Hybrid Zones in *Rhododendron* subsection *Taliensia*

Tobias Marczewski

Doctor of Philosophy
The University of Edinburgh,
Royal Botanic Garden Edinburgh
2011

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Tobias Marczewski
Musselburgh, February 2011

Acknowledgements

First of all I want to thank Jane Droop, for enduring endless hours of statistical thought experiments, supporting me during the hardest times of this PhD, and just being there for me at all times, and my Mother who supported me during all my previous education, and without whose support I would never have had the opportunity to start this PhD.

I am grateful to my supervisors Mary Gibby, David Chamberlain and Richard Milne, for giving me the opportunity to start this project, for always providing advice and support when needed, and for giving me the freedom to direct this project according to my interests.

David Rankin for intruducing me to the leaf waxes of rhododendrons, and providing support and the equipment to carry out this part of the project.

I am indebted to the members of The Royal Horticultural Society's Rhododendron, Camellia and Magnolia Group, and Phillip de Spoelberch for funding the project. I truly appreciate their enthusiasm for rhododendrons and their warm reception of my work.

Additionally, I thank The Davis Expedition Fund for providing supplementary funding for the fieldwork carried out in 2008.

It was a pleasure to be in the field with Lian-Ming Gao and Jie Liu. Without their support the two collection trips would not have been possible.

I also want to thank the following people: Alex Clark, Jane Squirrel and Michelle Hollingsworth for sharing their lab experience; Keith Gardner, Pete Hollingsworth, Michael Möller and Peter Cripps, for helpful discussions about statistics; Alastair Droop for geeking out over R; Catherine Kidner for sharing her broad knowledge of leaf morphology in hybrids; and all the other PhD students at RBGE for their humour and companionship.

Lastly, I want to thank all the people in the open source community, whose contribution makes software such as GIMP[1], Inkscape[2], LaTeX[3], and R[4] freely available. Without these tools, this thesis would not exist in its present form.

[1] `http://www.gimp.org/`
[2] `http://inkscape.org/`
[3] `http://www.latex-project.org/`
[4] `http://www.r-project.org/`

Abstract

The investigation of hybrid zones has proven to be one of the most promising approaches to advance our understanding of species barriers, and to elucidate evolutionary processes involved in speciation. Due to the improvement of molecular techniques it will soon be possible to investigate the genetic composition of non-model species in much greater detail, and also include species that defy investigation using controlled laboratory conditions. To be able to draw further reaching conclusions about the generality of certain evolutionary factors, it is crucial to investigate a wide spectrum of organisms differing in traits, life histories and relatedness. This study investigates patterns of hybridisation between two pairs of closely related species in the genus *Rhododendron*. AFLP data for 346 loci, from twelve populations in total comprising 390 individuals, were obtained. Additionally, the abundance of three alkane components in the leaf waxes of 115 individuals was determined. For the species pair *R. clementinae* and *R. roxieanum* low levels of recent hybridisation were found, however, the wax composition of *R. roxieanum* var. *cucullatum* suggests historical introgression. Two types of hybrid zones were found for *R. aganniphum* and *R. phaeochrysum*, one mainly comprising F1 individuals, and the other frequent backcrosses to *R. aganniphum*. Furthermore, evidence for genomic incompatibilities at several loci for the two species will be presented, and hybrid identity of *R. aganniphum* var. *flavorufum* and *R. phaeochrysum* var. *agglutinatum* is suggested.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Relationships of organisms and how we group them

Taxonomy Most people when hearing the word taxonomy, meaning here the classification of organisms, will undoubtedly think of the scientific discipline. However, it is likely that most organisms classify their surroundings and environment to a lesser or greater extent. For example, it is beneficial to tell the difference between a possible predator and prey, or whether food is toxic or harmless. As trivial as it sounds, this ability is of vital importance for the successful interaction with the environment. This classification is necessarily based on the formation of groups of "similar" individuals, characterised by certain traits, as only in this way can a prediction regarding the properties of a previously unencountered individual be made. Therefore, the determining characteristics of a group must be assumed to be present in all members of the group, and additionally only in the members of this group. The importance of this perception of groups for the interaction of organisms can be exemplified by a short story of a young rabbit, just in the process of forming its own "taxonomical" system. Having in general established the group "plants" as food, it is just about to encounter *Urtica dioica*. Assuming the rabbit is able to recognise different types of plants, it might reconsider attributing the category "food" to the group "plants" as a whole. As this subdivision is likely to be relatively imprecise, the newly formed group might, in addition to *Urtica*, comprise species of *Lamium sp.*, and this will influence the interaction of the growing rabbit with its environment.

Species concepts In more or less the same manner humans established groups for organisms, and for a very long time the main characteristics for classification were largely of morphological nature. This corresponds to the typological species concept (TSC), which relies on selective character choice to establish groups. However, with the refinement of techniques to investigate differences between groups it became apparent that some groups, inately different for certain characteristics, do not have morphological distinguishing features [19]; while other can host a large variation of morphological characters (e.g. examples in [133, 152]). Therefore it was complemented by the biological species concept (BSC) [114], which defines species as groups that do not interbreed with each other. Or, as Mayr stated referring to populations:
"*If their relationship is that of interbreeding they are conspecific; if it is that of being reproductively isolated, they are different species.*"
If we would test the crossfertility of members of *Urtica* and *Lamium* we would probably find that they do not produce offspring, and this seems to hold for most of our species. Reassuringly, the two concepts mostly do not produce different groupings, and completely agree in 83% of cases, as Mayr showed, using a local North-American flora [114]. Additionally, with reproductive isolation (RI) as criterion, the BSC provides us with a testable hypothesis; however, it does not give an explanation for the basis of the grouping. The theoretical basis of why species can be joined into groups, and the pattern underlying their relatedness is provided by Darwin, as summarized by Levinton ([99], page 496):
"[...] *the ability to classify organisms stems from their relationship by descent* [and] *the descent involves a series of evolutionary steps, or intermediate forms, from ancestor to descendant.*"
For example in contrast to the rabbit we were able not only to establish that *Lamium* is a different group and does not sting, but also to deduce an evolutionary explanation for why it lacks this ability, and ask why it seems so close to *Urtica* morphologically, despite not being closely related. This shows that although one might argue a certain grouping concept is better than another, in reality most is gained by considering all, and asking where and why they differ.

Speciation and reproductive isolation Congruence of the BSC and the concept of emergence of new species by descent implies the divergence of individuals from one initially interbreeding group into subgroups that do not interbreed with each other. As soon as the interbreeding stops and RI has been attained, the two new groups would be classified as species, which from this point onwards evolve independently from each other. On an evolutionary timescale

this event would be represented by a dichotomous split; this is the basis for the reconstruction of phylogenetic trees, in which any two species share exactly one common ancestor. Although RI is the key character, it is still to a certain extent unclear when and why it is attained in the process of speciation [121, 182]. Given the wide variety of possible barriers to geneflow [127, 139, 148] it is also unlikely that one mechanism will be the only important one for all species. Understanding their population structure helps in elucidating some of the important factors, as stated by Templeton ([170], page 24):

"*Thus, there is also no universal joint pattern relative to speciation. However, predictable patterns and differences do emerge for particular groups of organisms, and population-genetic considerations are apparently important determinants of these patterns.*"

Some of the mechanisms are easier to appreciate than others, and clarification of the relative importance of the following possible roles for RI is still lacking: is RI mostly responsible for the process of speciation, or is it often a by-product of the divergence of species? Scenarios that can lead to a genetic barrier are of fundamentally different nature, but the outcomes can be similar, and therefore difficult to distinguish. Separating populations in space, with a strong enough physical barrier, can impede interbreeding, and random genetic drift will eventually lead to significantly different gene pools. Darwin saw the process driven by adaptation to different habitats; this is still widely considered to be one major force, and under the right conditions, ecological requirements can be sufficient to drive populations apart [60, 97, 148]. At some stage during this ecological speciation RI has to be acquired, but the relationship of the ecological factors initiating speciation, and factors responsible for the development of RI are still unclear [26, 128], unless they are very closely linked; e.g. if the ecological variable is a pollinator, discriminating between the two groups [25]. If RI is strictly necessary, or selected for, in every case is, however, doubtful [65, 178], and the strength of RI found in nature can vary considerably [148].

Although the environment likely provides the selective pressures for groups of organisms to differentiate, this selection acts on differences in the genomic composition arising through mutations. Already Darwin realised that RI could not be selected for, as this would require "successive profitable degrees of sterility" (Darwin 1859, p. 245; as cited in [134]). A mutation leading to incompatibility, or hybrid sterility, necessarily arises first in a single individual. Due to the intrinsic effect of this mutation, however, it will by definition have a selective disadvantage, impeding it from spreading through a population, or resulting in a sterile individual to begin with [126]. For this reason incompatibility

cannot generally arise through the mutation of a single gene, and Bateson, Dobzhansky and Muller established that hybrid sterility and inviability are caused by sets of interacting complementary genes [126]. If two or more genes are involved in the incompatibilities, they can arise unhindered by a selective disadvantage. According to the Dobzhansky-Muller model, considering two loci with each two alleles (a/A and b/B; where a & b are ancestral), two scenarios are possible [134]. First, derived alleles go to fixation at different loci in two lineages; $aAbb \rightarrow AAbb \mid aaBb \rightarrow aaBB$ [$\overline{AA}bb \,/\, aa\overline{BB}$][1], where the two derived alleles (A and B)are incompatible. Second, after one derived allele (B) has become fixed in one lineage, another derived allele can arise at a second locus, which is incompatible to the ancestral allele (b); $aaBb \rightarrow aaBB \rightarrow AaBB \rightarrow AABB$ [$aa\overline{bb} \,/\, \overline{AA}BB$][1], where a derived allele (A) is incompatible to an ancestral one (b).

Bateson-Dobzhansky-Muller incompatibilities are likely *the* major mechanism for the emergence of hybrid sterility in animals, or as Orr stated ([126], p.1333): "*Although genetecists dissecting the basis of postzygotic isolation continue to squabble over many details, we all agree that hybrid sterility and inviability in animals is caused by sets of complementary genes.*"
They also certainly play an important role in hybrid incompatibilities in plants [134, 146, 172], however, in contrast to animals, their relative importance is not well understood, especially when compared to the role of chromosomal rearrangements [126, 146]. Furthermore, the actual functions of the genes involved, and reasons for the incompatibilities remain largely unknown.

Speciation and geneflow Sometimes a genetic event can lead to nearly instantaneous reproductive incompatibility of subgroups, which can then lead to the populations behaving as in allopatry, and drifting apart [40, 170]. Examples for this are chromosomal inversions or polyploidy [13, 40, 97]. The latter example taps into another problem, regarding reconstruction of the descent, which we defined as dichotomous. If the reproductive barriers between species are not always strong enough to impede the production of viable offspring and exchange of genetic material, or even allow the formation of a new reproductively isolated group, we have to refine the reconstruction, and definition, of species relationships regarding the descent. There is substantial evidence that reproductive barriers are frequently not strong enough to prevent interbreeding completely, for example 25 % of plant species are known to hybridise [111], and numerous examples for hybrid speciation have been reported as well [34]. If hybrid species arise

[1]The incompatible alleles are highlighted by a bar.

through polyploidy, the criterion of RI is mostly fulfilled, and these hybrids do not represent bridges for geneflow between parent species; hence they rather present a challenge for the interpretation of the descent rather than to the species criterion of RI. However, the evidence presented for homoploid hybrid speciation [59, 68, 142] and apparent geneflow between long established species [41, 111], suggests that there is still much more research required to understand all facets of species identity and possible evolutionary histories.

1.2 The role of hybridisation in the context of the evolutionary process

Historical background Whether or not hybridisation plays a significant role in the evolution of organisms comes down firstly to whether the hybrid offspring are viable and will contribute to generations that follow, and secondly whether this contribution occurs frequently enough, on an evolutionary timescale, to be important. Ever since Linnaeus raised the idea that hybridisation could play an evolutionary role [142], the debate about these questions has been ongoing. Hybridisation was widely assumed not to be an important force, particularly by zoologists [8]. Most advocates arguing for its importance were botanists [3, 158], which might not be too surprising taking into account that 25 % of plant species and only 10 % of animal species form hybrids [111]. Hybrids were seen by many biologists as a mere by-product without any further evolutionary consequences, as is illustrated by a statement from Darwin (1859 page 246, as cited in [8]):

"*Pure species have of course their organs of reproduction in a perfect condition, yet when intercrossed they produce either few or no offspring. Hybrids, on the other hand, have their reproductive organs functionally impotent.*"

That this interpretation was still adopted by scientists a century later is evidenced by a reaffirming statement from Mayr (1963 page 133 as cited in [8]):

"*... The majority of* [...] *hybrids are totally sterile.* [...] *Even those hybrids that produce normal gametes in one or both sexes are nevertheless unsuccessful in most cases and do not participate in reproduction* [...]. *When they do backcross to the parental species, they normally produce genotypes of inferior viability that are eliminated by natural selection. Successful hybridization is indeed a rare phenomenon among animals.*"

According to this interpretation, hybrid zones are generally narrow regions also called tension zones [16], maintained by a balance between random dispersal and selection against hybrids. Although some hybrid zones seem to conform to these dynamics [63], others do not [61, 69, 119].

The environment The importance of the environment for hybridization was first recognised by Kerner (1894-1895, as cited in [145]). He established the idea that by providing niches available for occupation by arising hybrids, the habitat is crucial to give them a possibility to establish themselves along with their parents. As organisms are normally very well adapted to their habitat, it has been argued that they would mostly outcompete hybrids and that therefore the latter are generally associated with unstable, newly arising, or disturbed environments [2, 3, 26, 86, 112, 145]. The hybrid zones between two (or more) species should in this case, depending on the amount of change of the habitat and the ecological amplitude of the species, be unstable over longer time scales [16, 27, 113], and will migrate within the species distribution range or eventually disappear. This ecospatial component can be crucial to understand hybrid zone occurrences and may be used to predict them [94]. However, some authors argue for a major importance of gene-interactions in this dynamic process rather than direct environment-dependence [16, 87, 92]. Nonetheless, the impact of the environment, as a main source of selective pressures is generally considered to play a, if not the, key role in speciation processes [148, 150] not only regarding hybridization [3, 9, 41, 101].

The observations made regarding the impact of the habitat led several scientists (e.g. [3, 4, 9, 100, 154]) to argue for an importance of hybridization in the evolutionary history of organisms; especially trying to explain speciation bursts seemingly occurring with changing environmental conditions and drastic disturbances along the palaeontological timescale [158]. One main point was to explain the observed rates of evolution and species radiations [154], as pointed out by Anderson and Stebbins ([4], page 378):

"*To the student of hybridization, however, another factor which may have contributed largely to these evolutionary bursts presents itself.* [...] *Populations*[2] *having very different genetic systems of adaptation* [..., and assuming that certain hybrids are viable,] *then new adaptive systems, adapted to new ecological niches, may arise relatively quickly* [through hybridization]."

Introgression The concept of introgression, the transfer of genes from one species to another through hybridization and repeated backcrossing, was first introduced by Anderson and Hubricht (1938, as cited in [3]). This discovery

[2]Hybridization can have several meanings for evolutionary biologists. The term "hybrid" can be restricted to organisms formed by cross-fertilization between individuals of different species. Alternatively, hybrids can be defined more broadly as the offspring between individuals from populations "which are distinguishable on the basis of one or more heritable characters" ([145], page 600)

suggests that the hybrid itself does not necessarily need to survive as a "new species" to affect the evolutionary process, because the transfer of genetic material could itself be beneficial through the enrichment of genetic variation within populations, and therefore affect evolutionary performance [3, 158]. As Anderson mentioned ([3], page 3):

"*Such hybridization is cryptic and only by very specialised techniques can we measure its exact importance.*"

Molecular methods offer such specialized techniques, and it is now accepted that introgression is a widespread phenomenon in plants (e.g. [41, 111, 143]). Furthermore it has been shown that, through hybridisation, fitness-relevant alleles can be transferred into a different genetic background [41, 112]. However, introgression is not always beneficial for both participants. For example a rare species can experience problems when hybridising with a more abundant one. Not only do individuals of the rarer species frequently experience a reduction of the potential to replace themselves through fertilization by conspecific pollen, therefore affecting the population growth rate [98], but they are also exposed to an invasion by foreign genes when hybrids contribute extraordinarily to fecundity [28]. This leads to a higher extinction probability when rare species are confronted with hybridization and/or can decrease the persistence time of the rare species before extinction [181]. This is especially alarming as human disturbance of habitats creates beneficial conditions for hybridisation [20] along with destruction of the species habitats.

Fitness and hybrid speciation Although both theory [15, 88, 104, 154], and some palaeontological data [140] indicate that hybridization might have played a role in the evolutionary history of many organisms, it is still not clear how important this impact is [9, 12]. Most case studies, of course, can only investigate contemporary phenomena; to extrapolate from these results, crucial factors in estimating the evolutionary importance are: how often can hybridization produce offspring that has a long-term fitness advantage; how likely are these events compared with other evolutionary mechanisms like adaptations arising through mutation; and, can they persist against random forces like drift [15]? Lewontin and Birch [100] were the first to show, with crossing experiments on two species of *Dacus*, that hybrids indeed can exhibit selective advantages over their parents under certain conditions and hence can be fitter than either parent in certain environments. Since then extensive research in the dynamics of hybrid zones has broadened our knowledge of the fitness behaviour of hybrids. Although offspring of hybrid origin are generally less

fit than either parent [15, 29, 145, 159], many exceptions have been reported where at least some genomic combinations, or introgression seems to present a selective advantage [9, 75, 145, 153, 162]. Furthermore, several studies provided evidence that introgression of alleles from one species to another could lead to a larger ecological amplitude, and therefore a range expansion of the recipient species [5, 41, 102, 112]. In some cases hybrids show extreme phenotypes with regards to their parents, also called transgressive segregation [143]. When this occurs, hybrids can exhibit completely new traits [104], not seen in the parents, enabling them to grow in completely new habitats, or at least have a considerably higher fitness in these than their parents [8, 30, 46]. This illustrates nicely one of the misconceptions regarding hybrids; that they are typically intermediate between their parents [141]. Furthermore, this suggests that an exclusively genic view of trait acquisition neither reflects all adaptive processes in hybrids, nor all possible mechanisms of speciation [12]. One of the most substantial impacts on the evolution of organisms is unquestionably the emergence of new species through hybridization. The main problem in this scenario is that hybrids usually occur sympatrically with regard to their parents and therefore it is believed that they have to develop mechanisms of reproductive isolation to allow for speciation [26, 40, 170]. One well documented mechanism in plants which achieves this nearly instantaneously is allopolyploidy (Karpechenko 1928 as cited in [40]). Indeed the frequency of polyploid hybridisation seems to be considerably higher than that of homoploid hybrid speciation [129], and exceeding a certain level of divergence between the parental species, it may be the only form of hybrid speciation still possible [34]. Interestingly, only herbaceous species are reported to form allopolyploids for reasons not really understood [129]. For that reason this form of hybrid speciation is not applicable to a wide range of angiosperm taxa, hence it is especially desirable to expand the knowledge about mechanisms governing homoploid hybrid speciation and their likeliness to occur in natural environments. Dobzhansky's view of allopolyploidy was that ([40] page 314):

"[...] *species formation by polyploidy is confined to certain groups of organisms, mostly plants. The more general method of speciation is through a gradual accumulation of genic and chromosomal changes.*"

Homoploid hybrid speciation could therefore be viewed as being closer to the general method, and it certainly offers more scientific opportunities to investigate general speciation mechanisms. However, relatively few examples are documented where homoploid hybrid ancestry is confirmed [68, 142], mainly because of a lack of molecular evidence. By far the best investigated cases are *Helianthus* [12, 88, 90, 104, 173], and *Iris* [6, 50, 78, 86, 112]. Both study systems suggest the overall

importance of selection for adaptive environmental traits [86, 88, 104], allowing the hybrid to establish itself in a new habitat, along with fertility selection [88] as main mechanism for hybrid speciation. Furthermore, it has been shown that when this process occurs it happens rather quickly [173].

Summary Natural hybridisation is a relatively common feature of vascular plant species [49, 111] and research in the field has made considerble advances in recent years, mainly through the deployment of newly arising techniques [12]. Several homoploid hybrid species have been confirmed [68], and studies of contemporary hybrid populations have been used to identify selective forces promoting their establishment [88, 104]. Furthermore, evidence has been presented that hybrids are not intrinsically less fit than their parents [7, 145, 153], and that introgression can be advantageous [89, 112]. Nonetheless, the overall importance of hybridization for evolutionary history and speciation is still debated [7, 12], especially as contemporary hybridisation rates vary significantly between plant families [180]. Although hybridisation is frequent among organisms, the strength of barriers impeding geneflow vary considerably, due to differnet interactions of environment, genetic composition, and population structure [41, 63, 75, 87, 101, 169]. The environment seems crucial to provide niches for the successful establishment of hybrids [68, 104], or permit geneflow [10, 171]. However, it is also a main factor restricting geneflow despite hybridisation [37, 87, 101, 119]. The conditions governing the likelihood of a certain outcome remain poorly understood, mainly due to the lack of information about genome-environment interactions [7, 178] and intrinsic genetical incompatibilities [11, 12, 90]. Furthermore, the impact of population structure remains unclear, as it interacts with other selective forces [41], and can lead to different outcomes in different contact zones of the same species [75, 169]. And if introgression is frequent, the question of whether it is mostly neutral, or often adptive, still remains [2]. As has been stated by Baack and Rieseberg (2007) [12] page 513):
"*The extent of introgression and hybrid speciation is unclear: occasional hybridization may not lead to permanent genetic exchange* [...]."
To answer this important question more information is needed regarding the impact of introgression for transfer of adaptive traits as opposed to other speciation processes like divergent selection [148]. Because hybridisation is related to the genetic divergence of the involved species [111, 134], it also is crucial to assess the impact of geneflow in newly arising species, for example during radiation events, as these are believed to have been a major mode of generating new diversity [154]. Does hybridisation facilitate this process, and how can young

species remain distinct and diverge despite recurrent geneflow? More case studies, deploying modern molecular techniques, are needed to clarify the the strengths and nature of species barriers. These studies should span a variety of plant families, because significant differences in hybridisation rates do occur between them [180]. This will enable us to assess the impact of hybridisation and inter-species geneflow on evolutionary processes and our present-day diversity, along with highlighting differences between taxa and understanding the reasons for these differences.

1.3 Introduction to the genus *Rhododendron*

The genus *Rhododendron* is widely-distributed throughout the world with the exception of Africa and South America, and has a centre of diversity in the Sino-Himalayan region, particularly the eastern Himalaya [33]. It is the largest genus in the family Ericaceae, with over 1000 described species [33], in eight recognised subgenera. Members of the genus are all woody perennials, however, they show a great variety of growth forms, from dwarf shrubs and epiphytic species to medium sized trees. Under normal circumstances they are obligate outcrossers and one study showed that members can suffer substantially from biparental inbreeding depression (negative reproductive impact of kinship) [77]. The species depend on insects as pollinators, mostly bumble bees. Occasional wind pollination seems unlikely, as groups of pollen grains are normally connected by sticky viscin threads [91], reducing the potential of wind dispersal. Furthermore, rhododendron flowers show elaborate syndromes to increase outcrossing. Pollen is largely selectively released in the presence of insects; this is achieved through buzz-pollination, in which a range of resonance frequencies promotes the release of pollen and viscin threads from the anthers [91]. Additionally flowers are protandrous, evidenced by dry emerging stigmas which become sticky and receptive after the flowers have been open for a few days [125].

The geology of the Himalayas The centre of diversity of the genus is located in the Himalayas, which is one of the 25 most species rich areas on the planet [123]. Its flora comprises over 12,000 plant species, 3500 of them endemic [123]. This phenomenon is tightly linked to the highly active tectonics in this region. The Indian plate collided here with the Eurasian plate in the early Tertiary, and the two plates still converge with an estimated speed of 50 mm/a [165]. Combining computer models with geological data

suggest that since approximately 30 Ma Before Present (BP) the uplift of the Himalayas progressed through three major stages (30–15 Ma BP, 10–4 Ma BP and 4 Ma BP –present) [165]. During the last of these stages the uplift rate reached values above 2 mm/a in the High Himalayas, resulting in an effective uplift of over 2 km/Ma [165]. This geological activity necessarily confronted organisms with extraordinary environmental changes. Not only were distribution ranges exposed to repeated fragmentation, along with frequent genesis of new habitats, but the climatic conditions and weather systems were also profoundly affected by the emergence of mountain ranges. For example, data suggests that monsoon circulations changed abrubtly three times after 7.6 Ma BP, with the last major change between 1.2–0.9 Ma BP, and that they experience continous weakening since 0.6 Ma BP [42, 163]. Not surprisingly, these factors have been shown to correlate strongly with repeated species composition changes in certain areas [183]. This geological history results in some contemporary groups of plants that have undergone a substantial radiation event [106, 176], or still seem to be actively speciating, as evidenced by morphologically often poorly defined species.

Subgenus Hymenanthes Subgenus Hymenanthes comprises 225 species, divided into 24 subsections [32]. Many of these occur exclusively in China, with a few reaching into Nepal, India, Japan and other Southeast Asian countries; the only exception to this pattern is subsection pontica. Certain sections of *Rhododendron* have posed problems, morphologically, as evidenced by significantly different treatments [32, 157], as well as phylogenetically. The genus *Rhododendron* forms a well-supported clade [95], and the placement of the subgenera is fairly robust [64, 96]. Resolution within the subgenera, however, is often poor, and within subgenus Hymenanthes most relationships remain unresolved [118]. Furthermore, gene trees obtained for the subgenus have been shown to be incongruent, possibly due to geneflow between species [47], or through a rapid radiation [118]. Additional support for potentially permeable species barriers comes from the extensive hybridisation that seems to occur in the wild within certain complexes [32]. Some of these hybrid zones have been confirmed with molecular methods [36, 110, 117, 184], and often seem to comprise large numbers of F1 hybrids [119, 185], but other cases have been reported [110]. Generally weak reproductive barriers are for example indicated by the occurrence of the hybrid *R. agastum* throughout the overlap of the parental species ranges [185]. Nonetheless F1 dominated hybrid zones might represent effective barriers to geneflow [119]. However, no conclusive research with regards

to the long-term permeability of reproductive barriers, or ecological factors, has been published so far.

Subsection *Taliensia* Subsection *Taliensia*, the focus of this study, is a conglomerate of 38 species, which were formerly grouped into four different subseries [32]. The result is "*a very diverse subsection in which some subdivision may be justified*" ([32], page 333). One of the difficulties opposing a more detailed morphological grouping is the absence of good qualitative characters separating the species. Floral morphology is varied but inconclusive, and hence the most important, and variable, characters available are the indumentum and shape of the leaves. The indumentum refers to the composition of hairs (different hair types) on the underside of the leaves. It can be unistrate or bistrate (if certain types of hairs are considerably longer, and form a second layer), and ranges from sparse to dense, and if dense often felted or compacted. Although these leaf characters normally allow easy recognition of "pure species types" in the field, the approach is rather holistic, as the characters and especially leaf shape can vary considerably within one species. Work with herbarium specimens is hindered because pressing can sometimes change the appearance of the indumentum and other leaf characteristics, and dichotomous keys can be ambiguous. Furthermore, many intermediates occur, and assessment of a single specimen, especially without knowledge of other species occurring in sympatry at the site, is very challenging. This occurrence of intermediate forms is, together with the scarcity of characters, the second major obstacle in establishing conclusive relationships between the species. Many of these intermediate forms are likely to result from cross-species fertilisation, as is common in the whole subgenus (see above). In subsection *Taliensia* the problem is exemplified by the description of, frequently several, varieties of the species; often these varieties exhibit intermediate characteristics of two distinct species, and hybrid origin seems likely.

1.4 General objectives of this study

Because phylogenetic approaches did not yield conclusive results in *Rhododendron*, a population genetic approach to establish species relationships within the subsection was suggested. As interspecific geneflow seems likely, one question was to establish whether or not the species can be robustly distinguished with genetic markers, and how species differentiation compares to population differentiation within species. Taxonomically, it was aimed to establish if the described varieties

are indeed of hybrid origin and if introgression was taking place between species involved in the hybrid zones. Furthermore, the question was asked, whether the introgression would be asymmetrical, or other patterns could be detected hinting to the architecture of the hybrid zones. Two different hybrid zones were investigated to establish if common patterns could be observed. As this is the first population study in the subsection, a further aim was to establish techniques and assess the value of the hybrid zones for future research.

To address these questions, populations of four species and some of their varieties were chosen to be investigated in detail. For reference their full names are given here, and all of them will be further introduced in the following section.

1. *Rhododendron aganniphum* Balfour f. & Kingdon Ward

   R. aganniphum var. *aganniphum* Balfour f. & Kingdon Ward

   R. aganniphum var. *flavorufum* (Balfour f. & Forrest) Chamberlain

2. *Rhododendron clementinae* Forrest

3. *Rhododendron roxieanum* Forrest

   R. roxieanum var. *cucullatum* (Handel-Mazzetti) Chamberlain

   R. roxieanum var. *oroneastes* Forrest

   R. roxieanum var. *roxieanum* Forrest

4. *Rhododendron phaeochrysum* Balfour f. & W. W. Smith

   R. phaeochrysum var. *agglutinatum* (Balfour f. & Forrest) Chamberlain

   R. phaeochrysum var. *phaeochrysum* Balfour f. & W. W. Smith

1.5 Species investigated in this study

The four species discussed here occur exclusively in South-West China (see Figure 1.1). However, it should be noted that the distribution maps are based on locality data from herbarium specimens [32]. Due to difficulties of access, many possible localities have never been collected; the maps therefore are likely to be incomplete. The species are predominantly found on mountain ranges above 3000 m, many of which are remote and difficult to access by any means of transport. The known distribution range of two of the species was extended by a certain amount, based solely on collections undertaken for this study.

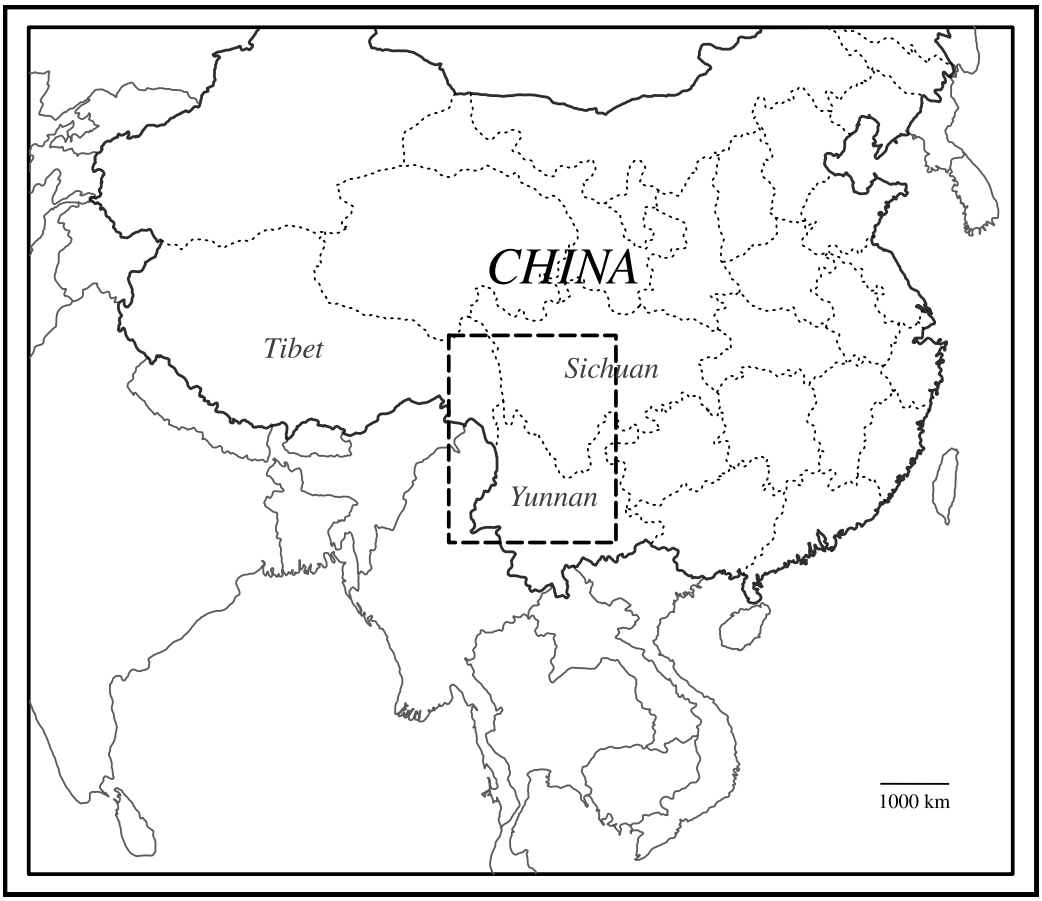

Figure 1.1: Collection area. Map showing the area where the species occur, and where populations were collected in 2007 and 2008. The square marked with dashed lines is the part of the map shown in detail in figures 1.2, 1.4 1.6 and 1.8.

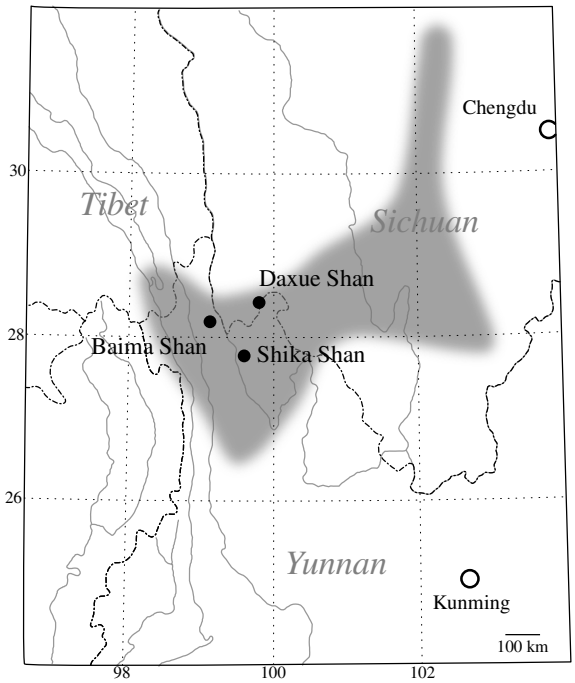

Figure 1.2: Distribution of *R. aganniphum* (grey area), changed from Chamberlain 1982 [32]; Locations of populations used in this study are marked with black circles.

1.5.1 *Rhododendron aganniphum*

The smallest of the investigated species in growth habit is *R. aganniphum* (at least from personal observations); this might rather be due to ecological than genetic reasons, as the species frequently occupies the upper regions of mountain ranges (>4000 m), and then only reaches the size of small, often appressed shrubs, which will be fully covered by snow during winter (Figure 1.3, E). If encountered at lower altitudes, individuals can grow to larger sizes, and although I personally never saw plants higher than 1.5 m, 3 m have been reported before [32]. The leaf shape is not a very good distinguishing character, and is mostly elliptic, but sometimes tending to broadly ovate. The indumentum is unistrate, composed of long hairs forming a seemingly continuous smooth upper layer; in appearance resembling *R. clementinae* (Figure 1.3, A). Hairs of young leaves are cream to light-yellow, and, in contrast to the other species, do not significantly change colour when older (*R. aganniphum* var. *aganniphum*, Figure 1.3, C). However, the two varieties described for *R. aganniphum*, var. *aganniphum* and var. *flavorufum*, are distinguished by the behaviour of the indumentum on maturation. *R. aganniphum* var. *aganniphum* behaves as described, while the indumentum in var. *flavorufum* changes colour to darker tones, often reaching

deep brown; coinciding with this colour change the continuous hair-layer starts splitting to different degrees, and the indumentum can finally become patchy (Figure 1.3, D). Because this study will present evidence for the admixed status of var. *flavorufum*, the pure species is not considered to exhibit this trait. The flowers of *R. aganniphum* are mostly white with purple spots inside the corolla; young flowers emerge pink from the bud and turn white later, but a variety of shades on the scale from pink to white are frequent (Figure 1.3, B). *R. aganniphum* is known to intergrade locally with *R. phaeochrysum* [32]. For the approximate geographical distribution see Figure 1.2.

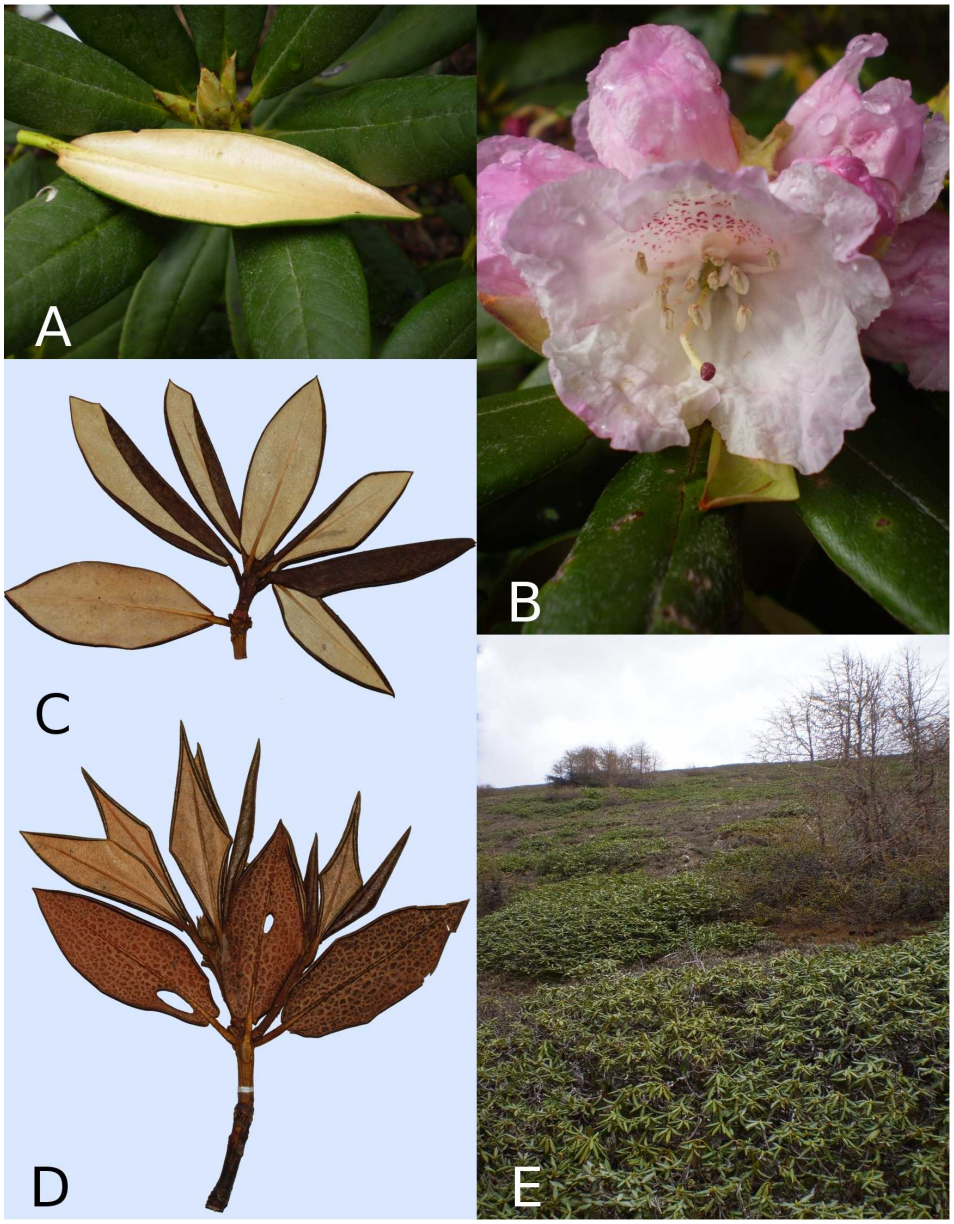

Figure 1.3: Morphology of *R. aganniphum*. **A.** Leaf of var. *aganniphum*; cream coloured thick indumentum, dense and compacted. **B.** Flower; flushed pink in bud, with crimson markings on dorsal petal. **C.** Typical specimen of var. *aganniphum*: indumentum remaining intact and cream coloured in older leaves. **D.** Typical specimen of var. *flavorufum*: indumentum intact and cream in young leaves, splitting and darkening during maturation. **E.** Growth habit: Appressed shrubs.

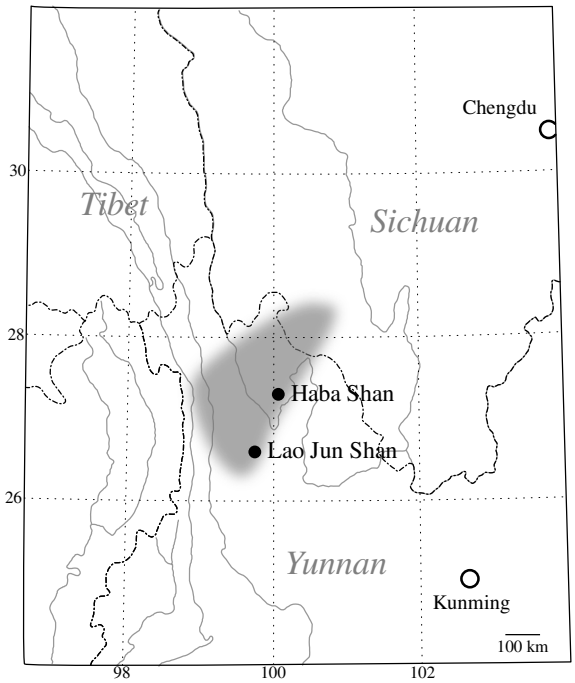

Figure 1.4: Distribution of *R. clementinae* (grey area), changed from Chamberlain 1982 [32]; Locations of populations used in this study are marked with black circles.

1.5.2 *Rhododendron clementinae*

Of the four species *R. clementinae* has the most restricted distribution range (Figure 1.4). Plants of the species are tall shrubs to small trees (1–3 m; Figure 1.5, A, C), and grow at altitudes between 3300–4100 m. The leaves are very broad, 1.5–2 × as long as broad, and the lower surface is covered with a thick bistrate indumentum; the upper layer is very dense and compacted, so that on first sight it seams like a continuous surface (Figure 1.5, D). In young leaves the indumentum is white to cream, often turning light-brown to milk-coffee coloured when older. When in flower, it is the most unambiguous species of subsection *Taliensia* to distinguish, as its consistently seven-lobed corolla is one of few available qualitative characters (Figure 1.5, B). Not different from the other species, the flowers are most commonly flushed pink when freshly emerging from the bud, but can turn completely white when older. Flowers of *R. clementinae* often have no or only very sparse purple markings on the petals. Although the seven-lobed corolla distinguishes pure specimens, this character apparently diminishes easily in hybrids, where six to five lobes are frequent (*R. clementinae* × *phaeochrysum*, Haba Shui Shan; personal observation). One subspecies has

been described for *R. clementinae*, ssp. *aureodorsale* [56]; it has been reported in Shaanxi Province only, which is far outside[3] the reported distribution range of ssp. *clementinae*, and the two 'subspecies' are probably not related (David Chamberlain, personal communiction). I will always refer to *R. clementinae* without considering this sub species.

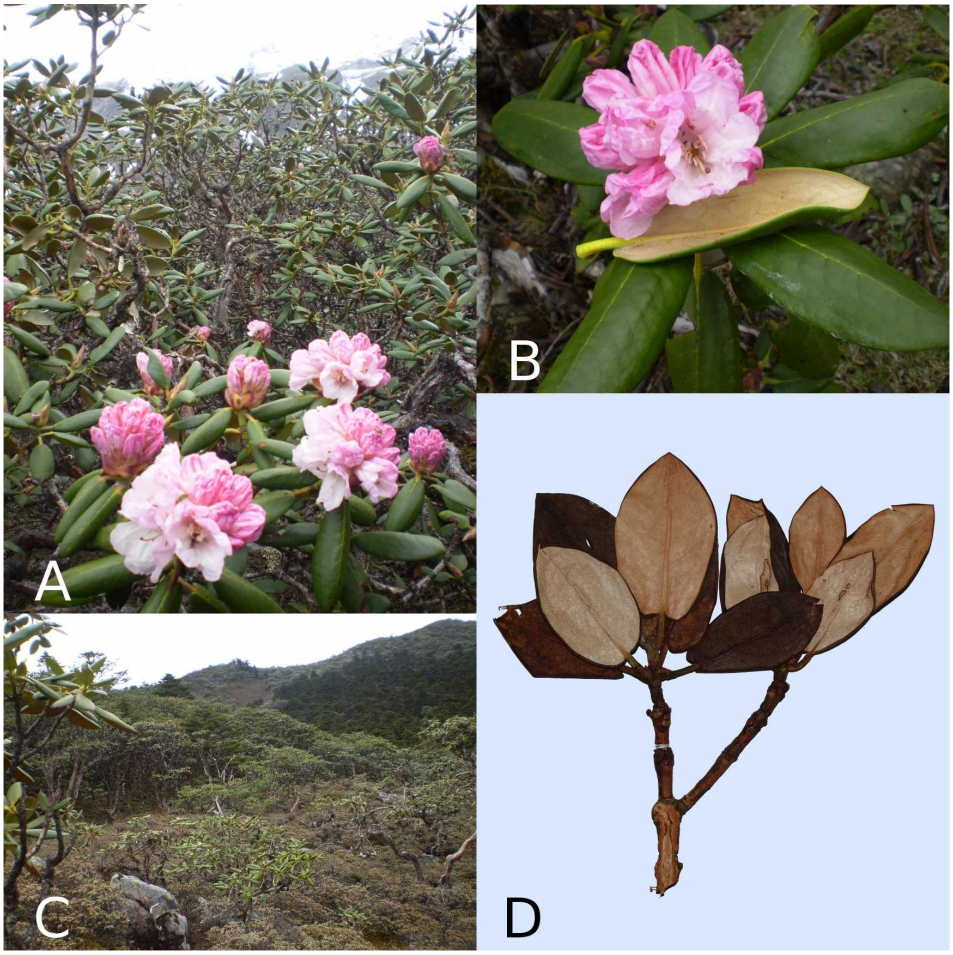

Figure 1.5: Morphology of *R. clementinae*. **A.** Dense stand of plants growing as tall shrubs; **B.** Flower, clearly showing the 7-lobed corolla, some scarce crimson markings on the dorsal petal. **C.** Juvenile and adult plants, front-center – shrub habit of young plant, back-left – forest of adult individuals with tree-like habit. **D.** Typical specimen: very broad leaves; indumentum thick, compacted, in young leaves whitish, changing to light-brown during maturation.

[3]Shaanxi Province is situated north-east of Sichuan Province, please compare with Figure 1.4.

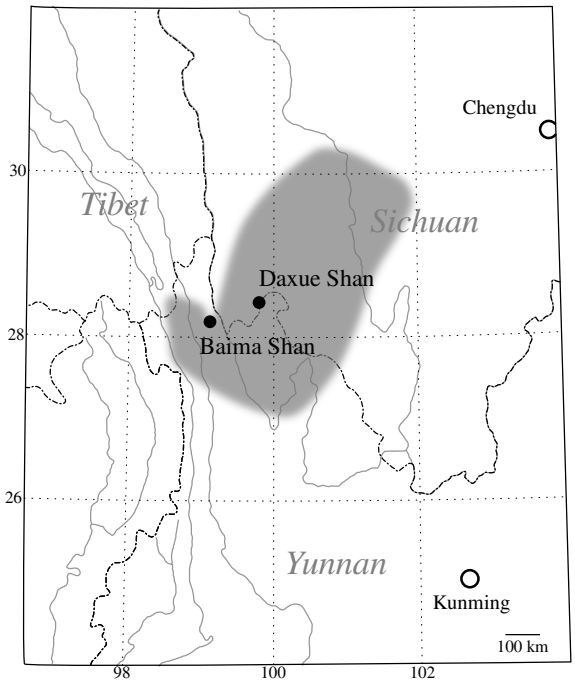

Figure 1.6: Distribution of *R. phaeochrysum* (grey area), changed from Chamberlain 1982 [32]; Locations of populations used in this study are marked with black circles.

1.5.3 *Rhododendron phaeochrysum*

Individuals of *R. phaeochrysum* are large shrubs to small trees (1.2–4.5 m; Figure 1.7, D), and the species can be found between 3300–4400 m. The leaves can vary considerably in shape; mostly elliptic, but often ovate-oblong or ovate (Figure 1.7, A). The indumentum is dense, and composed of short radiate hairs; these are whitish-cinnamon coloured when young and turn deep red-brown to brown in later stages of maturity (Figure 1.7, C). As in *R. aganniphum*, three varieties have been described for *R. phaeochrysum*, based on variation in the indumentum structure: var. *phaeochrysum*; var. *agglutinatum*; and var. *levistratum*; of which the first two will be discussed in this thesis. The type variety var. *phaeochrysum* represents the "ideal" state of the indumentum: uniform and felted. Individuals with agglutinated to sometimes splitting indumentum are referable to var. *agglutinatum*. The flowers emerge flushed pink and turn white quickly; often with a few crimson markings (Figure 1.7, B). The species is relatively wide spread, and sometimes common (Figure 1.6); it is known to intergrade locally at least with *R. aganniphum* [32], *R. beesianum* and *R. clementinae* (personal observation). The note of D. Chamberlain regarding

var. *agglutinatum* [32] (page 352) is interesting:

"*Closely resembling some forms of R. aganniphum but with a darker indumentum than var. aganniphum and a less patchy indumentum than var. flavorufum.*"

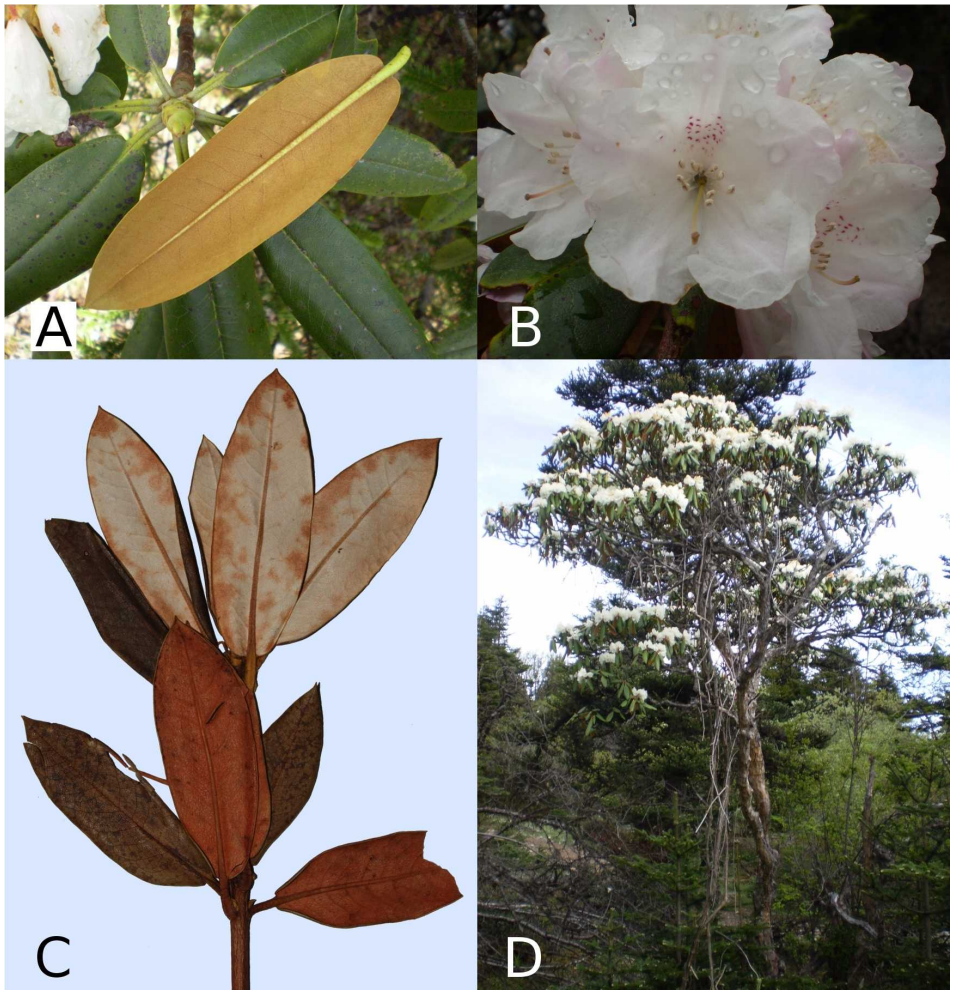

Figure 1.7: Morphology of *R. phaeochrysum*. **A.** Leaf, middle-aged indumentum, characteristic leaf shape, but more elliptic shapes are frequent. **B.** Flower with crimson markings on the dorsal petal. **C.** Typical specimen: indumentum thin; white in young leaves, turning deep red-brown in maturity. **D.** Growth habit of an adult individual.

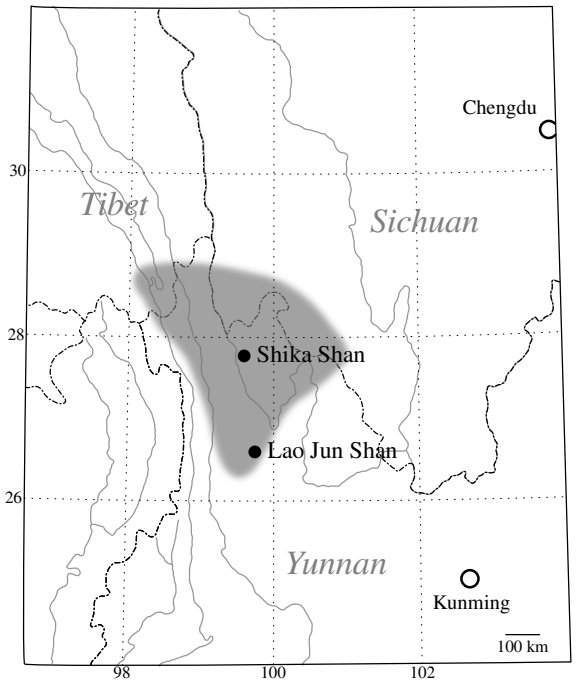

Figure 1.8: Distribution of *R. roxieanum* (grey area), changed from Chamberlain 1982 [32]; Locations of populations used in this study are marked with black circles.

1.5.4 *Rhododendron roxieanum*

The last of the species discussed in this thesis is *R. roxieanum*, for which three varieties have been described: var. *roxieanum*, var. *oroneastes*, and var. *cucullatum*. It can be found at altitudes between 3000–4300 m, and growth habits range from appressed shrubs at very high altitudes to small trees (4.5 m) at lower altitudes (Figure 1.9, H). As the other species it is found in north-western Yunnan and bordering provinces (Figure 1.8). Varieties of *R. roxieanum* are mainly distinguished by the shape of the leaves; on one extreme of the spectrum is var. *oroneastes* with very narrow linear leaves and on the other var. *cucullatum* with elliptic leaves (Figure 1.9, A, E, F). The underside of the leaves is covered with a thick bistrate indumentum, with the lower layer being compacted and composed of radiate hairs, and the upper layer, formed by the long hairs, being loose. On fresh leaves the indumentum is white, turning dark-red to red-brown (rufous) in older ones (Figure 1.9, E, G). Flowers open more or less intensively flushed pink, turning white; the petals are marked with crimson spots, and often the patterns are very dense, but as in all discussed species the floral colouring varies considerably (Figure 1.9, B–D). Transition between

the mentioned extremes of leaf shapes is gradual, and clear-cut assignment of varieties is frequently impossible. During the collection of the specimens for this study one consistent qualitative character was observed, but more samples are needed to confirm this. In morphologically clear individuals of var. *roxieanum* the indumentum above the midrib turns black in older leaves; the same can be observed in var. *oroneastes*, but never in var. *cucullatum* (Figure 1.9, A, B, E–G). Therefore, as already suggested by Chamberlain [32], *R. roxieanum* will be treated as having two varieties: var. *roxieanum* (including var. *oroneastes*) and var. *cucullatum*. This also makes sense as clear individuals of var. *cucullatum* can easily be distinguished from either var. *roxieanum* or var. *oroneastes*, but the latter two are morphologically too close together to ever avoid ambiguity. Chamberlain [32] mentions that:

"*Var. cucullatum is intermediate between var. roxieanum and R. proteoides and is almost certainly of hybrid origin.*"

However, in my opinion, although *R. proteoides* is almost certainly allied to *R. roxieanum* var. *roxieanum*, its leaves are very narrow, and var. *cucullatum* does not seem intermediate because of its considerably broader and longer leaves. The morphology of var. *cucullatum* is very variable between different localities, and might in fact be a conglomerate of intermediates between different species and *R. roxieanum*. Assuming some contribution of introgression to its morphology, the most probable involvement at the locality investigated in this study (Lao Jun Shan) seems to be *R. clementinae*. Evidence for introgression of genes from a different genetic background includes the occurrence of rugose leaves in some individuals (Figure 1.9, A), which can sometimes be observed in hybrids (Catherine Kidner, personal communication). Further support is the intermediate appearance of individuals of var. *cucullatum*: The leaves are considerably wider than in var. *roxieanum*, consistent with the very broad leaves of *R. clementinae*; the indumentum is mostly thicker than in var. *roxieanum*, and longer indumentum hairs are found in *R. clementinae* (Figure 1.10). Apart from the different forms of var. *cucullatum*, *R. roxieanum* has been seen to form hybrids with *R. traillianum* (at Lao Jun Shan, personal observation); and perhaps *R. aganniphum*, (*R. proteoides*, Baima Shui Shan, personal observation).

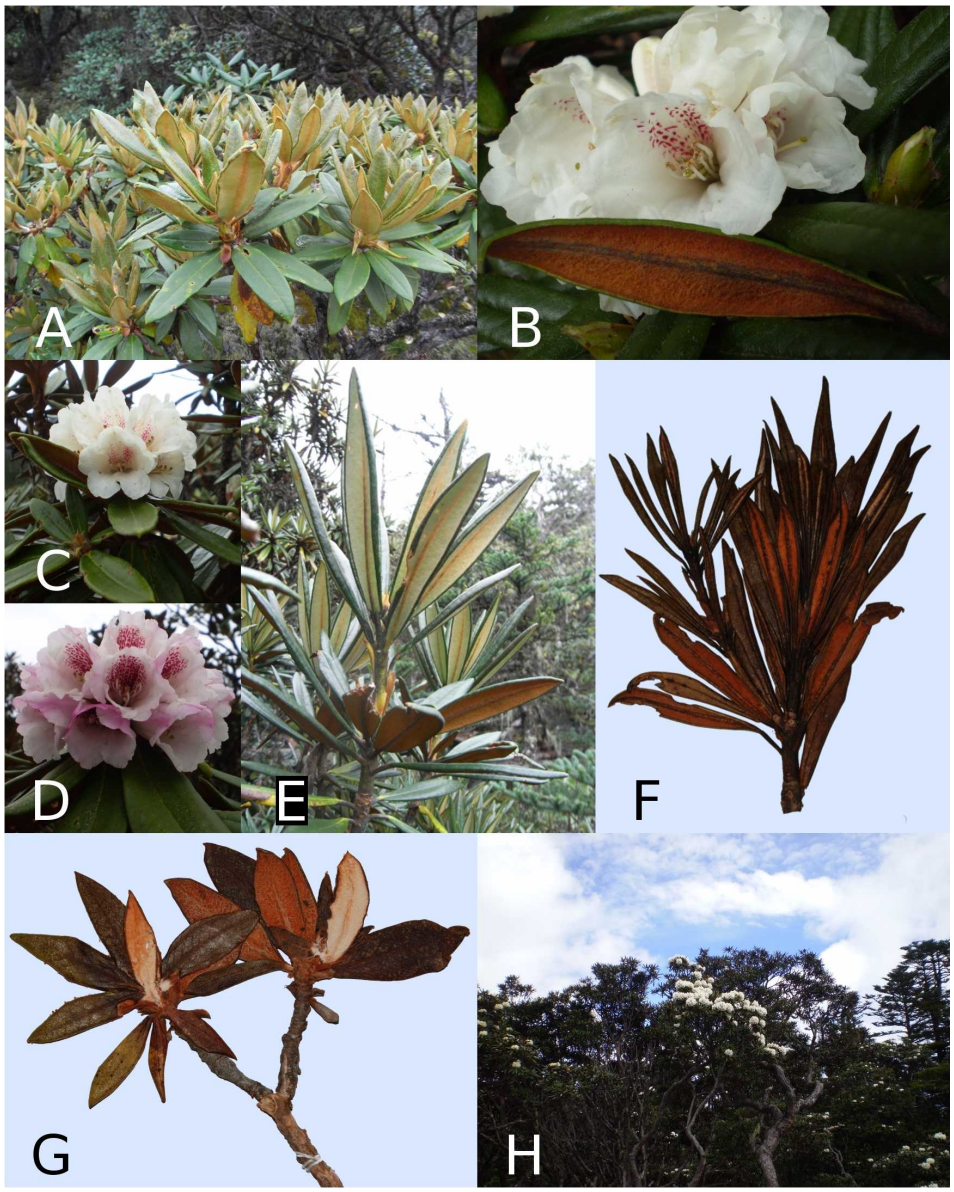

Figure 1.9: Morphology of *R. roxieanum*. **A, G.** Var. *cucullatum*. **A.** plant with rugose leaves; **G.** indumentum white in young leaves, turning red in maturity, always lacking a black midrib. **B, E, F.** Var. *roxieanum*; **B.** flower with crimson markings; **E.** young and old leaves changing indumentum colour; **F.** narrow leaved form, conforming to var. *oroneastes*; all with characteristic black midrib. **C, D.** Flowers of var. *cucullatum*; illustrating diversity in floral characters; this is an example using var. *cucullatum*, but the floral characteristics vary largely in all investigated species. **H.** Adult plants of var. *roxieanum* showing tree habit.

1.6 Sample localities

In this section the populations and localities, where the samples used in this study have been collected, will be introduced. All samples were collected during two fieldtrips that took place in late summer 2007[4] and spring 2008. More locations were visited and sampled, particularly small population samples of *R. phaeochrysum*, but as these were not included in the analysis, they will not be mentioned further. The first collecting expedition for this study took place from 25th August to 24th September 2007. It was undertaken together with David Chamberlain and Richard Milne, and we were accompanied by Jie Liu from the Kunming Institute of Botany (KIB) as counterpart. Autumn was chosen because of the possibility of collecting seeds, which would have allowed genetic analysis of offspring germinated from them at RBGE. Unfortunately the weather conditions during the flowering season that year had been exceptionally bad, and hardly any plants had managed to set seed; due to the scarcity of seedpods this idea was abandoned. The objective of the second trip, which was undertaken between 16th May to 16th June 2008, was to complement the sampling of the 2007 collections and to collect missing allopatric reference populations. The trip took place in spring to enable the collection of specimens during the flowering season. Members of this expedition included myself and two counterparts from KIB, Lian-Ming Gao and Jie Liu. One of the challenges regarding population sample collections in rhododendrons is the scarcity of information. Most herbarium specimens were collected before tools were availlable to provide accurate locality information (this includes maps, as well as GPS). Localities may be as vague as "The Big Mountain by The Lake". Additionally, even if an approximate area can be determined, missing the coordinates by 1 km mostly could mean being on the wrong mountain. For these reasons and general problems of accessibility in this mountainous region, success of the collections depended heavily on information from people who had relatively recently encountered populations of the desired species, and could give more detailed locality information. In 2007 David Chamberlain made the collections possible, and in 2008 vital information was obtained from David Rankin and Lian-Ming Gao. The collection dates were chosen mostly for the above given reasons, however, periods in which collections can be carried out are also heavily restricted by the climatic conditions. Two windows of accessibility for the area exist: after winter, when enough snow has melted to make the roads and mountain passes drivable; and after the rainy season, in which frequent landslides

[4] A short report on the collection trip in 2007 has been published in The Rhododendron, Camellia & Magnolia Group Bulletin [1], page 8.

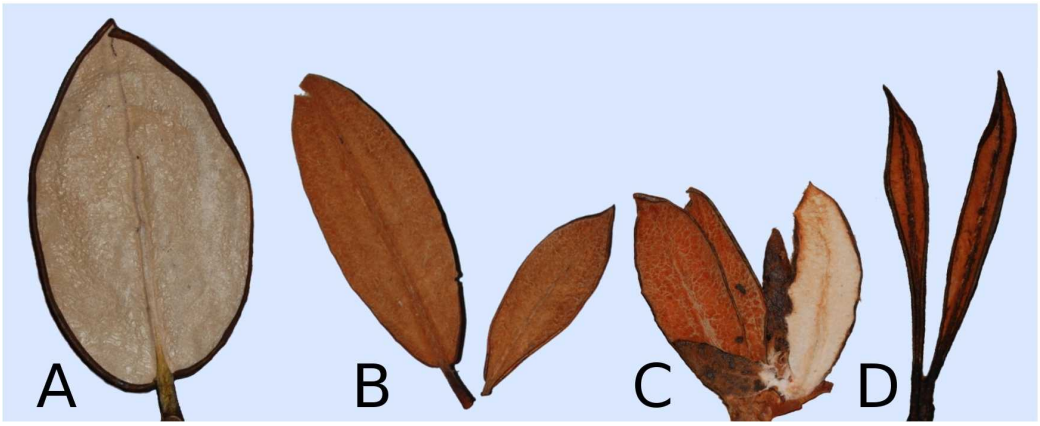

Figure 1.10: Leaf morphology of *R. clementinae* and *R. roxieanum*. **A.** *R. clementinae*; large broad leaves, indumentum hairs long but densely compacted. **B, C.** *R. roxieanum* var. *cucullatum*; leaves wider than var. *roxieanum*, but narrower than *R. clementinae*; indumentum not compacted but often thicker than in var. *roxieanum*; B. is the usual type of individuals found, C. belongs to individual `TXC09` and is not very common. **D.** *R. roxieanum* var. *roxieanum*; very narrow to linear leaves, indumentum hairs of medium length, not compacted. * leaves are not shown to same scale; C should be roughly as long as D; A & B should be slightly larger than depicted.

block roads and make travel unfeasible. Nonetheless the following localities were visited and populations sampled.

1.6.1 Lao Jun Shan

Both expeditions started at Lao Jun Shan, as it is the only locality that has accommodation very near by, and therefore allows acclimatisation to the altitude (for location see Figure 1.4, or 1.8). Situated between 3700–4000 m, several species of subsection *Taliensia* can be encountered here: *R. alutaceum*, *R. beesianum*, *R. clementinae*, *R. traillianum* and *R. roxieanum*.

R. traillianum* & *R. beesianum In several forest areas *R. traillianum* comprises the major part of the adult plant population, where it normally reaches a height of up to 4m, but younger plants can be found in mixed stands. In these *R. traillianum* dominated forests, dispersed small plants of *R. roxieanum*, and often *R. beesianum*, can be found. *R. beesianum* does not form stands on its own, but rather occurs frequently under the canopy of *R. traillianum*, with only few individuals reaching the height of *R. traillianum*. In 2007 one seedling recruitment area was found, where a large individual of *R. traillianum* had fallen over and left a gap in the canopy. Many of the emerging seedlings and young plants observed in this spot exhibited morphological characters intermediate between

R. traillianum and *R. beesianum*. All individuals in a plot of 20 by 20 meters around the recruitment area were sampled in a grid-like fashion. These species generally do not pose taxonomic difficulties, and no odd varieties have been described that would imply involvement of both. Furthermore, it was not clear if admixture during the seedling stage was normal behaviour, and the samples were not investigated further. However, in 2008 a similar recruitment area was discovered in a different part of the forest. Time constraints and study focus did not allow for extensive sample collection then, but the occurrence of hybrids during the juvenile stage seems frequent. Investigation of the composition of the seedling swarm, perhaps after resampling at a later point, and using the already extensive sampling as a reference, could yield valuable insights into potential genetic barriers due to competition, and assess temporal dynamics of the swarm composition.

R. alutaceum One species entirely composed of unclear varieties is *R. alutaceum*. At the beginning of the project it was chosen to be a focus of the study; however, field observations pointed to a potential hybrid origin involving *R. roxieanum* var. *cucullatum*, and it seemed more sensible to investigate the varieties of *R. roxieanum* first. In a large area around a lake, on rocky ground fragmented by numerous streams (Figure 1.11, A), many small groups of *R. alutaceum* var. *iodes* can be found amongst scattered plants and small stands of *R. clementinae*, *R. traillianum* and *R. roxieanum*. With near certainty all individuals are of hybrid origin, but due to the occurrence of the aforementioned species, parentage is unclear, although *R. traillianum* is very likely contributing. If at some future point sufficient genetic information for the other species should be available, this might well be an example case for a three-way hybrid. This population was extensively sampled, but will not be discussed further.

R. clementinae* & *R. roxieanum Finally this is the location for one of the potential hybrids investigated here. In the same area around the lake as mentioned before (3800 m), individuals of *R. clementinae*, *R. roxieanum* var. *roxieanum*, and *R. roxieanum* var. *cucullatum* grow interspersed in a mixed population. Additionally, in other areas on Lao Jun Shan, large, morphologically homogeneous stands of *R. clementinae*, *R. roxieanum* var. *roxieanum* and *R. roxieanum* var. *cucullatum* do occur, but the species are generally in close proximity. However, individuals attributable to *R. roxieanum* var. *cucullatum* are only found at higher altitudes, where *R. clementinae* also occurs (>3800 m).

The sympatric reference population of *R. roxieanum* var. *roxieanum* (X1)[5] was collected at a lower altitude (3700 m), mostly on a rock slide, where neither *R. clementinae* nor *R. roxieanum* var. *cucullatum* can be found (Figure 1.11, C). *R. roxieanum* var. *roxieanum* occurs at higher altitudes but not as abundantly as at lower ones. The sampling of *R. roxieanum* var. *cucullatum* (XC1) includes a few samples from around the lake, but mostly comprises individuals from a dense homogeneous population, covering a whole slope above Ninety-Nine-Dragon Pools (Figure 1.11, B). Individuals of the sympatric population of *R. clementinae* (C1) were collected not far along a path stretching about 300 m. Because more individuals were required for statistical reasons, following the collections of 2007, the area was re-sampled in 2008. To avoid recollecting individuals, areas were carefully chosen in which no plants had been collected the previous year. Given the co-occurrence of *R. roxieanum* var. *roxieanum* and *R. roxieanum* var. *cucullatum* over such a wide range, it is unclear how the varieties can retain morphological integrity, and the pattern of altitudinal distribution suggests that introgression might play a role in their divergence.

1.6.2 Shika Shan

Close to the city of Zhongdian is the mountain of Shika Shan, which has been recently developed for tourists (for location see Figure 1.2, or 1.8). The installation of a cable-car with two stations, one at the top and one half way up, makes this locality relatively easy to access. It is well known for its Rhododendrons, mostly not from subsection *Taliensia*, however, *R. aganniphum*, *R. phaeochrysum* and *R. roxieanum* can be found. Personally I only observed very few scattered individuals of *R. phaeochrysum*, and only populations of *R. aganniphum* and *R. roxieanum* were collected. This locality was visited in 2007 as well as 2008, because the intended reference population of *R. roxieanum* could not be located in 2007; this was due to no exact locality information and bad weather, which meant that the mountain had to be climbed as the cable-car was not functional.

R. aganniphum During the first attempt in 2007, despite heavy rain, at an altitude of 3700 m, a small population of *R. aganniphum* was collected. The individuals were growing in a gully, probably formed by a seasonal stream. This population was included as a third population of *R. aganniphum* (G1) in the

[5] The codes in brackets given after the populations refer to the population code, used in Table 1.1, page 49.

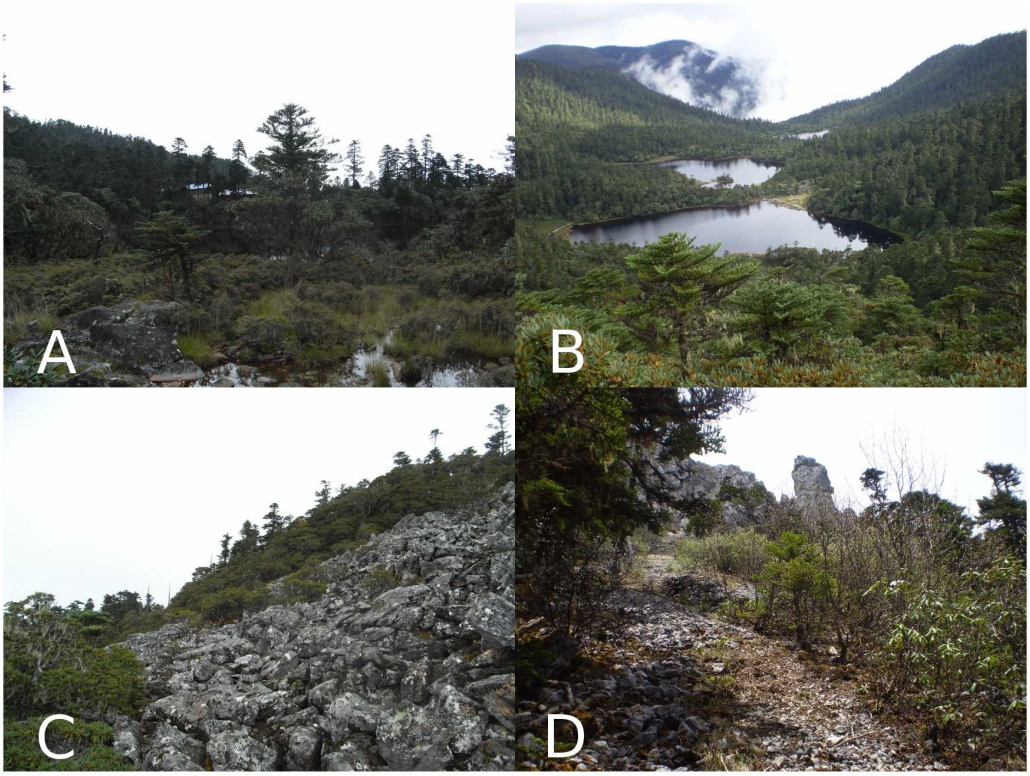

Figure 1.11: Collection sites I: A. Lao Jun Shan; heterogeneous habitat situated around lake. **B.** Lao Jun Shan; Ninety-Nine-Dragon Pools, bottom-border individuals of *R. roxieanum* var. *cucullatum* covering the whole slope; site for most samples of XC1. **C.** Lao Jun Shan; rocky slope with *R. roxieanum* var. *roxieanum* (left and top-right); site of X1. **D.** Shika Shan; *R. roxieanum* var. *roxieanum* growing on old rockslide, seemingly similar to C; site of X3.

analysis; rather small it seemed to be isolated, and even in 2008 only a single *R. aganniphum* was discovered elsewhere on the mountain.

R. roxieanum With more precise locality information, the *R. roxieanum* population (X3) could be located in 2008. The individuals were growing on a rockslide at about 4100 m, considerably higher than the sympatric population at Lao Jun Shan, but under comparable conditions (Figure 1.11, D). Most plants here could be attributed to *R. roxieanum* var. *oroneastes*, with a few tending to *R. roxieanum* var. *roxieanum*. This population was chosen as allopatric reference population as the morphology was very pure, no other species were detected nearby, and not a single intermediate form occurred.

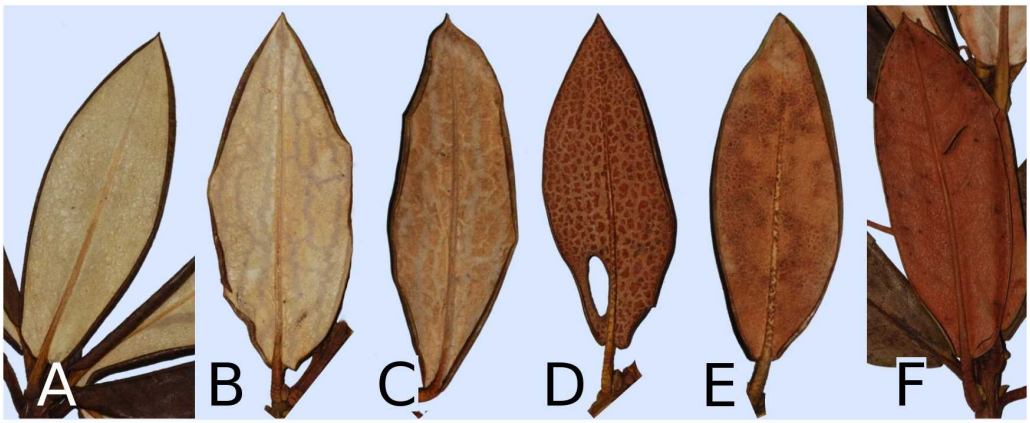

Figure 1.12: Indumentum morphology of *R. aganniphum* and *R. phaeochrysum*. Exemplary forms of indumentum type found amongst potential hybrid individuals. **A.** *R. aganniphum* var. *aganniphum*; indumentum remains white at maturity, long hairs, never splitting. **B, C.** Forms close to A; indumentum starts splitting at maturity but remains more or less white (B) or slightly changes colour (C). **D.** *R. aganniphum* var. *flavorufum*; indumentum with slightly shorter hairs, splitting in medium aged leaves, becoming patchy in maturity and turning deep brown. **E.** Forms close to F; indumentum with hairs visibly shorter than D, becoming patchy early, then later partly or entirely appearing agglutinated. **F.** *R. phaeochrysum*; indumentum turns dark red-brown at maturity, short hairs, cannot split due to missing longer hairs.

1.6.3 Baima Shan

This location includes two sampling sites in relative proximity, which were sampled separately in the consecutive years 2007 and 2008. The first collection at Baima Shan took place alongside the "Old Road", which is an abandoned road on the north-eastern flank of the Baima Shan range, situated between 3800–4300 m. For the second collection in 2008 we climbed Baima Shui Shan (Baima Shan), one of the major mountains in the massif, approximately 5 km from the "Old Road". (for location see Figure 1.2, or 1.6)

Old Road

There is no longer access for vehicles to the road, but the gravel is still intact, and therefore rather ruderal growth conditions dominate along the roadside. *R. aganniphum* grows here in sympatry with *R. phaeochrysum* and at lower altitudes some individuals of *R. beesianum* can be found (Figure 1.13, A).

R. beesianum Some individuals of *R. beesianum* were collected along the road, which do not seem to hybridise frequently with the other species. Hybrids

involving *R. beesianum* are generally easy detected, as the much larger leaves and characteristic shape (oblanceolate) of this species lead to significantly enlarged leaves in intermediates. Only one possible *R. beesianum* × *phaeochrysum* was found. This collection is not relevant for this study.

R. aganniphum* & *R. phaeochrysum For *R. aganniphum* and *R. phaeochrysum* a slight altitudinal gradient can be observed, with *R. aganniphum* mostly occupying the upper range uphill from the road, while *R. phaeochrysum* dominates the lower ranges. There is considerable overlap to both sides of the road, and in this zone some individuals of *R. aganniphum* var. *flavorufum*, along with numerous morphological intermediates between the two species are present, so that a whole gradient of indumentum types can be observed (Figure 1.12). Due to backcrosses apparently present, a high chance of introgression was assumed, and presented the possibility to test not only the hybrid status of *R. aganniphum* var. *flavorufum*, but also to investigate species barriers and possible gene exchange. This swarm was chosen as the second hybrid zone in this study and sampled extensively (GH2a). The sympatric reference populations of *R. aganniphum* (G2)[6] and *R. phaeochrysum* (P1) were sampled along the same road, and are composed of individuals that were deemed morphologically pure. Before ascending Baima Shui Shan in 2008, some additional samples of pure *R. aganniphum* were obtained by collecting slightly higher up the slope (Figure 1.13, B), to increase sample size, and avoid the use of uncertain individuals. They were included under the same population code as the site was identical.

Baima Shui Shan

The main reason for scaling Baima Shan in 2008 was because information gathered before the expedition hinted at many individuals of *R. roxieanum* growing near the top, and no reference population had been obtained for this species at this point. The lower part of Baima Shan (from 3500 m upwards) is covered with a light homogeneous forest of *R. beesianum*, plants frequently reaching about 4 m. Higher up (4000 m) individuals of *R. roxieanum* var. *roxieanum* start appearing, reaching about 2 m in size here, and widely taking over close to the treeline (∼4200 m), where they grow as low shrubs (Figure 1.13, C). Also at this altitude several areas are covered with plants morphologically resembling a potential

[6]This population also corresponds to G2a, before G2b was excluded from the analysis (chapter 3, section 3.1, page 96).

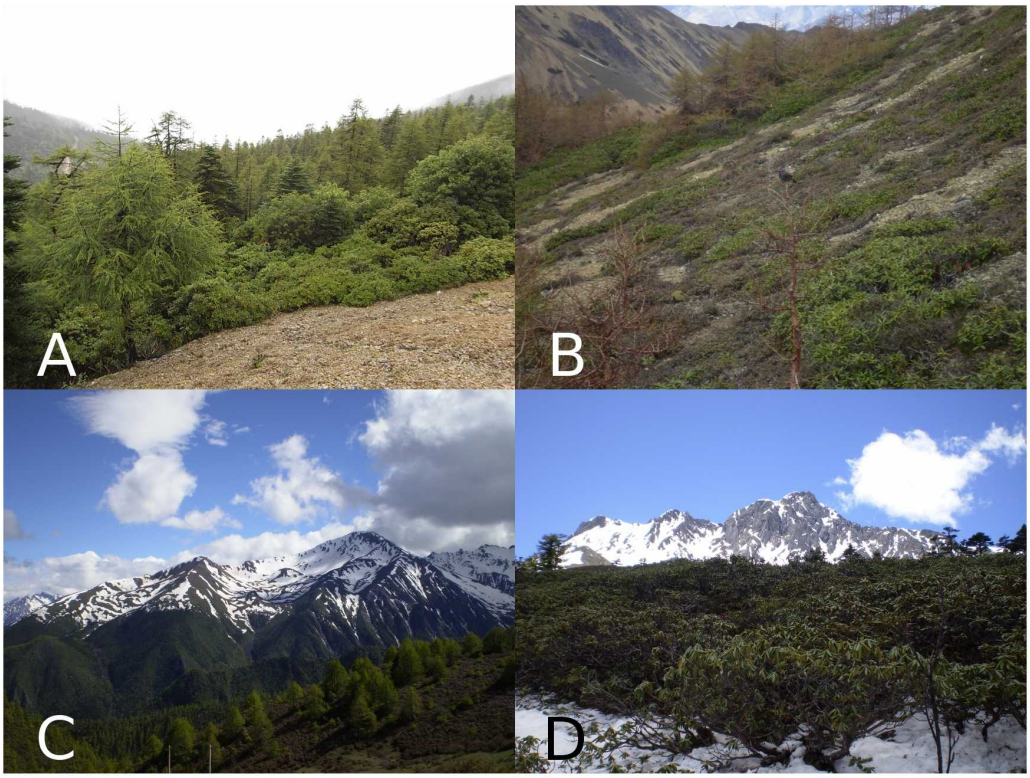

Figure 1.13: Collection sites II: A. Baima Shan; mixed population of *R. aganniphum* and *R. phaeochrysum* along the "Old Road", where many intermediates between the species can be found; site of GH2a and P1. **B.** Baima Shan; slope uphill from the "Old Road"; more or less scattered individuals of *R. aganniphum* amongst *Larix sp.*; site of G2(a). **C.** Baima Shui Shan; *R. beesianum* can be found amongst the forest on the slopes, *R. roxieanum* and *R. aganniphum* × *phaeochrysum* hybrids of morphotype E (Figure 1.12, E) start appearing at the tree line where the snow takes over. **D.** Baima Shui Shan; above the tree line; dense population of individuals with morphotype E (population GH2b) with very few plants of more or less pure *R. aganniphum* (population G2b).

hybrid of *R. aganniphum* and *R. phaeochrysum*. Amongst the populations of *R. roxieanum* var. *roxieanum* now and then there are individuals of *R. proteoides* which in my opinion seem intermediate between *R. aganniphum* and *R. roxieanum* (but see subsection 1.5.4).

R. roxieanum The population of *R. roxieanum* (X2)[7] was composed of morphologically homogeneous plants, and was sampled as allopatric reference population. It was finally not included in the study because of slight doubts regarding *R. proteoides*, and the population from Shika Shan (subsection 1.6.2), obtained shortly afterwards, seemed the better option.

[7]this population is not mentioned in Table 1.1, page 49

R. aganniphum* & *R. aganniphum* × *phaeochrysum Only few individuals of *R. aganniphum* were present at roughly the same altitude that is completely occupied by this species at the site of the "Old Road"; interestingly this was not only due to *R. roxieanum* occupying this region, but mostly the place of *R. aganniphum* seemed to be taken by plants whose morphology suggested a hybrid (Figure 1.13, D). Similar plants, morphologically corresponding to morphotype E (Figure 1.12, E), can also be found at the "Old Road". However, this morphological type is rather scarce in the latter population, and appeared to be only one of many intermediates in the hybrid swarm. Near the top of Baima Shan, the morphology of nearly all the individuals corresponded to this type, and the variety of indumentum intermediates was not observed. All this suggested a different dynamic in this hybrid population, and presented the opportunity to investigate whether there was evidence of a possible further reaching evolutionary impact of the apparently weak species barriers; or whether morphology was misleading. Therefore this population was sampled and included into the analysis as population GH2b of *R. aganniphum* × *phaeochrysum*. Additionally a few samples were obtained from the scarce sympatric *R. aganniphum* (G2b).

1.6.4 Dàxuě Shan

On the road from Weng Shui to Xiangcheng, near the border of the provinces of Yunnan and Sichuan, is the pass of Dàxuě Shan (for location see Figure 1.2, or 1.6). Situated at an altitude of 4300 m most slopes here are covered with *R. aganniphum*, and several individuals of *R. phaeochrysum* can be found (Figure 1.14, A). With the road creating a certain amount of disturbance, the habitat is apparently not dissimilar to the one at the "Old Road" on Baima Shan. However, the species here seem to be well separated and no intermediates are present. An extensive population sample of *R. aganniphum* (G3) was obtained here, as well as a good sample of *R. phaeochrysum* individuals (P2); although both species grow here in sympatry, these two populations were included in the study as allopatric reference populations for the hybrid zone on Baima Shan. To have the species in close proximity is not ideal for a reference, as introgression might have occurred, but due to time constraints no further attempts could be made to find other populations. More populations of *R. phaeochrysum* had been collected in 2007, but unfortunately none of the sample sizes were large enough to be used for a dominant marker technique as AFLP. Morphologically the species were pure, and seemed the best option available.

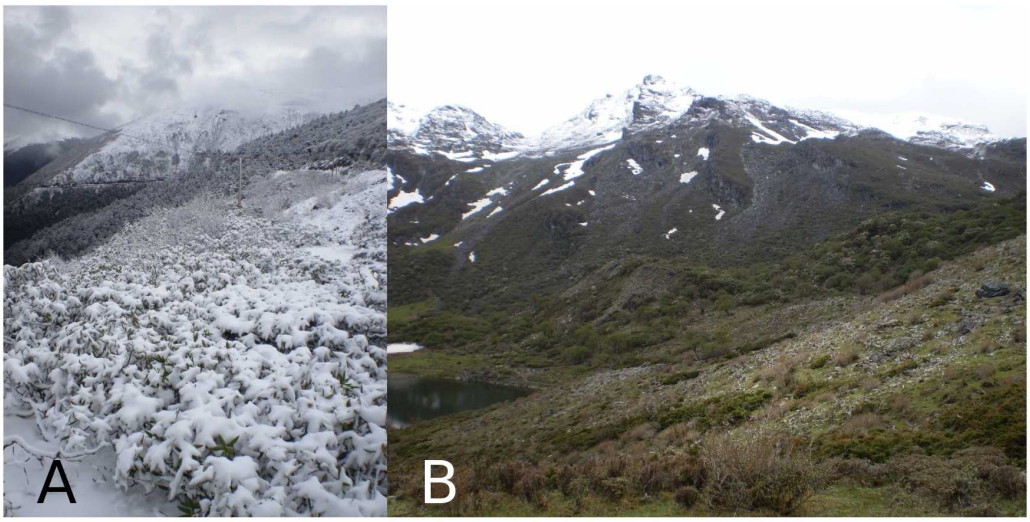

Figure 1.14: Collection sites III: A. Dàxuě Shan; dense population of *R. aganniphum* covering most of the slope; one third from the top of the picture the road over the pass can be seen, along which individuals of *R. phaeochrysum* can be found; site of G3 and P2. **B.** Haba Shan; Hi Hei Lake left in the picture; middle-right in the distance flowering individuals of *R. clementinae*; site of C2.

1.6.5 Haba Shan

The last missing reference population, that of *R. clementinae*, was collected in 2008 on Haba Shan (Haba Shui Shan, Figure 1.4). At an altitude of 4100 m, around Hi Hei Lake, a mixed population of *R. clementinae* and *R. phaeochrysum* can be found. The species hybridise here with each other at one end of the lake, where occasionally individuals of *R. clementinae* can be found in a population of *R. phaeochrysum*; on the other side of the lake is an apparently pure stand of *R. clementinae* (Figure 1.14, B). Although introgression between *R. phaeochrysum* and *R. clementinae* is possible here, *R. phaeochrysum* is not present on Lao Jun Shan, and is certainly not involved in *R. roxieanum* var. *cucullatum*. As there was no time remaining to search for other populations, and because morphology suggested pure individuals, a population sample of *R. clementinae* (C2) was obtained here as allopatric reference for the assessment of potential introgression at Lao Jun Shan.

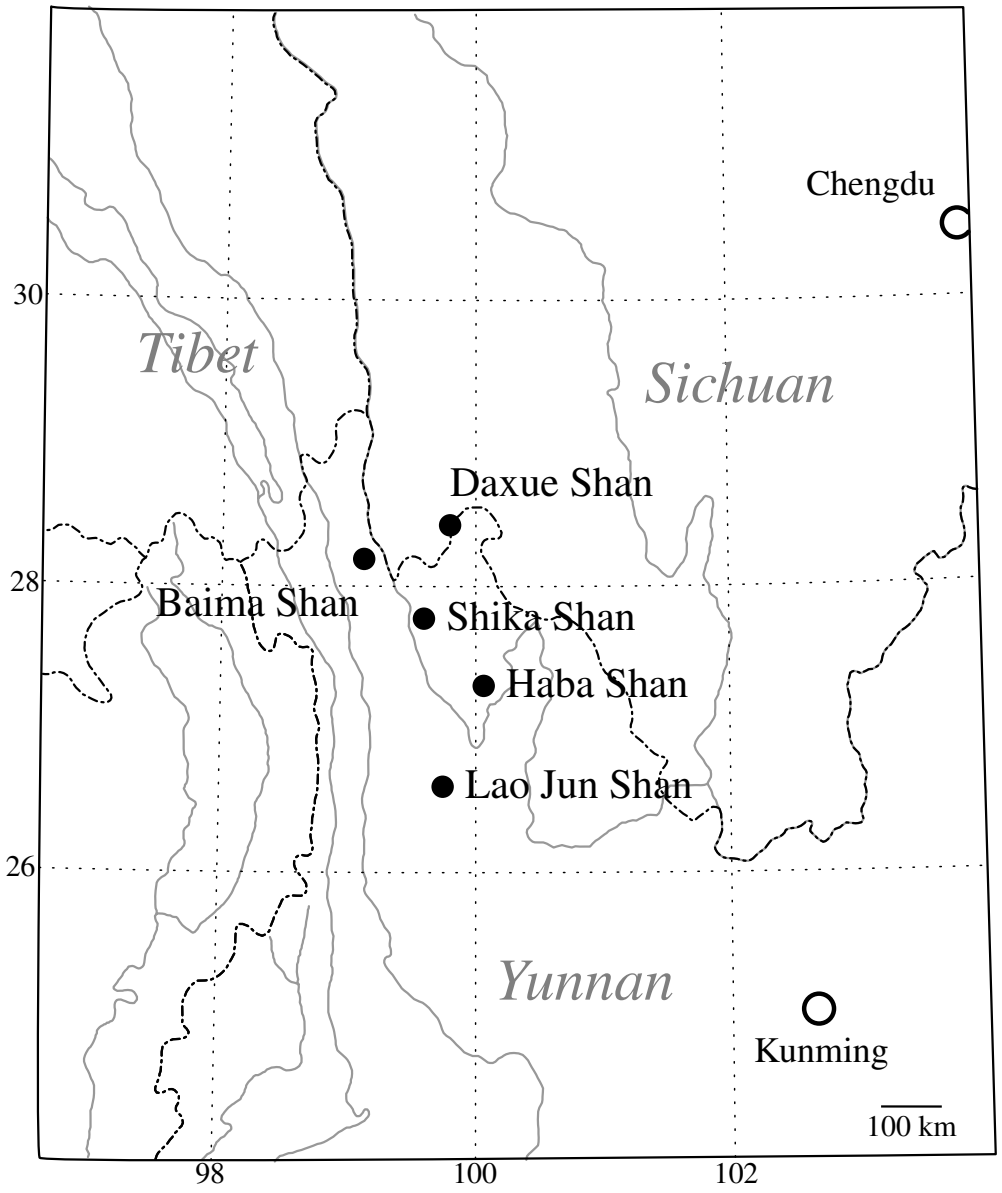

Figure 1.15: Sample localities map showing all localities where the populations used in this study were collected (black circles).

1.6.6 Summary of collected populations

In total 13 (12[8]) populations were sampled at five localities (Figure 1.15), that were included into the analyses (Table 1.1).

R. clementinae* and *R. roxieanum Five populations were collected for this species pair:

- *R. clementinae*

  C1 Lao Jun Shan, (3800–) 4000 m, sympatric, all individuals have a characteristic morphology, no apparent intermediates.

  C2 Haba Shan, 4080 m, allopatric (sympatric with *R. phaeochrysum*), individuals with pure morphology despite hybridisation with *R. phaeochrysum* at the other lake end.

  Distance between populations: **C1-C2** 88 km

- *R. roxieanum*

  X1 Lao Jun Shan, 3700–3850 (–3950) m, sympatric, several individuals intermediate between var. *roxieanum* and var. *cucullatum* in the population, only apparently pure ones were used for analyses.

  X3 Shika Shan, 4090 m, allopatric, pure individuals, with often characteristically narrow leaves.

  Distance between populations: **X1-X3** 129 km

- *R. roxieanum* var. *cucullatum*

  XC1 Lao Jun Shan, (3800–) 4000 m, most individuals have a fixed morphology, but several intermediates between the two varieties of *R. roxieanum* can be found in a mixed habitat.

[8] G2b was largely excluded from the analysis, see Table 1.1, and chapter 3, section 3.1, page 96

R. aganniphum* and *R. phaeochrysum Eight populations were collectetd for these species:

- *R. aganniphum*

  G1 Shika Shan, 3700 m, allopatric, small population in gully, apparently pure individuals, no splitting indumentum.

  G2 Baima Shan ("Old Road"), 4300–4350 m, sympatric, highly intergrading population, most samples were therefore collected higher up a slope were the morphology indicated no admixture.

  G2b Baima Shui Shan (near top), 4300 m, sympatric, not a real population sample, but few individuals that were found amongst population GH2b. (later excluded from the analyses)

  G3 Dàxuě Shan, 4300 m, sympatric with P2, but no hybrids present, called allopatric in the analyses for distinction to the intergrading sympatric population.

  Distance between populations:
  G1-G2 81 km, **G1-G3** 90 km, **G2-G3** 78 km

- *R. phaeochrysum*

  P1 Baima Shan, 4000–4250 m, sympatric, highly intergrading population, only individuals with characteristic morphology were used for analyses.

  P2 Dàxuě Shan, 4300 m, sympatric with G3, but no hybrids present, called allopatric in the analyses for distinction to the intergrading sympatric population.

  Distance between populations: **P1-P2** 78 km

- *R. aganniphum* × *phaeochrysum*

  GH2a Baima Shan ("Old Road"), 4250–4300 m, hybrid swarm, population with very diverse indumentum morphology, some of which are *R. aganniphum* var. *flavorufum* (see Figure 1.12).

  GH2b Baima Shui Shan (near top), 4300 m, with indumentum morphotype E (see Figure 1.12), putative hybrids with homogeneous morphology.

Table 1.1: Populations sampled for this study; collected in 2007 & 2008. Listed according to species and giving the locality, population code (pop), and sample size. For a more detailed table of individuals please see Table D.1, page 187.

			Samples	
Species	Locality	Pop	Initial[b]	Used[c]
R. clementinae	Lao Jun Shan	C1	37	37
	Haba Shui Shan	C2	37	36
R. aganniphum	Shika Shan	G1	13	12
	Baima Shan (road)	G2[a]	37	35
	Baima Shan (top)	G2b	9	–[a]
	Dàxuě Shan	G3	33	33
R. phaeochrysum	Baima Shan (road)	P1	22	22
	Dàxuě Shan	P2	25	24
R. roxieanum	Lao Jun Shan	X1	56	53
	Shika Shan	X3	33	33
R. aganniphum × *phaeochrysum*	Baima Shan (road)	GH2a	34	33
	Baima Shan (top)	GH2b	33	32
R. roxieanum var. *cucullatum*	Lao Jun Shan	XC1	43	40
Total number of samples			412	390

[a] This population also corresponds to G2a (for further detailes see chapter 3, section 3.1, page 96).
[b] Number of individuals that were used for labwork.
[c] Number of individuals that had sufficient data to be included in the analysis.

1.7 Concrete aims of the study

After completing the first sample collections in 2007 following problems were chosen to be addressed in this study:

- Populations of the species seem to be isolated on inselbergs and mountain ranges, which can present considerable barriers to insect mediated pollen-flow. Is this reflected in the genetic differentiation of the populations?
- Given possible geneflow between sympatric populations via hybridisation, are sympatric populations of different species genetically less differentiated than allopatric populations of the same species?
- The described varieties seem morphologically intermediate between species. Are the described varieties of hybrid origin?
- The fixed morphology of *R. roxieanum* var. *cucullatum* suggests an F1 hybrid, and F1 dominated hybrid zones might represent barriers to geneflow. Therefore, is var. *cucullatum* an F1 hybrid and hence is reproductive isolation maintained despite the hybrid zone, or is there evidence for introgression?
- The *R. aganniphum* × *phaeochrysum* population at the "Old Road" at Baima Shan shows a gradient in indumentum morphology consistent with a hybrid swarm. *R. aganniphum* var. *flavorufum* is not very abundant, but seems intermediate. Is var. *flavoruvum* the F1 hybrid of the cross *R. aganniphum* × *R. phaeochrysum*, and are the other morphologically intermediate individuals backcrosses in both directions to each parent?
- The putative hybrid population on Baima Shui Shan has a very homogeneus morphology, and seems to occupy the altitudinal range which is normally dominated by *R. aganniphum*. Are these individuals a stabilised, later generation, hybrid taxon, potentially invading the niche of *R. aganniphum*? [9]

[9] This last point was obviously added after discovering this population in 2008.

1.8 Techniques chosen to investigate the study organism

To answer the questions introduced in section 1.4 (p. 23), three techniques were chosen. Because of the evidence of geneflow between lineages, and hence the possibility that different parts of the genome behave differently with regard to the descent, it was desirable to use a technique that would produce a large number of markers distributed throughout the genome. AFLPs seemed the best choice available at the planning stage of the thesis. Additionally, to obtain information on the directionality of crossing, a maternally inherited marker was desirable; in plants the best and most widely accepted choice is to use parts of the chloroplast genome. Because studies indicating the extreme scarcity of polymorphism in the group were not available at the start of the project [118], Sanger sequencing of chloroplast markers was included initially.

The aforementioned techniques can be used to investigate a fair number of individuals within normal financial and time restrictions. However, if huge numbers of individuals have to be screened (e.g. for seedling analysis), they may, due to funding restrictions, not be feasible. Mostly with an eye to future research, a third method, the analysis of leaf waxes, was chosen for a subsample of the genetically genotyped individuals; the aim was a comparison of genetic composition with leaf wax profile, and to test robustness in a population biological setting.

1.8.1 Amplified Fragment Length Polymorphism

Amplified Fragment Length Polymorphism (AFLP) is a method of genetical fingerprinting first described by Vos et al. [175]. It is based on the digestion of genomic DNA with restriction enzymes, most commonly *Eco*RI and *Mse*I (1)[10], and subsequent semi-selective PCR amplification of certain fragments. This amplification is enabled by attaching predesigned oligonucleotides, so-called adapters, to the sticky ends of the restriction sites (2), and then using these adapter sequences as primer binding sites during a PCR reaction (3). Reduction of the number of fragments is achieved by adding additional bases at the 3‘-end of the primers and hence selectively amplifying a subsample of fragments. Generally two PCR amplifications are performed; for the first (preamplification) one additional base is added to the primers (*Eco*RI restriction site - E+1, *Mse*I

[10] For number references in brackets in the following explanation please see Figure 1.16

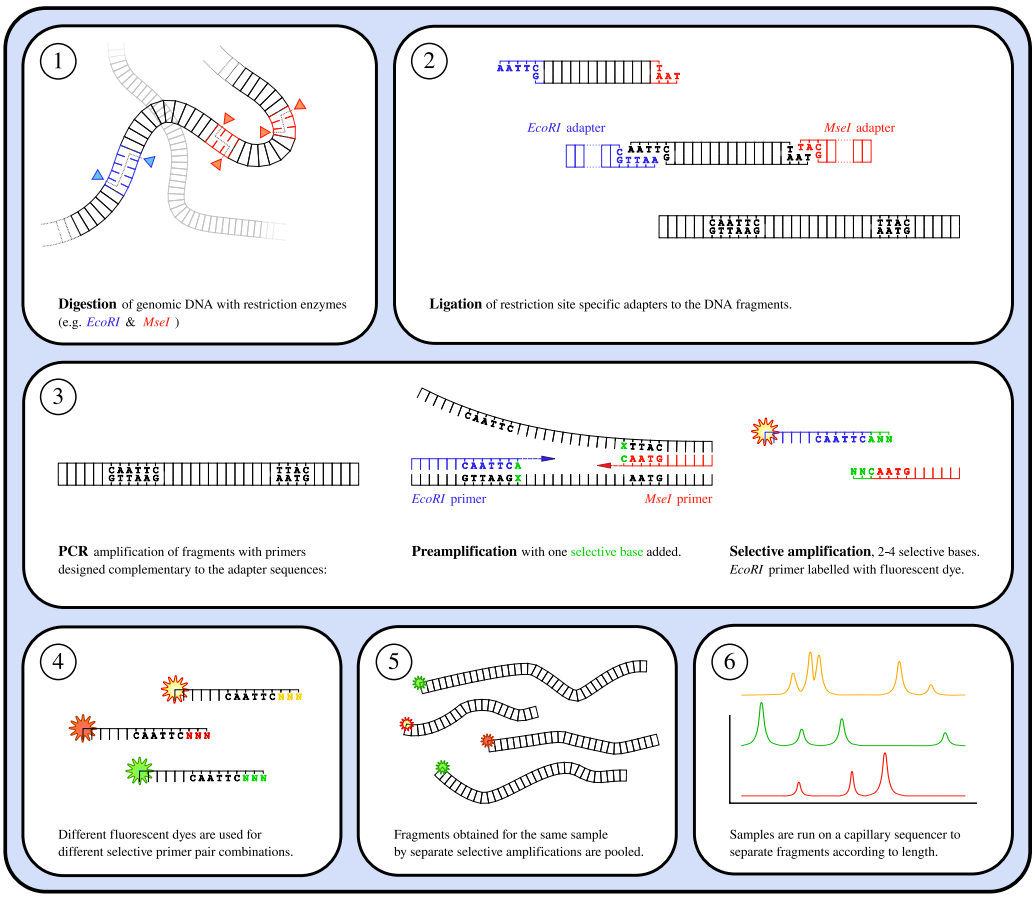

Figure 1.16: Amplified Fragment Length Polymorphism (AFLP), schematic showing the steps from cutting the genomic DNA to obtaining a profile.

restriction site M+1); for the second (selective amplification) three or more (*Eco*RI restriction site - E+3, *Mse*I restriction site M+3) are added. For the selective amplification dye labeled *Eco*RI primers are used (4), so that the amplified fragments resulting from at least one *Eco*RI restriction site can be detected on a capillary sequencer, and a chromatogram obtained for each sample. To improve the cost-effectiveness different dyes can be used in the selective amplification, and several primer-pair combinations subsequently combined in one sequencer run (5, 6).

The data generated are of dominant nature, so that at a given locus peak presence and absence is scored (coded as 1 and 0 respectively). This makes the direct detection of heterozygotes impossible, as they can not be distinguished from homozygous presence of the peak (1/1). This drawback is generally compensated for by the large amount of markers generated.

AFLPs are nowadays a well established molecular technique [115], and it

is thought that it is one of the most cost-effective and useful techniques to investigate questions of population structure and hybridisation in non-model organisms [18, 115]. The combination of restriction enzymes and adapter ligation allows the application of the technique without any preliminary knowledge of the genome of the organism in question. Furthermore the ideally random distribution of cutting sites throughout the genome minimizes the probability of only sampling from a small portion of the genome, which would be the case using classical sequencing techniques. Additionally the large amount of markers generated has more potential to resolve relationships of closely related species, and more power to detect backcrosses and underlying population structure.
AFLPs seemed to be the best available choice for following reasons:

- A marker system covering the whole genome is aimed for
- No whole genome information is available for *Rhododendron*
- Species seem to be extremely closely related
- Morphology suggests considerable backcrossing for one hybrid zone

1.8.2 Sequencing of chloroplast markers

Additionally to genotyping the individuals using AFLPs, which would give a whole genome image of the relatedness, a purely maternally inherited marker could provide valuable information about backcrossing patterns. The chloroplast normally fulfills this requirement in plants. Sequencing of chloroplast regions is widely established practice to reconstruct species relationships in plants, hence information for potentially suitable markers is easily obtained. Furthermore, RBGE has an extensive living collection of rhododendrons, which were available to establish suitable markers before population samples could be obtained from the wild.

1.8.3 Leaf waxes

In many plant species the epidermis is, additionally to the cuticle, covered by a layer of wax molecules. The morphology of these epicuticular waxes, or leaf waxes, can be of taxonomic importance in certain plant groups, and nowadays it is widely believed that the morphology of the waxes correlates with their chemistry [14]. One of the many molecule groups found in leaf waxes that is easily accessible for analysis are alkanes, and a relatively recent study [31] showed

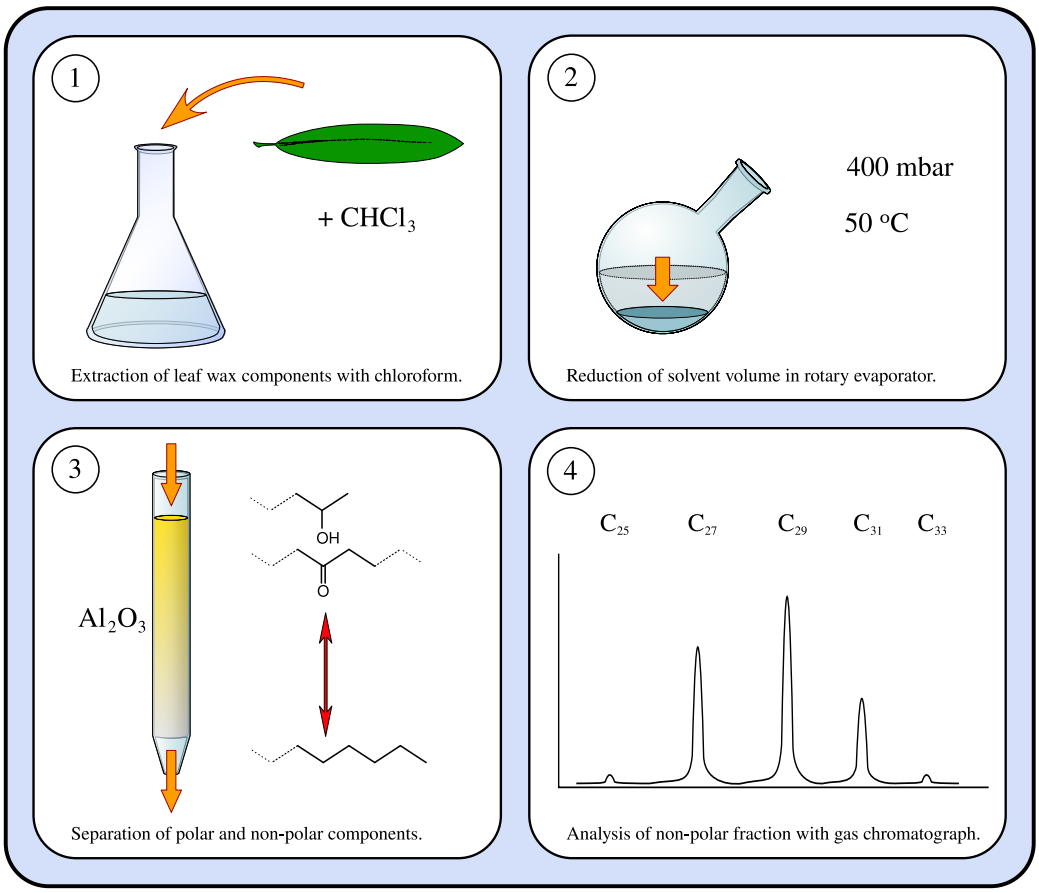

Figure 1.17: Leaf waxes, steps to extract waxes from leaf material.

that the distribution of n-alkanes can be a useful taxonomic tool in subsection *Taliensia*. In rhododendrons relative maxima of these saturated alkanes, with the general formula C_nH_{2n+2}, occur mostly for $C_{27}H_{56}$ and $C_{31}H_{64}$, and are distinctive for certain species. Furthermore, hybrids between two such species, with different maxima, exhibit an intermediate composition; often the amount of $C_{29}H_{60}$ is increased significantly, and hence the method can be used as an indicator of hybrid individuals. Additionally the technique is very cost effective, and has the potential to enable testing a large number of individuals for a hybrid background, if the conditions are right. The analysis is based on the extraction of the leaf waxes with an non-polar solvent, and subsequent detection of the quantity of several fractions of alkanes by gas chromatography (GC), see Figure 1.17. The amounts of the different extracted waxes are then normalised by defining the most abundant fraction as 100% (the maximum), and scaling all other fractions accordingly.

Although Chadwick at al. [31] were able to show that in rhododendrons the leaf wax composition is consistent in the same individual over time, and that it

that it seems to be characteristic for species, only a few individuals per species were surveyed, and no population samples were included. This marker was chosen in addition to the AFLP technique, as it offered the opportunity to detect hybrid individuals, and hint to patterns which a dominant marker based technique might not pick up (e.g. low levels of introgression). Furthermore, the setup of this study enabled the comparison with genetic data, and assessment of the consistency of the leaf waxes in population samples of the same species.

Chapter 2

Materials and Methods

2.1 Samples and explanation of codes

Samples Leaf samples taken in the field were put into sealable plastic zip-lock bags, and silica gel for faster drying was added to the bags the same day. For each population at least two morphologically representative individuals were chosen, and herbarium specimens prepared of these. Where possible it was aimed to collect at least 30 individuals per population, as this is considered to be an adequate sampling size for dominant marker techniques [23]. Sampling of individuals in close proximity to each other was avoided, as to minimise the chance of sampling the same clone, direct offspring or siblings. Of these collections 412 individuals were genotyped using AFLPs and for 115 additionally the composition of saturated *n*-alkanes (leaf waxes) was determined. Individuals of which specimens had been taken were preferred for the leaf wax analysis, to be able to consult morphology, or in case back-up leaf material would be needed; the rest was chosen randomly: 20 individuals each for *R. aganniphum*, *R. clementinae*, *R. phaeochrysum* and *R. roxieanum*; 20 potential hybrids (*R. aganniphum* × *phaeochrysum*), and 15 individuals of *R. roxieanum* var. *cucullatum*.

(see Table 1.1, page 49, for a summary of population samples, and Table D.1, page 187, for information about single individuals; individuals used for leaf wax analysis are given in Table D.2, page 209)

Label code Individuals were labelled adopting a code scheme introduced by Richard Milne; it was slightly changed to include new species and population sampling. The code generally consists of two letters followed by three ciphers. The first letter stands for the subsection in *Rhododendron*, the second for the species; the number formed by the three ciphers is the individual number. In

potential hybrids three letters and two ciphers were used, to include both potential parent species as letters two and three. The individual number in these cases therefore only consists of two ciphers. The first letter in the code is always a `T` for subsection *Taliensia*, as no species from other subsections were investigated. The species are coded with the letters given in Table 2.1. Special cases are the two varieties *R. aganniphum* var. *flavorufum* and *R. roxieanum* var. *cucullatum*, which use the hybrid letter pattern, but the third letter stands for the variety; `F` for var. *flavorufum*, and `C` for var. *cucullatum*. For clarification some examples:

`TX023` subsection *Taliensia* (`T`), *R. roxieanum* (`X`), individual `023`

`TGP04` subsection *Taliensia* (`T`), potential hybrid individual `04`, with assumed parents *R. aganniphum* (`G`) and *R. phaeochrysum* (`P`)

`TGF01` subsection *Taliensia* (`T`), *R. aganniphum* var. *flavorufum* (`GF`), individual `01`

`TXC10` subsection *Taliensia* (`T`), *R. roxieanum* var. *cucullatum* (`XC`), individual `10`

Table 2.1: Labelling of Samples. Letters used in species codes to denominate species.

Letter	Species
`A`	*R. alutaceum*
`B`	*R. beesianum*
`C`	*R. clementinae*
`G`	*R. aganniphum*
`O`	*R. proteoides*
`P`	*R. phaeochrysum*
`R`	*R. traillianum*
`S`	*R. sphaeroblastum*
`T`	*R. taliense*
`X`	*R. roxieanum*

2.2 Labwork

2.2.1 DNA extraction

DNA was extracted from silica dried leaf material, using the DNEasy kit from Qiagen with a slightly modified protocol. Apart from *R. phaeochrysum*, the species studied have a thick indumentum; Firstly, the formation of a soft hair-ball during mechanical break-up interferes with efficiency considerably, and secondly, in the following extraction steps the hairs would absorb too much liquid, and potentially bind DNA. Therefore it proved to be essential to remove these hairs, before weighing 20 mg of dried leaf material. The dried leaf material was put into a 2 ml eppendorf tube, and a pinch of Aluminium oxide (SIGMA, A2039-5006, Type 5) was added. The samples were then homogenised using a TissueLyser (Qiagen). Generally the protocol, as described in the handbook [137] (p. 24ff) was followed, with alteration to following steps:

4-6 The samples were ground at 20 Hz for 45 seconds, then the position of the tubes was reversed, and the samples were ground for another 45 seconds at 20 Hz. This procedure was repeated twice.

7-9 450 μl of AP1 were added to the, finely pulverized, plant material, and after mixing well by vortexing, the sample was incubated in a water bath at 65 °C. After 10 minutes 4 μl of RNaseA were added, and the solution mixed by stirring with a pipet tip. After further incubation at 65 °C for 15 min, once mixing by inverting the tube after 5–7 min, 150 μl of AP2 were added, and after vortexing vigorously, the sample was placed into the freezer for 10–15 min.

16-17 step 17 was a dry spin, without adding any Buffer AW, to dry the matrix.

18-19 Buffer AE was heated to 70 °C, before pipetting onto the column, and step 19 was skipped.

Success of the extractions was checked by running 5 μl of the eluted DNA at 80 V for 45 min on a 1.5 % agarose gel, stained with SybrSafe, using a 1kb+ ladder as size standard. The extraction was deemed successful if an intense enough band, larger than 12,000 kb, could be detected under UV light. The extracted DNA was then stored in the freezer at −20 °C.

2.2.2 Chloroplast markers

At the beginning of the project three individuals of *R. alutaceum*, three individuals of *R. traillianum*, and one individual of *R. roxieanum* where used to test three chloroplast regions (*trn*S-*trn*G, *trn*S-*trn*Fm and *trn*L) for polymorphisms, using primers already in stock at RBGE. As the regions were completely monomorphic new potential candidates, with high potentially informative character value (PIC sensu Shaw et al [155]) were identified in the literature [155, 156], and other primers in stock at RBGE were included. These newly selected regions were sequenced in seven new individuals, which were assumed to span a wider range of the subsection: two individuals of *R. alutaceum*, three of *R. roxieanum* and one each of *R. taliense* and *R. traillianum*. In total 18 different regions were sequenced for these individuals (for tested primers see Appendix A, Table A.1, page 174), but none was polymorphic enough to distinguish species from each other; although sporadic mutations did occur, all were singletons, and not representative for the species. It was concluded that chloroplast sequences do not yield information about relationships within subsection *Taliensia*, as the species are too closely related (also see [118] for resolution achieved with chloroplast markers). This part of the project was not followed further, and no data has been included in this thesis.

2.2.3 AFLPs

During the setup of the method it became apparent that repeatability of the AFLPs depended to a considerable degree on the amount of initial genomic DNA used. Therefore the concentration of the extracted genomic DNA was measured using a NanoVue spectrophotometer (GE Healthcare). The consistency of these measurements did not live up to the initial expectations, most likely due to interference at the measured wavelengths of compounds present in the AE buffer from the Qiagen kit the DNA was dissolved in. Owing to this problem, only the lowest concentration measurements were considered, and the amount of genomic DNA used for digestion by restriction enzymes chosen as to insure that at least 50 ng of DNA were used (but up to ∼250 ng), which agrees with what has been reported in the literature to yield consistent results [18]. Using this protocol the AFLP patterns were highly repeatable.

Digestion & Ligation

Often the digestion and ligation are performed in one single step (e.g. [175]), but as separate reactions were common lab practice at RBGE, this was adopted for the protocol used here.

Digestion Genomic DNA was digested with the restriction enzymes *Eco*RI and *Mse*I. Before use the enzymes where tested for activity by performing a digestion of four samples of genomic DNA with each enzyme separately; 10 μl of DNA were digested with 10 U of enzyme, and 5 μl run on an agarose gel to check for cutting activity. After verification of enzyme activity, genomic DNA of all samples was digested on the same day, using the same master mix to minimize treatment differences (see Appendix B, section B.2, page 179, for details).

Ligation The adapters used were the same as reported by Vos et al. [175] (see Table A.2, page 175). Adapters were prepared by bringing 400 ml water to boil in a 500 ml glass beaker; mixing the corresponding forward and reverse oligonucleotides with distilled water in a 1.5 ml eppendorf tube, and placing it in a floater into the boiled water. In the beaker they were left to cool down to room temperature (for more details see Appendix B, section B.1, page 177). Ligation was then performed by preparing a ligase-adapter mastermix for all samples, and adding 10 μl of the prepared mix to 15 μl of DNA, digested the same day; afterwards incubating over night. For consistency the mixture was not left at room temperature, but a thermocycler program with low temperature setting was used (see Appendix B, section B.3, page 179).

Preamplification

For the amplification steps a protocol suggested by Beckman Coulter Inc. [76] was followed, with slight adaptations owing to different chemicals used (for full protocol see Appendix B, section B.4, page 180). Preamplification was performed one day after ligation, and all samples were processed together using the same mastermix; afterwards they were incubated in several thermocyclers over night. Primers used for preamplification (E+1 and M+1) are given in Table A.3 (Appendix A, page 175). Success of the amplification was tested by running all samples on agarose gels, and checking for characteristic smear.

Selective amplification

To enhance the repeatability, performance, and scorability of resulting traces, several combinations of different preamplification dilutions and polymerases were tested for the selective amplification. DNA of one individual of each species (`TC060, TG083, TP079, TX113`) was used for the tests:

- two different preamplifications with dilutions between 1:5 and 1:10
- polymerases: *Taq*, Phire and GoTaq

(for product specifications see Appendix A, subsection A.2.1, page 171). Despite other protocols suggesting that higher grade and hot-start polymerases improve the traces (e.g. [76]), the normal *Taq* polymerase performed best; Furthermore dilution of the preamplification higher than 1:7.5 (normally 1:10 or higher is suggested [175, 76]) led to inconsistencies in some traces. Therefore a dilution level of 1:7.5 was adopted, and selective amplification performed using *Taq* polymerase (for the protocol see Appendix B, section B.5, page 180). After establishment of the protocol, 50 primer-pair combinations (*Eco*RI-*Mse*I) were tested for scorability, using one individual of each species (see Table A.4, Appendix A, page 176 for tested combinations). Among the ones with good traces and peak spacing, seven were chosen to be tested with 32 individuals (eight of each species) for polymorphisms and fixed differences. For cost reasons this second test was set up to use one 96-well plate; to label the *Eco*RI primers three different WellRED dyes are available for the CEQ system, so that a maximum of nine primer pair combinations could theoretically have been tested (96 wells $\div$ 32 individuals $\times$ 3 dyes). However, not all dye/primer combinations were available, and only the most promising combinations per dye were included; For the *Eco*RI primer E-AGC none of the previous combinations had fulfilled the requirements, but as space was still available for this particular dye (D3), a new combination was chosen to be tested; avoiding complete exclusion of this primer at this stage. As mentioned before, three primer pair combinations had to be chosen because of the restriction that three dyes are available to be pooled in one sequencer run. The three combinations with the best results were: E-ATC/M-CAG, E-ATC/M-CGA and E-ACT/M-CTA, which were therefore used in the study (for primer sequences see Table A.3, page 175).

After the setup, all selective amplifications were carried out using the same protocol (Appendix B, section B.5, page 180), one primer pair combination a day. All individuals were processed together on one day using the same master mix, and only singular non successful reactions repeated. The different dye labelled

reactions were then pooled, and run on a CEQ 8000 Genetic Analysis System (Beckman Coulter Inc.).

Scoring

The traces obtained were visualised with "CEQ DNA Analysis Software Version 8.0 for DNA sequencing and Fragment Analysis" (Beckman Coulter), and scored manually. Because of many known problems with the scoring of AFLP traces [22], and the missing functionality of the software to normalise traces, the closest peaks that were present in all traces were used to estimate the intensity of the trace; peaks were scored if they surpassed a certain height. Exceedingly low peaks were not scored, as mispriming might be a reason for their occurrence [23], and due to the used statistics less error would be introduced by false absence than false presence. To estimate the error rate of the AFLPs, four control individuals were run on every plate to estimate within individual variance, and the test runs for the used primer pair combinations were used to estimate overall error rate. During the scoring process loci were grouped into four categories according to clearness/height of peaks and separation from other peaks. Loci with many extremely low or merging peaks[1] were classified as 'C' and not used at all; category 'B' would include markers with some low peaks and well separated double peaks; if only one or two individuals had low peaks category 'AB' was applied, and loci with only very clear peaks were included in category 'A'.

2.2.4 Leaf waxes

Extraction

Silica dried leaf material was used for the extraction. The amount used per individual was roughly determined by estimating the leaf surface area through placement on a 10×10 cm square. The dry leaves were broken into smaller pieces, placed into a 100 ml Erlenmeyer flask, and covered with approximately 25 ml analysis grade chloroform. Samples were left for 15 minutes, gently swirling once after roughly 10 minutes. The solution was then filtered through a paper filter into a round-bottomed flask and evaporated to dryness using a rotary evaporator (50 °C, 400 mbar). To remove the polar fraction the residue was dissolved in 2–3 ml of chloroform, and passed through an alumina column (glass pipet filled with aluminium oxide); only recovering the first two-thirds of the flow through

[1] In some traces a clear double peak would be visible, while in others a single broad peak; latter frequently indistinguishable from a peak representing one band only.

to avoid eluting the polar fraction. This eluate was left to evaporate to dryness at room temperature, and the the non-polar fraction was weighed.

Run on gas chromatograph (GC)

Before analysis of the components with the gas chromatograph a standard solution (subsection A.2.3, page 173) was prepared and added to every sample. The amount added to each sample was chosen as to roughly obtain a concentration of 5 mg/ml of the extracted fraction, but at least 1 ml were added; this was the minimum the GC injector needle could reach. The samples were then run on an AutoSystem XL GC (Perkin Elmer), with a PE-5 column (N931-6076, length: 30 m, inner diameter: 250 μm, film: 0.25 μm); nitrogen (N_2) as carrier gas, and injecting 2.0 μl sample while using a split ratio of 50:1; the program settings were:

Initial temperature	80 °C
Hold	2 min
Ramp 1	30 °C / min to 220 °C
Hold	1 min
Ramp 2	2 °C / min to 295 °C
Total run time	45.17 min
Detector temperature	300 °C

2.3 Statistics

This section will give a summary of the statistics calculated in this thesis. Additionally, reasons for their use will be given and assumptions made discussed. It is also meant as a reference to the actual equations used to obtain these, mostly well-established, statistics, as sometimes more than one equation is available to obtain a similarly termed statistic, e.g. for F_{ST} [177]. Furthermore, where statistics were implemented and calculated in R [138] the equations represent back-translations of the functions used or written by the author. Although this means that the actual formulation of some equations deviates from the form in which they were originally presented in the cited publications, this approach was adopted to document more adequately how the final statistics were obtained, and to make it easier to relate them to the actual functions. The source code of R functions used for calculations is available in electronic form on the CD added to this thesis.

2.3.1 Dominant markers

Classically AFLP patterns were generated on gels [175], and at each locus a fragment (band on the gel) could be present or absent for a given individual. Nowadays patterns are generally obtained through automated detection of fragments labelled with fluorophores. Hence data are obtained by scoring peaks in a chromatogram. However, in published statistical methods the peaks are often referred to as bands. The caveat of the technique is now that if a band is present at a given locus, in a diploid individual, one can not distinguish if one allele is producing the band or both. This means that the genotypic state of an individual is not directly accessible, and only in the case of band absence are both alleles at the locus known with certainty. This property of a marker is referred to as dominance, as the presence dominates over the absence of a given allele. This is the most important property of AFLPs with regard to statistical approaches.

Because in the case of band presence a heterozygote can not be distinguished from a homozygote, two major approaches of analysing AFLP data are generally followed [23, 115]: One possibility is to treat the data effectively as haplotypic/phenotypic, as opposed to genotypic, and hence use methods that are only based on the observed presence or absence of bands. Although this disregards the potential heterozygous state at many loci, no assumptions have to be made, and statistics based on the genetic composition of single individuals can be used. Most of these band-based methods represent different types of distance measures [23],

or rather, statistics using these distance measures as a starting point. These will be further discussed in subsection 2.3.2 and subsection 2.3.4.

The second approach is to treat the data on a population-based level [115]. Using the information from several samples pertaining to the same population, the allele frequencies which should underlie the phenotypic composition can be inferred. However this improves the information content derived from the data, without additional information, certain assumptions about the proportions of homozygotes to heterozygotes have to be made, and the accuracy of inference is dependent on the (unknown) deviation from these assumptions [24]. These methods will be addressed in subsection 2.3.3 and subsection 2.3.6.

2.3.2 Distance measures

First we will have a look at distance measures for dominant data, which can be interpreted as measures of genetic distance, or equivalently genetic similarity [23]. These are the basis for statistics based on band counts.

Table 2.2: Types of matches between two AFLP profiles

		Individual i	
		band present (1)	band absent (0)
Individual j	band present (1)	a	b
	band absent (0)	c	d

Distance measures are obtained by first calculating a similarity coefficient and then subtracting it from 1 to convert it into a dissimilarity.

$$D = 1 - S$$

The similarity coefficients are calculated by summing over all possible types of matches between two individuals for all loci (see Table 2.2 and 2.3). This is done for all pairwise comparisons of individuals comprising the data. The coefficients differ in how these matches are weighted. Two different similarity coefficients were used in this study:

1. Jaccard coefficient, J (Jaccard 1908)

2. Simple-matching coefficient, SM (Sokal and Michener 1958)

Table 2.3: Pairwise matches of individuals i and j

	Locus									
Individual i	1	1	1	0	0	0	0	1	0	...
Individual j	1	0	1	1	0	1	0	1	1	...
Type of match	a	c	a	b	d	b	d	a	b	...

The Jaccard coefficient only takes comparisons into account in which at least one of the individuals has a band at a given locus. Therefore a shared absence of a band is not considered as a similarity. This avoids error through homoplasy [115], where the absence of the band is due to different mutations. Obviously, to account for homoplasy becomes more important the more distantly related the individuals are.

$$J_{ij} = \frac{\sum a_{ij}}{\sum a_{ij} + \sum b_{ij} + \sum c_{ij}}$$

Excluding negative matches makes the Jaccard coefficient an asymmetric measure, and the distance obtained $(1 - J)$ is therefore also called asymmetric binary (e.g. in R function `dist`). Gower and Legendre [67] (Table 4, p. 34) showed that the resolution obtained by the Jaccard coefficient is higher in comparison to the simple matching coefficient; it was therefore used to calculate the distances for statistics based on single individuals (MDS and NJ, subsection 2.3.4).

Table 2.4: J vs. SM influence on dissimilarity within groups.

	Locus 1	Locus 2
Individual A	1	0
Individual B	1	0
Individual C	1	0
Individual D	1	0
Individual E	1	0
Individual F	1	0
Individual G	1	0
Individual H	1	0
Individual I	1	0
Individual J	0	1
Dissimilarity*	$\frac{9}{45}$	$\frac{9}{9}$

* number of mismatches divided by number of counted comparisons

However, when estimates for subsets of the data are calculated, as for populations or species, the asymmetric behaviour of this coefficient introduces a considerable bias, overestimating dissimilarities within groups. This is especially pronounced when similarities are based on shared absences as in the case of low frequency markers. To illustrate the problem we consider one group of ten individuals at two different loci, which theoretically should give the same estimate of within-group dissimilarity, Table 2.4. Although the distance between single individuals is unaffected (only individual J has a distance of 1 to all other individuals, for each of the two loci separately) the within-group estimate for locus 2 is dominated by the low frequency allele as all other pairwise comparisons will not be considered (0—0 matches).

Furthermore, published data suggests that the species in this study are very closely related [118], and the error by assuming homology for shared absences can therefore be regarded negligible in comparison to the mentioned systematic bias when calculating group statistics. For this reason the Simple-matching coefficient which considers shared absences was used to calculate band-based group statistics (AMOVA, subsection 2.3.5).

$$SM_{ij} = \frac{\sum a_{ij} + \sum d_{ij}}{\sum a_{ij} + \sum b_{ij} + \sum c_{ij} + \sum d_{ij}}$$

Perhaps it is worth mentioning here that Bonin et al. [23] also recommend this coefficient for the AMOVA. However they give as reason (p. 3739):
"*This coefficient has interesting Euclidean metric properties that allow its use in an analysis of molecular variance ...*"
This is not, in this form, correct. Neither the Jaccard coefficient nor the Simple-matching coefficient (or their equivalent distance measures $1 - S$) are a Euclidean metric [67]. However, both can be transformed into one by taking their square root transformation ($\sqrt{1-S}$). Hence they do not differ in this property, and in my opinion the better reason to use SM over J is the aforementioned bias problem.

2.3.3 Estimating allele frequencies

Band-based methods do not take into account the real frequencies of the alleles present in the populations. However, population-based statistics, using approaches based on population genetic theory, require allele frequencies [188]. To partly overcome the limitations of the dominant nature of AFLP markers

Table 2.5: Expected genotype frequencies assuming Hardy-Weinberg proportions, and with inbreeding

Genotype	**Expectation**	$F = 0$	$F = 1$
1—1	$p^2(1-F)+pF$	p^2	p
1—0	$2pq(1-F)$	$2pq$	0
0—0	$q^2(1-F)+qF$	q^2	q

where $p + q = 1$,
$F = 0$ means random mating, and
$F = 1$ is complete inbreeding.
adapted from [74] p.144

one can estimate the allele frequencies (p, q) of the markers within groups. If the inbreeding coefficient (F), obtained from codominant data or crossing experiments [122], for the investigated population is available, it can be used to get an estimate of the frequencies according to theory, Table 2.5.

If this information is not available one has to consider if Hardy-Weinberg proportions can be assumed, and hence whether inbreeding is absent or negligible. Rhododendrons are mostly assumed to be obligate outcrossers, and even in the few cases where inbreeding occurs under extreme conditions, it has a negative effect on seed set or the offspring [77, 179], so that under normal conditions inbreeding in natural populations should not occur; deviations from Hardy-Weinberg equilibrium can therefore be considered unlikely. Making this assumption allows simplification of the equations in Table 2.5 by setting $F = 0$. Further assumptions that have to be made for Hardy-Weinberg equilibrium are that individuals within groups mate randomly, and that the loci used are not under selection. AFLP markers are randomly distributed throughout the genome, and are therefore assumed to be neutral [24]. Furthermore, the floral morphology of the species in this study is very similar, and there is no data that suggests that pollinators would distinguish between individuals of one species. However, care should be taken to estimate allele frequencies only for populations and not within species, as random mating cannot be assumed for geographically distant groups. It is important to note that all following approaches to estimate allele frequencies assume Hardy-Weinberg proportions within populations [109, 188], particularly as some publications state otherwise [115].

Using $F = 0$ in the equations in Table 2.5 would give an estimate of the null allele frequency q of:

$$\hat{q} = \sqrt{\frac{m}{n}} \tag{2.1}$$

where n is the total number of individuals, and m is the number of individuals that have no band at the given locus.

But, as the obtained data is nearly always a subset of the whole population Lynch and Milligan [109] showed that this estimate of q is downwardly biased due to sampling error, and that a better estimator is obtained by:

$$\hat{q} = \sqrt{\hat{x}} \times \left(1 - \frac{Var(\hat{x})}{8\hat{x}^2}\right)^{-1} \tag{2.2}$$

where $\hat{x} = \frac{m}{n}$, and $Var(\hat{x}) = \hat{x}(1-\hat{x})/n$

While this estimator always yields a better estimate for q than equation 2.1 it does not efficiently correct for bias when the null allele is rare ($\frac{m}{n} < 0.1$) [109].

A Bayesian approach, introduced by Zhivotovsky [188], addresses this problem, and is therefore most frequently used [23, 115]. This approach was also chosen to estimate the allele frequencies in this study. It was implemented as a function in R for the simulation of hybrids based on allele frequencies (see 2.3.8, page 87), and is one option in the program AFLP-SURV 1.0 (see 2.3.6, page 83). The frequency estimate is based on the expectation of q with respect to $Pr(q|m)$, (equation 3 in [188]):

$$\hat{q} = \int_0^1 Pr(q|m)\, q \, \mathrm{d}q = \frac{\int_0^1 Pr(m|q)\, Pr(q)\, q \, \mathrm{d}q}{\int_0^1 Pr(m|q)\, Pr(q) \, \mathrm{d}q} \tag{2.3}$$

where $Pr(q)$ is the prior distribution of q

Because q cannot be directly observed, a prior distribution of $\hat{x}$ (R in Zhivotovsky) is used, which is obtained from a beta distribution with positive parameters a and b, giving the estimator:

$$\hat{q} = \frac{\mathrm{beta}(m + a + 0.5,\ n - m + b)}{\mathrm{beta}(m + a,\ n - m + b)} \tag{2.4}$$

and its variance

$$s^2_{\hat{q}} = \frac{\mathrm{beta}(m + a + 1,\ n - m + b)}{\mathrm{beta}(m + a,\ n - m + b)} - \hat{q}^2 \tag{2.5}$$

(please see [188] for intermediate steps, as it was beyond my scope to fully understanding them)

If a uniform prior for q is used, $a = 0.5$ and $b = 1$ in equation 2.4. However, better estimates are often obtained when a and b are estimated from the data.

(equation 12 in [188]):

$$\bar{x} = \sum_{i=1}^{k} \frac{n_i}{n} x_i \,, \quad \sigma_x^2 = \sum_{i=1}^{k} \frac{n_i}{n} x_i^2 \; - \bar{x}^2 \tag{2.6}$$

where $x_i = \frac{m_i}{n_i}$, and $m_1, m_2, \ldots, m_k$ is the number of null homozygotes at a certain locus in k samples of size $n_1, n_2, \ldots, n_k$, respectively.

The estimators for the parameters of the beta distribution are then calculated by (equation 13 in [188]):

$$\hat{a} = \bar{x} \left(\frac{\bar{x}(1-\bar{x})}{\sigma_x^2} - 1 \right) \,, \quad \hat{b} = (1-\bar{x}) \left(\frac{\bar{x}(1-\bar{x})}{\sigma_x^2} - 1 \right) \tag{2.7}$$

When many populations are available the samples ($1 \ldots$ to k) can be obtained by treating every single population as a sample. This would require several populations of each species, which are not available in this study. However, a non-uniform prior distribution can be obtained by using the different loci of one population as samples and using the between locus values to obtain $\bar{x}$ and σ_x^2 (Note 4, p. 912 in [188]). Then equation 2.6 becomes:

$$\bar{x} = \frac{1}{L} \sum_{i=1}^{L} x_i \,, \quad \sigma_x^2 = \sum_{i=1}^{L} \frac{1}{L} x_i^2 \; - \bar{x}^2 \tag{2.8}$$

where L is the number of loci.

This approach was also adopted in the program AFLP-SURV [174], and a comparison of the frequency estimates obtained by the R function and AFLP-SURV showed that they are exactly the same.

2.3.4 Distance based clustering

One method to interpret the genetic distances obtained using the similarity measures described in subsection 2.3.2 is to display them graphically. However, the distance data obtained from AFLPs is generally multidimensional, as every locus has the potential to represent a dimension. In this form the data is not accessible to graphical analysis, and two methods where used to display the data in a two dimensional space:

1. Multidimensional Scaling (MDS)

2. Neighbour-Joining (NJ)

For both methods a dissimilarity matrix, obtained with the Jaccard coefficient (see subsection 2.3.2), was used.

Multidimensional Scaling (MDS)

Classical multidimensional scaling, also called principal co-ordinates analysis [66], can be used to reduce dimensionality by only displaying the dimensions that account for the largest proportion of the distances in the dataset. To extract these dimensions, the vectors corresponding to the pairwise distances of the individuals are normalised and their latent roots evaluated. The latent root, or eigenvector, can be understood as a dimension in which a certain amount of the overall distances are represented. Therefore the higher the eigenvalue of a latent root, the more of the overall distance within the data it represents. After rescaling the data accordingly, only the dimensions with the two (or three) largest eigenvalues are used for plotting. The fit of the data to the chosen dimensionality can be evaluated by comparing the eigenvalues of the plotted eigenvectors to the eigenvalues of the omitted eigenvectors. MDS was performed in R using the function `cmdscale`; and plotted using functions: `s.class` and `scatterutil.eti`, provided in the package ade4 [43]; and function `scatterplot3d`, package scatterplot3d [105].

Neighbour-Joining (NJ)

The neighbour-joining method is an agglomerative algorithm based on the minimum-evolution principle [62]. The algorithm proceeds as follows (taxa correspond in the beginning to single individuals):

1. One pair of taxa (the neighbours) is grouped and replaced by a node representing the group.
2. The distance matrix is simplified by replacing the taxa with the new node.
3. The node will then be treated as a single taxon and steps 1. and 2. are repeated.
4. This process will be followed until only three taxa remain.

Following the minimum-evolution principle, the tree with the smallest sum of branch lengths is then chosen. The NJ tree was calculated using the algorithm of Gascuel [62], implemented in the R package `ape`, as function `bionj`.

2.3.5 Analysis of Molecular Variance

The graphical methods described above are individual-based and only marginal information can be obtained about the characteristics of groups comprising the data. To obtain species and population based statistics from the distance measures an Analysis of Molecular Variance (AMOVA) was performed. AMOVA is a method to obtain variance components of different hierarchical levels of the data. It was developed by Excoffier et al. [54] for the analysis of molecular variation within a single species. The data is divided into predefined hierarchical groups classically corresponding to populations and regions, and the variance within and among groups is expressed as a percentage of the total. This allows the assessment of differentiation and variability among and within subsets of the data. Most studies use the program ARLEQUIN [53] written by Excoffier to perform this analysis, which is an obvious approach. However, in the case of AFLP data, the author himself stated that the program is not suited for dominant data. Citing a comment by Laurant Excoffier on the genetics-software-forum[2] : "*... first I need to remind people that AFLP data is not supported by Arlequin ...*""

A further problem with the analysis is that in its classical setup it is based on allele counts within groups, and evolutionary distances between these alleles. Therefore, loci with more than two alleles are needed to obtain reasonable estimates. One way around this is to pool all loci and treat them as one allele, but for the data in this study this is not feasible, as every individual will represent a unique allele. For these reasons I implemented an AMOVA in R that uses the distance matrix of the individuals directly, rather than allele counts; instead of geographic regions, the second hierarchical level is represented by species.

AMOVA for AFLPs in R

The function written in R is almost entirely based on the original work by Excoffier et al. [54], and equation references will refer to this publication; likewise, functions and packages will refer to R. Further ideas for the code were taken from the R function `amova` written by Sandrine Pavoine, in package `ade4` [43]. The main difference to a classical AMOVA is that the distance matrix D^2, obtained in Excoffier et al. [54] from evolutionary distances between haplotypes of one locus, is replaced by a matrix of inter individual distances obtained with the Simple-matching coefficient (see subsection 2.3.2).

[2] `http://gsf.gc.ucdavis.edu/viewtopic.php?f=8&t=1000`, accessed 22.07.2010.

The distance matrix was obtained using the function `dist.binary` in package `ade4` [43] using `(method = 2, upper = TRUE)`. Because this function returns the square root of D, the distances were squared before using them in the following calculations.

The sums of squared deviations (SSDs) were then calculated for the different subsets of the data. Firstly the within group (species, population) SSDs are calculated, which are named as follows:

$SSD_{(total)}$	SSDs for the whole dataset
$SSD_{(ws)}$	SSDs within species
$SSD_{(wp)}$	SSDs within populations

The equations to obtain these values are based on equations 5b and 8a–c in Excoffier et al.:

$$SSD_{(total)} = \frac{1}{2N} \sum_{j=1}^{N} \sum_{k=1}^{N} \delta_{jk} \tag{2.9}$$

where N is the total number of individuals, and δ_{jk} is the squared distance between individual j and k.

$$SSD_{(ws)T} = \sum_{i=1}^{S} SSD_{(ws)i}\,, \quad SSD_{(ws)i} = \frac{1}{2N_{si}} \sum_{j=1}^{N_{si}} \sum_{k=1}^{N_{si}} \delta_{jk} \tag{2.10}$$

where N_{si} is the number of individuals in species i, and S is the total number of species.

$$SSD_{(wp)T} = \sum_{i=1}^{P} SSD_{(wp)i}\,, \quad SSD_{(wp)i} = \frac{1}{2N_{pi}} \sum_{j=1}^{N_{pi}} \sum_{k=1}^{N_{pi}} \delta_{jk} \tag{2.11}$$

where N_{pi} is the number of individuals in population i, and P is the total number of populations. Next the SSDs between groups of one hierarchical level can be calculated by subtracting the SSDs within groups from the total of the respective level:

$SSD_{(ap)}$	SSDs among populations within species
$SSD_{(as)}$	SSDs among species

can then easily be obtained as follows:

$$SSD_{(ap)T} = \sum_{i=1}^{S} SSD_{(ap)si}\,, \quad SSD_{(ap)si} = SSD_{(ws)i} - \sum_{j=1}^{P_{si}} SSD_{(wp)j} \tag{2.12}$$

where P_{si} is the number of populations in species i, and $SSD_{(ap)si}$ are the sums of squared deviations among populations in species i.

$$SSD_{(as)T} = SSD_{(total)} - SSD_{(ws)T} \tag{2.13}$$

Because the SSDs are the sums of squared deviations from the centroid of a multidimensional space [54], which is dependent on the corresponding degrees of freedom of the subset, the mean squared deviations have to be calculated before the attributable variance components can be obtained.

The mean squared deviations (MSDs) are obtained by dividing the SSDs by their corresponding degrees of freedom (df), which are as follows:

$$\begin{aligned}
df_{(as)T} &= S - 1 \\
df_{(ap)T} &= P - S && = \sum_{i=1}^{S} df_{(ap)si}\,, \quad df_{(ap)si} = P_{si} - 1 \\
df_{(wp)T} &= N - P && = \sum_{i=1}^{P} df_{(wp)i}\,, \quad df_{(wp)i} = N_{pi} - 1 \\
df_{(total)} &= N - 1
\end{aligned} \tag{2.14}$$

where S is the number of species, P is the number of populations, $df_{(wp)i}$ are the degrees of freedom within population i, and $df_{(ap)si}$ are the degrees of freedom among populations in species i.

The MSDs are then:

$$\begin{aligned}
MSD_{(as)T} &= \frac{SSD_{(as)T}}{df_{(as)T}} \\
MSD_{(ap)T} &= \frac{SSD_{(ap)T}}{df_{(ap)T}} \\
MSD_{(wp)T} &= \frac{SSD_{(wp)T}}{df_{(wp)T}}
\end{aligned} \tag{2.15}$$

The correction for different sample sizes takes account of the proportion of pairwise comparisons within groups with regards to the total number of pairwise comparisons. This proportion is Cockerham's α [38]:

$$\alpha = \frac{\sum_{i=1}^{G}(N_i^2)}{\left(\sum_{i=1}^{G} N_i\right)^2} \tag{2.16}$$

where G is the number of groups, and N_i is the number of individuals in group i. The average sample size attributable to between group comparisons for a hierarchical level of size N_L is then:

$$n_L = \frac{N_L(1-\alpha_L)}{df_L} \tag{2.17}$$

Where $df = G - 1$ are the corresponding degrees of freedom.

If a further hierarchical level, below L, let it be called $L-1$, is considered, each of the groups of L will be split into subgroups; Each of the groups in L then has their own corresponding α, and the corresponding average sample size for between subgroup comparisons, taking into account the grouping in L, becomes:

$$n_{L-1} = \frac{\sum_{i=1}^{G_L}[N_i(1-\alpha_i)]}{df_{L-1}} \tag{2.18}$$

Where G_L is the number of groups at the hierarchical level L, and G_{L-1} is the total number of subgroups, N_i is the sample size of group i, α_i is the corresponding α for group i, and $df_{L-1} = G_{L-1} - G_L$ are the degrees of freedom for level $L-1$.

The contribution of the variance between subgroups $(L-1)$ to the variance between groups (L) is then dependent on the proportion of between subgroup comparisons in L not within groups. The total proportion of within subgroup comparisons in L is α_{L-1}, obtained by substituting groups with subgroups in equation 2.16. The proportion of between subgroup comparisons within groups is given by the numerator of equation 2.18 when scaled to the total sample size N_L. The proportion of within subgroup comparisons including between subgroup comparisons within groups is then:

$$\alpha'_{L-1} = \alpha_{L-1} + \sum_{i=1}^{G} \left[\frac{N_i}{N_L}(1 - \alpha_i) \right] \tag{2.19}$$

Hence the sample size for the expected variance between groups in L attributable to variance between subgroups (groups in $L-1$) is:

$$n'_{L \setminus L-1} = \frac{N_L(1 - \alpha'_{L-1})}{df_L} \tag{2.20}$$

For the data used in this study, populations represent level $L-1$, and species level L; With regards to Excoffier et al. [54], following equations correspond to equations 9a–c (n, n', n"); and following degrees of freedom correspond to the level degrees of freedom:

$$\begin{aligned} n_{L-1} &= n & df_{L-1} &= df_{(ap)} \\ n'_{L \setminus L-1} &= n' & & \\ n_L &= n'' & df_L &= df_{(as)} \end{aligned}$$

Although it has to be noted, that equation 9a in Excoffier et al. is not completely correct, as the degrees of freedom used there are the total number of subgroups and not $subgroups - groups$. This is probably just a typing error, as Sandrine Pavoine uses the data from this publication as example for her function `amova`, obtaining exactly the same results as in the paper; despite using the correct degrees of freedom. For the reformulations of equations 2.17, 2.18, and 2.20, see section E.1, page 217.

The variance components (σ^2) corresponding to different hierarchical levels of the data

- σ_a^2 variation among species
- σ_b^2 variation among populations (within species)
- σ_c^2 variation within populations

are then calculated by:

$$\sigma_c^2 = MSD_{(wp)T} \tag{2.21a}$$

$$\sigma_b^2 = \frac{MSD_{(ap)T} - \sigma_c^2}{n} \tag{2.21b}$$

$$\sigma_a^2 = \frac{MSD_{(as)T} - n'\,\sigma_b^2 - \sigma_c^2}{n''} \tag{2.21c}$$

$$\sigma_T^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 \tag{2.21d}$$

Percentages of variation attributable to different hierarchical levels are then obtained by dividing the different sigmas by the total:

$(\sigma_a^2/\sigma_T^2) \times 100$ % variation among species

$(\sigma_b^2/\sigma_T^2) \times 100$ % variation among populations within species

$(\sigma_c^2/\sigma_T^2) \times 100$ % variation within populations

With the obtained variance components so-called Φ-statistics can be calculated, which are called here slightly different from Excoffier et al. [54] to reflect the changed structure of hierarchical levels. They are renamed, so that the subscript T represents species, which normally stands for the whole dataset, and the whole dataset will be marked by the subscript G, with regard to Cockerham's [38] $\bar{\Theta}_g$. The meaning of the different Φ-statistics is then:

Φ_{SG} correlation of randomly drawn individuals within populations with regard to the whole dataset

Φ_{ST} correlation of randomly drawn individuals within populations relative to randomly drawn individuals within species

Φ_{TG} correlation of randomly drawn individuals within species with regard to the whole dataset

and they are calculated as follows:

$$\Phi_{SG} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2}, \quad \Phi_{ST} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}, \quad \Phi_{TG} = \frac{\sigma_a^2}{\sigma_T^2} \tag{2.22}$$

In this setup of the AMOVA these statistics give only an overview, and the usefulness of Φ_{SG} is rather doubtful; They have been mentioned here for

completeness, and can be seen as averages of the pairwise statistics which will be discussed now.

Additionally calculated AMOVA statistics

The AMOVA statistics so far are all taken from Excoffier et al.[54]. All these are overall measures of the data, and average over variance within hierarchical levels. To get a better insight into the structuring of the groups, further statistics where calculated.
Inspired by the second part of the GenAlEx tutorial [131, 132], pairwise Φ_{PT} measures where calculated for all pairs of populations; This gives a more detailed picture of population differentiation then Φ_{SG}. Furthermore Φ_{ST} values were obtained for species separately, which only adds information for *R. aganniphum* in this dataset; Because for all the other species only two populations were available, so that in these cases Φ_{ST} is not different from the pairwise Φ_{PT}. To evaluate differentiation at the species level, species were analysed in pairs, to calculate Φ_{TG}. In contrast to Φ_{PT} this measure accounts for substructure within groups (species), and values obtained are in excess of population differentiation.

For the pairwise population Φ_{PT} the SSDs within populations have already been calculated with equation 2.11, page 73, and only the total SSDs of the combined subsets of each pair of populations have to be obtained:

$$SSD_{(total)ij} = \frac{1}{2N_{pij}} \sum_{k=1}^{N_{pij}} \sum_{l=1}^{N_{pij}} \delta_{kl} \tag{2.23}$$

where $N_{pij} = N_{pi} + N_{pj}$ is the sum of individuals in populations i and j, and δ_{kl} is the squared distance between individual k and individual l of the pooled subsets. The SSDs among each population pair are then calculated with:

$$SSD_{(ap)ij} = SSD_{(total)ij} - \left(SSD_{(wp)si} + SSD_{(wp)sj}\right) \tag{2.24}$$

The according degrees of freedom within populations ($df_{(wp)}$) have already been calculated with equation 2.14, page 74, and for the pairwise comparisons the df among populations can be omitted, as

$$df_{(ap)ij} = (\text{number of populations} - 1) = 1$$

The MSDs are then as follows:

$$MSD_{(ap)ij} = SSD_{(ap)ij} \tag{2.25a}$$

$$MSD_{(wp)ij} = \frac{SSD_{((wp)i} + SSD_{((wp)j}}{df_{(wp)i} + df_{(wp)j}} \tag{2.25b}$$

and with these the variance components are calculated:

$$\sigma^2_{B,\,ij} = \frac{MSD_{(ap)ij} - MSD_{(wp)ij}}{n} \tag{2.26a}$$

$$\sigma^2_{C,\,ij} = MSD_{(wp)ij} \tag{2.26b}$$

where n is n_L (equation 2.17, page 75), with $N_L = N_i + N_j$ and $df_L = 1$.
The pairwise Φ_{PT} of population i and population j is then:

$$\Phi_{PT\,ij} = \frac{\sigma^2_{B,\,ij}}{\sigma^2_{B,\,ij} + \sigma^2_{C,\,ij}} \tag{2.27}$$

which can be interpreted as excess similarity among randomly chosen individuals within the same population relative to randomly chosen individuals from the whole subset [81].

The differentiation of populations within species (Φ_{ST}) is equivalent to the original Φ_{ST} from Excoffier et al. [54]. As mentioned in the introduction (subsection 2.3.5), the AMOVA was developed for the analysis of different hierarchical levels within a single species, considering regional differentiation as one of these levels. This level was replaced by the higher hierarchical level of species identity in this study, so that the Φ_{ST} calculated here corresponds to a classical AMOVA without considering regional differentiation. Conveniently the SSDs for this statistic have already been calculated with equations 2.10 and 2.11, page 73, and the according degrees of freedom with equation 2.14, page 74. Hence we can follow up with calculating the MSDs within and among populations, within single species:

$$MSD_{(wp)si} = \frac{\sum_{j=1}^{P_{si}} SSD_{(wp)j}}{\sum_{j=1}^{P_{si}} df_{(wp)j}} \tag{2.28a}$$

$$MSD_{(ap)si} = \frac{SSD_{(ws)i} - \sum_{j=1}^{P_{si}} SSD_{(wp)j}}{df_{(ap)i}} \tag{2.28b}$$

where P_{si} is the number of populations in species i, $MSD_{(wp)si}$ are the mean squared deviations within populations in species i, and $MSD_{(ap)si}$ are the mean squared deviations among populations within species i.
The variance components are then calculated as follows:

$$\sigma^2_{B,\,si} = \frac{MSD_{(ap)si} - MSD_{(wp)si}}{n} \tag{2.29a}$$

$$\sigma^2_{C,\,si} = MSD_{(wp)si} \tag{2.29b}$$

where n is n_L (equation 2.17, page 75), with $N_L = N_{si}$ and $df_L = df_{(ap)i}$.
The Φ_{ST} for populations within species i is then

$$\Phi_{ST\,si} = \frac{\sigma^2_{B,\,si}}{\sigma^2_{B,\,si} + \sigma^2_{C,\,si}} \tag{2.30}$$

which can be interpreted as excess similarity among randomly chosen individuals within the same population relative to the whole species [81].

The pairwise differentiation of species (Φ_{TG}) was introduced as a measure of distance between species. Because this measure is taking into account the differentiation already expected between populations, the statistic measures the distance between species in excess to population differentiation. The statistic is basically calculated with a normal AMOVA, only including one pair of species. Therefore only the total SSD has to be calculated additionally:

$$SSD_{(total)ij} = \frac{1}{2N_{sij}} \sum_{k=1}^{N_{sij}} \sum_{l=1}^{N_{sij}} \delta_{kl} \tag{2.31}$$

where $N_{sij} = N_{si} + N_{sj}$ is the sum of individuals in species i and j, and δ_{kl} is

the squared distance between individual k and individual l of the pooled subsets. With the according SSDs from equations 2.11 and 2.12, and degrees of freedom from 2.14, where $df_{(as)ij} = 1$, the MSDs are calculated as follows:

$$MSD_{(as)ij} = SSD_{(total)ij} - (SSD_{(ws)i} + SSD_{(ws)j}) \tag{2.32a}$$

$$MSD_{(ap)sij} = \frac{SSD_{(ap)si} + SSD_{(ap)sj}}{df_{(ap)si} + df_{(ap)sj}} \tag{2.32b}$$

$$MSD_{(wp)sij} = \frac{\sum_{k=1}^{P_{sij}} SSD_{(wp)k}}{\sum_{k=1}^{P_{sij}} df_{(wp)k}} \tag{2.32c}$$

where P_{sij} is the number of populations in species i and j together, $MSD_{(ap)sij}$ are the mean squared deviations among populations within species i and j, and $MSD_{(wp)sij}$ are the mean squared deviations within populations of species i and j. The variance components are then obtained according to equation 2.21, page 76:

$$\sigma^2_{C,\,sij} = MSD_{(wp)sij} \tag{2.33a}$$

$$\sigma^2_{B,\,sij} = \frac{MSD_{(ap)sij} - \sigma^2_{C,\,sij}}{n} \tag{2.33b}$$

$$\sigma^2_{A,\,sij} = \frac{MSD_{(as)ij} - n'\,\sigma^2_{B,\,sij} - \sigma^2_{C,\,sij}}{n''} \tag{2.33c}$$

$$\sigma^2_{T,\,sij} = \sigma^2_{A,\,sij} + \sigma^2_{B,\,sij} + \sigma^2_{C,\,sij} \tag{2.33d}$$

Where $n = n_{L-1}$ (equation 2.18, page 75), with $G_L = 2$ and $df_{L-1} = df_{(ap)si} + df_{(ap)sj}$; $n' = n'_{L\setminus L-1}$ (equation 2.20, page 76), and $n'' = n_L$ (equation 2.17, page 75), with $N_L = N_{sij}$ and $df_L = 1$.

To obtain the statistic Φ_{TG} the variance between species is divided by the total variance for the species pair:

$$\Phi_{TG,\,ij} = \frac{\sigma^2_{A,\,sij}}{\sigma^2_{T,\,sij}} \tag{2.34}$$

In accordance with the other Φ-statistics, this measure can be seen as excess similarity among randomly chosen individuals within the same species relative to the whole dataset of the species pair.

Testing the significance of the measures

Significance of the variance components and the Φ-statistics was tested as suggested by Excoffier et al. [54]. Null-distributions for the values were obtained by randomly permuting individuals or groups of the corresponding hierarchical level and performing an AMOVA on the randomised dataset. The null-distributions of the randomised datasets were then used to test the null-hypothesis that the observed values could be obtained with no structuring.

To generate null-distributions of pairwise Φ_{PT} and the corresponding variance components, individuals were randomly permuted within population pairs. For Φ_{ST} individuals were permuted within species, and to assess significance of Φ_{TG} whole populations were permuted within species pairs; 5000 permutations were used in all cases. When permuting individuals the sample size is large enough to yield sensible null-distributions, but in the case of pairwise species comparisons, mostly, only four populations are available for permutation. Therefore significance for Φ_{TG} is highly dependent on how often the populations belonging to one species are grouped together by chance. For the permutations to be meaningful more populations for each species would be required; Hence significance was assessed by eye, using the histograms of values obtained by randomisation of the datasets. If no more than one value of the randomisations (obtained by recreating the actual grouping by chance), was of the magnitude of the observed value, the observed value was assumed to be different from the null-expectation (also see chapter 3, section 3.4, page 112). Although this does not test significance, it ascertains that no other grouping in the dataset would yield a differentiation as high or higher.

Because these permutations only assess the significance of the deviation from the complete random case, which does not help when comparisons between the values of different groups are desired, 95 % confidence intervals for the obtained values were computed additionally. These were obtained by repeatedly drawing loci with replacement from the dataset, while keeping the number of loci constant (bootstrapping), and calculating an AMOVA on the bootstrapped dataset. The intervals were then calculated from 5000 bootstrap replicates using the extreme values of the probability interval 0.025 - 0.975; obtained with the function `quantile`, setting `probs = c(0.025, 0.975)`, in R.

2.3.6 Statistics using allele frequency estimates

Circumventing the dominant nature of AFLP markers by estimating allele frequencies within populations opens up the possibility to use population genetic theory initially developed for codominant markers. This theory is generally better developed, and enables the use of null hypotheses based on neutral assumptions. Although estimating the allele frequencies is not completely equivalent, with regards to accuracy, to obtaining the data directly, it enables the comparison of results with other studies using different types of markers.

To calculate frequency based measures of differentiation and variation within and between populations, the program AFLP-SURV 1.0 [174] was used. This program uses the method of Zhivotovsky (see subsection 2.3.3) to estimate the allele frequencies from the data. With these estimates different frequency based diversity and subdivision statistics, discussed in Lynch and Milligan (1994) [109], are calculated. The measures used in this study are the gene diversity within populations (H_j), and pairwise F_{ST} values. The following equations, and descriptions thereof, are a summary of the original publication [109].

Gene diversity within populations H_j can be understood as the probability that two alleles randomly drawn from population j differ at the ith locus [109]. This is equivalent to the expected heterozygosity under Hardy-Weinberg equilibrium ($2pq$, see Table 2.5, page 68), and, with correction for sampling error, is calculated as follows (equation 4a in [109]):

$$\hat{H}_{j,i} = 2\hat{q}_{j,i}\,(1 - \hat{q}_{j,i}) + 2\,s^2_{\hat{q}_{j,i}} \tag{2.35}$$

where $\hat{q}$ is the estimated null allele frequency (see equation 2.4, page 69), and $s^2_{\hat{q}}$ its variance. The mean gene diversity in population j averaged over all L loci is then (equation 5 in [109]):

$$\hat{H}_j = \frac{1}{L}\sum_{i=1}^{L} \hat{H}_{j,i} \tag{2.36}$$

Pairwise F_{ST} From the estimated allele frequencies within populations ($\hat{q}$) the probability that an allele at locus i drawn from population j differs from an allele drawn from population k, can be estimated as follows (equation 9a in [109]):

$$\hat{H}'_j k, i = \hat{q}_{j,i} + \hat{q}_{k,i} - 2\,\hat{q}_{j,i}\hat{q}_{k,i} \tag{2.37}$$

Because in the absence of population subdivision $\hat{H}'_{jk,i} = H_{j,i} = H_{k,i}$, a better measure is the heterozygosity between populations in excess of the heterozygosity within populations (equation 10a in [109]):

$$\hat{H}_{jk,i} = \hat{H}'_{jk,i} - \frac{\hat{H}_{j,i} + \hat{H}_{k,i}}{2} \tag{2.38}$$

The mean gene diversity between population j and population k is then defined as

$$\hat{H}_{jk} = \frac{1}{L}\sum_{i=1}^{L} \hat{H}_{jk,i} \tag{2.39}$$

and the F_{ST} between the populations

$$F_{ST} = \frac{\hat{H}_{jk}}{H_T} \tag{2.40}$$

where H_T, the total gene diversity is the sum of the average gene diversity within (equation 2.36), and between populations:

$$H_T = \left(\frac{\hat{H}_j + \hat{H}_k}{2}\right) + \hat{H}_{jk} \tag{2.41}$$

2.3.7 Bayesian inference of population structure

Methods described above in this chapter (MDS and NJ, subsection 2.3.4) use genetic distances between individuals, without assuming predefined structure. Although they can help identify clusters of individuals, they are only loosely connected to statistical procedures allowing the identification of homogeneous clusters of individuals [52]. The program STRUCTURE [135] implements a model-based clustering method to infer population structure and to assign individuals to K clusters. The clusters are defined by allele frequencies over all loci, assuming Hardy-Weinberg equilibrium within clusters, so that individuals will be grouped to minimise the violation of the equilibrium assumption. This allows the identification of groups of individuals between which geneflow is frequent, and which can be assumed to belong to the same population. Furthermore, individuals can be part of several clusters, allowing the occurrence and identification of

admixed individuals. The model does not assume a particular mutation process and from version 2.2 onwards, the program supports dominant data [55]. Version 2.3.1 was used in this study to investigate the identity of individuals, populations and possible admixture.

Model parameters The model implemented in STRUCTURE is based on several parameters, some of which have to be set initially, and others are either set by the user, or estimated during the simulation [136]. The options regarding these parameters are discussed in detail in the manual that can be obtained from the same web page as the program[3]. Only parameters and settings relevant to this study will be discussed here.

The model can allow individuals to be either admixed (part of several clusters), or a member of one cluster only. Because this study explicitly deals with hybridisation, only the admixture option was used. The parameter λ specifies the distribution from which allele frequencies within clusters are drawn. By default its value is set to one, which, according to the manual, should fit most data. However, if allele frequencies at many loci are skewed towards rarer variants, a lower value, which can be estimated from the data, might be more appropriate. Because many loci had markers at low frequency, duplicate simulations were performed using either $\lambda = 1$ or using λ estimated from the data. To estimate λ the whole dataset without hybrids was used, and ten independent simulations performed, with the number of clusters (K) set to one. The average λ, rounded up, obtained this way was then used for further runs. This approach was chosen because the STRUCTURE manual mentions that estimating multiple parameters from the data can lead to problems, and suggested the method with $K = 1$ (p. 12 in [136]). Allele frequencies within clusters can be either assumed to be independent from each other, or correlated and deviating by a F_{ST} analogous factor from an ancestral frequency common to all populations. The manual [136] suggests that correlated allele frequencies can improve the detection power of clusters in closely related species. Correlated allele frequencies (due to migration) can be expected in populations of the same species, but between species this assumption might not be appropriate. However, the differentiation of the species is very small, and in the case of hybridisation geneflow might occur over species barriers. Due to these considerations different hierarchical levels of the data can fall in either category, and both settings were tried.

To choose an appropriate length for the burn-in and run length of the Markov

[3] `http://pritch.bsd.uchicago.edu/structure.html`

chain, several runs where performed with the whole dataset. The behaviour of the admixture parameter α and the log probability [$\Pr(X|K)$] was monitored to estimate when equilibrium was reached. A chain length of 30,000 iterations with an additional burn-in of 20,000 was found to be sufficient and was used for all further simulations in this study.

Determining the number of clusters K To properly interpret the results of STRUCTURE, it is important to determine the right number of clusters (K) and hence the true number of homogeneous groups in the data. Pritchard et al. [135, 136] suggest to directly use the log probability of the data $\Pr(X|K)$, but Evanno et al. [52] proposed a more reliable method using a derivative measure termed ΔK. This measure was used to determine the most likely value of K and was calculated as follows (equations from [52]):

$$L'(K) = L(K) - L(K-1) \tag{2.42a}$$

$$|L''(K)| = |L'(K+1) - L(K)| \tag{2.42b}$$

$$\Delta K = \frac{|L''(K)|}{s[L(K)]} \tag{2.42c}$$

where $L(K)$ is the mean of $\Pr(X|K)$ for several independent simulations, and $s[L(K)]$ is its standard deviation. To obtain variances for the L measures, the mean was calculated from pairwise differences between successive runs, with the number of runs (R) being equal for each K.
$R(K=1) = R(K=2) = \ldots = R(K=n)$, e.g. for $L'(K)$:

$$L'(K) = \frac{1}{R^2} \sum_{\substack{i=1 \\ j>i}}^{R} [Pr(X|K)_i - Pr(X|K-1)_j] \tag{2.43a}$$

$$Var[L'(K)] = \frac{1}{R(R-1)} \sum_{\substack{i=1 \\ j>i}}^{R} \{[Pr(X|K)_i - Pr(X|K-1)_j] - L'(K)\}^2 \tag{2.43b}$$

The calculated means and variances were then plotted for every K and used to select the appropriate K for the dataset. Evanno et al. [52] pointed out that STRUCTURE always detects the topmost structure in the data. They used data generated from a hierarchical island model, consisting of five groups of populations, with each group containing four subpopulations. In their simulations STRUCTURE detected five clusters as opposed to 20; separate analysis of the

groups revealed the four subpopulations. The same was observed for the *Rhododendron* data, and the substructure in each species was analysed separately with a subset of the data.

2.3.8 Simulating hybrids

To investigate how well STRUCTURE would identify admixture in hybrids, and to obtain expectations of allele frequencies in hybrid offspring of the sympatric parental populations, several hybrid classes were simulated.

Expected allele frequencies in hybrids To simulate hybrids formed by interspecific random mating of the parents (RM-hybrids), the loci were assumed to be in linkage equilibrium, and because the populations within single species were considerably differentiated, only the sympatric parental populations were used to estimate allele frequencies. Using the method from Zhivotovsky described in subsection 2.3.3, allele frequencies were estimated and hybrid genotypes obtained by first generating an equal number of "gametes" for each parental population, and then combining them into "zygotes". Gametes were simulated by drawing alleles (0 or 1) for each locus from a binomial distribution with the probability of obtaining a '1' set to the estimated allele frequency p. These gametes were then combined into zygotes by taking one gamete from each parent. For later generation backcrosses the codominant state was used to calculate the allele frequencies in the simulated F1 (or BC1) population, and generate gametes as described before to combine them with newly generated gametes from the respective backcrossing parent. These codominant genotypes were then transformed into dominant phenotypes by setting each locus count[4] over one to one (see Table 2.6).

Table 2.6: Combining gametes into zygotes

	Locus									
gamete parent a	0	1	1	0	0	0	0	1	0	...
gamete parent b	1	0	1	1	0	1	0	1	1	...
zygote $(a + b)$	1	1	2	1	0	1	0	2	1	...
dominant phenotype	1	1	1	1	0	1	0	1	1	...

[4]possible counts are 0 = 0|0, 1 = 1|0 & 0|1, and 2 = 1|1

Bottleneck during hybridisation In the course of the data analysis, simulation results from STRUCTURE hinted towards a deviation of the hybrid population allele frequencies from genome wide expectations for *R. aganniphum* × *phaeochrysum*. The best explanation for this seemed a violation of the assumption of random mating, or in this case the formation of hybrids by certain parental individuals only. Hybrid populations with marker frequencies deviating from expectations, implied by assuming random mating, were simulated to investigate the cluster assignment of hybrid individuals by STRUCTURE that present a higher degree of linkage disequilibrium than expected. To this end hybrid populations were simulated in which only a small number of individuals from one of the parental populations contributed to the allele frequency pool. In other words, for one parent no restriction to hybridisation was assumed, while for the second parent a few individuals were chosen randomly to contribute to the hybrid genepool, hence assuming a bottleneck for the transmission of alleles. This approach was adopted because the gene diversity found in the hybrids was not lower than in the species[5], providing evidence against a bottleneck from both parental sides. Furthermore, asymmetric barriers to geneflow have been reported before [10, 41, 185], suggesting that unequal contribution represents a valid biological hypothesis. The bottleneck was assumed to take effect during the formation of first generation hybrids (F1s) only, and *R. aganniphum* was assumed to contribute without restrictions, as the data suggested more shared alleles between the hybrids and this species[6].

To simulate the hybrids, allele frequencies were estimated as mentioned above, and the gametes for the unrestricted parent were obtained as described before. For the generation of the "bottlenecked" gametes, between one and four individuals were chosen randomly from the parental population and allele frequencies estimated only from these. Gametes were then obtained as normal, and the zygotes and dominant phenotypes simulated as in the case assuming random mating.

2.3.9 Quantifying the deviation of allele frequencies

Analyses pointed to the deviation of several loci from the assumption of random transmission of alleles to the hybrid offspring in *R. aganniphum* × *phaeochrysum*. Because of this, simulations of expected allele frequencies in several hybrid

[5] Gene diversity values obtained for the populations are reported in chapter 3, Table 3.10, page 117.

[6] Much lower divergence was observed for TG-TGP opposed to TP-TGP, see chapter 3, Table 3.8 (page 115), and Table 3.9 (page 116).

classes were used to quantify this deviation and test for significance. Under the assumption of neutrality of all loci, allele frequencies, and hence marker counts in hybrids, should follow the principles of a binomial distribution. In the case of dominant markers possible alleles at loci are 0 and 1. The alleles found in gametes of one parental population should follow a binomial distribution with the probability of obtaining allele 1 equal to the population frequency of this allele. The expectation of allele counts in hybrid offspring can therefore be approximated by using the RM-hybrid simulation method, described in subsection 2.3.8, and averaging counts over multiple simulated populations. To reduce the number of assumptions a flat prior was used for the allele frequency estimation, using the method of Zhivotovsky (see subsection 2.3.3).

Simulating expected allele counts 10,000 datasets of expected marker counts were simulated per locus and compared to the marker count in the real hybrids. To allow direct comparison between simulated and observed counts, the number of simulated individuals for each locus was chosen to be the same as in the real hybrid population. The p-values for significant deviation from the expectation were then calculated as follows: The number of simulations in which a similar or more extreme count than observed was obtained were counted; this number was then expressed as a fraction of the total number of simulations. If the number of observed markers was higher than in both of the parents the upper tail of the simulated distribution was used, and values counted that were equal to the observed counts or higher. If the count in the hybrids was lower than in one of the parents, the lower tail was used and values counted that were equal to the observed count or lower. Because the class of the hybrids was not known, expected allele frequencies for first generation hybrids (F1s) as well as for backcrosses (BCs) were simulated, and the deviation from the observed count calculated. The p-values were adjusted for multiple testing using the method of Benjamini and Hochberg [17], as implemented in the function `p.adjust` in `R` with `(method = "BH")`. To increase stringency, for each locus all three[7] *p*-values were used to determine significance, as follows. The distortion of a locus in the hybrid population was only considered significant, if the deviation from the expectation was significant in all three cases. This method should be conservative even in the case of unknown hybrid class, as the distortion has to be significant for all classes, and should the real hybrid population be a mixture of F1s and BCs, counts would be expected to be intermediate and not more extreme. As a

[7] F1, backcross to *R. aganniphum*, and backcross to *R. phaeochrysum*

measure of transmission distortion into the hybrids (Δ_{obs}), the difference of the observed count to the median of the simulated values, normalised by the number of individuals n, was used. The median was chosen over the mean as it is less influenced by outliers.

$$\Delta_{obs} = \frac{\text{observed count} - \mathit{median}(\text{simulated count})}{n}$$

This measure can theoretically take values from -1 to 1; generally negative values mean that the dominant marker is under-represented in the hybrid, and positive values that it is over-represented.

As a distortion in the marker frequencies in the hybrids makes a classification of individual hybrid class impossible, an other measure was introduced to determine the most likely overall composition of the two *R. aganniphum* × *phaeochrysum* hybrid populations (GH2a, GH2b; see subsection 1.6.3, page 41). To this end the sum of the absolute differences of the observed values from the expectation for each of the simulated classes c over all loci L was calculated for each hybrid population separately:

$$\Delta'_c = \sum_{l=1}^{L} |\Delta_{obs,\,c,\,l}| \tag{2.44}$$

where c is one of the simulated classes (F1, BC1, BC2), and

$\Delta_{obs,\,c,\,l}$ is the observed difference at locus l for class c.

To compare these values the null assumption of F1s was used as reference point for each population, and the change in distortion calculated for all classes as follows:

$$\Delta_c = \frac{\Delta'_{F1} - \Delta'_c}{\Delta'_{F1}} \tag{2.45}$$

By definition $\Delta_{F1} = 0$; a positive value for Δ_c then indicates that class c represents a better fit for the population (less deviation from the expectation), and a negative value a worse fit.

Diagnostic value of markers To employ assignment tests that group putative hybrid individuals into different hybrid classes (F1, BC1, ...) markers should conform to neutral expectations. Neutral assumptions can be expected to be never fully met in natural populations, however, violation of this prerequisite should be more severe if especially diagnostic markers do not conform to the made

assumptions. The relative diagnostic value of a marker for assignment of hybrid individuals can be expected to increase with the differentiation of the parental populations at the locus in question; being fully diagnostic if it represents a fixed difference. On the other hand, the relative diagnostic value will be negatively affected the more the marker deviates from the assumptions made, hence the more it is subjected to a distortion in the hybrid offspring. To identify markers that are potentially diagnostic and do not experience distortion, the divergence of the parents at a given locus, measured as locus F_{ST}, was compared to the absolute distortion assuming F1s ($|\Delta_{obs,\,F1}|$). The F_{ST} was calculated from the estimated parental allele frequencies according to Hartl and Clark ([74], page 126, equation 4.6) based on the Wahlund's Principle (also see equation 8 in Weir and Cockerham [177]):

$$F_{ST} = \frac{\sigma^2}{\bar{p}\,\bar{q}} \tag{2.46}$$

where

$$\sigma^2 = \frac{(q_1 - q_2)^2}{4} \tag{2.47a}$$

$$\bar{p}\,\bar{q} = \frac{(q_1 + q_2)}{2} \times \left(1 - \frac{(q_1 + q_2)}{2}\right) \tag{2.47b}$$

Possible mechanisms for marker distortion Mechanisms leading to a deviation of marker frequencies from their respective expectations in hybrid offspring are likely to also be contributing to reproductive isolation between the parental taxa. It seems therefore appropriate to divide them into two major groups taking effect either at a prezygotic or postzygotic stage. An easily appreciated prezygotic barrier in plants is assortative mating mediated by pollinator discrimination or discrepancy in flowering times. As already mentioned in subsection 2.3.3, pollinator discrimination seems unlikely because of the similar floral morphology, and at least in 2008, *R. aganniphum* and *R. phaeochrysum* were flowering at exactly the same time. However, if in years of hybrid establishment the flowering periods overlapped only slightly, this could have resulted in few individuals from one parental species contributing, leading to a bottleneck effect. Post-pollination prezygotic effects that could affect marker transmission

include pollen competition and pollen-pistil interactions [10, 147]. Possible postzygotic interactions include endosperm-embryo interactions [172]. Incompatibilities occurring in the last three mentioned interaction types are mostly believed to result from Bateson-Dobzhansky-Muller incompatibilities (BDMI) [172]. Negative epistatic interactions between parental genes could therefore result in low fertilisation success or early embryo abortion for certain alleles and change the realised transmission into F1 hybrids. Furthermore, due to the genetic nature of these interactions [gametophytic(1n)-sporophytic(2n), and sporophytic(2n)-endosperm(3n)], the resulting incompatibilities are expected to be asymmetric [107, 172]. Cytonuclear interactions (interplay of nuclear genome and cytoplasmic organelles) [21] present other types of incompatibilities that might lead to a preferential transmission of certain alleles into hybrid offspring. Additionally, in later stages of the development, loci involved in adaptations or other BDMI can lead to intrinsic or ecological hybrid inviability [144], therefore removing certain alleles predominantly from the hybrid population. Chromosomal rearrangements are generally not considered to contribute majorly to F1 hybrid inviability, but are rather known to cause hybrid sterility [57]. Therefore they should not be expected to affect marker distortion in the F1 generation significantly. However, mechanisms have been suggested that lead to an enrichment of BDMIs and adaptive genes in chromosomal inversions (CI) [79], and evidence for this has been presented e.g. in *Mimulus* [108]. This could explain the occurrence of underdominance (homozygote advantage) observed for some CIs even during F1 hybrid formation [93]. To illustrate the effect of complete underdominance on the relationship of expected marker distortion in F1 hybrids, with regards to the parental F_{ST} at a given locus, we can compare it to the theoretical expectations under neutrality.

The expected genotype frequencies in F1 hybrid offspring, assuming interspecific random mating of the two parental populations and neutrality of alleles are as follows:

$$\begin{aligned} 1|1 &= p_1 \times p_2 \\ 1|0 &= p_1\, q_2 + q_1\, p_2 \\ 0|0 &= q_1 \times q_2 \end{aligned} \tag{2.48}$$

where p_1 and q_1 are the frequencies of the dominant and the recessive allele, respectively, in parental population 1, and p_2 and q_2 the respective frequencies in parental population 2; while for both $p_1 + q_1 = 1$, and $p_2 + q_2 = 1$.

At a locus experiencing complete underdominance the heterozygotes will be inviable and are therefore excluded; the expected frequencies of realised genotypes in the F1 offspring are then:

$$\begin{aligned} 1|1 &= p_1 \times p_2 \\ 1|0 &= missing \\ 0|0 &= q_1 \times q_2 \end{aligned} \tag{2.49}$$

The difference (Δ) between these two expectations is given by the fraction of heterozygotes, and because of $p_i + q_i = 1$, $p_i = 1 - q_i$[8] we can write:

$$\Delta = p_1\, q_2 + q_1\, p_2 \tag{2.50a}$$
$$= (1 - q_1)\, q_2 + q_1\, (1 - q_2) \tag{2.50b}$$
$$= q_1 + q_2 - 2q_1q_2 \tag{2.50c}$$

For a locus experiencing underdominance Δ can be regarded as expectation(s) of $|\Delta_{obs}|$ for a given value of F_{ST}. Although it does not reflect the underlying process adequately, Δ tends to be positively correlated with F_{ST} when the locus experiences underdominance (Figure 2.1).

In closely related species, markers linked to loci that are involved in adaptation due to directional selection, or some form of incompatibilities, as described above, might be expected to be more diverged [182]. However, not all diverged markers have to be linked to loci causing incompatibilities. If the observed bottleneck is mostly the result of few loci, or incompatibilities not linked to sampled markers, the effect should present itself in a mostly genome-wide manner, not affecting the loci systematically. If on the other hand many sampled markers are linked to loci producing the distortion, or most potentially diagnostic markers are, this should lead to a positive correlation of $|\Delta_{obs}|$ with the parental locus F_{ST}.

To test for a positive correlation of the two measures, Pearson's product-moment correlation coefficient (r) and its significance were calculated in R with the function `cor.test` setting `(method = "pearson", alternative = "greater")`.

Because of uncertainty that a higher differentiation of the parental frequencies might make it easier to detect distortion, leading to an artifact correlation of F_{ST} and $|\Delta_{obs,\,F1}|$, the random bottleneck case was simulated. To this end two random

[8]The recessive allele was chosen to represent the case, as experimentally only here homozygotes can be identified. In this theoretical example, however, the behaviour of the recessive and dominant allele are completely symmetrical as all heterozygotes are removed.

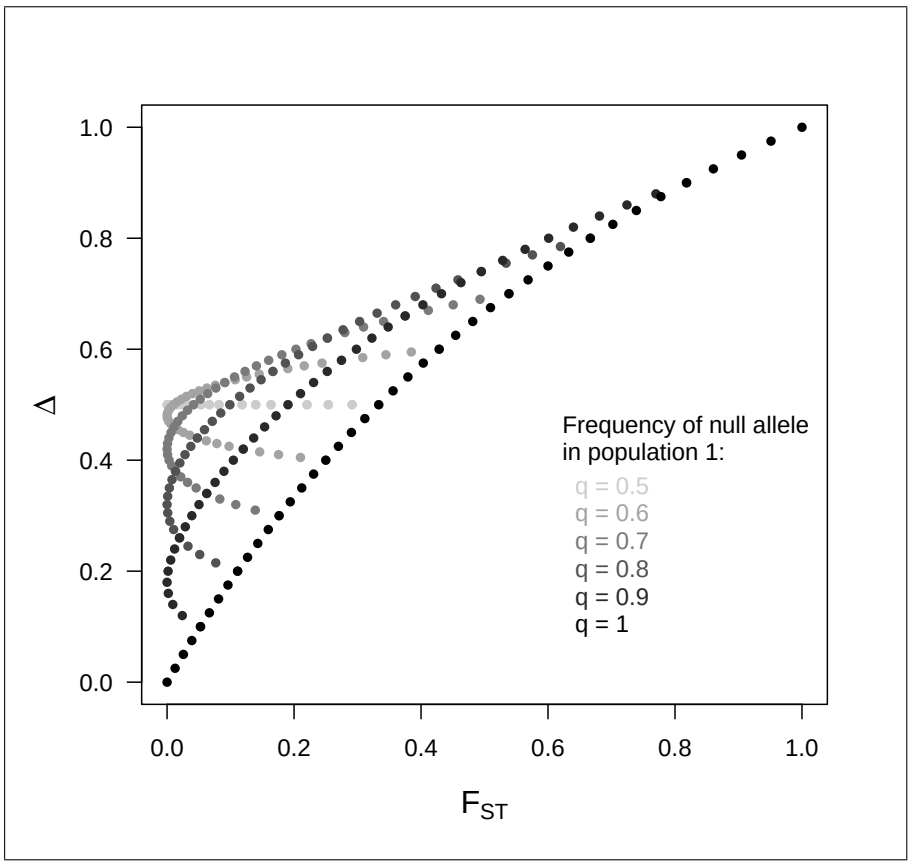

Figure 2.1: Δ and F_{ST} with underdominance. Theoretically expected deviation of the allele frequency at underdominant loci from the expectation of interspecific random mating compliant with Mendelian segregation (Δ), plotted against the parental locus F_{ST}. Values obtained by setting the allele frequency in population 1 to the given q and varying the frequency in the second population q_2 between 0 and 1 are shown. Values for $q_1 < 0.5$ are not shown as F_{ST}, and hence the relationship, is symmetrical with regards to $q_1 = 0.5$.

parental datasets were generated, comprising 40 individuals and 150 loci each. The allele counts for each locus were generated separately for each parent by choosing a random number between zero and one, and setting this number as the allele frequency from which 80 alleles (0 & 1) were drawn. These were then combined into 40 dominant phenotypes. As the loci frequencies were generated independently for each parental dataset, locus F_{ST} values between zero and one were obtained for the datasets, without having to specify it explicitly. From these parental datasets 40 RM-hybrids and 40 bottlenecked hybrids were simulated as described in subsection 2.3.8, resulting in a simulated dataset comprising two parental populations and two hybrid populations. 25 of these datasets were simulated, and the same analysis as above was performed with each to test if a general bottleneck would lead to a correlation between F_{ST} and $|\Delta_{obs,\,F1}|$.

2.3.10 Leaf waxes

Normalisation of area In GC analysis the area described by a peak, corresponding to a certain retention time, is proportional to the quantity of the substance eluted from the column. This quantity is dependend on the applied amount of sample, and may therefore vary for fractions between different traces only due to sample quantity difference. To be able to compare the profiles it is necessary to normalise them. This is easily achieved by scaling all fractions in a trace to the fraction with the largest observed area and using percentages:

$$x\% = 100 \times \frac{Area\,(fraction\,x)}{Area\,(largest\,fraction)}$$

The largest fraction will then always have 100 %, and the other fractions can be in the range of 0–100 %.

Displaying leaf wax data in combination with MDS To relate the leaf wax composition of individuals to genetic distance, coordinates for the individuals, obtained with MDS (subsection 2.3.4, page 70) based on the AFLP data were used. The relative contribution of the wax fractions to the overall leaf wax composition for single individuals was depicted using pie charts, plotted at the coordinates of the respective individual. Because of the non linear perception of humans regarding values that are represented as areas, as opposed to a linear manner [160, 161], the values were squared for this purpose.

Chapter 3

Results

3.1 Data overview

For the 412 accessions genotyped with AFLPs 423 loci were scored; 77 of these fell into category 'C' (bad quality, see chapter 2, section 2.2.3, page 62) and were not further considered, resulting in 346 usable markers.

Scoring error The error estimated for the different categories, was between 0.48–2.65%; this is at the lower range of error generally observed for AFLP profiles (2–5%, [23]). Two factors not easily distinguishable contribute to the error: error due to scoring and difference of profiles attributable to runs. Unless several persons, scoring the same data independently, can be found, the respective influence of the factors can not be determined. Category 'A' had a considerably lower error rate than 'B', which was expected as the groupings were intended to reflect fidelity. The average number of bands for a profile generated for the same individual was fairly consistent, with roughly a maximum of one band standard deviation for 100 loci (Table 3.1).

Missing data Because of missing data ten individuals and nine loci were removed. This was not due to the protocol not working, but because of a malfunctioning of some capillaries of the sequencer and the analysis software used. In some capillaries the traces broke down after reaching fragment sizes of >400 bp, while normally ∼450 bp were achieved; and in some profiles the analysis software failed to call the standard properly, resulting in shifted profiles which were fine, even the standard, but unscorable. Samples were repeated, but due to unreliability of hardware, and time issues, data for some individuals could not be obtained. Additionally the second sympatric population of *R. aganniphum*, G2b

was excluded from further analysis and G2a renamed to G2; The sample size of nine individuals for G2b was too small and furthermore many individuals seemed to be admixed (see Appendix C, Figure C.1, page 182), so that pooling with G2a did not seem reasonable.

Table 3.1: Scoring error estimated error introduced through variance between runs and/or scoring, as calculated from replicated samples. Listed according to category* and including the two datasets finally used: all except skew analysis (Dataset II), and skew analysis (Dataset III); monomorphic loci are included for the category estimates.

Category	Loci[a]	% mismatch[b]	Average bands[c]	SD[d]
A	136	0.48	50.44	0.503
A & AB	244	1.05	73.94	0.995
B	102	2.65	25.98	1.163
A, AB & B	346	1.52	99.92	1.542
Dataset II	127	1.42	22.89	0.820
Dataset III	105	2.42	30.69	0.966

[a] Number of loci falling into the category.
[b] Percent of average observed mismatch (number of mismatches / number of comparisons) between band profiles of the same individual; calculated from 31 individuals run twice.
[c] Average number of bands obtained per profile in the category.
[d] Standard deviation of a profile regarding number of bands; calculated from four individuals, for each performing five to six runs
* chapter 2, subsection 2.2.3, page 62.

3.1.1 Datasets

Three different datasets were used in the analyses, Datasets I–III, differing in the number of loci excluded. Loci were excluded to increase discriminatory power, and speed-up simulation analyses.

Dataset I The initial data, after removal of missing data and population G2b, comprised 393 individuals and 337 scored loci, hereafter referred to as Dataset I. Only two fixed differences between species were detected, both between *R. clementinae* and *R. roxieanum*; the number of polymorphic loci was roughly equal for *R. aganniphum*, *R. phaeochrysum* and *R. roxieanum*, but considerably lower in *R. clementinae* (Table 3.2).

Table 3.2: Datasets and summary of loci Number of polymorphic, monomorphic and singleton loci for the different species and potential hybrids. Overall statistics where calculated without subdivision, hence the number of singletons and monomorphic loci is smaller.

Dataset		Loci			
Species	Individuals	Polymorphic	Singleton	Monomorphic	Total
Dataset I[a]					
TC	73	121	36	180	337
TG	80	180	31	126	337
TGP	65	182	26	129	337
TP	47	164	42	131	337
TX	88	170	42	125	337
TXC	40	128	45	164	337
Overall	393	292	23	22	337
Dataset II[b]					
TC	73	56	14	57	127
TG	80	92	11	24	127
TGP	65	90	8	29	127
TP	46	81	10	36	127
TX	86	78	12	37	127
TXC	40	65	21	41	127
Overall	390	127	0	0	127
Dataset III[c]					
TG	80	87	7	11	105
TGP	65	90	5	10	105
TP	47	82	8	15	105
Overall	192	105	0	0	105

[a] All scored loci with missing data and population G2b removed.
[b] Data used for all analyses, excluding the analysis of skewed markers.
[c] Data used for the analysis of marker skew in *R. aganniphum* × *phaeochrysum*, pruned starting with A, AB & B, and excluding TC, TX and TXC.

Dataset II Many of the polymorphic markers were polymorphic in all species, and, because of low frequency differences, would not yield information about relationships. Furthermore, these loci would increase random noise [80], and therefore it was decided to keep potentially informative loci only. Additionally, as the error rates for the category 'B' markers were considerably higher, only markers falling into categories 'A' and 'AB' were included. Reducing the number of markers would also increase the speed of simulations, and enable more runs. After initial exploratory statistics pointed to a high noise level due to shared polymorphism, removing uninformative loci was considered to enhance resolution power for these apparently extremely closely related species. This was achived using three steps, removing following categories of markers:

1. Certainly uninformative: monomorphic and singleton loci.
2. Statistically uninformative: markers showing no frequency differences.
3. Only marginally informative: markers with low frequency differences.

Using only loci of categories 'A' & 'AB' (244 loci), first all monomorphic and singleton loci were removed (−38 loci); then polymorphic loci were removed for which no population marker frequency was significantly (p <0.01) different from the mean marker frequency of the whole dataset. The probability was calculated with function `binom.test` in R, setting: `x` (number of successes) to the observed count in the population; `n` (number of trials) to the number of individuals in the population; and `p` (probability of success) to the observed frequency of counts in the whole dataset (−26 loci). Finally, after removing loci and individuals with missing data (−5 loci), only loci were kept for which at least one population had a marker frequency >0.15, and additionally the highest frequency was at least 4× the lowest (−48 loci).

Dataset II comprised 127 loci and 390 individuals (Table 3.2). As intended the pruned data emphasised the differences between populations and species, but did not change the results of any analysis in a qualitative way (e.g. section 3.4, page 111).

Three individuals were erroneously removed from Dataset II because of unfortunate application of the pruning functions: when removing missing data, loci and individuals were removed at the same time, first removing individuals with missing data beyond a threshold, then loci. The threshold for the three individuals was met, because missing data was initially removed as the first step, not taking into account that certain loci would not meet the succeeding applied criteria as e.g. frequency thresholds. This was later corrected, but

several analyses, involving lengthy simulation runs had already been performed; the impact of the individuals was judged to not affect the results, and time considerations led to the decision that for consistency the dataset excluding these individuals would be used for all analyses.

Dataset III After discovering that markers in the *R. aganniphum* × *phaeochrysum* populations were behaving unusually in analyses performed with Dataset II, it was decided to include category 'B' for the species pair *R. aganniphum*-*R. phaeochrysum*. Because many of the 127 loci of Dataset II were mostly informative for the populations of *R. clementinae* and *R. roxieanum*, and hence uninformative for the investigation of *R. aganniphum* × *phaeochrysum*, it made sense to include as many loci as possible. Although higher than in 'A' and 'AB', the error for the category 'B' loci was not extraordinary, and for the analysis of the marker skew a new dataset was assembled including all three categories. After removing individuals belonging to *R. clementinae* and *R. roxieanum* the same procedure as for Dataset II was followed. Because markers exclusively informative for *R. clementinae* and *R. roxieanum*, according to the criteria applied above, were removed, the dataset comprised different loci, informative for *R. aganniphum* × *phaeochrysum*. Dataset III finally comprised 105 loci and 192 individuals (Table 3.2, page 98).

3.1.2 Putative clones

Some publications report clonal growth for *Rhododendron* [51, 119], and although the sampling was carried out avoiding individuals in close proximity, AFLP profiles were tested for exceedingly high similarity. For this purpose a distance matrix, calculated with the SM[1], was used to determine the mean distance of individuals in a population and the variance of this distance. Samples that showed a smaller distance than 2× the standard deviation were considered to be possible clones. This cut-of point was considered to be adequate to detect profiles that are likely the same (roughly 95% chance), while keeping the α-error (false positive) small. Five putative clone pairs were identified including loci from Dataset II only; two of these were still considered clones with all scored loci (Table 3.3). This analysis was carried out relatively late (after the unexpected behaviour of populations GH2a and GH2b), and the small number of putative clones, which should not affect the analyses, was not considered to justify removal

[1]chapter 2, subsection 2.3.2, page 65

Table 3.3: Putative clones individuals showing a lower than expected distance, using twice the standard deviation of the population as an indicator. Distances were calculated with the SM*.

		Dataset II (127 loci)		Dataset I (337 loci)	
Putative clone pairs	Population	Margin[a]	Distance[b]	Margin[a]	Distance[b]
TC043 - TC047	C2	0.0912	0.0000	0.0635	over margin[c]
TC060 - TC066	C2	0.0912	0.0887	0.0635	over margin[c]
TGP35 - TGP41	GH2b	0.1085	0.0000	0.0749	0.0545
TXC02 - TXC03	XC1	0.0966	0.0000	0.0653	over margin[c]
TXC38 - TXC39	XC1	0.0966	0.0000	0.0653	0.0545

[a] Distance below which the pair is assumed to be a clone; 2×standard deviation from the mean distance between individuals in the population.
[b] Distance between both individuals.
[c] This individual-pair is not considered a clone using 337 loci.
* chapter 2, subsection 2.3.2, page 65.

and repetition of most analyses at this stage. They are noted here, but none of the putative clone individuals have been excluded from the analysis.

Table 3.4: Eigenvalues obtained for MDS analysis. Values of the first five eigenvectors, and the respective percentage of variation explained (%) are shown.

		All species		TC - TX		TG - TP	
Index	Axis	Value	%	Value	%	Value	%
1	X	17.120	19.29	19.089	30.61	7.026	14.85
2	Y	11.176	12.59	3.825	6.13	3.365	7.11
3	Z*	5.691	6.41	2.722	4.37	2.138	4.52
4	not plotted	3.444	3.88	1.958	3.14	1.928	4.08
5	not plotted	2.766	3.12	1.757	2.82	1.708	3.61

* Only plotted for the MDS analysis including all species

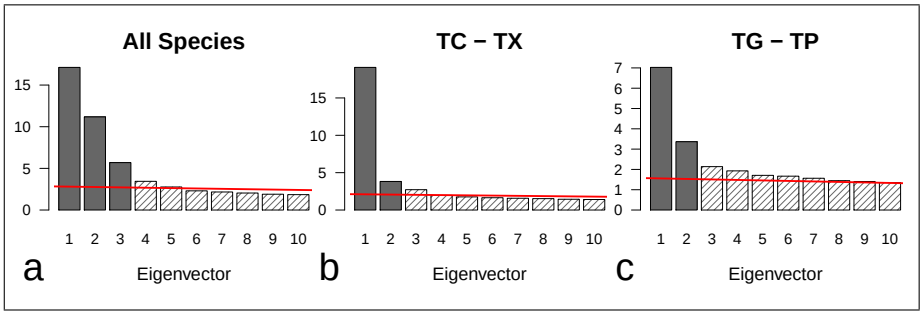

Figure 3.1: Eigenvalues obtained for MDS; shown are the ten largest eigenvalues, plotted eigenvectors are in dark grey; the red line represents the regression line through the 100 largest eigenvalues of the respective dataset.

3.2 Multidimensional Scaling

Eigenvalues obtained for the MDS analyses showed that the main axes separating groups do not explain the majority of the variation (Table 3.4). This is likely due to the extensive shared polymorphism between species, and the lack of fixed differences, as most variation is then found between individuals, not groups.

Combined analysis including all species For the analysis including all species, but no potential hybrids, the first two eigenvectors together represent 31.88% of the variation. Using a slope cutoff (scree test [70]), suggests that the first three eigenvectors should be plotted for the representation of the data (Figure 3.1, a). The four species then form distinct clusters, which are well separated, apart from a few individuals producing a slight overlap between *R. aganniphum* and *R. phaeochrysum* (Figure 3.2).

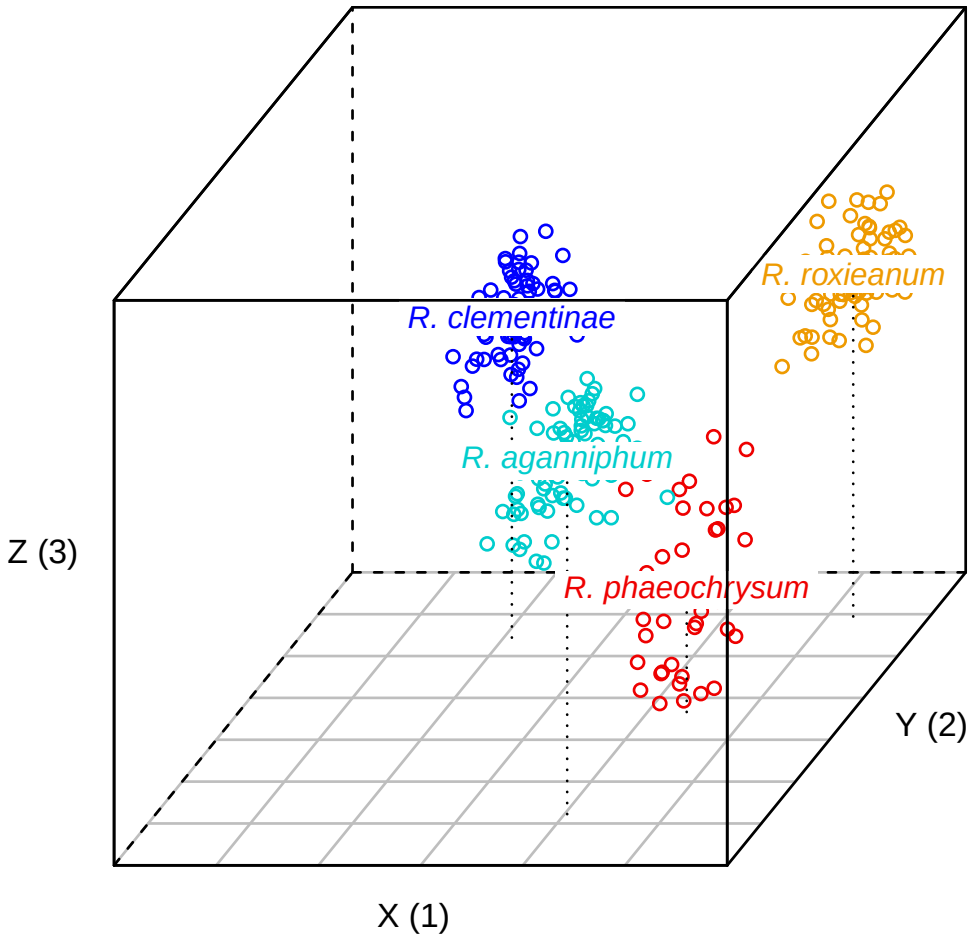

Figure 3.2: MDS including all species, potential hybrids were removed. the dotted lines connect the centroid of each species with the XY surface, hence represent shift in the Z dimension; to visualise separation in the XY space only, the point where the line meets the XY surface can be used as a indicator; for variation explained by the different axes see Table 3.4.

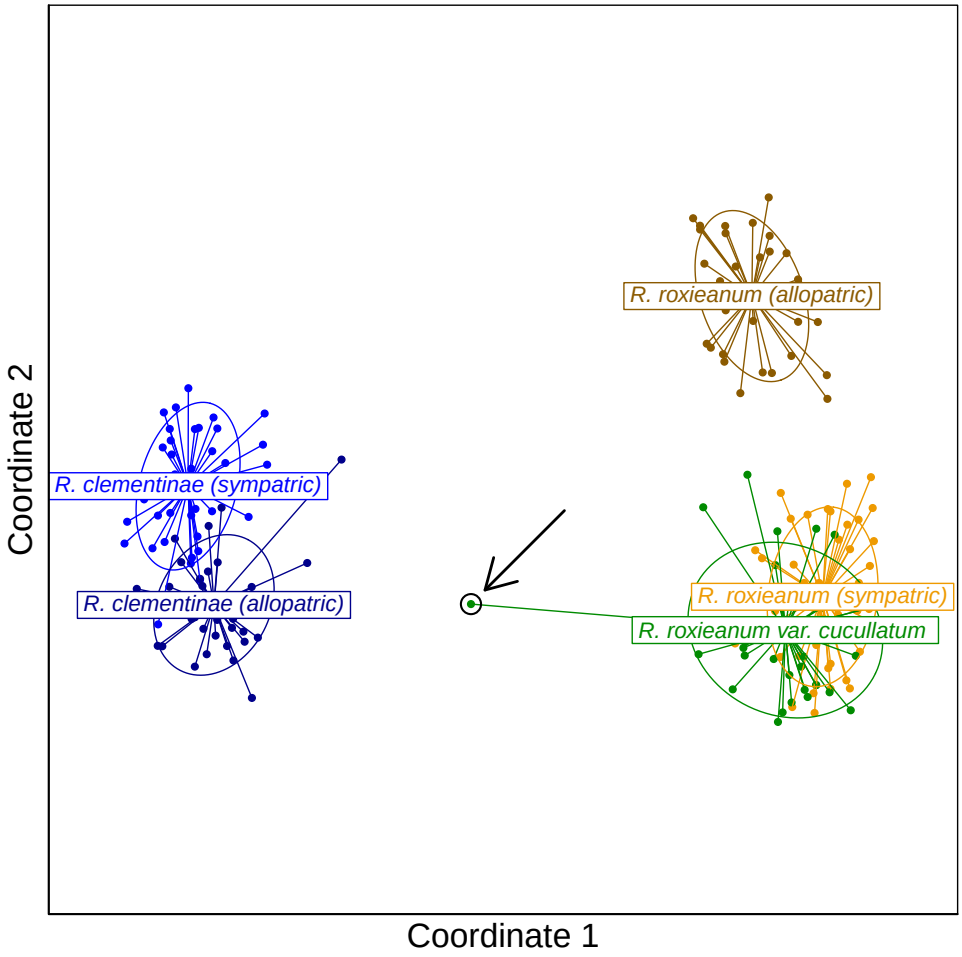

Figure 3.3: MDS *R. clementinae* — *R. roxieanum* including var. *cucullatum*. Samples are connected with lines to the centroid of the respective population, and ellipses represent 95% confidence areas for the populations. Individual `TXC09` is indicated by an arrow; for variation explained by the different axes see Table 3.4.

R. clementinae* and *R. roxieanum For the subset of the data including only species involved in hybrid zone one (*R. roxieanum* var. *cucullatum*), *R. clementinae* and *R. roxieanum*, a considerable amount of variation is explained by the largest eigenvector only (30.61%, Table 3.4). This coordinate corresponds to the difference between the species, which are clearly separated (Figure 3.3). The good distinction conforms with the observation that fixed differences were only detected for this species pair (section 3.1, page 99). The second eigenvector represents much less information (6.13%, Figure 3.1, b), but clearly separates the sympatric and allopatric populations of *R. roxieanum*. However, apart from one individual of *R. roxieanum* var. *cucullatum*, `TXC09` (indicated by an arrow in Figure 3.3), the variety is not markedly separated from the sympatric population of *R. roxieanum* var. *roxieanum*.

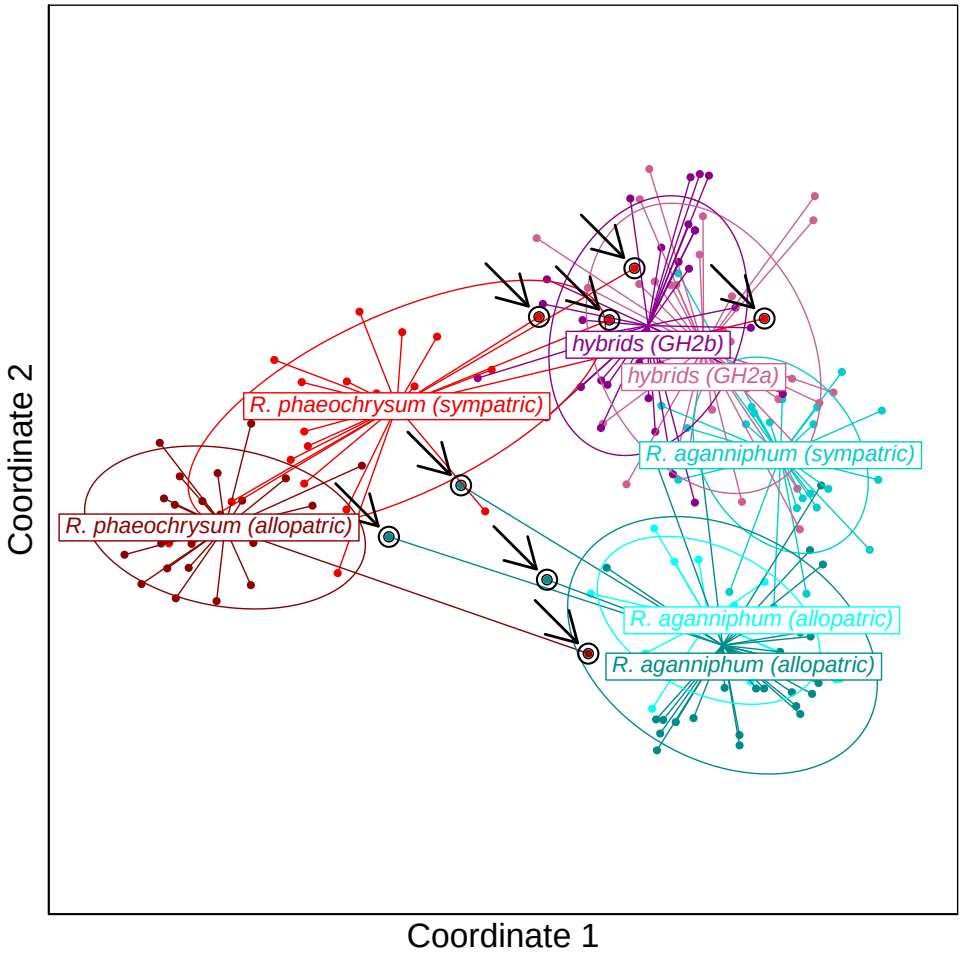

Figure 3.4: MDS *R. aganniphum* — *R. phaeochrysum* and putative hybrids. Samples are connected with lines to the centroid of the respective population, and ellipses represent 95% confidence areas for the populations. Four individuals of *R. phaeochrysum* var. *agglutinatum* clustering with hybrid population GH2b are indicated by arrows (top right), and four individuals from the allopatric populations that are markedly outside their population confidence (middle); for variation explained by the different axes see Table 3.4.

R. aganniphum* and *R. phaeochrysum The subset for the second hybrid zone around *R. aganniphum* × *phaeochrysum* conforms worse than the other cases to a low dimensional representation. Only 21.96% of the variation is explained by the two largest eigenvectors (Table 3.4), and probably some information is lost not plotting the third largest. However, noise levels are expected to be higher in these more closely related species, not separated by fixed differences. Furthermore, the difference in magnitude of the second eigenvalue to the third, with regards to the difference of the third to the fourth (Table 3.4, and Figure 3.1, c), justifies the use of only the two largest eigenvectors. Although the two dimensional representation separates the populations, considerable

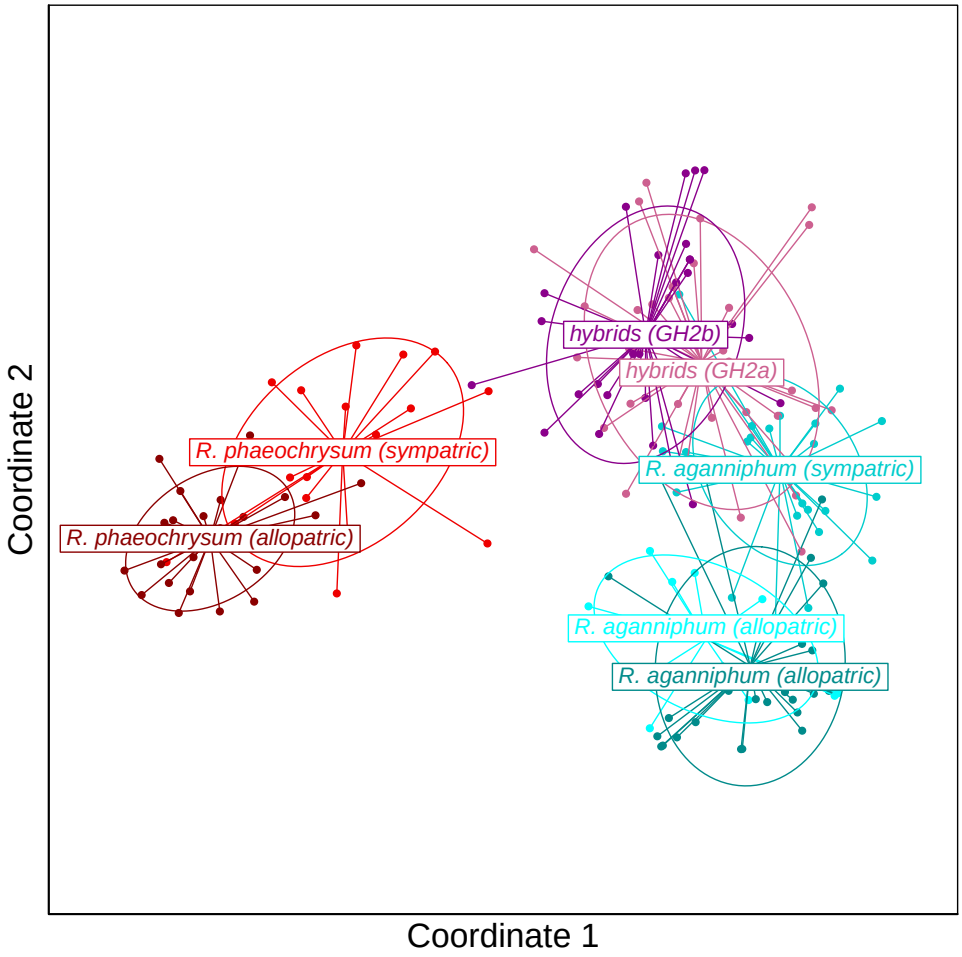

Figure 3.5: MDS *R. aganniphum* — *R. phaeochrysum* and putative hybrids, admixed individuals of parental populations were removed. Samples are connected with lines to the centroid of the respective population, and ellipses represent 95% confidence areas for the populations; for variation explained by the different axes see Table 3.4.

overlap exists (Figure 3.4). Two interesting patterns can be observed: four individuals of *R. phaeochrysum* (`TP027, TP029, TP030, TP031`), for which David Chamberlain remarked during the collections in 2007 that they would probably be attributable to var. *agglutinatum*, cluster with, or close to, hybrid population GH2b (Figure 3.4, individuals indicated by four arrows, top-right); one individual (`TP080`) of the allopatric population of *R. phaeochrysum* (Dàxuě Shan[2]) clusters with the population of *R. aganniphum* obtained at this locality, and in reverse three individuals of that population (`TG062, TG080, TG088`) tend towards *R. phaeochrysum* (Figure 3.4, individuals indicated by arrows, centre). After removing these eight individuals, the populations are better separated: the two populations of *R. phaeochrysum* form a cluster as well as

[2]chapter 1, subsection 1.6.4, page 44

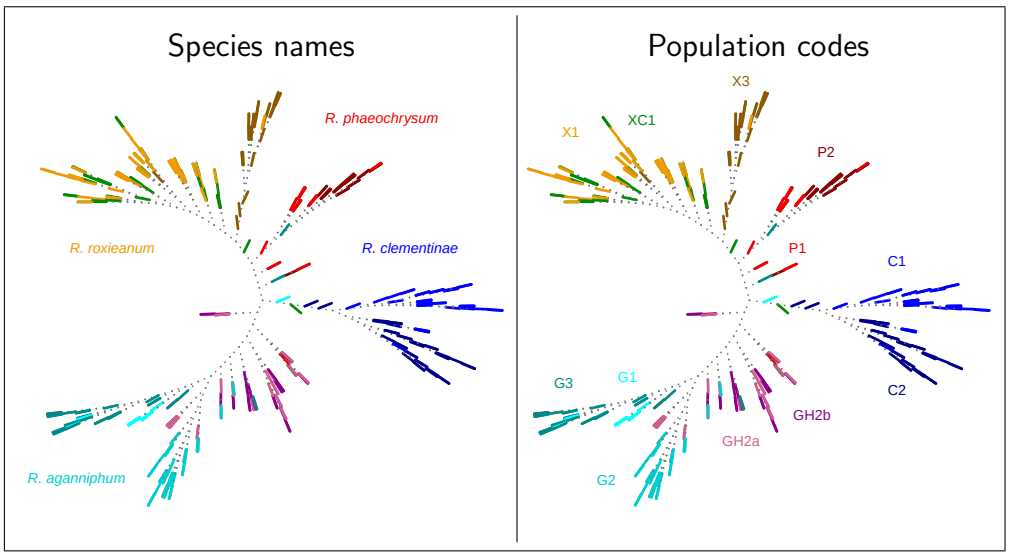

Figure 3.6: NJ tree including all species and potential hybrids. Terminal branch lengths are not shown; for further information about populations please see Table 1.1, page 49.

the two allopatric populations of *R. aganniphum*; the sympatric *R. aganniphum* population is closer to the hybrids, and while it seems slightly separated from hybrid population GH2b, individuals from GH2a cluster with both of them (Figure 3.5).

3.3 Neighbour-Joining trees

As mentioned before, the majority of the variation is found between individuals, and correspondingly most changes occurred on the terminal branches of the NJ trees; the groups only being separated by very few changes. This made meaningful visualisation difficult, and the terminal branch lengths were not plotted in the figures of this section. Already the MDS of all species suggested that the data would not conform well to a tree-like behaviour, as evidenced by three dimensions needed to show species relationships (section 3.2, Figure 3.2, page 103); only several nodes separating individuals are supported after bootstrapping (>90%, 1000 replicates) and most internal branches collapse, therefore no bootstrap support is shown.

Combined analysis including all species Because of the dichotomous clustering of NJ, potential hybrids could be included without compromising interpretability of the graphical representation. The results are not considerably different from the MDS analysis, and similar patterns can be observed

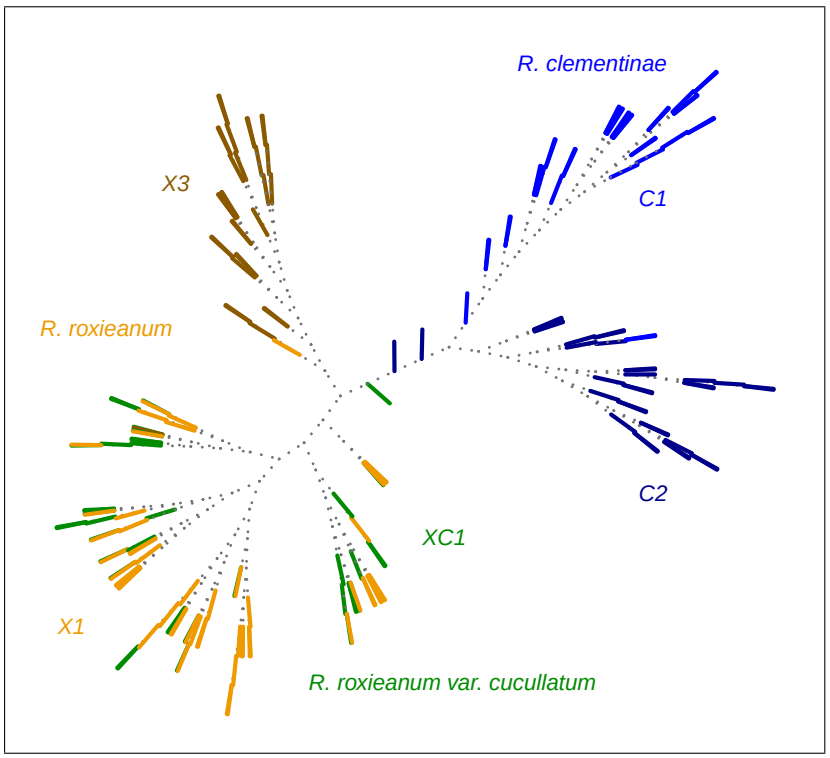

Figure 3.7: NJ tree for *R. clementinae* — *R. roxieanum*. Terminal branch lengths are not shown; for further information about populations please see Table 1.1, page 49.

(Figure 3.6). First, individuals of the same population generally cluster together, with good separation for the populations of *R. clementinae* (C1, C2), *R. roxieanum* (X1, X3), and allopatric *R. aganniphum* (G1, G3) versus sympatric *R. aganniphum* (G2). Second, *R. phaeochrysum* rather clusters by species than by population (P1, P2), and *R. roxieanum* var. *cucullatum* (XC1) clusters with the sympatric *R. roxieanum* var. *roxieanum* population (X3). Third, the putative *R. aganniphum* × *phaeochrysum* do not form a homogeneous cluster, but are included in the cluster encompassing all populations of *R. aganniphum*. Furthermore some of the individuals already behaving unusually in the MDS analysis can be identified again: one individual of *R. roxieanum* var. *cucullatum* clustering near *R. clementinae*; two individuals from population G3 clustering with *R. phaeochrysum*; and one individual from P1 clustering with GH2a/b.

R. clementinae* and *R. roxieanum The separate analysis of the subset including only *R. clementinae* and *R. roxieanum* shows a very clean clustering according to population for C1, C2 and X3 (Figure 3.7). Although X1 and XC1 form a cluster, it seems more sub-partitioned than the others. Furthermore, one individual (`TXC09`) is not included in the latter cluster, but is placed on the branch

connecting the two species, *R. clementinae* and *R. roxieanum*.

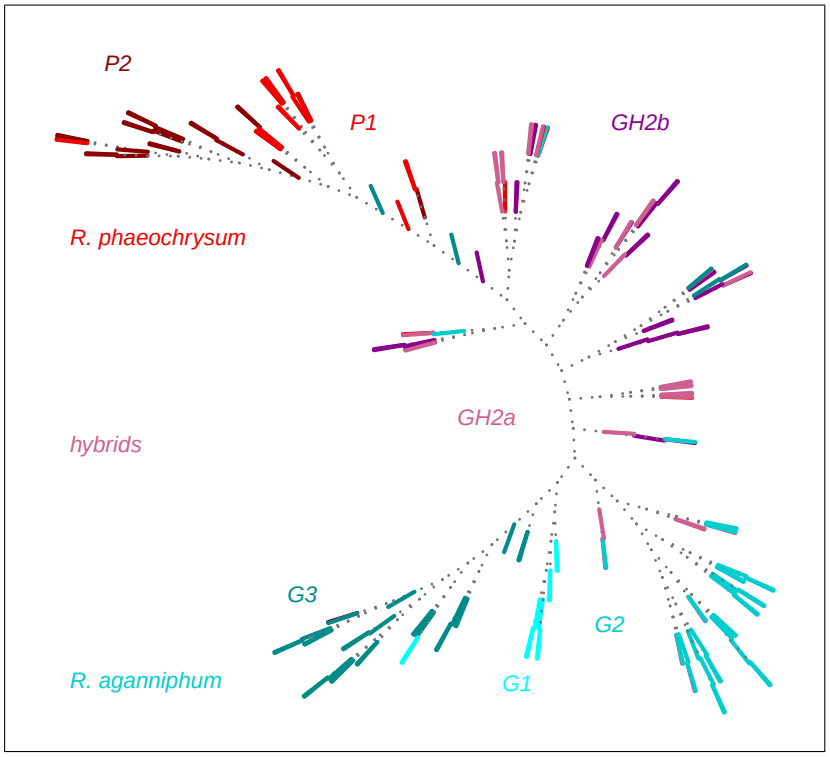

Figure 3.8: NJ tree for *R. aganniphum* — *R. phaeochrysum* and potential hybrids. Terminal branch lengths are not shown; for further information about populations please see Table 1.1, page 49.

R. aganniphum* and *R. phaeochrysum In the NJ tree for the subset of *R. aganniphum* and *R. phaeochrysum* the two species are at the extreme ends, with potential hybrids being placed in the middle (Figure 3.8). While the populations of *R. phaeochrysum* (P1, P2) do not form well defined clusters, the populations of *R. aganniphum* mostly do. The space in between the species clusters is widely dominated by individuals classified as putative hybrids; however, several individuals classified as a member of either species have been placed here as well: close to the cluster comprising the populations of *R. phaeochrysum*, individuals of the sympatric population P1 intergrade with the hybrids; additionally close to this boundry of the *R. phaeochrysum* cluster some individuals of G3 have been placed; several individuals of *R. aganniphum*, belonging to populations G2 and G3, cluster inbetween the species (*R. aganniphum* and *R. phaeochrysum*), together with the hybrids; finally some putative hybrids from the swarm population (GH2a) cluster near the node of population G2.

Table 3.5: AMOVA **variance components** obtained for an analysis of the whole dataset comprising four species and hybrids; hybrids were entered into the analysis as seperate species (*R. aganniphum* × *phaeochrysum* as one, and *R. roxieanum* var. *cucullatum* as another).

Dataset				Variance		% total		
Variance component	*df*	SSD	MSD	Observed	95% conf[a]	Observed	95% conf[a]	*p*-value[b]
Dataset II (127 loci)								
Between species	5	12.769	2.554	σ_a^2 0.033	0.024 – 0.042	26.1	20.2 – 32.1	<0.001
Between populations w/ species	6	2.741	0.457	σ_b^2 0.013	0.010 – 0.016	10.2	8.0 – 12.8	<0.001
Within populations	368	29.338	0.080	σ_c^2 0.080	0.071 – 0.089	63.6	58.3 – 68.8	<0.001
Total	379	44.848	0.118	σ_T^2 0.125	0.112 – 0.139	100.0		
Dataset I (337 loci)								
Between species	5	7.290	1.458	σ_a^2 0.018	0.014 – 0.023	21.6	17.6 – 25.6	<0.001
Between populations w/ species	6	1.677	0.279	σ_b^2 0.007	0.006 – 0.009	8.9	7.3 – 10.6	<0.001
Within populations	371	21.739	0.059	σ_c^2 0.059	0.052 – 0.065	69.5	65.7 – 73.4	<0.001
Total	382	30.705	0.080	σ_T^2 0.084	0.075 – 0.094	100.0		

[a] Calculated from 5000 bootstrap replicates.
[b] Calculated from 5000 permutations of the respective hierarchical level, for permuted levels see subsection 2.3.5, page 82; and for obtained null-distributions Appendix E, Figure E.1, page 239.

3.4 AMOVA

Excluding admixed individuals Certain individuals were excluded before performing the AMOVA. These were the admixed parental individuals identified with the MDS and NJ analyses, and confirmed with STRUCTURE; along with the unrepresentative individual from *R. roxieanum* var. *cucullatum* (`TXC09`, see e.g. Figure 3.3, page 104). This was considered appropriate, as the uncharacteristic individuals would distort this group-based analysis. The following individuals were removed: `TG008, TG062, TG080, TG088, TP027, TP029, TP030, TP031, TP080,` and `TXC09`. Additionally, *R. roxieanum* var. *cucullatum* and the two populations of *R. aganniphum* × *phaeochrysum* were included as species. To assess the impact of selectively removing uninformative loci from the initial dataset, an analysis was performed with Dataset II as well as Dataset I[3]; for pairwise comparisons, only results obtained with Dataset II will be shown.

Analysis of all species combined This overall analysis of the data was mainly used to assess differences potentially caused by removing loci. The comparison of the AMOVAs including all species and hybrids, calculated with Dataset I and Dataset II respectively, shows that they agree qualitatively (Table 3.5): most variation is found within populations (I = 69.5%/II = 63.6%); and substantially more between species as opposed to between populations within species. However, Dataset II pronounces between group variance, as was intended. First, this results in variation found between species using Dataset II (26.1%) being significantly higher than using Dataset I (upper confidence 25.6%). Second, the difference of between population variation within species is not affected significantly (Dataset II = 10.2%, upper confidence for Dataset I = 10.6%). Third, the shift of emphasis to between species variation is mainly obtained by removing within population variance, which is significantly lower for Dataset II (upper confidence limit 68.8%) than using Dataset I (69.5%). For both datasets differentiation at all hierarchical levels is highly significant (p <0.001).

Due to the removal of within population variance, all calculated Φ-statistics are significantly higher for Dataset II (Table 3.6), but yield qualitatively similar results, further justifying the the use of Dataset II in other analyses. The average differentiation found among populations within species ($\Phi_{ST} = 0.139$) is roughly half the average differentiation of species ($\Phi_{TG} = 0.261$), indicating that species identity is more important than population identity as a predictor of divergence between individuals.

[3]For composition of the different datasets see subsection 3.1.1, page 97.

Table 3.6: AMOVA **Φ-statistics** including all species and hybrids.

Dataset	Φ-value		
Statistic	Observed	95 % conf[a]	p-value[b]
Dataset II (127 loci)			
Φ_{SG}	0.364	0.312 – 0.417	<0.001
Φ_{ST}	0.139	0.111 – 0.169	<0.001
Φ_{TG}	0.261	0.202 – 0.321	<0.001
Dataset I (337 loci)			
Φ_{SG}	0.305	0.266 – 0.343	<0.001
Φ_{ST}	0.113	0.094 – 0.133	<0.001
Φ_{TG}	0.216	0.176 – 0.256	<0.001

[a] Calculated from 5000 bootstrap replicates.
[b] Calculated from 5000 permutations of the respective hierarchical level, for permuted levels see chapter 2, subsection 2.3.5, page 82; and for obtained null-distributions Figure E.1, Appendix E, page 239.

The separate calculation of Φ_{ST} for each species (Table 3.7)[4] reveals that no individual species Φ_{ST} is significantly different from the mean Φ_{ST}, therefore no major differences in population structure of the species are evident. However, the populations of *R. clementinae* are significantly more differentiated (Φ_{ST} = 0.198) than the *R. phaeochrysum* populations (upper confidence Φ_{ST} = 0.166). This likely reflects the different abundance of these two species: *R. clementinae* has the most restricted distribution range of the species investigated, and occurs only at higher altitudes, therefore leaving larger distances between populations; *R. phaeochrysum* can frequently be found at lower altitudes, decreasing effective distances between populations, and therefore potentially providing better bridges for geneflow. Lastly the differentiation of the two *R. aganniphum* × *phaeochrysum* hybrid populations is much lower than the species mean, nonetheless this differentiation is significant (Φ_{ST} = 0.034, p <0.001).

Pairwise comparisons (species) As mentioned above[5], the randomisations for the pairwise Φ_{TG}-values, to obtain p-values, were problematic because of the sample sizes (mostly four populations per species pair), and "significance" was judged differently as will be explained as follows. During randomizations of the hierarchical levels, the small sample sizes of populations per species led

[4]For variance components underlying the calculation of the Φ_{ST}-values see Appendix E, Table E.3, page 224.
[5]chapter 2, subsection 2.3.5 (page 82)

Table 3.7: AMOVA Φ_{ST}**-values** for individual species.

Species	Φ_{ST}	95% conf[a]	p-value[b]
R. clementinae (TC)	0.198	0.131 – 0.264	<0.001
R. aganniphum (TG)	0.167	0.109 – 0.233	<0.001
R. aganniphum × *phaeochrysum* (TGP)	0.034	0.016 – 0.052	<0.001
R. phaeochrysum (TP)	0.111	0.057 – 0.166	<0.001
R. roxieanum (TX)	0.164	0.113 – 0.220	<0.001

Note: TXC was not included here because only one population was available, so that no Φ_{ST} can be calculated.

[a] Calculated from 5000 bootstrap replicates.

[b] Calculated from 5000 permutations of the respective hierarchical level, for permuted levels see chapter 2, subsection 2.3.5, page 82; and for obtained null-distributions Appendix E, Figure E.2, page 240.

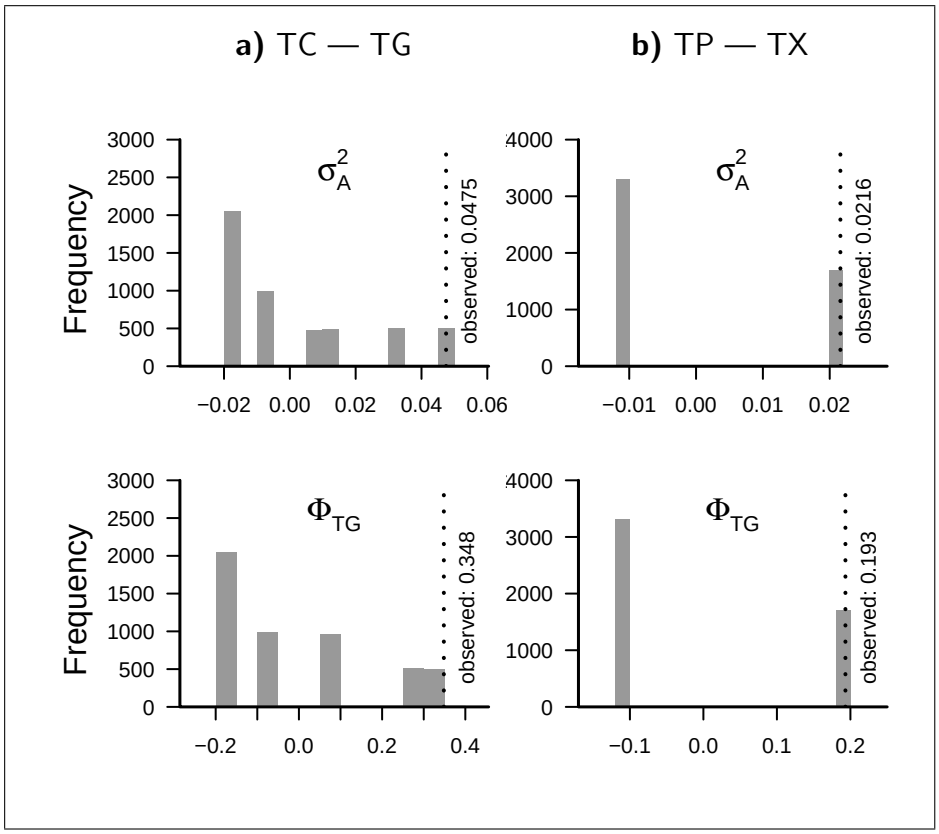

Figure 3.9: AMOVA **'significant' null-distributions for** Φ_{TG} and σ_A^2 for two selected species pairs; obtained by randomising whole populations within species pairs. Only one of the randomisation groups, representing the real grouping yields a value as high as the observed one.

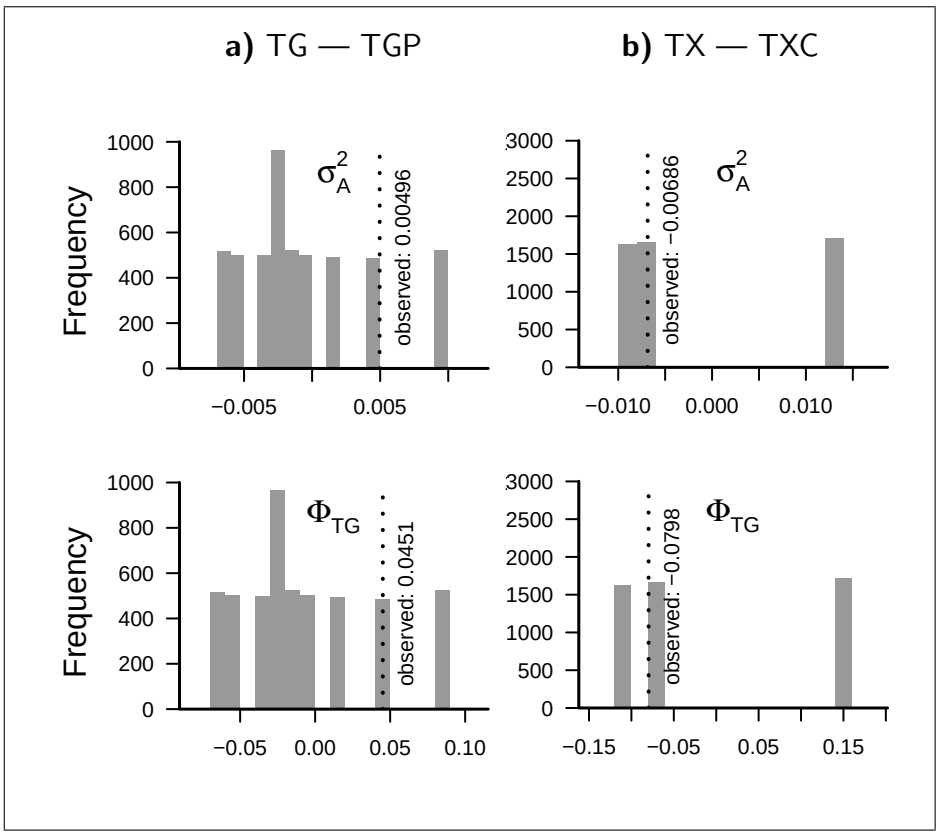

Figure 3.10: AMOVA **not 'significant' null-distributions for** Φ_{TG} and σ^2_A for two selected species pairs; obtained by randomising whole populations within species pairs. At least one random group pairing yields a higher value than the observed one.

to very frequent grouping of the populations belonging to one species. This necessarily resulted in the observed value of the statistic in the non randomised dataset. Additionally, the distribution of Φ_{TG}-values obtained by randomisations was clearly multimodal (Figure 3.9), showing that values obtained with most randomised combinations were far below the observed value. Often the two modes (randomly grouping populations belonging to the same species vs. random combinations) were so far apart, that the random groupings fell into one category only, when plotting frequency diagrams (Figure 3.9, **b**). Therefore, differentiation was assumed to be significant when no other combination of populations yielded a higher value than the observed one. In other words only one randomised value (representing the grouping of populations of the same species) was overlapping with the observed value as e.g. in Figure 3.9. If more randomised values overlapped with the observed value, or if some random combinations yielded a higher value (as in Figure 3.10), no significant differentiation of the species was

Table 3.8: AMOVA **Pairwise Φ_{TG}-values** for species comparisons. Observed values are shown in the lower triangle; the upper triangle contains 95% confidence intervals, obtained with 5000 bootstrap replicates.

	TC	TG	TGP	TP	TX	TXC
TC	•	$\frac{0.254}{0.436}$	$\frac{0.249}{0.415}$	$\frac{0.225}{0.436}$	$\frac{0.263}{0.466}$	$\frac{0.235}{0.477}$
TG	0.348	•	$\frac{-0.002}{0.099}$	$\frac{0.148}{0.310}$	$\frac{0.206}{0.374}$	$\frac{0.186}{0.367}$
TGP	0.333	0.045[N]	•	$\frac{0.132}{0.266}$	$\frac{0.162}{0.327}$	$\frac{0.199}{0.353}$
TP	0.334	0.231	0.197	•	$\frac{0.113}{0.273}$	$\frac{0.131}{0.297}$
TX	0.367	0.293	0.245	0.193	•	$\frac{-0.133}{-0.032}$
TXC	0.362	0.280	0.276	0.211	-0.080[N]	•

[N] These species pairs are not considered to be significantly differentiated; please see text for explanation, and Figure 3.10.

assumed[6]. Having defined significance in that way, all species are significantly differentiated from each other (Table 3.8). However, *R. aganniphum* (TG) is not differentiated from the putative hybrids (TGP), while *R. phaeochrysum* (TP) is (Φ_{TG} = 0.197); this is even in the range of normal species differentiation, e.g. TP/TX (Φ_{TG} = 0.193). Equally *R. roxieanum* (TX) and *R. roxieanum* var. *cucullatum* (TXC) are not differentiated (Φ_{TG} = -0.080); the negative value here can easily be explained by considering that Φ_{TG} represents species differentiation in excess of population differentiation, therefore, if populations within species are more differentiated than the species themselves, the value becomes negative. Furthermore, although all pairwise comparisons involving *R. clementinae* (TC) tend to be higher, most are not significantly higher than the mean (0.261, Table 3.6), when taking into account the confidence intervals. The only exception is *R. clementinae* (TC)/*R. roxieanum* (TX), which is marginally significant (lower confidence limit = 0.263). This is in accord with the only two fixed differences found in the dataset, which were for this species pair (section 3.1, page 99). Lastly, *R. phaeochrysum* is less differentiated from both, *R. aganniphum* and *R. roxieanum* than either is from *R. clementinae*.

[6]For all obtained null-distributions for pairwise species comparisons please see Figure E.3, Appendix E, pages 241 ff; and for variance components of the pairwise species datasets Table E.4, Appendix E, page 225

Table 3.9: AMOVA **Pairwise Φ_{PT}-values** for population comparisons. Observed values are shown in the lower triangle; the upper triangle contains 95% confidence intervals, obtained with 5000 bootstrap replicates.

	C1	C2	G1	G2	G3	GH2a	GH2b	P1	P2	X1	X3	XC1
C1	•	$\frac{0.131}{0.264}$	$\frac{0.374}{0.556}$	$\frac{0.384}{0.559}$	$\frac{0.378}{0.573}$	$\frac{0.321}{0.482}$	$\frac{0.322}{0.482}$	$\frac{0.302}{0.502}$	$\frac{0.369}{0.576}$	$\frac{0.401}{0.576}$	$\frac{0.377}{0.563}$	$\frac{0.375}{0.559}$
C2	0.198	•	$\frac{0.366}{0.547}$	$\frac{0.380}{0.554}$	$\frac{0.376}{0.566}$	$\frac{0.319}{0.478}$	$\frac{0.323}{0.478}$	$\frac{0.277}{0.483}$	$\frac{0.361}{0.560}$	$\frac{0.370}{0.558}$	$\frac{0.370}{0.563}$	$\frac{0.340}{0.542}$
G1	0.470	0.460	•	$\frac{0.052}{0.139}$	$\frac{0.087}{0.225}$	$\frac{0.119}{0.239}$	$\frac{0.148}{0.277}$	$\frac{0.208}{0.371}$	$\frac{0.291}{0.489}$	$\frac{0.332}{0.513}$	$\frac{0.304}{0.490}$	$\frac{0.321}{0.496}$
G2	0.476	0.471	0.094	•	$\frac{0.083}{0.242}$	$\frac{0.019}{0.111}$	$\frac{0.062}{0.223}$	$\frac{0.195}{0.375}$	$\frac{0.281}{0.468}$	$\frac{0.300}{0.480}$	$\frac{0.299}{0.461}$	$\frac{0.291}{0.464}$
G3	0.485	0.479	0.154	0.158	•	$\frac{0.098}{0.241}$	$\frac{0.115}{0.278}$	$\frac{0.212}{0.411}$	$\frac{0.281}{0.482}$	$\frac{0.348}{0.532}$	$\frac{0.344}{0.525}$	$\frac{0.341}{0.521}$
GH2a	0.403	0.401	0.178	0.061	0.165	•	$\frac{0.016}{0.052}$	$\frac{0.128}{0.249}$	$\frac{0.217}{0.373}$	$\frac{0.235}{0.399}$	$\frac{0.230}{0.369}$	$\frac{0.226}{0.377}$
GH2b	0.405	0.403	0.212	0.137	0.195	0.034	•	$\frac{0.131}{0.269}$	$\frac{0.194}{0.359}$	$\frac{0.256}{0.426}$	$\frac{0.229}{0.396}$	$\frac{0.239}{0.403}$
P1	0.406	0.385	0.293	0.286	0.316	0.188	0.197	•	$\frac{0.057}{0.166}$	$\frac{0.159}{0.324}$	$\frac{0.150}{0.325}$	$\frac{0.155}{0.304}$
P2	0.481	0.469	0.396	0.379	0.388	0.296	0.278	0.111	•	$\frac{0.263}{0.442}$	$\frac{0.258}{0.464}$	$\frac{0.251}{0.428}$
X1	0.493	0.470	0.427	0.393	0.445	0.317	0.343	0.240	0.356	•	$\frac{0.113}{0.220}$	$\frac{0.009}{0.050}$
X3	0.473	0.473	0.400	0.383	0.441	0.301	0.314	0.235	0.364	0.164	•	$\frac{0.124}{0.243}$
XC1	0.472	0.447	0.412	0.380	0.436	0.303	0.323	0.226	0.344	0.028	0.182	•

For all comparisons p <0.001; calculated by permuting individuals randomly within each population pair (for null-distributions see Appendix E, Figure E.4–E.4, pages 249 ff).

Pairwise comparisons (populations) The pairwise comparison of populations confirms observations made for the pairwise species comparisons (Φ_{TG}), but also adds information with regards to the relationship of the potential hybrids to the parental populations. *R. roxieanum* var. *cucullatum* (TXC) is well differentiated from the allopatric *R. roxieanum* (X3), $\Phi_{PT} = 0.182$, Table 3.9, but only slightly, although significantly, from the sympatric population (X1), $\Phi_{PT} = 0.028$, p <0.001. The two *R. aganniphum* × *phaeochrysum* hybrid populations (GH2a, GH2b) show a very interesting pattern: while GH2a shows low differentiation from the sympatric *R. aganniphum* (G2), $\Phi_{PT} = 0.061$, GH2b is considerably more differentiated $\Phi_{PT} = 0.137$, which is comparable with the level of population differentiation in *R. aganniphum* ($\Phi_{PT} = 0.094$–0.158). Furthermore, both hybrid populations are considerably differentiated from the sympatric *R. phaeochrysum* (P1), $\Phi_{PT} = 0.188$ and 0.197, which is significantly higher than population differentiation in *R. phaeochrysum* ($\Phi_{PT} = 0.111$, upper confidence limit = 0.166).

Table 3.10: Gene diversity (H_j) for all populations, calculated with the program AFLP-SURV; using the bayesian method of Zhivotovsky*, with non-uniform prior, to estimate allele frequencies.

		Dataset II (127 loci)		Dataset I (337 loci)	
Population	n	H_j	SD_{Hj}[a]	H_j	SD_{Hj}[a]
C1	37	0.104	0.013	0.100	0.0079
C2	36	0.099	0.013	0.098	0.0078
G1	11	0.180	0.016	0.157	0.0092
G2	35	0.163	0.016	0.142	0.0090
G3	30	0.152	0.015	0.130	0.0089
GH2a	33	0.169	0.015	0.146	0.0087
GH2b	32	0.164	0.015	0.140	0.0085
P1	18	0.127	0.013	0.140	0.0085
P2	23	0.125	0.014	0.134	0.0086
X1	53	0.123	0.014	0.116	0.0083
X3	33	0.126	0.014	0.120	0.0085
XC1	39	0.126	0.014	0.118	0.0082
Mean	12[b]	0.138	0.008	0.128	0.0053

[a] Standard deviation of the gene diversity estimates
[b] For the mean, n is the number of populations.
* See chapter 2, subsection 2.3.3, page 67

3.5 Gene diversity and F_{ST}

As for the AMOVA analysis Dataset II as well as Dataset I were analysed to ensure qualitative agreement of the results. To be consistent with the previous analyses, admixed parental individuals were removed, and the same data used as for the AMOVA (section 3.4, page 111).

Gene diversity The gene diversity estimates show small, but in *R. aganniphum* and *R. phaeochrysum* significant differences depending on which dataset was used for their calculation. However, generally the gene diversity is related to species rather than population: all populations of the same species do not have significantly different estimates (Table 3.10). *R. clementinae* (C1, C2) has significantly lower diversity than all other species, consistently in both datasets ($H_j = 0.098$–0.104). Although *R. aganniphum* (G1–3) has a significantly higher diversity than all other species using Dataset II ($H_j = 0.152$–0.180), *R. aganniphum* and *R. phaeochrysum* (P1, P2) are not different according to

estimates obtained with Dataset I. The diversity found in the populations of potential hybrids is in both cases in the range of an assumed parent: *R. roxieanum* var. *cucullatum* (TXC) is equal to *R. roxieanum* (∼0.126 for both); and the *R. aganniphum* × *phaeochrysum* populations (GH2a, GH2b) show the same level of diversity as *R. aganniphum* (Dataset II) or both, *R. aganniphum* and *R. phaeochrysum* (Dataset I). Because 'gene diversity' here is equivalent to 'expected heterozygosity', hybrids are expected to surpass the values of the parental populations. That this is not observed could be due to two main reasons. First, the allele frequencies are estimated from dominant data, assuming Hardy-Weinberg proportions; this clearly does not hold for hybrid populations, distorting the estimates. Second, most loci are highly polymorphic in all species; therefore the parental species showing polymorphism at more loci will contribute its heterozygosity at loci that are not polymorphic in the other parent. The parent with fewer polymorphic loci, however, does not significantly contribute to heterozygosity in the hybrids, as many of its loci are polymorphic as well in the other parent.

Pairwise population F_{ST} F_{ST} and Φ_{ST} (and in this study Φ_{PT}) are considered to be analogous [81], hence they should yield similar results. However, as slightly different approaches are used to obtain the estimates, they can be expected to differ quantitatively. This is exactly what can be observed when comparing the pairwise F_{ST} estimates (Table 3.11) with the pairwise Φ_{PT}-values (Table 3.9, page 116). The only notable exception, were they differ qualitatively, are the estimates for the populations of *R. aganniphum*: while the Φ_{PT}-values point to less differentiation of G1 and G2 (0.094 vs. 0.154 and 0.158), the F_{ST} estimates place G1 and G3 closer (0.086 vs. 0.111 and 0.115); the F_{ST}-values agree here better with the clustering analyses obtained with MDS and NJ (Figure 3.5, page 106; and Figure 3.8, page 109; respectively).

Table 3.11: Pairwise F_{ST}-values for population comparisons; calculated with the program AFLP-SURV; using the bayesian method of Zhivotovsky*, with non-uniform prior, to estimate allele frequencies. Values obtained with 127 loci (Dataset II) are shown in the lower triangle, values obtained with 337 loci (Dataset I) in the upper.

	C1	C2	G1	G2	G3	GH2a	GH2b	P1	P2	X1	X3	XC1
C1	•	0.069	0.219	0.264	0.276	0.214	0.219	0.245	0.292	0.321	0.295	0.302
C2	0.107	•	0.211	0.260	0.271	0.214	0.215	0.230	0.286	0.315	0.287	0.301
G1	0.272	0.265	•	0.087	0.059	0.087	0.094	0.138	0.185	0.223	0.189	0.209
G2	0.348	0.346	0.115	•	0.086	0.029	0.063	0.126	0.186	0.203	0.191	0.190
G3	0.351	0.347	0.086	0.111	•	0.092	0.099	0.175	0.212	0.275	0.255	0.261
GH2a	0.279	0.277	0.099	0.050	0.119	•	0.020	0.090	0.152	0.177	0.161	0.162
GH2b	0.275	0.277	0.112	0.093	0.123	0.017	•	0.088	0.137	0.187	0.163	0.171
P1	0.306	0.291	0.179	0.199	0.236	0.114	0.133	•	0.051	0.102	0.090	0.103
P2	0.374	0.362	0.247	0.265	0.280	0.196	0.195	0.080	•	0.181	0.174	0.181
X1	0.401	0.377	0.270	0.274	0.333	0.200	0.235	0.160	0.255	•	0.071	0.022
X3	0.368	0.365	0.234	0.251	0.320	0.179	0.205	0.144	0.260	0.107	•	0.095
XC1	0.381	0.356	0.266	0.268	0.329	0.193	0.224	0.145	0.245	0.012	0.124	•
	Dataset II: mean F_{ST} = 0.233, SD = 0.109, p-value[a]<0.001											
	Dataset I: mean F_{ST} = 0.179, SD = 0.113, p-value[a]<0.001											

[a] Calculated from 5000 bootstrap replicates.
* See chapter 2, subsection 2.3.3, page 67

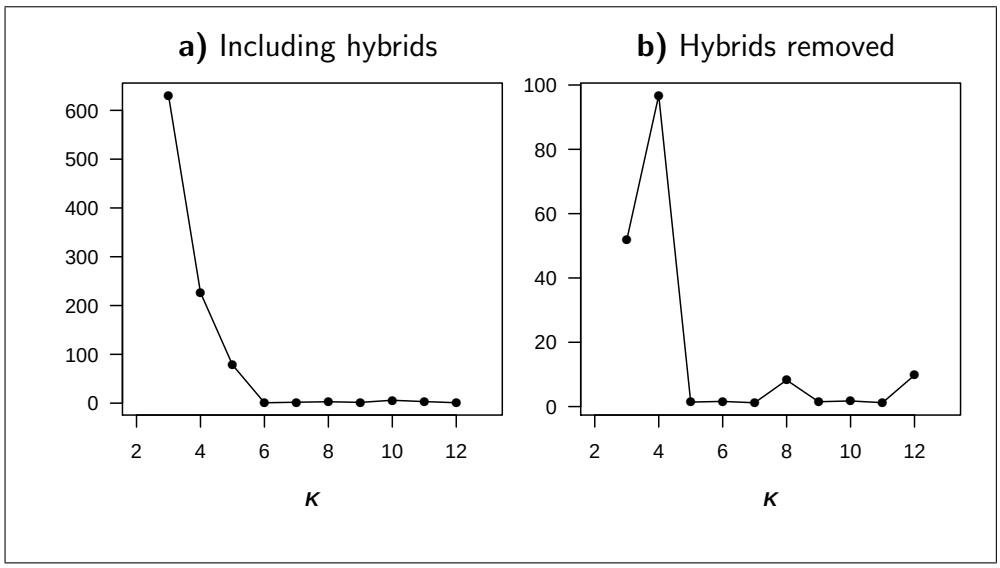

Figure 3.11: ΔK-values (all species); ΔK as a function of the number of assumed clusters K. Calculated from PrD returned by STRUCTURE, performing 10 runs for each K.

3.6 Bayesian clustering (STRUCTURE)

All of the analyses shown here have been carried out using Dataset II only.

Combined data of all parental species During the analysis of the dataset comprising all species it was discovered that hybrid populations interfered with the estimation of the number of clusters. This finding is consistent with observations made using simulated data, for which STRUCTURE had problems detecting the true number of clusters when datasets were generated using a contact zone population model [52]. Therefore it was decided to remove the hybrid populations to enable reliable identification of putative admixed parental individuals. For the dataset including the hybrids three clusters ($K = 3$) were identified as the most likely number, while the value for $K = 4$ was still elevated but much lower (Figure 3.11, **a**). Analysis of the data after the hybrids had been removed clearly identified four clusters (Figure 3.11, **b**), which corresponded to the four species (Figure 3.12), and therefore represent the topmost level of structuring. For the analysis using the parental populations only, *R. clementinae* (C1, C2) and *R. roxieanum* (X1, X3) do not show admixture, but in *R. aganniphum* and *R. phaeochrysum* certain individuals are grouped into two clusters simultaneously[7]: The individuals identified as admixed by STRUCTURE

[7]For a graph showing cluster membership of single individuals see Appendix C, Figure C.2, page 183.

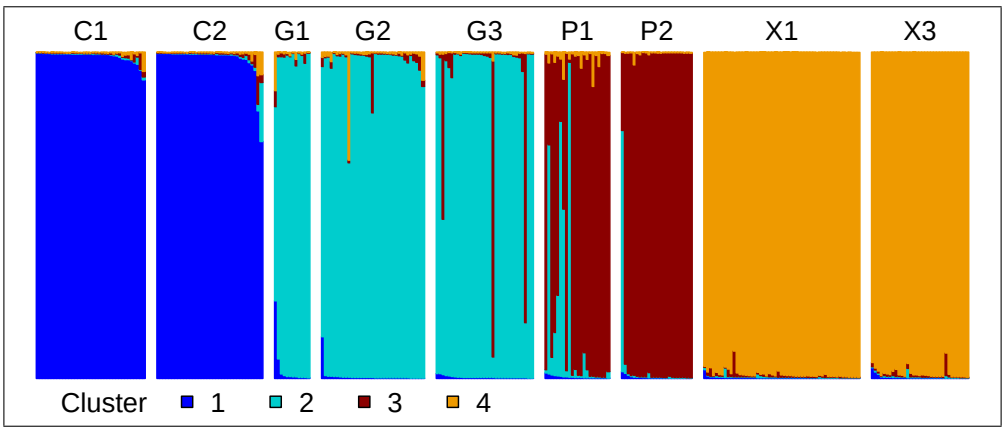

Figure 3.12: Cluster membership of populations for all parental species (K = 4); obtained with STRUCTURE; hybrids were excluded. Cluster/Species: **1** *R. clementinae*; **2** *R. aganniphum*; **3** *R. phaeochrysum*; **4** *R. roxieanum*.

agree with the results of the MDS analysis, discussed in section 3.2, page 105; in P1 the admixed individuals correspond to var. *agglutinatum*, and for populations G3 and P2 the same individuals show cluster membership for the other cluster respectively. The sympatric population of *R. aganniphum* (G2), however, does not show patterns of admixture.

Analysing parent species individually For the combined data STRUCTURE identified only the topmost clustering. This behaviour was mentioned in chapter 2, subsection 2.3.7 (page 84). Therefore, subsets comprising accessions from single species only were analysed seperately. For these mostly the number of detected clusters agreed with the number of populations sampled ($K = 2$, Figure 3.13), except for *R. aganniphum*, where three populations had been included, also $K = 2$ showed the highest ΔK. However, $K = 3$ has the next highest value (Figure 3.13, **b**), and shows clear grouping by population (Figure 3.17, **d**).

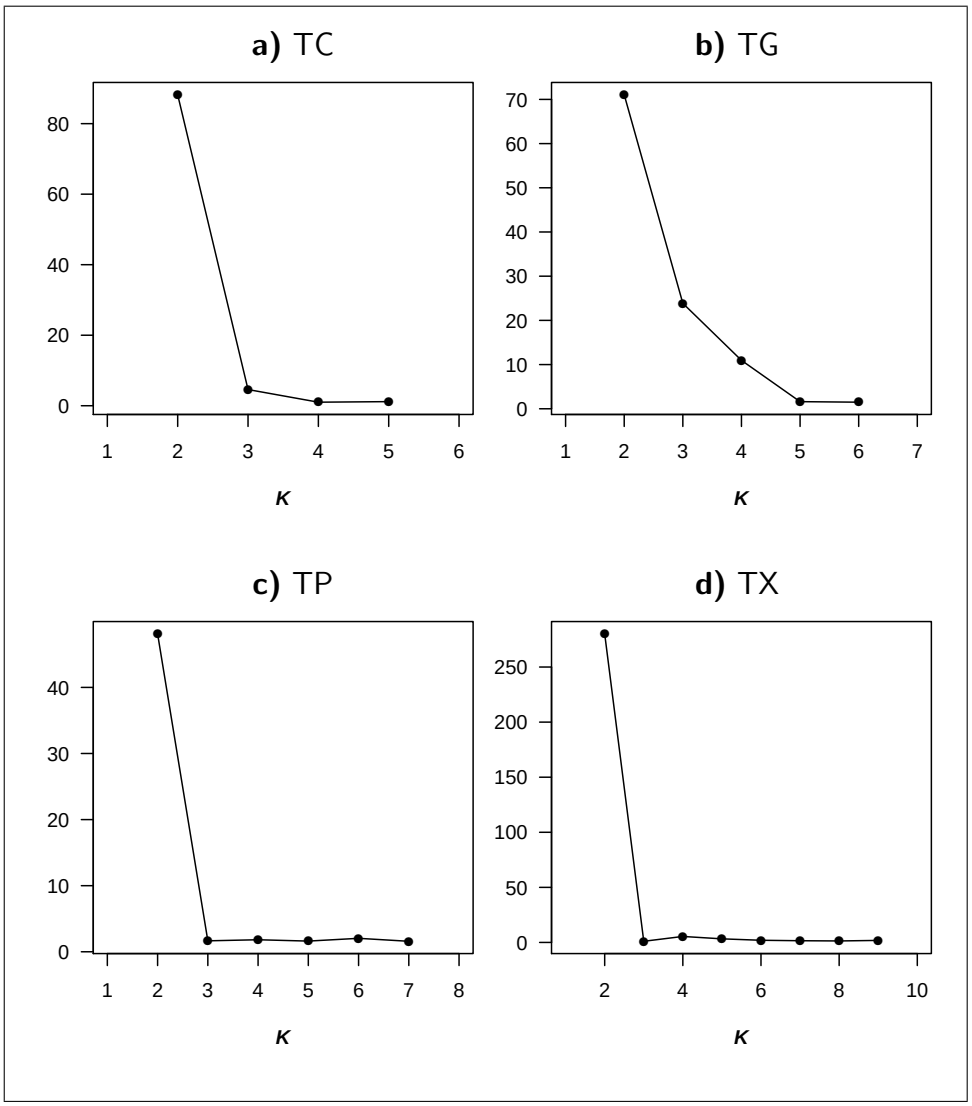

Figure 3.13: ΔK-values (individual species); ΔK as a function of the number of assumed clusters K. Calculated from PrD returned by STRUCTURE performing 15 runs for each K. Species: **a)** *R. clementinae*; **b)** *R. aganniphum*; **c)** *R. phaeochrysum*; **d)** *R. roxieanum*.

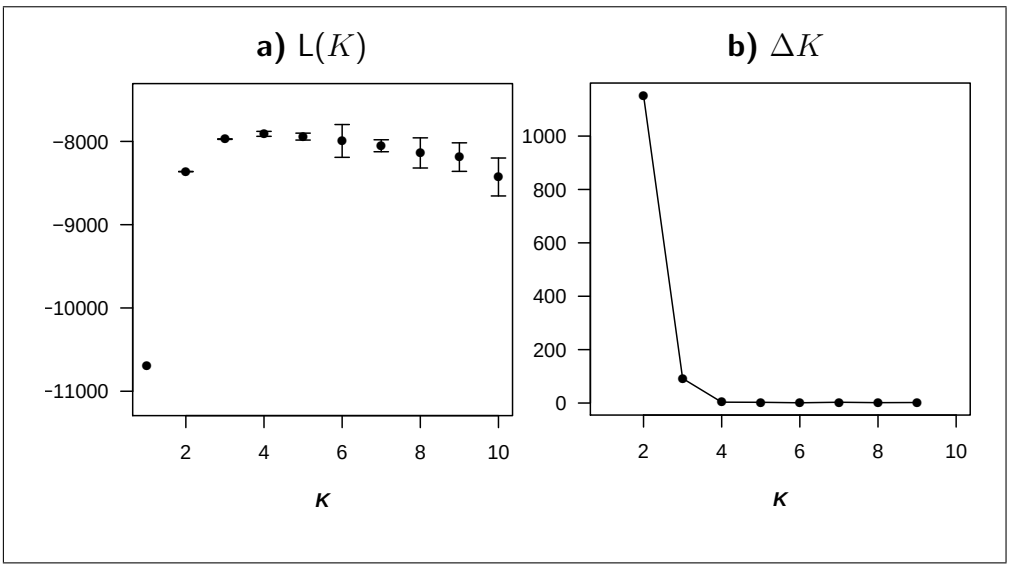

Figure 3.14: L(K) and ΔK-values (*R. clementinae* and *R. roxieanum*); a) L(K) (= PrD), as returned by STRUCTURE, and **b)** ΔK; both as a function of the number of assumed clusters K. Obtained by performing 15 runs for each K.

R. clementinae* and *R. roxieanum The analysis of the combined subsets of *R. clementinae* and *R. roxieanum* including *R. roxieanum* var. *cucullatum* clearly identified $K = 2$ as the number of clusters for the topmost level (Figure 3.14). The increment of L(K) from $K = 2$ to $K = 3$ was only very minor, and $K = 3$ would always detect the substructure in *R. clementinae* (two populations, as for the separate subset, not shown). As only admixed individual, `TXC09` is confirmed again (Figure 3.15, **a**, 'XC1'), but other individuals of *R. roxieanum* var. *cucullatum* (TXC) are placed in the same cluster as the sympatric *R. roxieanum* population (X1), Figure 3.15, **c**. Overall *R. clementinae* and *R. roxieanum* are placed in clear clusters by species, and separate analysis of species subsets places the two populations of each species in distinct clusters (Figure 3.15).

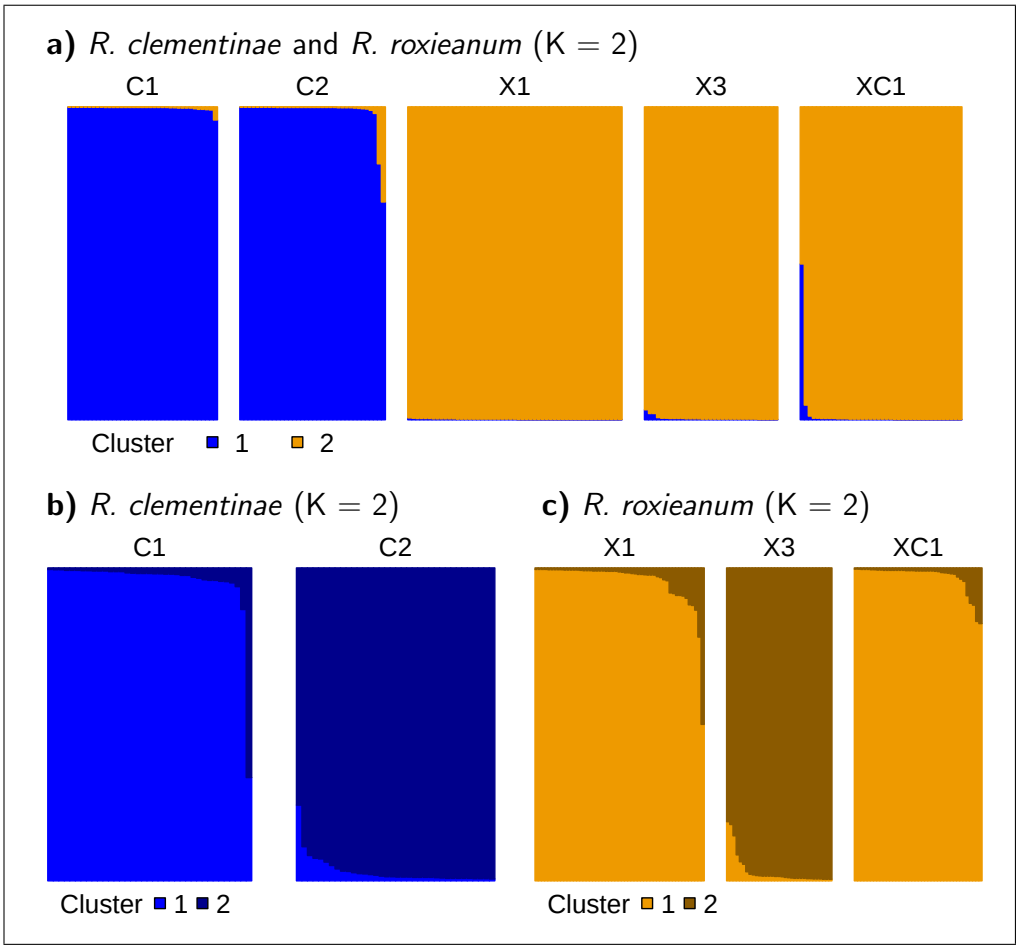

Figure 3.15: Cluster membership of populations for *R. clementinae* and *R. roxieanum* as estimated with the program STRUCTURE. **a)** Dataset of both species analysed combined, showing clustering by species; **b, c)** Dataset analysed for each species separately, revealing population structure.

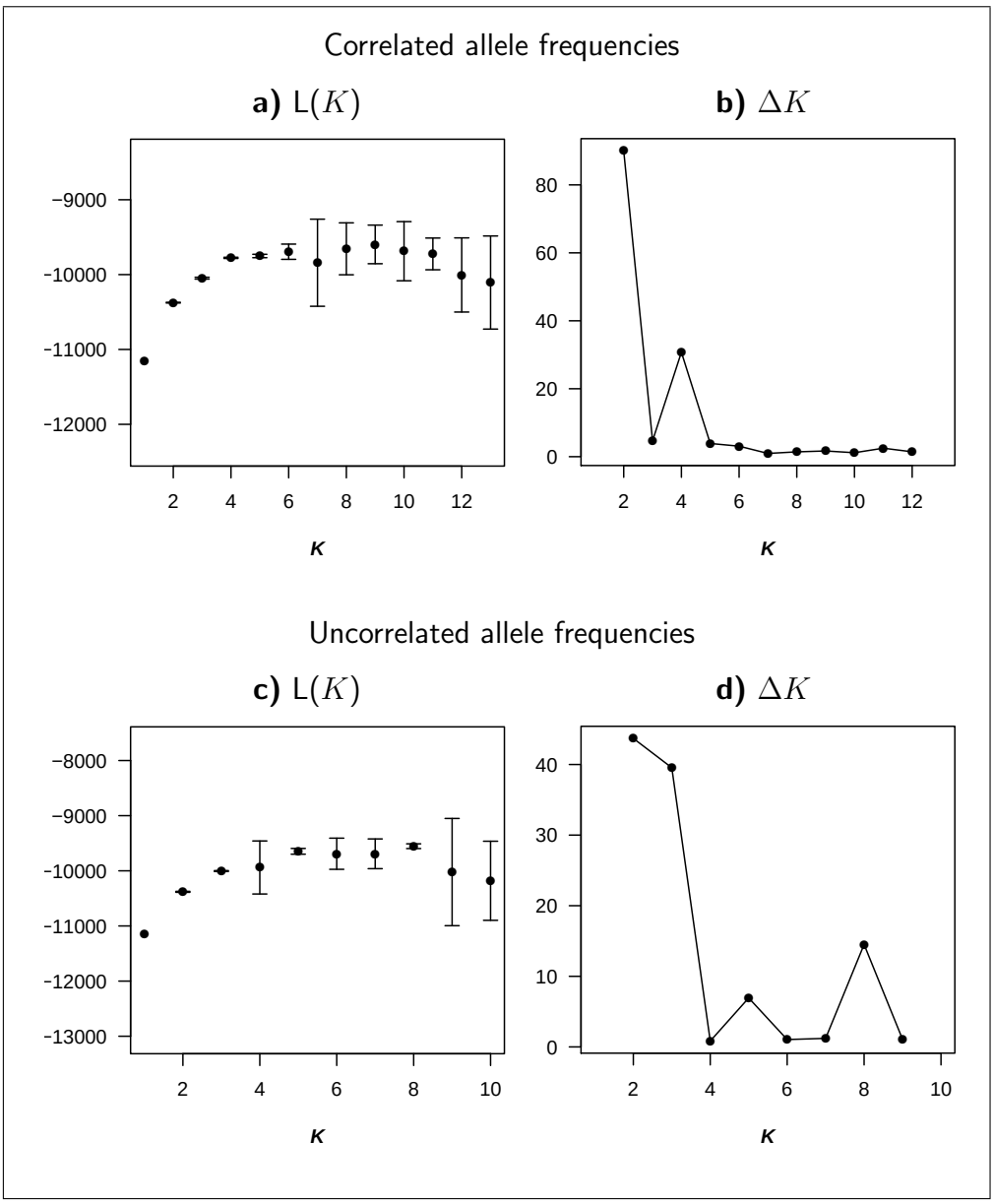

Figure 3.16: L(K) and ΔK-values (*R. aganniphum* and *R. phaeochrysum*); a, b) Allele frequencies (α's) correlated; **c, d)** Allele frequencies (α's) uncorrelated, λ set to 0.47. **a, c)** L(K) as returned by STRUCTURE, and **b, d)** ΔK; both as a function of the number of assumed clusters K. (Obtained by performing 10 runs for each K)

R. aganniphum* and *R. phaeochrysum As before, STRUCTURE identifies the species as clusters, if the dataset is not subdivided further ($K = 2$, Figure 3.16, **b**). The potential hybrid populations GH2a and GH2b nearly exclusively group then with *R. aganniphum* (Figure 3.17, **a**). Subdivision, again, identifies the correct number of populations in the parental species (Figure 3.13, **b** & **c**; Figure 3.17, **b**–**d**).

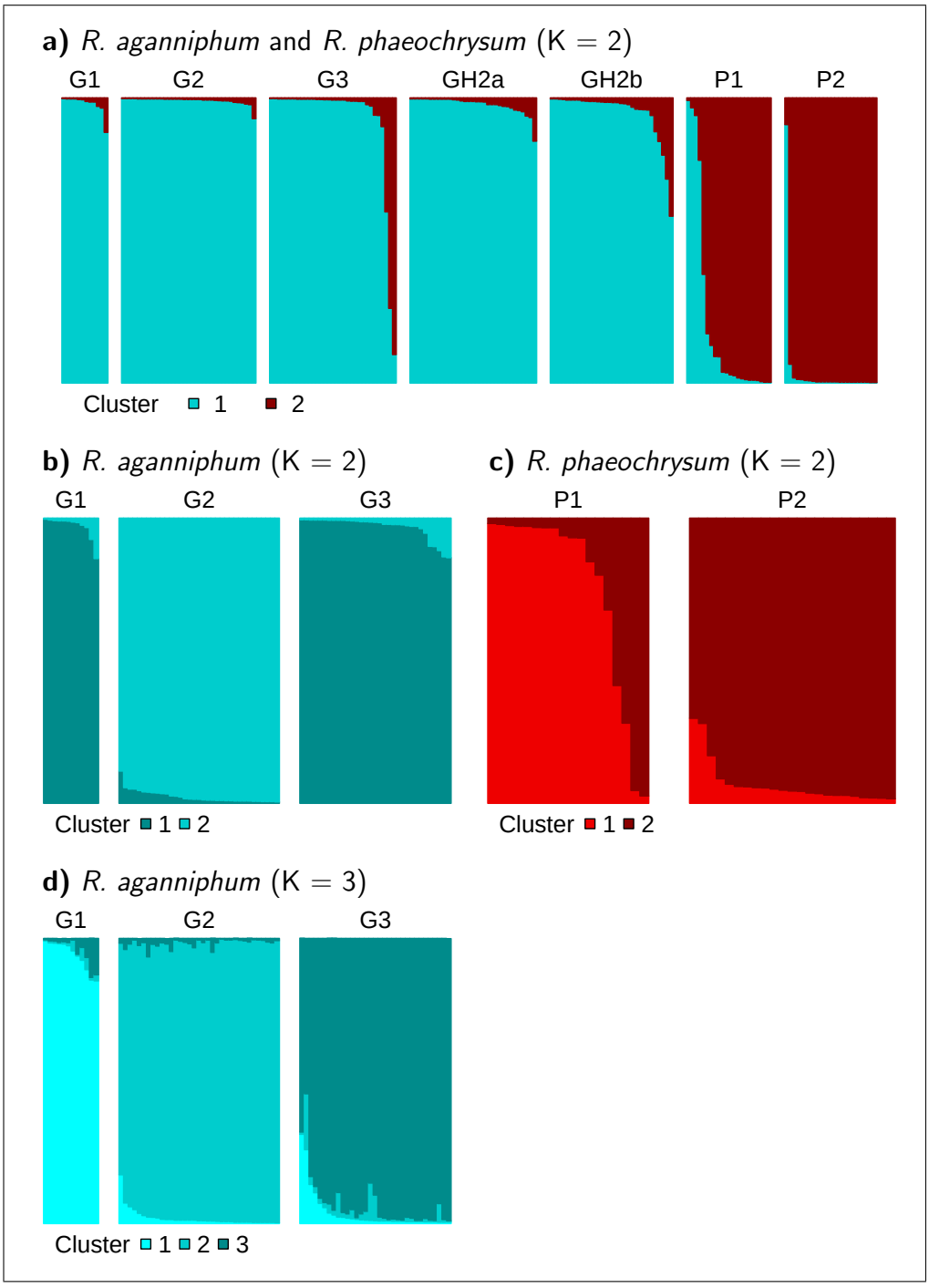

Figure 3.17: Cluster membership *R. aganniphum* and *R. phaeochrysum*, as estimated with the program STRUCTURE. **a)** Dataset of both species combined, showing clustering by species; **b–d)** Dataset analysed for each species separately, revealing population structure.

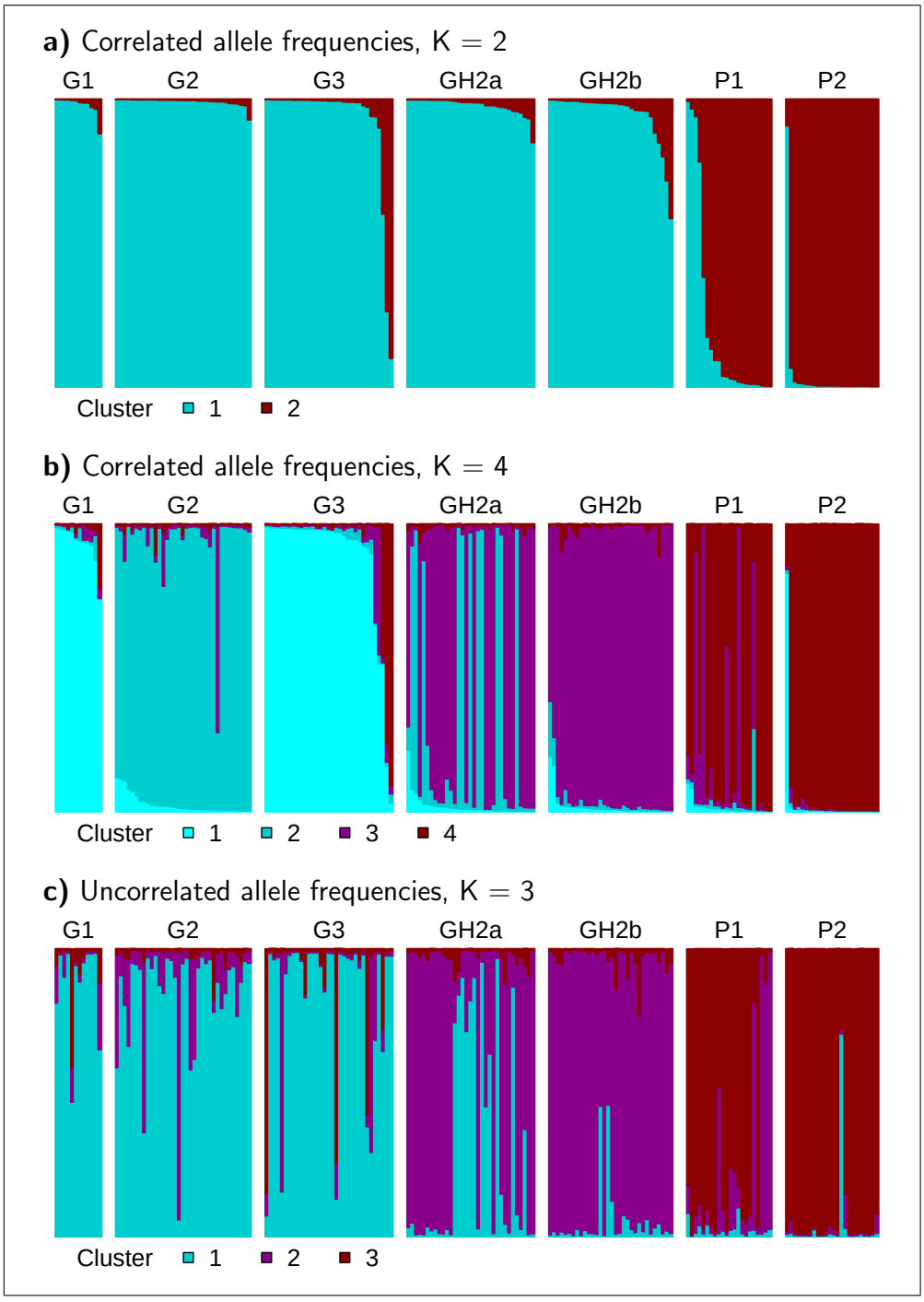

Figure 3.18: Cluster membership *R. aganniphum* × *phaeochrysum*, as estimated with the program STRUCTURE. **a)** When clusters represent species GH2a/b share a cluster with *R. aganniphum*; **b, c)** With larger K, GH2b forms its own cluster, and GH2a clusters with GH2b and G2 (*R. aganniphum*).

R. aganniphum* × *phaeochrysum The values for ΔK obtained for the combined datasets of the two species including hybrids (Figure 3.11), and the subset of *R. aganniphum* (Figure 3.13, **b**) showed that ΔK often does not have the highest value at the true number of clusters. This seems to be especially the case when the increment in L(K) is more gradual, and the largest slope-change occurs from $K = 1$ to $K = 2$. In this case the next larger value, after $K = 2$, frequently represents a better estimate for the most likely number of clusters. For the dataset of *R. aganniphum* and *R. phaeochrysum* including potential hybrids this is $K = 4$ (Figure 3.16, **b**). With this number of clusters the allopatric populations of *R. aganniphum* become separated into different clusters, while *R. phaeochrysum* retains is species cluster, and unexpectedly GH2b and several individuals of GH2a are joined into the fourth cluster (Figure 3.18, **b**). To see if this was due to inappropriate parameter settings that could influence the grouping, λ[8] was estimated from the data ($\lambda = 0.47$), and the setting for uncorrelated allele frequencies (α's) was used. Disregarding the biased $K = 2$, the most likely K was in this case $K = 3$ (Figure 3.16, **d**), but still identified the potential hybrids as a separate cluster (Figure 3.18, **c**), while joining all *R. aganniphum* populations into one cluster.

To test whether this behaviour might be expected for F1 hybrids, 50 hybrids were simulated from the sympatric parental populations[9], after removing admixed parents, and analysed combined with the populations of *R. aganniphum* and *R. phaeochrysum*. For this dataset $K = 3$ was identified as most likely number

[8] λ is the allele frequency prior parameter [136]

[9] The simulation of hybrids is explained in chapter 2, subsection 2.3.8, page 87.

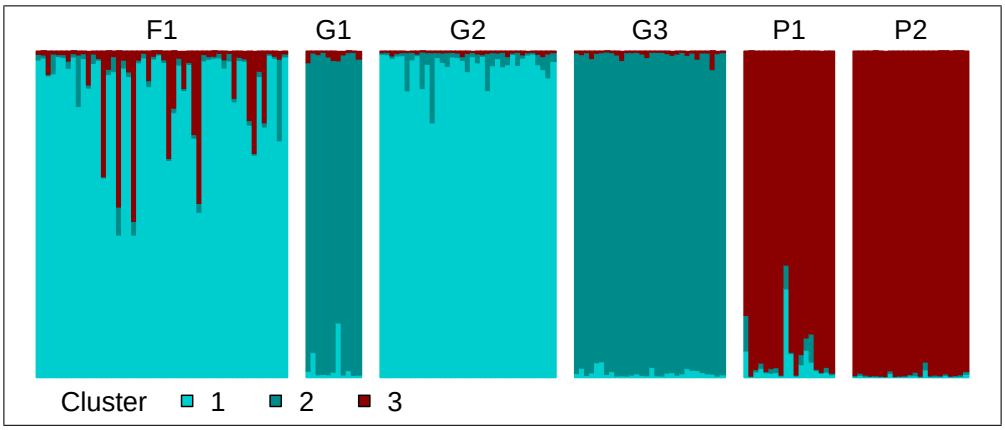

Figure 3.19: Cluster membership (RM-F1s) for $K = 3$, as estimated with the program STRUCTURE. All parental populations clearly belong to one respective species/population cluster; simulated F1 hybrids show low degree of admixture.

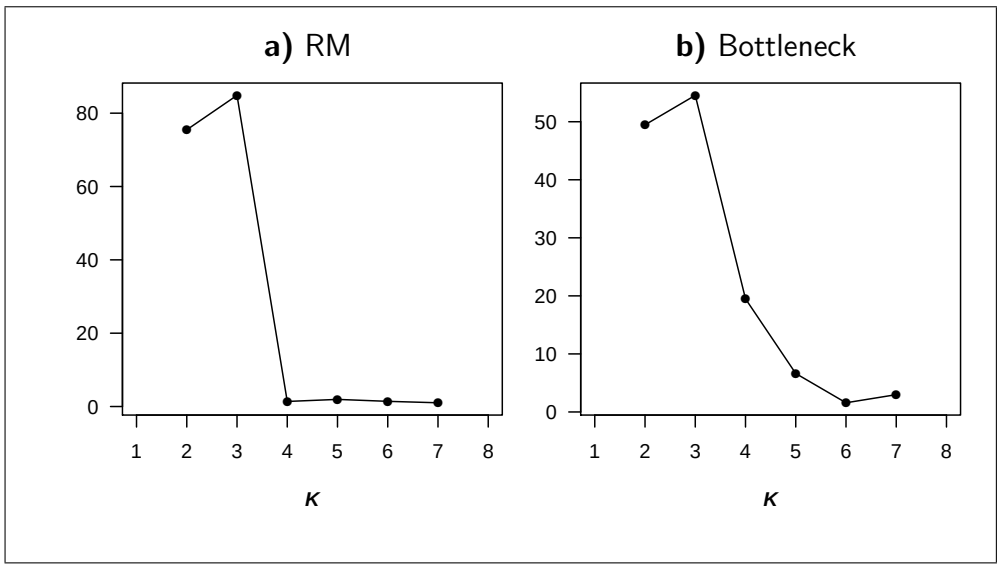

Figure 3.20: ΔK-values for a dataset comprising parental populations of *R. aganniphum* and *R. phaeochrysum* (admixed individuals were removed), and: **a)** 50 simulated RM-F1s; or **b)** 50 simulated bottleneck F1s. (For simulation of different hybrid classes see chapter 2, subsection 2.3.8, page 87)

of cluster, even higher than the value for the biased $K = 2$ (Figure 3.20, **a**). Using $K = 3$, STRUCTURE clustered the species separately, and differentiated allopatric from sympatric *R. aganniphum* populations. However, although many simulated F1s grouped exclusively with *R. aganniphum*, probably because the frequency of dominant alleles at many loci was considerably higher in this species opposed to *R. phaeochrysum*, many F1s showed the admixture pattern expected, and did not form their own cluster (Figure 3.19). Assuming $K = 4$, as used for the original data, did not change the general pattern, only the clustering became inconsistent over runs: often identifying the third *R. aganniphum* population as different cluster; sometimes separating the two *R. phaeochrysum* populations; and sometimes attributing a cluster to the F1s, though always heavily admixed with *R. aganniphum*, never forming a clear cluster on their own[10].

Because STRUCTURE forms clusters by 'fitting' them to Hardy-Weinberg equilibrium, one possible cause for the formation of a separate cluster is a significant deviation from this assumption. The simulations with the RM-F1s showed that normal hybridisation should not have a significant deviation as consequence. Therefore similar datasets were simulated assuming a bottleneck during the formation of the hybrids[9]: equally 50 individuals were simulated, and combined into one dataset with the parental species. However, as the bottleneck simulation is a process hugely governed by chance, an AMOVA was performed

[10]Examples are included in Appendix C, Figure C.3, page 185

on the datasets, and five sets selected in which the simulated hybrids showed comparable levels of variation with regards to the actual hybrid populations. Although the most likely number of clusters was again $K = 3$, which only separated the parental species and grouped the hybrids with *R. aganniphum* (not shown), the ΔK for $K = 4$ was considerably higher (Figure 3.20, **b**). A similar behaviour was observed for the data comprising the real populations GH2a and GH2b, only in this case $K = 2$ was much higher than the real $K = 4$ (Figure 3.16, **b**). Furthermore, these simulated hybrids form their own cluster, showing low levels of admixture with *R. aganniphum*, but no contribution of *R. phaeochrysum* is detected (Figure 3.21). These results provide evidence that a bottleneck scenario during F1 hybrid formation can explain observed patterns in populations GH2a and GH2b.

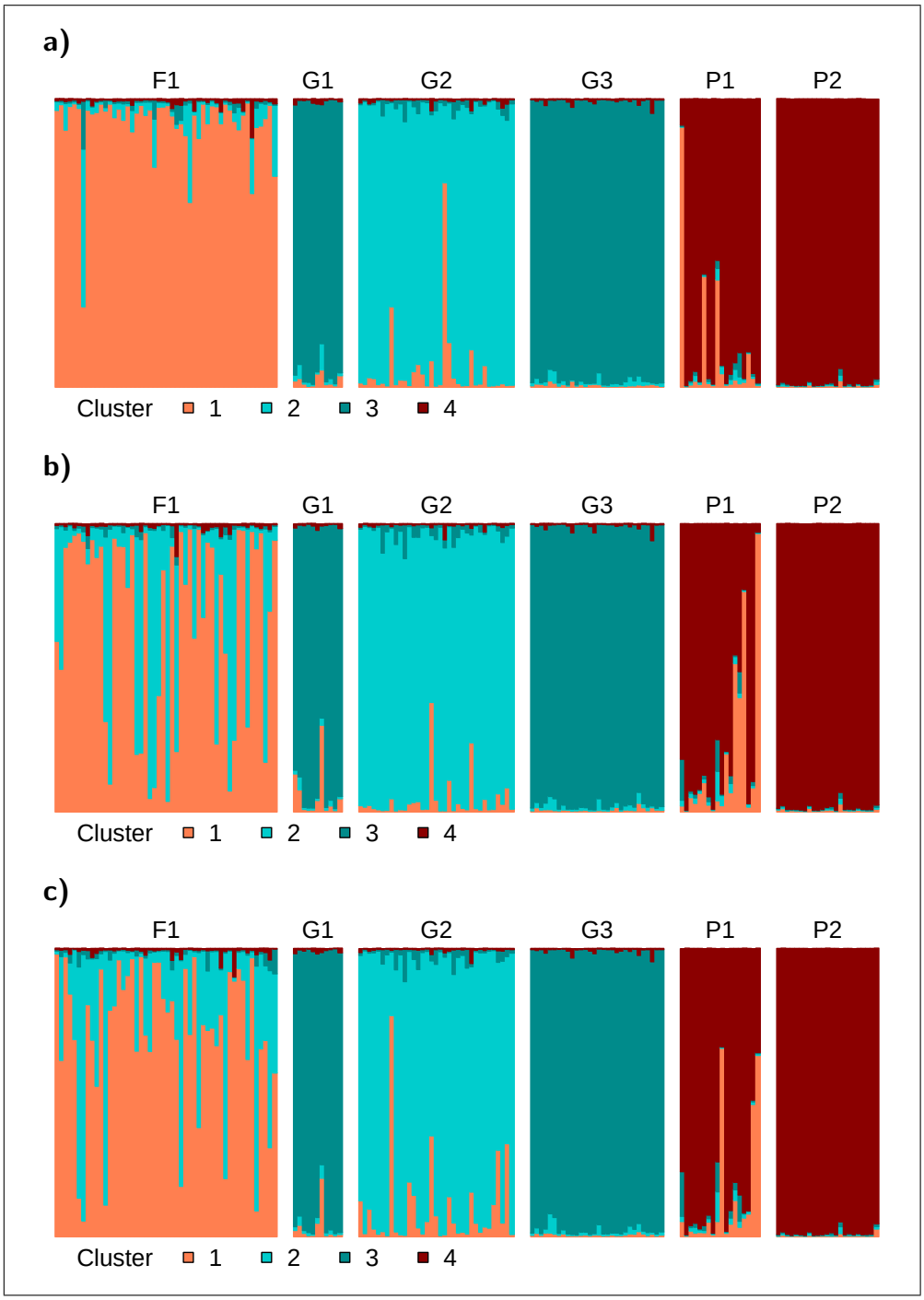

Figure 3.21: Cluster membership (bottleneck F1s) for $K = 4$, as estimated with the program STRUCTURE. Three of five simulated datasets are shown. F1s form their own cluster, which can be very homogeneous **a**, or show a certain degree of admixture with G2 **b, c**.

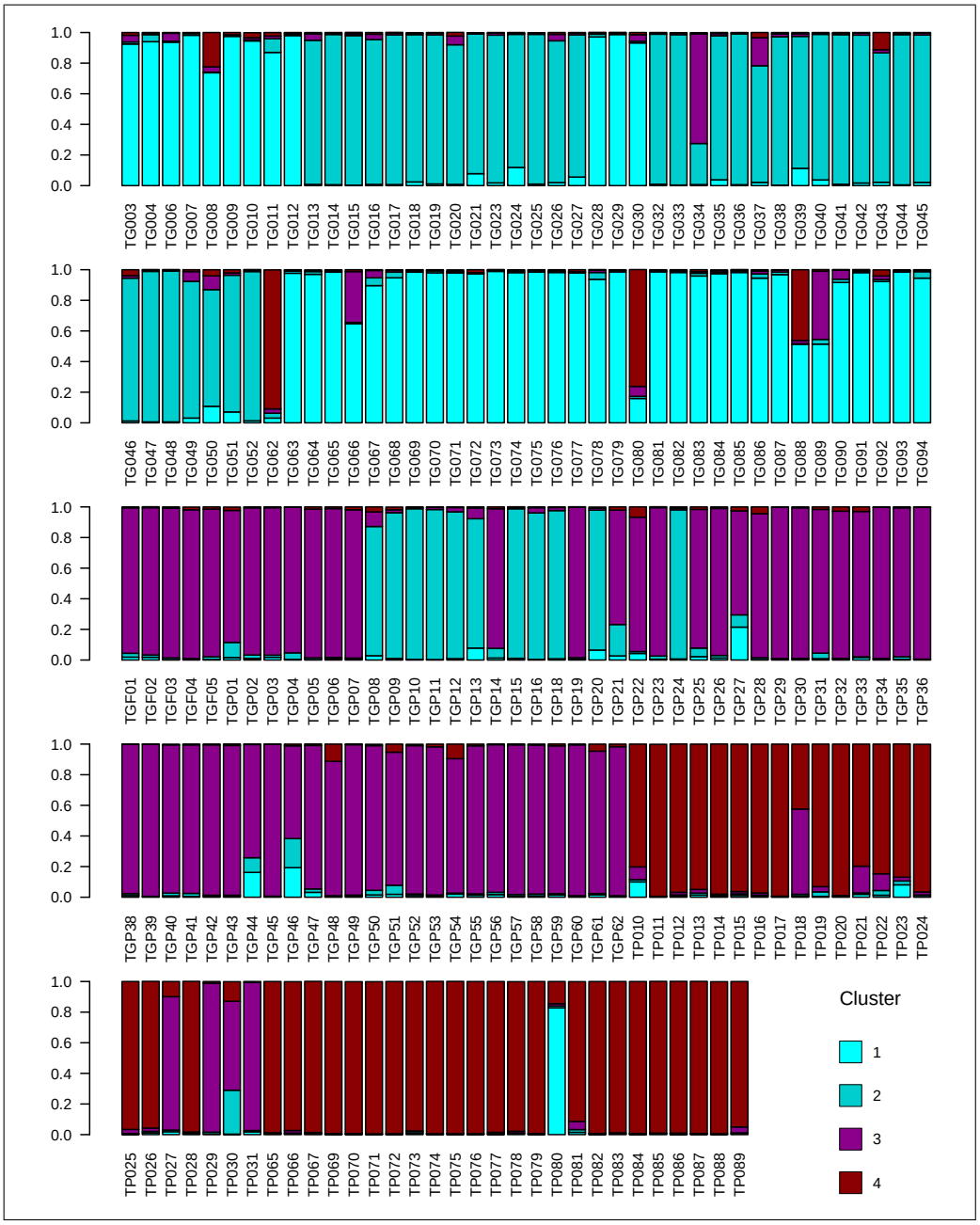

Figure 3.22: Cluster membership *R. aganniphum* and *R. phaeochrysum* (individuals) as estimated with the program STRUCTURE. The clusters correspond to following species/populations: **1** allopatric *R. aganniphum* (G1, G3); **2** sympatric *R. aganniphum* (G2); **3** potential hybrid (GH2b); **4** *R. phaeochrysum*.

Table 3.12: Summary of skewed loci. Number of loci with counts significantly (p <0.05[a]) deviating from the expectation assuming different types of hybrid class for the population. In total 105 loci were tested, and the percentage given refers to that number.

	Loci in GH2a			Loci in GH2b		
Assumed hybrid class	Skewed	%	Δ_c	Skewed	%	Δ_c
BC2 (*R. aganniphum*)	26	24.76	0.157	40	38.10	0.018
BC1 (*R. aganniphum*)	15	14.29	0.118	32	30.48	0.040
F1	32	30.48	0.000	48	45.71	0.000
BC1 (*R. phaeochrysum*)	30	28.57	-0.190	46	43.81	-0.051
BC2 (*R. phaeochrysum*)	46	43.81	-0.405	52	49.52	-0.105
Any[b]	7[c]	6.67		22[c]	22.86	

[a] After correcting for multiple testing with the method of Benjamini and Hochberg [17].
[b] Loci with significant deviation in the three hybrid classes F1, BC1-TG and BC1-TP (p for all three assumed hybrid classes <0.05).
[c] 1 locus of GH2a is significant on the 0.01 significance level for any cross type, and 14 loci for GH2b (p <0.01).

3.7 Deviation of marker counts (skew)

Datasets used for analysis To further investigate possible causes of the bottleneck observed in the *R. aganniphum* × *phaeochrysum* hybrid populations it would have been desirable to estimate co-transmission of markers. However, linkage disequilibrium is very difficult to test with dominant markers, especially in hybrids which should all be heterozygotic for the parental alleles. Therefore the deviation from the expectation of the marker counts in hybrids was used to provide further insights into the apparent bottleneck behaviour of the *R. aganniphum* × *phaeochrysum* hybrid populations. After a significant deviation of the counts was found with analyses using Dataset II[11], it was deemed sensible to include more markers, informative for this case. All analyses discussed here were finally carried out using Dataset III (section 3.1, page 100), comprising 105 loci, and only accessions of *R. aganniphum*, *R. phaeochrysum* and hybrids. To assess marker skew at individual loci, observed marker counts at each locus were compared to the expected counts obtained from simulations. The expected count was estimated by taking the median count of 10,000 simulated hybrid populations (for details see chapter 2, subsection 2.3.9, page 88). To assess behaviour of the method, and test for significant deviation of counts obtained by chance,

[11] For datasets see section 3.1, subsection 3.1.1, page 97.

25 datasets were simulated, each comprising: two parental populations; one population of RM-hybrids; and one of bottlenecked hybrids; all four populations with 40 individuals each (see chapter 2, subsection 2.3.9, page 88). Using these simulated datasets the same analysis was carried out as with the real data (Dataset III).

Deviating loci of simulated datasets As expected the analysis of the simulated RM-F1 hybrids showed no significant deviation from expected counts for any of the loci (Figure 3.23, **a**)[12]. The simulated bottleneck hybrids, however, showed significant deviation for a considerable number of loci (Figure 3.23, **b**). Mostly the dominant allele is over-represented (positive Δ_{obs}), which could be expected for dominant markers. The over-representation of the recessive allele is only likely to be detected at loci for which the populatons have a large frequency difference. Furthermore, the bottleneck has to happen in the population exhibiting a high frequency of the dominant allele. In case both populations have a relatively high frequency of the dominant allele, the over-represented recessive allele will mostly be masked by the dominant allele from the other parent. And if both populations have a low frequency of the dominant allele, there is effectively no over-representation of the recessive allele.

Number of deviating loci for GH2a and GH2b The analysis of the real hybrid populations (GH2a, GH2b) showed that in both a large number (often >30 of 105) of loci exhibited marker counts deviating significantly from simulated expectations. Furthermore, a significant deviation was detected regardless of which type of hybrid class was assumed (Table 3.12). Considerably more loci in population GH2b showed deviation from the genome-wide expectation than in population GH2a. After using a highly conservative estimate for the number of skewed loci, requiring sigificant deviation in F1 and BC1 classes, 6.67% of loci in GH2a showed deviation, and 22.86% in GH2b (Table 3.12). Dominant alleles were affetced symmetrically, equally often showing under-representation as over-representation (Figure 3.24). As the simulations showed, this is not expected for a bottleneck scenario were allele contribution from one parent only is randomly affected. The observed pattern could be explained if always one allele is preferentially transmitted to hybrid offspring. For example if at a given locus the recessive allele is selectively prefered, the dominant allele would be under-represented in the hybrid, even in the case of high frequency of the dominant

[12]Results of all 25 simulations can be found in Appendix F, page 263.

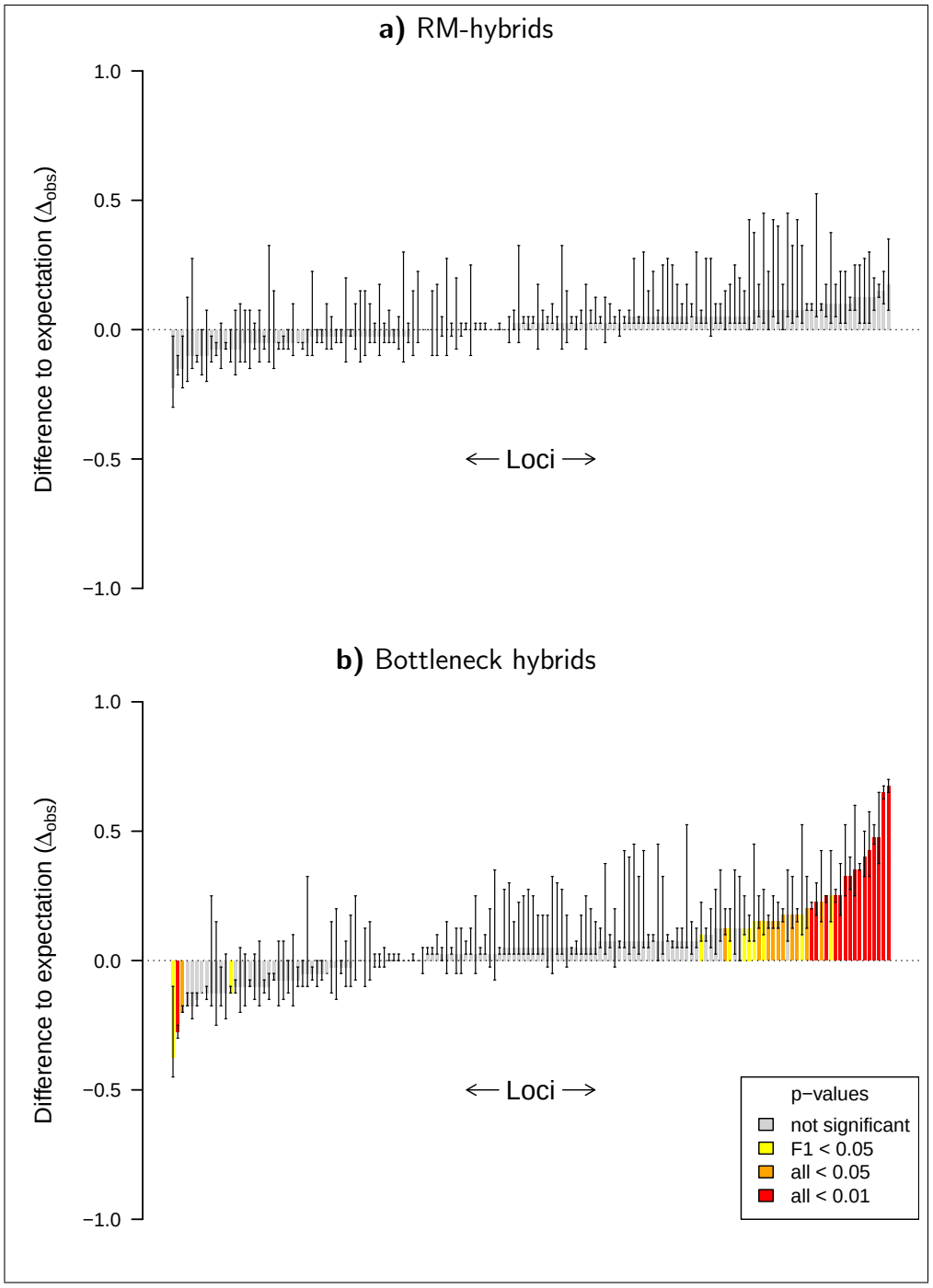

Figure 3.23: Δ_{obs}**.** Representative results of 25 simulated datasets showing expected deviation in RM and bottleneck hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parental species. **a)** no skew was observed assuming RM-hybrids; **b)** bottleneck hybrids show significant skew.

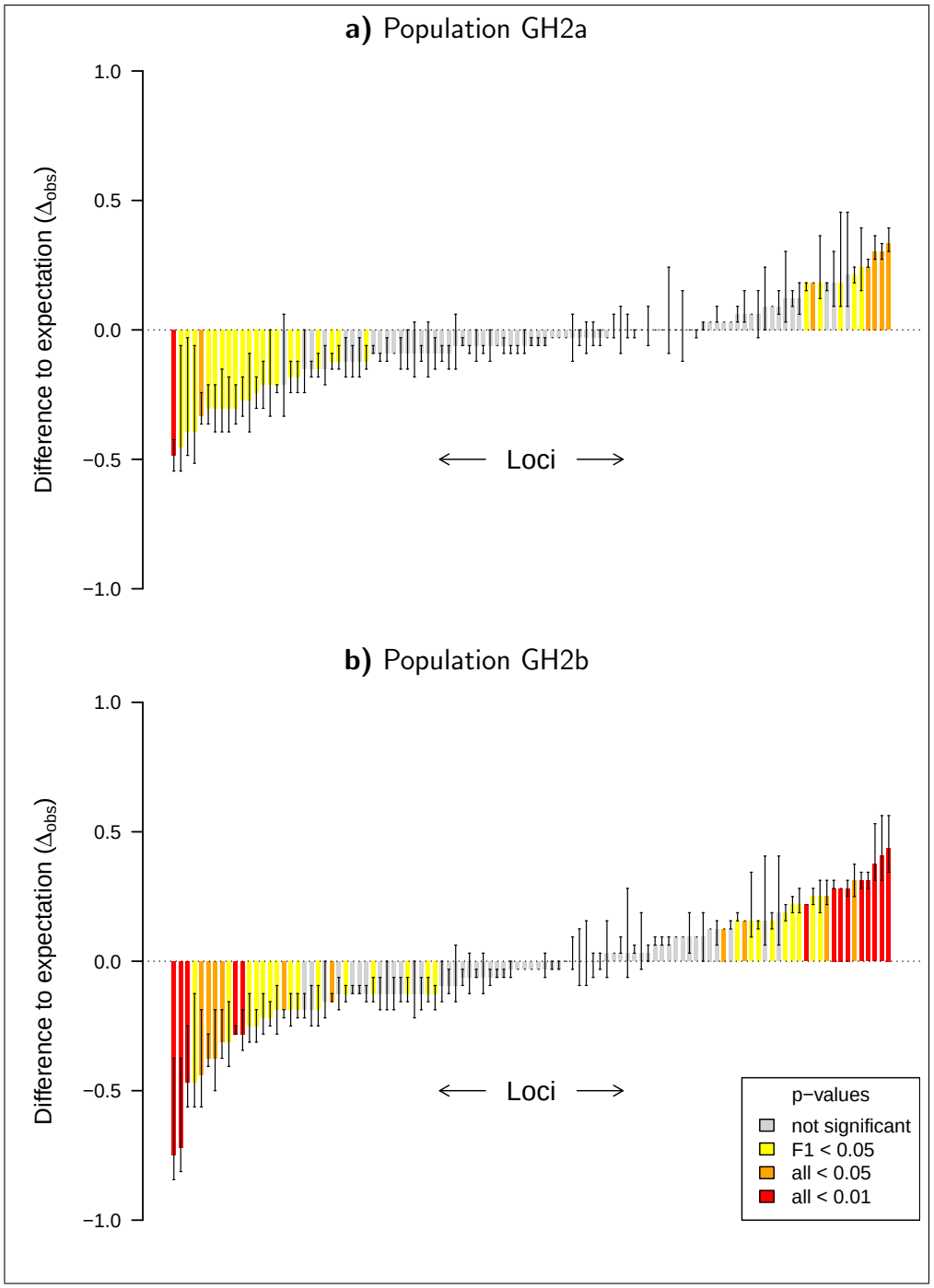

Figure 3.24: Δ_{obs}**.** Deviation of observed marker counts from simulated expectations for the two *R. aganniphum* × *phaeochrysum* hybrid populations; estimated by simulating 10,000 hybrid populations of the same size. 'Error bars' show the range of Δ_{obs} including BC1s to both parental species.

allele in both parents. Averaging over many loci this should favour dominant and recessive alleles equally, leading to a more or less symmetric distribution. This pattern therefore suggests that allele transmission from both parents into the hybrid is affected. Furthermore, it implies that selective forces act differently for several loci, and not that one major effect caused the bottleneck.

Deviation of marker count and locus F_{ST} Different transmission distortion for certain loci is expected if certain alleles are selectively advantageous, or if genetic incompatibilities interfere with viability of the hybrids [146]. Genetic incompatibilities often seem to be correlated to species divergence [182], and can therefore be expected to be correlated to F_{ST}. For this reason it was tested wether the absolute deviation ($|\Delta_{obs}|$) of the marker counts (strength of transmission distortion) is correlated to the parental F_{ST}. In both populations (GH2a, GH2b) a significant correlation of $|\Delta_{obs}|$ with the corresponding locus F_{ST} was detected ($r = 0.497$ and 0.581 respectively, both $p < 0.001$; Table 3.13; Figure 3.25, **a, b**).

Although the 25 simulated datasets show that a significant correlation of absolute Δ_{obs} and F_{ST} is expected for hybrids obtained assuming random-mating of the parents ($r = 0.062$–0.307; Figure 3.25, **c**)[13], this correlation is considerably weaker than observed in the data, and, furthermore, the observed significant marker skew is not expected in this case (Figure 3.23, **a**)[13]. The analysis of the simulated bottleneck hybrids, for which a significant marker skew was established (Figure 3.23, **b**), showed no correlation of $|\Delta_{obs}|$ and F_{ST} (Figure 3.25, **d**). This implies that a simple bottleneck scenario, leading to a major random distortion of several markers, can not explain the observed pattern adequately. Therefore, it strongly supports the hypothesis of selective forces acting on different loci simultaneously.

[13]Results of all 25 simulations can be found in Appendix F, page 263.

Table 3.13: Correlation of $|\Delta_{obs}|$ and locus F_{ST} for two potential hybrid populations; estimated with Pearson's product moment correlation coefficient.

Population	Statistic[a]	*df*	*r*	Conf int[b]	*p*-value
GH2a	5.814	103	0.497	0.365 – 1.0	<0.001
GH2b	7.251	103	0.581	0.463 – 1.0	<0.001

[a] Test statistic (z) for one-tailed (greater) t-distribution.
[b] Confidence interval of r, based on Fisher's Z transform.

Likely hybrid class of populations GH2a and GH2b The significant correlation of $|\Delta_{obs}|$ and F_{ST} shows that all potentially diagnostic markers are not transmitted into the hybrids as would be expected, if they behaved neutrally. Therefore no direct assignment of individuals to a hybrid class, based on genetic markers is possible. However, it is not too far fetched assuming unhindered transmission of these markers for later generation backcrosses. If this assumption holds to a considerable degree, the allele frequencies should converge towards the genome-wide expectation the further backcrossed the population is. Therefore, the overall detectable skew of markers should diminish in populations mostly comprising a different hybrid class from F1, when compared to simulated expectations of the respective class. To this end the difference between total deviations for each of the assumed classes (Δ_c) was calculated (for exact calculation of the statistic see chapter 2, subsection 2.3.9, page 88). Observing a lower total deviation of marker counts with respect to F1 hybrids (positive Δ_c) would be interpreted as indicating an overall better fit of the respective population to this class than F1.

For neither of the two hybrid populations (GH2a, GH2b) assuming backcrosses to *R. phaeochrysum* improves the fit ($\Delta_c = -0.190$ and -0.051 respectively, Table 3.12). Assuming GH2a comprises backcrosses to *R. aganniphum* improves the fit, which holds for both, BC1 and BC2 ($\Delta_c = 0.118$ and 0.157, respectively). For GH2b, although a better fit is obtained when assuming backcrosses to *R. aganniphum*, the improvement is only marginal and reverses from BC1 to BC2 ($\Delta_c = 0.040$ and 0.018, respectively). As there are no other studies available to compare the values to, only the much higher value obtained for GH2a, this improvement in fit for GH2b is not deemed significant. Therefore, the data is best explained by assuming GH2a comprises mostly backcrosses to *R. aganniphum*, and GH2b F1 hybrids.

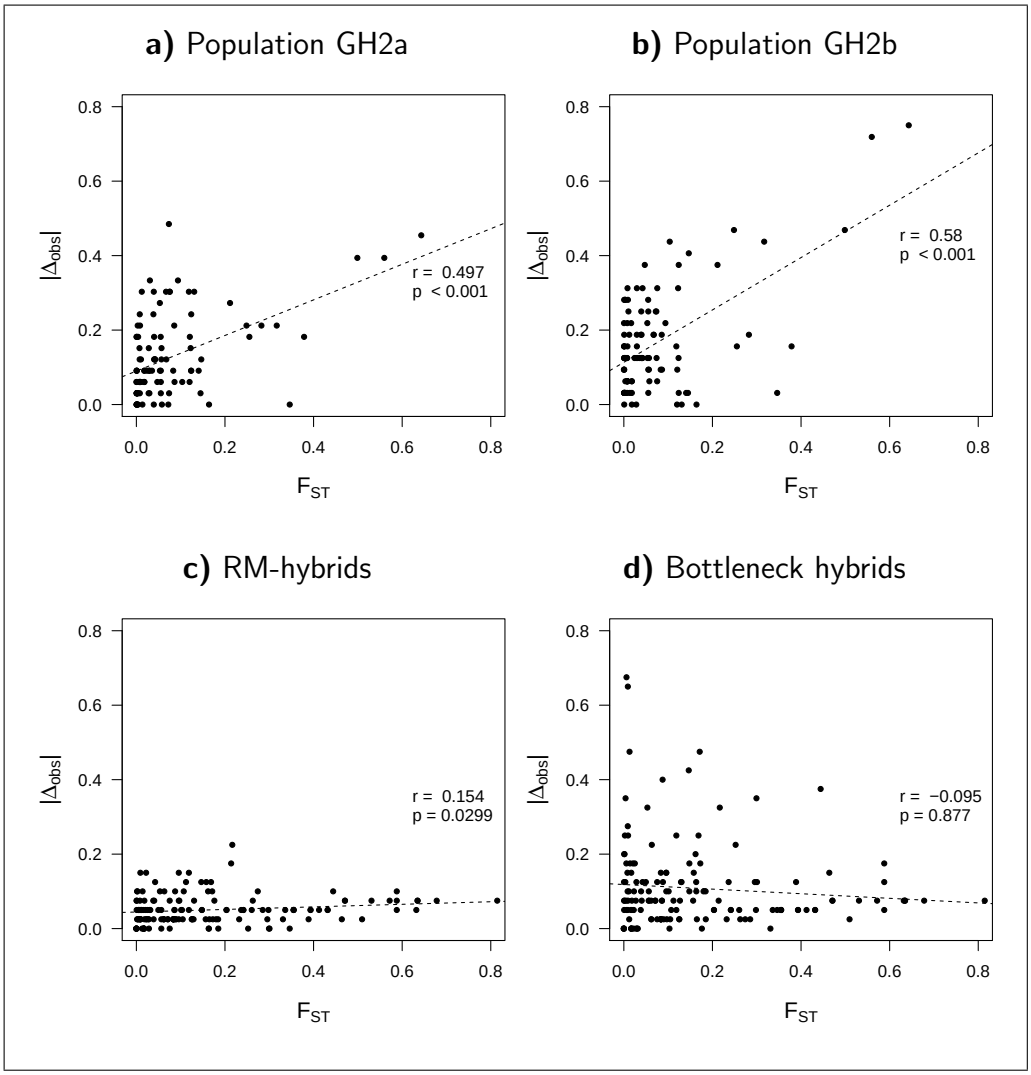

Figure 3.25: Correlation of F_{ST} and $|\Delta_{obs}|$. The deviation from expectation of marker counts in hybrids as a function of the parental locus F_{ST}. **c & d** are representative results of 25 simulations. **a, b)** significant correlation in *R. aganniphum* × *phaeochrysum* populations; **c)** in simulated hybrids assuming random mating $|\Delta_{obs}|$ is weakly, but significantly, correlated with F_{ST}; **d)** bottleneck hybrids show no significant correlation of $|\Delta_{obs}|$ and F_{ST}.

3.8 Leaf wax composition

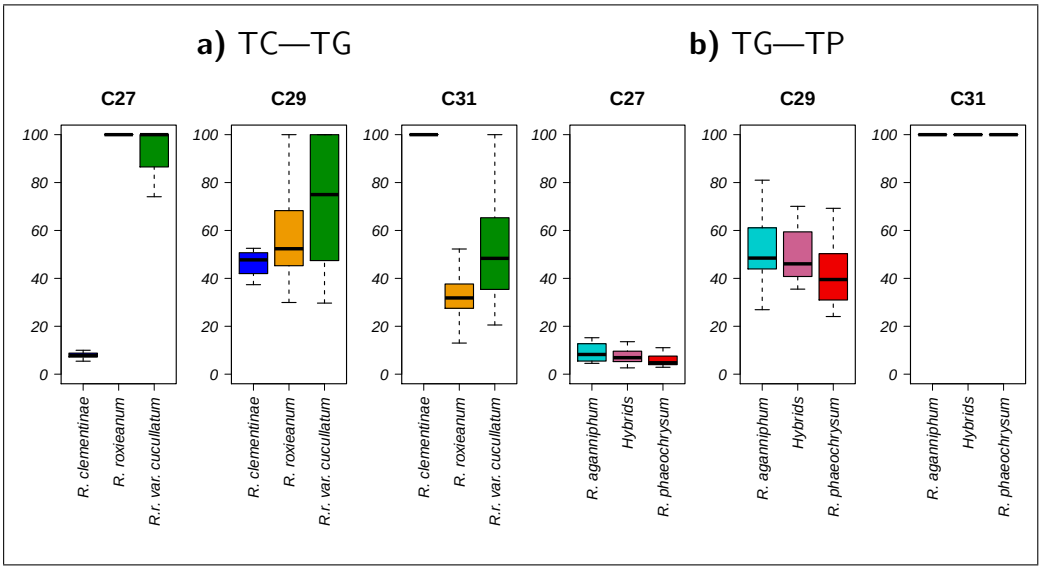

Figure 3.26: Leaf wax composition of the four parental species and hybrids. The size of each wax fraction, as given by the area of the fraction, was normalised to the largest fraction in the trace ($Area_{fraction\,x}/Area_{fraction\,max}$), and is shown as percentage (y-axes). **a)** *R. clementinae* and *R. roxieanum* have different maxima (C31, C27, respectively), *R. roxieanum* var. *cucullatum* shows elevated levels of C_{29} and C_{31}. **b)** *R. agganiphum* and *R. phaeochrysum* have a maximum for the same fraction (C31), hybrids therefore show no deviation.

Leaf wax profiles were obtained for 115 accessions. From the raw area readings obtained, only the fractions containing the *n*-alkanes $C_{27}H_{56}$ (C27), $C_{29}H_{60}$ (C29), and $C_{31}H_{64}$ (C31) were used. The most abundant fraction (maximum) in each trace was identified, and all fractions in the trace normalised so as to have the maximum at 100 (Appendix D, Table D.3, page 213). The values obtained showed that maxima were consistent within species: *R. agganiphum*, *R. clementinae* and *R. phaeochrysum* having C_{31} as most abundant fraction, and *R. roxieanum* C_{27}; the only exception were two *R. phaeochrysum* individuals, which showed a maximum at C_{27}[14]. This polymorphism for *R. phaeochrysum* has already been reported by Chadwick et al. [31], and in their paper the pattern seemed to be stable for two varieties, var. *phaeochrysum*-C_{31}, and var. *levistratum*-C_{27}, however, var. *agglutinatum* was polymorphic. Chadwick et al. [31] tested only very few samples, and therefore it is possible that *R. phaeochrysum* is generally polymorphic for leaf wax composition. In this case, allele frequency differences due to population structure might lead to fixation within certain populations or geographically separated varieties. The hypothesis of general

[14]For data obtained for individual accessions please see Appendix D, Table D.3, page 213.

polymorphism of leaf wax composition in this species is supported by the fact that the two individuals in this thesis showing an unusual maximum are not attributable to var. *agglutinatum*, but var. *phaeochrysum*. As *R. aganniphum* and *R. phaeochrysum* had both a maximum for C_{31}, leaf wax composition was not informative regarding the hybrids.

R. clementinae and *R. roxieanum* var. *roxieanum* showed distinct maxima for the leaf wax composition, and *R. roxieanum* var. *cucullatum* mostly conformed to the leaf wax maximum found in *R. roxieanum* var. *roxieanum* (C_{27}). However, several individuals of var. *cucullatum* showed elevated levels of C_{31} and C_{29}[15]. Therefore, overall, var. *cucullatum* has significantly higher levels of these two fractions than var. *roxieanum* (Figure 3.26).

The comparison of the leaf wax composition with genetic distance, using the coordinates obtained with MDS[16] as a representation of genetic relatedness, shows a weak pattern: the only confirmed hybrid between the two species (`TXC09`) is clearly intermediate, both genetically and in leaf wax composition; other individuals of *R. roxieanum* showing elevated levels of *R. clementinae* characteristic leaf waxes tend to be more differentiated from the allopatric *R. roxieanum* population, but not closer to *R. clementinae* (Figure 3.27, bottom-right).

[15]For data obtained for individual accessions please see Appendix D, Table D.3, page 213.

[16]The same coordinates, obtained for this species pair (TC-TX) with the MDS analysis described in section 3.2, page 104, were used.

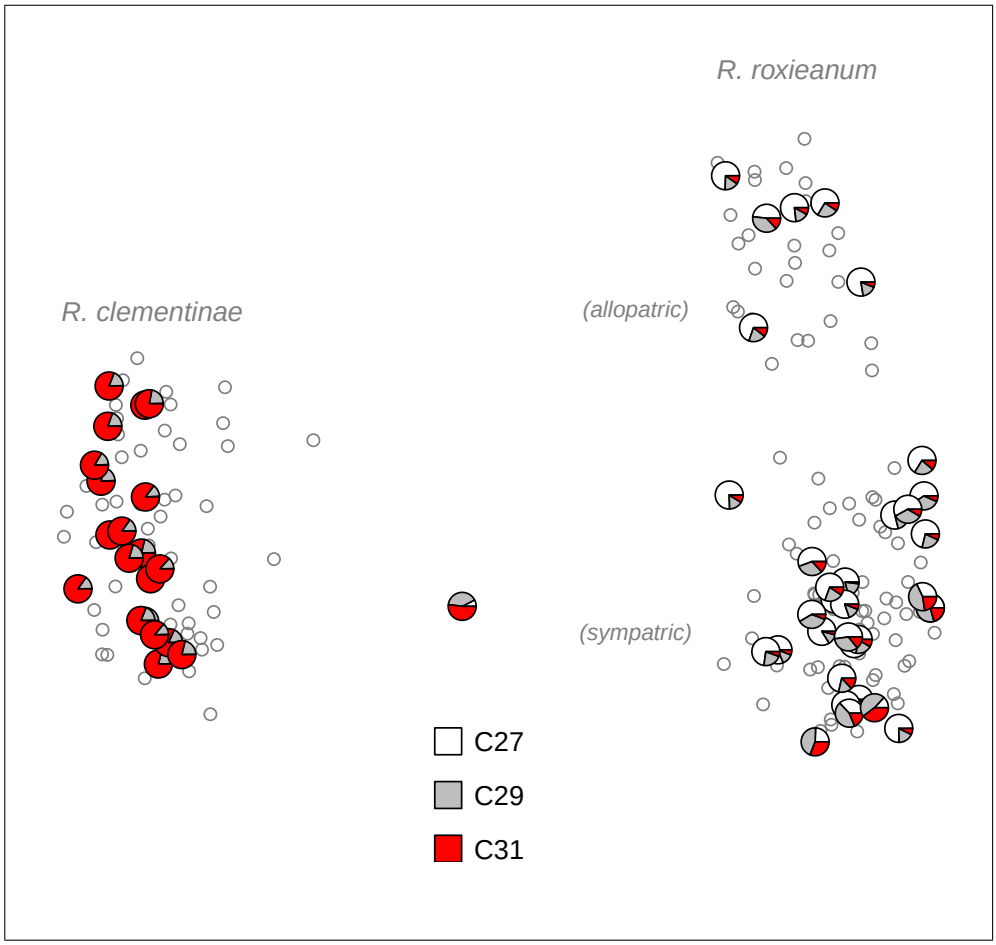

Figure 3.27: MDS overlaid with leaf wax data. Relative contribution of three different wax fractions (C27, C29, C31) to the overall leaf wax composition in 55 individuals of *R. clementinae* and *R. roxieanum* (pies; values of leaf wax fractions are depicted squared, see chapter 2, subsection 2.3.10, page 95). Coordinates of individuals were taken from a MDS analysis performed on 199 individuals shown earlier (Figure 3.3, page 104), which included the individuals used here. *R. clementinae* has a maximum for C_{31}, *R. roxieanum* for C_{27}; individual `TXC09` (center) is clearly intermediate; some sympatric *R. roxieanum* individuals, which tend to be further from the allopatric population, show elevated levels of C_{31} (bottom-right).

Chapter 4

Discussion

4.1 Population differentiation

For all species investigated most of the variation is found within populations (80–90%, Table E.3, page 224). This is in accordance with expectations for woody perennials [73]. Furthermore, populations of all species are significantly differentiated, and show comparable levels of differentiation (F_{ST} = 0.051–0.087, Table 3.11, page 119, 337 loci). The observed F_{ST}-values can be interpreted as moderate differentiation, as values lie between 0.05 and 0.15 ([74], page 119), and conform to differentiation reported for other woody taxa ($F_{ST} \sim$ 0.027–0.260, [44], supplementary data 1). Because geographic distances between all conspecific populations were approximately equal ($\sim$ 100 km, chapter 1, subsection 1.6.6, page 47), they were not considered to influence the F_{ST} estimates differently. Comparisons with published F_{ST}-values are generally complicated by the fact that distances between populations are often not given. However, the values are comparable with those of some wind pollinated species (*Populus*, F_{ST} = 0.117 [82])[1], and seem low considering the patchy distribution of the populations, which are often restricted to mountain ranges and inselbergs. Higher values would have been expected if pollen flow was affected considerably, and therefore this provides evidence that these factors do not present major barriers for gene flow. This agrees with a study of *Vitex*, in which pollen mediated geneflow by insects was not considerably hindered by the Yangtze river [187]. The similarity in the magnitude of population differentiation observed for all species probably also reflects their general similarity in reproductive biology.

[1] Note: The study included samples from all over Europe, and this F_{ST} value is considered high for a wind pollinated species. However, this comparison seems adequate as no major barriers should impact wind mediated pollen flow in Europe.

Parameters such as height, pollination type and lifespan have been shown to be correlated with F_{ST} [44], and the species seem to be comparable in these. Furthermore, they occupy similar altitudinal ranges, and, lacking significant floral trait differences, are likely to depend on the same insect pollinators. However, population differentiation seems to be affected by the abundance of populations throughout the distribution range. This is suggested by the significantly higher population differentiation of the least abundant species, *R. clementinae*, as opposed to the most abundant species, *R. phaeochrysum* ($\Phi_{ST} = 0.198$ and 0.111, respectively; Table 3.7, page 113). However, the population sampling for this study was not extensive, and a better population coverage would be needed to confirm this pattern.

4.2 Species differentiation

The species are very closely related, as evidenced by the absence of fixed differences for nearly all species pairs. Nonetheless, all species are significantly differentiated from each other (Φ_{TG} between 0.193 and 0.367, Table 3.8, page 115). However, although *R. clementinae* is slightly more differentiated from the other three species ($\Phi_{TG} = 0.334$–0.367), no significantly closest species pair can be identified, because the confidence intervals place them all in the same magnitude (Table 3.8, page 115). Therefore, to adequately represent species relationships a tetrahedral structure as in Figure 3.2 (page 103) is needed, disqualifying a phylogenetic approach based on dichotomies. This agrees with results from chloroplast data: no differences between the species were found using several gene regions (subsection 2.2.2, page 59), and chloroplast data also fails to resolve relationships within subgenus Hymenanthes [118]. This rather complicated pattern of species relationships could be caused by two mechanisms. As already suggested to explain resolution loss in the chloroplast data [118], the species could be in, or emerging from, a radiation event. In this case they would not have had enough time to accumulate species specific mutations, and would still show a considerable amount of shared, ancestral, polymorphism. On the other hand, ongoing hybridisation might keep the species from diverging, and destroy phylogenetic signal. These two scenarios are not mutually exclusive and both are likely to contribute to the observed patterns. This could also explain incongruencies in gene trees reported in *Rhododendron* [47]. Additionally, that the species do not conform to a tree-like structure at present suggests that, even in the evolutionary future, species relationships in the complex will never be

representable by dichotomies. As the species already seem formed, even if they diverge further, fixed shared markers between species will result from the random process of drift, and not represent more recent ancestry. The phylogenies are then either unresolved [83], or gene trees seem to resolve in a random manner [84].

Species integrity in sympatric populations One hypothesis was that sympatric populations could share considerable amount of polymorphism due to hybridisation. Combined with strong barriers to pollen flow between populations of the same species, this might lead to less differentiation between sympatric populations of different species. As already discussed above, barriers to geneflow between populations of the same species are not as strong as the distribution patterns of populations might suggest. Nonetheless, the sympatric *R. phaeochrysum* population (P1) is considerably less differentiated from *R. aganniphum* than the allopatric population (P2), although marginally not significant (Φ_{PT} P1-G1 = 0.293, lower confidence P2-G1 = 0.291; P1-G2 = 0.286, lower confidence P2-G2 = 0.281; P1-G3 = 0.316, lower confidence P2-G3 = 0.281; Table 3.8, page 115). This suggests a certain amount of geneflow across species barriers. Further evidence for occasional introgression between *R. aganniphum* and *R. phaeochrysum* is provided by the four admixed individuals in the populations at Dàxuě Shan (`TG062, TG080, TG088` and `TP080`; Figure 3.22, page 132). However, no sympatric population pair belonging to different species is less differentiated than populations belonging to the same species (Φ_{PT} C1-X1 = 0.493, P1-G2 = 0.286, P2-G3 = 0.388; Table 3.8, page 115; Φ_{ST} = 0.111–0.198; Table 3.7, page 113). Therefore there is no evidence that geneflow between sympatric populations of different species is affecting species integrity of any of the populations.

4.3 Varieties in subsection *Taliensia*

A major objective of the project was to investigate the potential hybrid status of described varieties of the species.

R. roxieanum* var. *cucullatum Because cases of stable F1 populations in *Rhododendron* had been reported before [119, 185], the homogeneous morphology of *R. roxieanum* var. *cucullatum* suggested that this could represent another such case. The AFLP marker analysis shows that this is certainly not the case. One F1 individual was identified which clearly clusters in the middle of the two

sympatric populations of *R. clementinae* and *R. roxieanum* (TXC09, Figure 3.3, page 104). Therefore the two species clearly do hybridise, but early generation hybrids seem to be very rare. Other individuals of *R. roxieanum* var. *cucullatum* cluster in all analysis with the sympatric population of *R. roxieanum* (chapter 3, 3.2, 3.3 and 3.6). Although the sympatric *R. roxieanum* var. *roxieanum* and var. *cucullatum* are significantly differentiated from each other ($\Phi_{PT} = 0.028$, p <0.001, Table 3.9, page 116), the value is extremely low, and might be caused by the fact that most individuals of the sympatric parental population were sampled further down the mountain in 2.5 km distance. However, due to the lack of fixed differences, the available markers would not be able to distinguish BC2, or later generation backcross individuals, from the sympatric population. Furthermore, the results from the leaf wax analysis show significantly elevated levels of a *R. clementinae* characteristic *n*-alkane in var. *cucullatum* (Figure 3.26, page 140). Patterns of hybridisation and introgression can follow substantially different patterns: F1s may be readily formed, but later backcrossing scarce [119]; or F1 hybrids may occur rarely, however, when formed, initiate large numbers of backcrosses [2, 145]. Hence the occasional occurrence of F1 hybrids could suffice to produce backcrosses, and *R. roxieanum* var. *cucullatum* might be a remnant of earlier introgression. However, only the leaf wax data hints to this scenario, and a more extensive marker approach would be needed to test this hypothesis. Therefore there is no good reason to abandon the variety status of *R. roxieanum* var. *cucullatum*.

R. phaeochrysum* var. *agglutinatum Morphologically individuals in population GH2b conform to *R. phaeochrysum* var. *agglutinatum*. Although the indumentum shows slight variation in this population, individuals prior to the analysis identified as var. *agglutinatum* (TP027, TP029, TP030, TP031) cluster with this population, confirmed with the MDS analysis as well as groupings identified by STRUCTURE (Figure 3.4, page 105; and Figure 3.22, page 132; respectively). The NJ clustering shows that individuals of GH2b are intermediate between *R. aganniphum* and *R. phaeochrysum* (Figure 3.8, page 109), providing evidence for the hybrid status of this population. However, the detected marker skew complicates assignment to a hybrid category, as neutrality of the markers is generally crucial for assignment tests[2]. Based on the homogeneous

[2]This apparently is rarely addressed in studies using assignment tests, and neutrality assumed, as no publications were found dealing with the issue. (Probably because in most cases only a low impact is expected, if the ratio of non-neutral markers to neutral ones is very low)

morphology, and differentiation from the parents, the possibility of a hybrid taxon should be considered. In this case, selection for certain combinations of genes might lead to a significant skew at certain loci. As the selection should be considerable to produce the marker skew quickly, probably genes involved in ecological traits or reproductive isolation should be primary candidates. In an other case, if the taxon resulted from initially very few early generation hybrids being reproductively isolated from the parents, a bottleneck effect could occur, also leading to the deviation of the marker frequencies in the newly formed taxon. The results from the simulations showed that a bottleneck, caused through contribution of only a few individuals of one parental population, leads to comparable skew in markers (Figure 3.23, page 135). However, under both these scenarios no correlation of allele frequency deviation at loci with their respective differentiation in the parents is expected, as observed for the hybrids investigated here. Furthermore, the newly formed taxon should be reproductively isolated from the sympatric parental species, which is not the case, as evidenced by the many intermediates found in the other hybrid population. The next best explanation for homogeneous morphology of a group of hybrids is then a population composed mostly of F1 hybrids, which has been reported previously in *Rhododendron* [119, 185]. Additionally, deviation of hybrid allele frequencies from genome-wide expectations due to parental effects should primarily manifest itself when the parental species cross initially, i.e. in the F1 hybrids. Excluding the possibility of a hybrid taxon, but allowing for backcrosses, population GH2b conforms equally well to F1 hybrids, or backcrosses to *R. aganniphum* ($\Delta_{F1} = 0$, $\Delta_{BC1} = 0.040$, $\Delta_{BC2} = 0.018$; Table 3.12, page 133). Taking into account the homogeneous morphology of the individuals, and the correlation of the marker skew with parental differentiation, F1 hybrids are the best explanation. Therefore it is proposed that *R. phaeochrysum* var. *agglutinatum* is most likely the F1 hybrid of the cross *R. aganniphum* × *phaeochrysum*.

R. aganniphum* var. *flavorufum In contrast to population GH2b, population GH2a, in which individuals exhibit a more diverse indumentum morphology, better conforms to later generation backcrosses to *R. aganniphum* ($\Delta_{F1} = 0$, $\Delta_{BC1} = 0.118$, $\Delta_{BC2} = 0.157$; Table 3.12, page 133). Furthermore, clustering places the population as intermediate between GH2b and the sympatric *R. aganniphum* population (Figure 3.5, page 106), or several individuals of the population very close to *R. aganniphum* (Figure 3.8, page 109). Therefore the composition of this population (GH2a) is best explained as mostly consisting

of backcrosses to *R. aganniphum*. This is in accordance with morphology, as a more diverse composition is expected in a hybrid swarm. Individuals of *R. aganniphum* var. *flavorufum* are members of this swarm, and are therefore considered to represent one possible morphology of backcross individuals between the F1 hybrids and *R. aganniphum*.

4.4 Barriers to geneflow in *Rhododendron*

Skewed transmission of markers into hybrid offspring To employ commonly used assignment tests to determine likely hybrid classes of individuals, diagnostic markers, or in other words loci for which different alleles have been fixed or have high frequency differences in the species investigated, are crucial. In the case of the *R. aganniphum* × *phaeochrysum* hybrid populations, for which a class assignment would have been desirable, no fixed differences were found and only a small number of loci with high frequency differences identified. This might have enabled the use of assignment algorithms if a relatively high power loss would have been tolerated. However, one further criterion to use these techniques was not met: the markers have to broadly follow the predictions made by theory, hence have to be transmitted according to expectations. The STRUCTURE analysis showed an unusual behaviour of the *R. aganniphum* × *phaeochrysum* hybrids by clustering them separately (Figure 3.18, page 127) while this is not expected for normal F1 hybrids Figure 3.19, page 128). Simulations showed that one possible scenario to cause this pattern is the occurrence of a bottleneck during the formation of hybrids (Figure 3.21, page 131). Only a scenario in which few individuals from one parental species contributed to the genepool was tested. Several possible mechanisms leading to distortion of marker frequencies were introduced in chapter 2, subsection 2.3.9 (page 91). All of them are based on incompatibilities occurring at different stages during the development. Pre-pollination mechanisms seem unlikely, or less important in the case of *R. aganniphum* × *phaeochrysum*, as the floral morphology is exceedingly similar, and flowering times largely overlap. Furthermore, interactions leading to incompatibilities at this early stage, should affect all marker frequencies in an equal manner, resulting in marker distortions, but no correlation between Δ_{obs} and parental locus F_{ST} (Figure 3.25, page 139, **d**). A significant correlation of these two measures in both hybrid populations (Figure 3.25, page 139, **a, b**) points to a more complex nature of the incompatibilities. The observed pattern is in accordance with deviations expected under underdominance (Figure 2.1

page 94), suggesting that chromosomal inversions might be involved. To achieve underdominance in F1 hybrids, BDMIs or adaptive gene complexes are also very likely to contribute[3]. Further evidence for the likely involvement of epistatic interactions is provided by the occurrence of backcrosses only towards *R. aganniphum*. Asymmetric incompatibilities generally seem to be caused by BDMIs that affect stages where different numbers of maternal and paternal genomes interact [e.g. gametophytic(1n)-sporophytic(2n), and sporophytic(2n)-endosperm(3n)] [172]. Additionally these types of postzygotic incompatibilities seem to be quite common in *Rhododendron*; frequently interspecific crosses between more distantly related taxa lead to spontaneous abortion or suffer from lack of endosperm formation, however, viable hybrids can be obtained via embryo rescue[4] [48]. To further elucidate the nature of the asymmetric incompatibilities more research including pollination experiments to assess the potential roles of pollen competition and hybrid seed viability would be required.

Coming back to the applicability of assignment tests, if enough diverged markers are not affected, removing the fraction behaving abnormal with regard to genome-wide expectations would still allow the employment of these techniques. However, the marker skew is systematic, affecting potentially diagnostic markers extraordinarily, as shown by the significant correlation of the level of distortion with the respective parental divergence ($r = 0.497$ and 0.581, $p < 0.001$; Table 3.13, page 137). Furthermore, the case for *R. aganniphum* and *R. phaeochrysum* is very extreme: the species *only* differ markedly at loci that are skewed, therefore not a single diagnostic marker, behaving according to expectations, is available.

This also brings up the question about the evolutionary history of these species. Apparently the species are only differentiated at loci, or markers closely linked to these, that are selected against during hybrid formation. Similar patterns, where only a few loci distinguish species while the remaining genome seems to be homogenised, have been reported for example in oaks, a group in which hybrids also occur frequently [103]. In this case divergent selection for loci under question, together with recurrent geneflow at neutral loci seems to explain the pattern best [103]. That functional divergence can drive allele frequencies apart at selected loci has even been shown to occur in populations of the same species [124]. However, whether this coincides with the emergence of incompatibilities is unclear. In *Drosophila* this seems to be the case and is

[3]chapter 2, subsection 2.3.9 (page 91)

[4]Cultivation of immature seeds on medium containing nutrients, therefore mostly avoiding dependency of the developing embryo on the endosperm.

possibly a first step in the process of speciation [182]. Furthermore, different potential for certain loci or genes under selection to cross species boundaries is apparently not uncommon [12, 103, 146, 182]. Therefore the species could be able to tolerate a certain degree of geneflow while retaining species integrity. This gene exchange would lead to a low divergence for most parts of the genome, but allowing certain regions under selection to diverge, leading to the pattern observed in *R. aganniphum* and *R. phaeochrysum*. However, information for *Rhododendron* is far from extensive, and in contrast to the above mentioned examples, for which more research has been conducted, similarity at most loci due to shared ancient polymorphism, as opposed to recurrent geneflow, cannot be ruled out completely. The build-up of incompatibilities in allopatry seems possible, even without invoking divergent selection [130]. After second contact this can lead to similar patterns of differentiation, which are generally difficult to distinguish [12]. However, overall, the frequent occurrence of hybrid zones in *Rhododendron* and the divergence of *all* potentially diagnostic markers from genome-wide expectations suggests a more substantial role for hybridisation, combined with divergent selection at certain loci, in shaping the pattern.

Hybrid zone types in *Rhododendron* The potential of species in the genus *Rhododendron* to hybridise has been confirmed in several studies [110, 119, 120, 185]. Furthermore, previous research on hybrids in subgenus Hymenanthes frequently found F1 dominated hybrid zones [119, 185]. This general pattern is not contradicted by this study, as population GH2b is predominantly, if not entirely, composed of F1 hybrids. The number of F1 individuals in all mentioned studies is fairly large, indicating high fitness for this hybrid class, and in one case they even have been shown to outcompete other hybrid classes and their parents under certain environmental conditions [119]. As long as no later generation hybrids are formed, these F1 hybrids do not present a bridge for geneflow between the two parental species.

Although population GH2b seems to conform to a F1 swarm, in the other population (GH2a) only few F1s are present, and predominantly backcrosses to *R. aganniphum* are found. Hence geneflow, however unidirectional from *R. phaeochrysum* to *R. aganniphum*, is likely to occur here. If F1s are fertile, which is apparently is the case for the *R. aganniphum* × *phaeochrysum* cross, the maintenance of species barriers due to F1s has often been attributed to ecological selection [119, 145]. Because this type of barrier is heavily dependent on habitat conditions, disturbance can lead to a breakdown of this barrier. Furthermore, it is

widely believed that intermediate or unstable environments favour hybridisation, and can lead to introgression [20, 30, 85, 112]. The environmental conditions at the site where the swarm population GH2a was collected, along a road, might represent such disturbance, and hence create conditions that allow the occurrence of later generation backcrosses, which would not occur otherwise. The systematic difference between anthropogenic and natural disturbance is unclear, however, as anthropogenic influence acts additionally to the existing natural disturbances, opportunities for the occurrence of hybrid zones should have increased in the recent past. Contemporary patterns of hybridisation might therefore not reflect historical extent of introgression, but result in elevated genetic exchange.

This scenario implies that environmental conditions are largely responsible for the maintenance of species integrity of the two taxa. Although important, environment alone cannot fully explain the patterns observed in *R. aganniphum* × *phaeochrysum*. Firstly, the two reference populations of *R. phaeochrysum* and *R. aganniphum* were growing sympatrically, under apparently similar conditions as the hybrid swarm population (GH2a), but no intermediates were present. One might rightly argue that it may be difficult to judge whether the environmental conditions are really similar enough to provide the conditions needed for backcrosses to occur, hence invalidating the previous point. However, some of the individuals showed a history of admixture, providing evidence that at an earlier stage not only were hybrids present, but that geneflow did occur, which requires the establishment of intermediates as observed at the "Old Road" at Baima Shan. Nonetheless, the two populations did not show any signs of deviation from what would be considered a morphologically pure species.

Secondly, during the formation of the F1 hybrids several loci deviate from the neutral expectation of transmission. The simulations not only suggest that a bottleneck occurs during their formation, which could be explained by environmental selection for adaptive genotypes[5], but that a crossing barrier exists for certain parts of the genome, as evidenced by the correlation of the selectivity for marker combinations in the hybrids with parental divergence. Therefore it is likely that local F1 formation is influenced by the population frequencies of the cross-compatible alleles at the respective loci. Additionally, even when backcrosses are present certain parts of the genome should experience considerably lower rates of geneflow.

Hence the barrier to geneflow between these two species comprises at least

[5]See e.g. [71] for a bottleneck caused by human selection for certain characteristics; the origin of the selection should not change the patterns created.

two components: a genetical component, based on incompatibilities in the genome, which impacts formation of F1 hybrids, and amounts of geneflow in different regions of the genome; and an environmental component, based on different selective forces depending on habitat, which may impede geneflow through non-establishment of backcrosses under certain habitat conditions. These two components are highly interdependent, as the interaction of the genome composition with the environment should mostly determine the fitness of hybrid offspring [9, 85].

4.5 Conclusions and further research

Conclusions The four species investigated are mostly found at high altitudes, on mountain ranges and inselbergs. Despite this patchy distribution, populations of the same species do not show unusually high differentiation, and form well-defined genetic continua, distinguished from the other species. Although introgression was detected between *R. aganniphum* and *R. phaeochrysum*, this does not affect species integrity. *R. roxieanum* var. *cucullatum* is not an F1 hybrid, but historical introgression from *R. clementinae* into *R. roxieanum* seems probable. However, its status as variety should be retained until further evidence is available. *R. aganniphum* var. *flavorufum* is not the F1 hybrid of the cross *R. aganniphum* × *phaeochrysum*, but a later generation backcross to *R. aganniphum*. The hybrid population found on Baima Shui Shan (GH2b) is not a stabilised hybrid taxon, but rather the F1 hybrid of the cross *R. aganniphum* × *phaeochrysum*. Furthermore, this F1 hybrid corresponds to *R. phaeochrysum* var. *agglutinatum*. Although *R. aganniphum* and *R. phaeochrysum* readily hybridise, genetic incompatibilities exist for these species which leads to differential permeability of certain parts of the genome.

The value of *Rhododendron* as a study system It can no longer be doubted that in many groups of organisms, not the whole genome is subjected to the process of speciation, but rather certain genes or groups of genes [102, 168, 182]. Barriers to geneflow at these loci are crucial to enable the formation of similar adapted groups which act as species. If the groups are not separated by complete reproductive isolation, this barrier has to manifest itself during hybrid formation. Hybrid zones therefore offer unique opportunities to investigate mechanisms underlying the process of speciation. With the rapid improvement of techniques that enable a more detailed investigation of hybrids and hybrid zone

architecture, several advances in understanding selective forces and identifying genic components have been made. However, the number of extensively studied hybrid zones, and organisms, especially in plants, is still extremely low. Most research has been done on annual or short lived species, which are easier accessible to experimental design and manipulation. By far the best investigated species are *Helianthus* [90, 146] and *Iris* [112, 166]. Only relatively recently scientists have started to include long-lived woody plant species, mostly wind pollinated trees that are important members of north-western hemisphere forests. Best represented here are *Quercus* [41, 103] and *Populus* [82, 101, 153].

If we want to identify the general mechanisms underlying the process of speciation, and understand how they might differ between organismal groups, it is crucial to investigate these processes in several different species. Although many mechanisms are likely to be common to most organisms, certain differences, as for example life history traits, might pronounce different aspects in different species. Often the importance of habitat disturbance is emphasised in relation to hybridisation [3, 20, 119]. If habitat disturbances occurred spasmodically, for example heavy rain or a landslide approximately every 50 years, an annual species will have several cycles of reproduction in between these disturbances, while individuals of long lived species might experience several in a single lifetime. This could alter the outcome of a fundamentally similar response dramatically in an evolutionary context.

The genus *Rhododendron* offers several advantages to investigate hybrid zones and processes of speciation. Many species pairs, within subsections as well as between subsections are known to hybridise. Often one species hybridises with several others, allowing comparison of strength of reproductive barriers involving species with differing amount of divergence. Additionally, if genes that are involved in incompatibility in one species-pair are identified, these can be compared to a different cross involving one of these species. Finally, due to the size of the genus, members can be found in a great variety of habitats, from the tropics to extreme conditions above the treeline in the Himalayas.

Questions that could be addressed in future studies In the case of *R. roxieanum* var. *cucullatum* the leaf wax data suggest a possible introgression event from *R. clementinae*, which, however, is not supported by the genetic data. Furthermore, the leafshape of var. *cucullatum* and that of var. *roxieanum* are considerably different. A morphometric analysis could further elucidate potential affinities between these taxa; in this analysis *R. proteoides* should also

be considered.

Two components have been described so far that can influence the formation of hybrids in *Rhododendron*: genomic incompatibilities and environmental conditions. It is still unknown which and how many genes or loci are involved in these incompatibilities. Genetic mapping will be required to identify these. The populations of *R. aganniphum* and *R. phaeochrysum* at Dáxuě Shan seem to have retained their integrity despite historical introgression, and no contemporary hybrids were found. Are the frequencies of alleles at loci under selection different in this population, so that no hybrids can be formed, or did environmental conditions change? When comparing different hybrid zones of the same species, are always the same loci responsible for incompatibilities or are there local differences? Are backcrosses really formed only in disturbed habitats, or is there a stronger genetic component? If they are only formed in disturbed habitats, what qualifies as such, and why? To better understand the amount of historical introgression it would be important to know how frequently in the evolutionary past disturbances have occurred. Furthermore, the question whether their occurrence has changed significantly due to anthropogenic influence could be addressed; and if so, does this make a difference, or is the genome through selection impermeable enough to tolerate more frequent occurrences of backcrossing?

Summary Previous population genetic work related to hybridisation in *Rhododendron* has mostly addressed questions of taxonomic or phylogenetic importance [36, 117, 184, 186], investigated the composition of hybrid swarms [110, 116, 185] and provided insights into the impact of environmental conditions for hybrid formation [116, 119].

In addition to tackling further taxonomic problems, this study for the first time presents evidence for genomic incompatibilities between two closely related species of the genus. While not contradicting previous observations of hybrid swarm dynamics, it adds a further element that has not been widely considered when investigating barriers to geneflow between species in *Rhododendron*.

References

[1] 2007. The Rhododendron, Camellia & Magnolia Group Bulletin. Volume 95.

[2] Abbott, R. J., J. K. James, R. I. Milne, and A. C. M. Gillies, 2003. Plant introductions, hybridization and gene flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358(1434):1123–1132.

[3] Anderson, E., 1948. Hybridization of the habitat. *Evolution* 2(1):1–9.

[4] Anderson, E. and J. Stebbins, G. L., 1954. Hybridization as an evolutionary stimulus. *Evolution* 8(4):378–388.

[5] Arnold, M. L., 1992. Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* 23:237–261.

[6] Arnold, M. L., 1993. *Iris nelsonii* (Iridaceae): Origin and genetic composition of a homoploid hybrid species. *American Journal of Botany* 80(5):577–583.

[7] Arnold, M. L., 2004. Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *The Plant Cell* 16(3):562–570.

[8] Arnold, M. L., M. R. Bulger, J. M. Burke, A. L. Hempel, and J. H. Williams, 1999. Natural hybridization: how low can you go and still be important? *Ecology* 80(2):371–381.

[9] Arnold, M. L., E. K. Kentner, J. A. Johnston, S. Cornman, and A. C. Bouck, 2001. Natural hybridisation and fitness. *Taxon* 50(1):93–104.

[10] Arnold, M. L., S. Tang, S. J. Knapp, and N. H. Martin, 2010. Asymmetric introgressive hybridization among louisiana iris species. *Genes* 1:9–22.

[11] Ayala, F. J. and M. Coluzzi, 2005. Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *Proceedings of the National Academy of Sciences USA* 102(Suppl 1):6535–6542.

[12] Baack, E. J. and L. H. Rieseberg, 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics and Development* 17(6):513–518.

[13] Barrier, M., B. G. Baldwin, R. H. Robichaux, and M. D. Purugganan, 1999. Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Molecular Biology and Evolution* 16(8):1105–1113.

[14] Barthlott, W., C. Neinhuis, D. Cutler, F. Ditsch, I. Meusel, I. Theisen, and H. Wilhelmi, 1998. Classification and terminology of plant epicuticular waxes. *Botanical Journal of the Linnean Society* 126:237–260.

[15] Barton, N. H., 2001. The role of hybridization in evolution. *Molecular Ecology* 10(3):551–568.

[16] Barton, N. H. and G. M. Hewitt, 1989. Adaptation, speciation and hybrid zones. *Nature* 341(6242):497–503.

[17] Benjamini, Y. and Y. Hochberg, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)* 57(1):289–300.

[18] Bensch, S. and M. Akesson, 2005. Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology* 14(10):2899–2914.

[19] Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das, 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22(3):148–155.

[20] Bleeker, W. and H. Hurka, 2001. Introgressive hybridization in *Rorippa* (Brassicaceae): gene flow and its consequences in natural and anthropogenic habitats. *Molecular Ecology* 10(8):2013–2022.

[21] Bomblies, K., 2010. Doomed Lovers: Mechanisms of Isolation and Incompatibility in Plants. *Annual Review of Plant Biology* 61(1):109–124.

[22] Bonin, A., E. Bellemain, P. Bronkeneidesen, F. Pompanon, C. Brochmann, and P. Taberlet, 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13:3261–3273.

[23] Bonin, A., D. Ehrich, and S. Manel, 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology* 16(18):3737–3758.

[24] Borowsky, R. L., 2001. Estimating nucleotide diversity from random amplified polymorphic DNA and amplified fragment length polymorphism data. *Molecular Phylogenetics and Evolution* 18(1):143–148.

[25] Bradshaw, H. D. J., K. G. Otto, B. E. Frewen, J. K. McKay, and D. W. Schemske, 1998. Quantitative trait loci affecting differences in floral morphology between two species of Monkeyflower (*Mimulus*). *Genetics* 149:367–382.

[26] Buerkle, C. A., 2003. Speciation: A rebirth. *New Phytologist* 160(1):14–17.

[27] Buggs, R. J. A. and J. R. Pannell, 2007. Ecological differentiation and diploid superiority across a moving ploidy contact zone. *Evolution* 61(1):125–140.

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

[29] Burke, J. M. and M. L. Arnold, 2001. Genetics and the fitness of hybrids. *Annual Review of Genetics* 35(1):31–52.

[30] Campbell, D. R., C. Galen, and C. A. Wu, 2005. Ecophysiology of first and second generation hybrids in a natural plant hybrid zone. *Oecologia* 144(2):214–225.

[31] Chadwick, M. D., D. F. Chamberlain, B. A. Knights, A. McAleese, S. Peters, D. W. H. Rankin, and F. Sanderson, 2000. Analysis of leaf waxes as a taxonomic guide to *Rhododendron* subsection *Taliensia*. *Annals of Botany* 86:371–384.

[32] Chamberlain, D. F., 1982. A Revision of Rhododendron, II. Subgenus Hymenanthes. *Notes from the Royal Botanic Garden Edinburgh* 39(2):209–486.

[33] Chamberlain, D. F., R. Hyam, G. Argent, G. Fairweather, and K. S. Walter, 1996. *The genus Rhododendron: Its classification and synonymy.* Royal Botanic Garden Edinburgh.

[34] Chapman, M. A. and J. M. Burke, 2007. Genetic divergence and hybrid speciation. *Evolution* 61(7):1773–1780.

[35] Chiang, T.-Y., B. A. Schaal, and C.-I. Peng, 1998. Universal primers for amplification and sequencing a noncoding spacer between the atpB and rbcL genes of chloroplast DNA. *Botanical Bulletin of Academia Sinica* 39:245–250.

[36] Chung, J.-D., T.-P. Lin, Y.-L. Chen, Y.-P. Cheng, and S.-Y. Hwang, 2007. Phylogeographic study reveals the origin and evolutionary history of a *Rhododendron* species complex in Taiwan. *Molecular Phylogenetics and Evolution* 42(1):14–24.

[37] Chung, M. Y., J. D. Nason, and M. G. Chung, 2005. Patterns of hybridization and population genetic structure in the terrestrial orchids *Liparis kumokiri* and *Liparis makinoana* (Orchidaceae) in sympatric populations. *Molecular Ecology* 14(14):4389–4402.

[38] Cockerham, C. C., 1973. Analyses of gene frequencies. *Genetics* 74:679–700.

[39] Demesure, B., N. Sodzi, and R. J. Petit, 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4:129–131.

[40] Dobzhansky, T., 1940. Speciation as a Stage in Evolutionary Divergence. *The American Naturalist* 74(753):312–321.

[41] Dodd, R. S. and Z. Afzal-Rafii, 2004. Selection and dispersal in a multispecies oak hybrid zone. *Evolution* 58(2):261–269.

[42] Donghuai, S., 2004. Monsoon and westerly circulation changes recorded in the late Cenozoic aeolian sequences of Northern China. *Global and Planetary Change* 41(1):63–80.

[43] Dray, S. and A. Dufour, 2007. The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* 22(4):1–20.

[44] Duminil, J., O. J. Hardy, and R. J. Petit, 2009. Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. *BMC Evolutionary Biology* 9:177.

[45] Dumolin-Lapegue, S., M.-H. Pemonge, and R. J. Petit, 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology* 6:393–397.

[46] Eckenwalder, J. E., 1998. Review: Hybridization as evolutionary creation. *American Journal of Botany* 85(7):1043–1045.

[47] Eckert, A. J. and B. C. Carstens, 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Molecular Phylogenetics and Evolution* 49:832–842.

[48] Eeckhaut, T., E. Keyser, J. Huylenbroeck, J. Riek, and E. Bockstaele, 2007. Application of embryo rescue after interspecific crosses in the genus *Rhododendron*. *Plant Cell, Tissue and Organ Culture* 89(1):29–35.

[49] Ellstrand, N. C., R. Whitkus, and L. H. Rieseberg, 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences USA* 93(10):5090–5093.

[50] Emms, S. K., S. A. Hodges, and M. L. Arnold, 1996. Pollen-tube competition, siring success, and consistent asymmetric hybridization in Louisiana irises. *Evolution* 50(6):2201–2206.

[51] Escaravage, N., S. Questiau, A. Pornon, B. Doche, and P. Taberlet, 1998. Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP markers. *Molecular Ecology* 7(8):975–982.

[52] Evanno, G., S. Regnaut, and J. Goudet, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14:2611–2620.

[53] Excoffier, L. and H. E. L. Lischer, 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Notes* 10(3):564–567.

[54] Excoffier, L., P. E. Smouse, and J. M. Quattro, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial dna restriction data. *Genetics* 131:479–491.

[55] Falush, D., M. Stephens, and J. K. Pritchard, 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes* 7(4):574–578.

[56] Fang, W. P., 1983. *Rhododendron clementinae* Forrest ssp. *aureodorsale* from Shaanxi Province. *Flora Tsinlingensis* 1:393.

[57] Faria, R. and A. Navarro, 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology and Evolution* 25(11):660–669.

[58] Fay, M. F., S. M. Swensen, and M. W. Chase, 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin* 52(1):111–120.

[59] Frajman, B., F. Eggens, and B. Oxelman, 2009. Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae) – A multigene phylogenetic approach with relative dating. *Systematic Biology* 58(3):328–345.

[60] Friesen, M. L., G. Saxer, M. Travisano, and M. Doebeli, 2004. Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution* 58(2):245–260.

[61] Fritsche, F. and O. Kaltz, 2000. Is the *Prunella* (Lamiaceae) hybrid zone structured by an environmental gradient? Evidence from a reciprocal transplant experiment. *American Journal of Botany* 87(7):995–1003.

[62] Gascuel, O., 1997. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* 14(7):685–695.

[63] Gay, L., P.-A. Crochet, D. A. Bell, and T. Lenormand, 2008. Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* 62(11):2789–2806.

[64] Goetsch, L., A. J. Eckert, and B. D. Hall, 2005. The molecular systematics of *Rhododendron* (Ericaceae): A phylogeny based upon RPB2 gene sequences. *Systematic Botany* 30(3):616–626.

[65] Gourbière, S. and J. Mallet, 2009. Are species real? the shape of the species boundary with exponential failure, reinforcement, and the "missing snowball. *Evolution* 64(1):1–24.

[66] Gower, J. C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 3(4):325–338.

[67] Gower, J. C. and P. Legendre, 1986. Metric and euclidean properties of dissimilarity coefficients. *Journal of Classification* 3:5–48.

[68] Gross, B. L. and L. H. Rieseberg, 2005. The ecological genetics of homoploid hybrid speciation. *Journal of Heredity* 96(3):241–252.

[69] Hairston, N. G. S., R. H. Wiley, C. K. Smith, and K. A. Kneidel, 1992. The dynamics of two hybrid zones in appalachian salamanders of the genus *Plethodon*. *Evolution* 46(4):930–938.

[70] Hakstian, A. R. and W. T. Rogers, 1982. The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research* 17:193–219.

[71] Hamblin, M. T., A. M. Casa, H. Sun, S. C. Murray, A. H. Paterson, C. F. Aquadro, and S. Kresovich, 2006. Challenges of detecting directional selection after a bottleneck: lessons from *Sorghum bicolor*. *Genetics* 173:953–964.

[72] Hamilton, M. B., 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8:521–523.

[73] Hamrick, J. L. and M. J. W. Godt, 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society London B: Biological Sciences* 351(1345):1291–1298.

[74] Hartl, D. L. and A. G. Clark, 1997. *Principles of Population Genetics*. Sinauer Associates, Inc., 3rd ed.

[75] Hauser, T. P., C. Damgaard, and R. B. Jørgensen, 2003. Frequency-dependent fitness of hybrids between oilseed rape (*Brassica napus*) and weedy *B. rapa* (Brassicaceae). *American Journal of Botany* 90(4):571–578.

[76] Hayashi, E., H.-C. Chi, S. K. Boyer, and D. W. Still. *Amplified Fragment Length Polymorphism Protocol for Plant Science on CEQ Series Genetic Analysis System*. California State Polytechnic University, Pomona; Beckman Coulter, Inc., a-2015a ed.

[77] Hirao, A. S., 2010. Kinship between parents reduces offspring fitness in a natural population of *Rhododendron brachycarpum*. *Annals of Botany* 105:637–646.

[78] Hodges, S. A., J. M. Burke, and M. L. Arnold, 1996. Natural formation of iris hybrids: Experimental evidence on the establishment of hybrid zones. *Evolution* 50(6):2504–2509.

[79] Hoffmann, A. A. and L. H. Rieseberg, 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics* 39(1):21–42.

[80] Hollingsworth, P. M. and R. A. Ennos, 2004. Neighbour joining trees, dominant markers and population genetic structure. *Heredity* 92:490–498.

[81] Holsinger, K. E. and B. S. Weir, 2009. Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics* 10(9):639–650.

[82] Ingvarsson, P. K., 2005. Nucleotide polymorphism and linkage disequilibrium within and among natural populations of european aspen (*Populus tremula* L., Salicaceae). *Genetics* 169:945–953.

[83] Jacobs, M. M. J., R. G. van den Berg, V. G. A. A. Vleeshouwers, M. Visser, R. Mank, M. Sengers, R. Hoekstra, and B. Vosman, 2008. AFLP analysis reveals a lack of phylogenetic structure within *Solanum* section Petota. *BMC Evolutionary Biology* 8:145.

[84] Jian, S., P. S. Soltis, M. A. Gitzendanner, M. J. Moore, R. Li, T. A. Hendry, Y.-l. Qiu, A. Dhingra, C. D. Bell, and D. E. Soltis, 2008. Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology* 57:38–57.

[85] Johansen-Morris, A. D. and R. G. Latta, 2008. Genotype by environment interactions for fitness in hybrid genotypes of *Avena barbata*. *Evolution* 62(3):573–585.

[86] Johnston, J. A., D. J. Grise, L. A. Donovan, and M. L. Arnold, 2001. Environment-dependent performance and fitness of *Iris brevicaulis, I. fulva* (Iridaceae), and hybrids. *American Journal of Botany* 88(5):933–938.

[87] Karrenberg, S. and A. Favre, 2008. Genetic and ecological differentiation in the hybridizing campions *Silene dioica* and *S. latifolia*. *Evolution* 62(4):763–773.

[88] Karrenberg, S., C. Lexer, and L. H. Rieseberg, 2007. Reconstructing the history of selection during homoploid hybrid speciation. *The American Naturalist* 169(6):725–737.

[89] Kim, M., M.-L. Cui, P. Cubas, A. Gillies, K. Lee, M. A. Chapman, R. J. Abbott, and E. Coen, 2008. Regulatory genes control a key morphological and ecological trait transferred between species. *Science* 322:1116–1119.

[90] Kim, S. C. and L. H. Rieseberg, 1999. Genetic architecture of species differences in annual sunflowers: Implications for adaptive trait introgression. *Genetics* 153(2):965–977.

[91] King, M. J. and S. L. Buchmann, 1995. Bumble bee-initiated vibration release mechanism of *Rhododendron* pollen. *American Journal of Botany* 82(11):1407–1411.

[92] Kirk, H., K. Vrieling, and P. G. L. Klinkhamer, 2005. Reproductive fitness of hybrids between *Senecio jacobaea* and *S. aquaticus* (Asteraceae). *American Journal of Botany* 92(9):1467–1473.

[93] Kirkpatrick, M. and N. Barton, 2006. Chromosome Inversions, Local Adaptation and Speciation. *Genetics* 173(1):419–434.

[94] Kohlmann, B., H. Nix, and D. D. Shaw, 1988. Environmental predictions and distributional limits of chromosomal taxa in the Australian grasshopper *Caledia captiva* (F.). *Oecologia* 75:483–493.

[95] Kron, K. A., 1997. Phylogenetic relationships of Rhododendroideae (Ericaceae). *American Journal of Botany* 84(8):973–980.

[96] Kurashige, Y., J.-I. Etoh, T. Handa, K. Takayanagi, and T. Yukawa, 2001. Sectional relationships in the genus *Rhododendron* (Ericaceae): Evidence from matK and trnK intron sequences. *Plant Systematics and Evolution* 228:1–14.

[97] Levin, D. A., 2001. 50 years of plant speciation. *Taxon* 50(1):69–91.

[98] Levin, D. A., J. Francisco-Ortega, and R. K. Jansen, 1996. Hybridization and the extinction of rare plant species. *Conservation Biology* 10(1):10–16.

[99] Levinton, J. S., 1982. Charles Darwin and Darwinism. *BioScience* 32(6):495–500.

[100] Lewontin, R. C. and L. C. Birch, 1966. Hybridization as a source of variation for adaptation to new environments. *Evolution* 20(3):315–336.

[101] Lexer, C., M. F. Fay, J. A. Joseph, M.-S. Nica, and B. Heinze, 2005. Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): The role of ecology and life history in gene introgression. *Molecular Ecology* 14(4):1045–1057.

[102] Lexer, C., B. Heinze, R. Alia, and L. H. Rieseberg, 2004. Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: lessons from wild annual sunflowers (*Helianthus* spp.). *Forest Ecology and Management* 197(1-3):49–64.

[103] Lexer, C., A. Kremer, and R. J. Petit, 2006. Shared alleles in sympatric oaks: Recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* 15(7):2007–2012.

[104] Lexer, C., M. E. Welch, J. L. Durphy, and L. H. Rieseberg, 2003. Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: Implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Molecular Ecology* 12(5):1225–1235.

[105] Ligges, U. and M. Mächler, 2003. Scatterplot3d - An R package for visualizing multivariate data. *Journal of Statistical Software* 8(11):1–20.

[106] Liu, J.-Q., Y.-J. Wang, A.-L. Wang, O. Hideaki, and R. J. Abbott, 2006. Radiation and diversification within the *Ligularia-Cremanthodium-Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution* 38(1):31–49.

[107] Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis, 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1506):3009–3021.

[108] Lowry, D. B. and J. H. Willis, 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* 8(9):e1000500.

[109] Lynch, M. and B. G. Milligan, 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3:91–99.

[110] Ma, Y., R. I. Milne, C. Zhang, and J. Yang, 2010. Unusual patterns of hybridization involving a narrow endemic *Rhododendron* species (Ericaceae) in Yunnan, China. *American Journal of Botany* 97(10):1749–1757.

[111] Mallet, J., 2005. Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* 20(5):229–237.

[112] Martin, N. H., A. C. Bouck, and M. L. Arnold, 2006. Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* 172(4):2481–2489.

[113] May, R. M., J. A. Endler, and R. E. McMurtrie, 1975. Gene frequency clines in the presence of selection opposed by gene flow. *The American Naturalist* 109(970):659–676.

[114] Mayr, E., 1992. A local flora and the biological species concept. *American Journal of Botany* 79(2):222–238.

[115] Meudt, H. M. and A. C. Clarke, 2007. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science* 12(3):106–117.

[116] Milne, R. I. and R. J. Abbott, 2008. Reproductive isolation among two interfertile *Rhododendron* species: Low frequency of post-F1 hybrid genotypes in alpine hybrid zones. *Molecular Ecology* 17(4):1108–1121.

[117] Milne, R. I., R. J. Abbott, K. Wolff, and D. F. Chamberlain, 1999. Hybridization among sympatric species of *Rhododendron* (Ericaceae) in Turkey: Morphological and molecular evidence. *American Journal of Botany* 86(12):1776–1785.

[118] Milne, R. I., C. Davies, R. Prickett, L. H. Inns, and D. F. Chamberlain, 2010. Phylogeny of *Rhododendron* subgenus *Hymenanthes* based on chloroplast DNA markers: Between-lineage hybridisation during adaptive radiation? *Plant Systematics and Evolution* 285:233–244.

[119] Milne, R. I., S. Terzioglu, and R. J. Abbott, 2003. A hybrid zone dominated by fertile F1s: Maintenance of species barriers in *Rhododendron*. *Molecular Ecology* 12:2719–2729.

[120] Morimoto, J., T. Kamichi, I. Mizumoto, S. Hasegawa, M. Nomura, and T. Kobayashi, 2005. Natural hybridization of Japanese *Rhododendron* section *Brachycaryx* in Mount Kintoki in eastern Japan and concerns for genetic diversity in restoring their habitat. *Landscape and Ecological Engineering* 1:149–156.

[121] Moyle, L. C., M. S. Olson, and P. Tiffin, 2004. Patterns of reproductive isolation in three angiosperm genera. *Evolution* 58(6):1195–1208.

[122] Mueller, U. G. and L. L. Wolfenbarger, 1999. AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* 14(10):389–394.

[123] Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent, 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.

[124] Nosil, P., S. P. Egan, and D. J. Funk, 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: isolation by adaptation and multiple roles for divergent selection. *Evolution* 62(2):316–336.

[125] Ono, A., I. Dohzono, and T. Sugawara, 2008. Bumblebee pollination and reproductive biology of *Rhododendron semibarbatum* (Ericaceae). *Journal of Plant Research* 121:319–327.

[126] Orr, H. A., 1996. Dobzhansky, Bateson, and the genetics of speciation. *Genetics* 144(4):1331–1335.

[127] Orr, H. A., 2005. The genetic basis of reproductive isolation: Insights from *Drosophila*. *Proceedings of the National Academy of Sciences USA* 102(Suppl 1):6522–6526.

[128] Orr, H. A., J. P. Masly, and D. C. Presgraves, 2004. Speciation genes. *Current Opinion in Genetics and Development* 14:675–679.

[129] Otto, S. P. and J. Whitton, 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34(1):401–437.

[130] Palmer, M. E. and M. W. Feldman, 2009. Dynamics of hybrid incompatibility in gene networks in a constant environment. *Evolution* 63(2):418–431.

[131] Peakall, R. and P. Smouse, 2009. *GenAlEx Tutorials-Part 2, Genetic Distance and Analysis of Molecular Variance (AMOVA)*. URL `http://www.anu.edu.au/BoZo/GenAlEx/`.

[132] Peakall, R. and P. E. Smouse, 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288–295.

[133] Pfennig, D. W., M. A. Wund, E. C. Snell-Rood, T. Cruickshank, C. D. Schlichting, and A. P. Moczek, 2010. Phenotypic plasticitys impacts on diversification and speciation. *Trends in Ecology and Evolution* 25(8):459–467.

[134] Presgraves, D. C., 2010. Darwin and the origin of interspecific genetic incompatibilities. In *The American Naturalist* [151], pages S45–S60.

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

[136] Pritchard, J. K., X. Wen, and D. Falush, 2009. *Documentation for structure software: Version 2.3*. Department of Human Genetics University of Chicago, Department of Statistics University of Oxford.

[137] QIAGEN, 2006. *DNeasy® Plant Handbook*. QIAGEN, 07/2006 ed.

[138] R Development Core Team, 2010. *R: A Language and Environment for Statistical Computing, reference index version 2.12.0*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org/`.

[139] Rice, W. R., J. E. Linder, U. Friberg, T. A. Lew, E. H. Morrow, and A. D. Stewart, 2005. Inter-locus antagonistic coevolution as an engine of speciation: Assessment with hemiclonal analysis. *Proceedings of the National Academy of Sciences USA* 102(Suppl 1):6527–6534.

[140] Richardson, J. B., 1996. Abnormal spores and possible interspecific hybridization as a factor in the evolution of Early Devonian land plants. *Review of Palaeobotany and Palynology* 93(1-4):333–340.

[141] Rieseberg, L. H., 1995. The role of hybridization in evolution: Old wine in new skins. *American Journal of Botany* 82(7):944–953.

[142] Rieseberg, L. H., 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 28:359–389.

[143] Rieseberg, L. H., M. A. Archer, and R. K. Wayne, 1999. Transgressive segregation, adaptation and speciation. *Heredity* 83(4):363–372.

[144] Rieseberg, L. H. and B. K. Blackman, 2010. Speciation genes in plants. *Annals of Botany* 106(3):439–455.

[145] Rieseberg, L. H. and S. E. Carney, 1998. Tansley review No. 102 Plant Hybridization. *New Phytologist* 140(4):599–624.

[146] Rieseberg, L. H., J. Whitton, and K. Gardner, 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152(2):713–727.

[147] Rieseberg, L. H. and J. H. Willis, 2007. Plant Speciation. *Science* 317(5840):910–914.

[148] Rundle, H. D. and P. Nosil, 2005. Ecological speciation. *Ecology Letters* 8:336–352.

[149] Sang, T., D. J. Crawford, and T. F. Stuessy, 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84(8):1120–1136.

[150] Schemske, D. W., 2010. Adaptation and the origin of species. In *The American Naturalist* [151], pages S4–S25.

[151] Schemske, D. W., 2010. Darwinian thinking: 150 years after the origin. *The American Naturalist* 176(S1):S1–S3.

[152] Schlichting, C. D., 1986. The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and Systematics* 17:667–693.

[153] Schweitzer, J. A., G. D. Martinsen, and T. G. Whitham, 2002. Cottonwood hybrids gain fitness traits of both parents: A mechanism for their long-term persistence? *American Journal of Botany* 89(6):981–990.

[154] Seehausen, O., 2004. Hybridization and adaptive radiation. *Trends in Ecology and Evolution* 19(4):198–207.

[155] Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small, 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast dna sequences for phylogenetic analysis. *American Journal of Botany* 92(1):142–166.

[156] Shaw, J., E. B. Lickey, E. E. Schilling, and R. L. Small, 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(3):275–288.

[157] Spethmann, W., 1987. A new infrageneric classification and phylogenetic trends in the genus *Rhododendron* (Ericaceae). *Plant Systematics and Evolution* 157(1987):9–31.

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

[159] Stebbins, G. L., 1985. Polyploidy, hybridization, and the invasion of new habitats. *Annals of the Missouri Botanical Garden* 72(4):824–832.

[160] Stevens, S. S., 1957. On the psychophysical law. *The Psychological Review* 64(3):153–181.

[161] Stevens, S. S. and E. H. Galanter, 1957. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology* 54(6):377–411.

[162] Stutz, H. C. and L. K. Thomas, 1964. Hybridization and introgression in *Cowania* and *Purshia*. *Evolution* 18(2):183–195.

[163] Sun, D., R. Su, J. Bloemendal, and H. Lu, 2008. Grain-size and accumulation rate records from Late Cenozoic aeolian sequences in northern China: Implications for variations in the East Asian winter monsoon and westerly atmospheric circulation. *Palaeogeography, Palaeoclimatology, Palaeoecology* 264(1-2):39–53.

[164] Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet, 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17:1105–1109.

[165] Takada, Y. and M. Matsu'ura, 2007. Geometric evolution of a plate interface-branch fault system: Its effects on the tectonic development of the Himalayas. *Journal of Asian Earth Sciences* 29(2-3):490–503.

[166] Tang, S., R. A. Okashah, S. J. Knapp, M. L. Arnold, and N. H. Martin, 2010. Transmission ratio distortion results in asymmetric introgression in Louisiana Iris. *BMC Plant Biology* 10:48.

[167] Tate, J. A. and B. B. Simpson, 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28(4):723–737.

[168] Teeter, K. C., B. A. Payseur, L. W. Harris, M. A. Bakewell, L. M. Thibodeau, J. E. O'Brien, J. G. Krenz, M. A. Sans-Fuentes, M. W. Nachman, and P. K. Tucker, 2008. Genome-wide patterns of gene flow across a house mouse hybridzone. *Genome Research* 18:67–76.

[169] Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, and P. K. Tucker, 2009. The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* 64(2):472–485.

[170] Templeton, A. R., 1981. Mechanisms of speciation – A population genetic approach. *Annual Review of Ecology and Systematics* 12:23–48.

[171] Tsukaya, H., T. Fukuda, and J. Yokoyama, 2003. Hybridization and introgression between *Callicarpa japonica* and *C. mollis* (Verbenaceae) in central Japan, as inferred from nuclear and chloroplast DNA sequences. *Molecular Ecology* 12:3003–3011.

[172] Turelli, M. and L. C. Moyle, 2007. Asymmetric postmating isolation: Darwin's corollary to Haldane's Rule. *Genetics* 176(2):1059–1088.

[173] Ungerer, M. C., S. J. E. Baird, J. Pan, and L. H. Rieseberg, 1998. Rapid hybrid speciation in wild sunflowers. *Proceedings of the National Academy of Sciences USA* 95(20):11757–11762.

[174] Vekemans, X., T. Beauwens, M. Lemaire, and I. Roldan-Ruiz, 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* 11:139–151.

[175] Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. v. H. M. Lee, A. Friters, J. Pot, J. Paleman, M. Kuiper, and M. Zabeau, 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* 23(21):4407–4414.

[176] Wang, Y.-J., A. Susanna, E. von Raab-Straube, R. Milne, and J.-Q. Liu, 2009. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the QinghaiTibetan Plateau. *Biological Journal of the Linnean Society* 97(4):893–903.

[177] Weir, B. S. and C. C. Cockerham, 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.

[178] West-Eberhard, M. J., 2005. Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences USA* 102(Suppl 1):6543–6549.

[179] Wheelwright, N. T., E. E. Dukeshire, J. B. Fontaine, S. H. Gutow, D. A. Moeller, J. G. Schuetz, T. M. Smith, S. L. Rodgers, and A. G. Zink, 2005. Pollinator limitation, autogamy and minimal inbreeding depression in insect-pollinated plants on a boreal island. *The American Midland Naturalist* 155(1):19–38.

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

[181] Wolf, D. E., N. Takebayashi, and L. H. Rieseberg, 2001. Predicting the risk of extinction through hybridization. *Conservation Biology* 15(4):1039–1053.

[182] Wu, C.-I., 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* 14:851–865.

[183] Wu, N., Y. Pei, H. Lu, Z. Guo, F. Li, and T. Liu, 2006. Marked ecological shifts during 6.2-2.4 Ma revealed by a terrestrial molluscan record from the Chinese Red Clay Formation and implication for palaeoclimatic evolution. *Palaeogeography, Palaeoclimatology, Palaeoecology* 233(3-4):287–299.

[184] Zha, H.-G., R. I. Milne, and H. Sun, 2008. Morphological and molecular evidence of natural hybridization between two distantly related *Rhododendron* species from the Sino-Himalaya. *Botanical Journal of the Linnean Society* 156(1):119–129.

[185] Zha, H.-G., R. I. Milne, and H. Sun, 2010. Asymmetric hybridization in *Rhododendron agastum*: A hybrid taxon comprising mainly F1s in Yunnan, China. *Annals of Botany* 105:89–100.

[186] Zhang, J.-L., C.-Q. Zhang, L.-M. Gao, J.-B. Yang, and H.-T. Li, 2007. Natural hybridization origin of *Rhododendron agastum* (Ericaceae) in Yunnan, China: Inferred from morphological and molecular evidence. *Journal of Plant Research* 120(3):457–463.

[187] Zhang, Z.-Y., X.-M. Zheng, and S. Ge, 2007. Population genetic structure of *Vitex negundo* (Verbenaceae) in three-gorge area of the Yangtze River: The riverine barrier to seed dispersal in plants. *Biochemical Systematics and Ecology* 35:506–516.

[188] Zhivotovsky, L. A., 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* 8:907–913.

Appendix A

Equipment used for laboratory work

A.1 Suppliers

- **New England Biolabs (NEB)**

  75–77 Knowl Piece

  Wilbury Way

  Hitchin

  Hertfordshire

  SG4 0TY

- **Promega UK Ltd. (Promega)**

  Delta House

  Southampton Science Park

  Southampton

  SO16 7NS

- **Sigma-Aldrich Co. Ltd. (Sigma)**

  Gillingham

  Dorset

  UK

  `http://www.sigma-aldrich.com`

- **QIAGEN Ltd. (Qiagen)**

  QIAGEN House

  Fleming Way

  Crawley

  West Sussex

  RH10 9NQ

- **Perkin Elmer Instruments LLC (Perkin Elmer)**

  710 Bridgeport Ave.

  Shelton

  Conneticut

  U.S.A.

- **ACROS ORGANICS (Acros)**

  Geel

  Belgium

  `http://www.acros.com`

A.2 Chemicals

A.2.1 Enzymes

Enzyme	**Supplier**
*Eco*RI	NEB
*Mse*I	NEB
T4 DNA Ligase	NEB
Taq DNA Polymerase	NEB
Phire^{TM} Hot Start DNA Polymerase	NEB
GoTaq® Hot Start Polymerase	Promega

A.2.2 Buffers

NEBuffer 2 (NEB)

50 mM	NaCl
10 mM	Tris-HCl
10 mM	$MgCl_2$
1 mM	dithiothreitol (DTT)

(pH 7.9 @ 25 °C)

Standard *Taq* (Mg-free) Reaction Buffer 1× (NEB)

10 mM	Tris-HCl
50 mM	KCl

(pH 8.3 @ 25 °C)

Ligase Reaction Buffer 1× (NEB)

50 mM	Tris-HCl
10 mM	$MgCl_2$
10 mM	dithiothreitol (DTT)
1 mM	ATP

(pH 7.5 @ 25 °C)

$TE^{0.1}$ (20 mM Tris-HCl, 0.1 mM EDTA pH 8.0)

(for 100 ml)

98 ml	ddH_2O
2 ml	Tris (1 M, pH 7.5)
20 μl	EDTA (0.5 M, pH 8.0)

A.2.3 Leaf waxes

GC specifications:

AutoSystem XL GC (Perkin Elmer)

Column: N931-6076 PE-5, Length: 30 m, I.D.: 0.25 mm, Film: 0.25 μm

Eicosane (99 %), Acros, CAS: 112-95-8

Hexatricontane (98 %), Sigma, CAS: 630-06-08

Alumina

Standard C_{20}/C_{36} (25 mg/ml)

(for 400 ml)	
100 mg	Eicosane ($C_{20}H_{32}$)
100 mg	Hexatricontane ($C_{36}H_{74}$)
fill up to 400 ml with chloroform	

A.3 Primers and adapters

A.3.1 Chloroplast

Table A.1: Tested chloroplast primers

Region	Primer Name	Sequence (5‘–. . . –3‘)	Reference
*rpo*C1-*trn*C	rpoC1-trnC.F	GCA CAA ATT CCR CTT TTT ATRGG	[45]
	rpoC1-trnC.R	CGA CAC CCR GAT TTG AAC TGG	[45]
*trn*D-*trn*T	trnDT.F	ACC AAT TGA ACT ACA ATC CC	[39]
	trnDT.R	CTA CCA CTG AGT TAA AAG GG	[39]
*psb*C-*trn*S	psbC-trnS.F	GGT CGT GAC CAA GAA ACC AC	[39]
	psbC-trnS.R	GGT TCG AAT CCC TCT CTC TC	[39]
*trn*S-*trn*Fm	trnSFm.F	GAG AGA GAG GGA TTC GAA CC	[39]
	trnSFm.R	CAT AAC CTT GAG GTC ACG GG	[39]
*atp*B-*rbc*L	atpB-rbcL.F	ACA TCK ART ACK GGA CCA ATAA	[35]
	atpB-rbcL.R	AAC ACC AGC TTT RAA TCC AA	[35]
*trn*S-*trn*G	trnSG.F	GCC GCT TTA GTC CAC TCA GC	[72]
	trnSG.R	GAA CGA ATC ACA CTT TTA CCAC	[72]
*trn*L-*trn*F	trnL.c	CGA AAT CGG TAG ACG CTA CG	[164]
	trnL.d	GGG GAT AGA GGG ACT TGA AC	[164]
	trnL.e	GGT TCA AGT CCC TCT ATC CC	[164]
	trnL.f	ATT TGA ACT GGT GAC ACG AG	[164]
*rbc*L	rbcL 1F	ATG TCA CCA CAA ACA GAA AC	[58]
	rbcL 724R	TCG CAT GTA CCT GCA GTA GC	[58]
*rpl*32-*trn*L	rpl32-trnL.F	CTG CTT CCT AAG AGC AGC GT	[156]
	rpl32-trnL.R	CAG TTC CAA AAA AAC GTA CTTC	[156]
*psb*D-*trn*T	psbD-trnT.F	CTC CGT ARC CAG TCA TCC ATA	[156]
	psbD-trnT.R	CCC TTT TAA CTC AGT GGT AG	[156]
3’*rps*16-5’*trn*K	3’rps16-5’trnK.F	AAA GTG GGT TTT TAT GAT CC	[156]
	3’rps16-5’trnK.R	TTA AAA GCC GAG TAC TCT ACC	[156]
*atp*I-*atp*H	atpI-atpH.F	TAT TTA CAA GYG GTA TTC AAGCT	[156]
	atpI-atpH.R	CCA AYC CAG CAG CAA TAA C	[156]
*psb*A-*trn*H	psbA3’f	GTT ATG CAT GAA CGT AAT GCTC	[149]
	trnHf	CGC GCA TGG TGG ATT CAC AATCC	[167]
*mat*K	trnK-5’_F	GGG TTG CTA ACT CAA CGG TAGAG	own design
	trnK-3’_R	CGG AAC TAG TCG GAT GGA GTAG	own design
*rpl*20-*rps*12	rpl20	TTT GTT CTA CGT CTC CGA GC	[72]
	5’rps12	GTC GAG GAA CAT GTA CTA GG	[72]
*psb*B-*psb*F	psb B	GTT TAC TTT TGG GCA TGC TTCG	[72]
	psb F	CGC AGT TCG TCT TGG ACC AG	[72]
*acc*D	accD.2f	GGR GCA CGT ATG CAA GAA GG	RBGE
	accD.4r	TCT TTT ACC CGC AAA TGC AAT	RBGE
*rpo*C1	rpoC1.2f	GGC AAA GAG GGA AGA TTT CG	RBGE
	rpoC1.4r	CCA TAA GCA TAT CTT GAG TTGG	RBGE

A.3.2 AFLPs

Table A.2: Adapter oligonucleotides used for AFLPs

Adapter	Oligonucleotide	Sequence (5‘–. . .–3‘)
*Eco*RI adapter	EcoRI-F	CTC GTA GAC TGC GTACC
	EcoRI-R	AAT TGG TAC GCA GTCTAC
*Mse*I adapter	MseI-F	GAC GAT GAG TCC TGAG
	MseI-R	TAC TCA GGA CTC AT

Table A.3: AFLP primers used for pre- and selective amplifications

Primer Name	Sequence (5‘–. . .–3‘)
*Eco*RI specific	GAC TGC GTA CCA ATTC
E+1	GAC TGC GTA CCA ATTC A
E+3	GAC TGC GTA CCA ATTC XXX
E-ACT^{D3}	GAC TGC GTA CCA ATTC ACT
E-ATC^{D2}	GAC TGC GTA CCA ATTC ATC
E-ATC^{D4}	GAC TGC GTA CCA ATTC ATC
*Mse*I specific	GAT GAG TCC TGA GTAA
M+1	GAT GAG TCC TGA GTAA C
M+3	GAT GAG TCC TGA GTAA XXX
M-CAG	GAT GAG TCC TGA GTAA CAG
M-CGA	GAT GAG TCC TGA GTAA CGA
M-CTA	GAT GAG TCC TGA GTAA CTA

^{D2} labelled with WellRED (Sigma) dye D2
^{D3} labelled with WellRED (Sigma) dye D3
^{D4} labelled with WellRED (Sigma) dye D4

Table A.4: Selective primer pair combinations. Primer pair combinations screened for usability in the selective amplification; only the selective nucleotides are given, for the full sequence please see Table A.3

EcoRI	MseI	Usability	EcoRI	MseI	Usability
AAC	CA	unusable	ACT	CTG	unusable
AAC	CT	unusable	ACT	CTT	unusable
AAC	CAA	good*	ACT	CGA	unusable
AAC	CAC	unusable	AGC	CA	unusable
AAC	CAG	bad	AGC	CT	unusable
AAC	CAT	bad	AGC	CAA	unusable
AAC	CTA	good*	AGC	CAC	unusable
AAC	CTC	O.K.*	AGC	CAG	unusable
AAC	CTG	unusable	AGC	CAT	bad
AAC	CTT	unusable	AGC	CTA	unusable
AAC	CGA	unusable	AGC	CTC	unusable
ACC	CA	unusable	AGC	CTG	unusable†
ACC	CT	unusable	AGC	CTT	unusable
ACC	CAA	unusable	AGC	CGA	bad
ACC	CAC	unusable	ATC	CA	unusable
ACC	CTG	unusable	ATC	CT	unusable
ACC	CGA	unusable	ATC	CAA	unusable
ACT	CA	unusable	ATC	CAC	unusable
ACT	CT	unusable	ATC	CAG	good*
ACT	CAA	unusable	ATC	CAT	unusable
ACT	CAC	unusable	ATC	CTA	unusable
ACT	CAG	unusable	ATC	CTC	unusable
ACT	CAT	O.K.	ATC	CTG	unusable
ACT	CTA	O.K.*	ATC	CTT	unusable
ACT	CTC	unusable	ATC	CGA	good*

* These combinations were selected for further testing with more individuals.
† This combination was also further tested, please see subsection 2.2.3 (page 61) for details.

Appendix B

AFLP protocols

B.1 Preparation

- **Restriction Enzyme Test**

***Eco*RI**	
2.9 μl	ddH_2O
1.5 μl	NEBuffer 2
0.5 μl	BSA (10 mg/ml)
0.1 μl	*Eco*RI (100 U/μl)
10 μl	genomic DNA
= 15 μl	

***Mse*I**	
2.8 μl	ddH_2O
1.5 μl	NEBuffer 2
0.5 μl	BSA (10 mg/ml)
0.2 μl	*Mse*I (50 U/μl)
10 μl	genomic DNA
= 15 μl	

Thermocycler program:

2 h	at 37 °C
30 min	at 65 °C
∞	at 10 °C

Run 5 μl on a 1.5 % agarose gel stained with SybrSafe (0.7 h, 85 V); a successful digestion will result in a smear (*Eco*RI $\sim$ 10,000–1500 bp, *Mse*I $\sim$ 1000–100 bp).

- **Adapter Preparation** (1 ml)

EcoRI (end concentration 5 μM)	
ddH_2O	980 μl
EcoRI-F (500 μM)	10 μl
EcoRI-R (500 μM)	10 μl

MseI (end concentration 50 μM)	
ddH_2O	800 μl
MseI-F (500 μM)	100 μl
MseI-R (500 μM)	100 μl

heat both to near boiling ($\sim$ 90 °C) and let cool down to room temperature.

B.2 Digestion

2.93 μl	ddH_2O
1.5 μl	NEBuffer 2 (10x)
0.5 μl	BSA (10 mg/ml)
0.02 μl	*Mse*I (50 U/μl)
0.05 μl	*Eco*RI (100 U/μl)
10 μl	Genomic DNA
= 15μl	

Thermocycler program:

3 h at 37 °C
∞ at 10 °C

B.3 Ligation

1.825 μl	ddH_2O
2.5 μl	Ligase Reaction Buffer (10x)
2.5 μl	EcoRI Adapter (5 μM)
2.5 μl	MseI Adapter (50 μM)
0.5 μl	BSA (10 mg/ml)
0.175μl	T4 DNA Ligase (400,000 cohesive end units/ml)
+15 μl	digested DNA (section B.2)
= 25 μl	

Thermocycler program:

1 h at 20 °C
4 h at 12 °C
∞ at 8 °C
(leave over night)

Dilute 1:4 with $TE^{0.1}$ (25 μl + 75 μl) before next step.

B.4 Preamplification

7.375 μl	dH_2O
1.5 μl	Standard *Taq* (Mg-free) Reaction Buffer (10x)
1.5 μl	dNTPs (2 mM each)
0.6 μl	$MgCl_2$ (50 mM)
0.45 μl	primer E+1 (10 μM)
0.45 μl	primer M+1 (10 μM)
0.125 μl	*Taq*
+ 3 μl	diluted DNA from section B.3
= 15 μl	

Thermocycler program:

72 °C	2 min	
94 °C	20 sec	25×
56 °C	30 sec	
72 °C	2 min	
72 °C	2 min	
60 °C	30 min	
4 °C	∞	

To check for successful preamplification: run 5 μl on 1.5 % agarose gel stained with SybrSafe (0.7 h, 85 V). Under UV light a smear should be visible between 100 bp and 2000 bp. Before next step dilute preamplification 1:7.5 (10 μl + 65 μl $TE^{0.1}$).

B.5 Selective amplification

4.3 μl	dH_2O
1 μl	Standard *Taq* (Mg-free) Reaction Buffer (10x)
1 μl	dNTPs (2 mM each)
0.4 μl	$MgCl_2$ (50 mM)
0.625 μl	primer E+3 (10 μM)
0.625 μl	primer M+3 (10 μM)
0.05 μl	*Taq*
+ 2 μl	diluted pre-amp (section B.4)
= 10 μl	

Thermocycler program:

94 °C	2 min	
94 °C	20 sec	10× (−1°C /cycle)
66 °C	30 sec	
72 °C	2 min	
94 °C	30 sec	25×
56 °C	30 sec	
72 °C	3 min	
60 °C	30 min	
4 °C	∞	

To check for amplification run 4 μl on 1.5 % agarose gel (0.7 h, 85 V) and check for presence of smear and bands between 100 bp and 2000 bp under UV light.

B.6 Run on sequencer

when testing:

35 μl	SLS
1 μl	Sample
0.5 μl	400bp size standard

when pooling samples:

35 μl	SLS
0.8 μl	Sample (each)
0.6 μl	400bp size standard

Appendix C

Figures not included in the text

C.1 MDS placement of population G2b

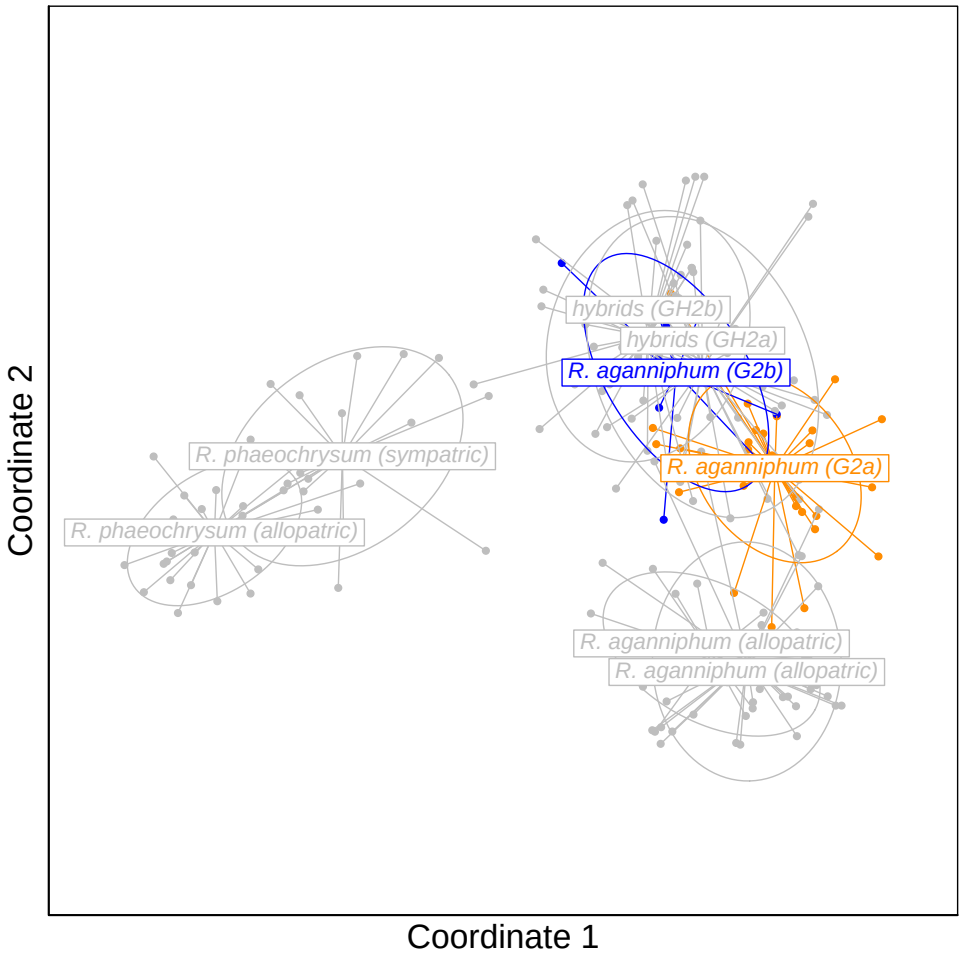

Figure C.1: MDS showing the placement of individuals from the excluded sympatric population G2b (blue) in relation to the used sympatric individuals of *R. aganniphum* (population G2a, renamed to G2; orange).

C.2 Additional figures for STRUCTURE

Cluster membership of individuals

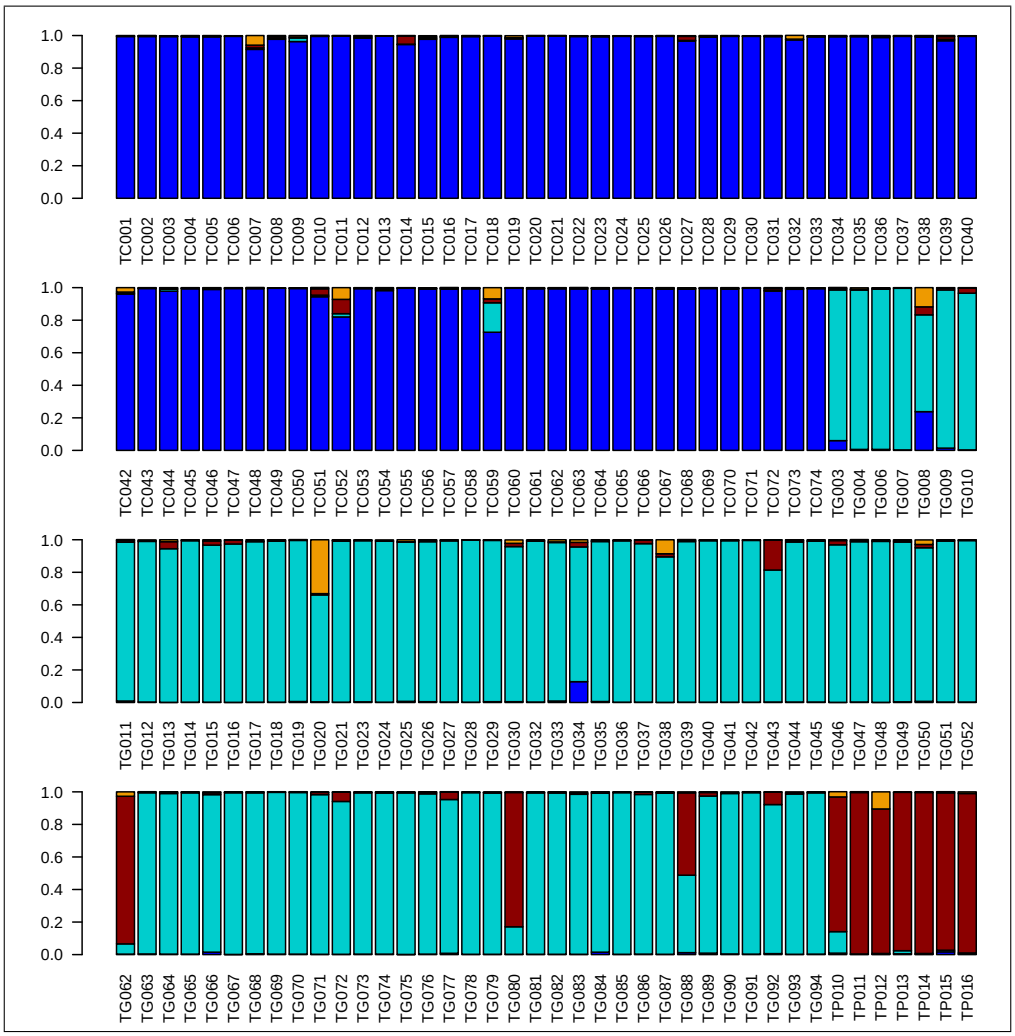

Figure C.2a: Cluster membership of single individuals as estimated with the program STRUCTURE. For a legend please see sub-figure **b**.

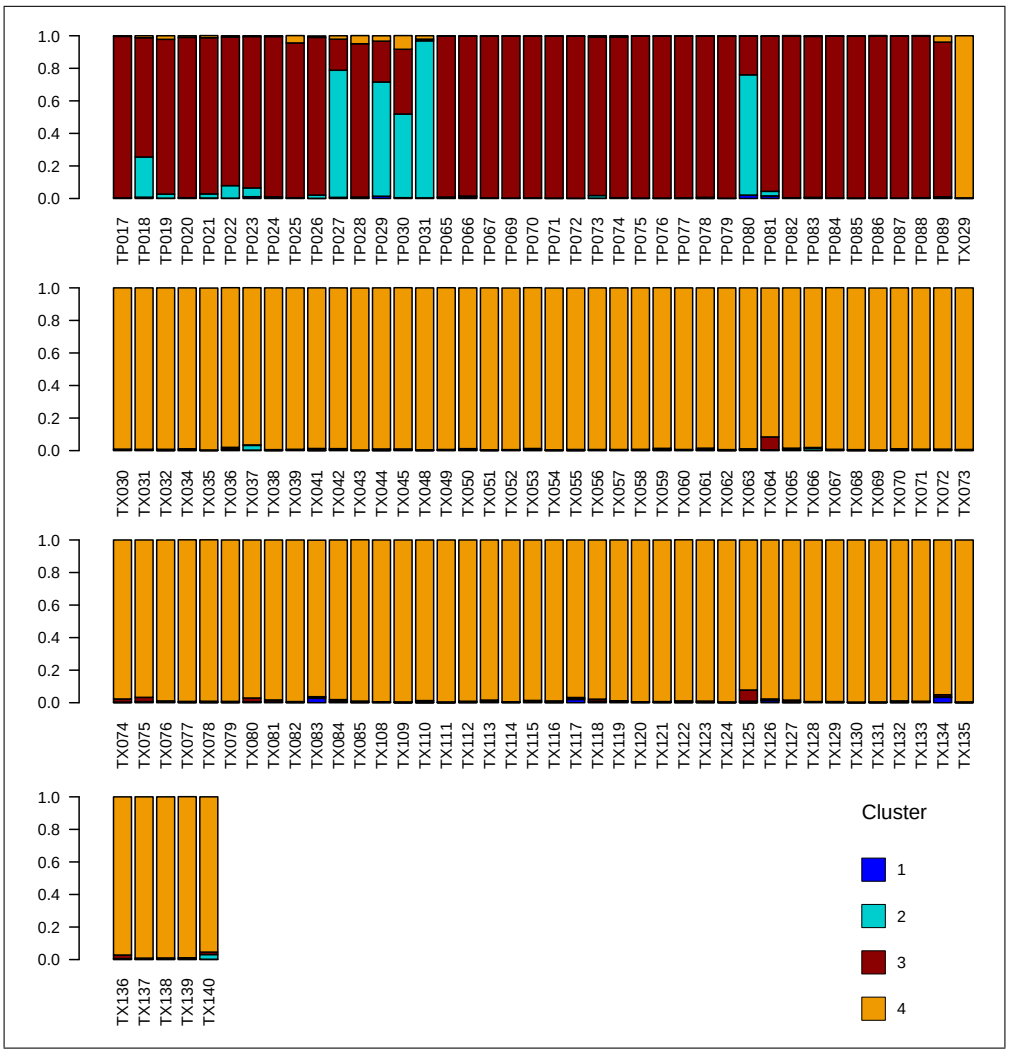

Figure C.2b: Cluster membership of single individuals as estimated with the program STRUCTURE. The clusters correspond to following species: **1** *R. clementinae*, **2** *R. aganniphum*, **3** *R. phaeochrysum*, **4** *R. roxieanum*.

Cluster membership of simulated RM-F1s

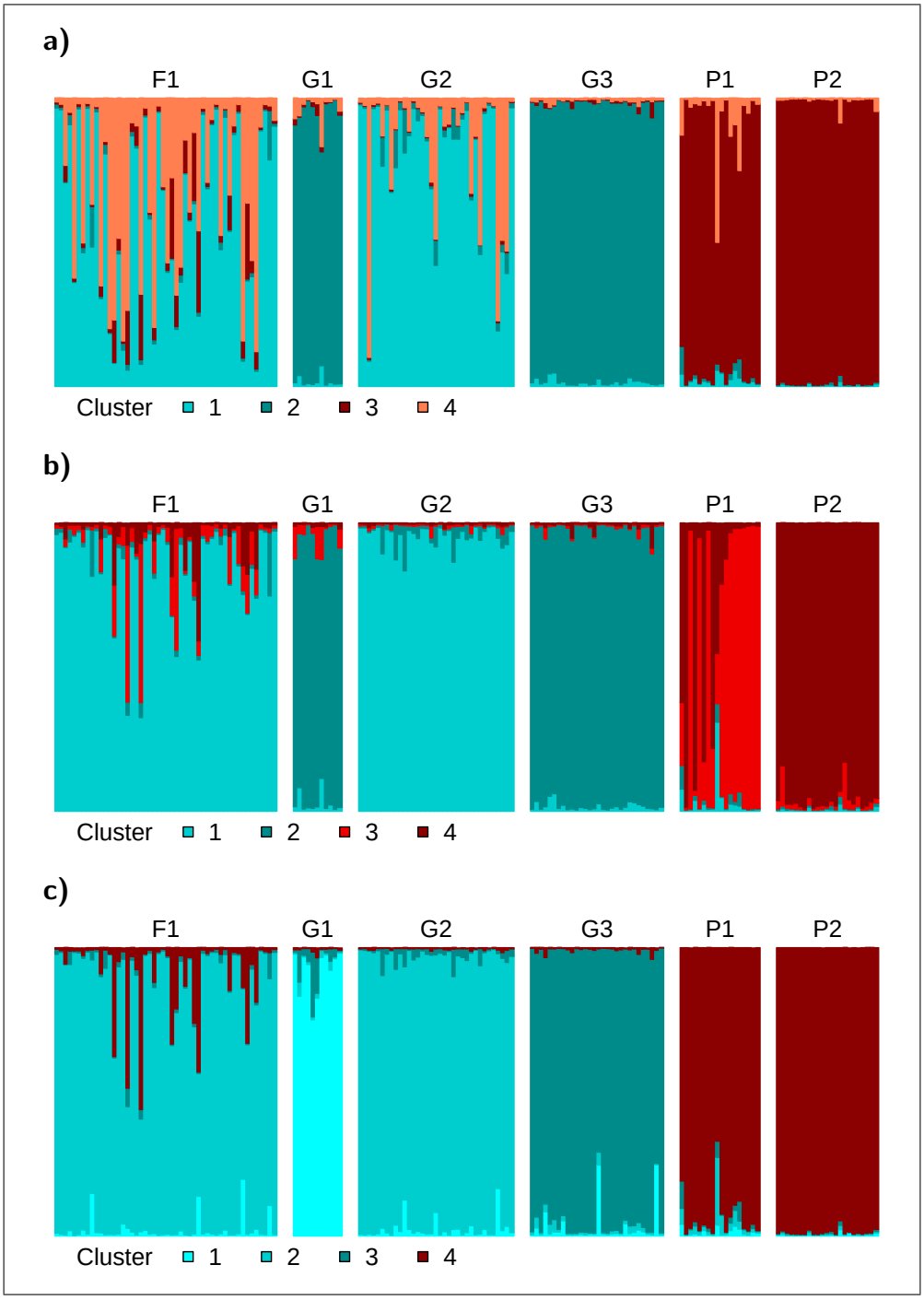

Figure C.3: STRUCTURE **simulated RM-F1s, K = 4.** Types of clustering obtained after different runs in STRUCTURE for the same dataset, comprising *R. aganniphum*, *R. phaeochrysum* and 50 simulated F1 hybrids (G2 × P1). Fourth cluster attributed to: **a)** F1s; **b)** P1/P2; **c)** G1/G3.

Appendix D

Tables not included in the text

D.1 Specimens that were analysed

(next page)

Table D.1: Specimens Information about plants collected in 2007 and 2008 that were analysed for this project; only samples collected from populations that were used have been included into this table. The naming scheme adopted for individuals is explained in section 2.1, page 56.

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TC001[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TC002[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TC003[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TC004[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TC005[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TC006[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TC007[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TC008	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	04.09.2007
TC009	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	04.09.2007
TC010	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	04.09.2007
TC011	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TC012	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TC013	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TC014	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TC015	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TC016[sp]	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TC017	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TC018	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TC019	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC020	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC021	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC022	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC023	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC024	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC025	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC026	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC027	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC028	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC029	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC030	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC031	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC032	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC033	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC034	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC035	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC036	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC037	C1	TC	*Rhododendron clementinae*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	22.05.2008
TC038	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TC039	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC040	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC041	C2	TC	*Rhododendron clementinae*	rem^{NA}	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC042	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC043	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC044	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC045	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC046	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC047	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC048^{sp}	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC049^{sp}	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC050	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC051	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC052	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC053	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC054	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC055	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC056	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC057	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC058	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TC059	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC060	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC061	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC062	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC063	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC064	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC065	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC066	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC067	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC068	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC069	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC070	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC071	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC072	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC073	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TC074	C2	TC	*Rhododendron clementinae*	yes	China, Yunnan, Haba Shan	27° 21' 32.2" N	100° 04' 18.3" E	4081 m	05.06.2008
TG003	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG004	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG005	G1	TG	*R. aganniphum* var. *aganniphum*	rem[NA]	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG006	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TG007	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG008	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG009	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG010	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG011[sp]	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG012[sp]	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG013	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG014	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG015	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG016	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG017	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG018	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG019	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG020	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG021	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG022	G2a	TG	*R. aganniphum* var. *aganniphum*	rem[NA]	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG023	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG024	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG025	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 05.8" N	099° 04' 43.8" E	4316 m	07.09.2007
TG026[sp]	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TG027	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TG028	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG029	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG030	G1	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Shika Shan	27° 47' 16.8" N	099° 37' 31.7" E	3723 m	06.09.2007
TG031	G2a	TG	*R. aganniphum* var. *aganniphum*	rem[NA]	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG032	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG033	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG034	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG035	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG036	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG037	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG038	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG039	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG040	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG041	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG042	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG043	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG044	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG045	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG046	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TG047	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG048	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG049	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG050	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG051	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG052	G2a	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Baima Shan	28° 20' 15.0" N	099° 04' 37.1" E	4334 m	28.05.2008
TG053[sp]	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG054[sp]	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG055	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG056	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG057	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG058	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG059	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG060	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG061	G2b	TG	*R. aganniphum* var. *aganniphum*	rem[G2]	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TG062[sp]	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG063[sp]	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG064[sp]	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG065	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG066	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TG067	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG068	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG069	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG070	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG071	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG072	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG073	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG074	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG075	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG076	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG077	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG078	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG079	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG080	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG081	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG082	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG083	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG084	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG085	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG086	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TG087	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG088	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG089	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG090	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG091	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG092	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG093	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG094	G3	TG	*R. aganniphum* var. *aganniphum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TG095	G3	TG	*R. aganniphum* var. *aganniphum*	no[en]	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TGF01[sp]	GH2a	TGF	*R. aganniphum* var. *flavorufum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGF02[sp]	GH2a	TGF	*R. aganniphum* var. *flavorufum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGF03[sp]	GH2a	TGF	*R. aganniphum* var. *flavorufum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGF04[sp]	GH2a	TGF	*R. aganniphum* var. *flavorufum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGF05	GH2a	TGF	*R. aganniphum* var. *flavorufum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP01[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP02[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP03[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP04[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP05[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP06[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TGP07[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP08	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP09	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP10	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP11[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP12	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP13	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP14	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP15[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP16	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP17	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	rem[NA]	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP18	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP19	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP20	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP21[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP22	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP23	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP24	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP25[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TGP26[sp]	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 28.0' N	099° 05' 35.3" E	4155 m	09.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TGP27	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP28	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP29	GH2a	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP30^{sp}	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP31^{sp}	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP32^{sp}	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP33^{sp}	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP34^{sp}	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP35	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP36	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP37	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	rem^{NA}	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP38	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP39	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP40	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP41	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP42	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP43	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP44	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP45	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP46	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TGP47	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP48	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP49	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP50	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP51	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP52	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP53	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP54	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP55	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP56	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP57	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP58	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP59	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP60	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP61	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TGP62	GH2b	TGP	*R. aganniphum* × *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 17' 25.7" N	099° 07' 15.0" E	4300 m	29.05.2008
TP010[sp]	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP011	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP012	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP013	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TP014	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP015	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP016	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP017	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP018	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP019	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 24' 03.3" N	098° 58' 57.2" E	4032 m	08.09.2007
TP020	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP021[sp]	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP022	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP023[sp]	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP024[sp]	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP025	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP026[sp]	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP027	P1	TP	*R. phaeochrysum* var. *agglutinatum*	yes	China, Yunnan, Baima Shan	28° 19' 52.9" N	099° 05' 00.7" E	4271 m	09.09.2007
TP028	P1	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Baima Shan	28° 19' 23.6" N	099° 05' 30.4" E	4229 m	09.09.2007
TP029[sp]	P1	TP	*R. phaeochrysum* var. *agglutinatum*	yes	China, Yunnan, Baima Shan	28° 19' 28.0' N	099° 05' 35.3" E	4155 m	09.09.2007
TP030	P1	TP	*R. phaeochrysum* var. *agglutinatum*	yes	China, Yunnan, Baima Shan	28° 19' 28.0' N	099° 05' 35.3" E	4155 m	09.09.2007
TP031	P1	TP	*R. phaeochrysum* var. *agglutinatum*	yes	China, Yunnan, Baima Shan	28° 19' 28.0' N	099° 05' 35.3" E	4155 m	09.09.2007
TP065[sp]	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP066[sp]	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TP067[sp]	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP068	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes[NA]	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP069	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP070	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP071	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP072	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP073	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP074	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP075	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP076	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP077	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP078	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP079	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP080	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP081	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP082	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP083	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP084	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP085	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP086	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TP087	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP088	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TP089	P2	TP	*R. phaeochrysum* var. *phaeochrysum*	yes	China, Yunnan, Dàxuě Shan	28° 34' 37.7" N	099° 49' 35.3" E	4298 m	02.06.2008
TX012	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 50.0" N	099° 43' 33.6" E	3828 m	01.09.2007
TX013	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 50.0" N	099° 43' 33.6" E	3828 m	01.09.2007
TX014	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 50.0" N	099° 43' 33.6" E	3828 m	01.09.2007
TX015[sp]	X1	TX	*R. roxieanum* var. ?	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TX016[sp]	X1	TX	*R. roxieanum* var. ?	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TX017[sp]	X1	TX	*R. roxieanum* var. ?	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TX018[sp]	X1	TX	*R. roxieanum* var. ?	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TX019[sp]	X1	TX	*R. roxieanum* var. ?	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TX020[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TX021[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	01.09.2007
TX022[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TX023[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TX024[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TX025[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TX026[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TX027[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 37' 57.3" N	099° 42' 48.6" E	3955 m	02.09.2007
TX028[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	26° 38' 07.9" N	099° 42' 59.3" E	3886 m	02.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TX029[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX030	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX031	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX032	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX033	X1	TX	*R. roxieanum* var. *roxieanum*	rem[NA]	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX034	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX035	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX036	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX037	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX038	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX039	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX040	X1	TX	*R. roxieanum* var. *roxieanum*	yes[NA]	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX041	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX042	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX043	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX044[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX045	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	04.09.2007
TX046[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes[NA]	China, Yunnan, Lao Jun Shan	27° 47' 24.6" N	099° 37' 21.4" E	3757 m	06.09.2007
TX047[sp]	X1	TX	*Rhododendron roxieanum*	no[en]	China, Yunnan, Lao Jun Shan	27° 47' 24.6" N	099° 37' 21.4" E	3757 m	06.09.2007
TX048[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TX049[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX050[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX051	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX052	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX053	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX054	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX055	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX056	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX057	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX058	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX059	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX060	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX061	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX062	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX063	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX064	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 39' 06.3" N	099° 44' 13.0" E	3736 m	23.05.2008
TX065	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 38' 50.8" N	099° 44' 02.7" E	3867 m	23.05.2008
TX066	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 38' 50.8" N	099° 44' 02.7" E	3867 m	23.05.2008
TX067	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 38' 50.8" N	099° 44' 02.7" E	3867 m	23.05.2008
TX068	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 38' 50.8" N	099° 44' 02.7" E	3867 m	23.05.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TX069	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 38' 50.8" N	099° 44' 02.7" E	3867 m	23.05.2008
TX070	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 38' 50.8" N	099° 44' 02.7" E	3867 m	23.05.2008
TX071	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX072[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX073	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX074[sp]	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX075	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX076	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX077	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX078	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX079	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX080	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX081	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX082	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX083	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX084	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX085	X1	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 52.6" N	099° 43' 05.5" E	3945 m	22.05.2008
TX108	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX109	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX110	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TX111	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX112	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX113	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX114	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX115	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX116	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX117	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX118	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX119	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX120	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX121	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX122	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX123	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX124	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX125	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX126	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX127	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX128	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX129	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX130	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TX131	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX132	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX133	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX134[sp]	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX135[sp]	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX136	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX137	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX138	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX139	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TX140	X3	TX	*R. roxieanum* var. *roxieanum*	yes	China, Yunnan, Shika Shan	27° 47' 52.2" N	099° 35' 42.4" E	4087 m	31.05.2008
TXC01[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TXC02[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TXC03[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TXC04[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TXC05[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TXC06[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	rem[NA]	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TXC07[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.1" N	099° 43' 30.3" E	3819 m	02.09.2007
TXC08[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC09[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC10[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TXC11[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC12[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	rem[NA]	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC13[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC14[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC15[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC16[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.3" N	099° 43' 30.6" E	3838 m	02.09.2007
TXC17[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC18	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC19	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC20	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC21[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC22[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC23	XC1	TXC	*R. roxieanum* var. *cucullatum*	rem[NA]	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC24[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC25	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC26	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC27	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC28	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC29	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007
TXC30	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 49.8" N	099° 42' 54.3" E	3996 m	03.09.2007

Continued on next page...

Table D.1 – *Continued*

Individual	Pop[a]	Sp[a]	Species name	Used[b]	Location	Latitude	Longitude	Altitude	Date
TXC31[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC32[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC33	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC34	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC35	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC36	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC37	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC38	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC39	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC40	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC41	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC42	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 59.8" E	3962 m	22.05.2008
TXC43[sp]	XC1	TXC	*R. roxieanum* var. *cucullatum*	yes	China, Yunnan, Lao Jun Shan	26° 37' 51.3" N	099° 42' 24.7" E	4000 m	22.05.2008

[sp] For these individuals a herbarium specimen has been collected.

[a] Code used to designate population (pop), and species (sp).

[b] yes = the individual has been used in this study; no = the individual has been excluded from the beginning, for reasons given below; rem = AFLPs have been generated, but the individual was then removed for reasons given.

[NA] Data was missing; if used = yes, this individual was in the base dataset, but erroneously removed from the final dataset.

[G2] The population G2b was considered too small for statistical analysis, and seemed too admixed to pool with G2a.

[en] Enough: already enough samples of this population were included, predominantly individuals with unclear morphology were then not used.

D.2 Tables for the leaf wax analysis

Table D.2: Individuals used for leaf wax extraction and analysis. Leaf area used for extraction, and amount of wax recovered after removal of the non-polar fraction.

Species	Individual	Leaf Area (cm^2)	Wax recovered (mg)
R. clementinae	TC001	50	3.5
	TC002	50	2.8
	TC003	50	4.0
	TC004	50	3.8
	TC005	50	3.3
	TC006	50	3.4
	TC014	50	3.6
	TC020	40	2.3
	TC022	60	2.9
	TC027	30	2.6
	TC047	40	4.3
	TC049	50	3.2
	TC053	50	4.3
	TC055	50	12.9
	TC058	50	3.7
	TC061	50	3.4
	TC064	60	3.0
	TC067	50	3.5
	TC070	30	3.2
	TC071	75	3.3
R. aganniphum	TG004	35	4.0
	TG007	30	12.4
	TG009	50	4.5
	TG016	25	9.0
	TG024	15	14.1
	TG035	15	15.3
	TG038	30	2.3
	TG044	30	2.4
	TG047	20	2.1

Continued on next page...

TableD.2 – Continued

Species	Individual	Leaf Area (cm^2)	Wax recovered (mg)
	TG052	20	2.0
	TG056	30	1.0
	TG061	20	0.8
	TG066	40	1.3
	TG071	35	0.9
	TG074	40	3.8
	TG077	45	0.9
	TG081	30	1.8
	TG083	50	0.7
	TG087	15	1.2
	TG092	30	1.6
R. aganniphum	TGF01	10	5.8
× *R. phaeochrysum*	TGF02	20	2.9
	TGF03	10	0.6
	TGF04	50	6.9
	TGF05	20	1.9
	TGP02	40	5.4
	TGP05	40	7.3
	TGP15	25	1.1
	TGP21	45	5.4
	TGP25	15	1.0
	TGP30	20	1.2
	TGP31	40	10.0
	TGP32	15	8.2
	TGP33	35	3.0
	TGP34	50	2.0
	TGP38	30	2.2
	TGP41	25	1.5
	TGP48	25	6.9
	TGP54	40	3.9
	TGP60	20	3.3

Continued on next page...

TableD.2 – Continued

Species	Individual	Leaf Area (cm^2)	Wax recovered (mg)
R. phaeochrysum	TP010	40	12.6
	TP011	20	0.8
	TP015	60	2.2
	TP017	40	1.3
	TP018	40	2.4
	TP021	40	2.3
	TP024	15	0.8
	TP027	10	2.0
	TP029	25	3.4
	TP030	25	8.9
	TP031	35	10.0
	TP066	50	1.6
	TP067	50	1.9
	TP070	50	11.6
	TP072	60	4.1
	TP073	50	3.1
	TP077	40	1.9
	TP081	70	3.2
	TP083	70	2.9
	TP086	90	3.1
R. roxieanum	TX029	15	3.6
	TX044	40	10.2
	TX048	10	2.6
	TX049	10	2.6
	TX050	20	6.6
	TX054	20	8.9
	TX058	20	7.1
	TX059	20	4.7
	TX062	15	3.8
	TX066	15	5.0
	TX072	20	5.4
	TX074	20	8.8
	TX081	15	4.6

Continued on next page...

TableD.2 – Continued

Species	Individual	Leaf Area (cm^2)	Wax recovered (mg)
	TX084	60	9.7
	TX110	10	4.5
	TX111	20	3.4
	TX117	20	7.2
	TX121	25	5.6
	TX134	25	4.5
	TX135	15	6.0
R. roxieanum	TXC01	40	5.6
var. *cucullatum*	TXC02	55	7.1
	TXC03	30	4.3
	TXC04	60	6.2
	TXC05	40	6.8
	TXC07	20	2.8
	TXC08	40	8.4
	TXC09	60	4.5
	TXC10	50	5.7
	TXC11	50	6.3
	TXC29	85	8.4
	TXC34	20	3.6
	TXC38	20	6.0
	TXC41	30	6.8
	TXC43	30	7.5

Table D.3: Leaf wax results; Size of wax fraction for each individual, given by the area obtained from the GC analysis of each sample, and respective percentage after scaling to the largest fraction of the trace.

		Area of fraction			% of largest area		
Individual	Species	C27	C29	C31	C27	C29	C31
TC001	TC	9574	69117	142749	6.7	48.4	100.0
TC002	TC	8110	51128	99550	8.1	51.4	100.0
TC003	TC	3511	14547	35977	9.8	40.4	100.0
TC004	TC	6585	34043	65985	10.0	51.6	100.0
TC005	TC	8962	57888	116106	7.7	49.9	100.0
TC006	TC	6549	37745	81998	8.0	46.0	100.0
TC014	TC	8540	51700	98386	8.7	52.5	100.0
TC020	TC	6192	35956	73776	8.4	48.7	100.0
TC022	TC	10702	54254	145204	7.4	37.4	100.0
TC027	TC	10256	45786	105586	9.7	43.4	100.0
TC047	TC	18111	73823	154788	11.7	47.7	100.0
TC049	TC	15413	102040	203778	7.6	50.1	100.0
TC053	TC	8501	54964	115106	7.4	47.8	100.0
TC055	TC	4401	34301	81219	5.4	42.2	100.0
TC058	TC	8334	47460	121962	6.8	38.9	100.0
TC061	TC	15181	111552	217148	7.0	51.4	100.0
TC064	TC	14034	73860	157544	8.9	46.9	100.0
TC067	TC	12351	82164	159882	7.7	51.4	100.0
TC070	TC	11369	57082	145052	7.8	39.4	100.0
TC071	TC	14806	89811	215080	6.9	41.8	100.0
TG004	TG	15164	65077	147581	10.3	44.1	100.0
TG007	TG	2223	12964	44071	5.0	29.4	100.0
TG009	TG	5587	48945	116670	4.8	42.0	100.0
TG016	TG	12291	90047	172135	7.1	52.3	100.0
TG024	TG	3669	29930	65643	5.6	45.6	100.0
TG035	TG	2321	14319	37236	6.2	38.5	100.0
TG038	TG	8136	77491	173457	4.7	44.7	100.0
TG044	TG	12994	45495	99341	13.1	45.8	100.0
TG047	TG	8706	41703	91910	9.5	45.4	100.0
TG052	TG	6182	59808	136660	4.5	43.8	100.0

Continued on next page...

Table D.3 – *Continued*

		Area of fraction			% of largest area		
Individual	Species	C27	C29	C31	C27	C29	C31
TG056	TG	6286	54942	107399	5.9	51.2	100.0
TG061	TG	9227	33498	65504	14.1	51.1	100.0
TG066	TG	9565	39981	64297	14.9	62.2	100.0
TG071	TG	10322	59472	73427	14.1	81.0	100.0
TG074	TG	13375	65722	108258	12.4	60.7	100.0
TG077	TG	8756	54555	94102	9.3	58.0	100.0
TG081	TG	10761	43490	70707	15.2	61.5	100.0
TG083	TG	13162	86130	136029	9.7	63.3	100.0
TG087	TG	7946	88524	129000	6.2	68.6	100.0
TG092	TG	12294	61814	229624	5.4	26.9	100.0
TGF01	TGF	4378	51754	96692	4.5	53.5	100.0
TGF02	TGF	8097	45512	106884	7.6	42.6	100.0
TGF03	TGF	2464	42805	93742	2.6	45.7	100.0
TGF04	TGF	28228	78577	208223	13.6	37.7	100.0
TGF05	TGF	8749	43269	77674	11.3	55.7	100.0
TGP02	TGP	10031	66383	142840	7.0	46.5	100.0
TGP05	TGP	3715	42998	113147	3.3	38.0	100.0
TGP15	TGP	10623	52155	80530	13.2	64.8	100.0
TGP21	TGP	32190	106220	151585	21.2	70.1	100.0
TGP25	TGP	6281	61052	92883	6.8	65.7	100.0
TGP30	TGP	5428	35154	83258	6.5	42.2	100.0
TGP31	TGP	5051	42563	64043	7.9	66.5	100.0
TGP32	TGP	4741	36228	83882	5.7	43.2	100.0
TGP33	TGP	10427	83062	131560	7.9	63.1	100.0
TGP34	TGP	8867	73537	207121	4.3	35.5	100.0
TGP38	TGP	14639	41407	114949	12.7	36.0	100.0
TGP41	TGP	8220	74567	168849	4.9	44.2	100.0
TGP48	TGP	7535	58528	112072	6.7	52.2	100.0
TGP54	TGP	8597	46642	118565	7.3	39.3	100.0
TGP60	TGP	5451	43758	91694	5.9	47.7	100.0
TP010	TP	3180	15777	55522	5.7	28.4	100.0

Continued on next page...

Table D.3 – *Continued*

		Area of fraction			% of largest area		
Individual	Species	C27	C29	C31	C27	C29	C31
TP011	TP	5013	62172	157691	3.2	39.4	100.0
TP015	TP	11603	87214	260141	4.5	33.5	100.0
TP017	TP	36632	98413	143557	25.5	68.6	100.0
TP018	TP	6159	52816	131890	4.7	40.0	100.0
TP021	TP	367528	162692	79666	100.0	44.3	21.7
TP024	TP	129137	125421	94524	100.0	97.1	73.2
TP027	TP	10026	63025	159141	6.3	39.6	100.0
TP029	TP	6911	67565	97572	7.1	69.2	100.0
TP030	TP	11073	68654	100165	11.1	68.5	100.0
TP031	TP	2449	34871	63363	3.9	55.0	100.0
TP066	TP	6991	59946	242802	2.9	24.7	100.0
TP067	TP	10197	79116	231058	4.4	34.2	100.0
TP070	TP	3153	29585	64914	4.9	45.6	100.0
TP072	TP	15705	124800	348250	4.5	35.8	100.0
TP073	TP	27889	83691	347747	8.0	24.1	100.0
TP077	TP	6503	57429	214634	3.0	26.8	100.0
TP081	TP	17265	113092	273436	6.3	41.4	100.0
TP083	TP	14290	144624	417392	3.4	34.6	100.0
TP086	TP	19471	123260	476167	4.1	25.9	100.0
TX029	TX	171437	204078	143670	84.0	100.0	70.4
TX044	TX	538347	232656	152032	100.0	43.2	28.2
TX048	TX	379227	288269	126929	100.0	76.0	33.5
TX049	TX	248355	189271	93829	100.0	76.2	37.8
TX050	TX	901858	445209	144652	100.0	49.4	16.0
TX054	TX	220124	128937	90195	100.0	58.6	41.0
TX058	TX	270486	211417	85083	100.0	78.2	31.5
TX059	TX	444015	224400	154290	100.0	50.5	34.7
TX062	TX	314736	173940	98512	100.0	55.3	31.3
TX066	TX	687718	205810	89315	100.0	29.9	13.0
TX072	TX	618452	234958	137050	100.0	38.0	22.2
TX074	TX	325479	134443	88802	100.0	41.3	27.3

Continued on next page...

Table D.3 – *Continued*

		Area of fraction			% of largest area		
Individual	Species	C27	C29	C31	C27	C29	C31
TX081	TX	509207	277620	141163	100.0	54.5	27.7
TX084	TX	501786	257720	76617	100.0	51.4	15.3
TX110	TX	309442	138073	99641	100.0	44.6	32.2
TX111	TX	406382	186715	112951	100.0	45.9	27.8
TX117	TX	587418	271677	212260	100.0	46.2	36.1
TX121	TX	310694	277222	162369	100.0	89.2	52.3
TX134	TX	329867	176075	127741	100.0	53.4	38.7
TX135	TX	303060	183524	113737	100.0	60.6	37.5
TXC01	TXC	190651	153220	99234	100.0	80.4	52.1
TXC02	TXC	190933	196836	127688	97.0	100.0	64.9
TXC03	TXC	89244	109279	71872	81.7	100.0	65.8
TXC04	TXC	197225	147866	95355	100.0	75.0	48.3
TXC05	TXC	412615	188129	141046	100.0	45.6	34.2
TXC07	TXC	93411	102221	66146	91.4	100.0	64.7
TXC08	TXC	101254	136679	113743	74.1	100.0	83.2
TXC09	TXC	50656	117465	133071	38.1	88.3	100.0
TXC10	TXC	87585	165789	152018	52.8	100.0	91.7
TXC11	TXC	270842	150039	99509	100.0	55.4	36.7
TXC29	TXC	350498	139816	95064	100.0	39.9	27.1
TXC34	TXC	271440	127809	115558	100.0	47.1	42.6
TXC38	TXC	362088	107433	74350	100.0	29.7	20.5
TXC41	TXC	345979	165917	106591	100.0	48.0	30.8
TXC43	TXC	313476	149741	114714	100.0	47.8	36.6

Appendix E

Additional material for AMOVA

E.1 Derivation of equations 9a–c from Excoffier et al.

The equations 9a–c (for n, n', n") in Excoffier et al. [54] are easily derived from equations 2.18, 2.20, and 2.17 in subsection 2.3.5. In accordance with [54] we call the number of groups G, and the number of subgroups in group g I_g. Because all hierarchical levels of the dataset have the same total sample size N, it follows that:

$$N = \sum_{g=1}^{G} N_g = \sum_{g=1}^{G} \sum_{i=1}^{I_g} N_{ig}$$

Equation 9a (n)

Equation 2.18, page 75, is equivalent to equation 9a in Excoffier et al. and is rewritten as follows:

$$n = n_{L-1} = \frac{\sum_{g=1}^{G} [N_g(1 - \alpha_g)]}{df_{L-1}} \tag{E.1a}$$

by substituting α with equation 2.16, page 75, and calling the number of subgroups in group g I_g we get:

$$n = \frac{\sum_{g=1}^{G}\left[N_g\left(1 - \frac{\sum_{i=1}^{I_g}(N_{ig}^2)}{(\sum_{i=1}^{I_g} N_{ig})^2}\right)\right]}{df_{L-1}} \tag{E.1b}$$

then, as $N_g = \sum_{i=1}^{I_g} N_{ig}$, we can write:

$$n = \frac{\sum_{g=1}^{G}\left[N_g - \frac{\sum_{i=1}^{I_g}(N_{ig}^2)}{\sum_{i=1}^{I_g} N_{ig}}\right]}{df_{L-1}} \tag{E.1c}$$

$$= \frac{\sum_{g=1}^{G} N_g - \sum_{g=1}^{G}\left(\frac{\sum_{i=1}^{I_g}(N_{ig}^2)}{\sum_{i=1}^{I_g} N_{ig}}\right)}{df_{L-1}} \tag{E.1d}$$

because $\sum_{g=1}^{G} N_g = \sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig}$,
and with the corresponding degrees of freedom ($df_{L-1} = G_{L-1} - G_L = \sum_{g=1}^{G}(I_g) - G$), we finally get equation 9a:

$$n = \frac{\sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig} - \sum_{g=1}^{G}\left(\frac{\sum_{i=1}^{I_g} N_{ig}^2}{\sum_{i=1}^{I_g} N_{ig}}\right)}{\sum_{g=1}^{G} I_g - G} \tag{E.1e}$$

In Excoffier et al. the denominator of equation 9a is incorrectly stated as the total number of subgroups ($\sum_{g=1}^{G} I_g$), but see paragraph 2.3.5, page 76.

Equation 9b (n')

Equation 2.20, page 76, is equivalent to equation 9b in Excoffier et al., and is reformulated as follows:

$$n' = n'_{L\setminus L-1} = \frac{N(1-\alpha'_{L-1})}{df_L} \tag{E.2a}$$

by replacing α'_{L-1} with equation 2.19, page 76 we get:

$$n' = \frac{N\left(1-\alpha_{L-1}-\sum_{g=1}^{G}\left[\frac{N_g}{N}(1-\alpha_g)\right]\right)}{df_L} \tag{E.2b}$$

$$= \frac{N-N\alpha_{L-1}-\sum_{g=1}^{G}[N_g(1-\alpha_g)]}{df_L} \tag{E.2c}$$

substituting the α's with equation 2.16, page 75, and terming the number of subgroups in group g I_g,

$$n' = \frac{N - N\left(\frac{\sum_{g=1}^{G}\sum_{i=1}^{I_g}(N_{ig})^2}{(\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig})^2}\right) - \sum_{g=1}^{G}\left[N_g\left(1-\frac{\sum_{i=1}^{I_g}(N_{ig})^2}{(\sum_{i=1}^{I_g}N_{ig})^2}\right)\right]}{df_L} \tag{E.2d}$$

because $N = \sum_{g=1}^{G} N_g = \sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig}$,
and $N_g = \sum_{i=1}^{I_g} N_{ig}$, we can rewrite equation E.2d as:

$$n' = \frac{N - \dfrac{\sum_{g=1}^{G}\sum_{i=1}^{I_g}(N_{ig})^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}} - \sum_{g=1}^{G}\left[N_g - \dfrac{\sum_{i=1}^{I_g}(N_{ig})^2}{\sum_{i=1}^{I_g}N_{ig}}\right]}{df_L} \tag{E.2e}$$

$$= \frac{N - \dfrac{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}} - \sum_{g=1}^{G}N_g + \sum_{g=1}^{G}\left[\dfrac{\sum_{i=1}^{I_g}N_{ig}^2}{\sum_{i=1}^{I_g}N_{ig}}\right]}{df_L} \tag{E.2f}$$

$$= \frac{\sum_{g=1}^{G}\left[\dfrac{\sum_{i=1}^{I_g}N_{ig}^2}{\sum_{i=1}^{I_g}N_{ig}}\right] - \dfrac{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}}}{df_L} \tag{E.2g}$$

By substituting the corresponding degrees of freedom $df_L = G - 1$, we obtain equation 9b:

$$n' = \frac{\sum_{g=1}^{G}\left(\dfrac{\sum_{i=1}^{I_g}N_{ig}^2}{\sum_{i=1}^{I_g}N_{ig}}\right) - \dfrac{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}}}{G-1} \tag{E.2h}$$

Equation 9c (n")

Equation 2.17 page 75 is equivalent to equation 9c in Excoffier et al. and is rewritten as follows:

$$n'' = n_L = \frac{N(1-\alpha_L)}{df_L} \tag{E.3a}$$

substituting α with equation 2.16 we get:

$$n'' = \frac{N\left(1 - \frac{\sum_{g=1}^{G}(N_g^2)}{(\sum_{g=1}^{G} N_g)^2}\right)}{df_L} \tag{E.3b}$$

because $N = \sum_{g=1}^{G} N_g = \sum_{g=1}^{G} \sum_{i=1}^{I_g} N_{ig}$,
and $N_g = \sum_{i=1}^{I_g} N_{ig}$, equation E.3b can be written as:

$$n'' = \frac{N - \frac{\sum_{g=1}^{G} N_g^2}{\sum_{g=1}^{G} N_g}}{df_L} \tag{E.3c}$$

$$= \frac{N - \frac{\sum_{g=1}^{G}\left(\sum_{i=1}^{I_g} N_{ig}\right)^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig}}}{df_L} \tag{E.3d}$$

$$n'' = \frac{\sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig} - \frac{\sum_{g=1}^{G}\left(\sum_{i=1}^{I_g} N_{ig}\right)^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig}}}{df_L} \tag{E.3e}$$

and with $df_L = G - 1$ formula 9c is obtained:
(see next page)

$$n'' = \frac{\sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig} - \frac{\sum_{g=1}^{G}\left(\sum_{i=1}^{I_g} N_{ig}\right)^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g} N_{ig}}}{G-1} \tag{E.3f}$$

E.2 Tables of variance components and Φ-statistics that were not included in the main text

Table E.1: AMOVA **pairwise Φ_{SG}-values** for species comparisons. For all comparisons p <0.001, calculated with 5000 datasets, by permuting individuals randomly in the whole subset of the species pair.

	TC	TG	TGP	TP	TX	TXC
TC	•	$\frac{0.385}{0.543}$	$\frac{0.329}{0.479}$	$\frac{0.347}{0.529}$	$\frac{0.393}{0.558}$	$\frac{0.378}{0.562}$
TG	0.467	•	$\frac{0.099}{0.209}$	$\frac{0.270}{0.423}$	$\frac{0.333}{0.484}$	$\frac{0.325}{0.476}$
TGP	0.404	0.151	•	$\frac{0.179}{0.311}$	$\frac{0.249}{0.395}$	$\frac{0.228}{0.378}$
TP	0.443	0.349	0.245	•	$\frac{0.235}{0.386}$	$\frac{0.216}{0.372}$
TX	0.479	0.412	0.323	0.312	•	$\frac{0.063}{0.130}$
TXC	0.475	0.403	0.304	0.297	0.095	•

Table E.2: AMOVA **pairwise Φ_{ST}-values** for species comparisons. For all comparisons p <0.001, calculated with 5000 datasets, by permuting individuals randomly within each species of the species pair.

	TC	TG	TGP	TP	TX	TXC
TC	•	$\frac{0.137}{0.231}$	$\frac{0.071}{0.143}$	$\frac{0.117}{0.210}$	$\frac{0.133}{0.227}$	$\frac{0.111}{0.236}$
TG	0.182	•	$\frac{0.074}{0.155}$	$\frac{0.107}{0.204}$	$\frac{0.127}{0.213}$	$\frac{0.109}{0.241}$
TGP	0.106	0.111	•	$\frac{0.036}{0.084}$	$\frac{0.073}{0.138}$	$\frac{0.019}{0.060}$
TP	0.164	0.153	0.059	•	$\frac{0.107}{0.189}$	$\frac{0.052}{0.164}$
TX	0.177	0.168	0.103	0.147	•	$\frac{0.110}{0.218}$
TXC	0.176	0.171	0.039	0.109	0.162	•

Table E.3: AMOVA **results for individual species (Φ_{ST}).**

Species					Variance		% total		
Variance component	*df*	SSD	MSD		Observed	95% conf[a]	Observed	95% conf[a]	*p*-value[b]
R. clementinae (TC)									
Between populations	1	0.580	0.580	σ_b^2	0.014	0.008 – 0.021	19.8	13.1 – 26.4	<0.001
Within populations	71	4.112	0.058	σ_c^2	0.058	0.045 – 0.072	80.2	73.6 – 86.9	<0.001
Total	72	4.692	0.065	σ_T^2	0.072	0.055 – 0.090	100.0		<0.001
R. aganniphum (TG)									
Between populations	2	0.982	0.491	σ_b^2	0.017	0.010 – 0.026	16.7	10.9 – 23.3	<0.001
Within populations	73	6.342	0.087	σ_c^2	0.087	0.073 – 0.101	83.3	76.7 – 89.1	<0.001
Total	75	7.323	0.098	σ_T^2	0.104	0.088 – 0.122	100.0		<0.001
R. aganniphum* × *phaeochrysum (TGP)									
Between populations	1	0.216	0.216	σ_b^2	0.004	0.002 – 0.006	3.4	1.6 – 5.2	<0.001
Within populations	63	6.359	0.101	σ_c^2	0.101	0.086 – 0.117	96.6	94.8 – 98.4	<0.001
Total	64	6.575	0.103	σ_T^2	0.104	0.089 – 0.121	100.0		<0.001
R. phaeochrysum (TP)									
Between populations	1	0.273	0.273	σ_b^2	0.010	0.005 – 0.016	11.1	5.7 – 16.6	<0.001
Within populations	39	3.023	0.078	σ_c^2	0.078	0.063 – 0.092	88.9	83.4 – 94.3	<0.001
Total	40	3.297	0.082	σ_T^2	0.087	0.070 – 0.104	100.0		<0.001
R. roxieanum (TX)									
Between populations	1	0.689	0.689	σ_b^2	0.015	0.009 – 0.022	16.4	11.3 – 22.0	<0.001
Within populations	84	6.436	0.077	σ_c^2	0.077	0.063 – 0.091	83.6	78.0 – 88.7	<0.001
Total	85	7.125	0.084	σ_T^2	0.092	0.075 – 0.109	100.0		<0.001

Note: TXC was not included because only one population was available.

[a] Calculated from 5000 bootstrap replicates.

[b] Calculated from 5000 permutations of the respective hierarchical level, see subsection 2.3.5, page 82.

Table E.4: AMOVA **results for pairwise species comparisons.**

Species pair					Variance		% total		
Variance component	*df*	SSD	MSD		Observed	95% conf[a]	Observed	95% conf[a]	*p*-value[b]
TC - TG									
Between species	1	4.144	4.144	σ_a^2	0.047	0.031 – 0.067	34.8	25.4 – 43.6	0.042*
Between populations w/ species	3	1.562	0.521	σ_b^2	0.016	0.011 – 0.022	11.9	8.5 – 15.8	<0.001
Within populations	144	10.454	0.073	σ_c^2	0.073	0.062 – 0.083	53.3	45.7 – 61.5	<0.001
Total	148	16.159	0.109	σ_T^2	0.136	0.115 – 0.160	100.0		<0.001
TC - TGP									
Between species	1	3.399	3.399	σ_a^2	0.044	0.029 – 0.060	33.3	24.9 – 41.5	0.149[a]
Between populations w/ species	2	0.796	0.398	σ_b^2	0.009	0.006 – 0.013	7.1	4.6 – 10.0	<0.001
Within populations	134	10.471	0.078	σ_c^2	0.078	0.068 – 0.089	59.6	52.1 – 67.1	<0.001
Total	137	14.666	0.107	σ_T^2	0.131	0.111 – 0.152	100.0		<0.001
TC - TP									
Between species	1	2.445	2.445	σ_a^2	0.039	0.023 – 0.057	33.4	22.5 – 43.6	0.156
Between populations w/ species	2	0.853	0.427	σ_b^2	0.013	0.008 – 0.018	11.0	7.4 – 14.9	<0.001
Within populations	110	7.136	0.065	σ_c^2	0.065	0.054 – 0.076	55.7	47.1 – 65.3	<0.001
Total	113	10.433	0.092	σ_T^2	0.117	0.095 – 0.138	100.0		<0.001

Continued on next page...

Table E.4 – *Continued*

Variance component	*df*	SSD	MSD	Variance		% total		
				Observed	95% conf[a]	Observed	95% conf[a]	*p*-value[b]
TC - TX								
Between species	1	4.448	4.448	σ_a^2 0.048	0.030 – 0.068	36.7	26.3 – 46.6	0.160
Between populations w/ species	2	1.269	0.635	σ_b^2 0.015	0.010 – 0.020	11.2	7.8 – 15.4	<0.001
Within populations	155	10.548	0.068	σ_c^2 0.068	0.058 – 0.078	52.1	44.2 – 60.8	<0.001
Total	158	16.265	0.103	σ_T^2 0.131	0.108 – 0.153	100.0		<0.001
TC - TXC								
Between species	1	2.913	2.913	σ_a^2 0.045	0.026 – 0.067	36.2	23.5 – 47.7	0.168
Between populations w/ species	1	0.580	0.580	σ_b^2 0.014	0.008 – 0.021	11.2	6.4 – 16.7	<0.001
Within populations	109	7.179	0.066	σ_c^2 0.066	0.055 – 0.077	52.5	43.8 – 62.2	<0.001
Total	111	10.672	0.096	σ_T^2 0.125	0.103 – 0.148	100.0		<0.001
TC - TGP								
Between species	1	0.803	0.803	σ_a^2 0.005	0.000 – 0.012	4.5	-0.2 – 10.0	0.241
Between populations w/ species	3	1.198	0.399	σ_b^2 0.012	0.007 – 0.017	10.6	6.9 – 15.0	<0.001
Within populations	136	12.701	0.093	σ_c^2 0.093	0.080 – 0.108	84.9	79.1 – 90.1	<0.001
Total	140	14.702	0.105	σ_T^2 0.110	0.094 – 0.127	100.0		<0.001

Continued on next page...

Table E.4 – *Continued*

Variance component	*df*	SSD	MSD		Variance		% total		
					Observed	95% conf[a]	Observed	95% conf[a]	*p*-value[b]
TG - TP									
Between species	1	2.025	2.025	σ_a^2	0.030	0.017 – 0.044	23.1	14.8 – 31.0	0.024
Between populations w/ species	3	1.255	0.418	σ_b^2	0.015	0.010 – 0.021	11.7	8.2 – 16.1	<0.001
Within populations	112	9.365	0.084	σ_c^2	0.084	0.071 – 0.096	65.1	57.7 – 73.0	<0.001
Total	116	12.645	0.109	σ_T^2	0.128	0.108 – 0.150	100.0		<0.001
TG - TX									
Between species	1	3.961	3.961	σ_a^2	0.041	0.025 – 0.057	29.3	20.6 – 37.4	0.055
Between populations w/ species	3	1.671	0.557	σ_b^2	0.016	0.012 – 0.022	11.8	8.7 – 15.7	<0.001
Within populations	157	12.777	0.081	σ_c^2	0.081	0.071 – 0.092	58.8	51.6 – 66.7	<0.001
Total	161	18.409	0.114	σ_T^2	0.138	0.118 – 0.158	100.0		<0.001
TG - TXC									
Between species	1	2.755	2.755	σ_a^2	0.040	0.024 – 0.057	28.0	18.6 – 36.7	0.077
Between populations w/ species	2	0.982	0.491	σ_b^2	0.017	0.010 – 0.026	12.3	7.6 – 18.3	<0.001
Within populations	111	9.408	0.085	σ_c^2	0.085	0.074 – 0.096	59.7	52.4 – 67.5	<0.001
Total	114	13.145	0.115	σ_T^2	0.142	0.121 – 0.163	100.0		<0.001

Continued on next page...

Table E.4 – *Continued*

Variance component	*df*	SSD	MSD		Variance Observed	Variance 95% conf[a]	% total Observed	% total 95% conf[a]	*p*-value[b]
TGP - TP									
Between species	1	1.449	1.449	σ_a^2	0.024	0.014 – 0.035	19.7	13.2 – 26.6	0.155
Between populations w/ species	2	0.490	0.245	σ_b^2	0.006	0.003 – 0.009	4.8	2.8 – 6.8	<0.001
Within populations	102	9.382	0.092	σ_c^2	0.092	0.079 – 0.105	75.5	68.9 – 82.1	<0.001
Total	105	11.321	0.108	σ_T^2	0.122	0.103 – 0.141	100.0		<0.001
TGP - TX									
Between species	1	2.800	2.800	σ_a^2	0.032	0.019 – 0.046	24.5	16.2 – 32.7	0.164
Between populations w/ species	2	0.906	0.453	σ_b^2	0.010	0.007 – 0.014	7.8	5.3 – 10.7	<0.001
Within populations	147	12.795	0.087	σ_c^2	0.087	0.076 – 0.098	67.7	60.5 – 75.1	<0.001
Total	150	16.500	0.110	σ_T^2	0.129	0.111 – 0.147	100.0		<0.001
TGP - TXC									
Between species	1	2.036	2.036	σ_a^2	0.037	0.024 – 0.052	27.6	19.7 – 35.3	0.149
Between populations w/ species	1	0.216	0.216	σ_b^2	0.004	0.002 – 0.006	2.8	1.4 – 4.4	<0.001
Within populations	101	9.426	0.093	σ_c^2	0.093	0.082 – 0.106	69.6	62.2 – 77.2	<0.001
Total	103	11.678	0.113	σ_T^2	0.134	0.115 – 0.154	100.0		<0.001

Continued on next page...

Table E.4 – *Continued*

Variance component	*df*	SSD	MSD	Variance		% total		
				Observed	95% conf[a]	Observed	95% conf[a]	*p*-value[b]
TP - TX								
Between species	1	1.658	1.658	σ_a^2 0.022	0.011 – 0.033	19.3	11.3 – 27.3	0.163
Between populations w/ species	2	0.963	0.481	σ_b^2 0.013	0.009 – 0.018	11.9	8.5 – 15.8	<0.001
Within populations	123	9.459	0.077	σ_c^2 0.077	0.065 – 0.089	68.8	61.4 – 76.5	<0.001
Total	126	12.080	0.096	σ_T^2 0.112	0.093 – 0.131	100.0		<0.001
TP - TXC								
Between species	1	1.317	1.317	σ_a^2 0.024	0.013 – 0.036	21.1	13.1 – 29.7	0.266
Between populations w/ species	1	0.273	0.273	σ_b^2 0.010	0.004 – 0.016	8.6	4.0 – 13.3	<0.001
Within populations	77	6.090	0.079	σ_c^2 0.079	0.068 – 0.091	70.3	62.8 – 78.4	<0.001
Total	79	7.680	0.097	σ_T^2 0.112	0.094 – 0.131	100.0		<0.001
TX - TXC								
Between species	1	0.326	0.326	σ_a^2 -0.007	-0.012 – -0.003	-8.0	-13.3 – -3.2	0.552
Between populations w/ species	1	0.689	0.689	σ_b^2 0.015	0.009 – 0.022	17.5	11.5 – 24.5	<0.001
Within populations	122	9.502	0.078	σ_c^2 0.078	0.064 – 0.092	90.5	87.0 – 93.7	<0.001
Total	124	10.517	0.085	σ_T^2 0.086	0.071 – 0.102	100.0		<0.001

[a] Calculated from 5000 bootstrap replicates.

[b] Calculated from 5000 permutations of the respective hierarchical level, see subsection 2.3.5, page 82.

Table E.5: AMOVA **results for pairwise population comparisons**

Population pair					Variance		% total	
Variance component	*df*	SSD	MSD		Observed	95% conf[a]	Observed	95% conf[a]
C1 - C2								
Between populations	1	0.580	0.580	σ_b^2	0.014	0.008 – 0.021	19.8	13.1 – 26.4
Within populations	71	4.112	0.058	σ_c^2	0.058	0.045 – 0.072	80.2	73.6 – 86.9
Total	72	4.692	0.065	σ_T^2	0.072	0.055 – 0.090	100.0	
C1 - G1								
Between populations	1	1.074	1.074	σ_b^2	0.059	0.041 – 0.079	47.0	37.4 – 55.6
Within populations	46	3.084	0.067	σ_c^2	0.067	0.055 – 0.080	53.0	44.4 – 62.6
Total	47	4.157	0.088	σ_T^2	0.126	0.103 – 0.151	100.0	
C1 - G2								
Between populations	1	2.490	2.490	σ_b^2	0.067	0.047 – 0.090	47.6	38.4 – 55.9
Within populations	70	5.186	0.074	σ_c^2	0.074	0.063 – 0.086	52.4	44.1 – 61.6
Total	71	7.676	0.108	σ_T^2	0.141	0.118 – 0.167	100.0	
C1 - G3								
Between populations	1	2.220	2.220	σ_b^2	0.065	0.044 – 0.088	48.5	37.8 – 57.3
Within populations	65	4.488	0.069	σ_c^2	0.069	0.058 – 0.081	51.5	42.7 – 62.2
Total	66	6.708	0.102	σ_T^2	0.134	0.110 – 0.161	100.0	
C1 - GH2a								
Between populations	1	1.961	1.961	σ_b^2	0.054	0.038 – 0.072	40.3	32.1 – 48.2
Within populations	68	5.427	0.080	σ_c^2	0.080	0.069 – 0.091	59.7	51.8 – 67.9
Total	69	7.388	0.107	σ_T^2	0.134	0.113 – 0.156	100.0	
C1 - GH2b								
Between populations	1	1.895	1.895	σ_b^2	0.053	0.037 – 0.071	40.5	32.2 – 48.2
Within populations	67	5.210	0.078	σ_c^2	0.078	0.067 – 0.089	59.5	51.8 – 67.8
Total	68	7.105	0.104	σ_T^2	0.131	0.109 – 0.153	100.0	
C1 - P1								
Between populations	1	1.183	1.183	σ_b^2	0.046	0.029 – 0.065	40.6	30.2 – 50.2
Within populations	53	3.565	0.067	σ_c^2	0.067	0.056 – 0.079	59.4	49.8 – 69.8
Total	54	4.748	0.088	σ_T^2	0.113	0.092 – 0.136	100.0	
C1 - P2								
Between populations	1	1.756	1.756	σ_b^2	0.060	0.038 – 0.082	48.1	36.9 – 57.6
Within populations	58	3.735	0.064	σ_c^2	0.064	0.053 – 0.076	51.9	42.4 – 63.1
Total	59	5.491	0.093	σ_T^2	0.124	0.100 – 0.149	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD		Variance		% total	
					Observed	95% conf[a]	Observed	95% conf[a]
C1 - X1								
Between populations	1	3.008	3.008	σ_b^2	0.067	0.047 – 0.089	49.3	40.1 – 57.6
Within populations	88	6.101	0.069	σ_c^2	0.069	0.058 – 0.081	50.7	42.4 – 59.9
Total	89	9.109	0.102	σ_T^2	0.137	0.112 – 0.161	100.0	
C1 - X3								
Between populations	1	2.190	2.190	σ_b^2	0.061	0.042 – 0.082	47.3	37.7 – 56.3
Within populations	68	4.612	0.068	σ_c^2	0.068	0.057 – 0.079	52.7	43.7 – 62.3
Total	69	6.802	0.099	σ_T^2	0.129	0.105 – 0.152	100.0	
C1 - XC1								
Between populations	1	2.458	2.458	σ_b^2	0.063	0.043 – 0.085	47.2	37.5 – 55.9
Within populations	74	5.205	0.070	σ_c^2	0.070	0.059 – 0.081	52.8	44.1 – 62.5
Total	75	7.663	0.102	σ_T^2	0.133	0.110 – 0.157	100.0	
C2 - G1								
Between populations	1	0.996	0.996	σ_b^2	0.055	0.038 – 0.075	46.0	36.6 – 54.7
Within populations	45	2.918	0.065	σ_c^2	0.065	0.052 – 0.078	54.0	45.3 – 63.4
Total	46	3.914	0.085	σ_T^2	0.120	0.097 – 0.144	100.0	
C2 - G2								
Between populations	1	2.372	2.372	σ_b^2	0.065	0.045 – 0.087	47.1	38.0 – 55.4
Within populations	69	5.021	0.073	σ_c^2	0.073	0.061 – 0.085	52.9	44.6 – 62.0
Total	70	7.392	0.106	σ_T^2	0.138	0.113 – 0.163	100.0	
C2 - G3								
Between populations	1	2.097	2.097	σ_b^2	0.062	0.041 – 0.084	47.9	37.6 – 56.6
Within populations	64	4.323	0.068	σ_c^2	0.068	0.056 – 0.080	52.1	43.4 – 62.4
Total	65	6.420	0.099	σ_T^2	0.130	0.105 – 0.155	100.0	
C2 - GH2a								
Between populations	1	1.889	1.889	σ_b^2	0.053	0.037 – 0.071	40.1	31.9 – 47.8
Within populations	67	5.262	0.079	σ_c^2	0.079	0.067 – 0.090	59.9	52.2 – 68.1
Total	68	7.151	0.105	σ_T^2	0.131	0.110 – 0.153	100.0	
C2 - GH2b								
Between populations	1	1.825	1.825	σ_b^2	0.052	0.036 – 0.069	40.3	32.3 – 47.8
Within populations	66	5.045	0.076	σ_c^2	0.076	0.065 – 0.088	59.7	52.2 – 67.7
Total	67	6.870	0.103	σ_T^2	0.128	0.107 – 0.150	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD		Variance Observed	Variance 95% conf[a]	% total Observed	% total 95% conf[a]
C2 - P1								
Between populations	1	1.048	1.048	σ_b^2	0.041	0.025 – 0.059	38.5	27.7 – 48.3
Within populations	52	3.400	0.065	σ_c^2	0.065	0.053 – 0.077	61.5	51.7 – 72.3
Total	53	4.448	0.084	σ_T^2	0.106	0.085 – 0.128	100.0	
C2 - P2								
Between populations	1	1.617	1.617	σ_b^2	0.055	0.036 – 0.076	46.9	36.1 – 56.0
Within populations	57	3.570	0.063	σ_c^2	0.063	0.052 – 0.075	53.1	44.0 – 63.9
Total	58	5.188	0.089	σ_T^2	0.118	0.094 – 0.142	100.0	
C2 - X1								
Between populations	1	2.667	2.667	σ_b^2	0.061	0.040 – 0.082	47.0	37.0 – 55.8
Within populations	87	5.936	0.068	σ_c^2	0.068	0.057 – 0.080	53.0	44.2 – 63.0
Total	88	8.603	0.098	σ_T^2	0.129	0.105 – 0.153	100.0	
C2 - X3								
Between populations	1	2.117	2.117	σ_b^2	0.060	0.040 – 0.081	47.3	37.0 – 56.3
Within populations	67	4.447	0.066	σ_c^2	0.066	0.055 – 0.078	52.7	43.7 – 63.0
Total	68	6.564	0.097	σ_T^2	0.126	0.103 – 0.149	100.0	
C2 - XC1								
Between populations	1	2.161	2.161	σ_b^2	0.056	0.036 – 0.077	44.7	34.0 – 54.2
Within populations	73	5.040	0.069	σ_c^2	0.069	0.058 – 0.081	55.3	45.8 – 66.0
Total	74	7.201	0.097	σ_T^2	0.125	0.102 – 0.149	100.0	
G1 - G2								
Between populations	1	0.262	0.262	σ_b^2	0.010	0.005 – 0.015	9.4	5.2 – 13.9
Within populations	44	4.202	0.095	σ_c^2	0.095	0.080 – 0.111	90.6	86.1 – 94.8
Total	45	4.464	0.099	σ_T^2	0.105	0.088 – 0.123	100.0	
G1 - G3								
Between populations	1	0.332	0.332	σ_b^2	0.015	0.008 – 0.024	15.4	8.7 – 22.5
Within populations	39	3.294	0.084	σ_c^2	0.084	0.070 – 0.099	84.6	77.5 – 91.3
Total	40	3.626	0.091	σ_T^2	0.100	0.083 – 0.118	100.0	
G1 - GH2a								
Between populations	1	0.461	0.461	σ_b^2	0.022	0.013 – 0.031	17.8	11.9 – 23.9
Within populations	42	4.233	0.101	σ_c^2	0.101	0.086 – 0.116	82.2	76.1 – 88.1
Total	43	4.693	0.109	σ_T^2	0.123	0.104 – 0.142	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD		Variance Observed	Variance 95% conf[a]	% total Observed	% total 95% conf[a]
G1 - GH2b								
Between populations	1	0.529	0.529	σ_b^2	0.026	0.017 – 0.037	21.2	14.8 – 27.7
Within populations	41	4.016	0.098	σ_c^2	0.098	0.084 – 0.113	78.8	72.3 – 85.2
Total	42	4.545	0.108	σ_T^2	0.124	0.106 – 0.144	100.0	
G1 - P1								
Between populations	1	0.584	0.584	σ_b^2	0.036	0.023 – 0.051	29.3	20.8 – 37.1
Within populations	27	2.371	0.088	σ_c^2	0.088	0.074 – 0.102	70.7	62.9 – 79.2
Total	28	2.955	0.106	σ_T^2	0.124	0.103 – 0.145	100.0	
G1 - P2								
Between populations	1	0.855	0.855	σ_b^2	0.052	0.033 – 0.073	39.6	29.1 – 48.9
Within populations	32	2.542	0.079	σ_c^2	0.079	0.067 – 0.092	60.4	51.1 – 70.9
Total	33	3.397	0.103	σ_T^2	0.132	0.109 – 0.155	100.0	
G1 - X1								
Between populations	1	1.153	1.153	σ_b^2	0.059	0.041 – 0.079	42.7	33.2 – 51.3
Within populations	62	4.907	0.079	σ_c^2	0.079	0.066 – 0.093	57.3	48.7 – 66.8
Total	63	6.061	0.096	σ_T^2	0.138	0.116 – 0.161	100.0	
G1 - X3								
Between populations	1	0.976	0.976	σ_b^2	0.054	0.036 – 0.073	40.0	30.4 – 49.0
Within populations	42	3.418	0.081	σ_c^2	0.081	0.068 – 0.095	60.0	51.0 – 69.6
Total	43	4.394	0.102	σ_T^2	0.136	0.113 – 0.157	100.0	
G1 - XC1								
Between populations	1	1.089	1.089	σ_b^2	0.059	0.041 – 0.078	41.2	32.1 – 49.6
Within populations	48	4.011	0.084	σ_c^2	0.084	0.071 – 0.097	58.8	50.4 – 67.9
Total	49	5.100	0.104	σ_T^2	0.142	0.121 – 0.165	100.0	
G2 - G3								
Between populations	1	0.605	0.605	σ_b^2	0.016	0.008 – 0.027	15.8	8.3 – 24.2
Within populations	63	5.397	0.086	σ_c^2	0.086	0.071 – 0.100	84.2	75.8 – 91.7
Total	64	6.001	0.094	σ_T^2	0.102	0.084 – 0.120	100.0	
G2 - GH2a								
Between populations	1	0.310	0.310	σ_b^2	0.006	0.002 – 0.012	6.1	1.9 – 11.1
Within populations	66	6.335	0.096	σ_c^2	0.096	0.081 – 0.112	93.9	88.9 – 98.1
Total	67	6.645	0.099	σ_T^2	0.102	0.085 – 0.119	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD		Variance		% total	
					Observed	95% conf[a]	Observed	95% conf[a]
G2 - GH2b								
Between populations	1	0.593	0.593	σ_b^2	0.015	0.006 – 0.026	13.7	6.2 – 22.3
Within populations	65	6.118	0.094	σ_c^2	0.094	0.079 – 0.110	86.3	77.7 – 93.8
Total	66	6.711	0.102	σ_T^2	0.109	0.091 – 0.127	100.0	
G2 - P1								
Between populations	1	0.924	0.924	σ_b^2	0.035	0.021 – 0.051	28.6	19.5 – 37.5
Within populations	51	4.474	0.088	σ_c^2	0.088	0.074 – 0.102	71.4	62.5 – 80.5
Total	52	5.398	0.104	σ_T^2	0.123	0.101 – 0.145	100.0	
G2 - P2								
Between populations	1	1.487	1.487	σ_b^2	0.051	0.033 – 0.071	37.9	28.1 – 46.8
Within populations	56	4.644	0.083	σ_c^2	0.083	0.070 – 0.096	62.1	53.2 – 71.9
Total	57	6.131	0.108	σ_T^2	0.134	0.110 – 0.157	100.0	
G2 - X1								
Between populations	1	2.303	2.303	σ_b^2	0.053	0.035 – 0.072	39.3	30.0 – 48.0
Within populations	86	7.009	0.082	σ_c^2	0.082	0.069 – 0.094	60.7	52.0 – 70.0
Total	87	9.312	0.107	σ_T^2	0.134	0.112 – 0.156	100.0	
G2 - X3								
Between populations	1	1.846	1.846	σ_b^2	0.052	0.036 – 0.070	38.3	29.9 – 46.1
Within populations	66	5.520	0.084	σ_c^2	0.084	0.072 – 0.096	61.7	53.9 – 70.1
Total	67	7.366	0.110	σ_T^2	0.136	0.113 – 0.158	100.0	
G2 - XC1								
Between populations	1	2.009	2.009	σ_b^2	0.052	0.035 – 0.071	38.0	29.1 – 46.4
Within populations	72	6.113	0.085	σ_c^2	0.085	0.073 – 0.097	62.0	53.6 – 70.9
Total	73	8.122	0.111	σ_T^2	0.137	0.115 – 0.159	100.0	
G3 - GH2a								
Between populations	1	0.667	0.667	σ_b^2	0.018	0.010 – 0.029	16.5	9.8 – 24.1
Within populations	61	5.638	0.092	σ_c^2	0.092	0.078 – 0.107	83.5	75.9 – 90.2
Total	62	6.305	0.102	σ_T^2	0.111	0.093 – 0.129	100.0	
G3 - GH2b								
Between populations	1	0.769	0.769	σ_b^2	0.022	0.012 – 0.035	19.5	11.5 – 27.8
Within populations	60	5.421	0.090	σ_c^2	0.090	0.077 – 0.105	80.5	72.2 – 88.5
Total	61	6.190	0.101	σ_T^2	0.112	0.094 – 0.132	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD		Variance Observed	Variance 95% conf[a]	% total Observed	% total 95% conf[a]
G3 - P1								
Between populations	1	0.937	0.937	σ_b^2	0.038	0.023 – 0.055	31.6	21.2 – 41.1
Within populations	46	3.776	0.082	σ_c^2	0.082	0.068 – 0.096	68.4	58.9 – 78.8
Total	47	4.713	0.100	σ_T^2	0.120	0.098 – 0.142	100.0	
G3 - P2								
Between populations	1	1.354	1.354	σ_b^2	0.049	0.030 – 0.069	38.8	28.1 – 48.2
Within populations	51	3.946	0.077	σ_c^2	0.077	0.065 – 0.090	61.2	51.8 – 71.9
Total	52	5.300	0.102	σ_T^2	0.126	0.103 – 0.150	100.0	
G3 - X1								
Between populations	1	2.473	2.473	σ_b^2	0.063	0.043 – 0.084	44.5	34.8 – 53.2
Within populations	81	6.312	0.078	σ_c^2	0.078	0.066 – 0.090	55.5	46.8 – 65.2
Total	82	8.785	0.107	σ_T^2	0.140	0.117 – 0.164	100.0	
G3 - X3								
Between populations	1	2.037	2.037	σ_b^2	0.062	0.042 – 0.083	44.1	34.4 – 52.5
Within populations	61	4.823	0.079	σ_c^2	0.079	0.067 – 0.091	55.9	47.5 – 65.6
Total	62	6.860	0.111	σ_T^2	0.141	0.118 – 0.164	100.0	
G3 - XC1								
Between populations	1	2.204	2.204	σ_b^2	0.063	0.043 – 0.084	43.6	34.1 – 52.1
Within populations	67	5.416	0.081	σ_c^2	0.081	0.070 – 0.092	56.4	47.9 – 65.9
Total	68	7.620	0.112	σ_T^2	0.143	0.121 – 0.167	100.0	
GH2a - GH2b								
Between populations	1	0.216	0.216	σ_b^2	0.004	0.002 – 0.006	3.4	1.6 – 5.2
Within populations	63	6.359	0.101	σ_c^2	0.101	0.086 – 0.117	96.6	94.8 – 98.4
Total	64	6.575	0.103	σ_T^2	0.104	0.089 – 0.121	100.0	
GH2a - P1								
Between populations	1	0.617	0.617	σ_b^2	0.022	0.014 – 0.032	18.8	12.8 – 24.9
Within populations	49	4.715	0.096	σ_c^2	0.096	0.082 – 0.111	81.2	75.1 – 87.2
Total	50	5.331	0.107	σ_T^2	0.119	0.099 – 0.138	100.0	
GH2a - P2								
Between populations	1	1.122	1.122	σ_b^2	0.038	0.025 – 0.053	29.6	21.7 – 37.3
Within populations	54	4.885	0.090	σ_c^2	0.090	0.077 – 0.103	70.4	62.7 – 78.3
Total	55	6.007	0.109	σ_T^2	0.129	0.108 – 0.150	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD		Variance		% total	
					Observed	95% conf[a]	Observed	95% conf[a]
GH2a - X1								
Between populations	1	1.716	1.716	σ_b^2	0.040	0.027 – 0.055	31.7	23.5 – 39.9
Within populations	84	7.250	0.086	σ_c^2	0.086	0.074 – 0.099	68.3	60.1 – 76.5
Total	85	8.967	0.105	σ_T^2	0.126	0.108 – 0.146	100.0	
GH2a - X3								
Between populations	1	1.367	1.367	σ_b^2	0.039	0.027 – 0.052	30.1	23.0 – 36.9
Within populations	64	5.761	0.090	σ_c^2	0.090	0.078 – 0.102	69.9	63.1 – 77.0
Total	65	7.129	0.110	σ_T^2	0.129	0.110 – 0.147	100.0	
GH2a - XC1								
Between populations	1	1.502	1.502	σ_b^2	0.039	0.026 – 0.054	30.3	22.6 – 37.7
Within populations	70	6.354	0.091	σ_c^2	0.091	0.079 – 0.103	69.7	62.3 – 77.4
Total	71	7.857	0.111	σ_T^2	0.130	0.111 – 0.150	100.0	
GH2b - P1								
Between populations	1	0.624	0.624	σ_b^2	0.023	0.014 – 0.034	19.7	13.1 – 26.9
Within populations	48	4.498	0.094	σ_c^2	0.094	0.080 – 0.109	80.3	73.1 – 86.9
Total	49	5.122	0.105	σ_T^2	0.117	0.097 – 0.136	100.0	
GH2b - P2								
Between populations	1	0.995	0.995	σ_b^2	0.034	0.021 – 0.049	27.8	19.4 – 35.9
Within populations	53	4.668	0.088	σ_c^2	0.088	0.075 – 0.101	72.2	64.1 – 80.6
Total	54	5.663	0.105	σ_T^2	0.122	0.102 – 0.143	100.0	
GH2b - X1								
Between populations	1	1.853	1.853	σ_b^2	0.044	0.029 – 0.061	34.3	25.6 – 42.6
Within populations	83	7.033	0.085	σ_c^2	0.085	0.073 – 0.097	65.7	57.4 – 74.4
Total	84	8.886	0.106	σ_T^2	0.129	0.109 – 0.150	100.0	
GH2b - X3								
Between populations	1	1.398	1.398	σ_b^2	0.040	0.026 – 0.056	31.4	22.9 – 39.6
Within populations	63	5.545	0.088	σ_c^2	0.088	0.075 – 0.101	68.6	60.4 – 77.1
Total	64	6.943	0.108	σ_T^2	0.128	0.109 – 0.148	100.0	
GH2b - XC1								
Between populations	1	1.578	1.578	σ_b^2	0.042	0.028 – 0.058	32.3	23.9 – 40.3
Within populations	69	6.138	0.089	σ_c^2	0.089	0.077 – 0.101	67.7	59.7 – 76.1
Total	70	7.715	0.110	σ_T^2	0.131	0.112 – 0.151	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD		Variance Observed	Variance 95% conf[a]	% total Observed	% total 95% conf[a]
P1 - P2								
Between populations	1	0.273	0.273	σ_b^2	0.010	0.005 – 0.016	11.1	5.7 – 16.6
Within populations	39	3.023	0.078	σ_c^2	0.078	0.063 – 0.092	88.9	83.4 – 94.3
Total	40	3.297	0.082	σ_T^2	0.087	0.070 – 0.104	100.0	
P1 - X1								
Between populations	1	0.742	0.742	σ_b^2	0.025	0.015 – 0.037	24.0	15.9 – 32.4
Within populations	69	5.389	0.078	σ_c^2	0.078	0.065 – 0.092	76.0	67.6 – 84.1
Total	70	6.130	0.088	σ_T^2	0.103	0.084 – 0.122	100.0	
P1 - X3								
Between populations	1	0.649	0.649	σ_b^2	0.024	0.014 – 0.037	23.5	15.0 – 32.5
Within populations	49	3.900	0.080	σ_c^2	0.080	0.066 – 0.093	76.5	67.5 – 85.0
Total	50	4.549	0.091	σ_T^2	0.104	0.086 – 0.122	100.0	
P1 - XC1								
Between populations	1	0.670	0.670	σ_b^2	0.024	0.015 – 0.035	22.6	15.5 – 30.4
Within populations	55	4.493	0.082	σ_c^2	0.082	0.069 – 0.095	77.4	69.6 – 84.5
Total	56	5.163	0.092	σ_T^2	0.106	0.088 – 0.124	100.0	
P2 - X1								
Between populations	1	1.407	1.407	σ_b^2	0.042	0.026 – 0.058	35.6	26.3 – 44.2
Within populations	74	5.559	0.075	σ_c^2	0.075	0.063 – 0.088	64.4	55.8 – 73.7
Total	75	6.966	0.093	σ_T^2	0.117	0.095 – 0.139	100.0	
P2 - X3								
Between populations	1	1.244	1.244	σ_b^2	0.043	0.026 – 0.062	36.4	25.8 – 46.4
Within populations	54	4.070	0.075	σ_c^2	0.075	0.063 – 0.089	63.6	53.6 – 74.2
Total	55	5.314	0.097	σ_T^2	0.118	0.097 – 0.141	100.0	
P2 - XC1								
Between populations	1	1.255	1.255	σ_b^2	0.041	0.026 – 0.057	34.4	25.1 – 42.8
Within populations	60	4.663	0.078	σ_c^2	0.078	0.066 – 0.090	65.6	57.2 – 74.9
Total	61	5.919	0.097	σ_T^2	0.118	0.098 – 0.141	100.0	
X1 - X3								
Between populations	1	0.689	0.689	σ_b^2	0.015	0.009 – 0.022	16.4	11.3 – 22.0
Within populations	84	6.436	0.077	σ_c^2	0.077	0.063 – 0.091	83.6	78.0 – 88.7
Total	85	7.125	0.084	σ_T^2	0.092	0.075 – 0.109	100.0	

Continued on next page...

Table E.5 – *Continued*

Variance component	*df*	SSD	MSD	Variance		% total	
				Observed	95% conf [a]	Observed	95% conf [a]
X1 - XC1							
Between populations	1	0.181	0.181	σ_b^2 0.002	0.001 – 0.004	2.8	0.9 – 5.0
Within populations	90	7.029	0.078	σ_c^2 0.078	0.063 – 0.094	97.2	95.0 – 99.1
Total	91	7.210	0.079	σ_T^2 0.080	0.065 – 0.096	100.0	
X3 - XC1							
Between populations	1	0.707	0.707	σ_b^2 0.018	0.011 – 0.025	18.2	12.4 – 24.3
Within populations	70	5.540	0.079	σ_c^2 0.079	0.066 – 0.093	81.8	75.7 – 87.6
Total	71	6.247	0.088	σ_T^2 0.097	0.080 – 0.114	100.0	

For all comparisons, and for all hierarchical levels $p < 0.001$, calculated by permuting individuals randomly over the whole dataset of the population pair (5000 iterations).

[a] Calculated from 5000 bootstrap replicates.

E.3 Null-distributions of variance components

Null-distributions obtained from randomisations of respective hierarchical levels and performing an AMOVA on the randomised dataset; only null-distributions obtained for analysis using the final data are shown. Always 5000 randomisations were performed for each level, and each pairwise comparison. For details of which levels where randomised for each statistic please subsection 2.3.5, page 82. The respective most important Φ-statistic calculated from the dataset is given in brackets, if applicable.

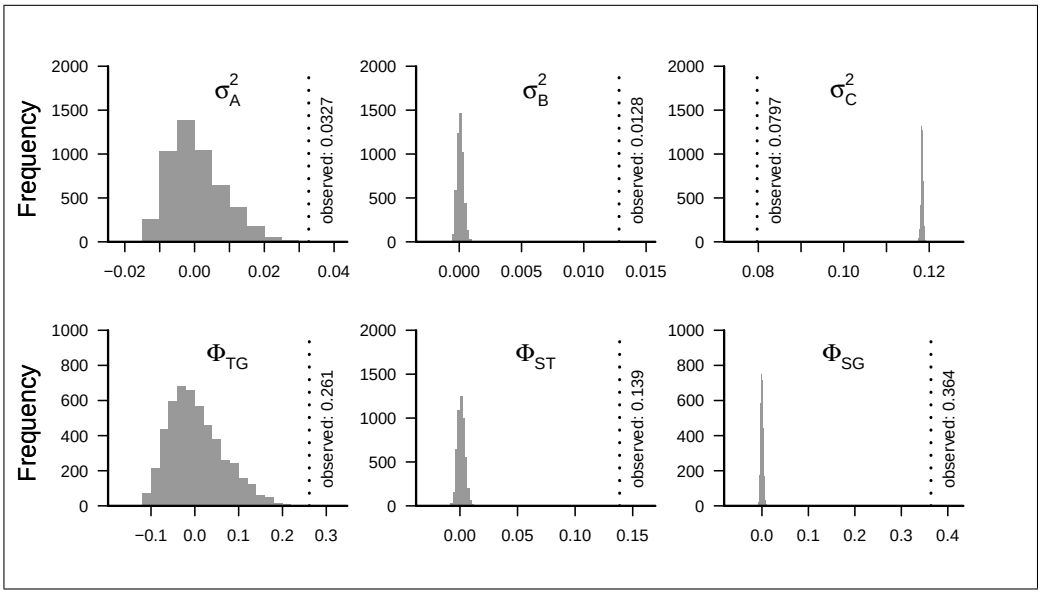

Figure E.1: AMOVA **null-distributions (whole dataset).** Null-distributions for all variance components and Φ-statistics when including the whole dataset (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

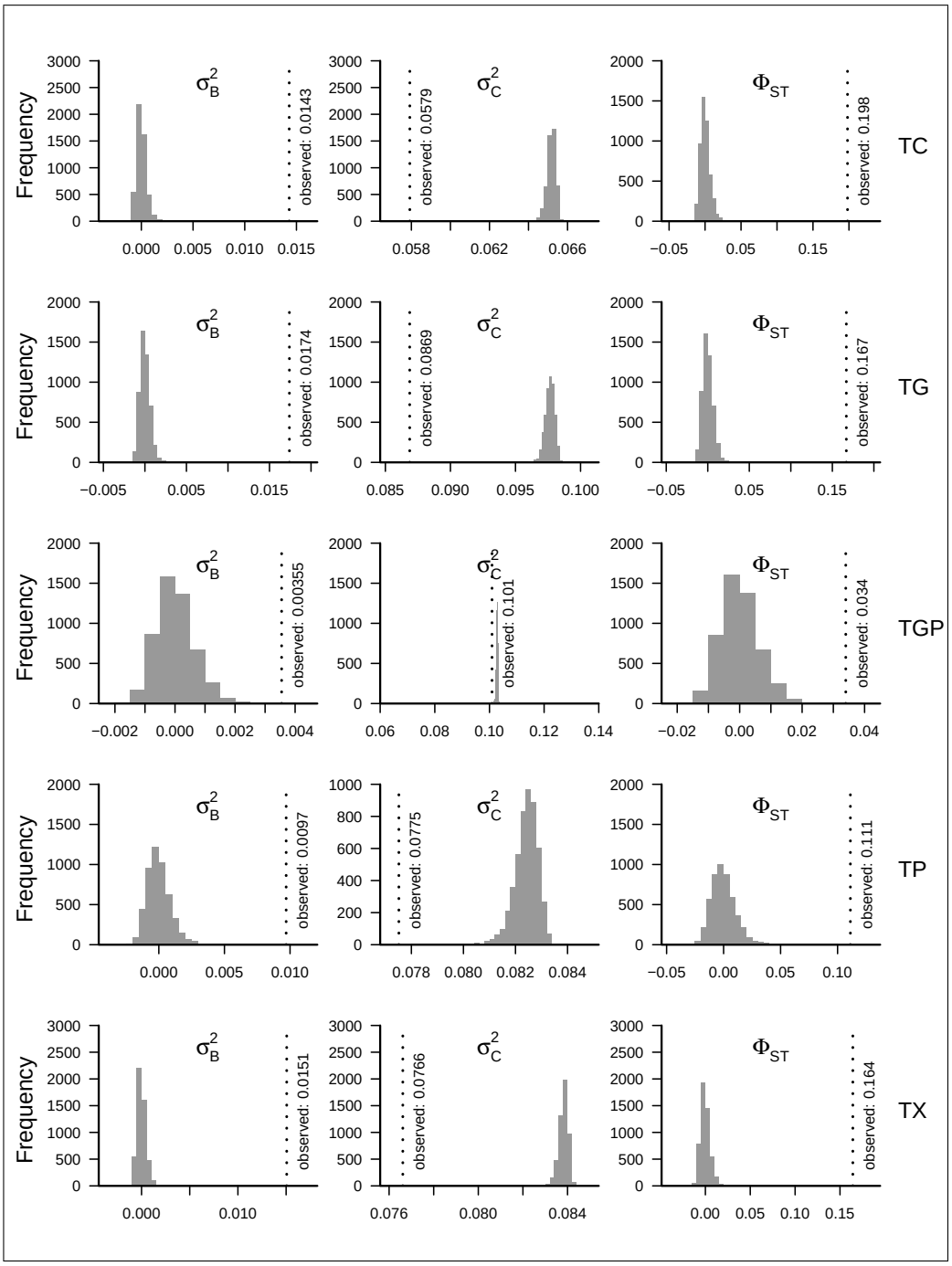

Figure E.2: AMOVA **null-distributions for individual species** (Φ_{ST}); Null-distributions for all variance components and Φ-statistics when analysing the respective subset of each species (variation within populations σ^2_C, variation between populations σ^2_B). The corresponding species for every row is given at the right margin.

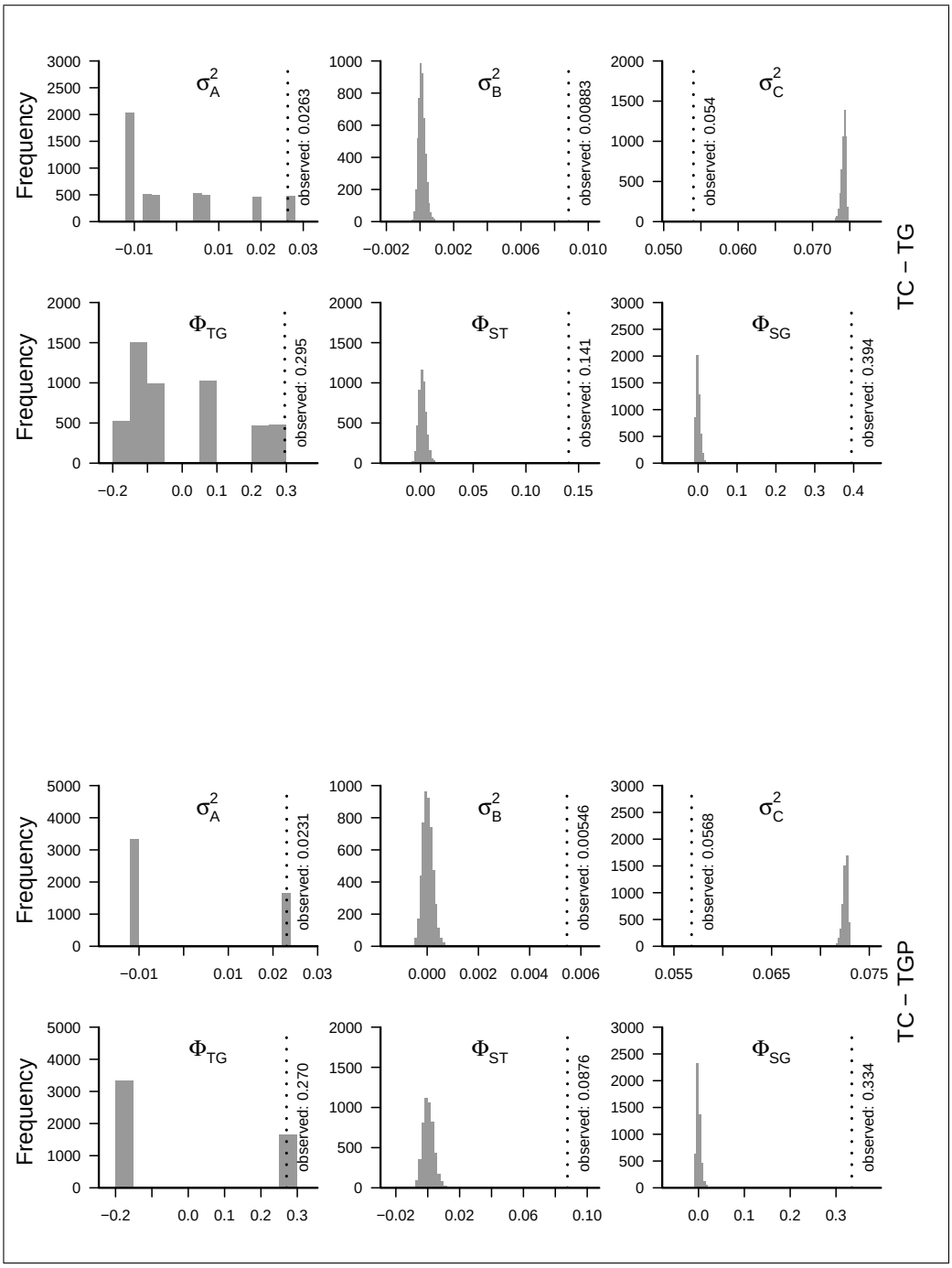

Figure E.3a: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

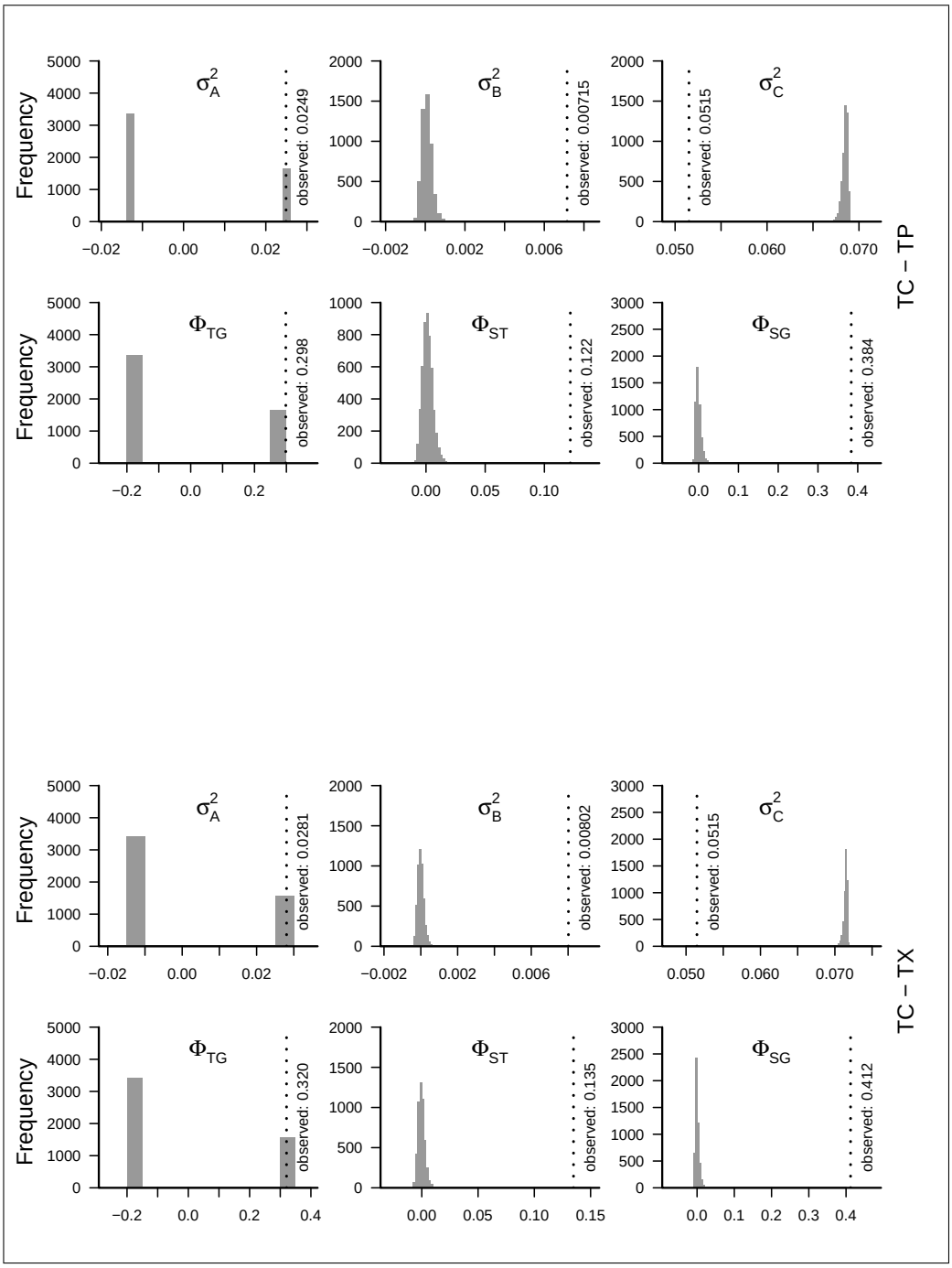

Figure E.3b: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

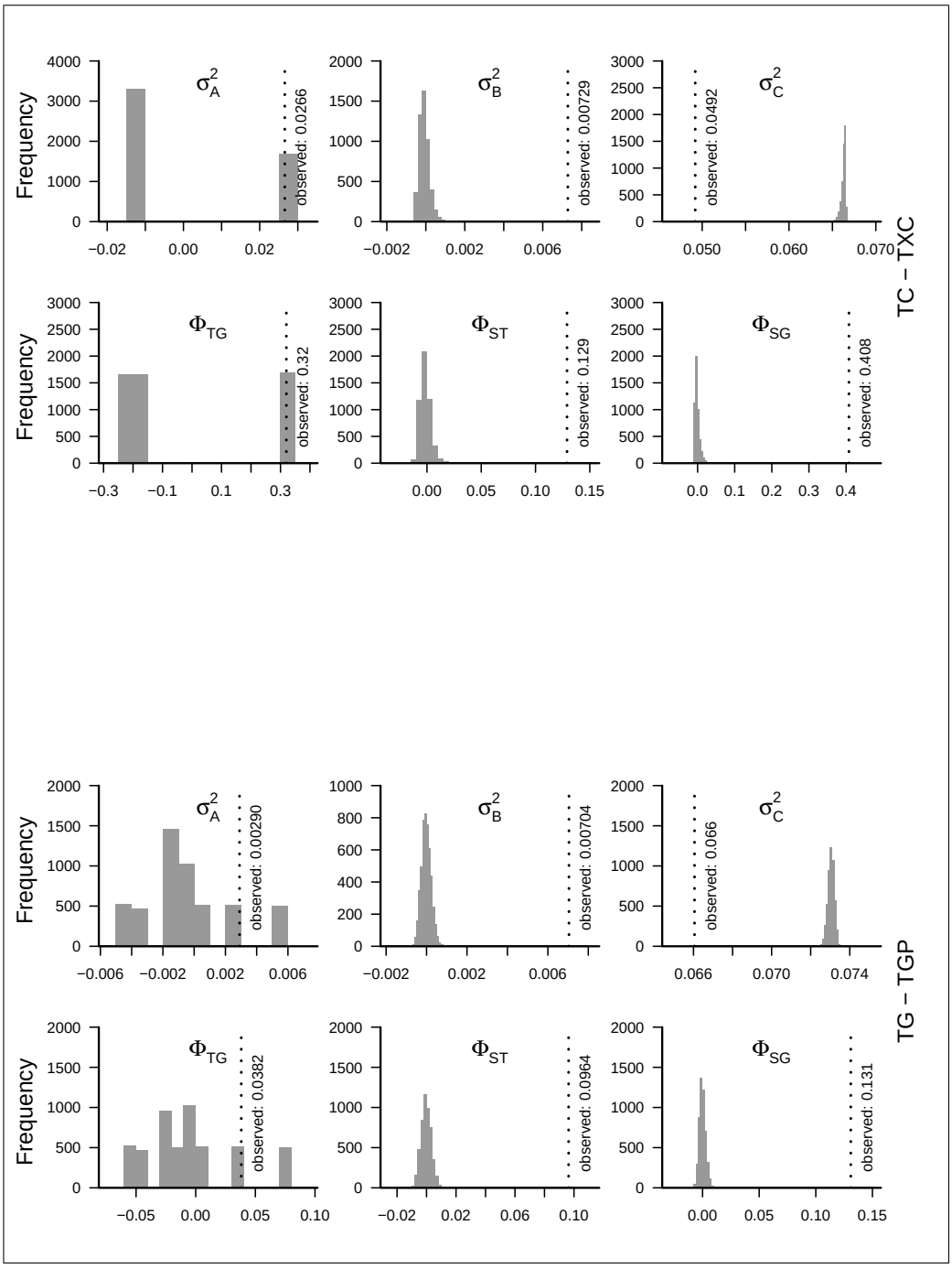

Figure E.3c: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

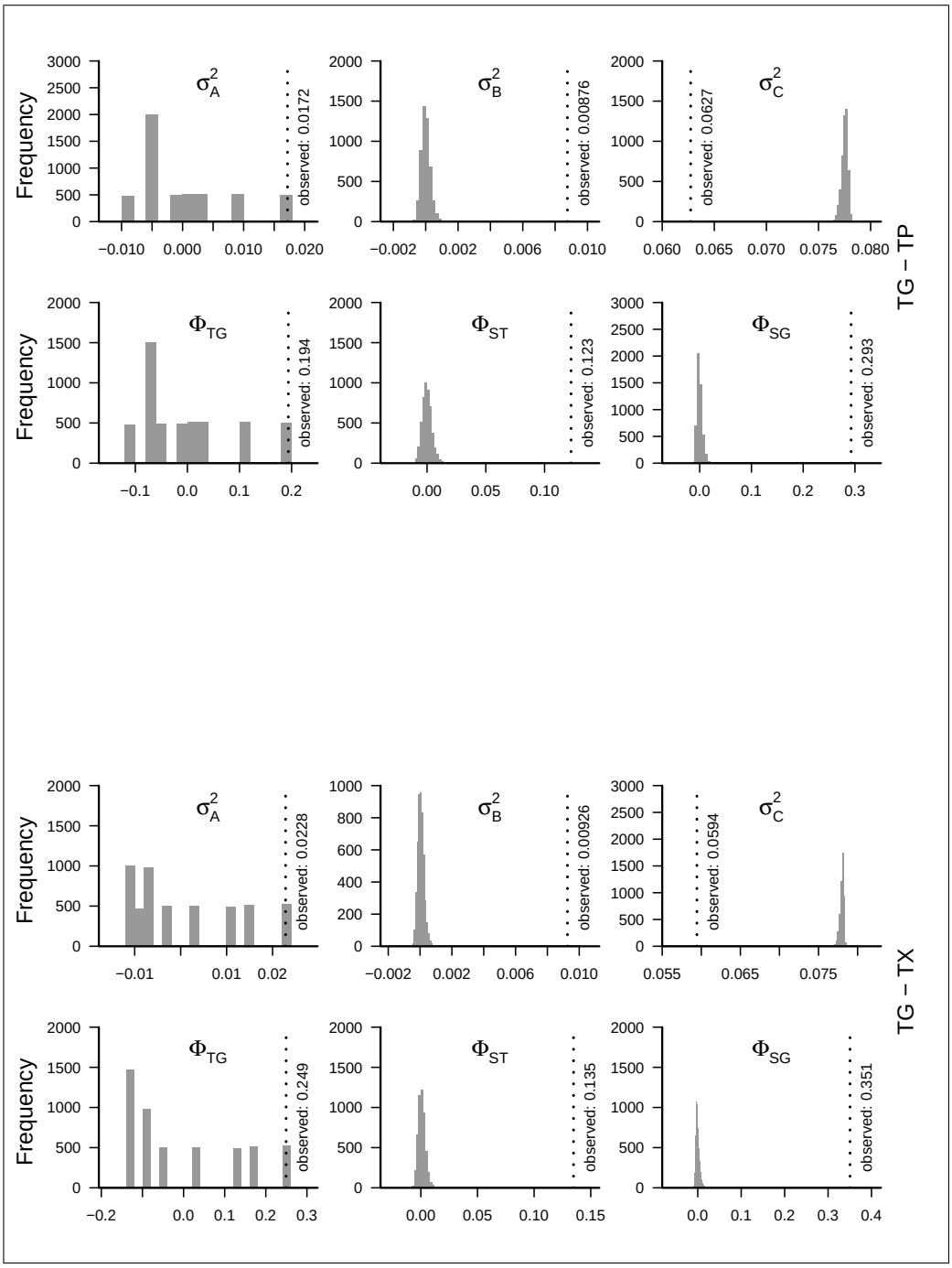

Figure E.3d: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

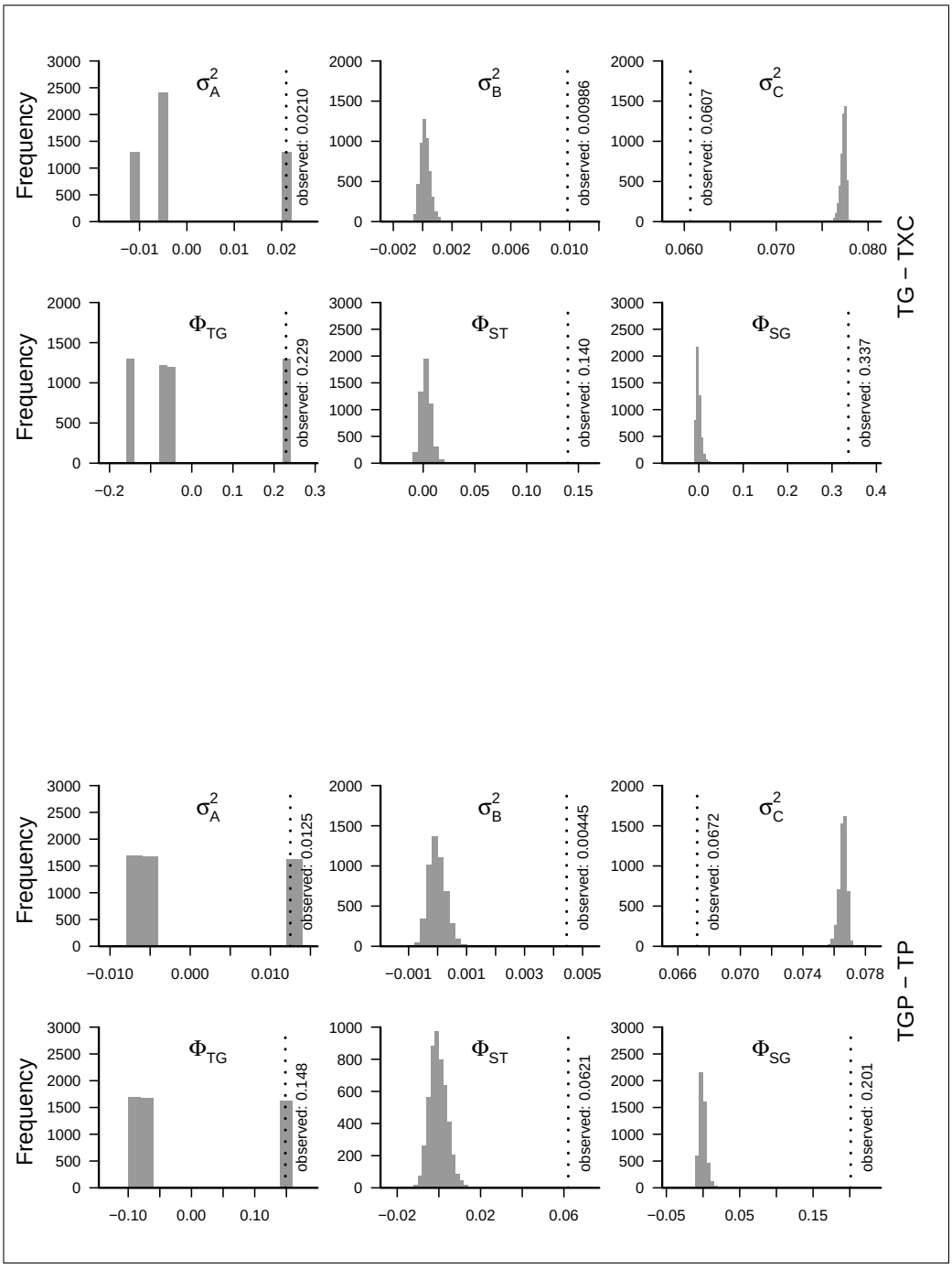

Figure E.3e: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

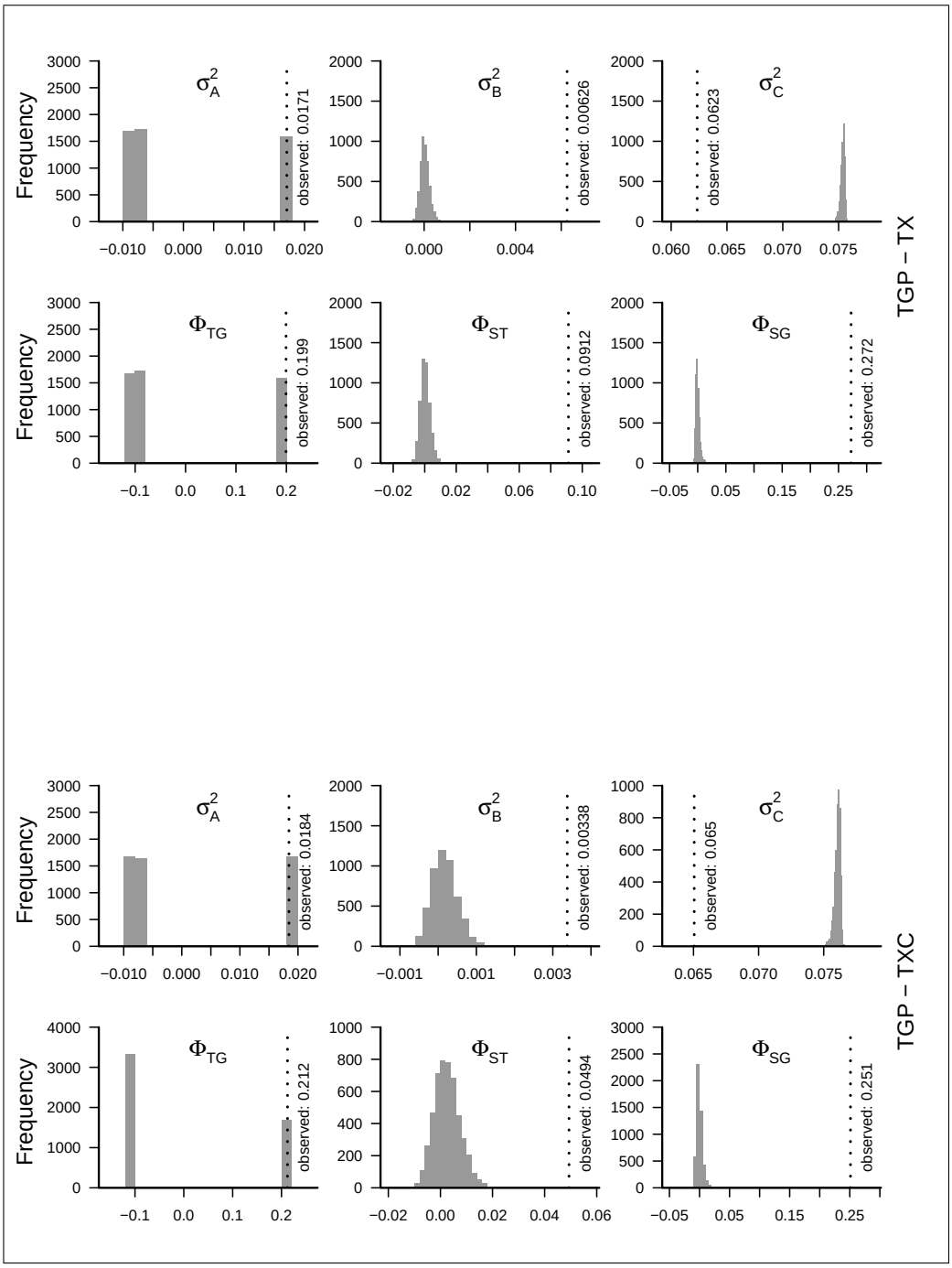

Figure E.3f: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

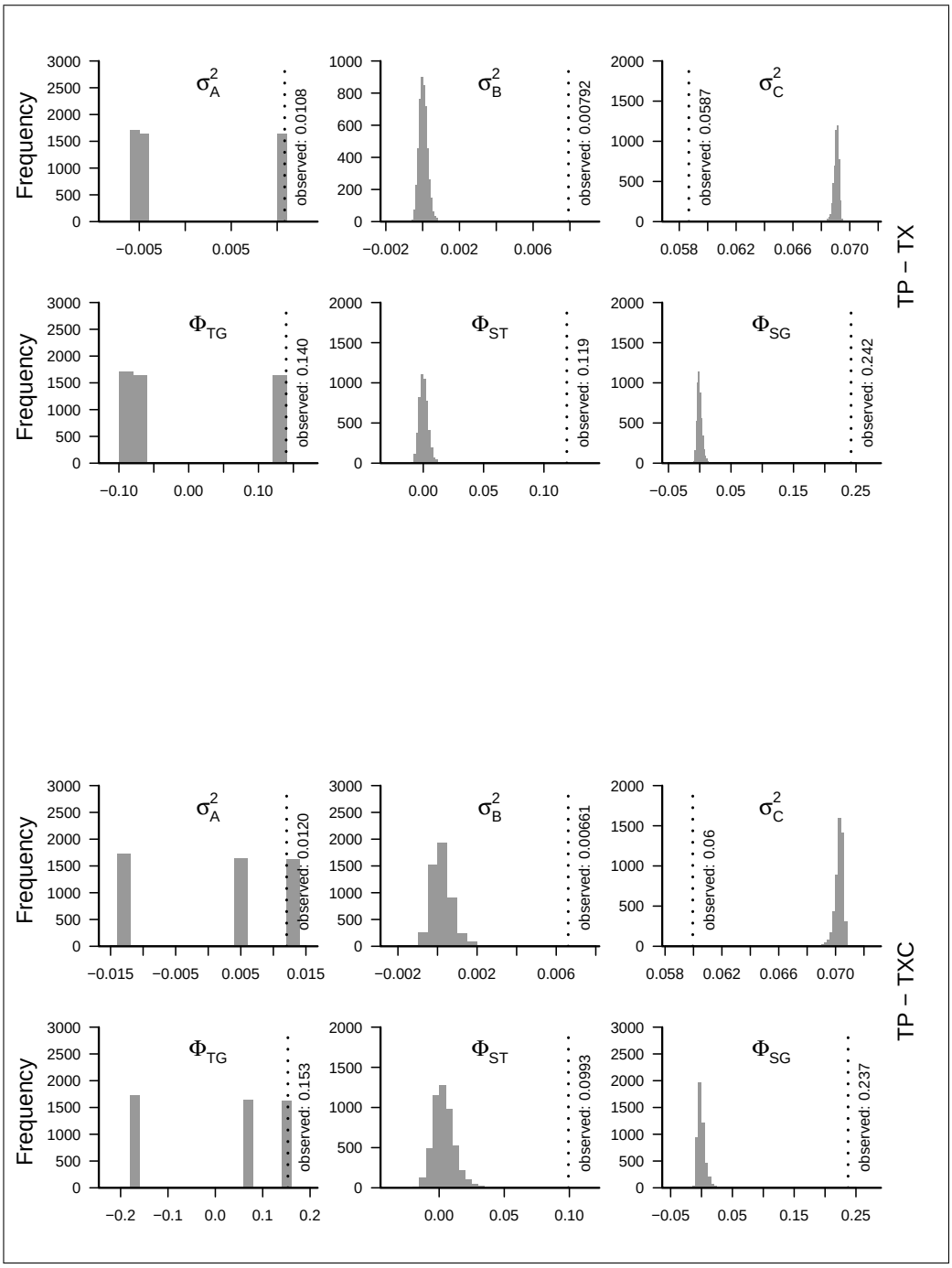

Figure E.3g: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

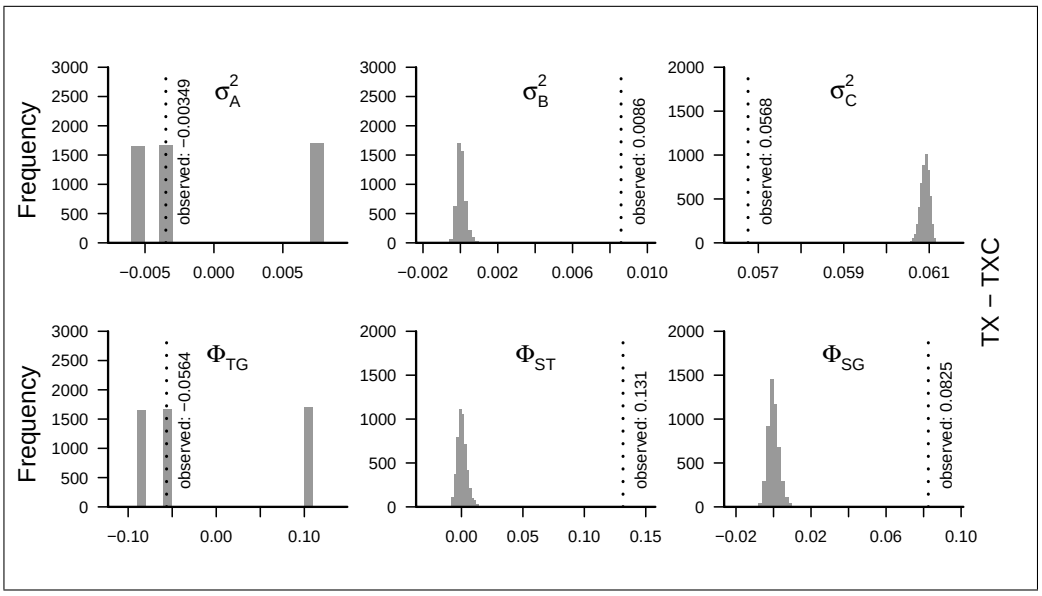

Figure E.3h: AMOVA **null-distributions (pairwise species comparisons,** Φ_{TG}**)** the corresponding species pair for always two rows is given at the right margin (variation within populations σ^2_C, variation between popualtions within species σ^2_B, and variation between species σ^2_A).

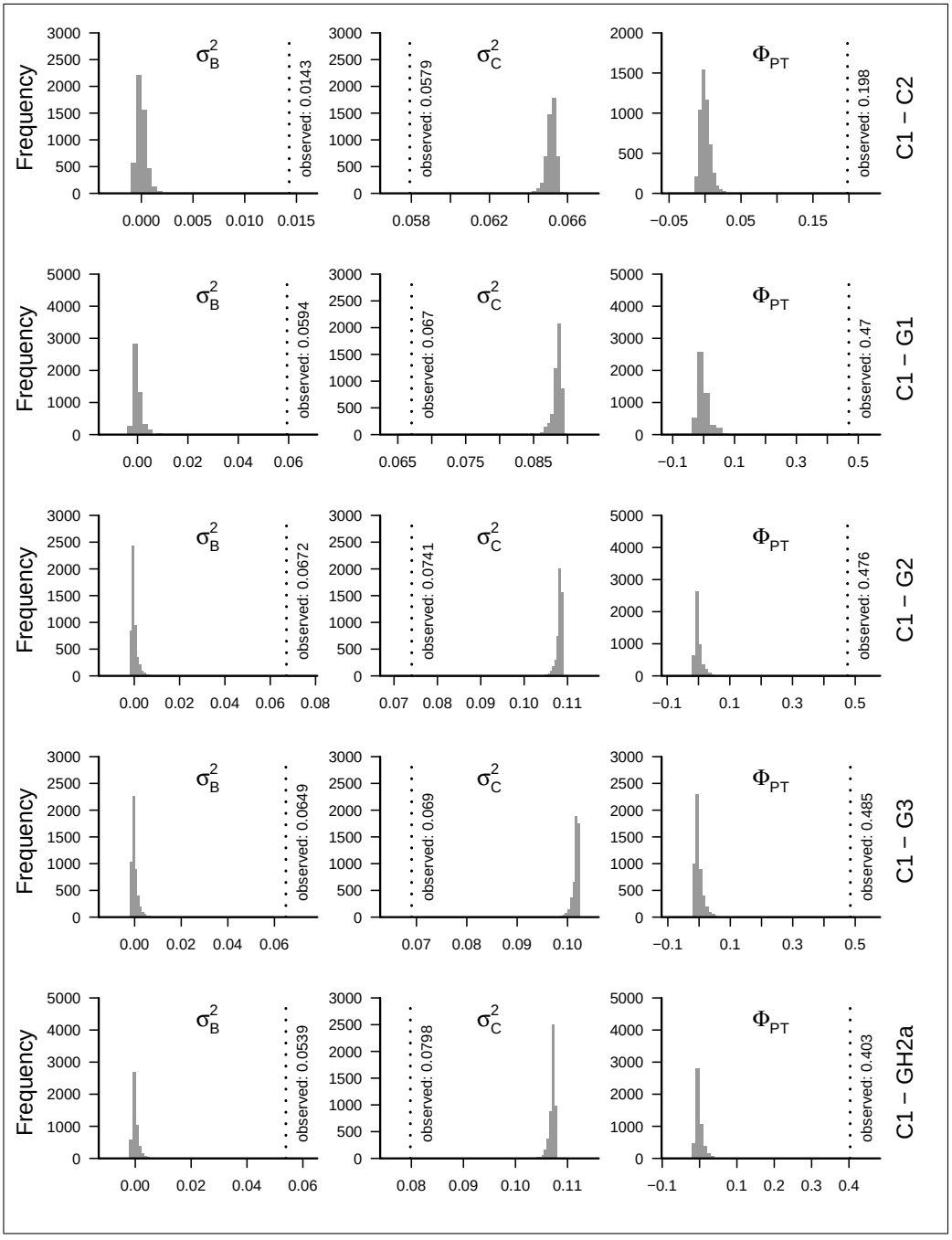

Figure E.4a: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

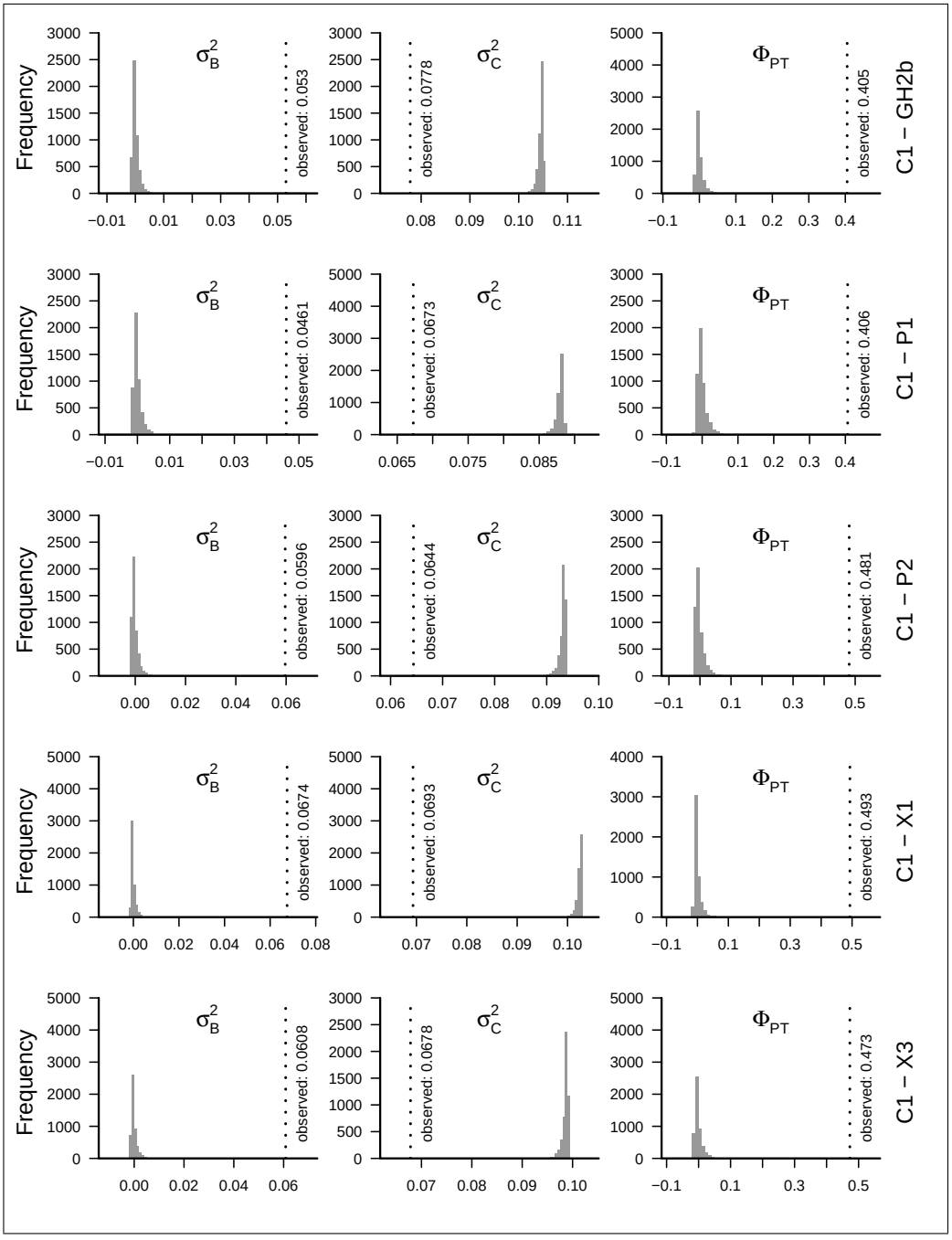

Figure E.4b: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

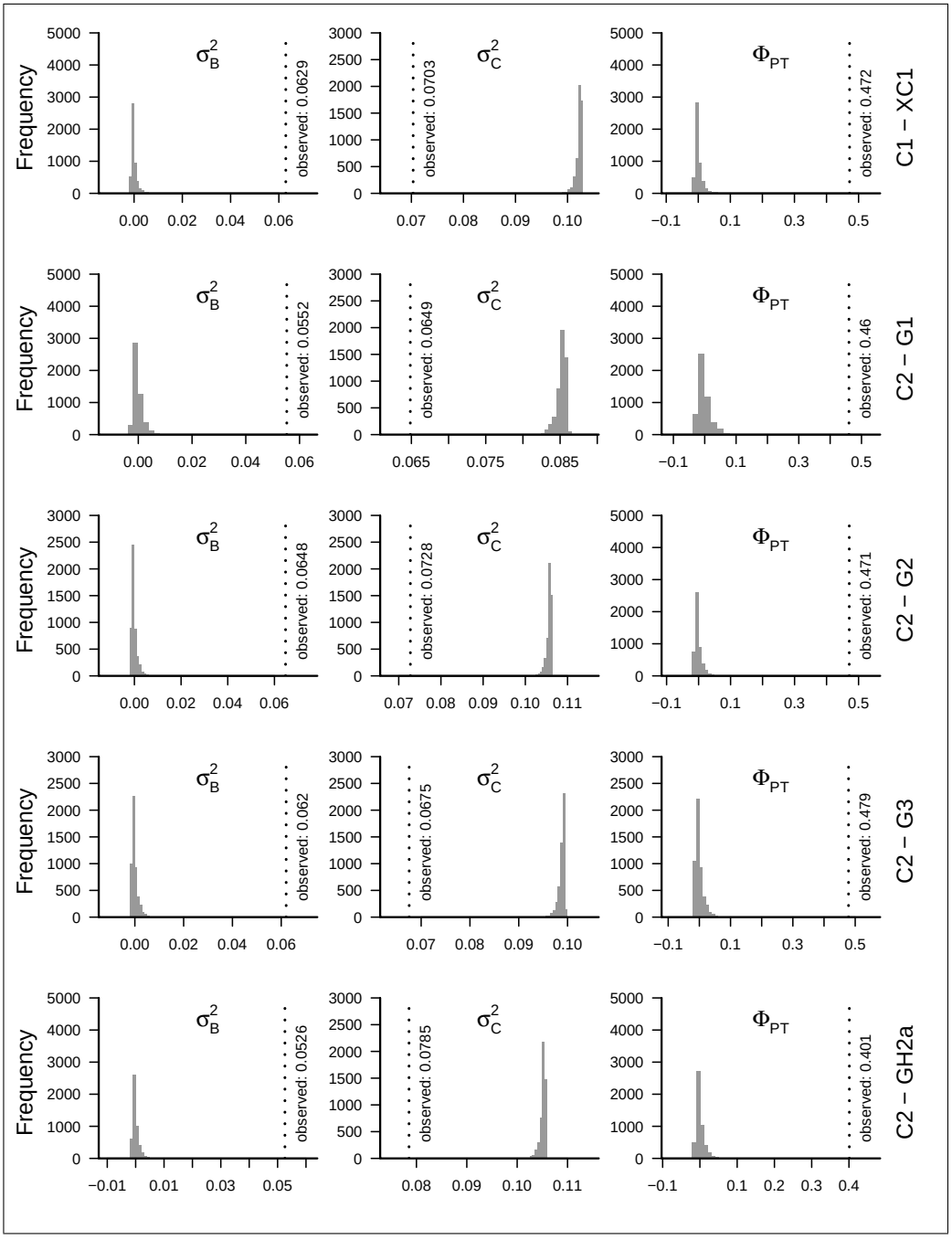

Figure E.4c: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

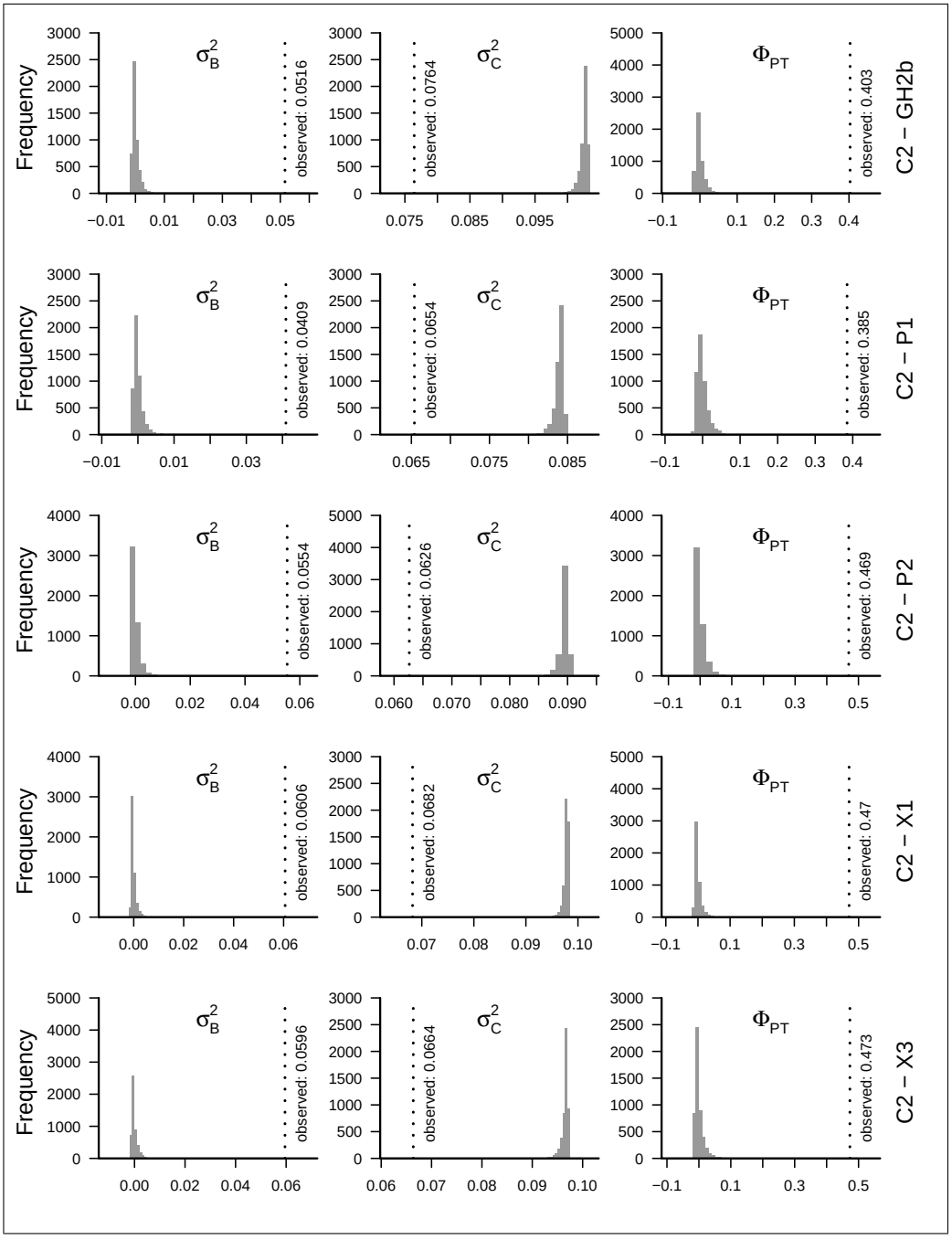

Figure E.4d: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

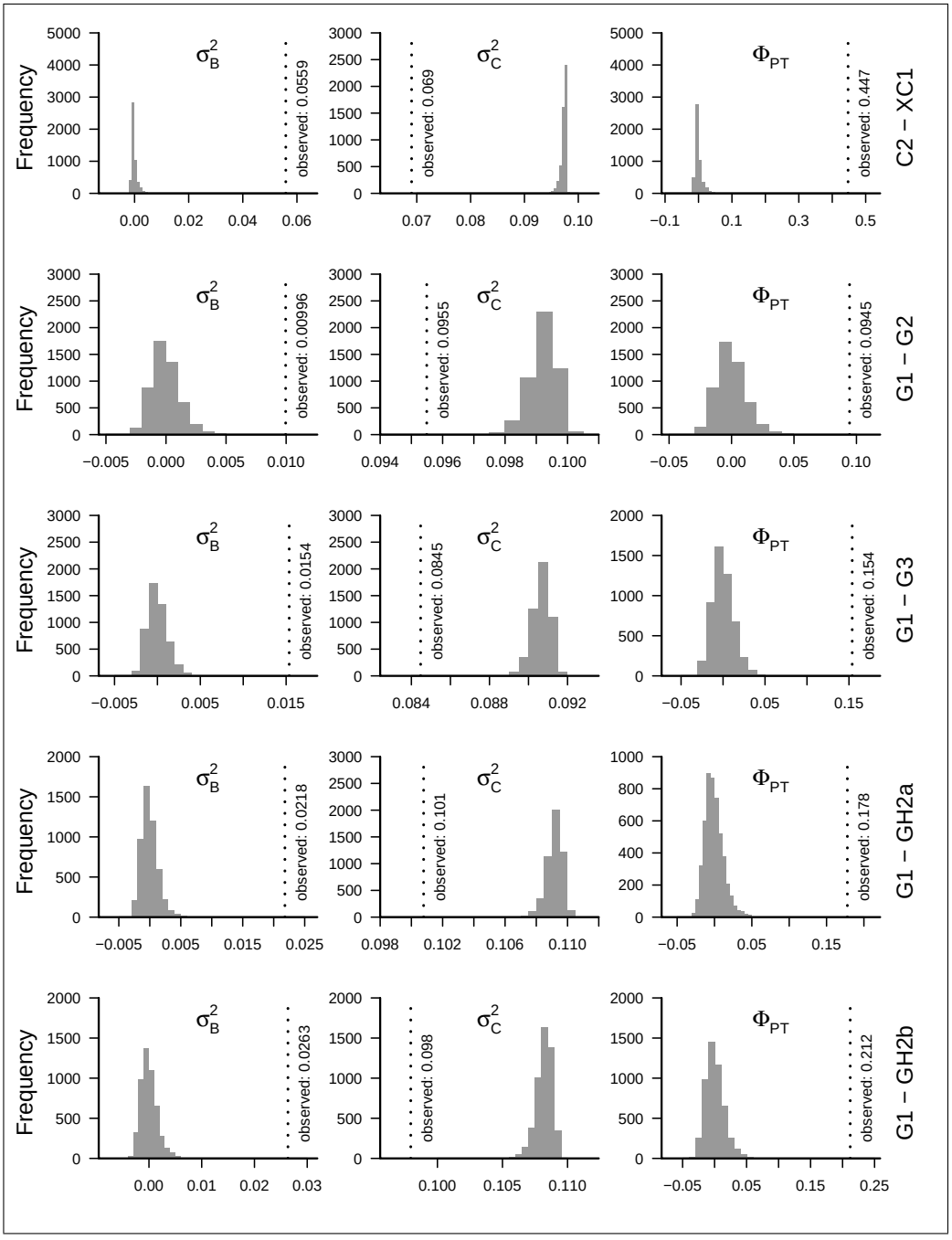

Figure E.4e: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

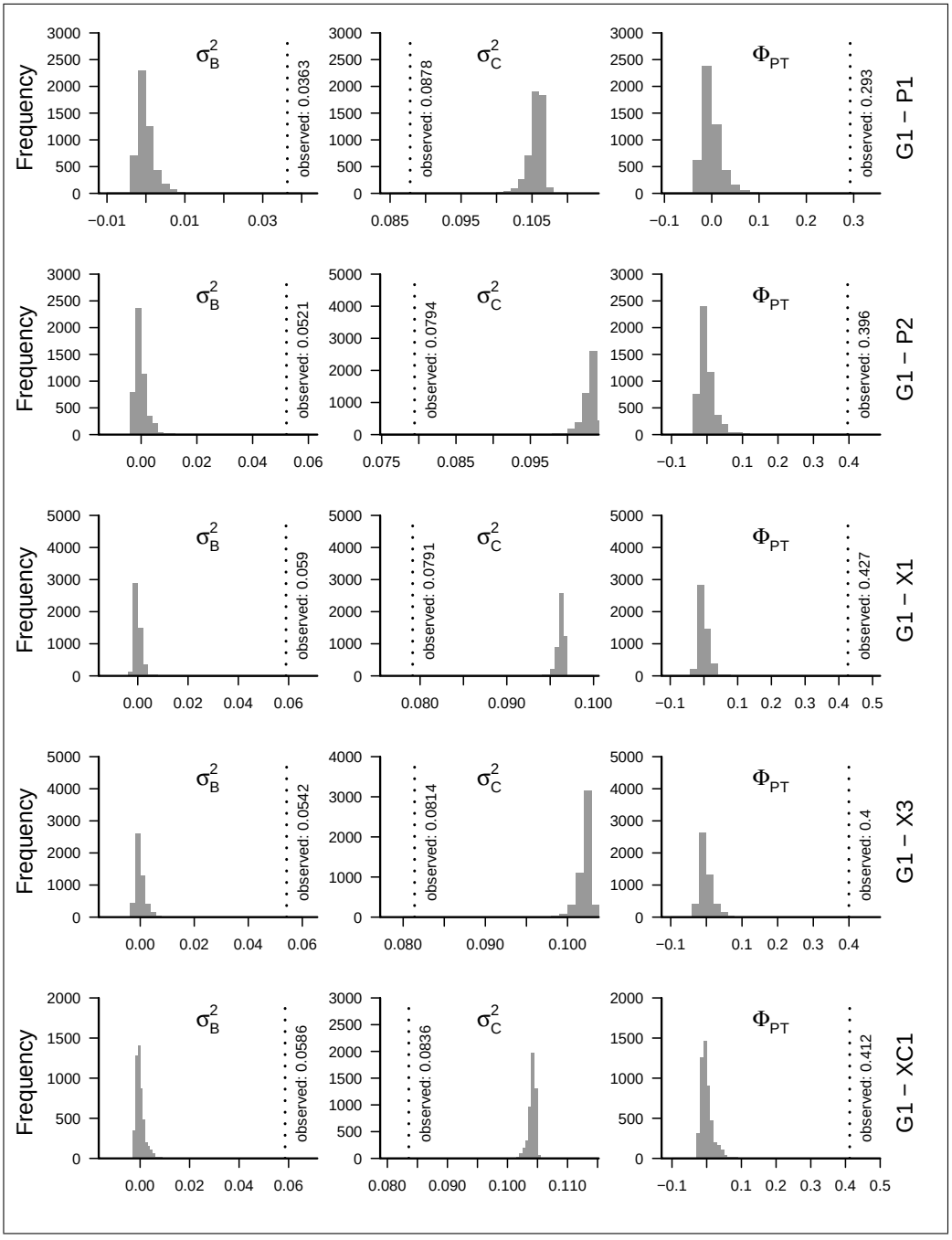

Figure E.4f: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

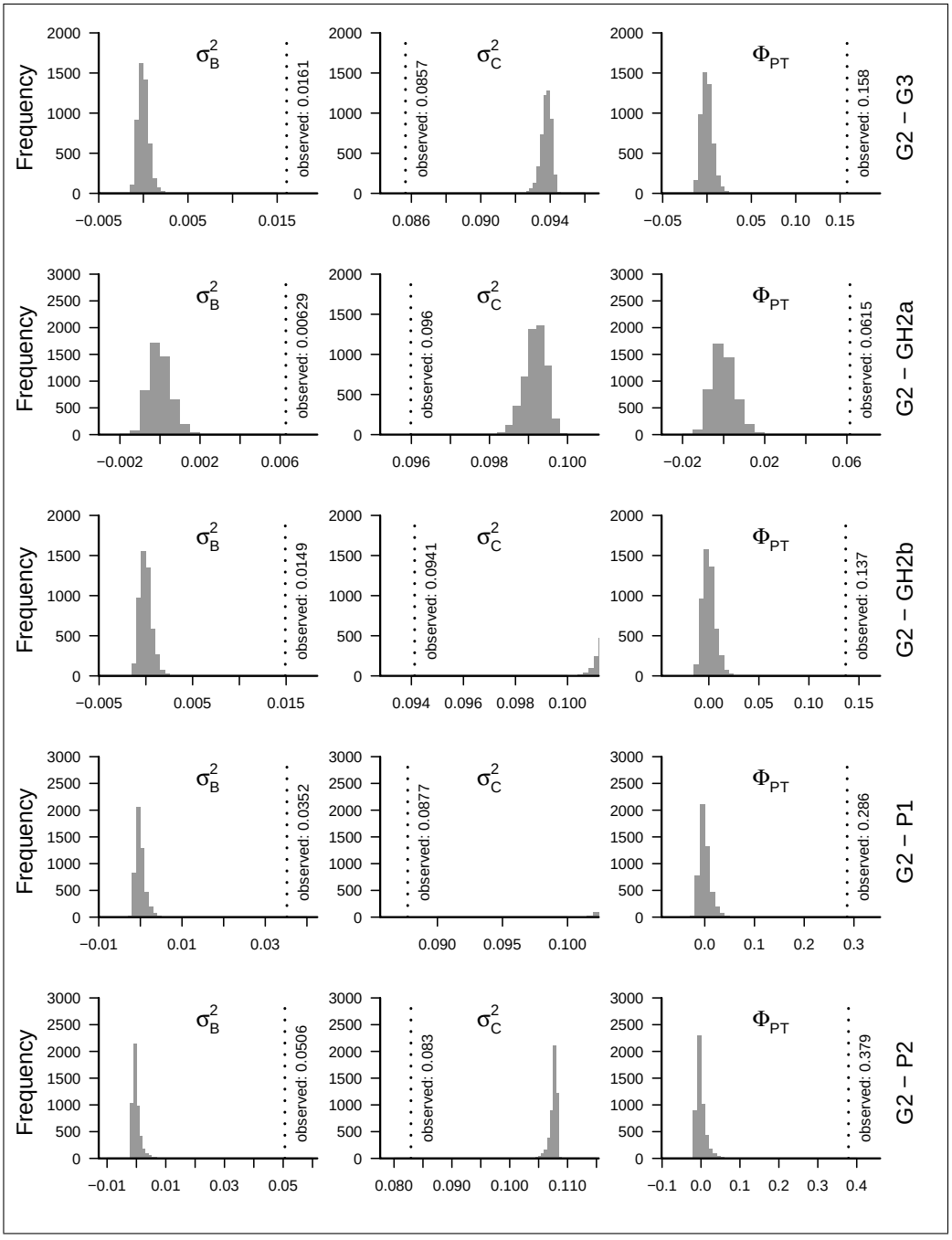

Figure E.4g: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

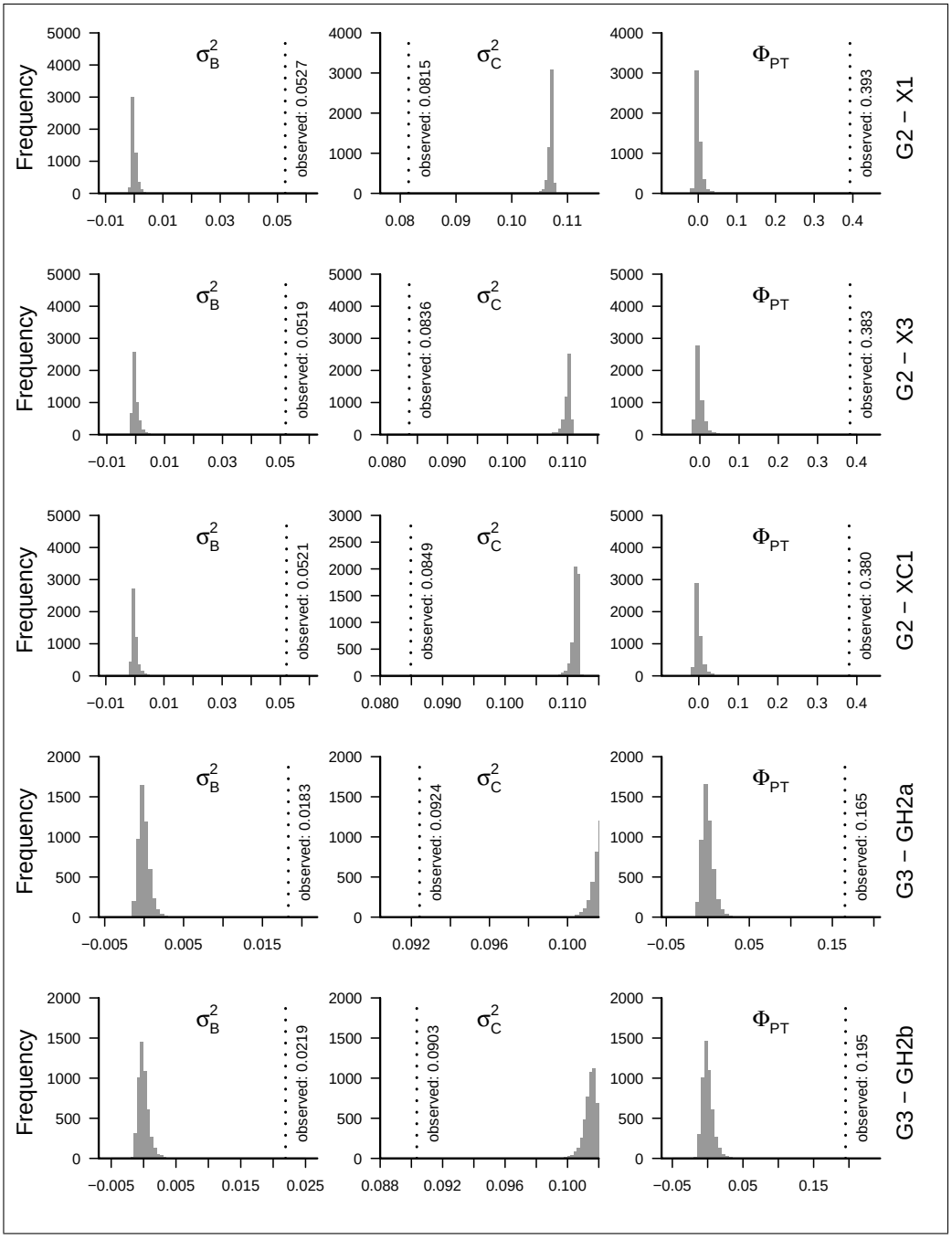

Figure E.4h: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

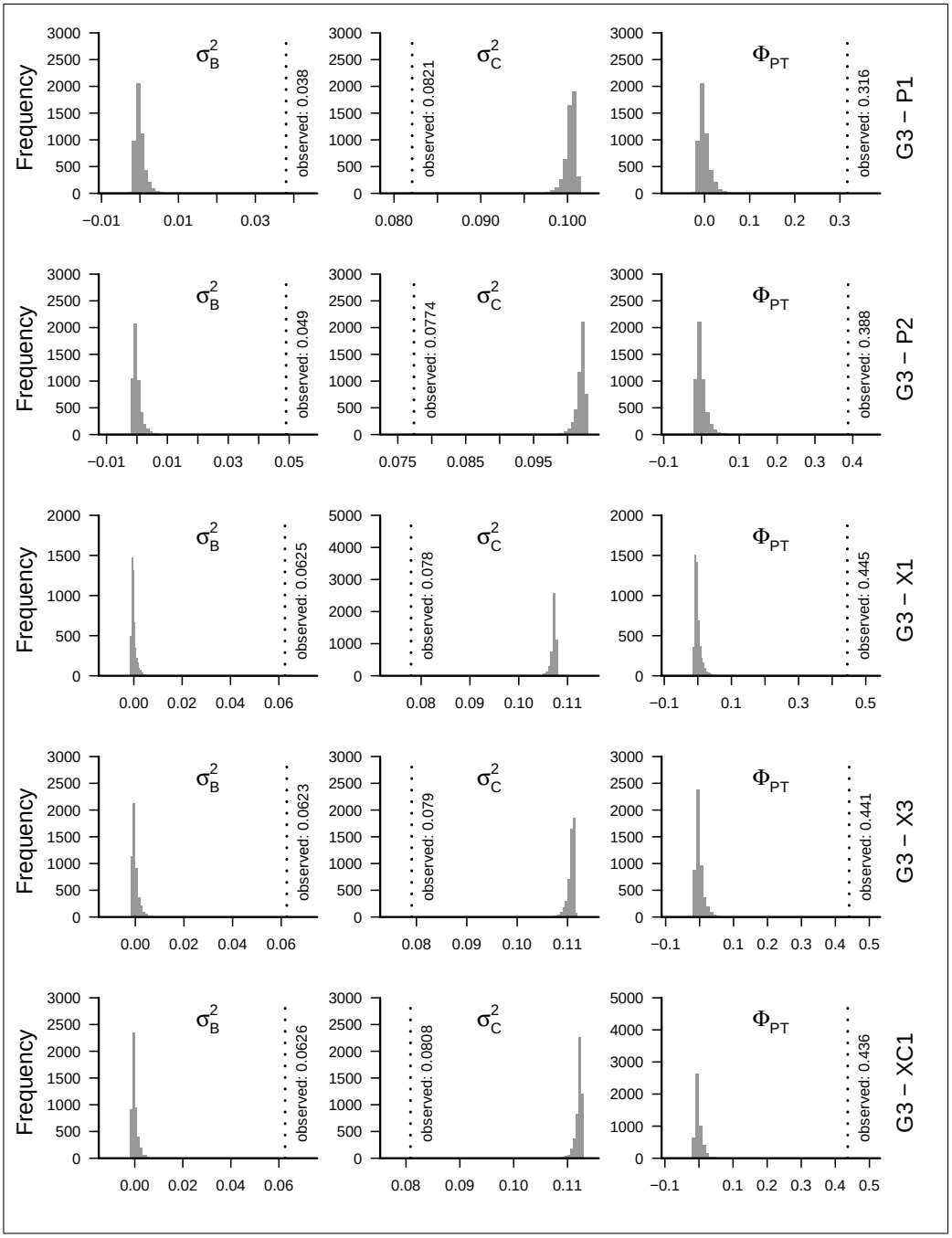

Figure E.4i: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

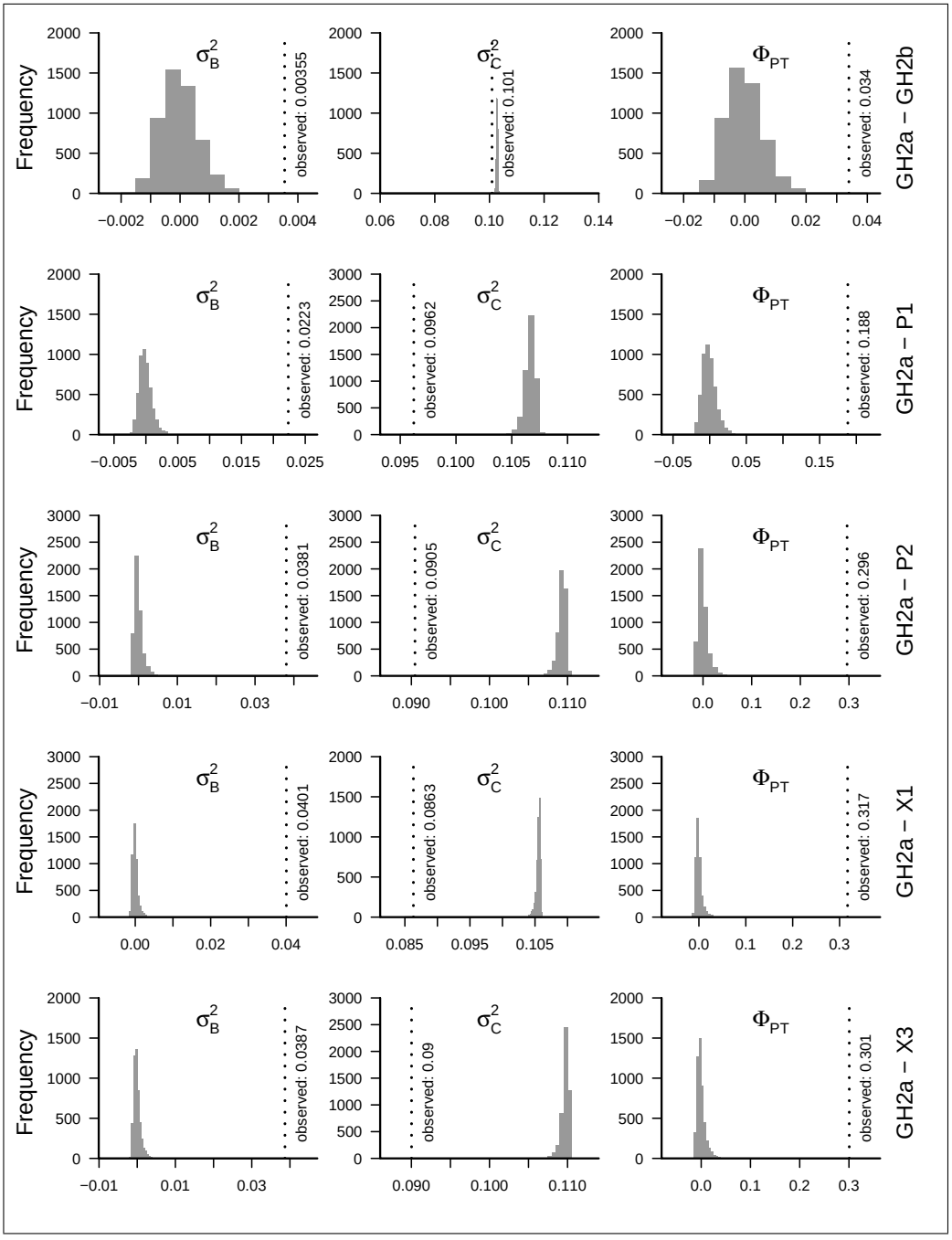

Figure E.4j: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

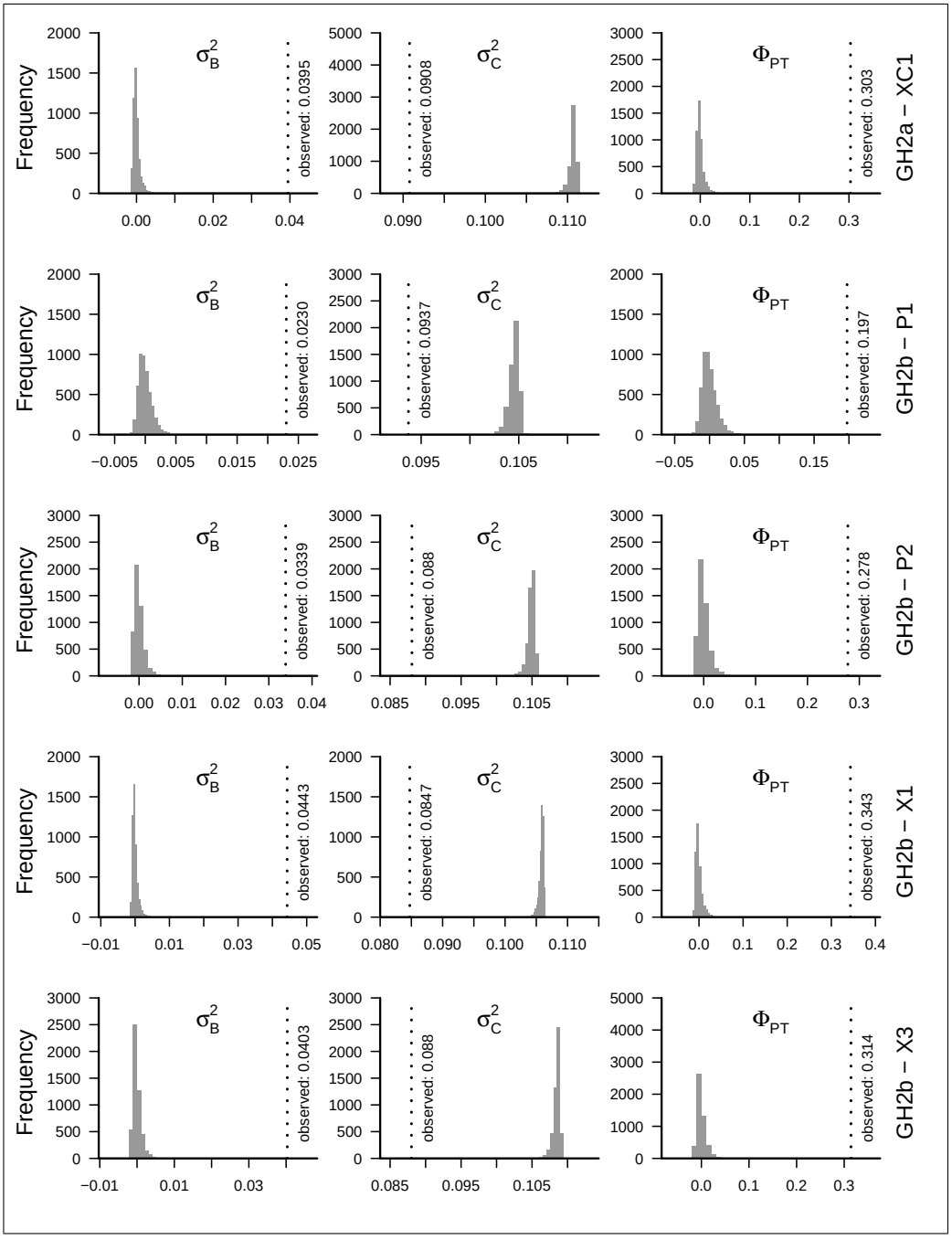

Figure E.4k: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

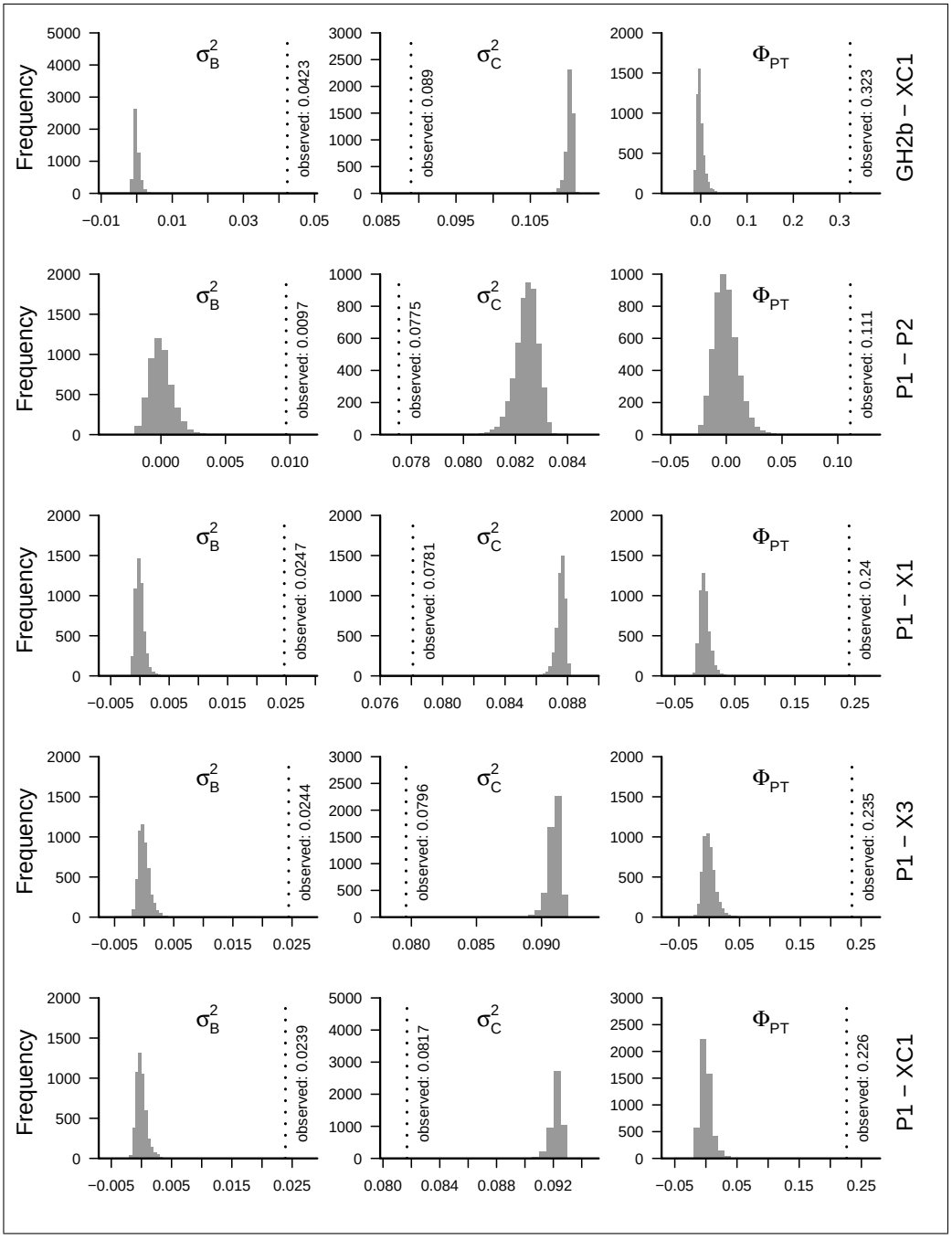

Figure E.4l: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

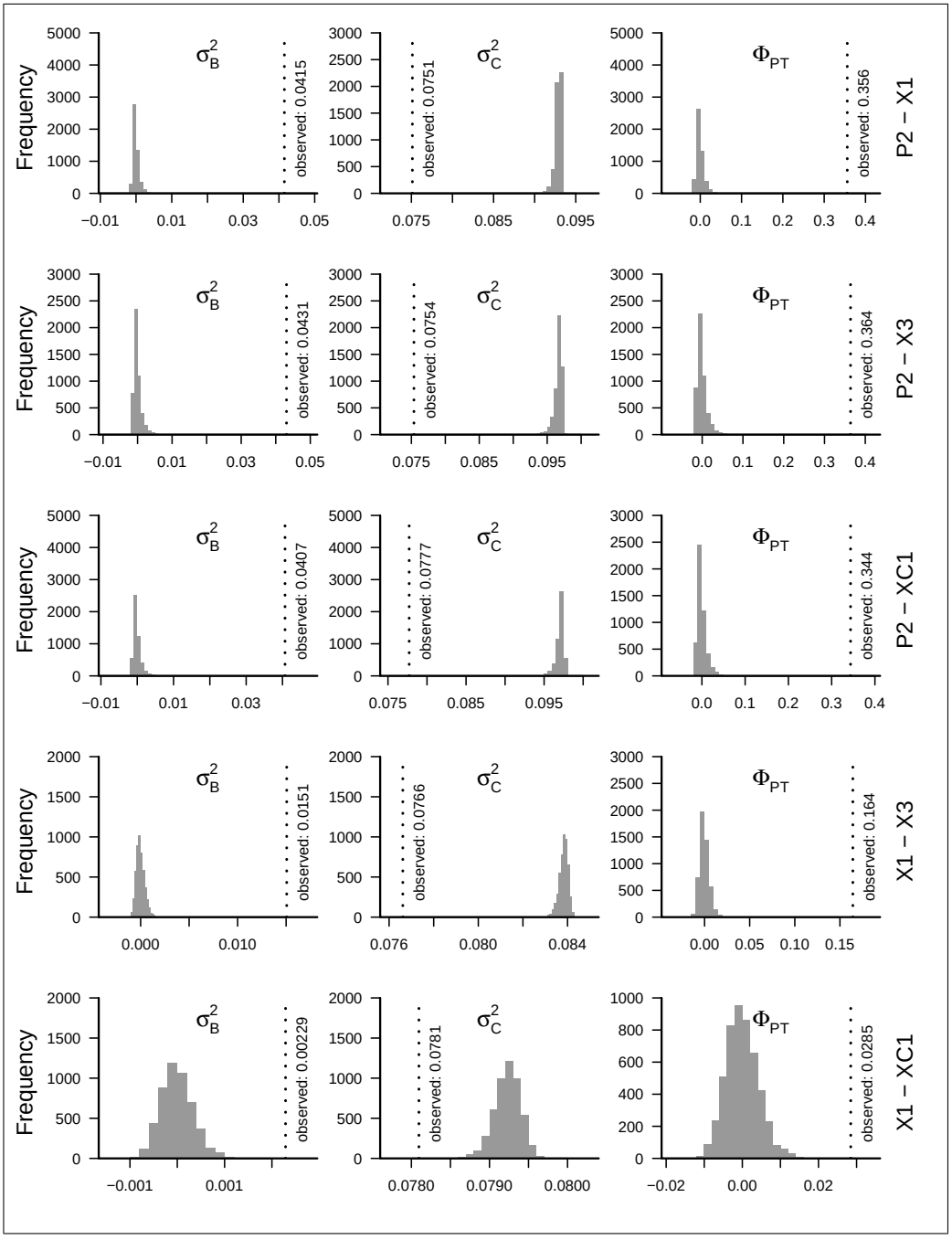

Figure E.4m: AMOVA **null-distributions (pairwise population comparisons,** Φ_{PT}**)** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

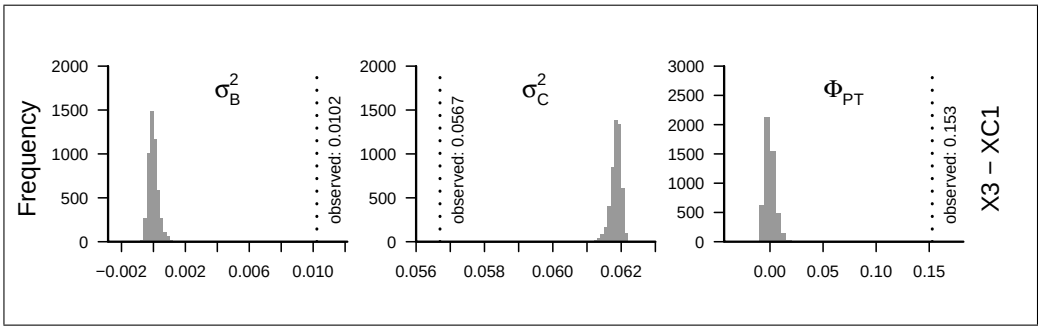

Figure E.4n: AMOVA **null-distributions (pairwise population comparisons, Φ_{PT})** obtained by randomising individuals over the dataset comprising the population pair; the corresponding pair for the row is given at the right margin (variation within populations σ^2_C, variation between populations σ^2_B).

Appendix F

Results of all 25 skew simulations

The graphs presented on the following pages are the results obtained by simulating datasets of dominant allele counts. In total 25 datasets (a–y), containing each 150 loci, were simulated; Every dataset comprises 40 individuals of each parent, 40 simulated RM-F1 hybrids, and 40 bottleneck F1 hybrids. (For a description of the simulation procedure please see chapter 2, subsection 2.3.8, page 87)

F.1 Marker deviation (Δ_{obs})

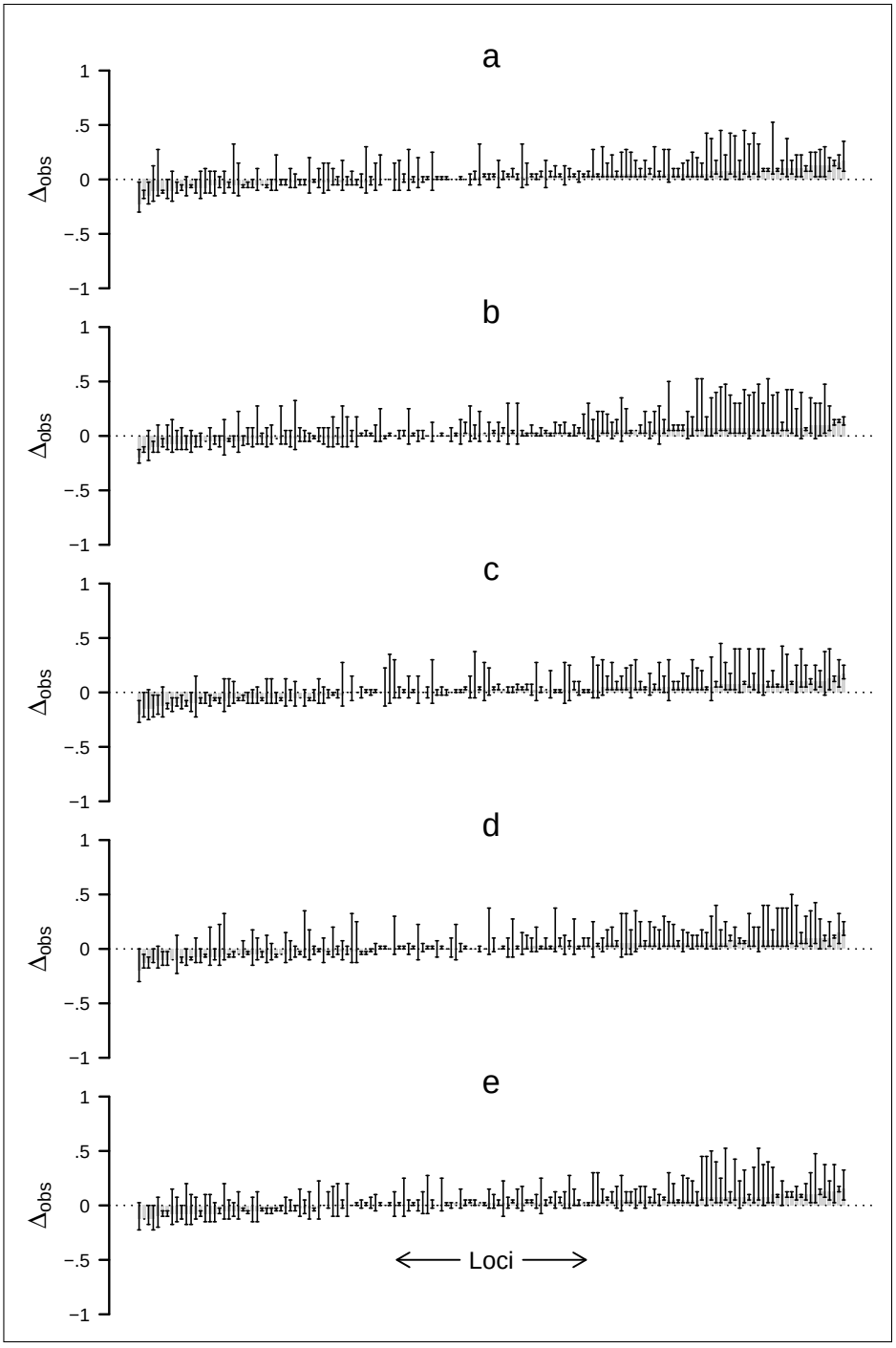

Figure F.1a: Δ_{obs}. Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and RM-F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

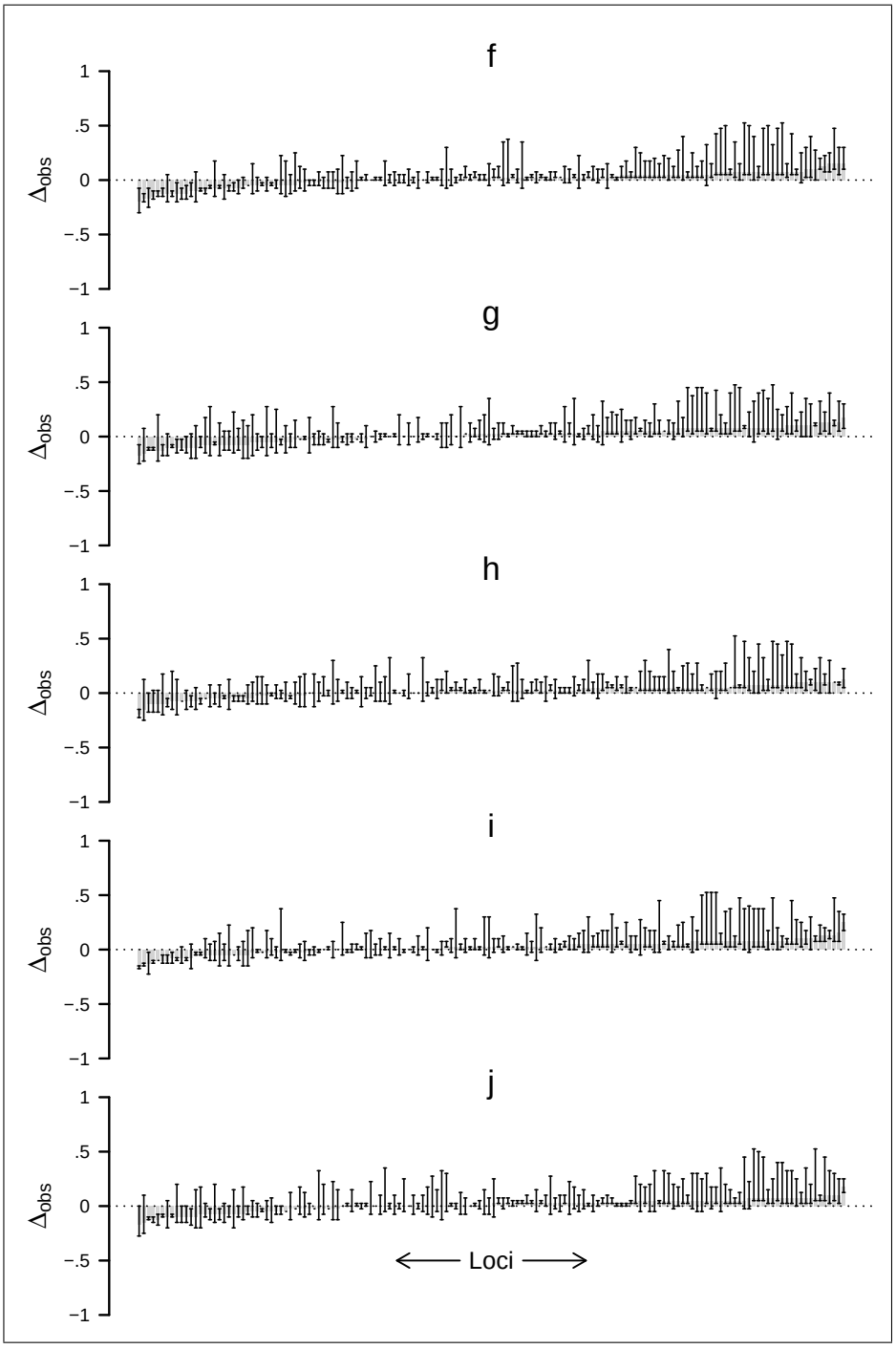

Figure F.1b: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and RM-F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

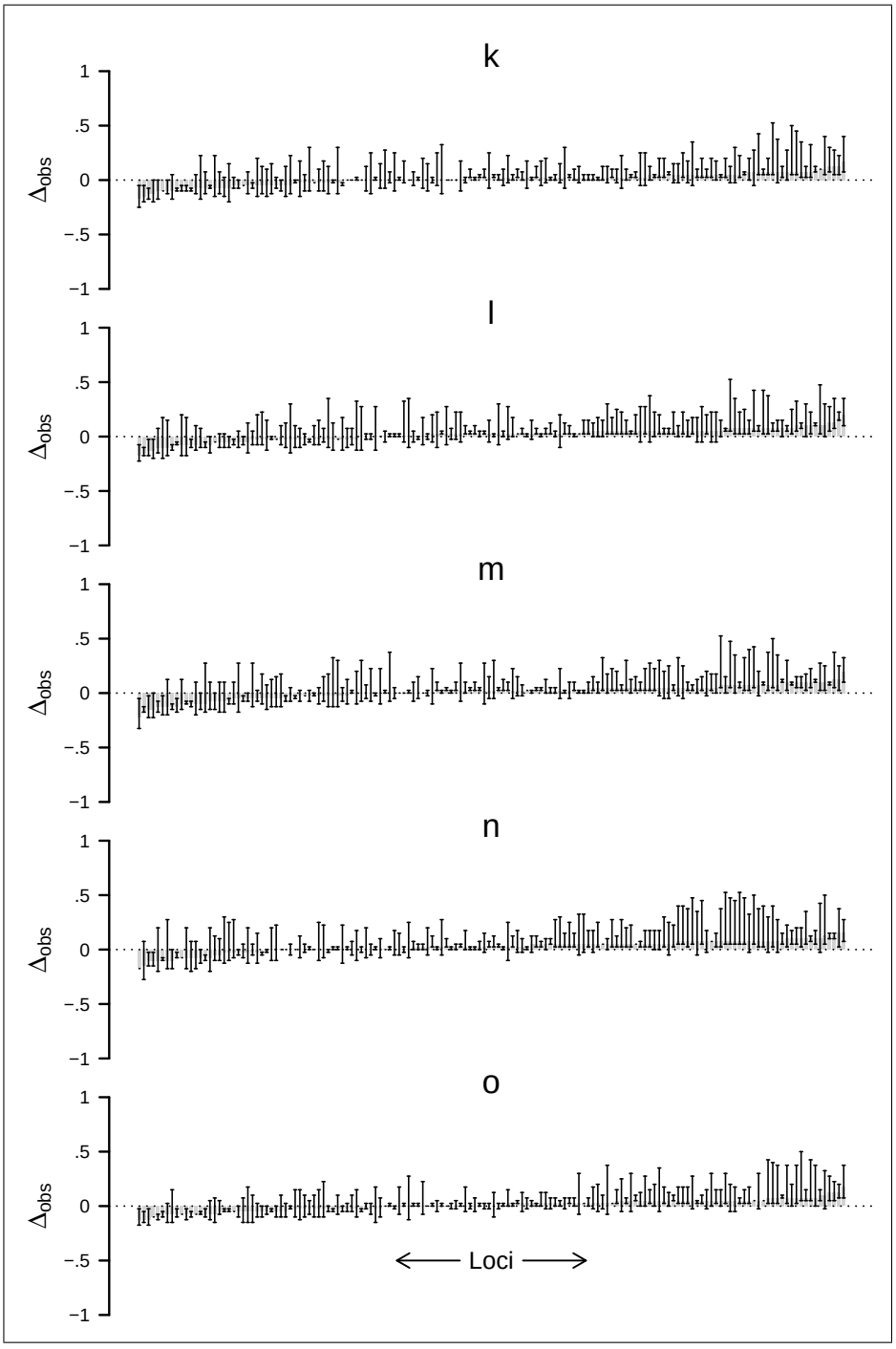

Figure F.1c: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and RM-F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

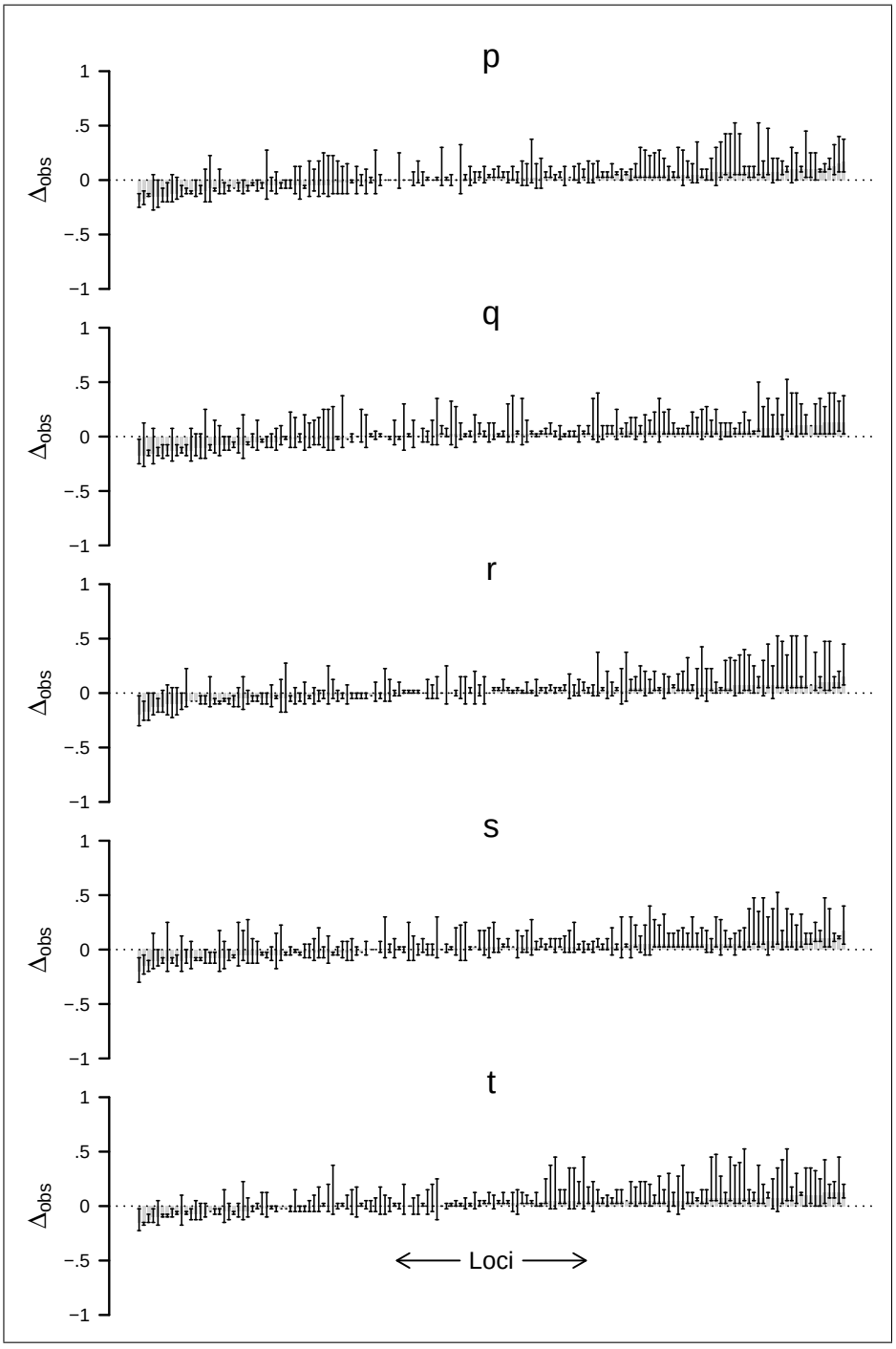

Figure F.1d: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and RM-F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

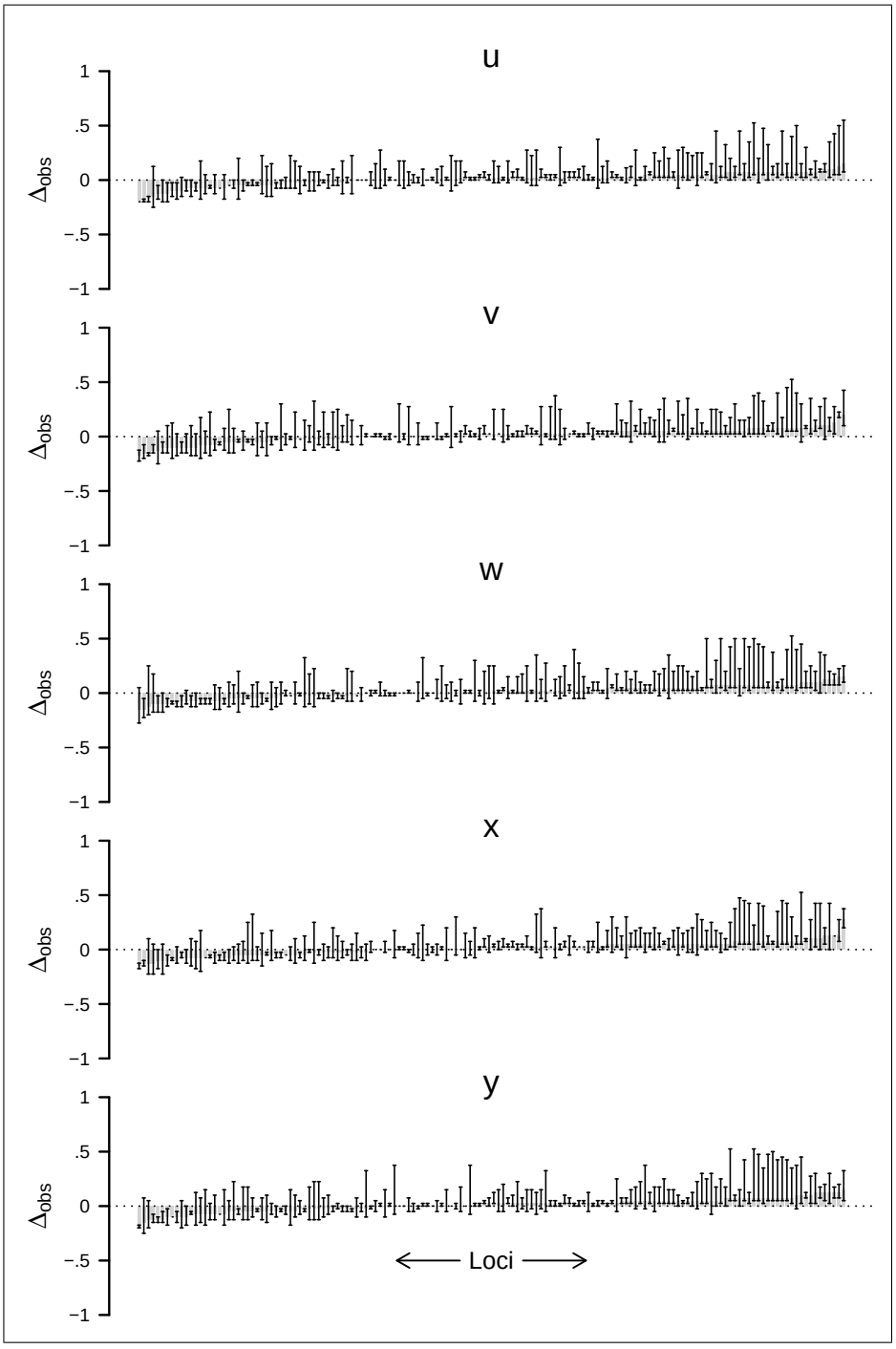

Figure F.1e: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and RM-F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

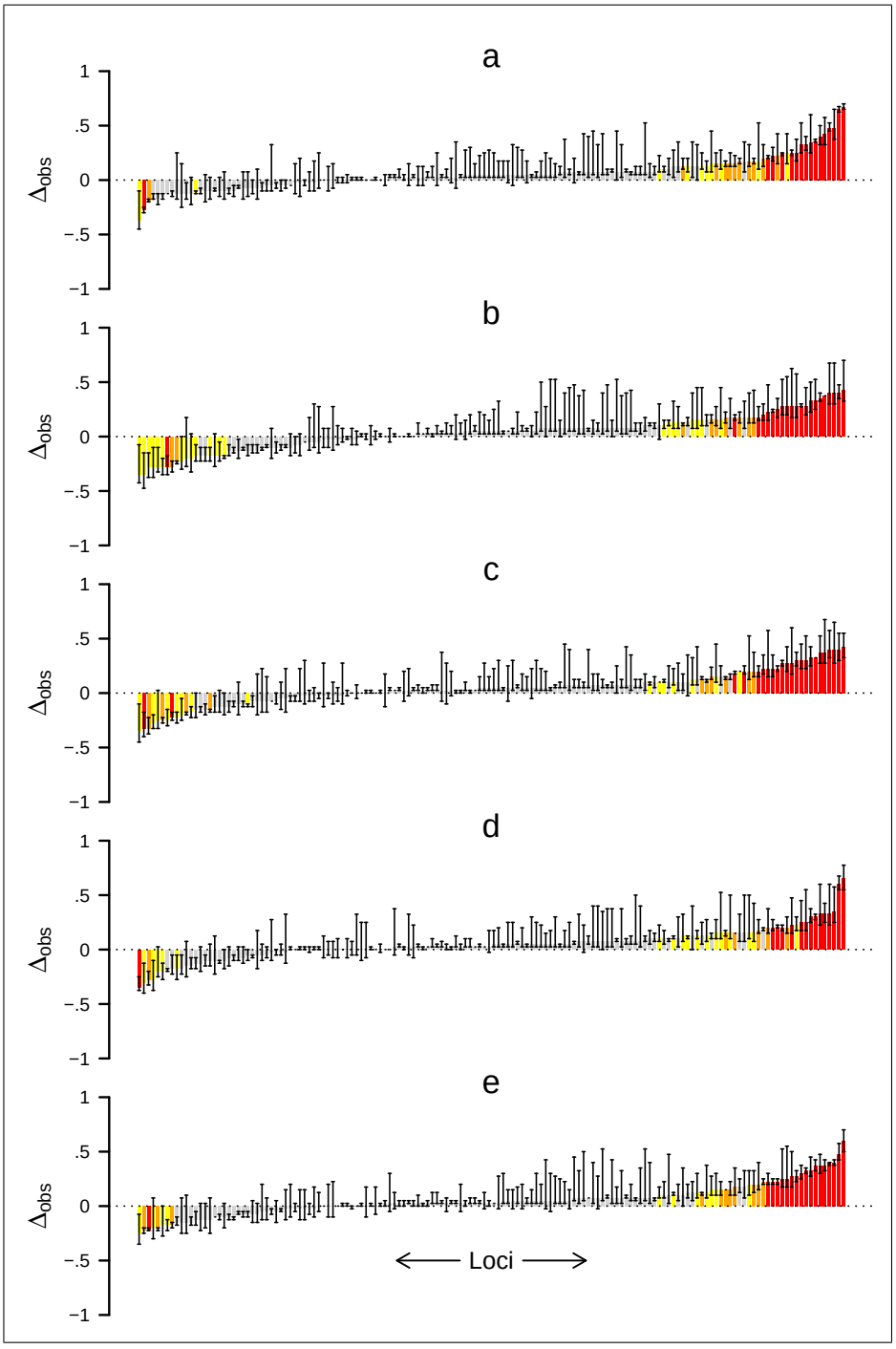

Figure F.2a: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and bottleneck F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

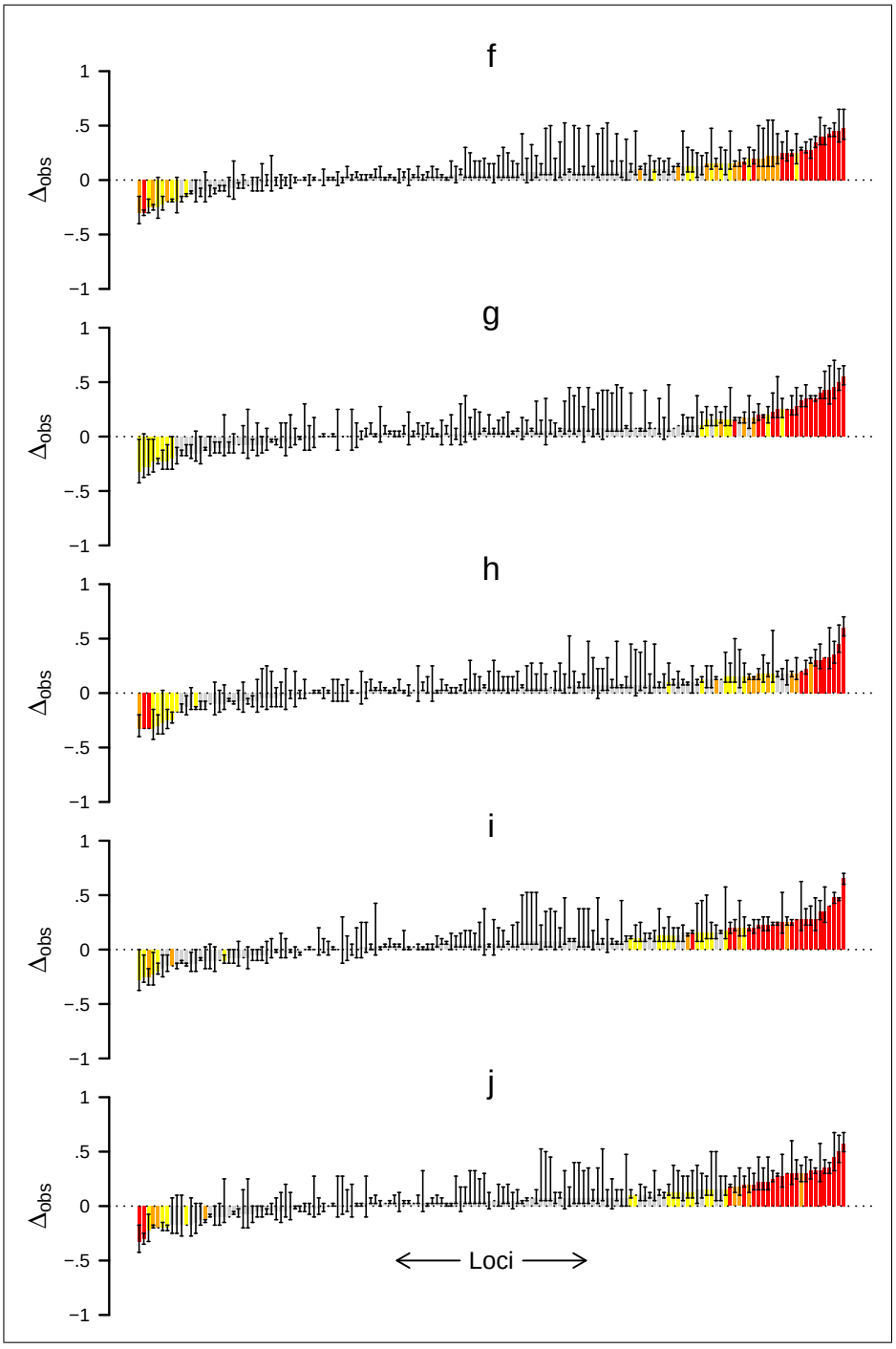

Figure F.2b: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and bottleneck F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

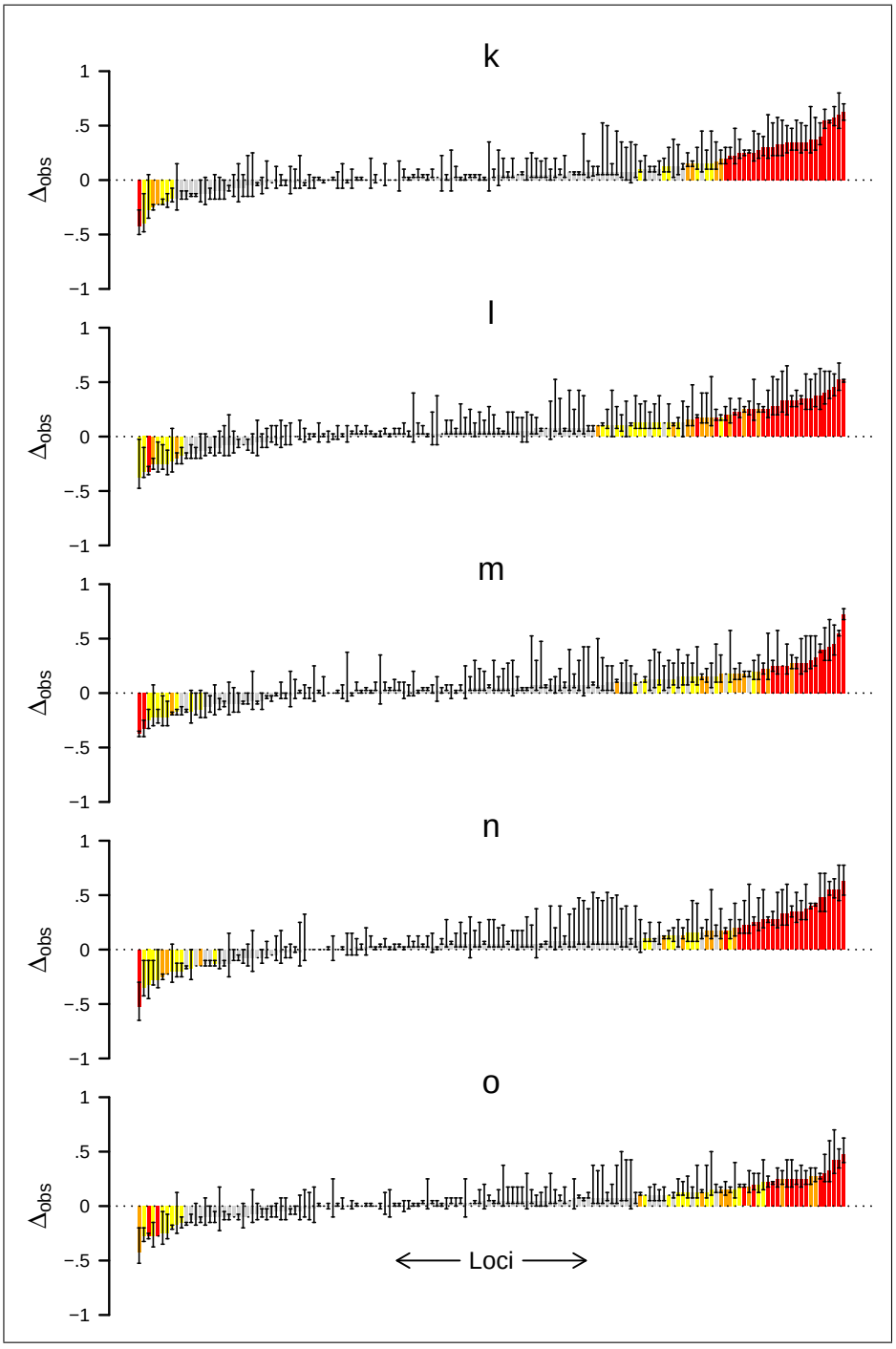

Figure F.2c: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and bottleneck F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

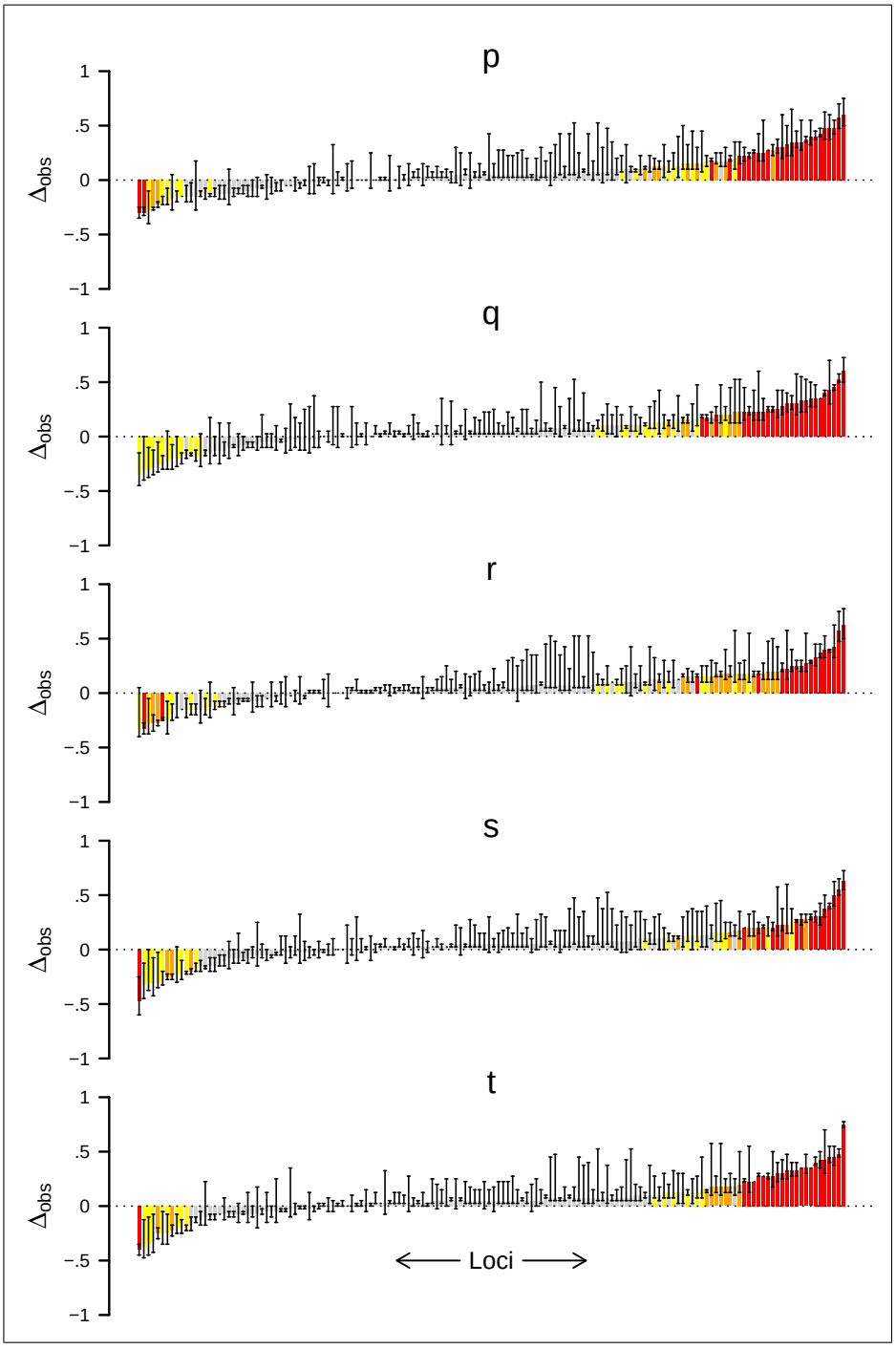

Figure F.2d: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and bottleneck F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

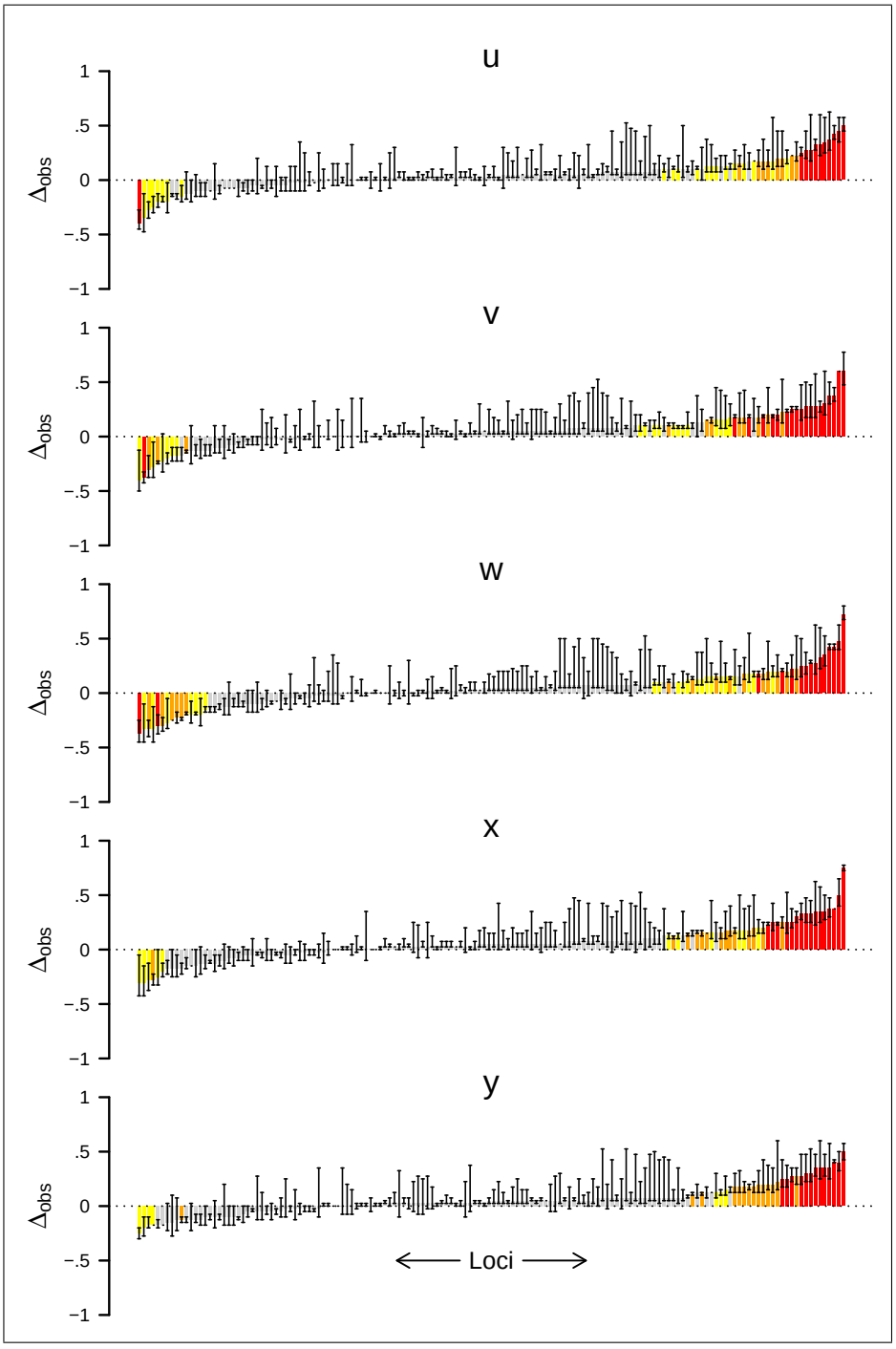

Figure F.2e: Δ_{obs}**.** Deviation of marker counts from simulated expectations for 25 simulations (a–y) of parents and bottleneck F1 hybrids. 'Error bars' show the range of Δ_{obs} including BC1s to both parents. (for a legend please compare to Figure 3.23, chapter 2, page 135)

F.2 Correlation of $|\Delta_{obs}|$ and F_{ST}

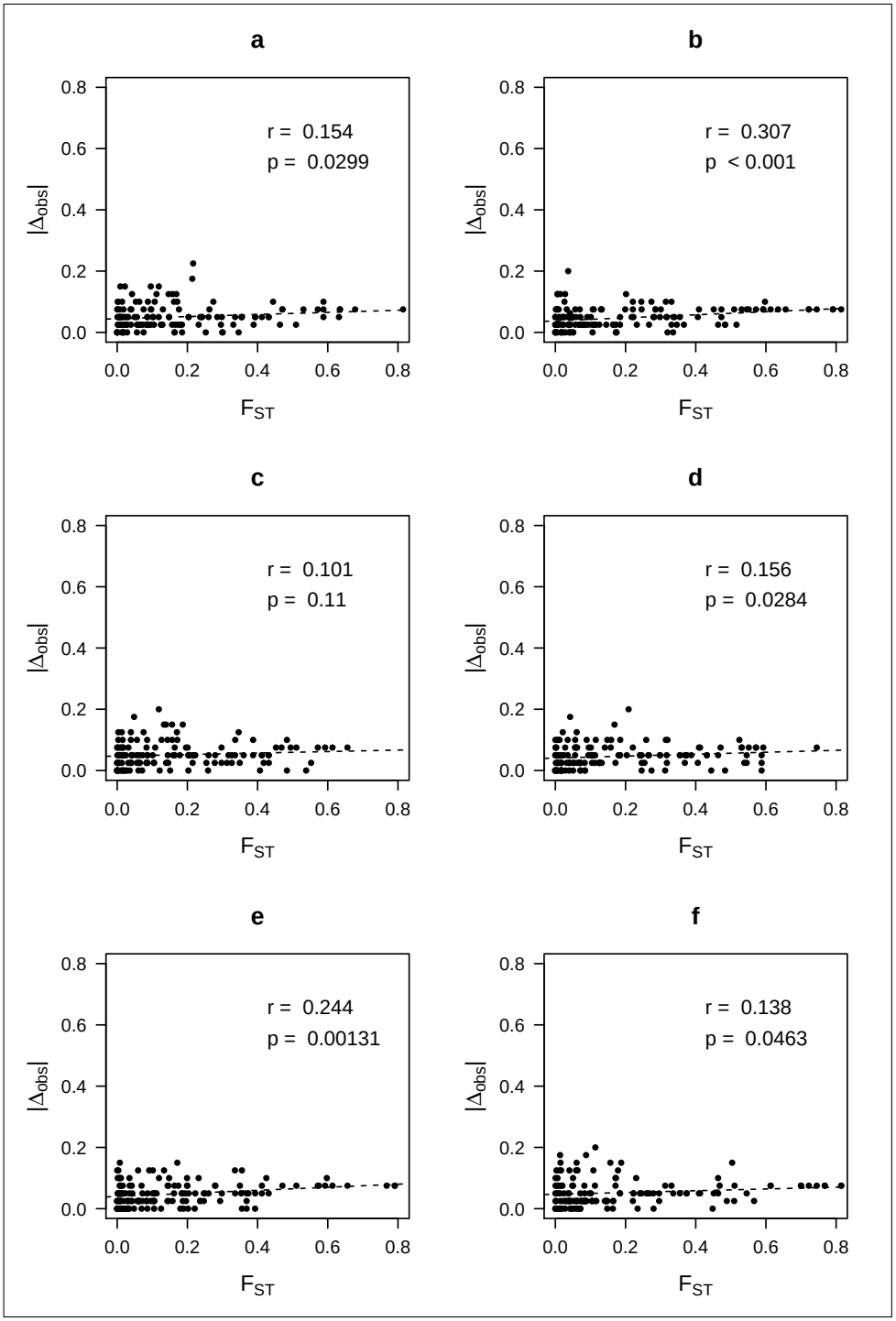

Figure F.3a: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and RM-F1 hybrids.

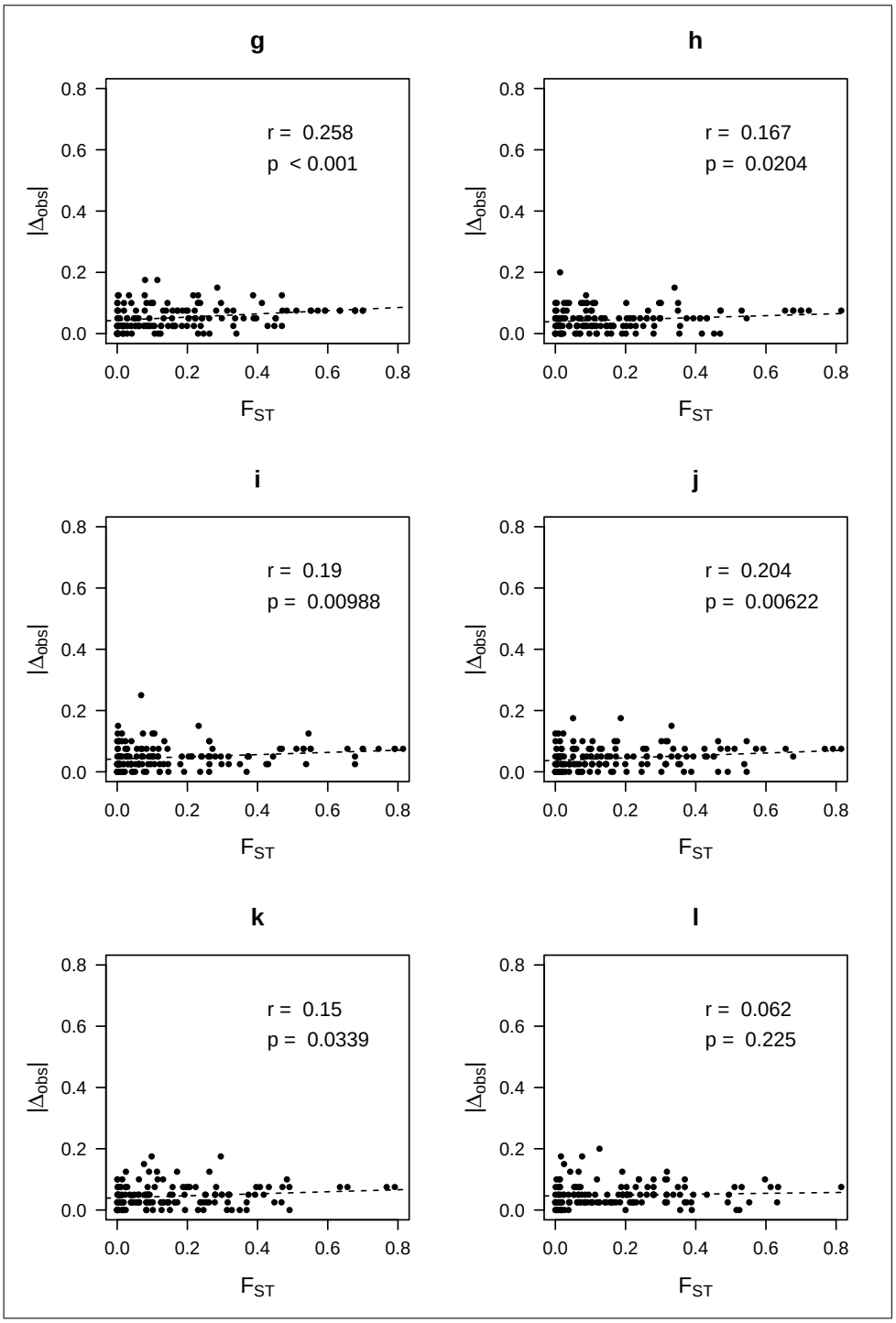

Figure F.3b: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and RM-F1 hybrids.

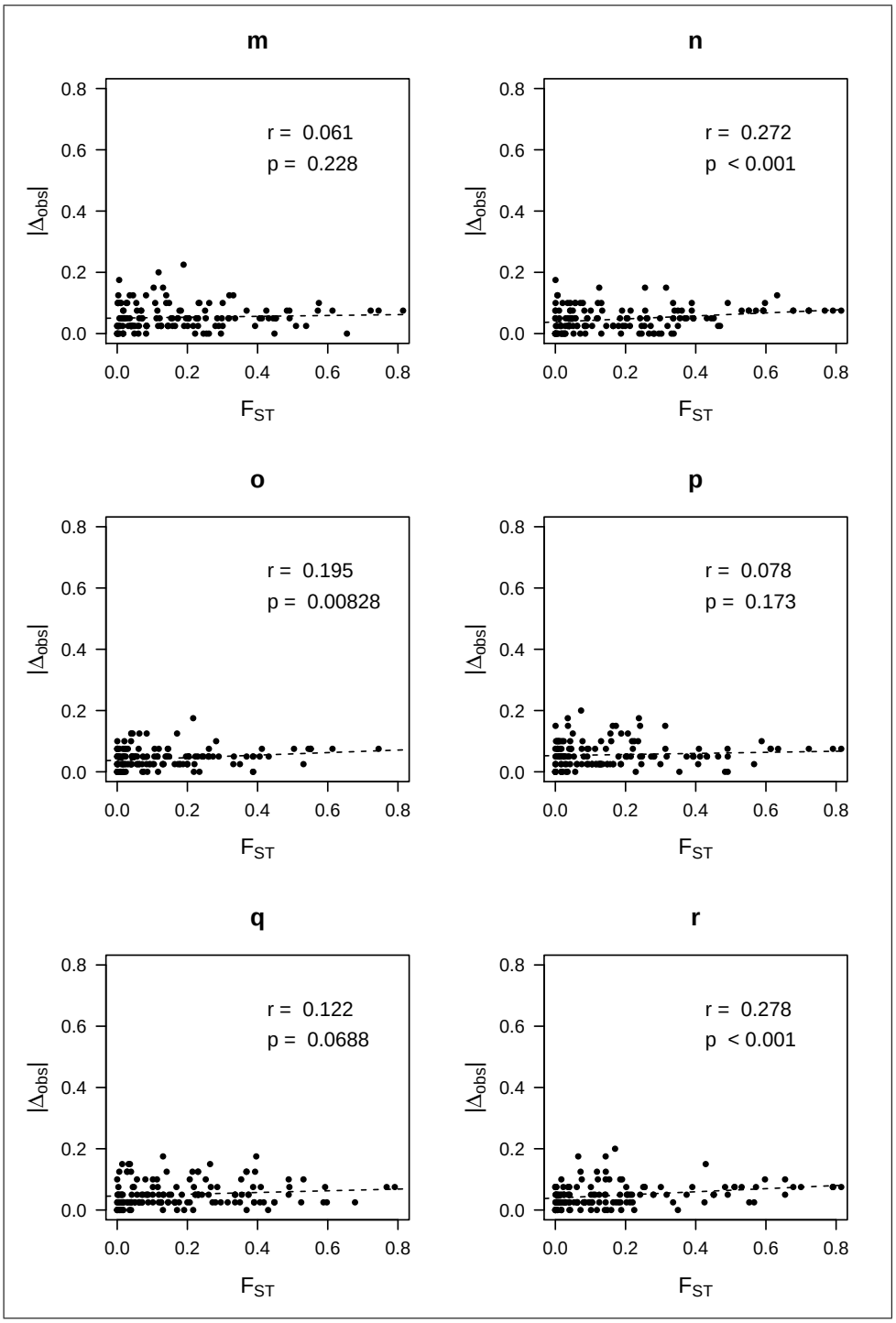

Figure F.3c: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and RM-F1 hybrids.

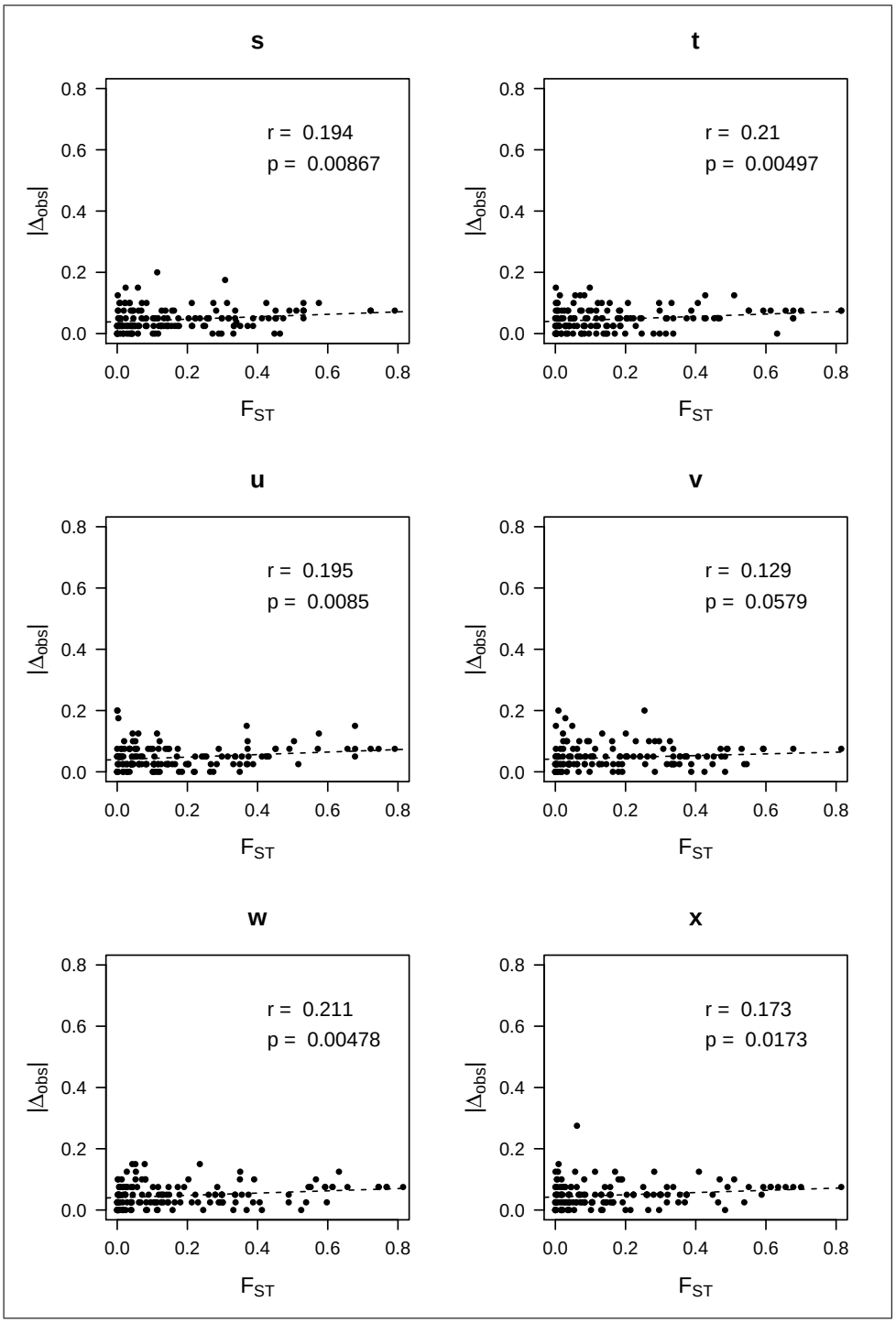

Figure F.3d: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and RM-F1 hybrids.

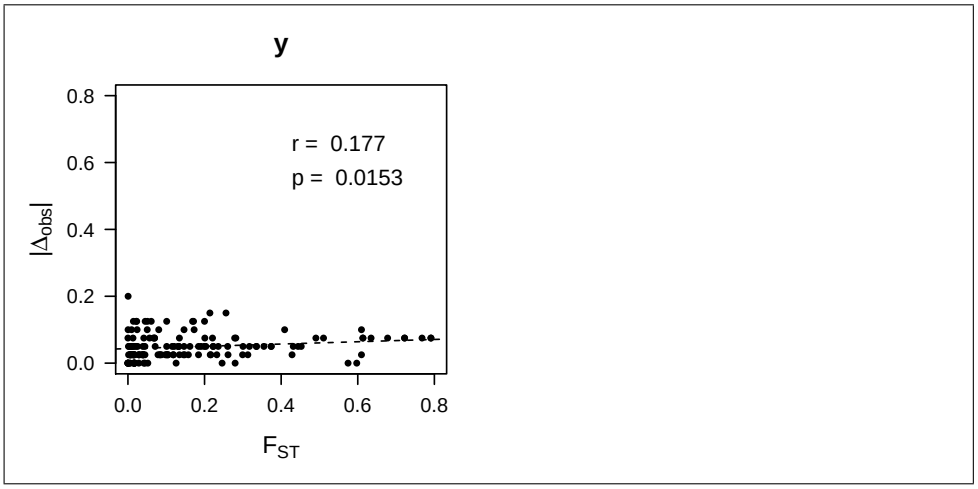

Figure F.3e: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and RM-F1 hybrids.

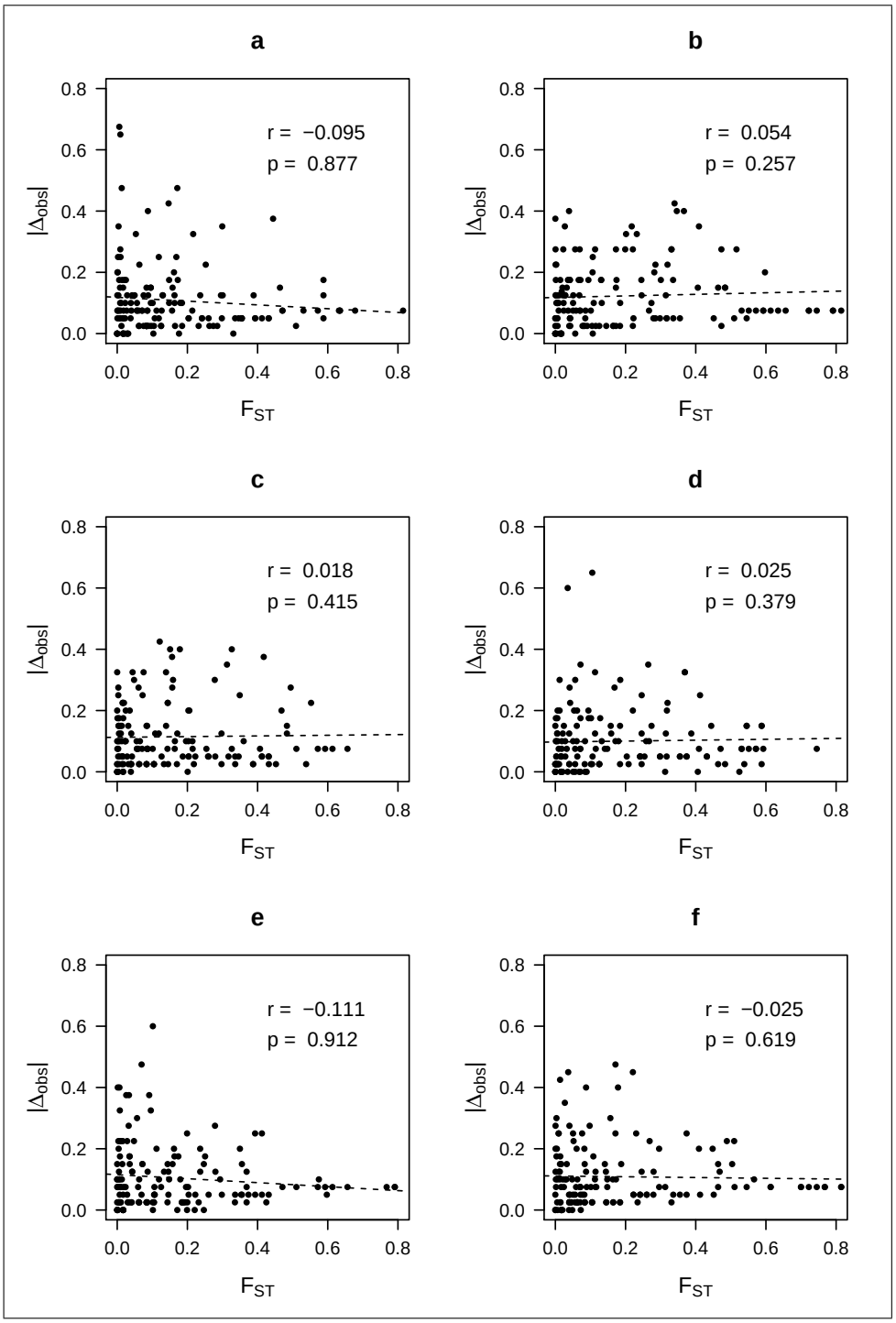

Figure F.4a: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and bottleneck F1 hybrids.

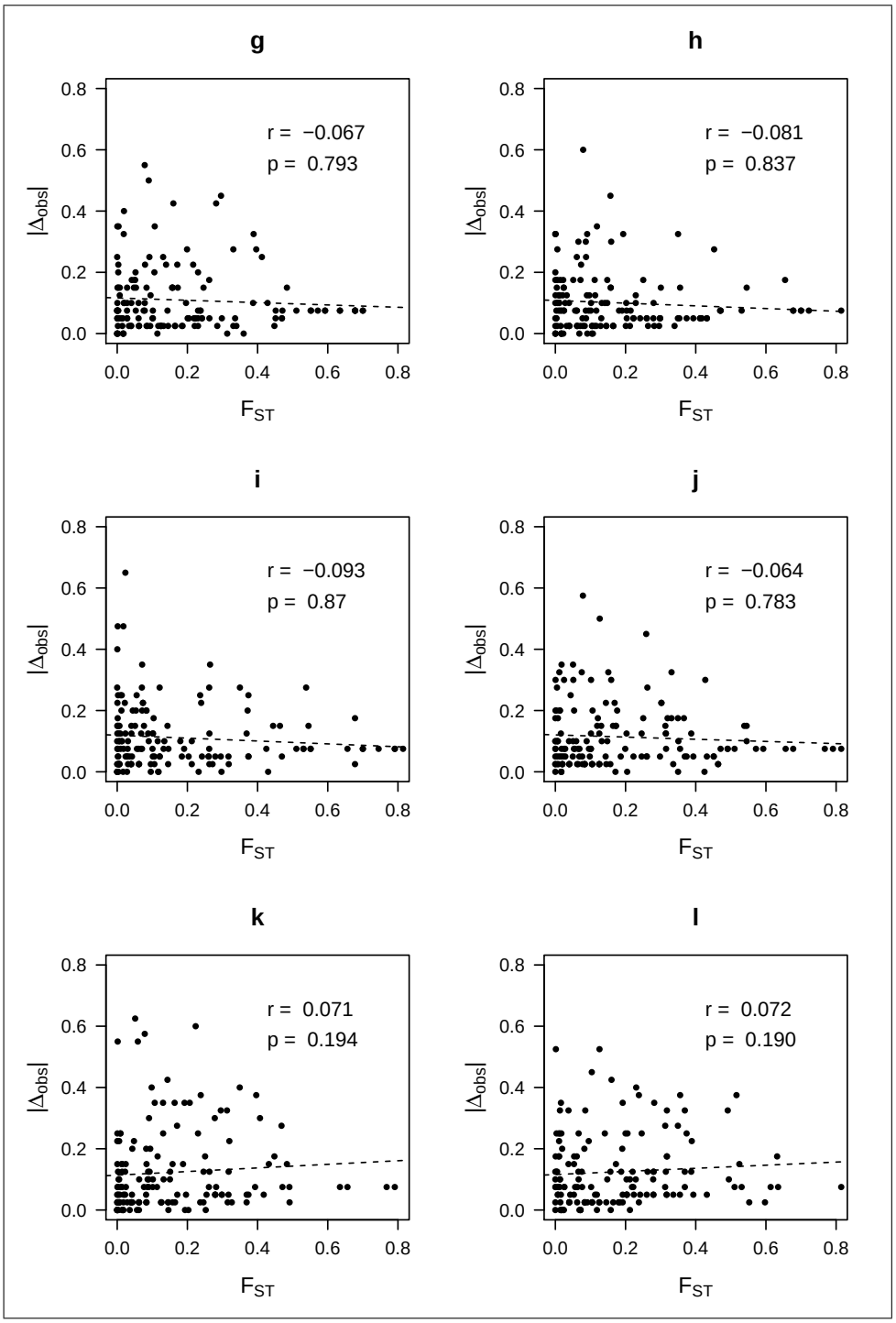

Figure F.4b: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and bottleneck F1 hybrids.

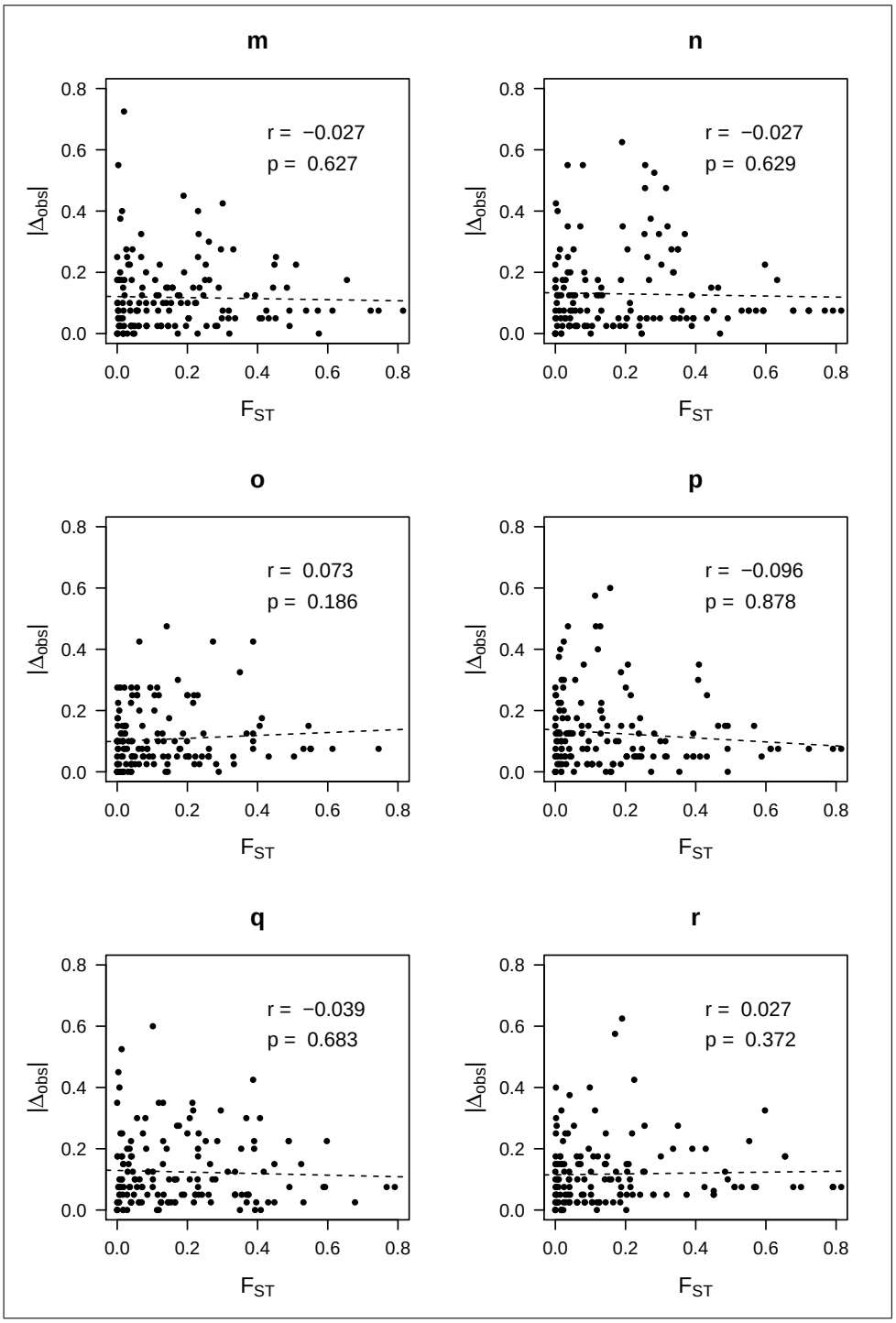

Figure F.4c: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and bottleneck F1 hybrids.

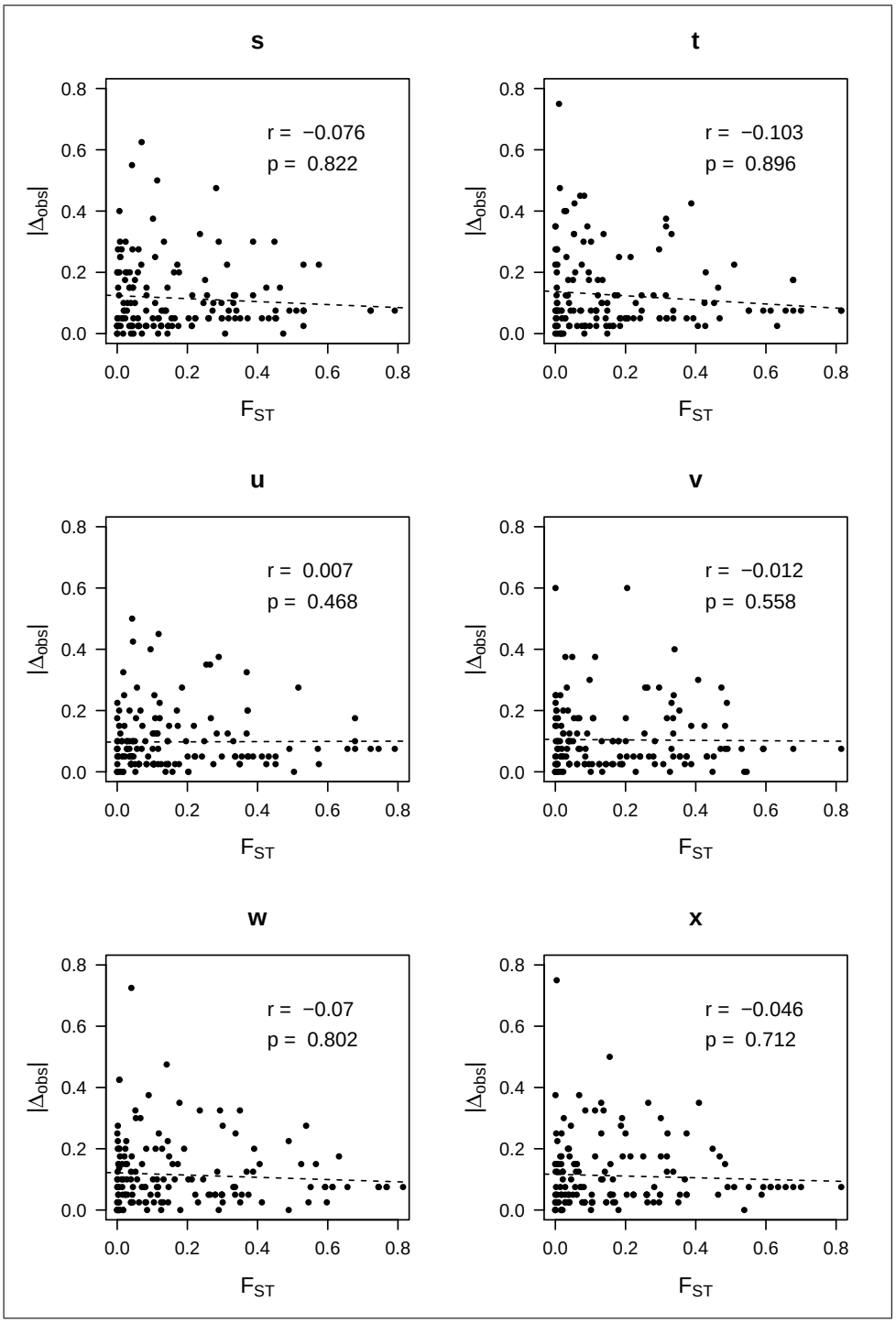

Figure F.4d: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and bottleneck F1 hybrids.

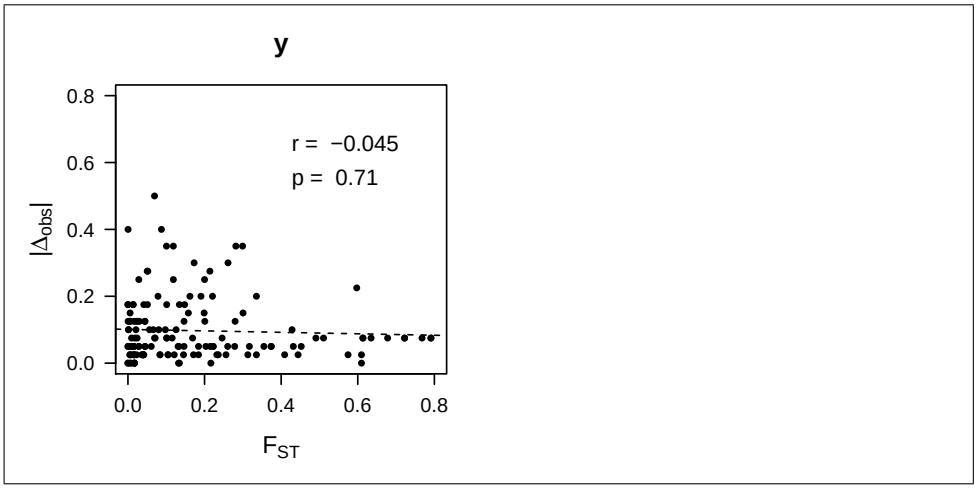

Figure F.4e: Correlation of F_{ST} and $|\Delta_{obs}|$ for 25 datasets (a–y) of simulated parents and bottleneck F1 hybrids.