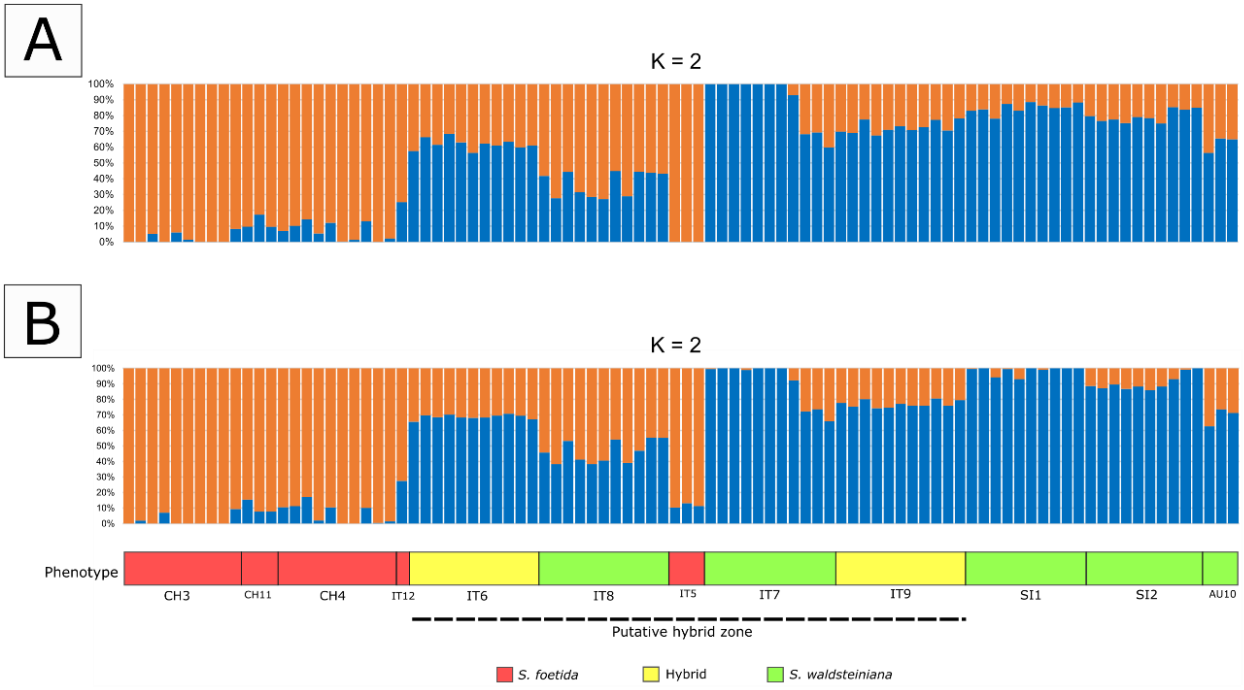

Figure 12: Side by side comparison of the results of *sNMF* analysis of 94 individuals for *K = 2* for **A) filtered *STACKS* assembly of 6,173 unlinked SNPs** and **B) *ipyrad* assembly of 46,695 unlinked SNPs.** Results are represented with stacked bar plots. Different colours of the plots represent posterior probabilies for different genetic groups as calculated in the analysis. Populations are ordered by their geographical location from west to east, labelled by their identifiers (Table 1) and coloured by phenotype as described in the legend below. The dotted line marks populations from the putative hybrid zone (contact zone).

Population structure with RADpainter and fineRADstructure

The analysis of the ancestry covariance matrix with *RADpainter* and *fineRADstructure* (Figure 13) showed high coancestry coefficients between all individuals of the *S. foetida* and *S. waldsteiniana* parental populations (dark orange colours in Figure 13). The only exceptions were two *S. waldsteiniana* individuals from population AU10 that were nested in one of the paraphyletic clades of the contact zone populations. For some populations, clear internal structuring with high coancestry coefficients (red colours in Figure 13) could be observed. Populations IT5 and IT7 showed high portions of individuals with extremely high coancestry coefficients (blue and black colours in Figure 13), while in other groups, single pairs of individuals exhibited such patterns. Similarly to the ML phylogenetic analysis (Figures 6 & 7), two other larger structural groups with higher coancestry coefficients could be observed, one including all parental *S. foetida* individuals, and populations IT8 and IT5 from the contact zone, and the second group including all parental *S. waldsteiniana* individuals, and populations IT6, IT9, and IT7 from the contact zone.

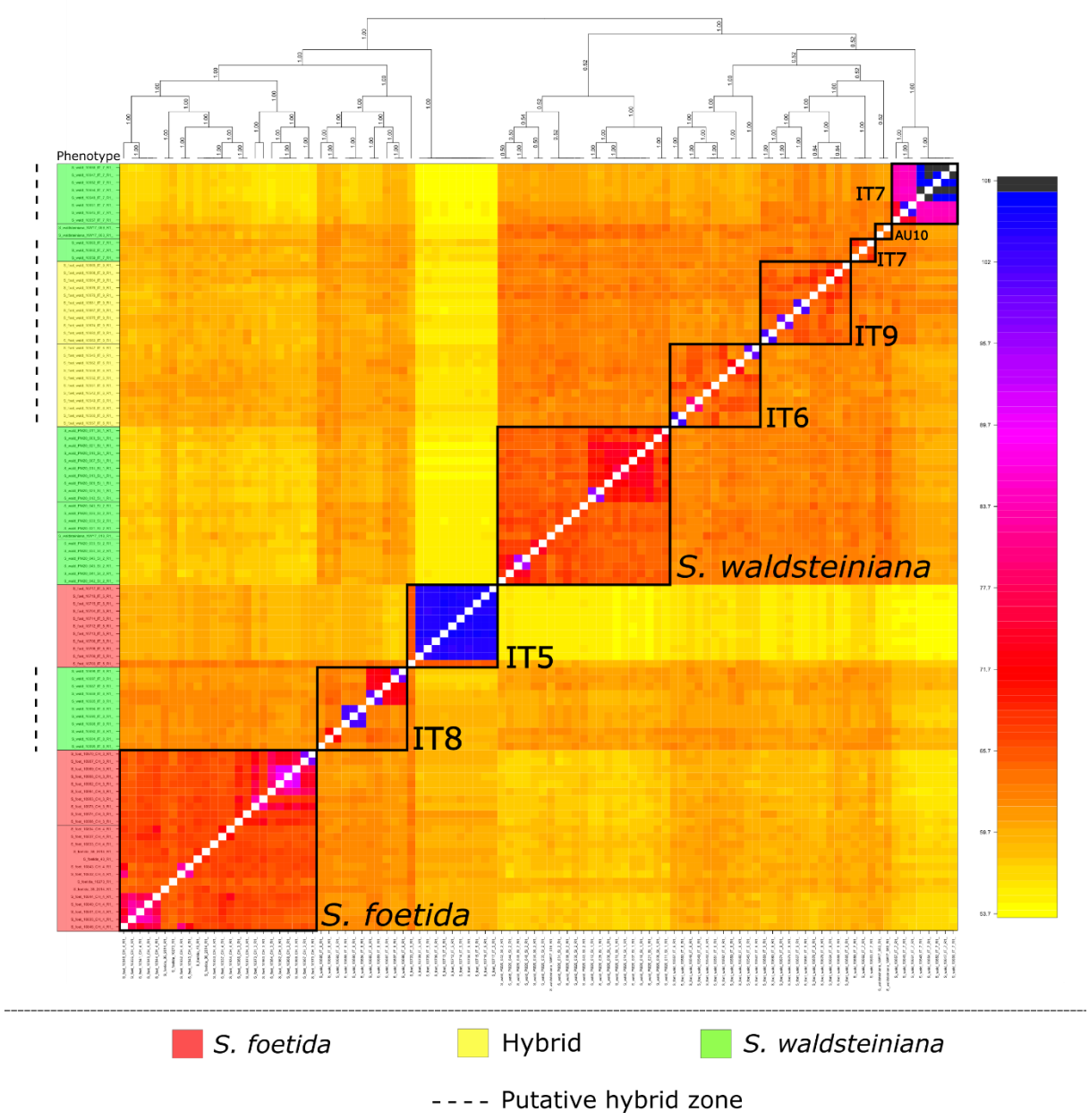

Figure 13: Results of *RADpainter* and *fineRADstructure* analysis represented in a heatmap plot of the calculated coancestry coefficient matrix. The scale is on the right (black, highest coancestry, yellow, lowest coancestry). On left, individuals are coloured by their identified phenotype as explained in the legend below. The dotted line marks populations located in the putative hybrid zone. Black squares on the coancestry matrix mark samples from the same population and are labelled accordingly (Table 1). The populations of *S. waldsteiniana* (with an exception of two samples from population AU10) and *S. foetida* from outside contact zones are collapsed into two large populations and labelled accordingly.

Assignment to parental and hybrid classes

The parental *S. foetida* individuals from sampling localities outside the putative hybrid zone were assigned as the first parent (*S. foetida*) with nearly maximum posterior probabilities (0.99 – 1.0). Only two individuals deviated from the assignment with slightly lower posterior probabilities for the first parent (*S. foetida*) class (0.86 and 0.85). Both were from populations of added single individuals (see Table 1), one from CH11 and the second one from IT12. *S. waldsteiniana* individuals sampled outside contact zones were classified as the second parent with nearly maximum posterior probabilities as well (0.89 – 1.0). Two of the additional single *S. waldsteiniana* individuals from AU10 were identified as F2 hybrids with high posterior probabilities (0.82 and 0.90), indicating great deviations from identified parental allelic frequencies. The three remaining individuals of the IT5 populations were classified as the first parent (*S. foetida)* with maximum posterior probabilities, which is in concordance with their phenotypic identification. Eight of the individuals of the IT7 population were classified as the second parent (*S. waldsteinana)*, concordant to phenotypic identification, with maximum posterior probabilities (0.98 – 1.0). The remaining three individuals from this population were classified as backcrosses to *S. waldsteinana* with high posterior probabilities (0.88 – 0.96). All individuals of the population IT8 that were identified as *S. waldsteiniana*, showed admixed allele frequencies and were classified as F2 hybrids with maximum posterior probabilities, except for one individual, which exhibited an admixture of *S. waldsteiniana* backcross (0.41) and F2 hybrid frequencies (0.59). Populations IT6 and IT9, identified as hybrids, showed heterogeneous assignment patterns between the individuals as well as within them. In population IT6, four individuals could be classified as F2 hybrids (0.75 - 1.0), five as backcrosses to *S. waldsteiniana* (0.45 – 0.91), and two as the second parent (*S. waldsteiniana*) (0.95, 0.97). In population IT9, three individuals could be classified as F2 hybrids (0.96 - 1.0), five as the second parent (*S. waldsteiniana*) (0.93 – 1.0). Three individuals were classified as *S. waldsteiniana* backcrosses, even if the posterior probabilites for this class were not prevalent (0.42 – 0.62), but still indicated a relatively high probability of admixture, associated with later generation backcrossing. No individuals could be classified as F1 hybrids or *S. foetida* backcrosses.

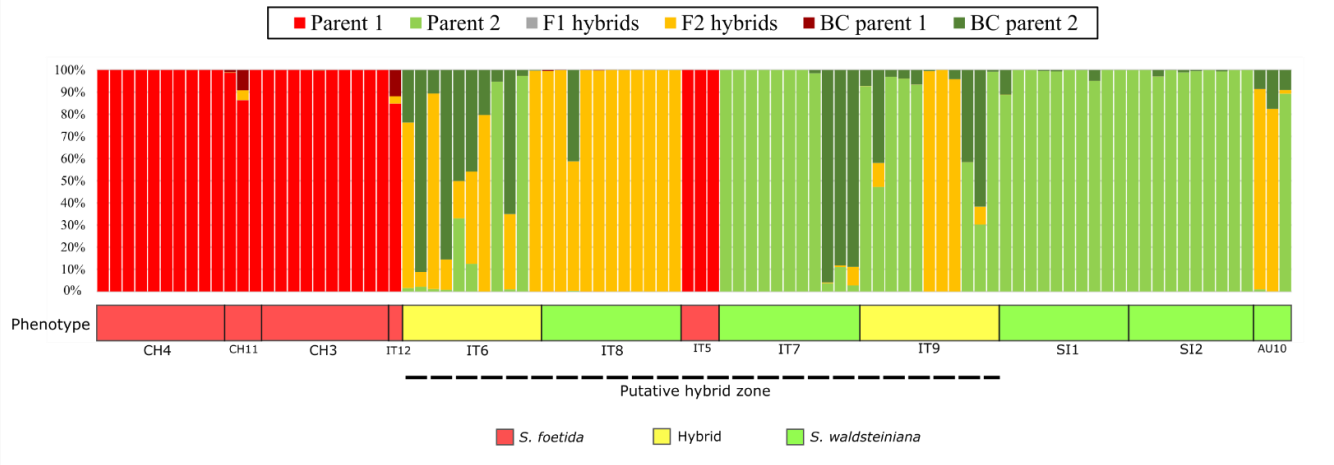

Figure 14: Results of the NewHybrids analysis represented in a stacked bar plot. Populations are geographically sorted from west to east and marked with corresponding identifiers (Table 1). The bar below the figure indicates identified phenotype as described in the legend below. The dotted line marks populations located in the putative hybrid zone (contact zone).

Multivariate analysis

PCoA was performed and reduced to the first three principle coordinates. For the analysis of all individuals grouped by sampling localities, PCo1 explained 10.48% of the total inertia and PCo2 6.34 % (Figure 15). Population IT5 stood out as an isolated group, while all other populations seemed to follow a pattern of graduate change from one parental species towards the other. Parental *S. waldsteiniana* populations SI1 and SI2, and parental *S. foetida* populations CH3, CH4, CH11, and IT12 were positioned on either end of the distribution of both two PCos. Population IT7 from the hybrid zone, identified as *S. waldsteiniana*, showed a wide dispersal of PCos, indicating a genetically separate subset of individuals. Population AU10, a supposed parental *S. waldsteiniana* population, was nested between populations IT9 and IT6 from the hybrid zone, which were identified as hybrids. Population IT8, identified as *S. waldsteiniana*, was positioned intermediately to both parental clusters, and showed a relatively large dispersal of calculated PCos, hinting at a high variability among the individuals.

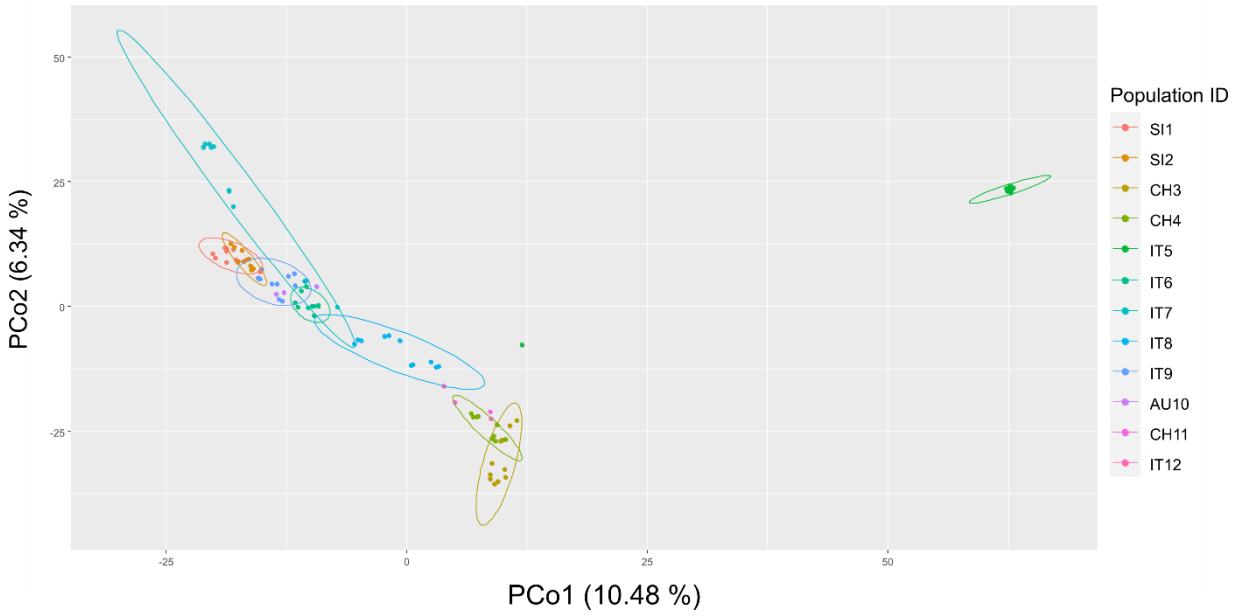

Figure 15: Plot of the first two principal coordinates of the PCoA analysis performed for all 102 individuals. The groups were assigned based on population locality (Table 1).

When class assignment by *NewHybrids* was used as the grouping on the reduced dataset of 94 individuals, PCo1 explained 7.96%, and PCo2 5.45% of the total inertia (Figure 16). The patterns of groups assigned by identified admixture patterns, followed the gradient from one species to another. *S. waldsteiniana* backcrosses were positioned at the border of *S. waldsteiniana* cluster with a large overlap. F2 hybrids cluster spanned intermediately between *S. waldsteiniana* and *S. foetida* cluster, and overlapped with *S. waldsteiniana* parental cluster, whereas there was no overlap with the *S. foetida* cluster. This cluster also overlapped with the

complete *S. waldsteiniana* backcross cluster. The isolating effect of population IT5 was diminished through the reduction of individuals.

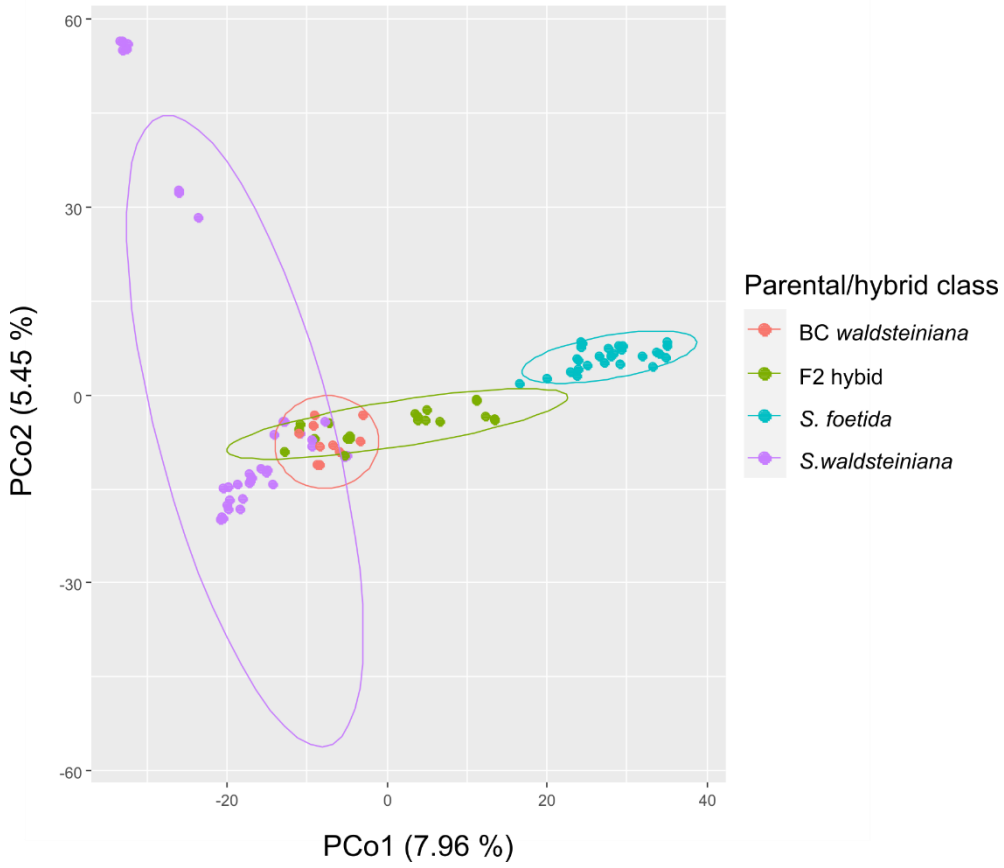

Figure 16: Plot of the first two principal coordinates of the PCoA analysis performed on 94 individuals (population IT5 reduced). Each individual was assigned a class, based on the results of the *NewHybrids* analysis (Figure 14).

F-statistics

The calculated F_{ST} values between tested populations ranged between 0.06 and 0.43 (Figure 17, Table 3). Parental *S. foetida* and *S. waldsteiniana* populations did not show the highest pairwise F_{ST}*-value* among all groups (0.16). The highest F_{ST}*-values* were observed between population IT5, identified as *S. foetida* (0.31 – 0.43), and all other populations. Of that, the lowest (0.31) was the F_{ST}*-value* with the parental *S. foetida* population and the highest (0.43) with population IT7, identified as *S. waldsteiniana*. Of all identified *S. waldsteiniana* populations, population IT7 was the most diverged to parental *S. foetida*, showing the highest F_{ST}*-value* (0.22), followed by the parental *S. waldsteiniana* population (0.16). Populations IT8, identified ad *S. waldsteiniana,* showed similar F_{ST}*-values* with both parental populations (*S. foetida* (0.10), *S. waldsteiniana* (0.11). Populations IT6, IT9, and IT7 had lower F_{ST}*-values* with parental *S. waldsteiniana* (0.060, 0.063, and 0.15, respectively), than with the parental *S. foetida* population (0.12, 0.15, and 0.22, respectively).

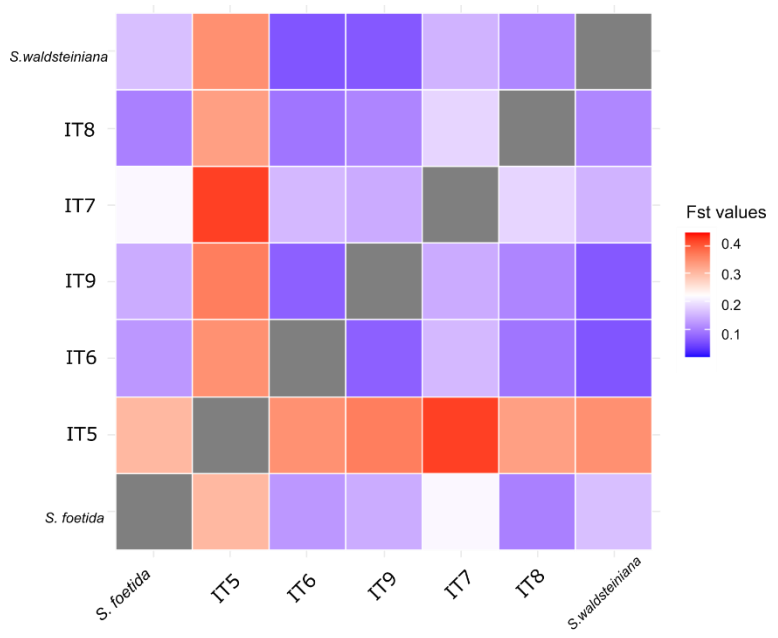

Figure 17: Heatmap plot of pairwise calculated F_{ST}-*values* of examined populations. All populations outside contact zone were collapsed into *S. waldsteiniana* (populations SI1, SI2 and AU10) or *S. foetida* (CH3, CH4, CH11 and CH12) population.

Table 3: Matrix of calculated pairwise F_{ST}-*values* between examined populations. Populations outside contact zone were collapsed into *S. waldsteiniana* (populations SI1, SI2 and AU10) or *S. foetida* (populations CH3, CH4, CH11 and IT12) population

Phenotype	**Population ID**	***S. foetida***	**IT5**	**IT6**	**IT9**	**IT7**	**IT8**	***S. waldsteiniana***
S. foetida	***S. foetida***	0	0.31	0.13	0.15	0.22	0.10	0.16
S. foetida	**IT5**	0.31	0	0.35	0.37	0.43	0.34	0.35
hybrid	**IT6**	0.13	0.35	0	0.07	0.16	0.09	0.06
hybrid	**IT9**	0.15	0.37	0.07	0	0.15	0.11	0.06
S. waldsteiniana	**IT7**	0.22	0.43	0.16	0.15	0	0.19	0.15
S. waldsteiniana	**IT8**	0.10	0.34	0.09	0.11	0.19	0	0.11
S. waldsteiniana	***S. waldsteiniana***	0.16	0.35	0.06	0.06	0.15	0.11	0

Isolation by distance

The calculated *p-value* for 1000 replicates, performed with the Mantel test, was 0.015, hinting at internalised structuring of different genetic distances in space. No significant correlation between genetic and geographic data of all individuals could be observed. The calculated linear model showed a slightly positive slope with a low goodness-of-fit-measure (R^2 = 0.025) indicating poor fit of the model (Figure 18). Individuals from the same localities (geographic distance = 0) displayed a wide range of calculated pairwise genetic distances. The majority was lower than 0.5, however, a noticeable portion was larger, spanning between 0.5 and 0.6, in some cases even exceeding the calculated genetic distance between the most geographically distant individuals. The observed extreme pairwise F_{ST}*-values* above 0.8 indicated a presence of a genetically very distant group of individuals, which can be assigned to population IT5, based on previous analyses.

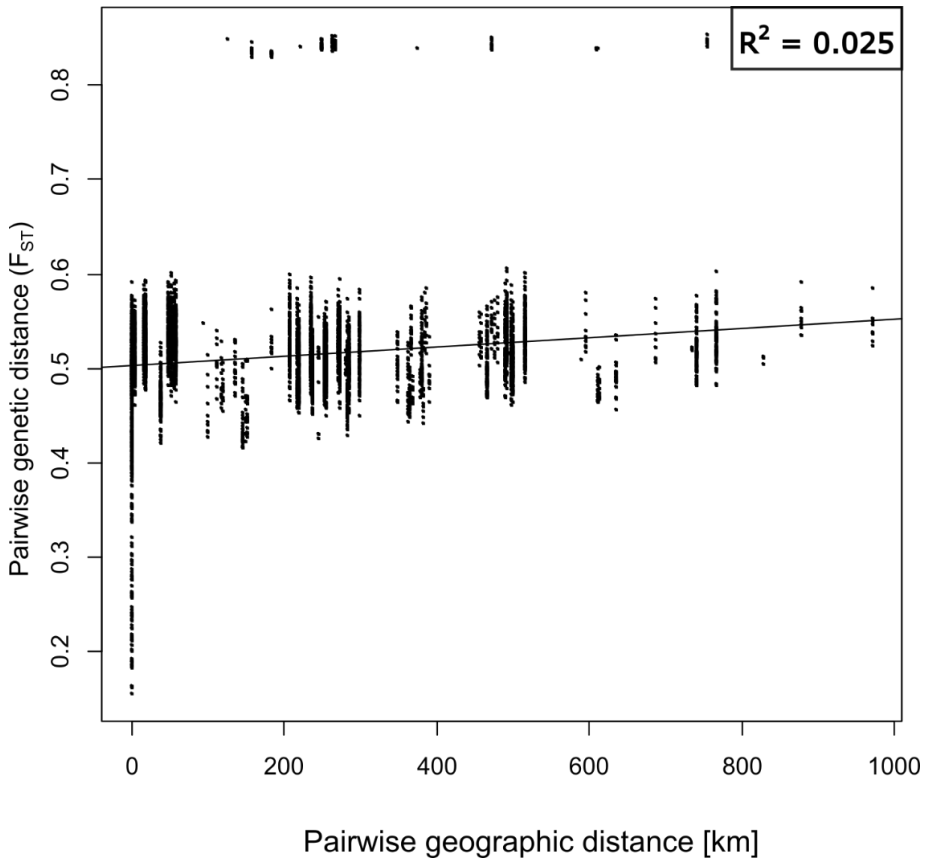

Figure 18: Isolation by distance model for all individuals used in the study. The x-axis represents pairwise geographic distance calculated in kilometres and the y-axis represents genetic distance calculated as the pairwise F_{ST}*-value* of all individuals. R^2 is the measure for goodness-of-fit of the calculated linear model displayed in the upper right corner.

3.2. Morphometric analysis

Multivariate PCA of morphometric traits, averaged by the individual, revealed, that the two groups of parental *S. foetida* and *S. waldsteiniana* individuals outside the contact zone showed separation on the axis of the first principal component (See Figure 19). PC1 describes 53.8% of the total variation and PC2 21.5%. PGMPs calculated in DAPC were 0.95 for *S. foetida*, 1.0 for *S. waldsteiniana*, and 0.975 for the correct assignment of a random individual.

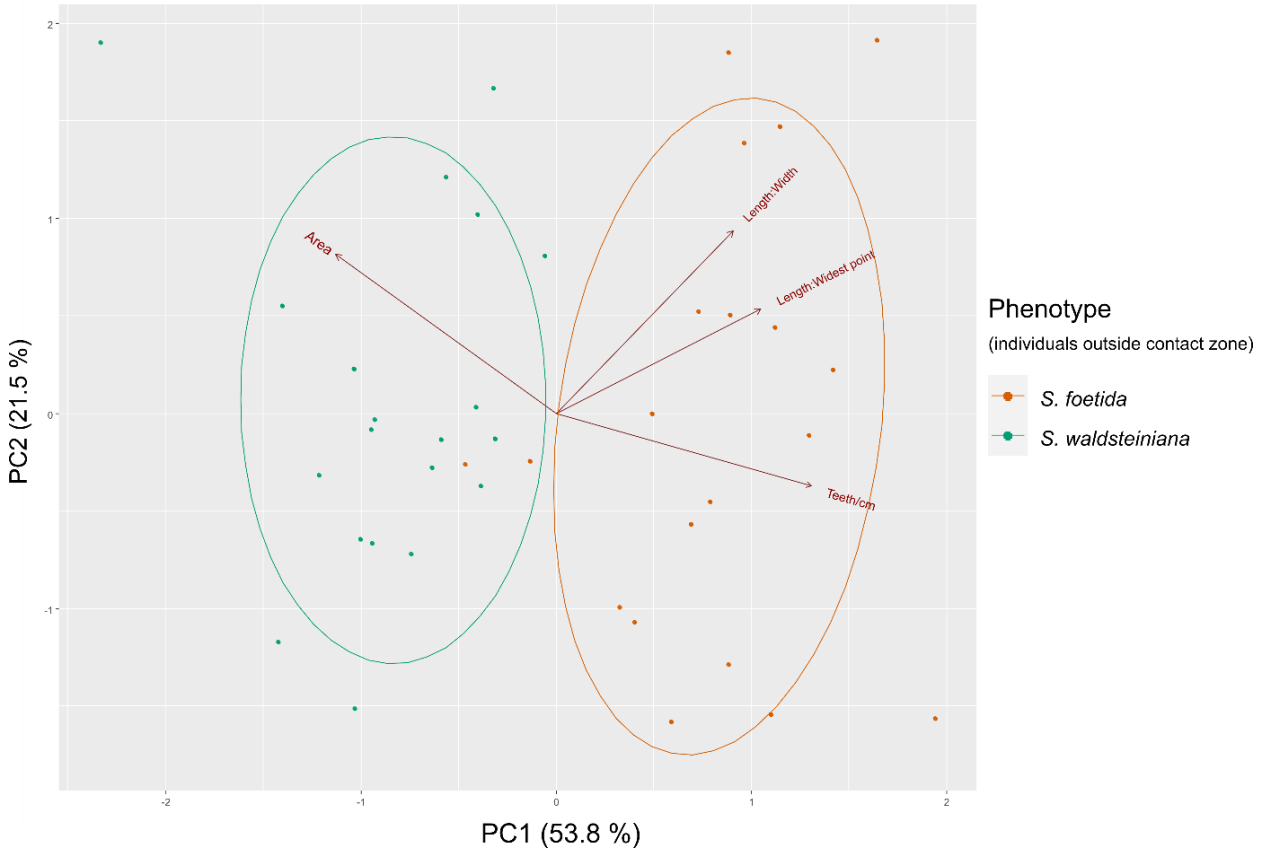

Figure 19: Plot of the first two principal components for the PCA of all purebred parental *S. foetida* (populations CH3 and CH4) and *S. waldsteiniana* (populations SI1 and SI2) individuals with 95% confidence ellipses.

When PCA was applied to all individuals, PC1 described 46,1% of the variation, followed by 25,9% described by PC2 (Figure 20). When the groups were assigned based on phenotypic identification, the two species showed an overlap, and the phenotypically intermediate hybrids formed an intermediate cluster that overlapped with both parental groups. Calculated PGMPs were 0.81 for *S. foetida*, 0.90 for *S. waldsteiniana*, 0.55 for hybrids, and 0.79 for the correct assignment of a random individual.

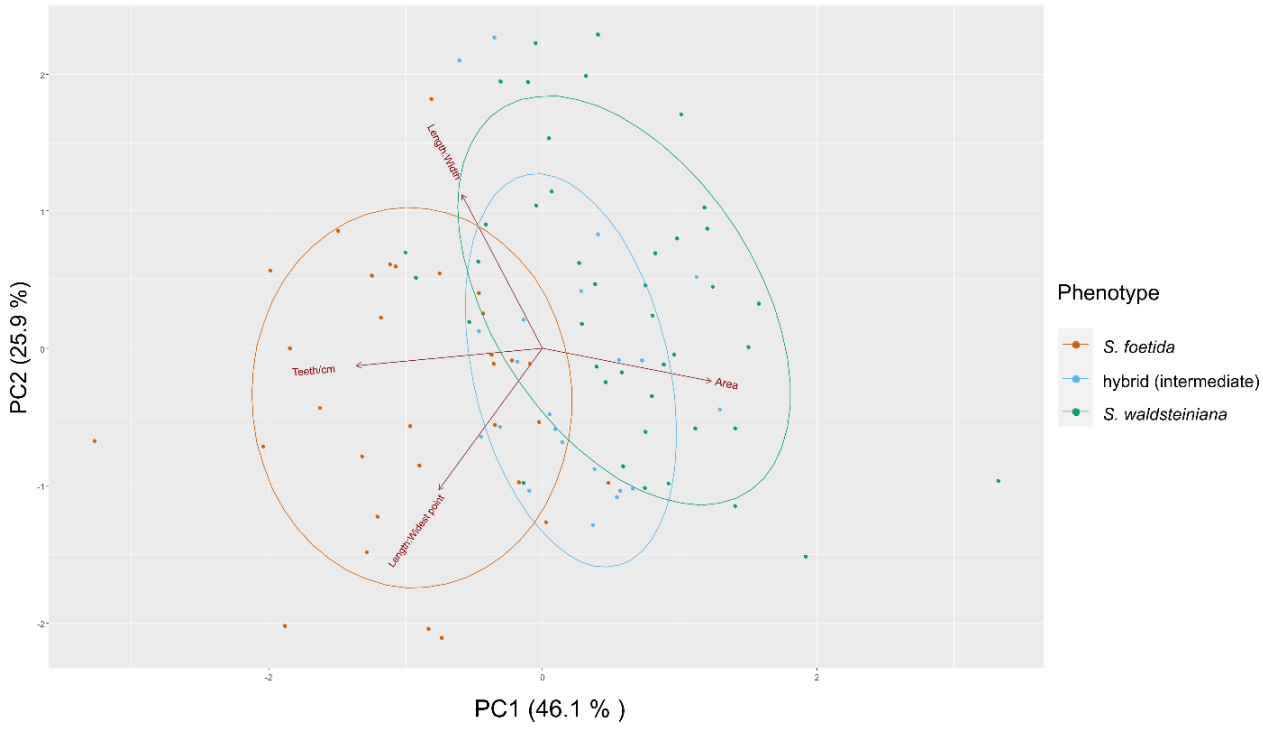

Figure 20: Plot of the first two principal components for the PCA of all 95 morphometrically measured individuals with 95% confidence ellipses. The groups are assigned by identification based on phenotype (*S. waldsteiniana*, *S. foetida*, hybrid (Table 1)).

When PCA was applied to 87 individuals (population IT5 reduced) that were assigned a parental/hybrid class in the *NewHybrids* analysis, calculated PC1 (46.5%) and PC2 (25.3%) were only slightly different for the reduced dataset (Figure 21). The two parental groups showed a lower overlap, and individuals classified as F2 hybrids or *S. waldsteiniana* backcrosses formed two strongly overlapping clusters, positioned intermediately between the parental clusters with a skew towards the *S. waldsteiniana* parental cluster. Calculated PGMPs were 0.78 for *S. foetida*, 0.92 for *S waldsteiniana*, 0.56 for F2 hybrids, and 0.70 for the correct assignment of a random individual. Not a single *S. waldsteiniana* backcross could be correctly assigned to the right group (PGMP = 0).

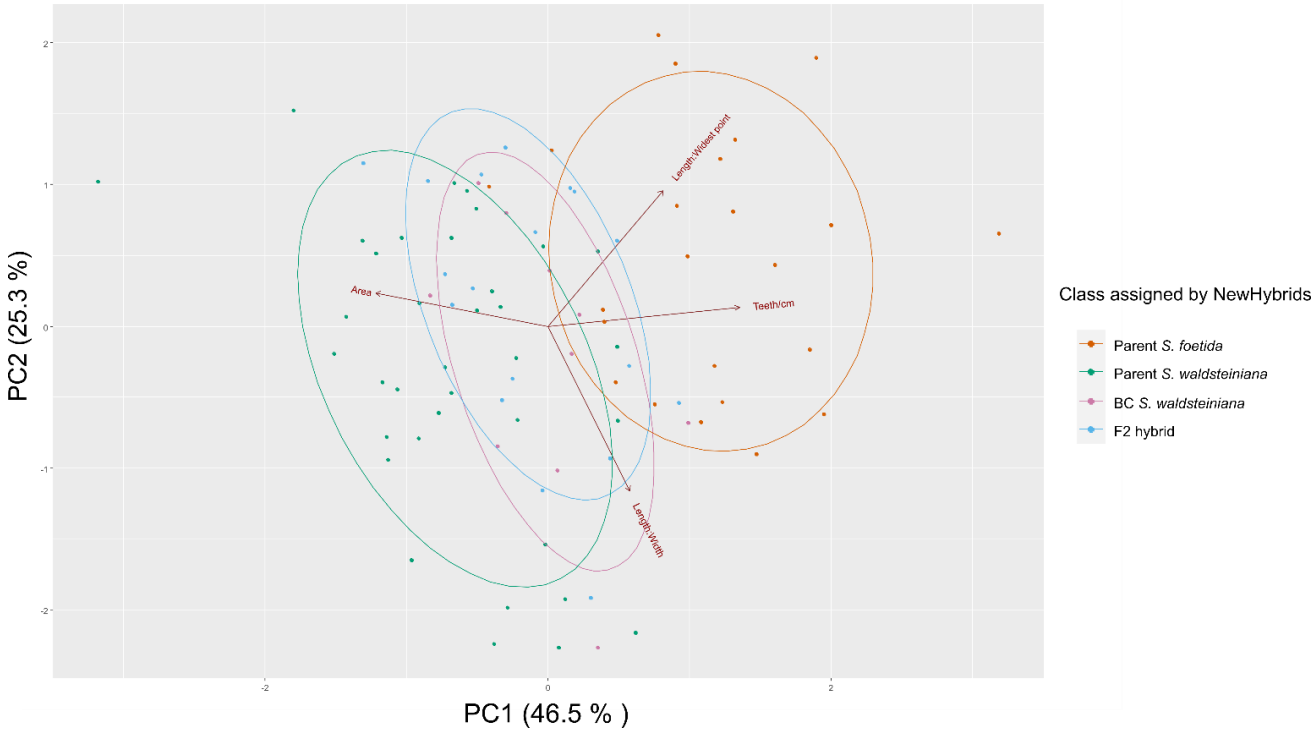

Figure 21: Plot of the first two principal components for the PCA of 87 morphometrically measured individuals (population IT5 reduced to three individuals) with 95% confidence ellipses. The groups were assigned based on the results of the *NewHybrids* analysis.

4. Discussion

4.1. Hybrid zone formation of *S. foetida* and *S. waldsteiniana*

With the use of RAD-seq technology, it was possible to detect signals for hybridization of *S. foetida* and *S. waldsteiniana* within their contact zone in the Alps.

The phylogenetic analysis and reconstruction of the ML tree and NN phylogenetic network revealed the division of the two species into clear monophyletic clusters, whereas populations sampled in the contact zone formed paraphyletic groups (Figures 6, 7 & 8). The presence of conflicting topologies, as a result of reticulate networks, was suggested by calculated *QS* scores for all clades of the contact zone populations (Figure 6). The discrepancy in phenotypic identification and phylogenetic topology could be observed. For example, population IT8, with identified *S. waldsteiniana* phenotype, showed greater genetic similarity with *S. foetida* populations than *S. waldsteiniana* populations in phylogenetic studies. An analysis of reticulate evolution with *HyDe* suggested that all populations from the contact zone resulted from admixture of both species (Figure 9), hinting at patterns of recent hybridization (in IT7, IT9, partially in IT8 and IT6) and introgression (in IT5, IT8, partially in IT6 and IT9). Calculated F_{ST}*-value* between all parental populations outside the contact zone was low, indicating little genetic divergence between the taxa, but was still higher than in pairwise comparisons of admixed populations IT6, IT9, and IT8, indicating admixture in these populations (Figure 17, Table 3). The general patterns observed in both phylogenetic inference and population genetics analysis indicate the presence of a broad hybrid zone of *S. foetida* and *S. waldsteinana* in the Alps, that is most likely linked to post glacial recolonization of the area, following the last glacial maximum (Hewitt, 2004, 2011). This is in accordance to several taxa that were found to

hybridize in the Alps (e.g. *Saxifraga* in Gugerli, 1997, *Euprhasia* in Liebst, 2008, *Tephroseris* in Pflugbeil et al., 2021).

Phylogenetic approaches provided a good overview of the reticulate relationships of the examined populations. However, the process of incomplete lineage sorting (ILS) may result in the same patterns and therefore cannot be excluded from the interpretation (Galtier & Daubin, 2008). The method implemented in *HyDe* tries to overcome ILS by testing for specific hybridization hypotheses (Elworth et al., 2019) and can be interpreted more confidently as signals of true hybridization. The phylogenetic invariant approach used in this software first models a coalescent tree along the tested parents, which accounts for ILS, and then tests for hybridization with *D*-statistics on coalescence independent sites (Kubatko & Chifman, 2019). Nevertheless, population genetics approaches are better suited to identify true hybridization patterns, and the results of these analyses showed a clear admixture of parental genomes in the contact zone. Still, it is important to note that the filtered, unlinked SNP dataset with 6,173 SNPs, which was utilised in population genetics, represents less than 1% of the complete, unfiltered *STACKS* SNP assembly. Phylogenetic tree reconstruction with *RAxML* and detection of hybridization with *HyDe*, were used on complete alignments of the nearly unfiltered RAD-loci assembly, and the NN network-building approach from *SplitsTree* was used on all aligned SNPs of this dataset. This was more than 47,000 loci, represented in an alignment of over 4.1 Mio nucleotides with more than 336,000 SNPs. However, in the case of RAD-loci that are assembled across the entire genome, more data does not necessarily mean results that are more accurate. The SNPs across alleles might be non-randomly associated through processes of linkage and selection, resulting in increased levels of linkage disequilibrium (Flint-Garcia & Thornsberry, 2003), which could affect the observed patterns. In the filtered and unlinked SNP assembly of *STACKS*, these effects were most likely reduced. The use of both datasets in this study acts reciprocally and adds weight to the interpretation.

4.2. Genetic structure of hybrid populations

The two sister species could be inferred to separate genetic clusters in the population genetics analyses, however, even individuals outside the contact zone showed some degree of admixture. Low F_{ST}-*value* between *S. waldsteiniana* and *S. foetida* populations outside the contact zone was an additional indicator of the low genetic divergence of both species. Intraspecific diversity of molecular characters may indicate historic hybridization and introgression of the two species, which has been documented in other *Salix* species (Hardig et al., 2000), as well as other plant genera (e.g. *Alnus glutinosa* in King & Ferris, 1998, *Quercus* in Petit et al., 1997) using cpDNA markers. Wagner et al. (2018) attributed the same processes to the observed topology in *Chamaetia/Vetrix* clade, after a phylogeny was inferred from the RAD-seq dataset. *S. foetida* and *S. waldsteiniana* are members of this clade. The whole section seems to have a shared genetic pool, resulting in low divergence across the whole clade, and hybridization is possible across different sections, not only within closely related species pairs (Wagner et al., 2018, Gramlich et al., 2018). In the *S. helvetica* and *S. purpurea* hybrid zone RAD-seq study, the calculated F_{ST}-*value* between the two species was contrastingly higher (F_{ST} -*value* = 0.53 for filtered loci in Gramlich et al., 2018), than for *S. foetida* and *S. waldsteinana* in this study (F_{ST}-*value* = 0.16 for filtered loci), and the two examined species were still capable of crossing. The genetic structure of populations showed that purebred *S. waldsteiniana* individuals contain higher portions of *S. foetida* genetic patterns. This proposes a hypothesis that gene flow between

these two species has historically been more directed towards *S. waldsteiniana.* The process might have been mitigated by a positive selection of ecologically favourable alleles, that were incorporated into the genome through admixture, as was observed in other plant genera (e.g. *Helianthus* in Rieseberg et al., 2003, *Iris* in Arnold et al., 2012). Indeed *S. waldsteiniana* seems to tolerate a wider palette of ecological niches. It can succeed in drier localities and is not strictly bound to carbonate bedrock soils, whereas *S. foetida* seems to inhabit a much narrower spectrum of habitats with sufficient moisture and strictly silicate bedrock substrate (Hörandl, 1992). It might be possible that adaptive alleles of *S. waldsteiniana,* allowing this species higher habitat flexibility, were obtained through ancient introgression.

Different degrees of genetic admixture in consistent patterns could be observed in populations of the contact zone throughout all population genetics analyses. Additionally, the *sNMF* analysis of the larger *ipyrad* assembly of unlinked SNPs showed a high congruence of results with the filtered *STACKS* assembly dataset (Figure 12). The observed gradient of divergence does not seem to follow a geographic continuity (Figure 18) but is characterised by locally established genetic patterns of admixture as a result of continuously ongoing hybridization. Thus, the results did not support the hypothesis by Hörandl (1992) that the genetically admixed populations solely consist of the F1 generation. Even though catkin deformations and lack of seeds in mature fruits hinted to lower fertility rates of intermediate forms (Hörandl, 1992), this study showed that the natural hybrids of *S. foetida* and *S. waldsteiniana* were capable of reproduction beyond the first hybrid generation. This is generally the case for naturally hybridizing plant taxa (Arnold, 1997, 2006, Mallet, 2005), and was frequently observed in *Salix* as well (Hardig et al., 2000, Fogelqvist et al., 2015, Gramlich et al., 2018). The state of admixture within local populations could be linked to the timing of post-colonisation succession of the local habitat patches, and possibly environmental factors. Populations IT6, IT9, and IT7 showed admixture with a dominance of the *S. waldsteiniana* genetic cluster and higher convergence with *S. waldsteiniana* than *S. foetida* parental population (Figures 10 & 12) In *NewHybrids,* these individuals exhibited allelic frequencies of the *S. waldsteiniana* parental class, backcrosses to *S. waldsteiniana* and F2 hybrids (Figure 14). The posterior probabilities for each class were variable, which further hints at a wide assembly of admixed genotypes present in the hybrid zone. In these populations, hybrids were shown to freely breed with *S. waldsteiniana,* but not *S. foetida*. In their study, Gramlich et al. (2018) observed a similar situation of more frequent backcrossing to *S. purpurea* compared to *S. helvetica,* and attributed it to a possibly greater flowering overlap of *S. purpurea* with the hybrids, pollen limitation of *S. helvetica* and stochastic factors, such as closer proximity of *S. purpurea* to the hybrid individuals in the sampling area. In the case of the studied species here, the flowering of *S. walsteiniana* and *S. foetida* occurs during the same time of the year (Hörandl, 1992), timed to fit within the very short alpine growing season (Körner, 1999), so the first interpretation seems unlikely. Because catkins of *S. waldsteiniana* are larger than catkins of *S. foetida,* they might produce more pollen, and pollen limitation of *S. foetida* might be a possible explanation for the lack of *S. foetida* backcrosses. Such instances were observed in *Morus* (Burgess et al., 2008) and possibly *Rhododendron* (Yan et al., 2017). Stochastic factors play a greater role in a limited sampling from a single population like in Gramlich et al. (2018), but in a metapopulational study with larger sampling, the stochastic effects should be diminished (Israel, 1992). The closer proximity of one parental species to the hybrids is more likely to result from differences

in relative abundance, which can affect the degree of backcrossing (Rieseberg, 1997), as was exemplified in *Morus,* where asymmetrical introgression to one of the parents was detected as a result of greater abundance (Burgess et al., 2005). The greater abundance of one parent might be ecologically mediated, meaning that *S. waldsteiniana* and hybrids might occupy more similar habitats than *S. foetida,* or that the fitness of uniparental backcrosses is larger in the given available niches (Abbott, 2017). Reproductive isolation of hybrids from *S. foetida* via genetic barriers is less likely than ecological isolation (Mallet, 2005, 2007, Yakimowski & Rieseberg, 2014). The absence of genetic incompatibilities seems to be especially true for genus *Salix* in general, whereas the roles of ecological factors as crossing barriers are poorly explored for alpine willows (Wagner et al., 2021). The prevalence of admixed genotypes in hybrid zone populations could be a sign of higher relative fitness of hybrids in instances when they occupy the same habitats as their parents (Arnold & Martin, 2010). Population IT8 showed the most equal rates of both parental genetic clusters (Figures 10 & 12) and relatively equal convergence with both parental populations (Figure 17, Table 3). In the *NewHybrids* analysis, the allelic frequencies within this population were recognised as strong F2 hybrid generation signals (Figure 14) with high posterior probabilities, except in one individual that exhibited a substantial portion of *S. waldsteiniana* backcross frequency patterns. It seems that this population might consist of a locally established hybrid lineage with potential for hybrid speciation. Since other hybrid zone populations of the two analysed species in this study show the ability to exchange genetic material with one of the parents, reproductive isolation in this population is more likely linked to the spatial or ecological isolation of a hybrid lineage. *Salix* hybrids were observed to exhibit higher tolerances to nutrient deficiency and soil acidity than either of the parental species (Gramlich et al., 2016), which could enable them to inhabit habitats that were unavailable for both parents. Locally isolated homoploid hybrid populations leading to hybrid speciation could be observed in Brochmann et al. (2000) where an endemic plant species *Argyranthemum sundingii* was observed to have originated from hybridization of two species followed by spatial separation. Adaptive divergence of lineages also plays an important role in processes of speciation (Abott & Comes, 2007), as exemplified in *Iris nelsonii,* a species of homoploid hybrid origin that is ecologically divergent from its progenitors (Taylor et al., 2011). Further surveys focusing on the investigation of spatial separation and ecological preferences of hybrid populations should be conducted to confirm the hypothesis of a possible speciation event.

Population IT5 (the remaining three individuals) and the majority of the population IT7 were seemingly classified as parental species' clusters with no (or only little) admixture in the *STRUCTURE* and *sNMF* analyses for $K = 2$ (Figures 10 & 12). High F_{ST}*-values* (Figure 17, Table 3) between these populations and parental populations, as well as PCoA analysis (Figure 15), showed a seemingly high divergence among these individuals. This was especially pronounced in population IT5. High coancestry coefficients between these individuals in the results of the *RADpainter* and *fineRADstructure* analyses (Figure 13), as well as short branches with low bootstrap support in the ML tree and NN phylogenetic network (Figures 6 & 8), hinted at very close relationships of these individuals, or even clonality. It is possible to assume that population IT5 is a remnant of a larger population that was reduced in size and restricted to a very small area, now surrounded by a ski slope (field observations from E. Hörandl). This would mean that the population has undergone a bottleneck event through the contraction of habitat

and individuals. The subsequent generations of this population came from a small number of survivors, resulting in high inbreeding and pronounced effects of genetic drift (Hedrick, 2011). The genetic patterns of such restricted and isolated populations show decreased allelic variability and increased homozygosity, resulting in departures from HWE (Hedrick, 2011), which affect the model approach in the *STRUCTURE* analysis and lead to overestimations of *K* (Pritchard et al., 2010). The approach in *sNMF* is better suited for inbred lineages because it avoids HWE assumptions, but it still predicts that genotyping frequencies are formed from ancestral frequencies without genetic drift (Frichot et al., 2014). In this study, the results of both analyses did not differ drastically and none was perfectly suitable for the constraints of this dataset. In both, estimation of a higher *K* was favoured (Appendices 7.5 & 7.6) as a result of patterns arising from high inbreeding within local populations. For *K = 2,* the less variable allelic patterns with higher homozygosity were recognised as stronger parental genetic entities. Especially the small numbered and spatially restricted population IT5 formed a strong genetic cluster. Such effects of inbreeding should be considered in the interpretation of the dataset. The patterns identified by *sNMF* and *STRUCTURE* seem to have arisen from these constraints, rather than from the true absence of admixture within the contact zone. The ability of vegetative dispersal via clones is frequent in plants and well documented in *Salix* (Skvortsov, 1999). This supports clonality as an interpretation for the individuals that show very high genetic similarity. Such individuals, when present among sexually reproducing lineages, probably bias the analysis even stronger than highly interbred ones, which is another explanation for the strong separation of the whole population IT5. For example, in population genetics of pathogens, a special emphasis is given to treating populations with clonal and sexual reproduction, which must be clone-corrected before the analysis for more accurate results (Grünwald et al., 2017).

Surprisingly, two *S. waldsteiniana* individuals from the population AU10, sampled outside the contact zone, could be classified as F2 hybrids in the *NewHybrids* analysis with high posterior probabilities (Figure 14). It is possible that treating these single individuals as a part of a larger population throughout the filtering steps of *STACKS* assembly for population genetics analyses, presented a bias, as investigated in (Graham et al., 2020). Two of the four additional single *S. foetida* individuals from populations CH11 and IT12 showed admixed patterns in the *NewHybrids* as well, as a contrast to all other purebred *S. foetida* individuals, which speaks favour of this interpretation. Further, the results of *NewHybrids* might be blurred due to the restricted number of loci that can be used in the analysis. However, individuals from AU10 consistently showed admixture patterns throughout the entire study, while this was not the case for the CH11 and IT12 *S. foetida* individuals. Even in the ML phylogenetic tree and NN network, where no grouping to populations had been done prior and no filtering was associated therewith, the individuals were positioned on long branches, and the *QS* scores of the AU10 population cluster favoured discordant topologies (Figures 7 & 8). In *RADpainter* and *fineRADstructure* two of them were positioned in the paraphyletic cluster with other admixed populations of the contact zone, whereas all *S. foetida* individuals outside the contact zone formed a monophyletic cluster, including the additional four single individuals (Figure 13). In *STRUCTURE,* as well as *sNMF* analysis, the ratios of the two parental clusters within AU10 individuals resembled patterns of introgression (Figures 10 & 12). Willows are capable of long distance dispersal of both pollen and seeds, so it is possible that the two species first came in contact in the central Alps during the latest post-glacial recolonization of the area, and admixed

individuals resembling *S. waldsteiniana* later dispersed towards the eastern-most parts of the Alps in the interior of Austrian territory. Not much is currently known about the biogeographical history of *Salix* in the Alps and what colonisation paths the species followed after the last glacial maximum (Wagner et al., 2021) to confirm or disprove this hypothesis. Given the complex, reticulate network within the genus *Salix*, especially within the *Chamaetia/Vetrix* clade (Wagner et al., 2018, 2020) with high interspecific hybridization and introgression rates, it is also possible that hybridization events with other members of this clade might have happened, presenting an additional noise in the analyses. These results call for a more complete sampling of *S. waldsteinana* populations in its eastern expansion range border in the Alps that would be needed to explain the observed patterns of admixture.

4.3. Morphological traits of parents and hybrids

Parental phenotypes, inferred from individuals outside the contact zone, could clearly be distinguished in the PCA analysis of the selected leaf traits. High PGMP confirmed the delimitation into two groups and phenotypic separation of two species based on the four examined leaf trait parameters. Leaf area and dentation density contributed most to interspecific variability, followed by length-to-widest point ratio and length-to-width ratio measurements (Figure 19). The results of the analysis are consistent with the current identification characters used in Hörandl et al. (2012). They also reveal that leaf area and dentation density should carry a greater weight in phenotypic identification, than the shape of the leaf, as they are more contrasting between the species.

The distinction of the two species was still possible when all populations from the contact zone were taken into account and assigned a group based on the phenotypic identification in the field. The variability of the second principal component, which describes most interspecific variability of the groups, was highly contributed by the variables explaining leaf shape (Figure 20), than leaf area and dentation density, hinting to higher intraspecific variability of leaf shape in the contact zone populations. Individuals identified as *S. foetida* and *S. waldsteinana* showed an overlap, as was also evident from the calculated average PGMP, which was lower for both species. These results suggest that phenotypes associated with each of the species are more similar and intermediate within the contact zone, than outside of it. Phenotypically identified hybrids nevertheless show the most intermediate characters and overlap with both parental groups. The overlap is slightly larger with the *S. waldsteiniana* phenotype. From the genetic perspective, this seems reasonable, as the analysis of molecular data displayed a tendency to higher rates of backcrossing and introgression with *S. waldsteinana*. However, the genetic analyses showed that morphologically identified phenotype was not always congruent with genetic data. Genetic identification of parents and hybrid classes, as defined by the results of the *NewHybrids* analysis, resulted in a lower overlap of the parental clusters in the PCA analysis. This suggested that genetically identified *S. waldsteiniana* individuals from populations IT6 (two individuals) and IT9 (six individuals) that were identified as hybrids, exhibit a less intermediate combination of morphometric characteristics, than individuals from populations IT8 (all eleven individuals) and IT7 (three individuals) that were phenotypically identified as *S. waldsteiniana*, but genetically classified as admixed. The reduction of population IT5, the only contact zone population identified as *S. foetida*, might contribute to a lesser overlap as well. The F2 hybrid cluster and *S. waldsteiniana* backcross cluster show a nearly complete overlap, which suggests that individuals with different ratios of admixture

cannot be phenotypically differentiated. All genetically admixed individuals show an overlap with both parental clusters, but it is larger with *S. waldsteiniana.* This is consistent to the pattern observed in the PCA analysis of groups based on field identification, even though several individuals were moved from *S. waldsteiniana* phenotype group to the admixed cluster and vice versa. Drawn from this is a conclusion that in the contact zone, admixed individuals are hard to distinguish from non-admixed *S. waldsteiniana* individuals on the basis of examined morphometric characters. Their phenotypic characteristics often fall within the phenotypical variability range of *S. waldsteiniana,* than exhibit perfect intermediary traits. Thompson et al. (2021) proposed that in hybridizing populations, individuals that are phenotypically similar to one parent should have a relatively high fitness and that phenotypic combinations of mismatches of parental characters are more detrimental. Congruent in this study is the observation of parental bias, but the multivariate approach does not allow for the investigation of single individual morphometric traits to search for combinations of mismatches. Instead, the multivariate PCA approach offers a more concise picture where intermediacy and parent bias can be assessed through a combination of all tested characters. The study of *S. eriocephala* and *S. sericea* hybrid zone by Hardig et al. (2000) has made very similar observations to this study. The analysed morphometric traits showed either intermediacy or were more similar to parental characteristics. Several phenotypically identified *S. eriocephala* individuals were found to be genetically admixed. If homoploid *Salix* hybridization events frequently result in parental-like phenotypes, as seen in this study, as well as in Hardig et al. (2000), it can be assumed that willow hybrid zones might be broader than the observed range of truly intermediate phenotypic forms, which calls for further studies with a more extensive sampling throughout the entire range of the contact zone, and not solely focused on areas of putatively intermediate forms.

The results of the morphometric studies might also be biased by the limited selection of morphometric traits (Lexer et al., 2009). Only four independent traits could be assessed in this study, which is comparably low to other morphometric analyses of plant hybrids (e.g. 11 traits for *Carex* in Blackstock & Ashton, 2010, 14 traits for *Quercus* in Kremer et al., 2002, 26 traits for *Castilleja* in Clay et al., 2012). However, the lack of morphometrically feasible traits seems to be the case in the genus *Salix* (e.g. 6 traits in Hardig et al., 2000). Additional possibly informative qualitative morphometric traits of *S. foetida* and *S. waldsteiniana* would assess the appearance of the leaf margin denticles in more detail. In field identification of the two species, an emphasis is given to the shape of the denticles and their glands as an identification character, especially when intermediate forms were observed (Hörandl, 1992). The typical *S. foetida* glands, unfortunately, lose their contrasting whitish colour when the sample is dried and become brown. To assess this qualitative character, it should be studied on living specimens. With more detailed microscopic imaging, the shapes of the denticles on the leaf margin could be captured and studied with geometric morphometrics approaches. Because the size and shape of the denticles are associated with the appearance of the glands, which are more prominent in *S. foetida* and minute in *S. waldsteiniana* (Hörandl, 1992, Hörandl et al., 2012), this might help to assess hybrid phenotypes. In a hybrid zone study of *S. eriocephala* and *S. sericea*, measuring the content of chemical compounds was proven to be more concordant with genetic classification, than morphometry (Hardig et al., 2000). Perhaps the investigation of secondary metabolites is more informative in disentangling relationships within the genus *Salix* (Nyman & Jukunen-Tiitto, 2005) and to identify hybrid individuals (Hardig et al., 2000, Oberprieler et

al., 2013) than morphological analyses, which are restricted to very few and very variable characters, as exemplified in this study and other studies on willows (Hardig et al., 2000, Wu et al., 2015, Wagner et al., 2018). Finally, Ecological surveys of the investigated populations might provide additional insight into relationships between phenotypes and ecological conditions of parental and hybrid habitats.

5. Conclusion

With the use of RAD-seq data and morphometrics, this study aimed to answer the questions regarding i) the establishment and extent of the hybrid zone in the overlapping range of the distribution area of *S. foetida* and *S. waldsteiniana* in the Alps, ii) genetic structure of admixed individuals and the extent of gene flow they share with the parental populations, and iii) the morphological traits of hybrid and parental phenotypes.

This study revealed that a geographically broad hybrid zone of *S. foetida* and *S. waldsteiniana* has evolved in the European Alps after the post-glacial colonisation of the area, characterised by patterns of introgression. A wide range of phylogenetic and population genetics analyses, performed on RAD-loci, assembled by two different software pipelines, showed high congruence of results. Since its establishment, the hybrid zone has progressed into a hybrid swarm with abundant admixture of later generation hybrids and backcrosses. Asymmetry of introgression to *S. waldsteiniana* was detected, with backcrosses to *S. foetida* being completely absent in the study area. Several factors might affect the establishment and succession of uniparental backcrosses, but the establishment of genetic barriers seems unlikely. Local potential for the establishment of a distinct hybrid lineage was detected. Phenotypes of individuals with admixed ancestry showed intermediacy with a bias to the *S. waldsteiniana* phenotype. This study complements current knowledge on willows and hybridization, and opens new questions regarding natural willow hybrid zones, in terms of ecological speciation, the transmission of adaptive alleles and biogeographical history of this species-rich taxon. It adds an important piece to the puzzle of plant hybrid zone investigation in the Alps, for being one of the few studies with a metapopulational approach covering the complete distribution range of *S. foetida* and *S. waldsteiniana* in this mountain range.

6. References

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*. **55**(4), 238-258.

Aeschimann, D., & Lauber, K. (2004). Flora alpina: ein Atlas sämtlicher 4500 Gefäßpflanzen der Alpen. Register (Vol. 3). *Haupt Verlag.*

Anderson, E. (1949). Introgressive hybridization. *John Wiley and Sons.*

Anderson, E. C., & Thompson, E. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. **160**(3), 1217-1229.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*. **17**(2), 81-92.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Argus GW. (1974). An experimental study of hybridization and pollination in *Salix* (willow). *Can J Bot*. **52**:1613–9.

Argus, G. W. (1997). Infrageneric classification of *Salix* (Salicaceae) in the new world. *Systematic botany monographs*, 1-121.

Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RAD seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular ecology*. **22**(11), 3179-3190.

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

Arnold, M. L. (2006). Evolution through genetic exchange. *Oxford University Press.*

Arnold, M. L., & Martin, N. H. (2010). Hybrid fitness across time and habitats. *Trends in Ecology & Evolution*. **25**(9), 530-536.

Arnold, M. L., Ballerini, E. S., & Brothers, A. N. (2012). Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. *Heredity*. **108**(3), 159-166.

Azuma, T., Kajita, T., Yokoyama, J., & Ohashi, H. (2000). Phylogenetic relationships of *Salix* (Salicaceae) based on rbcL sequence data. *American Journal of Botany*. **87**(1), 67-75.

Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: a Python package for genome-scale hybridization detection. *Systematic Biology*. **67**(5), 821-829.

Blumthaler, M. (2012). Solar radiation of the high Alps. In Lütz C. (2012) Plants in alpine regions: cell physiology of adaption and survival strategies. *Springer Science & Business Media*. 11-20.

Bryant, D., & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution*. **21**(2), 255-265.

Büchler, W. (1985). Neue Chromosomenzählungen in der Gattung *Salix*. *Botanica helvetica*. **95**(2), 165-175.

Burgess, K. S., Morgan, M., & Husband, B. C. (2008). Interspecific seed discounting and the fertility cost of hybridization in an endangered species. *New Phytologist*. **177**(1), 276-284.

Burgess, K. S., Morgan, M., Deverno, L., & Husband, B. C. (2005). Asymmetrical introgression between two Morus species (M. alba, M. rubra) that differ in abundance. *Molecular Ecology*. **14**(11), 3471-3483.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*. **22**(11), 3124-3140.

Clay, D. L., Novak, S. J., Serpe, M. D., Tank, D. C., & Smith, J. F. (2012). Homoploid hybrid speciation in a rare endemic *Castilleja* from Idaho (*Castilleja christii*, Orobanchaceae). *American journal of Botany*. **99**(12), 1976-1990.

Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P., & Wilkin, P. (2012). Plant species identification using digital morphometrics: A review. *Expert Systems with Applications*. **39**(8), 7562-7573.

Dobes, C. H., Hahn, B., & Morawetz, W. (1997). Chromosomenzahlen zur gefässpflanzen-flora Österreichs. *Linzer Biol. Beitr*. **29**(1), 5-43

Dray S. & Dufour A. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*. **22**(4), 1–20.

Eaton D.A.R. & Overcast I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics.*

Elworth, R. A., Ogilvie, H. A., Zhu, J., & Nakhleh, L. (2019). Advances in computational methods for phylogenetic networks in the presence of hybridization. In *Bioinformatics and Phylogenetics*, 317-360.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*. **14**(8), 2611-2620.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. *Genome research*. **8**(3), 186-194.

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. *Genome research*. **8**(3), 175-185.

Ferreyra, L. I., Vilardi, J. C., Verga, A., López, V., & Saidman, B. O. (2013). Genetic and morphometric markers are able to differentiate three morphotypes belonging to Section *Algarobia* of genus *Prosopis* (Leguminosae, Mimosoideae). *Plant Systematics and Evolution*. **299**(6), 1157-1173.

Fogelqvist, J., Verkhozina, A. V., Katyshev, A. I., Pucholt, P., Dixelius, C., Rönnberg-Wästljung, A. C., Lascoux, M., & Berlin, S. (2015). Genetic and morphological evidence for introgression between three species of willows. *BMC evolutionary biology*. **15**(1), 1-10.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*. **6**(8), 925-929.

Galtier, N., & Daubin, V. (2008). Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **363(**1512), 4023-4029.

Gotzek, D., Brady, S. G., Kallal, R. J., & LaPolla, J. S. (2012). The importance of using multiple approaches for identifying emerging invasive species: the case of the Rasberry crazy ant in the United States. *PLoS One.* **7**(9), e45314.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*. **53**(3-4), 325-338.

Graham, C. F., Boreham, D. R., Manzon, R. G., Stott, W., Wilson, J. Y., & Somers, C. M. (2020). How "simple" methodological decisions affect interpretation of population structure based on reduced representation library DNA sequencing: a case study using the lake whitefish. *PLoS One*. **15**(1), e0226608.

Gramlich, S., & Hörandl, E. (2016). Fitness of natural willow hybrids in a pioneer mosaic hybrid zone. *Ecology and Evolution*. **6**(21), 7645-7655.

Gramlich, S., Sagmeister, P., Dullinger, S., Hadacek, F., & Hörandl, E. (2016). Evolution in situ: hybrid origin and establishment of willows (*Salix* L.) on alpine glacier forefields. *Heredity*. **116**(6), 531-541.

Gramlich, S., Wagner, N. D., & Hörandl, E. (2018). RAD-seq reveals genetic structure of the F2-generation of natural willow hybrids (*Salix* L.) and a great potential for interspecific introgression. *BMC plant biology*. **18**(1), 1-12.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., … & Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*. **328**(5979), 710-722.

Gugerli, F. (1997). Hybridization of *Saxifraga oppositifolia* and *S. biflora* (Saxifragaceae) in a mixed alpine population. *Plant Systematics and Evolution*. **207**(3), 255-272.

Hardig, T. M., Brunsfeld, S. J., Fritz, R. S., Morgan, M., & Orians, C. M. (2000). Morphological and molecular evidence for hybridization and introgression in a willow (*Salix*) hybrid zone. *Molecular Ecology*. **9**(1), 9-24.

Hardig, T. M., Anttila, C. K., & Brunsfeld, S. J. (2010). A phylogenetic analysis of *Salix* (Salicaceae) based on matK and ribosomal DNA sequence data. *Journal of Botany*.

Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford surveys in evolutionary biology*. **7**, 69-128.

Harrison, R. G. (1993). Hybrid zones and the evolutionary process. *Oxford University Press on Demand*.

He, L., Wagner, N. D., & Hörandl, E. (2021). Restriction-site associated DNA sequencing data reveal a radiation of willow species (*Salix* L., Salicaceae) in the Hengduan Mountains and adjacent areas. *Journal of Systematics and Evolution*. **59**(1), 44-57.

Hedrick, P. W. (2011). Genetics of populations: 4th edition. *Jones & Bartlett Publishers*

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological journal of the Linnean Society*. **68**(1-2), 87-112.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. **359**(1442), 183-195.

Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. *Genetica*. **139**(5), 617-638.

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular ecology resources*. **11**, 117-122.

Hörandl, E. (1992). Die Gattung *Salix* in Österreich: mit Berücksichtigung angrenzender Gebiete (Vol. 27). *Abh. Der Zoologisch-Botanischen Gesellschaft in Österreich*. **27**.

Hörandl, E., Florineth, F., & Hadacek, F. (2012). Weiden in Österreich und angrenzenden Gebieten. 2. Auflage. *ViennaUniversity of Agriculture.*

Hörandl, E. (2022). Novel Approaches for Species Concepts and Delimitation in Polyploids and Hybrids. *Plants*. **11**(2), 204.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*. **24**(6), 417.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*. **23**(2), 254-267

Israel, G. D. (1992). Determining sample size. *University of Florida.*

Ivy-Ochs, S., Kerschner, H., Reuther, A., Preusser, F., Heine, K., Maisch, M., Kubik P.W., & Schlüchter, C. (2008). Chronology of the last glacial cycle in the European Alps. *Journal of Quaternary Science: Published for the Quaternary Research Association*. **23**(6-7), 559-573.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. **23**(14), 1801-1806.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. **24**(11), 1403-1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*. **11**(1), 1-15.

Jombart, T., & Collins, C. (2015). An introduction to adegenet 2.0.0. *Imperial College London, MRC Centre for Outbreak Analysis and Modelling.*

Kadereit, J. W. (2015). The geography of hybrid speciation in plants. *Taxon*. **64**(4), 673-687.

King, A.R., & Ferris C. (1998). Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular ecology*. **7**(9), 1151-1161.

Körner, C. (1989). The nutritional status of plants from high altitudes. *Oecologia*. **81**(3), 379-391.

Körner, C. (1999). Alpine plant life: functional plant ecology of high mountain ecosystems. *Springer*.

Kremer, A., Dupouey, J. L., Deans, J. D., Cottrell, J., Csaikl, U., Finkeldey, R., Espinel, S., Jensen J., Kleinschmit, J., & Badeau, V. (2002). Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science*. **59**(7), 777-787.

Kubatko, L. S., & Chifman, J. (2019). An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC evolutionary biology*. **19**(1), 1-13.

Liebst, B. (2008). Do they really hybridize? A field study in artificially established mixed populations of *Euphrasia minima* and *E. salisburgensis* (Orobanchaceae) in the Swiss Alps. *Plant Systematics and Evolution*. **273**(3), 179-18.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: population inference from RADseq data. *Molecular biology and evolution*. **35**(5), 1284-1290.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in ecology & evolution*. **20**(5), 229-237

Mallet, J. (2007). Hybrid speciation. *Nature*. **446**(7133), 279-283.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*. **27**(2 Part 1), 209-220.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular ecology resources*. **15**(1), 28-41.

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*. **74**(2), 560-564.

Mayr, E. (1963). Animal species and evolution. *Harvard University Press*.

McCartney-Melstad, E., Gidiş, M., & Shaffer, H. B. (2019). An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. *Molecular ecology resources*. **19**(5), 1195-1204.

Milla, R., & Reich, P. B. (2011). Multi-trait interactions, not phylogeny, fine-tune leaf size reduction with increasing altitude. *Annals of botany*. **107**(3), 455-465.

Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology*. **36**(2), 235-247.

Mosseler, A. (1990). Hybrid performance and species crossability relationships in willows (*Salix*). *Canadian Journal of Botany*. **68**(11), 2329-2338.

Neumann, A., & Polatschek, A. (1972). Cytotaxonomischer beitrag zur gattung *Salix*. *Annalen des Naturhistorischen Museums in Wien*. **76**, 619-633.

Oberprieler, C., Dietz, L., Harlander, C., & Heilmann, J. (2013). Molecular and phytochemical evidence for the taxonomic integrity of *Salix alba*, *S. fragilis*, and their hybrid *S.* × *rubens* (Salicaceae) in mixed stands in SE Germany. *Plant systematics and evolution*. **299**(6), 1107-1118.

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. **8**(10), 1360-1373.

Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., & Smith, S. A. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American journal of botany*. **105**(3), 385-403.

Percy, D. M., Argus, G. W., Cronk, Q. C., Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Husband, B. C., Newmaster S.G., Barret, S. C. H., & Graham, S. W. (2014). Understanding the spectacular failure of DNA barcoding in willows (*Salix*): Does this result from a trans-specific selective sweep?. *Molecular Ecology*. **23**(19), 4737-4756.

Petit, R.J., Pineau E., Demesure, B., Bacilieri, R., Ducousso, A., & Kremer, A. (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the USA*. **94**, 9996–10001.

Pflugbeil, G., Affenzeller, M., Tribsch, A., & Comes, H. P. (2021). Primary hybrid zone formation in *Tephroseris helenitis* (Asteraceae), following postglacial range expansion along the central Northern Alps. *Molecular ecology*. **30**(7), 1704-1720.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*. **155**(2), 945-959.

Pritchard, J. K., Wen, X., & Falush, D. (2010). Documentation for structure software: Version 2.3. *University of Chicago.*

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., … & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*. **81**(3), 559-575.

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org/.

Rambaut, A., & Drummond, A. J. (2012). FigTree version 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/.

Raymond, M. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity.* **86**, 248-249.

Rechinger, H. K. (1993). Flora Europaea 1. *Cambridge University Press*.

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**(3), 213-241.

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annual review of Ecology and Systematics*. **28**(1), 359-389.

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*. **83**(4), 363-372.

Rieseberg, L. H., Raymond, O., Rosenthal, 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**(5637), 1211-1216

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*. **12**(12), 2640-2659.

Ruiz, D. C., Machío, I. V., Nieto, A. H., & Feliner, G. N. (2021). Hybridization and cryptic speciation in the Iberian endemic plant genus *Phalacrocarpum* (Asteraceae-Anthemideae). *Molecular Phylogenetics and Evolution*. **156**, 107024.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*. **4**(4), 406-425.

Skvortsov, A. K. (1999). Willows of Russia and adjacent countries: taxonomical and geographical revision (transl. from: Skvortsov AK (1968) Willows of the USSR. Taxonomic and geographic revision. Nauka, Moscow). *Joensuu University, Joensuu.*

Soltis, D. E., Visger, C. J., & Soltis, P. S. (2014). The polyploidy revolution then… and now: Stebbins revisited. *American journal of botany*. **101**(7), 1057-1078.

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual review of plant biology*. **60**, 561-588

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**(9), 1312-1313.

Stebbins, G. L. (1950). Variation and evolution in plants. *Columbia University Press.*

Stebbins, G. L. (1959). The role of hybridization in evolution. *Proceedings of the American Philosophical Society*. **103**(2), 231-251.

Taylor, S. J., Willard, R. W., Shaw, J. P., Dobson, M. C., & Martin, N. H. (2011). Differential response of the homoploid hybrid species *Iris nelsonii* (Iridaceae) and its progenitors to abiotic habitat conditions. *American Journal of Botany*. **98**(8), 1309-1316.

Thompson, K. A., Urquhart-Cronish, M., Whitney, K. D., Rieseberg, L. H., & Schluter, D. (2021). Patterns, predictors, and consequences of dominance in hybrids. *The American Naturalist*. **197**(3), E72-E88

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Ralph, S., Rombauts S., … & Rokhsar, D. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. **313**(5793), 1596-1604.

Vu, V.Q. (2011). ggbiplot: A ggplot2 based biplot. https://rdrr.io/github/vqv/ggbiplot/.

Wagner, N. D., Gramlich, S., & Hörandl, E. (2018). RAD sequencing resolved phylogenetic relationships in European shrub willows (*Salix* L. subg. *Chamaetia* and subg. *Vetrix*) and revealed multiple evolution of dwarf shrubs. *Ecology and Evolution*. **8**(16), 8243-8255.

Wagner, N. D., He, L., & Hörandl, E. (2020). Phylogenomic relationships and evolution of polyploid *Salix* species revealed by RAD sequencing data. *Frontiers in plant science*, 1077.

Wagner, N. D., He, L., & Hörandl, E. (2021). The evolutionary history, diversity, and ecology of willows (*Salix* L.) in the European Alps. *Diversity*. **13**(4), 146.

Whitney, K. D., Ahern, J. R., Campbell, L. G., Albert, L. P., & King, M. S. (2010). Patterns of hybridization in plants. *Perspectives in Plant Ecology, Evolution and Systematics*. **12**(3), 175-182.

Wichura, M. (1865). Die Bastardbefruchtung im Pflanzenreich: erlaütert an den Bastarden der Weiden.

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York.*

Wright, S. (1938). Size of population and breeding structure in relation to evolution. *Science*. **87**, 430-431.

Wright, S. (1940). Breeding structure of populations in relation to speciation. *The American Naturalist*. **74**(752), 232-248.

Wu, J., Nyman, T., Wang, D. C., Argus, G. W., Yang, Y. P., & Chen, J. H. (2015). Phylogeny of *Salix* subgenus *Salix* sl (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC evolutionary biology*. **15**(1), 1-13.

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**(8), 1247-1258.

Yan, L. J., Burgess, K. S., Milne, R., Fu, C. N., Li, D. Z., & Gao, L. M. (2017). Asymmetrical natural hybridization varies among hybrid swarms between two diploid Rhododendron species. *Annals of Botany*. **120**(1), 51-61.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. **28**(24), 3326-3328.

7. Appendix

7.1. List of materials used

Table S1: List of all devices, kits, software and databases used in the thesis.

a) Devices and kits

Device	Label	Manufacturer
1-10 µL, 2-20 µL, 20-200 µL and 100-1000 µL pipettes	Pipetman Classic™	Gilson, Middleton, USA
DNA extraction kit	DNeasy Plant Mini Kit	Qiagen, Hilden, Germany
Fluorometer	Qubit 3.0	Thermo Fisher Scientific, Waltham, USA
Pipette tips for 10 µL, 200 µL and 1000 µL volume	Pipette tips	Sarstedt, Nümbrecht, Germany
Reaction tubes	1.5 ml and 2.5 ml safe-lock reaction tube	Sarstedt, Nümbrecht, Germany
Swing mill	TissueLyser II	Qiagen, Hilden, Germany
Steel beads	Stainless steel beads 5 ⌀ mm	Qiagen, Hiden, Germany
Tea bags	Profissimo dm Teefilter	dm-drogerie markt GmbH+ Co. KG, Karlsruhe, Germany
Scanner	CanoScan LiDe 220	Canon, Tokyo, Japan

b) Software applications

Application	Software	Citation/Manufacturer
Analysis of branch support	quartetsampling (v.1.3.1)	Pease et al., 2018
Assembly of RAD-loci	ipyrad (v.0.9.52)	Eaton & Overcast, 2020
Assembly of RAD-loci	STACKS (v.2.2)	Catchen et al., 2013
Assignment to parental and hybrid classes	NewHybrids (v.1.1. beta)	Anderson & Thompson, 2002
Detection of reticulate evolution	HyDe	Blischak et al., 2018
Geographic information system	QGIS (v.3.16.3)	QGIS developmental team, Open Source Geospatial Foundation Project
Image analysis software	Digimizer (v.5.7.2)	MedCalc Software Ltd, Ostend, Belgium

Parsing and summarizing output data from STRUCTURE	STRUCTURE HARVESTER (v.0.6.94)	Earl & vonHoldt, 2012
Permutation of clusters output	CLUMPP (v.1.1.2)	Jakobsson & Rosenberg, 2007
Phylogenetic network inference	SplitsTree (v.4.17.1)	Huson & Bryant, 2006
Phylogenetic tree inference	RAxML (v.8.2.4)	Stamatakis, 2014
Population structure analysis	STRUCTURE (v.2.3.4)	Pritchard et al., 2000
Population structure analysis	LEA (v.1.4.0)	Frichot & François, 2015
Population structure analysis	RADpainter and fineRADstructure	Malinsky et al., 2018
Quality check of sequenced, demultiplexed reads	FastQC (v.0.11.9)	Andrews, 2010
R Editor	RStudio (v.1.3.1093)	RStudio, Boston, USA
R package for biplot implementation with ggplot2	ggbiplot (v.0.55)	Vu, 2011
R package for declaratively creating graphics	ggplot2 (v.3.3.5)	Wickham, 2016
R package for multivariate analysis of genetic markers	ade4 (v.1.7.18)	Dray & Dufour, 2007
R package for multivariate analysis of genetic markers (extension of ade4)	adegenet (v.2.0.1)	Jombart, 2008
R package for parallel computation of relatedness and principal component analysis of SNP data	SNPRelate (v.12.2)	Zheng et al., 2012
R package for population genetics analyses	genepop (v.1.1.7)	Raymond, 1995
Spreadsheet program	Microsoft Office Excel (v.2016)	Microsoft Corporation, Redmond, USA
Vector graphics editor	Inkscape (v.1.0.3)	Inkscape project, 2020
Visualisation of phylogenetic trees	FigTree (v.1.4.4)	Rambaut, 2012

Whole genome data analysis toolset	PLINK (v.1.90)	Purcell et al., 2007

c) Databases

Type of data	Name	Location/Hyperlink
Herbaria collection	Herbarium GOET	University of Göttingen, Göttingen, Germany
Genetic sequence database	GenBank	https://www.ncbi.nlm.nih.gov/genbank

7.2. List of abbreviations

Table S2: List of abbreviations used in the thesis.

abbreviation	stands for
°C	degrees Celsius
µL	microlitre
BC	backcross
bp	base pair(s)
DAPC	discriminant analysis of principal components
DNA	deoxyribonucleic acid
F1	first generation
F2	second generation
g	gram
Hz	hertz
km	kilometre
mg	milligram
Mio	million
ML	maximum likelihood
mL	millilitre
mm	millimetre
ng	nanogram
NN	neighbour net
PC / PCA	principal component (analysis)
PCC	Pearson's correlation coefficient

PCo / PCoA	principal coordinate (analysis)
Q	phred score
RAD-seq	restriction site associated DNA sequencing
SNP	single nucleotide polymorphism

7.3. DNA extractions

Table S3: Table of all extracted samples. Population ID correspondent to Table 1. *before normalisation to 20 ng/µL

Individual ID	Species (identified phenotype)	Population ID	DNA concentration A [ng/µL]	DNA concentration B [ng/µL]	DNA concentration C [ng/µL]	DNA concentration used for RAD-seq*	comment
EH10542	hybrid	IT6	17.3	20		20	
EH10545	hybrid	IT6	21.8	33.7		33.7	
EH10546	hybrid	IT6	12.9	13.7	12.6	29.6	pooled
EH10547	hybrid	IT6	23.6	20.7		23.6	pooled
EH10549	hybrid	IT6	17.4	25.3		25.3	
EH10551	hybrid	IT6	29.9	16.1		29.9	
EH10552	hybrid	IT6	16.1	14.8		20.1	pooled
EH10554	hybrid	IT6	1.84	2.5			
EH10555	hybrid	IT6	1.9	2.27			
EH10556	hybrid	IT6	2.36	2.23			
EH10557	hybrid	IT6	3.22	18.9	14	22	pooled
EH10558	hybrid	IT6	11.1	10.2	10.5		
EH10559	hybrid	IT6	17.7	17		26.6	pooled
EH10560	hybrid	IT6	14.7	18.9		22	pooled
EH10562	hybrid	IT6	28	24.5		28	
EH10645	*Salix waldsteiniana*	IT7	28.1	25.3		28.1	
EH10647	*Salix waldsteiniana*	IT7	31.3	26.3		31.3	
EH10648	*Salix waldsteiniana*	IT7	28.1	30.6		30.6	
EH10650	*Salix waldsteiniana*	IT7	21	16.1			
EH10651	*Salix waldsteiniana*	IT7	20.4	20.6		20.6	
EH10652	*Salix waldsteiniana*	IT7	21.7	17.5		21.7	
EH10653	*Salix waldsteiniana*	IT7	18.6	20.2			
EH10654	*Salix waldsteiniana*	IT7	26.7	28.7		28.7	
EH10655	*Salix waldsteiniana*	IT7	21.5	18.2			
EH10656	*Salix waldsteiniana*	IT7	19.1	33.9		33.9	
EH10657	*Salix waldsteiniana*	IT7	22.2	21.1		22.2	
EH10659	*Salix waldsteiniana*	IT7	43.1	31		43.1	
EH10660	*Salix waldsteiniana*	IT7	13.3	18.5			
EH10662	*Salix waldsteiniana*	IT7	21.9	21.2		21.9	
EH10663	*Salix waldsteiniana*	IT7	26.3	33.2		33.2	
EH10664	hybrid	IT9	17.4	17.6		28.9	pooled
EH10665	hybrid	IT9	20.6	20.5		20.6	
EH10667	hybrid	IT9	20.6	22.6		22.6	
EH10668	hybrid	IT9	21.8	31.7		31.7	
EH10669	hybrid	IT9	15.4	18.3		21.2	pooled
EH10670	hybrid	IT9	1.9	2.38			
EH10672	hybrid	IT9	2.19	2.34			
EH10673	hybrid	IT9	2.14	2.35			
EH10674	hybrid	IT9	30.2	2.53		30.2	
EH10675	hybrid	IT9	23.1	34.7		34.7	
EH10676	hybrid	IT9	20.2	20.7		20.7	
EH10677	hybrid	IT9	8.34	10			

EH10678	hybrid	IT9	22.2	17.9		22.2	
EH10680	hybrid	IT9	16.7	19.6		25.7	evaporated
EH10681	hybrid	IT9	21.9	29.9		29.9	
EH10684	*Salix waldsteiniana*	IT8	16.3	14.8		24.2	pooled
EH10685	*Salix waldsteiniana*	IT8	21.9	22.3		22.3	
EH10687	*Salix waldsteiniana*	IT8	21.4	22.5		22.5	
EH10688	*Salix waldsteiniana*	IT8	21.4	27.2		27.2	
EH10689	*Salix waldsteiniana*	IT8	11.9	17.7		27	pooled
EH10690	*Salix waldsteiniana*	IT8	18.5	19.7		27.7	evaporated
EH10691	*Salix waldsteiniana*	IT8	13.3	14	15.3		
EH10692	*Salix waldsteiniana*	IT8	8.73	19.1	10.2	20.8	
EH10693	*Salix waldsteiniana*	IT8	7.56	9.12			
EH10694	*Salix waldsteiniana*	IT8	14.8	16.4		26.7	pooled
EH10695	*Salix waldsteiniana*	IT8	25.9	30.2		30.2	
EH10696	*Salix waldsteiniana*	IT8	11.3	7.48			
EH10697	*Salix waldsteiniana*	IT8	13.4	7.9	24.5	24.5	
EH10698	*Salix waldsteiniana*	IT8	25.4	18.5		25.4	
EH10699	*Salix waldsteiniana*	IT8	10.8	15.1	14.2		
EH10701	*Salix waldsteiniana*	IT8	10.7	6.56			
EH10703	*Salix foetida*	IT5	21	18.5		21	
EH10704	*Salix foetida*	IT5	11.5	15.7	28.2	28.2	
EH10705	*Salix foetida*	IT5	12.1	15.7	25.2	25.2	
EH10706	*Salix foetida*	IT5	19	11.3		22.1	pooled
EH10707	*Salix foetida*	IT5	11.5	11.2			
EH10708	*Salix foetida*	IT5	11.1	11.9			
EH10709	*Salix foetida*	IT5	13.6	14.9	26.5	26.5	
EH10710	*Salix foetida*	IT5	12.4	15.1			
EH10711	*Salix foetida*	IT5	18.8	15.6			
EH10712	*Salix foetida*	IT5	12.3	15	27.8	27.8	
EH10713	*Salix foetida*	IT5	14.7	14.8	29.8	29.8	
EH10714	*Salix foetida*	IT5	13.3	14.1	26.1	26.1	
EH10715	*Salix foetida*	IT5	14.5	12.7	23.9	23.9	
EH10716	*Salix foetida*	IT5	15.9	15.3	28.1	28.1	
EH10717	*Salix foetida*	IT5	14.9	15	24.3	24.3	
EH10718	*Salix foetida*	IT5	7.17	10.3			
EH10719	*Salix foetida*	IT5	11.7	10.3			
EH10720	*Salix foetida*	IT5	11.3	11.5			
EH10832	*Salix foetida*	CH4	15.9	22.70		22.70	
EH10833	*Salix foetida*	CH4	17.5	14.90		28	pooled
EH10834	*Salix foetida*	CH4	20.7	21.20		21.20	
EH10835	*Salix foetida*	CH4	21.8	20.60		21.8	
EH10836	*Salix foetida*	CH4	15.3	16.40			
EH10837	*Salix foetida*	CH4	16.6	15.20		22.6	pooled
EH10838	*Salix foetida*	CH4	16.4	14.10			
EH10839	*Salix foetida*	CH4	8.27	7.44			
EH10840	*Salix foetida*	CH4	18.5	14.10		22	pooled
EH10841	*Salix foetida*	CH4	18.8	15.90		26.3	pooled
EH10842	*Salix foetida*	CH4	11.5	16.70			
EH10843	*Salix foetida*	CH4	31.2	25.50		31.2	
EH10844	*Salix foetida*	CH4	25.2	25.70		25.70	
EH10845	*Salix foetida*	CH4	17.5	15.90			
EH10846	*Salix foetida*	CH4	18.7	20.10		20.10	
EH10958	*Salix foetida*	CH3	7.08	4.08			
EH10959	*Salix foetida*	CH3	10.9	8.97	14.7		
EH10960	*Salix foetida*	CH3	7.47	10.10	13.9		
EH10961	*Salix foetida*	CH3	5.37	7.31			
EH10962	*Salix foetida*	CH3	10.2	11.90	21.1	21.1	
EH10963	*Salix foetida*	CH3	13.9	15.90		32.4	pooled
EH10964	*Salix foetida*	CH3	22.5	18.40		22.5	
EH10965	*Salix foetida*	CH3	9.71	10.10	15.1		
EH10966	*Salix foetida*	CH3	21.2	20.60		21.2	

EH10967	*Salix foetida*	CH3	14.6	13.30		21.8	
EH10968	*Salix foetida*	CH3	30	24.40		30	
EH10969	*Salix foetida*	CH3	16.2	33.20		33.20	
EH10970	*Salix foetida*	CH3	19.8	21.00		21.00	
EH10971	*Salix foetida*	CH3	31.1	13.90			
EH10972	*Salix foetida*	CH3	8.65	8.65			
EH10973	*Salix foetida*	CH3	14.4	24.5		24.5	
EH10974	*Salix foetida*	CH3	39.5	48.8	21.8	48.8	
PM20_006	*Salix waldsteiniana*	SI1	29.9	17.9		29.9	
PM20_007	*Salix waldsteiniana*	SI1	16.3	14.3		56	pooled
PM20_008	*Salix waldsteiniana*	SI1	15.5	15.6		60	pooled
PM20_009	*Salix waldsteiniana*	SI1	15.2	17			
PM20_010	*Salix waldsteiiniana*	SI1	46.1	37.1		42.1	
PM20_011	*Salix waldsteiniana*	SI1	out of range	out of range		52	diluted
PM20_012	*Salix waldsteiniana*	SI1	50	44.7		50	
PM20_014	*Salix waldsteiniana*	SI1	38.7	42.8		42.8	
PM20_016	*Salix waldsteiniana*	SI1	32.9	27.7		32.9	
PM20_021	*Salix waldsteiniana*	SI1	37.3	44.7		47.7	
PM20_025	*Salix waldsteiniana*	SI1	45.1	45.3		45.3	
PM20_031	*Salix waldsteiniana*	SI2	31.8	36.6		36.6	
PM20_034	*Salix waldsteiniana*	SI2	16.8	23.9		23.9	
PM20_035	*Salix waldsteiniana*	SI2	20.3	39.6		39.6	
PM20_036	*Salix waldsteiniana*	SI2	28.5	29.4		29.4	
PM20_038	*Salix waldsteiniana*	SI2	35.9	26.7		35.9	
PM20_042	*Salix waldsteiniana*	SI2	18.7	25.1		25.1	
PM20_043	*Salix waldsteiniana*	SI2	27.4	25.3		27.4	
PM20_044	*Salix waldsteiniana*	SI2	14.2	18		28.7	pooled
PM20_045	*Salix waldsteiniana*	SI2	30	26.4		30	
PM20_046	*Salix waldsteiniana*	SI2	20.7	23.1		23.1	

7.4. Selection of the optimal clustering similarity

Description of the six metrics from McCartney-Melstadt et al. (2019) that were used to find the optimal similarity threshold for clustering of homologous RAD-loci.

Fraction of inferred paralogues: With an increase of the clustering threshold, the number of flagged (filtered) paralogues is expected to decrease moderately. But when the similarity threshold is set too high, this change is sharp and can be identified on the slope of the function of filtered loci against the similarity threshold setting. McCartney-Melstadt et al. (2019) propose, that with a split of real alleles into separate clusters, fewer paralogues are flagged and filtered, meaning that a strong decrease of filtered loci should be observed. However, each sharp change towards both, a decrease or an increase of filtered loci, should indicate a threshold set too high. It should also be avoided to choose a threshold with a too high number of filtered paralogs. Therefore, a range of the most fitting similarity thresholds lies in the shallow slope before the drastic change, where the best clusters composed of true homologues are formed. To investigate this, the numbers of filtered loci given in the output of step 5 of the *ipyrad* assembly and the numbers of filtered loci from *populations.log* of the *STACKS* assembly are visualised in a plot against the corresponding similarity threshold settings. Though the number of filtered loci obtained from *STACKS* in this manner also takes into account the loci filtered by the minimum percentage of samples (here set to 80%), *STACKS* generates no equivalent output to that of the *ipyrad* step 5. The relationship of filtered loci against the similarity threshold observed for the *ipyrad* and *STACKS* assembly was in congruence, which indicated that this approach is fitting.

Genetic variation (Heterozygosity and number of SNPs): Heterozygosity and the total number of SNPs across the different clustering similarity thresholds are both expected to follow a similar pattern, where they increase with an increase of the similarity threshold, up to a certain point where they are maximised, which is then followed by a decrease. For both of these metrics, the optimum values are at or near the maximisation point. Heterozygosity values were extracted from the step 5 output of the *ipyrad* assembly and from the summary statistics output of the *populations* output of *STACKS.* The umber of SNPs was extracted from the step 7 output of the *iyprad* assembly and from the *populations.log* file of the *STACKS* assembly.

Cumulative variance explained by the major principal components: If the assembly data is accurate and relevant, the noise of the main patterns of genetic differentiation should decrease and the proportion of variance explained by the first several principal components (PCs) should increase. A strong decrease in the variance proportion after optimal threshold values is a result of the over-splitting of true orthologues. To calculate the proportions of genetic variance for the first 8 PCs, a script, utilizing the functions of the *SNPRelate* (v.12.2) (Zheng et al., 2012) package was used. This script (as well as all the other scripts provided by the authors) were obtained from https://github.com/atcg/clustOpt/. Variant call format (VCF) outputs of both assembly runs were used as an input and the number of PCs was set to eight. Several other small numbers of PCs were investigated, but they all showed a consistent pattern.

Relationship between missingness and genetic divergence: Homologues loci of more distantly related individuals are more likely to over-split, compared to that of more closely related individuals because they tend to be more divergent. This incorrect split to divergent alleles increases pairwise missingness in the data. An investigation of the relationship between genetic similarity and its correlation with missingness across similarity clustering thresholds provides insight into the threshold where an incorrect splitting occurs. With the provided scripts, VCF files were used to calculate genetic similarity and pairwise data missingness. The result was a calculated Pearson's correlation coefficient between missingness and genetic similarity and the pairwise relationships were visualised with heatmaps.

The slope of isolation by distance: Generally, for individuals of the same species, genetic distance is expected to increase with geographic distance. When orthologous loci were over-split during the assembly with a similarity clustering threshold being set too high, this leads to a sharp reduction in the observed genetic distance of geographically distant samples. This effect was measured with provided scripts on the VFC outputs of both assemblies.

Maximisation of polymorphic loci at r = 80: For the *STACKS* assembly an additional metric was taken into account. This was the *r80* method recommended by the software developer that was described by Paris et al. (2017). In this approach, the most fitting threshold is that, where the number of polymorphic loci is maximised at *r80.* This filter is defined in the *populations* pipeline by the flag *–r,* which sets the minimal percentage of samples per population where a locus must be present to be retained. For the empirical investigation, all individuals were classified into one population, meaning *r 80* was equal to 80% of all investigated individuals.

7.5. Outputs of the *STRUCTURE* analysis

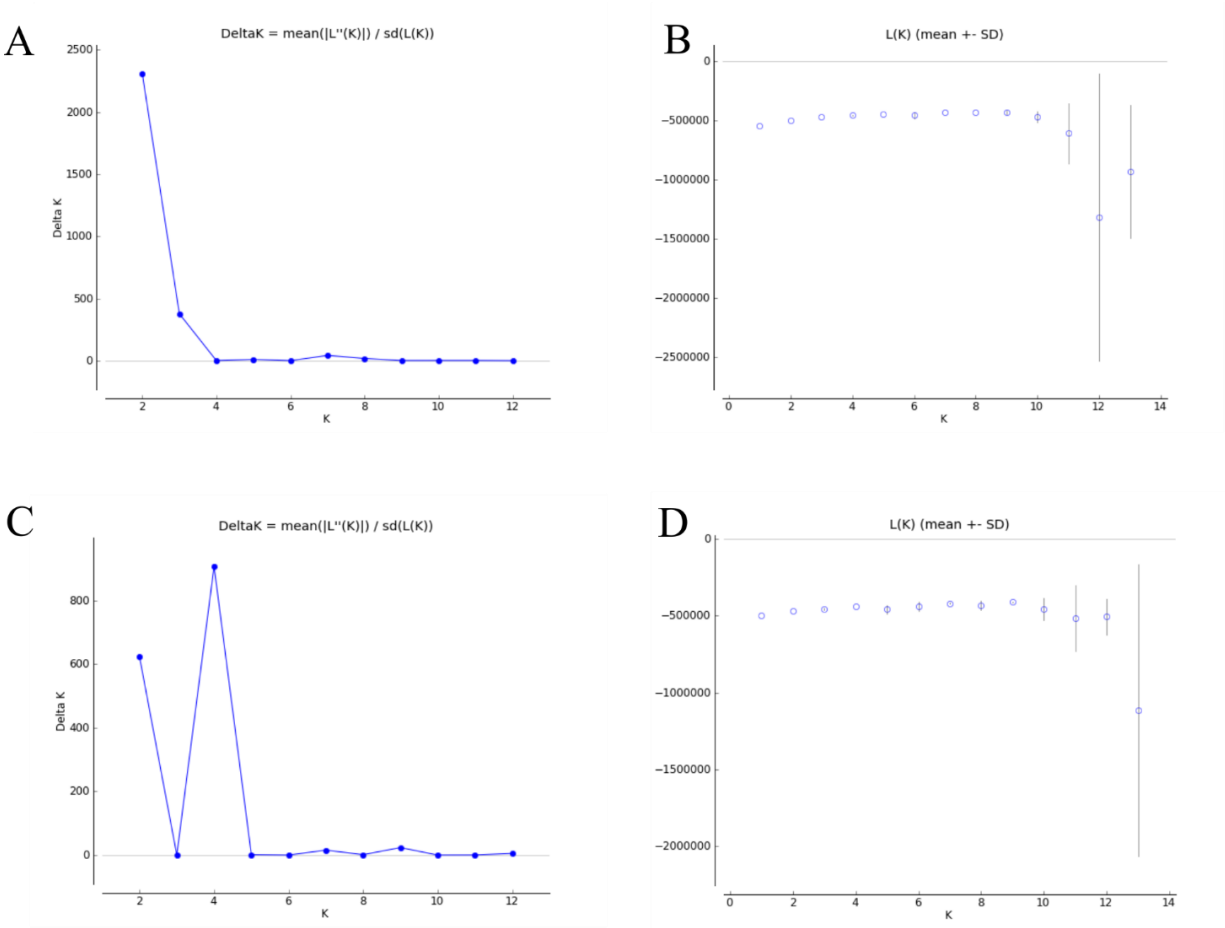

Figure S1: Evanno's Δ*K* and Ln PR(X/K) plots for *STRUCTURE* outputs of **A)** and **B) analysis of all 102 individuals, C)** and **D) analysis of 94 individuals (population IT5 reduced to three individuals).** *STRUCTURE* analysis was run for the *STACKS* filtered assembly for *K* 1 – 13. The outputs presented in this figure were generated in *STRUCTURE HARVESTER.*

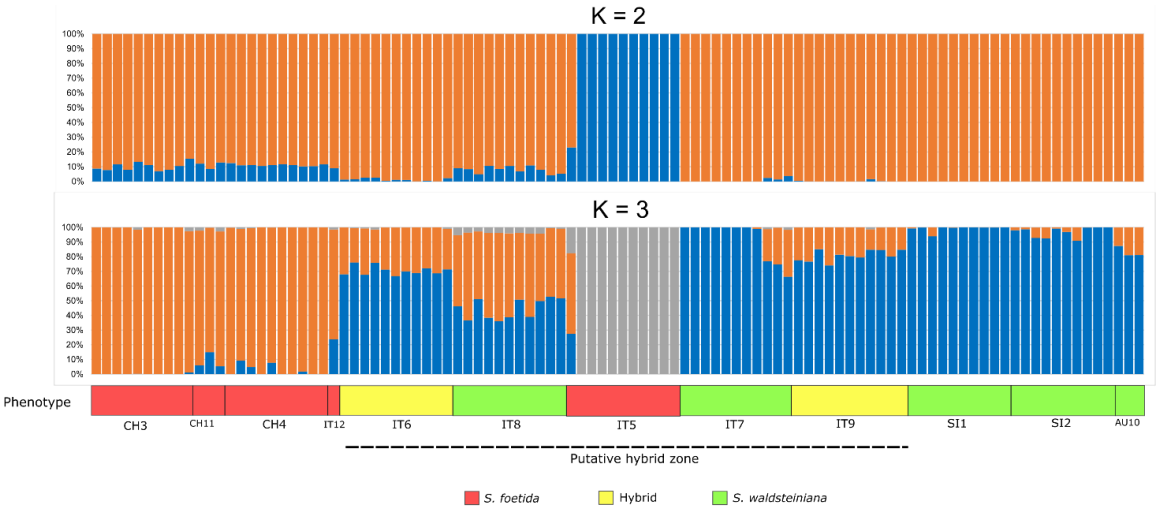

Figure S2: Results of the *STRUCTURE* analysis of all 102 individuals for the two *K* with the highest Δ*K* values**.** Results are represented with stacked bar plots. Different colours of the plots represent posterior probability for different genetic groups as calculated in the analysis. Populations are ordered by their geographical location from west to east, labelled by their identifiers (Table 1) and coloured by phenotype as described in the legend below. The dotted line marks populations of the putative hybrid zone (contact zone).

7.6 Outputs of the *sNMF* analysis

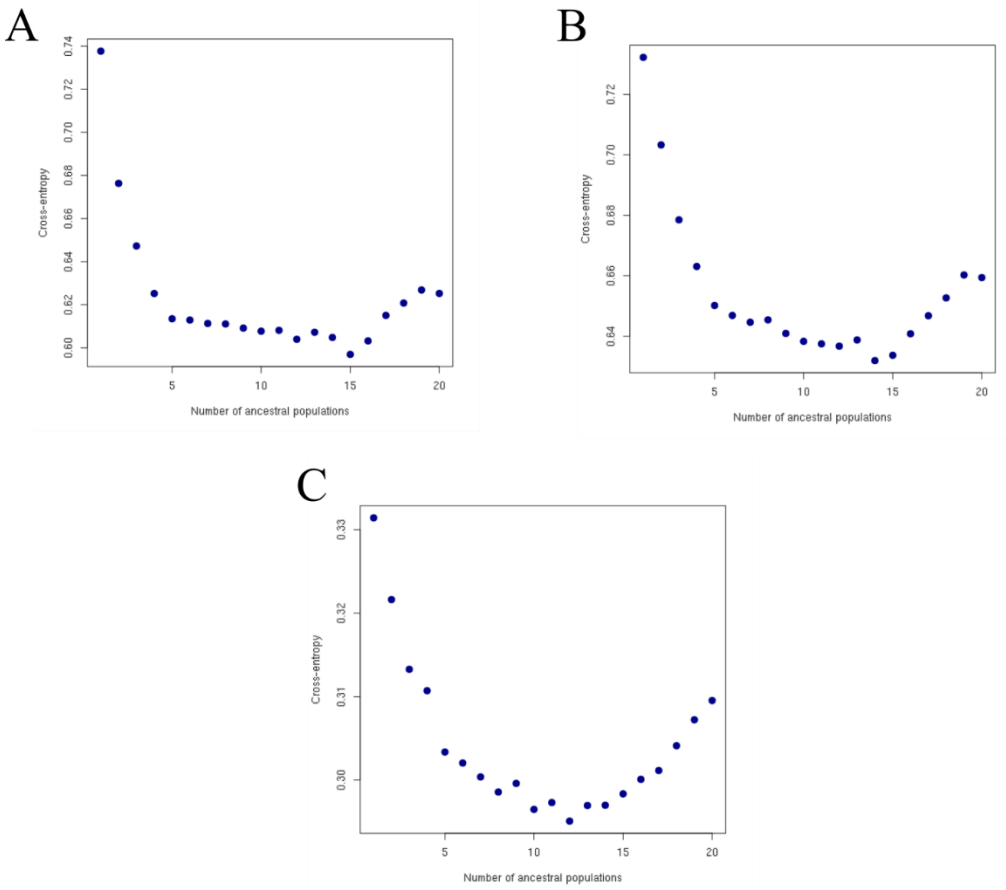

Figure S3: Plots of cross-entropy per *K* as calculated and generated with the *LEA* package in *R* for **A) all 102 individuals using *STACKS* filtered SNP assembly**, **B) the reduced dataset of 94 individuals using *STACKS* filtered SNP assembly** and **C) the reduced dataset of 94 individuals using *ipyrad* unlinked SNP assembly.**

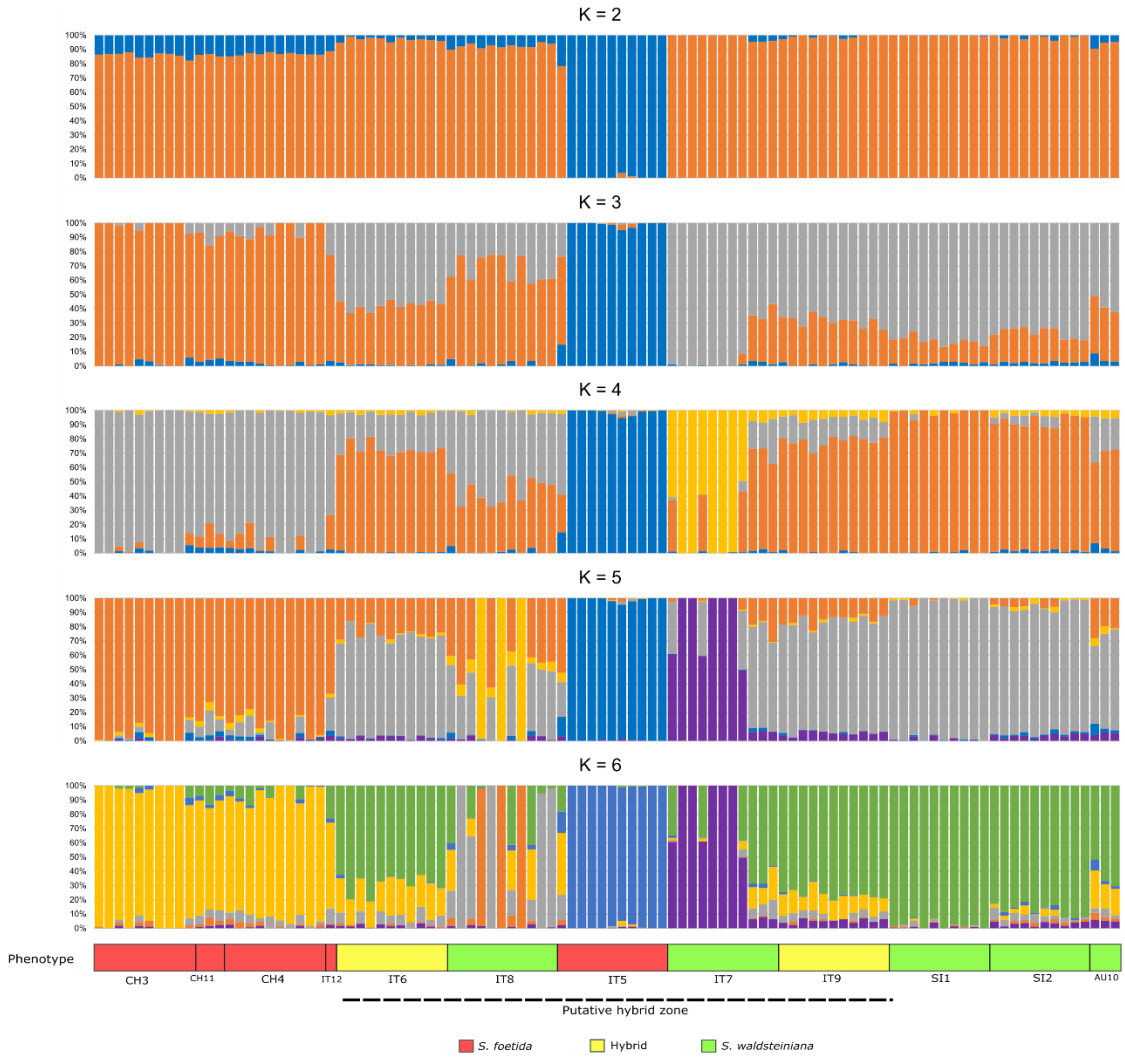

Figure S4: results of the *sMNF* analysis for *K* between 2 and 6 for all 102 individuals**.** Results are represented with stacked bar plots. Different colours of the plots represent posterior probabilities for different genetic groups as calculated in the analysis. Populations are ordered by their geographical location from west to east, labelled by their identifiers (Table 1) and coloured by phenotype as described in the legend below. The dotted line marks populations from the putative hybrid zone (contact zone).

7.7. Morphometric measurements

Table S4: Table of all morphometric measurements, averaged per individual. Population identifier is in correspondence to Table 1. The groups in the "*NewHybirds* class" column stand for: wald for *S. waldsteiniana*, foet for *S. foetida*, F2hyb for F2 hybrid, BCwald for *S. waldsteiniana* backcross and *N/A* represents a missing value (class not assigned). Individuals were assigned a group according to the results of the *NewHybrids* analysis.

Individual ID	Population ID	*NewHybrids* class	Leaf area	Leaf perimeter	Number of teeth	Length	Width	Distance to widest point	Teeth/cm	Length:Width	Length:Widest point
EH10542	IT6	F2hyb	215.70	57.50	30.80	24.60	12.40	14.20	1.52	1.98	1.74
EH10545	IT6	BCwald	263.60	68.20	38.80	30.70	11.80	15.30	1.59	2.64	2.04
EH10546	IT6	F2hyb	184.50	55.25	23.50	24.50	11.25	12.13	1.35	2.21	2.04
EH10547	IT6	BCwald	169.20	54.80	33.70	23.50	10.10	13.00	2.13	2.36	1.83
EH10549	IT6	BCwald	169.50	54.38	30.00	23.00	9.88	12.38	1.89	2.37	1.90
EH10551	IT6	BCwald	252.22	66.00	23.11	29.44	12.33	14.44	0.85	2.38	2.04
EH10552	IT6	F2hyb	214.30	63.30	44.50	28.10	12.20	16.20	2.20	2.32	1.74
EH10557	IT6	wald	176.38	55.13	33.00	24.50	10.50	12.50	1.93	2.34	1.98
EH10559	IT6	BCwald	240.00	62.40	23.50	27.30	13.40	14.10	1.04	2.05	1.96
EH10560	IT6	wald	184.86	54.29	22.57	23.71	10.71	12.14	1.35	2.21	1.94
EH10562	IT6	F2hyb	248.30	62.00	39.30	26.70	13.50	13.60	1.69	2.00	1.98
EH10645	IT7	wald	144.43	52.86	26.14	24.00	8.86	12.86	1.81	2.73	1.89
EH10647	IT7	wald	275.33	73.83	28.50	33.83	12.17	20.50	1.10	2.78	1.63
EH10648	IT7	wald	205.88	65.50	27.50	30.25	10.00	18.75	1.41	3.01	1.62
EH10651	IT7	wald	250.38	70.00	32.75	31.75	11.50	17.13	1.43	2.80	1.88
EH10652	IT7	wald	223.40	69.60	27.90	32.40	10.40	18.90	1.25	3.12	1.72
EH10654	IT7	wald	214.50	69.50	29.00	32.83	9.67	18.50	1.38	3.42	1.78
EH10656	IT7	wald	155.57	57.00	22.71	26.57	8.86	15.57	1.50	3.02	1.72
EH10657	IT7	wald	173.00	59.00	20.11	27.00	9.22	15.33	1.18	2.94	1.77
EH10659	IT7	BCwald	189.22	53.00	21.00	22.22	12.44	13.33	1.21	1.78	1.68
EH10662	IT7	BCwald	83.13	39.25	25.13	17.75	7.13	9.88	3.07	2.51	1.80
EH10663	IT7	BCwald	161.20	55.00	27.00	25.10	9.60	14.40	1.84	2.60	1.75
EH10664	IT9	wald	247.00	59.80	13.40	24.80	13.60	14.20	0.61	1.82	1.75
EH10665	IT9	BCwald	255.75	69.50	34.75	31.88	11.50	17.75	1.47	2.78	1.82
EH10667	IT9	wald	260.57	65.86	22.43	28.57	13.14	17.14	1.08	2.22	1.68
EH10668	IT9	wald	217.60	59.60	28.80	24.40	11.40	13.80	1.41	2.16	1.80
EH10669	IT9	wald	172.40	62.20	31.00	29.20	8.60	16.00	1.85	3.47	1.84
EH10674	IT9	F2hyb	266.43	64.43	43.00	27.71	13.71	14.43	1.71	2.02	1.94
EH10675	IT9	F2hyb	237.11	59.33	42.56	25.78	11.67	13.33	2.04	2.23	1.94
EH10676	IT9	F2hyb	148.00	50.10	28.50	22.10	9.80	11.20	2.00	2.26	1.98
EH10678	IT9	wald	315.00	73.20	36.80	32.60	13.90	16.20	1.19	2.35	2.03
EH10680	IT9	BCwald	157.56	58.56	29.67	27.33	8.56	16.00	1.94	3.22	1.71
EH10681	IT9	wald	138.50	49.60	27.20	22.30	9.10	11.80	2.09	2.45	1.89
EH10684	IT8	F2hyb	231.43	66.29	38.29	30.29	10.71	16.86	1.90	2.91	1.81
EH10685	IT8	F2hyb	290.25	63.50	41.00	26.50	14.75	14.25	1.84	1.84	1.84
EH10687	IT8	F2hyb	263.10	65.40	33.20	28.50	13.30	15.50	1.29	2.15	1.84
EH10688	IT8	F2hyb	173.75	58.75	30.50	27.00	9.63	13.63	1.78	2.81	2.00
EH10689	IT8	F2hyb	201.75	57.25	42.00	25.00	11.75	12.75	2.16	2.13	1.98
EH10690	IT8	F2hyb	286.00	75.10	40.90	34.50	12.30	18.30	1.51	2.83	1.90
EH10692	IT8	F2hyb	188.20	61.80	47.60	28.20	9.40	16.80	2.84	2.99	1.67
EH10694	IT8	F2hyb	169.22	57.22	31.78	26.44	9.33	14.22	2.03	2.85	1.86
EH10695	IT8	F2hyb	105.29	45.43	26.71	20.71	7.86	11.00	2.61	2.63	1.88
EH10697	IT8	F2hyb	153.90	47.30	20.90	19.40	11.00	11.40	1.44	1.76	1.70
EH10698	IT8	F2hyb	346.50	72.30	27.70	30.40	15.60	16.00	0.85	1.96	1.90
EH10703	IT5	foet	134.60	51.40	32.00	23.80	7.70	13.60	2.77	3.09	1.76
EH10704	IT5	*N/A*	151.70	49.90	30.10	22.00	10.00	11.40	2.10	2.20	1.93
EH10705	IT5	*N/A*	147.80	50.70	36.60	22.70	9.20	12.50	2.56	2.47	1.83
EH10706	IT5	*N/A*	86.20	39.40	27.70	17.60	7.00	9.70	3.31	2.53	1.82
EH10709	IT5	*N/A*	149.50	49.70	27.50	21.60	9.90	11.80	1.88	2.18	1.84
EH10712	IT5	*N/A*	128.40	46.50	31.90	20.60	9.00	11.50	2.52	2.30	1.80
EH10713	IT5	*N/A*	164.10	53.20	35.90	23.80	9.90	12.60	2.26	2.41	1.90
EH10714	IT5	*N/A*	195.90	57.10	36.10	25.20	11.40	13.10	2.05	2.19	1.91
EH10715	IT5	*N/A*	158.60	52.20	29.00	23.00	9.80	12.20	1.85	2.36	1.89
EH10716	IT5	foet	142.70	48.70	31.80	21.40	9.50	11.50	2.29	2.25	1.86
EH10717	IT5	*N/A*	182.50	54.90	30.00	23.80	10.90	11.90	1.68	2.19	2.01
EH10832	CH4	foet	118.50	44.13	42.75	19.13	8.63	8.63	4.18	2.21	2.24
EH10833	CH4	foet	201.50	60.50	59.50	27.20	10.30	12.10	3.11	2.64	2.26
EH10834	CH4	foet	148.90	54.20	49.50	24.70	9.00	12.90	3.49	2.74	1.95

EH10835	CH4	foet	159.50	58.30	59.00	26.80	8.40	12.60	3.76	3.19	2.16
EH10837	CH4	foet	90.40	40.30	31.50	17.90	7.00	9.70	3.53	2.57	1.85
EH10840	CH4	foet	147.90	51.90	45.20	23.30	9.00	12.70	3.11	2.60	1.84
EH10841	CH4	foet	98.29	43.71	45.00	19.71	6.86	10.29	4.80	2.88	1.94
EH10843	CH4	foet	109.13	48.38	40.38	22.25	7.63	11.75	3.73	2.92	1.90
EH10844	CH4	foet	249.20	67.60	56.70	30.10	12.00	13.20	2.40	2.54	2.33
EH10846	CH4	foet	191.40	62.00	60.80	28.40	9.40	14.60	3.41	3.03	1.96
EH10962	CH3	foet	121.00	43.38	40.50	18.25	9.00	9.63	3.38	2.03	1.93
EH10963	CH3	foet	148.88	48.75	56.13	21.00	9.88	10.38	4.03	2.13	2.03
EH10964	CH3	foet	130.38	43.38	30.75	17.75	10.25	8.63	2.54	1.76	2.10
EH10966	CH3	foet	78.20	36.00	44.50	15.70	7.40	8.20	5.78	2.12	1.92
EH10967	CH3	foet	215.50	58.25	43.75	24.75	12.50	12.50	2.13	1.99	1.98
EH10968	CH3	foet	53.25	29.75	43.25	12.75	5.50	6.50	8.28	2.33	1.97
EH10970	CH3	foet	237.89	60.00	36.89	24.89	12.78	13.22	1.67	1.96	1.91
EH10973	CH3	foet	78.60	35.60	35.60	15.60	7.20	8.20	4.64	2.19	1.91
EH10974	CH3	foet	178.00	56.20	55.70	25.20	10.00	12.50	3.23	2.53	2.03
EH10975	CH3	foet	101.71	40.00	44.29	17.00	9.57	9.00	4.55	1.97	1.89
PM20_006	SI1	wald	270.50	64.83	10.67	27.17	13.83	16.17	0.39	1.98	1.71
PM20_007	SI1	wald	188.13	55.13	1.25	23.75	11.63	12.13	0.06	2.08	1.97
PM20_008	SI1	wald	149.90	49.80	3.20	21.80	10.30	11.70	0.24	2.14	1.87
PM20_010	SI1	wald	287.25	67.25	5.13	28.50	14.25	15.50	0.16	2.02	1.84
PM20_011	SI1	wald	377.50	74.33	4.00	30.67	17.33	16.17	0.11	1.78	1.91
PM20_012	SI1	wald	256.00	68.33	16.89	30.56	11.89	16.89	0.70	2.58	1.84
PM20_014	SI1	wald	283.29	74.43	10.71	33.86	12.86	17.71	0.38	2.70	1.92
PM20_016	SI1	wald	291.17	74.00	23.67	33.83	13.17	19.17	0.86	2.66	1.77
PM20_021	SI1	wald	173.80	54.90	13.00	24.40	11.60	13.20	0.72	2.23	1.86
PM20_025	SI1	wald	188.80	58.40	8.60	26.40	10.40	13.80	0.45	2.57	1.92
PM20_031	SI2	wald	194.88	51.50	23.63	21.75	12.38	13.00	1.30	1.75	1.65
PM20_034	SI2	wald	219.70	58.50	10.50	24.90	12.80	12.90	0.48	1.96	1.94
PM20_035	SI2	wald	237.20	61.00	12.60	25.40	14.00	16.20	0.53	1.81	1.59
PM20_036	SI2	wald	170.00	53.40	9.80	23.40	10.80	14.00	0.59	2.17	1.67
PM20_038	SI2	wald	247.67	65.22	15.67	29.11	12.56	18.00	0.83	2.34	1.63
PM20_042	SI2	wald	204.50	55.70	0.00	23.70	12.90	13.10	0.00	1.85	1.82
PM20_043	SI2	wald	613.80	94.60	41.00	39.70	21.00	22.70	0.67	1.94	1.75
PM20_044	SI2	wald	220.70	59.40	17.80	25.70	12.90	13.80	0.82	2.00	1.86
PM20_045	SI2	wald	189.17	56.33	1.83	24.50	11.33	15.17	0.09	2.16	1.63
PM20_046	SI2	wald	231.33	62.33	2.33	27.00	12.33	15.83	0.13	2.20	1.71

7.8. Pearson's correlation for morphometric traits and altitude

Table S5: Matrix of calculated Pearson's correlation indices for measured morphometric traits and altitude.

	Altitude	Area	Perimeter	No. of teeth	Length	Width	Widest point	Teeth/cm	Length : Width	Length : Widest point
Altitude	1.00	-0.31	-0.22	0.63	-0.16	-0.41	-0.21	0.52	0.32	0.21
Area	-0.31	1.00	0.92	-0.14	0.83	0.92	0.78	-0.61	-0.19	-0.13
Perimeter	-0.22	0.92	1.00	-0.11	0.98	0.77	0.92	-0.65	0.14	-0.15
No. of teeth	0.63	-0.14	-0.11	1.00	-0.06	-0.29	-0.21	0.71	0.29	0.48
Length	-0.16	0.83	0.98	-0.06	1.00	0.64	0.94	-0.60	0.32	-0.14
Width	-0.41	0.92	0.77	-0.29	0.64	1.00	0.61	-0.69	-0.51	-0.14
Widest point	-0.21	0.78	0.92	-0.21	0.94	0.61	1.00	-0.64	0.29	-0.47
Teeth/cm	0.52	-0.61	-0.65	0.71	-0.60	-0.69	-0.64	1.00	0.20	0.36
Length : Width	0.32	-0.19	0.14	0.29	0.32	-0.51	0.29	0.20	1.00	-0.02
Length : Widest point	0.21	-0.13	-0.15	0.48	-0.14	-0.14	-0.47	0.36	-0.02	1.00

8. Acknowledgements

This Master's thesis is a product of a long time of effort and many wonderful people.

I would like to thank both supervisors, especially Prof. Dr. Elvira Hörandl, who has provided me with this project already back in 2020 and accepted me to the Department of Systematic Botany at the University of Göttingen. Here I have gained valuable experience in laboratory work, bioinformatics, and publishing of scientific work. Thank you for this wonderful challenge, the opportunity and for advising me throughout these years.

The next person I am extremely thankful for is Dr. Natascha Wagner, who is currently working on the *Salix* project in the department. She has mentored me from the first steps in the lab until the final moments before submission and has a special place in my heart as an exceptionally valuable teacher, advisor, and friend. It is her and Dr. Salvatore Tomasello, another dear friend and an excellent teacher, whom I have to thank for inviting me to participate in a project that turned out to be my first scientific publication.

Loïc Pitet has joined the *Salix* project a year ago and we've been friends ever since. He has helped me with challenges in bioinformatics and always tried to keep up my spirits when I was struggling. I see it now Loïc, it is finally here. The end, I mean.

I would also like to thank all the other members of the department for all the knowledge and great time we have spent together in the last two years. I will miss all of you and the pleasant atmosphere of the department.

If it wouldn't have been for Niklas and the Basler family, who were my first connection with Germany, I would have probably never started my studies in the charming city of Göttingen. Even though things didn't turn out the way we once thought they would, you will always remain an important piece of my life.

I am also indescribably grateful for my parents and the rest of my family, who have supported me through my studies, both financially and emotionally. Thank you for enabling me to study in a foreign country and for all the love you are giving me.

Aleksander, thank you for being my partner and my best friend. Your love and willingness to stay by my side despite the distance have made it all so much more unforgettable. I am glad I am closing this chapter of my life with you next to me.

And finally, I would like to acknowledge the wonderful city of Göttingen and every friend I've met here. The experience of studying here, though sometimes tough, was nevertheless beautiful.

I am submitting this thesis for my birthday, and it is probably the best gift I could have given myself.

9. Declaration of academic honesty

I, Pia Marinček, declare that I have authored the thesis

"**Evolution of a hybrid zone of willows (*Salix* L.) in the Alps analysed by RAD-seq and morphometrics"**

independently, that I have not used other that the declared sources/resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Göttingen, 20.05.2022 .....