

Snow White and Rose Red:
Studies on the contrasting evolutionary trajectories of the genera
Leucanthemum Mill. and *Rhodanthemum* B.H.Wilcox & al.
(Compositae, Anthemideae)



DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES DER
NATURWISSENSCHAFTEN (DR. RER. NAT.) DER FAKULTÄT FÜR
BIOLOGIE UND VORKLINISCHE MEDIZIN DER
UNIVERSITÄT REGENSBURG

vorgelegt von

Florian Wagner

aus Burgstall (Mitwitz)

Juli 2019

Das Promotionsgesuch wurde eingereicht am:

12.07.2019

Die Arbeit wurde angeleitet von:

Prof. Dr. Christoph Oberprieler

Unterschrift:

.....

Florian Wagner

Abstract

Plant systematics, the study of taxonomy, phylogeny and evolutionary processes in plants has undergone considerable progress in the last decades. The application of modern molecular approaches and DNA-sequencing techniques in the field has led to an improved inventory of biodiversity and a better understanding of evolutionary processes shaping the biological diversity on our planet. The increased availability of molecular and genomic data has particularly facilitated the investigation of shallowly diverged and taxonomically complex taxon-groups, which is challenging due to minor morphological differences, low genetic differentiation and/or hybridization among taxa. The present thesis investigates species delimitation, hybridization and polyploidization in the recently diverged genera *Leucanthemum* Mill. and *Rhodanthemum* B.H. Wilcox & al. of the subtribe Leucanthemeinae K.Bremer & Humphries (Compositae, Anthemideae) by applying Sanger-, 454-pyro-, and restriction site associated DNA (RAD) sequencing, as well as AFLP-fingerprinting and morphometric analyses. The first two parts are focusing on species delimitation and hybridization in the closely-knit taxon groups around *L. ageratifolium* Pau and *R. arundanum* B.H. Wilcox & al., respectively. Various analyses based on AFLP fingerprinting, RADseq and multi-locus sequence data demonstrate that the robustness of species delimitation results is considerably influenced by the intensity of hybridization among species and the number of hybrid individuals included. Therefore, a step-by-step approach is performed in both studies, with an initially step of identification and subsequent removal of hybrid individuals, followed by application of different species-delimitation methods. This strategy results in the reliable identification of independent species, subspecies and nothospecies in both taxonomically complex plant groups. The third part of the present thesis compares the contrasting evolutionary trajectories of diploid representatives of both genera in a more comprehensive phylogenetic study. Specific hypotheses for the formation of polyploids in plants are proposed and evaluated to find factors that promote polyploidization in certain plant groups (e.g., *Leucanthemum*) and not in others (e.g., *Rhodanthemum*). Multi-locus sequence data from 127 accessions of the subtribe Leucanthemeinae unveil a significantly higher genetic divergence and hybridization signal among diploid lineages of *Leucanthemum* compared to *Rhodanthemum*, in spite of a similar crown age and diversification pattern during the Quaternary. The study demonstrates the importance of genetic differentiation among diploid progenitors and their concurrent affinity for natural hybridization for the formation of a polyploid complex. Furthermore, the role of climate-induced range overlaps on hybridization and polyploid speciation during the Quaternary is discussed.

References of published and submitted manuscripts

The underlying thesis is composed of the following published or submitted manuscripts. The proposed nomenclatural changes/novelities are not intended being effectively published in the present thesis.

- A. **Wagner F, Härtl S, Vogt R, Oberprieler C. 2017.** 'Fix Me Another Marguerite!': Species delimitation in a group of intensively hybridizing lineages of ox-eye daisies (*Leucanthemum* Mill., Compositae-Anthemideae). *Molecular Ecology* **26**: 4260–4283.
- B. **Wagner F, Ott T, Schall M, Lautenschlager U, Vogt R, Oberprieler C.** Taming the Red Bastards: Hybridization and species delimitation in the *Rhodanthemum arundanum*-group (Compositae-Anthemideae). *Molecular Phylogenetics and Evolution* doi: 10.1016/ympev.2019.106702.
- C. **Wagner F, Ott T, Zimmer C, Reichhart V, Vogt R, Oberprieler C. 2019.** 'At the crossroads towards polyploidy': Genomic divergence and extent of homoploid hybridization are drivers for the formation of the ox-eye daisy polyploid complex (*Leucanthemum* Mill., Compositae-Anthemideae). *New Phytologist* **223**: 2039–2053.

In the course of my PhD, I contributed to further publications, which are not part of the thesis:

- D. **Konowalik K, Wagner F, Tomasello S, Vogt R, Oberprieler C. 2015.** Detecting reticulate relationships among diploid *Leucanthemum* Mill. (Compositae, Anthemideae) taxa using multilocus species tree reconstruction methods and AFLP fingerprinting. *Molecular Phylogenetics and Evolution* **92**: 308–328.
- E. **Oberprieler C, Wagner F, Tomasello S, Konowalik K. 2017.** A permutation approach for inferring species networks from gene trees in polyploid complexes by minimizing deep coalescences. *Methods in Ecology and Evolution* **8**: 835–849.
- F. **Hassanpour H, Zare-Maivan H, Sonboli A, Kazempour-Osaloo S, Wagner F, Tomasello S, Oberprieler C. 2018.** Phylogenetic species delimitation unravels a new species in the genus *Sclerorhachis* (Rech.f.) Rech.f. (Compositae, Anthemideae). *Plant Systematics and Evolution* **304**: 185–203.

- G. **Oberprieler C, Hassanpour H, Sonboli A, Ott T, Wagner F. 2019.** Multi-locus phylogenetic reconstructions reveal ample reticulate relationships among genera in Anthemideae subtribe Handeliinae (Compositae). *Plant Systematics and Evolution* doi: 10.1007/s00606-019-01588-0.
- H. **Oberprieler C, Schinhärl L, Wagner F, Hugot L, Vogt R.** Karyological and molecular-genetic analyses of *Leucanthemum* Mill. (Compositae, Anthemideae) in Corsica. Submitted for publication to *Willdenowia* (under review).

Personal contributions

Publication A

Florian Wagner (FW), Robert Vogt (RV), and Christoph Oberprieler (CO) conceived this study. FW and Sabine Härtl (SH) produced the sequence and AFLP fingerprint data, which were analyzed by FW. FW wrote a first draft of the study, which was complemented and partly rewritten by RV and CO.

Publication B

FW, RV, and CO conceived the present study and collected plant material. Maximilian Schall (MS) produced the Sanger sequence data, which were processed and analyzed together with the RADseq data by Tankred Ott (TO), FW, and Ulrich Lautenschlager (UL). A first draft of the paper was written by FW with input from RV and CO.

Publication C

FW, RV, and CO conceived the present study. FW, Claudia Zimmer (CZ) and Verena Reichhart (VR) produced the sequence data, which was processed by TO and FW and analysed by FW. A first draft of the paper was written by FW with input from RV and CO.

Contents

Abstract.....	v
References of published and submitted manuscripts	vi
Personal contributions.....	vii
Chapter 1: General introduction	1
1.1 Plant systematics in the era of next-generation sequencing.....	1
1.2 Species delimitation in the framework of the multi-species coalescent.....	2
1.3 What we still don't know about polyploidy.....	2
1.4 The subtribe Leucantheaminae	3
1.5 Thesis outline	4
Chapter 2: Fix Me Another Marguerite!	9
Abstract.....	9
2.1 Introduction.....	10
2.2 Materials and Methods.....	12
2.2.1 Plant material and DNA extraction	12
2.2.2 AFLP fingerprinting.....	13
2.2.3 Detection of potential hybrid individuals.....	17
2.2.4 Plastid and nuclear marker sequencing.....	18
2.2.5 Multiple sequence alignments and gene-tree reconstructions.....	18
2.2.6 MSC species-delimitation	19
2.3 Results.....	21
2.3.1 AFLP fingerprinting.....	21
2.3.2 Detection of potential hybrid individuals.....	22
2.3.3 Multiple sequence alignments and gene-tree reconstructions.....	25
2.3.4 MSC species delimitation	26
2.4 Discussion.....	30
2.4.1 Species delimitation and species concepts.....	30
2.4.2 Species delimitation and hybridization	32
2.4.3 Detection of hybridization patterns in the <i>L. ageratifolium</i> group.....	33
2.4.4 MSC species-delimitation	34
2.4.5 Phylogenetic considerations and taxonomic implications	36
2.5 Supplemental Figures and Tables	41
Chapter 3: Taming the Red Bastards	57
Abstract.....	57

3.1 Introduction	58
3.2 Materials and Methods	60
3.2.1 Taxon sampling and DNA extraction.....	60
3.2.2 ITS, ETS and plastid marker sequencing	64
3.2.3 Double digest restriction associated DNA (ddRAD) sequencing	64
3.2.4 Detection of hybrid individuals.....	66
3.2.5 Species delimitation analyses.....	67
3.3 Results	70
3.3.1 Sanger and ddRAD sequencing output	70
3.3.2 Comparison of multiple datasets to determine optimal parameter settings in IPYRAD	70
3.3.3 Detection of hybrid individuals.....	71
3.3.4 Species delimitation analyses.....	73
3.4. Discussion	77
3.4.1 Optimization of de-novo assembly parameters and evaluation of RADseq genotyping errors.....	77
3.4.2 Hybridization patterns in the <i>R. arundanum</i> -group.....	80
3.4.3 Evaluation of different species delimitation analyses	83
3.4.4 Conclusions and taxonomical/nomenclatural implications	84
3.5 Supplemental Figures and Tables	87
Chapter 4: At the crossroads towards polyploidy	101
Abstract	101
4.1 Introduction	102
4.2 Materials and Methods.....	105
4.2.1 Plant material and DNA extraction	105
4.2.2 Plastid and nuclear marker sequencing	105
4.2.3 Processing of 454 and Sanger sequence data	106
4.2.4 Multiple sequence alignments and model selection	107
4.2.5 Multi-species coalescent (MSC) species-delimitation	107
4.2.6 Inference of genetic divergence patterns.....	108
4.2.7 Inference of homoploid hybridization patterns	108
4.2.8 Divergence time estimation.....	110
4.3 Results	112
4.3.1 Multiple sequence alignments and model selection	112
4.3.2 MSC species delimitation.....	112
4.3.3 Genetic divergence patterns	112

4.3.4 Homoploid hybridization patterns.....	113
4.3.5 Divergence times estimation	114
4.4 Discussion.....	115
4.4.1 Does phylogenetic divergence drive polyploidization?	115
4.4.2 Homoploid hybridization in the evolution of <i>Leucanthemum</i> and <i>Rhodanthemum</i>	120
4.4.3 The role of climatic changes during the Quaternary	122
4.5 Supplemental Figures and Tables	124
4.6 Supplemental Methods and Notes.....	140
Chapter 5: Comprehensive summary, discussion and outlook.....	145
5.1 Comprehensive summary.....	145
5.2 Snow White, Rose Red and the seven veils	147
5.3 Outlook	149
References.....	151
Acknowledgments.....	174

Chapter 1: General introduction

1.1 Plant systematics in the era of next-generation sequencing

The field of (plant) systematics can be subdivided into three basic areas (Stuessy, 2014): (1) taxonomy, (2) the study of phylogeny, and (3) the study of processes of evolution. While taxonomy comprises the process of classification, i.e. grouping of individuals into taxa, the subsequent ranking and naming of taxa, and the identification of these, the study of phylogeny focuses on the mode, time and place of the divergence of a particular group. Finally, the study of evolutionary processes examine fundamental phenomena like differentiation of populations, speciation, and hybridization (Stuessy, 2009).

The present thesis covers all three fields of plant systematics by applying phylogenetic and species delimitation studies in the genera *Leucanthemum* Mill. and *Rhodanthemum* B.H. Wilcox & al. of the subtribe Leucantheaminae K.Bremer & Humphries (Compositae, Anthemideae): Species delimitation (field 1) and hybridization (field 3) within *Leucanthemum* and *Rhodanthemum* are in the focus of chapters 2 and 3, while phylogenetic relationships within and between both genera (field 2) and the search for polyploidy-promoting factors in *Leucanthemum* (field 3) are the scope of chapter 4.

Decreasing costs for DNA sequencing and the invention of next-generation sequencing (NGS) methods have undeniably influenced the field of plant systematics in the last decades (Hörandl and Appelhans, 2015). The investigation of relationships of closely related species was long dominated by the use of DNA fingerprinting (Weising et al., 2005; applied in chapter 2), and is more and more replaced by NGS-based ‘restriction site associated DNA’ (RAD) sequencing methods (Ree and Hipp, 2015; see chapter 3). Furthermore, the invention of high-throughput DNA-sequencing technologies like pyro-sequencing (Roche 454 system) or sequencing-by-synthesis (Illumina) has facilitated the generation of multi-locus sequencing data for phylogenetic and evolutionary studies of non-model organisms (as applied in chapter 4).

As a consequence, there is a continuous increase of the amount of molecular data used for answering current questions of plant systematics, which can be - on a small scale - retraced in the present thesis: While the first study (Wagner et al., 2017, chapter 2) is based on 207 Sanger sequences and 367 AFLP loci, the second paper (Wagner et al., 2019; chapter 4) includes already 77,067 quality filtered 454-sequencing reads. This abundance of data is even exceeded by the study described in chapter 3 (Wagner et al., under review), where a total of 485,075,916 quality filtered Illumina reads are used for delimiting species in the genus *Rhodanthemum*.

1.2 Species delimitation in the framework of the multi-species coalescent

Species delimitation is the process of determining the boundaries and numbers of species from empirical data (de Queiroz, 2007) and is in the focus of chapters 2 and 3 of the present thesis. While morphological characters have dominated the science of species delimitation for centuries, population genetics and phylogenetic methods are nowadays frequently applied to investigate species-level biological diversity (Choi, 2016). Molecular and genomic data sets are particularly useful for delimiting allopatrically distributed and morphologically similar, but distinctly differentiated populations of shallowly diverged plant groups (as it is the case in the here investigated genera *Leucanthemum* and *Rhodanthemum*). While early molecular species delimitation studies relied on single locus data and reciprocal monophyly or fixed differences among individuals as the main criteria for identifying species (Fujita et al., 2012), the focus has nowadays shifted to multi-locus sequence data, evaluated within the framework of the multi-species coalescent (MSC) (Rannala, 2015).

The MSC is a model of gene coalescence within a species tree and accounts for gene-tree incongruence (as an example see Figures S2.6-S2.10 of chapter 2) due to incomplete lineage sorting (Drummond and Bouckaert, 2015). Incorporated into Bayesian statistics [e.g. BPP (Rannala and Yang, 2013), DISSECT/STACEY (Jones et al., 2015; Jones 2017a), Bayes Factor delimitation (Grummer et al., 2014; Leaché et al., 2014a)], the MSC provides a powerful framework for determining boundaries among very recently diverged lineages (Fujita et al., 2012). Unfortunately, the MSC model assumes no gene flow after species divergence (Zhang et al., 2011), which is a common phenomenon in flowering plants (e.g., Blanco-Pastor et al., 2012; De Villiers et al., 2013; Scheunert and Heubl, 2014; Folk et al., 2018). Chapters 2 and 3 of the present thesis address this dilemma by performing a step-by-step approach, with an initial step of identification and subsequent removal of hybrid individuals, followed by application of different (MSC) species-delimitation methods exemplified in the genera *Leucanthemum* and *Rhodanthemum*.

1.3 What we still don't know about polyploidy

Polyploidy, the presence of more than two full chromosome sets in a nucleus, is a common phenomenon in flowering plants (Wood et al., 2009), resulting in a broad range of chromosome numbers [varying from $2n = 4$ to $2n = 640$ in angiosperms, Leitch and Leitch (2012)]. Due to the high level of polyploidy in vascular plants, it is assumed that 'polyploidy has been associated with speciation and is, therefore, of substantial evolutionary significance' (Stuessy and Weiss-Schneeweiss, 2019). Polyploids can be either formed by multiplication

of chromosome sets within a single species (autopolyploidy), or via merging of chromosome sets from different species due to hybridization (allopolyploidy). Although autopolyploidization is more common than realized previously (Parisod et al., 2010), it is supposed that hybridization between two parental species accompanied by chromosome doubling (allopolyploidization) is the more frequent mode of polyploid formation (Kadereit, 2015). The latter mechanism leads to polyploid species with genetic compositions different from their progenitors, which can be beneficial for the colonization of novel ecological niches (Ramsey, 2011).

Despite of considerable progress in recent years concerning the investigation of mechanisms and consequences of polyploidy, much less is known about the causes of polyploidy (Soltis et al., 2010) and it is poorly understood why the phenomenon is common in certain plant groups and not in others. The here investigated, closely related genera *Leucanthemum* and *Rhodanthemum* represent an attractive system for studying causes of polyploidy, as polyploidization is restricted to the former genus, albeit a similar number of diploids exists in both plant groups. Specific hypotheses for the formation of polyploids within *Leucanthemum* are proposed and evaluated within a phylogenetic context in chapter 4 of the present thesis. The applied approach includes (i) species delimitation analyses in *Leucanthemum* and *Rhodanthemum* (ii) evaluation of genetic divergence and homoploid hybridization patterns among delimited species, and (iii) divergence-time estimations in the subtribe Leucantheminae.

1.4 The subtribe Leucantheminae

The subtribe Leucantheminae comprises annual and perennial herbs or subshrubs and is part of the Mediterranean clade within the Eurasian grade of Compositae tribe Anthemideae (Oberprieler et al., 2009). Besides six unispecific or extremely small genera comprising only 2-4 species, the genera *Leucanthemum* and *Rhodanthemum* are the most prominent and species-rich genera of the subtribe (Table 1.1). The main distinctive feature for the circumscription of Leucantheminae is the achene anatomy of its members, which is i.a. characterized by myxogenic cells along the ribs and resin canals between the ribs of the pericarp (Bremer and Humphries, 1993). However, molecular analyses by Oberprieler et al. (2007) and Wagner et al. (2019) (chapter 4) argue for the extension of the subtribe by inclusion of three small genera (Table 1.1) devoid of the mentioned achene characteristics. This ‘extended subtribe’ has (i) a crown age of 11.86 Ma (8.71-15.38 Ma, see Figure 4.5 and Table 4.3), (ii) its origin in NW Africa (Oberprieler, 2005), and (iii) a recent distribution pattern covering the Mediterranean region, Macaronesia, Europe, and Asia (Table 1.1).

The genus *Leucanthemum* is a vast polyploid complex with 15 diploid and 25+ polyploid taxa (Euro+Med, 2019), showing chromosome numbers from $2n = 2x = 18$ to $2n = 22x = 108$ (Vogt, 1991). The genus has a crown age of 1.93 Ma (1.14-2.94 Ma, see Figure 4.5 and Table 4.3) and is distributed all over the European continent, with one species (*L. ircutianum*) reaching Siberia and some species being introduced to many temperate regions in the northern and southern hemisphere (Meusel and Jäger, 1992). According to Vogt (1991) and Marchi (1982), the centers of diversity of the genus are the Iberian and Apennine peninsulas. *Leucanthemum* taxa are traditionally delimited using morphological, karyological and chorological aspects (Vogt, 1991). Particularly important morphological features in this context are i.a. leaf shapes, shape and color of involucral bracts and achene characteristics (e.g., total length or length of corona, see Table 2.2). More recent studies have revealed new taxonomical insights into morphologically similar species-groups within the genus by additionally using molecular data [the *L. pluriflorum*-clan: Greiner et al. (2013); the *L. ageratifolium*-group: Wagner et al. (2017), chapter 2; the '*L. esterellense*-group': Oberprieler et al. (2018) and Vogt et al. (2018)]. Furthermore, *Leucanthemum* is considered being an interesting model system for studying reticulate evolution [(Oberprieler et al., 2011a, 2012, 2014; Greiner and Oberprieler, 2012; Greiner et al., 2012, 2013; Konowalik et al., 2015; Wagner et al., 2019 (chapter 4 of the present thesis)].

The genus *Rhodanthemum*, on the other hand, comprises 15 species with strictly diploid chromosome numbers ($2n = 2x = 18$) (Wilcox and Harcourt, 1982; Vogt and Oberprieler, 2008, 2012). The genus has a crown age of 1.29 Ma (0.88-1.87 Ma, see Figure 4.5 and Table 4.3) and is distributed in North Africa (Morocco and Algeria), with one species (*R. arundanum*) reaching southern Spain. Due to the uniform chromosome numbers, *Rhodanthemum* species are traditionally delimited using chorological aspects and morphological features like leaf shape and outline, involucral bracts or indumentum (e.g., Vogt, 1994). The two studies presented in chapter 3 and 4 are the first molecular surveys of the genus.

1.5 Thesis outline

The present thesis investigates micro- and macroevolutionary processes in the young and closely related genera *Leucanthemum* and *Rhodanthemum* by applying different molecular approaches (Sanger- and 454-sequencing, AFLP-fingerprinting and RAD-sequencing). While chapters 2 and 3 are dealing with species delimitation in two morphologically complex, shallowly diverged and intensively hybridizing taxon-groups within *Leucanthemum* and *Rhodanthemum*, respectively, chapter 4 is focusing on the contrasting evolutionary trajectories of both genera within the subtribe Leucantheminae.

The first study (chapter 2) evaluates the robustness of currently available species delimitation methods implemented in BEAST (BFD, BFD*, and DISSECT) in the closely-knit taxon-group around *L. ageratifolium*. Comprising five taxa being allopatrically distributed between northern Spain and southern Italy this study group shows signs of hybridization with the widespread and co-distributed species *L. vulgare* to various extent. As the applied species delimitation methods tend to underestimate species-level diversity in the presence of strong interspecific hybridization, a methodological pipeline for delimiting species despite ongoing gene flow is presented and applied to the empirical data.

In the second part (chapter 3), RAD- and Sanger-sequencing are conducted for delimiting species boundaries in the Ibero-Maghrebian *R. arundanum*-group, a group of four taxa with (i) morphologically differentiated populations or population groups, (ii) signs of interspecific hybridization and (iii) alternative taxonomic treatments based on morphology. RADseq data are assembled de-novo, after evaluation of genotyping errors and parameter optimization in the commonly used pipeline IPYRAD. Furthermore, a new method for delineating species boundaries based on RADseq data is presented and the performance of different species delimitation methods in the presence of hybridization and varying quantities of data is evaluated.

While chapters 2 and 3 are focusing on specific taxon groups within *Leucanthemum* and *Rhodanthemum*, respectively, chapter 4 compares the contrasting evolutionary trajectories of both genera in a more comprehensive phylogenetic study. The main question of this chapter is why the European genus *Leucanthemum* has built up a comprehensive polyploid complex with 25+ polyploid taxa while its North African counterpart *Rhodanthemum* strictly evolved on the diploid level. Genetic divergence and gene flow among diploid lineages of both genera are investigated to evaluate the role of genomic differentiation and hybridization for polyploid speciation. Furthermore, a time-calibrated phylogeny of the subtribe Leucanthemeinae is calculated, to test whether hybridization in *Leucanthemum* has been triggered by the geological conditions during its diversification.

Table 1.1 List of genera belonging to the subtribe Leucanthemeinae according to Bremer and Humphries (1993) plus three closely related genera *Daveau*, *Heteromera* and *Otospermum* according to Oberprieler et al. (2007) and Wagner et al. (2019). Information on number of species (*n*) and distribution area are taken from Euro+Med plantbase (2019) and Oberprieler et al. (2009), respectively.

	<i>n</i>	distribution
<i>Chlamydophora</i> Ehrenb. ex Less.	1	North Africa, Cyprus
<i>Chrysanthoglossum</i> B.H. Wilcox & al.	2	North Africa
<i>Coleostephus</i> Cass.	3	Mediterranean region, Macaronesia
<i>Glossopappus</i> Kunze	1	Southwest Europe, North Africa
<i>Leucanthemum</i> Mill.	42	Europe, Siberia
<i>Mauranthemum</i> Vogt & Oberprieler	4	North Africa, Southwest Europe
<i>Plagiis</i> L'Hèr. ex DC.	3	South Europe (Corsica, Sardinia), North Africa
<i>Rhodanthemum</i> (Vogt) B.H. Wilcox & al.	15	Northwest Africa, Southwest Europe
<i>Daveau</i> Willk. ex Mariz	1	Northwest Africa, Southwest Europe
<i>Heteromera</i> Pomel	2	North Africa
<i>Otospermum</i> Willk.	1	North Africa, Southwest Europe

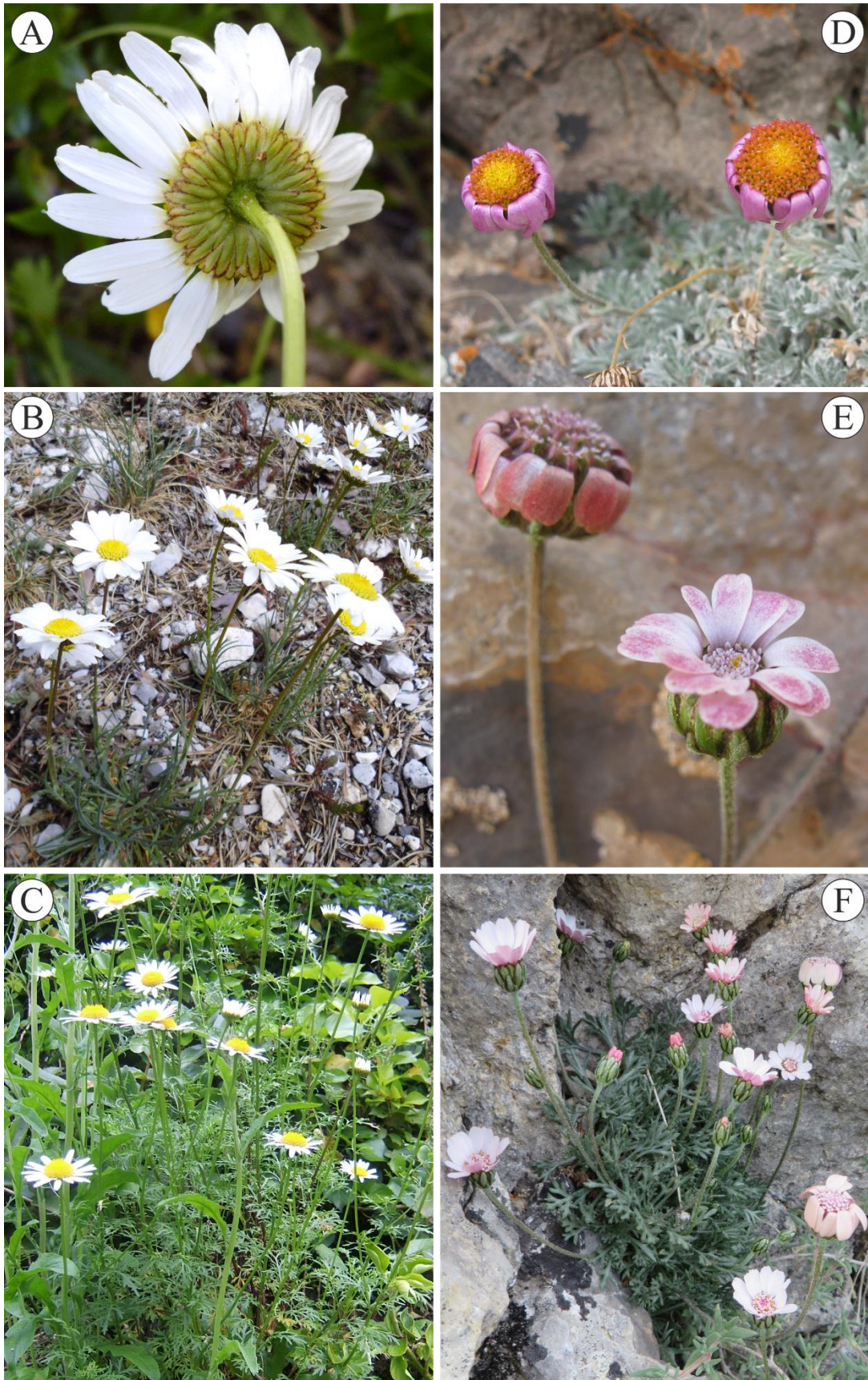


Figure 1.1 Snow White (*Leucanthemum*, left) and Rose Red (*Rhodanthemum*, right): A: *L. legraeanum*, B: *L. graminifolium*, C: *L. monspeliense*, D: *R. redieri* subsp. *humbertii*, E & F: *R. arundanum* s.l. [A-C, Florian Wagner; D, E, Christoph Oberprieler; F, Robert Vogt.

Chapter 2: Fix Me Another Marguerite!

“Fix Me Another Marguerite!”: Species delimitation in a group of intensively hybridizing lineages of ox-eye daisies (*Leucanthemum* Mill., Compositae-Anthemideae)

Florian Wagner, Sabine Härtl, Robert Vogt, Christoph Oberprieler
Molecular Ecology 26: 4260–4283. (2017)

Abstract

Delineating species boundaries in the framework of the multi-species coalescent (MSC) proves to be a reliable, objective, and reproducible method in an increasing number of studies. However, the underlying model assumes the lack of gene flow after speciation; an assumption which may be frequently violated in plant evolution. This study evaluates the robustness of currently available species delimitation methods implemented in BEAST (BFD, BFD*, and DISSECT) in the closely-knit ox-eye daisy group around *Leucanthemum ageratifolium* Pau. Comprising five taxa being allopatrically distributed between northern Spain and southern Italy this study group shows signs of hybridization with the widespread and codistributed species *Leucanthemum vulgare* (Vaill.) Lam. to various extent. As expected, our empirical analyses based on both AFLP fingerprinting and sequence data demonstrate that the robustness of species delimitation results is considerably influenced by the intensity of hybridization among species and the number of hybrid individuals included. Therefore, we set up a methodological pipeline with a first step of identification and subsequent removal of individuals showing admixed genetic patterns caused by actual interbreeding using AFLP-fingerprint and morphometric data, followed by application of different Bayesian MSC species delimitation methods based on the remnant individuals using both AFLP-fingerprint and sequence data (four nuclear markers, five concatenated intergenic spacer regions of the plastid genome). The results argue for acknowledgement of *Leucanthemum laciniatum*, *L. legraeum*, and *L. ligusticum* as independent species, show the close relationship of *L. ageratifolium*, *L. monspeliense*, and *L. vulgare*, and give rise to the description of three nothospecies new to science.

Keywords: Bayes factor delimitation, DISSECT, hybridization, marginal likelihoods, multi-species coalescent, species delimitation

2.1 Introduction

Species are routinely used as fundamental units in studies dealing with evolutionary biology, biogeography, ecology, and conservation biology (Camargo and Sites, 2013). However, defining these units by lumping populations into a single species or splitting populations into several species is not a trivial task, especially in the case of allopatric speciation processes and short divergence times (Carstens et al., 2013; Fujita et al., 2012). Using exclusively morphological traits to delimit species can lead to an over- as well as an underestimation of the true number of evolutionary independent lineages in a group of organisms, caused for instance by phenotypic plasticity (e.g., Flot et al., 2011) or cryptic speciation (e.g., Toprak et al., 2016). To prevent these problems and to delimit species in a more accurate and objective manner, a plethora of methods was developed in the last decades, which use molecular data for delineating species boundaries (e.g., Miralles and Vences, 2013). Among these, methods operating in the framework of the multi-species coalescent (MSC) model (Rannala and Yang, 2003) proved to be successful in an increasing number of studies that make use of multi-locus sequence or genomewide SNP data generated for the purpose of species delimitation (e.g., Aydin et al., 2014; Grummer et al., 2014; Hedin, 2015; Hedin et al., 2015; Leaché et al., 2014a; Toprak et al., 2016).

A very popular and frequently used approach in this context is the MSC species-delimitation method implemented in the software program BPP (Rannala and Yang, 2013; Yang and Rannala, 2010, 2014). BPP executes a reversible-jump Markov Chain Monte Carlo (rjMCMC) algorithm to move between different species-delimitation models using either a fixed guide tree or by simultaneously exploring alternative species phylogenies (Yang, 2015). Although this method was evaluated as performing quite well for simulated as well as empirical data sets (e.g., Zhang et al., 2011), one disadvantage of BPP is the lack of relaxed-clock models and sophisticated nucleotide substitution models. This constraint is removed when one performs MSC species-delimitation with the software package BEAST (Drummond et al., 2012), which offers the full range of substitution, frequency, site and clock models as well as different tree priors (Drummond and Bouckaert, 2015). Currently, two different species-delimitation methods are provided in BEAST: Bayes factor delimitation (BFD, Grummer et al., 2014; BFD*, Leaché et al., 2014a) and the threshold-based methods DISSECT/STACEY (Jones et al., 2015; Jones 2017a). When multi-locus sequence data are available, BFD can be performed within the species-tree estimation framework *BEAST (Heled and Drummond, 2010), whereas the package SNAPP (Bryant et al., 2012) has to be consulted in the case of SNP or AFLP data [BFD* (*with genomic data) in Leaché et al. (2014a)]. In both cases, marginal likelihoods are estimated for different species-delimitation scenarios and Bayes factors are calculated afterwards to evaluate the competing hypotheses.

In contrast to this approach, the recently developed BEAST package DISSECT explores the full space of possible clusterings of individuals (potential species) and tree topologies without the need of prior assignment of individuals to clusters/species. The method, which runs under the term STACEY in BEAST2 (Bouckaert et al., 2014), uses a Dirac delta function to bypass the need for reversible-jump MCMC (Jones et al., 2015) and was successfully used by Toprak et al. (2016) to reveal extensive cryptic speciation in the *Silene aegyptiaca* complex.

All MSC species-delimitation methods reviewed above consider incomplete lineage sorting (ILS, Maddison, 1997) as a source for incongruence among gene trees, but do not account for the blurring effect of gene flow among lineages on phylogenetic patterns (Slatkin and Maddison, 1989). Considering the high frequency of hybridization events in the plant kingdom (Mallet, 2005), the assumption of missing gene flow after species divergences may be easily violated in MSC based species-delimitation studies dealing with plants. In the present contribution, we address this dilemma by performing a step-by-step approach to investigate species delimitation in the close-knit *Leucanthemum ageratifolium* group: In a first step, potential hybrid individuals between the allopatrically distributed members of the *L. ageratifolium*-group with the sympatric species *L. vulgare* are identified based on AFLP-fingerprinting and morphometric data. The AFLP data and additional sequence information from five intergenic spacer regions of the plastid genome together with four nuclear markers are subsequently used for delimiting species by performing all currently available BEAST applications (DISSECT, BFD, BFD*) after removal of putative hybrid individuals from the data set. Furthermore, the robustness of the recently developed threshold-based method DISSECT is evaluated by performing all analyses with the complete sequence data set and with a reduced dataset excluding potential hybrid individuals.

The genus *Leucanthemum* Mill. ('Marguerites'; Compositae, Anthemideae) comprises 42 flowering plant species (Euro+Med, 2016) distributed all over the European continent and represents an attractive system for studying reticulate evolution on the diploid (Konowalik et al., 2015; Oberprieler et al., 2014) and polyploid (Greiner et al., 2012, 2013; Oberprieler et al., 2011a, 2014) level. In a recent next-generation sequencing study, Konowalik et al. (2015) investigated 19 diploid *Leucanthemum* species, which could be separated in two species groups with contrasting hybridization patterns: An early-diverging stock of morphologically clearly circumscribed species without evidence for recent hybridization events, and a second, morphologically elusive group characterized by a strong signal of gene flow among lineages. Despite extensive data acquisition and considerable methodological efforts, not all questions concerning the complex second group could be answered satisfactorily and especially the taxonomic rank and phylogenetic relationships of the recently described Ligurian species *L. ligusticum* remained unclear in this study. This was possibly due to poor taxon sampling [*L. ligusticum* was represented by only a single accession in Konowalik et al. (2015)] but was

surely also caused by the lack of the enigmatic species *L. legraeum*, described from S France (Bock and Tison, 2012) but recently also reported from locations in Liguria (Bernardello et al., 2015). As both taxa are characterized by strongly divided leaves, we concentrated in the current study on a group of *Leucanthemum* species sharing this leaf-shape feature (hereafter the *L. ageratifolium*-group, Figure 2.1). This study group comprises, in addition to the two already mentioned taxa, and the eponymous lineage *L. ageratifolium* from NE Spain, the diploid representatives of the S French species *L. monspeliense*, as well as the S Italian taxon *L. laciniatum*. Additionally, we included several populations of the widespread species *L. vulgare* in our sampling, because this taxon is codistributed with all members of the *L. ageratifolium*-group and therefore a proper candidate for potential hybridization events.

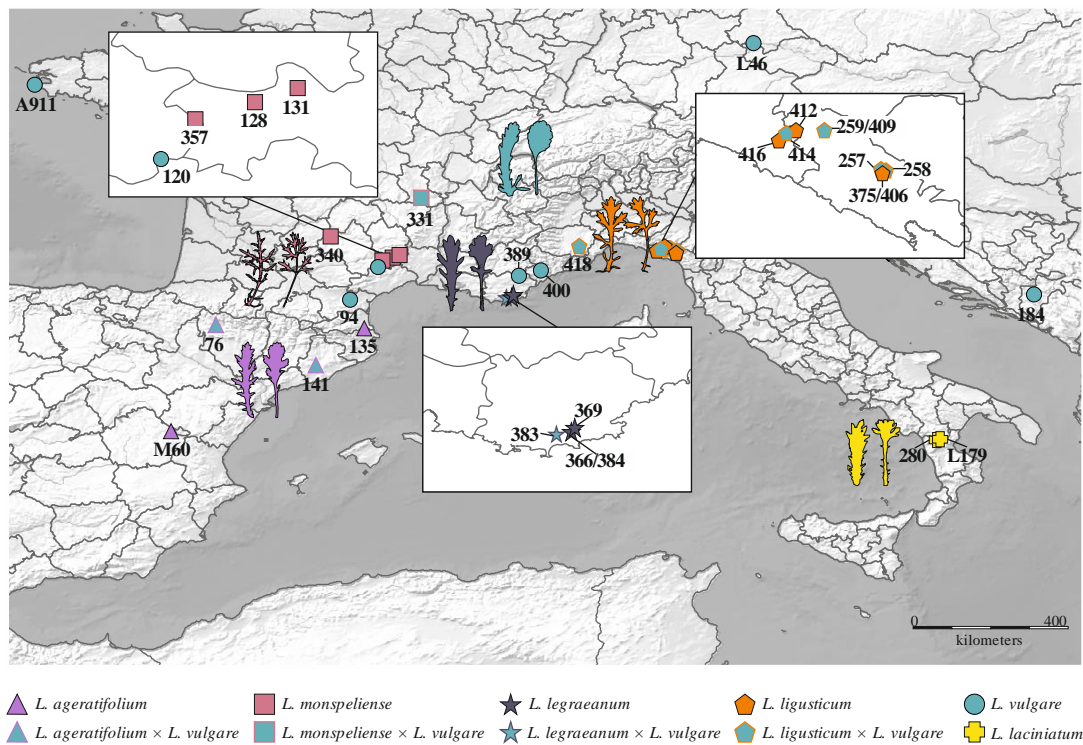


Figure 2.1 Map showing the locations of all examined *Leucanthemum* populations in the study. Populations considered being admixed according to the AFLP analyses are indicated by intermediate colours and shapes. In addition, each taxa of the *L. ageratifolium*-group is represented by digitized silhouettes of characteristic cauline and basal leaves next to its distributional range (leaves of *L. laciniatum* are obtained from Marchi (1982); leaves are not drawn to scale).

2.2 Materials and Methods

2.2.1 Plant material and DNA extraction

The majority of silica-dried leaf and herbarium material used in this study was collected during field trips in Spain, France, and Italy between 2007 and 2015. In total, 88 accessions

from 29 populations were included in the AFLP fingerprinting procedure and one representative of each population in the sequence-based analyses (see Figure 2.1 and Table 2.1 for accession information). For all molecularly analyzed samples, total genomic DNA was extracted using the CTAB DNA extraction protocol (Doyle and Dickson, 1987; Doyle and Doyle, 1987). Additionally to the molecular study, we analyzed at least one individual from each population morphologically by examining leaf dissection and achene characteristics. Voucher specimens are deposited in the herbarium of the Botanical Museum Berlin-Dahlem (see Table S2.4 for voucher information).

2.2.2 AFLP fingerprinting

The AFLP procedure followed the original protocol of Vos et al. (1995) with some minor modifications described in Konowalik et al. (2015). To evaluate the performance of the AFLP genotyping, we used 16 randomly selected replicates representing 18% of the total data set. After fragment detection on a CEQ8000 capillary sequencer (Beckman Coulter, Krefeld, Germany), raw CEQ trace files were checked manually in GENOGRAPHER v.1.6.0 (Benham et al., 1999) before automatic scoring of AFLP fragments was performed using GELCOMPAR II (Applied Maths, Sint-Martens-Latem, Belgium). In this step, 400 combinations of different values for minimal profiling (0.1, 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0), minimal area (0.1, 0.2, 0.3, 0.4, 0.5), matching tolerance (0.15, 0.2, 0.25, 0.3, 0.35), and the analyzed gel length (100–420 bp, 150–420 bp) were specified and evaluated using Python scripts developed by Holland et al. (2008), to find the best parameter combination. During this evaluation procedure Euclidean error rates and Jaccard distances were calculated for each replicate pair and each character matrix separately by: (a) dividing the number of incorrect calls $[N(0,1) + N(1,0)]$ by the sum over all possible calls $[N(0,0) + N(1,1) + N(0,1) + N(1,0)]$ (Euclidean error rates) and (b) using the same formula as in (a) but ignoring the (0,0) case (Jaccard distances). Each 0/1 matrix was subsequently used for bootstrap analyses carried out in PAUP* (Swofford, 2003) based on 1,000 replicates of neighbor-joining tree searches on uncorrected distances. Resolution scores were calculated subsequently by dividing the number of bootstrap scores over 50% by the maximum number of internal edges in each tree. Results of the bootstrap analyses were finally used for computing majority-rule consensus-trees to count the number of correctly paired replicates for each character matrix.

Table 2.1 Plant material used for the sequencing (Seq.-samples) and AFLP-fingerprinting (AFLP-samples) including information about population, location, and collector. Asterisks (*) refer to sequences from Konowalik et al. (2015). For herbarium voucher information see Table S2.4.

Taxon	Pop. code	Seq.- samples	AFLP-samples	Geographic location	Coord.	Collector	<i>trnL-trnF</i>	<i>trnC-petN</i>	<i>psbA-trnH</i>	<i>petN-psbM</i>	<i>trnQ-rps16</i>	A39	CI2	C33	D23
<i>Leucanthemum laciniatum</i> Huter, Porta & Rigo	L179	L179	L179	IT, Basilicata, Castrovillari, 1900–2100 m	39.91 N, 16.19 E	Vogt 15614	LN869035*	LN869085*	LN868985*	LN869135*	LN869184*	ERS758390*	ERS758390*	ERS758390*	ERS758390*
<i>Leucanthemum laciniatum</i> Huter, Porta & Rigo	280	280-1	280-1, 280-2, 280-3, 280-4, 280-5, 280-6, 280-7	IT, Calabria, Colle del Drogone, 1580 m	39.90 N, 16.11 E	Tomasello TS420	LN869036*	LN869086*	LN868986*	LN869136*	LN869185*	ERS758391*	ERS758391*	ERS758391*	ERS758391*
<i>Leucanthemum legraeum</i> (Rouy) B.Bock & J.-M.Tison	366/384	366-1	366-1, 384-2, 384-3, 384-4, 384-8	FR, Provence-Alpes-Côte d'Azur, Massif des Maures, 410 m	43.20 N, 06.31 E	Vogt 17189 / Vogt 17434, Oberprieler 10915 & Wagner	KY778058	KY778096	KY778077	KY778020	KY778039	KY778172 KY778173	KY778202 KY778203	KY778144	KY778115 KY778116
<i>Leucanthemum legraeum</i> (Rouy) B.Bock & J.-M.Tison	369	369-1	369-1	FR, Provence-Alpes-Côte d'Azur, Massif des Maures, 210 m	43.24 N, 06.34 E	Vogt 17192	KY778059	KY778097	KY778078	KY778021	KY778040	KY778174	KY778204 KY778205	KY778145	KY778117 KY778118
<i>L. legraeum</i> × <i>L. vulgare</i>	383	383-1	383-1, 383-2, 383-3, 383-4, 383-5	FR, Provence-Alpes-Côte d'Azur, Vallée du Pansard, 77m	43.19 N, 06.21 E	Vogt 17432, Oberprieler 10913 & Wagner	KY778060	KY778098	KY778079	KY778022	KY778041	KY778175 KY778176 KY778177	KY778206	KY778146 KY778147	KY778119 KY778120
<i>Leucanthemum ligusticum</i> Marchetti, R.Bernardello, Melai & Peruzzi	375/406	375-1	375-1, 406-1, 406-2	IT, Liguria, Rocchetta di Vara	44.25 N, 09.76 E	Marchetti s.n. / Vogt 17467, Oberprieler 10948 & Wagner	KY778061	KY778099	KY778080	KY778023	KY778042	KY778178 KY778179	KY778207	KY778148 KY778149	KY778121
<i>Leucanthemum ligusticum</i> Marchetti, R.Bernardello, Melai & Peruzzi	412	412-1	412-1, 412-2, 412-3	IT, Liguria, Rocche di Valletti, 700 m	44.36 N, 09.51 E	Vogt 17467, Oberprieler 10948 & Wagner	KY778062	KY778100	KY778081	KY778024	KY778043	KY778180	KY778208 KY778209	KY778150 KY778151	KY778122
<i>Leucanthemum ligusticum</i> Marchetti, R.Bernardello, Melai & Peruzzi	416	416-1	416-1, 416-2, 416-3	IT, Liguria, Ponte di Lagoscuro, 246 m	44.34 N, 09.46 E	Vogt 17471, Oberprieler 10952 & Wagner	KY778063	KY778101	KY778082	KY778025	KY778044	KY778181 KY778182	KY778210 KY778211	KY778152 KY778153	KY778123
<i>L. ligusticum</i> × <i>L. vulgare</i>	257	257-1	257-1, 257-2, 257-3	IT, Liguria, Rocchetta di Vara, 228 m	44.25 N, 09.76 E	Vogt 16943 & Oberprieler 10850	KY778064	KY778102	KY778083	KY778026	KY778045	KY778183 KY778184	KY778212	KY778154	KY778124

Table 2.1 Continued.

Taxon	Pop. code	Seq.- samples	AFLP-samples	Geographic location	Coord.	Collector	<i>trnL-trnF</i>	<i>trnC-petN</i>	<i>psbA-trnH</i>	<i>petN-psbM</i>	<i>trnQ-rps16</i>	A39	C12	C33	D23
<i>L. ligusticum</i> × <i>L. vulgare</i>	258	258-1	258-1, 258-2, 258-4, 258-9, 258-11, 258-12, 258-13, 258-14, 258-17, 258-25	IT, Liguria, Rocchetta di Vara, 228 m	44.25 N, 09.76 E	Vogt 16944 & Oberprieler 10851	LN869053*	LN869103*	LN869003*	LN869153*	LN869202*	ERS758392*	ERS758392*	ERS758392*	ERS758392*
<i>L. ligusticum</i> × <i>L. vulgare</i>	259/409	259-1	259-1, 409-1, 409-2	IT, Liguria, Varese Ligure, 341 m	44.37 N, 09.59 E	Vogt 16945 & Oberprieler 10852 / Vogt 17464, Oberprieler 10945 & Wagner	KY778065	KY778103	KY778084	KY778027	KY778046	KY778185	KY778213	KY778155 KY778156	KY778125
<i>L. ligusticum</i> × <i>L. vulgare</i>	414	414-1	414-1, 414-2, 414-3	IT, Liguria, Piani di Oneto, 829 m	44.36 N, 09.48 E	Vogt 17469, Oberprieler 10950 & Wagner	KY778066	KY778104	KY778085	KY778028	KY778047	KY778186	KY778214 KY778215	KY778157	KY778126
<i>L. ligusticum</i> × <i>L. vulgare</i>	418	418-1	418-1, 418-2, 418-3	IT, Piemonte, Mondovi, 492 m	44.35 N, 07.89 E	Vogt 17473, Oberprieler 10954 & Wagner	KY778067	KY778105	KY778086	KY778029	KY778048	KY778187 KY778188	KY778216 KY778217	KY778158 KY778159	KY778127 KY778128 KY778129
<i>Leucanthemum monspeliense</i> (L.) H.J.Coste	131	131-20	131-1, 131-2, 131-20	FR, Languedoc-Roussillon, St.-Andréde-Valborgne, 380 m	44.14 N, 03.73 E	Vogt 16716, Oberprieler 10671 & Konowalik	LN869019*	LN869069*	LN868969*	LN869119*	LN869168*	ERS758395*	ERS758395*	ERS758395*	ERS758395*
<i>Leucanthemum monspeliense</i> (L.) H.J.Coste	128	128-1	128-1	FR, Languedoc-Roussillon, l'Espérrou, 750 m	44.09 N, 03.58 E	Vogt 16712, Oberprieler 10667 & Konowalik	LN869020*	LN869070*	LN868970*	LN869120*	LN869169*	ERS758396*	ERS758396*	ERS758396*	ERS758396*
<i>Leucanthemum monspeliense</i> (L.) H.J.Coste	340	340-1	340-1, 340-2, 340-3	FR, Midi-Pyrénées, La Roque-Bouillac, 184 m	44.58 N, 02.18 E	Vogt 17156, Oberprieler 10881 & Wagner	KY778068	KY778106	KY778087	KY778030	KY778049	KY778189	KY778218 KY778219	KY778160	KY778130
<i>Leucanthemum monspeliense</i> (L.) H.J.Coste	357	357-1	357-1, 357-2, 357-3	FR, Midi-Pyrénées, Saint-Jean-du-Bruel, 571 m	44.03 N, 03.37 E	Vogt 17179, Oberprieler 10904 & Wagner	KY778069	KY778107	KY778088	KY778031	KY778050	KY778190 KY778191	KY778220	KY778161 KY778162	KY778131
<i>L. monspeliense</i> × <i>L. vulgare</i>	331	331-1	331-1, 331-2, 331-4	FR, Rhône-Alpes, Saint-Etienne, 404 m	45.47 N, 04.25 E	Vogt 17147, Oberprieler 10872 & Wagner	KY778070	KY778108	KY778089	KY778032	KY778051	KY778192 KY778193	KY778221	KY778163 KY778164	KY778132

Table 2.1 Continued.

Taxon	Pop. code	Seq.- samples	AFLP-samples	Geographic location	Coord.	Collector	<i>trnL-trnF</i>	<i>trnC-petN</i>	<i>psbA-trnH</i>	<i>petN-psbM</i>	<i>trnQ-rps16</i>	A39	CI2	C33	D23
<i>Leucanthemum ageratifolium</i> Pau	135	135-7	135-1,135-2, 135-7	FR, Pyrénées-Orientales, La Vallée Heureuse, 410 m	42.50 N, 02.96 E	Konowalik KK42 & Ogrodowczyk	LN869054*	LN869104*	LN869004*	LN869154*	LN869203*	ERS758411*	ERS758411*	ERS758411*	ERS758411*
<i>Leucanthemum ageratifolium</i> Pau	M60	M60-1	M60-1, M60-2, M60-3	ES, Castilla-La Mancha., Salinas de Manzano, 1157 m	40.10 N, 01.52 W	Cordel s.n.	LN869055*	LN869105*	LN869005*	LN869155*	LN869204*	ERS758412*	ERS758412*	ERS758412*	ERS758412*
<i>L. ageratifolium</i> × <i>L. vulgare</i>	141	141-1	141-1, 141-2, 141-3	ES, Catalunya, Montserrat, 645 m	41.61 N, 01.82 E	Konowalik KK48 & Ogrodowczyk	KY778071	KY778109	KY778090	KY778033	KY778052	KY778194 KY778195	KY778222	KY778165	KY778133 KY778134
<i>L. ageratifolium</i> × <i>L. vulgare</i>	76	76-2	76-2	ES, Aragon, Narvasa, 1020 m	42.53 N, 0.48 W	Höbl 76 & Himmelreich	KY778072	KY778110	KY778091	KY778034	KY778053	KY778196	KY778223 KY778224	KY778166	KY778135 KY778136
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	94	94-1	94-1	FR, Languedoc-Roussillon, Montlaur, 160 m	43.13 N, 02.61 E	Vogt 16641, Oberprieler 10592 & Konowalik	LN869050*	LN869100*	LN869000*	LN869150*	LN869199*	ERS758406*	ERS758406*	ERS758406*	ERS758406*
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	L46	L46-1	L46-1, L46-2, L46-3	DE, Bayern, Pittmannsdorf, 450 m	49.03 N, 11.88 E	Eder & Oberprieler s.n.	LN869051*	LN869101*	LN869001*	LN869151*	LN869200*	ERS758407*	ERS758407*	ERS758407*	ERS758407*
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	184	184-1	184-1	BA, Gacko, Ribari, 930 m	43.24 N, 18.34 E	Vogt 16806 & Prem-Vogt	LN869052*	LN869102*	LN869002*	LN869152*	LN869201*	ERS758408*	ERS758408*	ERS758408*	ERS758408*
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	120	120-20	120-1, 120-2, 120-20	FR, Midi-Pyrénées, La Pezade, 756 m	43.89 N, 03.25 E	Vogt 16699, Oberprieler 10654 Konowalik	KY778073	KY778111	KY778092	KY778035	KY778054	KY778197	KY778225	KY778167 KY778168	KY778137 KY778138
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	A911	A911	A911	FR, Bretagne, Point de Brézelle	48.06 N, 04.66 W	Stutz s.n.	KY778074	KY778112	KY778093	KY778036	KY778055	KY778198 KY778199	KY778226	KY778169	KY778139
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	389	389-1	389-1, 389-2, 389-3	FR, Provence-Alpes-Côte d'Azur, Draguignan, 774 m	43.67 N, 06.50 E	Vogt 17439, Oberprieler 10920 & Wagner	KY778075	KY778113	KY778094	KY778037	KY778056	KY778200	KY778227	KY778170	KY778140 KY778141
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	400	400-1	400-1, 400-2, 400-3	FR, Provence-Alpes-Côte d'Azur, Montagne du Cheiron, 918 m	43.79 N, 07.00 E	Vogt 17454, Oberprieler 10935 & Wagner	KY778076	KY778114	KY778095	KY778038	KY778057	KY778201	KY778228	KY778171	KY778142 KY778143

2.2.3 Detection of potential hybrid individuals

We used three different methods to identify potential hybrid individuals between the representatives of the *L. ageratifolium*-group on the one side and the widespread and codistributed species *L. vulgare* on the other: (a) The optimized and binary coded final AFLP-profile matrix (0/1-matrix) was split into five submatrices, each including all accessions of *L. vulgare* together with all accessions of only one representative of the *L. ageratifolium*-group. For each submatrix, an ordination of OTUs was performed by principal coordinates analysis (PCoA) based on Bray-Curtis pairwise distances calculated in MATLAB v.8.0.0.783 (R2012b) (The MathWorks inc., Natick, MA, USA) using the FATHOM toolbox (Jones, 2015). (b) The same data sets were used for calculating individual-wise maximum-likelihood hybrid indices with 95% confidence intervals as implemented in the R-package INTROGRESS (Gompert and Buerkle, 2010). For INTROGRESS analyses, individuals were assigned to pure parental populations based on the results of the PCoA analyses and taking also into account morphological and distributional evidence. (c) Neighbor-net networks were generated with SPLITSTREE v.4.13.1 (Huson and Bryant, 2006) based on the five submatrices, the total data set (aflpdata1) and a further data set without putative hybrid individuals (aflpdata2), according to the results of former analyses. For this purpose, pairwise distances among individual AFLP phenotypes were calculated according to Nei and Li's (1979) restriction-site distance coefficient as implemented in PAUP* (Swofford, 2003). The latter program was also used to obtain bootstrap support values via neighbor-joining tree searches (Saitou and Nei, 1987) performing 1,000 bootstrap replicates. All bootstrap values higher than 70% were finally plotted on the Neighbor-net networks based on the data sets aflpdata1 and aflpdata2.

In addition to the molecular studies, we performed also morphometric analyses for the purpose of hybrid detection. Basal and cauline leaves from a total of 58 herbarium specimens of *L. vulgare* and all of its codistributed taxa of the *L. ageratifolium* group, were digitized and analyzed with the software IMAGEJ v.1.50e (Schindelin et al., 2015). Both, lamina perimeter and total area were measured to calculate the dissection index (DI) for each leaf as defined in Kincaid and Schneider (1983). The DI of an outline is the ratio of its perimeter to the square root of its area standardized so that a circle has a value of 1.0 and a more complex outline is characterized by a higher value (McLellan, 1993). This dimensionless value was successfully used to describe the shape of leaves of herbs (McLellan, 1993), shrubs (McIntosh et al., 2014), and trees (McLellan and Endler, 1998) with a similar spectrum of dissection complexity as observed in our study group. To pinpoint populations with hybridization patterns, DI values were depicted for all *L. vulgare* specimens and all accessions of each member of the *L. ageratifolium*-group in separate scatterplots.

2.2.4 Plastid and nuclear marker sequencing

For 19 accessions of the study group, sequence data were generated for nine loci: five intergenic spacer regions of the plastid genome (*trnL-trnF*, *trnC-petN*, *psbA-trnH*, *petN-psbM*, *trnQ-rps16*) and four potentially unlinked and single-copy nuclear regions (*A39*, *C12*, *C33*, *D23*). Nuclear markers were developed by Chapman et al. (2007) for the sunflower family (Compositae) and proved to be variable and amplifiable for *Leucanthemum* species by Konowalik et al. (2015). PCR amplifications were performed with primers listed in Table S2.1 and *Taq* RED Polymerase (Ampliqon A/S, Odense, Denmark). We used AMPure magnetic beads (Agencourt Bioscience Corp., Beverly, MA, USA) to purify amplified products before sending them to MacroGen Inc. (Amsterdam, Netherlands) for Sanger sequencing. Electropherograms were checked manually for base-call errors using CHROMAS LITE v.2.0 (Technelysium Pty Ltd, South Brisbane, Australia) and in the case of one plastid marker (*trnC-petN*) a poly-A repeat was discarded to avoid misalignment. Nuclear sequences with more than one polymorphic site were treated as described below: (a) In the case of length-variable sequence copies ('alleles'), PCR products were resequenced from the reverse direction and CHAMPURU v.1.0 (Flot, 2007; Flot et al., 2006) was used for phase determination. (b) In the case of alleles of equal length, PCR products were cloned into a pJet cloning vector (Fermentas/Thermo Fisher Scientific Inc., Waltham, MA, USA) and transformed into NEB Turbo bacteria (New England Biolabs Inc., Ipswich, MA, USA). We finally picked and sequenced eight clones per accession to ensure a 0.95 probability of obtaining the two alleles expected for a diploid species (Joly et al., 2006). The resulting sequence data were united with sequence information of 10 individuals investigated by Konowalik et al. (2015) in a Roche 454 pyrosequencing study, to obtain a final data set (seqdata1), in which each of the 29 populations under study was represented by one accession (see Table 2.1). A second data set (seqdata2) was built by excluding all individuals that were identified as putative hybrids in the AFLP-based data analyses described above.

2.2.5 Multiple sequence alignments and gene-tree reconstructions

Sequences were sorted marker-wise, aligned manually in BIOEDIT (Hall, 1999), and passed to the program GAPPACODER (Young and Healy, 2003) for indel coding according to the simple gap-coding method of Simmons and Ochoterena (2000). Afterwards, all nucleotide and indel partitions of different plastid markers were concatenated by hand and subsequently treated as a single locus. For each alignment, we calculated the number of variable sites, parsimony informativeness, and consistency (CI) and retention index (RI) in PAUP*. Nucleotide substitution models for all loci of both data sets (seqdata1 and seqdata2) were selected using

the Akaike information criterion (AIC) in JMODELTEST v.2.1.7 (Darriba et al., 2012). Each alignment was also checked for evidence of recombination events by executing the ‘Genetic Algorithm for Recombination Detection’ (GARD; Kosakovsky Pond et al., 2006). Bayesian gene trees were estimated for both sequence data sets in BEAST v.1.8.3 (Drummond et al., 2012) using all allele sequences in the case of heterozygous individuals. We used the binary simple model for binary coded indel data and models calculated in JMODELTEST for sequence data. Priors for substitution models given by BEAUTI v.1.6.2 were accepted and models, which were not available in this application, were specified by hand in the xml files. Each marker was run separately with a strict and an uncorrelated relaxed-clock model (Drummond et al., 2006), using default priors in both cases. In all of the ten resulting xml files (five markers, two clock models) for both data sets (seqdata1 and seqdata2), a gamma prior with shape 2.0 and scale 0.002 was specified for the coalescent constant tree prior as in Aydin et al. (2014), before they were uploaded to the CIPRES web portal (Miller et al., 2010) to perform runs with 15 million generations and a sample frequency of 1,000. TRACER v.1.6.0 (Rambaut et al., 2014) was used to evaluate convergence and mixing for each run and only when all parameters showed ESS values higher than 200 it was accepted. If this criterion was not met, we performed additional runs with 150 million generations and a sample frequency of 10,000. Finally, we constructed a maximum clade credibility tree for each successful run using a burn-in of 10%, a posterior probability limit of 0.5, and the common ancestor heights algorithm in TREEANNOTATOR v.2.3.2. For the purpose of model comparison (strict vs. relaxed clock), we calculated marginal likelihood values via the path sampling method (under the term ‘thermodynamic integration’ in Lartillot and Philippe, 2006) using a chain length of 15 million generations and 100 path steps. Only in the case of a difference of more than 3 log-likelihood units, the more parameter-rich relaxed-clock model was preferred over a strict-clock model (following suggestions by Kass and Raftery, 1995).

2.2.6 MSC species-delimitation

Species-tree analyses without prior assignment of individuals to species were performed with DISSECT (Jones et al., 2015) using BEAST v.1.8.3. DISSECT analyses are similar to standard *BEAST (Heled and Drummond, 2010) analyses, in which all accessions are treated as separate species (designated as ‘minimal clusters’ in Jones et al., 2015). We used BEAUTI v.1.8.0 to prepare the xml files as described in detail for the *BEAST analyses below, and manipulated the xml files afterwards following the instructions of Jones et al. (2015) by replacing the usual birth-death model with a birth-death-collapse model and adding an operator for the origin height. Two additional parameters have to be specified in a DISSECT

analysis: (a) the ‘collapsing height’ value ϵ is a “compromise between exactly matching a particular model and the practicalities of computation” and should be set between $1e-4$ and $1e-5$ according to Jones et al. (2015); (b) the ‘collapsing weight’ parameter ω can be used to reflect prior knowledge about the number of species and can either be fixed to a specific value or estimated by adding a hyperprior [see Jones et al. (2015) for details]. We specified four different xml files with varying values for the parameters ϵ and ω : for ϵ , we specified either $1e-4$ or $1e-5$ to cover the two extremes of the range suggested by Jones et al. (2015). For ω , we used either a flat prior (beta distribution with parameters 1.0 and 1.0), or an informative prior with the highest probability density for 2 clusters (beta distribution with parameters 10 and 1.5). The latter prior distribution reflected our assumptions of the number of species in the data set after evaluating the AFLP data and the results of the gene-tree analyses. Each xml file was run twice with different seeds, a chain length of 100 million generations and a logging frequency of 5,000 in BEAST v.1.8.3. Convergence and ESS values were checked via TRACER v.1.6.0 and results from replicate runs were combined using LOGCOMBINER v.2.3.2 discarding the first 10% of each run as burn-in. The combined tree samples were processed with TREEANNOTATOR v.2.3.2 to calculate maximum clade credibility (MCC) trees with the same settings as in the individual gene-tree analyses. The same data sets were also analyzed with SPECIESDELIMITATIONANALYSER (Jones et al., 2015), discarding 10% as burn-in and using a ‘collapse-height’ equal to the specified ϵ value (see above). SPECIESDELIMITATIONANALYSER calculated the posterior frequencies of clusterings based on the species-tree distribution and produced tables of clusterings, which were afterwards used to generate and visualize similarity matrices by executing the R script provided by Jones et al. (2015). To test how DISSECT performs in the case of hybridization, all analyses were performed with (seqdata1) and without (seqdata2) individuals of putative hybrid origin.

We performed Bayes factor delimitation (BFD) based on the sequence and AFLP data sets without hybrid individuals (seqdata2 and aflpdata2), using eight different species delimitation scenarios, which were built on the basis of the results of all previous analyses (Figure 2.5a). For the sequence data, BEAUTI v.1.8.0 was used to specify one xml file for each species delimitation scenario, in which DNA and indel data were linked for tree and clock models and unlinked for substitution models. Following the marginal likelihood (ML) driven model comparison on the gene-tree level, a relaxed clock was set for the concatenated plastid markers and a strict clock on all other loci. Substitution models were specified as defined in Table S2.2 and all analyses were run with the Yule process as the species-tree prior, piecewise linear and constant root as the population size model, and UPGMA starting trees. To avoid improper priors, we followed Toprak et al. (2016) and Aydin et al. (2014) using a gamma distribution with shape 2.0 and scale 0.002 for the species population mean

hyperprior and a lognormal prior with mean 0.0 and stdev 1.0 was set for the Yule process birth rate. For all other priors, default values given by BEAUTI v.1.8.0 were accepted. We performed two separate runs for each species delimitation scenario with 500 million generations and a sample frequency of 50,000, which were checked, combined, and processed in the same way as described previously for the DISSECT analyses. Marginal likelihood (ML) values were calculated for each run using both, the path sampling (PS) and stepping-stone (SS) sampling method performed with 100 path steps and a total chain length of 10 million. The same species delimitation scenarios were also tested based on the AFLP data of 53 individuals without putative hybrid origin (aflpdata2). For this purpose we estimated ML values with SNAPP (Bryant et al., 2012) implemented in BEAST v.2.3.2 (Bouckaert et al., 2014) by conducting two separate runs of path sampling (PS) for each scenario using 60 steps, a chain length of 100,000 and a preburnin of 10,000. Priors for the Yule birth rate (λ), the population size parameter (θ), and the backward and forward mutation rates (μ , ν) were accepted as given by BEAUTI v.2.3.2. For the best scenario according to the Bayes factor calculation (see below), two additional MCMC runs were performed, each with 10 million states and a sample frequency of 1,000. Results of the MCMC runs were analyzed with TRACER, combined with LOGCOMBINER and a maximum clade credibility tree with a posterior probability limit of 0.5 was finally constructed with TREEANNOTATOR.

To enable comparisons among the different species delimitation hypotheses, we calculated scenario-wise Bayes factor values (2lnBFs) by conducting the following steps for each data set (seqdata2 and aflpdata2) and ML method (SS, PS) separately: (i) log-ML values were averaged across replicate runs; (ii) 2lnBF values were calculated by taking twice the difference between the averaged log-ML value of the best scenario and all other scenarios (see formula provided by Hedin et al., 2015); (iii) 2lnBF >10 was used as a ‘decisive’ criterion for discriminating between competing species delimitation hypotheses following recommendations of Kass and Raftery (1995).

2.3 Results

2.3.1 AFLP fingerprinting

Visual inspection of raw CEQ trace files showed that fragment detection worked well for all accessions except for one sample (389-3), which showed no analyzable band pattern and was therefore discarded from the following analyses. Automated band scoring and subsequent processing of 0/1-matrices yielded a final data set including 367 polymorphic loci in the range of 100 and 420 bp. Error rates, calculated with a Python script provided by Holland et

al. (2008), were comparable to values from a methodologically similar AFLP study of 19 diploid *Leucanthemum* species in Konowalik et al. (2015). In contrast to the mentioned study, the resolution score was found being quite low and only nine out of sixteen replicates were paired correctly. This result was not surprising as the current study investigated the close-knit *L. ageratifolium*-group while Konowalik et al. (2015) also included some clearly distinct *Leucanthemum* species (mainly members of their so-called group 1).

2.3.2 Detection of potential hybrid individuals

The outcome of the principal coordinates analyses (PCoA), the Neighbor-net network reconstructions, and the maximum-likelihood hybrid index calculations, all based on the same AFLP submatrices, are depicted in Figures 2.2 and S2.1. All analyses showed hybridization patterns between members of the *L. ageratifolium*-group on the one and the codistributed species *L. vulgare* on the other side, except in the case of the allopatric *L. laciniatum*, where no hybrids could be found. In the PCoA graphs, putative hybrid populations were either indicated by the intermediate position of their individuals between ‘pure’ parental populations together with a shift on the second axis (e.g., *L. ligusticum* and *L. monspeliense*), as previously observed in methodologically comparable studies (e.g., Hodkinson et al., 2002; Lihová et al., 2007; Takahashi and Hanyu, 2015), or by the position of single members of such populations in the *L. vulgare* cluster (e.g., *L. ageratifolium* and *L. legraeum*). The Neighbor-net networks, reconstructed on the basis of the same submatrices, showed a higher tendency of incompatible splits between hybrid individuals and ‘pure’ accessions in the case of *L. ageratifolium*, *L. legraeum*, and *L. monspeliense*, discernible by larger ‘boxes’ in the networks of Figure S2.1. However, this pattern was less clear for hybrids of *L. ligusticum*. Overall, the hybrid signal was most obvious when the results of the hybrid index calculations were taken into account (Figure 2.2, right panel). All members of populations with a probable hybrid background according to the PCoA analyses were characterized by intermediate maximum-likelihood hybrid index values. Besides giving evidence for hybrid patterns in the data set, PCoA results provided also useful information about the closeness of relationships between the taxa of the *L. ageratifolium*-group on the one and the widespread *L. vulgare* species on the other side. In the case of *L. laciniatum* and *L. legraeum*, high values for variation were explained by the first principal coordinate, while those values were considerable lower in the case of *L. ageratifolium*, *L. monspeliense* and *L. ligusticum*, indicating a closer relationship between the latter three taxa and *L. vulgare*. Results from network analyses based on AFLP fingerprint data of all 87 accessions (aflpdata1), and 53 individuals that showed no hybrid pattern in the former analyses (aflpdata2), are depicted in Figure 2.3. While the high level of background (hybridization)

noise in the total data set (aflpdata1, Figure 2.3a) resulted in short internal and long terminal branches of the Neighbor-net network, this pattern changed when putative hybrids were excluded (aflpdata2, Figure 2.3b): especially in the case of *L. ligusticum* and *L. legraeantum* internal branches got longer and individuals of the latter taxon were found to form a well-supported cluster (bootstrap value: 91%). It is also recognizable by comparing both networks that individuals of *L. ligusticum* only form a joint (yet unsupported) cluster when putative hybrids with *L. vulgare* were discarded.

Morphological analyses (summarized in Table 2.2 and explicitly depicted in Figures S2.2–S2.5) yielded similar hybridization patterns in the study-group as the molecular study: The majority of representatives of populations of the *L. ageratifolium*-group, identified as being influenced by gene flow with *L. vulgare* in the AFLP study, were found to be characterized by intermediate morphological traits compared to ‘pure’ parental individuals. These morphological features concern (i) less strongly dissected basal and/or cauline leaves, (ii) incompleteness of the corona of ray florets, and (iii) a combination of both characteristics. A detailed and quantitative analysis of cauline and basal leaf outlines yielded considerably lower leaf dissection indices (DI) for hybrid populations in the case of *L. monspeliense* and *L. ligusticum* (Table 2.2; Figures S2.4–S2.5). While ‘pure’ *L. monspeliense* populations showed mean DI values of $DI_c = 8.8$ (7.0–10.2) and $DI_b = 7.8$ (5.4–10.0) for cauline and basal leaves, respectively, specimens of the hybrid population 331 were found to have intermediate DI values [$DI_c = 4.2$ (4.1–4.3) and $DI_b = 4.0$ (3.8–4.3)] when taking measurements of *L. vulgare* into account [$DI_c = 2.8$ (2.6–3.0) and $DI_b = 1.5$ (1.3–1.6)]. A similar result was found for populations of *L. ligusticum* × *L. vulgare*, which showed considerably lower DI values compared to *L. ligusticum* [$DI_c = 3.1$ (2.2–3.9) vs. 5.4 (4.6–7.5) and $DI_b = 2.3$ (1.6–3.5) vs. 4.8 (3.8–6.4)] but higher values than *L. vulgare* (see above). Less obvious, but still discernible were the differences between the leaf shape measurements of specimens of *L. ageratifolium* [$DI_c = 3.8$ (3.4–4.2) and $DI_b = 2.3$ (2.3–2.4)], *L. ageratifolium* × *L. vulgare* [$DI_c = 2.9$ (2.8–3.1) and $DI_b = 1.9$ (1.7–2.1)] and *L. vulgare* (see Table 2.2 and Figure S2.2). However, we found no difference in DI values of cauline leaves between *L. legraeantum* populations 366/384 and 369 [$DI_c = 3.1$ (2.6–3.7)] in comparison with population 383 [$DI_c = 3.0$ (2.7–3.3)], although the latter one showed signs of hybridization with *L. vulgare* in the molecular study as described above. Nevertheless, three out of five specimens of population 383 were found to possess *L. vulgare*-like basal leaves although cauline leaves were similar to those of ‘pure’ *L. legraeantum* populations (see Table 2.2 and Figure S2.3).

2.3 Results

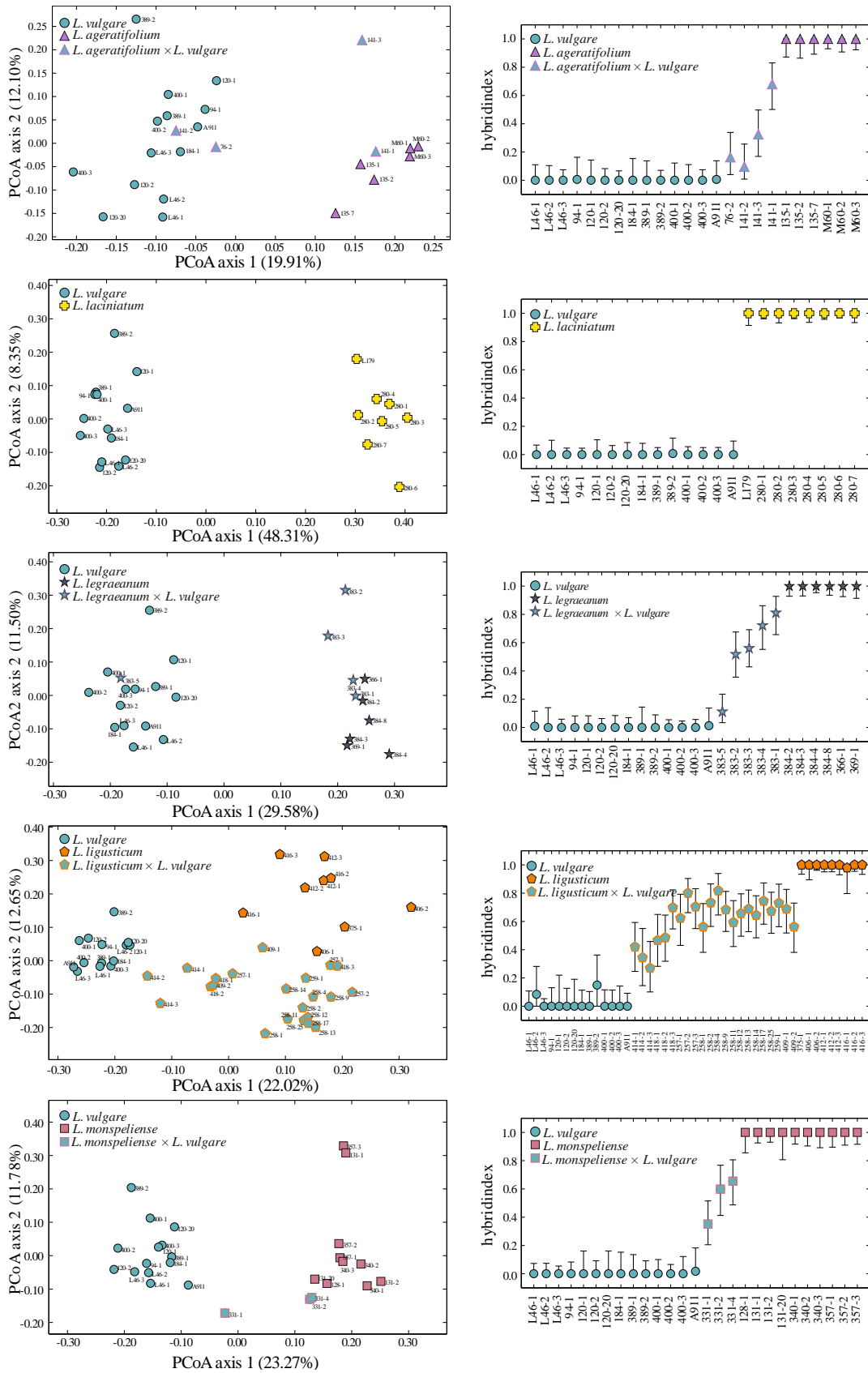


Figure 2.2 Identification of individuals resulting from current hybridization between sympatric *Leucanthemum* taxa using AFLP fingerprint data: Left diagrams show ordinations of taxa based on principal coordinates analyses (PCoA) using a Bray-Curtis dissimilarity matrix. Graphics on the right visualize the results of the maximum-likelihood hybrid index calculations with INTROGRESS. Bars on data points show 95% confidence intervals of hybrid indices.

2.3 Results

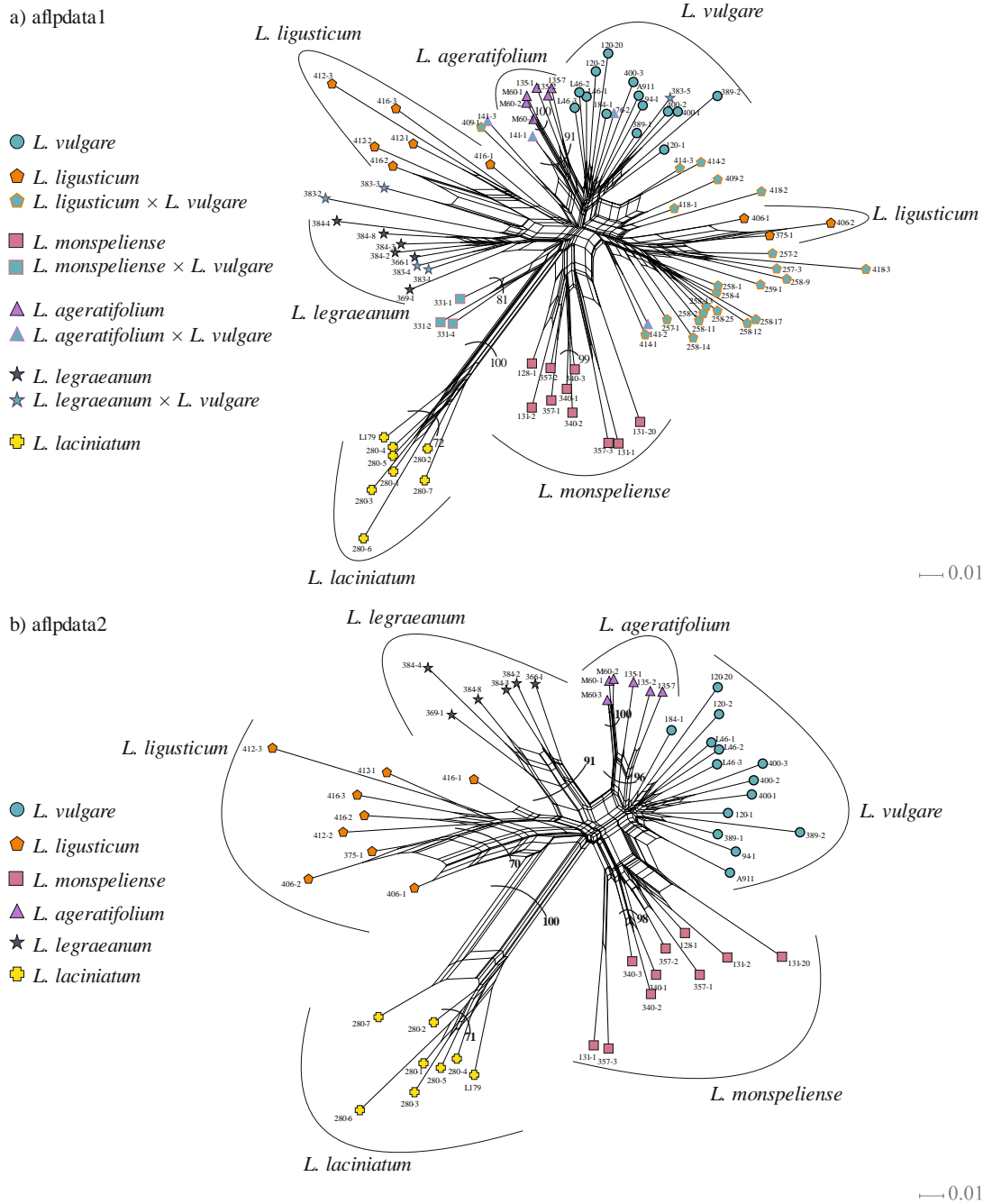


Figure 2.3 Results of network analyses based on AFLP fingerprint data of (a) all 87 accessions (aflpdata1) and (b) a reduced data set excluding 34 admixed individuals as indicated by the PCoA and hybrid index analyses (aflpdata2). Numbers next to the curved bars are support values obtained from neighbour-joining tree searches with 1,000 bootstrap replicates (only values >70 are shown).

2.3.3 Multiple sequence alignments and gene-tree reconstructions

GARD analyses showed no evidence of recombination within any of the studied loci. Alignments of nuclear loci ranged in size from 320 to 374 bp, containing 20–40 variable sites and 11–30 parsimony-informative characters (see Table S2.2). Consistency (CI) and

retention index (RI) analyses resulted in values between 0.75 and 0.98 with a slight tendency to higher values for the former index in the case of seqdata2 (potential hybrids excluded). The concatenated plastid markers had a total length of 2,107 bp and showed less variability with only 14 (9) variable and 11 (8) parsimony-informative sites for seqdata1 (seqdata2), but high consistency and retention indices (1.0 in all cases). The optimal nucleotide substitution models for all nuclear loci and the concatenated plastid markers are shown in Table S2.2.

Marginal likelihood calculations for different clock models using the path sampling (PS) technique in BEAST favored the strict-clock over the relaxed-clock model for nearly all loci in both data sets. Only when putative hybrids were excluded from the analysis of the concatenated plastid markers, the relaxed-clock model produced considerably better results (difference of >3 lnML units; see Table S2.3). The Bayesian gene-tree phylogenies (Figures S2.6–S2.10) varied in their topologies and support values for monophyletic groups of alleles even when potential hybrids were excluded (seqdata2). Nevertheless, there was a general trend noticeable that alleles of *L. laciniatum* show a higher tendency to form well-supported monophyletic groups compared to the alleles of all other taxa under study.

2.3.4 MSC species delimitation

Results from the DISSECT analyses using different data sets (seqdata1 and seqdata2), collapsing height (ϵ), and collapsing weight (ω) parameters are shown in Figures 2.4, S2.11, and S2.12. Varying ϵ and ω did not have any effects on the overall pattern of the similarity matrices produced by DISSECT (Figures S2.11 and S2.12), although the number of sampled clusters was slightly higher when a flat hyperprior for ω was used instead of an informative one. While parameters ϵ and ω had little influence on the analyses, including vs. excluding of hybrid individuals had a clear effect on the outcomes of different runs (Figure 2.4). Analyzing the total data set (seqdata1) resulted in two clearly separated and well-supported clusters (see Figure 2.4a): The first and most distinct cluster comprised the two accessions of the S Italian species *L. laciniatum* ($PP = 1.00$), while a second cluster encompassed hybrid and nonhybrid individuals of the S France lineage *L. legraeum* ($PP = 0.99$). A third and less supported cluster ($PP = 0.80$) was formed by all pure and hybrid individuals of *L. ligusticum*, *L. monspeliense*, *L. ageratifolium*, and *L. vulgare*, with an indication for the separation of the most eastern *L. vulgare* individuals (L46-1, 184-1) plus two ‘*vulgare* \times *ligusticum*’ hybrids (258-1, 259-1) from the remaining group. Excluding hybrids from the analyses (seqdata2) led to an additional, well-supported cluster ($PP = 0.91$) formed by all accessions of *L. ligusticum*. Furthermore, evidence for two more clusters were visible in the similarity matrix as well as in the species tree: One paraphyletic and not supported group formed by all accessions of *L. monspeliense* and *L. ageratifolium* plus a single

accession of *L. vulgare* (94-1) and a second comprising all remaining individuals of *L. vulgare* (Figure 2.4b). Results of Bayes factor delimitation, testing eight different species delimitation scenarios based on the sequence (seqdata2) and the AFLP data set (aflpdata2) without individuals of putative hybrid origin are reported in Table 2.3 and Figure 2.5, respectively. Replicate runs of stepping stone (SS) and path sampling (PS) applied to the sequence data set yielded similar results (Table 2.3) and favored both a five-species model (hypothesis G), which is congruent to the outcome of the DISSECT analyses of seqdata2 (Figure 2.4b). In this scenario, *L. laciniatum*, *L. ligusticum*, and *L. legraeantum* are considered being separate species, while *L. monspeliense* and *L. ageratifolium* are lumped together with the *L. vulgare* accession 94-1. All other *L. vulgare* individuals are united to form a fifth species in this model (Figure 2.5a). However, this hypothesis did not differ considerably from the quite similar five-species hypothesis D when Bayes factors were taken into account (Table 2.3) and using $2\ln\text{BF} < 10$ as a ‘decisive’ criterion. In this alternative scenario, *L. vulgare* and *L. ageratifolium* accessions are assigned to a single species while all other individuals are treated according to the morphological species concept. Species trees for the two just mentioned scenarios G and D both support speciation at the root of the tree into *L. laciniatum* and *L. legraeantum* (Figure 2.5c). However, relationships among the remaining taxa remain unclear due to poor posterior probability values.

Marginal likelihood estimations with the BEAST package SNAPP based on the AFLP data set coincided with the results of the sequence analysis for the scenarios A–D, but showed a contrary pattern concerning the scenarios E–H, which differ in the delimitation pattern of *L. monspeliense*, *L. ageratifolium*, and *L. vulgare* (see Figure 2.5b): While ln-marginal likelihood (lnML) values increased from E to G and sunk abruptly for the last scenario H in the case of the sequencing data, AFLP data resulted in an opposite trend with the highest lnML value being found for the last mentioned scenario. This six-species model, which reflects the traditional and morphology-based species concept of the study group, received ‘decisive’ support compared to all other scenarios tested with the AFLP data set when taking Bayes factor calculations into account (Table 2.3). The species tree calculated for this species-delimitation scenario H indicates a clear separation ($PP = 1.00$) between *L. laciniatum* and the remaining taxa, but again less internal structure, apart from a strongly supported ($PP = 0.99$) sister-group relationship between *L. vulgare* and *L. ageratifolium* (Figure 2.5c). Although excluding individuals of putative hybrid origin from this analysis, there was obviously still a lot of uncertainty in the AFLP data set concerning the relationships among *L. legraeantum*, *L. ligusticum*, and *L. monspeliense*.

Table 2.2 Leaf shape and achene characteristics of 27 population of the *L. ageratifolium*-group. Each population is represented by a characteristic pair of scanned-in leaf-silhouettes (not to scale).






























Taxon	Population code		Division of basal leaves	Leaf dissection index of basal leaves (DI _b)	Division of cauline leaves	Leaf dissection index of cauline leaves (DI _c)	Corona of achenes of ray florets	Adaxial/abaxial length of corona of achenes of ray florets [mm]	Length of achenes [mm]
<i>Leucanthemum ligusticum</i> Marchetti, R. Bernardello, Melai & Peruzzi	375		bipinnatisect	5.2	1-2-pinnatisect	4.9	complete	1.2/0.5	2
	406		N/A	4.5-6.4	1-2-pinnatisect	5.2-5.8	complete	1.5/0.8	2.0-2.3
	412		1-2-pinnatisect	3.9-4.7	1-2-pinnatisect	4.9-7.5	complete	1.2/0.8	2.2-2.3
	416		pinnatisect	3.8-3.9	pinnatisect	4.6-5.1	complete or incomplete	1.0/0.0	2.2-2.3
<i>L. ligusticum</i> × <i>L. vulgare</i>	257		pinnatifid to pinnatipartite	2.1	pinnatifid to pinnatipartite	3.3	complete or incomplete	0.8/0.5-0.1	2
	259/409		serrate to pinnatipartite	1.7-2.8	serrate to pinnatipartite	3.1-3.5	complete	1.0/1.0-0.5	1.8
	414		serrate to pinnatifid	1.7	pinnatipartite	3.4	adaxial scale	0.1/0.0	2
	418		serrate	1.8-2.4	serrate to pinnatifid	2.6-3.6	adaxial scale	0.2/0.0	1.9
	258		serrate to pinnatisect	1.6-3.5	serrate to pinnatisect	2.2-3.9	missing or incomplete or complete	0.0-1.0/0.0-0.3	1.6-2.0
<i>Leucanthemum legraeae</i> (Rouy) B. Bock & J.-M. Tison	366/384		pinnatifid to pinnatipartite	2.1-3.1	pinnatifid to pinnatipartite	2.6-3.3	complete-incomplete	1.5/0.5-0.0	2
	369		pinnatipartite to pinnatisect	2.8-3.2	pinnatipartite	3.5-3.7	complete-incomplete	1.5/0.3-0.0	2.3
<i>L. legraeae</i> × <i>L. vulgare</i>	383		pinnatifid to bipinnatipartite	1.5-2.8	pinnatifid to pinnatipartite	2.7-3.3	incomplete	0.6/0.0	2.0-2.2

Table 2.2 Continued.

Taxon	Population code		Division of basal leaves	Leaf dissection index of basal leaves (DI _b)	Division of cauline leaves	Leaf dissection index of cauline leaves (DI _c)	Corona of achenes of ray florets	Adaxial/abaxial length of corona of achenes of ray florets [mm]	Length of achenes [mm]
<i>Leucanthemum monspeliense</i> (L.) H.J.Coste	340		bipinnatisect	5.4-5.6	bipinnatisect	8.8-9.9	complete	1.0/0.8	2
	357		bipinnatisect	7.4-9.3	bipinnatisect	7.3-9.5	incomplete	1.2/0.0	N/A
	128		bipinnatisect	9.0	bipinnatisect	7.0	missing	0.0/0.0	1.8
	131		bipinnatisect	10.0	bipinnatisect	10.2	complete-incomplete	0.7/0.2-0.0	N/A
<i>L. monspeliense</i> × <i>L. vulgare</i>	331		pinnatipartite	3.8-4.3	pinnatipartite	4.1-4.3	incomplete	1.0/0.0	N/A
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	400		serrate	1.3	serrate	3.0	missing/adaxial scale	0.0-0.2/0.0	1.8 (unripe)
	389		serrate	1.6	serrate	2.6	missing	0.0/0.0	2
	120		serrate	1.3	serrate	2.9	missing	0.0/0.0	N/A
	L46		serrate to pinnatifid	1.6	serrate to pinnatifid	2.7	missing/adaxial scale	0.0-0.1/0.0	1.8 (unripe)
	184		serrate	1.6	serrate	2.8	incomplete	1.2/0.0	N/A
	94		serrate	1.5	serrate	2.7	missing	0.0/0.0	2.0 (unripe)
<i>Leucanthemum ageratifolium</i> Pau	135		pinnatifid to pinnatipartite	2.3	pinnatipartite	3.4	incomplete	1.0/0.0	1.7-1.8
	M60		pinnatipartite	2.3-2.4	pinnatipartite	3.8-4.2	missing/adaxial scale	0.0-0.1/0.0	N/A
<i>L. ageratifolium</i> × <i>L. vulgare</i>	141		serrate to pinnatifid	2.1	serrate to pinnatifid	2.8	missing/adaxial scale	0.0-0.1/0.0	N/A
	76		serrate	1.7	pinnatifid to pinnatipartite	3.1	missing	0.0/0.0	2

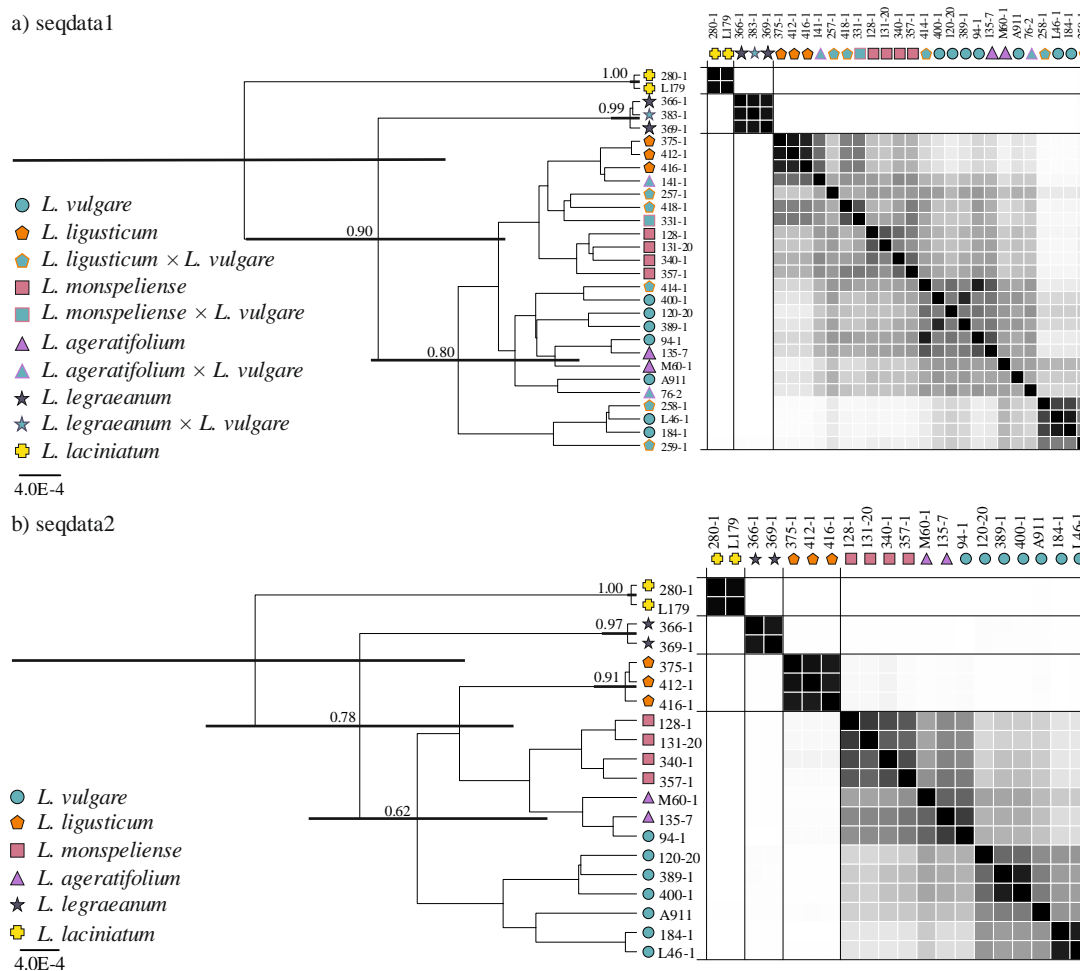


Figure 2.4 Results of the joint species-tree and clustering analyses using the BEAST application DISSECT based on (a) the complete data set (seqdata1) and (b) the nonhybrid data set (seqdata2), respectively. Similarity matrices to the right of the species trees visualize posterior probabilities (PP) for pairs of individuals to belong to the same cluster (black: $PP = 1.0$; white: $PP = 0.0$). Bars at nodes of the species-trees show the 95% highest posterior densities (HPD) for node heights and posterior probability values above 0.5 are shown besides the corresponding nodes. The calibration of the scale bar is substitutions per sites.

2.4 Discussion

2.4.1 Species delimitation and species concepts

With the biological species concept (Mayr, 1942) in mind, it seems to be contradictory to ask for species delimitation in a group of hybridizing species, as the main feature of this concept is the development of reproductive barriers in the speciation process. However, in his review dealing with species concepts and species delimitation, De Queiroz (2007) argued against a confusion of the issue of species delimitation with that of species conceptualization and provided a unified species concept, which defines the existence as a ‘separately evolving metapopulation lineage’ as the only necessary criterion for a species. Following this concept, characteristics like reproductive isolation, monophyly or ecological divergence, being the defining properties of the biological, phylogenetic (Rosen, 1979), and ecological (Paterson, 1985) species concept, respectively, are considered being only contingent properties evolving

in a successive, but randomly progressional manner, and may or may not be conjointly detectable in the continuous process of lineage divergence (De Queiroz, 2007, 2011). As a consequence of his concept, De Queiroz (2007) pleaded for a shift of awareness away from the ‘traditional species criteria’ to new methods for species delimitation, for example, in the framework of the multi-species coalescence (MSC) theory.

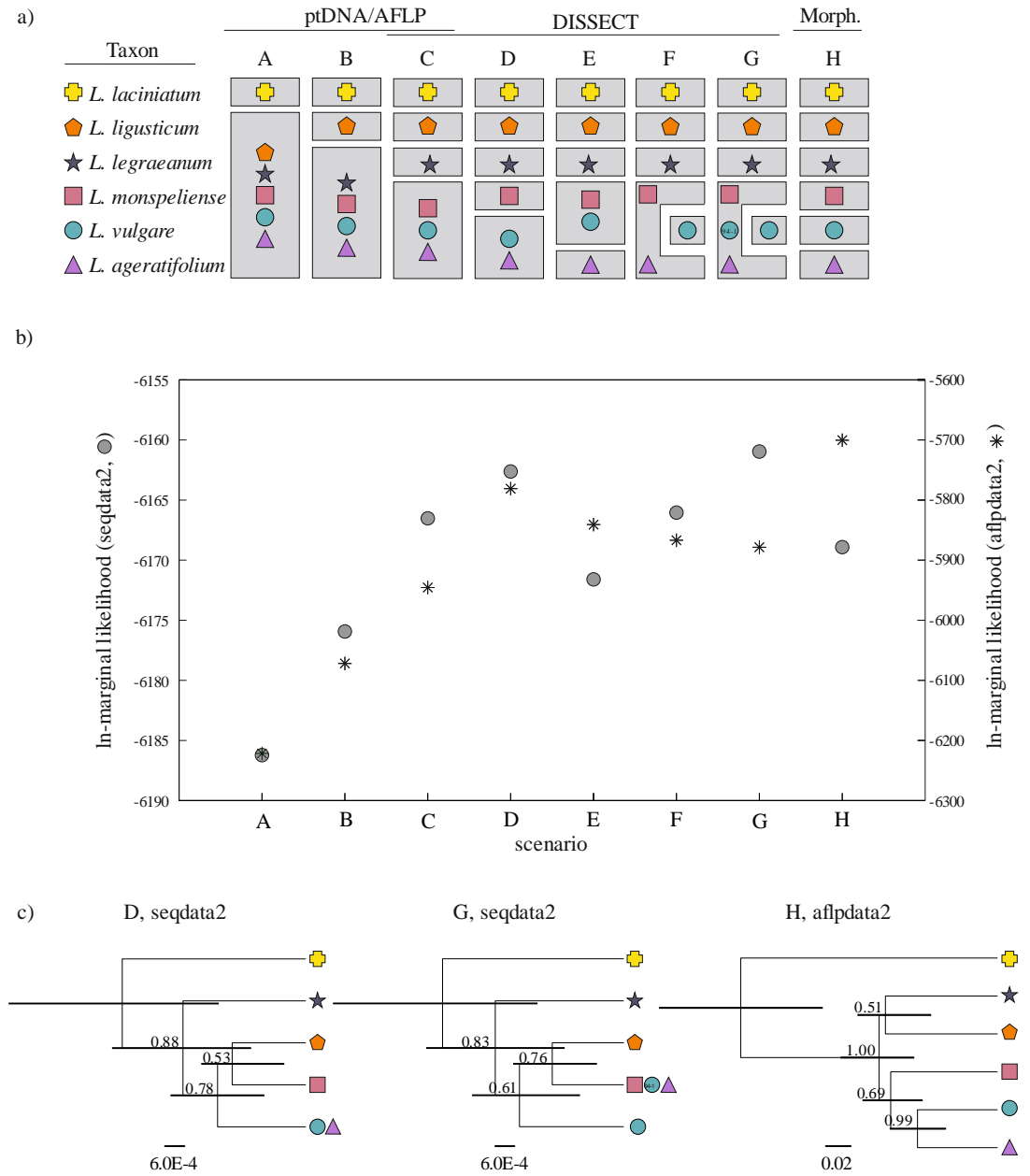


Figure 2.5 (a) Species delimitation hypotheses A-H (columns) and the corresponding combination of taxa (rows). Hypotheses A–C are based on the plastid gene-tree (ptDNA) and AFLP results, hypotheses C–G were generated according to the results from the DISSECT analysis and hypothesis H represents the traditional species concept based on morphological traits (Morph.). (b) Logarithmic marginal likelihood values for each scenario averaged over two replicate runs of path sampling (PS) for each data set (seqdata2: left axis, aflpdata2: right axis). (c) Species trees for the best scenarios according to Bayes factor analyses (see Table 2.3), including posterior probability values above 0.5 and 95% highest posterior densities (HPD) for node heights.

2.4.2 Species delimitation and hybridization

At the moment, the vast majority of coalescent-based species delimitation methods consider incomplete lineage sorting (ILS) via the coalescent model, but do not account for gene flow after divergence (Fujita et al., 2012). One exception of this disability is the study of Camargo et al. (2012), in which the authors used Approximate Bayesian Computation (ABC) to incorporate gene flow in the species delimitation process of an Argentinean lizard complex comprising parapatric and sympatric lineages. However, simulations performed in the same study, proved ABC showing only intermediate accuracy compared to other methods (e.g., BPP) despite of apparently being almost immune to the effects of gene flow for detecting lineage separation. Moreover, it was shown that ABC was the computationally least efficient species delimitation method evaluated in Camargo et al. (2012), which was due to the fact that it relies on the generation of simulated data. The present contribution describes a relatively simple approach for delimiting species in the presence of hybridization based on the a priori detection of potential hybrid individuals. Once all candidates for actual interbreeding are identified and removed, the full range of MSC methods can be used for delimiting species, without running into the risk of violating the model assumption of no genetic exchange after speciation.

Table 2.3 Logarithmic marginal likelihood (lnML) and Bayes factor (2lnBF) values for eight species delimitation hypotheses calculated on the basis of the hybrid excluded sequence- and AFLP-datasets (seqdata2 and aflpdata2) using replicate runs of stepping stone and path sampling. Species hypotheses are defined as in Figure 2.5 and supplemented by the number of comprising species (sp). Best scenarios (2lnBF<10) for each dataset and ML-method, following the ‘decisive’ criterion of Kass and Raftery (1995), are highlighted in bold.

	Stepping stone (seqdata2)				Path sampling (seqdata2)				Path sampling (aflpdata2)			
	lnML			2lnBF	lnML			2lnBF	lnML			2lnBF
	run 1	run 2	mean		run 1	run 2	mean		run 1	run 2	mean	
A (2 sp)	-6186.14	-6186.51	-6186.32	50.46	-6186.00	-6186.43	-6186.22	50.50	-6221.32	-6221.67	-6221.50	1042.23
B (3 sp)	-6176.12	-6176.19	-6176.15	30.12	-6175.90	-6175.97	-6175.94	29.94	-6072.80	-6070.90	-6071.85	742.94
C (4 sp)	-6166.86	-6166.72	-6166.79	11.38	-6166.61	-6166.44	-6166.52	11.11	-5946.07	-5945.09	-5945.58	490.40
D (5 sp)	-6162.78	-6162.96	-6162.87	3.54	-6162.50	-6162.76	-6162.63	3.33	-5781.23	-5780.93	-5781.08	161.39
E (5 sp)	-6172.18	-6171.32	-6171.75	21.31	-6171.92	-6171.28	-6171.60	21.27	-5841.17	-5840.57	-5840.87	280.97
F (5 sp)	-6165.84	-6166.46	-6166.15	10.10	-6165.74	-6166.35	-6166.04	10.16	-5866.46	-5867.42	-5866.94	333.12
G (5 sp)	-6161.24	-6160.96	-6161.10	N/A	-6161.05	-6160.88	-6160.97	N/A	-5878.86	-5878.19	-5878.52	356.29
H (6 sp)	-6169.06	-6169.09	-6169.07	15.95	-6168.95	-6168.89	-6168.92	15.90	-5700.23	-5700.54	-5700.38	N/A

2.4.3 Detection of hybridization patterns in the *L. ageratifolium* group

The AFLP-fingerprinting data set used for the detection of hybridization patterns in the *L. ageratifolium* group showed error rates that were comparable to values from a methodologically similar AFLP study of 19 diploid *Leucanthemum* species in Konowalik et al. (2015), calculated with the same Python script provided by Holland et al. (2008). However, the resolution score was found being quite low compared to Konowalik et al. (2015) and only nine out of sixteen replicates were paired correctly. This result was not surprising as the current study investigated the close-knit *L. ageratifolium*-group while Konowalik et al. (2015) also included some clearly distinct *Leucanthemum* species (mainly members of their so-called group 1).

We used a combination of multivariate statistics (PCoA), a maximum-likelihood-based hybrid index calculation and Neighbor-net network analyses for the identification of hybrid population formed by members of the *L. ageratifolium*-group and the widespread and codistributed species *L. vulgare*. Contrary to other studies dealing with hybridization (cf. Oberprieler et al., 2010, 2011b and 2013), we consciously decided not to use the admixture model implemented in the software STRUCTURE (Pritchard et al., 2000) for the detection of hybrid individuals, because this application proved to be not helpful for this purpose in an investigation of *Leucanthemum* diploids with a similar setup [see Konowalik et al. (2015) for a detailed discussion]. In contrast, there are several independent indications, that our here presented hybrid detection procedure provides reliable results: (i) all three applied AFLP-based methods (PCoA, hybrid index calculation, Neighbor-net analyses) show a highly congruent hybridization pattern in the investigated group and uncover the same populations as being affected by hybridization (Figures 2.2, 2.3 and S2.1). (ii) inclusion of sequences of individuals, which were identified as potential hybrids according to the AFLP data analyses, resulted in an expected homogenizing effect of hybridization on species delimitation carried out with the BEAST application DISSECT. This is especially true in the case of *L. ligusticum*, which shows an intense signal of gene flow with the codistributed *L. vulgare* (compare Figure 2.3, Figure 2.4 and the detailed discussion on DISSECT results below). (iii) the survey of leaf shapes [calculation of dissection indices (DI) for cauline and basal leaves] and pappus characteristics of ray achenes in the study group indicates, that nearly all representatives of populations, identified as being affected by hybridization by molecular means, also show intermediate morphological features compared to ‘pure’ individuals (see Table 2.2 and Figures S2.2–S2.5).

Our study provides clear evidences for hybridization between all members of the *L. ageratifolium*-group and the widespread species *L. vulgare*, with the only exception of *L. laciniatum*. All results from our study unambiguously demonstrate that this S Italian

endemic taxon is clearly distinct from the remaining lineages and a reproductive barrier may be already established in this species, which prevents hybridization events with *L. vulgare*. All other taxa have in common, that at least one population is affected by hybridization with the sympatrically distributed *L. vulgare*, which is in line with the strong signal of gene flow among diploids of *Leucanthemum* observed in Konowalik et al. (2015) and with crossing experiments carried out by Villard (1970) and Przywara (1974), which suggested the lack of intrinsic postzygotic isolation factors in the closely-knit taxon group around *L. vulgare* and *L. gaudinii*.

2.4.4 MSC species-delimitation

After removal of candidate hybrids from the sequence and AFLP data sets (seqdata2 and aflpdata2), we were able to uncover species boundaries in the study-group by executing all three currently available species delimitation methods implemented in BEAST (DISSECT, BFD, BFD*) without violating the assumption of no gene flow after speciation in the MSC model. We have consciously decided to use these applications and not the popular BPP approach, because this allowed us to adjust different clock models for the particular loci for the evaluation of the sequence data sets (DISSECT and BFD analyses), following the results of our marker-wise and marginal likelihood-based model comparison. Furthermore, by performing Bayes factor delimitation with the BEAST application *BEAST (BFD) and SNAPP (BFD*), we were able to evaluate the influence of different kind of data (AFLP and sequence data) on the results of our species delimitation analyses.

We used DISSECT for a first discovery analysis of the sequence data, because it works without prior assessment of individuals to species. This approach enabled us to identify three well-separated species, namely *L. laciniatum*, *L. legraeum*, and *L. ligusticum*, in contrast to the less distinct other members of the group, comprising *L. vulgare*, *L. ageratifolium*, and *L. monspeliense* (Figure 2.4b). Finding the two latter taxa being hardly distinguishable from the widespread species *L. vulgare* was rather surprising from a morphological point of view (see differences in the division of basal and cauline leaves in Table 2.2), but was in line with the results of the sequence-based BFD analysis, where the ambiguity concerning the delimitation of the three taxa resulted in two equal-supported scenarios, lumping either *L. vulgare* and *L. ageratifolium* or *L. monspeliense* and *L. ageratifolium* (plus one *L. vulgare* accession) together (Figure 2.5b). In contrast to the results of the sequence-based analyses, Bayes factor calculation using AFLP data (BFD*) led to a clear separation of all six taxa, which is in agreement with the traditional and morphology-based species classification.

We think that the equivocal results concerning species delimitation in our study group may be caused by differences in divergence times and effective population sizes of investigated

taxa as well as being due to unequal information content of sequence and AFLP data sets. In their empirical study addressing the influence of locus number and information content on species delimitation in the Mexican salamander species *Ambystoma ordinarium*, Hime et al. (2016) showed that shallowly diverged species can fail passing statistical validation via coalescent tests due to a lack of sufficient sequence information. In their species delimitation study using the software BPP and a varying number of loci [rank ordered by the number of parsimony-informative (PI) sites], the authors proofed that as few as $n = 10$ of the most informative loci (mean number of PI sites: 13.4) were enough to separate the clearly distinct western and eastern lineages of their study group, but that a considerably higher number of loci ($n = 30$, mean number of PI sites: 8.8 and $n = 50$, mean number of PI sites: 6.7) was necessary for the detection of the more shallowly diverged species within western (WE1–WE2) and eastern (EA1–EA2) localities [see Figure 5D and Table S3 in Hime et al. (2016)]. These findings may indicate that the total number of $n = 9$ sequenced loci [mean number of PI sites: 12.1 (seqdata1) and 10.4 (seqdata2), respectively] in our present study is indeed enough for delineating the clearly distinct species *L. laciniatum*, *L. legraeantum*, and *L. ligusticum*, but that the higher amount of loci generated via AFLP fingerprinting is necessary for separating the shallowly diverged group comprising *L. monspeliense*, *L. ageratifolium*, and *L. vulgare*.

The difficulty concerning the delimitation of the latter three taxa in the sequencing study is probably also caused by the simultaneous occurrence of low and high effective population size values N_e , in the study group, which can have an obscuring effect on species boundaries (Naciri and Linder, 2015). Considering the fact, that *L. laciniatum*, *L. legraeantum*, and also *L. ligusticum* are narrowly endemic species, which comprise only a few populations (Bock and Tison, 2012; Marchi, 1982; Melai et al., 2012) whereas *L. monspeliense*, *L. ageratifolium*, and especially *L. vulgare* show wider distribution ranges (see Figure 2.1) and higher population numbers (Vogt, 1991), differences in the amount of incomplete lineage sorting connected with N_e may explain the difficulty to delimit these three taxa from each other in contrast to the remaining members of the group.

The potential impact of gene flow on delineating species boundaries was already mentioned before and by analyzing the total data set, including also sequences of potential hybrid individuals, enabled us to gain insights into the effect of hybridization on the robustness of the recently developed DISSECT method. Our empirical results indicate that the accuracy of delimiting a particular species in DISSECT depends on how intensive it is affected by hybridization. While a low hybridization signal in the case of *L. legraeantum* (Figure 2.2) had no significant effect on either the species tree or the similarity matrix, we recognized a strong homogenizing effect in the case of *L. ligusticum*, where an intensive hybridization pattern (Figure 2.2) led to a complete obscuring of the species boundary (Figure 2.4). This behavior

of DISSECT appears to be consistent with that of BPP as evaluated in the simulation study of Zhang et al. (2011). In this study it was shown that low rates (<0.1 migrants per generation) of gene flow does not affect the accuracy of BPP even with a small sample configuration and only a few examined loci, while higher migration rates ($\gg 10$) have a homogenizing effect on Bayesian species delimitation under all conditions. More research in terms of simulation studies with varying intensities of gene flow, number of sequences, and different values for the MSC model parameters is needed to fully evaluate the performance of DISSECT in the presence of hybridization. We think, however, that the here presented study is a contribution to the understanding of the effect of gene flow on species delimitation studies working in the framework of the MSC and shows a possible way of how to deal with both phenomena, without violating model assumptions.

2.4.5 Phylogenetic considerations and taxonomic implications

Allopatrically distributed and morphologically similar, but distinctly differentiated populations or population groups pose a considerable problem to the taxonomist. On the one hand, actual natural interbreeding as a criterion for the application of a reproductive ('biological') species concept (BSC, Mayr, 1942) is logically inapplicable. On the other hand, getting information about potential interbreeding among members of allopatric populations is time-demanding, corrupted by experimentation under artificial common-garden conditions (Coyne and Orr, 2004), and often phylogenetically misleading, with closely related species being reproductively well-isolated while distantly related ones being often easily crossable even after extremely long times of divergence (and classification even in different genera; Stuessy, 2009). In particular, in higher plants, where evolutionary lineages may remain independent from each other despite gene flow through hybridization among them, multi-locus coalescent-based species delimitation methods could be extremely helpful in the process of evaluation of genetic independence and divergence of populations for backing taxonomic decisions on taxon circumscription and ranking.

The *Leucanthemum ageratifolium*-group was here defined by the possession of deeply dissected leaves, which is a quite uncommon feature in the genus (Vogt, 1991). This characteristic also occurs in the diploid *L. pluriflorum* Pau from NW Spain (Greiner et al., 2013; Oberprieler et al., 2014; Vogt, 1991), in the tetraploid *L. corsicum* subsp. *fenzlii* Gamisans (Marchi, 1982), and in the hexaploid *L. coronopifolium* Vill. (subsp. *ceratophylloides* and subsp. *tenuifolium*; Marchi, 1982) and *L. visianii* (Gjurasin) Vogt & Greuter. However, while the latter taxa were excluded from the present study due to their polyploid nature, *L. pluriflorum* was considered being unrelated to the other members of the

L. ageratifolium-group due to an unique plastid haplotype (Greiner et al., 2012) and a probable homoploid hybrid origin (Konowalik et al., 2015).

Despite the morphological similarities and the allegedly telltale allopatric distribution pattern, the members of the *L. ageratifolium*-group were not found constituting a monophyletic evolutionary lineage in the present study because of the closer relationship of *L. ageratifolium* with *L. vulgare* than with the other members of the group (Figure 2.5c). This corroborates a phylogenetic reconstruction for diploid *Leucanthemum* taxa made by Konowalik et al. (2015), where the five taxa of the present study were also found in three different lineages of a species tree based on ten gene-trees (from nine single-copy nuclear markers and spacers of the plastid genome). The mentioned study (Konowalik et al., 2015), however, should be considered preliminary; especially with regard to the *L. ageratifolium* group it has to be interpreted with restraint because of the low number of individuals/populations analyzed (usually 2–3 per taxon) and due to the fact that *L. ligusticum* (with only a single accession) was represented by an individual (accession 258–1) from a here uncovered hybrid swarm (population 258). With more accessions analyzed per taxon and the a priori elimination of hybrid individuals based on AFLP fingerprinting, we therefore consider our present phylogenetic reconstructions in the *L. ageratifolium*-group and its relationship to *L. vulgare* more trustworthy than that of Konowalik et al. (2015).

Despite proven occasional hybridization with *L. vulgare* in the two latter cases, our present analyses—both based on AFLP fingerprinting and sequence data—revealed the taxonomical independence and phylogenetic distance of *L. laciniatum*, *L. ligusticum*, and *L. legraeantum* from *L. vulgare* (Figure 2.5). This situation in the eastern part of the distribution range of the study group is obviously counterbalanced by less clear relationships among the three taxa found in the western part, where sequence-based species-delimitation methods are equivocal about the assignment of *L. ageratifolium* accessions to either *L. monspeliense* (Figure 2.5, scenario G) or *L. vulgare* (scenario D) on the one hand and AFLP data are supportive of a three-species scenario (scenario H) or the *L. ageratifolium*-*L. vulgare*-conspecificity scenario (scenario D) on the other. While equivocality of the two sequence-based scenarios (scenarios D and G) and leaf characteristics of *L. ageratifolium* being intermediate between *L. monspeliense* (bipinnatisect) and *L. vulgare* (serrate to pinnatifid) may argue for a hybrid origin of the former taxon, there are also arguments against that interpretation. The first comes from the sequence-based, multi-locus species-tree analysis of Konowalik et al. (2015) who found *L. ageratifolium* (sub *L. vulgare* subsp. *pujiulae* Sennen) exhibiting a relatively low hybrid index (gene-tree incongruence) score solely ascribable to the effects of incomplete lineage sorting (ILS). The second is the nonintermediate position of *L. ageratifolium* individuals in the networks based on AFLP-fingerprinting data (Figure 2.3), where closer relationships are found with *L. vulgare* than with *L. monspeliense*. As a

consequence, either acknowledgement of *L. ageratifolium* as an independent species or its treatment as a subspecies of *L. vulgare* (as *L. vulgare* subsp. *pujiulae* Sennen) are possible classification schemes here. As a consequence, genetically intermediate accessions found in accessions from populations 76 and 141 may then be treated as either hybrids between the two species or as just transient forms between two subspecies of *L. vulgare*. However, due to the lack of detailed information concerning the distribution of *L. vulgare* south of the Pyrenees (treated as *L. vulgare* s.l. by Vogt, 1991) it is unclear whether the two units are sympatric in NE Spain (arguing for independent, ecologically differentiated, but occasionally hybridizing species) or whether *L. ageratifolium* peripatrically substitutes *L. vulgare* at the SW fringe of its distribution (arguing for acknowledgement of the two taxa as subspecies of the same species). Only a denser sampling of these two taxa in the area for morphological and genetic analyses, preferably complemented by detailed ecological data of habitats and crossing experiments, may allow a final judgement. The invasiveness of *L. vulgare*, however, which is found growing on road embankments and in other anthropogenically influenced habitats (Vogt, 1991), may further hamper these analyses. For the time being, we consider the morphological differences of *L. ageratifolium* (pinnatifid to pinnatipartite leaves, involucre bracts with pale membranous margins) sufficient for its acknowledgement as an independent species, refraining however from describing morphologically and genetically transient forms as hybrids.

Owing to the fact that in all other cases genetically transient individuals are formed by taxa being phylogenetically more distant than the sister-taxa *L. ageratifolium* and *L. vulgare*, we formally describe the three observed hybrid combinations of *L. vulgare* on the one side and *L. ligusticum*, *L. legraeum*, and *L. monspeliense* on the other side as three nothospecies new to science. Commemorating the joint excursions of three of the present authors (CO, RV, FW) to southern France and Liguria during the last years hunting for *Leucanthemum* populations, we would like to devote these three hybrids to Alexandre Dumas' heroes in the novel *Les Trois Mousquetaires* (Dumas, 1844), Athos, Porthos, and Aramis. "Un pour tous, tous pour un!" (One for all, all for one!).

- (1) *Leucanthemum* × *athosii* Flor. Wagner, Vogt & Oberpr. in Mol. Ecol. 24: 4280. 2017. [*Leucanthemum vulgare* (Vaill.) Lam. × *L. monspeliense* (L.) H.J. Coste].

Type: France, Rhone-Alpes, Département Loire, Saint-Etienne, valley of river Loire near Essaloir between Chambles and Saint Rambert, steep slopes N of the dam of the "Barrage de Grangent", 45°28'4.0"N–04°14'56.9"E, 404 m, 03.06.2013, R. Vogt 17147, C. Oberprieler 10872 & F. Wagner [holotype: B (B100486652)].

Diagnosis: In terms of leaf dissection, with pinnatipartite to pinnatisect lower cauline leaves intermediate between *Leucanthemum vulgare* (Vaill.) Lam. (serrate to pinnatifid) and *L. monspeliense* (L.) H.J. Coste (bipinnatisect).

Notes: Presently this hybrid is only known from its *locus classicus* at the northern edge of the distribution range of *L. monspeliense*, which is restricted to the Massif Central in S France.

- (2) *Leucanthemum* × *porthosii* Flor. Wagner, Vogt & Oberpr. in Mol. Ecol. 24: 4280. 2017. [*Leucanthemum vulgare* (Vaill.) Lam. × *L. legraeantum* (Rouy) B. Bock & J.-M. Tison].

Type: France, Provence-Alpes-Côte d'Azur, Département Var, Massif des Maures, on road D88 in Vallée du Pansard north of La Londe-les-Maures, escarpments along the road in macchia and *Quercus suber* woodland near the creek, 43°11'10.2"N–06°12'45.2"E, 77 m, 30.05.2015, R. Vogt 17432, C. Oberprieler 10913 & F. Wagner [holotype: B (B100627807); isotypes: B (B100627805); B (B100627806); M; P].

Diagnosis: In terms of leaf dissection, with pinnatifid to pinnatipartite lower cauline leaves similar to *L. legraeantum* (Rouy) B. Bock & J.-M. Tison, but with shorter pappus on achenes of ray florets (adaxially 0.6 mm vs. 1.5 mm long, abaxially 0 mm vs. 0–0.5 mm long). As in *L. legraeantum* with pale to light-brown margins of involucre bracts [vs. dark-brown margins in *L. vulgare* (Vaill.) Lam.].

Notes: Presently this hybrid is only known from its *locus classicus* at the southern border of the Massif des Maures, where it grows together with its parental taxa at altitudes relatively low for *L. legraeantum*.

- (3) *Leucanthemum* × *aramisii* Flor. Wagner, Vogt & Oberpr. in Mol. Ecol. 24: 4280. 2017. [*Leucanthemum vulgare* (Vaill.) Lam. × *L. ligusticum* Marchetti, R. Bernardello, Melai & Peruzzi].

Type: Italy, Liguria, Province of La Spezia, Rocchetta di Vara, along Via Battaglione Vanni N of Rocchetta di Vara, waste places, 44°15'18"N–9°45'17"E, 228 m, 15.06.2011, R. Vogt 16943 & C. Oberprieler 10850 [holotype: B (B 10 0350184); isotype: FI].

Diagnosis: In terms of leaf dissection, with serrate to pinnatifid or 1-2-pinnatipartite lower cauline leaves intermediate between *Leucanthemum vulgare* (Vaill.) Lam. (serrate to pinnatifid) and *L. ligusticum* Marchetti, R. Bernardello, Melai & Peruzzi (pinnatisect to bipinnatisect).

Notes: This hybrid is considerably widespread in Liguria (NW Italy), where it grows both in close vicinity of its parental taxon *L. ligusticum* but also independently of it. In the latter case, the hybrid populations are only recognizable because of their basal and lower cauline leaves being more intensively dissected than *L. vulgare*. Presumably, some (if not all) indications of *L. legraeantum* for NW Italy by Briquet (1916; e.g., “env. de Mondovi” corresponding to population 418 of the present study) and Bernardello et al. (2015) relate to this nothospecies. This may indicate that *L. ligusticum* was once more broadly distributed in NW Italy, but lost terrain through hybridization with the invasive *L. vulgare* and is now found in ‘pure’ populations only in geographically (and possibly edaphically) restricted habitats.

2.5 Supplemental Figures and Tables

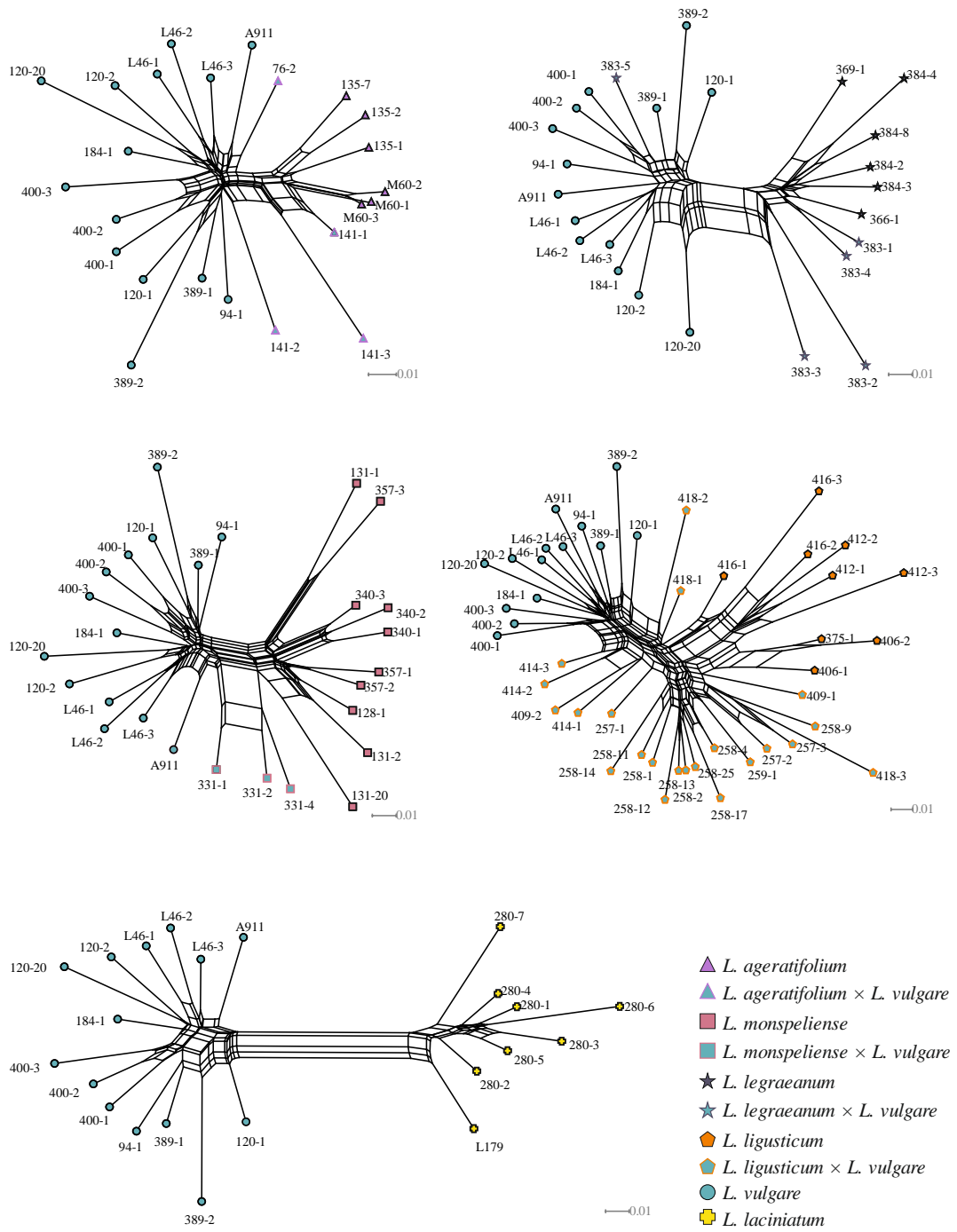


Figure S2.1 Identification of individuals resulting from current hybridization between sympatric *Leucanthemum* taxa using AFLP fingerprint data: Neighbor-net networks were calculated in SPLITSTREE v4.13.1 (Huson and Bryant, 2006) using Nei and Li's (1979) restriction-site distance coefficient as implemented in PAUP* (Swofford, 2003).

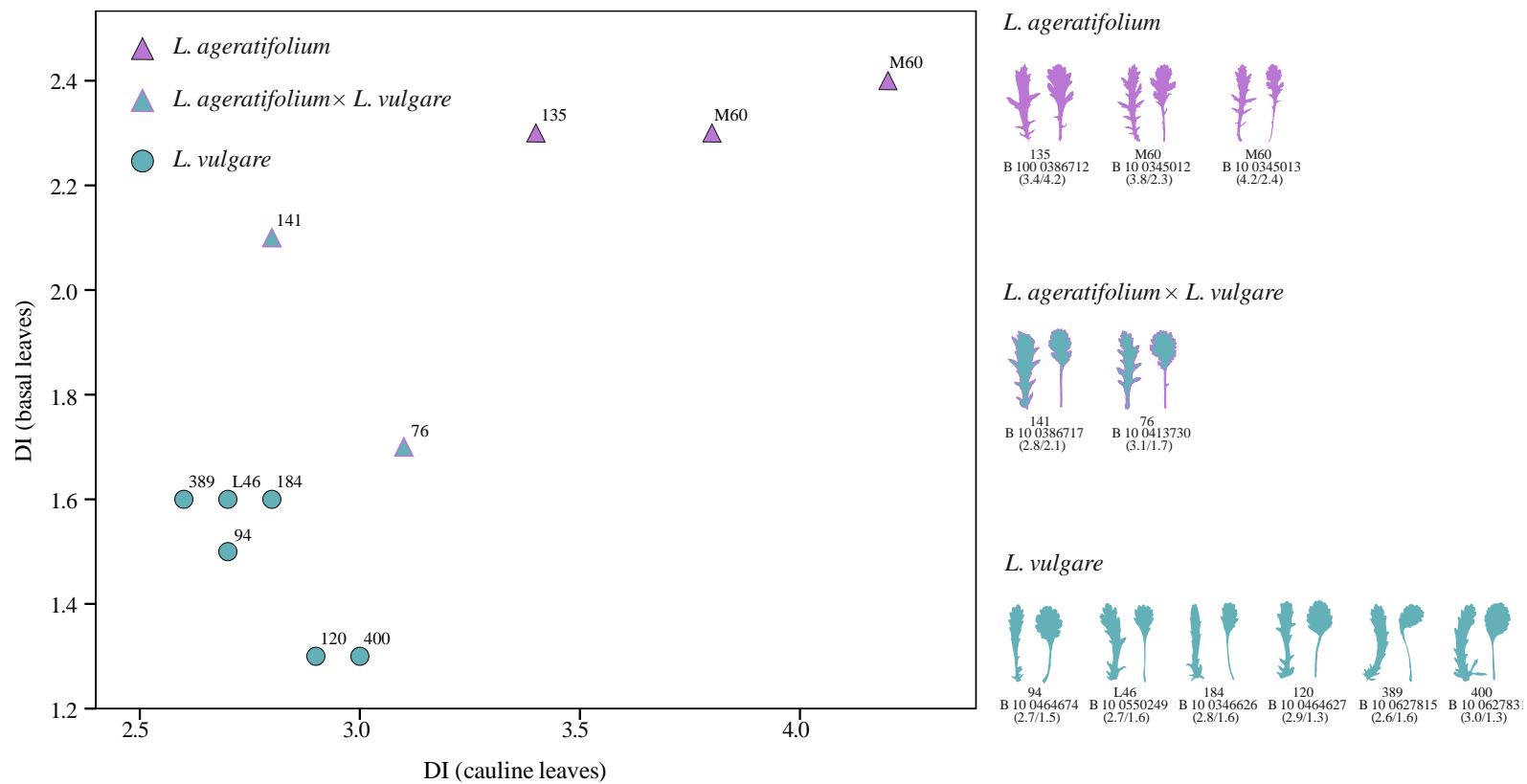


Figure S2.2 Dissection indices (DI) of cauline leaves plotted against DI values of basal leaves for herbarium specimens of *L. ageratifolium*, *L. ageratifolium* × *L. vulgare*, and *L. vulgare*. Silhouettes of leaves on the right side are labelled by population code, herbarium voucher, and DI values (cauline leaf/basal leaf). Leaves are not drawn to scale.

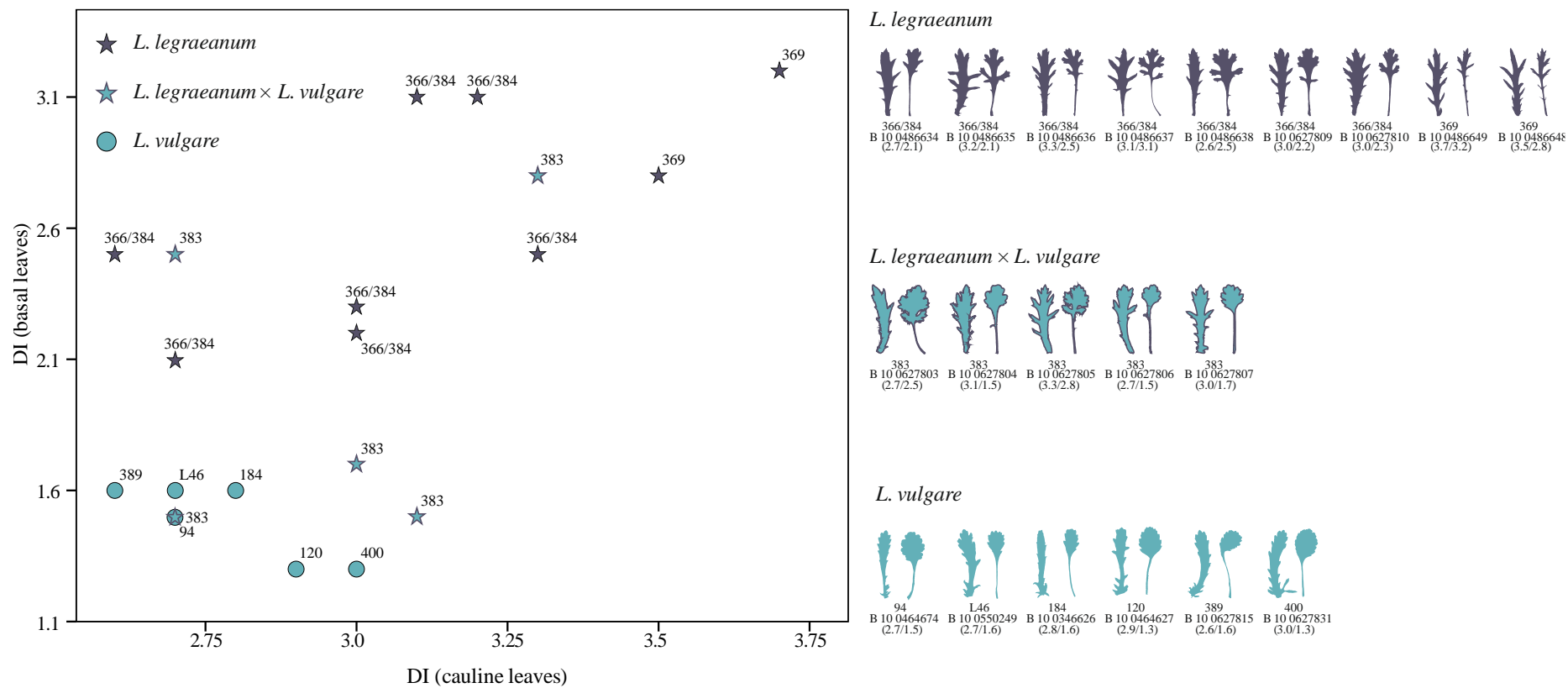


Figure S2.3 Dissection indices (DI) of cauline leaves plotted against DI values of basal leaves for herbarium specimens of *L. legraeantum*, *L. legraeantum* × *L. vulgare*, and *L. vulgare*. Silhouettes of leaves on the right side are labelled by population code, herbarium voucher, and DI values (cauline leaf/basal leaf). Leaves are not drawn to scale.

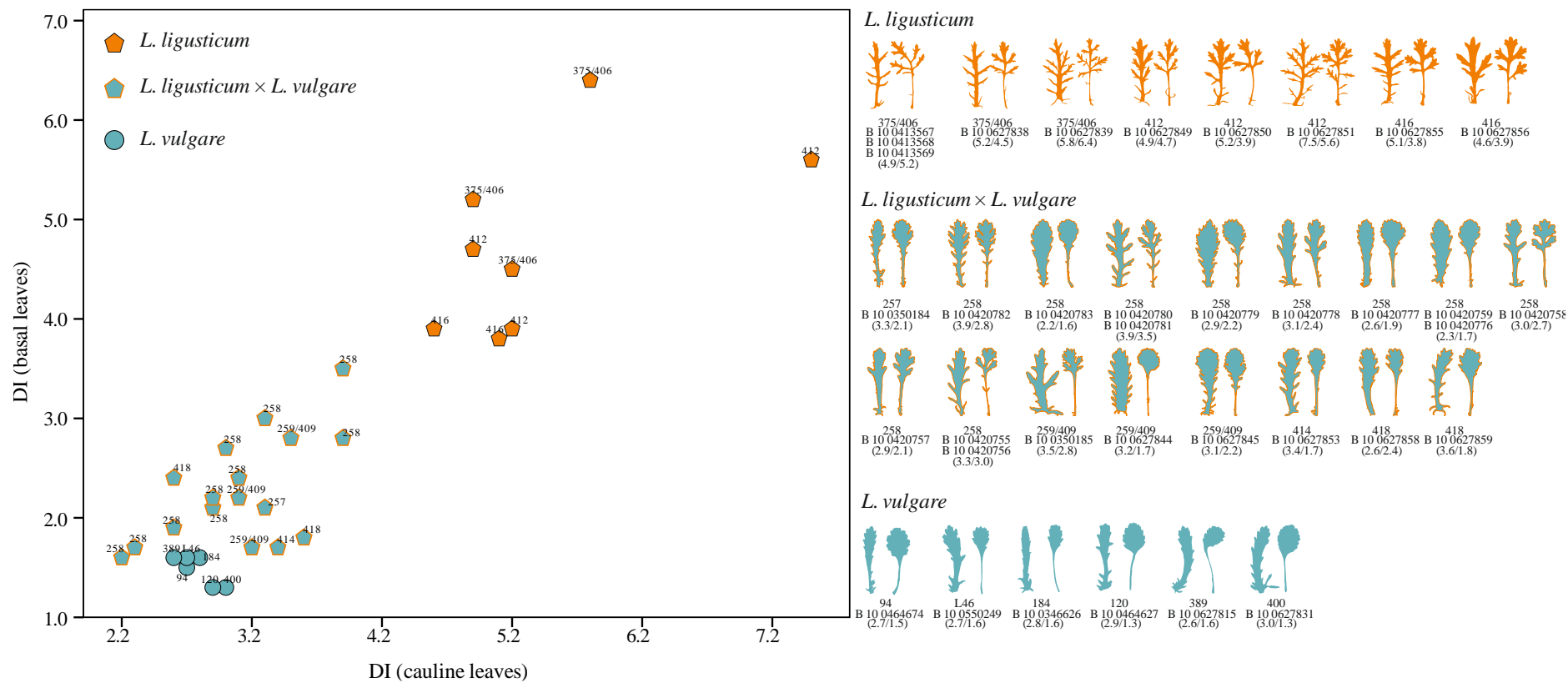


Figure S2.4 Dissection indices (DI) of cauline leaves plotted against DI values of basal leaves for herbarium specimens of *L. ligusticum*, *L. ligusticum* × *L. vulgare*, and *L. vulgare*. Silhouettes of leaves on the right side are labelled by population code, herbarium voucher, and DI values (cauline leaf/basal leaf). Leaves are not drawn to scale.

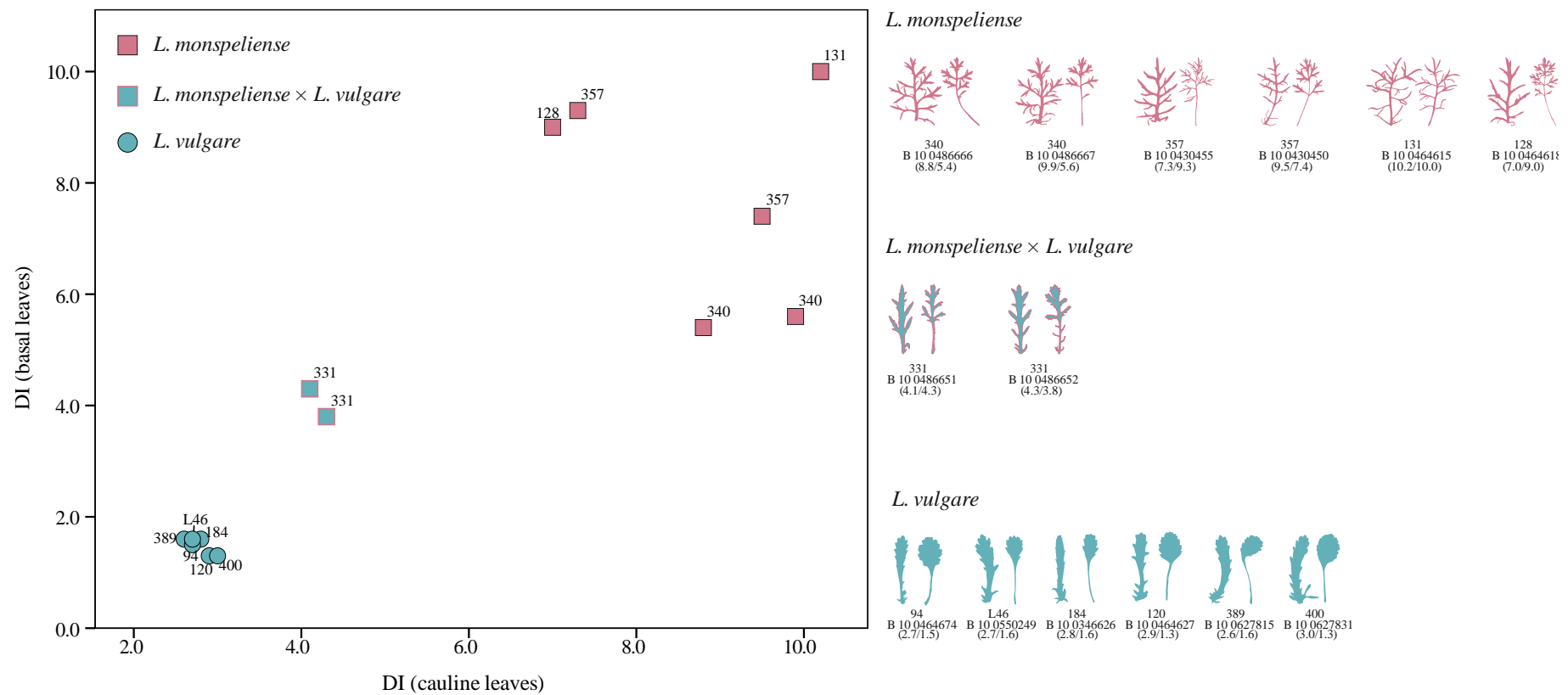
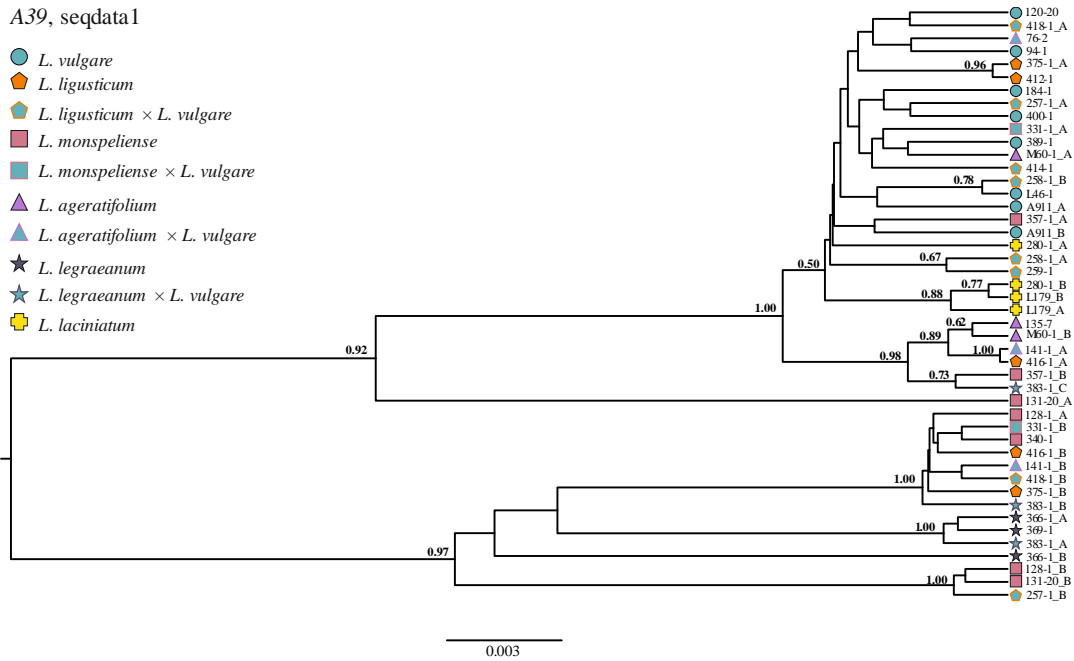


Figure S2.5 Dissection indices (DI) of cauline leaves plotted against DI values of basal leaves for herbarium specimens of *L. monspeliense*, *L. monspeliense* × *L. vulgare*, and *L. vulgare*. Silhouettes of leaves on the right side are labelled by population code, herbarium voucher, and DI values (cauline leaf/basal leaf). Leaves are not drawn to scale.

2.5 Supplemental Figures and Tables

A39, seqdata1

- *L. vulgare*
- *L. ligusticum*
- *L. ligusticum* × *L. vulgare*
- *L. monspeliense*
- *L. monspeliense* × *L. vulgare*
- ▲ *L. ageratifolium*
- ▲ *L. ageratifolium* × *L. vulgare*
- ★ *L. legraeianum*
- ★ *L. legraeianum* × *L. vulgare*
- *L. laciniatum*



A39, seqdata2

- *L. vulgare*
- *L. ligusticum*
- *L. monspeliense*
- ▲ *L. ageratifolium*
- ★ *L. legraeianum*
- *L. laciniatum*

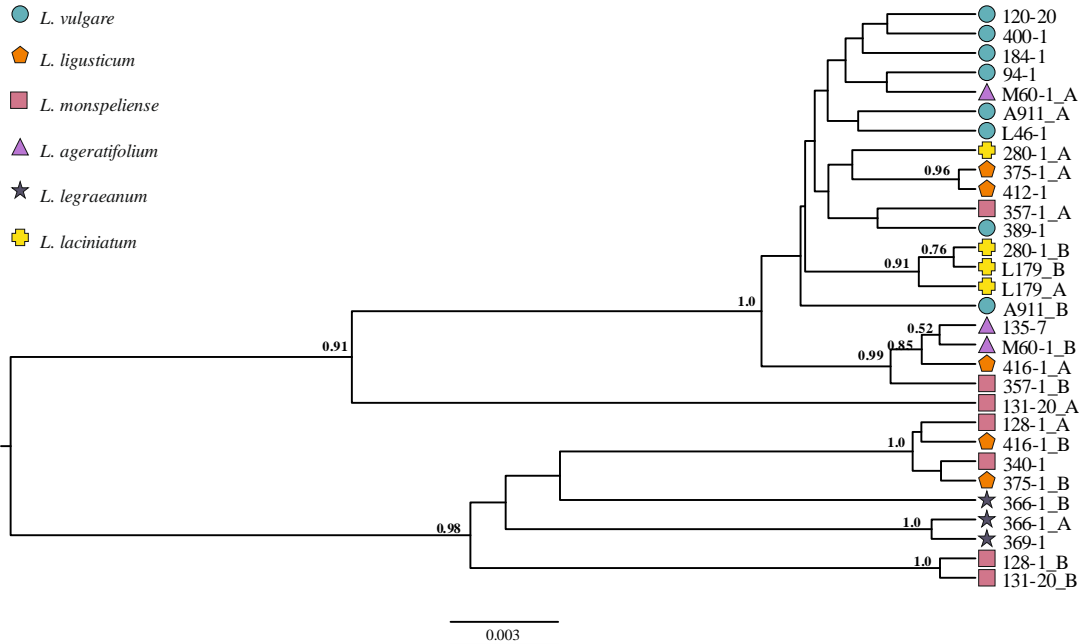
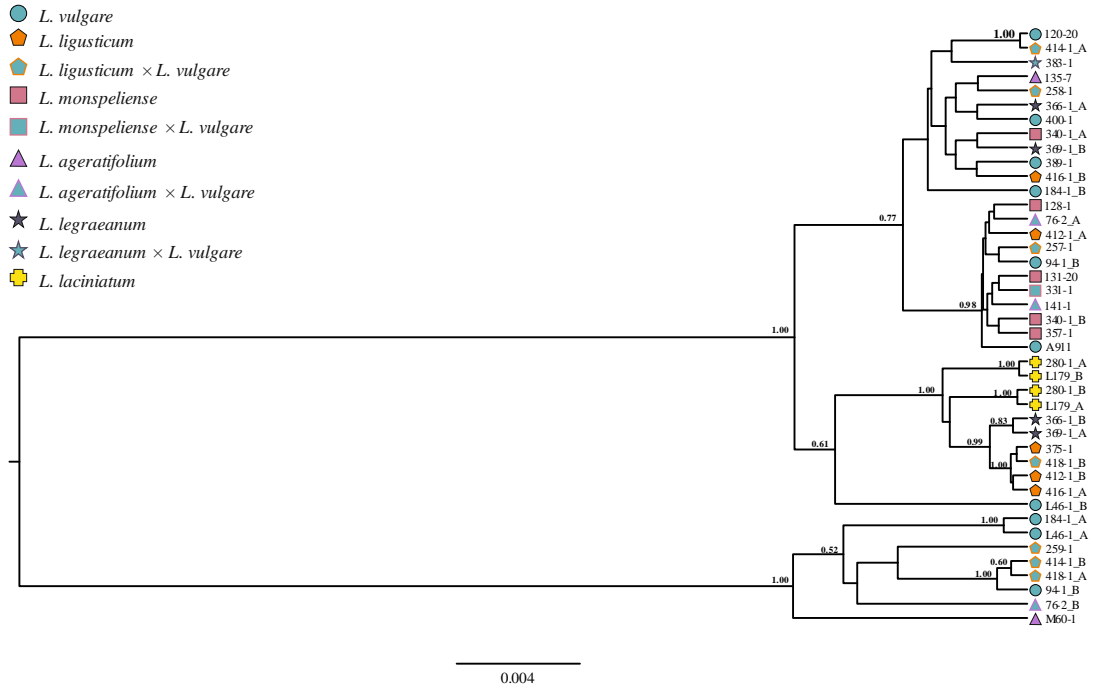


Figure S2.6 Gene trees based on sequence variation of marker A39 (Chapman et al., 2007) calculated in BEAST based on the total dataset (seqdata1) and on a dataset without putative hybrids (seqdata2). In the case of heterozygous individuals, alleles are labelled by alphabetic characters after accession code. Numbers to the left of nodes are posterior probabilities (only values >0.5 are shown).

2.5 Supplemental Figures and Tables

C12, seqdata1



C12, seqdata2

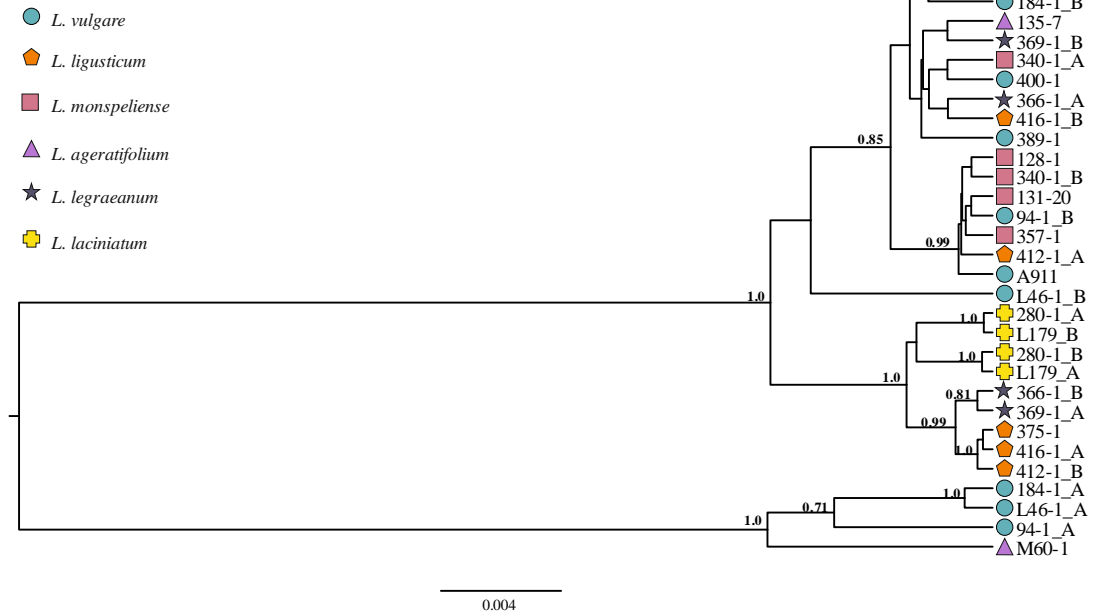
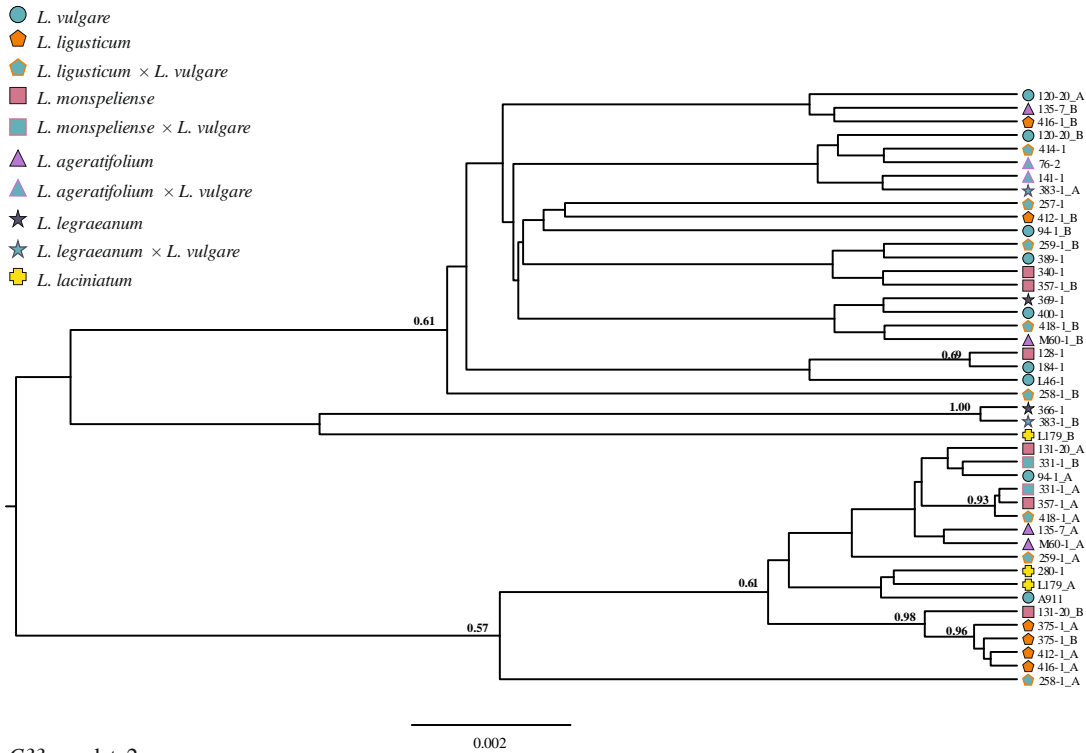


Figure S2.7 Gene trees based on sequence variation of marker *C12* (Chapman et al., 2007) calculated in BEAST based on the total dataset (seqdata1) and on a dataset without putative hybrids (seqdata2). In the case of heterozygous individuals, alleles are labelled by alphabetic characters after accession code. Numbers to the left of nodes are posterior probabilities (only values >0.5 are shown).

2.5 Supplemental Figures and Tables

C33, seqdata1



C33, seqdata2

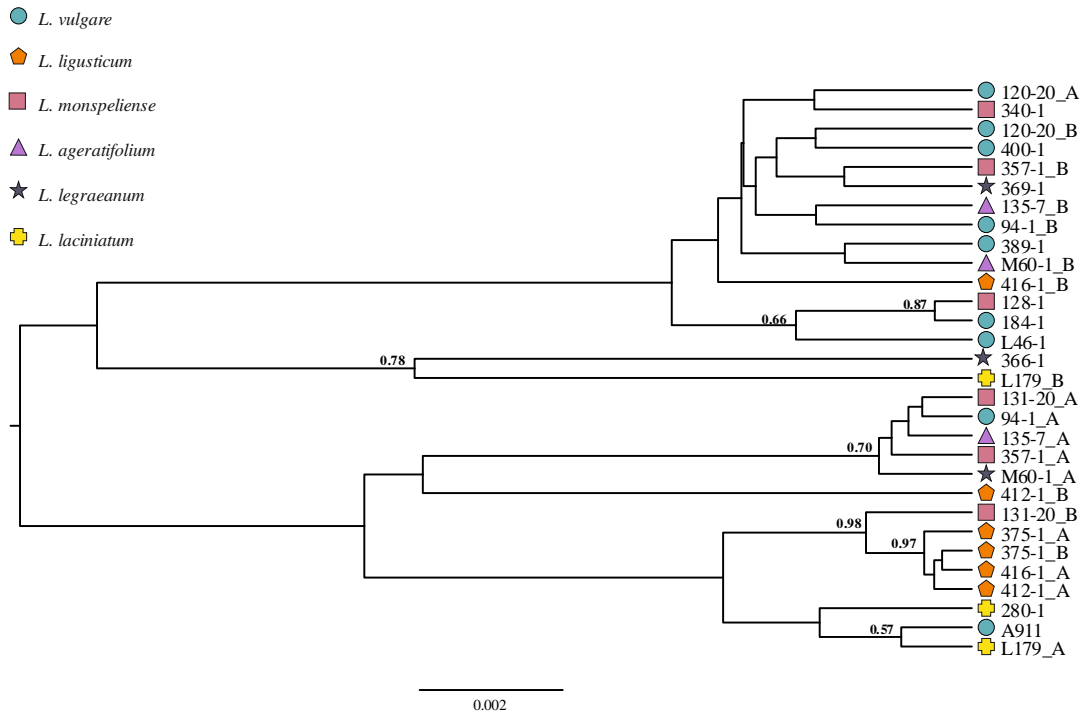


Figure S2.8 Gene trees based on sequence variation of marker *C33* (Chapman et al., 2007) calculated in BEAST based on the total dataset (seqdata1) and on a dataset without putative hybrids (seqdata2). In the case of heterozygous individuals, alleles are labelled by alphabetic characters after accession code. Numbers to the left of nodes are posterior probabilities (only values >0.5 are shown).

2.5 Supplemental Figures and Tables

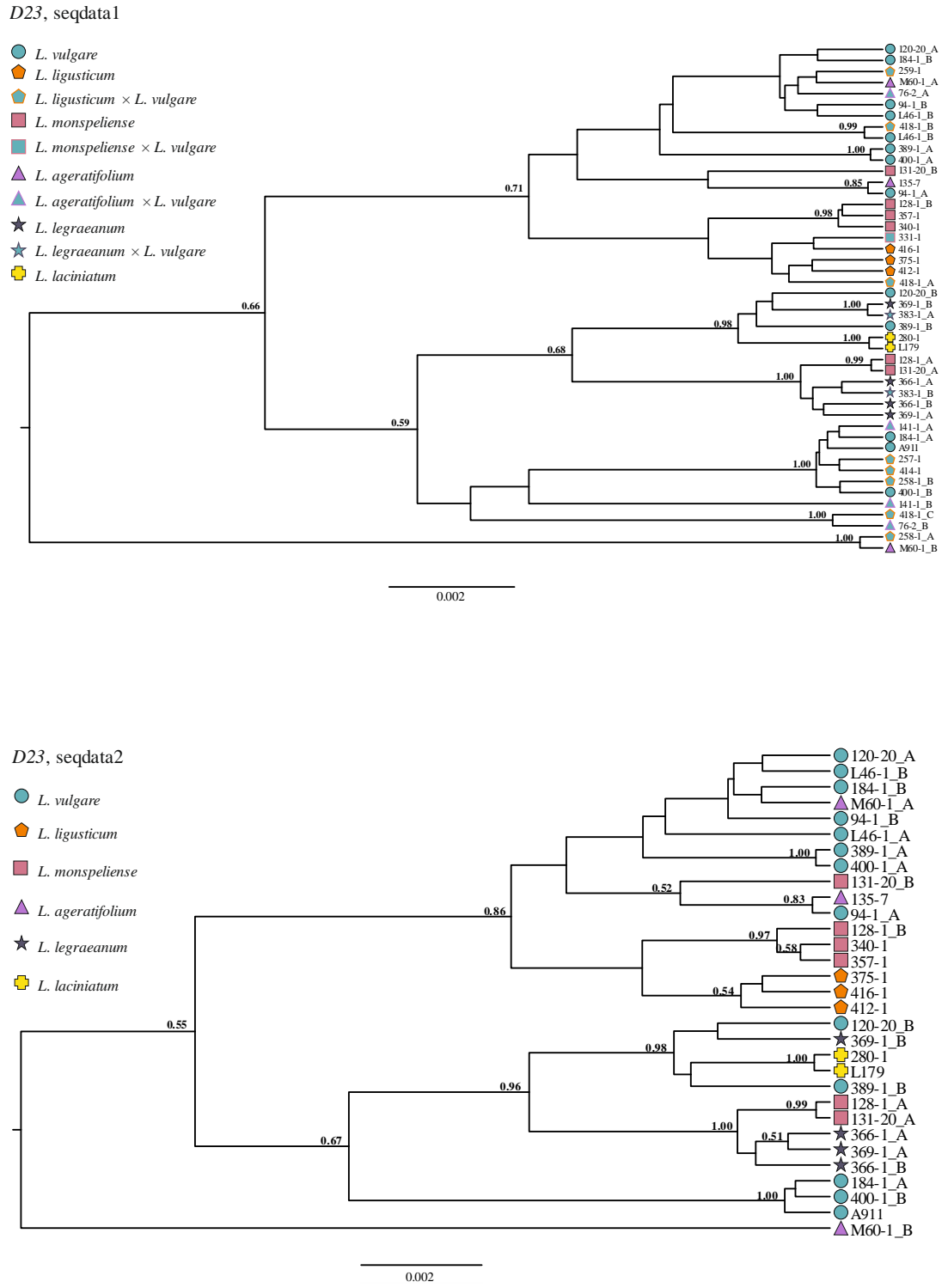
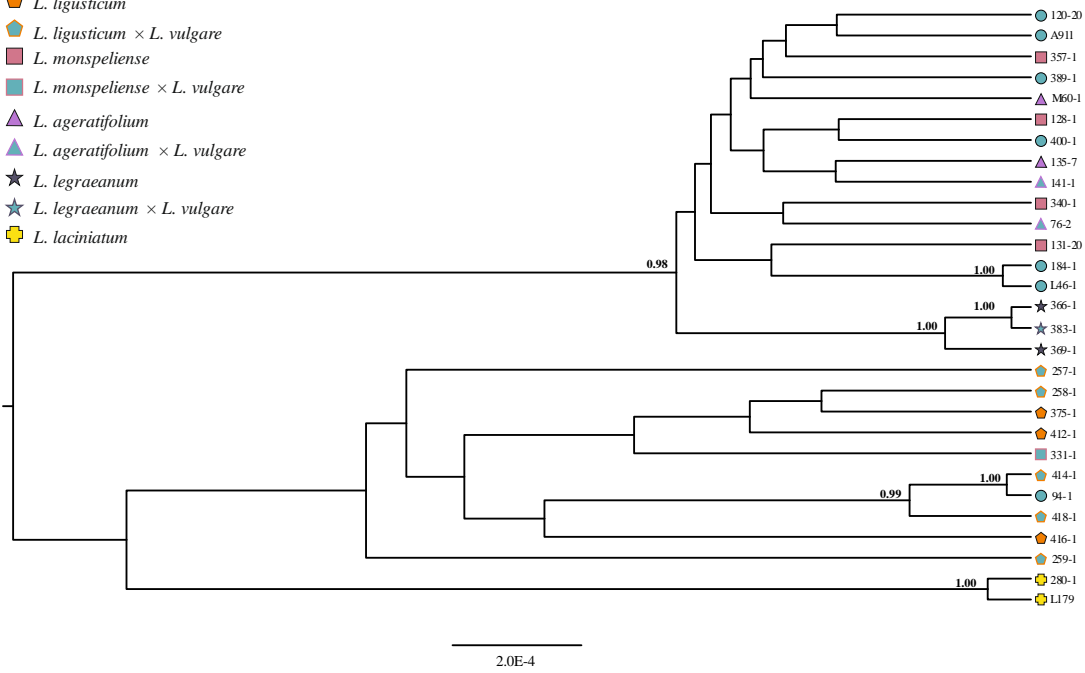


Figure S2.9 Gene trees based on sequence variation of marker *D23* (Chapman et al., 2007) calculated in BEAST based on the total dataset (seqdata1) and on a dataset without putative hybrids (seqdata2). In the case of heterozygous individuals, alleles are labelled by alphabetic characters after accession code. Numbers to the left of nodes are posterior probabilities (only values >0.5 are shown).

2.5 Supplemental Figures and Tables

ptDNA, seqdata1

- *L. vulgare*
- ◆ *L. ligusticum*
- ◆ *L. ligusticum* × *L. vulgare*
- *L. monspeliense*
- *L. monspeliense* × *L. vulgare*
- ▲ *L. ageratifolium*
- ▲ *L. ageratifolium* × *L. vulgare*
- ★ *L. legraeaeum*
- ★ *L. legraeaeum* × *L. vulgare*
- *L. laciniatum*



ptDNA, seqdata2

- *L. vulgare*
- ◆ *L. ligusticum*
- *L. monspeliense*
- ▲ *L. ageratifolium*
- ★ *L. legraeaeum*
- *L. laciniatum*

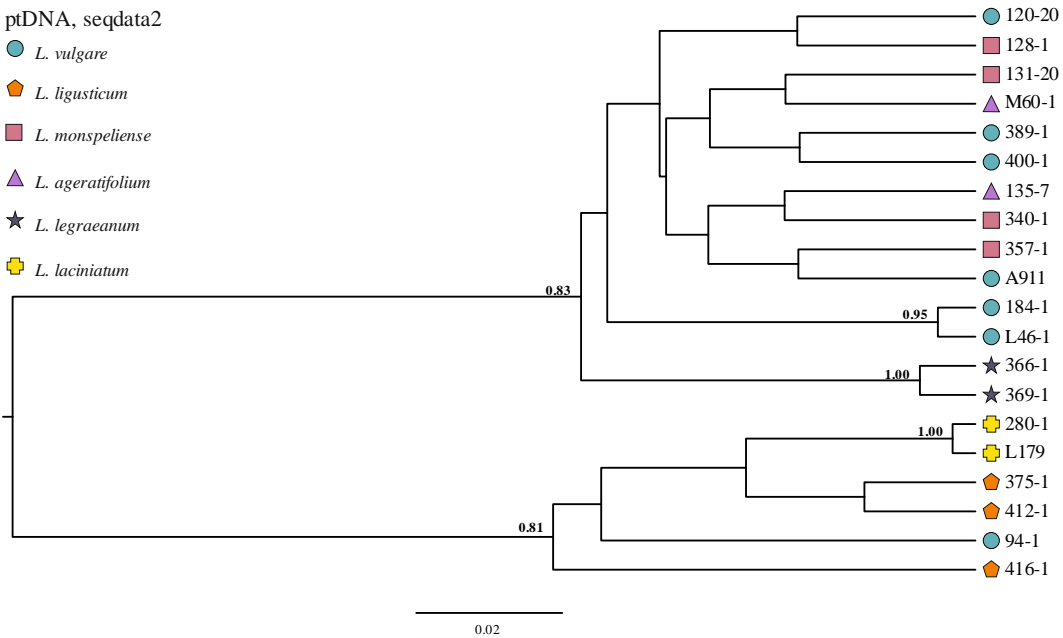


Figure S2.10 Gene trees based on sequence variation of concatenated plastid markers (ptDNA: *trnL-trnF*, *trnC-petN*, *psbA-trnH*, *petN-psbM*, *trnQ-rps16*) calculated in BEAST based on the total dataset (seqdata1) and on a dataset without putative hybrids (seqdata2). Numbers to the left of nodes are posterior probabilities (only values >0.5 are shown).

2.5 Supplemental Figures and Tables

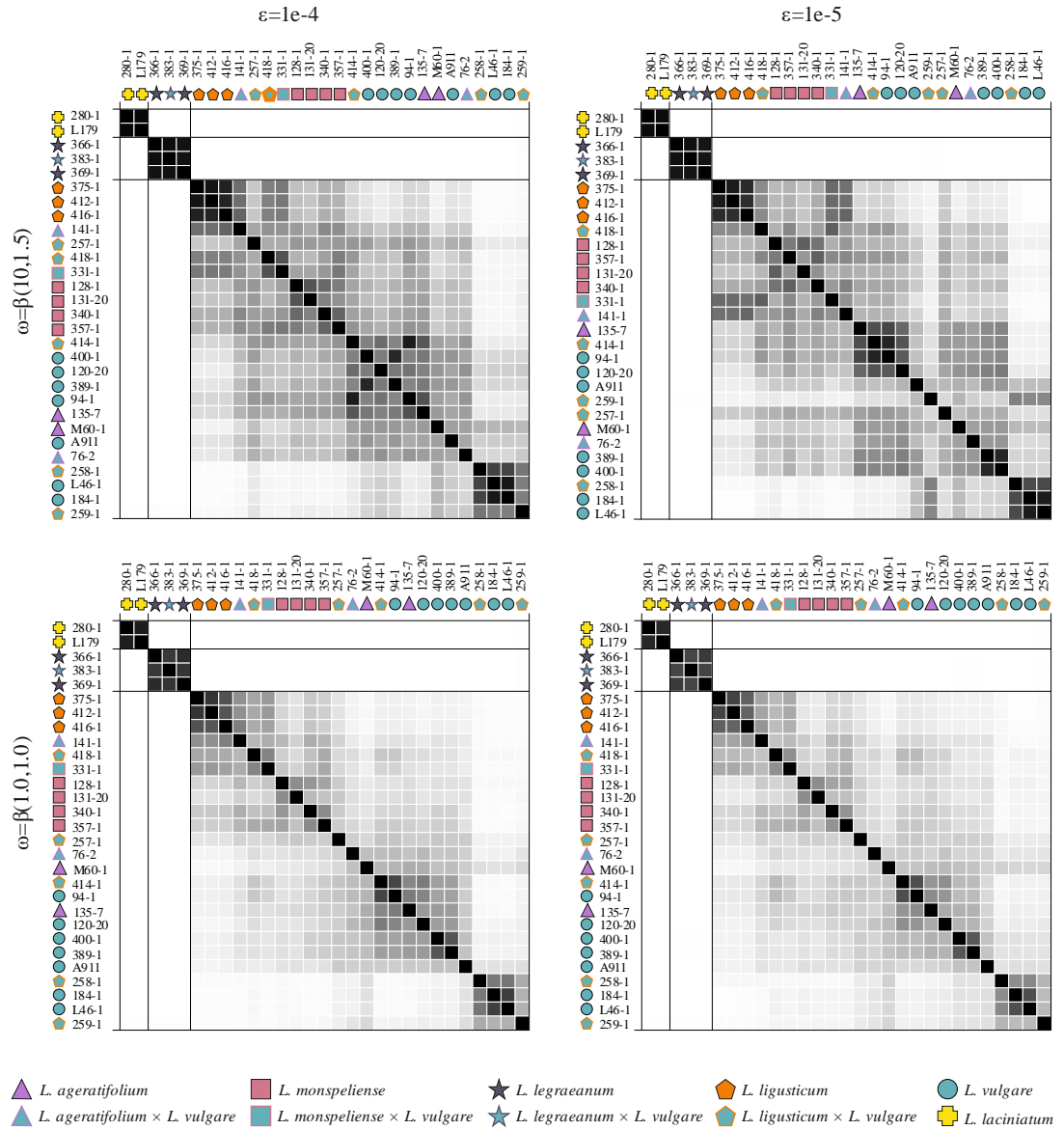


Figure S2.11 Similarity matrices summarizing DISSECT results for the total dataset (seqdata1) under various ‘collapsing height’ (ϵ) and ‘collapsing weight’ (ω) values. Squares represent posterior probabilities (white: $PP = 0$, black: $PP = 1.0$) for pairs of individuals belonging to the same cluster.

2.5 Supplemental Figures and Tables

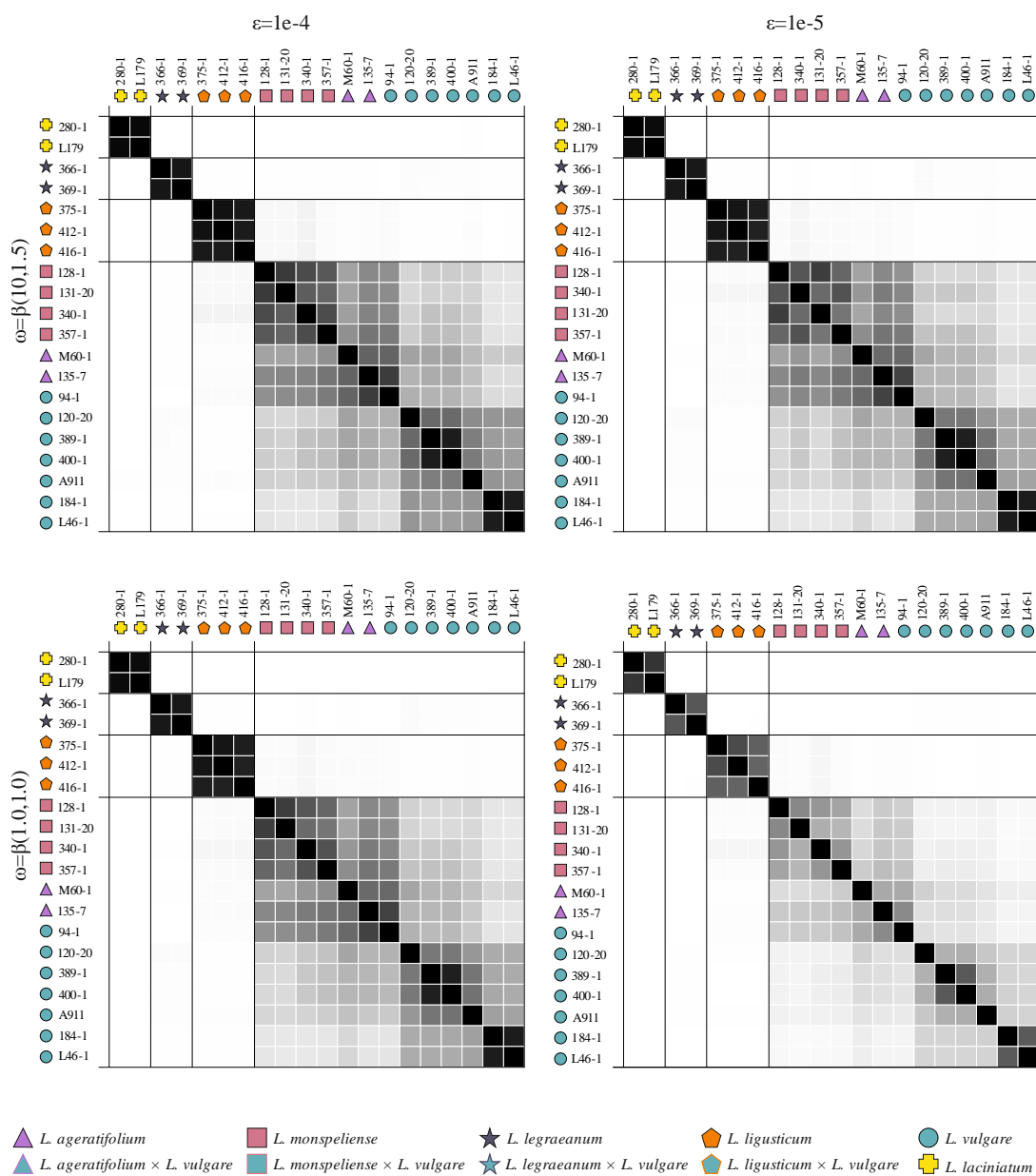


Figure S2.12 Similarity matrices summarizing DISSECT results for the dataset without putative hybrid individuals (seqdata2) under various ‘collapsing height’ (ε) and ‘collapsing weight’ (ω) values. Squares represent posterior probabilities (white: $PP = 0$, black: $PP = 1.0$) for pairs of individuals belonging to the same cluster.

2.5 Supplemental Figures and Tables

Table S2.1 Information about all primers used in the study, including marker and sequence information. For some samples PCR performed better when using tailed primers (M13/TitB) usually designed for 454 pyrosequencing library preparation described in Konowalik et al. (2015).

Primer name	Marker	Sequence	
trnL2(e)	<i>trnL-trnF</i>	GGTTC AAGTCCCTCTATCCC	Taberlet et al. (1991).
trnFr(f)	<i>trnL-trnF</i>	ATTGAACTGGTGACACGAG	
trnC	<i>trnC-petN</i>	CCAGTTCAAATCTGGGTGTC	Demesure et al. (1995)
petN1R	<i>trnC-petN</i>	CCCAAGCAAGACTTACTATATCC	Lee and Wen (2004)
psbA-HF	<i>psbA-trnH</i>	CGAAGCTCCATCTACAAATGG	Hamilton (1999)
trnH-HR	<i>psbA-trnH</i>	ACTGCCTTGATCCACTTGGC	
psbAf	<i>psbA-trnH</i>	GTTATGCATGAACGTAATGCTC	Sang et al. (1997)
trnHr	<i>psbA-trnH</i>	CGCGCATGGTGGATTCACAAATC	
petN1	<i>petN-psbM</i>	GGATATAGTAAGTCTTGCTGGG	Lee and Wen (2004)
psbM2R	<i>petN-psbM</i>	TTCTTGCAATTTATTGCTACTGC	
trnQ2	<i>trnQ-rps16</i>	GCGTGGCCAAGYGGTAAGGC	Shaw et al. (2007)
rps16x1_leu	<i>trnQ-rps16</i>	CAATCGAATTGTCAATGATGC	Konowalik et al. (2015)
A39f	A39	ACTAGTTGGCATYTRATGGTAACA	Chapman et al. (2007)
A39r	A39	GCCRACAAAATTGAGCTGAAGATC	
C12f	C12	TCTTGCACCACCAACTGYTTGGC	
C12r	C12	GACACCGCCTGGCTGC	
M13_C12_f	C12	CACGACGTTGTA AACGACTCTTGCACCACCAACTGYTTGGC	Konowalik et al. (2015)
TitB_C12_Leu350bp_r	C12	CTATGCGCCTTGCCAGCCCCTCAGGGACAATGTTCAATGCTG	
C33f	C33	ATTGGGAAAATYGGTGCKGCTAT	Chapman et al. (2007)
C33r	C33	ATATGHGTCATTGATGCTWGCCAA	
M13_C33_Leu350bp_f	C33	CACGACGTTGTA AACGACTCTACATCCAAAATACTACT	Konowalik et al. (2015)
TitB_C33_Leu350bp_r	C33	CTATGCGCCTTGCCAGCCCCTCAGATTCTGTTGACACATAAAC	
D23f	D23	AGAAGGGTGG AACAGARCATTTRGGGCT	Chapman et al. (2007)
D23r	D23	GGCATTRATYCCRATCTTGCAATTCWCCAGG	
M13_D23_f	D23	CACGACGTTGTA AACGACAGAAGGGTGG AACAGARCATTTRGGGCT	
TitB_D23_r	D23	CTATGCGCCTTGCCAGCCCCTCAGGGCATTRATYCCRATCTTGCAATTCWCCAGG	

2.5 Supplemental Figures and Tables

Table S2.2 Information about single markers of seqdata1 (total dataset) and seqdata2 (without potential hybrids), including aligned length, number and percentage of variable sites, number of coded indels, parsimony-informative sites (indels included), as well as consistency and retention indices calculated in PAUP*. Best fitting models of sequence evolution found in jMODELTEST and clock models according to marginal likelihood comparisons (see Table S2.3) are also itemized.

	Locus	Length (bp)	Variable sites (substitutions)	Indels	Parsimony-informative sites/indels	Consistency index	Retention index	Substitution model	Clock model
seqdata1	A39	320	37 (11.6%)	3	30	0.75	0.93	TIM2+I	strict
	C12	374	40 (10.7%)	3	32	0.94	0.98	HKY	strict
	C33	329	23 (7.0%)	5	14	0.78	0.94	F81	strict
	D23	362	23 (6.4%)	5	22	0.85	0.95	HKY+I	strict
	ptDNA	2107	14 (0.7%)	6	11	1.00	1.00	GTR	strict
seqdata2	A39	320	34 (10.6%)	3	30	0.76	0.90	TIM2+I	strict
	C12	374	35 (9.4%)	3	29	0.95	0.98	TPM1uf	strict
	C33	328	21 (6.4%)	4	11	0.86	0.96	HKY	strict
	D23	362	20 (5.5%)	5	16	0.86	0.93	HKY+I	strict
	ptDNA	2107	9 (0.4%)	5	8	1.00	1.00	TVM	relaxed

Table S2.3 Logarithmic marginal-likelihood values (lnML) for different loci, datasets (seqdata1: total dataset, seqdata2: excluding potential hybrids) and clock models (strict clock vs. relaxed clock) calculated with the path sampling method in BEAST. Best fitting clock models, using a difference of 3 lnML units as a threshold for accepting the more parameter-rich model (Kass and Raftery, 1995), are highlighted in bold.

	Locus	strict clock	relaxed clock
		lnML	lnML
seqdata1	A39	-945.86	-945.88
	C12	-945.95	-946.03
	C33	-805.27	-802.41
	D23	-876.94	-876.93
	ptDNA	-3024.25	-3021.43
seqdata2	A39	-858.44	-858.38
	C12	-863.93	-864.18
	C33	-734.37	-733.95
	D23	-797.61	-797.64
	ptDNA	-2944.42	-2932.20

2.5 Supplemental Figures and Tables

Table S2.4 Information about investigated *Leucanthemum* population comprising geographic location, coordinates, collectors and vouchers of all corresponding accessions deposited in the herbarium of the Botanical Museum Berlin-Dahlem (B).

Taxon	Pop. code	Geographic location	Coord.	Collector	Herbarium vouchers
<i>Leucanthemum laciniatum</i> Huter, Porta & Rigo	L179	IT, Basilicata, Castrovillari, 1900–2100 m	39.91 N, 16.19 E	Vogt 15614	B 10 0420805
	280	IT, Calabria, Colle del Drogone, 1580 m	39.90 N, 16.11 E	Tomasello TS420	B 10 0464203
<i>Leucanthemum legraeanum</i> (Rouy) B.Bock & J.-M.Tison	366/384	FR, Provence-Alpes-Côte d'Azur, Massif des Maures, 410 m	43.20 N, 6.31 E	Vogt 17189 / Vogt 17434, Oberprieler 10915 & Wagner	B 10 0486634, B 10 0486635, B 10 0486636, B 10 0486637, B 10 0486638, B 10 0627809, B 10 0627810
	369	FR, Provence-Alpes-Côte d'Azur, Massif des Maures, 210 m	43.24 N, 6.34 E	Vogt 17192	B 10 0486648, B 10 0486649
<i>L. legraeum</i> × <i>L. vulgare</i>	383	FR, Provence-Alpes-Côte d'Azur, Vallée du Pansard, 77m	43.19 N, 6.21 E	Vogt 17432, Oberprieler 10913 & Wagner	B 10 0627803, B 10 0627804, B 10 0627805, B 10 0627806, B 10 0627807
	375/406	IT, Liguria, Rocchetta di Vara	44.25 N, 9.76 E	Marchetti s.n. / Vogt 17460, Oberprieler 10941 & Wagner	B 10 0413569, B 10 0627838, B 10 0627839
<i>Leucanthemum ligusticum</i> Marchetti, R.Bernardello, Melai & Peruzzi	412	IT, Liguria, Rocce di Valletti, 700 m	44.36 N, 9.51 E	Vogt 17467, Oberprieler 10948 & Wagner	B 10 0627849, B 10 0627850, B 10 0627851
	416	IT, Liguria, Ponte di Lagoscuro, 246 m	44.34 N, 9.46 E	Vogt 17471, Oberprieler 10952 & Wagner	B 10 0627855, B 10 0627856
<i>L. ligusticum</i> × <i>L. vulgare</i>	257	IT, Liguria, Rocchetta di Vara, 228 m	44.25 N, 9.76 E	Vogt 16943 & Oberprieler 10850	B 10 0350184 B 10 0420782, B 10 0420783, B 10 0420780, B 10 0420781, B 10 0420779, B 10 0420778, B 10 0420777, B 10 0420759, B 10 0420776, B 10 0420758, B 10 0420757 B 10 0420756, B 10 0420755
	258	IT, Liguria, Rocchetta di Vara, 228 m	44.25 N, 9.76 E	Vogt 16944 & Oberprieler 10851	
	259/409	IT, Liguria, Varese Ligure, 341 m	44.37 N, 9.59 E	Vogt 16945 & Oberprieler 10852 / Vogt 17464, Oberprieler 10945 & Wagner	B 10 0350185, B 10 0627844, B 10 0627845
	414	IT, Liguria, Piani di Oneto, 829 m	44.36 N, 9.48 E	Vogt 17469, Oberprieler 10950 & Wagner	B 10 0627853
	418	IT, Piemonte, Mondovì, 492 m	44.35 N, 7.89 E	Vogt 17473, Oberprieler 10954 & Wagner	B 10 0627858, B 10 0627859
<i>Leucanthemum monspeliense</i> (L.) H.J.Coste	131	FR, Languedoc-Roussillon, St.-Andréde-Valborgne, 380 m	44.14 N, 03.73 E	Vogt 16716, Oberprieler 10671 & Konowalik	B 10 0464615
	128	FR, Languedoc-Roussillon, l'Espérou, 750 m	44.09 N, 03.58 E	Vogt 16712, Oberprieler 10667 & Konowalik	B 10 0464618
	340	FR, Midi-Pyrénées, La Roque-Bouillac, 184 m	44.58 N, 2.18 E	Vogt 17156, Oberprieler 10881 & Wagner	B 10 0486666, B 10 0486667
	357	FR, Midi-Pyrénées, Saint- Jean-du-Bruel, 571 m	44.03 N, 3.37 E	Vogt 17179, Oberprieler 10904 & Wagner	B 10 0430450, B 10 0430455
<i>L. monspeliense</i> × <i>L. vulgare</i>	331	FR, Rhône-Alpes, Saint- Etienne, 404 m	45.47 N, 4.25 E	Vogt 17147, Oberprieler 10872 & Wagner	B 10 0486652, B 10 0486651
<i>Leucanthemum ageratifolium</i> Pau	135	FR, Pyrénées-Orientales, La Vallée Heureuse, 410 m	42.50 N, 02.96 E	Konowalik KK42 & Ogrodowczyk	B 10 0386712
	M60	ES, Castilla-La Mancha, Salinas de Manzano, 1157 m	40.10 N, 01.52 W	Cordel s.n.	B 10 0345012, B 10 0345013
<i>L. ageratifolium</i> × <i>L. vulgare</i>	141	ES, Catalunya, Montserrat, 645 m	41.61 N, 1.82 E	Konowalik KK48 & Ogrodowczyk	B 10 0386717
	76	ES, Aragon, Narvasa, 1020 m	42.53 N, 0.48 W	Höbl 76 & Himmelreich	B 10 0413730
<i>Leucanthemum vulgare</i> (Vaill.) Lam.	94	FR, Languedoc-Roussillon, Montlaur, 160 m	43.13 N, 02.61 E	Vogt 16641, Oberprieler 10592 & Konowalik	B 10 0464674
	L46	DE, Bayern, Pittmannsdorf, 450 m	49.03 N, 11.88 E	Eder & Oberprieler s.n.	B 10 0550249
	184	BA, Gacko, Ribari, 930 m	43.24 N, 18.34 E	Vogt 16806 & Prem- Vogt	B 10 0346626
	120	FR, Midi-Pyrénées, La Pezade, 756 m	43.89 N, 03.25 E	Vogt 16699, Oberprieler 10654 Konowalik	B 10 0464627
	A911	FR, Bretagne, Point de Brézelle	48.06 N, 4.66 W	Stutz s.n.	B 10 0627815
	389	FR, Provence-Alpes-Côte d'Azur, Draguignan, 774 m	43.67 N, 6.50 E	Vogt 17439, Oberprieler 10920 & Wagner	B 10 0627815
400	FR, Provence-Alpes-Côte d'Azur, Montagne du Cheiron, 918 m	43.79 N, 7.00 E	Vogt 17454, Oberprieler 10935 & Wagner	B 10 0627831	

Chapter 3: Taming the Red Bastards

Taming the Red Bastards: Hybridization and species delimitation in the *Rhodanthemum arundanum*-group (Compositae-Anthemideae)

Florian Wagner, Tankred Ott, Maximilian Schall, Ulrich Lautenschlager, Robert Vogt, Christoph Oberprieler

submitted to Molecular Phylogenetics and Evolution

Abstract

Delineating species boundaries in a group of recently diverged lineages is challenging due to minor morphological differences, low genetic differentiation and the occurrence of gene flow among taxa. Here, we employ traditional Sanger sequencing and restriction-site associated DNA (RAD) sequencing, to investigate species delimitation in the close-knit Moroccan daisy group around *Rhodanthemum arundanum* B.H. Wilcox & al. that diverged recently during the Quaternary. After evaluation of genotyping errors and parameter optimization in the course of de-novo assembly of RADseq reads in IPYRAD, we assess hybridization patterns in the study group based on different data assemblies and methods (Neighbor-net networks, FASTSTRUCTURE and ABBA-BABA tests). RADseq data and Sanger sequences are subsequently used for delimitation of species, using both, multi-species coalescent methods (STACEY and SNAPP) and a novel approach based on consensus k -means clustering. In addition to the unveiling of two novel subspecies in the *R. arundanum*-group, our study provides insights into the performance of different species delimitation methods in the presence of hybridization and varying quantities of data.

Keywords: Consensus k -means clustering, RADseq, hybridization, IPYRAD, parameter optimization, species delimitation

3.1 Introduction

The science and art of species delimitation has been revolutionized by DNA-based approaches during the last decades (Rannala, 2015). While early species delimitation studies based on single-locus genetic sequences suffered from low genetic variability among species of recently evolving groups (e.g. Spooner, 2009), investigation of species-level biological diversity nowadays benefits from facilitated data acquisition via next-generation sequencing (NGS) (Camargo and Sites, 2013). Among different NGS-based techniques, restriction-site associated DNA (RAD) sequencing (Baird et al., 2008) has recently gained much attention in systematic biology, as it enables the discovery and genotyping of thousands of informative markers for many accessions in a short time (Ree and Hipp, 2015). Several recent studies have demonstrated the power of RADseq methods for resolving long-standing taxonomic problems and species boundaries in taxonomically complex groups (Leaché et al., 2014a; Pante et al., 2015, Anderson et al., 2017; Fernández-Mazuecos et al., 2018; Wagner et al., 2018; Spriggs et al., 2019).

Various RADseq protocols have been developed (e.g. ddRAD, ezRAD, GBS, 2bRAD), which differ in the use of one versus two restriction enzymes, varying types of adaptors or the performance and order of shearing, size-selection and amplification steps (reviewed in Andrews et al., 2016). A common feature of all RADseq methods is the generation of a large number of relatively short sequence reads from different loci, which are (i) widely distributed in the genomes under study, (ii) not characterized as paralogous or orthologous at the outset of a project and (iii) partly incomplete due to loci and allele dropout (Ree and Hipp, 2015). These characteristics constitute a major challenge for the processing and analyses of RADseq data in the course of phylogenetic and species delimitation studies, particularly in the absence of a reference genome.

A number of powerful pipelines for de-novo processing of RADseq data have been developed in the last years, such as STACKS (Catchen et al., 2011, 2013) or PYRAD/IPYRAD (Eaton and Ree, 2013; Eaton 2014; Eaton and Overcast, 2016). Several recent studies have shown that the quality of locus identification and orthology estimation in the course of these pipelines is strongly depending on the choice of reasonable core parameters throughout the different assembly steps (Mastretta-Yanes et al., 2015; Anderson et al., 2017; Paris et al., 2017; Shafer et al., 2017; Fernández-Mazuecos et al., 2018; McCartney-Melstad et al., 2019). Therefore, different strategies have been proposed for the optimization of parameter space, including error quantification based on sample replicates (Mastretta-Yanes et al., 2015; Anderson et al., 2017) or evaluation of core assembly metrics (Paris et al., 2017, McCartney-Melstad et al., 2019).

Once RADseq reads have been successfully assembled, hundreds to thousands of unlinked, putatively orthologous loci shared by many individuals become available for species delimitation analyses. Methods in the framework of the multi-species coalescent (MSC) model (Rannala and Yang, 2003), such as the BEAST2 application SNAPP (Bryant et al., 2012; Leaché et al., 2014a), have proven their ability to handle plenty of genome-wide SNPs with considerable power in identifying boundaries among recently diverged species (Leaché et al., 2014a). However, hybridization can affect the accuracy of MSC species delimitation results, as the underlying model assumes no hybridization after species divergence (Zhang et al., 2011). Apart from that, there is an ongoing debate about the ability of MSC methods to distinguish between genetic structure which is due to population-level processes on the one hand, or due to species boundaries on the other (Sukumaran and Knowles, 2017; Leaché et al., 2018).

In the present contribution, we use RAD sequencing to delimit species in the close-knit *Rhodanthemum arundanum*-group. After optimization of IPYRAD assembly parameters and evaluation of genotyping errors, species delimitation analyses are conducted using the MSC model and a novel approach based on consensus *k*-means clustering. The results of RADseq analyses are compared to those of a traditional Sanger sequencing survey of the same study group. Furthermore, we assess hybridization patterns and evaluate the influence of gene flow as well as different quantities of data on the accuracy of different species delimitation methods.

The genus *Rhodanthemum* B.H. Wilcox & al. ('Moroccan daisies'; Compositae, Anthemideae) comprises 21 taxa of flowering plants, distributed in Southern Spain, Morocco and Algeria (Euro+Med, 2019). The diversification of the genus has been dated back to the Quaternary, with a similar crown age (approximately 1.3 million years) as the closely related European ox-eye daisies (genus *Leucanthemum* Mill.; Wagner et al., 2019). In contrast to *Leucanthemum*, which has built up a comprehensive polyploid complex (Vogt, 1991), *Rhodanthemum* taxa have strictly evolved on the diploid level (Wilcox and Harcourt, 1982; Vogt and Oberprieler, 2008, 2012). Wagner et al. (2019) presented a phylogeny of the whole genus based on nine nuclear plus five plastid markers and 52 accessions assigned to 15 lineages. However, due to the young age, low morphological variability and the lack of a monograph of the genus (Vogt, in prep.), taxon boundaries are still partly uncertain and several new names and alternative taxonomic treatments have been proposed in recent time (Vogt, 1994; Gómiz 2000, 2001, 2014; Dobignard 2015).

Here, we focus on a group of *Rhodanthemum* taxa designated as the *R. arundanum*-group in Wagner et al. (2019). This group comprises the eponymous species *R. arundanum* (Boiss.) B.H. Wilcox & al., as well as *R. redieri* (Maire) B.H. Wilcox & al. and the recently described

taxon *R. quezelii* Dobignard & Duret [considered as *R. redieri* subsp. *soriae* Gómiz (Gómiz, 2014)]. Besides alternative taxonomic treatments of the latter taxon, uncertainty persists in the separation of *R. redieri* into two subspecies [*R. redieri* (Maire) B.H.Wilcox & al. subsp. *redieri* and *R. redieri* subsp. *humbertii* Gómiz, see Gómiz (2000) and Dobignard (2015)] and the taxonomic status of an enigmatic population from the High Atlas mountains [*R. spec.* in Wagner et al. (2019)].

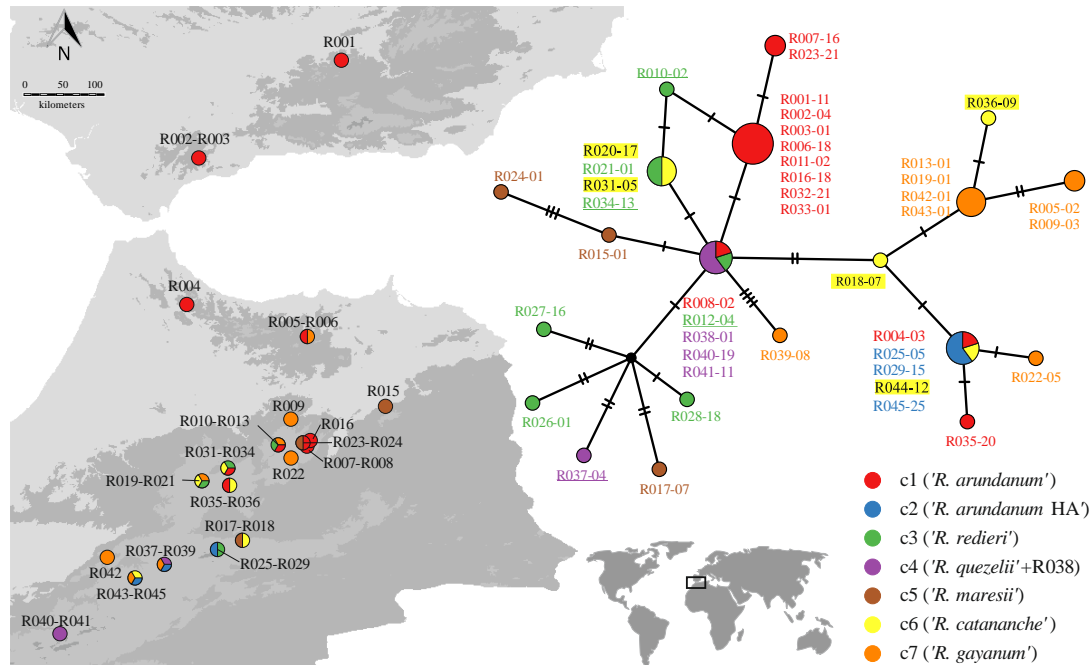


Figure 3.1 Map showing the locations of all examined *Rhodanthemum* populations (left). The TCS network on the right was inferred from intergenic spacer regions *trnC-petN* and *trnQ-rps16* of the plastid genome of one accession per population. Colors indicate the assignment of populations to clusters c1-c7 according to consensus *k*-means clustering of dataset *ct85ms112* (see Figure 3.5).

3.2 Materials and Methods

3.2.1 Taxon sampling and DNA extraction

Leaves of 102 accessions from 43 *Rhodanthemum* populations (Table 3.1, Figure 3.1) were collected and silica-dried during field trips to Spain (2016) and Morocco (2017). Our final sampling comprised all taxa of the *R. arundanum*-group according to Wagner et al. (2019), including *R. arundanum*, *R. redieri* subsp. *redieri*, *R. redieri* subsp. *humbertii*, *R. quezelii* and an enigmatic population (R038) from Djebel Bou Ijallabene of unknown taxonomic status. As outgroup, we included 12 accessions of the closely related '*R. maresii*' lineage, consisting of *R. maresii* (Coss.) B.H.Wilcox & al. and *R. mesatlanticum* (Emb. & Maire) B.H.Wilcox & al., and 17 accessions of two distantly related and codistributed species with regard to the *R. arundanum*-group, namely *R. gayanum* (Coss. & Durieu) B.H.Wilcox & al. and *R. catananche* (Ball) B.H.Wilcox & al. Total genomic DNA was extracted using the CTAB protocol (Doyle and Dickson, 1987; Doyle and Doyle, 1987).

Table 3.1 *Rhodanthemum* accessions used for RAD and nrDNA ITS/ETS sequencing including information about population, locality, collectors and corresponding herbarium specimen. Individual sample replicates in RAD procedure are bolded. Asterisks (*) refer to sequences from Wagner et al. (2019).

Taxon	Pop. code	ITS/ETS samples	GenBank (ITS; ETS)	RAD samples	GenBank (RADseq)	Locality	Coordinates	Collectors	Voucher
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R001	R001-11	MN182334; MN182395	R001-08, R001-11, R001-20	SAMN12288004, SAMN12288005, SAMN12288006	Spain, Jaén, Siera de Mágina, Cerro Carceles, 1965 m	37°44'28.3"N 03°28'57.7"W	Vogt 17524, Oberprieler 10961 & Wagner	B 10 0673421
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R002	R002-04	MN182335; MN182396	R002-01 , R002-04	SAMN12288007, SAMN12288008, SAMN12288009	Spain, Málaga, Sierra de las Nieves, Pilar de Tolox, 1748 m	36°41'28.5"N 05°00'19.5"W	Vogt 17528, Oberprieler 10963 & Wagner	B 10 0673422
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R003	R003-01	MN182336; MN182397	R003-01, R003-11	SAMN12288010, SAMN12288011	Spain, Málaga, Sierra de las Nieves, Pilar de Tolox, 1701 m	36°41'28.5"N 05°00'19.5"W	Vogt 17529, Oberprieler 10964 & Wagner	B 10 0673423
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R004	R004-03	MN182337; MN182398	R004-01, R004-02, R004-03	SAMN12288012, SAMN12288013, SAMN12288014	Morocco, Rif mountains, Chefchaouen, between Djebel Lakraa and Djebel Taloussisse, 1900 m	35°08'05.0"N 05°07'50.0"W	Vogt 17611, Oberprieler 10968 & Wagner	B 10 0754402
<i>R. gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R005	R005-02	MN182338; MN182399	R005-02	SAMN12288015	Morocco, Rif mountains, Taza, Djebel Azrou Akchar, 1535 m	34°47'41.5"N 03°48'57.8"W	Vogt 17625, Oberprieler 10982 & Wagner	B 10 0703548
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R006	R006-18	MN182339, MN182340, MN182400, MN182401	R006-01, R006-02, R006-18	SAMN12288016, SAMN12288017, SAMN12288018	Morocco, Rif mountains, Taza, Djebel Azrou Akchar, 1850 m	34°47'12.8"N 03°50'30.2"W	Vogt 17630, Oberprieler 10987 & Wagner	B 10 0703543
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R007	R007-16	MN182341; MN182402	R007-01, R007-16	SAMN12288019, SAMN12288020	Morocco, Middle Atlas, Boulmane, between Adrar Gueb er Rehal and Adrar bou Naseur, 1996 m	33°36'01.5"N 03°48'32.3"W	Vogt 17636, Oberprieler 10993 & Wagner	B 10 0703537
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R008	R008-02	MN182342; MN182403, MN182404	R008-01, R008-02	SAMN12288021, SAMN12288022	Morocco, Middle Atlas, Guercif, between Adrar Gueb er Rehal and Adrar bou Naseur, 2145 m	33°36'03.7"N 03°49'11.3"W	Vogt 17645, Oberprieler 11002 & Wagner	B 10 0704737
<i>R. gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R009	R009-03	MN182343; MN182405	R009-03	SAMN12288023	Morocco, Middle Atlas, Taza, between Maghraoua and Tmourhout, 899 m	33°54'29.7"N 04°01'50.9"W	Vogt 17647, Oberprieler 11004 & Wagner	B 10 0704734
<i>R. xpsudoredieri</i> Florian Wagner, Vogt & Oberpr.	R010	R010-02	MN182344, MN182345, MN182406, MN182407	R010-01, R010-02	SAMN12288024, SAMN12288025	Morocco, Middle Atlas, Guercif, Djebel Bou Iblane, 2275 m	33°38'44.5"N 04°09'17.8"W	Vogt 17651, Oberprieler 11008 & Wagner	B 10 0704730
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R011	R011-02	MN182346, MN182347, MN182408, MN182409	R011-01, R011-02	SAMN12288026, SAMN12288027	Morocco, Middle Atlas, Guercif, Djebel Bou Iblane, 2391 m	33°38'49.9"N 04°08'59.1"W	Vogt 17654a, Oberprieler 11011a & Wagner	B 10 0704729
<i>R. xpsudoredieri</i> Florian Wagner, Vogt & Oberpr.	R011	–	–	R011-16, R011-21	SAMN12288028, SAMN12288029	Morocco, Middle Atlas, Guercif, Djebel Bou Iblane, 2391 m	33°38'49.9"N 04°08'59.1"W	Vogt 17654b, Oberprieler 11011b & Wagner	B 10 1067612
<i>R. xpsudoredieri</i> Florian Wagner, Vogt & Oberpr.	R012	R012-04	MN182348; MN182410, MN182411	R012-01, R012-04	SAMN12288030, SAMN12288031	Morocco, Middle Atlas, Guercif, Djebel Bou Iblane, 2428 m	33°38'55.3"N 04°08'45.2"W	Vogt 17657, Oberprieler 11014 & Wagner	B 10 0704724
<i>R. gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R013	R013-01	MN182349, MN182350, MN182412, MN182413	R013-01	SAMN12288032	Morocco, Middle Atlas, Guercif, Djebel Bou Iblane, 2177 m	33°38'21.3"N 04°09'45.4"W	Vogt 17660, Oberprieler 11017 & Wagner	B 10 0704723
<i>R. mesatlanticum</i> (Emb. & Maire) B.H. Wilcox & al.	R015	R015-01	MK481574*, MK481575*, MN182414, MN182415	R015-01, R015-10, R015-16, R015-18	SAMN12288033, SAMN12288034, SAMN12288035, SAMN12288036	Morocco, Middle Atlas, Taourirt, Djebel Flouch, 1186 m	34°02'43.4"N 03°00'56.0"W	Vogt 17666, Oberprieler 11023 & Wagner	B 10 0704717
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R016	R016-18	MN182351, MN182352, MN182416, MN182417	R016-12, R016-18	SAMN12288037, SAMN12288038	Morocco, Middle Atlas, Guercif, Adrar Gelb er Rehal, 2900 m	33°37'04.2"N 03°49'40.6"W	Vogt 17673, Oberprieler 11030 & Wagner	B 10 0704793, B 10 0704792

Table 3.1 Continued

Taxon	Pop. code	ITS/ETS samples	GenBank (ITS; ETS)	RAD samples	GenBank (RADseq)	Locality	Coordinates	Collectors	Voucher
<i>R. maresii</i> (Coss.) B. H. Wilcox & al.	R017	R017-07	MN182353, MN182354, MN182418, MN182419	R017-04 , R017-07, R017-15, R017-21	SAMN12288039, SAMN12288040, SAMN12288041, SAMN12288042, SAMN12288043	Morocco, High Atlas, Midelt, Tizi-n-Talrhemt, 1700 m	32°37'25.9"N 04°32'18.9"W	Vogt 17683, Oberprieler 11040 & Wagner	B 10 0704783
<i>R. catananche</i> (Ball) B. H. Wilcox & al.	R018	R018-07	MN182355; MN182420, MN182421	R018-06, R018-07	SAMN12288044, SAMN12288045	Morocco, High Atlas, Midelt, Tizi-n-Talrhemt, 1900 m	32°35'33.4"N 04°32'04.1"W	Vogt 17687b, Oberprieler 11044b & Wagner	B 10 1067613
<i>R. gyanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R019	R019-01	MN182356; MN182422, MN182423	R019-01	SAMN12288046, SAMN12288047	Morocco, Middle Atlas, Ifrane, Djebel Ari Benij, 2034 m	33°15'04.4"N 04°57'18.2"W	Vogt 17688, Oberprieler 11045 & Wagner	B 10 0704777
<i>R. catananche</i> (Ball) B. H. Wilcox & al.	R020	R020-17	MN182357, MN182358; MN182424	R020-01 , R020-17	SAMN12288048, SAMN12288049, SAMN12288050	Morocco, Middle Atlas, Ifrane, Djebel Ari Benij, 2369 m	33°15'19.6"N 04°57'58.6"W	Vogt 17696, Oberprieler 11053 & Wagner	B 10 0704771
<i>R. redieri</i> (Maire) B. H. Wilcox & al. subsp. <i>redieri</i>	R021	R021-01	MK481576* MK481577*; MN182425	R021-01 , R021-02, R021-03, R021-26	SAMN12288051, SAMN12288052, SAMN12288053, SAMN12288054, SAMN12288055	Morocco, Middle Atlas, Ifrane, Djebel Ari Benij, 2369 m	33°15'19.6"N 04°57'58.6"W	Vogt 17699, Oberprieler 11056 & Wagner	B 10 0704774
<i>R. gyanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R022	R022-05	MN182359, MN182360; MN182426	R022-05	SAMN12288056	Morocco, Middle Atlas, Boulemane, NW of Ouled Ali Youssef, 1818 m	33°29'17.2"N 04°01'17.7"W	Vogt 17701, Oberprieler 11058 & Wagner	B 10 0760064
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R023	R023-21	MN182361; MN182427, MN182428	R023-12, R023-21	SAMN12288057, SAMN12288058	Morocco, Middle Atlas, Guercif, Tizi-n-Saft, 1880 m	33°36'42.6"N 03°52'03.2"W	Vogt 17702, Oberprieler 11059 & Wagner	B 10 0760063
<i>R. mesatlanticum</i> (Emb. & Maire) B. H. Wilcox & al.	R024	R024-01	MK481578*; MN182429, MN182430	R024-01 , R024-05, R024-11, R024-14	SAMN12288059, SAMN12288060, SAMN12288061, SAMN12288062, SAMN12288063	Morocco, Middle Atlas, Guercif, Tizi-n-Saft, 1880 m	33°36'42.6"N 03°52'03.2"W	Vogt 17703, Oberprieler 11060 & Wagner	B 10 0760062
<i>R. arundanum</i> subsp. <i>mairei</i> (Humbert) Florian Wagner, Vogt & Oberpr.	R025	R025-05	MN182362, MN182363; MN182431	R025-01, R025-05, R025-17	SAMN12288064, SAMN12288065, SAMN12288066	Morocco, High Atlas, Midelt, Ari Ajachi, 2816 m	32°31'46.8"N 04°48'27.4"W	Vogt 17704, Oberprieler 11061 & Wagner	B 10 0760061
<i>R. redieri</i> subsp. <i>humbertii</i> Gómiz	R026	R026-01	MN182364, MN182365; MN182432, MN182433	R026-01, R026-07	SAMN12288067, SAMN12288068	Morocco, High Atlas, Midelt, Ari Ajachi, 2816 m	32°31'46.8"N 04°48'27.4"W	Vogt 17707, Oberprieler 11064 & Wagner	B 10 0760058
<i>R. redieri</i> subsp. <i>humbertii</i> Gómiz	R027	R027-16	MN182366; MN182434, MN182435	R027-04, R027-09, R027-16	SAMN12288069, SAMN12288070, SAMN12288071	Morocco, High Atlas, Midelt, Ari Ajachi, 3000 m	32°31'26.3"N 04°48'23.5"W	Vogt 17713, Oberprieler 11070 & Wagner	B 10 0760052
<i>R. redieri</i> subsp. <i>humbertii</i> Gómiz	R028	R028-18	MN182367, MN182368; MN182436, MN182437	R028-05, R028-18	SAMN12288072, SAMN12288073	Morocco, High Atlas, Midelt, Ari Ajachi, 2877 m	32°31'15.2"N 04°48'42.4"W	Vogt 17714, Oberprieler 11071 & Wagner	B 10 0760051
<i>R. arundanum</i> subsp. <i>mairei</i> (Humbert) Florian Wagner, Vogt & Oberpr.	R029	R029-15	MN182369, MN182370; MN182438	R029-03, R029-05, R029-15	SAMN12288074, SAMN12288075, SAMN12288076	Morocco, High Atlas, Midelt, Ari Ajachi, 2877 m	32°31'15.2"N 04°48'42.4"W	Vogt 17715, Oberprieler 11072 & Wagner	B 10 0760050
<i>R. catananche</i> (Ball) B. H. Wilcox & al.	R031	R031-05	MN182371; MN182439	R031-05	SAMN12288077	Morocco, Middle Atlas, Boulemane, Djebel Tichoukt, 1992 m	33°23'17.4"N 04°41'41.2"W	Vogt 17718b, Oberprieler 11075b & Wagner	B 10 1067614
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R032	R032-21	MN182372, MN182373; MN182440, MN182441	R032-02, R032-05, R032-21	SAMN12288078, SAMN12288079, SAMN12288080	Morocco, Middle Atlas, Boulemane, Djebel Tichoukt, 1992 m	33°23'17.4"N 04°41'41.2"W	Vogt 17719, Oberprieler 11076 & Wagner	B 10 0760046

Table 3.1 Continued

Taxon	Pop. code	ITS/ETS samples	GenBank (ITS; ETS)	RAD samples	GenBank (RADseq)	Locality	Coordinates	Collectors	Voucher
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al	R033	R033-01	MN182374, MN182375; MN182442, MN182443	R033-01, R033-02	SAMN12288081, SAMN12288082	Morocco, Middle Atlas, Boulemane, Djebel Tichoukt, 1992 m	33°23'17.4"N 04°41'41.2"W	Vogt 17720b, Oberprieler 11077b & Wagner	B 10 1067611
<i>R. ×pseudoredieri</i> Florian Wagner, Vogt & Oberpr.	R033	–	–	R033-03	SAMN12288083	Morocco, Middle Atlas, Boulemane, Djebel Tichoukt, 1992 m	33°23'17.4"N 04°41'41.2"W	Vogt 17720a, Oberprieler 11077a & Wagner	B 10 0760045
<i>R. catananche</i> (Ball) B. H. Wilcox & al.	R033	–	–	R033-05	SAMN12288084	Morocco, Middle Atlas, Boulemane, Djebel Tichoukt, 1992 m	33°23'17.4"N 04°41'41.2"W	Vogt 17720c, Oberprieler 11077c & Wagner	B 10 1067610
<i>R. ×pseudoredieri</i> Florian Wagner, Vogt & Oberpr.	R034	R034-13	MN182376, MN182377; MN182444, MN182445	R034-02, R034-13, R034-15	SAMN12288085, SAMN12288086, SAMN12288087	Morocco, Middle Atlas, Boulemane, Djebel Tichoukt, 1992 m	33°23'17.4"N 04°41'41.2"W	Vogt 17724, Oberprieler 11081 & Wagner	B 10 0760025
<i>R. arundanum</i> (Boiss.) B. H. Wilcox & al.	R035	R035-20	MN182378, MN182379; MN182446, MN182447	R035-07, R035-08, R035-20	SAMN12288088, SAMN12288089, SAMN12288090	Morocco, Middle Atlas, Boulemane, Djebel Tamokrant, 2022 m	33°12'20"N 04°41'03.5"W	Vogt 17729, Oberprieler 11086 & Wagner	B 10 0760039
<i>R. catananche</i> (Ball) B. H. Wilcox & al.	R036	R036-09	MN182380, MN182381; MN182448	R036-05, R036-09	SAMN12288091, SAMN12288092	Morocco, Middle Atlas, Boulemane, Djebel Tamokrant, 2022 m	33°12'20"N 04°41'03.5"W	Vogt 17730, Oberprieler 11087 & Wagner	B 10 0760038
<i>R. arundanum</i> subsp. <i>mairei</i> × <i>R. quezelii</i> subsp. <i>ijallabenense</i>	R037	R037-04	MN182382; MN182449	R037-01, R037-04, R037-10	SAMN12288093, SAMN12288094, SAMN12288095	Morocco, High Atlas, Midelt, Djebel Bou Ijallabene, 1794 m	32°21'41.8"N 05°22'23.9"W	Vogt 17738, Oberprieler 11095 & Wagner	B 10 0760030
<i>R. quezelii</i> subsp. <i>ijallabenense</i> Florian Wagner, Vogt & Oberpr.	R038	R038-01	MK481579*, MK481580*; MN182450	R038-01, R038-03, R038-06, R038-10	SAMN12288096, SAMN12288097, SAMN12288098, SAMN12288099	Morocco, High Atlas, Midelt, Djebel Bou Ijallabene, 1794 m	32°21'41.8"N 05°22'23.9"W	Vogt 17739, Oberprieler 11096 & Wagner	B 10 0760029
<i>R. gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R039	R039-08	MN182383, MN182384; MN182451, MN182452	R039-08	SAMN12288100	Morocco, High Atlas, Midelt, Djebel Bou Ijallabene, 1794 m	32°21'41.8"N 05°22'23.9"W	Vogt 17740, Oberprieler 11097 & Wagner	B 10 0760028
<i>R. quezelii</i> Dobignard & Duret	R040/R041	R040-19, R041-11	MN182385, MN182386; MN182453, MN182454; MN182387, MN182388; MN182455, MN182456	R040-14, R040-19, R041-03, R041-11	SAMN12288101, SAMN12288102, SAMN12288103, SAMN12288104	Morocco, High Atlas, Azilal, Assif-n-Ait Bou Guemez S of Agouti, 1829 m	31°37'43.3"N 06°28'46.7"W	Vogt 17742, Oberprieler 11099 & Wagner	B 10 0760026
<i>R. gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R042	R042-01	MN182389, MN182390; MN182457, MN182458	R042-01	SAMN12288105	Morocco, Middle Atlas, Azilal, between Naour and Tagelft, 1141 m	32°26'18.3"N 05°59'15.0"W	Vogt 17746, Oberprieler 11103 & Wagner	B 10 0760019
<i>R. gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. s.l.	R043	R043-01	MN182391, MN182392; MN182459, MN182460	R043-01	SAMN12288106	Morocco, High Atlas, Midelt, Col Bab-n-Ouayad, 2661 m	32°13'02.2"N 05°41'17.4"W	Vogt 17752, Oberprieler 11109 & Wagner	B 10 0760012
<i>R. catananche</i> (Ball) B.H. Wilcox & al.	R044	R044-12	MN182393; MN182461	R044-02, R044-12	SAMN12288107, SAMN12288108	Morocco, High Atlas, Midelt, Col Bab-n-Ouayad, 2661 m	32°13'02.2"N 05°41'17.4"W	Vogt 17757, Oberprieler 11114 & Wagner	B 10 0760007
<i>R. arundanum</i> subsp. <i>mairei</i> (Humbert) Florian Wagner, Vogt & Oberpr.	R045	R045-25	MN182394; MN182462	R045-02, R045-06, R045-25	SAMN12288109, SAMN12288110, SAMN12288111	Morocco, High Atlas, Midelt, Col Bab-n-Ouayad, 2661 m	32°13'02.2"N 05°41'17.4"W	Vogt 17761, Oberprieler 11118 & Wagner	B 10 0760003

3.2.2 ITS, ETS and plastid marker sequencing

We sequenced the internal and external transcribed spacer (ITS and ETS) regions of the nuclear ribosomal repeat (nrDNA) and two intergenic spacer regions from the plastid genome (*trnC-petN*, *trnQ-rps16*) for one accession per population (43 accessions in total, Table 3.1). PCRs of ITS, ETS and plastid regions were carried out using *Taq* RED Polymerase (Ampliqon A/S, Odense, Denmark) and primers ITS-18SF (Rydin et al., 2004), ITS-26SR (Rydin, 2004), 18S-ETS (Baldwin and Markos, 1998), L-ETS (Lee et al., 2002), *trnC* (Demesure et al., 1995), *petN1R* (Lee and Wen, 2004), *trnQ2* (Shaw, 2007) and *rps16x1* (Shaw, 2007). After purification with AmpliClean™ magnetic bead-based PCR Cleanup (NimaGen, Nijmegen, Netherlands) all amplicons were sent to Macrogen Inc. (Amsterdam, The Netherlands) for Sanger sequencing in one or both directions.

Electropherograms were checked manually for base-call errors using CHROMAS LITE v2.0 (Technelysium Pty Ltd, South Brisbane, Australia). Sequences of plastid markers were concatenated, manually aligned and depicted in a TCS-network (Clement et al., 2002) with default settings in POPART v.1.7 (Leigh and Bryant, 2015). Whenever there was more than one ambiguous site in an individual electropherogram of the ITS or ETS region, we used the phasing software CHAMPURU v1.0 (Flot et al., 2006; Flot, 2007) or PHASE v2.1.1 (Stephens et al., 2001; Stephens and Scheet, 2005) to disentangle the underlying ITS/ETS copies as described in detail in Wagner et al. (2019). In two cases (R034-13 and R040-19), it was not possible to separate ITS copy types bioinformatically and PCR products were therefore cloned into a pJet cloning vector (Fermentas/Thermo Fisher Scientific Inc., Waltham, MA, USA). After transformation into NEB Turbo bacteria (New England Biolabs Inc., Ipswich, MA, USA), eight clones per accession were picked for colony PCR and sequenced as described above.

3.2.3 Double digest restriction associated DNA (ddRAD) sequencing

At least 300 ng of high molecular and RNA-free DNA of one to four accessions per population (102 accessions in total, Table 3.1) was sent to LGC Genomics (Berlin, Germany) for ddRAD sequencing (Poland et al., 2012). After restriction digestion with *PstI* and *ApeKI*, enzyme-specific adaptors were ligated to the fragmented DNA, including barcodes of different length (4-10 bp, Table S3.1). Individual samples were subsequently PCR-amplified and pooled, and the resulting library was normalized with the DSN enzyme from the Trimmer-2 cDNA normalization kit (Evrogen, Moscow, Russia) for reduction of abundant fragments. After gel-based size-selection (targeting 300-400 bp fragments), quality control and quantification of the library, paired-end sequencing (2 × 150 bp) was carried out on an Illumina NextSeq 500 instrument (Illumina Inc., San Diego, CA, USA). To enable detection

of genotyping errors and parameter optimization during de-novo assembly of reads, six randomly selected DNA samples were included twice in the ddRADseq procedure (individual sample replicates, Table 3.1).

3.2.3.1 Processing of RADseq data

Raw reads were quality-checked with FASTQC v.0.10.1 (Andrews, 2010) and demultiplexed according to their inline barcodes using BCL2FASTQ 2.17.1.14 (Illumina Inc., San Diego, CA, USA). Sequencing adaptor remnants were subsequently clipped from all reads before they were quality-filtered by (i) discarding reads with 5'-ends not matching the restriction enzyme site (ii) trimming of reads at their 3'-end, to ensure a minimum average phred quality score of >20 over a window of ten bases, and (iii) discarding reads with a final length <20 bp.

Pre-processed reads were passed through steps three to seven of the IPYRAD v.0.7.28 pipeline (Eaton and Overcast, 2016) for de-novo assembly of RADseq data. In the course of this pipeline, demultiplexed and quality filtered paired reads were initially merged and clustered de-novo within samples using VSEARCH v.2.6.0 (Rognes et al., 2016). During this step, we performed read clustering using different levels of stringency by varying the core parameter *clustering threshold* (*ct*) from 0.80 to 0.95 in incremental steps of 0.01. After cluster-wise alignment of reads with MUSCLE v.3.8.31 (Edgar, 2004), IPYRAD evaluated error rate and heterozygosity based on counts of site patterns across clustered reads. In this section of the workflow, we accepted a maximum of two alleles per site; a maximum of five uncalled bases was allowed for the following consensus base-calling step. Clustering across samples was subsequently performed with the same *ct* values as described above for the individual clustering step. To avoid paralogy, we finally discarded all loci that showed (i) heterozygous sites for more than 50% of the samples, (ii) more than two alleles per individual, or (iii) more than eight indels (default settings). Loci were additionally filtered based on the amount of missing data. For this purpose, the core parameter *minimal samples per locus* (*msl*) of IPYRAD assembly step 7 was varied from 4 to 108 in incremental steps of 8.

3.2.3.2 Comparison of multiple datasets to determine optimal parameter settings

We explored the effect of varying the key parameters *clustering threshold* (*ct*) and *minimum samples per locus* (*msl*) on de-novo assembly of reads within IPYRAD. For this purpose, we evaluated 224 datasets, generated with different combinations of *ct* and *msl* values (Table S3.2), by calculating error rates and information content for each dataset according to Mastretta-Yanes et al. (2015). Individual sample replicates were used to estimate three different error rates: (i) *locus error rates*, i.e. the ratio of loci present in only one of the

samples of a replicate pair to the total number of loci, (ii) *allele error rates*, i.e. the number of shared loci showing allele mismatches between replicate pairs by the total number of shared loci, and (iii) *SNP error rates*, i.e. the proportion of SNP mismatches between replicate pairs. Additionally, we assessed the number of RAD loci and the total number of single-nucleotide polymorphisms (SNPs) for each dataset. As a proxy for the phylogenetic structure present in the data, we further calculated the cumulative variation explained by the first two axes of a principal coordinates analysis (PCoA) for each dataset. Based on all measured quantities, three datasets with fixed (default) *clustering threshold* (*ct*), but varying *minimum samples per locus* (*msl*) values (12, 68, and 100) were selected for all following analyses (see chapter 3.3.2 for details).

3.2.4 Detection of hybrid individuals

We applied three different methods to our RADseq datasets to identify potential hybrid individuals in the study group. Neighbor-net networks were calculated using default settings in SPLITSTREE v.4.14.6 (Huson and Bryant, 2006) to get an impression of reticulate patterns in the study group. For this purpose, Kimura's two-parameter (K2P) genetic distance matrices were calculated in PAUP* v.4.0 (Swofford, 2003), based on concatenated SNPs of dataset *ct85msl12*, *ct85msl68* and *ct85msl100*, respectively.

Next, we adopted the strategy of Dillenberger and Kadereit (2017) based on Patterson's four-taxon D-statistics (Green et al., 2010, Durand et al., 2011) as implemented in IPYRAD v.0.7.28. In this method, hybridization is detected by evaluating alternative patterns (ABBA vs. BABA) of ancestral (A) and derived (B) alleles in quartet and pectinate topologies (((P1,P2),P3),O), where P1-P3 and O denote ingroup and outgroup taxa, respectively. As taxa circumscriptions in our study group were uncertain prior to the analyses, we initially clustered our accessions via consensus *k*-means clustering (Monti et al., 2003) and evaluated the optimal cluster number based on the Bayesian information criterion (BIC) for dataset *ct85msl12* (described in detail in chapter 3.2.5.1). The resulting seven clusters (c1-c7) were grouped into quartets by (i) defining *R. gyanum* (c6) and *R. catananche* (c7) as outgroup (O), (ii) building all possible pairs of 'ingroup clusters' (c1-c5), and (iii) assigning these cluster pairs to P1/P2 and P3, respectively (see Table 3.3). For each cluster combination, each possible combination of accessions was subsequently tested for ABBA vs. BABA patterns based on all SNPs of dataset *ct85msl12* in IPYRAD v.0.7.28. Resulting Z-scores, assessed from 1,000 bootstrap replicates, were used to test significance on a level of 0.01 after Holm-Bonferroni correction with the total number of tests to account for multiple testing. The percentage of significant tests was finally evaluated for each cluster combination to find

possible patterns of introgression among clusters according to Dillenberger and Kadereit (2017).

We used FASTSTRUCTURE v.1.0 (Raj et al., 2014) as a third method for detecting admixture in our study group. After exclusion of outgroup accessions (*R. gyanum* and *R. catananche*), we tested different numbers of clusters ($K = 1$ to 10) for each dataset (*ct85msl12*, *ct85msl68*, *ct85msl100*) using ten replicate runs per K and the simple prior in FASTSTRUCTURE v.1.0. The ‘chooseK’ algorithm (Raj et al., 2014) was subsequently applied to all runs for determining the optimal number of clusters (K) and CLUMPP v.1.1.2 (Jakobsson and Rosenberg, 2007) was used for combining results. CLUMPP was run with the greedy option, random input order and 1,000 repeats to combine replicate runs of optimal K for each dataset separately and to produce a combined (consensus) Q-matrix over all three datasets. Q-matrices were finally plotted with the R package POPHELPER v.2.2.6 (Francis, 2017) and all individuals with admixture proportions $>5\%$ in the consensus Q-matrix were treated as potential hybrids. To evaluate our hybrid detection approach, Neighbor-net and ABBA-BABA analyses were re-run after exclusion of putative hybrid individuals according to the FASTSTRUCTURE results.

3.2.5 Species delimitation analyses

3.2.5.1 Consensus *k*-means clustering

We used consensus *k*-means clustering (Monti et al., 2003; Wilkerson and Hayes, 2010) and the Bayesian information criterion (BIC) to investigate genetic structure and to find the ‘optimal’ number of species in our study group. In a first step, STRUCTURE files of IPYRAD assemblies *ct85msl12*, *ct85msl68* and *ct85msl100* were subjected to principal component analyses (PCA) with the R package adegenet v.2.1.1 (Jombart, 2008; Jombart and Ahmed, 2011). The resulting numerical data matrices were subsequently reduced by subsampling items (accessions) and features (characters) with a resampling rate of 0.8. Subsampled datasets were afterwards partitioned into k groups by *k*-means clustering. This process was repeated for 5,000 generations with k varying from 1 to 20 and a consensus matrix was finally generated for each k by calculating the proportion of clustering runs in which two accessions were grouped together. In the next step, the optimal number of clusters was determined with the Bayesian Information Criterion (BIC) (Schwarz, 1978). The ‘optimal’ consensus matrix for each dataset was finally sorted with the R package seriation v.1.2.3 (Hashler et al., 2008) and accessions were assigned to consensus clusters via UPGMA (unweighted pair-group method with arithmetic mean). To evaluate the effect of hybridization on consensus *k*-means clustering results, all analyses were repeated after exclusion of putative hybrid individuals according to chapter 3.2.4.

3.2.5.2 Multi-species coalescent (MSC) species delimitation

MSC species-delimitation analyses were conducted with the BEAST2 package STACEY v.1.2.4 (Jones et al., 2015; Jones, 2017a) based on the nrDNA ITS and ETS sequences. BEAUTI v.2.4.8 (Bouckaert et al., 2014) was used to generate two separate xml files using either all accessions or a reduced dataset without putative hybrid individuals. For each locus, we used a strict-clock model, a ploidy value of 2.0 and a site model according to the Akaike information criterion (AIC) in JMODELTEST v.2.1.10 (Darriba et al., 2012). All analyses were conducted with a Yule model and default priors as given in BEAUTI v.2.4.8, except for improper ones, which were changed according to the STACEY package documentation (Jones, 2017b). A lognormal distribution with mean (M) and standard deviation (SD) was used for the *growth rate* prior (M = 5.0 and SD = 2.0), the *clock rate* prior (M = 0.0 and SD = 1.0), and for the *scaling factor of the population size* prior (M = -7.0 and SD = 2.0). Furthermore, we assigned a flat prior to the possible number of species by defining a uniform distribution [0.0, 1.0] for the *collapseWeight* parameter. Two replicate runs were finally conducted for each xml file with 100 million generations and a sample frequency of 10,000 using BEAST v.2.5.1 on the CIPRES web portal (Miller et al., 2010). Convergence and ESS values were subsequently checked via TRACER v.1.7.1 (Rambaut et al., 2018) and replicate runs were combined with LOGCOMBINER v.2.4.8 after discarding 10% burn-in. Combined runs of each dataset were finally analyzed with TREEANOTATOR v.2.4.8 (*PP* limit = 0.5) and SPECIESDELIMITATIONANALYSER v.1.8.0 (*collapseheight* = 1.0e-4, *simcutoff* = 1.0) to obtain maximum clade credibility trees and tables of clusterings, which were visualized with FIGTREE v.1.4.3 and a customized R script provided by Jones et al. (2015).

As a second MSC method, we calculated marginal likelihoods for different species-delimitation methods with the BEAST2 package SNAPP v.1.4.2 (Bryant et al., 2012; Leaché et al., 2014a) and RADseq datasets *ct85msl12*, *ct85msl68*, and *ct85msl100*. Ten different species delimitation models (S01-S10) were generated by differentially lumping and splitting of ingroup taxa, populations and accessions based on plausible scenarios derived from prior analyses (see lower part of Figure 3.7a). Due to computational limitations, SNAPP was run without outgroup and replicate samples and with ten subsamples of each dataset including four randomly selected non-hybrid accessions of each ingroup taxon, respectively. SNAPP input files for different scenarios and datasets were prepared with the R library *phrynomics* (Barb Banbury, <http://github.com/bbanbury/phrynomics>) by (i) discarding non-binary SNPs, (ii) taking randomly a single SNP from each locus, and (iii) converting SNP data to the SNAPP binary format. For all 300 runs (three datasets, ten subsamples, and ten scenarios), a broad gamma distribution with *alpha* = 2 and *beta* = 200 was set in BEAUTI v.2.5.2 for the birth

rate prior of the Yule model (λ) and a gamma distribution with $\alpha = 1$ and $\beta = 700$ for the population size prior (γ). The mean of $\alpha/\beta \approx 0.0014$ of the latter distribution was estimated from pairwise p -distances among all individuals belonging to one taxon as recommended by the tutorial of Leaché and Bouckaert (2018). To account for different possibilities of assigning individuals to taxa, we averaged p -distances over all species delimitation scenarios and all three datasets. Path sampling analyses were finally conducted on the Athene HPC-cluster at the University of Regensburg to estimate marginal likelihoods for each species delimitation model of each dataset with $\alpha = 0.3$, $\text{chain length} = 100,000$, $\text{pre-burnin} = 50,000$, and 48 steps.

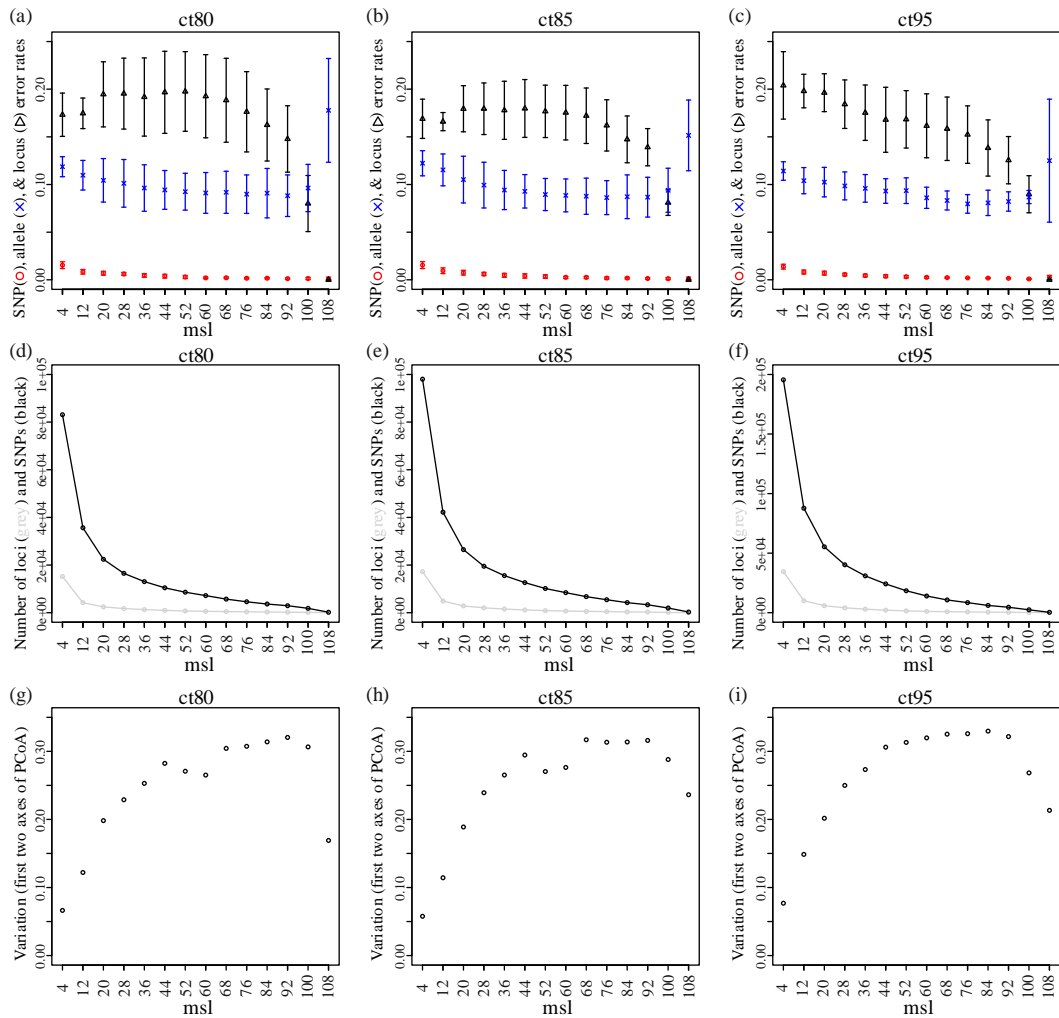


Figure 3.2 Error rates (a-c), number of loci and SNPs (d-f), and variation explained by the first two axes of PCoA (g-i) for different de-novo assemblies of RADseq reads generated with varying values of core parameters *clustering threshold* (ct) and *minimal samples per locus* (msl) in IPYRAD.

3.3 Results

3.3.1 Sanger and ddRAD sequencing output

Sanger sequencing of nrDNA ITS and ETS resulted in two alignments with a total length of 735 bp and 506 bp showing 38 and 31 parsimony informative sites (PIS), respectively. Concatenated plastid markers had a total length of 1,388 bp, including only 11 PIS and little phylogenetic resolution (see TCS-network in Figure 3.1). The number of de-novo assembled loci recovered by IPYRAD based on an average of 4,491,444 (SD = 2,245,722) pre-filtered reads per sample (Table S3.1) varied between 13 (*ct92msl108*) and 34,557 (*ct95msl4*) depending on the choice of assembly parameters (Table S3.2). The total number of recovered SNPs ranged between 183 (*ct92msl108*) and 195,466 (*ct95msl4*).

3.3.2 Comparison of multiple datasets to determine optimal parameter settings in IPYRAD

Evaluation of 224 different datasets revealed a contrasting impact of parameters *clustering threshold* (*ct*) and *minimal samples per locus* (*msl*) on quality and quantity of IPYRAD assemblies. While variation of *clustering thresholds* had little effect on error rates and amount of data, these quantities were strongly influenced by different *msl* values (Figure 3.2, Table 3.2 and Table S3.2). To account for this issue, we selected three datasets with varying *msl* values and default *ct* = 0.85: Dataset *ct85msl12* was selected due to its high amount of loci (4,888) and SNPs (42,204). This dataset showed, however, high *allele* (0.1153, SD: 0.0167) and *SNP error rates* (0.0096, SD: 0.0030) and a high percentage of missing data (70.85%), combined with little variation explained by the first two axes of PCoA (0.114). Although dataset *ct85msl4* showed even more loci and SNPs, we avoided this assembly due to its outstanding high *SNP error rate* of 0.0154 (SD: 0.0037). The second dataset selected (*ct85msl100*) was characterized by low *locus* (0.0809, SD: 0.0134), *allele* (0.1153, SD: 0.0167), and *SNP error rates* (0.0096, SD: 0.0030), but a low amount of loci/SNPs (154/1,977) with few missing data (3.92 %). We decided to choose *ct85msl100* instead of *ct85msl108*, as the latter dataset showed very high *allele error rates*, albeit *SNP* and *locus error rates* dropped (by definition) to zero for this parameter combination (Figure 3.2b). As a trade-off between *ct85msl12* and *ct85msl100*, we selected a third dataset (*ct85msl68*) with medium *allele/SNP error rates* (0.0877, SD: 0.0190; 0.0025, SD: 0.0011) and a medium amount of loci/SNPs (549/6,752) and missing data (17.30%). This dataset showed the highest percentage of variation explained by the first two axes of PCoA compared to all other datasets (Figure 3.2h).

3.3.3 Detection of hybrid individuals

Different datasets provided similar network topologies as depicted in Figures 3.3, S3.1, and S3.2. All clusters (c1-c7) that were reconstructed by consensus *k*-means clustering prior to ABBA-BABA tests, were found again in the Neighbor-net networks except for cluster c2 (*R. arundanum* individuals from High Atlas mountains), which was subdivided into two groups (Figures 3.3a, S3.1a, S3.2a). Individuals of ingroup clusters (c1-c4) in general showed more incompatible splits (illustrated by boxes in Neighbor-net networks) compared to outgroup clusters (c5-c7), pointing towards hybridization among ingroup taxa.

The percentage of significant ABBA-BABA tests conducted for 20 different combinations of clusters (c1-c5) varied between 0.00% and 42.47% (Table 3.3). The highest percentage of asymmetrical ABBA-BABA patterns was found for cluster combinations c2-c1 (42.47%), c3-c1 (36.10%), and c3-c2 (11.50%), hence involving populations of *R. arundanum* (c1 and c2) and *R. redieri* (c3).

The optimal cluster number of FASTSTRUCTURE runs varied between $K = 2$ and $K = 5$, depending on the dataset and the ‘chooseK’ metric (‘model complexity that maximizes marginal likelihood’ vs. ‘model components used to explain structure in data’). For the sake of comparability among datasets, we selected $K = 3$ for subsequent analyses, as this was the most frequently reconstructed optimal cluster number considering all runs (Table S3.3). Comparing FASTSTRUCTURE results for $K = 3$ among different IPYRAD assemblies (*ct85msl12*, *ct85msl68* and *ct85msl100*) revealed little variation in the assignment of individuals to clusters (Figure 3.4). In all three datasets, representatives of the ‘*R. maresii*’ lineage were clearly separated from ingroup accessions, which were in turn divided into two groups: A homogeneous group, comprising all 41 accessions of *R. arundanum* and a heterogeneous group, including 12 accessions of *R. redieri*, four individuals of *R. quezelii* plus all four members of the enigmatic population R038. Thirteen accessions from three different mountains showed admixture between both ingroup clusters with $PP > 0.05$ in the combined Q-matrix (see bar chart on top of Figure 3.4): (i) individuals R011-16 and R011-21 plus accessions of populations R010 and R012 from Djebel Bou Iblane, (ii) individual R033-03 and representatives of population R034 from Djebel Tichoukt, and (iii) all members of population R037 from Djebel Bou Ijallabene. These individuals were hereafter treated as potential hybrids.

Re-run of Neighbor-net and ABBA-BABA analyses after exclusion of putative hybrid individuals confirmed the efficiency of the above-described hybrid-detection approach: Neighbor-net networks showed less amount of incompatible splits (Figures 3.3b, S3.1b, and S3.2b) and the percentage of significant ABBA-BABA tests calculated for cluster combinations c2-c1, c3-c1 and c3-c2 decreased considerably (Table 3.3).

3.3 Results

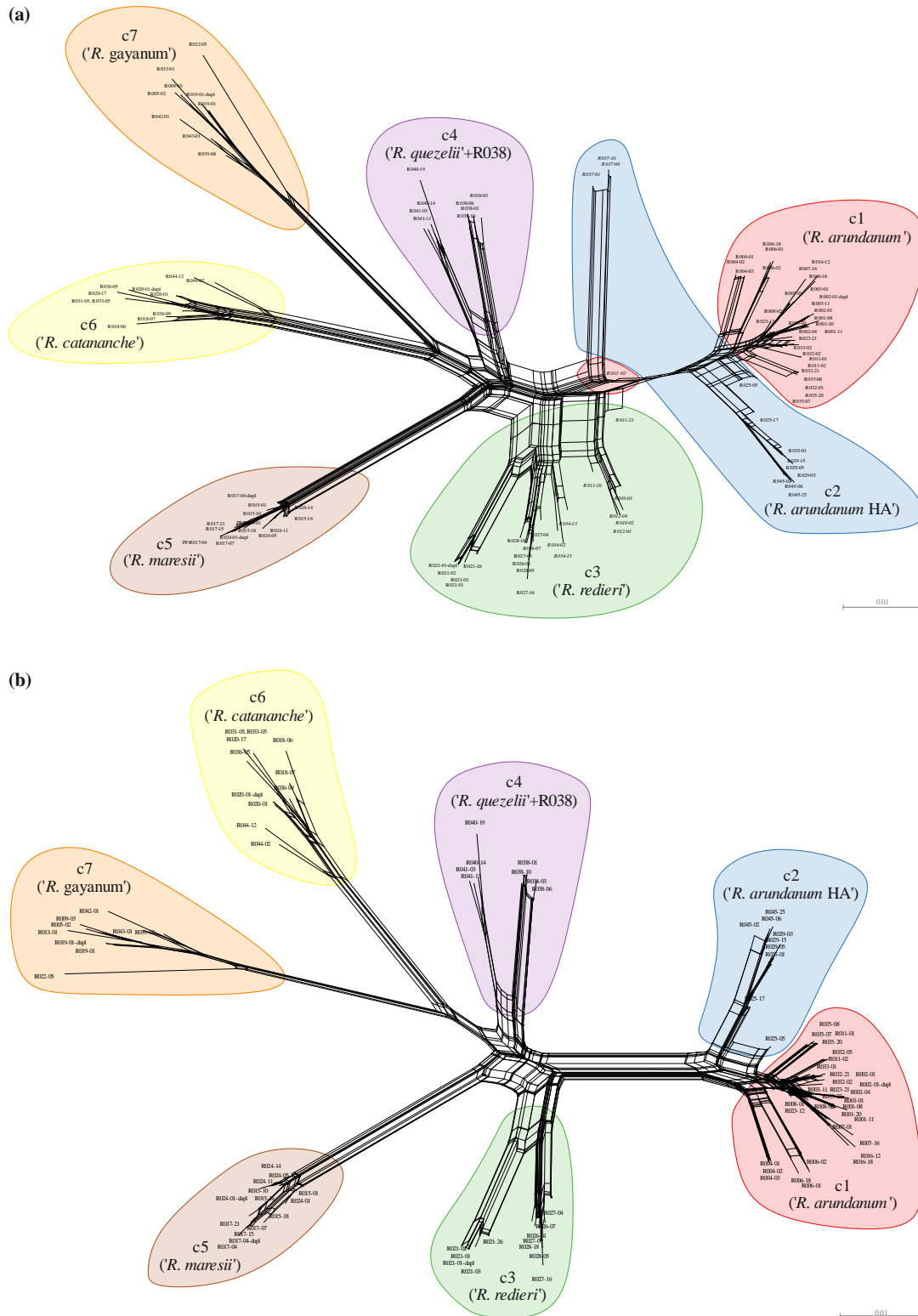


Figure 3.3 Neighbor-net networks based on concatenated SNPs of RADseq dataset *ct85ms/1100* including (a) all 108 samples and (b) a subset of 95 samples without signs of hybridization in FASTSTRUCTURE (see Figure 3.4). Colors indicate the assignment of populations to clusters c1-c7 according to consensus *k*-means clustering of dataset *ct85ms/12* (Figure 3.5), and potential hybrid individuals are italicized.

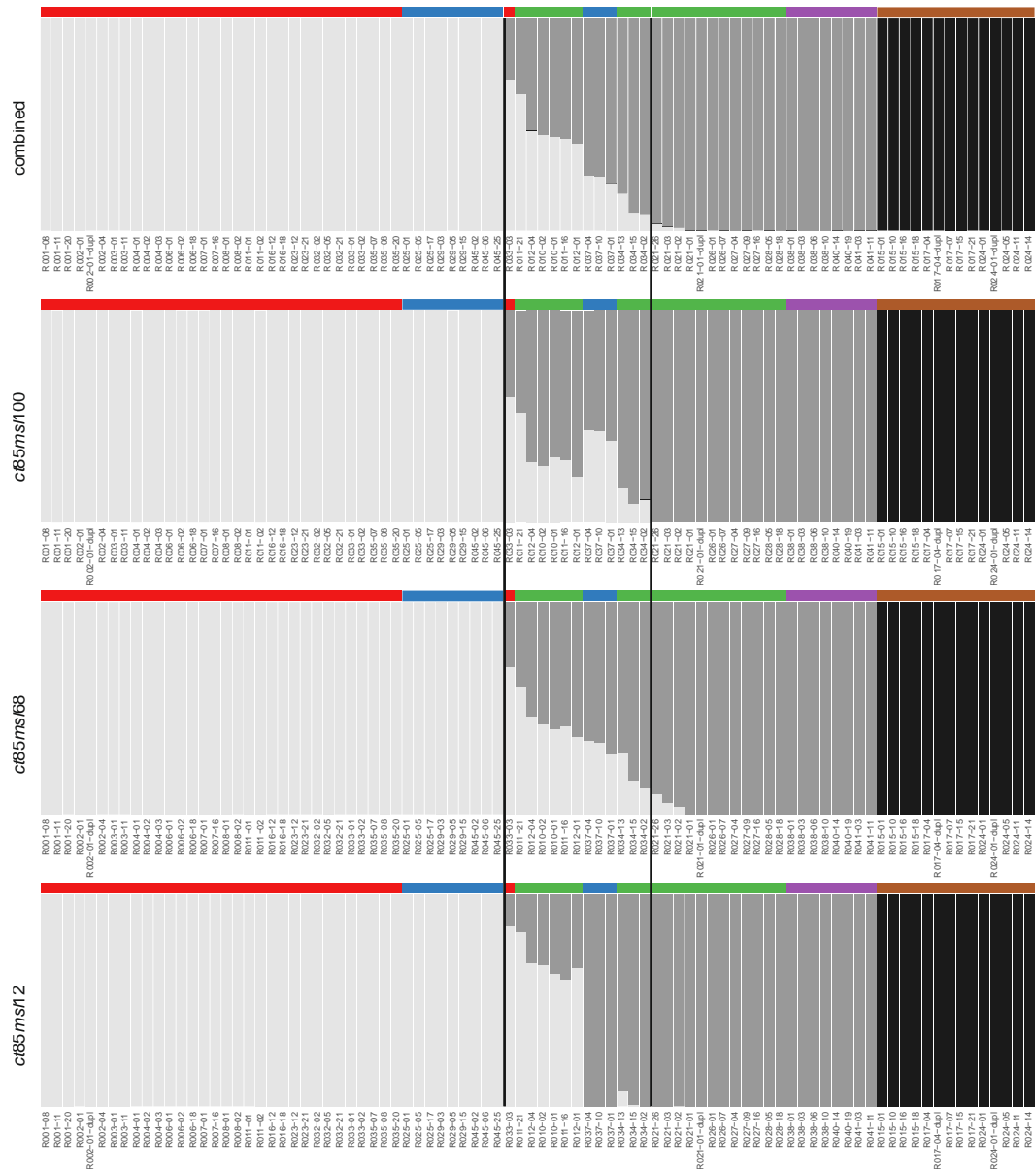


Figure 3.4 Results of FASTSTRUCTURE runs with $K = 3$ and three different RADseq assemblies *ct85msl12*, *ct85msl68*, and *ct85msl100*. Vertical lines indicate potential hybrid individuals showing admixture with $PP > 0.05$ in the combined Q-matrix over all datasets (top plot). Colors indicate the assignment of populations to clusters c1-c7 according to consensus k -means clustering of dataset *ct85msl12* (see Figure 3.5).

- c1 (*R. arundanum*)
- c2 (*R. arundanum* HA)
- c3 (*R. redieri*)
- c4 (*R. quezelii*+R038)
- c5 (*R. maresii*)

3.3.4 Species delimitation analyses

Results of consensus k -means clustering of different datasets (*ct85msl12*, *ct85msl68* and *ct85msl100*) are shown in Figures 3.5, S3.3, and S3.4. In all datasets, we found members of the same population being grouped into the same cluster, respectively, except for accessions of populations R011 (c1 and c3), R033 (c1, c3, and c6), and R025 (c1 and c2). Differences among datasets were found in the optimal number of clusters (k) according to the Bayesian Information Criterion (BIC). While IPYRAD assemblies *ct85msl12* and *ct85msl68* showed an

optimal cluster number of $k = 7$ (Figures 3.5a and S3.3a), *ct85msl100* resulted in 11 optimal clusters (Figure S3.4a). In the latter case, hybrid population R037 from Djebel Bou Ijallabene was split from cluster c2 and cluster c3 was subdivided into four sub-clusters: (i) a first cluster comprising all hybrid individuals from Djebel Bou Iblane (populations R010 and R012 plus R011-16 and R011-21), (ii) a second cluster including hybrid individuals from Djebel Tichoukt (population R034 plus R033-03) and (iii) a third and fourth cluster consisting of accessions of *R. redieri* subsp. *humbertii* (R026-R028) and subsp. *redieri* (R021), respectively. After exclusion of potential hybrids, consensus k -means clustering results of different IPYRAD assemblies converged noticeably with only minor discrepancies among datasets concerning the merging (*ct85msl12*, Figure 3.5) or splitting (*ct85msl68*, Figure S3.3b and *ct85msl100*, Figure S3.4b) of *R. arundanum* clusters c1 and c2.

Results from MSC species-delimitation analyses with the BEAST2 package STACEY based on nrDNA ITS/ETS sequences are depicted in Figure 3.6. Five groups of accessions could be delimited in the similarity matrix of the complete dataset with posterior probability (PP) values > 0.9 in the corresponding tree (Figure 3.6a). Outgroup accessions were assigned to three highly supported ($PP = 1.0$) groups in accordance to consensus k -means clustering results of RADseq data (cluster c5-c7). Another well-supported group ($PP = 0.98$) consisted of *R. quezelii* samples, accession R038-01 from Djebel Bou Ijallabene, and hybrid individual R037-04 from the same location. All remaining accessions, belonging either to *R. arundanum* or to *R. redieri*, were combined into a single group with $PP = 0.92$ in the tree and considerable substructure in the corresponding similarity matrix. After exclusion of putative hybrid individuals, we found not only a clear separation between *R. arundanum* and *R. redieri* accessions, but also a new sister-group relationship between individuals of the latter taxon and representatives of the *R. maresii*-lineage (although only weakly supported with $PP = 0.82$, Figure 3.6b).

Table 3.2 Error rates, amount of data, and variation explained by first two axes of PCoA for three different RADseq assemblies generated with fixed *clustering threshold* $ct = 0.85$ and varying values for the *minimum samples per locus (msl)* parameter in IPYRAD.

	<i>ct85msl12</i>	<i>ct85msl68</i>	<i>ct85msl100</i>
Mean locus error rate	0.1660 (SD 0.0095)	0.1722 (SD 0.0291)	0.0809 (SD 0.0134)
Mean allele error rate	0.1153 (SD 0.0167)	0.0877 (SD 0.0190)	0.0940 (SD 0.0231)
Mean SNP error rate	0.0096 (SD 0.0030)	0.0025 (SD 0.0011)	0.0011 (SD 0.0011)
Number of restriction site-associated DNA loci	4,888	549	154
Total number of SNPs	42,204	6,752	1,977
Total length of concatenated loci (bp)	914,749	91,994	29,140
Amount of missing data (%)	70.85	17.30	3.92
Variation explained by first two axes of PCoA	0.114	0.317	0.288

3.3 Results

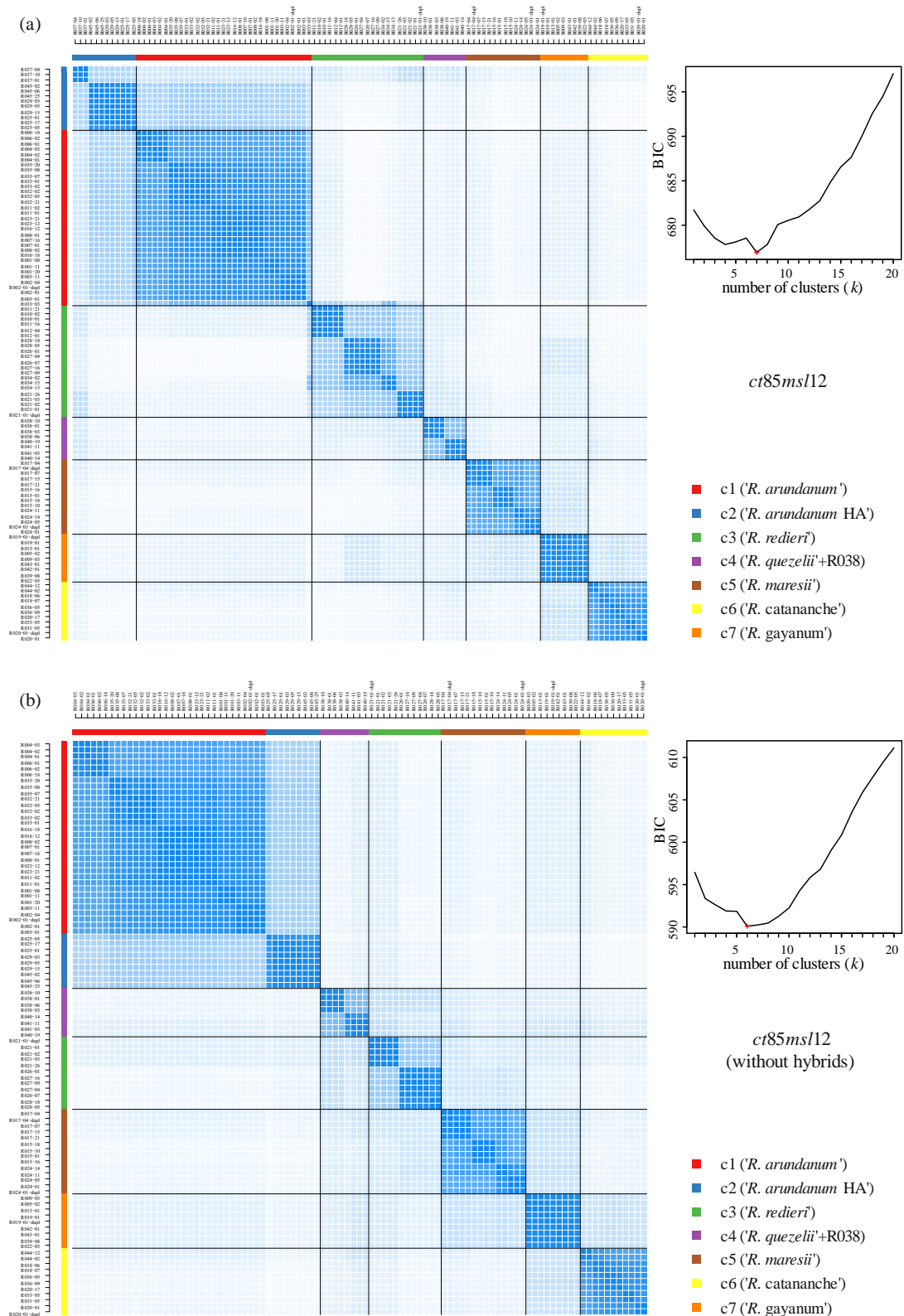


Figure 3.5 Consensus matrices resulting from consensus k -means clustering of RADseq dataset *ct85msl12*, including (a) all 108 samples and (b) a subset of 95 samples without signs of hybridization in FASTSTRUCTURE (see Figure 3.4). Pairwise amount of shared k -means clusters are indicated by shades of blue. Graphs on the right of the consensus matrices show the optimal k number of clusters determined via Bayesian Information Criterion (BIC).

3.3 Results

MSC species delimitation analyses with the BEAST2 package SNAPP based on different RADseq datasets are shown in Figures 3.7, S3.5, and S3.6 and Tables S3.4-S3.6. All datasets showed an ascending ranking of scenarios S01-S10 according to marginal likelihoods, which was particularly apparent in the case of assembly *ct85msl12* with the highest quantity of SNPs (Figure 3.7a). In all three IPYRAD assemblies, the six-species scenario S10 received the highest log-marginal likelihood values. In this scenario, populations of *R. arundanum* from the High Atlas mountains were separated from the remaining representatives of the same taxon and the two subspecies of *R. redieri* as well as *R. quezelii* and population R038 were treated as independent lineages. The two-species scenario S01, which was designed according to the FASTSTRUCTURE results (Figure 3.4) received by far the lowest support. Scenarios S02-S09 in between these extremes showed a clear tendency towards higher support for more complex (species-rich) hypotheses. Comparing scenarios with the same number of species (S03-S06 and S07-S09) revealed insights into the degree of divergence among delimited lineages: splitting of *R. arundanum* accessions into distinct species obtained lower marginal likelihood support than dividing *R. redieri* samples into two species or treating population R038 as an own lineage (independent of *R. quezelii*). This successive order of divergence was also apparent from increasing branch lengths of corresponding lineage pairs in the species tree of the ‘optimal’ scenario S10 (Figures 3.7b, S3.5b, S3.6b).

Table 3.3 Results from Patterson’s four-taxon D-statistics (ABBA-BABA test). While c6 (*R. catananche*) and c7 (*R. gayanum*) were fixed as outgroup, all combinations of clusters c1-c5 were assessed to P1/P2 and P3, respectively. Number of individual tests (*n*) and percentage of significant test [% *n*Sig (0.01)] after Holm-Bonferroni correction are given for the complete dataset *ct85msl12* and after exclusion of potential hybrids.

P1	P2	P3	complete		without hybrids	
			<i>n</i>	% <i>n</i> Sig (0.01)	<i>n</i>	% <i>n</i> Sig (0.01)
c2 (<i>R. arundanum</i> HA')	c2 (<i>R. arundanum</i> HA')	c1 (<i>R. arundanum</i>)	2178	42.47	1152	6.94
c2 (<i>R. arundanum</i> HA')	c2 (<i>R. arundanum</i> HA')	c3 (<i>R. redieri</i>)	1254	3.19	360	0.83
c2 (<i>R. arundanum</i> HA')	c2 (<i>R. arundanum</i> HA')	c4 (<i>R. quezelii</i> +R038)	528	2.08	288	5.21
c2 (<i>R. arundanum</i> HA')	c2 (<i>R. arundanum</i> HA')	c5 (<i>R. maresii</i>)	924	7.14	504	7.14
c1 (<i>R. arundanum</i>)	c1 (<i>R. arundanum</i>)	c2 (<i>R. arundanum</i> HA')	6336	4.59	4464	0.36
c1 (<i>R. arundanum</i>)	c1 (<i>R. arundanum</i>)	c3 (<i>R. redieri</i>)	10032	2.35	4960	1.05
c1 (<i>R. arundanum</i>)	c1 (<i>R. arundanum</i>)	c4 (<i>R. quezelii</i> +R038)	4224	4.85	3968	5.04
c1 (<i>R. arundanum</i>)	c1 (<i>R. arundanum</i>)	c5 (<i>R. maresii</i>)	7392	3.79	6944	4.15
c3 (<i>R. redieri</i>)	c3 (<i>R. redieri</i>)	c2 (<i>R. arundanum</i> HA')	2052	11.50	405	0.00
c3 (<i>R. redieri</i>)	c3 (<i>R. redieri</i>)	c1 (<i>R. arundanum</i>)	5643	36.10	1440	0.00
c3 (<i>R. redieri</i>)	c3 (<i>R. redieri</i>)	c4 (<i>R. quezelii</i> +R038)	1368	0.37	360	0.56
c3 (<i>R. redieri</i>)	c3 (<i>R. redieri</i>)	c5 (<i>R. maresii</i>)	2394	0.00	630	0.00
c4 (<i>R. quezelii</i> +R038)	c4 (<i>R. quezelii</i> +R038)	c2 (<i>R. arundanum</i> HA')	336	3.57	252	3.97
c4 (<i>R. quezelii</i> +R038)	c4 (<i>R. quezelii</i> +R038)	c1 (<i>R. arundanum</i>)	924	1.84	896	2.23
c4 (<i>R. quezelii</i> +R038)	c4 (<i>R. quezelii</i> +R038)	c3 (<i>R. redieri</i>)	532	4.70	280	1.79
c4 (<i>R. quezelii</i> +R038)	c4 (<i>R. quezelii</i> +R038)	c5 (<i>R. maresii</i>)	392	0.00	392	0.00
c5 (<i>R. maresii</i>)	c5 (<i>R. maresii</i>)	c2 (<i>R. arundanum</i> HA')	1092	1.56	819	1.71
c5 (<i>R. maresii</i>)	c5 (<i>R. maresii</i>)	c1 (<i>R. arundanum</i>)	3003	2.50	2912	2.51
c5 (<i>R. maresii</i>)	c5 (<i>R. maresii</i>)	c3 (<i>R. redieri</i>)	1729	0.40	910	0.66
c5 (<i>R. maresii</i>)	c5 (<i>R. maresii</i>)	c4 (<i>R. quezelii</i> +R038)	728	4.53	728	4.40

3.4. Discussion

3.4.1 Optimization of de-novo assembly parameters and evaluation of RADseq genotyping errors

Restriction-site associated DNA sequencing (RADseq) allows the collection of vast amounts of sequence data for non-model organisms, irrespective whether whole genome resources are available or not. The success of de-novo assembly of raw RADseq reads, however, is strongly dependent on the choice of reasonable core parameters throughout different steps of bioinformatic pipelines like STACKS (Catchen et al., 2011, 2013) or IPYRAD/IPYRAD (Eaton and Ree, 2013; Eaton 2014; Eaton and Overcast, 2016). Inappropriate sequence similarity thresholds (parameters *ct* in IPYRAD and *M* in STACKS) can lead to both, under- and over-merging of reads (Paris et al., 2017; McCartney-Melstad et al., 2019). If the *clustering threshold* is too low, paralogous and repetitive genomic regions are incorrectly assigned to one single cluster/locus. Setting the *clustering threshold* too high, on the other hand, may result in the splitting of true allelic variants of orthologous loci into different clusters/loci. Equally crucial is the handling of missing data by setting a minimum number of individuals (parameter *msl* in IPYRAD) necessary for keeping a given locus in the final dataset (Huang and Knowles, 2016).

Different strategies exist for optimizing parameter space in the course of de-novo assembly of RADseq reads. Paris et al. (2017) released a method for optimizing core parameters in STACKS, based on the maximization of the number of polymorphic loci present in >80% of the samples. Their ‘80% rule’ can be applied to find the trade-off between over- and under-merging by increasing the number of loci (e.g. by increasing the *clustering threshold*) until the splitting of alleles leads to a drop in the amount of loci due to the 80% filter criterion. A similar approach was performed in McCartney-Melstad et al. (2019), where a set of quantities (e.g. fraction of loci inferred as paralogs, percentage of heterozygous sites or phylogenetic resolution) was evaluated to find the upper bound for the *clustering threshold* in IPYRAD at which true alleles are incorrectly separated into distinct clusters. While both approaches were successfully applied to RADseq datasets including populations of single species, their performance remains unclear in the presence of strong population structure or species-level divergence, where high amounts of missing data are expected due to allele and locus dropout. Mastretta-Yanes et al. (2015) used individual sample replicates for optimizing de-novo assembly parameters within STACKS, by simultaneously minimizing error rates and maximizing the amount of informative loci. The central premise behind this approach is the assumption that replicates derived from the same DNA share the same genotype pattern. Differences between replicate pairs can be traced back either to errors introduced during wet laboratory procedure and sequencing or to improper adjustment of parameters during

assembly of reads. Hence, including sample replicates can be helpful for (i) optimizing core parameters for de-novo assembly of RADseq reads (ii) evaluating genotyping errors and (iii) comparing the results of different RADseq studies.

In the present survey, we used individual sample replicates to evaluate genotyping errors and for optimizing the core parameters *clustering threshold* (*ct*) and *minimum samples per locus* (*msl*) in IPYRAD. For this purpose, we calculated (i) *locus*, *allele* and *SNP error rates*, (ii) amount of loci and SNPs, and (iii) variation explained by the first two axes of PCoA for 224 RADseq data matrices, generated with different combinations of *ct* and *msl* values in IPYRAD. In concordance to the *Berberis* datasets of Mastretta-Yanes et al. (2015), we found *locus error rates* in our assemblies being characterized by high mean values (typically >0.1) and high standard deviations, regardless of which parameter values were used in IPYRAD (Figure 3.2a-c). As speculated in the former study, the high degree of non-perfect overlap of RADseq loci between samples from the same DNA source is most likely attributable to heterogeneous coverage among loci due to PCR and/or sequencing biases. Thus, locus errors are mostly introduced during the laboratory part of the RADseq process and seem to be relatively high regardless of which pipeline and parameters are used for processing of reads. Mismatches between alleles of replicate pairs (*allele error rates*), on the other hand, may be promoted by the acceptance of PCR or sequencing errors as allelic variation during de-novo assembly of RADseq reads and are therefore more likely to be pipeline-dependent. The slightly higher *allele error rates* in *Rhodanthemum* (>0.08) compared to *Berberis* (>0.05) are, however, probably rather a side effect of the increased length of utilized Illumina reads in our study (150bp paired-end vs. 100bp single-end), which have a higher chance of containing at least one erroneous nucleotide. In contrast to the slightly higher *allele error rates*, we found *SNP error rates* being consistently lower in *Rhodanthemum* (<0.02) as in *Berberis* (0.02 to 0.12), which indicates an overall high percentage of correctly called SNPs in our data matrices, regardless of which parameters were used in IPYRAD.

In contrast to the investigated error rates, which showed only small variations among datasets, we found the quantity of loci and SNPs considerably varying among different IPYRAD assemblies. Decreasing the threshold for missing data (by decreasing *msl*) led to an exponential increase of the amount of loci and SNPs in our assemblies (Figure 3.2d-f). This exponential connection between tolerance for missing data and data-matrix size points towards a shallow divergence of investigated taxa as demonstrated in the simulation study of Huang and Knowles (2016). This finding fits to the relatively recent diversification of the genus *Rhodanthemum* during the last 1.3 million years according to a dated phylogeny of the whole subtribe Leucantheinae in Wagner et al. (2019).

Overall, and in contrast to Mastretta-Yanes et al. (2015), we found no ‘optimal’ parameter combination in IPYRAD that simultaneously minimized error rates and maximized the amount

of informative loci and cumulative variance of PCoA. Furthermore, we found a higher impact of the threshold for missing data (*m**sl*) on the quality and quantity of RADseq assemblies as it was the case for the *clustering threshold* (*ct*). A possible explanation for the low influence of the parameter *ct* on the quality of our RADseq assemblies may lie in the reduction of abundant RAD fragments during library preparation (see chapter 3.2.3). This normalization step could have minimized the number of fragments from paralogous and repetitive genomic regions during the lab part of our RADseq procedure, reducing the necessity for finding an optimal condition for splitting reads of such regions during de-novo assembly of reads (mainly controlled by *ct*).

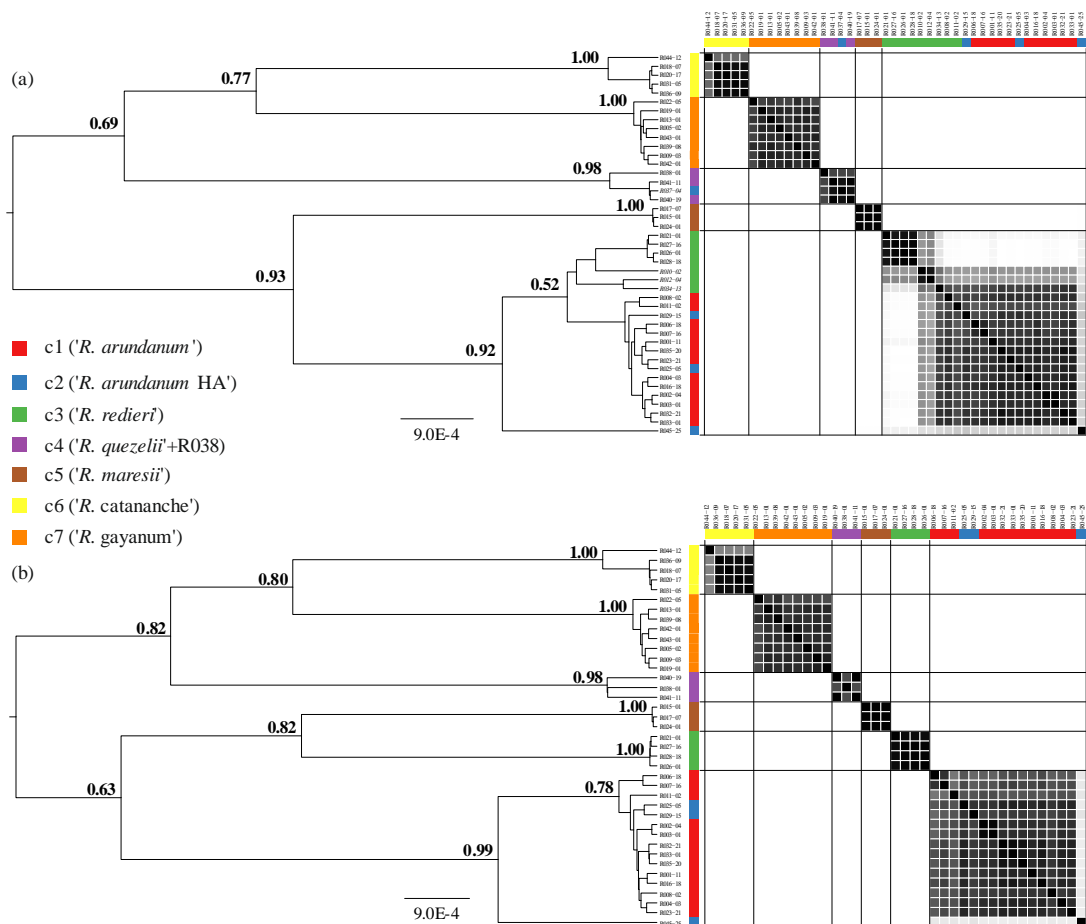


Figure 3.6 Results of joint species-tree (left) and clustering analyses (right) using the BEAST2 application STACEY and nrDNA ITS and ETS sequences of (a) all 43 accessions and (b) a subset of 39 accessions without signs of hybridization in FASTSTRUCTURE (see Figure 3.4). Similarity matrices visualize posterior probabilities (PP) for pairs of individuals belonging to the same cluster (black: $PP = 1.0$, white: $PP = 0.0$). Colors indicate the assignment of populations to clusters c1-c7 according to consensus *k*-means clustering of dataset *ct85msl12* (see Figure 3.5), and potential hybrid individuals are italicized.

3.4.2 Hybridization patterns in the *R. arundanum*-group

Numerous studies have shown that interspecific hybridization is a common phenomenon in the tribe Anthemideae of the Compositae family (e.g., Lo Presti et al., 2010; Himmelreich et al., 2014; Konowalik et al., 2015; Oberprieler et al., 2019). In *Rhodanthemum*, some evidence exists for the occurrence of interspecific hybridization, albeit a recent study of Wagner et al. (2019) showed that reticulate evolution played a much smaller role in the history of the genus compared to the closely related genus *Leucanthemum*.

Artificial crossing experiments in the early 1980's showed, that each of eleven investigated *Rhodanthemum* taxa (at that time part of a larger genus *Leucanthemum*) was potentially capable of exchanging genes with at least one other (Wilcox and Harcourt, 1982). Nevertheless, only two interspecific crosses of the latter study (*R. arundanum* × *R. redieri* and *R. arundanum* × *R. gyanum*) produced viable hybrids in a reciprocal manner. In accordance to the findings of Wilcox and Harcourt (1982), Dobignard (2015) suspected the occurrence of natural hybrids between *R. arundanum* and *R. redieri* in the Atlas mountains of Morocco. However, his assumption was only based on personal observations and was not yet evaluated by more detailed morphological or molecular studies.

In the present survey, we used RAD sequencing data for the detection of hybridization patterns among taxa of the *R. arundanum*-group by evaluating (i) Neighbor-net networks, (ii) ABBA-BABA tests and (iii) admixture patterns in FASTSTRUCTURE. Different analyses coherently identified 13 admixed individuals, including ten hybrids between *R. arundanum* and *R. redieri*, in accordance to the crossing experiments of Wilcox and Harcourt (1982) and assumptions by Dobignard (2015). In contrast to the former study, however, no indications for the existence of natural hybrids between *R. arundanum* and *R. gyanum* could be found in our RADseq assemblies.

Hybrids between *R. redieri* and *R. arundanum* were detected at two Middle Atlas mountains (Djebel Bou Iblane and Djebel Tichoukt), where they grow sympatrically with 'pure' representatives of *R. arundanum*. Surprisingly, we found no trace of 'pure' individuals of *R. redieri* at both locations in the course of our study. This imbalance could either be explained by a sampling bias due to the difficulty of locating *R. redieri*, which preferential grows on steep rock faces, or by the occurrence of asymmetric introgression between both taxa, with higher migration rates from *R. arundanum* into *R. redieri* than vice versa. This phenomenon has been frequently described for pairs of hybridizing plant species (e.g. *Helianthus annuus* and *H. debilis*, Scascitelli et al., 2010; *Orchis militaris* and *O. purpurea*, Jacquemyn et al., 2012) and can even culminate in the genetic swamping of one species by the other as shown for *Senecio ovatus* and *S. hercynicus* (Bog et al., 2017). The discrimination among the above-described scenarios, however, requires a broader sampling of the potential

hybrid swarms at both locations, which was not the focus of the underlying study. Interestingly, and in contrast to our observations in the Middle Atlas mountains, no signs of hybridization between *R. arundanum* and *R. redieri* were found in the High Atlas mountains, albeit both taxa co-occur at Djebel Ajachi (R025-R029, Figure 3.1). These findings may argue for the discrimination of two subspecies in both taxa with different affinities for producing natural hybrids with each other (see also discussion below).

All remaining hybrids in our survey were representatives of a single population (R037) from the High Atlas mountain Bou Ijallabene. This hybrid population is possibly the product of a hybridization event between *R. arundanum* and a yet undescribed taxon (represented by population R038) from the same location. This hypothesis is supported by (i) morphological similarities between *R. arundanum* accessions and individuals of population R037 (personal observation), (ii) their co-occurrence at the Djebel Bou Ijallabene (Figure 3.1 and Wagner et al., 2019) and (iii) genetic similarities between representatives of population R037 and accessions of *R. arundanum* from the High Atlas mountains according to consensus *k*-means clustering of RADseq data (Figure 3.5a). The involvement of population R038 as the second parental part, on the other hand, is evident from the high similarities of ITS and ETS sequences between individuals of R037 and R038 (Figure 3.6a), possibly due to concerted evolution after hybridization (Wendel et al., 1995). Anyhow, the hybrid origin of population R037 is less clear from our Neighbor-net, FASTSTRUCTURE and ABBA-BABA analyses compared to the above described hybridization between *R. arundanum* and *R. redieri* and therefore should be treated with caution.

Apart from cases of gene flow among representatives of the closely-knit *R. arundanum*-group, we found no further evidence for interspecific hybridization events in our study group. Particularly, we found no signs for hybridization between ingroup and outgroup taxa, in contrast to successful reciprocal crossings between *R. arundanum* and *R. gayanum* in the study of Wilcox and Harcourt (1982). This is in line with the overall low hybridization signal in the whole genus *Rhodanthemum* (Wagner et al., 2019) and indicates the existence of prezygotic and/or postzygotic isolation mechanisms among most *Rhodanthemum* taxa. The sharing of plastid haplotypes among different *Rhodanthemum* lineages in the study of Wagner et al. (2019) and in the present contribution (Figure 3.1) is therefore rather explained by incomplete lineage sorting (Maddison, 1997) and the low variability of plastid markers than by reticulate evolution.

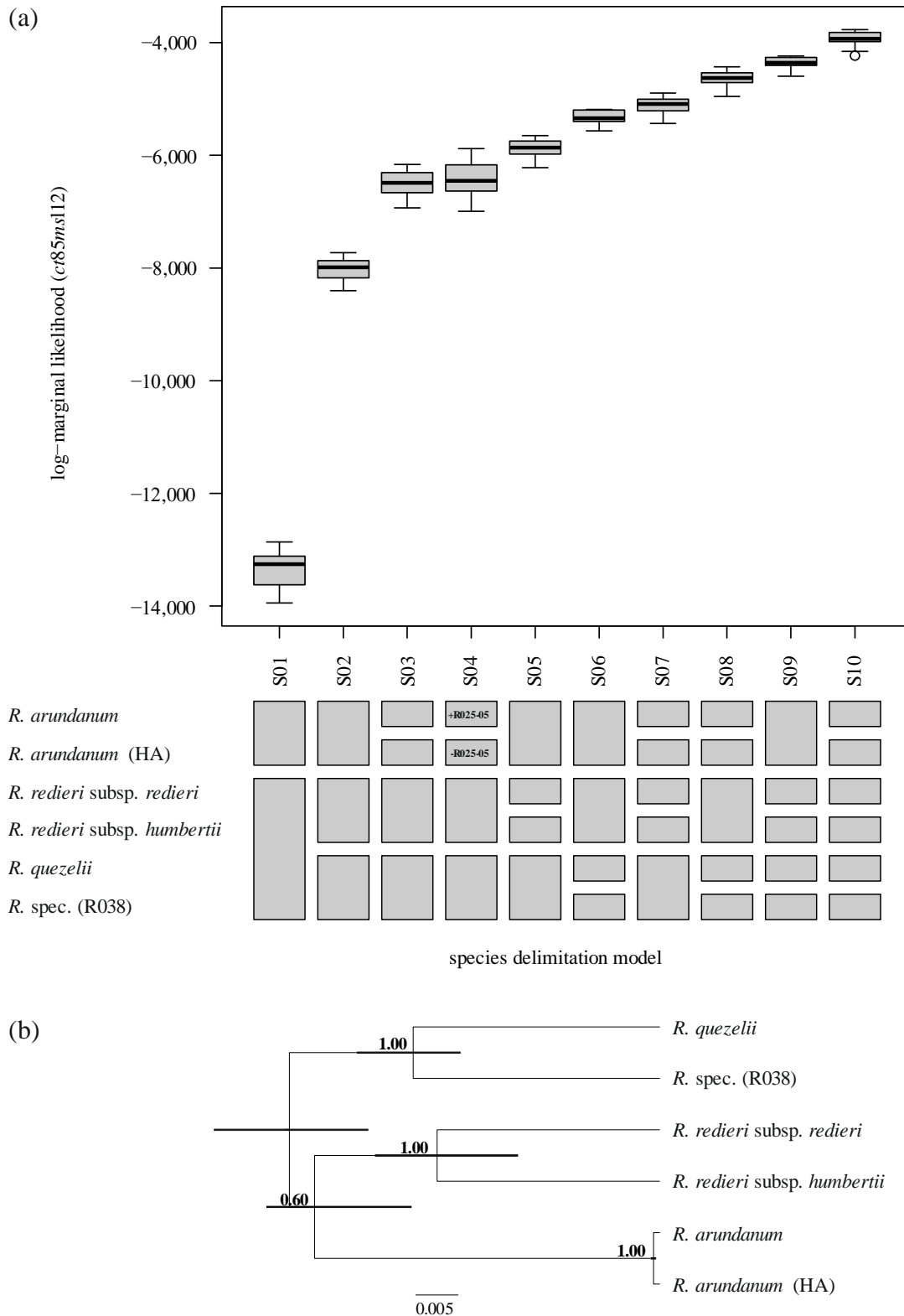


Figure 3.7 Comparison of different species delimitation models S01-S10 via marginal likelihood values calculated with the BEAST package SNAPP based on 3,837 SNPs of RADseq dataset *ct85msl12*. Box plots in (a) follow the standard convention, with solid lines reflecting the median, hinges the first and third quartiles, and whiskers the first and third quartiles plus $1.5 \times$ the interquartile range. Outliers are depicted by circles. The species tree in (b) corresponds to model S10 and includes posterior probability values above 0.5 and 95% highest posterior densities (HPD) for node heights.

3.4.3 Evaluation of different species delimitation analyses

We applied a wide range of species delimitation (SD) methods (consensus k -means clustering, STACEY and SNAPP) on different kinds of datasets (rDNA ITS/ETS data and multi locus RADseq assemblies, both with and without hybrid individuals) for delineating species boundaries in the recently diverged *R. arundanum*-group. This approach allowed us to evaluate the performance of different SD methods in the presence of hybridization and various quantities of data.

As shown in Zhang et al. (2011) and Wagner et al. (2017), multi-species coalescent (MSC) species-delimitation methods [like BPP (Yang and Rannala, 2010) or DISSECT/STACEY (Jones et al., 2015; Jones 2017a)] are prone to underestimate the number of species in the presence of strong interspecific hybridization. This blurring effect of gene flow on species delimitation is particularly apparent in the STACEY analyses of our ITS/ETS sequences, where the inclusion of hybrid individuals leads to the obscuring of species boundaries between *R. arundanum* and *R. redieri* (Figure 3.6a). Only after excluding these hybrids, we found both taxa being clearly separated from each other with high support (Figure 3.6b) and in accordance to the results of all RADseq-based analyses. This finding underlines the importance of carefully checking for potential violations of model assumptions (here: no gene flow after speciation), before applying coalescent models of species delimitation (see also Fujita et al., 2012 and Carstens et al., 2013). While STACEY underestimated the number of species in our study group due to hybridization, a contrasting effect was found for consensus k -means clustering of RADseq dataset *ct85msl100*. Here, we found an over-estimation of the number of entities (11 instead of 7 clusters) when hybrid individuals were included into the analysis. However, this over-splitting was probably also the result of difficulties in discriminating among different k -scenarios in assembly *ct85msl100* via the Bayesian information criterion, due to the low information content of this dataset (see plateau of BIC values in Figure S3.4a).

Comparing the results of different MSC-based species delimitation methods (STACEY and SNAPP) gives insights into the performance of the underlying model given various quantities of data. While STACEY analyses based on two loci (ITS and ETS) resulted in the delimitation of three ingroup species (*R. arundanum*, *R. redieri* and *R. quezelii*), SNAPP analyses based on hundreds to thousands of RADseq loci supported a scenario with twice as much entities (Figure 3.7). This observation is in line with the assumption of Sukumaran and Knowles (2017) that in the light of genomic data, increasingly finer-scaled genetic structure can be detected under the multi-species coalescent model. Moreover, the authors of the latter study suspect, that the MSC model rather delimits population structure than species divergences when data of many loci are analyzed (but see also Leaché et al., 2018). Due to the detection

of a considerable higher number of species compared to all other analyses of our study, we believe that SNAPP has rather detected population-level structure in our RADseq data than species boundaries. Anyhow, MSC-based methods like SNAPP are still worthwhile tools for studying divergence patterns, as they unveil meaningful genetic structure in (genomic) data, which can be valuable for taking taxonomic decisions on infraspecific levels (see taxonomical conclusions).

3.4.4 Conclusions and taxonomical/nomenclatural implications

Following the recommendation of Carstens et al. (2013), we place our trust in those species-delimitation results that are congruent across different methods. After exclusion of hybrids, STACEY analysis of ITS and ETS sequences and consensus *k*-means clustering of RADseq dataset *ct85msl12* consistently identified three in-group species in our study group, namely *R. arundanum*, *R. redieri*, and *R. quezelii* (including the enigmatic population R038). This consensus is remarkable due to the contrasting amounts of loci/SNPs of both datasets (2/88 vs. 4,888/42,204) and the contrasting nature of the applied analysis methods (multi-species coalescent vs. consensus *k*-means clustering). Furthermore, we found evidence for the recognition of two subspecies in *R. arundanum*, showing differences in (i) their geographical distribution (High Atlas mountains vs. remaining study area), their genetic constitution (see consensus *k*-means clustering of RADseq datasets *ct85msl68* and *ct85msl100* as well as SNAPP analyses), and (iii) their affinity for hybridization with the closely related *R. redieri*. Furthermore, our analyses support the distinction of two subspecies of the latter species (subsp. *redieri* in the Middle Atlas mountains and subsp. *humbertii* in the High Atlas mountains) as proposed by Gómiz (2000). Finally, the genetic similarity of population R038 to accessions of *R. quezelii* (STACEY analyses and consensus *k*-means clustering), despite differences in morphology (see above), argue for the designation of the former population as a subspecies (subsp. *ijallabenense*) of the latter taxon. As a consequence of our analyses, we propose the acknowledgment of two new taxa on subspecies rank and two new hybrid names/combinations:

- (1) *Rhodanthemum arundanum* subsp. *mairei* (Humbert) Flor. Wagner, Vogt & Oberpr., **comb. nov.** \equiv *Leucanthemum mairei* Humbert in Bull. Soc. Hist. Nat. Afrique N. 15: 201. 1924 \equiv *Leucanthemum arundanum* var. *mairei* (Humbert) Maire in Bull. Soc. Hist. Nat. Afrique N. 28: 362. 1937 \equiv *Leucanthemum arundanum* subsp. *mairei* (Humbert) Cuatrec. in Cavanillesia 1: 43. 1928.

Lectotype (designated by R. Vogt in Taxon 54: 482. 2005): H. Humbert, pl. du Maroc (1923), No. 884, Grand Atlas oriental, Ari Ayachi, escarpements et rocailles calcaires entre Tittasuine et le culminant, 2500-3500 m, 11.7., *H. Humbert* (P! [P00486669]; isolectotypes: BC!, MPU-Afrique du Nord! [MPU001117], P! [P00486670], RAB! [RAB034979]).

Notes: *Rhodanthemum arundanum* subsp. *mairei* is distributed in the mountains of the High Atlas in Morocco.

(2) *Rhodanthemum quezelii* subsp. *ijallabenense* Flor. Wagner, Vogt & Oberpr., **subsp. nov.**

Holotype: Morocco, Region Drâa-Tafilalet, Province Midelt, High Atlas, Sidi Yahia Ou Youssef, Assaka, gorge Arhbalou-n-Oussaka between Djebel Bou Ijallabene and Irhil ou Abbar, limestone cliffs, 32°21'41.8" N – 05°22'23.9" W, 1794 m, 16.06.2017, *R. Vogt 17739*, *C. Oberprieler 11096* & *F. Wagner* (B! [B100760029]; isotype: RAB!).

Diagnosis: Differs from *Rhodanthemum quezelii* subsp. *quezelii* by its 1-2-pinnatisect and sparsely hairy (with basifixed hairs) to glabrescent leaves.

Notes: Presently *Rhodanthemum quezelii* subsp. *ijallabenense* is only known from the area around its *locus classicus* in the gorge Arhbalou-n-Oussaka between Djebel Bou Ijallabene and Irhil ou Abbar in the High Atlas, where it grows on steep limestone cliffs. The new taxon was collected for the first time in July 1989 by Ch. Oberprieler [B100550383]. The name *ijallabenense* refers to the Djebel Bou Ijallabene in the Moroccan High Atlas.

(3) *Rhodanthemum* × *pseudoredieri* Flor. Wagner, Vogt & Oberpr., **nothosp. nov.**
[*Rhodanthemum arundanum* (Boiss.) B. H. Wilcox & al. subsp. *arundanum* × *R. redieri* (Maire) B. H. Wilcox & al. subsp. *redieri*]

Holotype: Morocco, Region Fès-Meknès, Province Guercif, Middle Atlas, Djebel Bou Iblane, surroundings of Tizi Bou Zabel, limestone cliffs, 33°38'44.5"N – 04°09'17.8"W, 2275 m, 09.06.2017, *R. Vogt 17651*, *C. Oberprieler 11008* & *F. Wagner* [B! (B100704730); isotype: RAB!]

Diagnosis: Genetically and in terms of morphological characters of indumentum and leaf outline intermediate between *Rhodanthemum arundanum* (Boiss.) B. H. Wilcox & al. subsp. *arundanum* and *R. redieri* (Maire) B. H. Wilcox & al. subsp. *redieri*.

Notes: This hybrid occurs in the joint distribution range of *Rhodanthemum arundanum* and *R. redieri* in the mountains of the Middle Atlas. Presently it is known from Djebel Tichoukt and Djebel Bou Iblane.

- (4) *Rhodanthemum arundanum* subsp. *mairei* (Humbert) Flor. Wagner, Vogt & Oberpr.
× *R. quezelii* subsp. *ijallabenense* Flor. Wagner, Vogt & Oberpr.

Notes: Presently this hybrid is only known from the gorge Arhbalou-n-Oussaka between Djebel Bou Ijallabene and Irhil ou Abbar in the High Atlas, where it grows together with its parental taxa on limestone cliffs and stony slopes. It is morphologically indistinguishable from *R. arundanum* subsp. *mairei* and its hybrid character is only evident from molecular investigations.

3.5 Supplemental Figures and Tables

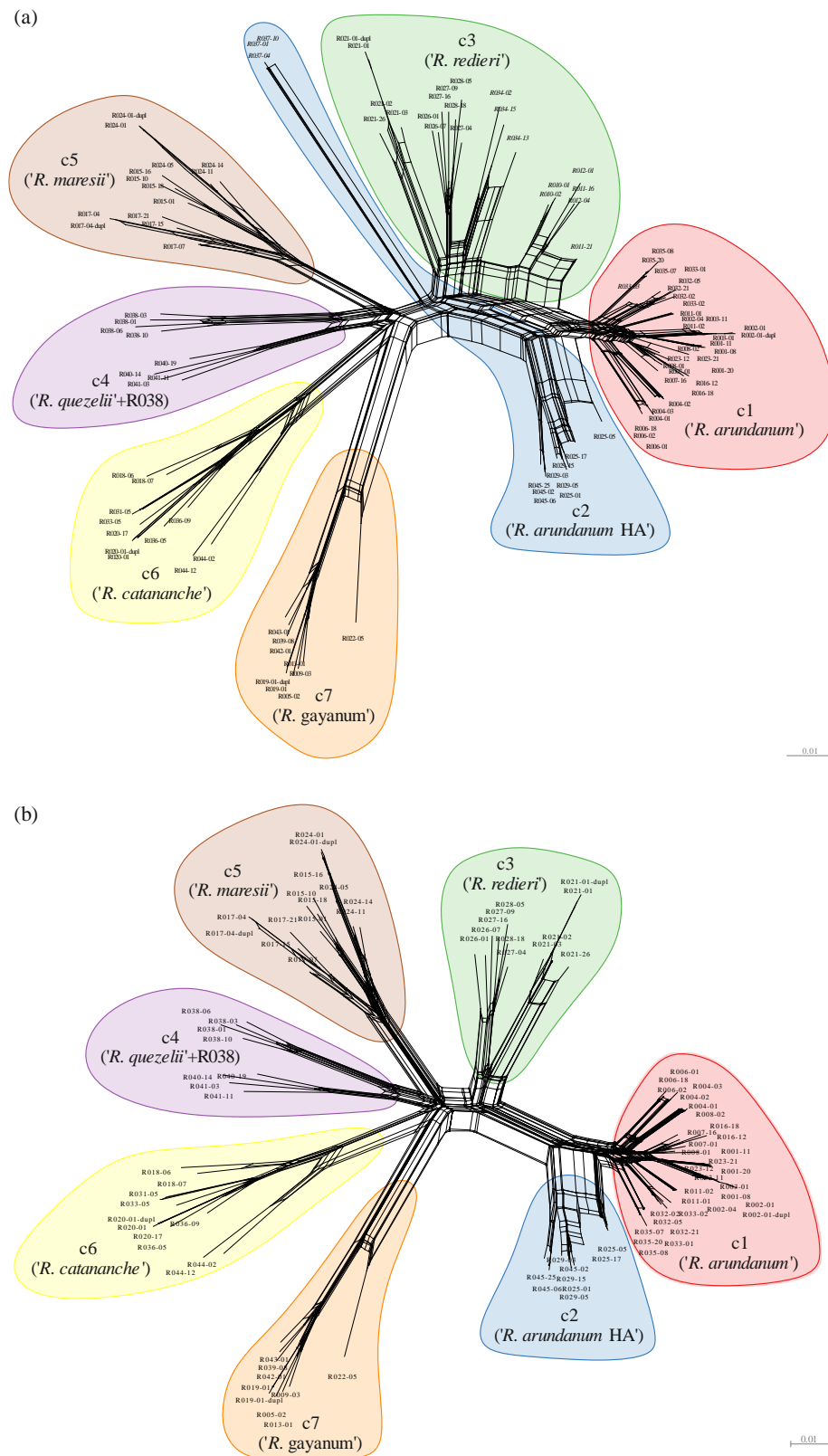


Figure S3.1 Neighbor-net networks based on concatenated SNPs of RADseq dataset *ct85msl12* including (a) all 108 samples and (b) a subset of 95 samples without signs of hybridization in FASTSTRUCTURE (see Figure 3.4). Colors indicate the assignment of populations to clusters c1-c7 according to consensus *k*-means clustering of dataset *ct85msl12* (see Figure 3.5), and potential hybrid individuals are italicized.

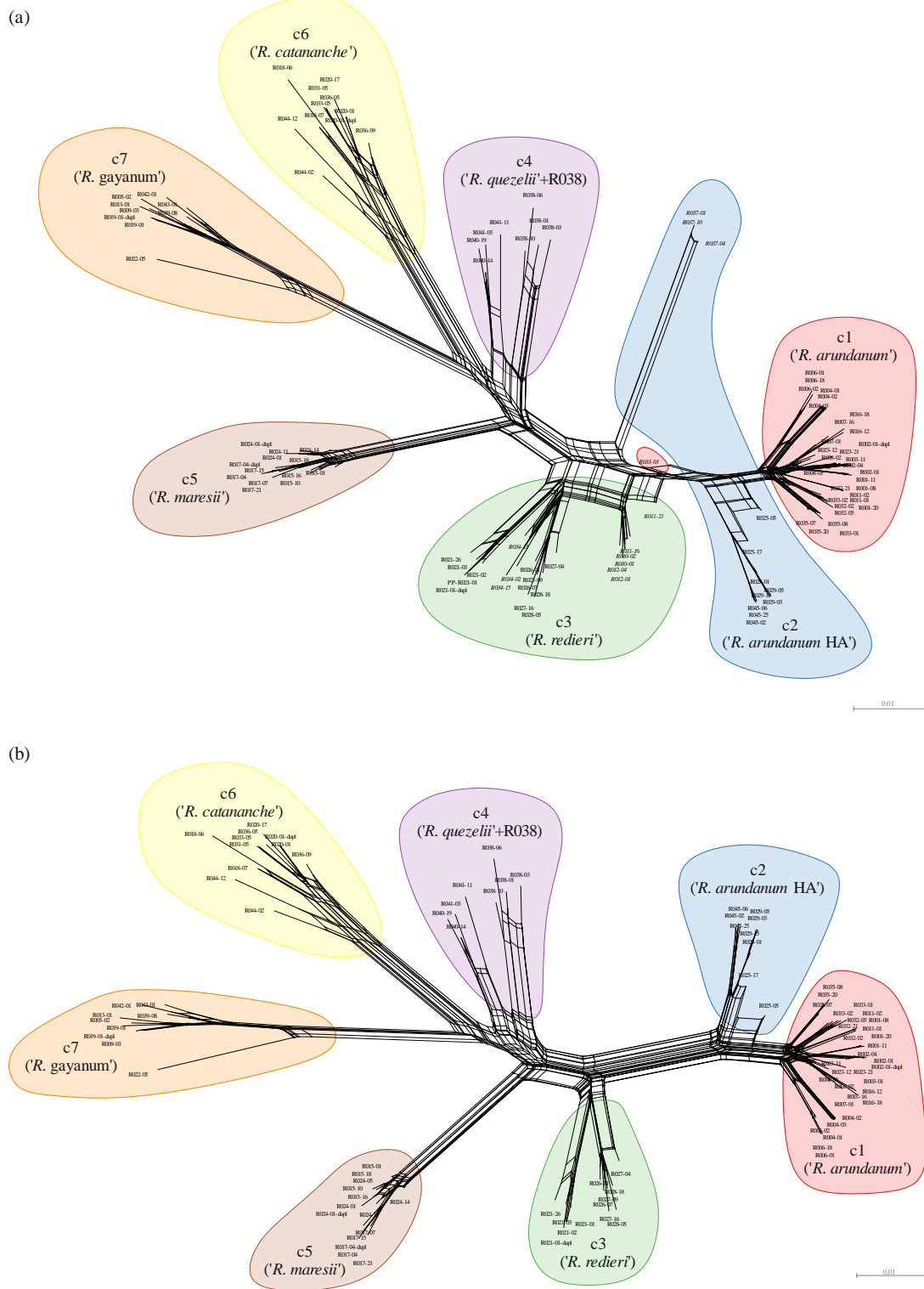


Figure S3.2 Neighbor-net networks based on concatenated SNPs of RADseq dataset *ct85msl68* including (a) all 108 samples and (b) a subset of 95 samples without signs of hybridization in FASTSTRUCTURE (see Figure 3.4). Colors indicate the assignment of populations to clusters c1-c7 according to consensus *k*-means clustering of dataset *ct85msl12* (see Figure 3.5), and potential hybrid individuals are italicized.

3.5 Supplemental Figures and Tables

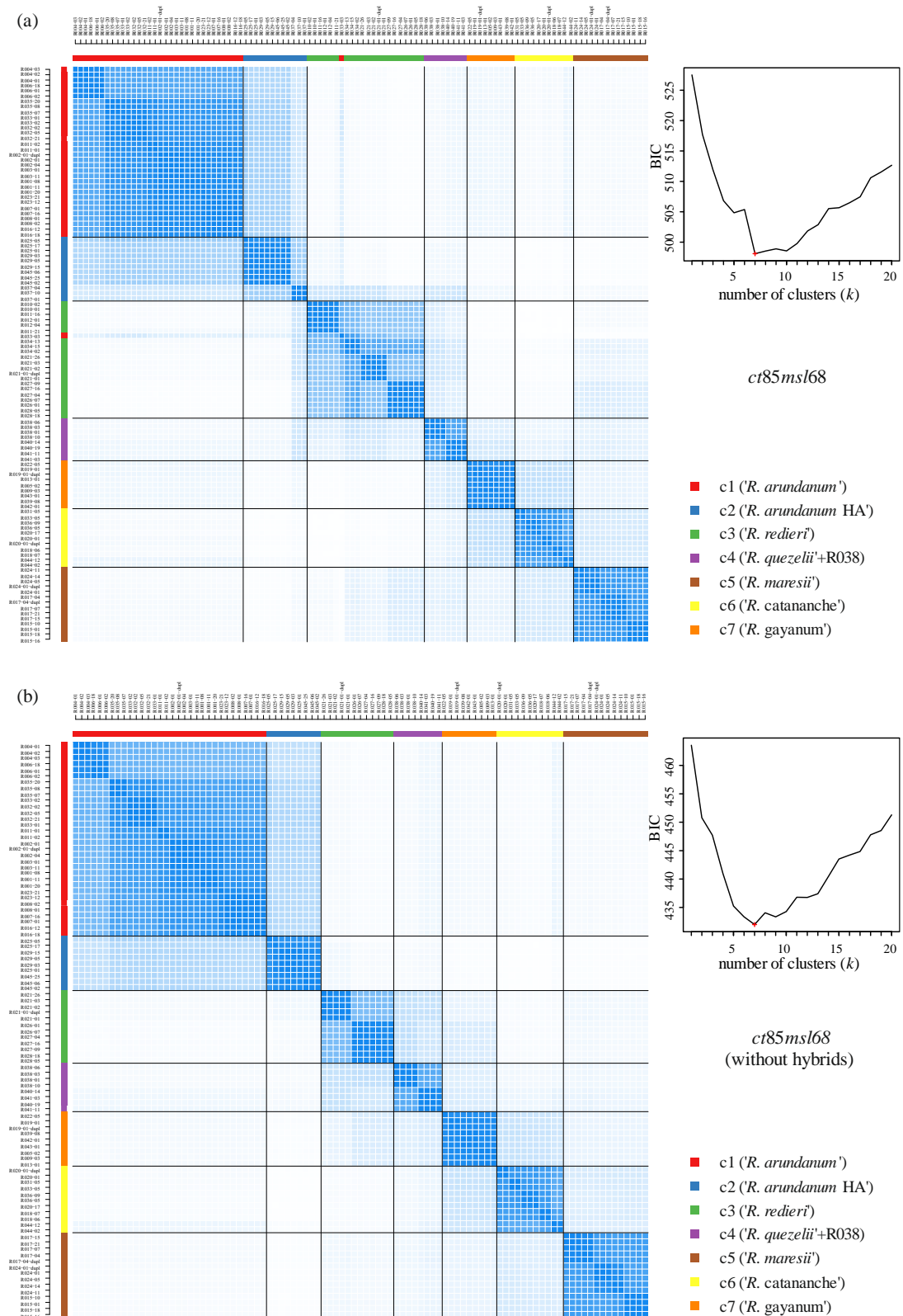


Figure S3.3 Consensus matrices resulting from consensus k -means clustering of RADseq dataset *ct85msl68*, including (a) all 108 samples and (b) a subset of 95 samples without signs of hybridization in FASTSTRUCTURE (see Figure 3.4). Pairwise amount of shared k -means clusters are indicated by shades of blue. Graphs on the right of the consensus matrices show the optimal number of clusters determined via Bayesian Information Criterion (BIC). Colors indicate the assignment of populations to clusters c1-c7 according to consensus k -means clustering of dataset *ct85msl12* (see Figure 3.5).

3.5 Supplemental Figures and Tables

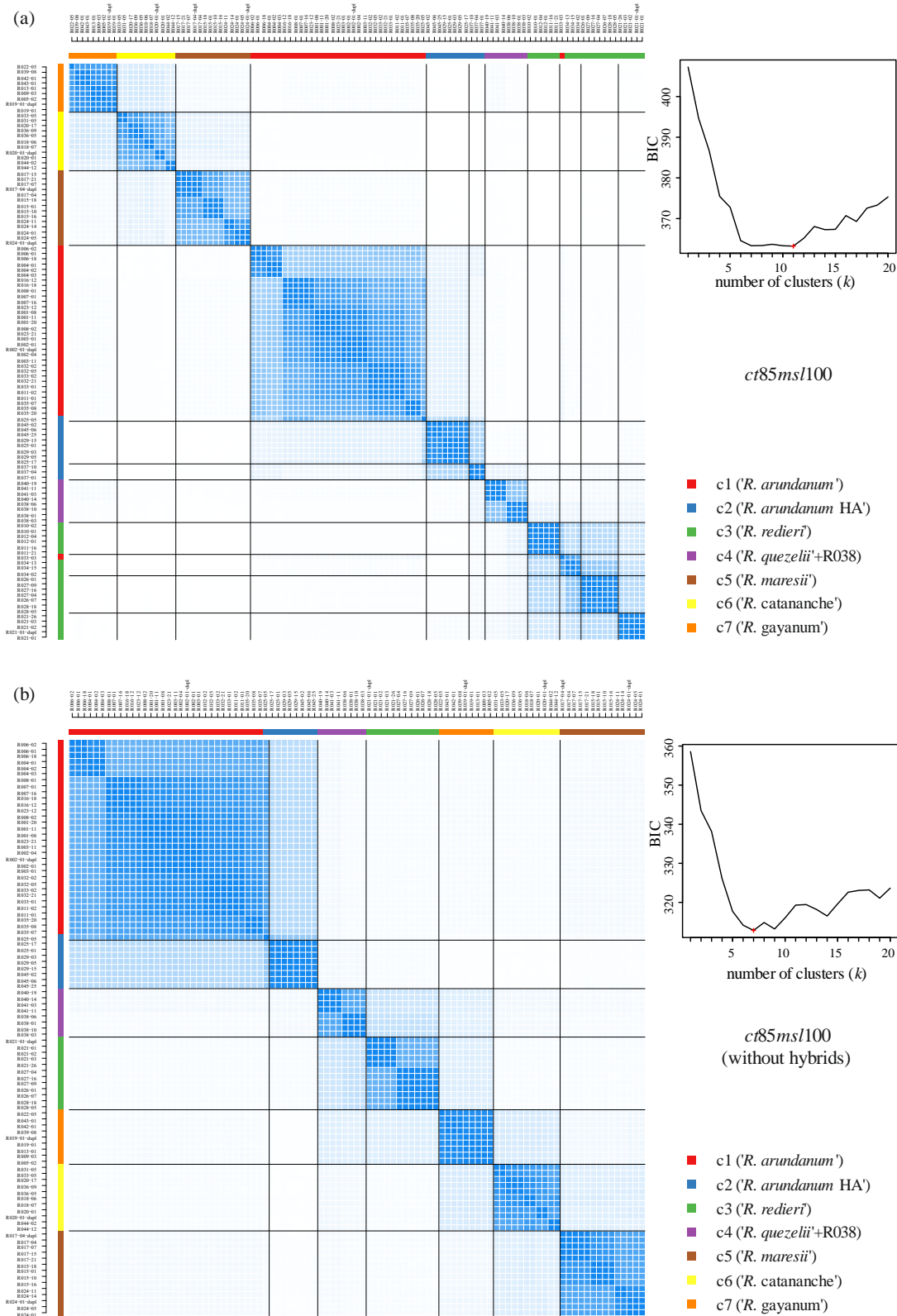


Figure S3.4 Consensus matrices resulting from consensus k -means clustering of RADseq dataset *ct85msl100*, including (a) all 108 samples and (b) a subset of 95 samples without signs of hybridization in FASTSTRUCTURE (see Figure 3.4). Pairwise amount of shared k -means clusters are indicated by shades of blue. Graphs on the right of the consensus matrices show the optimal number of clusters determined via Bayesian Information Criterion (BIC). Colors indicate the assignment of populations to clusters c1-c7 according to consensus k -means clustering of dataset *ct85msl12* (see Figure 3.5).

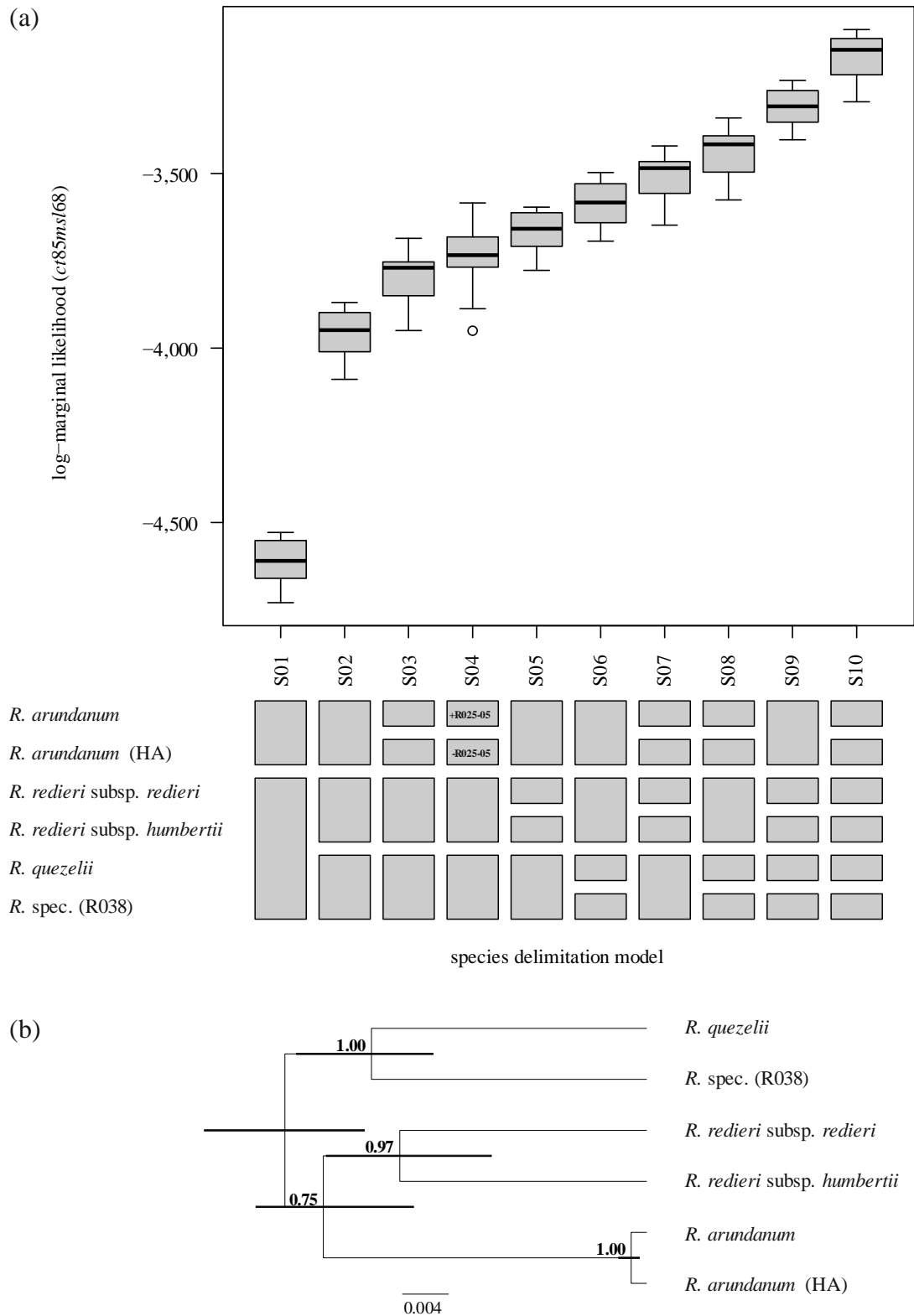


Figure S3.5 Comparison of different species delimitation models S01-S10 via marginal likelihood values calculated with the BEAST package SNAPP based on 537 SNPs of RADseq dataset *ct85msl68*. Box plots in (a) follow the standard convention, with solid lines reflecting the median, hinges the first and third quartiles, and whiskers the first and third quartiles plus $1.5 \times$ the interquartile range. Outliers are depicted by circles. The species tree in (b) corresponds to model S10 and includes posterior probability values above 0.5 and 95% highest posterior densities (HPD) for node heights.

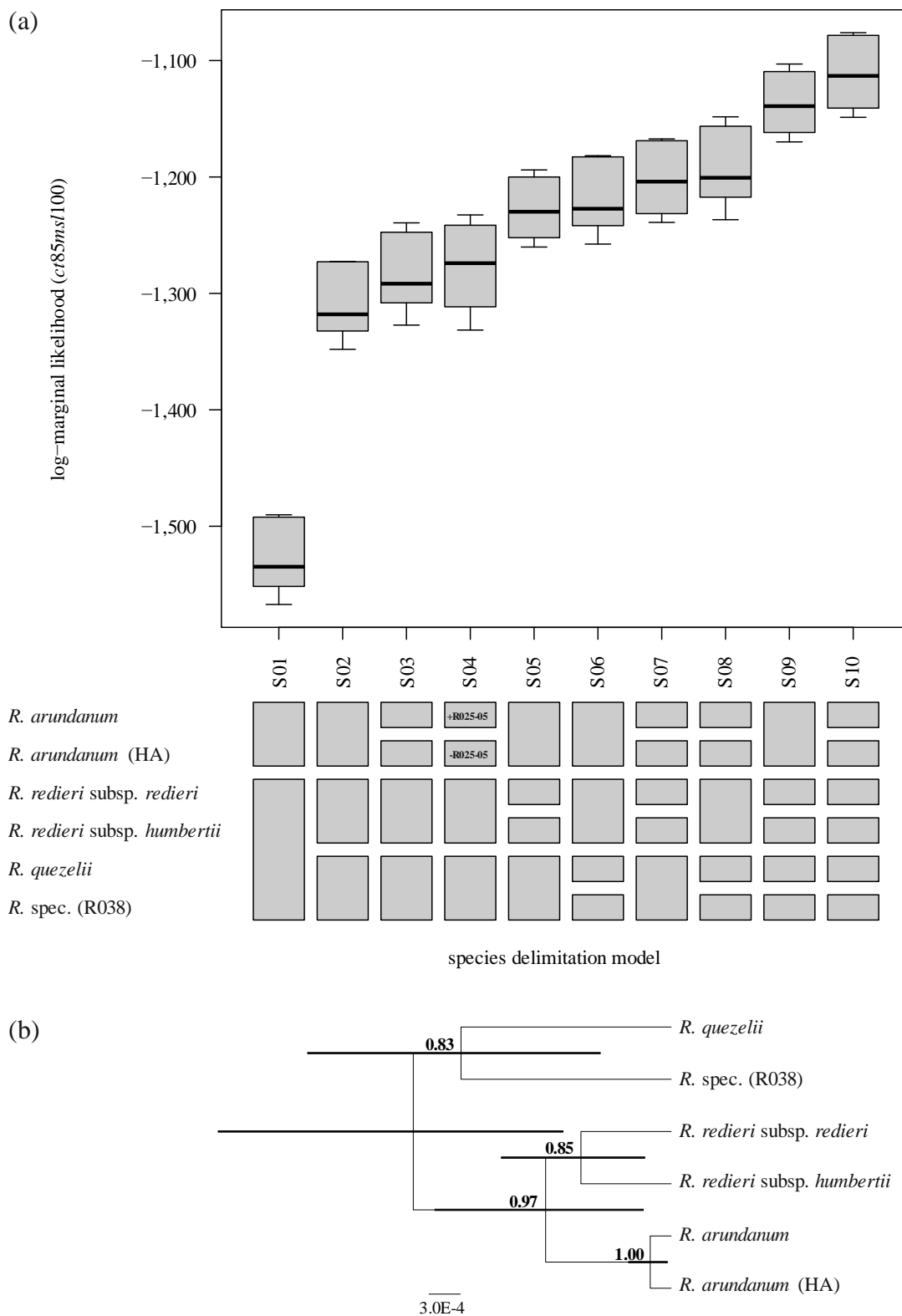


Figure S3.6 Comparison of different species delimitation models S01-S10 via marginal likelihood values calculated with the BEAST package SNAPP based on 153 SNPs of RADseq dataset *ct85ms/100*. Box plots in (a) follow the standard convention, with solid lines reflecting the median, hinges the first and third quartiles, and whiskers the first and third quartiles plus $1.5 \times$ the interquartile range. The species tree in (b) corresponds to model S10 and includes posterior probability values above 0.5 and 95% highest posterior densities (HPD) for node heights.

3.5 Supplemental Figures and Tables

Table S3.1 *Rhodanthemum* accessions used for ddRAD sequencing including information about Illumina plate well, barcode and number of reads before and after trimming during quality-filtering.

Sample Name	Plate well	Barcode	Raw total reads	Quality trimmed reads
R001-08	A06	TCCGCA	11,457,154	11,385,882
R001-11	B06	TCTCA	7,026,788	6,990,890
R001-20	C06	TGCGAGA	5,508,204	5,480,724
R002-01	A11	CTTGA	7,522,290	7,482,444
R002-01-dupl	G11	ATATCGCCA	8,475,074	8,412,474
R002-04	D06	TGCTGAA	5,071,582	5,046,260
R003-01	E06	TGGC	9,942,886	9,893,550
R003-11	F06	TTGACCAG	9,670,760	9,619,252
R004-01	G06	TTGCGTCT	3,234,414	3,216,578
R004-02	H06	TTGTAG	7,692,014	7,655,132
R004-03	A07	TGACGCCA	3,686,396	3,666,946
R005-02	B07	CAGATA	6,643,474	6,601,550
R006-01	C07	GAAGTG	4,610,280	4,585,920
R006-02	D07	TAGCGGAT	3,505,900	3,487,152
R006-18	E07	TATTGCGCAT	6,373,164	6,335,722
R007-01	F07	ATAGAT	3,145,694	3,117,332
R007-16	G07	CCGAACA	4,521,014	4,482,294
R008-01	H07	GGAAGACAT	2,752,340	2,736,620
R008-02	A08	GGCTTA	3,233,092	3,217,492
R009-03	B08	AACGCACATT	6,412,916	6,362,786
R010-01	C08	CCTTGCCATT	1,693,626	1,684,408
R010-02	D08	GGTATA	1,791,864	1,781,424
R011-01	E08	TCTTGG	3,978,224	3,961,306
R011-02	F08	GGTGT	3,970,102	3,951,588
R011-16	G08	GGATA	5,040,430	5,012,104
R011-21	H08	CTAAGCA	1,966,564	1,954,404
R012-01	A09	GCGCTCA	3,886,682	3,866,828
R012-04	B09	ACTGCGAT	2,848,144	2,830,912
R013-01	C09	TTCGTT	5,465,162	5,439,180
R015-01	D09	ATATAA	4,542,820	4,496,948
R015-10	E09	GCCTACCT	4,478,694	4,455,748
R015-16	F09	AATTAG	6,630,128	6,566,454
R015-18	G09	GGAACGA	4,399,410	4,376,678
R016-12	H09	ACAAC	4,102,376	4,071,916
R016-18	A10	ACTGCT	3,995,676	3,972,186
R017-04	B11	GCGTCTCT	4,818,308	4,793,872
R017-04-dupl	H11	CTCTA	7,130,248	7,094,090
R017-07	B10	CGTGGACAGT	1,507,952	1,499,962
R017-15	C10	TGGCACAGA	2,739,558	2,724,912
R017-21	D10	GCAAGCCAT	5,393,090	5,363,492
R018-06	E10	CGCACCAATT	1,436,434	1,427,144
R018-07	F10	CTCGCGG	1,706,242	1,696,076
R019-01	C11	CCACTCA	6,415,068	6,376,006
R019-01-dupl	A12	GGTGACATT	6,167,582	6,129,858
R020-01	D11	TCACGGAAG	7,388,334	7,346,904
R020-01-dupl	B12	TCCGAG	8,137,800	8,093,526
R020-17	G10	AACTGG	4,529,740	4,501,064
R021-01	E11	TATCA	12,373,750	12,308,440
R021-01-dupl	C12	TAGATGA	8,814,210	8,764,448
R021-02	H10	ATGAGCAA	983,500	976,950
R021-03	A11	CTTGA	2,005,350	1,993,894
R021-26	B11	GCGTCTCT	2,628,226	2,615,460
R022-05	C11	CCACTCA	1,313,856	1,305,870
R023-12	D11	TCACGGAAG	2,986,080	2,970,052
R023-21	E11	TATCA	4,090,372	4,065,172
R024-01	F11	TAGCCAA	6,969,660	6,920,196
R024-01-dupl	D12	CGCAACCAAGT	4,322,358	4,289,752

3.5 Supplemental Figures and Tables

Table S3.1 Continued.

Sample Name	Plate well	Barcode	Raw total reads	Quality trimmed reads
R024-05	F11	TAGCCAA	3,275,426	3,256,034
R024-11	G11	ATATCGCCA	3,938,254	3,906,548
R024-14	H11	CTCTA	2,763,212	2,749,320
R025-01	A12	GGTGCACATT	6,303,750	6,267,976
R025-05	B12	TCCGAG	7,832,826	7,790,966
R025-17	C12	TAGATGA	4,227,190	4,201,468
R026-01	D12	CGCAACCAGT	4,085,170	4,056,296
R026-07	E12	ATCTGT	2,880,932	2,863,364
R027-04	F12	AAGACGCT	2,218,090	2,200,558
R027-09	G12	CATCTGCCG	4,023,122	3,990,618
R027-16	H12	TAGCAG	4,008,184	3,984,024
R028-05	A06	TCCGCA	5,197,146	5,167,720
R028-18	B06	TCTCA	6,667,768	6,638,430
R029-03	C06	TGCGAGA	4,818,038	4,795,222
R029-05	D06	TGCTGAA	3,747,278	3,731,268
R029-15	E06	TGGC	2,864,858	2,852,544
R031-05	F06	TTGACCAG	3,003,966	2,989,654
R032-02	G06	TTGCGTCT	3,503,030	3,486,508
R032-05	H06	TTGTAG	4,996,430	4,975,386
R032-21	A07	TGACGCCA	3,023,062	3,008,126
R033-01	B07	CAGATA	970,942	965,036
R033-02	C07	GAAGTG	2,781,190	2,767,886
R033-03	D07	TAGCGGAT	2,453,900	2,442,100
R033-05	E07	TATTCGCAT	4,783,540	4,757,182
R034-02	F07	ATAGAT	6,859,044	6,799,678
R034-13	G07	CCGAACA	3,613,044	3,593,124
R034-15	H07	GGAAGACAT	3,263,270	3,246,366
R035-07	A08	GGCTTA	3,605,172	3,589,424
R035-08	B08	AACGCACATT	4,822,426	4,788,836
R035-20	C08	CCTTGCCATT	3,153,822	3,139,008
R036-05	D08	GGTATA	5,395,720	5,367,602
R036-09	E08	TCTTGG	5,366,964	5,345,042
R037-01	F08	GGTGT	1,228,808	1,224,180
R037-04	G08	GGATA	6,761,880	6,729,078
R037-10	H08	CTAAGCA	4,631,176	4,606,176
R038-01	A09	GCGCTCA	5,015,784	4,992,818
R038-03	B09	ACTGCGAT	2,774,750	2,759,386
R038-06	C09	TTCGTT	3,304,696	3,290,554
R038-10	D09	ATATAA	6,649,994	6,589,758
R039-08	E09	GCCTACCT	3,548,680	3,533,310
R040-14	F09	AATTAG	10,652,346	10,556,900
R040-19	G09	GGAACGA	1,653,510	1,645,234
R041-03	H09	ACAACCT	2,343,670	2,324,382
R041-11	A10	ACTGCT	3,136,744	3,119,526
R042-01	B10	CGTGGACAGT	2,740,582	2,728,228
R043-01	C10	TGGCACAGA	3,065,330	3,051,438
R044-02	D10	GCAAGCCAT	3,612,130	3,593,472
R044-12	E10	CGCACCAATT	1,917,802	1,906,242
R045-02	F10	CTCGCGG	1,746,356	1,736,736
R045-06	G10	AACTGG	4,905,068	4,876,068
R045-25	H10	ATGAGCAA	4,670,156	4,642,892

3.5 Supplemental Figures and Tables

Table S3.2 Mean and standard deviation (SD) of error rates, total number of SNPs/loci, and variation explained by first two axes of PCoA for 224 RADseq assemblies generated with varying values for *clustering threshold (ct)* and *minimum samples per locus (msl)* in IPYRAD.

dataset	clustering threshold (ct)	minimum samples per locus (msl)	total number of SNPs	number of RADseq loci	mean locus error rate	SD locus error rate	mean allele error rate	SD allele error rate	mean SNP error rate	SD SNP error rate	variation (first two axes of PCoA)
ct80msl4	80	4	83088	15197	0.1732	0.0227	0.1187	0.0105	0.0153	0.0036	0.0664
ct80msl12	80	12	35645	4225	0.1747	0.0161	0.1097	0.0155	0.0084	0.0024	0.1220
ct80msl20	80	20	22427	2439	0.1946	0.0341	0.1044	0.0228	0.0069	0.0020	0.1983
ct80msl28	80	28	16508	1744	0.1954	0.0372	0.1012	0.0250	0.0061	0.0016	0.2289
ct80msl36	80	36	13052	1296	0.1918	0.0410	0.0963	0.0244	0.0046	0.0019	0.2530
ct80msl44	80	44	10458	979	0.1967	0.0430	0.0943	0.0203	0.0038	0.0020	0.2823
ct80msl52	80	52	8603	739	0.1976	0.0418	0.0924	0.0193	0.0029	0.0016	0.2708
ct80msl60	80	60	7187	595	0.1926	0.0436	0.0911	0.0214	0.0021	0.0008	0.2652
ct80msl68	80	68	5721	471	0.1884	0.0440	0.0918	0.0221	0.0022	0.0009	0.3043
ct80msl76	80	76	4593	381	0.1763	0.0421	0.0899	0.0202	0.0016	0.0008	0.3075
ct80msl84	80	84	3648	297	0.1624	0.0378	0.0909	0.0260	0.0017	0.0007	0.3139
ct80msl92	80	92	2944	235	0.1478	0.0348	0.0882	0.0218	0.0012	0.0008	0.3204
ct80msl100	80	100	1851	143	0.0800	0.0295	0.0963	0.0248	0.0013	0.0012	0.3065
ct80msl108	80	108	201	15	0.0000	0.0000	0.1778	0.0544	0.0008	0.0020	0.1691
ct81msl4	81	4	84710	15481	0.1733	0.0227	0.1205	0.0107	0.0149	0.0037	0.0656
ct81msl12	81	12	36347	4304	0.1749	0.0157	0.1125	0.0140	0.0086	0.0030	0.1208
ct81msl20	81	20	22690	2467	0.1941	0.0329	0.1047	0.0230	0.0068	0.0024	0.1974
ct81msl28	81	28	16647	1757	0.1943	0.0371	0.1006	0.0240	0.0060	0.0019	0.2278
ct81msl36	81	36	13138	1300	0.1920	0.0419	0.0978	0.0233	0.0045	0.0022	0.2522
ct81msl44	81	44	10599	991	0.1961	0.0436	0.0967	0.0205	0.0038	0.0021	0.2844
ct81msl52	81	52	8671	747	0.1976	0.0410	0.0956	0.0204	0.0030	0.0017	0.2769
ct81msl60	81	60	7210	597	0.1944	0.0412	0.0939	0.0214	0.0020	0.0008	0.2736
ct81msl68	81	68	5696	469	0.1907	0.0432	0.0952	0.0236	0.0021	0.0009	0.3058
ct81msl76	81	76	4514	375	0.1783	0.0399	0.0935	0.0210	0.0016	0.0009	0.3075
ct81msl84	81	84	3549	290	0.1639	0.0358	0.0950	0.0267	0.0016	0.0008	0.3268
ct81msl92	81	92	2881	232	0.1482	0.0317	0.0934	0.0238	0.0012	0.0008	0.3229
ct81msl100	81	100	1756	137	0.0785	0.0254	0.1015	0.0229	0.0014	0.0012	0.3050
ct81msl108	81	108	234	18	0.0000	0.0000	0.1667	0.0609	0.0007	0.0017	0.1775
ct82msl4	82	4	87671	15832	0.1743	0.0209	0.1211	0.0116	0.0158	0.0038	0.0646
ct82msl12	82	12	37369	4402	0.1720	0.0121	0.1134	0.0145	0.0088	0.0030	0.1188
ct82msl20	82	20	23415	2536	0.1893	0.0310	0.1060	0.0243	0.0069	0.0023	0.2038
ct82msl28	82	28	17305	1815	0.1933	0.0360	0.1014	0.0251	0.0061	0.0017	0.2303
ct82msl36	82	36	13689	1350	0.1922	0.0411	0.0953	0.0240	0.0046	0.0020	0.2675
ct82msl44	82	44	10983	1024	0.1957	0.0427	0.0938	0.0197	0.0038	0.0020	0.2926
ct82msl52	82	52	8932	771	0.1957	0.0407	0.0929	0.0194	0.0030	0.0016	0.2809
ct82msl60	82	60	7411	616	0.1946	0.0407	0.0913	0.0224	0.0021	0.0009	0.2689
ct82msl68	82	68	5880	486	0.1921	0.0444	0.0927	0.0233	0.0022	0.0010	0.3086
ct82msl76	82	76	4678	390	0.1814	0.0425	0.0921	0.0217	0.0017	0.0010	0.3116
ct82msl84	82	84	3701	303	0.1653	0.0369	0.0918	0.0267	0.0017	0.0010	0.3123
ct82msl92	82	92	3023	242	0.1501	0.0337	0.0905	0.0251	0.0012	0.0008	0.3172
ct82msl100	82	100	1885	147	0.0883	0.0305	0.0970	0.0211	0.0013	0.0011	0.3026
ct82msl108	82	108	225	17	0.0000	0.0000	0.1765	0.0644	0.0007	0.0018	0.1742
ct83msl4	83	4	89922	16160	0.1708	0.0213	0.1225	0.0106	0.0155	0.0041	0.0628
ct83msl12	83	12	38447	4493	0.1671	0.0118	0.1134	0.0144	0.0087	0.0029	0.1228
ct83msl20	83	20	24129	2593	0.1850	0.0289	0.1055	0.0237	0.0066	0.0021	0.2014
ct83msl28	83	28	17839	1858	0.1857	0.0322	0.1033	0.0260	0.0058	0.0015	0.2370
ct83msl36	83	36	14104	1387	0.1854	0.0344	0.0974	0.0220	0.0044	0.0019	0.2680
ct83msl44	83	44	11313	1049	0.1891	0.0363	0.0963	0.0191	0.0039	0.0020	0.2944
ct83msl52	83	52	9138	785	0.1879	0.0323	0.0942	0.0186	0.0031	0.0016	0.2814
ct83msl60	83	60	7636	629	0.1863	0.0332	0.0931	0.0210	0.0023	0.0010	0.2716
ct83msl68	83	68	6051	494	0.1830	0.0351	0.0948	0.0218	0.0024	0.0010	0.3132
ct83msl76	83	76	4815	398	0.1717	0.0315	0.0925	0.0213	0.0019	0.0013	0.3095
ct83msl84	83	84	3786	310	0.1555	0.0272	0.0910	0.0259	0.0018	0.0010	0.3108
ct83msl92	83	92	3130	251	0.1470	0.0242	0.0892	0.0249	0.0013	0.0011	0.3162
ct83msl100	83	100	1898	150	0.0854	0.0196	0.0972	0.0263	0.0012	0.0012	0.2945
ct83msl108	83	108	261	20	0.0000	0.0000	0.1500	0.0548	0.0006	0.0016	0.2225
ct84msl4	84	4	93922	16706	0.1684	0.0216	0.1221	0.0117	0.0153	0.0039	0.0609
ct84msl12	84	12	40570	4714	0.1674	0.0111	0.1126	0.0175	0.0087	0.0030	0.1156
ct84msl20	84	20	25345	2726	0.1818	0.0283	0.1043	0.0262	0.0070	0.0021	0.1948
ct84msl28	84	28	18680	1955	0.1813	0.0319	0.1013	0.0287	0.0060	0.0013	0.2443
ct84msl36	84	36	14880	1463	0.1793	0.0338	0.0944	0.0253	0.0046	0.0019	0.2722
ct84msl44	84	44	12001	1109	0.1833	0.0362	0.0931	0.0224	0.0039	0.0020	0.2961
ct84msl52	84	52	9695	830	0.1830	0.0352	0.0920	0.0207	0.0031	0.0016	0.2793
ct84msl60	84	60	8019	660	0.1798	0.0359	0.0911	0.0221	0.0024	0.0010	0.2811
ct84msl68	84	68	6418	525	0.1758	0.0385	0.0913	0.0232	0.0025	0.0010	0.3181
ct84msl76	84	76	5122	423	0.1651	0.0358	0.0894	0.0214	0.0018	0.0010	0.3092
ct84msl84	84	84	3979	327	0.1493	0.0307	0.0903	0.0270	0.0018	0.0008	0.3026
ct84msl92	84	92	3205	259	0.1421	0.0245	0.0888	0.0251	0.0013	0.0010	0.3135
ct84msl100	84	100	1879	150	0.0829	0.0219	0.0991	0.0272	0.0013	0.0011	0.2905
ct84msl108	84	108	232	19	0.0000	0.0000	0.1491	0.0517	0.0007	0.0018	0.1907
ct85msl4	85	4	98020	17263	0.1689	0.0206	0.1223	0.0132	0.0154	0.0037	0.0577
ct85msl12	85	12	42204	4888	0.1660	0.0095	0.1153	0.0167	0.0096	0.0030	0.1143
ct85msl20	85	20	26486	2842	0.1794	0.0242	0.1051	0.0245	0.0074	0.0025	0.1888

3.5 Supplemental Figures and Tables

Table S3.2 Continued.

dataset	clustering threshold (ct)	minimum samples per locus (msl)	total number of SNPs	number of RADseq loci	mean locus error rate	SD locus error rate	mean allele error rate	SD allele error rate	mean SNP error rate	SD SNP error rate	variation (first two axes of PCoA)
ct85msl28	85	28	19500	2048	0.1795	0.0271	0.0993	0.0240	0.0061	0.0018	0.2392
ct85msl36	85	36	15571	1537	0.1779	0.0305	0.0942	0.0206	0.0047	0.0021	0.2652
ct85msl44	85	44	12635	1170	0.1798	0.0303	0.0928	0.0176	0.0040	0.0023	0.2946
ct85msl52	85	52	10173	870	0.1768	0.0275	0.0895	0.0168	0.0034	0.0018	0.2703
ct85msl60	85	60	8421	691	0.1754	0.0286	0.0885	0.0173	0.0026	0.0012	0.2764
ct85msl68	85	68	6752	549	0.1722	0.0291	0.0877	0.0190	0.0025	0.0011	0.3170
ct85msl76	85	76	5443	443	0.1620	0.0269	0.0862	0.0178	0.0018	0.0012	0.3134
ct85msl84	85	84	4247	343	0.1474	0.0245	0.0871	0.0229	0.0018	0.0009	0.3137
ct85msl92	85	92	3346	267	0.1391	0.0195	0.0867	0.0209	0.0014	0.0009	0.3161
ct85msl100	85	100	1977	154	0.0809	0.0134	0.0940	0.0231	0.0012	0.0011	0.2880
ct85msl108	85	108	298	22	0.0000	0.0000	0.1515	0.0371	0.0011	0.0017	0.2363
ct86msl4	86	4	102751	17912	0.1679	0.0202	0.1216	0.0112	0.0151	0.0037	0.0593
ct86msl12	86	12	44354	5134	0.1629	0.0088	0.1159	0.0152	0.0092	0.0026	0.1143
ct86msl20	86	20	27968	2994	0.1729	0.0226	0.1056	0.0251	0.0072	0.0022	0.1882
ct86msl28	86	28	20537	2154	0.1716	0.0255	0.0999	0.0244	0.0062	0.0021	0.2367
ct86msl36	86	36	16295	1613	0.1712	0.0299	0.0954	0.0208	0.0048	0.0021	0.2571
ct86msl44	86	44	13178	1223	0.1735	0.0312	0.0939	0.0182	0.0040	0.0021	0.2855
ct86msl52	86	52	10672	912	0.1721	0.0287	0.0914	0.0160	0.0034	0.0017	0.2702
ct86msl60	86	60	8798	722	0.1683	0.0286	0.0904	0.0168	0.0027	0.0010	0.2881
ct86msl68	86	68	7004	570	0.1653	0.0329	0.0902	0.0176	0.0025	0.0009	0.3214
ct86msl76	86	76	5664	463	0.1549	0.0307	0.0869	0.0184	0.0017	0.0009	0.3201
ct86msl84	86	84	4412	357	0.1432	0.0283	0.0872	0.0231	0.0018	0.0009	0.3137
ct86msl92	86	92	3474	277	0.1351	0.0222	0.0843	0.0188	0.0014	0.0010	0.3122
ct86msl100	86	100	2073	161	0.0807	0.0186	0.0926	0.0196	0.0012	0.0011	0.2789
ct86msl108	86	108	298	22	0.0000	0.0000	0.1515	0.0371	0.0011	0.0017	0.2358
ct87msl4	87	4	108145	18692	0.1681	0.0210	0.1202	0.0106	0.0139	0.0030	0.0649
ct87msl12	87	12	47017	5423	0.1608	0.0112	0.1132	0.0118	0.0087	0.0024	0.1248
ct87msl20	87	20	29774	3188	0.1720	0.0221	0.1035	0.0199	0.0069	0.0019	0.2030
ct87msl28	87	28	22004	2287	0.1715	0.0267	0.0994	0.0195	0.0061	0.0018	0.2422
ct87msl36	87	36	17500	1721	0.1717	0.0305	0.0954	0.0169	0.0049	0.0021	0.2574
ct87msl44	87	44	14057	1300	0.1730	0.0314	0.0924	0.0157	0.0044	0.0022	0.2910
ct87msl52	87	52	11380	970	0.1718	0.0283	0.0898	0.0135	0.0036	0.0019	0.2703
ct87msl60	87	60	9273	759	0.1669	0.0279	0.0900	0.0130	0.0028	0.0012	0.2904
ct87msl68	87	68	7392	602	0.1657	0.0274	0.0897	0.0153	0.0026	0.0011	0.3113
ct87msl76	87	76	5867	481	0.1537	0.0305	0.0875	0.0138	0.0018	0.0009	0.3110
ct87msl84	87	84	4627	374	0.1424	0.0280	0.0850	0.0195	0.0018	0.0008	0.3119
ct87msl92	87	92	3662	291	0.1328	0.0221	0.0825	0.0145	0.0013	0.0008	0.3112
ct87msl100	87	100	2145	164	0.0836	0.0224	0.0922	0.0161	0.0012	0.0009	0.2796
ct87msl108	87	108	294	21	0.0000	0.0000	0.1587	0.0577	0.0011	0.0018	0.2253
ct88msl4	88	4	113973	19514	0.1666	0.0192	0.1187	0.0093	0.0152	0.0038	0.0627
ct88msl12	88	12	49761	5712	0.1572	0.0118	0.1108	0.0141	0.0094	0.0026	0.1206
ct88msl20	88	20	31548	3369	0.1691	0.0253	0.1042	0.0219	0.0075	0.0023	0.1913
ct88msl28	88	28	23334	2424	0.1691	0.0298	0.1001	0.0221	0.0070	0.0024	0.2354
ct88msl36	88	36	18475	1805	0.1696	0.0336	0.0950	0.0204	0.0055	0.0025	0.2713
ct88msl44	88	44	14868	1364	0.1696	0.0341	0.0934	0.0202	0.0048	0.0025	0.3057
ct88msl52	88	52	11942	1009	0.1704	0.0316	0.0904	0.0188	0.0041	0.0022	0.2828
ct88msl60	88	60	9573	777	0.1658	0.0303	0.0905	0.0181	0.0033	0.0016	0.3002
ct88msl68	88	68	7550	612	0.1675	0.0309	0.0892	0.0186	0.0028	0.0012	0.3088
ct88msl76	88	76	6068	496	0.1571	0.0298	0.0872	0.0141	0.0020	0.0009	0.3134
ct88msl84	88	84	4782	389	0.1470	0.0276	0.0850	0.0181	0.0019	0.0008	0.3231
ct88msl92	88	92	3726	296	0.1362	0.0242	0.0810	0.0106	0.0014	0.0008	0.3205
ct88msl100	88	100	2115	163	0.0907	0.0199	0.0903	0.0125	0.0013	0.0009	0.2839
ct88msl108	88	108	306	21	0.0000	0.0000	0.1667	0.0398	0.0016	0.0027	0.2208
ct89msl4	89	4	121931	20517	0.1668	0.0184	0.1179	0.0118	0.0154	0.0037	0.0653
ct89msl12	89	12	53218	6025	0.1553	0.0084	0.1112	0.0146	0.0095	0.0026	0.1279
ct89msl20	89	20	33964	3582	0.1669	0.0208	0.1035	0.0204	0.0081	0.0021	0.1878
ct89msl28	89	28	25291	2597	0.1665	0.0236	0.1017	0.0192	0.0076	0.0021	0.2280
ct89msl36	89	36	19855	1929	0.1658	0.0271	0.0965	0.0188	0.0060	0.0019	0.2590
ct89msl44	89	44	15799	1446	0.1662	0.0277	0.0935	0.0181	0.0053	0.0018	0.3014
ct89msl52	89	52	12655	1064	0.1680	0.0279	0.0918	0.0142	0.0045	0.0014	0.2990
ct89msl60	89	60	10112	819	0.1644	0.0246	0.0907	0.0133	0.0039	0.0013	0.2952
ct89msl68	89	68	7965	643	0.1680	0.0239	0.0896	0.0123	0.0035	0.0014	0.3052
ct89msl76	89	76	6430	521	0.1574	0.0253	0.0871	0.0109	0.0029	0.0017	0.3029
ct89msl84	89	84	4968	403	0.1504	0.0240	0.0864	0.0174	0.0027	0.0015	0.3135
ct89msl92	89	92	3826	303	0.1407	0.0229	0.0820	0.0133	0.0024	0.0021	0.3183
ct89msl100	89	100	2141	165	0.0888	0.0271	0.0866	0.0165	0.0013	0.0009	0.2823
ct89msl108	89	108	289	19	0.0000	0.0000	0.1404	0.0430	0.0017	0.0029	0.1909
ct90msl4	90	4	129895	21698	0.1667	0.0184	0.1183	0.0090	0.0150	0.0033	0.0677
ct90msl12	90	12	57010	6416	0.1531	0.0111	0.1097	0.0143	0.0091	0.0029	0.1256
ct90msl20	90	20	36233	3814	0.1639	0.0210	0.1028	0.0187	0.0076	0.0024	0.1921
ct90msl28	90	28	27309	2773	0.1632	0.0275	0.1012	0.0169	0.0069	0.0025	0.2451
ct90msl36	90	36	21458	2072	0.1623	0.0318	0.0972	0.0161	0.0057	0.0023	0.2657
ct90msl44	90	44	17049	1546	0.1625	0.0319	0.0948	0.0159	0.0049	0.0021	0.2982
ct90msl52	90	52	13523	1134	0.1630	0.0332	0.0917	0.0136	0.0037	0.0016	0.3114

3.5 Supplemental Figures and Tables

Table S3.2 Continued.

dataset	clustering threshold (ct)	minimum samples per locus (msl)	total number of SNPs	number of RADseq loci	mean locus error rate	SD locus error rate	mean allele error rate	SD allele error rate	mean SNP error rate	SD SNP error rate	variation (first two axes of PCoA)
ct90msl60	90	60	10943	881	0.1588	0.0308	0.0886	0.0145	0.0031	0.0010	0.3078
ct90msl68	90	68	8444	679	0.1582	0.0297	0.0886	0.0125	0.0027	0.0005	0.3145
ct90msl76	90	76	6631	541	0.1480	0.0280	0.0847	0.0095	0.0020	0.0005	0.3097
ct90msl84	90	84	5080	414	0.1362	0.0239	0.0872	0.0142	0.0018	0.0008	0.3177
ct90msl92	90	92	3933	314	0.1275	0.0174	0.0836	0.0112	0.0017	0.0006	0.3127
ct90msl100	90	100	2165	170	0.0854	0.0209	0.0840	0.0119	0.0012	0.0009	0.2942
ct90msl108	90	108	299	20	0.0000	0.0000	0.1500	0.0447	0.0017	0.0028	0.1883
ct91msl4	91	4	139467	23154	0.1714	0.0210	0.1146	0.0076	0.0146	0.0033	0.0677
ct91msl12	91	12	61317	6858	0.1581	0.0087	0.1086	0.0128	0.0091	0.0026	0.1359
ct91msl20	91	20	38849	4058	0.1650	0.0191	0.1024	0.0192	0.0077	0.0019	0.1949
ct91msl28	91	28	29053	2943	0.1639	0.0236	0.0983	0.0201	0.0069	0.0020	0.2502
ct91msl36	91	36	22677	2182	0.1620	0.0277	0.0941	0.0186	0.0055	0.0021	0.2672
ct91msl44	91	44	17994	1619	0.1647	0.0332	0.0921	0.0197	0.0046	0.0020	0.3126
ct91msl52	91	52	14417	1200	0.1641	0.0344	0.0904	0.0157	0.0038	0.0015	0.3069
ct91msl60	91	60	11468	919	0.1591	0.0286	0.0881	0.0147	0.0030	0.0011	0.3171
ct91msl68	91	68	8827	706	0.1597	0.0280	0.0872	0.0127	0.0025	0.0007	0.3122
ct91msl76	91	76	7034	568	0.1548	0.0266	0.0847	0.0109	0.0019	0.0006	0.3213
ct91msl84	91	84	5293	431	0.1408	0.0241	0.0820	0.0168	0.0017	0.0008	0.3171
ct91msl92	91	92	4106	327	0.1336	0.0195	0.0794	0.0127	0.0016	0.0005	0.3005
ct91msl100	91	100	2158	169	0.0835	0.0179	0.0795	0.0209	0.0010	0.0006	0.2713
ct91msl108	91	108	306	20	0.0000	0.0000	0.1250	0.0524	0.0011	0.0027	0.2133
ct92msl4	92	4	156028	25069	0.2411	0.0743	0.1173	0.0086	0.0178	0.0029	0.0701
ct92msl12	92	12	66952	7404	0.2288	0.0784	0.1060	0.0117	0.0091	0.0029	0.1322
ct92msl20	92	20	42550	4393	0.2370	0.0865	0.1003	0.0188	0.0079	0.0023	0.1919
ct92msl28	92	28	31210	3130	0.2317	0.0938	0.0977	0.0198	0.0069	0.0022	0.2585
ct92msl36	92	36	23962	2285	0.2255	0.0965	0.0947	0.0194	0.0053	0.0020	0.2681
ct92msl44	92	44	18869	1661	0.2234	0.0992	0.0916	0.0190	0.0043	0.0018	0.2917
ct92msl52	92	52	14753	1211	0.2231	0.1017	0.0903	0.0148	0.0038	0.0015	0.2906
ct92msl60	92	60	11588	919	0.2177	0.1038	0.0885	0.0151	0.0030	0.0011	0.2983
ct92msl68	92	68	8883	698	0.2165	0.1083	0.0863	0.0110	0.0027	0.0006	0.2850
ct92msl76	92	76	6918	554	0.2073	0.1062	0.0840	0.0111	0.0022	0.0006	0.2854
ct92msl84	92	84	5233	420	0.2043	0.1107	0.0802	0.0162	0.0020	0.0011	0.3006
ct92msl92	92	92	3803	300	0.1948	0.1047	0.0817	0.0186	0.0014	0.0008	0.3110
ct92msl100	92	100	2092	162	0.1496	0.1142	0.0751	0.0193	0.0005	0.0004	0.2748
ct92msl108	92	108	183	13	0.0000	0.0000	0.1154	0.0807	0.0000	0.0000	0.1265
ct93msl4	93	4	171030	27712	0.1972	0.0329	0.1166	0.0088	0.0168	0.0028	0.0730
ct93msl12	93	12	74202	8179	0.1829	0.0139	0.1045	0.0118	0.0083	0.0024	0.1526
ct93msl20	93	20	47228	4834	0.1882	0.0175	0.0996	0.0145	0.0069	0.0016	0.2047
ct93msl28	93	28	34628	3442	0.1821	0.0260	0.0970	0.0135	0.0060	0.0017	0.2582
ct93msl36	93	36	26624	2521	0.1763	0.0304	0.0934	0.0118	0.0049	0.0017	0.2741
ct93msl44	93	44	21146	1842	0.1706	0.0323	0.0937	0.0105	0.0041	0.0016	0.2911
ct93msl52	93	52	16627	1350	0.1715	0.0325	0.0943	0.0108	0.0037	0.0016	0.3081
ct93msl60	93	60	12929	1008	0.1630	0.0302	0.0923	0.0111	0.0030	0.0010	0.3120
ct93msl68	93	68	9785	757	0.1562	0.0309	0.0921	0.0082	0.0027	0.0007	0.3185
ct93msl76	93	76	7716	603	0.1503	0.0287	0.0899	0.0048	0.0025	0.0007	0.3077
ct93msl84	93	84	5924	461	0.1438	0.0286	0.0871	0.0097	0.0019	0.0007	0.3073
ct93msl92	93	92	4302	333	0.1351	0.0248	0.0879	0.0072	0.0016	0.0006	0.3015
ct93msl100	93	100	2315	178	0.0837	0.0248	0.0913	0.0071	0.0006	0.0005	0.2663
ct93msl108	93	108	311	21	0.0000	0.0000	0.1111	0.0492	0.0000	0.0000	0.2145
ct94msl4	94	4	181037	30484	0.1992	0.0353	0.1147	0.0088	0.0159	0.0025	0.0788
ct94msl12	94	12	79447	8922	0.1867	0.0159	0.1026	0.0117	0.0081	0.0022	0.1630
ct94msl20	94	20	50435	5267	0.1895	0.0156	0.0973	0.0149	0.0071	0.0016	0.2313
ct94msl28	94	28	36816	3717	0.1796	0.0259	0.0937	0.0158	0.0061	0.0012	0.2747
ct94msl36	94	36	28527	2709	0.1721	0.0279	0.0922	0.0155	0.0053	0.0012	0.2972
ct94msl44	94	44	22328	1958	0.1679	0.0311	0.0895	0.0138	0.0044	0.0009	0.3089
ct94msl52	94	52	17310	1415	0.1652	0.0307	0.0907	0.0126	0.0039	0.0009	0.3307
ct94msl60	94	60	13412	1049	0.1582	0.0330	0.0892	0.0104	0.0028	0.0011	0.3195
ct94msl68	94	68	10094	788	0.1535	0.0302	0.0860	0.0083	0.0025	0.0009	0.3176
ct94msl76	94	76	7941	624	0.1462	0.0288	0.0852	0.0075	0.0023	0.0009	0.3200
ct94msl84	94	84	6080	474	0.1381	0.0276	0.0842	0.0086	0.0020	0.0012	0.3153
ct94msl92	94	92	4496	347	0.1270	0.0221	0.0840	0.0116	0.0019	0.0014	0.2994
ct94msl100	94	100	2428	185	0.0844	0.0188	0.0826	0.0180	0.0013	0.0020	0.2532
ct94msl108	94	108	302	21	0.0000	0.0000	0.1032	0.0557	0.0006	0.0014	0.2076
ct95msl4	95	4	195466	34557	0.2040	0.0354	0.1141	0.0097	0.0138	0.0026	0.0769
ct95msl12	95	12	87708	10085	0.1980	0.0176	0.1040	0.0137	0.0081	0.0021	0.1486
ct95msl20	95	20	55311	5877	0.1963	0.0199	0.1025	0.0155	0.0070	0.0021	0.2017
ct95msl28	95	28	40230	4106	0.1843	0.0254	0.0985	0.0147	0.0055	0.0014	0.2500
ct95msl36	95	36	30941	2983	0.1753	0.0290	0.0959	0.0145	0.0045	0.0015	0.2733
ct95msl44	95	44	24179	2138	0.1679	0.0341	0.0932	0.0129	0.0038	0.0015	0.3061
ct95msl52	95	52	18446	1527	0.1683	0.0301	0.0935	0.0135	0.0032	0.0014	0.3132
ct95msl60	95	60	14046	1121	0.1616	0.0335	0.0861	0.0111	0.0026	0.0009	0.3198
ct95msl68	95	68	10719	837	0.1585	0.0332	0.0832	0.0100	0.0023	0.0006	0.3252
ct95msl76	95	76	8486	668	0.1523	0.0302	0.0797	0.0097	0.0020	0.0003	0.3259
ct95msl84	95	84	6132	487	0.1383	0.0295	0.0807	0.0133	0.0018	0.0005	0.3297
ct95msl92	95	92	4598	361	0.1255	0.0248	0.0822	0.0102	0.0016	0.0005	0.3217
ct95msl100	95	100	2452	191	0.0898	0.0194	0.0869	0.0069	0.0009	0.0006	0.2684
ct95msl108	95	108	340	24	0.0000	0.0000	0.1250	0.0645	0.0025	0.0022	0.2133

3.5 Supplemental Figures and Tables

Table S3.3 Optimal cluster numbers of ten FASTSTRUCTURE runs for each dataset (*ct85msl100*, *ct85msl68*, *ct85msl12*) using the ‘chooseK’ metrics ‘model complexity that maximizes marginal likelihood’ and ‘model components used to explain structure in data’.

run	<i>ct85msl100</i>		<i>ct85msl68</i>		<i>ct85msl12</i>	
	model complexity that maximizes marginal likelihood	model components used to explain structure in data	model complexity that maximizes marginal likelihood	model components used to explain structure in data	model complexity that maximizes marginal likelihood	model components used to explain structure in data
1	2	3	2	3	3	5
2	2	3	2	3	3	4
3	2	3	2	3	3	4
4	2	3	2	3	3	5
5	2	3	2	3	3	4
6	2	3	2	3	3	4
7	2	3	2	3	3	4
8	2	3	2	3	3	5
9	2	3	2	3	3	4
10	2	3	2	3	3	5

Table S3.4 Logarithmic marginal likelihood values for ten species delimitation models (S01-S10) and ten runs calculated with the BEAST package SNAPP based on 3,837 SNPs of RADseq dataset *ct85msl12*. The number of taxa (n), as well as mean and standard deviation (SD) of marginal likelihood values are listed for each species delimitation model.

	n	run01	run02	run03	run04	run05	run06	run07	run08	run09	run10	mean	SD
S01	2	-13281.6	-12861.8	-13230.7	-13946.4	-13636.4	-13114.0	-13137.0	-13593.5	-13622.6	-13110.1	-13353.4	314.2
S02	3	-8022.1	-7732.8	-7953.3	-8401.8	-8177.0	-7929.1	-7871.1	-8132.0	-8306.7	-7726.8	-8025.3	216.4
S03	4	-6492.7	-6163.9	-6482.1	-6934.2	-6512.2	-6443.4	-6308.0	-6664.7	-6886.0	-6252.1	-6513.9	240.4
S04	4	-6608.3	-5881.6	-6168.9	-6993.3	-6623.1	-6260.0	-6299.5	-6634.9	-6985.5	-6027.1	-6448.2	362.0
S05	4	-5810.8	-5679.2	-5747.7	-6218.1	-5976.4	-5917.2	-5779.0	-5977.0	-5994.7	-5651.3	-5875.1	165.0
S06	4	-5343.8	-5185.1	-5316.2	-5566.8	-5399.0	-5373.4	-5194.5	-5342.0	-5562.5	-5197.0	-5348.0	130.8
S07	5	-5093.2	-4895.3	-5006.0	-5435.0	-5085.9	-5107.0	-4970.6	-5210.9	-5302.9	-5007.2	-5111.4	155.2
S08	5	-4623.4	-4429.7	-4610.9	-4953.9	-4636.0	-4709.6	-4457.5	-4636.7	-4932.1	-4537.2	-4652.7	166.2
S09	5	-4345.2	-4238.2	-4262.6	-4595.5	-4383.0	-4404.5	-4290.2	-4370.1	-4490.2	-4243.8	-4362.3	108.6
S10	6	-3947.5	-3768.2	-3820.6	-4229.9	-3932.1	-3984.0	-3814.3	-3936.1	-4154.8	-3903.9	-3949.1	138.7

3.5 Supplemental Figures and Tables

Table S3.5 Logarithmic marginal likelihood values for ten species delimitation models (S01-S10) and ten runs calculated with the BEAST package SNAPP based on 537 SNPs of RADseq dataset *ct85ms/68*. The number of taxa (n), as well as mean and standard deviation (SD) of marginal likelihood values are listed for each species delimitation model.

	n	run01	run02	run03	run04	run05	run06	run07	run08	run09	run10	mean	SD
S01	2	-4551.5	-4616.2	-4570.6	-4710.5	-4607.1	-4613.4	-4528.7	-4547.3	-4730.1	-4659.8	-4613.5	65.2
S02	3	-3898.1	-3950.0	-3931.1	-4055.3	-3946.7	-3983.2	-3869.7	-3888.5	-4089.6	-4010.4	-3962.3	68.6
S03	4	-3757.5	-3753.2	-3769.4	-3900.5	-3770.0	-3815.9	-3685.2	-3696.1	-3949.5	-3849.9	-3794.7	80.4
S04	4	-3719.7	-3681.3	-3703.9	-3887.3	-3747.1	-3754.2	-3584.2	-3662.9	-3950.2	-3768.0	-3745.9	101.0
S05	4	-3601.7	-3659.7	-3640.0	-3765.8	-3656.2	-3695.3	-3595.9	-3612.1	-3777.0	-3708.7	-3671.2	61.2
S06	4	-3517.3	-3566.7	-3539.4	-3674.2	-3600.4	-3599.1	-3497.1	-3528.9	-3693.3	-3640.5	-3585.7	64.1
S07	5	-3465.6	-3467.8	-3481.6	-3617.5	-3486.9	-3531.3	-3420.6	-3421.8	-3647.6	-3556.5	-3509.7	73.5
S08	5	-3393.5	-3393.4	-3391.7	-3536.8	-3438.3	-3448.3	-3340.7	-3352.4	-3575.5	-3495.9	-3436.6	74.2
S09	5	-3232.9	-3295.4	-3261.7	-3396.9	-3319.2	-3326.1	-3236.8	-3263.7	-3402.7	-3352.6	-3308.8	58.5
S10	6	-3113.0	-3125.5	-3118.5	-3264.5	-3164.9	-3180.9	-3086.9	-3092.7	-3293.8	-3216.8	-3165.8	68.7

Table S3.6 Logarithmic marginal likelihood values for ten species delimitation models (S01-S10) and ten runs calculated with the BEAST package SNAPP based on 153 SNPs of RADseq dataset *ct85ms/100*. The number of taxa (n), as well as mean and standard deviation (SD) of marginal likelihood values are listed for each species delimitation model.

	n	run01	run02	run03	run04	run05	run06	run07	run08	run09	run10	mean	SD
S01	2	-1563.2	-1492.2	-1525.9	-1543.8	-1490.2	-1551.7	-1500.1	-1492.2	-1567.2	-1551.5	-1527.8	29.9
S02	3	-1348.0	-1272.6	-1308.2	-1327.7	-1272.8	-1332.3	-1278.6	-1272.5	-1346.9	-1328.7	-1308.8	30.2
S03	4	-1327.3	-1239.5	-1281.7	-1308.1	-1246.1	-1301.7	-1249.3	-1247.5	-1325.6	-1305.8	-1283.3	33.1
S04	4	-1331.5	-1232.6	-1263.4	-1311.5	-1245.2	-1284.7	-1240.4	-1241.4	-1329.1	-1289.5	-1276.9	35.9
S05	4	-1252.2	-1200.5	-1214.9	-1257.3	-1195.9	-1250.2	-1200.0	-1194.0	-1260.2	-1244.9	-1227.0	26.7
S06	4	-1257.6	-1181.7	-1217.5	-1237.0	-1182.7	-1241.8	-1188.2	-1181.7	-1256.8	-1238.2	-1218.3	30.3
S07	5	-1231.4	-1167.3	-1188.2	-1237.8	-1168.8	-1219.8	-1170.5	-1168.9	-1239.0	-1221.8	-1201.4	29.7
S08	5	-1236.7	-1148.4	-1190.6	-1217.4	-1155.4	-1210.8	-1158.3	-1156.4	-1235.4	-1214.9	-1192.4	33.2
S09	5	-1161.8	-1109.5	-1124.3	-1166.8	-1105.5	-1159.9	-1109.5	-1103.0	-1169.9	-1154.1	-1136.4	26.9
S10	6	-1140.8	-1076.1	-1097.2	-1147.1	-1078.4	-1129.1	-1079.6	-1078.1	-1148.8	-1131.0	-1110.6	29.8

Chapter 4: At the crossroads towards polyploidy

‘At the crossroads towards polyploidy’: Genomic divergence and extent of homoploid hybridization are drivers for the formation of the ox-eye daisy polyploid complex (*Leucanthemum*, Compositae-Anthemideae)

Florian Wagner, Tankred Ott, Claudia Zimmer, Verena Reichhart, Robert Vogt, Christoph Oberprieler

New Phytologist (2019) doi: 10.1111/nph.15784

Abstract

- Polyploidy plays a paramount role in phytodiversity, but the causes for this evolutionary pathway require further study. Here, we use phylogenetic methods to examine possible polyploidy-promoting factors by comparing diploid representatives of the comprehensive European polyploid complex *Leucanthemum* with members of its strictly diploid North African counterpart *Rhodanthemum*.
- We investigate genetic divergence and gene flow among all diploid lineages of both genera to evaluate the role of genomic differentiation and hybridization for polyploid speciation. To test whether hybridization in *Leucanthemum* has been triggered by the geological conditions during its diversification, we additionally generate a time-calibrated phylogeny of 46 species of the subtribe Leucanthemeinae.
- *Leucanthemum* shows a significantly higher genetic divergence and hybridization signal among diploid lineages compared to *Rhodanthemum*, in spite of a similar crown age and diversification pattern during the Quaternary.
- Our study demonstrates the importance of genetic differentiation among diploid progenitors and their concurrent affinity for natural hybridization for the formation of a polyploid complex. Furthermore, the role of climate induced range overlaps on hybridization and polyploid speciation during the Quaternary is discussed.

Keywords: Darlington’s rule, genetic divergence, homoploid hybridization, molecular dating, multi-species coalescent, polyploidy, Quaternary.

4.1 Introduction

Despite some controversy concerning the evolutionary significance of polyploidization for the longer-term diversity of higher plants (Fawcett and Van de Peer, 2010; Arrigo and Barker, 2012; Mayrose et al., 2011; but see Soltis et al., 2014 for an opposite view), polyploid speciation plays a paramount role in phytodiversity on smaller time-scales. Wood et al. (2009) showed that about 35% of vascular plant species are recent polyploids. These ‘neopolyploids’ have been formed since their genus arose in contrast to ‘paleopolyploids’, which have often lost their polyploid status during their long course of evolution [‘diploidization’, e.g. Soltis et al. (2015)]. As a consequence, neopolyploid species and species-rich polyploid complexes are undeniable building blocks of the actual plant biodiversity and are therefore highly important drivers of ecological processes at the community and global levels. Although enormous progress has been made in recent years concerning the consequences of polyploidy (see reviews of Otto, 2007; Parisod et al., 2010; Parisod, 2012; and Weiss-Schneeweiss et al., 2013 among many others), the causes of polyploidy are less studied (Soltis et al., 2010). Therefore, we are still unable to pinpoint and evaluate the importance of individual genetic, organismal, and environmental factors that promote polyploidization.

In order to give rise to a polyploid lineage, polyploid individuals (a) need to be formed in an auto- or allopolyploid manner, (b) need to establish themselves besides (and often in competition with) their diploid progenitor populations, and (c) need to persist as an independent lineage for an evolutionary significant period of time. In two review articles, Ramsey and Schemske (1998, 2002) have summarized present knowledge about factors influencing the formation of new cytotypes and the processes that govern the establishment of new polyploid populations. Studies have shown that the formation of polyploids by somatic chromosome doubling is very rare (Nasrallah et al., 2000; Grant, 2002) and that the formation and fusion of unreduced gametes is by far the more common pathway (de Wet, 1980; Bretagnolle and Thompson, 1995; Ramsey and Schemske, 1998). The understanding of the genetic and molecular mechanisms involved, however, is presently still in its infancy (i.e., only known in *Arabidopsis* to some extent; see Brownfield and Köhler, 2011). Factors that contribute to the establishment and long-term success of a polyploid population were rarely identified and hypotheses were rarely tested (Thompson and Lumaret, 1992; Ramsey and Schemske, 2002). Most of the suggested mechanisms are based on alleged relationships between polyploidy and various measures of ecological success (Stebbins, 1947, 1950; Ehrendorfer, 1980; Lewis, 1980), mainly connected with habitat differentiation among cytotypes and varying fitness values of cytotypes in different environments. A number of individual case studies [e.g., *Fragaria* spp. (Hancock and Bringhurst, 1981); *Dactylis*

glomerata (Lumaret, 1984; Lumaret et al., 1987); *Anthoxanthum odoratum* (Felber, 1988); *Antennaria* spp. (Bayer et al., 1991); *Heuchera grossulariifolia*, (Oswald and Nuismer, 2010); *Achillea borealis* (Ramsey et al., 2008; Ramsey, 2011)] has led to the generalised notion that a greater variability in polyploids for morphological, demographic, and ecological traits relative to their diploid progenitors may form the prerequisite for habitat differentiation and – as a consequence – avoidance of the ‘minority cytotype exclusion’ disadvantage (Levin, 1975) of newly formed polyploids.

While some very specific hypotheses on the causes of polyploid formation, establishment, and long-term-success may only be addressed by painstaking, long-lasting, microevolutionary ‘magnifying glass’ experiments leading to very plant-group specific, erratic results, some more general hypotheses are accessible for testing through a macroevolutionary, phylogeny-based, retrospective ‘spyglass’ approach (Via, 2009). Closely related genera with different evolutionary pathways in respect of polyploidy – ideally sister-groups exclusively consisting of diploid species on the one hand and diploids as well as polyploids on the other – may therefore be helpful to reach more comprehensive conclusions regarding the causes of polyploidy in the angiosperm branch of life. Hypotheses on polyploidy-promoting factors which are accessible through phylogeny-based approaches may address temporal, spatiotemporal, biogeographical, eco-climatological, and/or genomic correlates. Biogeographical and eco-climatological reconstructions based on dated phylogenies may be used to detect possible correlations between the occurrence of polyploid speciation and higher latitudinal or elevational distributions and/or climatic oscillations during the Pleistocene or Holocene. Finally, phylogenetic distances (as a proxy for evolutionary divergence) between progenitors of (allo)polyploid species are suitable for testing hypotheses based on assumptions proposed by Winge (1917), Darlington (1937), or Grant (1981) that one of the prerequisites for (allo)polyploid species formation is a ‘Goldilocks’ condition of divergence among the diploid genomes involved, being not too different to still allow natural hybridization of ancestral diploids, but being sufficiently diverged to avoid reduced fertility of polyploids caused by meiotic multivalent formation (‘Darlington’s rule’; Buggs et al., 2011).

In the present study we selected diploid representatives of the two closely-related genera *Leucanthemum* Mill. and *Rhodanthemum* B.H.Wilcox & al. for investigating causes of polyploidy in a phylogeny-based framework. Together with some unispecific or extremely small genera comprising only 2-4 species, the two genera *Leucanthemum* and *Rhodanthemum* form the closely-knit subtribe Leucantheminae K.Bremer & Humphries in tribe Anthemideae of the sunflower family (Oberprieler et al., 2006, 2009). Despite their sister-group relationship (Oberprieler, 2005), both genera demonstrate a clearly contrasting pattern in their evolutionary trajectories: while the European genus *Leucanthemum* has built

up a comprehensive polyploid complex with 25+ polyploid taxa (Euro+Med, 2018) ranging from tetraploid ($4x$) to docosaploid ($22x$) chromosome numbers (Vogt, 1991), its N African counterpart *Rhodanthemum* strictly evolved on the diploid level (Wilcox and Harcourt, 1982; Vogt and Oberprieler, 2008, 2012). Therefore, the two genera represent an attractive system for studying causes of polyploidy in a comparative phylogenetic manner. For this purpose our present study addresses the following three questions: (1) Is there a difference between *Leucanthemum* and *Rhodanthemum* with respect to the genetic distances among their diploid lineages and does phylogenetic divergence drive polyploidization in the former genus in accordance with ‘Darlington’s rule’? (2) Are there indications of gene flow among diploid lineages of *Leucanthemum* and *Rhodanthemum*, respectively, and do both genera show differences in hybridization patterns, the latter being the prerequisite for allopolyploid species formation? (3) Into which geological time scale falls the diversification of each genus and was their evolution shaped by climatic oscillations during the Pleistocene and Holocene?

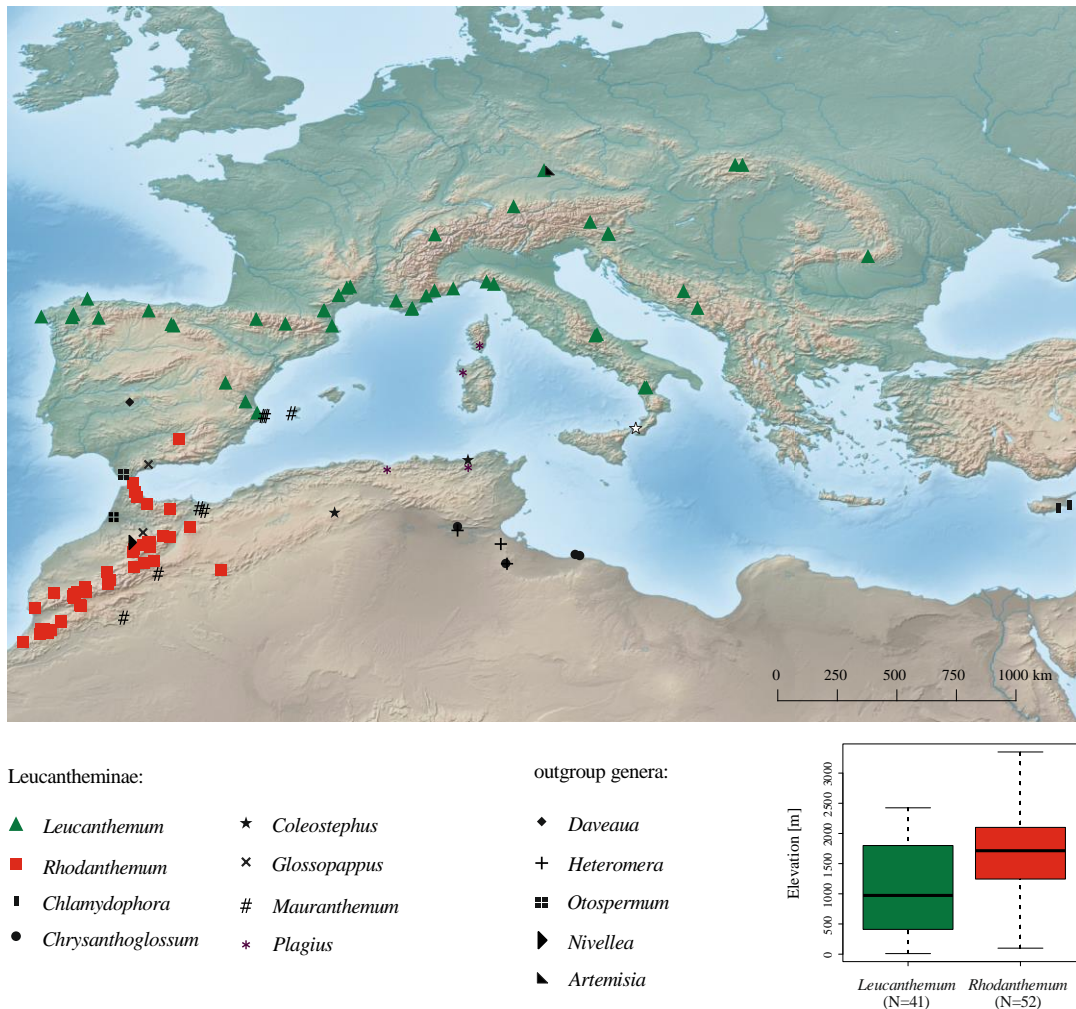


Figure 4.1 Map showing the locations of all examined *Leucanthemum* (green triangles) and *Rhodanthemum* (red squares) accessions together with the investigated samples from the remaining genera of the Leucanthemeinae and three outgroup genera. Elevations of investigated *Leucanthemum* and *Rhodanthemum* individuals are graphically depicted in the boxplots on the bottom right. Boxplots follow the standard convention, with solid lines reflecting the median, hinges the first and third quartiles, and whiskers the first and third quartiles plus $1.5 \times$ the interquartile range.

4.2 Materials and Methods

4.2.1 Plant material and DNA extraction

The majority of plant material was sampled from herbarium specimens deposited in the herbaria in Berlin (B), León (LEB), Madrid (MA), Munich (MSB), Reading (RNG), and Salamanca (SALA). Additionally, silica-dried material was obtained during excursions to Morocco in 2017 and 2018 (see Supporting Information Table S4.1). The final sampling comprised at least two accessions for almost all diploid species and subspecies of the subtribe Leucanthemeinae, with an especially dense sampling for the focal genera *Leucanthemum* and *Rhodanthemum* (Figure 4.1). As outgroup, we selected members of three genera (*Daveaua* Mariz, *Heteromera* Pomel, and *Otospermum* Willk.) with a close relationship to subtribe Leucanthemeinae according to Oberprieler et al. (2007) together with three phylogenetically more distant representatives of tribe Anthemideae (*Nivellea* B.H. Wilcox & al., *Artemisia* L., and *Ursinia* Gaertn.). For all accessions total genomic DNA was extracted using the CTAB DNA extraction protocol described by Doyle and Doyle (1987) and Doyle and Dickson (1987).

4.2.2 Plastid and nuclear marker sequencing

We sequenced a set of eight nuclear markers (*A39*, *B12 B20*, *C12*, *C20*, *D18*, *D23*, and *D27*), which were characterized as putative single-copy regions for the sunflower family (Compositae) by Chapman et al. (2007). Additionally, we sequenced the nuclear ribosomal transcribed spacer region (nrDNA ITS) and five intergenic spacer regions from the plastid genome (*petN-psbM*, *psbA-trnH*, *trnC-petN*, *trnL-trnF* and *trnQ-rps16*).

Nuclear single-copy markers were sequenced via Roche 454 sequencing for the majority of the investigated accessions after library preparation following Konowalik et al. (2015) with some minor modifications. For the two PCR steps, described in detail in the mentioned study, we used the proofreading PCRBIO HiFi Polymerase (Nippon Genetics, Düren, Germany) and all amplicons were purified with AmpliClean™ Magnetic Bead-based PCR Cleanup (NimaGen, Nijmegen, Netherlands). All barcoded and multiplexed amplicons were finally outsourced for Roche 454 sequencing to a contract sequencing company (microBIOMix GmbH, Regensburg, Germany).

In rare cases, where the 454-sequencing procedure failed to produce suitable reads and for all nrDNA ITS and plastid regions, we used *Taq* RED Polymerase (Ampliqon A/S, Odense, Denmark) and primers specified in Supporting Information Table S4.2 to produce PCR amplicons for direct sequencing. When nuclear sequences were unreadable due to an overlap of different alleles, we cloned the corresponding amplicons into a pJet cloning vector

(Fermentas/Thermo Fisher Scientific Inc., Waltham, MA, USA.). All cloning vectors were transformed into NEB Turbo bacteria (New England Biolabs Inc., Ipswich, MA, USA.) and eight clones per accession were picked for colony PCR, to ensure a 0.95 probability of obtaining the two alleles expected for a diploid species (Joly et al., 2006). Purified PCR products were subsequently sent to Macrogen Inc. (Amsterdam, Netherlands) for Sanger sequencing in one or both directions.

4.2.3 Processing of 454 and Sanger sequence data

Raw 454-reads were united with 454-reads from a study of the genus *Leucanthemum* by Konowalik et al. (2015) and further processed using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (Caporaso et al., 2010) as described in detail in Supporting Information Methods S4.1. If the pipeline failed to find alleles, we collapsed all polymorphic sites in the reads using the International Union of Pure and Applied Chemistry (IUPAC) nucleotide code. Resulting consensus sequences were incorporated into the phasing procedure with the software PHASE v2.1.1 (Stephens et al., 2001; Stephens and Scheet, 2005) as described below (see also Supporting Information Figure S4.1).

Electropherograms obtained from Sanger sequencing were checked manually for base-call errors using CHROMAS LITE v2.0 (Technelysium Pty Ltd, South Brisbane, Australia) and a poly-A repeat was discarded in the sequences of the plastid marker *trnC-petN* to avoid misalignment and homoplasy. All cloned nuclear sequences were assigned to alleles and screened for chimeric sequences using the Neighbor-net approach in SplitsTree v4.14.6 (Huson and Bryant, 2006) following Bertrand et al. (2015). Whenever we found more than one polymorphic site in a directly sequenced nuclear sequence, we used one of the following phasing methods depending on the relative length of the two underlying alleles: (i) in the case of length differences between overlying sequence copies, CHAMPURU v1.0 (Flot et al., 2006; Flot, 2007) was applied to the forward and reverse sequences, while (ii) PHASE v2.1.1 was used for disentangling alleles of equal length. As PHASE is a Bayesian method that uses Gibbs sampling to calculate the posterior distribution of unknown haplotype pairs given known genotypes, all sequences from homozygous individuals and already known haplotype pairs (from QIIME, cloning, or CHAMPURU allele phasing) were combined and processed together with all unphased sequences from Sanger and 454 sequencing in PHASE (Supporting Information Figure S4.1). The webtool SEQPHASE (Flot, 2010) was used to process PHASE input and output files and all PHASE analyses were run genus- and marker-wise with default settings (phase threshold = 90%, 100 iterations, thinning interval = 1, burn-in = 100) as in Hedin (2015) and Hedin et al. (2015).

4.2.4 Multiple sequence alignments and model selection

In total, 25 alignments were generated using the online service of MAFFT v.7.402 with default (--auto) settings (Kuraku et al., 2013; Katoh et al., 2017). We created a ‘total dataset’ of five alignments by aligning sequences of all 129 accessions for the markers *A39*, *B20*, *D27*, nrDNA ITS, *trnC-petN*, *trnL-trnF*, and *trnQ-rps16*, separately. All plastid loci were subsequently concatenated and treated as one single marker. Two further datasets consisting of 10 alignments each (all eight Chapman markers, nrDNA ITS, and the five concatenated plastid loci) were generated by selecting only *Leucanthemum* (42 accessions, ‘*Leucanthemum* dataset’) and *Rhodanthemum* individuals (52 accessions, ‘*Rhodanthemum* dataset’), respectively. For each alignment, the number of variable sites, parsimony informativeness, consistency (CI) and retention index (RI) was calculated in PAUP* v.4.0 (Swofford, 2003) and the best fitting nucleotide substitution model was determined using the Akaike information criterion (AIC) in JMODELTEST v.2.1.10 (Darriba et al., 2012). Bayesian gene-tree estimation and marginal-likelihood calculations were performed for each marker of each dataset to determine the appropriate clock model. For this purpose, we ran each alignment separately with a strict and an uncorrelated relaxed-clock model (Drummond et al., 2006) in BEAST v.1.8.4 (Drummond and Rambaut, 2007; Suchard and Rambaut, 2009). All xml files were generated in BEAUTI v.1.8.4 using a gamma distribution with shape 2.0 and scale 0.002 for the coalescent constant tree prior and otherwise default priors. Substitution models that were not available in BEAUTI were manually specified in the xml files, before they were uploaded to the CIPRES web portal (Miller et al., 2010) to perform two runs each with 100 million generations and a sample frequency of 10,000. For the purpose of clock model selection (strict vs. relaxed), marginal likelihoods were calculated using the path-sampling (Baele et al., 2012) and stepping-stone (Baele et al., 2013) methods with a chain length of 1 million generations and 100 path steps. Marginal-likelihood values were averaged over replicate runs and only in the case of a difference of more than three mean log-likelihood values was the more parameter-rich relaxed clock accepted (Kass and Raftery, 1995).

4.2.5 Multi-species coalescent (MSC) species-delimitation

To assign individuals to distinct lineages in the two focal genera *Leucanthemum* and *Rhodanthemum* as a prerequisite for all subsequent analyses, species delimitation analyses were conducted with the BEAST2 (Bouckaert et al., 2014) package STACEY v.1.2.4 (Jones, 2017a) as in Wagner et al. (2017). We used BEAUTI v.2.4.8 to generate two independent xml files for the ‘*Leucanthemum* dataset’ and ‘*Rhodanthemum* dataset’, respectively. Substitution and clock models for the ten unlinked loci of each dataset were specified in accordance with

the results of the above described model-selection part and a uniform distribution from zero to one was chosen for the collapse weight parameter in order to give an uninformative prior information about the likely number of species. To prevent improper priors as suggested in the software documentation of STACEY, we changed the scaling factor for the population size (popPriorScale) to an exponential distribution with a mean of 0.1 and the growth rate prior (bdcGrowthRate) to a lognormal distribution (mean 5, standard deviation 2) following Barley et al. (2018). Each xml file was subsequently executed three times independently for 500 million generations and a sample frequency of 50,000 on the CIPRES Science Gateway using BEAST v.2.4.8. After checking convergence and ESS values via TRACER (Rambaut et al., 2018), all log- and tree-files were combined in LOGCOMBINER v.2.4.8 with a burn-in of 10%. The combined tree samples were subsequently analyzed with TREEANOTATOR v.2.4.8 and SPECIESDELIMITATIONANALYSER v.1.8.0 (collapseheight = 1.0e-4, simcutoff = 1.0) and the resulting maximum clade credibility trees and tables of clusterings were visualized with FIGTREE v.1.4.3 and a customized R script provided by Jones et al. (2015).

4.2.6 Inference of genetic divergence patterns

To investigate genetic divergence patterns in both focal genera, we calculated phylogenetic Bray-Curtis distances (PBC; Göker and Grimm, 2008) among all delimited lineages in *Leucanthemum* and *Rhodanthemum*, respectively. For this purpose, the software POFAD (Joly and Bruneau, 2006) was applied to all ten sequence alignments of the ‘*Leucanthemum* dataset’ and ‘*Rhodanthemum* dataset’, separately using allele mappings according to the outcome of the above described STACEY analyses. The resulting PBC distance matrices were subsequently tested for significant deviations from each other using a Mann-Whitney U test in R v.3.4.5 (R Development Core Team, 2017).

4.2.7 Inference of homoploid hybridization patterns

We calculated the genealogical sorting index (*gsi*, Cummings et al., 2008) for all markers and lineages delimited in the above-described STACEY analyses for the genera *Leucanthemum* and *Rhodanthemum*. As a frequently used statistic for detection of hybridization in plants (De Villiers et al., 2013; Konowalik et al., 2015; Meeus et al., 2016), the *gsi* evaluates the degree of monophyly of all accessions of a predefined group by ranging from 0 to 1, with higher values corresponding to more phylogenetic exclusivity (Winter et al., 2016). A marker-wise *gsi* value for each lineage of *Leucanthemum* and *Rhodanthemum* was calculated based on an ensemble of trees (gsi_T). For this purpose LOGCOMBINER v.2.4.8 was applied to the resulting gene-tree distributions of the two replicate BEAST runs with the highest mean marginal likelihood value (see model selection earlier) using a burn-in of 10%

and a re-sample frequency of 1.8 million. The resulting 100 gene-tree topologies for each marker were subsequently used to calculate gene-wise gsi_T values for each ‘STACEY-lineage’ with the PYTHON script GSI.py (Kryvokhyzha, 2017). Furthermore, we performed the accompanying permutation test that assesses the probability of calculating a gsi_T value equal to or greater than the observed one by chance under a null-hypothesis that there is no significant association among the leaves of each lineage, using 5,000 permutations.

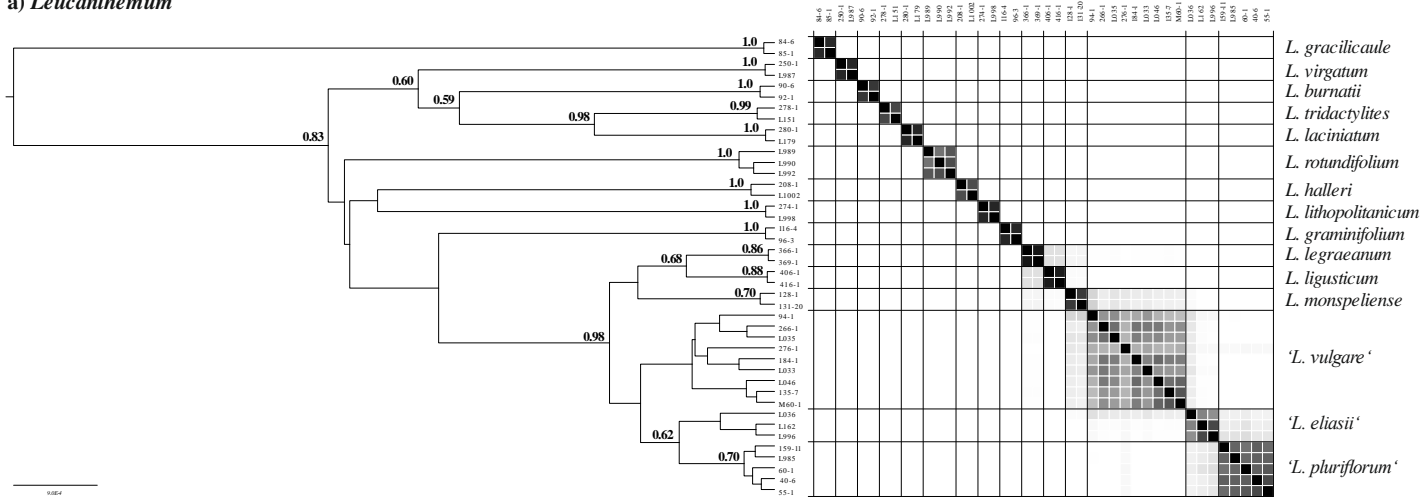
Posterior predictive checking with the software JML (Joly, 2012) was conducted as a second approach for detecting hybridization patterns among diploid lineages in the two genera under study. In contrast to the gene-tree based gsi method, JML utilizes species trees to predict hybridization events in a given dataset. This is done by using the posterior distribution of species trees with branch lengths and population sizes from an appropriate species-tree analysis (mostly from *BEAST) for the simulation of sequences via gene trees under the coalescent-with-no-migration model. The minimum pairwise sequence distances (*minDist*) among simulated sequences of two species are finally used as a null-hypothesis of a strictly bifurcating evolution, and a subsequent test of empirical values against the null-hypothesis distributions allows for identification of potential hybrids.

We conducted species-tree analyses in BEAST v.1.8.0 (Drummond and Rambaut, 2007; Suchard and Rambaut, 2009) for both genera to obtain species-tree distributions that are suitable for genus-wise hybrid detection via JML. Species assignments, substitution and clock models were set in the *BEAST template of BEAUTI v.1.8.0 according to the results of the species-delimitation and model-selection procedures described above. The ‘autosomal nuclear’ ploidy type was selected for nuclear markers, the ‘mitochondrial’ type for the concatenated plastid sequences and a ‘piecewise constant’ population-size model for all loci as suggested in the JML manual. A gamma prior ($shape = 2$, $scale = 0.002$) was applied to the ‘population size’ and the ‘species tree birth rate’ following Aydin et al. (2014), and four independent runs for each genus with 500 million generations and a sample frequency of 50,000 were conducted on the CIPRES platform. Replicate runs were subsequently checked for convergence and proper ESS values in TRACER and trimmed (10% burn-in), combined and resampled in LOGCOMBINER to obtain a final set of 9,000 species trees for each genus. The resulting species-tree distributions and the underlying sequence alignments were subsequently analyzed with JML v.1.3.0 in a marker-wise fashion. ‘Locusrates’, ‘heredityscalars’, and substitution models in JML were set according to the *BEAST analyses and a thinning of 2 (i.e., using only every second species tree for simulation) was adjusted to the plastid datasets to reduce computational complexity. For all twenty JML analyses (i.e., two genera and ten markers), a significance level of 0.05 was assumed.

4.2.8 Divergence time estimation

In order to elucidate the phylogenetic relationships and the divergence times of the two focal genera *Leucanthemum* and *Rhodanthemum* within the subtribe Leucantheinae, a *BEAST (Heled and Drummond, 2010) analysis was conducted based on the ‘total dataset’ and two calibration points following Tomasello et al. (2015). Owing to the observation that hybridization may bias downstream phylogenetic analyses (Leaché et al., 2014b; Meyer et al., 2017; Smith et al., 2018), we have omitted markers that would influence crown-age determination for the two genera by contributing an incongruence signal (especially marker *B12* in *L. gracilicaule* based on the earlier JML results). The first calibration point was the age of the node at the split between *Ursinia* and the Asian-southern African grade (represented by *Artemisia*) plus the Euro-Mediterranean clade (represented by all remaining accessions) of the Anthemideae. The time range of 28-38 Ma for this calibration point has been estimated in a re-calibration analysis of Tomasello et al. (2015) based on a *ndhF* dataset of Kim and Jansen (1995) for the whole family of Compositae. The estimated range was incorporated in our present analysis by setting a normal distribution (mean: 33.8 Ma; SD: 3 Ma) for the most recent common ancestor (MRCA) prior of the root in BEAUTI v.2.4.8. A second calibration point was set by defining a lognormal prior (mean: 2.7, SD: 0.5) with an offset of 23.05 Ma for the split between *Artemisia* and all accessions of the Euro-Mediterranean clade. This prior is based on the age of *Artemisia* calculated from fossilized pollen records of the Lower and Upper Oligocene (Wang, 2004). Substitution models in BEAUTI were chosen according to the results of JMODELTEST using the BEAST2 package SSM v.1.0.1 (Bouckaert and Xie, 2017) and a log normal relaxed-clock model (Drummond et al., 2006) was set for each of the five unlinked partitions following the marginal-likelihood-based model-comparison as described above. A ‘Yule model’ was chosen for the species tree together with a ‘linear with constant root’ model for the population size and all default priors given by BEAUTI v.2.4.8 were accepted. We performed four independent runs on CIPRES under BEAST v.2.4.8, each with a length of 500 million generations and a sample frequency of 50,000 together with an additional run without data (sample from prior) to check for spurious prior distribution interactions. All tree- and log-files of the four replicate runs were checked for convergence and ESS values in TRACER and finally combined using a 10% burn-in and a resample frequency of 200,000 in LOGCOMBINER v.2.4.8. The combined log-file was compared to the outcome of the ‘sample from the prior’ run, to check whether the priors might overwhelm the signal in the data, and a maximum-clade-credibility tree with a posterior-probability limit of 0.5 was calculated with TREEANOTATOR v.2.4.8.

a) *Leucanthemum*



b) *Rhodanthemum*

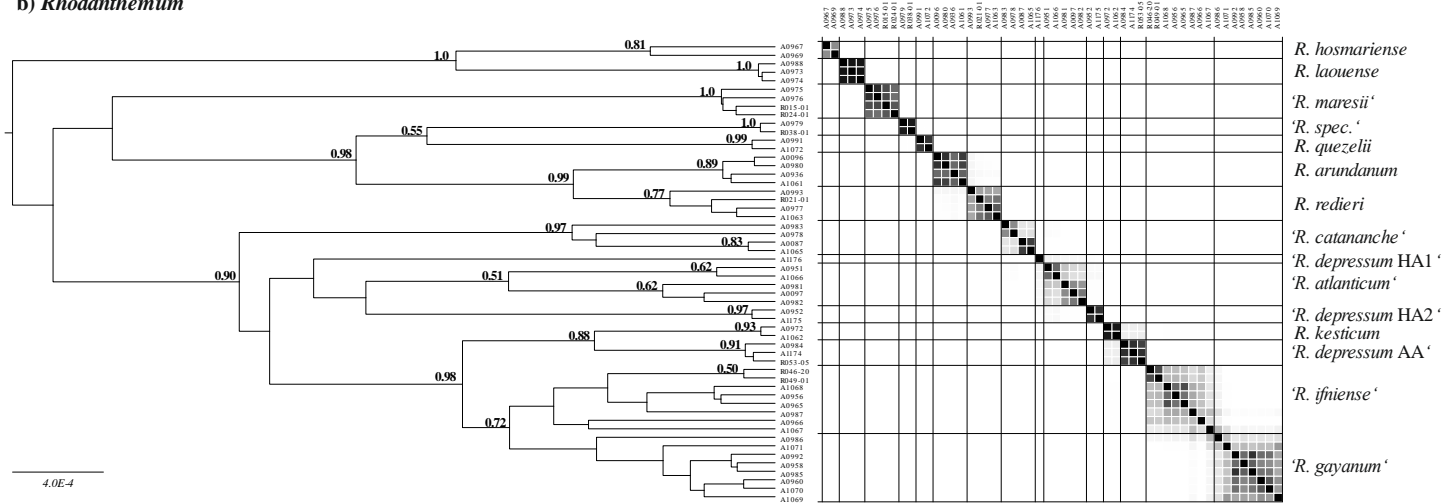


Figure 4.2 Results of species delimitation analyses with the BEAST application STACEY for the genera *Leucanthemum* (a) and *Rhodanthemum* (b). Similarity matrices to the right of the accession trees visualize posterior probability (PP) values for each pair of individuals to belong to the same cluster (black, PP = 1.0; white, PP = 0.0). The resulting 15 lineages for each genus are given to the right of the similarity matrices. All lineages that result from merging or splitting of morphologically described species into new units are shown in quotation marks.

4.3 Results

4.3.1 Multiple sequence alignments and model selection

After checking for barcode errors and primer mismatches followed by a quality filtering step of raw Roche 454 pyrosequencing data, we retrieved a total of 77,067 quality-filtered reads, of which 57,598 were generated in the course of the present study and 19,469 came from Konowalik et al. (2015; see Supporting Information Table S4.3 for details). Extraction of alleles and addition of sequence information obtained via Sanger sequencing resulted in 25 alignments varying in size, length, information content, and best-fitting substitution and clock models as shown in Supporting Information Tables S4.4 and S4.5.

4.3.2 MSC species delimitation

Results from species delimitation analyses for the genera *Leucanthemum* and *Rhodanthemum* are shown in Figure 4.2. Evaluation of both output formats from STACEY analyses (accession trees and similarity matrices) in conjunction with consideration of geographical aspects (described in detail in Supporting Information Notes S4.1 and Table S4.1) led to the discrimination of 15 units in both genera, which were used as fixed ‘lineages’ for all following analyses.

4.3.3 Genetic divergence patterns

Distributions of Phylogenetic Bray-Curtis (PBC) distances of all lineage pairs for *Leucanthemum* and *Rhodanthemum* are depicted in Figure 4.3. The two genera show significantly different genetic divergence patterns (PBC-distance distributions) according to a Mann-Whitney U test ($P < 0.001$), with lineages of *Leucanthemum* exhibiting on average higher PBC-distances (mean = 0.47, SD = 0.11) than those found among *Rhodanthemum* lineages (mean = 0.39, SD = 0.10).

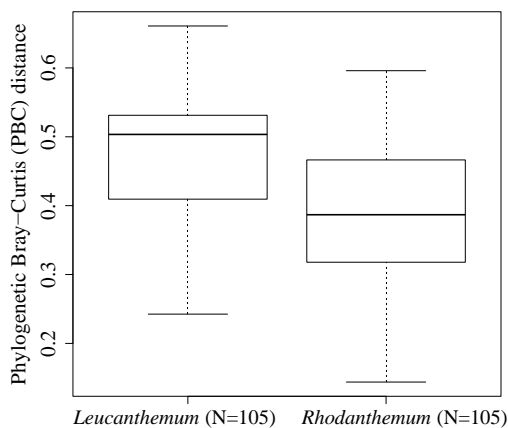


Figure 4.3 Box plots of the distribution of Phylogenetic Bray-Curtis distances of all lineage pairs for *Leucanthemum* and *Rhodanthemum*. The two genera show significantly different genetic divergence patterns (Mann-Whitney U-test, $P < 0.001$; see text for details), with lineages of *Leucanthemum* being significantly more divergent from each other than those of *Rhodanthemum*. Boxplots follow the standard convention, with solid lines reflecting the median, hinges the first and third quartiles, and whiskers the first and third quartiles plus $1.5 \times$ the interquartile range.

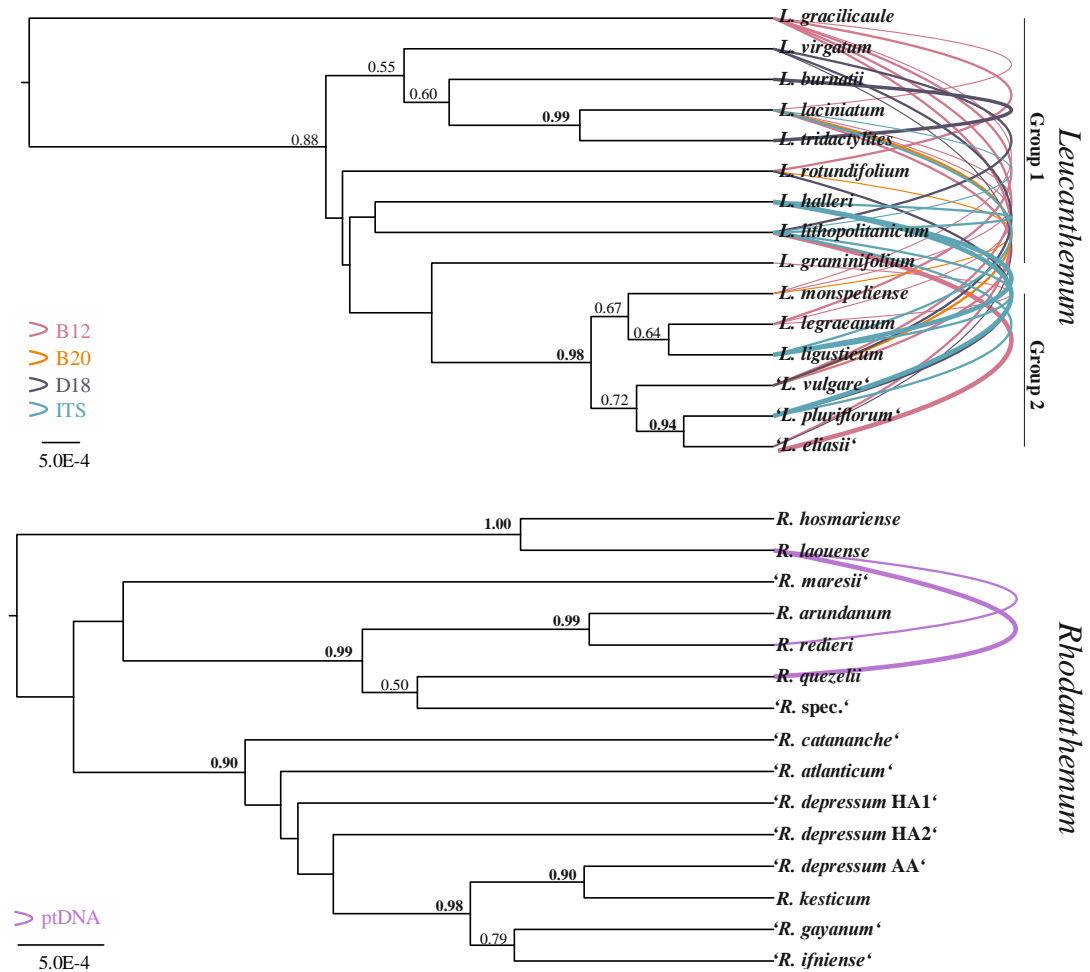


Figure 4.4 Results of *BEAST and JML analyses for the genera *Leucanthemum* and *Rhodanthemum* based on nine nuclear and five plastid loci. All posterior probability values > 0.5 are shown in the species trees and values ≥ 0.9 are indicated in bold. All significant ($P < 0.05$) hybridization events between accessions detected via JML (see also Table 4.2) are plotted by drawing curves between corresponding lineages with colors indicating the relevant marker concerned and with line widths proportional to the number of hybridization events inferred.

4.3.4 Homoploid hybridization patterns

The gsi_T values calculated for *Leucanthemum* and *Rhodanthemum* on the basis of 100 gene-tree topologies per marker fall within the range of 0.0772-1.0000 and 0.0575-1.0000, respectively (Table 4.1). Despite a slight tendency towards higher gsi_T values in *Leucanthemum* compared to *Rhodanthemum* (mean 0.4122 vs. 0.3352), we found a comparable percentage (65% vs. 67%) of gsi_T values significantly differing from zero in both genera, indicating a similar pattern in the non-random distribution of accessions/alleles of lineages across gene trees. Similar trends in both datasets were also found on the marker level, with the highest mean gsi_T values calculated for the plastid and nrDNA ITS loci (Table 4.1).

Results from *BEAST species tree reconstructions (Supporting Information Notes S4.2) and subsequent tests for hybridization in JML are given in Figure 4.4 and Table 4.2. In total, we

found 49 cases in *Leucanthemum*, where the observed minimum distance between empirical sequences of two individuals of different lineages was not adequately predictable via simulations under the coalescent with a no-migration model. In *Rhodanthemum*, however, only six cases were found (see Table 4.2 and Figure 4.4). In detail, all *Leucanthemum* lineages show at least one hybridization event, with a maximum of ten for the lineages *L. gracilicaule*, *L. laciniatum*, and *L. halleri* and a minimum of one for *L. graminifolium*. Furthermore, all significant hybridization events were attributable to nuclear markers, with a strong hybridization pattern in the markers *B12*, nrDNA ITS, and *D18* (19, 17 and 10 events, respectively), but only three hits in *B20*. All other markers, including the concatenated plastid loci, yielded no significant results ($P > 0.05$). In contrast, we found signs of hybridization between lineages of *Rhodanthemum* exclusively in the concatenated plastid sequences for the lineages *R. laouense*, *R. quezelii*, and *R. redieri*.

4.3.5 Divergence times estimation

Results from the dated *BEAST analysis based on the ‘total dataset’ are shown in Figure 4.5 and Table 4.3. Monophyly of the included genera was generally highly supported ($PP = 1.0$) in the maximum-clade-credibility (MCC) tree except for *Mauranthemum* ($PP = 0.74$), *Plagius* ($PP < 0.5$), and the two species of the genus *Coleostephus*, for which no sister-group relationship was found at all. All genera of the Leucanthemeinae sensu Bremer and Humphries (1993) were part of an unsupported monophyletic group, while a more comprehensive clade additionally contained *Davea*, *Heteromera*, and *Otospermum* with very high support ($PP = 1.0$). The crown age of this ‘extended subtribe’ (clade C in Figure 4.5) falls into the Miocene (8.71-15.38 Ma) and coincides with the estimation of Oberprieler (2005) for the same group of taxa. In contrast to the latter study, however, we found the unspecific genus *Chlamydophora* not only being nested within the subtribe Leucanthemeinae but also in a well-supported monophyletic group ($PP = 0.98$) together with *Leucanthemum* and *Rhodanthemum*. Diversification of the two latter genera falls into the Quaternary, with a slightly older crown age for *Leucanthemum* [clade D: 1.93 Ma (1.14-2.94 Ma)] compared to *Rhodanthemum* [clade E: 1.29 Ma (0.88-1.87 Ma)].

4.4 Discussion

Table 4.1 Genealogical sorting index (*gsir*) values for *Leucanthemum* and *Rhodanthemum* lineages. Bold numbers indicate *gsir* values significantly different from zero.

Lineage	individuals	A39	B12	B20	C12	C20	D18	D23	D27	ITS	ptDNA	Lineage mean
<i>L. gracilicaule</i>	2	0.5079	0.2756	0.3174	0.5296	0.3103	0.1017	0.8515	0.2167	1.0000	1.0000	0.5111
<i>L. virgatum</i>	2	0.4374	0.9249	0.0994	1.0000	1.0000	0.1720	0.6196	0.1488	0.8058	1.0000	0.6208
<i>L. burnatii</i>	2	0.5520	0.1328	0.1249	0.4313	0.9938	0.1507	0.4916	0.1944	1.0000	0.3196	0.4391
<i>L. laciniatum</i>	2	0.2934	0.1283	0.5638	0.6423	1.0000	1.0000	0.4872	0.1591	0.6612	0.6184	0.5554
<i>L. tridactylites</i>	2	0.2144	0.7411	0.1159	0.1842	0.3325	0.1562	0.4855	1.0000	0.2216	0.5802	0.4032
<i>L. rotundifolium</i>	3	0.3866	0.1978	0.3205	1.0000	0.2715	0.1736	0.1395	0.2491	0.5452	0.9930	0.4277
<i>L. halleri</i>	2	0.0872	1.0000	0.1337	0.9880	0.9053	0.1149	0.2879	0.2165	0.2261	0.4323	0.4392
<i>L. lithopolitanicum</i>	2	0.2479	1.0000	0.9526	0.3564	1.0000	0.2558	1.0000	0.1624	0.2362	1.0000	0.6211
<i>L. graminifolium</i>	2	1.0000	0.0848	1.0000	0.4264	0.7247	0.2305	1.0000	1.0000	1.0000	0.3175	0.6784
<i>L. monspeliense</i>	2	0.1692	0.0943	0.0777	0.2300	0.1020	0.1826	0.2237	0.0966	0.1451	0.2728	0.1594
<i>L. legraeum</i>	2	0.2789	0.0772	0.1554	0.1886	0.0795	0.1052	0.3197	0.2172	1.0000	0.6413	0.3063
<i>L. ligusticum</i>	2	0.3965	0.2353	0.1087	0.9966	0.1919	0.1192	0.0914	0.1915	0.1172	0.3136	0.2762
' <i>L. vulgare</i> '	9	0.2415	0.2495	0.1878	0.2494	0.2231	0.2235	0.3197	0.1784	0.3160	0.2768	0.2466
' <i>L. pluriflorum</i> '	5	0.1730	0.1520	0.1585	0.1878	0.1389	0.2679	0.2802	0.2583	0.2075	0.2509	0.2075
' <i>L. eliasii</i> '	3	0.1160	0.1246	0.2013	0.2535	0.2716	0.1398	0.1660	0.7337	0.7510	0.1501	0.2908
Marker mean		0.3401	0.3612	0.3012	0.5109	0.5030	0.2262	0.4509	0.3348	0.5489	0.5444	
<i>R. hosmariense</i>	2	0.1843	0.1378	0.1294	0.4501	0.1237	0.6956	0.2477	0.1494	0.9949	0.1638	0.3277
<i>R. laouense</i>	3	0.1509	0.1267	0.1496	0.4432	0.2192	1.0000	0.8043	0.9916	1.0000	0.4076	0.5293
' <i>R. maresii</i> '	4	0.2422	0.2193	0.1722	0.3895	0.6668	0.2399	0.2908	0.3230	1.0000	0.3547	0.3899
<i>R. arundanum</i>	4	0.7656	0.4089	0.1516	0.2909	0.5480	0.1775	0.1840	0.4479	0.7908	0.1713	0.3936
<i>R. redieri</i>	4	0.2611	0.2738	0.1712	0.3751	0.3042	0.2650	0.3589	0.3115	0.4652	0.2034	0.2989
<i>R. quezelii</i>	2	0.1527	0.1271	0.3787	0.2430	0.1152	0.1132	0.1368	0.1772	0.8126	1.0000	0.3256
' <i>R. spec.</i> '	2	0.3727	0.1396	0.1368	0.2545	0.1307	0.1183	0.1592	0.1976	0.9804	0.8080	0.3298
' <i>R. catananche</i> '	4	0.2026	0.1838	0.2084	0.9933	0.1975	0.1646	1.0000	0.3417	0.4068	0.1346	0.3833
' <i>R. atlanticum</i> '	5	0.2117	0.2173	0.1341	0.2051	0.1695	0.2239	0.2006	0.2134	0.5119	0.2280	0.2315
' <i>R. depressum</i> HA1'	1	N/A	N/A	0.0575	0.9949	N/A	N/A	0.1043	N/A	0.1813	N/A	0.3345
' <i>R. depressum</i> HA2'	2	0.1126	0.9916	0.9924	0.1479	1.0000	0.1461	0.2112	0.2667	0.3030	0.1008	0.4272
' <i>R. depressum</i> AA'	3	0.2197	0.1478	0.1578	0.2924	0.1212	0.1591	0.1675	0.1017	0.4586	0.7506	0.2576
<i>R. kesticum</i>	2	0.1095	0.2009	0.1089	0.1225	0.0972	0.1085	0.1087	0.0759	0.1670	1.0000	0.2099
' <i>R. gayanum</i> '	8	0.3186	0.5250	0.1919	0.3027	0.2158	0.1458	0.4187	0.2515	0.3421	0.3263	0.3038
' <i>R. ifniense</i> '	8	0.2887	0.1685	0.4765	0.3662	0.1578	0.2012	0.1816	0.2665	0.4288	0.3095	0.2845
Marker mean		0.2566	0.2763	0.2411	0.3914	0.2905	0.2685	0.3050	0.2940	0.5896	0.4256	

4.4 Discussion

4.4.1 Does phylogenetic divergence drive polyploidization?

One of the oldest theories concerning factors promoting polyploidy is the idea that successful polyploid formation and establishment is correlated with the genomic divergence of diploid progenitors (e.g., Darlington, 1937; Grant, 1981). This idea, often cited as 'Darlington's rule', was primarily based on functional aspects of genome interactions, but was recently revisited in molecular studies using genetic distances between parental species of allopolyploids as a proxy for evolutionary and cytogenetic divergence (Buggs et al., 2011). Chapman and Burke (2007) used DNA sequences from the internal transcribed spacer (ITS) region to assess the Kimura's two-parameter (K2P) genetic distances between parental species of 12 homoploid and 26 polyploid hybrids. Subsequent comparison of all homoploid

hybrid versus allopolyploid parental pairs provided a significantly larger divergence between the parents of allopolyploids. In contrast to these findings, Buggs et al. (2008) found that the phylogenetic divergence between parents of polyploids in eight different genera calculated via node-based and clade-based methods was not significantly different from the divergence expected under the null-hypothesis that hybridization occurs at random among all species of a genus. Paun et al. (2009) on the other hand developed a customized ‘genetic divergence index’ (GDI) to test Darlington’s rule in a survey based on p- and K2P-distances from ITS and/or low-copy nuclear gene sequences of 32 allopolyploid species with known progenitors. The GDI was calculated by dividing the parental divergence for each polyploid representative by the average genetic distance between all species pairs in the genus concerned. This approach provided significantly higher GDI values for parents of polyploids compared to those of homoploid hybrids and parents of polyploids were generally more divergent than the average intrageneric distance. However, the result of Paun et al. (2009) was not reproducible in a re-analysis of the same data by Buggs et al. (2009), where the authors used the average divergence between all species pairs in a genus as a null-hypothesis for the expected divergence between parents of allopolyploids. In a critical survey of all above-mentioned studies, Buggs et al. (2011) concluded that ‘there is not currently persuasive evidence that hybridization between divergent parents serves as a driver for polyploidization’.

Two main problems were pointed out in the last mentioned review article that are connected with sampling strategies of all phylogenetic studies testing ‘Darlington’s rule’: the negligence of autopolyploids and the uncertainty connected with species delimitation in hybridizing plant groups. Further limitations of the mentioned studies are (i) the incorporation of only few phylogenetic markers (mostly nrDNA ITS), (ii) the inclusion of hybrid-parents with different base chromosome numbers (Chapman and Burke, 2007), or (iii) the uncertainty concerning the parentage of investigated polyploid species (Buggs et al., 2008). In the present study, we have tried to bypass these problems with a new approach: while the studies of Chapman and Burke (2007), Buggs et al. (2008, 2009), and Paun et al. (2009) are all based on prior-determined parental pairs of allopolyploid species, the mode of origin (allo- or autopolyploid formation) and the putatively involved parental species are mostly unknown in the investigated genus *Leucanthemum*. Instead of searching for differences in the divergence patterns of polyploid versus homoploid progenitors (Chapman and Burke, 2007, Paun et al., 2009) or between polyploid progenitors and all remaining species pairs in a genus (Buggs et al., 2008, 2009), we more generally investigate differences in the genetic distances among all diploid representatives of a polyploid complex (*Leucanthemum*) compared to those in a closely related, strictly diploid genus (*Rhodanthemum*). Furthermore, the Phylogenetic Bray Curtis (PBC) distance as

implemented in the software POFAD was used for all pairwise genetic distance calculations instead of p- or K2P-distances to simultaneously incorporate allelic variation, multiple sequences per species and multiple markers. Additionally, we performed species delimitation analyses with the multi-species coalescent method STACEY prior to all genetic distance calculations to account for uncertainty concerning the assignment of accessions to evolutionary entities ('lineages') in the morphologically closely-knit genera *Leucanthemum* and *Rhodanthemum*.

Our data show that diploid lineages in the polyploid complex *Leucanthemum* are significantly more divergent compared to lineages of the strictly diploid genus *Rhodanthemum* (Figure 4.3). This observation is in line with the longstanding assumption that the probability for (allo)polyploid formation and establishment is positively correlated with the genomic divergence of diploid progenitors (e.g., Darlington, 1937; Grant, 1981, Sang et al. 2004). Two possible mechanisms are discussed to explain this correlation (e.g., Sang et al., 2004; Paun et al., 2009). A first idea goes back to Grant (1981), who predicted that the probability of unreduced gamete formation due to meiotic abnormalities is increased in homoploid hybrids between more distantly related parents. Van Tuyl et al. (1989) confirmed this assumption in an empirical study of *Lilium*, where an increased frequency of unreduced gamete production in wide interspecific hybrids was demonstrated. The high probability for unreduced gamete formation in wide interspecific hybrids, in turn, might lead to an increased frequency of allopolyploid speciation via the 'triploid bridge' pathway (Ramsey and Schemske, 1998). A second explanation for the positive correlation of parental divergence and (allo)polyploidization was proposed by Darlington (1937) and is rather connected with the successful establishment of polyploids than with their formation. Darlington (1937) reasoned that high parental divergence leads to a decrease of meiotic abnormalities (multivalent formation and uneven segregation) in allopolyploids and thus to an increase of fitness.

Table 4.2 Results from posterior predictive checking with the software JML for the genera *Leucanthemum* and *Rhodanthemum*. Listed are all cases, where the minimum pairwise sequence distances (observed distance) among empirical sequences of two lineages are significantly smaller ($P < 0.05$) than expected under a null hypothesis of a strictly bifurcating evolution.

Gene	Lineage comparison	Individual 1	Individual 2	Obs. distance	P-value
B12	<i>L. laciniatum</i> - <i>L. gracilicaule</i>	280.1	84.6	0.0106	0.0404
B12	<i>L. legraeianum</i> - <i>L. gracilicaule</i>	369.1	84.6	0.0080	0.0211
B12	<i>L. legraeianum</i> - <i>L. gracilicaule</i>	369.1	85.1	0.0106	0.0475
B12	<i>L. legraeianum</i> - <i>L. graminifolium</i>	366.1	96.3	0.0053	0.0499
B12	<i>L. legraeianum</i> - <i>L. laciniatum</i>	369.1	280.1	0.0027	0.0101
B12	<i>L. monspeliense</i> - <i>L. gracilicaule</i>	131.2	84.6	0.0080	0.0282
B12	<i>L. monspeliense</i> - <i>L. laciniatum</i>	131.2	280.1	0.0027	0.0137
B12	<i>L. rotundifolium</i> - <i>L. gracilicaule</i>	L990	84.6	0.0000	0.0010
B12	<i>L. rotundifolium</i> - <i>L. gracilicaule</i>	L990	85.1	0.0080	0.0305
B12	' <i>L. eliasii</i> ' - <i>L. gracilicaule</i>	L996	84.6	0.0000	0.0013
B12	' <i>L. eliasii</i> ' - <i>L. gracilicaule</i>	L996	85.1	0.0080	0.0402
B12	' <i>L. eliasii</i> ' - <i>L. rotundifolium</i>	L996	L990	0.0000	0.0131
B12	' <i>L. eliasii</i> ' - <i>L. rotundifolium</i>	L996	L989	0.0000	0.0131
B12	' <i>L. eliasii</i> ' - <i>L. rotundifolium</i>	L996	L990	0.0000	0.0131
B12	' <i>L. eliasii</i> ' - <i>L. rotundifolium</i>	L996	L992	0.0000	0.0131
B12	' <i>L. vulgare</i> ' - <i>L. gracilicaule</i>	94.1	84.6	0.0027	0.0076
B12	' <i>L. vulgare</i> ' - <i>L. gracilicaule</i>	94.1	85.1	0.0053	0.0215
B12	' <i>L. vulgare</i> ' - <i>L. laciniatum</i>	94.1	280.1	0.0027	0.0331
B12	' <i>L. vulgare</i> ' - <i>L. laciniatum</i>	L035	280.1	0.0027	0.0331
B20	<i>L. rotundifolium</i> - <i>L. monspeliense</i>	L990	128.1	0.0000	0.0233
B20	' <i>L. vulgare</i> ' - <i>L. laciniatum</i>	L035	280.1	0.0000	0.0244
B20	' <i>L. vulgare</i> ' - <i>L. laciniatum</i>	L035	L179	0.0000	0.0244
D18	<i>L. rotundifolium</i> - ' <i>L. pluriflorum</i> '	L992	40.6	0.0000	0.0337
D18	<i>L. rotundifolium</i> - ' <i>L. pluriflorum</i> '	L992	40.6	0.0000	0.0337
D18	<i>L. tridactylites</i> - <i>L. burnatii</i>	L151	90.6	0.0000	0.0454
D18	<i>L. tridactylites</i> - <i>L. burnatii</i>	L151	90.6	0.0000	0.0454
D18	<i>L. tridactylites</i> - <i>L. burnatii</i>	L151	92.1	0.0000	0.0454
D18	<i>L. virgatum</i> - <i>L. lithopolitanicum</i>	L987	274.1	0.0031	0.0425
D18	<i>L. virgatum</i> - <i>L. lithopolitanicum</i>	L987	L998	0.0031	0.0425
D18	<i>L. virgatum</i> - ' <i>L. eliasii</i> '	L987	L996	0.0000	0.0180
D18	' <i>L. vulgare</i> ' - <i>L. virgatum</i>	184.1	L987	0.0000	0.0319
D18	' <i>L. vulgare</i> ' - <i>L. virgatum</i>	L033	L987	0.0000	0.0319
ITS	<i>L. ligusticum</i> - <i>L. halleri</i>	406.1	208.1	0.0000	0.0184
ITS	<i>L. ligusticum</i> - <i>L. halleri</i>	406.1	L1002	0.0000	0.0184
ITS	<i>L. ligusticum</i> - <i>L. halleri</i>	416.1	208.1	0.0000	0.0184
ITS	<i>L. ligusticum</i> - <i>L. halleri</i>	416.1	L1002	0.0000	0.0184
ITS	<i>L. ligusticum</i> - <i>L. laciniatum</i>	406.1	L179	0.0013	0.0364
ITS	<i>L. ligusticum</i> - <i>L. laciniatum</i>	416.1	L179	0.0013	0.0364
ITS	<i>L. lithopolitanicum</i> - <i>L. halleri</i>	274.1	208.1	0.0000	0.0375
ITS	<i>L. lithopolitanicum</i> - <i>L. halleri</i>	274.1	L1002	0.0000	0.0375
ITS	<i>L. lithopolitanicum</i> - <i>L. laciniatum</i>	274.1	L179	0.0013	0.0456
ITS	<i>L. lithopolitanicum</i> - <i>L. ligusticum</i>	274.1	406.1	0.0000	0.0158
ITS	<i>L. lithopolitanicum</i> - <i>L. ligusticum</i>	274.1	416.1	0.0000	0.0158
ITS	' <i>L. pluriflorum</i> ' - <i>L. halleri</i>	L985	208.1	0.0000	0.0421
ITS	' <i>L. pluriflorum</i> ' - <i>L. halleri</i>	L985	L1002	0.0000	0.0421
ITS	' <i>L. pluriflorum</i> ' - <i>L. halleri</i>	60.1	208.1	0.0000	0.0421
ITS	' <i>L. pluriflorum</i> ' - <i>L. halleri</i>	60.1	L1002	0.0000	0.0421
ITS	' <i>L. pluriflorum</i> ' - <i>L. lithopolitanicum</i>	L985	274.1	0.0000	0.0368
ITS	' <i>L. pluriflorum</i> ' - <i>L. lithopolitanicum</i>	60.1	274.1	0.0000	0.0368
ptDNA	<i>R. laouense</i> - <i>R. quezelii</i>	A0973	A0991	0.0004	0.0347
ptDNA	<i>R. laouense</i> - <i>R. quezelii</i>	A0973	A1072	0.0004	0.0347
ptDNA	<i>R. laouense</i> - <i>R. quezelii</i>	A0974	A0991	0.0004	0.0347
ptDNA	<i>R. laouense</i> - <i>R. quezelii</i>	A0974	A1072	0.0004	0.0347
ptDNA	<i>R. redieri</i> - <i>R. laouense</i>	A1063	A0973	0.0004	0.0484
ptDNA	<i>R. redieri</i> - <i>R. laouense</i>	A1063	A0974	0.0004	0.0484

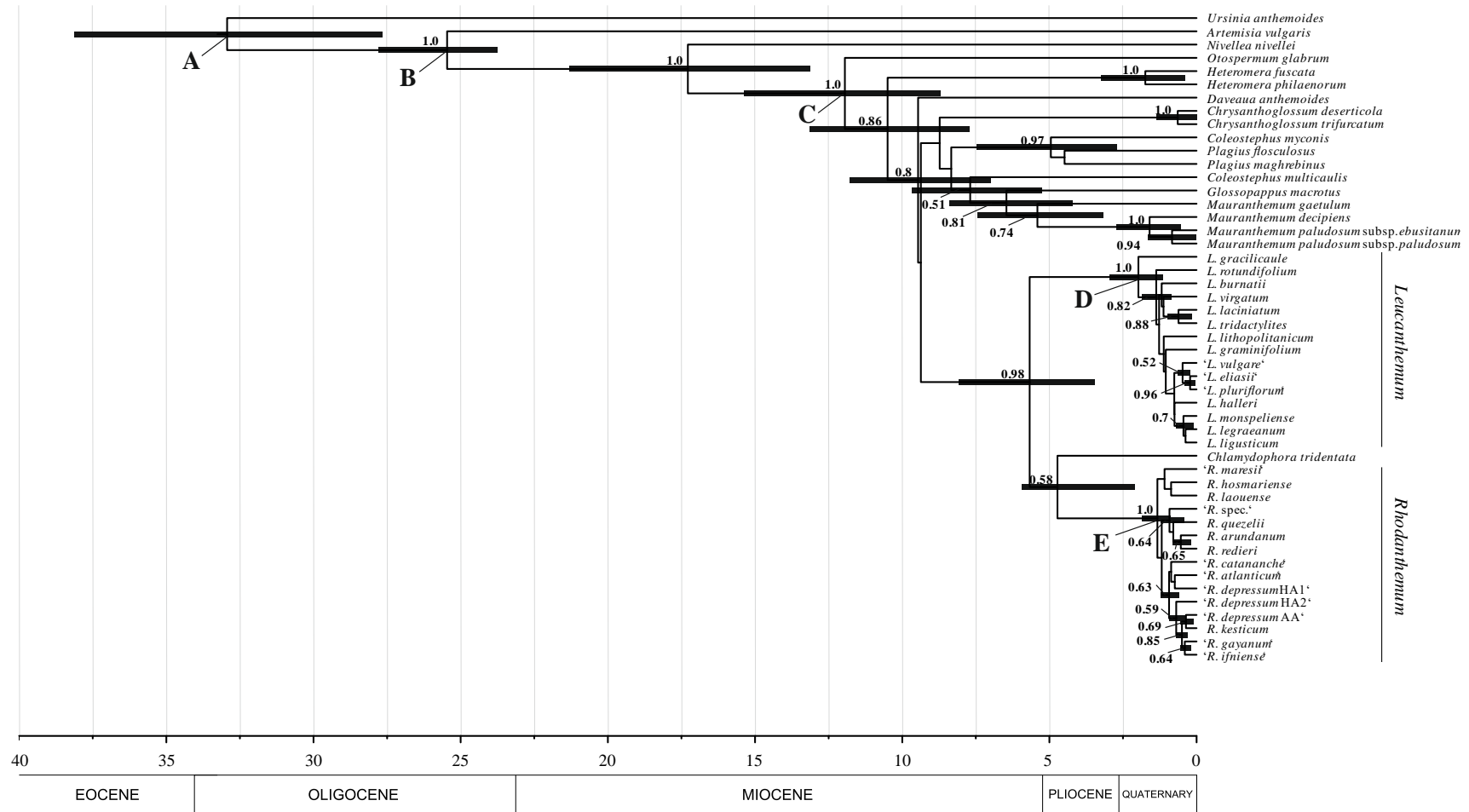


Figure 4.5 *BEAST chronogram inferred for the subtribe Leucanthemeinae based on four nuclear markers (*A39*, *B20*, *D27*, *ITS*), three plastid loci (*trnC-petN*, *trnL-trnF*, *trnQ-rps16*) and two calibration points (A, B). Speciation processes within the two target genera fall into the Quaternary, with a slightly older crown age for the genus *Leucanthemum* [D: 1.93 million years ago (Ma) (1.14–2.94 Ma)] compared with *Rhodanthemum* [E: 1.29 (0.88–1.87 Ma)]. All posterior probabilities > 0.5 are given and divergence time estimates as well as 95% highest posterior density (black bars) of important nodes (A–E) are summarized in Table 4.3.

4.4.2 Homoploid hybridization in the evolution of *Leucanthemum* and *Rhodanthemum*

Another longstanding theory concerning polyploidy-promoting factors is the idea that whole-genome doubling is linked with hybridization (Winge, 1917). In the present study, we investigated homoploid hybridization patterns in *Leucanthemum* and *Rhodanthemum* in a two-fold manner, by using the genealogical sorting index (*gsi*) in combination with posterior predictive checking via JML. The *gsi* has already been used as an indicator for gene-tree incongruence and potential hybridization events in the genus *Leucanthemum* (Konowalik et al., 2015) and in other plant groups like the southern African genus *Streptocarpus* (De Villiers et al., 2013). The latter authors provided three possible explanations for low *gsi* values and non-significance of association of alleles within a species: (i) the young age of a species resulting in a high degree of incomplete lineage sorting, (ii) the involvement of its members in hybridization events, and (iii) an incorrect definition of species boundaries. To rule out the last point, we conducted all *gsi* calculations on an assignment of accessions to species ('lineages') based on the results of species delimitation analyses in STACEY. For discriminating between incomplete lineage sorting and hybridization, coalescent simulations assuming no migration and subsequent comparison between empirical and simulated sequences were conducted via posterior predictive checking implemented in JML. This gene-centered method (i.e., a method asking whether a particular gene tree is expected under a given species tree) was successfully conducted in recent studies of hybridization in several genera of the Compositae family [e.g. *Picris* (Slovák et al., 2014); *Tolpis* (Gruenstaeudl et al., 2017); *Diplostephium* (Vargas et al., 2017), *Sclerorhachis* (Hassanpour et al., 2018)].

Our *gsi* and JML analyses unveil a contrasting pattern in the two genera *Leucanthemum* and *Rhodanthemum* concerning hybridization and incomplete lineage sorting: In spite of similar percentages of non-significance of association within lineages and equally low mean *gsi* values for all members of both plant groups, JML recovers considerably more hybridization events in *Leucanthemum* compared to *Rhodanthemum* (Figure 4.4). Hence, the topological inconsistencies among gene trees, resulting in similarly low *gsi* values for both genera, are rather explained by incomplete lineage sorting (ILS) alone in the case of *Rhodanthemum*, while both effects (ILS and hybridization) have to be considered as important mechanisms in the evolution of diploid representatives of *Leucanthemum*. This result fits with other phylogenetic studies of Mediterranean plants, where ILS and hybridization occur either alone or combined as drivers for incongruence among gene trees [see Table 8 in Blanco-Pastor et al. (2012)].

The hybridization signal among *Leucanthemum* diploids in the nuclear markers is in line with the study of Konowalik et al. (2015), where a similar sequence data set was analyzed with an

alternative hybrid-detection approach. By conducting coalescent simulations in the framework of a customized hybrid-index method based on likelihood calculations via PHYLONET (Than et al., 2008), the mentioned study found only five out of 19 diploid *Leucanthemum* species to be non-hybrid (*L. gaudinii*, *L. ageratifolium*, *L. monspeliense*, *L. graminifolium*, and *L. burnatii*). The fact that the present JML analyses show a weak, but measurable hybridization signal even for these taxa, may be explained by the difference between the two methods (hybrid-index calculations vs. JML) concerning the capability to detect asymmetrical hybridization. While taxon-centered methods like the hybrid index calculations of Konowalik et al. (2015) or network inference approaches like PHYLONET ask whether a particular OTU or ancestral branch in a phylogeny is of hybrid origin, gene-centred methods like JML queries whether a particular gene genealogy is expected under a given species tree. Hence, in the case of asymmetrical hybridization, gene-centered methods may outperform taxon-centered methods in the detection of hybrids that contain only few transmitted genes (Folk et al., 2018).

In summary, our study shows a high affinity for hybridization among diploid *Leucanthemum* lineages, which might be, in addition to the high genetic divergence in the genus, a plausible reason for the formation of a polyploid complex. This observation is in accordance to former considerations of Grant (1981), who hypothesized that naturally occurring hybridization is, besides a long-lived perennial growth habit and the existence of diploid species carrying different genomes or subgenomes, an important ‘primary factor’ for polyploid speciation. Conversely, the contrasting low hybridization signal in *Rhodanthemum*, with only three lineages showing a weak signal of horizontally transmitted plastid genes [possibly chloroplast capture events (Soltis and Kuzoff, 1995)] may explain the strictly diploid nature of this genus.

Table 4.3 Absolute divergence times and 95% highest posterior density (HPD) intervals for important nodes (C–E) in the evolution of the subtribe Leucantheinae from the *BEAST chronogram (see Figure 4.5), based on four nuclear markers (*A39*, *B20*, *D27*, ITS), three plastid loci (*trnC-petN*, *trnL-trnF*, *trnQ-rps16*) and two calibration points (A, B).

Node	Description	Prior distribution		Posterior distribution	
		Median age	95% HPD interval	Median age	95% HPD interval
A	Root age	33.78	28.85 - 38.72	32.86	27.66 - 38.12
B	Split between <i>Artemisia</i> and all accessions of the Euro-Mediterranean clade	25.43 (offset = 23.05)	24.10 - 28.47	25.22	23.74 - 27.79
C	crown age of Leucantheinae + <i>Daveaua</i> , <i>Heteromera</i> & <i>Otospermum</i>			11.86	8.71 - 15.38
D	crown age of <i>Leucanthemum</i>			1.93	1.14 - 2.94
E	crown age of <i>Rhodanthemum</i>			1.29	0.88 - 1.87

4.4.3 The role of climatic changes during the Quaternary

With a crown age of 1.9 Ma and a fast radiation around 1.3 Ma ago (see Figure 4.5), it seems likely that the diversification of *Leucanthemum* was influenced by climatic changes during the Quaternary over the past 1.6 Ma in Europe. It is well known that alternating glacial and interglacial periods during the Quaternary resulted in changes of distribution ranges of plant populations leading to homoploid and polyploid hybrid species formation (Klein and Kadereit, 2016; Marques et al., 2016; Folk et al., 2018). An appropriate scenario in this context was recently described by Kadereit (2015) based on a plethora of examples from different plant genera of the northern temperate regions, comprising (1) climate-induced range shifts of species during the Quaternary, (2) secondary contact of formerly allopatrically distributed species in refugial or re-colonized areas resulting in formation of interspecific hybrids, (3) re-colonization of the originally allopatric ranges by parental species, and (4) hybrids remaining in the area of secondary contact along with geographical isolation from their parents. The temporal placement of the *Leucanthemum* radiation into the Quaternary, the strong hybridization signal on the diploid level, and the high number of polyploid species makes the above mentioned spatial-temporal scenario conceivable for this genus. Another indicator is the present distribution of *Leucanthemum*, whose species cover many of the proposed refugial areas of Quaternary glaciations (Comes and Kadereit, 1998; Tribsch and Schönswetter, 2003; Schönswetter et al., 2005; Gómez and Lunt, 2006), such as the Carpathians, the Dinaric and Maritime Alps, as well as the southern European peninsulas. Furthermore, eco-climatological modelling for *Leucanthemum* representatives of the Iberian Peninsula indicated the existence of contact zones during the last glacial maximum (LGM) for most of the currently allopatrically distributed diploid lineages in that area [see Figure 7 in Oberprieler et al. (2014)]. Alternative scenarios like the radiation of several polyploid *Leucanthemum* species from the same ancestral event appear unlikely in the light of several preceding studies in the genus documenting the independent formation of polyploid species (e.g., Oberprieler et al., 2011a; Greiner et al., 2012, 2013; Oberprieler et al., 2014) and even the repeated formation of the same species under reciprocal parentage (e.g., Oberprieler et al., 2018).

In *Rhodanthemum*, we found no evidence for a reticulate history in the above described manner, in spite of a similar crown age of the genus compared to *Leucanthemum*. One possible reason for the different impact of Quaternary climatic changes on *Leucanthemum* and *Rhodanthemum* species concerning hybridization is possibly connected with the contrasting distribution patterns of both genera (Figure 4.1). Occupying almost exclusively higher elevations of the Rif, Middle Atlas, High Atlas, and Anti-Atlas mountains of Morocco, *Rhodanthemum* populations may have compensated climatic shifts during the Pleistocene

mainly by vertical migration as expected for mountain-dwelling organisms (Guralnick, 2007; García-Aloy et al., 2017). Adaptation for climatic changes causing elevational rather than latitudinal shifts might have prevented secondary contact among *Rhodanthemum* species and thus may have resulted in a low homoploid hybridization rate and the lack of polyploidy in the genus. Conversely, *Leucanthemum* occupies lowland and montane habitats throughout Europe and changing environments during the Pleistocene were probably accompanied by latitudinal and elevational shifts of populations in the genus. As indicated above, secondary contact of formerly allopatrically distributed species in refugial or re-colonized areas may have led to the pronounced homoploid hybridization pattern and the formation of a polyploid complex. The latter phenomenon was presumably supported by the high level of differentiation among lineages in the genus.

4.5 Supplemental Figures and Tables

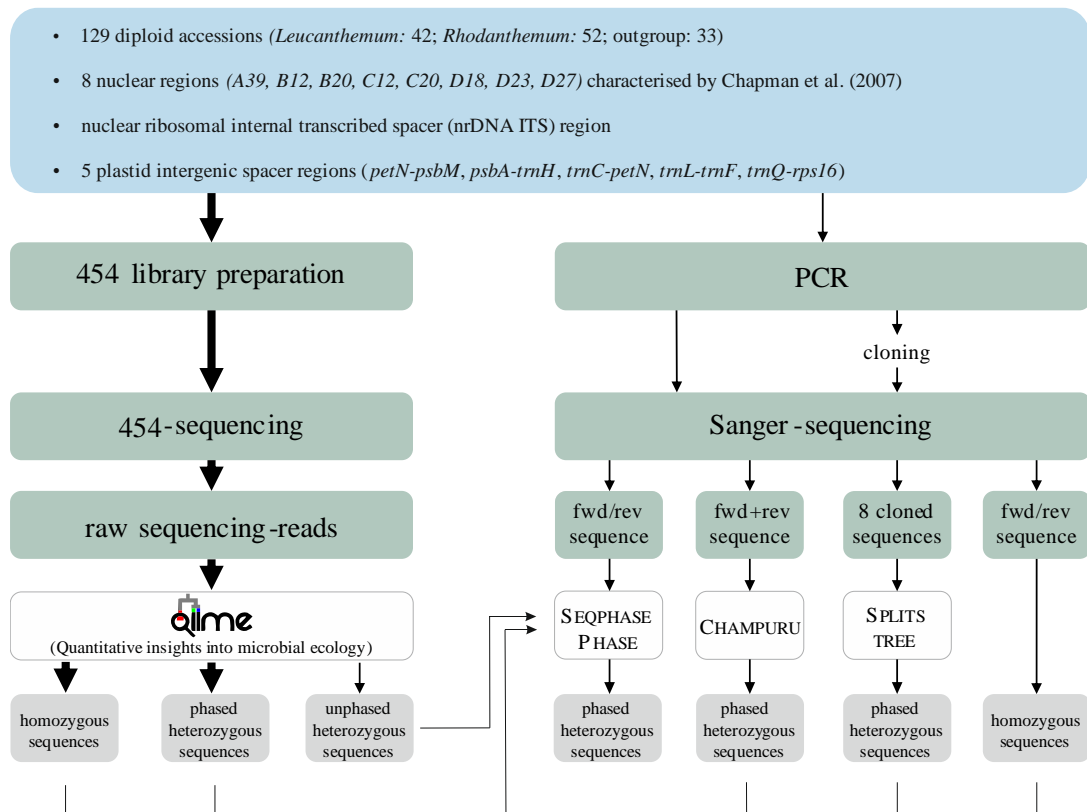


Figure S4.1 Processing of 454 and Sanger sequence data in the course of the underlying study. The amount of accessions passed through the different stages of the workflow is indicated by the thickness of the arrows.

Table S4.1 Accessions used in the present study including information on lineage assignment, location, collector and herbarium voucher. Asterisks (*), crosses (†), arrows (↓) and hashes (#) refer to sequences from Konowalik et al. (2015), Wagner et al. (2017), Himmelreich et al. (2008) and Oberprieler and Vogt (2000). Countries are abbreviated according to the ISO 3166 standard.

Taxon	Lineage	ID	Geographic location	Coord.	Collector	Voucher	A39	B12	B20	C12	C20	D18	D23	D27	ITS	trnL-trnF	trnC-petN	psbA-trnH	petN-psbM	trnQ-rps16
<i>Artemisia vulgaris</i> L.	<i>Artemisia vulgaris</i>	A0838	DE, Bavaria, Regensburg	49.02 N, 12.10 E	Konowalik s.n.	WRS1	LN869206*	–	MK465292 MK465293	–	–	–	–	MK465207	MK481440 MK481441 MK481442	LN869056*	LN869106*	–	–	LN869205*
<i>Chlamydomorpho tridentata</i> (Delile) Less.	<i>Chlamydomorpho tridentata</i>	A1020	CY, Larnaka, Cape Greco	34.97 N, 34.07 E	Vogt 8005	B 10 0673090	SAMN10845483	–	SAMN10845483	–	–	–	–	SAMN10845483	MK481510	MK481635	MK481717	–	–	MK481853
<i>Chlamydomorpho tridentata</i> (Delile) Less.		A0795	CY, Larnaka, Meneou	34.85 N, 33.61 E	Vogt 8120	B 10 0550374	ERS758365*	–	ERS758365*	–	–	–	–	ERS758365*	MK481442	LN869008*	LN869058*	–	–	MK481808
<i>Chrysanthoglossum deserticola</i> (Murb.) B. H. Wilcox & al.	<i>Chrysanthoglossum deserticola</i>	A1012	TN, Tozeur, Kariz.	34.05 N, 08.24 E	Vogt 16586, Oberprieler 10529 & Gst001	B 10 0673095	SAMN10845480	–	SAMN10845480	–	–	–	–	SAMN10845480	MK481507	MK481632	MK481714	–	–	MK481850
<i>Chrysanthoglossum deserticola</i> (Murb.) B. H. Wilcox & al.		A0791	TN, Tataouine, Tataouine - Remada, 450 m	32.50 N, 10.27 E	Vogt 13038 & Oberprieler 7343	B 10 0550241	ERS758364*	–	ERS758364*	–	–	–	–	ERS758364*	MK481430 MK481431	MK481591	LN869057*	–	–	LN869157*
<i>Chrysanthoglossum trifurcatum</i> (Desf.) B. H. Wilcox & al.	<i>Chrysanthoglossum trifurcatum</i>	A1015	LY, Tripolis, Tagiura	32.83 N, 13.38 E	Boramüller 1933	B 10 0673093	SAMN10845481	–	SAMN10845481	–	–	–	–	SAMN10845481	MK481508	MK481633	MK481715	–	–	MK481851
<i>Chrysanthoglossum trifurcatum</i> (Desf.) B. H. Wilcox & al.		A1017	LY, Tripolis	32.88 N, 13.19 E	Baschant s.n.	B 10 0673091	SAMN10845482	–	SAMN10845482	–	–	–	–	–	SAMN10845482	MK481509	MK481634	MK481716	–	–
<i>Coleostephus multicaulis</i> (Desf.) Durieu	<i>Coleostephus multicaulis</i>	A1000	N/A (BG Copenhagen, cultivated since 1894)	N/A	N/A (cult. hort. bot. Berlin-Dahlem, Vogt s.n., 9.7.1992)	B 10 0673089	SAMN10845476	–	SAMN10845476	–	–	–	–	SAMN10845476	MK48150	MK481628	MK481710	–	–	MK481846
<i>Coleostephus multicaulis</i> (Desf.) Durieu		A1042	DZ, Wilaya de Djelfa, 1200 m	34.63 N, 03.07 E	Dubuis, Maurel & Rhamoun 1855	MSB	SAMN10845487	–	SAMN10845487	–	–	–	–	–	SAMN10845487	MK481514	MK481639	MK481721	–	–
<i>Coleostephus myconis</i> (L.) Rchb. f.	<i>Coleostephus myconis</i>	A0996	TN, Jendouba, Babouch	36.80 N, 08.68 E	Vogt 13703 & Oberprieler 8008	B 10 0673102	SAMN10845474	–	SAMN10845474	–	–	–	–	SAMN10845474	MK481498	MK481626	MK481708	–	–	MK481844
<i>Coleostephus myconis</i> (L.) Rchb. f.		A0792	IT, Calabria, Gallico-Gambarie	38.17 N, 15.74 E	Vogt 13976 & Oberprieler 8281	B 10 0550362	ERS758366*	–	ERS758366*	–	–	–	–	ERS758366*	MK481432	LN869009*	MK481670	–	–	MK481806
<i>Daveaua anthemoides</i> Mariz	<i>Daveaua anthemoides</i>	A1113	ES, Cáceres, Las Chamizas, 425 m	39.28 N, 05.55 W	Rico ER-8082	SALA 143862	SAMN10845502	–	SAMN10845502	–	–	–	–	SAMN10845502	MK481538	MK481654	MK481736	–	–	MK481872
<i>Daveaua anthemoides</i> Mariz		A1114	ES, Cáceres, Las Chamizas	39.28 N, 05.55 W	Rico s.n.	SALA 135877	SAMN10845503	–	SAMN10845503	–	–	–	–	–	SAMN10845503	MK481539 MK481540	MK481655	MK481737	–	–
<i>Glossopappus macrotus</i> (Durieu) Briq. & Cavill.	<i>Glossopappus macrotus</i>	A0998	ES, Andalucía, Coin	36.66 N, 04.76 W	Prem s.n.	B 10 0673101	SAMN10845475	–	SAMN10845475	–	–	–	–	SAMN10845475	MK481499	MK481627	MK481709	–	–	MK481845
<i>Glossopappus macrotus</i> (Durieu) Briq. & Cavill.		A0790	MA, Sefrou, Immouzer du Kandar, 1090 m	33.80 N, 04.99 W	Vogt 12028	B 10 0550375	ERS758367*	–	ERS758367*	–	–	–	–	ERS758367*	MK481429	MK481590	LN869060*	–	–	LN869159*
<i>Heteromera fuscata</i> (Desf.) Pomel	<i>Heteromera fuscata</i>	A0937	TN, Gabès, Toujane, 469 m	33.48 N, 10.06 E	Vogt 16547, Oberprieler 10490 & Gst001	B 10 0673058	SAMN10845443	–	SAMN10845443	–	–	–	–	SAMN10845443	MK481444 MK481445	MK481595	MK481677	–	–	MK481813
<i>Heteromera fuscata</i> (Desf.) Pomel		A0796	TN, Tozeur, 65 m	34.05 N, 08.24 E	Vogt 16585, Oberprieler 10528 & Gst001	B 10 0216212	ERS758368*	–	ERS758368*	–	–	–	–	ERS758368*	MK481435	LN869011*	MK481673	–	–	MK481809
<i>Heteromera philaenorum</i> Maire & Weiller	<i>Heteromera philaenorum</i>	A0944	TN, Tataouine, Bir Thlethine, 420 m	32.65 N, 10.32 E	Vogt 13007 & Oberprieler 7312	B 10 0673044	SAMN10845444	–	SAMN10845444	–	–	–	–	SAMN10845444	MK481446 MK481447	MK481596	MK481678	–	–	MK481814
<i>Leucanthemum burnatii</i> Briq. & Cavill.	<i>Leucanthemum burnatii</i>	90-6	FR, Provence-Alpes-Côte d'Azur, Grasse, 1235 m	43.76 N, 06.92 E	Vogt 16615, Oberprieler 10566 & Konowalik	B 10 0464678	ERS758374*	ERS758374*	ERS758374*	ERS758374*	ERS758374*	ERS758374*	ERS758374*	ERS758374*	MK481418	LN869017*	LN869067*	LN868967*	LN869117*	LN869166*
<i>Leucanthemum burnatii</i> Briq. & Cavill.		92-1	FR, Provence-Alpes-Côte d'Azur, Mgne Ste-Victoire, 650–750 m	43.55 N, 05.66 E	Vogt 16618, Oberprieler 10569 & Konowalik	B 10 0464676	ERS758375*	ERS758375*	ERS758375*	ERS758375*	ERS758375*	ERS758375*	ERS758375*	ERS758375*	ERS758375*	MK481419	LN869018*	LN869068*	LN868968*	LN869118*
<i>Leucanthemum gracilicaule</i> (Dufour) Pau	<i>Leucanthemum gracilicaule</i>	84-6	ES, Valencia, Benirrama, 296 m	38.84 N, 00.19 W	Konowalik KK20 & Ogrodowczyk	B 10 0386704	ERS758384*	ERS758384*	ERS758384*	ERS758384*	ERS758384*	ERS758384*	ERS758384*	ERS758384*	MK481416	LN869029*	LN869079*	LN868979*	LN869129*	LN869178*
<i>Leucanthemum gracilicaule</i> (Dufour) Pau		85-1	ES, Valencia, Altury, 337 m	39.31 N, 00.68 W	Konowalik KK25 & Ogrodowczyk	B 10 0386702	ERS758385*	ERS758385*	ERS758385*	ERS758385*	ERS758385*	ERS758385*	ERS758385*	ERS758385*	MK481417	LN869030*	LN869080*	LN868980*	LN869130*	LN869179*

Table S4.1 Continued.

Taxon	Lineage	ID	Geographic location	Coord.	Collector	Voucher	A39	B12	B20	C12	C20	D18	D23	D27	ITS	trnL-trnF	trnC-petN	psbA-trnH	petN-psbM	trnQ-rps16
<i>Leucanthemum graminifolium</i> (L.) Lam.	<i>Leucanthemum graminifolium</i>	116-4	FR, Languedoc-Roussillon, Roqueredonde, 802 m	43.78 N, 03.24 E	Vogt 16693, Oberprieler 10648 & Konowalik Vogt 16656,	B 10 0464684	ERS758386*	ERS758386*	ERS758386*	ERS758386*	ERS758386*	ERS758386*	ERS758386*	ERS758386*	MK481381 MK481382	LN869031*	LN869081*	LN868981*	LN869131*	LN869180*
<i>Leucanthemum graminifolium</i> (L.) Lam.		96-3	FR, Languedoc-Roussillon, Roc de L'Aigle, 560–600 m	43.15 N, 02.63 E	Oberprieler 10607 & Konowalik	B 10 0464663	ERS758387*	ERS758387*	ERS758387*	ERS758387*	ERS758387*	ERS758387*	ERS758387*	ERS758387*	ERS758387*	MK481421 MK481422	LN869032*	LN869082*	LN868982*	LN869132*
<i>Leucanthemum halleri</i> (Vitman) Ducoum	<i>Leucanthemum halleri</i>	L1002	AT, Tirol, Tannheim, 1840 m	47.51 N, 10.60 E	Vogt 16874	B 10 0420901	ERS758388*	ERS758388*	ERS758388*	ERS758388*	ERS758388*	ERS758388*	ERS758388*	ERS758388*	MK481552 MK481553 MK481554	LN869033*	LN869083*	LN868983*	LN869133*	LN869182*
<i>Leucanthemum halleri</i> (Vitman) Ducoum		208-1	CH, Valais, Sion, 2320 m	46.33 N, 07.29 E	Tomasello TS65	B 10 0386672	ERS758389*	ERS758389*	ERS758389*	ERS758389*	ERS758389*	ERS758389*	ERS758389*	ERS758389*	ERS758389*	MK481390 MK481391 MK481392	LN869034*	LN869084*	LN868984*	LN869134*
<i>Leucanthemum luciniatum</i> Huter & al.	<i>Leucanthemum luciniatum</i>	L179	IT, Basilicata, Castrovillari, 1900–2100 m	39.91 N, 16.19 E	Vogt 15614	B 10 0420805	ERS758390*	ERS758390*	ERS758390*	ERS758390*	ERS758390*	ERS758390*	ERS758390*	ERS758390*	MK481558 MK481559	LN869035*	LN869085*	LN868985*	LN869135*	LN869184*
<i>Leucanthemum luciniatum</i> Huter & al.		280-1	IT, Calabria, Colle del Drognone, 1580 m	39.90 N, 16.11 E	Tomasello TS420	B 10 0464203	ERS758391*	ERS758391*	ERS758391*	ERS758391*	ERS758391*	ERS758391*	ERS758391*	ERS758391*	ERS758391*	MK481405 MK481406	LN869036*	LN869086*	LN868986*	LN869136*
<i>Leucanthemum legracianum</i> (Rouy) B. Bock & J.-M. Tison	<i>Leucanthemum legracianum</i>	366-1	FR, Provence-Alpes-Côte d'Azur, Massif des Maures, 410 m	43.20 N, 06.31 E	Vogt 17189	B 10 0486634	SAMN10845436	SAMN10845436	SAMN10845436	SAMN10845436	SAMN10845436	SAMN10845436	SAMN10845436	SAMN10845436	MK481407	KY778058†	KY778096†	KY778077†	KY778020†	KY778039†
<i>Leucanthemum legracianum</i> (Rouy) B. Bock & J.-M. Tison		369-1	FR, Provence-Alpes-Côte d'Azur, Massif des Maures, 210 m	43.24 N, 06.34 E	Vogt 17192	B 10 0486648	SAMN10845437	SAMN10845437	SAMN10845437	SAMN10845437	SAMN10845437	SAMN10845437	SAMN10845437	SAMN10845437	SAMN10845437	MK481408	KY778059†	KY778097†	KY778078†	KY778021†
<i>Leucanthemum ligusticum</i> Marchetti & al.	<i>Leucanthemum ligusticum</i>	406-1	IT, Liguria, Rochetta di Vara,	44.25 N, 09.77 E	Vogt 17460, Oberprieler 10941 & Wagner	B 10 0627838	SAMN10845438	SAMN10845438	SAMN10845438	SAMN10845438	SAMN10845438	SAMN10845438	SAMN10845438	SAMN10845438	MK481411	KY778061†	KY778099†	KY778080†	KY778023†	KY778042†
<i>Leucanthemum ligusticum</i> Marchetti & al.		416-1	IT, Liguria, Ponte di Lagoscuro, 246 m	44.25 N, 09.46 E	Vogt 17471, Oberprieler 10952 & Wagner	B 10 0627855	SAMN10845439	SAMN10845439	SAMN10845439	SAMN10845439	SAMN10845439	SAMN10845439	SAMN10845439	SAMN10845439	MK481412	KY778063†	KY778101†	KY778082†	KY778025†	KY778044†
<i>Leucanthemum lithopolitanicum</i> (E. Mayer) Polatschek	<i>Leucanthemum lithopolitanicum</i>	L998	SI, Kamnik, Kamniška Bistrica, 1880–2120 m	46.35 N, 14.61 E	Hörandl, Hadaček, M. & jun. s.n.	W 1999-3533	ERS758393*	ERS758393*	ERS758393*	ERS758393*	ERS758393*	ERS758393*	ERS758393*	ERS758393*	MK481570 MK481571	LN869037*	LN869087*	LN868987*	LN869137*	LN869186*
<i>Leucanthemum lithopolitanicum</i> (E. Mayer) Polatschek		274-1	AT, Kärnten, Lesnik, 1999 m	46.38 N, 14.57 E	Oberprieler 10864	B 10 0413013	ERS758394*	ERS758394*	ERS758394*	ERS758394*	ERS758394*	ERS758394*	ERS758394*	ERS758394*	ERS758394*	MK481397 MK481398 MK481399	LN869038*	LN869088*	LN868988*	LN869138*
<i>Leucanthemum monspeliense</i> (L.) H. J. Coste	<i>Leucanthemum monspeliense</i>	131-20	FR, Languedoc-Roussillon, St.-André-de-Valborgne, 380 m	44.14 N, 03.73 E	Vogt 16716, Oberprieler 10671 & Konowalik Vogt 16712,	B 10 0464615	ERS758395*	ERS758395*	ERS758395*	ERS758395*	ERS758395*	ERS758395*	ERS758395*	ERS758395*	MK481384 MK481385	LN869019*	LN869069*	LN868969*	LN869119*	LN869168*
<i>Leucanthemum monspeliense</i> (L.) H. J. Coste		128-1	FR, Languedoc-Roussillon, l'Espéron, 750 m	44.09 N, 03.58 E	Oberprieler 10667 & Konowalik	B 10 0464618	ERS758396*	ERS758396*	ERS758396*	ERS758396*	ERS758396*	ERS758396*	ERS758396*	ERS758396*	ERS758396*	MK481383	LN869020*	LN869070*	LN868970*	LN869120*
<i>Leucanthemum rotundifolium</i> (Willd.) DC.	<i>Leucanthemum rotundifolium</i>	L990	RO, Prahova, Buzeni, 1000–1500 m	45.42 N, 25.51 E	Hörandl 9063, Hadaček & Costea	W1999-05366	ERS758399*	ERS758399*	ERS758399*	ERS758399*	ERS758399*	ERS758399*	ERS758399*	ERS758399*	MK481564	LN869041*	LN869091*	LN868991*	LN869141*	LN869190*
<i>Leucanthemum rotundifolium</i> (Willd.) DC.		L989	BA, Fojnica, Paljike, 1800 m	43.95 N, 17.75 E	Horvat s.n.	ZA	ERS758400*	ERS758400*	ERS758400*	ERS758400*	ERS758400*	ERS758400*	ERS758400*	ERS758400*	ERS758400*	MK481563 MK481565	LN869042*	LN869092*	LN868992*	LN869142*
<i>Leucanthemum rotundifolium</i> (Willd.) DC.	L992	PL, Podkarpackie, Zakopane, 1290 m	49.26 N, 19.92 E	Jasiewicz & Peřkos s.n.	W 1970-12192	ERS758401*	ERS758401*	ERS758401*	ERS758401*	ERS758401*	ERS758401*	ERS758401*	ERS758401*	ERS758401*	MK481566 MK481567	LN869043*	LN869093*	LN868993*	LN869143*	LN869192*
<i>Leucanthemum tridactylites</i> (A. Kern. & Huter) Huter & al.	<i>Leucanthemum tridactylites</i>	L151	IT, Abruzzo, Passo di San Leonardo, 1500–1800 m	42.08 N, 14.03 E	Vogt 14050 & Oberprieler 8355	B 10 0420849	ERS758402*	ERS758402*	ERS758402*	ERS758402*	ERS758402*	ERS758402*	ERS758402*	ERS758402*	MK481555 MK481556	LN869044*	LN869094*	LN868994*	LN869144*	LN869193*
<i>Leucanthemum tridactylites</i> (A. Kern. & Huter) Huter & al.		278-1	IT, Abruzzo, Blockhaus alla Fonte dell'Acquaviva, 2080 m	42.14 N, 14.11 E	Tomasello TS417	B 10 0464207	ERS758403*	ERS758403*	ERS758403*	ERS758403*	ERS758403*	ERS758403*	ERS758403*	ERS758403*	ERS758403*	MK481402 MK481403 MK481404	LN869045*	LN869095*	LN868995*	LN869145*
<i>Leucanthemum virgatum</i> (Des.) Clos	<i>Leucanthemum virgatum</i>	L987	FR, Alpes Maritimes, Vésubie, 1013 m	43.98 N, 07.27 E	Saatkamp s.n.	B 10 0603654	ERS758404*	ERS758404*	ERS758404*	ERS758404*	ERS758404*	ERS758404*	ERS758404*	ERS758404*	MK481561 MK481562	LN869048*	LN869098*	LN868998*	LN869148*	LN869197*

Table S4.1 Continued.

Taxon	Lineage	ID	Geographic location	Coord.	Collector	Voucher	A39	B12	B20	C12	C20	D18	D23	D27	ITS	trnL-trnF	trnC-petN	psbA-trnH	petN-psbM	trnQ-rps16
<i>Leucanthemum virgatum</i> (Des.) Clos	<i>Leucanthemum virgatum</i>	250-1	IT, Liguria, Pogli to Onzo, 215 m	44.06 N, 08.06 E	Vogt 16932 & Oberprieler 10839	B 10 0350169	ERS758405*	ERS758405*	ERS758405*	ERS758405*	ERS758405*	ERS758405*	ERS758405*	ERS758405*	MK481393 MK481394	LN869049*	LN869099*	LN868999*	LN869149*	LN869198*
<i>Leucanthemum vulgare</i> Lam.	<i>Leucanthemum vulgare</i>	94-1	FR, Languedoc-Roussillon, Montaur, 160 m	43.13 N, 02.61 E	Vogt 16641, Oberprieler 10592 & Konowalik	B 10 0464674	ERS758406*	ERS758406*	ERS758406*	ERS758406*	ERS758406*	ERS758406*	ERS758406*	ERS758406*	MK481420	LN869050*	LN869100*	LN869000*	LN869150*	LN869199*
<i>Leucanthemum vulgare</i> Lam.		L046	DE, Bayern, Pittmannsdorf, 450 m	49.03 N, 11.88 E	Eder & Oberprieler s.n.	B 10 0550249	ERS758407*	ERS758407*	ERS758407*	ERS758407*	ERS758407*	ERS758407*	ERS758407*	ERS758407*	MK481551	LN869051*	LN869101*	LN869001*	LN869151*	LN869200*
<i>Leucanthemum vulgare</i> Lam.		184-1	BA, Gacko, Ribari, 930 m	43.24 N, 18.34 E	Vogt 16806 & Prem-Vogt	B 10 0346626	ERS758408*	ERS758408*	ERS758408*	ERS758408*	ERS758408*	ERS758408*	ERS758408*	ERS758408*	MK481389	LN869052*	LN869102*	LN869002*	LN869152*	LN869201*
<i>Leucanthemum pyrenaicum</i> Vogt, Konowalik & Oberpr		L035	ES, Catalunya, Punta Brulle, 2350–2500 m	42.58 N, 01.00 E	Vogt 5125 & Prem	B 10 0216900	ERS758378*	ERS758378*	ERS758378*	ERS758378*	ERS758378*	ERS758378*	ERS758378*	ERS758378*	MK481548 MK481549	LN869021*	LN869071*	LN868971*	LN869121*	LN869170*
<i>Leucanthemum pyrenaicum</i> Vogt, Konowalik & Oberpr		266-1	ES, Aragon, Balneario de Panticoosa, 2150 m	42.78 N, 00.23 W	Tomasello TS382	B 10 0464208	ERS758379*	ERS758379*	ERS758379*	ERS758379*	ERS758379*	ERS758379*	ERS758379*	ERS758379*	MK481395 MK481396	LN869022*	LN869072*	LN868972*	LN869122*	LN869171*
<i>Leucanthemum gaudinii</i> Dalla Torre		L033	SK, Prešovský kraj, Siroké sedlo, 1700 m	49.25 N, 20.23 E	Knoph & Schrüfer s.n.	B 10 0216898	ERS758382*	ERS758382*	ERS758382*	ERS758382*	ERS758382*	ERS758382*	ERS758382*	ERS758382*	MK481547	LN869025*	LN869075*	LN868975*	LN869125*	LN869174*
<i>Leucanthemum gaudinii</i> Dalla Torre		276-1	AT, Kärnten, Falkert, 2270 m	46.86 N, 13.82 E	Oberprieler 10866	B 10 0413015	ERS758383*	ERS758383*	ERS758383*	ERS758383*	ERS758383*	ERS758383*	ERS758383*	ERS758383*	MK481400 MK481401	LN869026*	LN869076*	LN868976*	LN869126*	LN869175*
<i>Leucanthemum aegeratfolium</i> Pau		135-7	FR, Pyrénées-Orientales, La Vallée Heureuse, 410 m	42.50 N, 02.96 E	Konowalik KK42 & Ogródowczyk	B 10 0386712	ERS758411*	ERS758411*	ERS758411*	ERS758411*	ERS758411*	ERS758411*	ERS758411*	ERS758411*	MK481386	LN869054*	LN869104*	LN869004*	LN869154*	LN869203*
<i>Leucanthemum aegeratfolium</i> Pau		M60-1	ES, Castilla-La Mancha, Salinas de Manzano, 1157 m	40.10 N, 01.52 W	Cordel s.n.	B 10 0345012	ERS758412*	ERS758412*	ERS758412*	ERS758412*	ERS758412*	ERS758412*	ERS758412*	ERS758412*	MK481572 MK481573	LN869055*	LN869105*	LN869005*	LN869155*	LN869204*
<i>Leucanthemum eliasii</i> (Sennen & Pau) Vogt, Konowalik & Oberpr.		L996	ES, Burgos, San Pantaleón del Páramo, 973 m	42.56 N, 03.80 W	Cela 1433 & Lopez	B 10 0420857	ERS758409*	ERS758409*	ERS758409*	ERS758409*	ERS758409*	ERS758409*	ERS758409*	ERS758409*	MK481568 MK481569	LN869046*	LN869096*	LN868996*	LN869146*	LN869195*
<i>Leucanthemum eliasii</i> (Sennen & Pau) Vogt, Konowalik & Oberpr.	L162	ES, Burgos, Ubierna, 887 m	42.50 N, 03.70 W	Cela 465PG & Martin	B 10 0420851	ERS758410*	ERS758410*	ERS758410*	ERS758410*	ERS758410*	ERS758410*	ERS758410*	ERS758410*	MK481557	LN869047*	LN869097*	LN868997*	LN869147*	LN869196*	
<i>Leucanthemum cucuminis</i> Vogt, Konowalik & Oberpr.	L036	ES, Cantabria, Pozas de Lloroza, 1830 m	43.13 N, 04.75 W	Bayón 2132, Llanquiza & Villanueva	B 10 0420752	ERS758380*	ERS758380*	ERS758380*	ERS758380*	ERS758380*	ERS758380*	ERS758380*	ERS758380*	MK481550	LN869023*	LN869073*	LN868973*	LN869123*	LN869172*	
<i>Leucanthemum cucuminis</i> Vogt, Konowalik & Oberpr.	60-1	ES, Galicia, Piomedo, 1530 m	42.83 N, 06.86 W	Höbl 60	B 10 0413746	ERS758381*	ERS758381*	ERS758381*	ERS758381*	ERS758381*	ERS758381*	ERS758381*	ERS758381*	MK481415	LN869024*	LN869074*	LN868974*	LN869124*	LN869173*	
<i>Leucanthemum galleacium</i> Rodr. Oubiña & S. Ortiz	159-11	ES, Galicia, Sierra de Basadre, 375 m	42.85 N, 07.99 W	Konowalik, Rodríguez Oubiña & Ortiz s.n.	B 10 0386789	ERS758376*	ERS758376*	ERS758376*	ERS758376*	ERS758376*	ERS758376*	ERS758376*	ERS758376*	MK481387 MK481388	LN869027*	LN869077*	LN868977*	LN869127*	LN869176*	
<i>Leucanthemum galleacium</i> Rodr. Oubiña & S. Ortiz	L985	ES, Galicia, Paradelá, 672 m	42.98 N, 07.92 W	Rodríguez Oubiña s.n.	no voucher	ERS758377*	ERS758377*	ERS758377*	ERS758377*	ERS758377*	ERS758377*	ERS758377*	ERS758377*	MK481560	LN869028*	LN869078*	MK481794	LN869128*	LN869177*	
<i>Leucanthemum pluriflorum</i> Pau	40-6	ES, Galicia, Cabo Fisterra, 100 m	42.88 N, 09.27 W	Höbl 40	B 10 0413758	ERS758397*	ERS758397*	ERS758397*	ERS758397*	ERS758397*	ERS758397*	ERS758397*	ERS758397*	MK481409 MK481410	LN869039*	LN869089*	LN868989*	LN869139*	LN869188*	
<i>Leucanthemum pluriflorum</i> Pau	55-1	ES, Galicia, Cangas de Foz, 10 m	43.63 N, 07.33 W	Höbl 55	B 10 0413749	ERS758398*	ERS758398*	ERS758398*	ERS758398*	ERS758398*	ERS758398*	ERS758398*	ERS758398*	MK481413 MK481414	LN869040*	LN869090*	LN868990*	LN869140*	LN869189*	
<i>Mauranthemum decipiens</i> (Pomel) Vogt & Oberpr.	Mauranthemum decipiens	A0048	MA, Berkane, Monts de Beni-Snassen, 850-900 m	34.82 N, 02.40 W	Vogt 10748 & Oberprieler 5196	B 10 0673084	SAMN10845440	–	SAMN10845440	–	–	–	–	SAMN10845440	MK481423	MK481586	MK481666	–	–	MK481802
<i>Mauranthemum decipiens</i> (Pomel) Vogt & Oberpr.	Mauranthemum decipiens	A1004	MA, Nador, Selouane - Berkane, 70 m	34.93 N, 02.63 W	Bayón, Oberprieler & Vogt 5247	B 10 0673085	SAMN10845477	–	SAMN10845477	–	–	–	–	SAMN10845477	MK481501 MK481502	MK481629	MK481711	–	–	MK481847
<i>Mauranthemum gaetulum</i> (Batt.) Vogt & Oberpr.	Mauranthemum gaetulum	A1005	MA, Er-Rachidia, Tunnel de Legionnaire, 1250 m	32.18 N, 04.35 W	Bayón, Oberprieler, Vogt 5400	B 10 0673086	SAMN10845478	–	SAMN10845478	–	–	–	–	SAMN10845478	MK481503 MK481504	MK481630	MK481712	–	–	MK481848
<i>Mauranthemum gaetulum</i> (Batt.) Vogt & Oberpr.	Mauranthemum gaetulum	A1056	MA, d'Ourazate, Zagora, Jbel Adafane, 900-1009 m	30.33 N, 05.80 W	Ewald & Schuberwerk 90881	B 10 0673087	SAMN10845489	–	SAMN10845489	–	–	–	–	SAMN10845489	MK481517 MK481518	MK481641	MK481723	–	–	MK481859
<i>Mauranthemum paludosum</i> (Poir.) Vogt & Oberpr. subsp. <i>paludosum</i>	Mauranthemum paludosum subsp. <i>paludosum</i>	A1008	ES, Alicante, Jesus Pobre - Denia, 300 m	38.81 N, 00.09 E	Pedrol & Vogt 2982	B 10 0673099	SAMN10845479	–	SAMN10845479	–	–	–	–	SAMN10845479	MK481505 MK481506	MK481631	MK481713	–	–	MK481849
<i>Mauranthemum paludosum</i> (Poir.) Vogt & Oberpr. subsp. <i>paludosum</i>	Mauranthemum paludosum subsp. <i>paludosum</i>	A0798	ES, Valencia, Javea - El Mongio, 210 m	38.80 N, 00.16 E	Konowalik KK9 & Ogródowczyk	WRSL	ERS758370*	–	ERS758370*	–	–	–	–	ERS758370*	MK481436 MK481437	LN869013*	MK481674	–	–	MK481810

Table S4.1 Continued.

Taxon	Lineage	ID	Geographic location	Coord.	Collector	Voucher	A39	B12	B20	C12	C20	D18	D23	D27	ITS	trnL-trnF	trnC-petN	psbA-trnH	petN-psbM	trnQ-rps16
<i>Mauranthemum paludosum</i> subsp. <i>ebastianum</i> (Vogt) Vogt & Oberpr.	<i>Mauranthemum paludosum</i> subsp. <i>ebastianum</i>	A0799	ES, Ibiza, Atalaria de Sant Josep, 163 m	38.92 N, 01.26 E	Lloréns s.n.	B 10 0550369	ERS758369*	–	ERS758369*	–	–	–	–	ERS758369*	MK481438 MK481439	MK481593	MK481675	–	–	MK481811
<i>Nivellea nivellei</i> (Braun-Blanquet & Maire) B. H. Wilcox & al.	<i>Nivellea nivellei</i>	A1111	MA, Ifrane, Azrou, 1550 m	33.37 N, 05.25 W	Vogt 9418 & Oberprieler 3856	Herb. Oberprieler	SAMN10845501	–	SAMN10845501	–	–	–	–	SAMN10845501	MK481537	MK481653	MK481735	–	–	MK481871
<i>OospERMum glabrum</i> (Lag.) Wilk.	<i>OospERMum glabrum</i>	A1024	MA, Fès, Moulay-Yakoub, 400 m	34.50 N, 06.23 W	Bayón, Oberprieler & Vogt 5984	Herb. Oberprieler	SAMN10845484	–	SAMN10845484	–	–	–	–	SAMN10845484	MK481511	MK481636	MK481718	–	–	MK481854
<i>OospERMum glabrum</i> (Lag.) Wilk.	<i>OospERMum glabrum</i>	A1053	ES, Cádiz, Algeciras - Cádiz, 320 m	36.31 N, 05.81 W	Vogt 9282 & Vogt	Herb. Oberprieler	SAMN10845488	–	SAMN10845488	–	–	–	–	SAMN10845488	MK481515 MK481516	MK481640	MK481722	–	–	MK481858
<i>PlagiUS flosculosus</i> (L.) Alavi & Heywood	<i>PlagiUS flosculosus</i>	A1038	FR, Corse-du-Sud, Zonza, 600 m	41.75 N, 09.17 E	Lambinon 91/Co310	MSB	SAMN10845486	–	SAMN10845486	–	–	–	–	SAMN10845486	MK481513	MK481638	MK481720	–	–	MK481856
<i>PlagiUS flosculosus</i> (L.) Alavi & Heywood	<i>PlagiUS flosculosus</i>	A0793	IT, Sardinia, Sassari	40.61 N, 08.47 E	Zedda s.n.	B 10 0550 370	ERS758371*	–	ERS758371*	–	–	–	–	ERS758371*	MK481433	LN869014*	MK481671	–	–	LN869163*
<i>PlagiUS maghrebinus</i> Vogt & Greuter	<i>PlagiUS maghrebinus</i>	A1036	DZ, Bejaia, Bordj-Mira, 170 m	36.53 N, 05.28 E	Podlech 39385	MSB	SAMN10845485	–	SAMN10845485	–	–	–	–	SAMN10845485	MK481512	MK481637	MK481719	–	–	MK481855
<i>PlagiUS maghrebinus</i> Vogt & Greuter	<i>PlagiUS maghrebinus</i>	A0794	TN, Jendouba, Ain Draham, 950 m	36.77 N, 08.69 E	Vogt 13696 & Oberprieler 8001	B 10 0550371	ERS758372*	–	ERS758372*	–	–	–	–	ERS758372*	AJ2964038 AJ2964388	MK481592	MK481672	–	–	MK481807
<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	A0996	ES, Jaen, Sierra de Magina, Cerro Carceles, 1900 m	37.72 N, 03.46 W	Vogt 3362	B 10 0550797	SAMN10845441	SAMN10845441	SAMN10845441	SAMN10845441	SAMN10845441	SAMN10845441	SAMN10845441	SAMN10845441	MK481426	MK481587	MK481667	MK481748	MK481884	MK481803
<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	A0980	MA, Taza, Jebel Azrou-Achkar, 1850 m	34.79 N, 03.85 W	Gómez s.n.	MA 703921	SAMN10845462	SAMN10845462	SAMN10845462	SAMN10845462	SAMN10845462	SAMN10845462	SAMN10845462	SAMN10845462	MK481478	MK481614	MK481696	MK481768	MK481904	MK481832
<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	A1061	MA, Fes, Jebel Tamokrant, 2000 m	33.20 N, 04.68 W	Gómez FG-7959	LEB	SAMN10845490	SAMN10845490	SAMN10845490	SAMN10845490	SAMN10845490	SAMN10845490	SAMN10845490	SAMN10845490	MK481519 MK481520	MK481642	MK481724	MK481780	MK481916	MK481860
<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	A0936	MA, Midelt, Jebel Bou Ijallabene - Jebel Masker, 2250-2500 m	32.37 N, 05.38 W	Oberprieler 3414 (2003)	B 10 0550788	MK465318	MK465303 MK465304	MK465284	MK465268	MK465252	MK465240	MK465221 MK465222	MK465201	MK481443	MK481594	MK481676	MK481750	MK481886	MK481812
<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	A0997	MA, Al Haouz, Jebel Toukhal, 3350 m	31.06 N, 07.91 W	Kreisch 92/0599	B 10 0469571	SAMN10845442	SAMN10845442	SAMN10845442	SAMN10845442	SAMN10845442	SAMN10845442	SAMN10845442	SAMN10845442	MK481427 MK481428	MK481588	MK481668	MK481749	MK481885	MK481804
<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	A0981	MA, Al Haouz, Oukaïmeden, 2750 m	31.19 N, 07.85 W	Herrero & al., AH 3096	MA 746560	SAMN10845463	SAMN10845463	SAMN10845463	SAMN10845463	SAMN10845463	SAMN10845463	SAMN10845463	SAMN10845463	MK481479 MK481480 MK481481	MK481615	MK481697	MK481769	MK481905	MK481833
<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	A0982	MA, Quarzazate, Jebel Siroua, 3200 m	30.70 N, 07.62 W	Gonzalo & al., 1158	MA 801046	SAMN10845464	SAMN10845464	SAMN10845464	SAMN10845464	SAMN10845464	SAMN10845464	SAMN10845464	SAMN10845464	MK481482	MK481616	MK481698	MK481770	MK481906	MK481834
<i>Rhodanthemum briquetii</i> (Maire) B. H. Wilcox & al.	<i>Rhodanthemum briquetii</i> (Maire) B. H. Wilcox & al.	A0951	MA, Marrakech, Tizi-n-Tichka, 1900 m	31.33 N, 07.37 W	Oberprieler 3580 (2169)	B 10 0273219	SAMN10845445	SAMN10845445	SAMN10845445	SAMN10845445	SAMN10845445	SAMN10845445	SAMN10845445	SAMN10845445	MK481448 MK481449	MK481597	MK481679	MK481751	MK481887	MK481815
<i>Rhodanthemum briquetii</i> (Maire) B. H. Wilcox & al.	<i>Rhodanthemum briquetii</i> (Maire) B. H. Wilcox & al.	A1066	MA, Azilal, Tambda, 2200 m	31.79 N, 06.38 W	Gómez, FG-7934	LEB	SAMN10845494	SAMN10845494	SAMN10845494	SAMN10845494	SAMN10845494	SAMN10845494	SAMN10845494	SAMN10845494	MK481526 MK481527	MK481646	MK481728	MK481784	MK481920	MK481864
<i>Rhodanthemum catananche</i> (Ball) B. H. Wilcox & al.	<i>Rhodanthemum catananche</i> (Ball) B. H. Wilcox & al.	A0983	MA, Quarzazate, Askaun-Ansal, 2466 m	30.77 N, 07.66 W	Quintanar & al., AQ3627	MA 799727	SAMN10845465	SAMN10845465	SAMN10845465	SAMN10845465	SAMN10845465	SAMN10845465	SAMN10845465	SAMN10845465	MK481483 MK481484	MK481617	MK481699	MK481771	MK481907	MK481835
<i>Rhodanthemum catananche</i> (Ball) B. H. Wilcox & al.	<i>Rhodanthemum catananche</i> (Ball) B. H. Wilcox & al.	A0087	MA, Boulmane, Ait-Ameur-Ouïd - Boulmane	33.40 N, 04.80 W	Vogt 10332 & Oberprieler 4780	B 10 0550836	ERS758373*	ERS758373*	ERS758373*	ERS758373*	ERS758373*	ERS758373*	ERS758373*	ERS758373*	MK481424 MK481425	LN869016*	LN869066*	LN868966*	LN869116*	LN869165*
<i>Rhodanthemum pseudocatananche</i> (Maire) B. H. Wilcox & al.	<i>Rhodanthemum pseudocatananche</i> (Maire) B. H. Wilcox & al.	A0978	MA, Ouarzazate, Tizi-n-Tichka, 2220 m	31.29 N, 07.38 W	Alexander & Kapicha s.n.	BM	SAMN10845460	SAMN10845460	SAMN10845460	SAMN10845460	SAMN10845460	SAMN10845460	SAMN10845460	SAMN10845460	MK481474 MK481475	MK481612	MK481694	MK481766	MK481902	MK481830
<i>Rhodanthemum pseudocatananche</i> (Maire) B. H. Wilcox & al.	<i>Rhodanthemum pseudocatananche</i> (Maire) B. H. Wilcox & al.	A1065	MA, Boulmane, Jebel Tichhoukt, 2100 m	33.39 N, 04.70 W	Gómez FG-5108	LEB	SAMN10845493	SAMN10845493	SAMN10845493	SAMN10845493	SAMN10845493	SAMN10845493	SAMN10845493	SAMN10845493	MK481524 MK481525	MK481645	MK481727	MK481783	MK481919	MK481863

Table S4.1 Continued.

Taxon	Lineage	ID	Geographic location	Coord.	Collector	Voucher	A39	B12	B20	C12	C20	D18	D23	D27	ITS	trnL-trnF	trnC-petN	psbA-trnH	petN-psbM	trnQ-rps16	
<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.	' <i>Rhodanthemum depressum</i> HA1'	A1176	MA, Al Haouz, Oukaimeden, 2660 m	31.20 N, 07.90 W	B.M. Exped. 721	RNG 9511203	MK465317	MK465302	MK465282 MK465283	MK465264 MK465265	MK465251	MK465238	MK465219 MK465220	MK465200	MK481545 MK481546	MK481658	MK481740	MK481793	MK481929	MK481876	
<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.	' <i>Rhodanthemum depressum</i> HA2'	A0952	MA, Al Haouz, Imhli, 1600-1700 m	31.14 N, 07.92 W	Vogt 15687	B 10 0550826	SAMN10845446	SAMN10845446	SAMN10845446	SAMN10845446	SAMN10845446	SAMN10845446	SAMN10845446	SAMN10845446	MK481450	MK481598	MK481680	MK481752	MK481888	MK481816	
<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.		A1175	MA, Al Haouz, Tizi-n-Tastiane, 2860 m	31.19 N, 07.91 W	B.M. Exped. 712	RNG 9511199	MK465316	MK465301	MK465281	MK465267	MK465250	MK465237	MK465217 MK465218	MK465199	MK481543 MK481544	MK481657	MK481739	MK481792	MK481928	MK481875	
<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.	' <i>Rhodanthemum depressum</i> AA'	A0984	MA, Tiznit, Col de Kerdouz, 1100 m	29.55 N, 09.34 W	Gómez s.n.	MA 703919	SAMN10845466	SAMN10845466	SAMN10845466	SAMN10845466	SAMN10845466	SAMN10845466	SAMN10845466	SAMN10845466	MK481485 MK481486	MK481618	MK481700	MK481772	MK481908	MK481836	
<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.		A1174	MA, Tiznit, Jebel Imzi, 1250 m	29.75 N, 09.29 W	Gómez 6325	LEB	MK465314 MK465315	MK465300	MK465279 MK465280	MK465263	MK465249	MK465235 MK465236	MK465215 MK465216	MK465198	MK481541 MK481542	MK481656	MK481738	MK481791	MK481927	MK481874	
<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.		R053-05	MA, Tiznit, Jebel Kest, 1250 m	29.76 N, 09.14 W	Vogt 17831, Oberprieler & Wagner	B 10 0704568	MK465323	MK465309	MK465290 MK465291	MK465273	MK465257	MK465244	MK465227 MK465228	MK465206	MK481584 MK481585	MK481665	MK481747	MK481801	MK481936	MK481883	
<i>Rhodanthemum maroccanum</i> (Batt.) B. H. Wilcox & al.	' <i>Rhodanthemum gayanum</i> '	A0992	MA, Khéouf, Aguelmam Azegza, 1440 m	32.97 N, 05.43 W	Blanché & al. s.n.	B 10 0673105	SAMN10845472	SAMN10845472	SAMN10845472	SAMN10845472	SAMN10845472	SAMN10845472	SAMN10845472	SAMN10845472	MK481495 MK481496	MK481624	MK481706	MK481778	MK481914	MK481842	
<i>Rhodanthemum maroccanum</i> (Batt.) B. H. Wilcox & al.		A1070	MA, Khéouf, Source de l'Oum-er-Rhia	33.12 N, 05.35 W	Vogt 12011	B 10 0550768	SAMN10845498	SAMN10845498	SAMN10845498	SAMN10845498	SAMN10845498	SAMN10845498	SAMN10845498	SAMN10845498	SAMN10845498	MK481532 MK481533	MK481650	MK481732	MK481788	MK481924	MK481868
<i>Rhodanthemum maroccanum</i> (Batt.) B. H. Wilcox & al.		A1071	MA, Marrakech, Imi-n-Tanoute, 1100 m	31.25 N, 08.75 W	Gómez, FG-6281	LEB	SAMN10845499	SAMN10845499	SAMN10845499	SAMN10845499	SAMN10845499	SAMN10845499	SAMN10845499	SAMN10845499	SAMN10845499	MK481534	MK481651	MK481733	MK481789	MK481925	MK481869
<i>Rhodanthemum gayanum</i> subsp. <i>denmatense</i> (Murb.) Vogt		A0960	MA, Beni Melhal, Afourer - Bin-el-Quidane, 1240 m	32.14 N, 06.53 W	R. Vogt 11941	B 10 0550756	SAMN10845449	SAMN10845449	SAMN10845449	SAMN10845449	SAMN10845449	SAMN10845449	SAMN10845449	SAMN10845449	SAMN10845449	MK481454 MK481455	MK481601	MK481683	MK481755	MK481891	MK481819
<i>Rhodanthemum gayanum</i> subsp. <i>denmatense</i> (Murb.) Vogt		A0985	MA, Chefchaouen, Bab Berret - Kétama, 1372 m	34.99 N, 04.82 W	Santos & al. MS 1056	MA 782160	SAMN10845467	SAMN10845467	SAMN10845467	SAMN10845467	SAMN10845467	SAMN10845467	SAMN10845467	SAMN10845467	SAMN10845467	MK481487	MK481619	MK481701	MK481773	MK481909	MK481837
<i>Rhodanthemum gayanum</i> subsp. <i>denmatense</i> (Murb.) Vogt		A1069	MA, Agadir, Imouzzer, 1030 m	30.64 N, 09.52 W	Gómez, FG-4383	LEB	SAMN10845497	SAMN10845497	SAMN10845497	SAMN10845497	SAMN10845497	SAMN10845497	SAMN10845497	SAMN10845497	SAMN10845497	MK481531	MK481649	MK481731	MK481787	MK481923	MK481867
<i>Rhodanthemum gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. subsp. <i>gayanum</i>		A0958	MA, Marrakech, street to Oukaimeden, 1150 m	31.30 N, 07.77 W	Oberprieler 3612	B 10 0550777	SAMN10845448	SAMN10845448	SAMN10845448	SAMN10845448	SAMN10845448	SAMN10845448	SAMN10845448	SAMN10845448	SAMN10845448	MK481452 MK481453	MK481600	MK481682	MK481754	MK481890	MK481818
<i>Rhodanthemum gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. subsp. <i>gayanum</i>		A0986	MA, Marrakech, Toufiat, 1450 m	31.49 N, 07.44 W	Jury & al., SL 19560	MA 698292	SAMN10845468	SAMN10845468	SAMN10845468	SAMN10845468	SAMN10845468	SAMN10845468	SAMN10845468	SAMN10845468	SAMN10845468	MK481488	MK481620	MK481702	MK481774	MK481910	MK481838
<i>Rhodanthemum gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. subsp. <i>gayanum</i>		A1067	MA, Figuig, Jbel Araira, 1400 m	32.24 N, 01.72 W	Gómez, FG-5877	LEB	SAMN10845495	SAMN10845495	SAMN10845495	SAMN10845495	SAMN10845495	SAMN10845495	SAMN10845495	SAMN10845495	SAMN10845495	MK481528	MK481647	MK481729	MK481785	MK481921	MK481865
<i>Rhodanthemum ifniense</i> (Font Quer) Ibn Tattou		R046-20	MA, Sidi-Ifni, Jebel Boumesguidam, 1164 m	29.20 N, 10.02 W	Vogt 17781, Oberprieler & Wagner	B 10 0745356	MK465321	MK465306	MK465287 MK465288	MK465269 MK465270	MK465255	MK465241	MK465225	MK465203 MK465204	MK481581	MK481663	MK481745	MK481799	MK481934	MK481881	
<i>Rhodanthemum ifniense</i> (Font Quer) Ibn Tattou	R049-01	MA, Sidi-Ifni, Jebel Sidi-Tual, 1126 m	29.20 N, 10.00 W	Vogt 17800, Oberprieler & Wagner	B 10 0745379	MK465322	MK465307 MK465308	MK465289	MK465271 MK465272	MK465256	MK465242 MK465243	MK465226	MK465205	MK481582 MK481583	MK481664	MK481746	MK481800	MK481935	MK481882		
<i>Rhodanthemum gayanum</i> subsp. <i>antiatlanticum</i> (Emb. & Maire) Vogt & Greuter	A0956	MA, Taroudant, Kemis d Tsafen - Igherm, 1600 m	30.10 N, 08.46 W	Kilian 3538	B 10 0550783	SAMN10845447	SAMN10845447	SAMN10845447	SAMN10845447	SAMN10845447	SAMN10845447	SAMN10845447	SAMN10845447	SAMN10845447	MK481451	MK481599	MK481681	MK481753	MK481889	MK481817	
<i>Rhodanthemum gayanum</i> subsp. <i>antiatlanticum</i> (Emb. & Maire) Vogt & Greuter	A0987	MA, Taroudant, Igherm, 1729 m	30.06 N, 08.44 W	Buim & Calvo s.n.	MA 758001	SAMN10845469	SAMN10845469	SAMN10845469	SAMN10845469	SAMN10845469	SAMN10845469	SAMN10845469	SAMN10845469	SAMN10845469	MK481489 MK481490	MK481621	MK481703	MK481775	MK481911	MK481839	
<i>Rhodanthemum gayanum</i> subsp. <i>fallax</i> (Maire & Weiller) Vogt	A1068	MA, Tiznit, Tafraoute, 1550 m	29.56 N, 09.00 W	Gómez, FG-4171	LEB	SAMN10845496	SAMN10845496	SAMN10845496	SAMN10845496	SAMN10845496	SAMN10845496	SAMN10845496	SAMN10845496	SAMN10845496	MK481529 MK481530	MK481648	MK481730	MK481786	MK481922	MK481866	
<i>Rhodanthemum gayanum</i> subsp. <i>fallax</i> (Maire & Weiller) Vogt	A0966	MA, Tiznit, Tioouit - Tanalt, 1550-1570 m	29.80 N, 09.13 W	Vogt 11828 & Oberprieler 6276	B 10 0673013	SAMN10845451	SAMN10845451	SAMN10845451	SAMN10845451	SAMN10845451	SAMN10845451	SAMN10845451	SAMN10845451	SAMN10845451	MK481458 MK481459	MK481603	MK481685	MK481757	MK481893	MK481821	
<i>Rhodanthemum gayanum</i> subsp. <i>fallax</i> (Maire & Weiller) Vogt	A0965	MA, Tiznit, Tizi-n-Mil, 1650 m	29.72 N, 08.85 W	Vogt 5694, Bayon & Oberprieler 2366	B 10 0673008	SAMN10845450	SAMN10845450	SAMN10845450	SAMN10845450	SAMN10845450	SAMN10845450	SAMN10845450	SAMN10845450	SAMN10845450	MK481456 MK481457	MK481602	MK481684	MK481756	MK481892	MK481820	
<i>Rhodanthemum kesticum</i> Gómez	A0972	MA, Tiznit, Jebel Kest, 1700 m	29.79 N, 09.11 W	Gómez s.n.	B 10 0484217	SAMN10845454	SAMN10845454	SAMN10845454	SAMN10845454	SAMN10845454	SAMN10845454	SAMN10845454	SAMN10845454	SAMN10845454	MK481462 MK481463	MK481606	MK481688	MK481760	MK481896	MK481824	
<i>Rhodanthemum kesticum</i> Gómez	A1062	MA, Tiznit, Jebel Kest, 1750 m	29.79 N, 09.11 W	Gómez, FG-5485	LEB	SAMN10845491	SAMN10845491	SAMN10845491	SAMN10845491	SAMN10845491	SAMN10845491	SAMN10845491	SAMN10845491	SAMN10845491	MK481521 MK481522	MK481643	MK481725	MK481781	MK481917	MK481861	

Table S4.1 Continued.

Taxon	Lineage	ID	Geographic location	Coord.	Collector	Voucher	A39	B12	B20	C12	C20	D18	D23	D27	ITS	trnL-trnF	trnC-petN	psbA-trnH	petN-psbM	trnQ-rps16
<i>Rhodanthemum hosmarieense</i> (Ball) B. H. Wilcox & al. <i>Rhodanthemum hosmarieense</i> (Ball) B. H. Wilcox & al.	<i>Rhodanthemum hosmarieense</i>	A0969	MA, Tanger-Asilah, Mont de Beni Hosmar, 600-750 m	35.50 N, 05.33 W	Vogt 10049, Oberprieler 4497	B 10 0550953	SAMN10845453	SAMN10845453	SAMN10845453	SAMN10845453	SAMN10845453	SAMN10845453	SAMN10845453	SAMN10845453	MK481461	MK481605	MK481687	MK481759	MK481895	MK481823
		A0967	MA, Tanger, Jebel Mousa, 100 m	35.90 N, 05.41 W	Deil 5780	B 10 0550956	SAMN10845452	SAMN10845452	SAMN10845452	SAMN10845452	SAMN10845452	SAMN10845452	SAMN10845452	SAMN10845452	SAMN10845452	MK481460	MK481604	MK481686	MK481758	MK481894
<i>Rhodanthemum laouense</i> Vogt	<i>Rhodanthemum laouense</i>	A0973	MA, Tanger-Asilah, Oued Laou, 350 m	35.28 N, 05.23 W	Vogt 10065 & Oberprieler 4513	B 10 0550954	SAMN10845455	SAMN10845455	SAMN10845455	SAMN10845455	SAMN10845455	SAMN10845455	SAMN10845455	SAMN10845455	MK481464 MK481465	MK481607	MK481689	MK481761	MK481897	MK481825
<i>Rhodanthemum laouense</i> Vogt		A0974	MA, Tanger-Asilah, Oued Laou, 350 m	35.28 N, 05.23 W	R. Vogt 9633 & Ch. Oberprieler 4069	B 10 0550961	SAMN10845456	SAMN10845456	SAMN10845456	SAMN10845456	SAMN10845456	SAMN10845456	SAMN10845456	SAMN10845456	MK481466 MK481467	MK481608	MK481690	MK481762	MK481898	MK481826
<i>Rhodanthemum laouense</i> Vogt		A0988	MA, Tanger-Asilah, Oued Laou, 144 m	35.28 N, 05.23 W	Buira, Calvo & Hanison, TB1150	MA 807350	SAMN10845470	SAMN10845470	SAMN10845470	SAMN10845470	SAMN10845470	SAMN10845470	SAMN10845470	SAMN10845470	MK481491 MK481492	MK481622	MK481704	MK481776	MK481912	MK481840
<i>Rhodanthemum maresii</i> (Coss.) B. H. Wilcox & al. <i>Rhodanthemum maresii</i> (Coss.) B. H. Wilcox & al.	' <i>Rhodanthemum maresii</i> '	A0975	MA, Midelt, Tizi-n-Talthem, 1630 m	32.62 N, 04.54 W	Vogt 14630 & Oberprieler 8939	B 10 0550969	SAMN10845457	SAMN10845457	SAMN10845457	SAMN10845457	SAMN10845457	SAMN10845457	SAMN10845457	SAMN10845457	MK481468 MK481469	MK481609	MK481691	MK481763	MK481899	MK481827
		A0976	MA, Midelt, Tizi-n-Talthem, 1820-1900 m	32.62 N, 04.54 W	Oberprieler 3326	B 10 0550976	SAMN10845458	SAMN10845458	SAMN10845458	SAMN10845458	SAMN10845458	SAMN10845458	SAMN10845458	SAMN10845458	SAMN10845458	MK481470 MK481471	MK481610	MK481692	MK481764	MK481900
<i>Rhodanthemum mesatlanticum</i> (Emb. & Maire) B. H. Wilcox & al.		R015-01	MA, Taourirt, Debou, 1186 m	34.05 N, 03.02 W	Vogt 17666, Oberprieler & Wagner	B 10 0704717	MK465310	MK465296	MK465274 MK465275	MK465258 MK465259	MK465245	MK465229 MK465230	MK465210 MK465211	MK465195	MK481574 MK481575	MK481659	MK481741	MK481795	MK481930	MK481877
<i>Rhodanthemum mesatlanticum</i> (Emb. & Maire) B. H. Wilcox & al.		R024-01	MA, Guercif, Beni Maqbal, 1880 m	33.61 N, 03.87 W	Vogt 17703, Oberprieler & Wagner	B 10 0760062	MK465312 MK465313	MK465297 MK465298	MK465277 MK465278	MK465261 MK465262	MK465247 MK465248	MK465233 MK465234	MK465213 MK465214	MK465197	MK481578	MK481661	MK481743	MK481797	MK481932	MK481879
<i>Rhodanthemum redieri</i> (Maire) B. H. Wilcox & al. s.l.	<i>Rhodanthemum redieri</i>	A0977	MA, Guercif, Jebel Bou Iflane, 2210 m	33.65 N, 04.17 W	Oberprieler 1973	B 10 0550379	SAMN10845459	SAMN10845459	SAMN10845459	SAMN10845459	SAMN10845459	SAMN10845459	SAMN10845459	SAMN10845459	MK481472 MK481473	MK481611	MK481693	MK481765	MK481901	MK481829
<i>Rhodanthemum redieri</i> subsp. <i>humbertii</i> Góniz		A0993	MA, Midelt, Jebel Ayachi, 2200 m	32.50 N, 04.94 W	Staudinger & Finckh s.n.	B 10 0673106	SAMN10845473	SAMN10845473	SAMN10845473	SAMN10845473	SAMN10845473	SAMN10845473	SAMN10845473	SAMN10845473	MK481497	MK481625	MK481707	MK481779	MK481915	MK481843
<i>Rhodanthemum redieri</i> (Maire) B. H. Wilcox & al. subsp. <i>redieri</i>		R021-01	MA, Hrane, Jebel Ari Beni, 2369 m	33.26 N, 04.97 W	Vogt 17699, Oberprieler & Wagner	B 10 0704774	MK465311	MK465299	MK465276	MK465260	MK465246	MK465231 MK465232	MK465212	MK465196	MK481576 MK481577	MK481660	MK481742	MK481796	MK481931	MK481878
<i>Rhodanthemum redieri</i> (Maire) B. H. Wilcox & al. s.l.		A1063	MA, Boulmane, Jebel Tichchoukt, 2100 m	33.39 N, 04.70 W	Góniz, FG-5110	LEB	SAMN10845492	SAMN10845492	SAMN10845492	SAMN10845492	SAMN10845492	SAMN10845492	SAMN10845492	SAMN10845492	MK481523	MK481644	MK481726	MK481782	MK481918	MK481862
<i>Rhodanthemum quezelii</i> Dobignard & Duret	<i>Rhodanthemum quezelii</i>	A0991	MA, Azilal, Agouti, 1850 m	31.63 N, 06.48 W	Góniz & Prieto s.n.	MA 883456	SAMN10845471	SAMN10845471	SAMN10845471	SAMN10845471	SAMN10845471	SAMN10845471	SAMN10845471	SAMN10845471	MK481493 MK481494	MK481623	MK481705	MK481777	MK481913	MK481841
<i>Rhodanthemum quezelii</i> Dobignard & Duret		A1072	MA, Azilal, Agouti, 1850 m	31.63 N, 06.48 W	Góniz, FG-9837	LEB	SAMN10845500	SAMN10845500	SAMN10845500	SAMN10845500	SAMN10845500	SAMN10845500	SAMN10845500	SAMN10845500	MK481535 MK481536	MK481652	MK481734	MK481790	MK481926	MK481870
<i>Rhodanthemum</i> spec.	' <i>Rhodanthemum</i> spec.'	R038-01	MA, Midelt, Jebel Bou Ijallabene - Jebel Masker, 1794 m	32.36 N, 05.37 W	Vogt 17739, Oberprieler & Wagner	B 10 0760029	MK465319 MK465320	MK465305	MK465285 MK465286	MK465266	MK465253 MK465254	MK465239	MK465223 MK465224	MK465202	MK481579 MK481580	MK481662	MK481744	MK481798	MK481933	MK481880
<i>Rhodanthemum</i> spec.		A0979	MA, Midelt, Jebel Bou Ijallabene - Jebel Masker, 1950-2000 m	32.37 N, 05.37 W	Oberprieler 3392 (1982)	B 10 0550383	SAMN10845461	SAMN10845461	SAMN10845461	SAMN10845461	SAMN10845461	SAMN10845461	SAMN10845461	SAMN10845461	MK481476 MK481477	MK481613	MK481695	MK481767	MK481903	MK481831
<i>Ursinia anthemoides</i> subsp. <i>vesicolor</i> (DC.) Prassler	<i>Ursinia anthemoides</i>	A0436	ZA, Cape, Kamieskroon, 800-1000 m	30.20 S, 17.97 E	Strid & Strid 37382	S	MK465324 MK465325	-	MK465294 MK465295	-	-	-	-	MK465208 MK465209	AM774473↓	MK481589	MK481669	-	-	MK481805

Table S4.2 Information about all primers used in the study, including marker and sequence information. Original Chapman markers were either modified (M13/TitB) for 454 pyro-sequencing library preparation described in Konowalik et al. (2015) or redesigned for the genus *Leucanthemum* (e.g. A39_leu350bp_f) or *Rhodanthemum* (e.g. RhoD27r).

Primer name	Marker	Sequence	Source
trnL2(e) trnFr(f)	<i>trnL-trnF</i> <i>trnL-trnF</i>	GGTTCAAGTCCCTCTATCCC ATTTGAACTGGTGACACGAG	Taberlet et al. (1991)
trnC	<i>trnC-petN</i>	CCAGTTCAAATCTGGGTGTC	Demesure et al. (1995)
petN1R	<i>trnC-petN</i>	CCCAAGCAAGACTTACTATATCC	Lee and Wen (2004)
psbA-HF trnH-HR	<i>psbA-trnH</i> <i>psbA-trnH</i>	CGAAGCTCCATCTACAAATGG ACTGCCTTGATCCACTTGGC	Hamilton (1999)
psbAf trnHr	<i>psbA-trnH</i> <i>psbA-trnH</i>	GTTATGCATGAACGTAATGCTC CGCGCATGGTGGATTACAAAATC	Sang et al. (1997)
petN1 psbM2R	<i>petN-psbM</i> <i>petN-psbM</i>	GGATATAGTAAGTCTTGCTTGGG TTCTTGCAATTTATTGCTACTGC	Lee and Wen (2004)
trnQ2	<i>trnQ-rps16</i>	GCGTGGCCAAGYGGTAAGGC	Shaw et al. (2007)
rps16x1_leu	<i>trnQ-rps16</i>	CAATCGAATTGTCAATGATGC	Konowalik et al. (2015)
ITS-18SF ITS-26SR	ITS ITS	GAACCTTATCGTTTAGAGGAAGG CCGCCAGATTTTCACGCTGGGC	Rydin et al. (2004)
ITS1 - P2	ITS	CTCGATGGAACACGGGATTCTGC	Ochsmann (2000)

Table S4.2 Continued.

Primer name	Marker	Sequence	Source
ITS2 - D	ITS	CTCTCGGCAACGGATATCTCG	Blattner (1999)
ITS2 - P3	ITS	GCATCGATGAAGAACGCAGC	White et al. (1990)
ITS1-P1B (ITS 5A)	ITS	GGAAGGAGAAGTCGTAACAAGG	Funk et al. (2004)
A39f	A39	ACTAGTTGGCATYTRATGGTAACA	Chapman et al. (2007)
A39r	A39	GCCRACAAAATTGAGCTGAAGATC	
A39_leu350bp_f	A39	AATGGTGTTC AATTGGTTTTC	Konowalik et al. (2015)
A39_leu350bp_r	A39	CCAACTCCAACAAGTAGGAG	
M13_A39_Leu350bp_f	A39	CACGACGTTGTAAAACGACAATGGTGTTC AATTGGTTTTC	
TitB_A39_Leu350bp_r	A39	CTATGCGCCTTGCCAGCCCGCTCAGCCA AACTCCAACAAGTAGGAG	
B12f	B12	CAAGTGGCTGCAGCCATGGG	Chapman et al. (2007)
B12r	B12	ACATCRGGMACCATTCWCCGGTGT	
M13_B12_f	B12	CACGACGTTGTAAAACGACCAAGTGGCTGCAGCCATGGG	
TitB_B12_Leu350bp_r	B12	CTATGCGCCTTGCCAGCCCGCTCAGACGTAGTAGTTGATCAACTG	
B20f	B20	AGTGGWATYAGTGGKGCTAGTTACT	
B20r	B20	CCACCACGHACAAGMAGCCAAAG	
M13_B20_f	B20	CACGACGTTGTAAAACGACAGTGGWATYAGTGGKGCTAGTTACT	
TitB_B20_r	B20	CTATGCGCCTTGCCAGCCCGCTCAGCCACCACGHACAAGMAGCCAAAG	
C12f	C12	TCTTGCACCACCAACTGYTTGGC	
C12r	C12	GACACCGCCTTGCTGC	

Table S4.2 Continued.

Primer name	Marker	Sequence	Source	
RhoC12f	<i>C12</i>	GCAAAGGTCTTGGATGAGGAATCG	this study	
RhoC12r	<i>C12</i>	GCTCTRGCTCTCCTTAAATCCCTG		
M13_C12_f	<i>C12</i>	CACGACGTTGTAAAACGACTCTTGACCACCAACTGYTTGGC	Chapman et al. (2007)	
TitB_C12_Leu350bp_r	<i>C12</i>	CTATGCGCCTTGCCAGCCCGCTCAGGGGACAATGTTCAATGCTG	Konowalik et al. (2015)	
C20f	<i>C20</i>	TTCTTCAATGCKKCTGCTTCTCA	Chapman et al. (2007)	
C20r	<i>C20</i>	AGCCAGTTGAATGAYAGCTCA		
M13_C20_f	<i>C20</i>	CACGACGTTGTAAAACGACTTCTTCAATGCKKCTGCTTCTCA		
TitB_C20_r	<i>C20</i>	CTATGCGCCTTGCCAGCCCGCTCAGAGCCAGTTGAATGAYAGCTCA		
D18f	<i>D18</i>	GGAAGRCTHCTWAGATATGACCCWCC		
D18r	<i>D18</i>	CTGCAACAATCAATWGCHACCCAA		
M13_D18_f	<i>D18</i>	CACGACGTTGTAAAACGACGGAAGRCTHCTWAGATATGACCCWCC		
TitB_D18_r	<i>D18</i>	CTATGCGCCTTGCCAGCCCGCTCAGCTGCAACAATCAATWGCHACCCAA		
D23f	<i>D23</i>	AGAAGGGTGGAACAGARCATTTRGGGCT		
D23r	<i>D23</i>	GGCATRATYCCRATCTTGCATTCWCCAGG		
M13_D23_f	<i>D23</i>	CACGACGTTGTAAAACGACAGAAGGGTGGAACAGARCATTTRGGGCT		
TitB_D23_r	<i>D23</i>	CTATGCGCCTTGCCAGCCCGCTCAGGGCATRATYCCRATCTTGCATTCWCCAGG		
D27f	<i>D27</i>	ATGATYAGTGAAAAGGAGCTYCT		
D27r	<i>D27</i>	GGWACAAAATGAGCMGTYACVACAGC		
RhoD27f	<i>D27</i>	GTCAATAGGTAACRTATCTTGC		this study
RhoD27r	<i>D27</i>	GGGAATCCTGCATTGTCCARAAC		
M13D27_f	<i>D27</i>	CACGACGTTGTAAAACGACATGATYAGTGAAAAGGAGCTYCT	Chapman et al. (2007)	
TitB_D27_r	<i>D27</i>	CTATGCGCCTTGCCAGCCCGCTCAGGGWACAAAATGAGCMGTYACVACAGC		

Table S4.3 Number of quality filtered reads obtained after checking for barcode errors, primer mismatches and phred quality-scores of raw Roche 454 pyrosequencing data.

ID	Barcode	Taxa	Project	Number of quality filtered reads							
				A39	B12	B20	C12	C20	D18	D23	D27
A1020	CGCT	<i>Chlamydomophora tridentata</i> (Delile) Less.	this study	93	-	24	-	-	-	-	80
A0795	ACTG	<i>Chlamydomophora tridentata</i> (Delile) Less.	Konowalik et al. (2015)	41	-	44	-	-	-	-	37
A1012	TAGG	<i>Chrysanthoglossum deserticola</i> (Murb.) B. H. Wilcox & al.	this study	84	-	156	-	-	-	-	121
A0791	ACAG	<i>Chrysanthoglossum deserticola</i> (Murb.) B. H. Wilcox & al.	Konowalik et al. (2015)	36	-	31	-	-	-	-	15
A1015	TAGT	<i>Chrysanthoglossum trifurcatum</i> (Desf.) B. H. Wilcox & al.	this study	60	-	44	-	-	-	-	116
A1017	CCAT	<i>Chrysanthoglossum trifurcatum</i> (Desf.) B. H. Wilcox & al.	this study	213	-	58	-	-	-	-	138
A1000	AAAG	<i>Coleostephus multicaulis</i> (Desf.) Durieu	this study	47	-	83	-	-	-	-	127
A1042	ACTG	<i>Coleostephus multicaulis</i> (Desf.) Durieu	this study	85	-	115	-	-	-	-	108
A0996	ATCG	<i>Coleostephus myconis</i> (L.) Rchb. f.	this study	115	-	78	-	-	-	-	91
A0792	AATG	<i>Coleostephus myconis</i> (L.) Rchb. f.	Konowalik et al. (2015)	38	-	54	-	-	-	-	35
A1113	TTCG	<i>Daveaua anthemoides</i> Mariz	this study	97	-	61	-	-	-	-	158
A1114	CCAG	<i>Daveaua anthemoides</i> Mariz	this study	68	-	48	-	-	-	-	120
A0998	CCCA	<i>Glossopappus macrotus</i> (Durieu) Briq. & Cavill	this study	37	-	46	-	-	-	-	90
A0790	ATGG	<i>Glossopappus macrotus</i> (Durieu) Briq. & Cavill	Konowalik et al. (2015)	35	-	28	-	-	-	-	34
A0937	AGCG	<i>Heteromera fuscata</i> (Desf.) Pomel	this study	182	-	84	-	-	-	-	207
A0796	AAAG	<i>Heteromera fuscata</i> (Desf.) Pomel	Konowalik et al. (2015)	20	-	58	-	-	-	-	32
A0944	CCTC	<i>Heteromera philaenorum</i> Maire & Weiller	this study	139	-	94	-	-	-	-	167
90-6	AAAG	<i>Leucanthemum burnatii</i> Briq. & Cavill.	Konowalik et al. (2015)	55	53	81	67	100	58	56	90
92-1	ACCG	<i>Leucanthemum burnatii</i> Briq. & Cavill.	Konowalik et al. (2015)	92	57	32	53	50	63	51	45
159-11	AACG	<i>Leucanthemum gallaenicum</i> Rodr. Oubiña & S. Ortiz	Konowalik et al. (2015)	49	60	83	68	71	43	60	90
L985	AGAG	<i>Leucanthemum gallaenicum</i> Rodr. Oubiña & S. Ortiz	Konowalik et al. (2015)	33	58	59	50	54	37	55	42
L035	AAGG	<i>Leucanthemum pyrenaicum</i> Vogt, Konowalik & Oberpr	Konowalik et al. (2015)	66	52	83	58	90	65	80	82
266-1	AGTG	<i>Leucanthemum pyrenaicum</i> Vogt, Konowalik & Oberpr	Konowalik et al. (2015)	38	56	59	47	26	29	36	59
L036	ATAG	<i>Leucanthemum cacuminis</i> Vogt, Konowalik & Oberpr.	Konowalik et al. (2015)	65	75	87	93	87	29	77	93
60-1	AGCG	<i>Leucanthemum cacuminis</i> Vogt, Konowalik & Oberpr.	Konowalik et al. (2015)	45	51	59	35	57	19	45	22
L033	ATTG	<i>Leucanthemum gaudinii</i> Dalla Torre	Konowalik et al. (2015)	38	66	42	28	37	39	27	23

Table S4.3 Continued.

ID	Barcode	Taxa	Project	Number of quality filtered reads							
				A39	B12	B20	C12	C20	D18	D23	D27
276-1	AGGG	<i>Leucanthemum gaudinii</i> Dalla Torre	Konowalik et al. (2015)	36	66	42	52	45	32	49	57
84-6	ATCG	<i>Leucanthemum gracilicaule</i> (Dufour) Pau	Konowalik et al. (2015)	62	57	75	70	65	50	69	85
85-1	TATG	<i>Leucanthemum gracilicaule</i> (Dufour) Pau	Konowalik et al. (2015)	59	48	73	71	48	28	40	57
116-4	ATGG	<i>Leucanthemum graminifolium</i> (L.) Lam.	Konowalik et al. (2015)	33	23	49	53	38	55	36	61
96-3	ATTC	<i>Leucanthemum graminifolium</i> (L.) Lam.	Konowalik et al. (2015)	33	62	44	76	50	53	58	51
L1002	ACAG	<i>Leucanthemum halleri</i> (Vitman) Ducommun	Konowalik et al. (2015)	72	51	86	67	54	67	30	90
208-1	ATTT	<i>Leucanthemum halleri</i> (Vitman) Ducommun	Konowalik et al. (2015)	67	54	63	72	39	38	54	67
L179	ACTG	<i>Leucanthemum laciniatum</i> Huter & al.	Konowalik et al. (2015)	45	62	83	74	78	94	49	66
280-1	AACA	<i>Leucanthemum laciniatum</i> Huter & al.	Konowalik et al. (2015)	46	63	52	63	37	46	49	45
366-1	ATAG	<i>Leucanthemum legraeum</i> (Rouy) B. Bock & J.-M. Tison	this study	72	138	31	53	151	147	53	180
369-1	ACAA	<i>Leucanthemum legraeum</i> (Rouy) B. Bock & J.-M. Tison	this study	122	100	21	63	-	117	42	178
406-1	TACG	<i>Leucanthemum ligusticum</i> Marchetti & al.	this study	94	97	35	49	-	107	50	154
416-1	TTGC	<i>Leucanthemum ligusticum</i> Marchetti & al.	this study	120	78	29	82	364	141	133	161
L998	ACCG	<i>Leucanthemum lithopolitanicum</i> (E. Mayer) Polatschek	Konowalik et al. (2015)	63	61	89	87	55	71	83	114
274-1	AAGA	<i>Leucanthemum lithopolitanicum</i> (E. Mayer) Polatschek	Konowalik et al. (2015)	40	69	54	82	53	76	46	29
131-20	AATG	<i>Leucanthemum monspeliense</i> (L.) H. J. Coste	Konowalik et al. (2015)	39	69	84	57	79	93	73	100
128-1	ACGG	<i>Leucanthemum monspeliense</i> (L.) H. J. Coste	Konowalik et al. (2015)	44	40	38	53	37	32	43	53
40-6	AGAG	<i>Leucanthemum pluriflorum</i> Pau	Konowalik et al. (2015)	53	63	107	79	83	78	73	86
55-1	ATAT	<i>Leucanthemum pluriflorum</i> Pau	Konowalik et al. (2015)	41	65	71	42	64	68	44	51
L990	AGTG	<i>Leucanthemum rotundifolium</i> (Willd.) DC.	Konowalik et al. (2015)	60	41	85	53	74	90	39	88
L989	ATTA	<i>Leucanthemum rotundifolium</i> (Willd.) DC.	Konowalik et al. (2015)	63	37	31	47	30	68	32	53
L992	ATCA	<i>Leucanthemum rotundifolium</i> (Willd.) DC.	Konowalik et al. (2015)	47	63	60	86	44	59	25	34
L151	AGCG	<i>Leucanthemum tridactylites</i> (A. Kern. & Huter) Huter & al.	Konowalik et al. (2015)	50	31	52	94	65	56	44	127
278-1	ATGT	<i>Leucanthemum tridactylites</i> (A. Kern. & Huter) Huter & al.	Konowalik et al. (2015)	43	68	50	69	47	89	63	68
L987	AGGG	<i>Leucanthemum virgatum</i> (Desr.) Clos	Konowalik et al. (2015)	70	50	71	85	55	53	57	102
250-1	ACAA	<i>Leucanthemum virgatum</i> (Desr.) Clos	Konowalik et al. (2015)	46	63	40	78	51	56	44	45
94-1	TAAG	<i>Leucanthemum vulgare</i> Lam.	Konowalik et al. (2015)	75	59	105	85	100	66	99	97
L046	ACTA	<i>Leucanthemum vulgare</i> Lam.	Konowalik et al. (2015)	29	67	53	49	51	74	58	54
184-1	ACCT	<i>Leucanthemum vulgare</i> Lam.	Konowalik et al. (2015)	63	90	35	51	45	68	58	67
L996	TATG	<i>Leucanthemum eliasii</i> (Sennen & Pau) Vogt, Konowalik & Oberpr.	Konowalik et al. (2015)	65	49	69	47	52	52	69	70
L162	AGAA	<i>Leucanthemum eliasii</i> (Sennen & Pau) Vogt, Konowalik & Oberpr.	Konowalik et al. (2015)	38	60	34	74	68	53	47	35

Table S4.3 Continued.

ID	Barcode	Taxa	Project	Number of quality filtered reads							
				A39	B12	B20	C12	C20	D18	D23	D27
135-7	ATTC	<i>Leucanthemum ageratifolium</i> Pau	Konowalik et al. (2015)	60	41	91	67	58	87	74	77
M60-1	AGTT	<i>Leucanthemum ageratifolium</i> Pau	Konowalik et al. (2015)	40	63	62	80	64	74	52	53
A0048	AACA	<i>Mauranthemum decipiens</i> (Pomel) Vogt & Oberpr.	this study	60	-	82	-	-	-	-	91
A1004	ATGT	<i>Mauranthemum decipiens</i> (Pomel) Vogt & Oberpr.	this study	145	-	82	-	-	-	-	151
A1005	CATA	<i>Mauranthemum gaetulum</i> (Batt.) Vogt & Oberpr.	this study	118	-	37	-	-	-	-	66
A1056	CTAA	<i>Mauranthemum gaetulum</i> (Batt.) Vogt & Oberpr.	this study	230	-	63	-	-	-	-	272
A1008	TAAT	<i>Mauranthemum paludosum</i> (Poir.) Vogt & Oberpr. subsp. <i>paludosum</i>	this study	123	-	43	-	-	-	-	110
A0798	AACG	<i>Mauranthemum paludosum</i> (Poir.) Vogt & Oberpr. subsp. <i>paludosum</i>	Konowalik et al. (2015)	40	-	52	-	-	-	-	42
A0799	AAGG	<i>Mauranthemum paludosum</i> subsp. <i>ebusitanum</i> (Vogt) Vogt & Oberpr.	Konowalik et al. (2015)	47	-	62	-	-	-	-	40
A1111	TTAG	<i>Nivellea nivellei</i> (Braun-Blanq. & Maire) B. H. Wilcox & al.	this study	157	-	98	-	-	-	-	126
A1024	AGGA	<i>Otospermum glabrum</i> (Lag.) Willk.	this study	100	-	156	-	-	-	-	119
A1053	TGCA	<i>Otospermum glabrum</i> (Lag.) Willk.	this study	95	-	91	-	-	-	-	146
A1038	CACT	<i>Plagius flosculosus</i> (L.) Alavi & Heywood	this study	795	-	73	-	-	-	-	95
A0793	ATCG	<i>Plagius flosculosus</i> (L.) Alavi & Heywood	Konowalik et al. (2015)	46	-	59	-	-	-	-	28
A1036	TCCC	<i>Plagius maghrebinus</i> Vogt & Greuter	this study	202	-	66	-	-	-	-	110
A0794	ATTG	<i>Plagius maghrebinus</i> Vogt & Greuter	Konowalik et al. (2015)	52	-	49	-	-	-	-	32
A0096	AGAG	<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	this study	210	51	71	255	270	189	54	166
A0980	CAGC	<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	this study	150	43	64	250	235	92	28	150
A1061	AAGA	<i>Rhodanthemum arundanum</i> (Boiss.) B. H. Wilcox & al.	this study	306	14	71	284	265	114	34	130
A0097	AACG	<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	this study	131	22	73	171	222	241	67	80
A0981	TCTT	<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	this study	344	42	72	227	173	150	86	150
A0982	AAGG	<i>Rhodanthemum atlanticum</i> (Ball) B. H. Wilcox & al. subsp. <i>atlanticum</i>	this study	281	37	65	198	235	89	38	126
A0951	ACGG	<i>Rhodanthemum briquetii</i> (Maire) B. H. Wilcox & al.	this study	182	34	81	145	199	140	39	109
A1066	CCGT	<i>Rhodanthemum briquetii</i> (Maire) B. H. Wilcox & al.	this study	279	21	48	206	116	161	32	37
A0983	TTAA	<i>Rhodanthemum catananche</i> (Ball) B. H. Wilcox & al.	this study	364	28	86	311	311	159	56	123
A0087	ATAG	<i>Rhodanthemum catananche</i> (Ball) B. H. Wilcox & al.	Konowalik et al. (2015)	44	25	65	65	40	55	43	43
A0952	ACCG	<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.	this study	150	55	79	228	329	199	83	147
A0984	AAAC	<i>Rhodanthemum depressum</i> (Ball) B. H. Wilcox & al.	this study	478	60	79	177	347	277	49	200
A0956	ACAG	<i>Rhodanthemum gayanum</i> subsp. <i>antiatlanticum</i> (Emb. & Maire) Vogt & Greuter	this study	251	60	82	372	218	190	37	36

Table S4.3 Continued.

ID	Barcode	Taxa	Project	Number of quality filtered reads							
				A39	B12	B20	C12	C20	D18	D23	D27
A0987	TTTA	<i>Rhodanthemum gayanum</i> subsp. <i>antiatlanticum</i> (Emb. & Maire) Vogt & Greuter	this study	384	64	99	191	343	128	63	228
A1068	TACA	<i>Rhodanthemum gayanum</i> subsp. <i>antiatlanticum</i> (Emb. & Maire) Vogt & Greuter	this study	143	60	88	198	141	142	43	125
A0958	TATA	<i>Rhodanthemum gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. subsp. <i>gayanum</i>	this study	120	26	68	182	228	121	53	132
A0986	AGTT	<i>Rhodanthemum gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. subsp. <i>gayanum</i>	this study	186	70	92	202	145	152	71	132
A1067	AGTG	<i>Rhodanthemum gayanum</i> (Coss. & Durieu) B. H. Wilcox & al. subsp. <i>gayanum</i>	this study	186	50	86	217	111	97	39	83
A0960	ATTA	<i>Rhodanthemum gayanum</i> subsp. <i>dennatense</i> (Murb.) Vogt	this study	314	108	75	222	288	150	75	137
A0985	AGAA	<i>Rhodanthemum gayanum</i> subsp. <i>dennatense</i> (Murb.) Vogt	this study	301	82	66	235	205	141	62	167
A1069	AATG	<i>Rhodanthemum gayanum</i> subsp. <i>dennatense</i> (Murb.) Vogt	this study	174	42	50	137	122	84	18	90
A0966	ATGG	<i>Rhodanthemum gayanum</i> subsp. <i>fallax</i> (Maire & Weiller) Vogt	this study	162	26	72	207	199	134	40	141
A0965	ATCA	<i>Rhodanthemum gayanum</i> subsp. <i>fallax</i> (Maire & Weiller) Vogt	this study	238	52	59	145	249	140	63	156
A0969	CGGA	<i>Rhodanthemum hosmariense</i> (Ball) B. H. Wilcox & al.	this study	250	35	77	252	142	364	26	97
A0967	CGAC	<i>Rhodanthemum hosmariense</i> (Ball) B. H. Wilcox & al.	this study	271	25	62	218	282	373	37	182
A0972	TGGT	<i>Rhodanthemum kesticum</i> Gómiz	this study	127	28	54	202	178	184	72	209
A1062	CGTA	<i>Rhodanthemum kesticum</i> Gómiz	this study	102	34	72	192	235	201	21	55
A0973	TGAT	<i>Rhodanthemum laouense</i> Vogt	this study	230	48	54	250	258	114	41	45
A0974	CTCC	<i>Rhodanthemum laouense</i> Vogt	this study	153	25	68	335	213	125	37	46
A0988	TCAA	<i>Rhodanthemum laouense</i> Vogt	this study	211	35	50	311	233	167	61	118
A0975	CAAA	<i>Rhodanthemum maresii</i> (Coss.) B. H. Wilcox & al.	this study	203	4	47	136	117	82	125	170
A0976	AATT	<i>Rhodanthemum maresii</i> (Coss.) B. H. Wilcox & al.	this study	80	5	55	147	221	110	36	173
A0992	ACTA	<i>Rhodanthemum maroccanum</i> (Batt.) B. H. Wilcox & al.	this study	232	52	96	252	264	176	75	274
A1070	ACGC	<i>Rhodanthemum maroccanum</i> (Batt.) B. H. Wilcox & al.	this study	293	56	68	253	526	319	53	213
A1071	ACCT	<i>Rhodanthemum maroccanum</i> (Batt.) B. H. Wilcox & al.	this study	278	43	85	184	167	117	41	84
A0978	AGCC	<i>Rhodanthemum pseudocatananche</i> (Maire) B. H. Wilcox & al.	this study	233	-	47	263	206	153	32	161
A1065	TGTC	<i>Rhodanthemum pseudocatananche</i> (Maire) B. H. Wilcox & al.	this study	227	37	52	99	159	101	24	101
A1063	ATAT	<i>Rhodanthemum redieri</i> (Maire) B. H. Wilcox & al. s.l.	this study	163	35	54	107	171	122	41	109
A0993	CTTT	<i>Rhodanthemum redieri</i> subsp. <i>humbertii</i> Gómiz	this study	153	21	52	115	245	101	21	107
A0977	TTCT	<i>Rhodanthemum redieri</i> (Maire) B. H. Wilcox & al. s.l.	this study	260	58	114	205	310	139	48	117
A0991	TATG	<i>Rhodanthemum quezelii</i> Dobignard & Duret	this study	110	32	50	159	112	86	23	76
A1072	CTGA	<i>Rhodanthemum quezelii</i> Dobignard & Duret	this study	88	5	30	220	788	119	20	106
A0979	TAAG	<i>Rhodanthemum</i> spec.	this study	227	66	52	219	256	118	69	124
total				15240	4292	7598	11653	12790	9336	4398	11760
Mean				130.5	50.9	65.6	136.1	152.9	109.7	51.5	101.2
SD				112.8	22.5	23.7	85.1	125.0	69.7	21.0	54.2

Table S4.4 Information about single markers of the ‘total dataset’, ‘*Leucanthemum* dataset’ and ‘*Rhodanthemum* dataset’, including aligned length, number and percentage of variable sites, number and percentage of parsimony-informative sites, as well as consistency (CI) and retention (RI) indices calculated in PAUP*. Best fitting models of sequence evolution found in JMODELTEST and clock models according to marginal likelihood comparisons are also itemized.

	locus	length (bp)	variable sites	parsimony-informative sites	CI	RI	nucleotide substitution model	clock model
total dataset	<i>A39</i>	385	161 (41.8%)	119 (30.9%)	0.57	0.87	GTR+I+G	relaxed
	<i>B20</i>	371	187 (50.4%)	140 (37.7%)	0.60	0.90	TPM2uf+G	relaxed
	<i>D27</i>	307	166 (54.1%)	120 (39.1%)	0.76	0.95	TPM3uf+I+G	relaxed
	ITS	806	267 (33.1%)	210 (26.1%)	0.54	0.90	SYM+I+G	relaxed
	ptDNA	1565	321 (20.5%)	224 (14.3%)	0.85	0.94	TVM+G	relaxed
<i>Leucanthemum</i> dataset	<i>A39</i>	319	44 (13.8%)	36 (11.3%)	0.73	0.96	TIM1+G	strict
	<i>B12</i>	376	86 (22.9%)	62 (16.5%)	0.54	0.91	TrN+I+G	strict
	<i>B20</i>	322	74 (23.0%)	51 (15.8%)	0.76	0.91	TPM3uf+G	strict
	<i>C12</i>	373	57 (15.3%)	39 (10.5%)	0.74	0.95	GTR+G	strict
	<i>C20</i>	307	21 (6.8%)	13 (4.2%)	0.91	0.97	TIM2	strict
	<i>D18</i>	323	46 (14.2%)	26 (8.0%)	0.58	0.81	TIM2+I+G	strict
	<i>D23</i>	367	43 (11.7%)	28 (7.6%)	0.90	0.98	HKY+G	strict
	<i>D27</i>	266	14 (5.3%)	8 (3.0%)	1.00	1.00	F81	strict
	ITS	761	77 (10.1%)	45 (5.9%)	0.89	0.94	TIM3ef+G	relaxed
	ptDNA	2319	47 (2.0%)	30 (1.3%)	0.96	0.98	TVM+G	relaxed
<i>Rhodanthemum</i> dataset	<i>A39</i>	289	23 (8.0%)	9 (3.1%)	0.85	0.95	TIM3+G	strict
	<i>B12</i>	385	28 (7.3%)	16 (4.2%)	0.83	0.91	TPM3uf+I	strict
	<i>B20</i>	317	29 (9.1%)	8 (2.5%)	0.91	0.96	TPM1uf+G	strict
	<i>C12</i>	302	38 (12.6%)	17 (5.6%)	0.95	0.97	TPM2uf	strict
	<i>C20</i>	276	26 (9.4%)	11 (4.0%)	0.90	0.91	TrN+G	strict
	<i>D18</i>	306	31 (10.1%)	19 (6.2%)	0.97	0.97	TPM3uf+I	strict
	<i>D23</i>	358	46 (12.8%)	21 (5.9%)	0.94	0.96	TPM2uf	strict
	<i>D27</i>	294	37 (12.6%)	16 (5.4%)	0.97	0.98	HKY	strict
	ITS	768	75 (9.8%)	42 (5.5%)	0.76	0.94	TIM1ef+I+G	strict
	ptDNA	2442	48 (2.0%)	26 (1.1%)	0.85	0.94	TVM+I+G	strict

Table S4.5 Logarithmic marginal-likelihood values for all loci of three different datasets using either an uncorrelated relaxed-clock or a strict clock model in path sampling (PS) and stepping stone (SS) analyses in BEAST v.1.8.4. The more parameter-rich relaxed-clock model was preferred over a strict-clock in the case of a mean difference of >3 log-likelihood values.

	locus	relaxed clock model						strict clock model						mean difference (relaxed vs. strict)		best clock
		PS			SS			PS			SS			PS	SS	
		run1	run2	mean	run1	run2	mean	run1	run2	mean	run1	run2	mean			
total dataset	A39	-3797.7	-3795.8	-3796.7	-3800.5	-3796.9	-3798.7	-3805.0	-3806.8	-3805.9	-3805.8	-3807.3	-3806.5	9.2	7.8	relaxed
	B20	-4276.5	-4276.2	-4276.4	-4278.4	-4277.5	-4278.0	-4293.6	-4292.2	-4292.9	-4294.5	-4294.1	-4294.3	16.6	16.3	relaxed
	D27	-2807.4	-2805.7	-2806.6	-2808.3	-2806.9	-2807.6	-2819.6	-2816.2	-2817.9	-2820.0	-2816.7	-2818.4	11.3	10.8	relaxed
	ITS	-5982.3	-5983.1	-5982.7	-5984.6	-5985.3	-5985.0	-5988.5	-5988.7	-5988.6	-5990.9	-5990.9	-5990.9	5.9	5.9	relaxed
	ptDNA	-5395.8	-5394.2	-5395.0	-5396.0	-5393.9	-5395.0	-5416.4	-5415.8	-5416.1	-5416.8	-5416.0	-5416.4	21.0	21.4	relaxed
Leucantheum dataset	A39	-1090.2	-1091.7	-1090.9	-1090.2	-1091.7	-1090.9	-1090.6	-1091.8	-1091.2	-1090.6	-1091.8	-1091.2	0.3	0.3	strict
	B12	-1917.8	-1918.3	-1918.1	-1917.8	-1918.3	-1918.0	-1920.3	-1920.4	-1920.3	-1920.3	-1920.4	-1920.4	2.2	2.3	strict
	B20	-1508.7	-1512.7	-1510.7	-1509.0	-1512.8	-1510.9	-1512.9	-1513.2	-1513.0	-1512.9	-1513.0	-1512.9	2.3	2.1	strict
	C12	-1336.3	-1337.4	-1336.9	-1337.6	-1338.1	-1337.9	-1337.1	-1335.5	-1336.3	-1338.0	-1335.5	-1336.8	-0.6	-1.1	strict
	C20	-740.3	-739.5	-739.9	-740.4	-739.5	-739.9	-740.2	-739.2	-739.7	-740.2	-739.2	-739.7	-0.2	-0.2	strict
	D18	-1285.9	-1284.4	-1285.1	-1285.8	-1284.5	-1285.1	-1284.5	-1283.9	-1284.2	-1284.6	-1284.0	-1284.3	-0.9	-0.8	strict
	D23	-1107.3	-1107.8	-1107.5	-1107.3	-1107.8	-1107.5	-1107.5	-1107.6	-1107.6	-1107.4	-1107.6	-1107.5	0.0	0.0	strict
	D27	-562.6	-562.3	-562.5	-562.6	-562.3	-562.5	-562.5	-562.4	-562.5	-562.5	-562.4	-562.5	0.0	0.0	strict
	ITS	-1962.3	-1959.7	-1961.0	-1962.5	-1959.8	-1961.2	-1966.0	-1966.7	-1966.4	-1966.1	-1966.7	-1966.4	5.3	5.2	relaxed
	ptDNA	-3667.2	-3664.4	-3665.8	-3669.1	-3666.0	-3667.6	-3674.8	-3675.9	-3675.4	-3676.8	-3677.7	-3677.3	9.6	9.7	relaxed
Rhodantheum dataset	A39	-805.6	-806.3	-806.0	-805.7	-806.4	-806.1	-803.3	-806.2	-804.8	-803.6	-806.2	-804.9	-1.2	-1.1	strict
	B12	-969.1	-967.3	-968.2	-969.1	-967.4	-968.3	-968.0	-968.9	-968.5	-968.1	-968.9	-968.5	0.3	0.3	strict
	B20	-900.9	-900.5	-900.7	-901.0	-900.7	-900.9	-899.3	-898.7	-899.0	-899.4	-898.7	-899.0	-1.8	-1.8	strict
	C12	-968.1	-967.8	-968.0	-968.2	-967.9	-968.1	-969.4	-969.0	-969.2	-969.4	-969.1	-969.3	1.2	1.2	strict
	C20	-748.6	-748.5	-748.6	-748.7	-748.5	-748.6	-748.4	-747.9	-748.2	-748.4	-747.9	-748.2	-0.4	-0.4	strict
	D18	-867.0	-866.6	-866.8	-867.2	-866.6	-866.9	-867.0	-867.4	-867.2	-867.0	-867.4	-867.2	0.4	0.3	strict
	D23	-1103.7	-1104.5	-1104.1	-1103.7	-1104.4	-1104.0	-1103.4	-1103.4	-1103.4	-1103.5	-1103.5	-1103.5	-0.7	-0.6	strict
	D27	-861.0	-861.1	-861.1	-861.0	-861.2	-861.1	-861.3	-861.2	-861.2	-861.3	-861.2	-861.3	0.2	0.2	strict
	ITS	-2196.8	-2197.8	-2197.3	-2196.9	-2197.8	-2197.4	-2199.4	-2198.8	-2199.1	-2199.5	-2198.9	-2199.2	1.8	1.9	strict
	ptDNA	-3882.7	-3883.7	-3883.2	-3882.7	-3883.8	-3883.2	-3881.8	-3882.9	-3882.4	-3881.9	-3883.0	-3882.4	-0.8	-0.8	strict

4.6 Supplemental Methods and Notes

Methods S4.1 Detailed description of processing 454-sequence data with the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (Coparaso et al., 2010).

As the name implies, QIIME was developed for the analysis of microbial communities. Nevertheless, the methods and scripts used by QIIME to extract sequences from raw NGS data can be transferred also to botanical phylogenetic studies, and can be used to pick alleles from accession-wise reads. The following six QIIME scripts/modules were used in an automated workflow based on customized Python scripts for the preparation of accession-wise allele alignments for each Chapman marker: (1) 'process_sff.py' for deserialising the binary flowgram files (.sff) and for converting them into .fasta and .qual files, (2) 'split_libraries.py' for extracting and sorting reads according to barcodes and markers, allowing no barcode mismatch but five mismatches in the primer sequences, along with a quality filtering step conducted by retaining only those reads with a minimum sequence length of 100 bp and an average quality phred-score above 30, (3) 'split_sequence_file_on_sample_ids.py' for preparing barcode-wise .fasta files, (4) 'denoiser.py' (Reeder and Knight, 2010) for a crucial three-step de-noising of 454-pyrosequencing data, and finally, (5) 'inflate_denoiser_output.py' and (6) 'pick_otus.py' for clustering reads allele-wise [using the 'usearch' algorithm, Edgar (2010)] and for removing of chimeric sequences. Clustering threshold, similarity threshold, and minimal cluster size of 'usearch' were chosen in a fashion that allowed retrieving single nucleotide polymorphisms (SNPs) among sequences and picking only those alleles represented by more than 20 percent of the reads. We used MAFFT v7.205 (Kato et al., 2005) for aligning resulting alleles together with the underlying quality filtered reads obtained in step 3 in order to check for errors regarding allele annotation.

Notes S4.1 Detailed description of species delimitation in *Leucanthemum* and *Rhodanthemum* based on analyses with the BEAST2 (Bouckaert et al., 2014) package STACEY (Jones, 2017a).

Nine out of the 15 delimited lineages in the genus *Leucanthemum* were represented by groups of 2-3 accessions showing strongly supported ($PP = 0.99-1.00$) monophyly in the tree and clear segregating patterns in the similarity matrix of Figure 4.2a. These lineages correspond to a group of morphologically clearly circumscribed and allopatrically distributed species [the so-called 'group 1' in Konowalik et al. (2015)]. All remaining *Leucanthemum* accessions of our study were part of a monophyletic group [$PP = 0.98$] with less noticeable substructure ['group 2' of Konowalik et al. (2015)]. Nevertheless, all individuals determined as *L. legraeum*, *L. ligusticum*, and *L. monspeliense* formed sharp and distinct clusters in the

similarity matrix in accordance with their morphological assignment, albeit with lower support in the tree ($PP = 0.70-0.88$). Following this result, we acknowledged these taxa as independent entities in accordance with a recent species-delimitation study of Wagner et al. (2017), based on sequence, AFLP fingerprinting, and morphometric data.

A less clear picture emerged for the 17 remaining individuals of the second group, illustrated by lighter grayscales (lower posterior frequencies of clustering) in the similarity matrix and a lack of significant support values in the corresponding accession tree. We merged these representatives of eight morphologically circumscribed taxa into three lineages, also considering geographical aspects given in Vogt (1991): The '*L. pluriflorum*' lineage comprised a weakly supported group ($PP = 0.7$) of individuals from Galicia (NW Spain) morphologically assigned to either *L. pluriflorum* (accessions 55-1, 40-6), *L. gallaecicum* (159-11, L985), or *L. cacuminis* (60-1). The second accession of the latter taxon from Cantabria (L036) showed higher genetic similarity to both representatives of the Cantabrian *L. eliasii* (L162, L996) and was consequently included into the '*L. eliasii*' lineage. The remaining accessions, representing the widespread taxa *L. vulgare*, *L. pyrenaicum*, *L. gaudinii*, and *L. ageratifolium*, were all pooled together in a '*L. vulgare*' lineage following the tree topology and the posterior frequencies of clusterings visualized in the similarity matrix of Figure 4.2.

In *Rhodanthemum*, 15 lineages were revealed and found being clustered into two main groups with a faint geographical pattern (Figure 4.2b): (a) an early-diverging group of seven lineages mainly from Spain, the Rif and Middle Atlas mountains and (b) a monophyletic group ($PP = 0.9$) comprising eight lineages that are either widespread throughout Morocco or mainly distributed in the High and Anti-Atlas mountains.

The first group contains all accessions of the Rif mountain taxa *R. hosmariense* and *R. laouense* in close relationship to each other ($PP = 1.0$), but still separated ($PP = 0.81$ and $PP = 1.0$) in correspondence with their discriminating morphological features [leaf shape and outline/indumentum of involucre bracts, as described in Vogt (1994)]. All individuals of *R. maresii* and *R. mesatlanticum* formed a strongly supported ($PP = 1.0$) unit without signs of internal differentiation in the similarity matrix and were consequently treated as representatives of a single '*R. maresii*' lineage in spite of contrasting morphological patterns (the former taxon exhibits yellow, the latter white-reddish ligules). The remaining accessions of the first group were part of a strongly supported monophyletic group subdivided into four lineages: (i) a *R. quezelii* lineage including representatives of a taxon that was recently acknowledged on subspecies or even species level due to its characteristic spatulate leaves [*R. redieri* subsp. *soriae* in Gómiz (2014), *R. quezelii* in Dobignard (2015)], (ii) a '*R. spec.*'

lineage comprising individuals from a population of yet unknown taxonomic status¹ from the Jebel Bou Ijallabene (High Atlas mountains), (iii) a *R. arundanum* lineage, and (iv) a *R. redieri* lineage comprising all subspecies of this taxon.

The second group comprised a strongly supported [$PP = 0.97$] cluster of *R. catananche* and *R. pseudocatananche* individuals from the Middle and High Atlas mountains. We treated these accessions as members of a single ‘*R. catananche*’ lineage, ignoring a weak substructure in the similarity matrix, showing rather a geographical pattern than a taxonomic separation. Signs of a weak separation were also detectable for accessions of the sympatrically distributed taxa *R. atlanticum* and *R. briquetii*. Following a conservative approach, we decided to merge individuals of both taxa into a single ‘*R. atlanticum*’ lineage, due to weak support values ($PP = 0.62$) and the presence of genetic overlap in the similarity matrix.

Surprisingly, we found the six accessions of *R. depressum* being separated into three different clusters: Accessions from the High Atlas mountains were assigned to two different units (‘*R. depressum* HA1’ and ‘*R. depressum* HA2’) both in the tree and in the similarity matrix. Finally, all individuals of *R. depressum* from the Anti-Atlas mountains were part of a large monophyletic group (the ‘*R. depressum* AA’ lineage; $PP = 0.98$). The latter clade was found being further subdivided into (i) a *R. kesticum* lineage containing representatives of a recently described species from the Jebel Kest (Gómez, 2001), (ii) a large ‘*R. ifniense*’ lineage comprising all representatives of the Anti-Atlas taxa *R. ifniense*, *R. gayanum* subsp. *fallax*, and *R. gayanum* subsp. *antiatlanticum* plus one individual of *R. gayanum* subsp. *gayanum* from the Saharan Atlas (accession A1067), and (iii) all remaining accessions of the widespread taxa *R. gayanum* subsp. *gayanum*, *R. gayanum* subsp. *demnatense*, and *R. maroccanum* (subsumed under a ‘*R. gayanum*’ lineage).

Notes S4.2 Detailed description of species trees from * BEAST analyses for *Leucanthemum* and *Rhodanthemum*.

The species tree reconstructed for the 15 lineages in *Leucanthemum* was largely unresolved with only three nodes supported by $PP \geq 0.9$. Strong support ($PP = 0.98$) was found for the bipartition of *Leucanthemum* lineages into a paraphyletic ‘group 1’ and a monophyletic ‘group 2’ (see STACEY analyses above), in accordance with the “Minimize Deep Coalescence” (MDC) based species tree in Konowalik et al. (2015). Beyond that, we found two well-supported sister-group relationships with a strong geographical correlate: (i) a node ($PP = 0.94$) connecting the Spanish lineages ‘*L. eliasii*’ (Cantabrian mountains) and

¹ The taxonomic status of this lineage (*R. quezelii* subsp. *ijallabenense* Florian Wagner, Vogt & Oberpr.) has been clarified in the course of the species delimitation study described in chapter 3)

'*L. pluriflorum*' (Galicia) and (ii) a sister-group relationship ($PP = 0.99$) between the lineages *L. laciniatum* and *L. tridactylites* from C and S Italy.

The species tree reconstructed for *Rhodanthemum* provided twice as much nodes with trustworthy support ($PP \geq 0.9$). Similar to the STACEY results, we found three species groups: (a) One monophyletic group ($PP = 0.9$) comprised all lineages either widespread throughout Morocco ('*R. catananche*', '*R. gayanum*') or mainly distributed in the High Atlas ('*R. atlanticum*', '*R. depressum* H1' and '*R. depressum* H2') and Anti-Atlas mountains ('*R. depressum* AA', *R. kesticum*, '*R. ifniense*'). Two further supported relationships were found being nested in this group: (i) a well-supported monophyletic clade ($PP = 0.98$) of all Anti-Atlas lineages plus '*R. gayanum*' and (ii) a sister-group relationship ($PP = 0.9$) between the sympatrically distributed lineages '*R. depressum* AA' and *R. kesticum*. (b) An unsupported group consisting of '*R. maresii*' together with a well-supported ($PP = 0.99$) group of lineages morphologically characterized by a deviating number of achene ribs (5-6 vs. 10). This '*R. arundanum* group' comprised the eponymous lineage *R. arundanum*, its sister lineage ($PP = 0.99$) *R. redieri*, as well as *R. quezelii* and the enigmatic population from the Bou Ijallabene ('*R. spec.*'). (c) A well-supported ($PP = 1.0$) group comprising the Rif mountain taxa *R. hosmariense* and *R. laouense*.

Chapter 5: Comprehensive summary, discussion and outlook

5.1 Comprehensive summary

The present thesis investigates micro- and macroevolutionary processes in the young and closely related genera *Leucanthemum* and *Rhodanthemum* from the subtribe Leucanthemeinae (Anthemideae, Compositae). The first two parts are focusing on species delimitation and hybridization in the closely-knit taxon groups around *L. ageratifolium* and *R. arundanum*, respectively, while the third part comprises a more comprehensive phylogenetic study of both genera with regard to their contrasting evolutionary trajectories concerning polyploid speciation.

The first study (chapter 2) evaluates the influence of hybridization on currently available species delimitation methods implemented in BEAST (BFD, BFD*, and DISSECT) using a group of five allopatrically distributed *Leucanthemum* taxa between northern Spain and southern Italy as a model system (the so-called *L. ageratifolium*-group). Analyses based on AFLP fingerprinting and morphometric data consistently identified 34 hybrid individuals between members of the *L. ageratifolium*-group on the one and the codistributed species *L. vulgare* on the other side, except in the case of the allopatrically distributed, S Italian *L. laciniatum*, where no hybrids could be detected (possibly due to the existence of reproductive barriers). The study showed that the robustness of applied species delimitation analyses based on AFLP fingerprinting and multi-locus sequence data was considerably influenced by the intensity of hybridization among species and the number of hybrid individuals included. Particularly the strong interspecific hybridization signal between *L. ligusticum* and *L. vulgare* resulted in the underestimation of species-level diversity and only after removal of individuals showing admixed genetic patterns, *L. ligusticum* populations were acknowledged as representatives of an independent species. In contrast to *L. ligusticum*, *L. laciniatum*, *L. legraeum*, and *L. monspeliense*, the taxonomic treatment of *L. ageratifolium* as either independent species or subspecies of *L. vulgare* remained uncertain in the course of this study.

The second study (chapter 3) infers species boundaries in the Ibero-Maghrebian *R. arundanum*-group, a group of four taxa with (i) morphologically differentiated populations or population groups, (ii) signs of interspecific hybridization and (iii) alternative taxonomic treatments based on morphology. Instead of AFLP fingerprinting, a modern restriction site associated DNA (RAD) sequencing approach was applied to 102 accessions of the study group, which provided up to 42,204 SNPs from 4,888 informative loci after *de-*

novo assembly and parameter optimization in IPYRAD (Eaton and Overcast, 2016). The assessment of different RADseq assemblies revealed 13 individuals showing admixed genetic patterns between *R. arundanum* on the one, and *R. redieri* or *R. quezelii* on the other side. Similar to the former study in the *L. ageratifolium*-group, the reliance of species-delimitation analyses were negatively influenced by gene flow among lineages and only after exclusion of hybrid individuals several methods and datasets consistently delimited three independent species, namely *R. arundanum*, *R. redieri*, and *R. quezelii*. Additionally, multi-species coalescent (MSC) species-delimitation analyses based on genomic RADseq data revealed genetic structure on the infraspecific level confirming a recently described subspecies (*R. redieri* subsp. *humbertii*) and arguing for the acknowledgment of two further taxa on subspecies rank (*R. quezelii* subsp. *ijallabenense* and *R. arundanum* subsp. *maireri*) new to science.

To determine factors that influence propensity toward polyploidization, diploid representatives of the European polyploid complex *Leucanthemum* are compared to members of its strictly diploid North African counterpart *Rhodanthemum* in a comprehensive phylogenetic study, described in the third part of the present thesis. Genetic differentiation among all lineages of both genera was evaluated to test the hypothesis that (allo)polyploidization is more common among species, which are genetically similar enough for successful crossings but genetically distinct enough to prevent homeologous chromosome pairing and multivalent formation during meiosis in offspring (Darlington, 1937). Phylogenetic Bray-Curtis genetic distances (Göker and Grimm, 2008) among all species of both genera were calculated as a proxy for genomic divergence using eight nuclear single-copy markers plus internal transcribed spacer (nrDNA ITS) and five plastid intergenic spacer regions. Results demonstrated that diploid *Leucanthemum* species are clearly more divergent among each other than those in *Rhodanthemum*, arguing for the importance of genetic divergence as a stimulus for polyploidization. Furthermore, investigation of hybridization patterns in both genera using both, species-tree (JML) and gene-tree (genealogical sorting index, *gsi*) approaches showed that diploid species of *Leucanthemum* carry more genomic signatures of past interspecific hybridization events than do those of *Rhodanthemum*. Both results demonstrate the importance of genetic differentiation among diploid progenitors and their concurrent affinity for natural hybridization for the formation of a polyploid complex. Furthermore, a time-calibrated phylogeny of 46 species of the subtribe Leucanthemeinae suggested that hybridization on the diploid and polyploid level was probably triggered by climate-induced range overlaps during the diversification of *Leucanthemum* in the Quaternary.

5.2 Snow White, Rose Red and the seven veils

In their article on species delimitation and relationships, Naciri and Linder (2015) noted, that ‘taxonomists lives would be simple if a clean phylogenetic signal right down to species could be obtained, so that sequence data can be used to build a phylogeny to species level’. The mentioned authors reviewed seven processes (‘veils’) that can obscure species delimitation and relationships especially in young plant groups, namely hybridization, incomplete lineage sorting, genome organization, intergenomic transfer, phylogeographic structure, demography, and selection.

Advances in theory and phylogenomic data have demonstrated that hybridization has an important impact on diversification and genome organization and occurs frequently both at shallow and deep taxonomic levels (Folk et al., 2018). Genetic fingerprinting and genotyping data in the course of the present thesis revealed patterns of recent interspecific hybridization in *Leucanthemum* (chapter 2) and *Rhodanthemum* (chapter 3) and due to the blurring effect of gene-flow on the reconstruction of species boundaries (particularly in the framework of the multi-species coalescent), hybrid individuals were discarded from species delimitation analyses. Furthermore, signatures of past interspecific hybridization were found among *Leucanthemum* species in chapter 4 and markers that would have influenced crown-age determination for this genus by contributing an incongruence signal, have been omitted in molecular dating analyses (Figure 4.5). The exclusion of individuals and markers, showing signs of hybridization was unavoidable in the present investigations, as alternative approaches that explicitly take into account gene flow after speciation (e.g. Camargo et al., 2012; Than et al., 2008) failed due to their computational complexity in the present study groups.

Incongruent or unresolved relationships among closely related species can also result from incomplete lineage sorting (ILS), the discordance between gene tree and species tree due to the stochastic segregation of alleles at a polymorphic locus at time of speciation (Naciri and Linder, 2015). ILS is particularly likely if the branches of the species tree are short (in terms of generations) and wide (in terms of effective population sizes) (Maddison, 1997) and the resulting effects on gene trees are hardly distinguishable from patterns caused by hybridization (see Figures S2.6-S2.10 of chapter 2). In the present thesis, different strategies have been applied to consider ILS and to distinguish between ILS and hybridization: (i) In all phylogenetic analyses, different loci were analyzed separately in the framework of the multi-species coalescent (MSC), which accounts for lineage-sorting stochasticity. (ii) In chapter 2 and 3, MSC species delimitation methods (DISSECT/STACEY, BFD* or BFD) were applied after exclusion of individuals showing admixed genetic patterns caused by actual interbreeding to account for both, ILS and hybridization. (iii) In chapter 3, both phenomena

were explicitly disentangled by conducting simulations under the coalescent-with-no-migration model in the course of a posterior predictive checking approach with the software JML (chapter 4.2.7).

In addition to hybridization and ILS, aspects connected with genome organization or genomic structure of investigated species can influence species delimitation analyses and phylogenetic reconstructions (Naciri and Linder, 2015). Modification of genome organization can be induced by whole genome duplication, translocations and chromosome fusions (Schneider and Grosschedl, 2007) and ultimately complicates the distinction between paralog and ortholog loci. Whilst no polyploid species were investigated in the present thesis¹, the distinction between paralog and ortholog loci was non-trivial due to the large genome sizes of diploid *Leucanthemum* and *Rhodanthemum* species (Pustahija et al., 2013; Oberprieler et al., 2018; Schall, 2019 unpublished) and several rounds of whole genome duplications in the evolution of Compositae (Badouin et al., 2017). To reduce the amount of (paralog) loci, AFLP fingerprinting in *Leucanthemum* was conducted with additional selective nucleotides during amplification of fragments (chapter 2) and a normalization step was included in ddRADseq library preparation for *Rhodanthemum* (chapter 3). Additionally, several paralog-filtering steps were performed during de-novo assembly of raw RADseq reads in chapter 3 and nuclear markers of chapter 2 and 4 were selected due to their single-copy nature in Compositae according to Chapman et al. (2007). Anyhow, due to the large and probably complex genomes of *Leucanthemum* and *Rhodanthemum* species and the absence of a reference genome, paralogy of investigated loci in the present thesis cannot be completely ruled out.

Plastid genomes are, on the other hand, considerably smaller and less complex compared to nuclear ones. Therefore, plastid markers have to be treated differently in species delimitation and phylogenetic studies compared to nuclear loci (Naciri and Linder, 2015). Due to the lack of intra-molecular recombination within organelle genomes, plastid markers were concatenated in all three studies of the present thesis. Furthermore, effective population sizes of plastid loci were scaled by a factor of 0.25 relative to the nuclear ones in all coalescent-based analyses to account for the haploid nature and the uniparental (maternal) inheritance of the plastome. A misinterpretation of species boundaries due to non-identification of paralog copies of plastid DNA transferred into the nucleus (intergenomic transfers, NuPt) as

¹ Cytometric investigations in the *R. arundanum* group (Schall, 2019 unpublished) revealed a potential tetraploid ploidy level for two populations of *Rhodanthemum mesatlanticum* (R015 and R024) investigated in chapter 3. Due to the lack of reliable chromosome counts, this finding was not included in the present thesis. If future studies will confirm that *R. mesatlanticum* is indeed a polyploid taxon, it can be assumed that it is originated from the diploid *R. maresii* via autopolyploidization (see results of chapters 3 and 4). This finding would confirm that (auto)polyploidization is generally possible in *Rhodanthemum*, but unlikely possibly due to the reasons stated in chapter 4.

stated by Naciri and Linder (2015) is rather unlikely in the present studies due to the low divergence found among plastid haplotypes of *Leucanthemum* and *Rhodanthemum* taxa, respectively (Table S2.2, Figure 3.1, Table S4.4).

While many of the ‘obscuring processes’ reviewed in Naciri and Linder (2015) have been considered in the present thesis by using appropriate methods and assumptions for species delimitation and phylogenetic reconstructions, phenomena like phylogeographic structure, demography and selection are more difficult to reconcile. Phylogeographic processes and demographic changes, such as bottlenecks, founder events or range expansions, can have a strong influence on effective population sizes (N_e) of investigated species and hence on species delimitation and phylogenetic analyses (Naciri and Linder, 2015). Similarly, changes in selection intensity is expected to affect N_e and so the coalescence depth. Phylogenetic reconstructions incorporating changes in N_e during evolution of populations and species are rare (but see Cornille et al., 2016), probably due to methodological limits, both in terms of sequencing and computational analyses. However, facilitated data acquisition via next-generation sequencing (NGS) and increasing computational power may overcome this limitation in future studies.

5.3 Outlook

In his book titled *Species Concepts in Biology*, Zachos (2016) reviewed a total of 32 existing species concepts, which can be roughly assigned to three categories, namely genealogy, ecology, and morphology (Figure 5.1). Species delimitation studies in the course of the present thesis are using multi-locus sequencing-, AFLP fingerprinting, and RAD-sequencing data to delimit species boundaries in the genera *Leucanthemum* and *Rhodanthemum*, and are therefore focusing on genealogy. Future studies in both genera should additionally incorporate ecological and morphological aspects to delimit biological meaningful units. Morphometric data can be collected by measuring leaf shapes of *Leucanthemum* and *Rhodanthemum* accessions using either the relatively simple leaf-dissection approach of chapter 2 or more advanced Fourier analysis techniques (e.g., Kuhl and Giardina, 1982). Ecological data, on the other hand, may be obtained by applying eco-climatological niche modelling as already conducted for *Leucanthemum* taxa from the Iberian Peninsula in Oberprieler et al. (2012). Morphological and eco-climatological niche data can be finally combined with genealogical datasets and jointly evaluated by using the integrative species delimitation method IBP&P of Solís-Lemus et al. (2015). An additional challenge will be the incorporation of polyploid *Leucanthemum* taxa into species delimitation analyses, which is probably hampered by (i) multiple formation of a polyploid species from the same parental

species, (ii) reciprocal formation of a polyploid species from the same parental species, and (iii) repeated hybridization between the same parental species followed by polyploidization. Polyploidy promoting factors have been studied in chapter 4 of the present thesis and the specific hypothesis for the formation of allopolyploids within *Leucanthemum* should be tested in other genera as suggested by Stuessy and Weiss-Schneeweiss (2019). Examining closely related genera, one consisting of exclusively diploid species and the other containing both diploid and polyploid species in a phylogenetic framework has proved to be a helpful approach for this purpose. Additionally, genetic and ecological factors that are responsible for the formation of polyploids or even polyploidy complexes may be evaluated by conducting eco-climatological niche reconstructions, crossing experiments among diploid species and the creation of artificial auto- and allopolyploids in both genera.

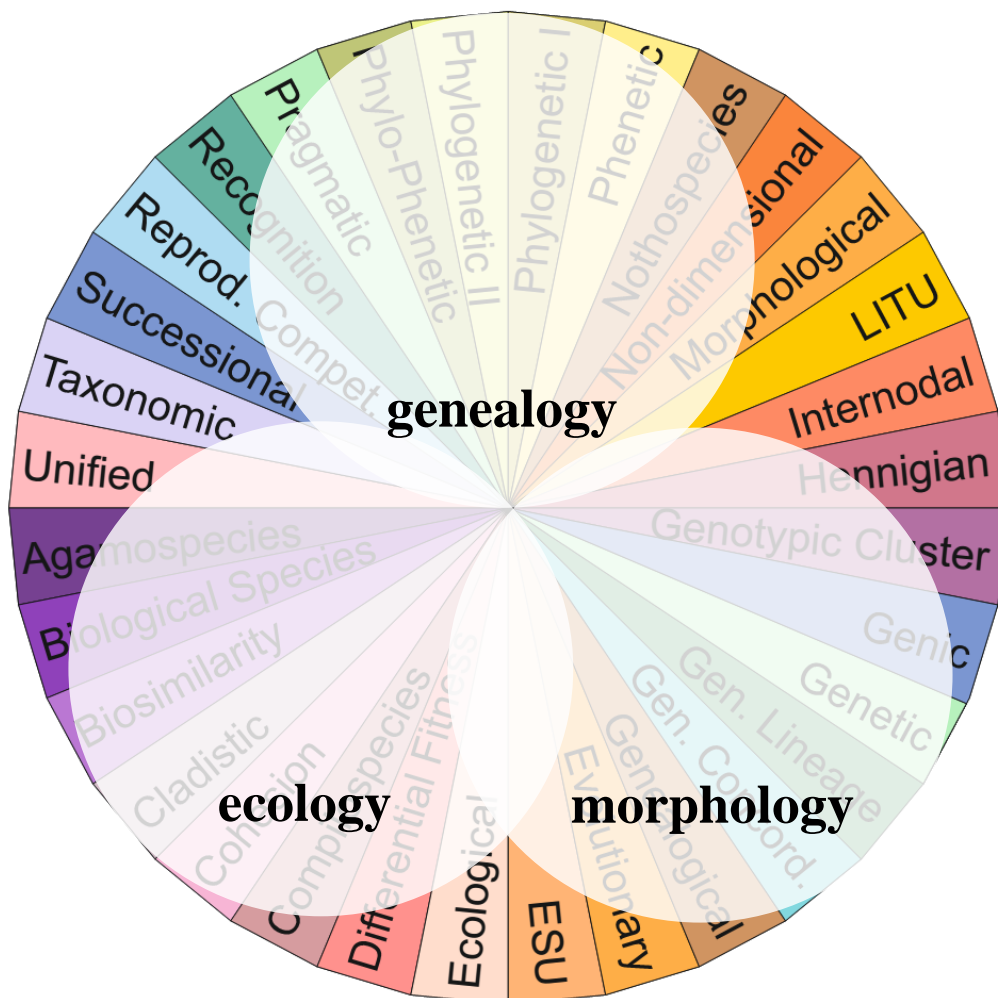


Figure 5.1 32 species concepts reviewed in Zachos (2016) can be roughly assigned to three categories, namely genealogy, ecology and morphology. While the present thesis concentrates on genealogical aspects of species delimitation in *Leucanthemum* and *Rhodanthemum*, future studies should also incorporate ecological and morphological data.

References

- Anderson BM, Thiele KR, Krauss SL, Barrett MD. 2017.** Genotyping-by-sequencing in a species complex of Australian Hummock Grasses (*Triodia*): methodological insights and phylogenetic resolution. *PLoS ONE* **12**: e0171053.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016.** Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**: 81–92.
- Andrews S. 2010.** *FastQC: A Quality Control Tool for High Throughput Sequence Data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. [accessed 28 March 2019].
- Arrigo N, Barker MS. 2012.** Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* **15**: 140–146.
- Aydin Z, Marcussen T, Ertekin AS, Oxelman B. 2014.** Marginal likelihood estimate comparisons to obtain optimal species delimitations in *Silene* sect. *Cryptoneurae* (Caryophyllaceae). *PLoS ONE* **9**: e106990.
- Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais- Brière C, Owens GL, Carrère S, Mayjonade B et al. 2017.** The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **546**: 148–152.
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012.** Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29**: 2157–2167.
- Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013.** Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* **30**: 239–243.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008.** Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**: e3376.
- Baldwin BG, Markos S. 1998.** Phylogenetic utility of the external transcribed spacer (ETS) of 18S–26S nrDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution* **10**: 449–463.

- Barley AJ, Brown MB, Thomson RC. 2018.** Impact of model violations on the inference of species boundaries under the multispecies coalescent. *Systematic Biology* **67**: 269–284.
- Bayer RJ, Purdy BG, Lebedyx DG. 1991.** Niche differentiation among eight sexual species of *Antennaria* Gaertner (Asteraceae: Innuleae) and *A. rosea*, their allopolyploid derivative. *Evolutionary Trends in Plants* **5**: 109–123.
- Benham J, Jeung JU, Jasieniuk M, Kanazin V, Blake T. 1999.** GENOGRAPHER: A graphical tool for automated fluorescent AFLP and microsatellite analysis. *Journal of Agricultural Genomics* **4**: 4.
- Bernardello R, Fiorini G, Marchetti D, Tison J-M. 2015.** *Leucanthemum legraeum* (Rouy) B. Bock et J.-M. Tison (Asteraceae) in Liguria, a new species for Italian Flora. *Annali del Museo Civico di Rovereto* **30**: 325–329.
- Bertrand Y, Scheen AC, Marcussen T, Pfeil BE, de Sousa F, Oxelman B. 2015.** Assignment of homoeologues to parental genomes in allopolyploids for species tree inference, with an example from *Fumaria* (Papaveraceae). *Systematic Biology* **64**: 448–471.
- Blanco-Pastor JL, Vargas P, Pfeil BE. 2012.** Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. *PLoS ONE* **7**: e39089.
- Blattner FR. 1999.** Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. *BioTechniques* **27**: 1180–1186.
- Bock B, Tison J-M. 2012.** Révisions nomenclaturales et taxonomiques (note no 2). *Bulletin de la Société Botanique du Centre-Ouest* **43**: 209–220.
- Bog M, Bässler C, Oberprieler C. 2017.** Lost in the hybridisation vortex: high-elevation *Senecio hercynicus* (Compositae, Senecioneae) is genetically swamped by its congener *S. ovatus* in the Bavarian Forest National Park (SE Germany). *Evolutionary Ecology* **31**: 401–420.
- Bouckaert RR, Heled J, Kuehnert D, Vaughan TG, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**: e1003537.
- Bouckaert R, Xie D. 2017.** Standard nucleotide substitution models v1.0.1. doi: 10.5281/zenodo.995740.
- Bremer K, Humphries CJ. 1993.** Generic monograph of the Asteraceae-Anthemideae. *Bulletin of the British Museum (Natural History), Botany* **23**: 71–177.
- Bretagnolle F, Thompson JD. 1995.** Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytologist* **129**: 1–22.

- Briquet J. 1916.** Composées. In: Briquet J, Cavillier F, eds. *E. Burnat, Flore des Alpes Maritimes* 6 (1). Genève, Bâle, Lyon.
- Brownfield L, Köhler C. 2011.** Unreduced gamete formation in plants: mechanisms and prospects. *Journal of Experimental Botany* **62**: 1659–1668.
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg N, RoyChoudhury A. 2012.** Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* **29**: 1917–1932.
- Buggs RJA, Soltis PS, Mavrodiiev EV, Symonds VV, Soltis DE. 2008.** Does phylogenetic distance between parental genomes govern the success of polyploids? *Castanea* **73**: 74–93.
- Buggs RJA, Soltis PS, Soltis DE. 2009.** Does hybridization between divergent progenitors drive whole-genome duplication? *Molecular Ecology* **18**: 3334–3339.
- Buggs RJA, Soltis PS, Soltis DE. 2011.** Biosystematic relationships and the formation of polyploids. *Taxon* **60**: 324–332.
- Camargo A, Morando M, Avila LJ, Sites J. 2012.** Species delimitation with ABC and other coalescent-based methods: A test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwini* complex (Squamata: Liolaemidae). *Evolution* **66-9**: 2834–2849.
- Camargo A, Sites JJ. 2013.** Species delimitation: a decade after the renaissance. In: Pavlinov IY, ed. *The Species Problem – Ongoing Issues*. New York, USA: InTech, 225–247.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA *et al.* 2010.** QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–336.
- Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013.** How to fail at species delimitation. *Molecular Ecology* **22**: 4369–4383.
- Catchen JM, Amores PA, Hohenlohe PA, Cresko WA, Postlethwait JH. 2011.** STACKS: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**: 171–182.
- Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013.** STACKS: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140.
- Chapman MA, Burke JM. 2007.** Genetic divergence and hybrid speciation. *Evolution* **61**: 1773–1780.

- Chapman MA, Chang J, Weisman D, Kesseli RV, Burke JM. 2007.** Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theoretical and Applied Genetics* **115**: 747–755.
- Choi SC. 2016.** Methods for delimiting species via population genetics and phylogenetics using genotype data. *Genes and Genomics* **38**: 905–915.
- Clement M, Snell Q, Walker P, Posada D, Crandall K. 2002.** TCS: Estimating gene genealogies. *Proc 16th Int Parallel Distrib Process Symp* **2**: 184.
- Comes HP, Kadereit JW. 1998.** The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science* **3**: 432–438.
- Cornille A, Salcedo A, Kryvokhyzha D, Glémin S, Holm K, Wright SI, Lascoux M. 2016.** Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Molecular Ecology* **25**: 616–629.
- Coyne JA, Orr HA. 2004.** *Speciation*. Sunderland, UK: Sinauer.
- Cummings MP, Neel MC, Shaw KL. 2008.** A genealogical approach to quantifying lineage divergence. *Evolution* **62**: 2411–2422.
- Darlington CD. 1937.** *Recent Advances in Cytology*, 2nd edn. Philadelphia, PA, USA: Blakiston.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012.** JMODELTEST 2: More models, new heuristics and parallel computing. *Nature Methods* **9**: 772.
- De Queiroz K. 2007.** Species concepts and species delimitation. *Systematic Biology* **56**: 879–886.
- De Queiroz K. 2011.** Branches in the line of descent: Charles Darwin and the evolution of the species concept. *Biological Journal of the Linnean Society* **103**: 19–35.
- De Villiers MJ, Pirie MD, Hughes M, Möller M, Edwards TJ, Bellstedt DU. 2013.** An approach to identify putative hybrids in the 'coalescent stochasticity zone', as exemplified in the African plant genus *Streptocarpus* (Gesneriaceae). *New Phytologist* **198**: 284–300.
- De Wet JMJ. 1980.** Origins of polyploids. In: Lewis WH, ed. *Polyploidy: Biological Relevance*. New York, USA: Plenum Press, 3–15.
- Demesure B, Sodzi N, Petit RJ. 1995.** A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* **4**: 129–134.
- Dillenberger MS, Kadereit JW. 2017.** Simultaneous speciation in the European high mountain flowering plant genus *Facchinia* (*Minuartia* s.l., Caryophyllaceae) revealed by genotyping-by-sequencing. *Molecular Phylogenetics and Evolution* **112**: 23–35.

- Dobignard A. 2015.** À propos de 3 taxons critiques pour la flore du Maroc observés lors de la 49^e session extraordinaire dans le Grand Atlas marocain. *Evaxiana* **2**: 253–266.
- Doyle JJ, Dickson EE. 1987.** Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon* **36**: 715–722.
- Doyle JJ, Doyle JS. 1987.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.
- Doyle JJ. 1992.** Gene trees and species trees: Molecular markers as one-character taxonomy. *Systematic Botany* **17**: 144–163.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006.** Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.
- Drummond AJ, Rambaut A. 2007.** BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012.** Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.
- Drummond AJ, Bouckaert RR. 2015.** *Bayesian Evolutionary Analysis with BEAST*. Cambridge, UK: Cambridge University Press.
- Durand EY, Patterson N, Reich D, Slatkin M. 2011.** Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* **28**: 2239–2252.
- Eaton DAR, Ree RH. 2013.** Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* **62**: 689–706.
- Eaton DAR. 2014.** PYRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**: 1844–1849.
- Eaton DAR, Overcast I. 2016.** *IPYRAD: Interactive Assembly and Analysis of RADseq Data Sets*. <http://ipyrad.readthedocs.io/>. [accessed 28 March 2019].
- Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Edgar RC. 2010.** Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Ehrendorfer F. 1980.** Polyploidy and distribution. In: Lewis WH, ed. *Polyploidy: Biological Relevance*. New York, USA: Plenum Press, 45–60.
- Euro+Med. 2016.** *Euro+Med PlantBase – The Information Resource for Euro-Mediterranean Plant Diversity*. <http://ww2.bgbm.org/EuroPlusMed/>. [accessed 7 November 2016].

- Euro+Med. 2018.** *Euro+Med PlantBase – The Information Resource for Euro-Mediterranean Plant Diversity*. <http://ww2.bgbm.org/EuroPlusMed/>. [accessed 19 October 2018].
- Euro+Med. 2019.** *Euro+Med PlantBase – The Information Resource for Euro-Mediterranean Plant Diversity*. <http://ww2.bgbm.org/EuroPlusMed/>. [accessed 14 May 2019].
- Fawcett JA, Van de Peer Y. 2010.** Angiosperm polyploids and their road to evolutionary success. *Trends in Evolutionary Biology* **2**: e3.
- Felber F. 1988.** Phenologie de la floraison de populations diploïdes et tétraploïdes d'*Anthoxanthum alpinum* et d'*Anthoxanthum odoratum*. *Canadian Journal of Botany* **66**: 2258–2264.
- Fernández-Mazuecos M, Mellers G, Vigalondo B, Sáez L, Vargas P, Glover BJ. 2018.** Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. *Systematic Biology* **67**: 250–268.
- Flot JF, Tillier A, Samadi S, Tillier S. 2006.** Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**: 627–630.
- Flot JF. 2007.** CHAMPURU 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Molecular Ecology Notes* **7**: 974–977.
- Flot JF. 2010.** SEQPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources* **10**: 162–166.
- Flot JF, Blanchot J, Charpy L, Cruaud C, Licuanan WY, Nakano Y, Payri C, Tillier S. 2011.** Incongruence between morphotypes and genetically delimited species in the coral genus *Stylophora*: Phenotypic plasticity, morphological convergence, morphological stasis or interspecific hybridization? *BMC Ecology* **11**: 22.
- Folk RA, Soltis PS, Soltis DE, and Guralnick R. 2018.** New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany* **105**: 364–375.
- Francis R. 2017.** Pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* **17**: 27–32.
- Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. 2012.** Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecology and Evolution* **27**: 480–488.
- Funk VA, Chan R, Keeley SC. 2004.** Insights into the evolution of the tribe Arctoteae (Compositae: subfamily Cichorioideae s.s.) using *trnL-F*, *ndhF*, and ITS. *Taxon* **53**: 637–655.

- García-Aloy S, Vitales D, Roquet C, Sanmartín I, Vargas P, Molero J, Kamau P, Aldasoro JJ, Alarcón A. 2017.** North-west Africa as a source and refuge area of plant biodiversity: a case study on *Campanula kremeri* and *Campanula occidentalis*. *Journal of Biogeography* **44**: 2057–2068.
- Göker M, Grimm GW. 2008.** General functions to transform associate data to host data, and their use in phylogenetic inference from sequences with intra-individual variability. *BMC Evolutionary Biology* **8**: 86.
- Gómez A, Lunt DH. 2006.** Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, eds. *Phylogeography of Southern European Refugia*. Dordrecht, NLD: Springer, 155–188.
- Gómiz GF. 2000.** Notas sobre flora de Marruecos. II. *Anales del Jardín Botánico de Madrid* **58**: 199–200.
- Gómiz GF. 2001.** Notas sobre flora de Marruecos. III. *Anales del Jardín Botánico de Madrid* **58**: 364–365.
- Gómiz GF. 2014.** Nuevo nombre en *Rhodanthemum* (Compositae, Anthemideae). *Acta Botánica Malacitana* **39**: 290–293.
- Gompert Z, Buerkle CA. 2010.** INTROGRESS: A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**: 378–384.
- Grant V. 1981.** *Plant Speciation*. 2nd edn. New York, NY, USA: Columbia University Press.
- Grant V. 2002.** Frequency of spontaneous amphiploids in *Gilia* (Polemoniaceae) hybrids. *American Journal of Botany* **89**: 1197–1202.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY *et al.* 2010.** A draft sequence of the Neandertal genome. *Science* **328**: 710–722.
- Greiner R, Oberprieler C. 2012.** The role of inter-ploidy block for reproductive isolation of the diploid *Leucanthemum pluriflorum* Pau (Compositae, Anthemideae) and its tetra- and hexaploid relatives. *Flora* **207**: 629–635.
- Greiner R, Vogt R, Oberprieler C. 2012.** Phylogenetic studies in the polyploid complex of the genus *Leucanthemum* Mill. (Compositae, Anthemideae) based on cpDNA sequence variation. *Plant Systematics and Evolution* **298**: 1407–1414.
- Greiner R, Vogt R, Oberprieler C. 2013.** Evolution of the polyploid north-west Iberian *Leucanthemum pluriflorum* clan (Compositae, Anthemideae) based on plastid DNA sequence variation and AFLP fingerprinting. *Annals of Botany* **111**: 1109–1123.

- Gruenstaedl M, Carstens BC, Santos-Guerra A, Jansen RK. 2017.** Statistical hybrid detection and the inference of ancestral distribution areas in *Tolpis* (Asteraceae). *Biological Journal of the Linnean Society* **121**: 133–149.
- Grummer JA, Bryson RW, Reeder TW. 2014.** Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* **63**: 119–133.
- Guralnick R. 2007.** Differential effects of past climate warming on mountain and flatland species distributions: a multispecies North American mammal assessment. *Global Ecology and Biogeography* **16**: 14–23.
- Hahsler M, Hornik K, Buchta C. 2008.** Getting things in order: an introduction to the R package seriation. *Journal of Statistical Software* **25**: 1–34.
- Hall TA. 1999.** BIOEDIT: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.
- Hamilton MB. 1999.** Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8**: 521–523.
- Hancock JF, Bringham RS. 1981.** Evolution of California populations of diploid and octoploid *Fragaria* (Rosaceae): a comparison. *American Journal of Botany* **68**: 1–5.
- Hassanpour H, Zare-Maivan H, Sonboli A, Kazempour-Osaloo, Wagner F, Tomasello S, Oberprieler C. 2018.** Phylogenetic species delimitation unravels a new species in the genus *Sclerorhachis* (Rech.f.) Rech.f. (Compositae, Anthemideae). *Plant Systematics and Evolution* **304**: 185–203.
- Hedin M, Carlson D, Coyle F. 2015.** Sky island diversification meets the multispecies coalescent – divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. *Molecular Ecology* **24**: 3467–3484.
- Hedin M. 2015.** High-stakes species delimitation in eyeless cave spiders (*Cicurina*, Dictynidae, Araneae) from central Texas. *Molecular Ecology* **24**: 346–361.
- Heled J, Drummond AJ. 2010.** Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**: 570–580.
- Hime PM, Hotaling S, Grewelle RE, O’Neill EM, Voss SR, Shaffer HB, Weisrock DW. 2016.** The influence of locus number and information content on species delimitation: An empirical test case in an endangered Mexican salamander. *Molecular Ecology* **25**: 5959–5974.

- Himmelreich S, Källersjö M, Eldenäs P, Oberprieler C. 2008.** Phylogeny of southern hemisphere Compositae-Anthemideae based on nrDNA ITS and cpDNA *ndhF* sequence information. *Plant Systematics and Evolution* **272**: 131–153.
- Himmelreich S, Breitwieser I, Oberprieler C. 2014.** Phylogenetic relationships in the extreme polyploid complex of the New Zealand genus *Leptinella* (Compositae: Anthemideae) based on AFLP data. *Taxon* **63**: 883–898.
- Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA. 2002.** The use of DNA sequencing (ITS and *trnL-F*), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). *American Journal of Botany* **89**: 279–286.
- Holland BR, Clarke AC, Meudt HM. 2008.** Optimizing automated AFLP scoring parameters to improve phylogenetic reconstruction. *Systematic Biology* **57**: 347–366.
- Hörandl E, Appelhans M. 2015.** *Next Generation Sequencing in Plant Systematics*. Königstein, Germany: Koeltz Scientific Books.
- Huang H, Knowles LL. 2016.** Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology* **65**: 357–365.
- Huson DH, Bryant D. 2006.** Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.
- Jacquemyn H, Brys R, Honnay O, Rolda'n-Ruiz I. 2012.** Asymmetric gene introgression in two closely related *Orchis* species: evidence from morphometric and genetic analyses. *BMC Evolutionary Biology* **12**: 178.
- Jakobsson M, Rosenberg NA. 2007.** CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.
- Joly S, Bruneau A. 2006.** Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. *Systematic Biology* **55**: 623–636.
- Joly S, Starr JR, Lewis WH, Bruneau A. 2006.** Polyploid and hybrid evolution in roses east of the rocky mountains. *American Journal of Botany* **93**: 412–425.
- Joly S. 2012.** JML: testing hybridization from species trees. *Molecular Ecology Resources* **12**: 179–184.
- Jombart T. 2008.** Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

- Jombart T, Ahmed I. 2011.** ADEGENET 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**: 3070–3071.
- Jones DL. 2015.** *FATHOM Toolbox for MATLAB: Software for Multivariate Ecological and Oceanographic Data Analysis*. <http://www.marine.usf.edu/user/djones/>. [accessed 7 November 2016].
- Jones G, Aydin Z, Oxelman B. 2015.** DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* **31**: 991–998.
- Jones G. 2017a.** Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology* **74**: 447–467.
- Jones G. 2017b.** *STACEY Package documentation: Species Delimitation and Species Tree Estimation in BEAST2*. <http://www.indriid.com/>. [accessed 28 March 2019].
- Kadereit JW. 2015.** The geography of hybrid speciation in plants. *Taxon* **64**: 673–687.
- Kass RE, Raftery AE. 1995.** Bayes factors. *Journal of the American Statistical Association* **90**: 773–795.
- Katoh K, Kuma K, Toh H, Miyata T. 2005.** MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**: 511–518.
- Katoh K, Rozewicki J, Yamada D. 2017.** MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. doi: 10.1093/bib/bbx108.
- Kim KJ, Jansen RK. 1995.** *NdhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences, USA* **92**: 10379–10383.
- Kincaid DT, Schneider RB. 1983.** Quantification of leaf shape with a microcomputer and Fourier transform. *Canadian Journal of Botany* **61**: 2333–2342.
- Klein JT, Kadereit JW. 2016.** Allopatric hybrids as evidence for past range dynamics in *Sempervivum* (Crassulaceae), a western Eurasian high mountain oreophyte. *Alpine Botany* **126**:119–133.
- Konowalik K, Wagner F, Tomasello S, Vogt R, Oberprieler C. 2015.** Detecting reticulate relationships among diploid *Leucanthemum* Mill. (Compositae, Anthemideae) taxa using multilocus species tree reconstruction methods and AFLP fingerprinting. *Molecular Phylogenetics and Evolution* **92**: 308–328.
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006.** GARD: A genetic algorithm for recombination detection. *Bioinformatics* **22**: 3096–3098.

- Kryvokhyzha D. 2017.** *Phylogenetic-analyses: Genealogical Sorting Index*. *GitHub Repository*. <https://github.com/evodify/phylogenetic-analyses/tree/master/GSI>. [accessed 1 November 2018].
- Kuhl F, Giardina C. 1982.** Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* **18**: 236–258.
- Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013.** aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research* **41**: 22–28.
- Lartillot N, Philippe H. 2006.** Computing bayes factors using thermodynamic integration. *Systematic Biology* **55**: 195–207.
- Leaché AD, Fujita MK, Minin VN, Bouckaert R. 2014a.** Species delimitation using genome wide SNP data. *Systematic Biology* **63**: 534–542.
- Leaché AD, Harris RB, Rannala B, Yang Z. 2014b.** The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology* **63**: 17–30.
- Leaché AD, Bouckaert R. 2018.** *Species Trees and Species Delimitation with SNAPP: A Tutorial and Worked Example*. <http://www.beast2.org/bfd/>. [accessed 28 March 2019].
- Leaché AD, Zhu T, Rannala B, Yang Z. 2018.** The spectre of too many species. *Systematic Biology* **68**: 168–181.
- Lee C, Wen J. 2004.** Phylogeny of *Panax* using chloroplast *trnC–trnD* intergenic region and the utility of *trnC–trnD* in interspecific studies of plants. *Molecular Phylogenetics and Evolution* **31**: 894–903.
- Lee J, Baldwin BG, Gottlieb LD. 2002.** Phylogeny of *Stephanomeria* and related genera (Compositae–Lactuceae) based on analysis of 18S–26S nuclear rDNA ITS and ETS sequences. *American Journal of Botany* **89**: 160–168.
- Leigh JW, Bryant D. 2015.** POPART: full-feature software for haplotype network construction. *Methods Ecology and Evolution* **6**: 1110–1116.
- Leitch AR, Leitch IJ. 2012.** Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytologist* **194**: 629–646.
- Levin DA. 1975.** Minority cytotype exclusion in local plant populations. *Taxon* **24**: 35–43.
- Lewis WH. 1980.** Polyploidy in species populations. In: Lewis WH, ed. *Polyploidy: Biological Relevance*. New York, USA: Plenum Press, 241–268.
- Lihová J, Kučera J, Perný M, Marhold K. 2007.** Hybridization between two polyploid *Cardamine* (Brassicaceae) species in north-western Spain: Discordance between morphological and genetic variation patterns. *Annals of Botany* **99**: 1083–1096.

- Lo Presti RM, Oppolzer S, Oberprieler C. 2010.** A molecular phylogeny and a revised classification of the Mediterranean genus *Anthemis* s.l. (Compositae, Anthemideae) based on three molecular markers and micromorphological characters. *Taxon* **59**: 1441–1456.
- Lumaret R. 1984.** The role of polyploidy in the adaptive significance of polymorphism at the Got-1 locus in the *Dactylis glomerata* complex. *Heredity* **52**: 153–169.
- Lumaret R, Guillerm JL, Delay J, Loutfi AAL, Izco J, Jay M. 1987.** Polyploidy and habitat differentiation in *Dactylis glomerata* L. from Galicia (Spain). *Oecologia* **73**: 436–446.
- Maddison WP. 1997.** Gene trees in species trees. *Systematic Biology* **46**: 523–536.
- Mallet J. 2005.** Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* **20**: 229–237.
- Marchi P. 1982.** 795. *Leucanthemum* Miller – Margherita (9341). In Pignatti S, ed. *Flora d'Italia Vol. 3*. Bologna: Edagricole. Mayr E (1942). *Systematics and the origin of species*. New York, USA: Columbia University Press, 89–96.
- Marques I, Draper D, López-Herranz ML, Garnatje T, Segarra-Moragues JG, Catalán T. 2016.** Past climate changes facilitated homoploid speciation in three mountain spiny fescues (*Festuca*, Poaceae). *Scientific Reports* **6**: 36283.
- Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Pineros D, Emerson BC. 2015.** RAD sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* **1**: 28–41.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011.** Recently formed polyploid plants diversify at lower rates. *Science* **333**: 1257.
- McCartney-Melstad E, Gidiş M, Shaffer HB. 2019.** An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. *Molecular Ecology Resources*. doi:10.1111/1755-0998.13029.
- McIntosh EJ, Rosetto M, Weston PH, Wardle GM. 2014.** Maintenance of strong morphological differentiation despite ongoing natural hybridization between sympatric species of *Lomatia* (Proteaceae). *Annals of Botany* **113**: 861–872.
- McLellan T. 1993.** The roles of heterochrony and heteroblasty in the diversification of leaf shapes in *Begonia dregei* (Begoniaceae). *American Journal of Botany* **80**: 796–804.
- McLellan T, Endler JA. 1998.** The relative success of some methods for measuring and describing the shape of complex objects. *Systematic Biology* **47**: 264–281.

- Meeus S, Janssens S, Helsen K, Jacquemyn H. 2016.** Evolutionary trends in the distylous genus *Pulmonaria* (Boraginaceae): Evidence of ancient hybridization and current interspecific gene flow. *Molecular Phylogenetics and Evolution* **98**: 63–73.
- Meusel H, Jäger EJ. 1992.** *Vergleichende Chorologie der zentraleuropäischen Flora. Vol. 3.* Jena, Germany: Gustav Fischer Verlag.
- Melai M, Marchetti D, Bernardello R, Peruzzi L. 2012.** A new diploid species of *Leucanthemum* (Asteraceae, Anthemideae) from Liguria (northwestern Italy). *Phytotaxa* **66**: 27–37.
- Meyer BS, Matschiner M, Salzburger W. 2017.** Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Systematic Biology* **66**: 531–550.
- Miller MA, Pfeiffer W, Schwartz T. 2010.** *Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees.* New Orleans, LA, USA: Gateway Computing Environments Workshop (GCE).
- Miralles A, Vences M. 2013.** New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PLoS ONE* **8**: e68242.
- Monti S, Tamayo P, Mesirov JP, Golub TR. 2003.** Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning* **52**: 91–118.
- Naciri Y, Linder HP. 2015.** Species delimitation and relationships: The dance of the seven veils. *Taxon* **64**: 3–16.
- Nasrallah ME, Yogeewaran K, Snyder S, Nasrallah JB. 2000.** *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiology* **124**: 1605–1614.
- Nei M, Li WH. 1979.** Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA* **76**: 5269–5273.
- Oberprieler C, Vogt R. 2000.** The position of *Castrilanthemum* Vogt & Oberprieler and the phylogeny of Mediterranean Anthemideae (Compositae) as inferred from nrDNA ITS and cpDNA *trnL/trnF* IGS sequence variation. *Plant Systematics and Evolution* **225**: 145–170.
- Oberprieler C. 2005.** Temporal and spatial diversification of Circum-Mediterranean Compositae-Anthemideae. *Taxon* **54**: 951–966.

- Oberprieler C, Vogt R, Watson LE. 2006.** XVI. Tribe Anthemideae Cass. In: Kadereit JW, Jeffrey C, eds. *The Families and Genera of Vascular Plants, Vol. VIII: Flowering Plants, Eudicots, Asterales*. Berlin-Heidelberg, DE: Springer, 342–374.
- Oberprieler C, Himmelreich S, Vogt R. 2007.** A new subtribal classification of the tribe Anthemideae (Compositae). *Willdenowia* **37**: 89–114.
- Oberprieler C, Himmelreich S, Källersjö M, Vallès J, Watson LE, Vogt R. 2009.** Tribe Anthemideae Cass. In: Funk VA, Susanna A, Stuessy T, Bayer R, eds. *Systematics, Evolution, and Biogeography of the Compositae*. Washington DC, USA: International Association for Plant Taxonomy, 631–666.
- Oberprieler C, Barth A, Schwarz S, Heilmann J. 2010.** Morphological and phytochemical variation, genetic structure, and phenology in an introgressive hybrid swarm of *Senecio hercynicus* and *S. ovatus* (Compositae, Senecioneae). *Plant Systematics and Evolution* **286**: 153–166.
- Oberprieler C, Eder C, Meister J, Vogt R. 2011a.** AFLP fingerprinting suggests the allopolyploid origin of two members of the *Leucanthemum vulgare* aggregate (Compositae, Anthemideae) in Central Europe. *Nordic Journal of Botany* **29**: 370–377.
- Oberprieler C, Hartl S, Schauer K, Meister J, Heilmann J. 2011b.** Morphological, phytochemical, and genetic variation in mixed stands and a hybrid swarm of *Senecio germanicus* and *S. ovatus* (Compositae, Senecioneae). *Plant Systematics and Evolution* **293**: 177–191.
- Oberprieler C, Konowalik K, Altpeter S, Siegert E, Lo Presti RM, Greiner R, Vogt R. 2012.** Filling of eco-climatological niches in a polyploid complex – A case study in the plant genus *Leucanthemum* Mill. (Compositae, Anthemideae) from the Iberian Peninsula. *Flora* **207**: 862–867.
- 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**: 1107–1118.
- Oberprieler C, Greiner R, Konowalik K, Vogt R. 2014.** The reticulate evolutionary history of the polyploid NW Iberian *Leucanthemum pluriflorum* clan (Compositae, Anthemideae) as inferred from nrDNA ETS sequence diversity and eco-climatological niche-modelling. *Molecular Phylogenetics and Evolution* **70**: 478–491.
- Oberprieler C, Konowalik K, Fackelmann A, Vogt R. 2018.** Polyploid speciation across a suture zone: Phylogeography and species delimitation in S French *Leucanthemum*

- Mill. representatives (Compositae-Anthemideae). *Plant Systematics and Evolution* **304**: 1141–1155.
- Oberprieler C, Hassanpour H, Sonboli A, Ott T, Wagner F. 2019.** Multi-locus phylogenetic reconstructions reveal ample reticulate relationships amongst genera in Anthemideae subtribe Handeliinae (Compositae). *Plant Systematics and Evolution*. doi: 10.1007/s00606-019-01588-0.
- Ochsmann J. 2000.** Morphologische und molekularsystematische Untersuchungen an der *Centaurea stoebe* L.-Gruppe (Asteraceae-Cardueae) in Europa. *Dissertationes Botanicae* **324**.
- Oswald BP, Nuismer SL. 2010.** Neopolyploidy and diversification in *Heuchera grossulariifolia*. *Evolution* **65**: 1667–1679.
- Otto SP. 2007.** The evolutionary consequences of polyploidy. *Cell* **131**: 452–462.
- Pante E, Abdelkrim J, Viricel A, Gey D, France SC, Boisselier MC, Samadi S. 2015.** Use of RAD sequencing for delimiting species. *Heredity* **114**: 450–459.
- Paris JR, Stevens JR, Catchen JM. 2017.** Lost in parameter space: a road map for STACKS. *Methods Ecology and Evolution* **8**: 1360–1373.
- Parisod C, Holderegger R, Brochmann C. 2010.** Evolutionary consequences of autopolyploidy. *New Phytologist* **186**: 5–17.
- Parisod C. 2012.** Polyploids integrate genomic changes and ecological shifts. *New Phytologist* **193**: 297–300.
- Paterson HEH. 1985.** The recognition concept of species. In: Vrba ES, ed. *Species and Speciation*. Pretoria: Transvaal Museum, 21–29.
- Paun O, Fay M, Soltis DE, Chase MW. 2009.** Hybrid speciation in angiosperms: Parental divergence drives ploidy. *New Phytologist* **182**: 507–518.
- Poland JA, Brown PJ, Sorrells ME, Jannink JL. 2012.** Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7**: e32253.
- Pritchard J, Stephens M, Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **6**: 945–959.
- Przywara L. 1974.** Badania biosystematyczne nad gatunkiem zbiorowym *Leucanthemum vulgare* z obszaru Polski. *Fragmenta Floristica et Geobotanica* **20**: 413–470.
- Pustahija F, Brown SC, Bogunic F, Basic N, Muratovic E, Ollier S, Hidalgo O, Bourge M, Stevanovic V, Siljak-Yakovlev S. 2013.** Small genomes dominate in plants growing on serpentine soils in West Balkans, an exhaustive study of 8 habitats covering 308 taxa. *Plant and Soil* **373**: 427–453.

- R Development Core Team. 2017.** *R: a language and environment for statistical computing.*, Vienna, Austria: Foundation for Statistical Computing.
- Raj A, Stephens M, Pritchard JK. 2014.** FASTSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**: 573–589.
- Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014.** *TRACER v1.6*. <http://beast.bio.ed.ac.uk/Tracer>. [accessed 7 November 2016].
- Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA. 2018.** Posterior summarisation in Bayesian phylogenetics using TRACER 1.7. *Systematic Biology* **67**: 901–904.
- Ramsey J, Schemske DW. 1998.** Pathways, mechanisms, and rates of polyploidy formation in flowering plants. *Annual Review of Ecology, Evolution, and Systematics* **29**: 467–501.
- Ramsey J, Schemske DW. 2002.** Neopolyploidy in flowering plants. *Annual Review of Ecology, Evolution, and Systematics* **33**: 589–639.
- Ramsey J, Robertson A, Husband B. 2008.** Rapid adaptative divergence in new world *Achillea*, an autopolyploid complex of ecological races. *Evolution* **62**: 639–653.
- Ramsey J. 2011.** Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences, USA* **108**: 7096–7101.
- Rannala B, Yang ZH. 2003.** Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**: 1645–1656.
- Rannala B, Yang Z. 2013.** Improved reversible jump algorithms for Bayesian species delimitation. *Genetics* **194**: 245–253.
- Rannala B. 2015.** The art and science of species delimitation. *Current Zoology* **61**: 846–853.
- Ree RH, Hipp AL. 2015.** Inferring phylogenetic history from restriction site associated DNA (RADseq). In: Hörandl E, Appelhans MS, eds. *Next-Generation Sequencing in Plant Systematics*. Bratislava, Slovakia: International Association for Plant Taxonomy, 1–24.
- Reeder J, Knight R. 2010.** Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nature Methods* **7**: 668–669.
- Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016.** VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.
- Rosen DE. 1979.** Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. *Bulletin of the American Museum of Natural History* **162**: 267–376.

- Rydin C, Pedersen KR, Friis EM. 2004.** On the evolutionary history of *Ephedra*: Cretaceous fossils and extant molecules. *Proceedings of the National Academy of Sciences, USA* **101**: 16571–16576.
- Saitou N, Nei M. 1987.** The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.
- Sang T, Crawford D, Stuessy T. 1997.** Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* **84**: 1120–1136.
- Sang T, Pan J, Zhang D, Ferguson D, Wang C, Paun K-Y, Hong D-Y. 2004.** Origins of polyploids: an example from peonies (*Paeonia*) and a model for angiosperms. *Biological Journal of the Linnean Society* **82**: 561–571.
- Scascitelli M, Whitney KD, Randell RA, King M, Buerkle CA, Rieseberg LH. 2010.** Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Molecular Ecology* **19**: 521–541
- Schall M. 2019.** Cytometric investigations in the *Rhodanthemum arundanum*-group. Unpublished.
- Scheunert A, Heubl G. 2014.** Diversification of *Scrophularia* (Scrophulariaceae) in the Western Mediterranean and Macaronesia – phylogenetic relationships, reticulate evolution and biogeographic patterns. *Molecular Phylogenetics and Evolution* **70**: 296–313.
- Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. 2015.** The IMAGEJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development* **82**: 518–529.
- Schönswetter P, Stehlik I, Holderegger R, Tribsch A. 2005.** Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology* **14**: 3547–3555.
- Schneider R, Grosschedl R. 2007.** Dynamics and interplay of nuclear architecture, genome organization and gene expression. *Genes and Development* **21**:3017–3043.
- Schwarz G. 1978.** Estimating the dimension of a model. *Annals of Statistics* **6**: 461–464.
- Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2017.** Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecology and Evolution* **8**: 907–917.
- Shaw J, Lickey EB, Schilling EE, Small RL. 2007.** Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in

- angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**: 275–288.
- Simmons MP, Ochoterena H. 2000.** Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49**: 369–381.
- Slatkin M, Maddison WP. 1989.** A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**: 603–613.
- Slovák M, Kučera J, Závěská E, Vďačný P. 2014.** Dealing with discordant genetic signal caused by hybridisation, incomplete lineage sorting and paucity of primary nucleotide homologies: A case study of closely related members of the genus *Picris* subsection *Hieracioides* (Compositae). *PLoS ONE* **9**: e104929.
- Smith SA, Brown JW, Walker JF. 2018.** So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLoS ONE* **13**: e0197433.
- Solís-Lemus C, Knowles LL, Ané C. 2015.** Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* **69**: 492–507.
- Soltis DE, Kuzoff RK. 1995.** Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* **49**: 727–742.
- Soltis DE, Buggs RJA, Doyle JJ, Soltis PS. 2010.** What we still don't know about polyploidy. *Taxon* **59**: 1387–1403.
- Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA. 2014.** Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al.* (2011). *New Phytologist* **202**: 1105–1117.
- Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015.** Polyploidy and genome evolution in plants. *Current Opinion in Genetics and Development* **35**: 119–125.
- Spooner D. 2009.** DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. *American Journal of Botany* **96**: 1177–1189.
- Spriggs EL, Eaton DAR, Sweeney PW, Schlutius C, Edwards EJ, Donoghue MJ. 2019.** Restriction-site-associated DNA sequencing reveals a cryptic *Viburnum* species on the North American Coastal Plain. *Systematic Biology* **68**: 187–203.
- Stebbins GL. 1947.** Types of polyploids: their classification and significance. *Advances in Genetics* **1**: 403–429.
- Stebbins GL. 1950.** *Variation and Evolution in Plants*. New York, USA: Columbia University Press.

- Stephens M, Smith NJ, Donnelly P. 2001.** A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**: 978–989.
- Stephens M, Scheet P. 2005.** Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics* **76**: 449–462.
- Stuessy TF. 2009.** *Plant Taxonomy – The Systematic Evaluation of Comparative Data*. New York, USA: Columbia University Press.
- Stuessy TF, Crawford DJ, Soltis DE, Soltis PS. 2014.** *Plant Systematics: the Origin, Interpretation, and Ordering of Plant Biodiversity*. Königstein, Germany: Koeltz Scientific Books.
- Stuessy TF, Weiss-Schneeweiss H. 2019.** What drives polyploidization in plants? *New Phytologist*. doi: 10.1111/nph.15784.
- Suchard MA, Rambaut A. 2009.** Many-core algorithms for statistical phylogenetics. *Bioinformatics* **25**: 1370–1376.
- Sukumaran J, Knowles LL. 2017.** Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences, USA* **114**: 1607–1612.
- Swofford DL. 2003.** *PAUP* v4.0: Phylogenetic Analysis Using Parsimony (*and other methods)*. Sunderland, MA, USA: Sinauer Associates.
- Taberlet P, Gielly L, Pautou G, Bouvet J. 1991.** Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105–1109.
- Takahashi K, Hanyu M. 2015.** Hybridization between alien species *Rumex obtusifolius* and closely related native vulnerable species *R. longifolius* in a mountain tourist destination. *Scientific Reports* **5**: 13898.
- Than C, Ruths D, Nakhleh L. 2008.** PHYLONET: A software package for analyzing and reconstructing reticulate evolutionary histories. *BMC Bioinformatics* **9**: 322.
- Thompson JD, Lumaret R. 1992.** The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends in Ecology and Evolution* **7**: 302–307.
- Tomasello S, Álvarez I, Vargas P, Oberprieler C. 2015.** Is the extremely rare Iberian endemic plant species *Castrilanthemum debeauxii* (Compositae, Anthemideae) a ‘living fossil’? Evidence from a multi-locus species tree reconstruction. *Molecular Phylogenetics and Evolution* **82**: 118–130.
- Toprak Z, Pfeil BE, Jones G, Marcussen T, Ertekin AS, Oxelman B. 2016.** Species delimitation without prior knowledge: DISSECT reveals extensive cryptic speciation

- in the *Silene aegyptiaca* complex (Caryophyllaceae). *Molecular Phylogenetics and Evolution* **102**: 1–8.
- Tribsch A, Schönswetter P. 2003.** Patterns of endemism and comparative phylogeography confirm palaeo-environmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon* **52**: 47–497.
- Van Tuyl JM, De Vries JN, Bino RJ, Kwakkenbos TAM. 1989.** Identification of 2n-pollen producing interspecific hybrids of *Lilium* using flow-cytometry. *Cytologia* **54**: 737–745.
- Vargas OM, Ortiz EM, Simpson BB. 2017.** Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist* **214**: 1736–1750.
- Via S. 2009.** Natural selection in action during speciation. *Proceedings of the National Academy of Sciences, USA* **106**: 9939–9946.
- Villard M. 1970.** Contribution à l'étude cytotaxonomique et cytogénétique du genre *Leucanthemum* Adans. ex Briq. et Cav. *Berichte der Schweizerische Botanische Gesellschaft* **80**: 96–188.
- Vogt R. 1991.** Die Gattung *Leucanthemum* Mill. (Compositae-Anthemideae) auf der Iberischen Halbinsel. *Ruizia* **10**: 1–261.
- Vogt R. 1994.** *Rhodanthemum laouense* (Compositae, Anthemideae), a new species from Morocco. *Willdenowia* **24**: 91–96.
- Vogt R, Oberprieler C. 2008.** Chromosome numbers of North African phanerogams. VIII. More counts in Compositae. *Willdenowia* **38**: 497–519.
- Vogt R, Oberprieler C. 2012.** Chromosome numbers of North African phanerogams. X. Plants collected during Iter Mediterraneum V of OPTIMA in Morocco. *Annalen des Naturhistorischen Museums in Wien Serie B* **113**: 193–221.
- Vogt R, Konowalik K, Oberprieler C. 2018.** Karyological analysis reveals two new polyploid marguerite taxa (*Leucanthemum*, Compositae–Anthemideae) in S France and NW Italy. *Willdenowia* **48**: 221–226.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M et al. 1995.** AFLP: A new technique for DNA fingerprinting. *Nucleic Acid Research* **23**: 4407–4414.
- Wagner F, Härtl S, Vogt R, Oberprieler C. 2017.** “Fix Me Another Marguerite!”: Species delimitation in a group of intensively hybridizing lineages of ox-eye daisies (*Leucanthemum* Mill., Compositae-Anthemideae). *Molecular Ecology* **26**: 4260–4283.

- Wagner F, Ott T, Zimmer C, Reichhart V, Vogt R, Oberprieler C. 2019.** 'At the crossroads towards polyploidy': Genomic divergence and extent of homoploid hybridisation are drivers for the formation of the ox-eye daisy polyploid complex (*Leucanthemum*, Compositae-Anthemideae). *New Phytologist*. doi: 10.1111/nph.15784.
- Wagner ND, 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**: 8243–8255.
- Wang W. 2004.** On the origin and development of *Artemisia* (Asteraceae) in the geological past. *Botanical Journal of the Linnean Society* **145**: 331–336.
- Weising K, Nybom H, Wolff K, Kahl G. 2005.** *DNA Fingerprinting in Plants: Principles, Methods, and Applications*. Boca Raton, USA: CRC Press.
- Weiss-Schneeweiss H, Emadzade K, Jang TS, Schneeweiss GM. 2013.** Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetic and Genome Research* **140**: 137–150.
- Wendel JF, Schnabel A, Seelanan T. 1995.** Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences, USA* **92**: 280–284.
- White TJ, Bruns T, Lee S, Taylor J. 1990.** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR Protocols: A Guide to Methods and Applications*. San Diego, USA: Academic Press, 315–322.
- Wilcox BH, Harcourt P. 1982.** Cytological and hybridization studies in *Leucanthemum* (Compositae-Anthemideae) from North Africa. *Plant Systematics and Evolution* **139**: 179–195.
- Wilkerson MD, Hayes DN. 2010.** ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**: 1572–1573.
- Winge Ø. 1917.** The chromosomes: Their number and general importance. *Comptes-rendus des travaux du Laboratoire Carlsberg* **13**: 131–275.
- Winter DJ, Trewick SA, Waters JM, Spencer HG. 2016.** The genealogical sorting index and species delimitation. *bioRxiv*. doi: <http://dx.doi.org/10.1101/036525>.
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009.** The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* **33**: 13875–13879.

- Yang Z, Rannala B. 2010.** Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences, USA* **107**: 9264–9269.
- Yang Z, Rannala B. 2014.** Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution* **31**: 3125–3135.
- Yang Z. 2015.** The BPP program for species tree estimation and species delimitation. *Current Zoology* **61**: 854–865.
- Young ND, Healy J. 2003.** GAPCODER automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* **4**: 6.
- Zachos FE. 2016.** *Species Concepts in Biology: Historical Development, Theoretical Foundations and Practical Relevance*. Basel, Switzerland: Springer International Publishing.
- Zhang C, Zhang DX, Zhu T, Yang Z. 2011.** Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology* **60**: 747–761.

Acknowledgments

No PhD thesis stands alone without support of many people!

First of all, I would like to thank my supervisor Prof. Dr. Christoph Oberprieler and my second mentor Dr. Robert Vogt for introducing me to the fascinating field of plant systematics, including lab-, computer-, herbarium- and field-work. Thank you for teaching me the art and science of plant hunting in the course of several excursions to France (2013, 2015), Italy (2015), Spain (2016) and Morocco (2017, 2018).

For their professional contribution to the present thesis, I would like to thank my first mentor Prof. Dr. Rainer Merkl, my colleagues Dr. Agnes Scheunert, Tankred Ott and Ulrich Lautenschlager and all students who accompanied me over the last years, namely Sabine Härtl, Verena Reichhart, Claudia Zimmer and Maximilian Schall.

Likewise, I would like to thank Anja Heuschneider, Nicole Schmelzer and Gudrun Karch for their technical assistance in the molecular laboratory of the Evolutionary and Systematic Botany Group at Regensburg University.

I sincerely thank Prof. Dr. Günther Heubl, who kindly accepted to act as referee of my thesis, as well as Prof. Dr. Christoph Schubart, Prof. Dr. Klaus Grasser and Prof. Dr. Christoph Reisch for their participation in the examination committee.

For providing financially support, I thank the German Research Foundation and the IPID4all program of the German Academic Exchange Service (DAAD).