universität wien

MASTERARBEIT / MASTER'S THESIS

Title of the Master's Thesis

„Phylogenetic relationships of the *Galium lucidum* complex in the Eastern Alps"

verfasst von / submitted by

Lukas Paul Grossfurthner, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science
MSc

Vienna, 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 832
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Master's degree programme Botany
Betreut von / Supervisor:	Assoz. Prof. Mag. Dr. Gerald Schneeweiss

Table of Contents

Abstract

Alpine endemic species are great study systems to infer the influence of Pleistocene glaciation on evolution and distribution of mountain plants. A particularly well-suited group is *Galium* (Rubiaceae). This genus is represented by 46 species in the Alps, ten of which are endemic to this mountain range. Many of the widespread *Galium* species show big variation in morphological characters and ploidy level, which renders it difficult to assess their relationships to range-restricted species that may be anything from clearly distinct species to mere ecotypes of widespread species. In this study, phylogenetic relationships of two endemics from the north-eastern Alps – the diploid *G. truniacum* and the tetraploid *G. meliodorum* – to the tetraploid, but widespread *G. lucidum* are investigated. Thus, we test the hypothesis, implied by current taxonomy, that each of these species represents a distinct lineage, separated from each other by ploidy (diploid *G. truniacum* versus tetraploid *G. meliodorum* and *G. lucidum*) and/or by ecology (high elevation *G. meliodorum* versus low elevation *G. lucidum*). Applying RAD-sequencing data on individuals with known ploidy, determined using flow-cytometry, we found a clear separation into two groups corresponding to ploidy level, but no separation within the tetraploid group. The strong morphological and ecological differentiation between the two tetraploids might be due to few adaptive loci or elevated phenotypic plasticity. Additionally, we found signals of shared ancestry between ploidy levels. Therefore, a re-evaluation of the taxonomic status of *G. meliodorum* is proposed.

Zusammenfassung

Alpine Endemiten sind hervorragend geeignet, um den Einfluss pleistozäner Vergletscherungen auf die Evolution und Verbreitung alpiner Pflanzen zu untersuchen. Eine besonders gut geeignete Gruppe stellt die Gattung *Galium* (Rubiaceae) dar, die mit 46 Arten, zehn davon endemisch, in den Alpen vertreten ist. Viele der weit verbreiteten *Galium*-Arten sind hinsichtlich Morphologie und Zytologie sehr variabel. Das erschwert die Einschätzung der Verwandtschaftsverhältnisse von engerräumig verbreiteten Sippen, die gut abgrenzbare Arten oder auch nur Ökotypen weitverbreiteter Arten sein können. In der vorliegenden Arbeit werden die Verwandtschaftsverhältnisse der beiden Nordostalpen-Endemiten, dem diploiden *G. truniacum* und dem tetraploiden *G. meliodorum,* gegenüber der weit verbreiteten,

ebenfalls tetraploiden Art *G. lucidum* untersucht. Wir testeten die auf der aktuellen Taxonomie basierende Hypothese, dass es sich bei allen drei Sippen um eigenständige Arten handelt, die entweder durch unterschiedliche Ploidiestufen (diploides *G. truniacum* gegen tetraploide *G. lucidum* und *G. meliodorum*) oder durch deren Ökologie (obermontan-subalpines *G. meliodorum* gegen submontan-montanes *G. lucidum*) getrennt sind. Mit Hilfe von RAD-Sequenzdaten von Individuen bekannter Ploidiestufe, ermittelt durch Durchflusszytometrie, konnten wir eine klare Trennung durch die Ploidiestufe, aber nicht durch die Ökologie nachweisen. Die dennoch gut ausgeprägte morphologische und ökologische Differenzierung innerhalb der tetraploiden Gruppe könnte durch wenige adaptive Loci oder erhöhte phänotypische Plastizität begründet sein. Zudem konnten Hybridisierungen sowohl zwischen den Ploidiestufen als auch innerhalb der tetraploiden Gruppe nachgewiesen werden. Aus diesem Grund wird eine Re-Evaluierung des taxonomischen Status von *G. meliodorum* vorgeschlagen.

Keywords: Narrow endemics*, Galium lucidum* complex, polyploidy, RAD sequencing

Introduction

Range-restricted species are suitable study systems to infer evolutionary-biogeographic processes such as range shifts and adaptation to new habitats imposed by Pleistocene climate oscillations. A geographical model system for addressing such questions are the European Alps (hereinafter referred to simply as Alps). Although for the Alps survival in interior refugia (nunataks) has been shown for some species (Stehlik *et al.,* 2002; Escobar Garcia *et al.,* 2012), the majority of plants are expected to have survived in ecologically and environmentally more stable peripheral glacial refugia. Species, whose current distribution is associated with these refugia, may be remnants of a formerly wider distribution (relics) or may have differentiated from their more widely distributed ancestor in phases of geographic isolation during Pleistocene range shifts (Tribsch & Schönswetter, 2003; Comes & Kadereit, 1998; Rabitsch & Essl, 2009).

An excellent system to address such questions is the *Galium lucidum* complex (Rubiaceae), which in the eastern Alps consists of the three species *G. lucidum*, *G. meliodorum* and *G. truniacum*. Whereas *G. lucidum* is widespread, the latter two species are restricted to the north-eastern limestone Alps: *Galium meliodorum,* distributed between Schneeberg and

Hochschwab, and *G. truniacum*, distributed between Ötscher and Berchtesgaden Alps (Kästner & Ehrendorfer, 2016). In the context of studying the *G. mollugo* complex, to which the *G. lucidum* complex belongs, Krendl (1967) used morphometric analyses and chromosome counts to delimit species within this complex and to infer their relationships. He found *G. truniacum* to be diploid (2*n*=2*x*=22) and *G. meliodorum* as well as *G. lucidum* to be tetraploid (2*n*=4*x*=44). Morphometric analysis revealed good separation between the species, but Krendl (1967) also reported overlapping traits, which could indicate hybridisation between both diploids and tetraploids as well as between tetraploids, particularly in contact areas such as Ötscher. Nevertheless, ecological and morphological differences as well as different evolutionary hypotheses for each entity led Krendl (1967) to treat them as species, a taxonomy that is still accepted, although questioned by Kästner and Ehrendorfer (2016). Pleistocene glaciations are thought to have played a major role in shaping the distribution of these taxa. Specifically, diploid *G. truniacum* is considered to be an ancestor with relictual distribution in or close to the prominent peripheral refugium in the north-eastern Alps (Merxmüller, 1952, 1953, 1954; Schönswetter *et al.,* 2005) of the tetraploid taxa of the *G. lucidum* complex (Krendl 1967). Concerning *G. meliodorum*, even more strongly associated with the north-eastern Alpine refugium than *G. truniacum*, two contrasting evolutionary hypotheses exist. Krendl (1967) considered *G. truniacum* as a likely ancestor of *G. meliodorum,* thus implicitly suggesting ploidy level change as a main driver of divergence, whereas Kästner and Ehrendorfer (2016) consider the latter species to be a postglacially evolved alpine ecotype of the widespread tetraploid *G. lucidum*, thus implicitly suggesting ecological differentiation as main driver of divergence. However, none of these hypotheses have been tested so far.

Here we investigate the phylogenetic relationships of members of the *G. lucidum* complex from the eastern Alps, allowing the relative roles of ploidy shift and ecology for lineage divergence to be assessed. To this end, we applied cytological methods (flow-cytometric ploidy level determination calibrated by chromosome counts) as well as a state-of-the-art molecular-genetic method, RAD sequencing (Miller *et al.,* 2007, Baird *et al.,* 2008). Specifically, we tested (i) whether the three species are genetically distinct lineages as implied by current taxonomy and (ii) whether there is gene flow among these species, particularly in areas of sympatry, as suggested by overlapping morphological traits.

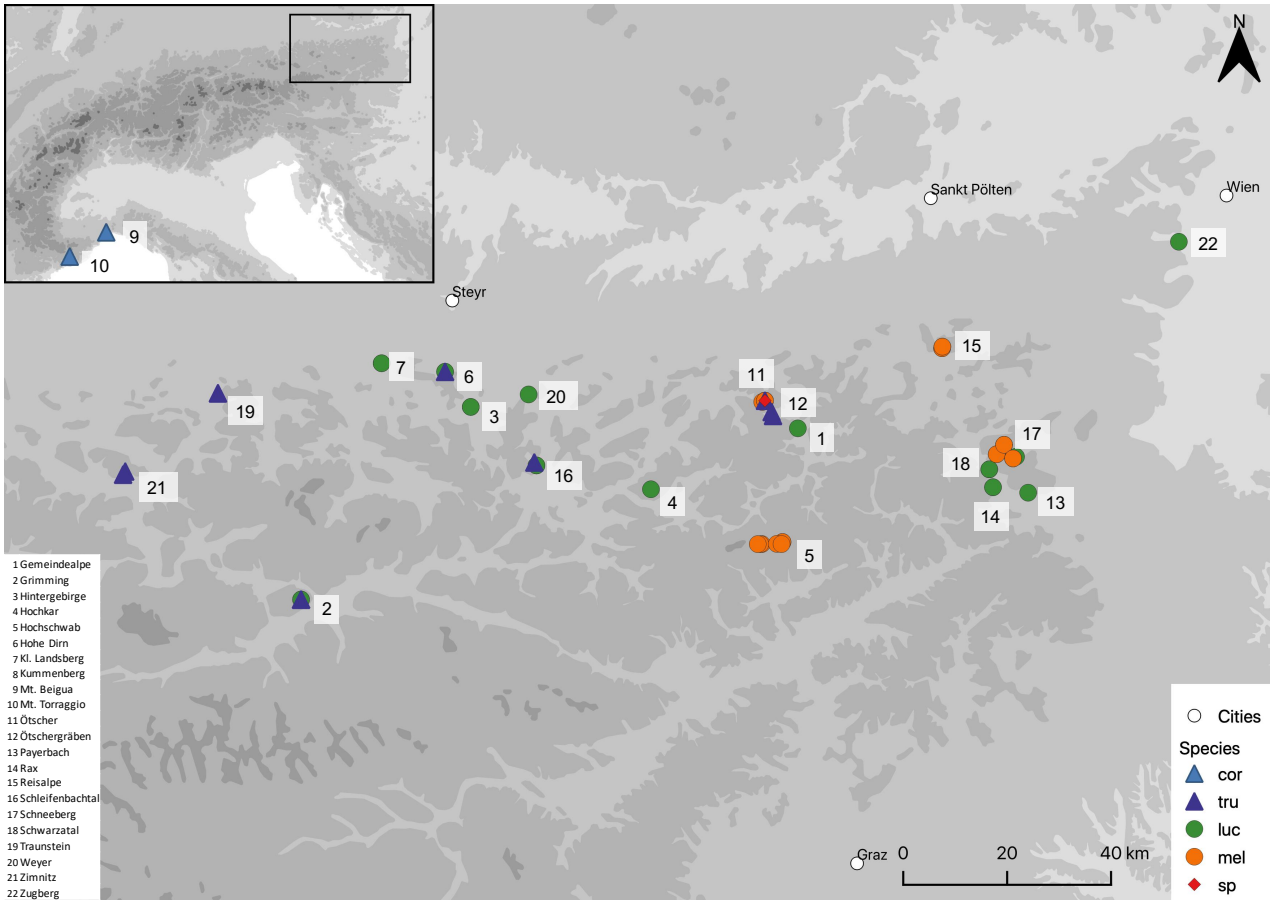


Figure 1: Sampling sites; Abbreviations: cor = G. corrudifolium, tru = G. truniacum, luc = G. lucidum, mel = G. meliodorum, sp = sp11_1

Table 1: Sampling sites (location and geographical coordinates) and ploidy level estimations.

Location	LocID	range	Lat	Lon	Ploidy Level, Relative Flourescence[a]±CV
Gemeindealpe	1		47°48'54.77"	15°17'17.92"	4*x* : RF: 1.371±2.8
Grimming	2		47°31'24.67"	14°1'49.28"	2*x* : RF: 0.654±2.57, RF: 0.665±2.34; 4*x* : 1.457±2.39
Hintergebirge	3		47°51'5.59"	14°27'36.13"	4*x* : RF:1.440±3.48
Hochkar	4		47°42'41.91"	14°54'58.52"	4*x* : RF: 1.398±4.23
Hochschwab	5	from to	47°37'18.17" 47°37'5.44"	15°14'58.06" 15°11'35.96"	4*x* : RF_{DWHK}[b]: 1.336±3.14, RF_{P}: 1.336±4.52, RF_{JG}: 1.398±5.08
Hohe Dirn	6		47°54'40.58"	14°23'41.93"	2*x* : RF: 0.6629±2.89; 4*x* : 1.4034±2.64
Kl. Landsberg	7		47°55'32.38"	14°14'3.63"	-
Mt. Beigua	9		44°25'40.45"	8°37'7.57"	2*x* : RF: 0.6910±9.24
Mt. Torraggio	10		43°58'7.11"	7°39'18.52"	2*x* : RF: 0.668±9,09; RF: 0.676, CV±9.18
Ötscher	11	from to	47°51'47.90" 47°51'18.30"	15°12'35.99" 15°11'9.77"	2*x* : RF_{RK}: 0.660±6.54, RF: 0.654±3.97; 4*x* : RF_{RK}: 1.242±3.73, RF_{AS}: 1.32±2.58
Ötschergräben	12	from to	47°50'17.40" 47°51'8.89"	15°13'24.90" 15°17'17.84"	2*x* : RF: 0.660±3.49, RF: 0.656±3.11
Payerbach	13		47°42'20.51"	15°52'16.82"	4*x* : RF: 1.361±3.61
Rax	14		47°42'54.08"	15°46'56.67"	-
Reisalpe	15	from to	47°57'13.79" 47°57'3.97"	15°39'15.21" 15°39'12.56"	4*x* : RF: 1.408±3.69; RF:1.353±2.52; RF: 1.390±2.05
Schleifenbachtal	16	from to	47°45'51.45" 47°45'7.55"	14°36'23.87" 14°37'35.08"	2*x* : RF: 0.656±4.5, RF: 0.641±3.31; 4*x* : 1.427±2.93, RF:1.430±3.29
Schneeberg	17	from to	47°47'16.60" 47°46'16.25"	15°48'38.45" 15°47'30.04"	4*x* : RF_{HS}: 1.369±2.71, RF_{KR}:1.396±3.58, RF_{AS}: 1.358±2.9 RF: 1.409±4.18 RFFS: 1.4321±3.98, RFFSM: 1.367±4.06
Schwarzatal	18	from to	47°44'43.87" 47°44'5.92"	15°46'22.86" 15°47'32.82"	4*x* : RF: 1.377±2.96
Traunstein	19	from to	47°52'29.38" 47°51'48.25"	13°49'11.77" 13°49'14.48"	-
Weyer	20		47°52'21.97"	14°36'23.61"	4*x* : RF: 1.414±2.94
Zimnitz	21		47°44'11.75"	13°34'46.90"	2*x* : RF: 0.652±5.46, RF: 0.639±3.33, RF: 0.648,±3,37
Zugberg	22		48°7'52.47"	16°15'11.03"	4*x* : RF: 1.4180±2.03

a: Relative flourescence based on 3 to 6 pooled samples and given as G1 peak ratio, comparing the objects with the internal standard Solanum pseudocapsicum.

b: Abrbreviations of sample pools: Location ID 5: DWHK=Dullwitz+Höllkampl, P=Parkplatz, JG=Jochgraben; 11: RK=Rauer Kamm, AS=Abstieg; 17: HSM=Herminensteig mel, KRm=Krumme Riess mel, AS=Abstieg FSM= Fadensteig mel

Methods

Sampling and DNA Isolation

Species were identified using the determination key in Fischer, Oswald & Adler (2008) as well as the detailed descriptions and drawings of Kästner & Ehrendorfer (2016). Additional diagnostic characters, which were recognized during fieldwork, and the currently used diagnostic characters are summarised in Table 1. Diagnostic characters and habitus as well as typical habitats are shown in Fig. 2.

Plant material from *Galium lucidum, G. truniacum, and G. meliodorum* was collected at 21 sites throughout the Austrian northern limestone Alps; as outgroup the Mediterranean *G. corrudifolium*, considered closely related to *G. lucidum* (Krendl, 1967), was collected at two localities in the Ligurian Alps in Italy (Fig. 1). The plant material was dried and stored in silica gel. Vouchers are deposited at WU. Wherever possible, living plants were wrapped in moist tissue for subsequent flow cytometric analysis. The dataset consisted of 69 individuals, ranging from one to six individuals per species per site. Samples were collected at different altitudes, therefore, distances between individuals ranged from 0.1 km (Location 7) up to 8 km (Location 5).

For DNA extraction, c. 20 µg of silica gel dried leaves were frozen with liquid nitrogen and ground for two cycles of 2 min with Tissue Lyser II (Qiagen, Hilden, Germany). Secondary metabolites were removed during an overnight incubation in 400 µl sorbitol wash buffer (Russel *et al.,* 2010) at 4°C; subsequently, two additional sorbitol washing steps, comprising 10 min incubation on ice and centrifugation at 4°C at 10.000 rpm, were performed. The final steps of DNA extraction were performed with the Stratec Invisorb Spin Plant Mini Kit (Stratec Molecular, Berlin, Germany) following the manufacturer's instruction. DNA was quantified with the Qubit dsDNA HS Assay Kit (Thermofisher Scientific, Waltham, MA, USA) following the manufacturer's protocols.

Genome size and ploidy level

About one cm^2 of fresh leaf tissue was chopped together with fresh leaf tissue from the species used as internal standard, i.e., *Pisum sativum* (Greilhuber & Ebert, 1994) or *Solanum pseudocapsicum* (Praca-Fontes *et al.,* 2011), in 550 µl Otto's Buffer I (Otto, 1990). Before filtering the cell wall remnants through a 30 µm mesh, additionally 550 µl Otto's Buffer I were

added. To clean the filtered nuclei isolate from RNA, the solution was treated with 50 µl RNase A and incubated for 30 min at 37°C. Approximately 400 µl of the resulting solution were transferred and stored as backup, the remaining solution was stained with propidium iodide (PI) in Otto's Buffer II (Otto, 1990) and incubated at 4°C for at least 60 min. Genome size was calculated based on the fluorescence intensity of 9,999 particles of fresh plant material of three individuals each per species. 1C-values were calculated using the following formula (Doležel & Bartoš 2005):

$$1C\text{-value}_{object} = \left(\frac{\text{Mean G1object}}{\text{Mean G1standard}}\right) * 1C\text{-value}_{standard}$$

Ploidy level (from fresh material) and DNA ploidy level (from silica-gel dried material, prepared the same way as aforementioned for fresh material) were estimated by pooling up to five individuals of the same species and location and comparing their relative fluorescence. Measurements were repeated on single individuals if an odd G1:G2 peak ratio and a high coefficient of variation (CV; >5% for fresh and >10% for silica-gel dried material, thresholds determined based on initial measurements) occurred. All measurements were performed in a CyFlowML flow cytometer (Partec, Münster, Germany).

Chromosome analysis

Root tips of *Galium lucidum, G. truniacum and G. meliodorum* were taken from living plants collected in additional field trips in winter 2018 (Hohe Dirn, Schneeberg, and Zugberg) and cultivated in the experimental garden of the University of Vienna.

Root tips were washed in tap water and treated with 8-hydroxyquinolin (0.002mol/l 8-hydroxyquinolin). After 2.5 hr incubation at room temperature and 2.5 hr at 4°C in darkness, the samples were washed with tap water and fixed in a 3:1 ethanol:glacial acetic acid for 1 hr. For chromosome staining, root tips were incubated in 5 M HCl at room temperature for 18 min, subsequently washed with water, and finally incubated for 60 min in Schiff's reagent. Stained root tips were transferred to a slide with 60% acetic acid, dissected with preparation needles, and squashed with rubber gum to properly disperse cells. Pictures were taken with a Zeiss Axiocam 702 mono (Zeiss, Oberkochen, Germany) and edited with Zeiss-ZEN software 2.3 (Zeiss, Oberkochen, Germany). From the same individuals, genome size was estimated as

described above allowing the correspondence of distinct levels of relative fluorescence with karyologically confirmed ploidy levels to be established.

RAD-seq library preparation

A single-digest RAD library was prepared using the PstI restriction enzyme (New England Biolabs, Ipswitch, MA, USA) following the protocol of Paun *et al.,* (2016) with minor modifications. Specifically, to account for dosage compensation, half the DNA amount was used for the diploids as for the tetraploids, i.e., 100 ng for the diploid species and 200 ng DNA for the tetraploid species. After an overnight ligation of 2 µl 100 nM P1 adapters for the diploid species and 4 µl 100 nM P1 adapters for the tetraploid species, shearing by sonication was performed with a Bioruptor Pico (Diagenode Inc., Denville, NJ, USA) with three cycles, each of 30 s "on" and 60 s "off", at 4°C. Prior to sequencing, fragment size was selected using Pippin Prep (Sage Science, Beverly, MA, USA) aiming for fragments between 220 bp to 800 bp long and evaluated with Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). The library was sequenced at VBCF Vienna (https://www.vbcf.ac.at/facilities/next-generation-sequencing/) on an Illumina HiSeq 2500 using 100 bp single-end reads.

Table 2: Summary of the determining characters. Asterisks indicate characters newly found in this study.

G. truniacum	***G. meliodorum***	***G. lucidum***
Leaves narrow, thin, margin smooth	Leaf narrow, fleshy, margin smooth	Leaf thin, margin with emergences facing forward. Fig. 1G
Leaves of the main axis sometimes withered, in that case with numerous vegetative shoots*.	Leaves of the main axis similar to G. *lucidum*, but without emergences, leaves of the vegetative shoots narrow and bent tipwards* (Fig. 1D)	Leaves of the main axis and the vegetative shoots do not differ in appearance
Panicle narrow, few flowering shoots.	Panicle narrow cone shaped, many flowers	Panicle narrow cone shaped, many flowers

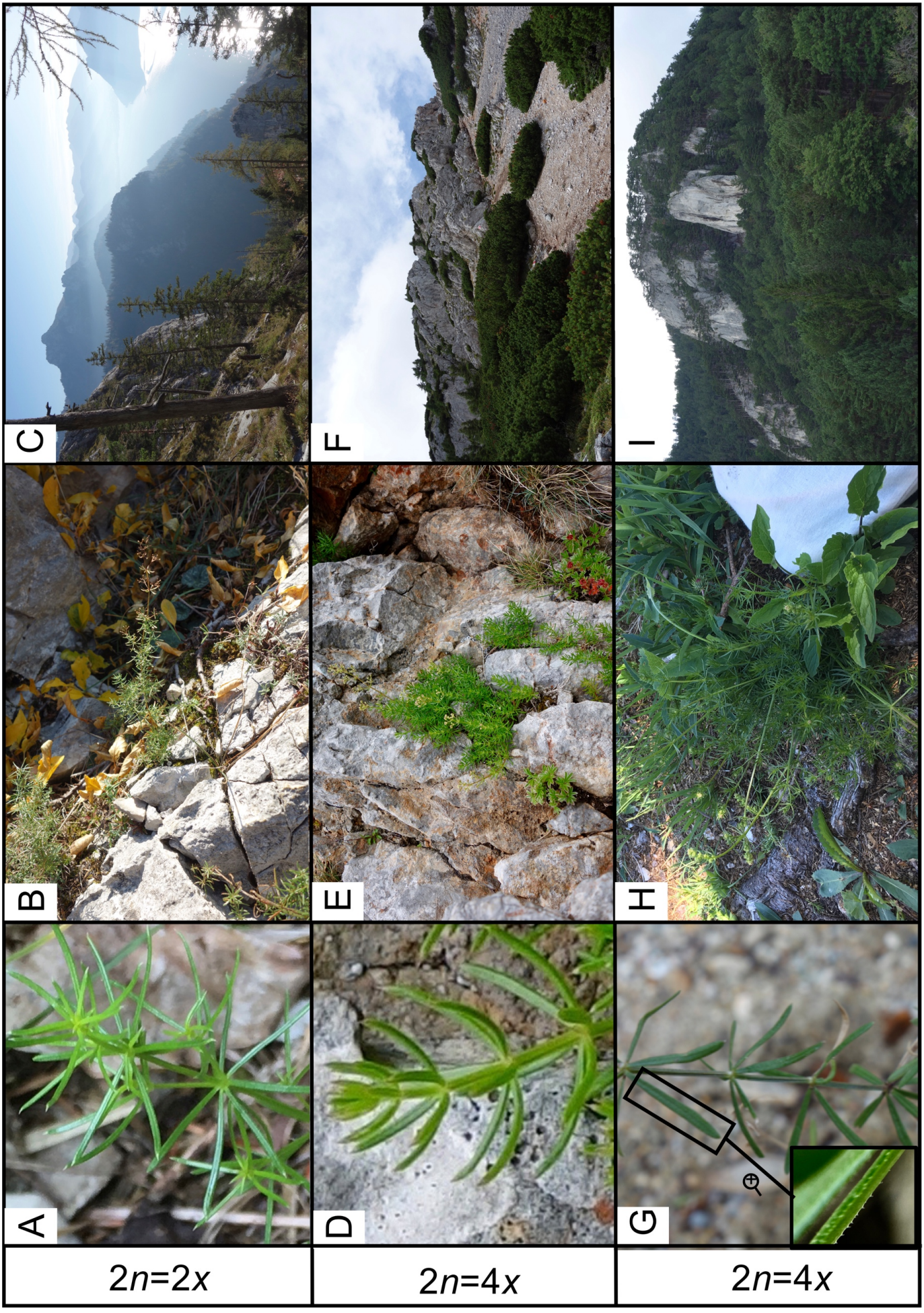


Figure 2 A-C: Galium truniacum, leave characters, habitus and habitat. D-F: Galium meliodorum, G-I: Galium lucidum, G: emergences magnified

Identification of RAD-loci and SNP calling

Raw Illumina reads were first demultiplexed based on index barcodes using BamIndexDecoder (available from https://github.com/wtsi-npg/illumina2bam), allowing for barcodes with maximally two mismatches to expectations. Subsequently, further demultiplexing based on inline barcodes and filtering for uncalled bases (-c) and low quality (-q) defined by the quality score phred33 (-E) was performed with process_radtags, from STACKS 1.48 (Catchen, 2013).

Loci were initially assembled using STACKS' denovo_map.pl wrapper, requiring a minimum coverage of five reads to create a stack (-m), allowing for up to one mismatch between stacks within individuals to create a locus (-M) and up to one mismatch between stacks between individuals (-n). Testing several settings, the 5-1-1 setting was eventually chosen as it resulted in the highest number of loci with between one and ten SNPs that were covered in at least 60% of the individuals.

To increase the information content, a pseudoreference was constructed for further use in a mapping approach. Briefly, based on the *de novo* built loci, a haplotype consensus was extracted with export_sql.pl for those polymorphic loci covered in at least half of the individuals (pare_l=35) and containing maximally ten (snps_l=1 and snps_u=10) SNPs in order to prevent combining paralogs into loci (Paun *et al.,* 2016).

Following the approach by Heckenhauer *et al*. (2018), samples were mapped against the pseudoreference using the -M option of BWA MEM 0.7.15 (Li & Durbin, 2010), followed by sorting the bam files by reference coordinates, as well as addition of read groups with picard tools 2.17.11 (available from https://broadinstitute.github.io/picard/). Finally, the bam files were realigned around indels using the Genome Analysis Toolkit 3.8.1 (McKenna *et al*., 2010).

Two analytical approaches have been further used to extract polymorphic positions from the mapping files. First, a genotype-based approach as implemented in STACKS' ref_map.pl was used to call SNPs with default settings. In this genotype-based approach, we prepared a whitelist of loci with export_sql.pl with the same settings as above. STACKS' populations software was further used to mildly filter the dataset for a maximum of 0.65 observed heterozygosity (--max_obs_het), to remove any putative cases of combined paralogs and to calculate the number of private alleles and nucleotide diversity π, as well as to generate a vcf output file. The resulting vcf file was further filtered with vcftools (Danecek *et al.,* 2011) for a minor allele frequency of 0.3, 5% missing data and removing indels. Heterozygosity and F statistics were calculated in vcftools (argument --het) and plotted in R (R Core Team 2016)

using the ggplot2 package (available from https://cran.r-project.org/web/packages/ggplot2/index.html).

Two datasets were prepared: One, hereinafter called *all5miss*, consisting of the ingroup taxa and the outgroup *G. corrudifolium*, and the second, hereinafter called *excor5miss*, containing only the ingroup taxa. File format conversions into software specific formats were performed with PGD spider 2.1.1.4 (Lischer & Excoffier, 2012).

As an alternative to the genotype-based approach of STACKS, genotype posterior probabilities were estimated directly from the mapping files with ANGSD v. 0.921-26 (Korneliussen *et al.,* 2014) based on a Genome Analysis Toolkit (GATK) model. This genotype-free approach is particularly recommended for processing low to intermediate NGS datasets (Korneliussen *et al.,* 2014). As it retains genotype uncertainties in further analyses, this approach may be more suitable for analyzing mixed ploidy datasets. The settings used for the genotype-free inference in ANGSD allowed for a maximum of 50% missing data in order to process a position, a mapping quality threshold of 20 (minMapQ 20), a base quality threshold of 20 (minQ = 20) and retained only variable positions with a high likelihood (p-value < 1e-6).

Data analyses

Differences in observed heterozygosity (H_O), the inbreeding coefficient F, nucleotide diversity π, and private alleles among the investigated taxa (*G. lucidum*, *G. meliodorum*, *G. truniacum*) were tested for using Mann-Whitney U tests with Bonferroni corrections employing SPSS 24.0 (IBM, Armonk, NY, USA).

Phylogenetic analysis was performed using RAxML AVX 8.2.10 (Stamatakis, 2014). Using a streamline editor (sed -e), all polymorphic loci of the *all5miss* dataset were removed as required for the ascertainment bias correction, resulting in a dataset including 1105 SNPs. The analysis was performed using the GTR-CAT model, as suggested to be used for datasets over 50 Individuals (Stamatakis, 2014), Lewis ascertainment bias correction and 1000 rapid bootstrap replicates. The result was visualized in FigTree 1.4.3 (available from http://tree.bio.ed.ac.uk/software/figtree/). In order to visualize reticulations between groups and to identify putative hybrids, for both datasets, NeighborNet networks were generated with SplitsTree 4.14.6 (Huson and Bryant, 2006), using uncorrected P distances.

A recently developed approach, fineRADstructure (Malinsky *et al.,* 2018), was performed to gain further insight into the ancestry of the individuals by generating a co-ancestry matrix. Advantages of this software are its capability to work with haplotype data and that it deals with polyploid data. Haplotypes present in 98% of the individuals (pare_l=63) were extracted using export_sql.pl from STACKS.

Finally, to assess clustering patterns among accessions, the genotype posterior probability estimated in ANGSD were used to perform a PCA analysis with PCAngsd v. 0.95 (Meisner and Albrechtsen 2018) and the resulting eigenvectors were plotted in R v. 3.4.4.

Results

Genome size and chromosome counts

DNA amounts (1C-values) are 0.8468 pg for the diploid *G. truniacum*, 1.8275 pg for the tetraploid *G. lucidum* and 1.7992 pg for the tetraploid *G. meliodorum*. The 1C*x* values are 0.9137 for *G. lucidum* and 0.8996 for *G. meliodorum*. The CVs for the measured genome size values are in all cases below the 3% threshold. In a second set of individuals not included in the genetic analysis, but used for chromosome number determination, genome size is 1C=0.8558 pg for *G. truniacum,* 1.8108 pg for *G. lucidum,* and 1.8067pg for *G. meliodorum*. Thus, the 1C*x* values are 0.9054 for *G. lucidum* and 0.9033 for *G. meliodorum*. The differences between the independent measurements are lower than differences that occurred within a single individual.

During ploidy level estimation of pooled samples, occasionally high CVs of up to 9% occurred. Re-measuring each of these individuals separately shows that high CVs in a sample pool (>5% on a single re-measured fresh individual or >8% on re-measured dried individuals) can be caused by a single individual. Nevertheless, the relative position of the standard G1 peak always allows the correct assignment of the ploidy levels. The highest CVs are always found in dried material.

Chromosome numbers are 2*n*=2*x*=22 for *G. truniacum* (Fig. 3a), and 2*n*=4*x*=44 for *G. lucidum* (Fig. 3b) and 2*n*=2*x*=44 for *G. meliodorum* (figure not shown).

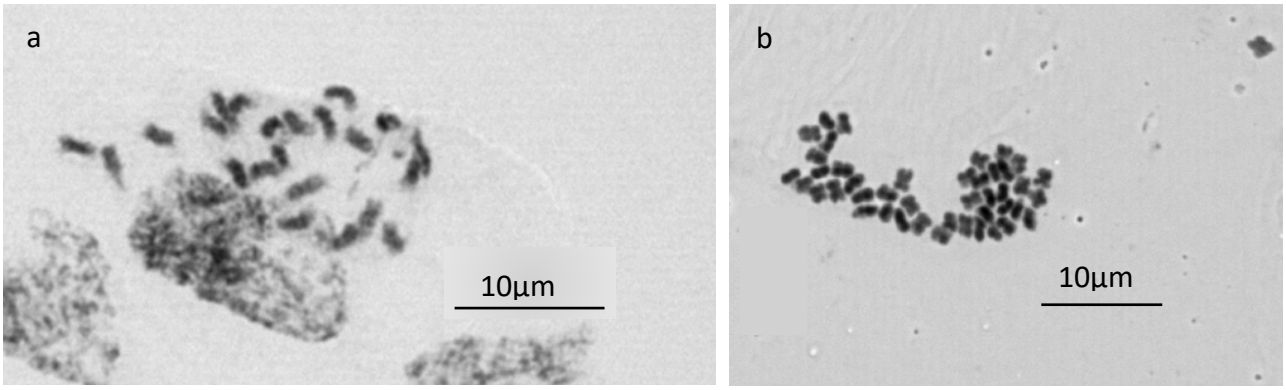


Figure 3: Mitotic chromosomes from (A) G. truniacum (2n=2x=22) from Hohe Dirn, Upper Austria, and (B) from G. lucidum (2n=4x=44) from Zugberg, Lower Austria.

Sequencing and SNP calling

After demultiplexing and filtering, an average of 1.29±0.9 million high quality reads are retained per sample. The number of retained reads varies across samples from 333,954 (tru16_2) to 2,900,098 (mel5_2). After optimization of *de novo* catalog building and filtering as indicated in the material and methods section, 5,000 polymorphic loci of 94 bp each are retained in the synthetic reference. On average, the mapping success to this reference ranges from 52.21% to 58.27%, resulting in final coverage of an average of 20.02 reads.
After filtering for missing data, removing indels, and accounting for minor allele frequency, the dataset *all5miss* contains 5,828 SNPs and the dataset *excorr5miss* contains 6,537 SNPs. The genotype-free approach implemented in ANGSD yields information at 290,629 variable sites after filtering.

Data analyses

In comparison to diploid *G. truniacum*, the observed heterozygosity is higher in the tetraploids *G. lucidum* and *G. meliodorum* as well as in the unidentified individual sp11_1, resulting in an elevated inbreeding coefficient F in *G. truniacum* (Fig. 4). Diversity measures (observed heterozygosity, π) as well as private alleles are significantly higher in the tetraploid taxa than in the diploid ($p<0.001$; Table 3). The per-individual inbreeding coefficient F is also significantly negative and takes lower values in the tetraploids than in the diploids. Differences between the tetraploid taxa are statistically not significant (H_O: p=0.074, F: p=0.063, π: p=0.173, private alleles: p=0.434).

The maximum likelihood tree shows a clear and well supported separation of the ingroup and the outgroup as well as of the ploidies (Fig. 5a). The diploid group has moderate support and indicates geographic clustering, but with very low bootstrap support. The backbone of the tetraploid group shows neither any geographic clustering nor any bootstrap support. Although it is tetraploid, individual sp11_1 falls in the diploid group.
The NeighborNet network of the *all5miss* dataset is consistent with the maximum likelihood tree and shows a clear split between ingroup and outgroup as well as between the two ploidy levels (App. 1). The NeighborNet network of the *excorr5miss* dataset reveals a clear separation

of the two ploidy levels, yet with some individuals of both *G. lucidum* and *G. meliodorum* and the individual sp11_1 showing splits connecting them to *G. truniacum* (Fig. 5b).

The first three eigenvectors of the PCA built based on genotype likelihoods explain 8.9% of the variation and indicate a clear separation by ploidy, but no structure within the tetraploids (Fig. 5c).

Table 3: Summary of diversity measurements. Given are mean values of respective species.

Group	H_o	*F*	π	Private Alleles
G. truniacum	0.0451	0.6042	0.0315	253
G. lucidum	0.1535	-0.3518	0.0934	1213
G. meliodorum	0.1399	-0.2317	0.0885	1109
sp11_1	0.1194	-0.0504	0.0817	753

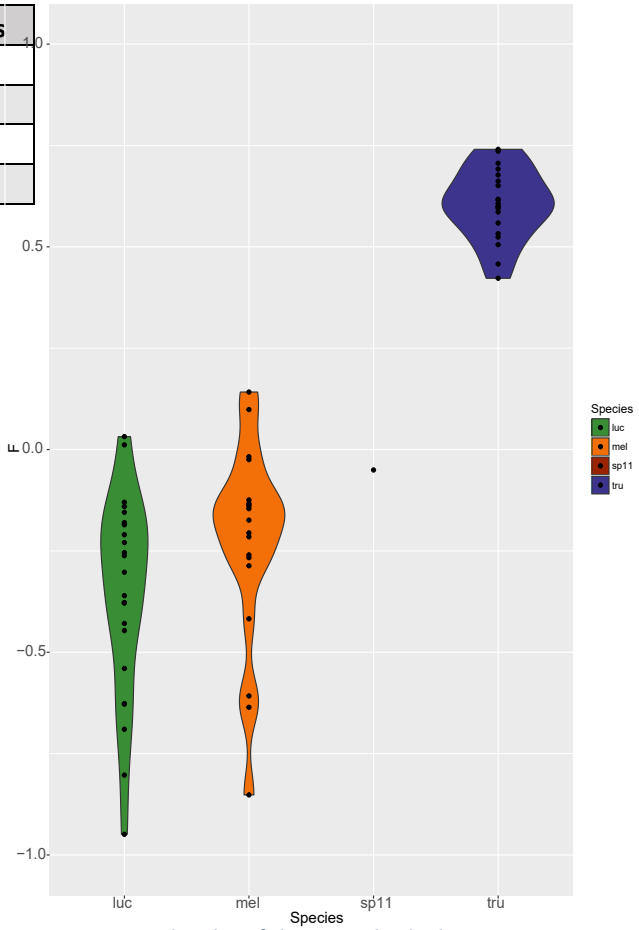


Figure 4: Violin plot of the per-individual inbreeding coefficient F.

Analysis of the data set *excorr5miss* using fineRADstructure supports the clear separation of the di- and tetraploid groups, although most tetraploid individuals have some shared ancestry with diploids and evidence for admixture is found in at least four tetraploid individuals (sp11_1, mel11_1, mel5_6, mel15_2; Fig. 6). There is clear geographic structure within the diploid group (individuals from the same populations show higher co-ancestry), whereas this is not the case for the tetraploids, where neither populations nor taxa (*G. lucidum* versus G. *meliodorum*) form cohesive groups. Within the tetraploid individuals, a group comprising seven individuals that have the lowest levels of co-ancestry, is found (Fig. 6), but it does not show any geographic or taxonomic structure.

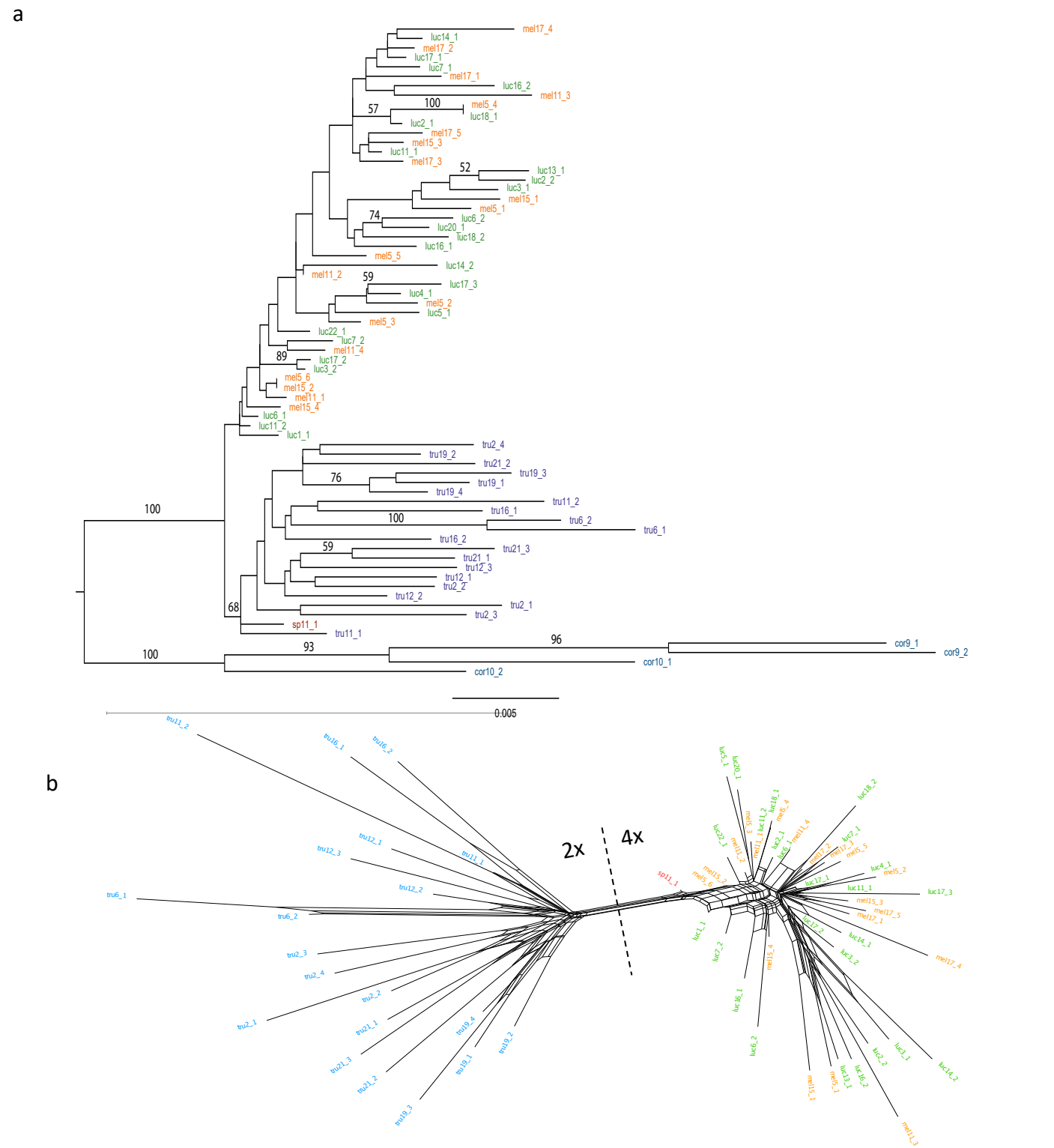


Figure 5. A: Maximum likelihood tree, B: NeighborNet Network C:and D: PCA based on genotype free inference

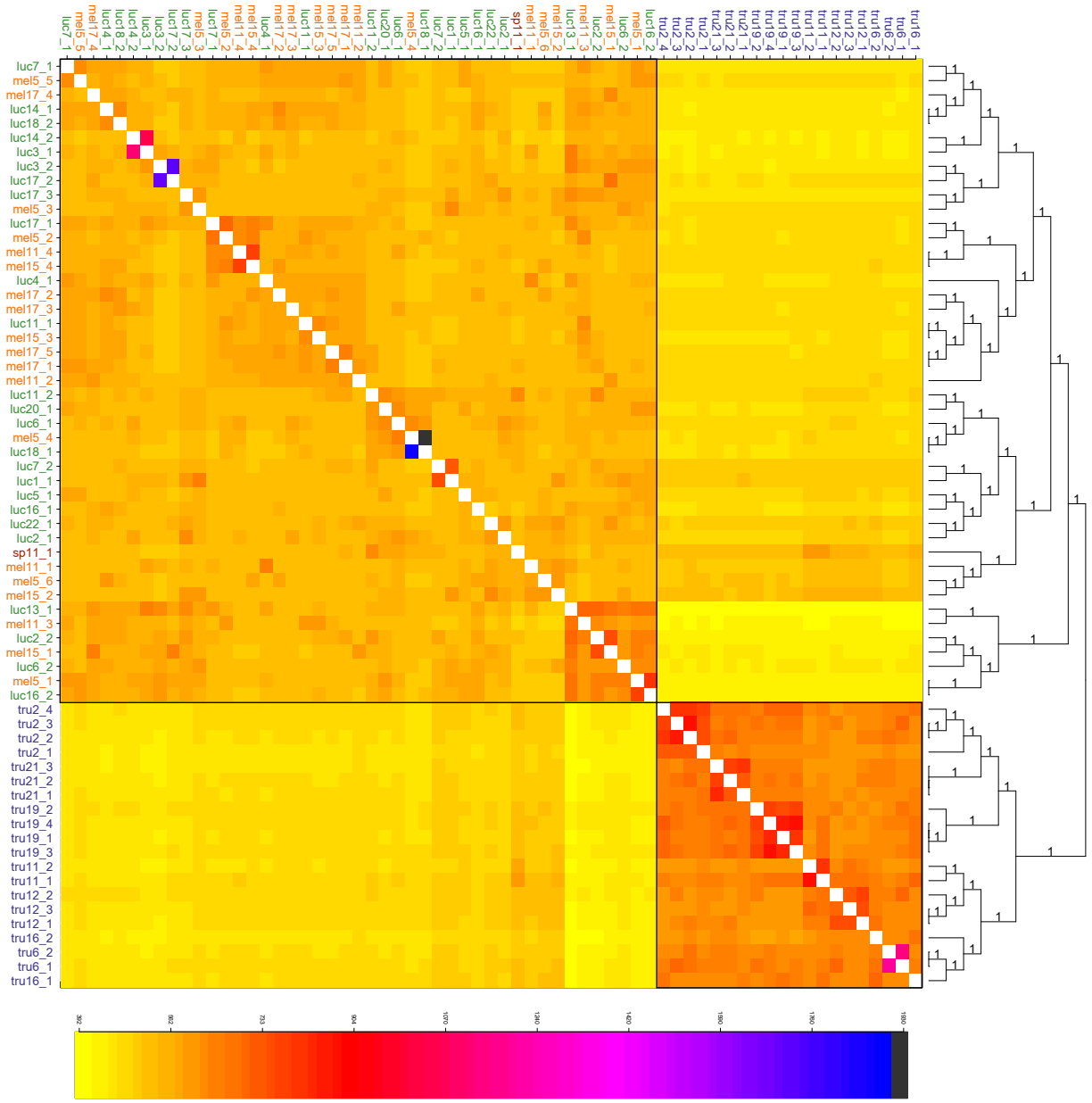


Figure 6: fineRADstructure plot: ploidies are separated by solid lines.

Discussion

Clear species boundary by ploidy level

The studied individuals of *G. truniacum*, *G. lucidum,* and *G. meliodorum* fall into two well separated groups differing in their ploidy, i.e., diploids and tetraploids. This separation is consistently inferred by all analyses, ranging from the maximum likelihood analysis (Fig. 5a) with the lowest amount of data due to the removal of all ambiguous sites, to the principal component analysis of the largest data set generated by ANGSD (Fig. 5c, d). Ploidy levels are recognized as an important factor in conferring reproductive isolation (Petit, Bretagnolle & Felber, 1999. Carrió & Güemes, 2014). Separation may have been additionally enhanced by geographic isolation. The distribution range of *G. truniacum* only partially overlaps with that of the tetraploids, but this overlap may have been smaller during Pleistocene glacial periods, forcing *Galium* species into peripheral refugia. Such a historical component could also explain the strong geographic structure among populations in *G. truniacum* (Fig. 5b, 6). Nevertheless, there is evidence for gene flow from diploids into tetraploids (position of some tetraploid individuals in the NeighborNet network and elevated co-ancestry of some tetraploid individuals: Figs. 5b, 6). Gene flow into tetraploids is easy to achieve via unreduced gametes (Kreiner, Kron & Husband, 2016; Kolar, 2017).

The lacking separation of *G. lucidum* and *G. meliodorum* does not support the evolutionary hypothesis of Krendl (1967), who hypothesized that *G. lucidum* and *G. meliodorum* were of independent origin from *G. truniacum*. Instead, our data are consistent with the hypothesis of Kästner & Ehrendorfer (2016), that *G. meliodorum* is a high elevation derivative of *G. lucidum*. This should also be reflected in the used taxonomic ranks. Originally described as *G. lucidum* var. *meliodorum* by Beck (1884), it was raised to species rank by Fritsch (1909); based on our genetic data, but also taking the clear morphological and ecological differences into account, *G. meliodorum* is best to be treated as subspecies of *G. lucidum*, as already proposed by Kästner & Ehrendorfer (2016).

The relationship of *G. truniacum* to the tetraploids cannot be clearly resolved by our data. The elevated co-ancestry between diploids and most tetraploid individuals, which has been revealed by fineRADstructure, may indicate (recent) admixture between the two ploidy levels. However, the elevated co-ancestry might also suggest a contribution of *G. truniacum* to the

formation of the tetraploid group, thus supporting the hypothesis of an allopolyploid origin (Krendl, 1967) of the tetraploids.

Ecological differentiation despite shallow genetic differentiation

Despite the high genetic similarity of *G. meliodorum* to *G. lucidum*, the question remains why the very specific morphological and ecological traits of *G. meliodorum* (Table 1) are (only) restricted to populations in the prominent Pleistocene refugium in the northeastern limestone Alps. The morpho- and ecotype of *G. meliodorum* may be determined by only a few loci, which will remain undetected in a genome-wide SNP data set obtained from RADseq loci. The great influence of a few adaptive loci on phenotype and thus, ecological separation, is known from Darwin finches (Chaves *et al.,* 2016). Elevated phenotypic plasticity in polyploids might also lead to rapid adaptation and formation of distinct morphological traits as it was reported from a European dry grassland species (Hahn, van Kleunen & Müller-Scherrer, 2018).

This does, however, not explain why *G. meliodorum* is restricted to the northeastern limestone Alps. Possibly, Pleistocene range shifts and geographic separation during cold periods permitted increased divergence between genotypes, also under selection pressure of alpine habitats not encountered elsewhere. Although thus differentiated populations likely experienced gene flow during warm periods, as is evidently the case now, repeated cycles of isolation might have been sufficient for differentiation in those loci that are responsible for the morphological and ecological differences. Following the scheme of Wu (2001), the studied tetraploids of the *G. lucidum* complex might have reached stage II – that is, transition between race and species. At this stage, both fusion or division are still possible. Since the majority of the genes in stage II are expected to be undifferentiated (Wu, 2001), negatively affecting the power to detect differences using genome-wide RAD-seq derived SNP data, the acquired differences might be weakened by (re-)established gene flow or might be maintained by strong selection pressure in mountain environments.

Morphological plasticity and hybridization

All members of the *G. lucidum* complex show morphological variation under different ecological conditions (Krendl, 1967). This morphological plasticity increases the risk of misidentifications, particularly with narrow-leaved *G. album* (Kästner & Ehrendorfer, 2016). This might be the case for a group of seven tetraploid individuals showing low levels of shared ancestry with *G. truniacum* (Fig. 6b), although including individuals (luc2_2, luc6_2 and mel11_3) which were growing in closest proximity to *G. truniacum*. Morphological re-evaluation of these individuals shows that they have a decreased leaf length : leaf width ratio, putting the leaf shapes closer to those observed in the closely related *G. mollugo* complex. Therefore, these seven individuals might be introgressed from tetraploid *G. album*. Krendl (1967) suspects that *G. album* does hybridize with *G. lucidum* in contact areas. Indeed, on the sampling sites, *G. album* has been observed close by.

Future research

To gain further insights in the phylogenetic relationships as well as the origin and evolution of the *Galium lucidum* complex, extended sampling within this complex (especially diploid *G. montis-arerae*, a local endemic of the southern Alps, and *G. cinereum* from the southwestern Alps) and inclusion of closely related species (the *G. mollugo* complex, comprising diploid *G. mollugo* and tetraploid *G. album* and the *G. verum* complex), that do hybridize with members of the *G. lucidum* complex, will be necessary. In depth analysis of the gradient within the *G. lucidum* s.l. using both genetic tools (like RNAseq) and reciprocal transplantation experiments should help to understand the nature and genetic architecture underlying ecological and morphological differentiation in this plant group.

Acknowledgements

First and foremost, I want to thank Birgit Orthofer for providing transportation as well as help during field work and material preparation, without her, for me, the thesis would not have been possible.

I am thankful for inspiring discussions and for solving determination problems with Luise and Friedrich Ehrendorfer as well as Clemens Pachschwöll for support and insights in biogeography of refugial endemics.

Furthermore, I want to thank Norma Rivera for lab briefing and procedure optimization as well as Hanna Schneeweiss, Eva M. Temsch, Michael H.J. Barfuss, Elfriede Grasserbauer and Da Pan for advices in cytogenetic and molecular lab work.

I want to thank Ovidiu Paun for his great advices, endless time and help in data analysis and error diagnostics, which without him would have not been detected, as well as my brother Martin Grossfurthner for help statistical analysis and support during fieldwork.

Last but not least I want to thank Gerald M. Schneeweiss for accepting me as a Master student and enabling such a resource consuming project.

References:

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

Beck-Mannagetta, G. (1884). Flora von Hernstein in Niederösterreich und der weiteren Umgebung. Holzhausen. Wien.

Baird N.A., Etter P.D., Atwood T.S., Currey M.C., Shiver A.L., Lewis Z.A., Selker E.U., Cresko W.A., Johnson E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Mol Ecol, 22: 3124-3140. doi:10.1111/mec.12354

Carrió, E., Güemes, J. (2014). The effectiveness of pre- and post-zygotic barriers in avoiding hybridization between two snapdragons (*Antirrhinum* L.: Plantaginaceae), *Botanical Journal of the Linnean Society*, Volume 176, Issue 2, pp 159–172, https://doi.org/10.1111/boj.12201

Chaves, J. A., Cooper, E. A., Hendry, A. P., Podos, J., De León, L. F., Raeymaekers, J. A., MacMillan, W. and Uy, J. A. (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. Mol Ecol, 25: 5282-5295. doi:10.1111/mec.13743

Comes H.P., Kadereit, J.W. (1998). The effect of Quaternary climatic changes on plant distribution and evolution. Trends Plant Sci 3: 432–438

Danecek, P., Auton A., Abecasis G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R., Lunter, G., Marth, G., Sherry, S.T, McVean, G., Durbin, R. and 1000 Genomes Project Analysis Group (2011). The Variant Call Format and VCFtools, Bioinformatics

Dolezel, J., & Bartoš, J. (2005). Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. Annals of Botany, 95(1), 99–110. http://doi.org/10.1093/aob/mci005

Escobar García, P., Winkler, M., Flatscher, R., Sonnleitner, M. , Krejčiková, J., Suda, J., Hülber, K., Schneeweiss, G. M. & Schönswetter, P. (2012). Extensive range persistence in peripheral and interior refugia characterizes Pleistocene range dynamics in a widespread Alpine plant species (*Senecio carniolicus*, Asteraceae). Molecular Ecology, 21: 1255-1270. doi:10.1111/j.1365-294X.2012.05456.x

Fischer, M.A., Oswald, K., Adler, W., (2008). Exkursionsflora für Österreich, Liechtenstein, Südtirol. 3. Auflage. Biologiezentrum der oberösterreichischen Landesmuseen, Linz.

Fritsch, K. (1909). Exkursionsflora für Österreich (mit Ausschluss von Galizien, Bukowina und Dalmatien). 2. Auflage. Verlag von Gerolds Sohn. Wien

Greilhuber, J., Ebert, I., (1994). Genome size variation in *Pisum sativum*. Genome 37: 646-655.

Hahn M.A., van Kleunen, M., Müller-Schärer, H. (2012). Increased Phenotypic Plasticity to Climate May Have Boosted the Invasion Success of Polyploid *Centaurea stoebe*. PLoS ONE 7(11): e50284. https://doi.org/10.1371/journal.pone.0050284

Heckenhauer, J., Samuel, R., Ashton, P.S., Salim, K.A., Paun, O. (2018). Phylogenomics resolves evolutionary relationships and provides insights T into floral evolution in the tribe Shoreeae (Dipterocarpaceae), Molecular Phylogenetics and Evolution 127 (2018) 1–13

Huson, D. H., and Bryant D., Application of Phylogenetic Networks in Evolutionary Studies, Mol. Biol. Evol., 23(2):254-267, 2006

Kadereit, J.W. (2017). The role of in situ species diversification for the evolution of high vascular plant species diversity in the European Alps—A review and interpretation of phylogenetic studies of the endemic flora of the Alps. Perspectives in Plant Ecology, Evolution and Systematics 26 (2017) 28-38

Krendl, F., (1967). Cytotaxonomie der *Galium mollugo*-Gruppe in Mitteleuropa (Zur Phylogenie der Gattung *Galium,* VIII). Österr. Bot. Zeitschrift, Bd. 114 H. 4/5

Li H. and Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics, Epub

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup (2009). The Sequence alignment/map (SAM) format and SAMtools, *Bioinformatics* 25(16) 2078-9

Lischer H.E.L. & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28: 298-299

Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data, *Molecular Biology and Evolution*, Volume 35, Issue 5, https://doi.org/10.1093/molbev/msy023

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky A., Garimella, K., David Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297-1303; doi:10.1101/

Meisner J., Albrechtsen A. (2018). Inferring population structure and admixture proportions in low depth NGS data. Genetics in press. doi:10.1534/genetics.118.301336

Merxmüller H. (1952). Untersuchungen zur Sippengliederung und Arealbildung in den Alpen. I. Jahrbuch des Vereins zum Schutze der Alpenpflanzen und –tiere. 17:96–133

Merxmüller H. (1953). Untersuchungen zur Sippengliederung und Arealbildung in den Alpen. II. Jahrbuch des Vereins zum Schutze der Alpenpflanzen und –tiere. 18:138–158

Merxmüller H. (1954). Untersuchungen zur Sippengliederung und Arealbildung in den Alpen. III. Jahrbuch des Vereins zum Schutze der Alpenpflanzen und –tiere. 19:97–139

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site as- sociated DNA (RAD) markers. Genome Ressour. 17, 240–248

Novembre, J., Pritchard, J.K, Stephens, M., and Donnelly, P., (2016) on population structure. Genetics 204, 391–393

Otto, F. (1990). DAPI staining of fixed cells for high resolution of flow Cytometry of nuclear DNA. Methods in Biology 33, Flow Cytometry, pp. 105- 110. San Diego, Calif.: Academic Press.

Paun, O., Turner, B., Trucchi, E., Munzinger, J., Chase, M. W., & Samuel, M. R. (2016). Processes Driving the Adaptive Radiation of a Tropical Tree (Diospyros, Ebenaceae) in New Caledonia, a Biodiversity Hotspot. Systematic Biology, 65(2), 212-227. DOI: 10.1093/sysbio/syv076

Pachschwöll, C., Winkler, M., Escobar García, P., Schneeweiss, G. M., & Schönswetter, P. (2017). Evolution of high mountain plant species in the Alps and Carpathians – the “hairy” case of the Doronicum clusii aggregate (Asteraceae). Studia Universitatis Babes-Bolyai Biologia, 62(Special Issue), 56-57

Praca-Fontes, M., Carvalho C.R., Clarindo W.R. Cruz, C.D, (2011). Revisiting the DNA C-values of the genome size-standards used in plant flow cytometry to choose the “best primary standards”. Plant Cell Reports 30/7: Pp. 1183-1191

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

Stamatakis, A., (2014). "RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies". *Bioinformatics* 30(9)*:* 1312–1313

Rabitsch, W., Essl, F., (2009). Endemiten: Kostbarkeiten aus Österreichs Pflanzen- und Tierwelt. Naturwissenschaftlicher Verein für Kärnten. Klagenfurt

Kästner A., Ehrendorfer, F., (2016). Gustav Hegi – Illustrierte Flora von Mitteleuropa VI Teil 2B/ Spermatophyta: Angiospermae: Dicotyledones 4 (2/2); VI/2B, Jäger, E.J., Weissdorn Verlag, Jena

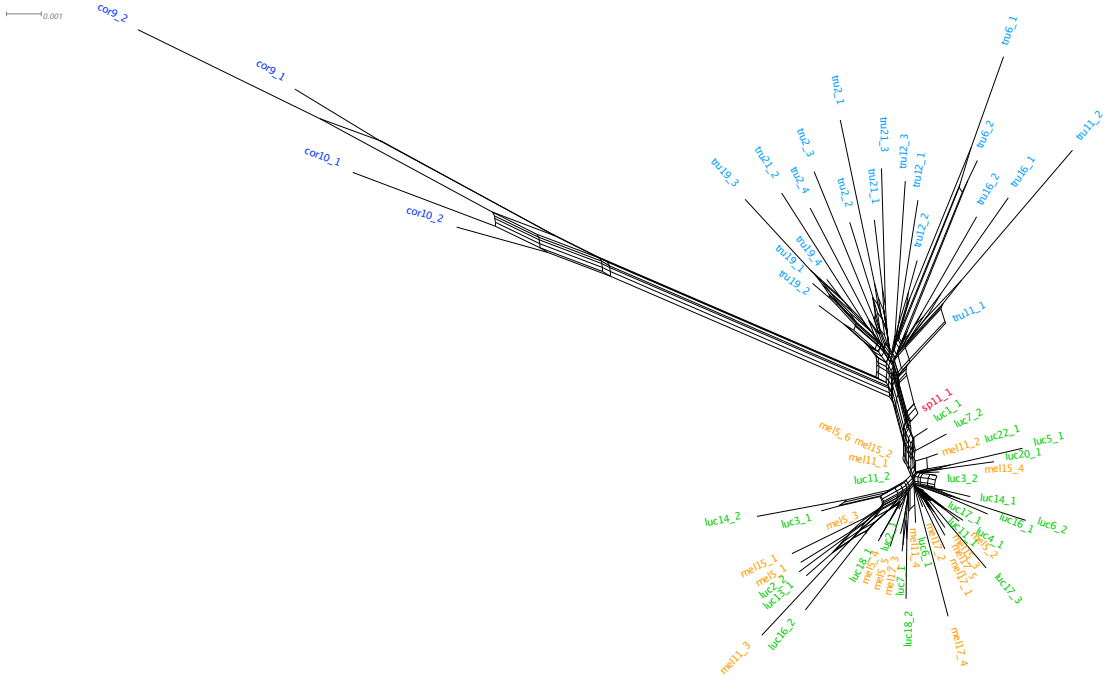
Korneliussen T.S., Albrechtsen A., Nielsen R. (2014). ANGSD: Analysis of next generation sequencing data. BMC Bioinf. 15:356

Russell, A., Samuel, R., Rupp, B. and Barfuss M.H.J. (2010). Phylogenetics and cytology of a pantropical orchid genus *Polystachya* (Polystachyinae, Vandeae, Orchidaceae): Evidence from plastid DNA sequence data. *Taxon* 59: 389- 404

Stehlik, I., Blattner, F. R., Holderegger, R. and Bachmann, K. (2002). Nunatak survival of the high Alpine plant Eritrichium nanum (L.) Gaudin in the central Alps during the ice ages. Molecular Ecology, 11: 2027-2036. doi:10.1046/j.1365-294X.2002.01595.x

Tribsch A, Schönswetter P (2003). Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for Pleistocene refugia in the Eastern Alps. Taxon 52: 477–497. PMID: 12857639

Appendix



Appendix 1: Neighbor-joining network of the all5miss dataset