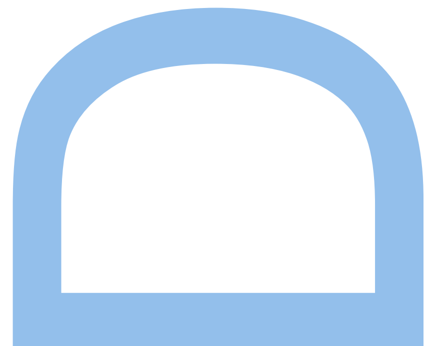
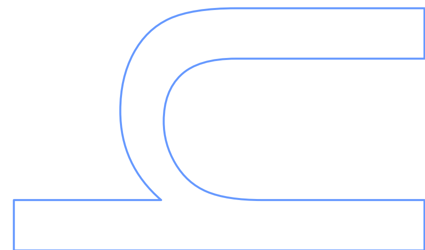
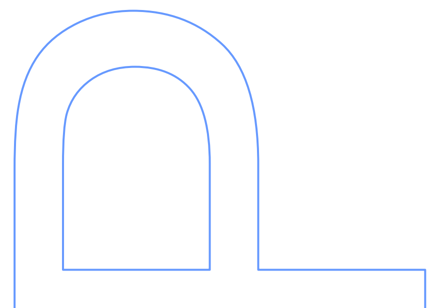
The role of inter-island gene-flow during colonization and speciation processes on archipelagos: analysis of the *Micromeria* Benth. (Lamiaceae) on the Canary Islands

Manuel António Cardoso Curto
Biodiversidade, genética e evolução
Departamento de Biologia
2017

Orientador
Harald Meimberg, University of Natural Resources and Life Sciences, Vienna







Nota prévia

Na elaboração desta dissertação, e nos termos do número 2 do Artigo 4º do Regulamento Geral dos Terceiros Ciclos de Estudos da Universidade do Porto e do Artigo 31º do D.L. 74/2006, de 24 de Março, com a nova redação introduzida pelo D.L. 230/2009, de 14 de Setembro, foi efetuado o aproveitamento total de um conjunto coerente de trabalhos de investigação já publicados ou submetidos para publicação em revistas internacionais indexadas e com arbitragem científica, os quais integram alguns dos capítulos da presente tese. Tendo em conta que os referidos trabalhos foram realizados com a colaboração de outros autores, o candidato esclarece que, em todos eles, participou ativamente na sua conceção, na obtenção, análise e discussão de resultados, bem como na elaboração da sua forma publicada. O candidato também gostaria de acrescentar que estes artigos só foram incluídos nesta dissertação. Os artigos aqui apresentados que foram publicados estão representados de forma integral com alterações de formatação.

Este trabalho foi apoiado pela Fundação para a Ciência e Tecnologia (FCT) através da atribuição da bolsa de doutoramento (SFRH/BD/79010/2011).

Acknowledgments

First I would like to thank my advisor Harald Meimberg for providing me with the opportunity to work on this project. He has broadened my horizons by encouraging and enabling me to work outside of Portugal, and his concern and advice since I started working with him has proved invaluable.

I also thank Pamela Puppo for all her contributions and hard work. This thesis would have not been possible without her significant contributions to the sampling, laboratory work, and her co-authorship of several papers. But above all, I would like to thank her for her advice and friendship.

I wish to thank everyone from CIBIO for providing me with an environment where I could develop myself as scientist. I would like to thank the people from CTM for all the help and support during the time I spent doing my lab work there. In particular, I would like to thank Susana Lopes for her availability in helping me with all protocol optimizations.

I would like to thank the people from the Technical University of Munich that helped me during the time I spent there. I would like to thank Christina Schachtler, Maximiliane Schuemann and Sarah Sturm for contributing their work concerning the preparation of RAD-sequencing and for making me feel welcome. I also would like to thank Saurabh Dilip for introducing me to a lot of the bioinformatics analyses that I used in this thesis.

I would also like to thank Christian Bräuchler and Pedro Luis Perez-de-Paz for providing some of the samples used in this thesis and Andreas Brachmann for helping with all Illumina sequencing.

Warmest thanks go to everyone from Colorado State University for welcoming me there. In particular I would like to thank Mark Simmons for advising me and for sharing his knowledge and wisdom with me. Additional thanks go to everyone from his laboratory for their friendship while I was there.

I thank Peter Sykacek for hosting and advising me during my first year in BOKU. I would like to thank the people from INF for providing help when it was needed.

Finally, I would like to thank my parents for their love, advice and support.

Resumo

A maioria da diversidade encontrada em ilhas oceânicas operou-se através de radiações adaptativas. Este processo é promovido pela grande quantidade de nichos vazios disponíveis neste tipo de sistemas. “Hybrid swarms” são populações que resultam da hibridação de diferentes linhagens. Este processo potencia probabilidade de ocorrência de radiações adaptativas porque a combinação de novos genótipos pode aumentar a capacidade de adaptação das populações envolvidas.

Desde o ano 2000 vários estudos descreveram uma diversidade genética superior ao esperado nas Ilhas Canárias o que levou à formulação da hipótese sobre a existência de “surfing syngameon”. Segundo esta hipótese, tal observação resultou da criação de “hybrid swarms ” através de múltiplos eventos de colonização do arquipélago. As populações daí resultantes divergem em espécies que estão ligadas por fluxo génico e evolvem dependentes umas das outras. Este tipo de espécies são conhecidas por "syngameon". Apesar da hipótese das “surfing syngameon” ter sido só aplicada para explicar a colonização de arquipélagos esta também poderá ser aplicada na colonização de cada ilha. Na tese aqui apresentada, testei se a esta hipótese esta correta usando as espécies de *Micromeria* endémicas das Ilhas Canarias como sistema.

Micromeria distribui-se por todas as ilhas e habitats das Ilhas Canárias e apresenta uma grande diversidade morfológica. Este género contem um grupo de espécies que são morfologicamente idênticas, sendo o exemplo mais relevante as espécies previamente descritas como *M. varia* s.l. De acordo com estudos de filogenia molecular elas têm origens independentes, o que é congruente com um processo de evolução convergente. No entanto, considerando os processos de evolução reticulada descritos em ilhas oceânicas, estes resultados também podem ser uma consequência de introgressão de outras espécies da mesma ilha impedindo a deteção de uma origem comum usando apenas alguns marcadores.

Nesta tese, foquei-me nas hipóteses relacionadas com a extensão da “syngameon hypothesis” para a colonização de cada ilha; e uma possível origem única do fenótipo de *M. varia* s.l. A primeira hipótese foi testada através da análise de padrões de estruturação genética, diversidade genética e fluxo-genético, numa perspetiva abrangente (arquipélago completo) e local (ilha de Tenerife). Para testar a segunda hipótese, dados genómicos foram usados para avaliar relações filogenéticas das espécies previamente descritas como *M. varia* s.l. Os objetivos visados só foram só concretizados através do desenvolvimento de novos marcadores moleculares. Três tipos de marcadores moleculares foram desenvolvidos: “exon primed and intron

spanning sequencing markers" (EPIC), microssatélites, e RAD-sequencing. Aqui, discuto o tipo de informação e de enviesamentos associados com os diferentes tipos de marcadores.

Os microssatélites mostraram uma diversidade genética mais elevada do que esperada em populações ocupando ilhas/regiões mais novas e a uma maior diferenciação genética em taxa ocupando ilhas mais antigas. Adicionalmente, foram detetados sinais de recente e histórico fluxo genético. Estes resultados são congruentes com um cenário onde cada ilha colonizada diversas vezes. Diversidade genética de diferentes origens e combinada em populações bastante variáveis através de hibridação. Este processo impede a perda de diversidade genética por efeito fundador criando condições ideais para a ocorrência de processo de radiação adaptativa. A especiação dessas populações resultam em syngameons. Em regiões mais antigas o fluxo genético não e ao eficiente na manutenção de conectividade entre espécies. Assim barreiras para o fluxo génico podem ser estabelecidas resultando na exclusão dessas espécies da syngameon.

Os dados de RAD-sequencing confirmara origens múltiplas para a morfologia *M. varia* s.l. suportando a ideia de evolução convergente. Adicionalmente, o sinal filogenético foi congruente com as zonas ecológicas presentes no arquipélago. Estes resultados em conjunto que a prevalência de fluxo-génico inter-especifico detetado com os microssatélites, levaram à criação da hipótese que este fenótipo apareceu apenas uma vez. Introgressão com espécies da mesma ilha contribuiu para a perda da maioria do genoma ancestral de *M. varia* s.l. Os genes responsáveis por esta morfologia foram selecionados durante o processo de adaptação.

Esta tese disponibiliza novos recursos moleculares que podem ser usados quer em *Micromeria* quer em géneros filogeneticamente próximos. Adicionalmente, o quadro teórico desenvolvido nesta tese pode ser aplicado não só em Micromeria ou em ilhas oceânicas mas também noutros sistemas onde hibridação promove radiação adaptativa ou adaptação local.

Palavras-chave: Ilhas oceânicas, Ilhas Canarias, *Micromeria*, radiação adaptativa, syngameon de espécies, microssatélites, RAD-seq, introgressão, hibridação, filogeografia, filogenómica, marcadores moleculares.

Abstract

Most of the diversity found in oceanic islands is produced by adaptive radiation events. This process is promoted by the high availability of free niches and isolation from the mainland typical for these systems. Hybrid swarms, where different lineages are combined through hybridization, are more likely to result in adaptive radiation events because new combinations of genotypes increase the adaptive ability of populations.

Since the year 2000 several studies suggested that genetic diversity was higher than expected in the Canary Islands, which led to the creation of the surfing syngameon hypothesis. This hypothesis states that this pattern of high genetic diversity is a consequence of the creation of hybrid swarms after multiple colonization events of the archipelago. These populations diverge into species connected by gene-flow and evolve dependently from each other, also known as species syngameon. Although the surfing syngameon hypothesis was originally developed to explain the colonization of the entire archipelago, the same process might be applied to the colonization process of each individual island. In the present thesis I intend to test if this is the case by using the *Micromeria* species endemic to the Canary Islands as a system.

Micromeria is distributed throughout all islands and ecological zones of the Canary Islands, and it has a high morphological variation. This genus contains groups of morphologically identical species, the most remarkable example being the species previously classified as *M. varia* s.l. Based on molecular phylogenetic analyses they show independent origins, and since they occupy similar habitats on different islands, a case of convergent evolution is indicated. Alternatively, in light of the reticulate patterns described on oceanic islands, these results can also be a consequence of introgression with other species from the same islands. In this situation, the detection of a single origin is prevented if only a few markers are used.

In this thesis I focus on and compare the hypothesis regarding the extension of the surfing syngameon hypothesis to the colonization of each island and the possible single origin of the *M. varia* s.l. phenotype. The first hypothesis was tested by analyzing genetic structure, gene-flow, and diversity patterns in wider (entire archipelago) and local (Tenerife) scales, including all *Micromeria* species. To test the second hypothesis, a genomic dataset was used to evaluate the phylogenetic relationships of the species previously classified as *M. varia.* These objectives could only be addressed by developing new sets of molecular markers. Three marker sets were developed: exon primed and intron spanning sequencing markers (EPIC); microsatellites; and RAD

sequencing markers. The type of information they provide and the biases associated with different types of markers are discussed.

Microsatellite loci showed a higher genetic diversity than expected in populations occupying younger areas/islands, whereas there was a higher genetic differentiation among taxa occupying older regions. Additionally, signs of current and historical gene-flow between species from the same and different islands were detected. These results are congruent with a scenario where each island is colonized multiple times. Genetic diversity from multiple sources is then combined into a highly diverse population through hybridization. This process prevents the loss of genetic diversity through the founder effect by creating ideal conditions for adaptive radiation. When these populations speciate the resulting species remain connected by gene-flow contributing for the expansion of the species syngameon. In older regions gene-flow stops being as effective in the remaining connected species, resulting in the establishment of reproductive isolation and thereby excluding them from the syngameon.

The RAD-sequencing dataset, confirmed multiple origins for the species showing the *M. varia* phenotype, supporting the idea of convergent evolution. Moreover, the first time phylogenetic analyses was congruent with the distribution of ecological zones in the archipelago. These results together with the high prevalence of inter-specific gene-flow detected with microsatellites, led to the creation of the hypothesis that this morphology appeared just once and was spread through the archipelago. Introgression with other species from the same archipelago contributed to the loss of most of the ancestral *M. varia* s. l. genome. The genes responsible for this morphology were selected during the adaptation process.

This thesis provides new molecular resources that can be used both in *Micromeria* species and in some closely related genera. Theoretical framework developed here may not only be applied for research focusing in *Micromeria* or in oceanic islands, but also in other systems where adaptive radiation or local adaptation is promoted by hybridization.

Key words: Oceanic islands, Canary Islands, *Micromeria*, adaptive radiations, species syngameon, microsatellites, RAD-seq., introgression, hybridization, phylogeographic, phylogenomics, molecular markers.

Table of contents

Resumo ..... **i**

Abstract ..... **iii**

Table of contents ..... **v**

List of tables ..... **vii**

List of figures ..... **x**

Chapter 1: General Introduction **1**

1.1. Adaptive radiation and hybrid speciation 3

1.2. Islands as models to study evolution 4

1.3. Study system 7

Canary Islands *7*

Micromeria species from the Canary Islands *9*

Micromeria phylogenetic patterns and evolutionary predictions *12*

1.4. Molecular markers: applications and development 15

1.5. Objectives and thesis outline. 16

1.6. References 18

Chapter 2: Molecular markers development **27**

2.1. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae) 29

2.3. Evaluation of microsatellites of Catha edulis (qat; Celastraceae) identified using pyrosequencing 50

2.4. Influence of past agricultural fragmentation to the genetic structure of *Juniperus oxycedrus* in a Mediterranean landscape 68

Chapter 3: *Micromeria* evolution in the Canary Islands **93**

3.1. Genetic structure of *Micromeria* (Lamiaceae) in Tenerife, the imprint of geological history and hybridization on within-island diversification. 95

3.2. Genetic diversity and differentiation patterns of *Micromeria* from the Canary Islands are congruent with multiple colonization dynamics and the establishment of species syngameons 127

Chapter 4: Distribution of adaptive traits during adaptive radiation. **157**

4.1. Phylogenetic analyses of RAD data to infer the evolution of *Micromeria* in the Canary Islands. 159

Chapter 5: General discussion **181**

5.1 New insights into evolution on oceanic islands 183

Evidence of evolution with gene-flow and its impact on genetic variation *183*

Island geological dynamics has an impact on hybridization and diversification patterns *185*

5.2. Effect of syngameon dynamics on species definition and morphological traits 188

Maintenance of species boundaries despite gene-exchange *188*

The appearance and maintenance of morphologically similar species with independent origins *189*

5.3. Some remarks about molecular markers 192

Comparison of information obtained among marker systems *192*

Is it still worthy to invest in classical methods? *194*

5.4. Future directions 195

5.5. Summary of the major conclusions 197

5.6. References 198

Appendix 205

Appendix I: Supplementary material from section 2.3. 207

Appendix II: Supplementary material from section 2.4. 213

Appendix III: Supplementary material from section 3.2. 221

Appendix IV: Supplementary material from section 3.3. 227

Appendix V: Supplementary material from section 4.2. 239

List of tables

Chapter 1: General Introduction

Table 1. List of Micromeria species described in the Canary Islands according to their current and previous classifications. Reclassified species are in bold...................... 10

Chapter 2: Molecular markers development

Paper I

Table 1. Primer information for the loci presenting positive amplification results. The information comprises locus name and primer sequence, accession number of the sequence from which the primer was designed and respective organism and gene. For loci indicated with asterisk (∗) more than one primer pair was designed............ 33

Table 2. Amplification and sequencing results of primer pairs tested in *Micromeria*. The different numbers correspond to different levels of amplification success: 0 – no product; 1 – amplification and sequencing success; 2 – amplification of multiple products; and 3 – successful amplification but no readable sequences. Note that Ba: Baleares, Te: Tenerife, Ma: Madeira, La1: Lanzarote sample one, La2: Lanzarote sample two and GC: Gran Canaria.y: yes, n: no.. 39

Table 3. Information on the phylogenetic analyses, marker variability in the ingroup and outgroup; ingroup sequence alignment length and percentage of exon region in that alignment. The phylogenetic analyses information corresponds to the trees obtained using different genera as outgroup and include number of trees, tree length, CI (consistency index) and RI (retention index). The variability measures shown are the number of polymorphic positions (left value) and number of informative positions (right value). 41

Table 4. Test of different phylogenetic hypotheses using the 22 markers that were used in step 4 (Fig. 1), based on interpretation of the phylogeny and the SH test. Indicated are four classes of support, 1: the topology supporting the hypotheses (bootstrap value is shown), 2: the topology contradicting the hypothesis (–), 3: the topology is not resolved enough to support or contradict the hypothesis (0), 4: taxon sampling is not sufficient to test the hypothesis (?). The SH test was performed using the maximum likelihood trees. The SH test marked with an asterisk used the combination of the east–west split and Gran Canaria monophyly as null hypotheses. 42

Paper II

Table 1. List of 23 qat samples. The population number was defined according to location where the samples were collected. ... 53

Table 2. Number of sequences retained after each of the selection steps. The Sciroko fraction contains all sequences comprising any kind of SSR motif including interrupted repeats. This fraction was analyzed for sequences only containing perfect repeats, at least six perfect repeats and sequences that also contained flanking regions longer than 30 bp (optimal). From the optimal fraction we designed 63 primer pairs (synthesized primers) from which 43 show successful amplification. ... 56

Table 3. Amplification, allele calling and variability of the final set of 27 loci. ... 59

Paper III

Table 1. Number of individuals sampled for genetic analysis per a priori population. For each population, the table also shows the number of individuals measured for trunk perimeter, minimum and maximum measures, the average and standard deviation of the trunk perimeter, and the number of unmeasured individuals with trunk perimeter visually estimated to be > 100 cm ... 73

Table 2. Markers variability, amplification success and Hardy-Weinberg deviation statistics. ... 77

Table 3. Summary results of AMOVA analyses for individuals grouped according to the sampled population, trunk perimeter (<60 cm vs > 60 cm), STRUCTURE analyses with and without population priors, and Geneland (See main text for details). ... 79

Table 4. Genetic diversity and HWE deviation test results for the populations defined and categories of trunks perimeter, computed using the data set either with (17) or without (11) markers under Hardy-Weinberg equilibrium. ... 80

Table 5. Pairwise F_{ST} for all pair of a priori populations, computed using the data set either with (17) or without (11) markers under Hardy-Weinberg equilibrium. ... 81

Chapter 3: *Micromeria* evolution in the Canary Islands

Paper I

Table 1. Genetic variation statistics per regions and species of *Micromeria*. This table contains information regarding number of populations (Pops.); average number of individuals (Ind.); average number of alleles (N), observed (H_O) and expected heterozygosity (H_E); portion of private alleles (Priv. Al.), and total heterozygosity (H_T). ... 105

Table 2. AMOVA analyses of four groupings calculated using RST. The results presented in a percentage form correspond to the amount of variation explained by differences within and among groups ... 107

Table 3. Contemporary migration rates calculated by BAYESASS between species of *Micromeria*. Results are presented in percentage (%) of individuals from species on top originated from the species in the left; values for migration within taxa are highlighted in gray; values above 10% are presented in bold; standard deviation values are in parentheses.. 112

Table 4. Historical migration rates calculated by MIGRATE between species of Micromeria. Results are presented in average number of individuals per generation; values correspond to migration rates from the species in the left to the species on top; values for migration within taxa are highlighted in gray; values above 10 migrants per generation are presented in bold; values for 95% confidence intervals are in parentheses 113

Paper II

Table 1. Summary of NewHybrid analyses showing the number of individuals with assignment above 50% of being F1 and F2 hybrids, and backcrosses with either species one or two. 140

Table 2. Results of variance analyses of island age over genetic diversity for three datasets: whole archipelago, M. hyssopifolia and taxa belonging to the M. varia complex; and taxa belonging to the western lineage (Tenerife, La Palma and El Hierro and *M. pedro-luisii* from La Gomera). In the Tests column the letters correspond to the tests classes used. I: Islands; S: Species; O/N: old/new. * corresponds to p-values < 0.001. 142

Table 3. Results of variance analyses of island age over genetic differentiation for three datasets: all archipelago, *M. hyssopifolia* and taxa belonging to the *M. varia* complex; and taxa belonging to the western lineage (Tenerife, La Palma and El Hierro and *M. pedro-luisii* from La Gomera). I: Islands; S: Species; O/N: old/new. * corresponds to p-values < 0.001. 143

Chapter 4: Distribution of adaptive traits during adaptive radiation.

Paper I

Table 1. Alignment information content from the two missing data extractions including: number of loci used in the alignment; number of total, variable and informative characters, and portion of branches with a posterior probability above 0.7. 168

List of figures

Chapter 1: General Introduction

Fig. 1. Representation of island ontology and how species richness and speciation is affected by this process according to the GDM model from Whittaker et al. (2007, 2008). 5

Fig. 2. Schematic representation of the “surfing syngameon” hypothesis proposed by Caujapé-Castells (2011). Arrows correspond to likely colonization events and the symbols to different alleles of the same locus. Islands closer to the mainland receive multiple colonization events preventing the loss of genetic diversity due to the founder effect. These are integrated in the island’s genepool through hybridization and speciation results in species syngameon. Colonization of the remaining islands is made from the syngameon. Because colonization from the mainland is unlikely, the founder effect contributes to a loss of genetic diversity, and together with selection more genetic differences are established. 7

Fig. 3. Map of Macaronesia region. The bottom right map represents the seven Canary Islands with their names and ages in millions of years (according to Juan et al. 2000). Picture was taken from Puppo (2015). 8

Fig. 4. *Micromeria* species distribution and pictures based on Pérez de Paz (1978) and Puppo and Meimberg (2015a, 2015b). *Micromeria* photos were taken by Pamela Puppo, Harald Meimberg and Pedro Luis Perez-de-Paz. The photo from *M. mahanensis* was taken from http://www.floradecanarias.com 11

Fig. 5. Phylogeny of *Micromeria* genus in the Canary Islands and Madeira based on the work from Puppo et al. (2015a). The tree is not a true phylogenetic analysis, but rather a representation of the conclusions drawn in their publication. The terminals do not correspond to real individuals, and branch lengths to not correspond to real degree of differentiation among species. 13

Chapter 2: Molecular markers development

Paper I

Fig. 1. Schematic summary of marker development strategy used to develop EPIC markers for Micromeria using publicly available sequence information of DNA or RNA from different genera of the Nepetoideae (Lamiaceae). 35

Fig. 2. Phylogenetic tree calculated using a combined matrix of all 19 loci retained. Maximum likelihood and maximum parsimony trees showed congruent results, therefore only the maximum likelihood is shown. 43

Paper II

Fig. 1. a) Relative frequency distribution of sequence length for all 60,678 sequences compared to SSR containing sequences retained in the Sciroko fraction and sequences containing only perfect SSRs that are at least six repeats long. The frequency for each size category was calculated by calculating the percentage of sequences in with that particular in relation to the total number of sequences of that fraction. b) Comparison of average PIC and He for dinucleotide SSRs, TTA/AAT SSRs and all other trinucleotide SSRs. Error bars indicate standard deviation. Both values for non-TTA trinucleotides are significantly lower than for the TTA ones in the t-test (alpha = 0.05). 57

Fig. 2. STRUCTURE analysis of Ethiopian and Kenyan Catha edulis samples alternately assuming two (K = 2) or seven (K = 7) clusters. Asterisks (*) corresponds to cultivated or feral samples that were not directly transplanted from wild populations. 60

Paper III

Fig. 1. Map of the study area, showing sampling localities and the area to be flooded by the Sabor dam. The background gray pallet corresponds to the altitudinal gradient. The rectangles correspond to the populations defined a priori. 72

Fig. 2 Plots from the cluster assignment probability obtained in Geneland analyses for a maximum number of three groups. The plots were sorted according three different criteria: a) *a priori* populations; b) trunks perimeter and *a priori* populations; c) separate analyses of trees below 60 cm and trees above 60 cm. 78

Chapter 3: *Micromeria* evolution in the Canary Islands

Paper I

Fig. 1. Maps of Tenerife showing: (A) *Micromeria* sampling localities; long-dashed lines indicate remnants of paleo-islands; short-dashed lines indicate major valleys; dotted polygons indicate regions formed by geographically close populations (see Table 1); symbol shapes and colors correspond to different species of *Micromeria*; numbers on symbols indicate collection localities (see Table 1); (B) distribution of *Micromeria* species. Species distributions were obtained by converting a point per quadrant dataset from Pérez de Paz (1978) into continuous ranges. Individuals of *M. varia* on the central north coast had been assigned to *M. hyssopifolia* in the meanwhile (Puppo et al. 2014); (C) genetic diversity for each species calculated as H_E (upper right), H_O (middle right), and H_T (bottom right). The boxplots showing H_E and H_O were made from single values estimated per population. 98

Fig. 2. UPGMA of pairwise unbiased uNei distances and R_{ST} among population groups of *Micromeria* according to island regions; regions are those showed in Fig 1 and Table 1. 106

Fig. 3. Principal Coordinates Analyses (PCoA) of pairwise distances of individuals of *Micromeria* implemented in GeneAlEx for codominant datasets. Shown are the first two coordinates of analyses including: A. all species; B. only paleo-island species; C. only central area species; D. only *M. lasiophylla* and *M. lachnophylla*; E. only *M. varia* divided in samples from Anaga (A) and Teno (T); F. only *M. hyssopifolia* divided in samples from Adeje (Old) and the remaining samples. 107

Fig. 4. STRUCTURE analyses of the species of *Micromeria* present in Tenerife showing blots of assignment probability from K values ranging from K = 2 to K = 9; optima K according to the Evanno method are indicated in red: K = 3 for all runs and K = 9 when only K = 4 to K = 15 are analyzed. 108

Fig. 5. Suboptimum K (K = 5) for the analyses including only central area species and assignment probability plotted per population on the map. The structure plot is shown to provide a context for the colors shown in the map. 109

Fig. 6. STRUCTURE analyses of potential hybrid individuals of *Micromeria* between four species pairs*: M. densiflora* and *M. varia*; *M. rivas-martinezii* and *M. varia; M. teneriffae* and *M. hyssopifolia* from the South; and *M. teneriffae* and *M. varia* from Anaga. Only the optimal K values according to the Evanno method are presented. The individuals marked with * were identified as morphological hybrids. 111

Paper II

Fig. 1. Map of the Canary Islands showing sampled localities and recognized species per island and group (East/West) to which they belong. 132

Fig. 2. Principal coordinates analyses and uNeiD UPGMA dendrogram including all localities with at least four individuals. For the PCoA, three analyses were performed: all populations; populations be-longing to the group composed by the eastern islands; and the ones belonging to the group composed by the western islands and *M. lepida* from La Gomera. The vertical and horizontal axis explained respectively: 33.1% and 8.3% of the variation for the all island analyses, 31.9% and 10.4% for the Gran Canaria analyses, and 17.6% and 13.9% for the western group analyses. Islands are represented by the following abbreviations: La (Lanzarote), GC (Gran Canaria), LG (La Gomera); LP (La Palma); EH (El Hierro). 137

Fig. 3. Bayesian clustering analyses with STRUCTURE. A: Analyses including all samples; best value of K was K=13, second best was K=19. B: Analyses including only the western and eastern lineages; for both the optimum K value was K=2; the

suboptimum was K=11 and K=17 for the eastern and western lineages, respectively. 138

Fig. 4. Representation of recent and historical migration rates calculated with BayesAss and migrate, respectively. All arrows correspond to migration rates with 95% confidence intervals above 0.001 for BayesAss and 0 for Migrate. Arrow thickness is proportional to the migration rate. Panels A and B correspond to contemporary migration rates calculated between islands using species and islands as groups, respectively. Panel C shows historical migration rates between islands groups calculated with Migrate. These were the only significant values obtained from this program. In Panel D, contemporary migration rates between species within islands are shown.. 139

Fig. 5. Representation of expected heterozygosity (HE) and FST per population used as proxy of genetic diversity and differentiation. For each measure there is a graph per island, per species, and per age class. Here only the test including all samples is shown. Graphs for the other tests are shown in supplementary material. 144

Chapter 4: Distribution of adaptive traits during adaptive radiation.

Paper I

Fig. 1. Up: Map of the Canary Islands showing the distribution of the species of *Micromeria* present in the Canary Islands. Down: species classification according to both Pérez de Paz 1978 (1978) and Puppo and Meimberg (2015 a, 2015b) indicating to which lineage (east or west) each species is assigned to according to molecular studies. Species showing the *M. varia* phenotype are indicated in bold. 161

Fig. 2. Schematic summary of the RAD-sequencing library preparation method developed applied in the current study. This was divided into four steps, further details of which can be found in the text beside each figure section. 165

Fig. 3. Mr.Bayes tree calculated with the dataset including only loci with a maximum missing data of 90%. Bars in the nodes correspond to posterior probability values >90. 169

Chapter 5: General discussion

Fig. 1. Schematic summary of how the surfing syngameon hypothesis can be applied to single-island colonization based on the data presented in this thesis. Unless indicated otherwise, dashed circles represent species. Symbols represent alleles for a specific locus. A: Multiple colonization on a newly emerged island resulting in the creation of a hybrid swarm. Arrows represent colonization events. B: The hybrid swarm differentiates into two species through an ecological speciation process. Alleles are

sorted according to local selective pressures. C: Connectivity through gene-flow of the species syngameon resulted from the multiple colonization process. Species occurring on younger islands are more connected than the ones occurring in older ones. This process results in an expansion of the syngameon in the same direction of the addition of new land. 186

Fig. 2. Representation of how a single genetic origin of a phenotype is congruent with a convergent evolution scenario due to the multiple colonization dynamics. Dashed circles represent different species. Green dashed circles show the convergent phenotype (ex. *M. varia* morphology). Arrows represent colonization events and green arrows correspond to the establishment of a colonizer containing target phenotype. 191

Fig. 3. Distribution and pictures of *M. herpyllomorpha* from La Palma (blue) *and M. lasiophylla* (yellow).from Tenerife. The regions surrounded by the black line correspond to the high-altitude populations. Species distributions are based in Perez-de-Paz (1976). Photos were taken by Pamela Puppo. 192

Chapter 1: General Introduction

1.1. Adaptive radiation and hybrid speciation

Adaptive radiation events are diversification processes responsible for much of the biological diversity found today (Rundell and Price 2009). Some of the main examples of this evolutionary process are the Darwin Finches in the Galapagos (Grant and Grant 2002), the *Anolis* lizards in the Caribbean (Losos et al. 1998), and the cichlids from the African lakes (Seehausen 2006). This process occurs when an ancestral lineage comes into contact with a heterogeneous habitat and new species are created. Adaptive radiation is driven by divergent selection triggered by local adaptation, and results in an extremely rapid process of ecological speciation. Ecological speciation is the process in which populations occupying divergent niches differentiate from one another, becoming gradually reproductively isolated until they are considered different species (Schluter 2001).

From the classical view, the main factors triggering adaptive radiation events are the presence of high amount of free niches (either by the appearance of new resources or the extinction of species previously occupying them), heterogeneous environments, and high competition (Losos 2010). Competition forces individuals to occupy new niches and adapt to new ecological conditions. During this process populations are subjected to different selective pressures resulting in rapid ecological speciation events. Descriptions of adaptive radiation mechanics are mostly based in inferences constructed from empirical data. Experimental work using microorganisms as a model confirmed the expectations formulated based on empirical data. Using *Pseudomonas fluorescens* as a model, Meyer and Kassen (2007) showed that competition and predation triggers radiation. Also using *P. fluorescens*, Rainey and Travisano (1998), and Brockhurs et al. (2007) showed that the availability of heterogeneous conditions and empty niches are limiting factors for adaptive radiation events.

During biological invasions, multiple introductions contribute to the maintenance of high genetic diversity and promote local adaptation (Kolbe et al. 2004). The same is expected to happen during adaptive radiation events. High genetic diversity facilitates the adaptation of populations to new habitats promoting their differentiation (Arnold 2006, 2015). With hybridization, new combinations of genotypes appear, which theoretically would expand populations' ability to occupy new niches (Dyer and Rice 1999, Meimberg et al. 2009). This was the reasoning behind the creation of the hybrid swarm hypothesis from Seehausen (2004), where divergent lineages are combined in hybrid swarms which result in adaptive radiation events due to their high adaptive ability. Again, this theory has been supported by biological invasion studies. By studying the invasion process of the three spine stickleback species in Switzerland, Roy et al. (2015) found that the

creation of hybrid swarms contributed to the adaptation and diversification of this organism into the new habitat. Seehausen's theory assumes that hybridization is common and that it plays a major role in the evolutionary process, which was previously not thought to be the case. The description of several examples of speciation with gene-flow support this idea (i.e. Peters et al. 2007; Alves et al. 2008; Reid et al. 2012).

The genetic consequences of introgression have been studied in detail by focusing on hybrid zones dynamics (i.e. Teeter et al. 2010; Nosil et al. 2012; Larson et al. 2014). These studies showed that some regions of the genome were more permeable to genetic-exchange than others. This is a consequence of strong, stabilizing selection, which maintains genotypes responsible for local adaptation. If this dynamic occurs, genomic regions being selected by local conditions will be highly differentiated between the two parent species, while neutral-behaving regions will be homogeneous. Over longer time scales, with multiple introgression events, genomes become mosaics where different regions show different evolutionary patterns (Linder and Reiseberg 2004). Lineage sorting also produces a similar incongruence pattern, but because it is random the resulting phylogenic signals cancel each other out (Durand et al. 2011). By excluding lineage sorting signals the remaining incongruence should be caused by introgression (Eaton and Rice 2013). This assumption is widely used to estimate the amount and direction of introgression (Jeffroy et al. 2006; Martin et al. 2013). In the present study I evaluated the effects of hybridization during adaptive radiation events.

1.2. Islands as models to study evolution

Oceanic islands are usually created by volcanic activity, formed over oceanic plates that have never been connected to a continent (Whittaker and Fernandez-Palacios 2007). These systems are considered to be natural laboratories to study evolution (Losos and Ricklefs 2009). They are generally small with well-defined boundaries and simplified biota, facilitating the interpretation of evolutionary patterns. Moreover, we know that all biota arrived there through colonization, having had to adapt to new conditions. The fact that oceanic islands have a volcanic origin and are relatively young makes them particularly interesting for geological studies, and a lot of information on that matter is available (i.e. Carlquist 1980; Holcomb 1981; Ancochea et al. 1990; Carracedo et al. 2002). By incorporating information such as island age, and frequency and time of geological events it is possible to have a better perspective of time and tempo of the evolutionary processes. Islands can work as replicates of themselves where

similar process can be studied in parallel on multiple archipelagos and islands (Losos and Ricklefs 2009).

Many factors influence evolutionary patterns on oceanic islands. The main one are geomorphological dynamics associated with volcanic activity and erosion (Fernández-Palacios et al. 2011). Volcanic activity contributes to the emergence of islands and increases their altitudinal gradient. When it stops, erosion is the main factor shaping island morphology (Figure 1). Initially this process increases topographical complexity, but ultimately islands become flat and sink. There are several models relating these dynamics with speciation rate and species richness (Whittaker and Fernandez-Palacios 2007; Whittaker et al. 2008; Chen and He 2009; Rosindell and Phillimore 2011). The most famous is the general dynamic model of island biogeography (GDM), where speciation rate is dependent on an island's carrying capacity, which is in turn dependent on island area and topographic complexity (Whittaker and Fernandez-Palacios 2007; Whittaker et al. 2008). According to GDM, speciation rate increases with an island's altitudinal gradient and it reaches its peak concurrent with the island's maximum topological complexity, which results in the increase of species richness. When islands start to sink, extinction overcomes the speciation rate and leads to a decrease in species richness. This model was later supported by empirical data relating endemism with topographic data in the Canary Islands (Otto et al. 2015). Additional models were proposed, where the main difference consisted of the weight given to immigration rate (Chen and He 2009; Rosindell and Phillimore 2011). Chen and He (2009) suggested that speciation rate does not increase as quickly alongside island carrying capacity if the immigration rate from the mainland is still high, because gene-flow will prevent differentiation and consequently speciation. This provides evidence that geology is not the only factor shaping evolution on islands.

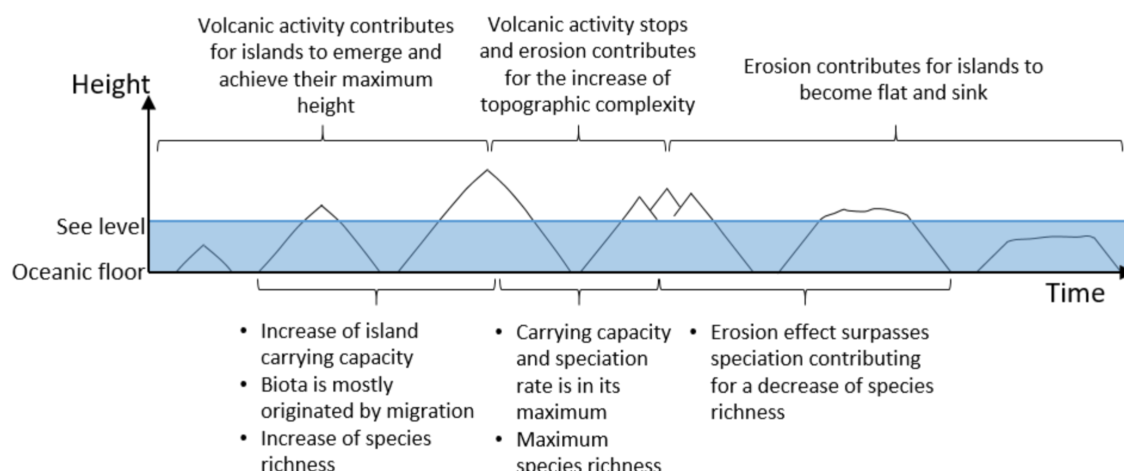


Fig. 1. Representation of island ontology and how species richness and speciation is affected by this process according to the GDM model from Whittaker et al. (2007, 2008).

Their isolation from the mainland and the high availability of free niches make oceanic islands particularly prone to adaptive radiation events (Heaney 2000; Whittaker and Fernandez-Palacios 2007; Losos and Ricklefs 2009). These same factors may contribute to a high prevalence of hybridization (Francisco-Ortega et al. 1996), which, according to Seehausen's hybrid swarm theory, promotes adaptive radiation events. Herben et al. (2005) and Saunders and Gibson (2005) proposed that a process of multiple colonization events followed by hybridization could be a frequent occurrence in archipelagos close to the mainland and could promote species adaptation. Nevertheless, until recently, colonization on islands was thought to occur through single introduction events. According to Silvertown (2004) and Silvertown et al. (2005), after the first colonization the resulting individuals would occupy their respective niche, out-competing all subsequent colonizers.

When compared with their mainland relatives, Canarian flora is genetically more diverse than expected (Francisco-Ortega et al. 2000; García-Verdugo et al. 2015). This is an indication that the founder effect is not very pronounced, which can only be explained by multiple colonization events (García-Verdugo et al. 2015). This finding resulted in the creation of the surfing syngameon hypothesis by Caujapé-Castells (2011). He suggested that islands work as allelic sinks from the mainland, where individuals from multiple sources are combined in hybrid swarms (Fig. 2). Species resulting from these highly diverse populations are connected through gene-flow and evolve as unit (Pérez de Paz and Caujapé-Castells 2013), which is the definition of a syngameon (Grant 1981). According to this hypothesis, islands closer to the mainland constantly experience colonization from the mainland, maintaining high levels of diversity and low levels of differentiation (Fig. 2). The other islands are colonized mainly by the syngameon, resulting in the loss of genetic diversity due to the founder effect. According to Caujapé-Castells (2011), this will allow for genetic differences to be accumulated in a process similar to allele surfing from an expanding population (Excoffier and Ray 2008). Thus the surfing syngameon hypothesis predicts that different evolutionary processes occur in parallel in one archipelago, such that on some islands the accumulation of genetic diversity is more prevalent, while on others differentiation events are more dominant. These expectations have been met by Pérez de Paz and Caujapé-Castells (2013) and García-Verdugo et al. (2015) where, in the Canary Islands, genetic diversity was positively correlated with distance to the mainland. This thesis will focus on genetic diversity and differentiation patterns of a group of taxa from the Canary Islands.

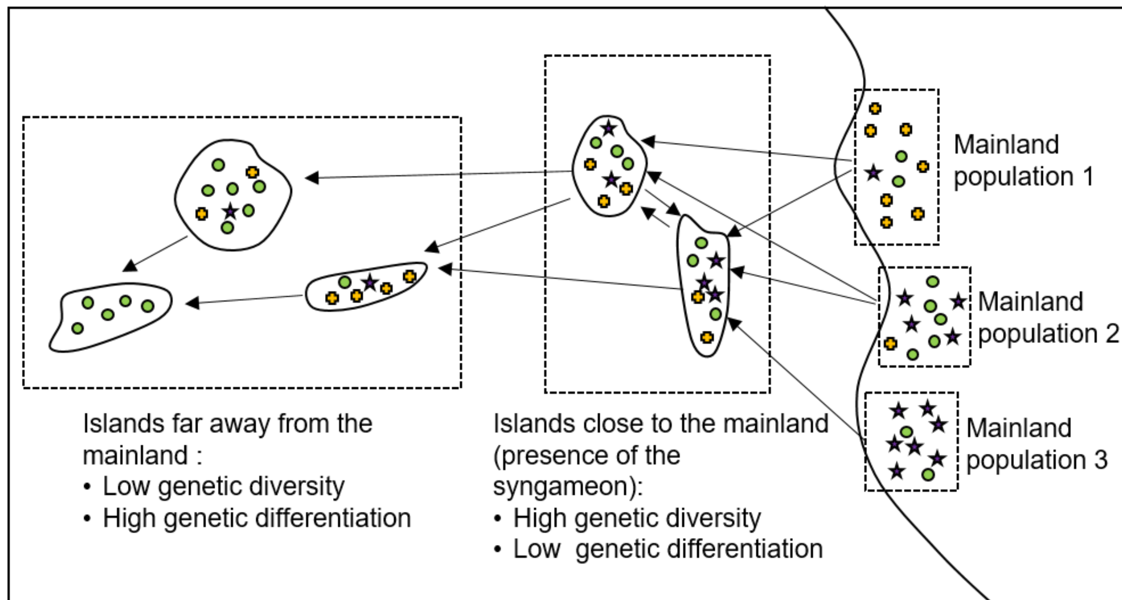


Fig. 2. Schematic representation of the "surfing syngameon" hypothesis proposed by Caujapé-Castells (2011). Arrows correspond to likely colonization events and the symbols to different alleles of the same locus. Islands closer to the mainland receive multiple colonization events preventing the loss of genetic diversity due to the founder effect. These are integrated in the island's genepool through hybridization and speciation results in species syngameon. Colonization of the remaining islands is made from the syngameon. Because colonization from the mainland is unlikely, the founder effect contributes to a loss of genetic diversity, and together with selection more genetic differences are established.

1.3. Study system

Canary Islands

The Canary Islands archipelago comprises seven islands located between 100 Km to 490 Km away from the West Saharan coast. The islands were created by a hotspot movement from east to west resulting in an age gradient in the same direction (Fig. 3). The eastern most islands Lanzarote and Fuerteventura, formerly a single island, are 15.5 Ma and 20 Ma, respectively (Fernández-Palacios et al. 2011). Gran Canaria, the third oldest island with around 15 Ma, was formed by a subsequent addition of land, resulting in two parts with different ages (del-Arco et al. 2002). The southwestern half of Gran Canaria is mostly composed of Miocene substrates (Palaeo-canaria) while the northeastern part of Pliocene rocks (Neo-canaria). Tenerife is also composed of subtracts with different ages, but they resulted from different processes. In the late Miocene there were three islands in that region (Ancochea et al. 1990): Adeje (11.6-3.5 Ma), Teno (6.7-4.5 Ma), and Anaga (6.5-3.5 Ma). These were later connected by volcanic eruptions occurring in the late Miocene-Pliocene (around 2 Ma), producing Tenerife's current shape. As a result, Tenerife is composed of substrates from different ages and origins. The remaining islands do not have such complex geological histories, and vary in age from 10 Ma (La Gomera) to 2 Ma and 1 Ma (La Palma and El Hierro, respectively).

The Canary Islands are in different stages of the GDM (Whittaker et al. 2008, Fernández-Palacios et al. 2011). Lanzarote and Fuerteventura are highly eroded and are almost in the sinking stage. Volcanic activity has ceased on Gran Canaria and La Gomera, so erosion processes are the major factors shaping these islands. Tenerife already achieved its maximum altitudinal range, whereas the other islands are still increasing their altitudinal range.

The Canary Islands are part of the phytobiogeographic region of Macaronesia, which is known for high levels of endemism. In fact, just in the Canary Islands, a total of 3857 endemic terrestrial fungi, plants and animal species are described (Arechavaleta et al. 2009). These high levels of endemism are related to the high diversity of ecological zones created by the large altitudinal range found in these islands (Juan 2000). These ecological zones vary in their composition and location depending on the island slope. The northern slopes are wetter because they receive moist air from the trade winds. The following ecological zones are found (Juan 2000): coastal desert; arid to semi-arid shrubland, laurel forest, pine forest, and subalpine desert. The southern slopes are drier and the vegetation belt limit extends to around 300 m higher in altitude, where it is too dry for laurel forest to establish (Juan 2000).

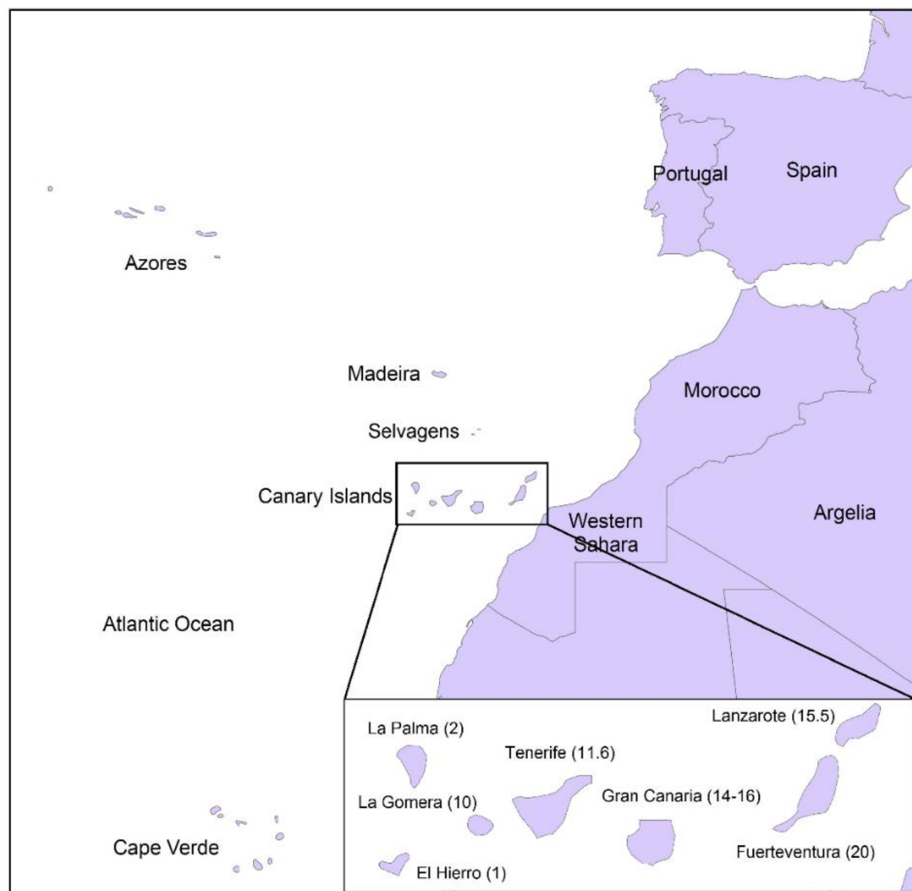


Fig. 3. Map of Macaronesia region. The bottom right map represents the seven Canary Islands with their names and ages in millions of years (according to Juan et al. 2000). Picture was taken from Puppo (2015).

Flora in the Canary Islands has been described to be mostly of western Mediterranean origin (Juan et al. 2000; Carine et al. 2004). However, African and America origins have been reported for several genera (i.e. Olmstead and Palmer 1997; Juan et al. 2000; Galbany-Casals et al. 2009; Jaén-Molina et al. 2009).

The fact that the Canary Islands have a well-known geological history makes them an ideal system to test evolutionary biology hypothesis, as it is possible to relate species diversification patterns with the main geological events. Moreover, the fact that we know when these events occurred gives us a better idea of their diversification time and tempo. As a result there are many studies focusing on the evolution of both animal and plant taxa in the Canary Islands (i.e. Francisco-Ortega et al. 2001, Jorgensen and Olesen 2001). Most of the phylogenetic studies done in the archipelago show a monophyletic origin, which is unexpected given its proximity to the mainland. However, as mentioned in the last section, this may be due to hybridization and the low amount of molecular markers used per study, making it impossible for multiple origins to be detected (Herben et al. 2005; Caujapé-Castells 2011). Inter-island colonization has been described as following mostly a stepping stone model (Sanmartin et al. 2008), which also may not be the case when an island is occupied by lineages from multiple sources that hybridize after establishment. In the present work, some of these questions are discussed using the *Micromeria* genus in the Canary Islands as a model.

***Micromeria* species from the Canary Islands**

Micromeria Benth. is a genus of around 52 species distributed throughout the Mediterranean basin and Macaronesia, and in some parts of Africa and Asia (Bräuchler et al. 2008). It belongs to the family Lamiaceae and it is composed of shrubs, subshrubs and herbs with monoecious flowers pollinized by insects. Fruits are mostly dispersed by wind, but ants and water can also contribute to this process. Macaronesia presents a hotspot of biodiversity for this genus, with 23 currently described species all of them being endemic to the archipelagos of Canary Islands, Madeira and Cape Verde (Perez de Paz 1978; Bräuchler et al. 2008; Puppo and Meimberg 2015a, 2015b). *Micromeria forbesii,* the only species described in Cape Verde, has an origin independent from the remaining Macaronesia taxa (Bräuchler et al. 2005), whereas the remaining species on the Canary Islands and Madeira have been reported to be monophyletic (Bräuchler et al. 2005; Meimberg et al. 2006; Puppo et al. 2015a).

On the Canary Islands, *Micromeria* was initially described in 16 species (Perez de Paz 1978) using only morphological characters (Table 1). This classification included three species found on more than one island: *M. varia* (all with exception of La Palma);

M. hyssopifolia (Tenerife and El Hierro); and *M. lasiophylla* (Tenerife and La Palma). The remaining species were single-island endemics. With the inclusion of molecular approaches, it was verified that these widespread species corresponded to multiple lineages (Meimberg et al. 2006; Puppo et al. 2015). Based on this observation, Puppo and Meimberg (2015a, 2015b) reclassified this genus, resulting in the following changes (Table 1): *M. varia* was divided into several single-island endemics (with the exception of Lanzarote and Fuerteventura, where the same species is found); *M. lasiophylla* from La Palma was considered to be *M. herpyllomorpha*; and *M. varia* and *M. hyssopifolia* from El Hierro were classified into a new species (*M. hierrensis*).

Table 1. List of Micromeria species described in the Canary Islands according to their current and previous classifications. Reclassified species are in bold.

Island	Current classification (Puppo & Meimberg 2015a, 2015b)	Previous classification (Pérez de Paz 1978)
Lanzarote	***M. mahanensis***	***M. varia rupestris***
Gran Canaria	*M. pineolens*	*M. pineolens*
	M. leucantha	*M. leucantha*
	M. helianthemifolia	*M. helianthemifolia*
	M. benthamii	*M. benthamii*
	M. lanata	*M. lanata*
	M. tenuis	*M. tenuis*
	M. canariensis	***M. varia canariensis/meridialis***
La Gomera	***M. gomerensis***	***M. varia gomerensis***
	M. lepida	*M. lepida*
	M. pedro-luisii	***M. varia varia***
Tenerife	*M. rivas-martinezii*	*M. rivas-martinezii*
	M. glomerata	*M. glomerata*
	M. teneriffae	*M. teneriffae*
	M. lasiophylla	*M. lasiophylla*
	M. lachnophylla	*M. lachnophylla*
	M. hyssopifolia	*M. hyssopifolia*
	M. varia	***M. varia***
	M. densiflora	*M. densiflora*
La Palma	***M. herpyllomorpha***	***M. herpyllomorpha / M. lasiophylla***
El Hierro	***M. hierrensis***	***M. varia hierrensis / M. hyssopifolia***

Micromeria species distribution patterns vary from island to island (Fig. 4). *Micromeria mahanensis* has a patchy distribution in both Lanzarote and Fuerteventura. El Hierro and La Palma also have only one species, but these are evenly distributed throughout the island. La Gomera has three species with highly overlapping distribution ranges. In fact, there are populations where the three species grow together, and

morphological intermediates of *M. lepida* and *M. pedro-luisii* are found. Gran Canaria and Tenerife have the highest number of species with eight and seven, respectively.

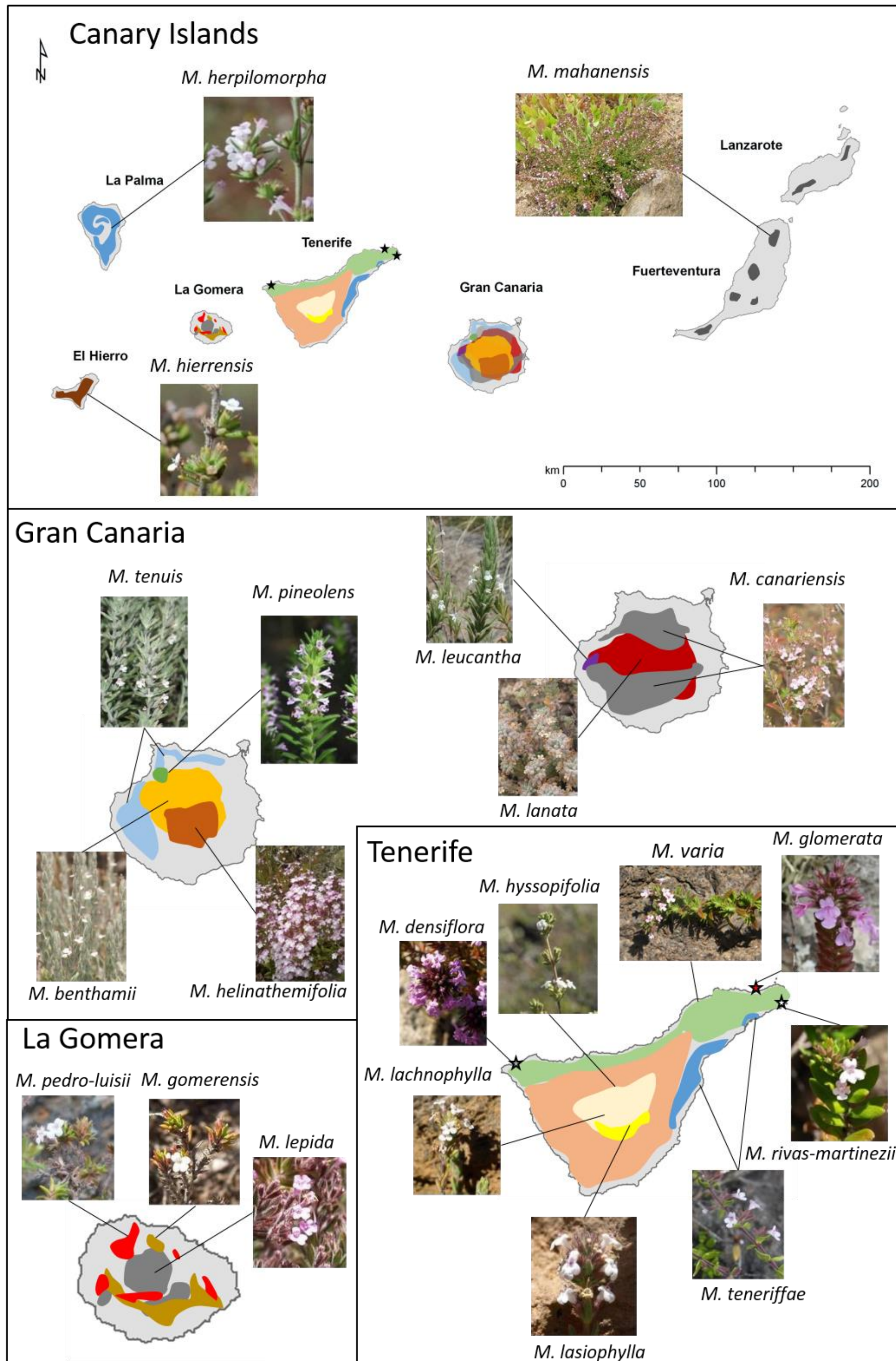


Fig. 4. *Micromeria* species distribution and pictures based on Pérez de Paz (1978) and Puppo and Meimberg (2015a, 2015b). *Micromeria* photos were taken by Pamela Puppo, Harald Meimberg and Pedro Luis Perez-de-Paz. The photo from *M. mahanensis* was taken from http://www.floradecanarias.com

In Tenerife, species have a mostly parapatric distribution with some degree of overlap (Fig. 4) where morphological hybrids can be found. On this island, there are three narrowly distributed species restricted to the older part of the islands: *M. densiflora* restricted to Teno, and *M. glomerata* and *M. rivas-martinezii* restricted to Anaga. The following species were included in the Spanish IUCN red list: *M. rivas-martinezii* and *M. glomerata* as critically endangered, and *M. densiflora* as endangered. *Micromeria teneriffae* is distributed in fragmented patches from Anaga to the southern coast of the island. The distribution of the remaining species seems to be related to habitat composition. *Micromeria varia* is found in the Northern coast from Anaga to Teno. *Micromeria hyssopifolia* occupies most of the central part of the island, spanning habitats from coastal desert to laurel forest. *Micromeria lachnophylla* and *M. lasiophylla* replace *M. hyssopifolia* at high altitudes, the former occupying pine forest up to high desert, the latter being found in the Las Cañadas cliffs at the skirts of the Teide volcano.

Species in Gran Canaria overlap in most of their distributions, and many morphological intermediate forms are found where populations from different species grow together (Fig. 4). Two species, *M. leucantha and M. pineolens*, have a highly restricted distribution. They are morphologically very different from the other species and are included in the IUCN red list as endangered.

Micromeria species in the Canary Islands seem to have resulted from an adaptive radiation event, given that the distribution of most species is correlated with ecological zone composition (Meimberg et al. 2006). However, many of the species distribution seems to have been influenced by geological events as well (Puppo et al. 2014, 2015).

Micromeria is an ideal system for investigating the evolutionary biology of biota on the Canary Islands. First, it occupies all islands allowing archipelago-wide inferences to be made. Second, it occupies all ecological zones, therefore showing a wide range of adaptations that cannot be studied in a more restricted genera. Third, publications focusing on the evolution of *Micromeria* in the Canary Islands using molecular data have been available since 2006 (Meimberg et al. 2006), providing a theoretical framework and a wide variety of available molecular resources.

***Micromeria* phylogenetic patterns and evolutionary predictions**

According to sequence data from both nuclear and plastid genes, *Micromeria* is divided into two lineages (Meimberg et al. 2006; Puppo et al. 2015a; Fig. 5): an eastern lineage composed of the taxa found on the islands of Lanzarote, Fuerteventura, and Gran Canaria; and a western lineage composed of taxa found on Tenerife, La Palma and El Hierro. Species from both lineages are found on La Gomera, *M. lepida* and *M. gomerensis* from the eastern lineage, and *M. pedro-luisii* from the western lineage.

Within these lineages, Tenerife and Gran Canaria are paraphyletic relative to the other islands, indicating that they were probably the main source of colonization (Puppo et al. 2015a). The closest related species from the mainland has not been found yet, thus a close outgroup was not included in phylogenetic analyses so far. For that reason the directionality in this inference is not clear, and final conclusions regarding the starting point of *Micromeria* colonization have not yet been reached.

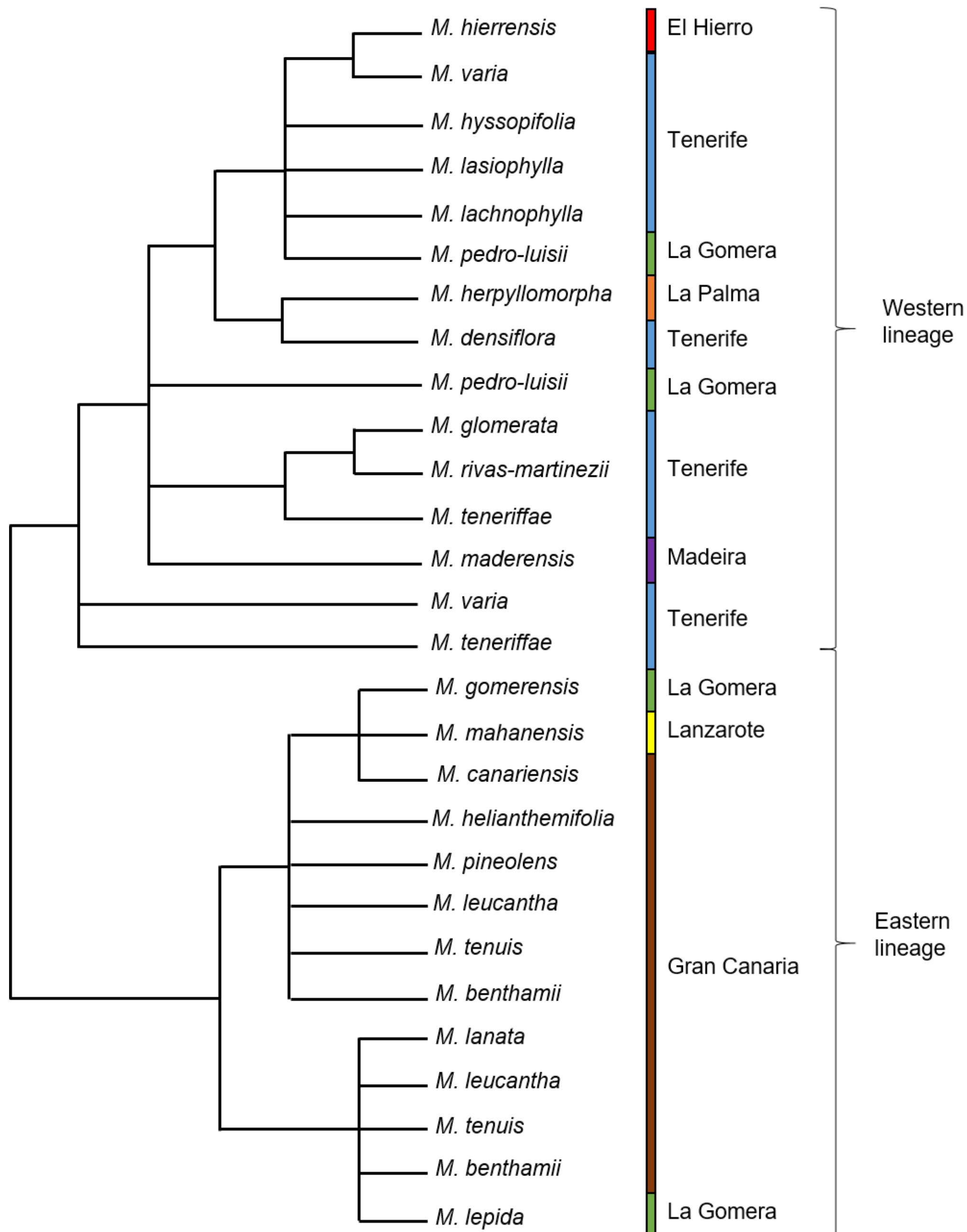


Fig. 5. Phylogeny of *Micromeria* genus in the Canary Islands and Madeira based on the work from Puppo et al. (2015a). The tree is not a true phylogenetic analysis, but rather a representation of the conclusions drawn in their publication. The terminals do not correspond to real individuals, and branch lengths to not correspond to real degree of differentiation among species.

In Puppo et al. (2015a), the species previously classified as *M. varia* from the eastern group seem to have a common origin (Fig. 5). That *M. lepida* has an independent origin from *M. gomerensis* indicates that La Gomera was colonized at least three times. This finding supports the hypothesis that multiple colonizations can occur. The remaining islands seem to have been colonized only once: El Hierro from central Tenerife and La Palma, although with low support, from the region of Teno. The molecular markers used by Puppo et al. (2015a) were not powerful enough to recover potential origins of *M. pedro-luisii.* According to the predictions from Herben et al. (2005) and Caujapé-Castells (2011), the apparent single origin found on most islands can be a consequence of introgression directly after colonization from multiple sources.

Puppo et al. (2014) found that the phylogeographic patterns in Tenerife were congruent with the island's geological history. The species restricted to the older parts of the island and *M. teneriffae* corresponded to older lineages (between 6.7 to 4.5 Ma), while the central species resulted from a more recent divergent event (around 4 Ma). This pattern is congruent with the colonization of the central part of Tenerife by species from the older regions. This could have resulted in a hybrid swarm that differentiated in the central species. Considering the predictions made by Caujapé-Castells (2011), the resulting species are most likely connected by a syngameon. Additionally, given their parapatric distribution of species in Tenerife, hybrid zone dynamics may be established in their contact zones.

Another consequence of the phylogenetic patterns recovered in Puppo et al. (2014, 2015a) is the independent origin of several morphological forms, *M. varia* s.l. being the most prominent. This phenotype seems to have appeared once in the eastern lineage and possibly several times in the western islands, raising several questions regarding the origin of this phenotype. Parallel or convergent evolution are likely scenarios, however if gene-flow plays a role in *Micromeria* diversification other hypotheses are plausible. This morphology may actually have appeared only once during early stages of *Micromeria* evolution, and it is undetectable because of high introgression with other lineages during *Micromeria* dispersal throughout the archipelago.

All of the predictions outlined above will be tested within this thesis. Prior to this thesis, all related work had been done with eight exon primed intron covered markers (EPIC), which are likely under selection and thus less permeable to gene-flow between species. In order to better test hypotheses involving hybridization neutral markers need to be used. In the case of *Micromeria*, this became possible by the development of 16 microsatellite markers by Puppo et al. (2015b). The implications of the uses of each set of markers will also be discussed in the present work.

1.4. Molecular markers: applications and development

The use of molecular markers have revolutionized studies of evolutionary biology (Avise 2012). With them it is possible to access genetic information directly at its source of variability. Among others, some of the advantages of using molecular markers compared to morphological data are: their ability to assess homology more accurately; their ability to access the genetic basis of phenotypic variation; their almost unlimited source of information, since all nucleotides combinations can potentially work as characters.

Several types of molecular markers were developed across time depending on the technology available (Schlötterer 2004). Among them it is possible to find markers accessing: protein information (allozymes); DNA variability without previous sequence knowledge (RAPD, ISSRs, and AFLPs); DNA variability requiring previous sequence knowledge (microsatellites, SNPs, and sequence of orthologous regions). The last group of markers became more popular because the analyzed genomic region is known, and due to the fact that they are able to recover codominant information. Their appropriate application also depends on the information being sought. For example, microsatellites are preferentially used for comparisons within the same species (Ellegren 2004), while sequencing markers for deeper taxonomic comparisons (Thomson et al. 2010). Additionally, because microsatellites are mostly neutral, they are widely used to access neutral processes such as demographic patterns (Selkoe and Toonen 2006). Sequencing markers and SNPs occurring in coding regions can be used to test the effect of selection on evolutionary patterns (Morin et al. 2004). In the present study we used microsatellites to evaluate structure and connectivity patterns of *Micromeria,* thus we give a particular emphases to these markers.

Marker development often requires previous knowledge of the genome sequencing. Accessing this information with Sanger technology is very expensive and labor intensive (i.e. Zane et al. 2002; Selkoe and Toonen 2006). With the appearance of second generation sequencing platforms, also known as next generation sequencing (NGS), some of these limitations were surpassed (Davey et al. 2011; Gardner et al. 2011). These technologies allowed the sequencing of millions of base pairs at a relatively low price, allowing the assessment of genomic sequence information that could be used for primer design. For microsatellite searches, the most common approach was the use of the 454 platform from both enriched and non-enriched libraries (i.e. Allentoft et al. 2009; Santana et al. 2009; Csencsics et al. 2010). In the present study I present examples of both approaches. Nevertheless, the genotyping of these markers was still

done with traditional methods, such as capillary electrophoresis, which presents a limitation to the amount of throughput produced.

Recently new genotyping approaches incorporating NGS were developed, allowing the screening of thousands of markers for non-model organisms. All of these require a previous reduction of genome complexity before sequencing (Cronn et al. 2012). The most common one uses restriction enzymes to sequence flanking regions associated to their recognition sites. This is called restriction associated DNA (RAD), and it was first presented by Baird et al. (2008). Since then several variations of this approach were developed, creating a new class of molecular markers. Alternative methods using NGS to sequence amplicons (Egan et al. 2012) or libraries previously enriched for specific genes (McCormack et al. 2013) were also developed, but they require some *a priori* knowledge regarding the genome sequence. With RADs this is not necessary, and both coding and non-coding regions are sequenced. In this thesis is presented a variation of the RAD protocol and its ability in recovering highly detailed phylogenetic patterns for *Micromeria.*

1.5. Objectives and thesis outline.

The main goal of this thesis is to evaluate several hypotheses regarding the effect of continuous gene-flow during the evolution of *Micromeria* on the Canary Islands. More specifically, I will:

I) test the influence of island geological history in *Micromeria* evolution;

II) evaluate the possibility of multiple colonization events during the dispersion of the genus throughout the Canary Islands;

III) investigate the role of gene-flow and hybridization in genetic diversity and diversification ability;

IV) study the relation between *Micromeria* ecological requirements and phylogenetic relationships;

V) focus on the development of molecular resources to test the hypotheses outlined in the previous objectives.

The work developed during my thesis is presented in six papers, four of them already published. These are divided in three chapters: Molecular marker development (Chapter 2); Evolutionary patterns of *Micromeria* in the Canary Islands (Chapter 3); and Distribution of adaptive traits during adaptive radiation (Chapter 4). The work presented in these chapters goes from marker development to their application.

Chapter 2 focuses on molecular marker development. It consists of three published papers in which molecular markers are developed for several taxa. In the first paper, several EPIC markers were developed to be used with *Micromeria* and other Lamiaceae. These were then used in the work from which the hypotheses tested in this thesis are based on (Puppo et al. 2014, 2015; Puppo 2015). The other two papers focus on microsatellite marker development for *Catha edulis* and *Juniperus oxycedrus*. The *Catha* paper was the first in a series of three microsatellite development papers addressing NGS shotgun sequencing data. The selection approach was described in detail in this publication, and it was extended in the *Juniperus* paper. The microsatellite markers for *Micromeria* were developed in the context of another PhD thesis (Puppo 2015), using the same approach presented in the *Juniperus* and *Catha* papers. By discussing the papers presented in this section it will be possible to better understand: 1) how molecular marker selection affects future inferences; 2) marker efficiency in recovering phylogenetic and phylogeographic patterns; 3) marker utility in the genomic era.

Chapter 3 is the core of this thesis and where most of the proposed objectives are addressed. It is composed of two publications, one of them previously published (Puppo et al. 2016). Here we tested our hypothesis of how island geological features and gene-flow influence the evolution of *Micromeria*. This was done in two different geographical scales: on Tenerife and on the entire archipelago. In this section the following questions are discussed: 1) are the phylogeographic patterns of *Micromeria* related to an island's geological features; 2) is multiple colonization an important process in the occupation of new islands; 3) does hybridization and gene-flow play a role in *Micromeria* evolution.

Chapter 4 focuses on the relation between ecological conditions with the phylogenetic patterns of *Micromeria*. The results are interpreted in a scenario where introgression can play a role in the distribution of adaptations. This is done by studying the phylogenetic relationships of the species previously classified as *M. varia*. This chapter aims to answer the following questions: 1) are there new *M. varia* s.l. lineages; 2) did this phenotype appear completely independently of the known lineages; 3) does multiple colonization and introgression play a role in the distribution of *Micromeria* ecological traits.

Chapter 5 is dedicated to the discussion of the main findings of the separate papers and the overall relationships between them. This begins with a discussion of the insights that this thesis contributed toward a better understanding of evolution on oceanic islands. The evidences of evolution with gene-flow and its impacts in *Micromeria* evolution are examined, as well as how this process can be influenced by an island's

geological history. Afterwards, the influence of genetic-exchange on morphological traits is addressed in two parts: first, in how species morphological integrity can be maintained despite hybridization; and second, how the mechanisms outlined in the surfing syngameon hypothesis can contribute to the appearance of morphologically identical species, such as the ones showing the *M. varia* phenotype. The final discussion point centers on molecular markers and the different biases and information associated with them. This opens up an opportunity to discuss new genomic approaches and whether it is still worthy to invest in classical genotyping strategies, as well as new research ideas that relate to the main themes of this thesis.

Through this comprehensive approach I hope to have covered most of the factors shaping the evolutionary patterns of *Micromeria* in the Canary Islands, from both methodological and genetic perspectives.

1.6. References

Allentoft, M. E., Schuster, S., Holdaway, R., Hale, M., McLay, E., Oskam, C. L., ... & Bunce, M. (2009). Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques*, *46*(3), 195-200.

Alves, P. C., Melo-Ferreira, J., Freitas, H., & Boursot, P. (2008). The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus Lepus. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363(1505), 2831-2839.

Ancochea, E., Fuster, J., Ibarrola, E., Cendrero, A., Coello, J., Hernan, F., ... & Jamond, C. (1990). Volcanic evolution of the island of Tenerife (Canary Islands) in the light of new K-Ar data. Journal of Volcanology and Geothermal Research, 44(3), 231-249.

Arechavaleta, M., Rodríguez, S., Zurita, N., & García, A. (2010). Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres. 2009. Gobierno de Canarias, 579.

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

Arnold, M. L. (2015). Divergence with genetic exchange. Oxford University Press.

Avise, J. C. (2012). Molecular markers, natural history and evolution. Springer Science & Business Media.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one, 3(10), e3376.

Bräuchler, C., Meimberg, H., Abele, T., & Heubl, G. (2005). Polyphyly of the genus Micromeria (Lamiaceae)—evidence from cpDNA sequence data. Taxon, 54(3), 639-650.

Bräuchler, C., Ryding, O., & Heubl, G. (2008). The genus Micromeria (Lamiaceae), a synoptical update. Willdenowia, 363-410.

Brockhurst, M. A., Colegrave, N., Hodgson, D. J., & Buckling, A. (2007). Niche occupation limits adaptive radiation in experimental microcosms. PLoS One, 2(2), e193.

Carine, M. A., Russell, S. J., Santos-Guerra, A., & Francisco-Ortega, J. (2004). Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in Convolvulus (Convolvulaceae). American Journal of Botany, 91(7), 1070-1085.

Carlquist, S. (1980). Hawaii: a natural history. geology, climate, native flora and fauna above the shoreline (Vol. 6). Honolulu: SB Printers, Inc. for Pacific Tropical Botanical Garden (xii), 468p.-illus., col. illus., maps.. En Icones, Maps. Geog.

Carracedo, J. C., Pérez Torrado, F. J., Ancochea, E., Meco, J., Hernán, F., Cubas, C. R., ... & Ahijado, A. (2002). Cenozoic volcanism II: the Canary islands. Geological Society of London.

Caujapé-Castells, J. (2011). Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. The biology of island floras, 284-324.

Chen, X. Y., & He, F. (2009). Speciation and endemism under the model of island biogeography. Ecology, 90(1), 39-45.

Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., & Udall, J. (2012). Targeted enrichment strategies for next-generation plant biology. American Journal of Botany, 99(2), 291-311.

Csencsics, D., Brodbeck, S., & Holderegger, R. (2010). Cost-effective, species-specific microsatellite development for the endangered Dwarf Bulrush (Typha minima) using next-generation sequencing technology. Journal of Heredity, esq069.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12(7), 499-510.

de Paz, J. P., & Caujapé-Castells, J. (2013). A review of the allozyme data set for the Canarian endemic flora: causes of the high genetic diversity levels and implications for conservation. Annals of botany, 111(6), 1059-1073.

de Paz, P. L. P. (1978). Revisión del género micromeria Bentham (lamiaceaestachyoideas) en la región mecaronésica.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Molecular biology and evolution, 28(8), 2239-2252.

Dyer, A. R., & Rice, K. J. (1999). Effects of competition on resource availability and growth of a California bunchgrass. Ecology, 80(8), 2697-2710.

Eaton, D. A., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Systematic Biology, 62(5), 689-706.

Egan, A. N., Schlueter, J., & Spooner, D. M. (2012). Applications of next-generation sequencing in plant biology. American journal of botany, 99(2), 175-185.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nature reviews genetics, 5(6), 435-445.

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. Trends in ecology & evolution, 23(7), 347-351.

Fernández-Palacios, J. M., de Nascimento, L., Otto, R., Delgado, J. D., García-del-Rey, E., Arévalo, J. R., & Whittaker, R. J. (2011). A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. Journal of Biogeography, 38(2), 226-246.

Francisco-Ortega, J., Jansen, R. K., & Santos-Guerra, A. (1996). Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. Proceedings of the National Academy of Sciences, 93(9), 4085-4090.

Francisco-Ortega, J., Santos-Guerra, A., Kim, S. C., & Crawford, D. J. (2000). Plant genetic diversity in the Canary Islands: a conservation perspective. American Journal of Botany, 87(7), 909-919.

Francisco-Ortega, J., Barber, J. C., Santos-Guerra, A., Febles-Hernández, R., & Jansen, R. K. (2001). Origin and evolution of the endemic genera of Gonosperminae (Asteraceae: Anthemideae) from the Canary Islands: evidence from nucleotide sequences of the internal transcribed spacers of the nuclear ribosomal DNA. American Journal of Botany, 88(1), 161-169.

Galbany-Casals, M., Garcia-Jacas, N., Sáez, L., Benedí, C., & Susanna, A. (2009). Phylogeny, biogeography, and character evolution in Mediterranean, Asiatic, and Macaronesian Helichrysum (Asteraceae, Gnaphalieae) inferred from nuclear phylogenetic analyses. International journal of plant sciences, 170(3), 365-380.

García-Verdugo, C., Sajeva, M., La Mantia, T., Harrouni, C., Msanda, F., & Caujapé-Castells, J. (2015). Do island plant populations really have lower genetic variation

than mainland populations? Effects of selection and distribution range on genetic diversity estimates. Molecular ecology, 24(4), 726-741.

Gardner, M. G., Fitch, A. J., Bertozzi, T., & Lowe, A. J. (2011). Rise of the machines–recommendations for ecologists when using next generation sequencing for microsatellite development. Molecular Ecology Resources, 11(6), 1093-1101.

Grant, P. R., & Grant, B. R. (2002). Adaptive radiation of Darwin's finches: Recent data help explain how this famous group of Galapagos birds evolved, although gaps in our understanding remain. American Scientist, 90(2), 130-139.

Grant, V. (1981). Plant speciation. New York: Columbia University Press xii, 563p.-illus., maps, chrom. nos.. En 2nd edition. Maps, Chromosome numbers. General (KR, 198300748).

Heaney, L. R. (2000). Dynamic disequilibrium: a long-term, large-scale perspective on the equilibrium model of island biogeography. Global Ecology and Biogeography, 9(1), 59-74.

Herben, T., Suda, J., & Munclinger, P. (2005). The ghost of hybridization past: niche pre-emption is not the only explanation of apparent monophyly in island endemics. Journal of Ecology, 93(3), 572-575.

Holcomb, R. T. (1981). Kilauea Volcano, Hawaii: Chronology and morphology of the surficial lava flows.

Jaén-Molina, R., Caujapé-Castells, J., Reyes-Betancort, J. A., Akhani, H., Fernández-Palacios, O., de Paz, J. P., ... & Marrero-Rodríguez, Á. (2009). The molecular phylogeny of Matthiola R. Br.(Brassicaceae) inferred from ITS sequences, with special emphasis on the Macaronesian endemics. Molecular phylogenetics and evolution, 53(3), 972-981.

Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H. (2006). Phylogenomics: the beginning of incongruence?. TRENDS in Genetics, 22(4), 225-231.

Jorgensen, T. H., & Olesen, J. M. (2001). Adaptive radiation of island plants: evidence from Aeonium (Crassulaceae) of the Canary Islands. Perspectives in Plant Ecology, Evolution and Systematics, 4(1), 29-42.

Juan, C., Emerson, B. C., Oromí, P., & Hewitt, G. M. (2000). Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends in Ecology & Evolution, 15(3), 104-109.

Kolbe, J. J., Glor, R. E., Schettino, L. R., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. Nature, 431(7005), 177-181.

Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. Molecular ecology, 23(7), 1668-1678.

Linder, C. R., & Rieseberg, L. H. (2004). Reconstructing patterns of reticulate evolution in plants. American journal of botany, 91(10), 1700-1708.

Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., & Rodríguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. Science, 279(5359), 2115-2118.

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. Nature, 457(7231), 830-836.

Losos, J. B. (2010). Adaptive radiation, ecological opportunity, and evolutionary determinism. The American Naturalist, 175(6), 623-639.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Research, 23(11), 1817-1828.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution, 66(2), 526-538.

Meimberg, H., Abele, T., Bräuchler, C., McKay, J. K., de Paz, P. L. P., & Heubl, G. (2006). Molecular evidence for adaptive radiation of Micromeria Benth.(Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data. Molecular phylogenetics and evolution, 41(3), 566-578.

Meimberg, H., Rice, K. J., Milan, N. F., Njoku, C. C., & McKay, J. K. (2009). Multiple origins promote the ecological amplitude of allopolyploid Aegilops (Poaceae). American Journal of Botany, 96(7), 1262-1273.

Meyer, J. R., & Kassen, R. (2007). The effects of competition and predation on diversification in a model adaptive radiation. Nature, 446(7134), 432-435.

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution, 19(4), 208-216.

Nosil, P., Parchman, T. L., Feder, J. L., & Gompert, Z. (2012). Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in Timema stick insects. BMC evolutionary biology, 12(1), 1.

Olmstead, R. G., & Palmer, J. D. (1997). Implications for the phylogeny, classification, and biogeography of Solanum from cpDNA restriction site variation. Systematic Botany, 19-29.

Otto, R., Whittaker, R. J., Gaisberg, M., Stierstorfer, C., Naranjo-Cigala, A., Steinbauer, M. J., ... & Fernández-Palacios, J. M. (2016). Transferring and implementing the general dynamic model of oceanic island biogeography at the scale of island

fragments: the roles of geological age and topography in plant diversification in the Canaries. Journal of Biogeography.

Peters, J. L., Zhuravlev, Y., Fefelov, I., Logie, A., & Omland, K. E. (2007). Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (Anas spp.). Evolution, 61(8), 1992-2006.

Puppo, P. (2015). Diversification patterns of Micromeria Benth. (Lamiaceae) in the Canary Islands: genetic and morphological implications of inter-island colonization. PhD Thesis, University of Porto, Porto, Portugal.

Puppo, P., Curto, M., Velo-Antón, G., Pérez de Paz, P. L., & Meimberg, H. (2014). The influence of geological history on diversification in insular species: genetic and morphological patterns of Micromeria Benth.(Lamiaceae) in Tenerife (Canary archipelago). Journal of Biogeography, 41(10), 1871-1882.

Puppo, P., Curto, M., Gusmão-Guedes, J., Cochofel, J., de Paz, P. L. P., Bräuchler, C., & Meimberg, H. (2015a). Molecular phylogenetics of Micromeria (Lamiaceae) in the Canary Islands, diversification and inter-island colonization patterns inferred from nuclear genes. Molecular phylogenetics and evolution, 89, 160-170.

Puppo, P., Curto, M., & Meimberg, H. (2015b). Development and characterization of 16 microsatellite markers for Micromeria (Lamiaceae) from Tenerife (Canary Islands, Spain) using 454 sequencing. Conservation Genetics Resources, 7(3), 743-749.

Puppo, P., & Meimberg, H. (2015a). New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa, 230(1), 1-21.

Puppo, P., & Meimberg, H. (2015b). Corrections to Phytotaxa 230: New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa, 239(2), 197-198.

Rainey, P. B., & Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. Nature, 394(6688), 69-72.

Reid, N., Demboski, J. R., & Sullivan, J. (2012). Phylogeny estimation of the radiation of western North American chipmunks (Tamias) in the face of introgression using reproductive protein genes. Systematic biology, 61(1), 44-62.

Rosindell, J., & Phillimore, A. B. (2011). A unified model of island biogeography sheds light on the zone of radiation. Ecology Letters, 14(6), 552-560.

Roy, D., Lucek, K., Walter, R. P., & Seehausen, O. (2015). Hybrid ‘superswarm’leads to rapid divergence and establishment of populations during a biological invasion. Molecular ecology, 24(21), 5394-5411.

Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends in Ecology & Evolution, 24(7), 394-399.

Sanmartín, I., Van Der Mark, P., & Ronquist, F. (2008). Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. Journal of Biogeography, 35(3), 428-449.

Santana, Q. C., Coetzee, M. P., Steenkamp, E. T., Mlonyeni, O. X., Hammond, G. N., Wingfield, M. J., & Wingfield, B. D. (2009). Microsatellite discovery by deep sequencing of enriched genomic libraries. Biotechniques, 46(3), 217.

Saunders, N. E., & Gibson, D. J. (2005). Breeding system, branching processes, hybrid swarm theory, and the humped-back diversity relationship as additional explanations for apparent monophyly in the Macaronesian island flora. Journal of Ecology, 93(4), 649-652.

Schlötterer, C. (2004). The evolution of molecular markers—just a matter of fashion?. Nature Reviews Genetics, 5(1), 63-69.

Schluter, D. (2001). Ecology and the origin of species. Trends in ecology & evolution, 16(7), 372-380.

Seehausen, O. (2004). Hybridization and adaptive radiation. Trends in Ecology & Evolution, 19(4), 198-207.

Seehausen, O. (2006). African cichlid fish: a model system in adaptive radiation research. Proceedings of the Royal Society of London B: Biological Sciences, 273(1597), 1987-1998.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology letters, 9(5), 615-629.

Silvertown, J. (2004). The ghost of competition past in the phylogeny of island endemic plants. Journal of Ecology, 92(1), 168-173.

Silvertown, J., FRANCISCO-ORTEGA, J., & Carine, M. (2005). The monophyly of island radiations: an evaluation of niche pre-emption and some alternative explanations. Journal of Ecology, 93(4), 653-657.

Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W., & Tucker, P. K. (2010). The variable genomic architecture of isolation between hybridizing species of house mice. Evolution, 64(2), 472-485.

Thomson, R. C., Wang, I. J., & Johnson, J. R. (2010). Genome-enabled development of DNA markers for ecology, evolution and conservation. Molecular Ecology, 19(11), 2184-2195.

Whittaker, R. J., & Fernández-Palacios, J. M. (2007). Island biogeography: ecology, evolution, and conservation. Oxford University Press.

Whittaker, R. J., Triantis, K. A., & Ladle, R. J. (2008). A general dynamic theory of oceanic island biogeography. Journal of Biogeography, 35(6), 977-994.

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. Molecular ecology, 11(1), 1-16.

Chapter 2: Molecular markers development

Paper I: Curto, M. A., Puppo, P., Ferreira, D., Nogueira, M., & Meimberg, H. (2012). Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). Molecular phylogenetics and evolution, 63(3), 758-767.

Paper II: Curto, M. A., Tembrock, L. R., Puppo, P., Nogueira, M., Simmons, M. P., Meimberg H. (2013). Evaluation of microsatellites of Catha edulis (qat; Celastraceae) identified using pyrosequencing. Biochemical Systematics and Ecology, 49, 1–9.

Paper III: Curto, M., Nogueira, M., Beja, P., Amorim, F., Schümann, M., & Meimberg, H. (2015). Influence of past agricultural fragmentation to the genetic structure of Juniperus oxycedrus in a Mediterranean landscape. Tree Genetics & Genomes, 11(2), 1-13.

.

2.1. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae)

Manuel A. Curto, Pamela Puppo, Diana Ferreira, Marcos Nogueira, Harald Meimberg
CIBIO, Research Center for Biodiversity and Genetic Resources, University of Porto, Campus Vairão, R. Monte-Crasto, 4485-661 Vairão, Portugal

Abstract

The use of single copy nuclear markers is of increasing importance in plant phylogenetics. The generally higher level of variability compared to chloroplast DNA and the ability to use incongruence in a multilocus analysis to determine reticulation patterns makes these kinds of sequence based markers especially useful for species level investigations. However, the prevalence of gene duplication that results from the high frequency of polyploidization events during the evolution of higher plants can impede marker development especially for groups lacking model organisms.

Here, we present the strategy and results of marker development for phylogenetic analysis in *Micromeria*, using publicly available DNA sequences and ESTs from related genera from Lamiaceae, subfamily Nepetoideae. By eliminating markers with signatures of duplication during four steps of marker development, we were able to select 19 primer pairs that resulted in orthologous products for all the species studied. This corresponds to 23% of the initial 84 primer pairs designed.

Using an initial sampling of eight individuals, we tested the markers for support of phylogenetic hypotheses related to the evolution of Micromeria on the Canary Islands. While some hypotheses were supported by all markers, an east west split, with a closer relationship between the species of Tenerife and Madeira on one hand and the ones from Gran Canaria and the eastern islands on the other is supported by 12 markers but contradicted by the remaining seven. This indicates that reticulation and inter-island gene flow played a role in the evolution of Micromeria.

Keywords: Micromeria; Lamiaceae; Nepetoideae; EPIC marker; Single copy gene

Introduction

The use of nuclear gene sequence information is of increasing importance to resolve deep or species level phylogenies. Low-copy nuclear genes are not only useful

because of their potentially rapid evolutionary rate (Sang, 2002, Small et al., 2004 and Choi et al., 2006) but also because they constitute a nearly unlimited source of markers on different variability levels and are essential for detecting introgression, hybridization, and allopolyploidization events (Wendel and Doyle, 1998, Sang, 2002 and Duarte et al., 2010). Depending on the level of variability, nuclear gene derived markers can be classified into two classes: (1) nuclear protein coding regions (NPLCs) and (2) intron regions, which are flanked by exons that provide conserved primer binding sites (EPIC; Thomson, 2010). Coding sequences tend to be conserved, so NPLC markers are preferentially used in phylogenies that cover a broad range of taxa. EPIC markers, however, have higher variability and are useful for studies using closely related species and intra-species comparisons. On the species level, their high variability results in more resolved phylogenies and can constitute a significant improvement compared to organelle DNA phylogenetic inferences (Peters et al., 2005, Beltran et al., 2002 and Sang, 2002). EPIC markers can also be applied to determine allele frequencies using the sequences in haplotype or SNP analysis and have been shown to have a high potential in population genetic studies (Backström et al., 2008 and Thomson, 2010).

Especially in cases of rapid speciation events, genomic DNA may not be sufficiently diverged to resolve a phylogeny with only one locus (Seehausen et al., 2003). In such cases, multiple independent nuclear loci can then provide the variability necessary for reliable phylogenetic analysis (Beltran et al., 2002 and Sang, 2002). Methods developed in recent years allow for determination of species trees from datasets using coalescence or Bayesian approaches (Murphy et al., 2001, Li et al., 2007, Rowe et al., 2008 and Edwards, 2009) and provide resources to apply multiple, highly variable EPIC markers for phylogenetic analysis.

Success in the development of EPIC markers depends highly on the genomic resources available for a certain taxonomic group. One approach to develop such markers is the comparison between two genomes and the design of primers in conserved exon regions. This approach results in markers theoretically suitable for the taxonomic range between the species used for the initial genomic comparison (Li et al., 2007) and can be especially applied to develop universal NPLC markers. For EPIC marker development, it is desirable to include species closely related to the target group in order to identify polymorphisms in the potential primer binding sites. Increasing dissimilarity between the source genomic sequences used for primer design and the target genomic sequences will result in a decrease of PCR amplification.

While there is an ever-increasing number of newly sequenced genomes (130 in progress at Genbank with 30 completed or in assembly as of April 2011), for most non-model plants whole genome sequences are not available. One way to overcome this lack

of complete sequencing data is to use transcript libraries for marker creation. Since primer binding sites should be preferentially located in conserved DNA regions, transcripts can be used for comparisons and subsequent marker development, for example as cDNA libraries (Whittall et al., 2006) or ESTs (Choi et al., 2006 and Townsend et al., 2008).

The use of nuclear genes in phylogenetic analyses of plants is impeded by a high frequency of gene duplication. Only a small subset of genes exist in low copy numbers across the genome, due to independent gene duplication events or polyploidization, i.e., whole genome duplication (Small et al., 2004 and Wu et al., 2006). Polyploidization is one of the major evolutionary forces in Angiosperms (Adams and Wendel, 2005) and it is assumed that all recent angiosperms underwent at least one polyploidization event during their evolution, leading to an increased level of duplicated loci and the existence of gene families for most genes (Bodt et al., 2005). The genome tends to lose redundant regions by a process known as diploidization, which explains the small sizes of some genomes even after polyploidization (Comai, 2005). The most prominent example is Arabidopsis thaliana (L.) Heynh. that experienced several polyploidization events but has one of the smallest genomes in angiosperms. Nevertheless, plants with polyploidy in their evolutionary history show elevated levels of duplicated genes even after reduction of genome size (Adams and Wendel, 2005). Consequently, more effort is required in marker development and special attention must be given to verify and exclude paralogous genes that could confound phylogenetic analyses.

It has been estimated that around 10% of the genes in plant genomes are single copy, presumably because of a detrimental dosage effect after duplication (Duarte et al., 2010). These dosage affected duplicated genes would likely return to a single copy state within a few generations. Comparisons between whole genome sequences allowed the discovery of single copy dosage sensitive genes, which were then used to develop conserved primers of variable loci for a broad range of taxa (Duarte et al., 2010). However, if a locus is not maintained in a single copy state by a biological mechanism, it is not possible to conclusively test a priori if a marker is duplicated or not without testing multiple sequences and screening the phylogeny or deviation of Hardy Weinberg Equilibrium (HWE). The high prevalence for gene duplication in plants has several limitations on marker development as error sources in phylogenies (Thomson, 2010): (I) Most markers developed will target multiple paralogous copies in the PCR product preventing direct sequencing. This is a consequence of length mutations in the intron and can be detected in gel electrophoresis if there is a sufficient length difference. (II) All heterozygotes detected in duplicated genes will constitute polymorphisms between the paralogs rather than alleles, leading to a faulty haplotype reconstruction and increasing

the need for a large amount of clones to differentiate the different paralog forms. This is not feasible if a multitude of loci and samples are used. (III) Post-duplication: paralogs could be sorted into different lineages, or different copies may be lost in different individuals. This will lead to incorrect phylogenies where the topology reflects the split between groups of paralogs rather than the actual phylogenetic relationships.

Here we present the strategy and results of a study designed to develop EPIC markers for multi-locus intrageneric investigations in the Nepetoideae, a subfamily of Lamiaceae. The Lamiaceae, or the mint family, comprises more than 200 genera and 7000 species subdivided into seven subfamilies (Bräuchler et al., 2010). The largest subfamily, the Nepetoidoeae contains several economically important genera such as *Salvia* L., *Mentha* L., Thymus L., *Rosmarinus* L., *Origanum* L., *Lavandula* L., and *Ocimum* L. We used ESTs available for single species of Nepetoideae for primer design after comparison to published whole genome sequences. These primers were used for amplification of samples belonging to different taxonomic levels in the Nepetoideae. The results were used to discriminate low copy genes for use as phylogenetic markers in an infrageneric study of the Nepetoideae genus *Micromeria* Benth. on the Canary Islands. We developed a final set of EPIC markers that mainly consisted of single copy genes and are likely to produce orthologous, rather than paralogous products. These loci were tested for potential duplication using phylogenetic analyses. Our strategy for marker development is a rapid and cost effective way to develop a set of multiple and applicable markers.

Materials and methods

Plant material and DNA isolation

Representatives from six genera of the Lamiaceae were used in the study. We used one species each of *Lavandula*, *Origanum*, Salvia, and *Ocimum*, two species from *Mentha*, and eight samples from six species of *Micromeria* (Table 1). Plant material was collected as fresh leaves from cultivated material that was either the subject of previous studies or from cultures for commercial use (Fa. Ervital). One sample of *Mentha* was collected in the field. Leaves were dried over silica and grounded in 2 ml tubes in a Retsch Mill (MM400), at maximum force (30 Hz) for 10 min using 2 mm steel beads.

Table 1. Primer information for the loci presenting positive amplification results. The information comprises locus name and primer sequence, accession number of the sequence from which the primer was designed and respective organism and gene. For loci indicated with asterisk (*) more than one primer pair was designed.

Marker name	Source sequence accession number	primer sequence		organism	gene
		Forward	Reverse		
O.oni.007	EF558371.1	TGTAAGTCCCAACAAATGTGTC	TGTTCGCCATAAATGGGTT	Origanum onites	mitogen activated protein kinase 1 (MAPKK1)
L.ang.001	DQ886905.1	GGTACAAATAAACCATCCCAT	TCCACGTGGCACTTCATGAG	Lavandula angustifolia	putative alcohol acyltransferase 2 (AAT2)
L.ang.002	DQ886905.1	ATGAGCTCATGAAGTGCC	GCCTTCTACTACAACCAGAAG	Lavandula angustifolia	putative alcohol acyltransferase 2 (AAT2)
M.pip.002	AF116825.2	CCAGCAGCATACGCTAGATC	AAGCTGGAAAGGACATTGCT	Mentha x piperita	1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)
M.pip.006	AY288138.1	AGCTAACTGCCTGCATATCT	TCACCAACAATTGTCAATGT	Mentha x piperita	mentholdehydrogenase
M.pip.008	AJ249324.1	TGTAAACTTCACCAGTGGG	GAGAAGATGGATTCCATGAT	Mentha x piperita	isopentenyl monophosphate kinase (IPK)
M.pip.013	AW254715	CAGATGGCTCGTACCAAGCAAAC	CACAGCATGGCTCTGGAAA	Mentha x piperita	histone H3
M.pip.014	AW254724	AGATGTTCATCGGAGATGGA	TCCTCGGTGGGGTGAGGAA	Mentha x piperita	Mg-dependent ATPase 1 (LeMA-1)
M.pip.015	AW254748	GCAGAGCTTCAGAGGCTTCT	GGAATAAAGGTAATGGAGGAAG	Mentha x piperita	PHD finger protein-realated (ING2)
M.pip.016	AW254751	TTGCTGAGGAATTGAAGCG	TCTCAATCACCCGAAACTCT	Mentha x piperita	rac-GTP binding protein (GTPBP1)
M.pip.017	AW254759	GATCCCTATTTCATGCGCAAC	TCAATTTCTGTGCTGGGAACC	Mentha x piperita	hydroxyproline-rich glycoprotein family protein (HRGP1)
M.pip.024	AW255338	GAGTATGGTGATATGCAGCTG	AGCCACCTTTCCAGATTC	Mentha x piperita	6-phosphogluconate dehydrogenase(PGD)
M.pip.025	AW255366	GAAATAGCAGAGAGCATGATG	CAGTAGGTTCTACGCCTGAAC	Mentha x piperita	mitochondrial processing peptidase (PMPC)
M.pip.027	AW255375	TAGAGCCAGATCCGAAGCT	CACCGCTTCAATGAGCATCTGT	Mentha x piperita	calcium-dependent protein kinase (CaM-KK 1)
M.pip.028	AW255391	TCGACTCGAAGGCTGCAG	CAACAAGGCATAGCTAACTCAAG	Mentha x piperita	enolase (ENO)
M.pip.029	AW255394	CCTGATTACTACTTCCGCATCAC	TTTGGAGGTGGGCGTCG	Mentha x piperita	chalcone synthase (CHS)
M.pip.031	AW255423	AATGGTCTTTGTGGTCAAGAG	TACATGTCGGAGGTGTCG	Mentha x piperita	putative aldo/keto reductase (AKR)
M.pip.035	AW255478	TGTCCTGCGATCCTTACATG	AGCAGTCATACCAGGCATACC	Mentha x piperita	quinone oxidoreductase (NQO)
M.pip.037	AW255529	GAGAAGGAAGAAGTTCGAGTTC	GGCATCAACAAGGTATATTGC	Mentha x piperita	3'-5' exonuclease (Exo1)
M.pip.038	AW255571	CCTCAAAATGCTTAAGCACG	GTTCAGCCCATGAATCAAAG	Mentha x piperita	26S proteasome, non-ATPase regulatory subunit (PSMD)
M.pip.041	AW255629	AATCAATAAACCAGTCGTTGC	GACATTGGAACACCAGCATA	Mentha x piperita	ATP-citrate synthase
M.pip.042	AW255663	AAGTCTACACCACGAGCAAG	TTTTCAGAGCCTTTCCTTTC	Mentha x piperita	aldo/keto reductase
M.pip.043	AW255671	GGGATTTCAAGAGGTTGCAG	GGGTCACACAGCAATGACTG	Mentha x piperita	ubiquitin conjugating (UBE)
M.pip.044	AW255729	CCAGATTCTCAAGTATCCTCTC	TGATTCCGATTTTGTTGG	Mentha x piperita	60S ribosomal protein L23A
M.pip.046	AW255789	CGCGGTAGTAGTTGTTGCAG	GCACGTGGACCTAACAACAG	Mentha x piperita	unknown protein
M.pip.047	AW255802	GCTCCATTTATCTGTGAGTTC	CAACTGTGATAAAAATCCACC	Mentha x piperita	adenosine kinase (adk)
M.pip.050	AW255849	ATGGTGCTCTTGTTCGCGTC	GGACCCTCCAGCAAAACGTG	Mentha x piperita	Rab GDP dissociation inhibitor (GDI)
M.pip.054	AW255887	TCATTGCTTTATTTCAGAGGG	TGATACCAATGCCCGTTC	Mentha x piperita	pectin acetylesterase
M.pip.056	AW255892	GTTTGGCATTGTTGAGGGTCT	AGTGCAATTCCAGCCTTGG	Mentha x piperita	glyceraldehyde 3-phosphate dehydrogenase (gapdh)
M.pip.057	AW255966	GCTACAGCTACTTCTTTGAAGC	GTCCAGAATCATCTGAGATAG	Mentha x piperita	NAC domain protein (SlNAC1)
M.pip.058	AW255972	CCAACGTAGAACTGGTCAACT	GAGAGTTACAAGGACCAATGC	Mentha x piperita	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (ispG)
M.pip.059	AW256000	GGCTTCCTCTCGAAGCTAAC	CAACAGCATGAAGAAGTCCAA	Mentha x piperita	ubiquitin c-terminal hydrolase 3 (UCH3)
M.pip.061	EL342293	GTTGGCCGTGTTGCTATG	GCCAATCTTCCCACAAATAACT	Mentha x piperita	Photosystem II, 22 kDa protein (PSBS1)
M.pip.062	AW254789	GATGTCTATGGTACCTTAGAAG	GAGGTTCCCAAATCATCAGC	Mentha x piperita	limonene synthase
O.basi.002	DY344636	TTGGCCATCAAGGATGAAG	AGCGGCAACCATGTTCTC	Ocimum basilicum	cobalamine-independent methionine synthase
O.basi.005	DY344623	GTAGCTTCCAACAGATGCTGC	TCCTTACCTCCCCTGACCTC	Ocimum basilicum	selenocysteine methyltransferase
S.milt.001	GQ370517	CTCGAATGTGTTCCTGCAG	CACATCCCTCTTAGTCCCATAC	Salvia miltiorrhiza	Cold acclimation protein (COR)
S.milt.003	GQ245764	TGTCTTCCCTGTCCATGTTC	CCATGGCCTACTACGTCATC	Salvia miltiorrhiza	galactinol synthase (GOLS1)
S.milt.004	GQ249111	ATCAATGGATCAATGAGTTTC	GCTCATGTTTGCTCAATTC	Salvia miltiorrhiza	phenylalanine ammonia-lyase (PAL2)
S.milt.009	FJ540907	CTGTCCACCAGACAAAAAG	TTCAATACCTGAGTGGTGTG	Salvia miltiorrhiza	fructose-bisphosphate aldolase-like
S.milt.010	FB335864	GAACTCCTCGTAGTTGATCTGC	AGGAAGCCTTCTCGCTATTC	Salvia miltiorrhiza	calmodulin
S.milt.013	EF377337	GTGGTTCACTAGCTCAGCAATG	ATGACGGTGCCGTTCTTC	Salvia miltiorrhiza	cinnamate 4-hydroxylase (C4H)
M.pip.030*	AW255395	TCAATGAAGGACTGGAGAGG	ACAGACTGCATCTACCATTAGG	Mentha x piperita	glyceraldehyde-3-phosphate dehydrogenase

Table 1. Continued

Marker name	Source sequence accession number	primer sequence		organism	gene
		Forward	Reverse		
M.pip.033*	AW255473	TCTGGTCGGTTGGATCGT	AAGAACTCCTTTTGGTGGACG	Mentha x piperita	Mg-dependent ATPase 1
S.milt.002*	FJ476255	CATTGGCTGCACTAAGAAC	GCTCTAAAGCATCCTAACTGG	Salvia miltiorrhiza	1-deoxy-d-xylulose 5-phosphate reductoisomerase
S.milt.006*	FJ768961	CAGCAGCATAAGCTAGATCC	TAGCAGATGCTCTAAAGCATC	Salvia miltiorrhiza	1-deoxy-d-xylulose 5-phosphate reductoisomerase
S.milt.012*	EF666999	GTTGTTGATCTTGGTGGTGC	TCGACACCTCCGACATGTAC	Salvia miltiorrhiza	putative aldo/keto reductase 2
S.milt.015*	DQ991431	GTAATCTTCTTTCCCATATTCC	AGACAAGCCTGAAATCATTC	Salvia miltiorrhiza	1-deoxy-d-xylulose 5-phosphate reductoisomerase
S.milt.017*	AY657030	AACTCCTCGTAGTTGATCTG	GAAGCCTTCTCGCTATTC	Salvia miltiorrhiza	calmodulin

DNA extraction followed the protocol described by Alexander et al. (2006). For 20 mg of powdered dry material, 320 µl homogenization buffer, 80 µl lyses buffer and 40 µl RNAse (10 mg/ml) were added and incubated for 10 min at 65 °C. For protein precipitation 130 µl potassium acetate (7.5 M) was added, incubated 5 min on ice and centrifuged differentially, 1 min at 1000 rpm and subsequently 8 min at 14,000 rpm. The supernatant was mixed with 1.5 times its volume in binding buffer and the resulting solution passed through a silica membrane column (EconoSpin™) by centrifugation for 1 min at 14,000 rpm. After washing the membrane two times with 500 µl 70% ethanol, DNA was eluted with 50 µl 10 mM Tris buffer pH 7.5 and used subsequently as DNA solution in the PCR.

Marker development strategy and Primer design

We used a four step selection process to identify loci from public databases for suitability of amplifying the desired ingroup: the species from *Micromeria* of the Canary Islands (Fig. 1): Step 1: Comparison of publicly available *mRNA*-derived sequences of representatives from the Nepetoideae with the Arabidopsis genome and selection of loci that span introns and showed no sign of duplication in the Arabidopsis genome, i.e. that produced one single hit in the BLAST search; Step 2: amplification and annealing temperature optimization using one individual congeneric to the source species of the EST or sequence where failed amplification indicates autapomorphies or sequencing errors in the primer binding sites; Step 3: amplification of several genera of Nepetoideae including the ingroup to determine the level of applicability of the primers positive in step 2; Step 4: amplification of multiple samples from *Micromeria* with primers able to amplify the ingroup in step 3 to verify the applicability to multiple species.

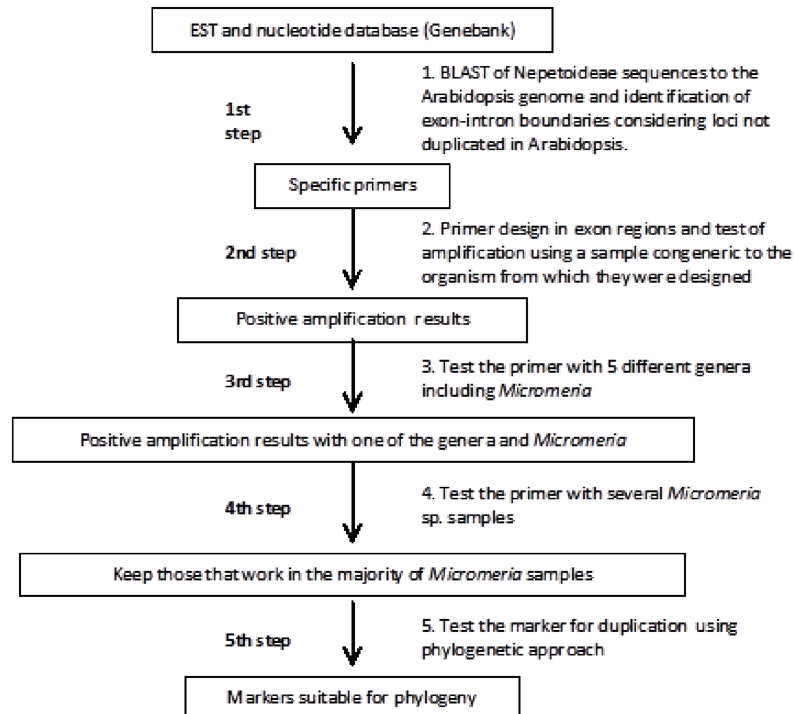


Fig. 1. Schematic summary of marker development strategy used to develop EPIC markers for Micromeria using publicly available sequence information of DNA or RNA from different genera of the Nepetoideae (Lamiaceae).

At each step, primer pairs that amplified paralogs that differed in size of about 25 bp were identified by gel electrophoresis. All amplification products of markers amplifying multiple samples in step four were sequenced from PCR products. Markers excluded were those that showed signs of duplication as heterozygous indels fixed within the sampling and an excess of heterozygous positions.

For primer design, ESTs and a low number of genomic sequences were available from Genbank for five genera of Nepetoideae, the majority for *Mentha* and *Salvia*, a lesser amount for *Lavandula*, *Origanum* and *Ocimum*. In total, about 1000 ESTs and mRNAs from the nucleotide database were chosen randomly if they were longer than 500 bp, primarily for *Mentha*. *Mentha* is more closely related to our target group, *Micromeria*, than the other genera (Bräuchler et al., 2010) and mainly ESTs from the cultivated species *Mentha x piperita* L. were available. Using the local blast function of BioEdit version 7.0.5 (Hall, 1999) the ESTs were blasted against the complete *Arabidopsis thaliana* genome. Tabular output was screened for sequences that (1) produced highly significant hits of at least 40 bp, (2) were not aligned to more than one region in the Arabidopsis genome and (3) showed highly significant hits of at least two different regions of the EST not more than 1000 bp apart on the Arabidopsis genome. The latter criterion was chosen because regions between two hits could constitute intron regions for the development of the desired variable markers. ESTs satisfying these

criteria were selected for primer design. Primers were picked for regions of significant blast hits that comprised not more than two mismatches in the potential primer region between the genome and query sequence. The sequence from the EST was used as primer sequence. Primer design was performed using Primer3Plus (Untergasser et al., 2007) under manual control, targeting a length of the primers to 20 bp, with 50% GC content and not more than three identical nucleotides in a row. For nine loci, multiple primers were designed to achieve a higher coverage of the target gene and increase likelihood of positive amplification.

Optimization of PCR and selection of primers

All PCRs were performed with Qiagen Mastermix (QIAGEN – HotStarTaq Plus Master Mix Kit). This system was used because of the high level of specificity and the reduced need for PCR optimization. All primers designed were initially tested in an eight steps temperature gradient ranging from 7 °C below to 7 °C above the optimal calculated annealing temperature (typically 48–62 °C), using a sample from the genus the EST was derived from. The final reaction volume was 10 µl per well and contained 1 U Taq DNA polymerase (QIAGEN), 75 nmol MgCl2, 1 nmol of each dNTP, 4 nmol of each primer, about 20 ng template corresponding typically to 0.5 DNA solution. The following temperature profile was used: initial denaturation and polymerase activation at 95 °C for 15 min, followed by 40 cycles of 95 °C for 30 s, annealing for 1 min, and extension at 72 °C for 1 min, with a final extension step of 72 °C for 10 min.

Optimal annealing temperatures were defined as the temperature where only one specific product was indicated after gel electrophoresis (2% agarose with a resolution of approximately 25 bp). Primers that showed no amplification product or more than one band in the gel electrophoresis over the whole range of temperature were discarded from subsequent steps.

Primer pairs with positive amplification within the temperature gradient were tested as a PCR on all the remaining genera using the optimal annealing temperature, and primer pairs that produced a clear band for the *Micromeria* sample were used subsequently to test the applicability of the primer within this genus using the remaining *Micromeria* samples. Primers that were able to produce a product for the majority of the samples were chosen for further investigation.

Sequencing and phylogenetic analysis

Amplification products of the primers chosen in the last step of primer screening were sequenced for all samples investigated. PCR products were cleaned using Exo/Sap digestion in a final volume of 8 µl containing 4 U Exonuclease I (Fa), 1 U Shrimp alkaline

Phosphatase (Fa) at 37 °C for 15 min followed by 15 min at 85 °C for inactivation. The resulting solution was used for sequencing using the BigDye 3.1 kit according the manufacturer's protocol, performed at the central sequencing facility at the University of Munich.

Sequences were tested for specificity by BLAST searches against the Genbank sequence used for primer design. Sequences of one locus were aligned using the ClustalW algorithm as implemented in BioEdit (Hall, 1999) under manual control. Heterozygous point mutations were included using equate macros (W, R, Y, S, K, and M). In sequences comprising heterozygote length polymorphisms, unreadable parts of the sequence were treated as missing data.

Haplotype reconstruction was performed using the PHASE algorithm as implemented in DNASP ver. 5.10 (Librado and Rozas, 2009). Informative indels were included in the analysis by treating them as nucleotide information to avoid losing information and to limit computational effort.

Phylogenetic reconstruction was based on maximum parsimony and maximum likelihood inference using the exhaustive search algorithm implemented in PAUP v4.0 (Swofford, 2002). Parsimony analyses were done considering unordered character states and equal weighting with gaps treated as missing data. When exhaustive searches became computationally intractable, a heuristic search was used with 100 random stepwise-addition replicates, TBR branch swapping, and MULTREES optimization. The consistency index (CI) and retention index (RI) were calculated (Farris, 1989 and Kluge and Farris, 1969) and branch support was determined by bootstrap analysis of 1000 replicates (Felsenstein, 1985) using a heuristic search with identical settings as above. The support values are shown on the consensus tree. Maximum likelihood analysis was performed under the GTR model of sequence evolution. Maximum likelihood and parsimony reconstruction led to comparable results so only the likelihood analysis is shown.

To determine the variability of markers at different sampling levels, we calculated three estimates for the number of polymorphic positions: (1) within Nepetoideae: between *Micromeria* species and other Lamiaceae genera, (2) within *Micromeria* from Canary Islands and the outgroup *Micromeria inodora* (Desf.) Benth. from Baleares, and (3) within the Canary Islands only. All measures were obtained using DNAsp (Librado and Rozas, 2009). Pair wise and absolute distances were calculated as a measure of genetic differentiation on these levels. The distance matrices were obtained using PAUP v4.0 (Swofford, 2002) and the average values for each different level were calculated using Excel. These measures were used as indicators for suitability of each marker at different taxonomic levels.

Exon boundaries were determined by BLAST searches of the sequences against the mRNA used for primer design. The exon regions were searched for synonymous and non-synonymous substitutions using DNAsp (Librado and Rozas, 2009). This approach required the definition of the open reading frame (ORF) of each gene, which was obtained using the ORF finder function at NCBI.

Results

PCR amplification and sequencing

In total, 84 primer pairs were designed corresponding to 76 loci, 59 primer pairs from ESTs of *Mentha x piperita*, 20 pairs from ESTs of different Salvia species, five pairs from ESTs of *Ocimum basilicum* L. and one pair from ESTs of *Origanum onites* L. Twenty-six pairs exhibited no product and eight pairs exhibited multiple products over the whole range of the temperature gradient and were therefore discarded from further analysis. Fifty of the 84 primer pairs successfully amplified a clear distinct product using a sample congeneric to the EST (step 2, Fig. 1 and Table 1). Of these, only nine were unable to amplify at least one of the other genera investigated in step 3. Amplification success of the different Nepetoideae genera was slightly higher with increasing phylogenetic relatedness (according Bräuchler et al., 2010) to the source sequence. Of the 34 primers derived from *Mentha*, 33 amplified a product from the *Mentha* species used as template in this study and between 14 and 22 products for the other species included.

A total of 32 primer pairs were applicable for *Micromeria* within step 3 (Fig. 1). These primers were subsequently used for amplification of multiple samples from the Canary Islands (the desired ingroup) and the outgroup species, *M. inodora* from the Baleares. The majority of primer pairs were suitable to amplify most of the *Micromeria* samples investigated (Table 2), and 14 primers amplified one or two of the additional samples. The outgroup taxon, *M. inodora*, was positively amplified by only six markers. This is likely due to the fact that in the initial amplification in step 3, a sample from *Micromeria* from the Canary Islands was used, creating a desired bias for amplification success within the ingroup. Within the ingroup, 24 primer pairs were able to amplify the majority of samples.

Table 2. Amplification and sequencing results of primer pairs tested in *Micromeria*. The different numbers correspond to different levels of amplification success: 0 – no product; 1 – amplification and sequencing success; 2 – amplification of multiple products; and 3 – successful amplification but no readable sequences. Note that Ba: Baleares, Te: Tenerife, Ma: Madeira, La1: Lanzarote sample one, La2: Lanzarote sample two and GC: Gran Canaria.y: yes, n: no.

Marker names	M. varia GC	M. lanata GC	M. tenuis GC	M. varia La2	M. varia La1	M. varia. Ma	M. hyssopifolia Te	M. inodora Ba	Nr. positives	Used in step four
M.pip.017	1	1	1	1	1	1	1	1	8	y
M.pip.044	1	1	1	1	1	1	1	1	8	y
M.pip.056	1	1	1	1	1	1	1	1	8	y
M.pip.002	1	1	1	1	1	1	1	0	7	y
M.pip.024	1	1	1	1	0	1	1	1	7	y
M.pip.027	1	1	1	1	1	1	1	3	7	y
M.pip.038	1	1	1	1	1	1	1	0	7	y
M.pip.047	0	1	1	1	1	1	1	1	7	y
O.basi.005	1	1	1	1	1	1	1	0	7	y
S.milt.010	1	1	1	1	1	1	1	0	7	y
M.pip.006	1	1	0	1	1	1	1	0	6	y
M.pip.014	1	1	1	1	0	1	1	0	6	y
M.pip.050	1	1	1	1	0	1	1	0	6	y
M.pip.057	1	1	1	1	1	0	1	0	6	y
M.pip.058	1	1	1	1	1	0	1	0	6	y
O.oni.007	1	1	1	1	1	1	3	0	6	y
S.milt.003	1	1	1	1	0	0	1	1	6	y
M.pip.013	1	3	1	2	1	1	1	0	5	y
M.pip.016	1	1	1	1	3	0	1	0	5	y
M.pip.031	1	0	1	1	1	0	1	0	5	y
S.milt.001	0	1	1	1	1	0	1	2	5	y
S.milt.013	1	1	1	0	0	1	1	0	5	y
M.pip.015	1	1	1	0	0	1	3	0	4	n
M.pip.041	3	0	1	0	1	1	1	3	4	n
L.ang.002	1	3	3	1	0	1	3	0	3	n
M.pip.008	3	0	3	3	3	1	3	3	1	n
M.pip.025	3	0	0	0	0	0	1	0	1	n
M.pip.029	0	0	0	2	0	2	1	0	1	n
M.pip.062	0	1	0	0	0	0	0	0	1	n
S.milt.012	3	3	0	3	3	3	3	0	0	n

The absolute and relative amount of primers amplifying multiple products decreased as expected during the selection steps. A total of 54 primers were unable to differentiate between duplicated loci in the PCR. After sequencing of the candidate markers identified in step 4, an additional three loci had to be excluded because of fixed heterozygous indels indicated in the sequences for all samples investigated. Finally, a total of 22 markers were selected according to these criteria based on the results of PCR amplifications and sequencing.

Sequence analysis

Marker sequences were between 350 bp and 1600 bp and in agreement with expected lengths relative to the corresponding Arabidopsis sequences, which were between 300 bp and 2714 bp. Marker sequence length deviated by 50–250% of the genomic sequence used in the BLAST search. In *Micromeria*, the markers chosen for the study were between 361 bp and 1689 bp (in *M. hyssopifolia* Webb & Berthel. from

Tenerife) and comprised between six and 133 polymorphic positions. Absolute numbers of polymorphic positions were significantly correlated to length of the locus, however several outliers were observed. Maximum average pairwise distances between the sequences were between 0.0064 and 0.1548 and nucleotide diversity within the ingroup was between 0.00474 and 0.09179.

From the 22 markers in the final dataset, 20 contained introns. Intron regions, as expected, were more variable than exon regions and were responsible for 81% of the sequence variation in intron containing markers. This is partly because intron sequences cover a larger part of the loci (63% on average of the alignment; sd = 15). The level of variability in exon regions ranged between 16% and 77% of intron variability (averages 37%; sd = 19). Variability of the entire set of markers in the ingroup ranged from 1.2% to 6.6%, in 8 to 73 positions, respectively (average 27; sd = 17).

Polymorphic positions were detected as mutations between sequences in addition to heterozygous position within one sequence in the chromatogram. Within 14 of the markers within the ingroup, there was at least one polymorphic position. Additionally, there were heterozygous and homozygous positions for both states.

Some markers showed an excess of heterozygous positions with all of the sequences heterozygous at a specific position. Only four of the loci exhibited a higher rate of heterozygous polymorphisms than homozygote polymorphisms, indicating low quality of sequencing or multiple binding sites. One locus, S.milt.001, showed a heterozygous indel for most sequences of the ingroup resulting in a very low amount of information. It could therefore not be included in further analyses.

A phylogenetic analysis was performed for all the 22 loci that could be used for comparative sequencing. Results of parsimony analysis, likelihood scores and alignment characteristics are summarized in Table 3. Most phylogenies were sufficiently resolved for the ingroup to test phylogenetic hypotheses. Phylogenetic analyses of the majority of the loci investigated produced congruent results. We tested four hypotheses of the evolution of the group as indicated in Meimberg et al., 2006 and Bräuchler et al., 2005 and according to the species concept of the Canary Islands *Micromeria* (Pérez de Paz, 1978): monophyly of *Micromeria*; monophyly of the Canary Island representatives; an east–west split of the island samples, with closer relationships of samples from Gran Canaria and Fuerteventura on one side, and Tenerife La Palma and Madeira on the other; and finally, the hypothesis of polyphyly of *Micromeria* varia Benth., a species distributed on several islands.

Table 3. Information on the phylogenetic analyses, marker variability in the ingroup and outgroup; ingroup sequence alignment length and percentage of exon region in that alignment. The phylogenetic analyses information corresponds to the trees obtained using different genera as outgroup and include number of trees, tree length, CI (consistency index) and RI (retention index). The variability measures shown are the number of polymorphic positions (left value) and number of informative positions (right value).

Locus	Number of trees	Tree length	CI	RI	Variability		Heterozygote positions in the ingroup	Alignment length	% Exon
					Ingroup	Outgroup			
M.pip.017	38	268	1	1	41|6	265|7	23	1427	27
M.pip.044	6	51	1	1	8|2	56|40	4	425	58
M.pip.056	5	472	0.9068	0.7427	24|10	338|115	7	766	52
M.pip.002	12	368	0.9511	0.8125	28|6	312|68	1	979	52
M.pip.024	1	115	0.9478	0.9381	12|5	117|42	2	536	100
M.pip.027	4	138	0.9783	0.9348	20|8	130|38	2	750	42
M.pip.038	1	345	0.9884	0.907	57|27	328|32	3	1533	29
M.pip.047	3	289	0.9931	0.9836	20|1	251|120	5	749	45
O.basi.005	12	500	0.942	0.7339	26|9	406|94	6	1311	17
S.milt.010	3	158	1	1	18|3	156|6	0	1325	21
M.pip.006	1	449	0.9866	0.875	41|16	412|43	7	1346	31
M.pip.014	2	40	1	1	16|4	40|5	8	745	40
M.pip.050	2	284	0.9894	0.8696	17|2	277|21	5	918	43
M.pip.057	1	67	0.9254	0.902	14|2	61|27	4	427	100
M.pip.058	2	240	0.9958	0.975	21|9	222|32	3	810	42
M.pip.013	2	94	0.9787	0.75	15|5	95|6	3	544	50
M.pip.016	5	217	1	1	13|4	214|4	5	929	23
M.pip.031	3	516	0.9961	0.8667	49|8	503|14	5	1717	33
S.milt.013	2	243	0.9918	0.9333	73|21	234|22	0	1635	17
O.oni.007	40	44	0.9767	0.6667	39|1	74|3	26	594	74
S.milt.003	156	130	0.9923	0.9914	133|113	137|113	18	660	65
S.milt.001	72	9	1	1	6|1	14|1	4	362	38

The majority of the phylogenies supported these hypotheses. Seven markers contradicted an east–west split and three markers contradicted a monophyly of Gran

Canaria samples (Table 4). The marker S.milt.003, while showing no heterozygous positions, had a highly unlikely phylogeny: four *Micromeria* samples were very similar to the outgroup with 0.5–2.5% pairwise differences. The two remaining ingroup samples comprised highly diverging sequences with up to 19% differences. This could be an indicator of a duplicated locus where different paralogs were amplified in closely related individuals and the marker was therefore not regarded as suitable for phylogenetic analysis. In O.oni.007 a very high number of heterozygous signals were observed in the ingroup which resulted in a low amount of informative positions and poor resolution. The remaining 19 markers were good candidates for use in a multi-locus analysis. Combined into a single matrix, phylogenetic analysis resulted in a well-supported and resolved phylogeny, indicating monophyly of *Micromeria*, the east–west split and polyphyly of *Micromeria varia*. Gran Canaria samples were shown to be paraphyletic with respect to Lanzarote samples. Here, *M. varia* from Gran Canaria formed the sister clade to *M. varia* from Lanzarote (Fig. 2).

Table 4. Test of different phylogenetic hypotheses using the 22 markers that were used in step 4 (Fig. 1), based on interpretation of the phylogeny and the SH test. Indicated are four classes of support, 1: the topology supporting the hypotheses (bootstrap value is shown), 2: the topology contradicting the hypothesis (–), 3: the topology is not resolved enough to support or contradict the hypothesis (0), 4: taxon sampling is not sufficient to test the hypothesis (?). The SH test was performed using the maximum likelihood trees. The SH test marked with an asterisk used the combination of the east–west split and Gran Canaria monophyly as null hypotheses.

Loci	Micromeria monophyly	SH test	East–west split	Gran Canaria monophyly	SH test*	M. varia monophyly	SH test
M.pip. 017	100	1	0	0	0.067	–	0.08
M.pip. 044	100	1	73	0	0.489	–	0.163
M.pip.031	96	1	100	0	0.556	–	0.015
M.pip. 002	100	1	–	–	0.019	–	0.015
M.pip. 024	84	1	62	0	1	–	0.008
M.pip. 027	100	1	–	0	0.216	–	0.005
M.pip.038	100	1	100	99	0.517	–	0
M.pip. 047	100	1	100	90	0.755	–	0.436
S.milt.010	100	1	100	73	0.353	–	0.053
M.pip. 006	100	1	98	98	1	–	0.049
M.pip. 014	100	1	66	0	0.807	–	0.072
M.pip. 050	100	1	–	65	0.081	–	0.081
M.pip. 057	85	1	98	66	1	–	0.087
M.pip. 058	100	1	–	–	0.004	?	?
O.basi.005	95	1	99	0	0.27	0	0.427
M.pip. 013	100	1	–	–	0.259	–	0.265
M.pip. 016	100	1	–	0	0.02	0	0.01
M.pip. 056	100	1	–	?	0.165	–	0.73
S.milt.013	100	1	100	99	1	–	0
O.oni.007	100	1	0	0	0.734	0	1
S.milt.003	–	0	–	–	0	0	0.573
S.milt.001	0	0.779	0	0	0.062	?	?

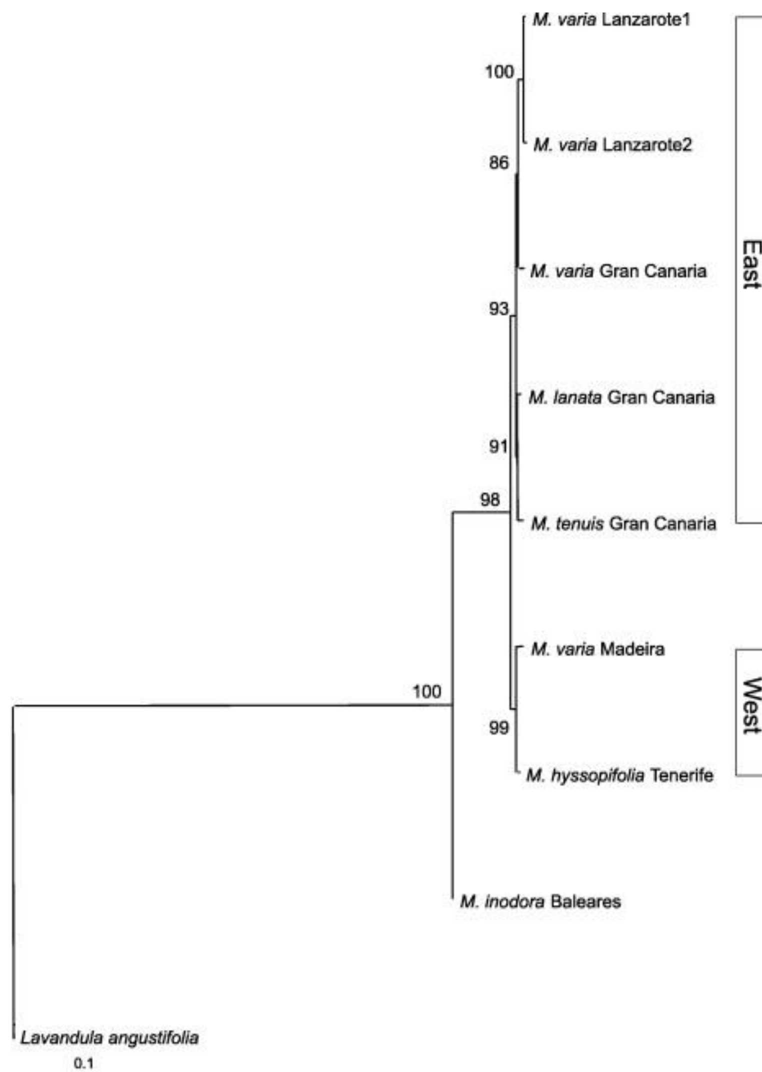


Fig. 2. Phylogenetic tree calculated using a combined matrix of all 19 loci retained. Maximum likelihood and maximum parsimony trees showed congruent results, therefore only the maximum likelihood is shown.

Discussion

The high frequency of polyploidization events during evolution of higher plants and segmental gene duplication complicates the application of nuclear loci in phylogenetic studies in angiosperms. In principle, it is also possible in plants with duplicated genomes to differentiate between paralogs by selective primers within the PCR (Whittall et al., 2006). In population genetic investigations using nuclear markers such as microsatellites, primer pairs that lead to multiple products are generally excluded during marker development. This can result in the majority of markers included in a study to be specific for orthologous sequences even within a polyploid plant that is duplicated for all loci. Marker specificity is indicated by unelevated levels of heterozygosity or deviations from HWE (e.g. Meimberg et al., 2010). Similar approaches can be used to identify nuclear sequence based markers to develop primers that produce orthologous products for sequence analysis. In this respect, positive amplification for a comparative

study results not only from the existence of a specific primer binding site in all individuals but also on the ability of the primers to be selective for one specific ortholog. Depending on the source of the initial comparison, either sequence information of species within the ingroup to be investigated or sequences from outgroup taxa, different conservation patterns of the primer binding sites can be assumed. Ingroup derived primers will increasingly show mismatches with the outgroup or the phylogenetically more distantly related taxa of the ingroup and thus show lower amplification success over a certain range of taxa, considering the increasing likelihood of divergence in primer binding sites with increasing phylogenetic distance. On the other hand, the possibility of primers amplifying only orthologous sequences is higher here, because primer binding sites can be in regions where paralogs have diverged. This divergence was exploited to locate one primer of a primer pair in the more variable 5′ UTR region (Whittall et al., 2006). In contrast, if only outgroup sequences can be used, as in our case, primers must be designed from regions conserved between different genera and should be more prone to co-amplifications of paralogs, especially for loci that became duplicated after the split between the outgroup and the ingroup. The necessity to use outgroup taxa for primer development for *Micromeria* resulted in a comparably high number of initial failed amplifications of about 30% of the constructed primers which amplified only one or none of the species used in the study.

Our approach constitutes a four step, mainly amplification based selection process that allows to eliminate loci with signatures for duplication events and that were unable to amplify the desired ingroup, *Micromeria*. The absolute and relative numbers of loci excluded at each step decreased, indicating a slight enrichment effect towards primer pairs amplifying only orthologous sequences during the procedure. From the 54 primers amplifying at least two samples, 22 markers which did not show obvious signatures of duplication were selected during this process for comparative sequencing of multiple ingroup samples in the last step.

Next to the scan for duplication signal, in different steps of marker development as used here and in other studies (e.g., Li et al., 2007), diagnostic characters or sequence analysis can be used to avoid the inclusion of duplicated loci in a phylogenetic analysis. If diagnostic characters are identified to distinguish different paralogs they can not only be used to construct specific primers, as described above, but also to differentiate orthologous copies after amplification, for example, if length differences between paralogs exist that are large enough to be detected in agarose gel electrophoresis. In such a case, both bands can be gel extracted and analyzed independently (Thomson et al., 2008). However, in gene families two or more paralogs could comprise the diagnostic character and could be co-isolated.

We tested the sequence information of the 22 markers selected by focusing on excessive heterozygous signals in the chromatograms and by a phylogenetic approach. Of these markers only three proved to be unsuitable for comparative analysis. An obvious problem with a phylogenetic approach is if reticulation or interspecific geneflow is studied, phylogenetic signals that are incongruent to the majority of the markers can contain useful and necessary information. Using incongruence to exclude certain markers will a priori bias the dataset towards a certain hypothesis. We therefore considered also the markers that showed a deviating phylogeny from the majority of the markers as suitable.

Intron containing markers present higher variability than coding sequences because in the latter only the synonymous positions are not conserved. Therefore, and especially in closely related species where alignment can be done unambiguously, intron markers will contain more information (Sang, 2002). For our target group, *Micromeria* from the Canary Islands, the markers were reasonably variable for phylogenetic analysis: levels of variation were in all cases higher than comparable chloroplast sequences (Meimberg et al., 2006) and showed between 4 and 14 times greater variability. Variability level of the intron markers is therefore higher than some of the most variable regions in the chloroplast DNA. Phylogenetic analysis showed for the majority of the markers congruent phylogenetic hypotheses. Consequently, the combined analysis resulted in a strongly supported phylogeny. An east-west split, a close phylogenetic relationship between the eastern islands, Gran Canaria and Lanzarote and between the western islands, Tenerife and Madeira was supported by the majority of the markers. For *Micromeria*, this east–west split was already described using chloroplast and nuclear markers (Meimberg et al., 2006) and could be a consequence of the direction of the archipelago's colonization. This colonization route (from east to west) might be quite common and it has been suggested that it applies to about 25% of plant groups studied so far in the Canary Islands (Caujapé-Castells, 2011). The same pattern was found for other systems such as Gonosperminae (Francisco-Ortega et al., 2001) and *Olea europaea* L. (García-Verdugo et al., 2009). Furthermore, Caujapé-Castells (2011) describes a decrease of genetic diversity from east to west for several species from the Canary Islands which is most likely a consequence of an expansion starting in the eastern most islands. This colonization route is facilitated by the proximity of Lanzarote and Fuerteventura to the continent (less than 100 km) and by the fact that these islands are the oldest of the archipelago. However, even though an east–west split for *Micromeria* is supported by the majority of the markers, it is contradicted by some of them. In systems influenced by reticulation, such contradicting phylogenies could

indicate a signature for adaptive introgression or past hybridization and are worthy of more intense study. Future work will concentrate on this topic.

Conclusion

Multi-locus analysis using nuclear gene derived markers is of high importance for phylogenetic analysis of plants. With our approach we were able to develop a large number of suitable EPIC markers for our group of interest that are also applicable for other genera in the Nepetoideae. In total, we identified 33 primers that are able to amplify a specific product from four out of the five genera investigated, which can also be used as starting point for investigations of related genera, such as *Thymus*, *Satureja* L. or *Rosmarinus*. Using the sequence information generated, we are able to further refine our primers to achieve a higher success rate in future studies, including the remaining loci with poor amplification success. We are using the generated markers not only for sequence analysis but also to design genotyping markers to measure the state of indels in the loci for determination of multi-locus genotypes. Primers with a 100% success rate can then be easily designed from this initial sequence information.

Acknowledgments

We would like to thank John MacAdams from Ervital and Christian Bräuchler from LMU Munich for providing samples. Pedro Luis Pérez de Paz from the University of La Laguna helped during field trip in the islands. The research is funded by FCT (Fundação para a Ciência e Tecnologia, Portugal) with the projectsPTDC/BIA-BEC/108866/2008 to HM and a PhD fellowship, SFRH/BD/74747/2010 to PP.

References

Adams, K.L., Wendel, J.F., 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. 8, 135–141.

Alexander, P.J., Rajanikanth, G., Bacon, C., Bailey, D., 2006. Recovery of plant DNA using a reciprocating saw and silica-based columns. Mol. Ecol. Notes 7, 5–9.

Backström, N., Fagerberg, S., Ellegren, H., 2008. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. Mol. Ecol. 17, 964–980.

Beltran, M., Jiggins, C.D., Bull, V., Linares, M., Mallet, J., McMillan, W.O., Bermingham, E., 2002. Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating Heliconius butterflies. Mol. Biol. Evol. 19, 2176–2190.

Bodt, S., Maere, S., Van de Peer, Y., 2005. Genome duplication and the origin of angiosperms. Trends Ecol. Evol. 20, 591–597.

Bräuchler, C., Meimberg, H., Abele, T., Heubl, G., 2005. Polyphyly of the genus Micromeria (Lamiaceae): evidence from cpDNA sequence data. Taxon 54, 639–650.

Bräuchler, C., Meimberg, H., Heubl, G., 2010. Molecular phylogeny of Menthinae (Lamiaceae, Nepetoideae, Mentheae) – taxonomy, biogeography and conflicts. Mol. Phylogen. Evol. 55, 501–523.

Caujapé-Castells, J., 2011. Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. In: Bramwell, D., Caujapé-Castells, J. (Eds.), The Biology of Islands Floras. Cambridge University Press, Cambridge, pp. 284–324.

Comai, L., 2005. The advantages and disadvantages of being polyploid. Nat. Rev. Gen. 6, 836–846.

Choi, H.-K., Luckow, M.A., Doyle, J., Cook, D., 2006. Development of nuclear genederived molecular markers linked to legume genetic maps. Mol. Gen. Genom. 276, 56–70.

Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., Leebens-Mack, J., dePamphilis, C.W., 2010. Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol. Biol. 10, 61–79.

Edwards, S., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19. Farris, J.S., 1989. The retention index and homoplasy excess. Syst. Zool. 38, 406–407.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using bootstrap. Evolution 39, 783–791.

Francisco-Ortega, J., Barber, J.C., Santos-Guerra, A., Febles-Hernández, R., Jansen, R.K., 2001. Origin and evolution of the endemic genera of Gonosperminae (Asteraceae: Anthemideae) from the Canary Islands: evidence from nucleotide sequences of the internal transcribed spacers of the nuclear ribosomal DNA. Am. J. Bot. 88, 161–169.

García-Verdugo, C., Fay, M.F., Granado-Yela, C., Rubio de Casas, R., Balaguer, L., Besnard, G., Vargas, P., 2009. Genetic diversity and differentiation processes in the ploidy series of Olea europaea L.: a multiscale approach from subspecies to insular populations. Mol. Ecol. 18, 454–467.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Sympos. Ser. 41, 95–98.

Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18, 1–32.

Li, C., Ortí, G., Zhang, G., Lu, G., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 7, 44.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Meimberg, H., Abele, T., Bräuchler, C., McKay, J.K., Pérez de Paz, P.L., Heubl, G., 2006. Molecular evidence for adaptive radiation of Micromeria Benth. (Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data. Mol. Phylogen. Evol. 41, 566–578.

Meimberg, H., Milan, N.F., Karatassiou, M., Espeland, E.K., McKay, J.K., Rice, K.J., 2010. Patterns of introduction and adaptation during the invasion of Aegilops triuncialis (Poaceae) into Californian serpentine soils. Mol. Ecol. 19, 5308–5319.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryderk, O.A., O'Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.

Pérez de Paz, P.L., 1978. Revisión del género Micromeria Bentham (Lamiaceae-Stachyoideae) en la Región Macaronésica. Instituto de Estudios Canarios. Monografías, 16, 306, p. +32 Pl.

Peters, J.,McCracken, K.G., Zhuravlev, Y., Lu, Y.,Wilson, R., Johnson, K., Omland, K., 2005. Phylogenetics of wigeons and allies (Anatidae: Anas): the importance of sampling multiple loci andmultiple individuals.Mol. Phylogen. Evol. 35, 209–224.

Rowe, K.C., Reno, M.L., Richmond, D.M., Adkins, R.M., Steppan, S.J., 2008. Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): multilocus systematics of the old endemic rodents (Muroidea: Murinae). Mol. Phylogen. Evol. 47, 84–101.

Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37, 121–147.

Seehausen, O., Koetsier, E., Schneider, M.V., Chapman, L.J., Chapman, C.A., Knight, M.E., Turner, G.F., Van Alphen, J.J.M., Bills, R., 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species Xock. Proc. Roy. Soc. Lond. Ser. B: Biol. Sci. 270, 129–137.

Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. Aust. Syst. Bot. 17, 145–170.

Swofford, D.L., 2002. PAUP⁄. Phylogenetic Analysis Using Parsimony (and other methods). Version 4.0 Sinauer Associates, Sunderland, Massachusetts.

Thomson, R.C., Shedlock, A.M., Edwards, S.V., Shaffer, H.B., 2008. Developing markers for multilocus phylogenetics in non-model organisms: a test case with turtles. Mol. Phylogen. Evol. 49, 514–525.

Thomson, R.C., 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. Mol. Ecol. 2184–2195.

Townsend, T.M., Alegre, R.E., Kelley, S.T., Wiens, J.J., Reeder, T.W., 2008. Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. Mol. Phylogen. Evol. 47, 129–142.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucl. Acids Res. 35, 71–74.

Wendel, J.F., Doyle, J.J., 1998. Phylogenetics incongruence: window into genome history and molecular evolution. In: Soltis, et al. (Eds.), Molecular Systematics of Plants II, pp. 265–296.

Whittall, J.B., Medina-marino, A., Zimmer, E.A., Hodges, S.A., 2006. Generating single-copy nuclear gene data for a recent adaptive radiation. Mol. Phylogen. Evol. 39, 124–134.

Wu, F., Mueller, L., Crouzillat, D., Pétiard, V., Tanksley, S.D., 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174, 1407–1420.

2.3. Evaluation of microsatellites of Catha edulis (qat; Celastraceae) identified using pyrosequencing

Manuel A. Curto [1, 2], Luke R. Tembrock [2], Pamela Puppo [1], Marcos Nogueira [1], Mark P. Simmons [2], Harald Meimberg [1, 3]

[1] CIBIO, Research Center for Biodiversity and Genetic Resources, University of Porto, Campus Vairão, R. Monte-Crasto, 4485-661 Vairão, Portugal

[2] Colorado State University, Department of Biology, 1878 Campus Delivery, Fort Collins, CO, 80523-1878, USA

[3] Restoration Ecology, Technical University Munich, Emil Raman Str. 6, 85350 Freising, Germany

Abstract

The use of Next Generation Sequencing (NGS) techniques to identify microsatellite markers has replaced more time intensive methods such as molecular cloning. The main advantage of NGS over traditional methods of identifying microsatellite markers is the generation of many more sequences with less effort. It is possible to design primers from unenriched DNA, thereby further reducing the workload and also allowing the use of SSRs that are difficult to enrich (e.g., TA/AT and TAA/ATT). We present microsatellite primer pairs that may be used for phylogeographic analysis as well as to infer the geographical origin of traded material of Catha edulis, which contains two amphetamines that are controlled substances in many counties. We used data from two partial 454 pyrosequencing runs that generated about 2000 sequences containing microsatellites (3% of all sequences) as well as flanking regions sufficient for primer design. Using 23 samples of C. edulis we identified 27 single-copy markers that were broadly amplified across the sampled individuals; 18 showed polymorphism information content (PIC) higher than 0.5. The genetic structure in wild individuals is concordant with their geographic origins; wild samples from northern Kenya are more closely related to Ethiopian samples than are other wild samples from Kenya. The geographic differences in allele frequencies indicate that microsatellite analysis can be used to determine the geographic source of cultivated and wild collected material.

Keywords: Catha edulis; Cultivation; Khat; Phylogeography; Qat

Introduction

Catha edulis (Vahl) Forssk. ex Endl., commonly known as khat or qat, is a species of evergreen shrubs and trees from the Celastraceae. It grows naturally at altitudes between 1000 m and 2500 m in eastern Africa (Robson, 1966) and is distributed from Ethiopia to the Cape Province in South Africa. Qat is also found in Yemen, although it is not clear if it is native there (Friis, 1992).

Qat has variously been considered to be a “social plague” (Al-Thani, 1983) or as a “custom [that] gives positive results in the daily gatherings for cultural, historical and literary debate” (Morghem and Rufat, 1983). Unlike alcohol, use of qat is not explicitly prohibited by Sharia (Islamic religious law; Kennedy, 1987). Some Muslims have considered qat to be a holy gift that allows them to pray throughout the night (Trimingham, 1965), whereas some Islamic scholars have interpreted it to be banned under Sharia (Al-Ghdaian, 1983).

Qat is openly traded in Djibouti, Ethiopia, Kenya, Somalia, Uganda, and Yemen but it is illegal in Eritrea, Kuwait, Saudi Arabia, Sudan, Tanzania, and Zambia (Gerstle, 2007). Qat is now “… an integral part of the Somali culture” (Odenwald et al., 2005). Somalia has perhaps the highest percentage of qat users in the world; most Somali qat is imported from Kenya (Perlez, 1992).

Qat use has recently expanded to Europe and North America based on the demand created by Ethiopian, Somali, and Yemeni immigrants (Browne, 1991). Recently it has been used by non-immigrants in Europe (Gebissa, 2010). The demand is supplied by qat flown in from Ethiopia, Kenya, and Yemen (Gough and Cookson, 1984). Qat is still legal in Great Britain (Elliott, 2006), but it is a controlled substance in most other countries in Europe, Canada, and the U.S. (DEA, 2006).

Qat contains three stimulant alkaloids that are structurally similar to amphetamine: norephedrine, cathine (Wolfes, 1930) and cathinone (UN Narcotics Laboratory, 1975). These alkaloids are concentrated in the young leaves and stems of qat. Cathinone is unstable and is only present in young leaves and stems; it is enzymatically reduced into cathine as the leaves mature or dry after picking (Schorno and Steinegger, 1979). Norephedrine, cathine, and cathinone have been isolated from only one other plant genus: Ephedra (Grue-Sorensen and Spenser, 1994). In the U.S., cathinone is a schedule I controlled substance and cathine is a schedule IV controlled substance (Department of Justice, 1988, 1993).

Because of its economic and social impact, qat is an interesting model for phylogeographic studies. In particular, the different ways of cultivation, by transplanting wild individuals or taking seeds or cuttings from cultivated individuals, allows us to test the effect of these cultivation activities on the genetic structure of the species. Given the

widespread use of cultivated qat, investigations on genetic structure might also enable identification of the source of traded material. So far, there is only one study published focusing on qat phylogeography and population genetics, which only sampled cultivated Yemeni qat for RAPDs without broader conclusions over the entire range of the species (Al-Thobhani et al., 2008). Questions that may be addressed by sampling both cultivated and wild individuals across the range of qat include: where and how many times qat was brought into cultivation, the relationships between named cultivated varieties, gene-flow patterns within and between cultivated and wild qat, and the preferred means of propagating cultivated qat.

Microsatellite markers have been one of the preferred markers employed in phylogeographic and population genetics studies (Sunnucks, 2000; Haasl and Payseur, 2011). In non-model species, development of microsatellites generally requires the establishment of genomic resources, traditionally in the form of a genomic library, enriched for certain SSR (short sequence repeats) motifs, which is used for sequencing and tagging of the SSR and the flanking primer sites (Zane et al., 2002). With Next Generation Sequencing (NGS), the construction of a genomic library and sequencing of individual colonies is replaced by sequencing fragmented genomic DNA. The NGS technology most commonly used for SSR discovery is pyrosequencing as implemented by the 454 GS-FLX platform (Roche, Branford, CT, U.S.A.).

Since 2009 several publications have described marker-development approaches using this technology (e.g., Santana et al., 2009) and have shown a significant improvement in the time, cost and amount of markers compared to traditional methodologies. The general strategy using pyrosequencing consists of DNA fragmentation, sequencing the fragments, and retaining sequences containing SSR motifs with flanking regions on both sides that are long enough for primer design. In this strategy an enrichment step is frequently added, using SSR-motif-specific probes (Santana et al., 2009; Jennings et al., 2011; Malausa et al., 2011). This approach can be advantageous because it increases the number of sequences containing SSRs but it also limits the diversity of microsatellite types found.

In this paper we present the results of a study designed to produce a set of polymorphic SSR markers for phylogeographic and forensic studies on qat. We used a 454-shotgun-genomic-sequencing approach without an enrichment step. The resulting loci were tested for allelic richness and population structure using multiple samples from both Ethiopian and Kenyan populations. In particular we compared the performance of SSRs containing AT/TA vs. TAA/ATT SSRs. These SSRs are difficult to use as targets for traditional enrichment methods and therefore they are generally not considered

during marker development despite being very frequent microsatellites in most plant species (Ellegren, 2004).

Materials and methods

Samples used and DNA isolation

We used one sample of C. edulis for 454 shotgun sequencing: an individual of uncertain provenance collected from a tree that is part of the United States Botanic Garden collection in Washington D.C. (accession 94-0091). The DNA from this sample was isolated following Alexander et al. (2006) method.

Table 1. List of 23 qat samples. The population number was defined according to location where the samples were collected.

Population	Accession	Voucher	Country	Latitude	Longitude	Altitude (ft)	Status
1	2083	ETHa	Ethiopia	N5.22932	E39.61806	4743	Wild
1	2085	ETHa	Ethiopia	N5.22995	E39.61503	4981	Wild
1	2088		Ethiopia	N5.22935	E39.61574	4958	Wild
1	2095		Ethiopia	N5.22811	E39.61510	4950	Wild
1	2111		Ethiopia	N5.22546	E39.60992	4988	Wild
1	2119		Ethiopia	N5.22828	E39.61129	5051	Wild
2	2203		Ethiopia	N7.29380	E36.47435	6766	Wild or feral
3	2296	ETHa	Ethiopia	N11.28550	E39.83878	7176	Feral
4	2314		Ethiopia	N6.38444	E37.71507	4861	Wild
5	2348	ETHa	Ethiopia	N6.37552	E36.73054	5133	Wild
6	2389		Kenya	S0.70634	E37.15706	4042	Wild
6	2390		Kenya	S0.70644	E37.15696	4430	Wild
6	2391	EAb	Kenya	S0.70644	E37.15693	4422	Wild
6	2392	EAb	Kenya	S0.70654	E37.15699	4420	Wild
6	2394		Kenya	S0.70409	E37.15348	4478	Wild
6	2396		Kenya	S0.70406	E37.15369	4461	Wild
7	2434a	EAb	Kenya	N1.222	E36.555	7961	Cultivated from wild
8	2541		Kenya	S0.22954	E37.75226	4391	Wild
9	2622	EAb	Kenya	S2.52772	E37.78793	4387	Wild
10	2657a	EAb	Kenya	S3.409	E38.364	4340	Cultivated from Ethiopia
11	2776		Kenya	N1.07331	E34.82217	6166	Wild
12	2823		Kenya	N0.361	E34.724	4700	Cultivated from Tanzania
13	2945b	EAb	Kenya	S0.286	E36.117	6184	Cultivated from Yemen

Microsatellite variability was tested using 23 qat samples collected from five locations in Ethiopia and eight locations in Kenya (Fig.S1 in supplementary material, Table 1). Of these 23 samples four are vouchered at the National Herbarium of Addis Ababa University (ETH) and six are vouchered at the National Museums of Kenya (EA; Table 1). Six individuals were sampled from population 1 in Ethiopia and six individuals

were sampled from population 6 in Kenya; we included a single sample from each of the other populations. This strategy was applied to broadly sample genetic diversity of the Ethiopian and Kenyan populations as well as check for variability within individual populations. Four samples were taken from cultivated plants in Kenya (albeit reported to have been originally obtained from Ethiopia, Kenya, Tanzania, or Yemen, respectively) and one feral specimen was sampled from Ethiopia (Table 1). Groups of plants that were originally cultivated but abandoned by the date of sampling were considered feral. The DNA from all 23 of the qat samples was isolated according the protocol of Alexander et al. (2006) with the following modifications: extraction and lyses buffers were formulated according to De La Cruz et al. (1995), and ascorbic acid and DIECA (Diethyldithiocarbamic acid sodium salt) were included in the extraction buffer following Couch and Fritz (1990).

Pyrosequencing

The shotgun sequencing (Anderson, 1981), including library preparations, was conducted in a 454 GS FLX platform as a service provided by Microsynth (Balgach, Switzerland). For library construction 1 µg of DNA was used and nebulized to an estimated average fragment length of 400 bp. Sequences were obtained by two independent 1/16th picotitre-plate runs.

SSR discovery, primer construction and testing

The sequences obtained were screened for the presence of microsatellite motifs using a text editor (UltraEdit Professional, v16, Fa. IDM Computer Solutions, Inc. U.S.A.) or Sciroko ver. 3.4 (Kofler et al., 2007). Sequences were filtered in three steps using the default values of Sciroko, which require a minimum score of 15 (i.e., microsatellites containing at least 15 bp of perfect repeats). First, all sequences that Sciroko retained with a minimum score of 15 were reported (Sciroko fraction). Second, all sequences that contained imperfect or interrupted repeats were discarded (perfect-repeat fraction). Third, all sequences containing less than 6 repeats were excluded (6-rep. fraction). Fourth, only sequences ≥ 200 bp long and with flanking regions of ≥30 bp long on either end of the microsatellite were considered for primer design (optimal fraction).

Primers were designed using Primer3Plus ver. 1.1.0 (Untergasser et al., 2007) under manual control using the following settings: optimal melting temperature of 60 °C, a GC content ranging from 47 to 53%, an optimal oligo length between 19 and 21 bp, and the amplification product between 150 and 300 bp.

For capillary-electrophoresis genotyping each primer was tagged at the 5′ end with one of four different universal primers using the M13-tailed primer method (Oetting

et al., 1995). The four universal primers tagged with different dyes were: 6-FAM (TGTAAAACGACGGCCAGT), VIC (TAATACGACTCACTATAGGG), NED (TTTCCCAGTCACGACGTTG), and PET (GATAACAATTTCACACAGG). The sequences of these primers were chosen to complement the sequence that was added to the 5′ end of the forward SSR primer (Godinho et al., 2011). The primers were tested in two steps: the single qat sample used for microsatellite discovery, and an additional 23 qat samples (Table 1).

All primer pairs were tested using QIAGEN Multiplex PCR Master Mix (Qiagen, Valencia, CA, U.S.A.) in a 10 µL singleplex reaction containing 5 µL of QIAGEN Multiplex PCR Master Mix, 3.3 µL of water, 0.5 µL of DNA (about 40 ng/µL), 0.4 µL of each primer with 10 mM reverse and the universal florescent primer and 1 mM forward primer using the temperature profile described below.

Primers that worked in the first step were multiplexed in a single PCR reaction of four markers that each incorporated a unique dye during amplification. The multiplex primer combination consisted of 4 nmol of each forward primer, 40 nmol of each reverse primer and the florescent universal primer. The 10 µL PCR reactions contained 5 µL of QIAGEN Multiplex PCR Master Mix, 1 µL of primer mix and 0.5 µL of template DNA (about 40 ng/µL). The thermolcycler conditions for the PCR reactions was: initial denaturation/activation step of 15 min; 95 °C for 30 s; touchdown starting at 58 °C to 55 °C, decreasing 0.5 °C per cycle for 45 s; extension at 72 °C for 30 s; followed by 25 cycles at 55 °C and eight cycles at 54 °C. Amplification success was checked using 2% agarose gels. Genotyping was performed with an internal size standard (Genescan-500 LIZ; Applied Biosystems, Inc., Foster City, CA, U.S.A.) in an ABI3130xl automatic sequencer (Applied Biosystems, Inc.). Alleles were called using GeneMapper ver. 4.0 (Applied Biosystems, Inc.).

Variability analyses

Variability of the SSRs was assessed by counting the number of alleles per locus and by estimating the PIC (polymorphism information content) and expected heterozygosity (He) using Cervus ver. 3.0.3 (Kalinowski et al., 2007).

To check if the set of SSRs obtained were informative to discriminate geographical patterns we tested whether genetic structure reflects the geographical distribution of the samples by using STRUCTURE ver. 2.3.3 (Hubisz et al., 2009) without prior identification of populations. We allowed the number of clusters to range from K = 2 to K = 9. For each K, we ran the program for 1,000,000 MCMC generations after an initial burn-in of 100,000 iterations. The most likely K was chosen by analyzing the variation of log probability of the data (ΔK) between successive K values (Evanno et al.,

2005). To confirm how the variability is distributed across the clusters defined by STRUCTURE, and which SSRs contributed the most to this pattern, we performed a locus-by-locus AMOVA by using Arlequin ver. 3.11 (Excoffier et al., 2005).

Results

SSR filtering

Table 2. Number of sequences retained after each of the selection steps. The Sciroko fraction contains all sequences comprising any kind of SSR motif including interrupted repeats. This fraction was analyzed for sequences only containing perfect repeats, at least six perfect repeats and sequences that also contained flanking regions longer than 30 bp (optimal). From the optimal fraction we designed 63 primer pairs (synthesized primers) from which 43 show successful amplification.

SSR	Sciroko	Perfect	6 Repeat	Optimal	Synthesized primers	Successful amplification
Mononucleotide	2911	353	353	159	0	0
Dinucleotide	4545	1895	1895	642	34	25
TA/AT	3149	1341	1341	456	9	6
CT/GA	1127	441	441	138	19	13
GT/CA	266	112	112	48	6	6
CG/GC	3	1	1	0	0	0
Trinucleotide	1303	551	275	199	26	19
AAG/CTT	292	125	62	43	6	5
CCA/TGG	71	25	11	10	0	0
CAA/TTG	146	57	12	10	2	2
ATT/AAT	468	207	128	96	13	9
GCT/ACG	14	1	0	3	0	0
GGA/TCC	104	44	23	16	3	2
GTA/CAT	94	49	20	14	2	1
GTC/CAG	28	17	6	2	0	0
GAT/CTA	81	24	13	5	0	0
GGC/CCG	5	2	0	0	0	0
Tetranucleotide	771	377	21	18	3	2
Pentanucleotide	1048	540	7	4	0	0
Hexanucleotide	1328	241	4	3	0	0
Total	11906	3957	2555	1025	63	46

The two independent 1/16th picotitre plate runs on the 454 platform resulted in 65,401 sequences from which 60,678 high quality sequences were considered in the subsequent analyses. From these, 11,906 sequences contained SSR motifs that were recognized in the Sciroko fraction (Table 2). Excluding mononucleotide SSRs, 8995 sequences (15% of all sequences) contained an SSR motif. Of these, 3604 sequences had perfect SSR motifs and 2202 were at least six repeats long. In the optimal fraction, 866 (1.4% of the 60,678 sequences) were retained for primer design based on the presence of ≥30 bp flanking regions on either side of dinucleotide or longer SSRs without ambiguous bases. Mononucleotide SSRs were present in 2911 of the SSR-containing sequences (24%) but only 12% of these (353 sequences) were perfect repeats.

Mononucleotide SSRs were present in a similar number of sequences as dinucleotide SSRs in the Sciroko fraction, yet only 159 of those sequences (compared to 642) were retained in the optimal fraction. Mononucleotide SSRs were not considered for primer design because of the high slippage error associated with these loci (Ellegren, 2004).

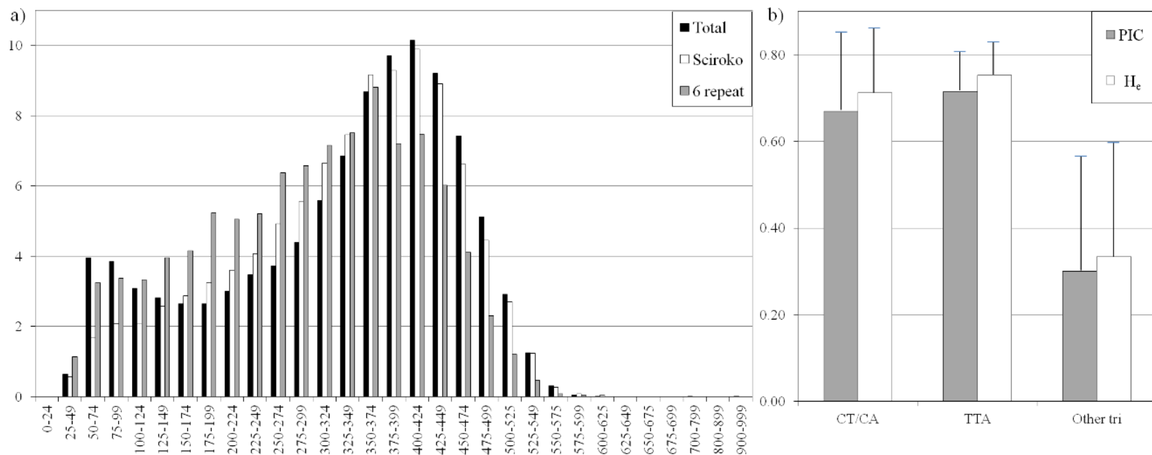


Fig. 1. a) Relative frequency distribution of sequence length for all 60,678 sequences compared to SSR containing sequences retained in the Sciroko fraction and sequences containing only perfect SSRs that are at least six repeats long. The frequency for each size category was calculated by calculating the percentage of sequences in with that particular in relation to the total number of sequences of that fraction. b) Comparison of average PIC and He for dinucleotide SSRs, TTA/AAT SSRs and all other trinucleotide SSRs. Error bars indicate standard deviation. Both values for non-TTA trinucleotides are significantly lower than for the TTA ones in the t-test (alpha = 0.05).

The length of all high quality sequences obtained after initial filtration ranged from 40 to 969 bp with an average of 329 bp. The sequences retained in the Sciroko fraction had a maximum and average length of 607 and 335 bp, respectively, while the sequences containing perfect SSRs (perfect-repeat fraction) with a minimum of six repeats had a maximum and average length of 595 and 294 bp, respectively (Fig. 1a). The average length of the sequences used for primer design (optimal fraction) was 354 bp. This increase in average length is a consequence of the elimination of sequences shorter than 200 bp. Dinucleotide repeats were the most frequent SSR type retained in the Sciroko fraction (4545 sequences). From these the most frequent SSR was TA/AT in 3149 sequences. About 14% of the dinucleotide-SSR-containing sequences retained in the Sciroko fraction were suitable for primer design and regarded as optimal by our filtering steps. Most of the sequences containing SSR motifs were excluded because they did not contain perfect repeats (only 33% of the Sciroko-fraction sequences were retained in the perfect fraction). In addition to motif type and the presence of perfect repeats, the number of repeat units was employed in filtering sequences for primer development. For tetra-, penta-, and hexa-nucleotide SSRs the lower bound for the

exclusion step was six repeats, which retained only 2.8% of the sequences in the perfect fraction. For mononucleotide and dinucleotide SSRs all sequences containing less than six repeats were excluded in the Sciroko fraction. In contrast to the tetra-, penta-, and hexanucleotide SSRs, the optimal fraction for mono- and di-nucleotide SSRs resulted in retaining only 45% and 34%, respectively, of the sequences in the 6-repeat fraction. The contrast in filtration percentages among the motif types is caused by the difference between the average length of mono- and di-nucleotide-SSR-containing-sequences (314 bp) versus tetra-, penta-, or hexanucleotide SSRs (381 bp). We attribute the shorter length of dinucleotide-containing sequences to the often terminal position of the SSR in the sequence: 33% of the perfect dinucleotide SSR sequences that included at least six repeats did not include downstream flanking regions longer than 30 bp, while only 3% of the dinucleotide-containing-sequences did not contain long enough upstream flanking regions. Upon completion of filtration a total of 866 sequences were found suitable for primer design. The 866 sequences consist of 642 dinucleotide SSRs, 199 trinucleotide SSRs, 159 mononucleotide SSRs, and 25 sequences in the tetra-, penta-, or hexanucleotide class of SSRs (Table 2).

Primer success and SSR variability

From the 866 optimal sequences with dinucleotide or longer SSRs we designed 63 primer pairs: 34 for dinucleotide SSRs, 26 for trinucleotide SSRs, and three for tetranucleotide SSRs (Table 2). The number of repeats in the SSRs varied from six to 21 (Table S1 in supplementary material). From the 63 primer pairs, 46 produced detectable PCR products from the single initial sample of qat (Table S1). In the second step we applied those 46 primer pairs to 23 samples of qat from a diverse geographic sampling. After the second step we excluded 14 primer pairs that failed to amplify bands from ≥ four samples as well as five primer pairs that may have amplified paralogous loci as indicated by fixed heterozygosity or amplification of more than two alleles in any sample. We retained 27 primer pairs to test for population structure (Table 3).

The number of alleles per locus varied between two and 16 (Table 3). The PIC value/expected heterozygosity (He) varied from 0.08/0.08 (CE39) to 0.89/0.90 (CE37; Table 3). Genetic diversity from the two populations that were each represented by six samples was higher for the Ethiopian population (He = 0.40) than for the Kenyan population (He = 0.29).

Table 3. Amplification, allele calling and variability of the final set of 27 loci.

Locus	SSR	Failed amplifications	Heterozygotes	Allele length variation	# alleles	PIC	H_e
CE3	$(CT)_{12}$	0	12	174-198	12	0,8	0,81
CE4	$(AAG)_{8}$	0	5	204-210	3	0,36	0,39
CE8	$(ATT)_{8}$	2	5	197-227	6	0,68	0,72
CE15	$(AG)_{9}$	0	15	170-183	3	0,39	0,46
CE16	$(AAG)_{8}$	0	8	242-258	4	0,28	0,3
CE22	$(AG)_{10}$	0	7	243-253	6	0,67	0,71
CE23	$(ATC)_{13}$	0	12	171-192	8	0,78	0,81
CE24	$(AC)_{11}$	0	11	216-224	5	0,53	0,58
CE29	$(AAT)_{7}$	0	4	246-258	5	0,6	0,65
CE30	$(AGG)_{7}$	0	3	245-254	3	0,33	0,36
CE31	(AC)9	0	4	189-199	4	0,51	0,6
CE33	$(CTT)_{7}$	0	4	215-233	2	0,33	0,42
CE34	$(CT)_{12}$	1	9	179-218	12	0,85	0,87
CE37	$(ATCT)_{11}$	0	12	208-276	14	0,89	0,9
CE39	$(AAC)_{6}$	0	2	171-185	3	0,08	0,08
CE40	$(GT)_{10}$	0	1	167-171	3	0,5	0,59
CE41	$(AG)_{15}$	0	9	225-265	16	0,83	0,84
CE42	$(GT)_{9}$	0	6	245-261	7	0,76	0,79
CE43	$(ATT)_{10}$	0	16	157-193	9	0,82	0,83
CE45	$(AG)_{11}$	0	10	268-290	8	0,81	0,83
CE47	$(GTT)_{8}$	0	1	255-264	3	0,12	0,12
CE50	$(CT)_{12}$	0	16	167-191	10	0,85	0,86
CE56	$(GA)_{6}$	0	4	208-218	3	0,48	0,56
CE57	$(TTA)_{13}$	0	8	194-209	6	0,72	0,75
CE58	$(TTA)_{13}$	0	3	174-180	3	0,49	0,57
CE59	$(TTA)_{15}$	1	8	191-213	10	0,81	0,83
CE64	$(CT)_{11}$	0	5	251-263	8	0,72	0,75

Population genetic structure

In the STRUCTURE analyses the most likely number of clusters was K = 7 (Ln P (D)= -1172). The first clear division (K = 2) is between a group composed of all samples from Ethiopia and four from Kenya (samples 2434a and 2776 from northern Kenya and samples 2657a and 2945b, which were originally obtained from Ethiopia and Yemen, respectively) and the other group consisting of the remaining Kenyan samples, with only two samples of >0.8 mixed affinity (Fig. 2). These STRUCTURE results were supported by AMOVA analyses, which indicated that 21.8% of variation is explained by differences between these two clusters.

For K = 7 most of the 23 samples were assigned to a single cluster with probability of 0.8 (Fig. 2). Overall, the STRUCTURE analysis differentiated all wild Ethiopian and Kenyan samples from each other. Of the cultivated Kenyan samples, 2657 and 2945b (originally obtained from Ethiopia and Yemen, respectively) cluster together with a feral sample from Ethiopia (2296).

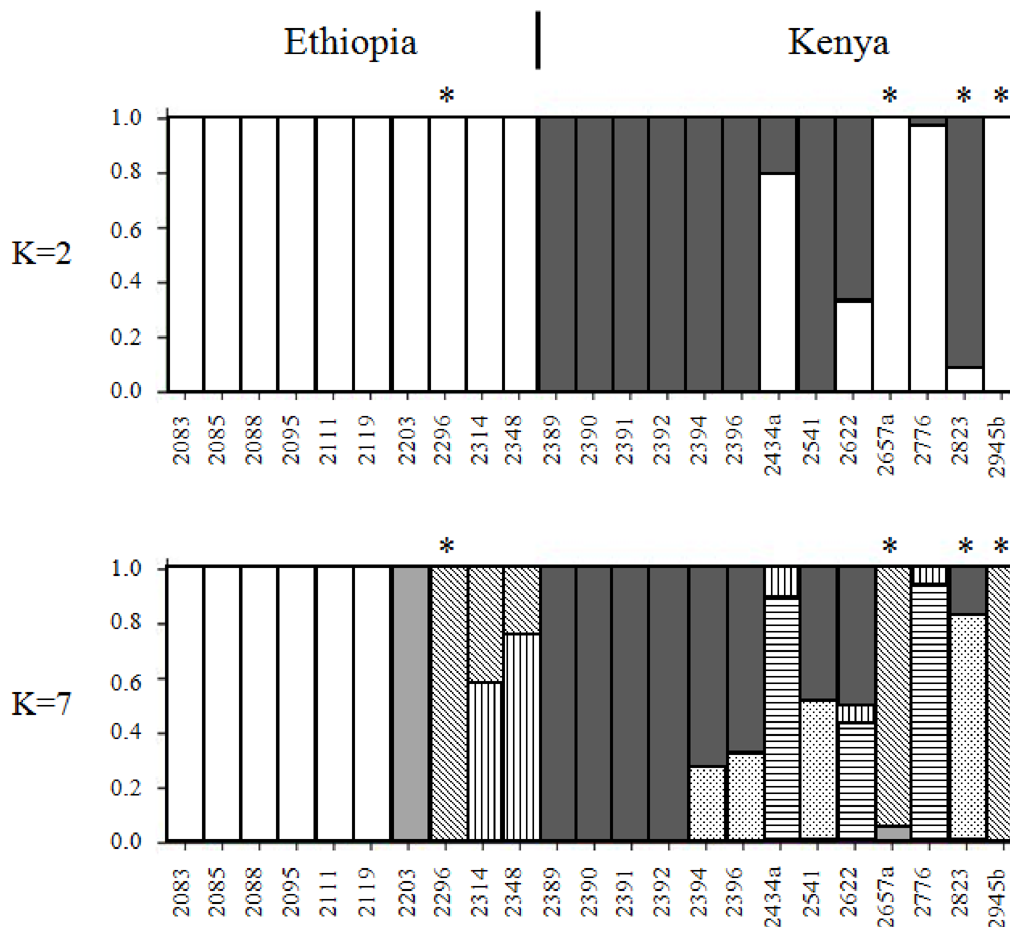


Fig. 2. STRUCTURE analysis of Ethiopian and Kenyan Catha edulis samples alternately assuming two (K = 2) or seven (K = 7) clusters. Asterisks (*) corresponds to cultivated or feral samples that were not directly transplanted from wild populations.

TA, TAA and TTAA SSRs

We designed 23 primers targeting microsatellite SSRs exclusively composed of A and T: 9 dinucleotides, 13 trinucleotides and one tetranucleotide. From the nine primer pairs designed for TA-dinucleotide SSRs, only four were able to successfully amplify at least 12 of our 23 qat samples. These included two primer pairs that amplified more than two alleles from some samples. This corresponds to a lower amplification success than the others dinucleotide SSRs (17 out of 25 amplified at least 12 samples). This indicates a slight bias against TA SSRs in amplification success. For trinucleotide SSRs, six of the 13 primers were successfully optimized irrespective of the motif. The TTA/AAT SSRs have a level of variability similar to all dinucleotide SSRs, both of which are significantly higher than the other trinucleotide SSRs (both comparisons are significantly different based on t-tests with alpha = 0.05; Fig. 1b). The TTA/AAT SSRs contribute more to the genetic structure than the other trinucleotide loci. An AMOVA of the two clusters indicated by STRUCTURE analysis for K = 2, using only wild samples, indicated that

44.7% of the TTA/AAT variation is explained by between-group differences, while the remaining trinucleotide SSRs explained only 13.2% of the variation. When the same analysis was performed using only the two populations composed of six samples similar results were obtained. For TAA/AAT SSRs, 63.5% of the variability is explained by differences between these two populations compared to 28.3% for the remaining trinucleotide SSRs.

Discussion

The use of Next Generation Sequencing (NGS) techniques to identify SSR markers is increasingly important and has become more efficient for microsatellite discovery than traditional methods. The most obvious advantage of NGS over older methods is the replacement of compiling genomic libraries with time-intensive methods by a NGS run. Additionally it is possible, when using NGS, to design primers from unenriched DNA, thereby further reducing the workload and also allowing the use of SSRs that are difficult to enrich (e.g., TA and TAA). Because of the lower structural stability of TA and TAA SSRs, they might be more variable and thus especially useful for phylogeographic studies with closely related individuals or populations.

In this study, we were able to obtain 866 sequences that contained six or more perfect repeats with flanking sequences longer than 30 bp, and therefore potentially suitable to design PCR primers. Compared with other studies using similar approaches, our 866 sequences is considered high. For example, Abdelkrim et al. (2009) and Csencsics et al. (2010) used total genomic DNA and they obtained 231 and 307 sequences containing microsatellites. This difference might be partly because sequence length obtained with NGS increased in recent years. In addition it has been shown that the fraction of these repetitive motifs varies according to organism (Ellegren, 2004). Even though many studies using an enrichment step present higher or similar numbers of sequences containing SSRs, we were able to obtain more SSRs than some studies using this strategy (Jennings et al., 2011). Other studies that used a larger portion of the picotitre 454 plate or an enrichment step obtained many more sequences with SSRs (e.g., Boomer and Stow, 2010 [4362 SSRs] and Malausa et al., 2011 [28,336 SSRs]. Nevertheless, using our simplified approach we were able to obtain a sufficient amount of SSR markers for phylogeographic studies in qat.

Jennings et al. (2011) used the Illumina platform (Bentley et al., 2008) for SSR development and identified 356,958 sequences containing SSRs with 11,650 sequences suitable for primer design. Even though the overall number of SSRs containing sequences good enough for primer design is high, only 3.3% contain SSRs, which is lower than in our study (8.6%), probably because of the shorter sequence length

generated by the Illumina platform. An alternative would be to use the Ion Torrent method (Rothberg et al., 2011), which produces average sequence lengths of 200 bp and was expected to have achieved 400 bp by 2012 (Life Technologies Corporation, 2011) which would be sufficient to obtain most of the 866 sequences in our optimal fraction while being more cost effective than our pyrosequencing approach (Glenn, 2011).

The average length of the sequences containing optimal-fraction SSRs obtained in our study was just 294 bp, which is shorter than the 329 bp average for all sequences obtained. One factor that caused this discrepancy was a drop in sequence quality downstream of dinucleotide SSRs. This drop in sequence quality resulted in sequences potentially useful for SSR primer design being shorter, on average, than the others as well as SSRs lacking sufficient flanking regions to design primer binding sites more frequent than expected by chance. Only 25% of dinucleotide containing sequences could be used for primer design even though 70% contained a sufficient number of perfect repeats. Similar to Sanger sequencing, emulsion PCR prior to pyrosequencing may create slippage artifacts that diminish sequence quality downstream of a SSR motif. This effect would be overcome by the use of paired-end reads on a given DNA strand, as may be obtained using the Illumina technology (Straub et al., 2012).

In our study we did not use an enrichment strategy to increase the amount of SSR-containing sequences but we were nevertheless able to identify a large number of SSRs. This way we were able to save time and money obtaining enough SSRs that can be used in subsequent studies. The success of designing PCR primers for SSRs depends not only on the total number of SSR-containing sequences generated, but also on the subsequent success of amplification or optimization and the information content from each locus. Unenriched libraries, which contain a relatively large number of TA/AT and TTA/AAT SSRs, may provide greater information content per SSR locus than studies that include an enrichment step for other SSRs, as was shown in our study.

We retained just 25 sequences containing tetra-, penta-, or hexa-nucleotide SSRs in our optimal fraction (Table 2). If these SSRs are targeted to increase genotyping performance then an enrichment step is advisable. Most of our SSR-containing sequences that are suitable for primer design correspond to TA/AT SSRs, the most frequent SSR in plants (Ellegren, 2004). Success in amplification was lower for TA/AT SSRs in comparison to the other SSR types and we were unsuccessful in developing any markers with this SSR in our study. Yet one advantage of our approach of using an unenriched library was the ability to identify TTA/AAT trinucleotide SSRs. In our study this SSR type showed information content comparable to dinucleotide SSRs but with the advantage of genotyping a trinucleotide SSR (i.e., less stutter).

We successfully developed 27 primer pairs that had high amplification success and do not show evidence of amplifying paralogous loci, which corresponds to 43% of the 63 primer pairs tested. We further tested the loci using the genetic-diversity measures of allele number, PIC and *He* to determine the amount of information that can be obtained from each locus. PIC and *He* are diversity measures that take into account not only the variability of each locus but also the frequency distribution of alleles across the 23 qat samples. Referencing these measures enables investigators to avoid polymorphic SSRs where all of the variability is found in a few samples, which is not very informative.

The STRUCTURE analyses successfully differentiated between many of the qat samples collected from different regions. Wild qat samples from Ethiopia and northern Kenya are differentiated from those collected in southern Kenya. In addition, one of the Kenyan samples (2823) was originally obtained from Tanzania and it is assigned to the same clusters as the samples from southern Kenya, albeit with different probabilities (Fig. 2). These results show that the approach of using microsatellites to differentiate qat collected from different regions is promising.

One of the clusters obtained by STRUCTURE with K = 7 consisted of sample 2296 from northern Ethiopia and the samples 2657 and 2945b that were cultivated in Kenya but originated from Ethiopia and Yemen, respectively. This clustering is consistent with the assertion that cultivated qat plants in Yemen originated from Ethiopia rather than Kenya. This finding demonstrates that these SSRs can be used in the future to clarify the origin of Yemeni qat.

Although only 23 samples of qat were included in this study we were able to obtain preliminary inferences concerning the phylogeographic patterns in qat. The use of many more samples from a wider geographic area and multiple samples per population will be necessary to make definitive conclusions about population structure. The SSR primer pairs that we developed have a high potential to be informative in future studies about phylogeographic patterns, genetic structure between cultivated and wild qat, as well as determining variation in genetic diversity across the native and cultivated distribution of the species. These SSR markers may also be used for forensic identification of illegally traded qat.

Acknowledgments

We thank members of the National Herbarium at Addis Ababa University and the East African Herbarium of the National Museums of Kenya for assistance with fieldwork as well as the Republic of Kenya National Council for Science and Technology for research authorization. M.A.C., P.P. and M.N. were supported by Ph.D. fellowships funded by Fundação para a Ciência e Tecnologia, Portugal with the references

SFRH/BD/79010/2011, SFRH/BD/74747/2010, and SFRH/BD/64225/2009, respectively.

References

Abdelkrim, J., Robertson, B.C., Stanton, J.L., Gemmell, N.J., 2009. Fast, cost- effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques 46, 185–192.

Alexander, P.J., Rajanikanth, G., Bacon, C.D., Bailey, C.D., 2006. Recovery of plant DNA using a reciprocating saw and silica-based columns. Mol. Ecol. Notes 7, 5–9.

Al-Ghdaian, A.A., 1983. Khat in the Shari'ah (Islamic Law), in: Shahandeh, B., Geadah, R., Tongue, A., Tongue, E., Rolli, J. (Eds.), The Health and Socio-Economics of Khat Use. International Council on Alcohol and Addictions, Lausanne, pp. 231-239.

Al-Thani, I.M., 1983. Development: the Saudi solution for the problem of khat, in: Shahandeh, B., Geadah, R., Tongue, A., Tongue, E., Rolli, J. (Eds.), The Health and Socio-Economics of Khat Use. International Council on Alcohol and Addictions, Lausanne, pp. 181-194.

Al-Thobhani, M.A.H., Sathyanarayana, B.N., Simon, L., Sondur, S.N., 2008. First comparative genotypic study on khat (*Catha edulis* Forsk.) genotypes from Yemen. Middle E. Russ. J. Plant Sci. Biotechnol. 2, 1–8.

Anderson, S., 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucleic Acids Res. 9, 3015-3027

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K. P., et al., 2008. Accurate whole genome sequencing using reversible terminator chemistry. Nature 456, 53–59.

Boomer, J., Stow, A., 2010. Rapid isolation of the first set of polymorphic microsatellite loci from the Australian gummy shark, *Mustelus antarcticus*, and their utility across divergent shark taxa. Conserv. Genet. Resour. 2, 393–395.

Browne, D.L., 1991. Qat use in New York City. NIDA Res. Monogr. 105, 464–465.

Couch, J.A., Fritz, P.J., 1990. Isolation of DNA from plants high in polyphenolics. Plant. Mol. Biol. Rep. 8, 8–12.

Csencsics, D., Brodbeck, S., Holderegger, R., 2010. Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. J. Hered. 101, 789–793.

De La Cruz, M., Whitkus, R., Mota-Bravo, L., 1995. Tropical tree DNA isolation and amplification. Mol. Ecol. 4, 787–789.

DEA (Drug Enforcement Agency), 2006. Khat: AKA: *Catha edulis*. http://www.usdoj.gov/dea/pubs/pressrel/pr072606a.html Accessed 14 February 2007.

Department of Justice, 1988. Schedules of controlled substances; temporary placement of cathine ((+)-norpseudoephedrine), fencamfamin, fenproporex and mefenorex into schedule IV. Federal Register 53, 17459–17460.

Department of Justice, 1993. Schedules of controlled substances: placement of cathinone and 2,5-Dimethoxy-4-ethylamphetamine into schedule I. Federal Register 58, 4316–4318.

Ellegren, H., 2004. Microsatellites: Simple sequences with complex evolution. Nat. Rev. Genet. 5, 435–445.

Elliott, J., 2006. Dramatic warning about khat misuse. BBC News. http://news.bbc.co.uk/2/hi/uk_news/4720262.stm.

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

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol. Bioinform. 1, 47–50.

Friis, I., 1992. Forests and forest trees of northeast tropical Africa: their natural habits and distribution patterns in Ethiopia, Djibouti and Somalia. London: Her Majesty's Stationery Office.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Resour. 11, 759–769.

Godinho, R., Blanco, J.C., Lopes, S., Álvares, F., García, E.J., Palacios V., Cortés, Y., Talegón, J., Ferrand, N., 2011. Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. Mol. Ecol. 20, 5154–5166.

Gebissa, E., 2010. Khat in the Horn of Africa: Historical perspectives and current trends. J. Ethnoparmacol. 132, 607–614.

Gerstle, D.J., 2007. Taming the African Horn's khat trade. SIPA News 2007, 27–29.

Gough, S.P., and Cookson, I. B., 1984. Khat-induced schizophreniform psychosis in UK. Lancet 8373, 455.

Grue-Sorensen, G., and Spenser, I.D., 1994. Biosynthetic route to the *Ephedra* alkaloids. J. Am. Chem. Soc. 116, 6195–6200.

Haasl, R.J., Payseur, B.A., 2011. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity 106, 158–171.

Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9, 1322–1332.

Jennings, T.N., Knaus, B.J., Mullins, T.D., Haig, S.M., 2011. Multiplexed microsatellite recovery using massively parallel sequencing. Mol. Ecol. Resour. 11, 1060–1067.

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16, 1099–1006.

Kennedy, J.G., 1987. The flower of paradise: the institutional use of the drug qat in North Yemen. Dordrecht, D. Reidel Publishing Company.

Kofler, R., Schlötterer, C., Lelley, T., 2007. Sciroko: a new tool for whole genome microsatellite search and investigation. Bioinformatics 231683–1685.

Life Technologies Corporation, 2011. Application note: Ion personal genome machine™ performance overview. http://ioncommunity.lifetechnologies.com/servlet/JiveServlet/downloadBody/1453-102-1-2586/performance_overview_application_note_041211.pdf

Malausa, T., Gilles, A., Meglécz, E., Blanquart, H., Duthoy, S., Costedoat, C., Dubut, V., Pech, N., Castagnone-Sereno, P., Délye, C., Feau, N., Frey, P., Gauthier, P., Guillemaud, T., Hazard, L., Corre, V., Lung-Escarmant, B., Malé, P.G., Ferrreira, S., Martin, J., 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. Mol. Ecol. Resour. 11, 638–644.

Morghem, M.M., and Rufat, M.I., 1983. Cultivation and chewing of khat in the Yemen Arab Republic, in: Shahandeh, B., Geadah, R., Tongue, A., Tongue, E., Rolli, J. (Eds.), The Health and Socio-Economics of Khat Use. International Council on Alcohol and Addictions, Lausanne, pp. 215-218.

Odenwald, M., Neuner, F., Schauer, M., Elbert, T., Catani, C., Lingenfelder, B., Hinkel, H., Häfner, H., Rockstroh, B., 2005. Khat use as a risk factor for psychotic disorders: a cross-sectional and case-control study in Somalia. BMC Med. 12, 3–5.

Oetting, W.S., Lee, H.K., Flanders, D.J., Wiesner, G.L., Sellers, T.A., King, R.A., 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30, 450–458.

Perlez, J., 1992. Guns, greed and khat define a once-graceful Somali city. The New York Times. http://www.nytimes.com/1992/12/07/world/mission-somalia-reporter-s-notebook-guns-greed-khat-define-once-graceful-somali.html

Robson, N.K.B., 1966. Celastraceae (incl. Hippocrateaceae). Flora Zambesiaca, ed., A. W. Exell, A. Fernandes and H. Wild. London, Crown Agents for Oversea Governments and Administrations 2, 355–418.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., et al., 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348–352.

Santana, Q.C., Coetzee, M.P.A., Steenkamp, E.T., Mlonyeni, O.X., Hammond, G.N.A., Wingfield, M.J., Wingfield, B.D., 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. DNA Seq. 16, 217–223.

Schorno, X., and Steinegger, E., 1979. CNS-active phenylpropylamines of *Catha edulis* FORSK. (Celastraceae) of Kenyan origin. Experientia 15, 572–574.

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot., 99, 349–364.

Sunnucks, P., 2000. Efficient genetic markers for population biology. Trends. Ecol. Evol. 15, 199–203.

Trimingham, J.S., 1965. Islam in Ethiopia. New York, Barnes & Noble, Inc.

United Nations Narcotics Laboratory. 1975. Studies on the chemical composition of khat: III. Investigations on the phenylalkylamine fraction. Geneva, United Nations Narcotics Laboratory.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucl. Acids Res. 35, 71–74.

Wolfes, O., 1930. Über das Vorkommen von d-nor-iso-Ephedrin in *Catha edulis*. Arch. Pharm. 268, 81–83.

Zane, L., Bargelloni, L., Patarnello, T., 2002. Strategies for microsatellite isolation: a review. Mol. Ecol. 11, 1–16.

2.4. Influence of past agricultural fragmentation to the genetic structure of *Juniperus oxycedrus* in a Mediterranean landscape

Manuel Curto[1,3], Marcos Nogueira[1,2], Pedro Beja[1], Francisco Amorim[1], Maximiliane Schümann[2,3] and Harald Meimberg[1,2,3]

[1]CIBIO, Research Center in Biodiversity and Genetic Resources, InBio Associate Laboratory, University of Porto, Campus Vairão, Vairão 4485-661, Portugal

[2]Chair of Restoration Ecology, Dept. Ecology and Ecosystem Management, Technische Universität München, Emil-Ramann-Straße 6, 85354 Freising, Germany

[3]Institute for Nature Conservation Research, University of Natural Resources and Life Sciences (Boku), Vienna A-1180, Austria.

Abstract

Habitat fragmentation can have a profound effect on the genetic diversity of forest species. These effects are especially interesting when forests previously fragmented by agriculture start to reconnect due to land abandonment. In this study, we investigate the genetic structure and diversity patterns of *Juniperus oxycedrus* populations from the Sabor river valley in Northeast Portugal. We developed 17 microsatellite markers using pyrosequencing technology as implemented in the 454 platform. As expected, among population differentiation was low with high variability within populations. There was no strong pattern of genetic structure in our analyses (F_{ST} = 0.018) suggesting that the individuals analyzed here belong to one population. The genetic structure seems to be equally explained by locality and by tree age. We hypothesize that this is a consequence of the land use history from the region. After the abandonment of cultivated fields these terrains were probably colonized by individuals from a few older *J. oxycedrus* populations. Thus the genetic structure pattern found may best be explained by this recent expansion. This expansion may be currently influenced by the construction of two hydroelectric dams that will flood areas with older individuals of the species.

Keywords: *Juniperus oxycedrus*, microsatellites, genetic diversity, land abandonment, habitat fragmentation, population structure

Introduction

Habitat fragmentation may strongly affect plant population genetics by resulting in decreased effective population sizes (Ellstrand and Elam 1993) and reduced gene

flow among populations (Schaal and Leverich 1996; Couvet 2002), thereby potentially causing inbreeding effects and the loss of genetic diversity (Keller and Waller 2002). Ultimately this can put species survival at risk and is therefore considered one of the major threats to biodiversity (Young et al. 1996). Fragmentation has a higher impact on organisms with low dispersal ability and organisms that are obligate outcrossing. Autogamous plants are expected to be little affected by fragmentation, while exclusively cross pollinated plants (self-incompatible dioecious plants) would be highly sensitive to fragmentation effects (Berge et al. 1998; Larson and Barrett 2000; Lennartsson 2002). In this group, wind pollinated plants are less affected than animal pollinated plants (Berge et al. 1998; Weidema et al. 2000).

The Mediterranean basin is a hotspot of biodiversity, despite the millennial influence of human populations (Cowling et al. 1996; Blondel et al. 2010). Since the introduction of agriculture, continuous expanses of natural forests were transformed into a mosaic of cultivated fields and forest patches. With time many species adapted to this kind of conditions, which allowed them to persist in highly fragmented cultural landscapes (Blondel et al. 2010). After the middle of the 20th century, with modernization and intensification of agriculture, cultivation of arable fields that were not suitable for high productivity and machinery have been progressively abandoned (Sluiter and de Jong 2007). This resulted in the recolonization of trees and shrubs from the surrounding areas which changed the habitats significantly (Chauchard et al. 2007). A number of studies have evaluated the effect of this development on the biodiversity of those regions (e.g. Sluiter and de Jong 2007; Porto et al. 2011; Santana et al. 2011). Nevertheless, past fragmentation and recolonization may also have an effect on the genetic diversity of tree and shrub populations especially, but to the best of our knowledge this has never been analyzed (but for comparable studies in other regions see Jacquemyn et al. 2009; Leite et al. 2014).

The Sabor region from Northeast Portugal is a good example of past fragmentation caused by agriculture (Hoelzer 2003). Human population and agriculture strongly expanded in the area from the mid-19th century to the 1950s. Then a process of human population decline and land abandonment started and it is ongoing until today. Until the mid-20th century almost all the area was cultivated, with remnants of native woody vegetation being confined to rock outcrops and the steepest slopes of the river valley. In the second half of the last century there was a progressive recovery of shrub land and forest vegetation that now covers vast areas across the region. A peculiarity of this region are two hydroelectric dams built between 2009 and 2013 which started flooding the valley in the winter of 2013/14. As a consequence, the dams will contribute to destroying at least some of the forest patches near the river that may be the origin of

the colonization of the abandoned farmlands. This makes our study of high priority because understanding the spatial genetic structures of populations destroyed by the dams will help to understand the impacts of these constructions in the future.

In this study we investigate the genetic structure of stands of *Juniperus oxycedrus* L. (prickly juniper, Cupressaceae). We expect that the structure patterns found would reflect the impact of the past fragmentation followed by land abandonment and subsequent expansion of woody species. The species was considered particularly suitable to study this process, because its expansion in the region is occurring through natural processes, whereas afforestation has facilitated the expansion of other dominant trees such as oaks *Quercus* spp. *Juniperus oxycedrus* are small trees or big bushes with a distribution from Portugal (West) to Iran (East) and from Morocco (South) to France (North) (Amaral 1986). It is a drought resistant, light demanding species that together with its low soil quality requirements makes it a common pioneer of areas affected by fire and deforestation (Bondi 1990; Cano et al. 2007). These characteristics make it more likely to establish quickly after abandonment of agricultural fields. Sluiter and de Jong (2007) found that when the use of agriculture fields stopped, pioneer vegetation occupies it immediately. J. oxycedrus is a dioecious and wind pollinated plant with low fertility (Ortiz et al. 1998). This means it may not be as affected by fragmentation effects like the other exclusive cross-pollinated plants, though its low reproductive success may make it more susceptible. *J. oxycedrus* has a low production of viable seeds (Juan et al. 2003), and studies with its close relatives *J. macrocarpa* and *J. communis* showed seed viability values of 12% and 3% (Juan et al. 2003; Verheyen et al. 2005). The seeds are mainly dispersed by small mammals and birds, making J. oxycedrus a long dispersal species (Flynn et al. 2006). In a *J. oxycedrus* population in Italy Baldoni et al. (2004) found that these trees achieve maturity between 17 to 21 years of age and the oldest individual was 83 years old.

Former studies on *Juniperus oxycedrus* had mainly a morphological (e.g. Klimko et al. 2007; Brus et al. 2011), phytosociological (e.g. Cano et al. 2007) and demographic (e.g. Baldoni et al. 2004) focus. The only molecular studies were made at the European scale, focusing on species and subspecies differentiation (e.g. Adams et al. 2005; Boratyński et al. 2014). Population genetics studies have been made mainly on other juniper species like *Juniperus communis* and *Juniperus macrocarpa*. Most of them used chloroplast (Provan et al. 2008; Juan et al. 2012) and dominant markers like AFLP and isozymes (Lewandowski et al. 1996; Van Der Merwe et al. 2000; Oostermeijer and Knegt 2004; Vanden-Broeck et al. 2011). Michalczyk et al. (2006) was able to develop five microsatellite loci for *J. communis* that were later used by Provan et al. (2008) in *J. communis* from populations from Ireland, and by Boratyński et al. (2014) for genetic

differentiation of several juniper species. Although the use of microsatellites is an improvement in relation to the other markers, using such a low number of markers is not powerful enough to detect structure in small geographical scales.

In this study we identified a set of microsatellite loci suitable to identify small scale genetic structure in *J. oxycedrus*. We then investigated genetic structure and diversity patterns at these loci in *J. oxycedrus* trees in the Sabor region to determine the main source of genetic structure in the area, in order to estimate signs of past and present impacts of the agriculture land use on the vegetation. Specifically, we aim to answer the following questions: 1) Is there a signature of past fragmentation in the current genetic structure of the population? 2) Does genetic structure depend on geographical distance? or 3) Does it depends rather on the age structure of the trees and thus reflect former recolonization pathways? The study is part of a Long Term Ecological Research (LTER) project designed to determine the impact of fragmentation caused by agriculture and infrastructure development. Monitoring of *J. oxycedrus* populations in this region will be made during the lifetime of the hydroelectric dams (about 65 years). This way it will be possible to observe the development of genetic structure after the construction event.

Material and methods

Study area

Field sampling was carried out in 2011 in the valley of the lower reaches of the Sabor River. Climate of the region is Mediterranean with a subcontinental character; in the river valley and slopes the mean annual temperature is 14-16ºC and annual precipitation is 400-600 mm, whereas in the surrounding plateau the mean annual temperature is 12-14ºC and the annual precipitation is 500-800 mm (Hoelzer 2003). The precipitation is concentrated in the wet and cold semester (October-March), and it is virtually nil in summer. The bedrock of the study area is dominated by schists, whereas the soil types prevailing are leptosols and anthrosols (Hoelzer 2003). Human population is low and concentrated in a few scattered villages, and the declining agriculture is dominated by olive and almond groves, and extensive livestock grazing (mainly sheep and goats). Natural vegetation is recovering after the peak of agricultural expansion in the 1950s, and include forest patches dominated by cork oak Quercus suber or holm oak Quercus rontundifolia, sometimes in combination with prickly juniper, tall scrubs often dominated by prickly juniper, and shrublands with species of, for instance, Cistus, Genista and Cytisus (Costa et al. 1998; Hoelzer 2003).

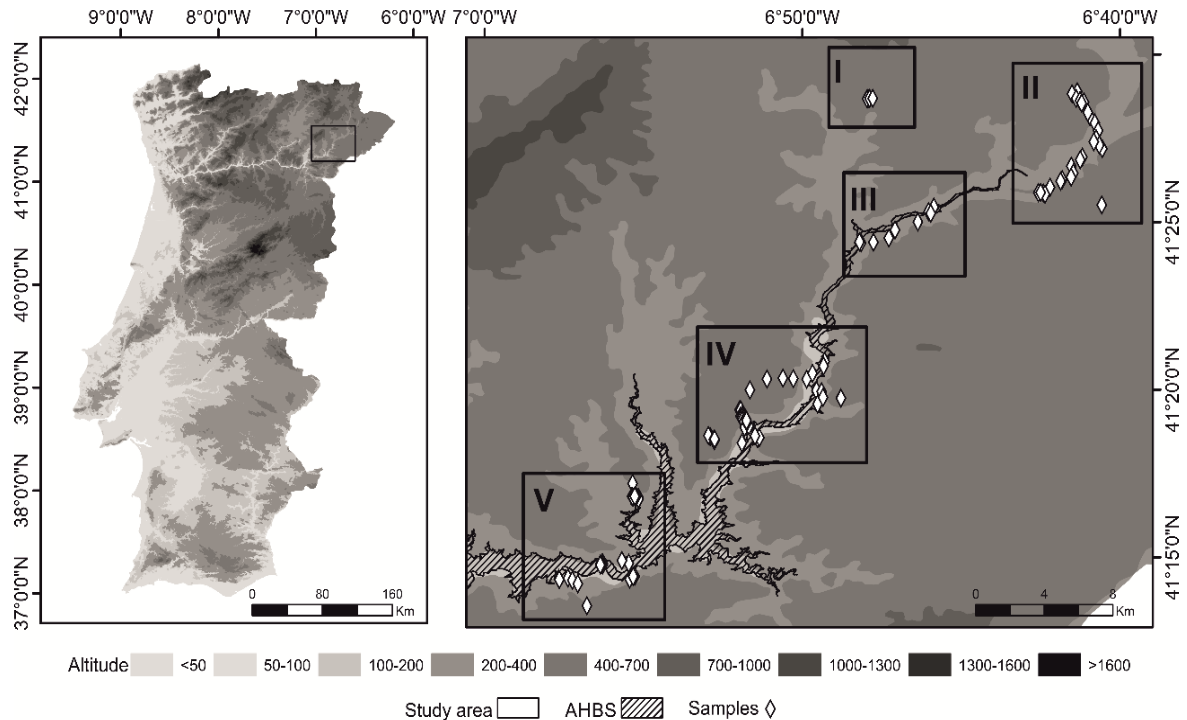


Fig. 1. Map of the study area, showing sampling localities and the area to be flooded by the Sabor dam. The background gray pallet corresponds to the altitudinal gradient. The rectangles correspond to the populations defined a priori.

Sampling strategy and DNA isolation

We collected leaves from 126 individuals corresponding to five populations defined *a priori* according to the geographic clustering of the individuals (Table S1; Fig. 1). This corresponds to a subset of the existing *J. oxycedrus* in the region, though it is not possible to estimate what portion of the total population was sampled because no census study is currently available. This study is part of a broader project aiming to study the influence of the construction of the two dams on the genetic diversity of *J. oxycedrus* in the region. Because of that, the sampling was made in a way to cover localities alongside the river on both sides of the dam construction sites (populations II to V). One population (I) was sampled to represent regions away from the river. We measured tree trunk perimeter for 55 individuals (Table 1), which was used as a proxy to determine the relation between tree age and genetic structure. Because of high vegetation density and other accessibility problems in some regions, the trunk perimeter could not be measured for all individuals sampled. Some of the unmeasured trees were estimated visually to have trunk perimeters > 100cm, and were thus classified as "> 100 cm" and included in all the analyses using trunk perimeter (Table 1; Table S1). A sample of *J. oxycedrus* from the UTAD (Universidade de Trás-os-Montes e Alto Douro) botanical garden was collected for marker discovery (Reference: D7D8J7).

For DNA isolation, leaves were stored and dried in silica gel. Twenty mg of leaf tissue was ground by stainless steel beads in two mL tubes using a Retsch Mill (MM400), at maximum force (30 Hz). The remaining procedure was performed using the protocol described by Alexander et al. (2006) using a CTAB based Lysis buffer and a 96 well format like described earlier (Curto et al. 2013).

Table 1. Number of individuals sampled for genetic analysis per a priori population. For each population, the table also shows the number of individuals measured for trunk perimeter, minimum and maximum measures, the average and standard deviation of the trunk perimeter, and the number of unmeasured individuals with trunk perimeter visually estimated to be > 100 cm.

Population	Total number of individuals	Individuals measured	Average trunk perimeter	Standard deviation	Minimum	Maximum	Unmeasured individuals > 100 cm
I	6	3	36.67	10.27	25	50	0
II	27	13	70.62	26.54	35	132	1
III	11	0	NA	NA	NA	NA	2
IV	48	28	71.83	30.95	25	133	1
V	34	11	37.73	23.49	15	100	6
Total	126	55	62.81	31.45	15	133	12

Marker discovery, screening and data production

Microsatellite discovery was performed by pyrosequencing of a genomic library enriched for microsatellite motifs in a Roche 454 GS-FLX platform, subsequent marker screening, and primer design as a service by Genoscreen (Lille Cedex, France). To do so the company used 1 µg of DNA for library construction. This library was subsequently enriched for the following motifs: TG, TC, AAC, AAG, AGG, ACG, ACAT, and ACTC. The result from this step was sequenced using only 1/12th of the 454 platform capacity. The resulting sequences were filtered so that they contained microsatellite motifs that were suitable for primer design. Forward primers were tagged with an extra oligonucleotide sequence on their 5' end according to the M13-tailed primer method (Oetting et al. 1995) as described earlier (Curto et al. 2013). Those corresponded to four universal primer tails with complementary sequence to a third primer that had a sequence specific florescent dye: 6-FAM (TGTAAAACGACGGCCAGT), VIC (TAATACGACTCACTATAGGG), NED (TTTCCCAGTCACGACGTTG), and PET (GATAACAATTTCACACAGG). A GTTT tail was added to the 5' end of the reverse primer to minimize the polymerase stuttering effect. In this three primer assay the forward and reverse primers were used to amplify the region of interest and the tail primer to tag

the resulting fragment with a fluorescent chromatophore. The different universal primers allowed multiplexing in the PCR.

All primers were first tested in a simplex on their ability to amplify the sample used for marker discovery. With this step, primers that had mismatches presumably due to sequencing errors were sorted out. Amplification was done using QIAGEN Multiplex PCR Master Mix (Qiagen, Valencia, CA, U.S.A.) in 10 µL reactions containing 5 µL of QIAGEN Multiplex PCR Master Mix, 3.3 µL of water, 0.5 µL of DNA, 0.4 µL of each primer solution with the following concentrations: 10 mM for the reverse and the universal florescent primer, and 1 mM for the forward primer. The temperature profile is described below. All primers that showed positive results were multiplexed in four combinations of 50 µL solutions containing 4 nmol of each forward primer, and 40 nmol of each reverse and florescent universal primers. Each multiplex contained a maximum of four primers pairs corresponding to the four dyes used. Those mixtures were then used to amplify all 126 samples using the PCR protocol described above with 1 µL of primer mix and adjusted amount of water.

All PCR reactions were executed according to the following temperature profile: initial denaturation/activation step of 15 min; 95 °C for 30 s; touchdown starting at 62 °C to 56 °C, decreasing 0.5 °C per cycle for 60 s; extension at 72 °C for 30 s; followed by 20 cycles at 54 °C and eight cycles at 53 °C. Amplification success was evaluated by electrophoresis on a 2% agarose gel. Nevertheless, all amplification results were also genotyped with an internal size standard (Genescan-500 LIZ; Applied Biosystems, Inc., Foster City, CA, U.S.A.) in an ABI3130xl automatic sequencer (Applied Biosystems, Inc.). Alleles were called using GeneMapper ver. 4.0 (Applied Biosystems, Inc.).

Variability and genetic structure detection

The markers with positive amplification for most of the samples were screened for variability and information content. Variability was assessed by the number of alleles per marker and information content by calculating average polymorphism information content (PIC) and expected heterozygosity (He) per marker using the program Cervus ver. 3.0.3 (Kalinowski et al. 2007). Monomorphic loci were not included in further analyses. Deviations from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) estimations were tested using the software Genepop v 4.2 (Rousset 2008) and FSTAT (Goudet 1995), respectively. Frequency of null alleles was estimated using the program FreeNA (Chapuis and Estoup, 2007).

Genetic structure patterns were assessed with and without location coordinates of the individuals as prior information using the software Geneland v.4.0.4 (Guillot et al. 2005) and STRUCTURE v. 2.3.4 (Hubisz et al. 2009), respectively. Since the populations

studied are in a small area and gene flow among them is expected, STRUCTURE analysis was performed using the admixture model and correlated allele frequencies among populations. To find the best value of K, the program was run using K values of 1 to 20 (ten replicates each) for 200,000 replications, and eliminating the first 100,000. The best value of K was defined by Evanno's et al. (2005) method as implemented on Structure Harvester v0.6.93 (Earl and vonHoldt 2012). A STRUCTURE analysis with the best K was performed with the same parameters but letting the program run for 1,000,000 replications excluding the first 500,000. All the analyses made with STRUCTURE were performed with and without using populations as a prior. Initially Geneland was run for 100,000 replications for a maximum number of populations ranging from 2 to 20. The results were recorded every 100th iteration and the first 50,000 replications were excluded. The optimal number of populations was defined by choosing the higher number of clusters where most of the individuals showed an assignment larger than 0.5 to one of the clusters. As in STRUCTURE analyses, allele frequencies were considered to be correlated. Tests with and without the spatial and null allele model were performed as described above with the best maximum number of populations found. The model combination that resulted in the least ambiguous cluster assignment was used in an additional run, where 1,000,000 iterations were performed and results were recorded every 100 replicate. As in the previous tests, the first half of the results were excluded.

Genetic structure was estimated for populations circumscribed as explained above and two age classes. Trunk perimeter was used to separate "younger" from "older" individuals. Studies from J. oxycedrus in Italy indicate that a trunk perimeter of 60 cm corresponds to trees about 70 years old (Baldoni et al. 2004). This value was used as threshold to divide the samples in a group of individuals that were likely present before the peak of land use abandonment (about 1940-50) and a group that established later. This threshold also resulted in highest genetic differentiation as estimated from AMOVA when testing groupings with thresholds set at 10 cm intervals from perimeters 30cm to 100 cm.

An AMOVA analysis was also used to evaluate which grouping explained the highest amount of difference among groups. AMOVA was done using the software Genalex v. 6.5 (Peakall and Smouse 2012) with the following grouping: populations, trunk perimeter, and STRUCTURE and Geneland clusters. Groups according to STRUCTURE and Geneland, corresponded to the three clusters indicated by these analyses respectively. Hereby, samples with assignment probability higher than 0.5 for a certain cluster were considered. The AMOVA analyses was performed using F_{ST} instead of R_{ST} because it has been shown that this measure is more efficient in detecting

genetic structure in population with high degree of admixture (Balloux and Goudet 2002), which is expected due to the relatively small distance among individuals.

Genetic diversity variation among populations and tree size categories was also determined using the software FSTAT by calculating the average expected heterozygosity (He) and allelic richness (Ar) per grouping. As an additional measure of differentiation, populations' pairwise F_{ST} was calculated using the program FreeNA. This software was used because it can incorporate a null allele correction. The software Colony (Wang 2004) was used to find possible family relationships in the dataset containing trunk perimeter information. Trees with perimeter above 60 cm were assumed as potential parents and trees with perimeter below 60 cm as potential offspring. The mating system was considered to be polygamous and because we are studying a confound area that was previously fragmented we allowed inbreeding to be possible. A medium length, full likelihood run was performed with medium precision allowing allele frequencies to be updated. Because there is no knowledge of family structure, no siblingship prior was used. Since information about the gender of the individuals was not collected the potential parents were considered to be both potential fathers and mothers.

Results from BOTTLENECK v. 1.2.02 (Cornuet and Luikart 1997) were used to evaluate the possibility that demographic expansions or declines or founding effect led to the current structure. This was done assuming the Stepwise Mutation Model (SMM) since all the alleles were separated according to their repeat motif. Significant deviations from the mutation-drift equilibrium were calculated using Wilcoxon signed-rank test because it is the most reliable when low number of loci are used (Piry et al. 1999). As a complementary measure to this analysis, HWE deviations for all populations and trunk perimeter categories were estimated using the program Genepop v 4.2.

Results

Pyrosequencing of the enriched library resulted in 32,883 sequences from which 6,293 contained microsatellite motifs. From all the potential primer pairs analyzed, 241 passed Genoscreen bioinformatics validation. From those, 42 primer pairs were constructed and tested for amplification success, resulting in 18 markers that could be used for genotyping all samples (Table S2).

All primers amplified most of the samples and only two of them (Joxy28 and Joxy35) had an amplification failure larger than 10%. Joxy11 was the only monomorphic marker and for that reason was the only marker excluded from further analyses. The remaining markers had between two and eleven alleles (Table 2). The marker with lowest PIC and He values (Joxy13; PIC=0.08; He=0.09) was not the least polymorphic with 3 alleles. The most polymorphic and informative marker (Joxy8) had a PIC and He values

of 0.83 and 0.85 respectively. According to the LD deviation test using a Bonferroni correction, none of the markers were linked. Six markers deviated significantly from Hardy-Weinberg equilibrium due to heterozygotes deficiency (F_{IS} of these markers ranged from 0.27 to 0.89). In addition, these six markers were the only ones presenting a frequency of null alleles greater than 0.1. To test the utility of those markers the analyses evaluating population genetic diversity and structure were performed with two datasets. The first one was composed by all markers that were variable (17 marker dataset) and the second one was composed by all markers that did not deviated from HWE (11 marker dataset).

Table 2. Markers variability, amplification success and Hardy-Weinberg deviation statistics.

Locus	Number of alleles	Amplification failure (%)	HObs	HExp	PIC	HWE p-value	F_{IS}	NA
Joxy1	5	8	0.23	0.67	0.60	HS	0.65	0.26
Joxy3	4	0	0.44	0.46	0.41	0.261	0.03	0.00
Joxy4	7	1	0.57	0.58	0.50	0.998	0.02	0.01
Joxy8	11	0	0.78	0.85	0.83	0.528	0.09	0.04
Joxy10	3	0	0.28	0.46	0.39	0.000	0.40	0.13
Joxy11	1	0	-	-	-	-	-	-
Joxy12	4	0	0.33	0.34	0.29	0.201	0.03	0.00
Joxy13	3	1	0.08	0.09	0.08	0.329	0.06	0.00
Joxy14	3	0	0.42	0.45	0.37	0.678	0.06	0.01
Joxy17	3	0	0.27	0.26	0.23	0.993	-0.02	0.00
Joxy20	4	0	0.64	0.62	0.53	0.449	-0.05	0.00
Joxy22	2	1	0.21	0.26	0.23	0.265	0.20	0.06
Joxy23	5	1	0.51	0.62	0.56	0.122	0.17	0.06
Joxy28	4	16	0.10	0.26	0.25	HS	0.61	0.17
Joxy31	3	0	0.25	0.31	0.28	0.214	0.19	0.05
Joxy35	5	44	0.06	0.54	0.48	HS	0.89	0.32
Joxy37	2	2	0.23	0.50	0.37	0	0.54	0.18
Joxy42	4	1	0.33	0.45	0.42	0.003	0.27	0.11

STRUCTURE analysis indicated K=3 as optimal according to Evanno's Method. The analyses with and without a priori knowledge of populations had similar results. A priori knowledge of population assignment improved the AMOVA but the structure found did not correspond to any geographical pattern, with individuals from the same populations being assigned to different clusters (results not shown). The analysis without the markers deviating from HWE showed the same assignment probability to all the clusters for all individuals.

In Geneland analysis, the optimal number of populations was also three. However, when the non-spatial and null allele models were used, all the samples were

assigned to the same cluster and so these models were not used in further analyses. No structure among individuals was found for the analyses using 11 markers, so only results with the 17 markers are shown. For the analysis using the spatial model, all the populations showed assignment to more than one cluster (Fig. 2a). Cluster 1 (white in Fig. 2a) was more frequently assigned in the south and Cluster 3 (black) in the north, resulting in a slight signature of geographic structure. Nevertheless, all populations contained individuals assigned to at least two clusters. When Geneland's cluster assignment probability was plotted according to trunk perimeter as proxy for tree age class a similar result was obtained (Fig. 2b). Although all age classes show some degree of assignment to all clusters, trees larger than 60 cm showed more assignment to the northern cluster while smaller trees to the southern cluster.

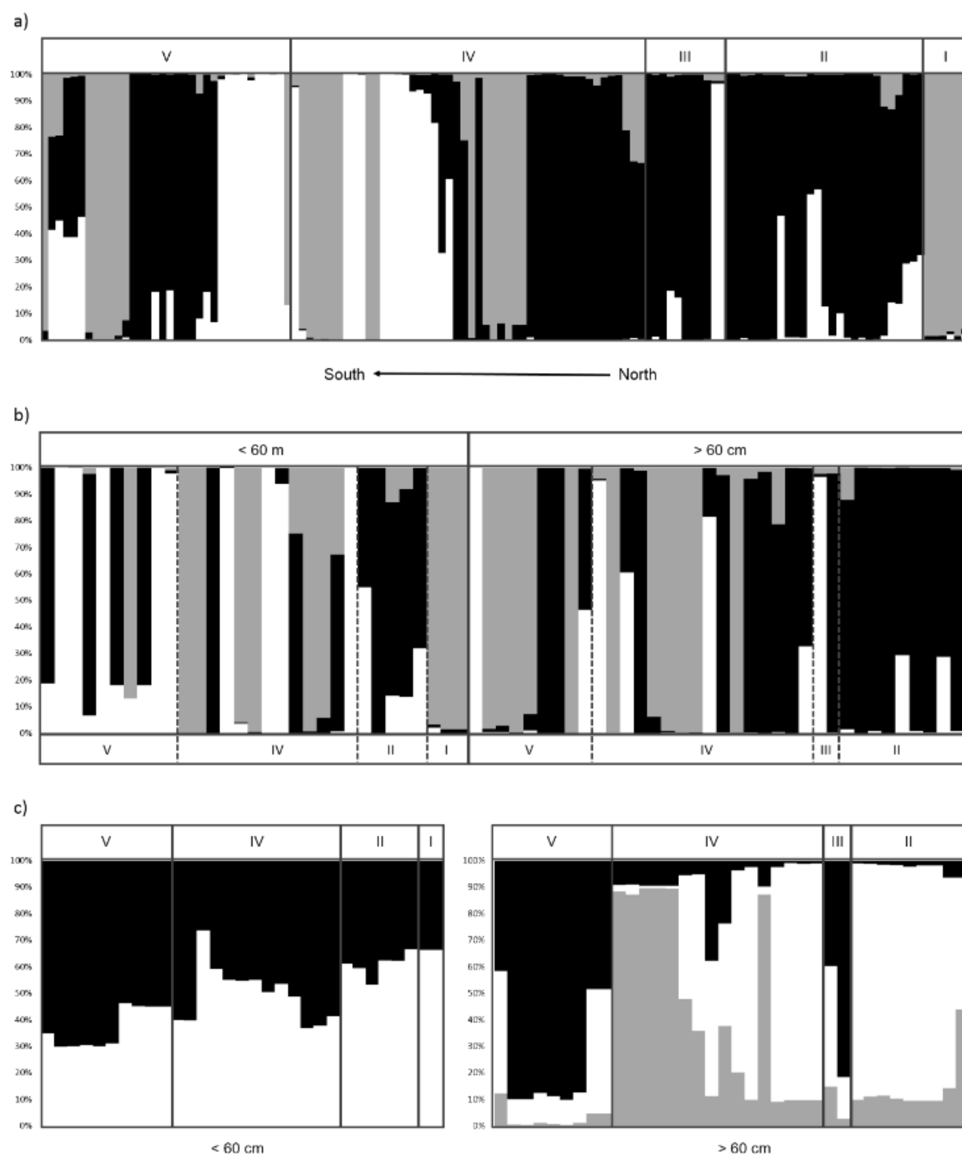


Fig. 2 Plots from the cluster assignment probability obtained in Geneland analyses for a maximum number of three groups. The plots were sorted according three different criteria: a) *a priori* populations; b) trunks perimeter and *a priori* populations; c) separate analyses of trees below 60 cm and trees above 60 cm.

Two additional Geneland runs were performed using either "younger" or "older" trees to evaluate whether there was any spatial pattern within age classes. In these new GENELAND analysis, the best grouping was two for "younger" trees, and three for "older" trees. For "younger" trees there was almost no structure with just a very small tendency for a North to South gradient in the relative prevalence of the two groups in individuals genotyped (Fig. 2c). For older trees we see a clear differentiation among populations, with one of the groups dominating in population V and III, another represented primarily in population IV, and another dominating in population II but also occurring in population IV (Fig. 2c).

Table 3. Summary results of AMOVA analyses for individuals grouped according to the sampled population, trunk perimeter (<60 cm vs > 60 cm), STRUCTURE analyses with and without population priors, and Geneland (See main text for details).

Test	Nr of Groups	Nr of individuals	Variation (%)		F_{ST} variance		F_{ST}
			Among Groups	Within Groups	Among Groups	Within Groups	
Sampled populations							
17 markers	5	126	1.8%	98.2%	0.075	4.014	0.018*
11 markers	5	126	0.6%	99.4%	0.014	2.425	0.006
Trunk perimeter							
17 markers	2	67	1.4%	98.6%	0.058	4.134	0.014*
11 markers	2	67	0.7%	99.3%	0.018	2.445	0.007
STRUCTURE							
with population prior	3	108	10.8%	89.2%	0.457	3.776	0.108*
without population prior	3	113	9.2%	90.8%	0.391	9.867	0.092*
Geneland	3	124	4.5%	95.5%	0.187	3.974	0.045*

*Significant value (p value<0.01)

AMOVA showed low levels of variation among groups and low values of F_{ST}, from which only the ones with the 17 marker dataset were significant (Table 3). The analyses considering different populations showed both slightly higher F_{ST} and slightly higher percentage of variation explained by differences among groups (F_{ST} = 0.018 and1.8%, respectively) when comparing to the division made with trunk perimeter (F_{ST} = 0.014 and 1.4%, respectively). The STRUCTURE test showed higher F_{ST} values and higher variation among groups (11% and 0.108, respectively) than the Geneland analyses (4% and 0.045, respectively). However, for the STRUCTURE analyses fewer individuals where used because there was more ambiguity in the assignment of the individuals.

Table 4. Genetic diversity and HWE deviation test results for the populations defined and categories of trunks perimeter, computed using the data set either with (17) or without (11) markers under Hardy-Weinberg equilibrium.

Grouping	Category	Nr individuals	17 Marker			11 Marker		
			He	AR	HWE p-value	He	AR	HWE p-value
Populations	I	6	0.48	2.25	0.03	0.48	2.70	0.60
	II	27	0.45	2.22	0.00	0.43	2.65	1.00
	III	11	0.38	2.03	0.16	0.38	2.46	0.88
	IV	48	0.47	2.30	HS	0.45	2.64	0.09
	V	34	0.44	2.19	0.00	0.44	2.58	0.08
Trunks Perimeter	< 60 cm	31	0.42	3.55	HS	0.46	3.28	0.03
	> 60 cm	36	0.46	3.81	HS	0.48	3.49	0.23

Expected heterozygosity and allelic richness per a priori population only varied a little, ranging between 0.38 to 0.48, and 2.03 to 2.28, respectively (Table 4). For both estimates the least diverse was population III, which may be a consequence of its small sampling size. A similar result was found for the 11 marker set (exclusion of Markers with significant deviation from HWE). Overall there was no clear geographical pattern in diversity. The He and AR estimates were slightly higher for older trees in both datasets. (Table 4). Using BOTTLENECK, no population or trunk perimeter category defined showed significant deviations from the mutation-drift equilibrium ($p > 0.01$), meaning that there was no genetic signature of the major population expansion occurring since about the 1950s. Deviations from HWE were only verified for the 17 markers dataset ($p > 0.01$; see Table 4). More specifically for populations I and III and all trunk perimeter categories. Pairwise F_{ST} values for the 11 markers dataset were smaller than the ones from 17 markers dataset ranging from -0.014 to 0.017 and from -0.004 to 0.089, respectively (Table 5). These values only corresponded to geographical patterns for the 17 marker dataset, where population I was more distant in relation to the others and, with exception of the comparison between population III and V, F_{ST} values seemed to be related with geographical proximity when compared to the other populations. For example, populations II and III and populations V and IV seem to be more related to each other than to the others. Several negative values were observed which can be a consequence of continuous gene flow among these groups.

The family structure analyses performed with Colony found eight potential relationships for the 17 marker dataset and 13 for the 11 marker dataset, corresponding only to a small portion of the individuals analyzed (Table S3). Most of the relationships were between population IV and V and within them. Besides those, there was only one relationship between population II and V. Each of the relationship found corresponded to a different potential offspring individual and to four potential parents from populations IV and V.

Table 5. Pairwise F_{ST} for all pair of a priori populations, computed using the data set either with (17) or without (11) markers under Hardy-Weinberg equilibrium.

Population		17 markers	11 markers
I	II	0.064	-0.001
I	III	0.089	-0.002
I	IV	0.027	0.017
I	V	0.021	-0.001
II	III	-0.004	-0.001
II	IV	0.008	0.007
II	V	0.015	0.005
III	IV	0.019	0.005
III	V	-0.001	-0.014
IV	V	0.001	0.003

Discussion

Marker development

In this study we were able to develop 17 variable microsatellite markers for *Juniperus oxycedrus*. This improvement, compared with previous microsatellite studies on *Juniperus* (five markers for *Juniperus communi*s, Michalczyk et al. 2006), is likely to be a consequence of the high throughput sequencing technology used. The higher number of sequences generated, several thousand compared to several hundred sequences using a subcloning strategy, facilitates the process of microsatellite motifs discovery. Thus using Next Generation Sequencing is becoming rapidly the method of choice for marker discovery (Ekblom and Galindo 2011). With pyrosequencing (the 454 technology), typically around 100,000 sequences can be produced (Loman et al. 2012) and, when an enrichment step is applied, most of the microsatellite loci from the genome can be represented on those sequences. In our case, although we did not use the full capacity of a run, we produced 32,883 sequences from which 6,293 contained microsatellite motifs and 241 were good enough for primer design. Even when an approach without enrichment is used, a high number of potential loci are obtained. For example, Csencsics et al. (2010) and Curto et al. (2013) obtained 307 and 866, respectively, good quality microsatellite containing sequences from a total of 76,692 and 65,401 initial sequences. Nevertheless, using an enrichment step resulted in 19% of the loci containing a motif, compared to 1.3% in our earlier study (Curto et al. 2013). In this study there was very small number of sequences suitable for primer design when compared to the microsatellite containing sequences (only 19%). The reads excluded in this step did not contain sufficiently long flanking regions for primer design. This may be a consequence of the presence of some fragmented DNA that is preferably amplified in

the emulsion PCR step of the library construction due to its smaller size. This may be creating overrepresentation of sequences containing microsatellite motifs in the beginning or in the end of the reads. Recently, alternative sequencing technologies like Illumina (Castoe et al. 2010), Ion torrent (Elliott et al. 2014) and PacBio (Grohme et al. 2013) platforms have been used for microsatellite discovery with similar or better results than pyrosequencing. Illumina and Ion torrent platforms produce smaller reads (maximum of 250 and 300 bp, respectively) than 454, however they have higher throughput making them generally cheaper methods. PacBio produces longer reads (around 2000 bp) increasing the likelihood of obtaining a microsatellite containing read with flanking regions big enough for primer design.

The high number of sequences obtained by NGS allows pooling different samples during sequencing and screening for markers that are variable already within the sequencing step. We used only one individual mainly because we used only a comparably small subset of the 454 run for marker discovery, and we were interested in variation on a relatively small spatial scale. For both reasons, the amount of duplicated loci in the group of polymorphic markers might be increased if multiple non-barcoded individuals are used. Nevertheless, from the primers designed, 18 amplified most of the samples and only one of them was monomorphic, but in general the markers presented a low number of alleles. Only one of them had more than 10 alleles and most of them had less than five alleles. The sampling used was geographically very restricted (the most distant sampling localities were around 35 Km apart) so low variation is expected. For future studies using a larger study area, the variability may be higher and the monomorphic locus may be variable in that situation. In addition, less variable markers may be especially useful when different species are used. Theoretically, if they are less variable in their motif, their primer binding regions might be more conserved and thus more likely to be cross amplified in different species. In Boratyński et al. (2014) work, the authors had a problem in cross-amplifying microsatellites markers in multiple *Juniperus* species, ending up with a dataset of only three markers. The more conserved markers described here may contribute for improvements in this kind of studies.

Six markers deviated from Hardy-Weinberg equilibrium because their heterozygosity was lower than expected. Of those, only one amplified in all the samples, and all of them had a high frequency of null alleles, in terms of missing data or determined by the program FreeNA. The genetic diversity measures could therefore be underestimated in our analysis. For the STRUCTURE and Geneland analyses, when these markers were discarded the differentiation signal was lost. Excess of null alleles in microsatellite analysis is generally attributed to mutations in the primers' binding sites. The variability in the primer binding site may reflect a general higher variability of these

loci which is ultimately contributing to its higher information content. This might explain why their inclusion in the analyses lead to detection of a weak genetic structure patterns which cannot be resolved using only the 11 marker dataset. No structure was obtained using the complete dataset in Geneland when the null allele model was used. Geneland tends to overestimate null allele frequencies in case missing data are present (Guillot 2012) which could have influenced the inference.

Genetic structure

The dataset contains a weak signal of genetic structure, which was determined using Bayesian clustering methods as implemented in the programs STRUCTURE and Geneland. For the STRUCTURE analysis, the assignment within each population was very heterogeneous and there was no differentiation among the populations studied. For the Geneland analysis, the assignment was more homogeneous within each population, and there were clear differences among populations. AMOVA supported higher differentiation among groups for the STRUCTURE analyses. Geneland analyses uses spatial information as a prior and the algorithm is not sensitive to loci deviating from Hardy-Weinberg equilibrium (Guillot et al. 2005). Contrary, STRUCTURE assigns individuals by optimizing HWE within each cluster (Hubisz et al. 2009) and the use of loci not in agreement with this assumption can result in a wrong inference. In the case of Geneland none of the assumptions were violated, making it more suitable for this particular dataset. Moreover, due to the low variability found in the data, the spatial information may have been valuable to detect any weak signal of structure. The Geneland analyses implementing the spatial model recovered better results than when this model was not used. The non-spatial model does not consider geographical information, thus making the analyses less reliable in recovering structural patterns for populations with weak genetic structure (Guillot et al., 2008).

Similar to our analysis, other *Juniperus* species in larger geographical scales showed high variability within populations and low genetic differentiation. This was, for example, the case of *Juniperus communis* populations from Northwest Europe (Van Der Merwe et al. 2000; Oostermeijer and Knegt 2004; Michalczyk et al. 2006; Provan et al. 2008; Vanden-Broeck et al. 2011) and *Juniperus macrocarpa* from Southern Iberia (Juan et al. 2012). *Juniperus* are dioecious wind pollinated plants, making gene flow among fairly distant populations likely. This pattern was found in populations that were way more distant from each other than the ones analyzed in here. Thus the gene flow in the populations analyzed may be extremely high making it likely for those five populations to be in fact one. This would explain the low F_{ST} values, the AMOVA results, and the fact that we found paternity relationships between distant populations (i.e. between

populations IV and I). Moreover, our results show that genetic structure found was generally weak and it can be explained by geographical patterns and by tree age. This is mainly supported by the AMOVA and Geneland analyses. In our AMOVA result, the division by populations provided only slightly better results than the division by trunk perimeter. In the Geneland results, one of the clusters was more frequently assigned in the south and to younger trees, and the other in the north and to older trees. This can be a signal of a weak isolation by distance pattern and a week differentiation among young and old trees. The isolation by distance pattern is also supported by the pairwise F_{ST} results, for which the more distant population (population I) is also the one more genetically dissimilar, whereas geographically close populations like IV and V or populations II and III showed the lowest F_{ST} values. The weak genetic structure found has also been reported for other *Juniperus* populations (e.g. Van der Merwe et al. 2000; Oostermeijer and Knegt 2004; Juan et al. 2012).

When Geneland analysis was ran separately for young and old trees, we found a clear geographical structure in older trees. No structure was found in younger trees showing that there is no bias created by our sampling. This analyses is furthermore congruent with a scenario of past fragmentation followed by recolonization. Genetic structure between populations of older trees indicate some restrictions to gene flow among these populations in the past. More recently, these barriers ceased to exist allowing for a population expansion. As it is shown from the results of Colony, some individuals appeared to have contributed more for the expansion of the *J. oxycedrus* than others, resulting in some alleles to be more frequent in younger than older trees due to founding effects. This ultimately resulted in the lack of structure in younger trees, which contributed to the differentiation observed between old and young trees, and the weakening of spatial structure. Because all the parental individuals are from populations IV and V we could assume that those are the major contributors for the genetic pool of younger trees. However we were only able to find family relationships in a small set of the samples used, especially when populations I and II are considered. Thus, younger trees may have been originated from populations that were not sampled and a more intensive sampling of the region and the surrounding areas is necessary in order to take conclusions about this matter. Nevertheless, the fact that a structure according to age exists is an indication that some populations are contributing more for the recolonization than others. This can be a consequence of demographic reasons, where some populations were larger and older when expansion started, making them numerically more likely to reproduce successfully. Selection of pioneering features may be also behind this finding. Some individuals may have presented characters that were advantageous to the establishment of pioneer vegetation being favorably selected. To

test these hypotheses a study assessing pioneering features such as seed production and viability is necessary. If no difference in these features is found among populations the demographic hypotheses would be the most likely explanation.

This past fragmentation scenario fallowed by expansion of the populations is congruent with the land use history of the region. After arable fields were abandoned, the expansion of natural vegetation could take place and the population of *J. oxycedrus* increased. A similar pattern was found by Vellend et al. (2004) in the forest associated herb *Trillium grandiflorum*, where individuals from younger secondary forest were genetically more divergent from individuals from older primary forest. Other studies using forest species found a loss of genetic diversity in recent forests that colonized abandoned agriculture fields (e. g. Jacquemyn et al. 2009; Leite et al. 2014). This loss of genetic diversity of younger plants is expected because populations undergone a genetic bottleneck due to the fragmentation and recolonization event. However, *Juniperus* plants are dioecious making them particularly sensitive to barriers to gene flow. Vanden-Broeck et al. (2011) found that younger trees had lower genetic diversity in fragmented populations of *J. communis*. Although we found a greater genetic diversity in older trees, the difference in relation to younger trees was small. In addition, no signal of deviation from HWE and mutation-drift equilibrium was found for both groups. This is an indication that the founding effect was not very severe, which may be a consequence of how recent the land abandonment began (around 60 years ago). Alternatively, as described above, the genetic diversity of *Juniperus* populations is highly heterogeneous being possible for one population to retain the same degree of genetic diversity when compared with a larger group of populations. The high variability within populations plus how recent the event is make in this case the founding effects not to create a drastic reduction of genetic diversity.

We found that the best trunk perimeter threshold to assess genetic differentiation between younger and older trees was 60 cm. This corresponds to trees about 70 years old, as inferred from a regression equation between trunk diameter and tree age for a *Juniperus oxycedrus* populations from Italy (Baldoni et al. 2004), considering that a perimeter of 60 cm corresponds to 19 cm of diameter in a perfectly circular section of trunk. Taking into account that the maturity age for *J. oxycedrus* is around 20 years, and that land abandonment started at around 60 year ago, it is likely that the trees with current perimeter above 60 cm were the only ones able to reproduce at the time. This made them the major contributors of the population's expansion, which explains why the value of 60 cm was obtained as the best trunk perimeter division. Nevertheless, to be sure of this conclusion a dendrochronological study would be needed to have a more accurate estimate of tree ages and the local relationship between trunk diameter and age.

Ideally, a fine-scale genetic structure analyses would have been performed, focusing not only on genetic diversity and structure patterns, but also on gene flow. Only a small set of the populations and localities present in the region were sampled, therefore some other populations may be contributing for the genetic poll of the region. For that reason, we did not want to take conclusions in the matter of gene flow patterns and relations among populations. Debout et al. (2011) did a spatial genetic structure analyses (SGS) for *Distemonanthus benthamianus*, a tree from Western Africa rain forest, which consisted in finding correlations between the coefficient of kinship and the spatial distance, allowing to find small scale genetic structure patterns. Using that analyses they were able to estimate seed dispersal distances. The application of this analyses in *J. oxycedrus* populations would be mainly advantageous in a larger geographical scale. This way it would be possible to study the main barriers to gene flow in this species and not only in the current study area. In addition, it would also be interesting to see if similar patterns are found in other populations that were and were not fragmented in the past. That particular study would help to validate the findings in this paper. However, due the high intensity of agricultural and pastoral activities in the Mediterranean region it would be impossible or nearly impossible to find an area suitable for this study.

The finding that genetic structure in *J. oxycedrus* in the Sabor region is best explained by age, has several implications: the recovery of the population after the past fragmentation might still be ongoing. Trees that established during the recolonization are not yet fully admixed with older trees, even though we would expect this to happen in the future. This has some implications for the impact of the dam, because it has caused the loss of part of the remnant trees that survived the peak of agricultural expansion in steep slopes of the river valley. Therefore, these trees will not be available as source of genetic variation in the future, and the current pattern of genetic structure might be stabilized. The data presented here are a first step in our attempt to study this context.

Acknowledgements

The study was funded by the Portuguese Science Foundation through project LTER/BIA-BEC/0004/2009. The same foundation also supports M.C., M.N., and F.A. through Ph.D. fellowships with the reference SFRH/BD/79010/2011, SFRH/BD/64225/2009, and PD/BD/52606/2014, respectively. PB was supported by EDP Biodiversity Chair.

Data archiving statement

The sequences used for primer design were submitted to Genbank and they will be soon available with the reference number KM013316 to KM013333. A more detail

view of the of the sequence reference number attribution can be found in the Table S2. The genotypes obtained for the microsatellite dataset used in this study are available in the Online Resources (http://link.springer.com/article/10.1007/s11295-015-0861-2).

References

Adams RP, Morris JA, Pandey RN, Schwarzbach AE (2005) Cryptic speciation between Juniperus deltoids and Juniperus oxycedrus (Cupressaceae) in the Mediterranean. Biochem Sys Ecol 33:771-787

Alexander PJ, Rajanikanth G, Bacon CD, Bailey CD (2006) Recovery of plant DNA using a reciprocating saw and silica-based columns. Mol Ecol Notes 7:5-9

Amaral FJ (1986) Juniperus. In: S. Castroviejo et al. (eds) Flora iberica. Plantas vasculares de la península Ibérica e islas Baleares. Real Jardín Botánico de Madrid, CSIC, Madrid, pp 181-188

Baldoni M, Biondi E, Ferrante L (2004) Demographic and spatial analysis of a population of Juniperus oxycedrus L. in an abandoned grassland. Plant Biosyst 138:89-100

Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Mol. Ecol. 11:771-783

Berge G, Nordal I, Hestmark G (1998) The effect of breeding systems and pollination vectors on the genetic variation of small plant populations within an agricultural landscape. Oikos 81:17-29

Blondel J, Aronson J, Bodiou J-Y, Boeuf G (2010) The Mediterranean Region Biological Diversity in Space and Time. Oxford University Press, USA, pp 392

Bondi E (1990) Population characteristics of Juniperus oxycedrus L. and their importance to vegetation dynamics. Giorn Bot Ital 124:330-337

Boratyński A, Wachowiak W, Dering M, Boratyńska K, Sękiewicz K, Sobierajska K, Jasińska AK, Klimko M, Montserrat JM, Romo A, ; Ok T, Didukh Y (2014) The biogeography and genetic relationships of Juniperus oxycedrus and related taxa from the Mediterranean and Macaronesian regions. Bot J Linn Soc 174:637-653

Brus R, Ballian D, Zhelev P, Pandža M, Bobinac M, Acevski J, Raftoyannis Y, Jarni K (2011) Absence of geographical structure of morphological variation in Juniperus oxycedrus L. subsp. oxycedrus in the Balkan Peninsula. Eur J For Res 130:657-670

Cano E, Rodriguez-Torres A, Gomes CP, Garcia-Fuentes A, Torres JA, Salazar C, Ruiz-Valenzuela L, Cano-Ortiz A, Montilla RJ (2007) Analysis of the Juniperus oxycedrus L. communities in the centre and south of the Iberian peninsula (Spain and Portugal). Acta Bot Gallica 154:79-99

Castoe TA, Poole AW, Gu W, Jason de Koning AP, Daza JM, Smith EN, Pollock DD (2010) Rapid identification of thousands of copperhead snake (Agkistrodon contortrix) microsatellite loci from modest amounts of 454 shotgun genome sequence. Mol Ecol Resour, 10:341-347

Chapuis, MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. 24: 621-631.

Chauchard S, Carcaillet C, Guibal F (2007) Patterns of land-use abandonment control tree-recruitment and forest dynamics in Mediterranean mountains. Ecosystems 10:936-948.

Cornuet JM, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014

Costa JC, Aguiar C, Capelo JH, Lousã M, Neto C (1998) Biogeografia de Portugal continental. Quercetea 1:5-56

Couvet D (2002) Deleterious effects of restricted gene flow in fragmented populations. Conserv Biol 16:369-376

Cowling RW, Rundel PW, Lamont BB, Arroyo MK, Arianoutou M (1996) Plant diversity in Mediterranean-climate regions. Trends Ecol Evol 11:362-366

Csencsics D, Brodbeck S, Holderegger R (2010) Cost-Effective, Species-Specific Microsatellite Development for the Endangered Dwarf Bulrush (Typha minima) Using Next-Generation Sequencing Technology. J Hered 101:789-793

Curto MA, Tembrock LR, Puppo P, Nogueira M, Simmons MP, Meimberg H (2013) Evaluation of microsatellites of Catha edulis (qat; Celastraceae) identified using pyrosequencing. Biochemical Systematics and Ecology 49:1-9

Debout GDG, Doucet JL, Hardy OJ (2011) Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, Distemonanthus benthamianus (Caesalpinioideae). Heredity 106:88-99

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1-15

Elliott CP, Enright NJ, Allcock RJN, Gardner MG, Meglécz E, Anthony J, Krauss SL (2014) Microsatellite markers from the Ion Torrent: a multi-species contrast to 454 shotgun sequencing. Mol Ecol Resours 14:554-568

Ellstrand NC, Elam DC (1993) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24:217-242

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

Flynn S, Turner RM, Stuppy WH (2006) Seed information database (release 7.0, October 2006). http://www.kew.org/ data/sid. Accessed 2 Nov 2014

Goudet J (1995) Fstat version 1.2: a computer program to calculate Fstatistics. J Hered 86:485-486

Grohme MA, Soler RF, Wink M, Frohme M (2013) Microsatellite marker discovery using single molecule real-time circular consensus sequencing on the Pacific Biosciences RS. BioTechniques 55:253–256

Guillot G, Mortier F, Estoup D (2005) Geneland: a computer package for landscape genetics. Mol Ecol Notes 170:1261-1280

Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. Bioinformatics 24: 1406-1407

Guillot, G (2012) Population genetic and morphometric data analysis using R and the Geneland program. Available at http://www2.imm.dtu.dk/~gigu/Geneland/Geneland-Doc.pdf

Hoelzer A (2003) Vegetation ecological studies at the lower course of Sabor River (Tras-os-Montes, NE-Portugal). Dissertation, University of Bremen

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322-1332

Jacquemyn H, Vandepitte K, Roldan-Ruiz I, Honnay O (2009) Rapid loss of genetic variation in a founding population of Primula elatior (Primulaceae) after colonization. Ann Bot 103:777-783

Juan A, Fay MF, Pastor J, Juan R, Fernández I, Crespo MB, (2012) Genetic structure and phylogeography in Juniperus oxycedrus subsp. macrocarpa around the Mediterranean and Atlantic coasts of the Iberian Peninsula, based on AFLP and plastid markers. Eur J For Res 131:845-856

Juan R, Pastor J, Fernández I, Diosdado JC (2003) Relationships between mature cone traits and seed viability in Juniperus oxycedrus. Acta Biol. Cracoviensia Ser. Bot. 45: 69-78

Kalinowski ST, Taper ML, Marshall TC, (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099-1106

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230-241

Klimko M, Boratyńska K, Montserrat JM, Didukh Y, Romo A, Gómez D, Kluza-Wieloch M, Marcysiak K, Boratyński A (2007) Morphological variation of Juniperus oxycedrus subsp. oxycedrus (Cupressaceae) in the Mediterranean region. Flora 202:133-147

Larson BMH, Barrett SCH (2000) A comparative analysis of pollen limitation in flowering plants. Biol J Linnean Soc 69:503-520

Leite FAB, Brandão RL, de Oliveira Buzatti RS, de Lemos-Filho JP, Lovato MB (2014) Fine-scale genetic structure of the threatened rosewood Dalbergia nigra from the Atlantic forest: comparing saplings versus adults and small fragment versus continuous forest. Tree Genet Genomes 10:307-316.

Lennartsson T (2002) Extinction thresholds and disrupted plant-pollinator interactions in fragmented plant populations. Ecology 83:3060-3072

Lewandowski A, Boratyński A, Mejnartowicz L (1996) Low level of isozyme variation in an island population of Juniperus oxycedrus subsp. macrocarpa (Sm. ex Sibth.) Ball Acta Soc Bot Pol 65:335-338

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30:434-439

Michalczyk IM, Sebastiani F, Buonamici A, Cremer E, Mengel C, Ziegenhagen B, Vendramin GG (2006) Characterization of highly polymorphic nuclear microsatellite loci in Juniperus communis L. Mol Ecol Notes, 6, 346-348

Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA, King RA (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30450-458

Oostermeijer JGB, De Knegt B (2004) Genetic population structure of the wind-pollinated, dioecious shrub Juniper communis in fragmented Dutch heathlands. Plant Species Biol 19:175-184

Ortiz PL, Arista M, Talavera S (1998) Low reproductive success in two subspecies of Juniperus oxycedrus L. Int J Plant Sci 159:843-847

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539

Piry S, Luikart G, Cornuet J-M (1999) Bottleneck: a program for detecting recent reductions in effective population size reductions from allele frequency data. J. Heredity 86:502-503

Porto M, Correia O, Beja P (2011) Long-term consequences of mechanical fuel management for the conservation of Mediterranean forest herb communities. Biodivers Conserv 20: 2669-2691

Provan J, Beatty GE, Hunter AM, McDonald RA, McLaughlin E, Preston SJ, Wilson S (2008) Restricted gene flow in fragmented populations of a wind-pollinated tree. Conserv Genet 9:1521-1532

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources 8: 103-106.

Santana J, Porto M, Reino L, Beja P (2011) Long-term understory recovery after mechanical fuel reduction in Mediterranean cork oak forests. Forest Ecol Manag 261:447-459

Schaal BA, Leverich WJ (1996) Molecular variation in isolated plant populations. Plant Species Biol 11:33-40

Sluiter R, de Jong SM (2007) Spatial patterns of Mediterranean land abandonment and related land cover transitions. Landscape Ecol 22:559-76

Van der Merwe M, Winfield MO, Arnold GM, Parker JS (2000) Spatial and temporal aspects of the genetic structure of Juniperus communis populations. Mol Ecol 9:379-386

Vanden-Broeck A, Gruwez R, Cox K, Adriaesnssens S, Michalczyk IM, Verheyen K (2011) Genetic structure and seed-mediated dispersal rates of an endangered shrub in a fragmented landscape: a case study for Juniperus communis in northwestern Europe. BMC Genet 12:73

Vellend M (2004) Parallel effects of land-use history on species diversity and genetic diversity of forest herbs. Ecology 85:3043–3055.

Verheyen K, Schreurs K, Vanholen B, Hermy M (2005) Intensive management fails to promote recruitment in the last large population of Juniperus communis (L.) in Flanders (Belgium). Biol Conserv 124:113–121

Wang, J (2004) Sibship reconstruction from genetic data with typing errors. Genetics 166: 1963-1979.

Weidema IR, Magnussen LS, Philipp M (2000) Gene flow and mode of pollination in a dry-grassland species, Filipendula vulgaris (Rosaceae). Heredity 84:311-320

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11:413-418

Chapter 3: *Micromeria* evolution in the Canary Islands

Paper I: Puppo, P., Curto, M., Meimberg, H (2016) Genetic structure of *Micromeria* (Lamiaceae) in Tenerife, the imprint of geological history and hybridization on within-island diversification. Ecology and Evolution. 2045-7758

Paper II: Curto M., Puppo P., Meimberg H. (submitted) Genetic diversity and differentiation patterns of *Micromeria* from the Canary Islands are congruent with multiple colonization dynamics and the establishment of species syngameons.

3.1. Genetic structure of *Micromeria* (Lamiaceae) in Tenerife, the imprint of geological history and hybridization on within-island diversification.

Pamela Puppo[1,2], Manuel Curto[1,2], Harald Meimberg[2]
[1]CIBIO, Research Center in Biodiversity and Genetic Resources/InBio Associated Laboratory, University of Porto, Vairão, Portugal
[2]Institute for Integrative Nature Conservation Research, University of Natural Resources and Life Sciences, A-1180 Vienna, Austria.

Abstract

Geological history of oceanic islands can have a profound effect on the evolutionary history of insular flora, especially in complex islands such as Tenerife in the Canary Islands. Tenerife results from the secondary connection of three paleo-islands by a central volcano, and other geological events that further shaped it. This geological history has been shown to influence the phylogenetic history of several taxa, including genus *Micromeria* (Lamiaceae). Screening 15 microsatellite markers in 289 individuals representing the eight species of *Micromeria* present in Tenerife, this study aims to assess the genetic diversity and structure of these species and its relation with the geological events on the island. In addition, we evaluate the extent of hybridization among species and discuss its influence on the speciation process. We found that the species restricted to the paleo-islands present lower levels of genetic diversity but the highest levels of genetic differentiation suggesting that their ranges might have contracted over time. The two most widespread species in the island, *M. hyssopifolia* and *M. varia*, present the highest genetic diversity levels and a genetic structure that seems correlated with the geological composition of the island. Samples from *M. hyssopifolia* from the oldest paleo-island, Adeje, appear as distinct while samples from *M. varia* segregate into two main clusters corresponding to the paleo-islands of Anaga and Teno. Evidence of hybridization and intraspecific migration between species was found. We argue that species boundaries would be retained despite hybridization in response to the habitat's specific conditions causing postzygotic isolation and preserving morphological differentiation

Keywords: Genetic structure, hybrid zones, island evolution, Macaronesia, microsatellites, oceanic islands, paleo-islands, SSR.

Introduction

Speciation is traditionally seen as the accumulation of differences between two populations in allopatry, with geographic distance as barrier to gene flow. In general, gene-flow will prevent differentiation, so continuous migration and hybridization events will counteract speciation processes (Yeaman and Whitlock 2011) and potentially also homogenize formerly differentiated species when they come secondarily into contact and are not reproductively isolated. However, it had been shown that speciation can occur by adaptation and divergent selection also with gene-flow (Seehausen et al. 2014) and several new concepts had been developed that explain the context between genetic diversity, selection, and gene flow, e.g., the hybrid swarm – (Seehausen 2004) or the surfing syngameon hypothesis (Caujapé-Castells 2011). These hypotheses postulate that populations can work as sink of genetic diversity through hybridization which furthermore could buffer effects of genetic drift and could increase the level of diversity for selection to act upon and could thus foster differentiation by adaptation. This context had become known during the last year as "speciation-with-gene-flow" especially in zoology. A recent paper published by Roy et al. (2015), showed how hybridization in contact zones can transform between-lineage variation into within-population genetic diversity increasing the population's potential for adaptation, ultimately favoring adaptive radiations in a short period of time. Overall, hybridization might enhance genetic and phenotypic variation facilitating further divergence and adaptation to changing environmental conditions (Pavarese et al. 2013; Seehausen et al. 2014).

Hybridization might also be able to explain peculiarities of insular radiations, i.e., adaptive evolution on oceanic islands. It can be hypothesized that because of the restricted space available on islands, alleles not under selection might rapidly drift throughout all subpopulations of hybridizing species. In case the selection regime does not stabilize both species, the small ranges will cause the two species to rapidly become one morphospecies. This will be especially pronounced after secondary contact, e.g., by frequent dispersal between current islands or land bridges between paleo-islands (Puppo et al. 2014, 2015a).

This scenario might explain the comparable high levels of genetic diversity (Pérez de Paz and Caujapé-Castells 2013; García-Verdugo et al. 2015). In addition, hybridization can be quite frequent on islands. For example, Kim (2007) found that 34% of the genome in *Sonchus* (Asteraceae) had been exchanged between two species where hybridization has been observed, but the remaining genome had been hypothesized to be stabilized by selection.

Volcanic archipelagos present an ontogeny that is composed of different phases beginning with the growth of a sea mount above the sea level, its continuous building

until it reaches its maximum area and height, and its reduction below the sea level by erosion or other catastrophic events such as caldera collapsing and landslides created by volcanic activity (Fernández-Palacios et al. 2011). This continuous change in profile directly affects speciation opportunities by increasing or diminishing habitat availability as explained by the theory of island biogeography (MacArthur and Wilson 1967) and by the general dynamic model of oceanic island evolution (Whittaker et al. 2007, 2008).

One example of a volcanic archipelago is the Canary Islands, composed of seven islands located ca. 100 km off the western coast of Morocco in the Atlantic Ocean. The islands have each an independent origin, being oldest in the east and youngest toward the west (Carracedo 1994; Juan et al. 2000; Fernández-Palacios et al. 2011). Among the Canaries, Tenerife presents the most complex geological history and is currently the highest and largest island of the archipelago. Tenerife used to be three islands: Adeje (11.6–3.5 Ma), Teno (6.7–4.5 Ma) and Anaga (6.5–3.6 Ma), that got secondarily connected during the late Miocene—Pliocene due to successive volcanic activity (Ancochea et al. 1990). There is the possibility that Teno and Adeje created their own island but the three island hypothesis is more accepted (i.e., Ancochea et al. 1990; Guillou et al. 2004; Fernández-Palacios et al. 2011). Tenerife reached its current shape ca. 2 Ma (Ancochea et al. 1990) and parts of the paleo-islands remain in Tenerife today and exhibit distinct geomorphological and geological characteristics (Fernández-Palacios et al. 2011; Fig. 1). They also harbor unique floral elements: at least 55 plant species are endemic to at least one paleo-island (Trusty et al. 2005): 16 on Anaga, 25 on Teno, and 14 on the smallest paleo-island region, Adeje (Martín et al. 1999). The floristic differences between the paleo-island regions might have been further intensified by additional volcanic activity and catastrophic landslide events that might have reisolated parts of the island thus disconnecting existing populations (i.e., Mairal et al. 2015; Otto et al. 2016). From the many landslides occurred during the geological history of Tenerife, three massive ones stand out for creating the three major valleys in Tenerife. Güímar in the southeast and La Orotava in the northeast were formed between 800–600 ka and isolated Anaga from the rest of the island (Ancochea et al. 1990; Watts and Masson 1995; Juan et al. 2000; Fig. 1). Likewise, the valley of Las Cañadas in the north-center was formed less than 200 ka and reisolated Anaga and Teno (Ancochea et al. 1990; Fig. 1). The Teide volcano filled Las Cañadas becoming the highest point of Tenerife today (3718 m; Fig 1).

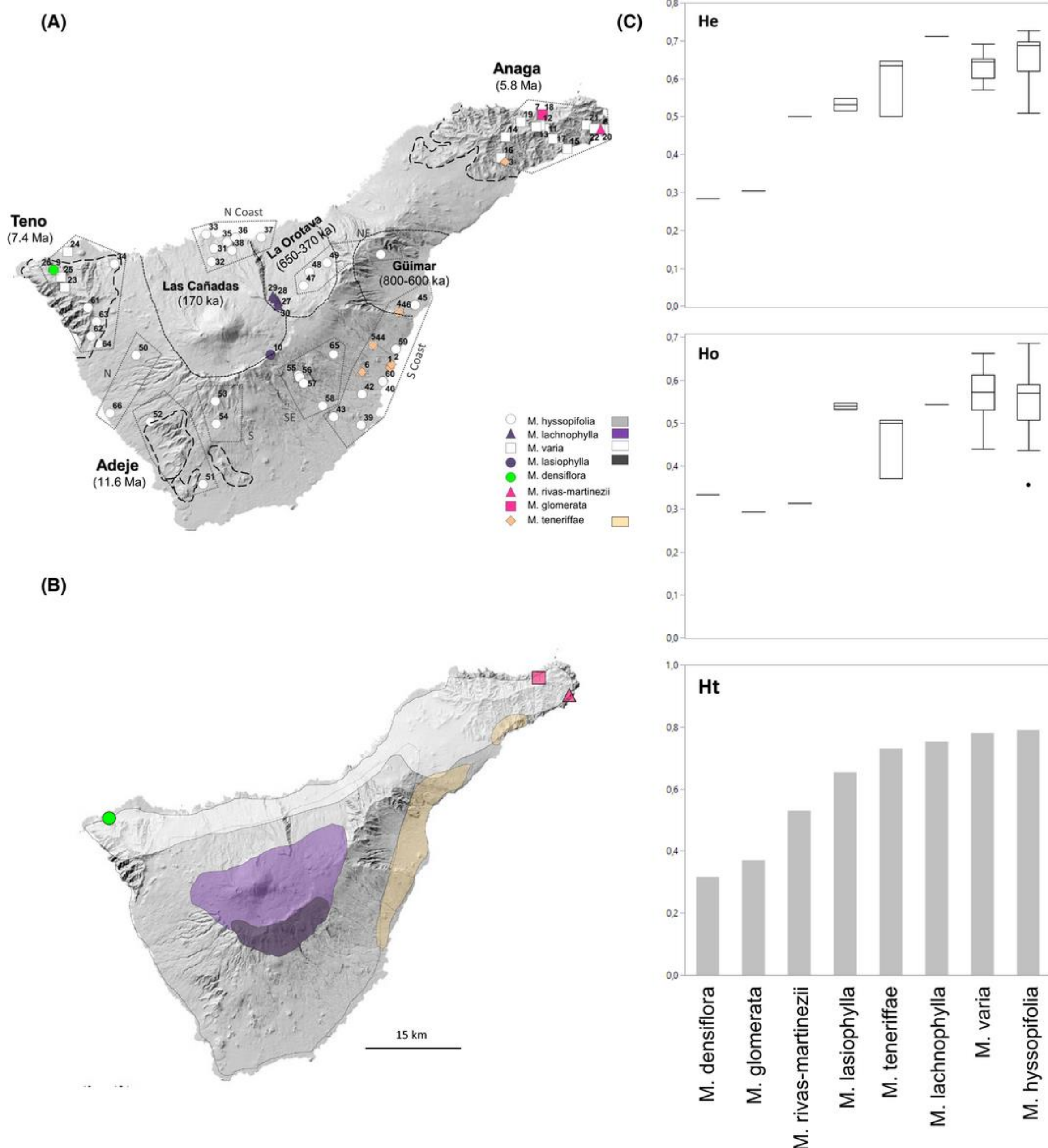


Fig. 1. Maps of Tenerife showing: (A) *Micromeria* sampling localities; long-dashed lines indicate remnants of paleo-islands; short-dashed lines indicate major valleys; dotted polygons indicate regions formed by geographically close populations (see Table 1); symbol shapes and colors correspond to different species of *Micromeria*; numbers on symbols indicate collection localities (see Table 1); (B) distribution of *Micromeria* species. Species distributions were obtained by converting a point per quadrant dataset from Pérez de Paz (1978) into continuous ranges. Individuals of *M. varia* on the central north coast had been assigned to *M. hyssopifolia* in the meanwhile (Puppo et al. 2014); (C) genetic diversity for each species calculated as H_E (upper right), H_O (middle right), and H_T (bottom right). The boxplots showing H_E and H_O were made from single values estimated per population.

The geomorphological history of Tenerife has not only had a strong influence on the composition of the regional flora but there are also examples that show its influence on population differentiation within species and potential impact on speciation. Examples are mainly from animals, where haplotype diversity seems correlated with the paleo-

islands with high haplotype divergence between Teno and Anaga (Gübitz et al. 2000; Brown et al. 2006; Macías-Hernández et al. 2013), though studies with plants are increasing over the last years (i.e., van Hengstum et al. 2012; Rumeu et al. 2014; Jones et al. 2014; Mairal et al. 2015). It had been postulated that this high divergence and patterns of genetic structure are not only explained by the geomorphological history such as secondary contact and reisolation by landslides and lava streams. Rather, the populations have been probably stabilized by selection, with reduced gene flow between genotypes characterized by the haplotypes and the different ecological conditions. For example, in the case of Gallotia lizards and *Tarentola* geckos, this is supported by different color patterns and other traits (Gübitz et al. 2000; Brown et al. 2006). In these examples, since differences are being maintained, the selection regime must be stabilizing the different species preventing them from forming a single morphotype.

Furthermore, in geologically complex islands such as Tenerife species ranges previously disrupted by volcanic activity, landslides, and other geological events could have later come into contact forming small-scale hybrid zones. Hybrid zones usually develop at zones of secondary contact between interbreeding species. In these zones, hybridization could be somewhat frequent, with introgression and backcross probability decreasing in both directions. The occurrence of hybrid introgression can be masked when sequence-based genetic markers are used to investigate the phylogeny of species (Herben et al. 2005). Multilocus investigations on insular species groups are comparably rare, only a few examples exist where dominant marker sets had been used (e.g., Meimberg et al. 2006; Mairal et al. 2015). Codominant markers are the method of choice to investigate genetic structure, gene flow and differentiation between populations because they allow determining the heterozygote state at one locus. Microsatellites or simple sequence repeat (SSR) are loci that show high level of length polymorphisms and constitute the method of choice for population genetic analyses, normally used for within species investigations. For species groups, they are more rarely applied because even though cross species applicability is observed, application can be technically challenging (Barbará et al. 2007). However, if markers can be identified that successfully amplify across a wider range of species, the use of this marker system allows determining gene-flow and differentiation between species (González-Pérez et al. 2009; Sosa et al. 2013; Turini et al. 2014).

In this paper, we are studying the context of geological history and population differentiation using multiple populations of the species of *Micromeria* Benth. on Tenerife, a genus that comprises paleo-island endemic representatives next to species that are widely distributed on the island. We use a set of 15 microsatellite markers able to cross amplify all *Micromeria* species from Tenerife (Puppo et al. 2015b), to investigate the

genetic structure of the species of *Micromeria* present in this island. With this, we aim to understand the diversification process of this genus in Tenerife, in particular, if the genetic structure can be related to the major geological events that occurred on the island. This is of particular interest for the central area species *M. hyssopifolia*, *M. lachnophylla*, *M. lasiophylla*, and *M. varia*. In addition, we investigate the role of hybridization in the evolution of *Micromeria* in Tenerife since natural hybrids had been described for most of the species of the genus occurring in this island. Introgression after hybridization could have combined Teno and Anaga genotypes and could have facilitated the adaptation to the different ecological niches.

The use of codominant markers and the possibility to determine gene flow within species allow outlining different hypotheses about the influence of hybridization on evolutionary patterns on oceanic islands. This will contribute to create a new perspective on speciation dynamics in oceanic islands: an interaction of gene flow and selection driven by geologic and climatic factors might shape evolutionary processes in these systems.

Materials and Methods

Study system, DNA isolation and genotyping

Micromeria is a genus of the mint family Lamiaceae, subfamily Nepetoideae, and is composed of ca. 54 species distributed in parts of Africa and Asia, the Mediterranean basin and Macaronesia (Bräuchler et al. 2008). *Micromeria* is present in the Canary Islands with 21 species, presenting the highest diversity on Tenerife and Gran Canaria, with 8 and 7 species, respectively (Puppo and Meimberg 2015). In Tenerife, three species are narrowly restricted to the paleo-islands, one to Teno (*M. densiflora*) and two to Anaga (M. glomerata and *M. rivas-martinezii*). *Micromeria teneriffae* also grows in Anaga but its range extends toward the southeast up to Fasnia and Güímar (Fig. 1). In the paleo-islands, these four species grow on old rocks and in the southeast, *M. teneriffae* inhabits the coastal desert. In a phylogenetic analysis of multiple nuclear genes and morphometric analysis, the species associated to the paleo-islands are not only highly morphologically different from those occupying the central area of the island, but are also older (Puppo et al. 2014). Contrary to this, relations among the common species, i.e., those distributed in the younger parts of the island (*M. varia*, *M. hyssopifolia*, *M. lachnophylla*, and *M. lasiophylla*), are less well supported in the phylogeny and further conclusions about their relationships could not be drawn (Puppo et al. 2014). *Micromeria varia* is distributed along the north part of the island from Teno to Anaga, *M. lachnophylla* grows in the central highland of the island above 2000 m, and *M. lasiophylla* is restricted to the southeast rock cliffs of Las Cañadas, above 2000 m (Fig. 1). The species with the

widest distribution, *M. hyssopifolia* occurs throughout the island from 0–2000 m and shows a high level of variability growing from costal desert in the south to the pine forest belt and the middle altitude wet regions in the north (Fig. 1). The species inhabiting this central part come into contact in zones where their distributions overlap and it is possible that hybrid zones between all the species exist.

In total, we included 289 samples of *Micromeria* in the present study representing all currently recognized species in Tenerife. Two to twelve individuals were collected in each of the 66 locations sampled (Table S1; Fig. 1). Collection was conducted in Tenerife during the years of 2010 and 2012 and leaves were conserved in silica gel in the field for subsequent DNA analysis.

Dried leaves were ground and DNA was extracted using the Macherey-Nagel Plant DNA Extraction Kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. The 289 samples were amplified with the 16 microsatellite markers developed for *Micromeria* by Puppo et al. (2015b). Each primer was tagged at the 5′- end with one of four different universal primers using the M13-tailed primer method as described in Curto et al. (2013) and Puppo et al. (2015b). The 16 primers were multiplexed in different polymerase chain reactions (PCR) as in Puppo et al. (2015b) using HotStarTaq Plus Master Mix Kit (Qiagen, Valencia, CA). The multiplex primer combination consisted of 4 nmol of each forward primer, 40 nmol of each reverse primer, and the florescent universal primer. The final volume reaction was 10 µL and contained: 5 µL of QIAGEN Multiplex PCR Master Mix (Qiagen), 1 µL of primer mix and 0.5 µL of template DNA (about 40 ng/µL), and 3 µL of water. PCR was performed using the following cycle profile: 95°C for 15 min; 7 cycles of 95°C for 30 sec; touchdown from 58°C to 55°C, decreasing 0,5°C per cycle for 45 sec; 72°C for 30 sec; 25 cycles of 95°C for 30 sec; 55°C for 45 sec; 72°C for 30 sec; 8 cycles of 95°C for 30 sec; 54°C for 45 sec; 72°C for 30 sec; and a final extension step of 60°C for 30 min. Amplification success was confirmed using 2% agarose gels stained with GelRed (Biotium, Hayward, CA). Genotyping was performed with an internal size standard (Genescan-500 LIZ; Applied Biosystems, Inc., Foster City, CA) in an ABI3130xl automatic sequencer (Applied Biosystems, Inc.). Alleles were called using GeneMapper ver. 4.0 (Applied Biosystems, Inc.). To check for reproducibility of the data, the amplification and scoring of 96 individuals were independently repeated for all primers mixes.

Data analyses

For the population level analyses, only localities with at least four individuals sampled were considered. To better understand how the estimates vary across the

island regions geographically close localities within the same habitat were considered as one population for some analyses (Fig. 1).

Microsatellite quality was evaluated by quantifying the frequency of null alleles and searching for evidence of genotyping errors such as scoring of stuttering bands and large allele drops. This was performed with the program Micro-Checker (Van Oosterhout et al. 2004) and only populations with at least five individuals with less than 50% missing data for all markers were used. Additionally, we tested if they followed all assumptions from Hardy–Weinberg Equilibrium (HWE) using the program GenAlEx 6.41 (http://biology-assets.anu.edu.au/GenAlEx/).

Genetic diversity per population was estimated by calculating the total number of alleles (N), expected and observed heterozygosities (H_E and H_O), and portion of private alleles. To prevent biases due to population size, the total unbiased H_E per species and regions was calculated (H_T). Genetic differentiation was estimated by calculating pairwise F_{ST}, R_{ST}, and Nei distance; R_{ST}, to include the information about allele size when using microsatellites in the distance estimate. This allows to have a better perspective of the evolutionary relationships among groups (Balloux and Goudet 2002). The pairwise matrices for genetic differentiation measures were represented by an UPGM dendrogram, calculated using the program NTSys pc (Rohlf 1993). Deviations from Hardy–Weinberg equilibrium (HWE) were estimated for each population. All these statistics were calculated using the program GenAlEx. The existence of changes in population sizes was evaluated with the program BOTTLENECK v. 1.2.02 (Cornuet and Luikart 1997) under the Stepwise Mutation Model (SMM). Since a low number of loci were used, significant deviations from the mutation-drift equilibrium were calculated using Wilcoxon signed-rank (Piry et al. 1999).

Analyses of Molecular Variance (AMOVA) were conducted in GenAlEx 6.41 using RST as the measure of differentiation. This was done to access the distribution of genetic variation within and among several species groupings. The different groupings that had been considered are: paleo-island species versus central species; different species within paleo-islands; different species within the central region.

Genetic structure between and within species was also investigated using the Bayesian clustering algorithm implemented in the program STRUCTURE ver. 2.3.3 (Hubisz et al. 2009) and using Principal Coordinates Analysis (PCoA) calculated in GenAlEx 6.41. Creating prior decisions of how taxa are structured may lead to circular conclusions. For these reasons, STRUCTURE was run assuming an admixture model of population structure with default settings for inferring alpha and without any location or population priors. Moreover, it was run with and without considering the allele frequencies to be correlated among populations. To determine the number of K

(unknown) genetic clusters, K was set to range from 1 to 15; the program was run as 10 iterations of 500,000 MCMC generations with a burn-in of 100,000 generations for each K. The most likely K was selected by analyzing the second-order rate of change of the posterior probability of the data (DK) between successive K values (Evanno et al. 2005) using Structure Harvester v.0.6.9.3 (http://taylor0.biology.ucla.edu/structureHarvester/). Additionally, the suboptimal value of K was searched by redoing the DK test without the optimal and smaller values of K. This allowed us to investigate more detailed structure signal shown by our data. All 10 iterations were combined using the greedy algorithm from the program CLUMPP (Jakobsson and Rosenberg 2007) For better interpretation of the results, this analysis was performed for three datasets: a first one containing all samples; a second one containing only central species considered by Puppo et al. (2014) as young lineages (*M. varia*, *M. hyssopifolia*, *M. lasiophylla*, and *M. lachnophylla*), and a third one containing only *M. varia* and *M. hyssopifolia*.

We calculated historical and contemporary migrations rates between all species pairs as proxy of gene flow using the programs MIGRATE v3.2.1 (Beerli and Felsenstein 2001) and BAYSASS v3.0 (Wilson and Rannala 2003), respectively. MIGRATE estimates the number of migrants per generation while BAYSASS calculates the portion of individuals originated from the foreigner population. Because of the genetic structure and spatial distance between M. varia from Teno and Anaga, these were considered as two distinct groups. Two independent replicates were performed for each analysis and the average migration rate values are presented. For MIGRATE, these migration rate corresponds to the number of individual migrants per generation from the source population. While for BAYSASS, these correspond to the portion of migrant individuals in the sink originating from the source population. We considered a high migration rate to be above 10 individuals per generation for MIGRATE and 10% for BAYSASS in accordance to previous studies (i.e., Bertrand et al. 2014; Conflitti et al. 2014; Peacock et al. 2015).

MIGRATE was run considering the data under the Brownian motion model and implementing a Bayesian search strategy. One long chain was run saving 25,000 generations with sampling increments of 100 generations after a burnin step of 10,000 generations. We defined the maximum prior boundaries of theta and migration rate to be 200 and 1000, respectively. As recommended by Beerli and Palczewski (2010), a static heating scheme was applied with four temperatures of 1, 1.5, 3, and 1 × 106.

Several test runs were performed with BAYSASS to optimize the acceptance rates and the number of generations that should be excluded in the burnin step as recommended in the program's manual. For each run, trace files were saved and analyzed using the program TRACER v1.5.0 (Rambaut and Drummond 2007). In the

final analyses, BAYESASS ran for 20,000,000 generations with a burnin of 2,000,000 and sampling increment of 200. The experimental run with the best acceptance rates (below 0.6) had the DeltaA and DeltaF parameter set to 0.4 and DeltaM to 0.1. For this reason, we used these values for the main analyses.

Because some morphological hybrids were found in our sampling, we tested for the likelihood of them being real hybrids by doing a STRUCTURE analyses with the individuals from the same localities in which they were found. With this approach, we expect that hybrid individuals will show an equal assignment to the clusters from the parent species. This result is only considered to be valid if both species are clearly differentiated (K = 2). Morphological intermediate individuals were found in the field between *M. densiflora* and *M. varia* in Teno, *M. rivas-martinezii*, and *M. varia* in Anaga, *M. teneriffae* and *M. varia* in Anaga, and *M. teneriffae* and *M. hyssopifolia* in the south coast. We performed a STRUCTURE analysis for each species pair with the parameters described above.

Results

Genetic diversity

From the 16 microsatellite markers included, one (5978) presented low amplification success (<50%), so only 15 SSRs were used for further analysis. The remaining markers comprised between 11 and 25 alleles, giving a total of 273 analyzed alleles. None of the analyzed populations deviated significantly from Hardy–Weinberg equilibrium for most of the loci. A few deviations were indicated with near marginal P values ($P < 0.05$) and only for a few loci and single populations. No locus deviated from Hardy–Weinberg equilibrium across the majority of populations meaning that all its assumptions such as neutrality were met. The same was observed the other way around: no population deviated from HWE for most of the loci analyzed (Table S2). There was no evidence of scoring errors and none of the markers constantly showed high frequency of null alleles in the populations analyzed. The 15 loci investigated were therefore retained in the analysis.

Across all populations, mean number of alleles (N) varied from 4.20 (*M. densiflora*, *M. lasiophylla*) to 11.27 (*M. lachnophylla*), H_O from 0.29 (*M. glomerata*) to 0.62 (*M. hyssopifolia*), and H_E from 0.28 (*M. densiflora*) to 0.71 (*M. lachnophylla*), H_T from 0.32 (*M. densiflora*) to 0.81 (*M. hyssopifolia*) (Table 1). Expected heterozygosity increased with range size (Fig. 1), i.e., smaller diversity was found in the restricted paleo-island species and highest diversity was found in the most widespread species *M. lachnophylla*, *M. varia*, and *M. hyssopifolia*. Genetic diversity of populations and groups of populations were generally similar within one species. Slight differences were found

in *M. hyssopifolia* which seems to have the highest genetic diversity in the southern coast (H_E = 0.70, H_O = 0.62, and H_T = 0.81). In *M. teneriffae*, the populations from the Southern coast showed slightly lower diversity (H_E = 0.57, H_O = 0.44, and H_T = 0.72) than the population from Anaga (H_E = 0.65, H_O = 0.51, and H_T = 0.72). No differences in genetic diversity were found between the two regions (Anaga and Teno) where *M. varia* grows.

Table 1. Genetic variation statistics per regions and species of *Micromeria*. This table contains information regarding number of populations (Pops.); average number of individuals (Ind.); average number of alleles (N), observed (H_O) and expected heterozygosity (H_E); portion of private alleles (Priv. Al.), and total heterozygosity (H_T).

Region	**Anaga**			**Teno**				**Adeje**
Species	***M. rivas-martinezii***	***M. teneriffae***	***M. varia***	***M. glomerata***	***M. densiflora***	***M. hyssopifolia***	***M. varia***	***M. hyssopifolia***
Nr. Pops.	1	1	8	1	1	3	3	2
Av. Nr. Ind.	11	6	6.5	5	5	6.33	7	9.5
N	9.8	5.27	5.98	4.27	4.2	6.07	6.42	9.03
H_O	0.31	0.51	0.57	0.29	0.33	0.46	0.57	0.54
H_E	0.5	0.65	0.64	0.31	0.28	0.63	0.63	0.61
P. Priv. Al.	–	–	0.15	0.13	0.27	0.07	0.11	0.07
H_T	0.53	0.72	0.75	0.37	0.32	0.75	0.76	0.7

Table 1. Continued

Region	**Teide**		**Southern coast**		**Northeast**	**Northwest**	**Northern coast**	**Southeast**
Species	***M. lachnophylla***	***M. lasiophylla***	***M. teneriffae***	***M. hyssopifolia***	***M. hyssopifolia***			
Nr. Pops.	1	2	2	3	3	1	8	3
Av. Nr. Ind.	12	4.5	5	5.67	6	6	1.63	8
N	11.27	4.2	4.57	5.42	5.44	5.87	–	7.78
H_O	0.54	0.54	0.44	0.62	0.59	0.52	–	0.54
H_E	0.71	0.53	0.57	0.7	0.67	0.65	–	0.67
P. Priv. Al.	0.13	0.1	0.07	0.089	0.111	0.267	–	0.133
H_T	0.75	0.65	0.72	0.81	0.8	0.72	0.81	0.75

The number of alleles private to a particular species was generally low (Table 1) and no correlation to species range was obvious. Only in *M. densiflora* and *M. hyssopifolia* from the Northwest, more than 20% of alleles were private. *Micromeria rivas-martinezii* and *M. teneriffae* from Anaga did not show any private allele. The private alleles found within a species also tended to be rare. For example, only private alleles in *M. densiflora* and *M. lasiophylla* had a frequency within species above 10% not shown. Frequency of the remaining alleles private to a species was below 10% with an average of 3.5%.

Four of the analyzed populations significantly deviated from the mutation-drift equilibrium ($P < 0.05$) suggesting that they went through a bottleneck event (Table S2). These were the populations from *M. glomerata* and *M. densiflora*, one population from *M. varia* from Teno, and one population from *M. hyssopifolia* from the Southeast.

Genetic Structure

For all populations, the pairwise F_{ST} values were highly significant ($P < 0.001$), varying from 0.042 to 0.500 (Table S3). F_{ST} was correlated to species age, with the older species (*M. glomerata*, *M. rivas-martinezii*, and *M. densiflora*) presenting higher pairwise F_{ST} values than the youngest (*M. varia* and *M. hyssopifolia*). The pairwise unbiased Nei (uNei) distance showed similar patterns to the F_{ST} values. R_{ST} was calculated among island regions and used to evaluate genetic distance patterns among them (Fig. 2). As expected, the paleo-island species were the most dissimilar. *Micromeria lasiophylla* and *M. lachnophylla* appear as sister branches to the remaining central species. *M. varia* and *M. hyssopifolia* were mostly grouped according to geographical position. For example, the populations from both species from Teno grouped together.

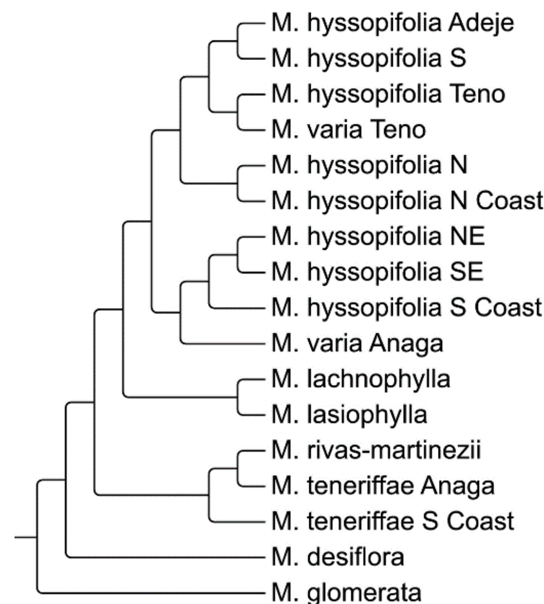


Fig. 2. UPGMA of pairwise unbiased uNei distances and R_{ST} among population groups of *Micromeria* according to island regions; regions are those showed in Fig 1 and Table 1.

We performed four independent AMOVA tests using different groupings: all species; paleo-island species versus central species; different species within paleo-islands; different species within the central region (Tables 2). The highest amount of variation among groups was explained by differences among paleo-island species (29%) followed by differences among all species (11%). Difference between paleo-islands species and central species was explained by 8% of variation. Difference among central

species was explained by the lowest amount of variation in the dataset (3%). These results are concordant with the analyses of pairwise FST and RST, where higher differentiation is found among paleo-island species and lower among central species.

Table 2. AMOVA analyses of four groupings calculated using RST. The results presented in a percentage form correspond to the amount of variation explained by differences within and among groups

Grouping	Number of groups	Number of individuals	Among groups (%)	Within groups (%)
Among all species	8	289	11	89
Among central species	4	245	3	97
Among Paleo-island species	4	44	29	71
Central species versus Paleo-island species	2	289	8	92

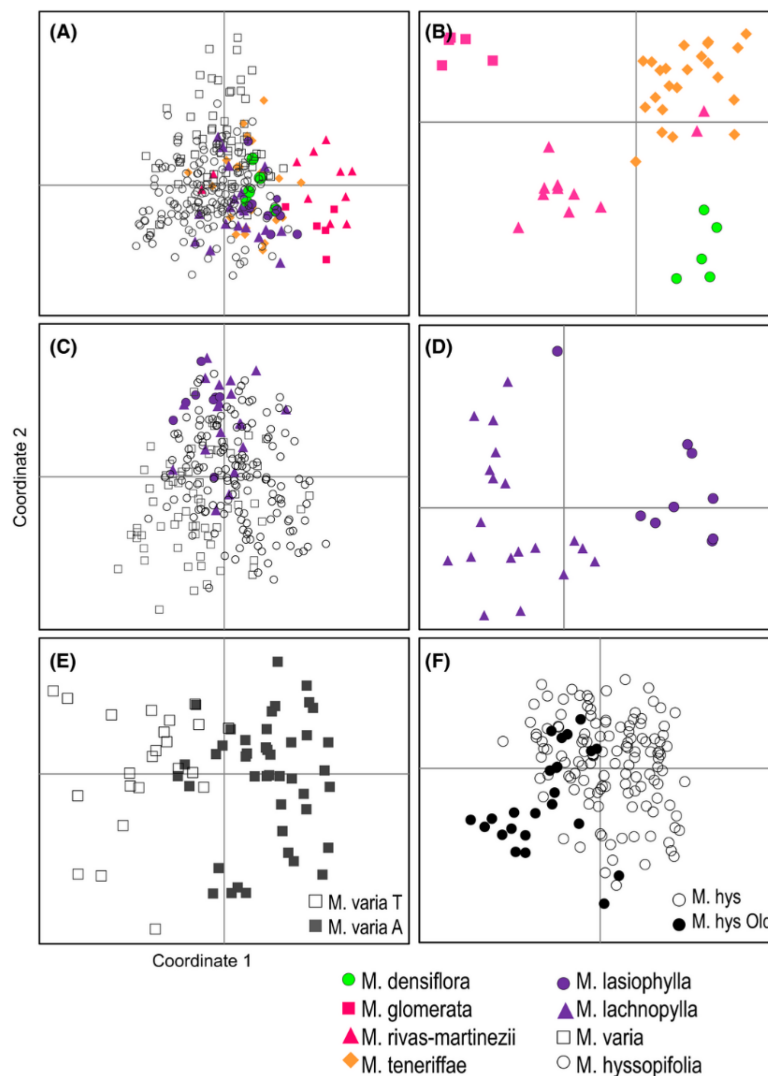


Fig. 3. Principal Coordinates Analyses (PCoA) of pairwise distances of individuals of *Micromeria* implemented in GeneAlEx for codominant datasets. Shown are the first two coordinates of analyses including: A. all species; B. only paleo-island species; C. only central area species; D. only *M. lasiophylla* and *M. lachnophylla*; E. only *M. varia* divided in samples from Anaga (A) and Teno (T); F. only *M. hyssopifolia* divided in samples from Adeje (Old) and the remaining samples.

When pairwise differences are visualized by PCoA, *M. glomerata* and *M. rivas-martinezii* are separating from the others (Fig. 3A). When only the paleo-island species are included (*M. teneriffae*, *M. glomerata*, *M. rivas-martinezii*, and *M. densiflora*), the PCoA shows four clusters corresponding to each species (Fig. 3B). The analysis including only the central species (*M. lasiophylla*, *M. lachnophylla*, *M. varia*, and *M. hyssopifolia*) shows no separation of the samples (Fig. 3C). When only the central species with narrow range (*M. lasiophylla* and *M. lachnophylla*) are analyzed, there is a distinction among them (Fig. 3D). When M. varia is analyzed separately, samples from Anaga slightly segregate from the rest (Fig. 3E). Although the analysis including only *M. hyssopifolia* shows no obvious subdivisions of the samples, there is a weak signal of subdivision between individuals located in older and younger parts of the island (Fig. 3F).

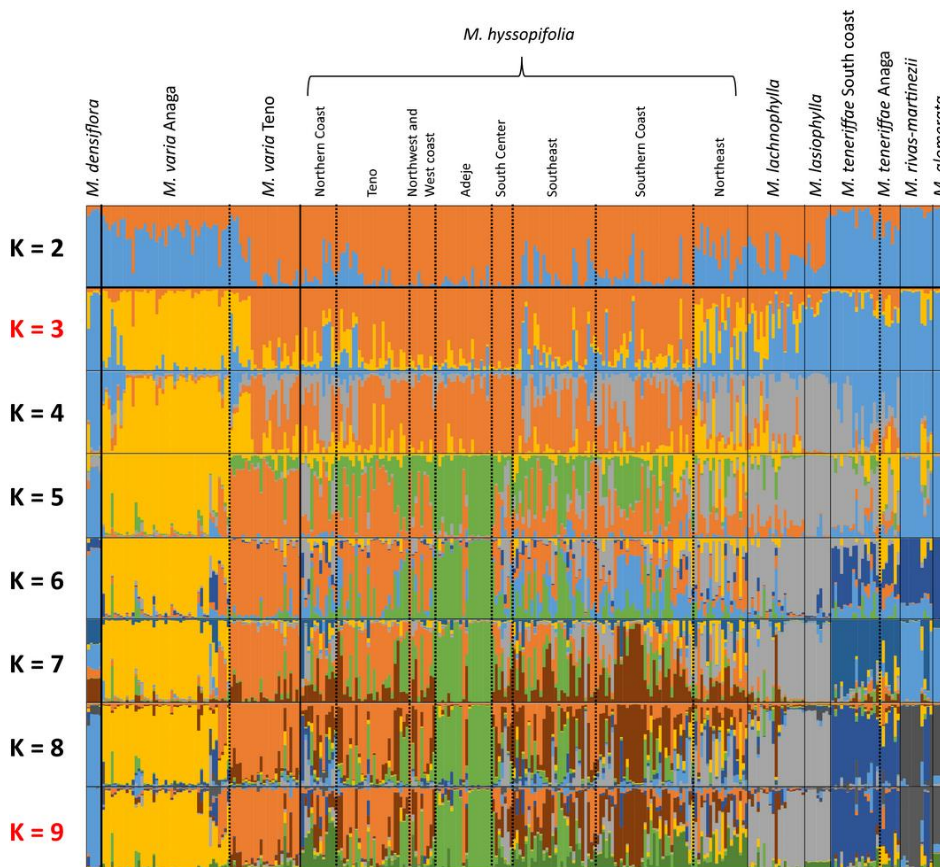


Fig. 4. STRUCTURE analyses of the species of *Micromeria* present in Tenerife showing blots of assignment probability from K values ranging from K = 2 to K = 9; optima K according to the Evanno method are indicated in red: K = 3 for all runs and K = 9 when only K = 4 to K = 15 are analyzed.

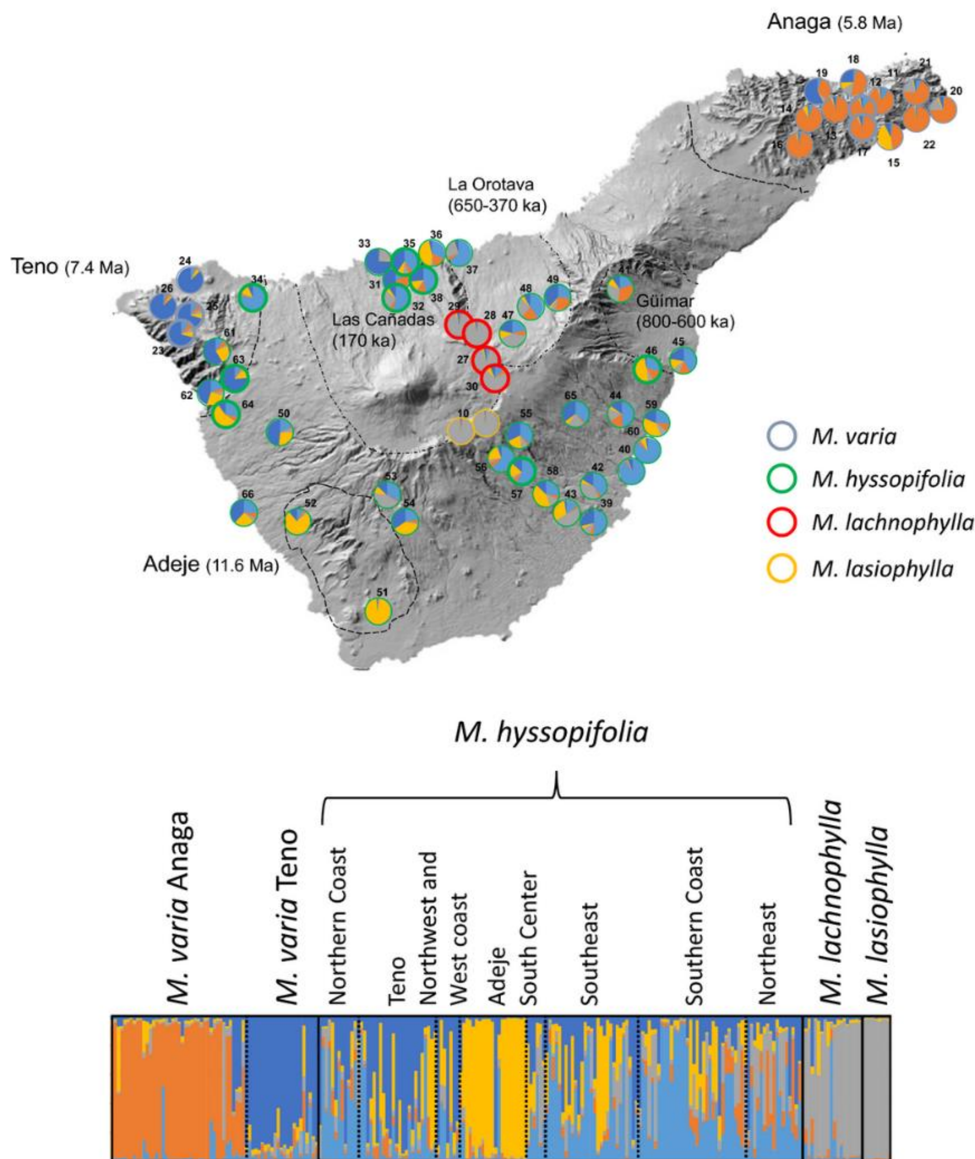


Fig. 5. Suboptimum K (K = 5) for the analyses including only central area species and assignment probability plotted per population on the map. The structure plot is shown to provide a context for the colors shown in the map.

In the STRUCTURE analysis, an optimal K = 3 was obtained according to Evanno et al. (2005) method. If results between K = 4 and K = 15 are tested, optimal K is K = 9. At K = 9, STRUCTURE analysis resolves all species with the exception of *M. lachnophylla* and *M. lasiophylla*. The results from the structure analysis at different values of K are summarized in Figure 4. The Delta K plots obtained with STRUCTURE Harvester for all STRUCTURE tests performed are included in Fig. S1. At K = 2, *M. varia* from Teno, *M. lachnophylla*, *M. lasiophylla* and *M. hyssopifolia* are forming one of the clusters. At K = 3, M. varia samples collected in Anaga are forming an additional cluster. At K = 5*, M. hyssopifolia* samples from Adeje are forming their own cluster, and with increasing K, *M. hyssopifolia* becomes more and more subdivided. When the central species (*M. varia*, *M. hyssopifolia*, *M. lachnophylla* and *M. lasiophylla*) are analyzed

independently, this differentiation within *M. hyssopifolia* is clearer. For example, for K values higher than 7 one of the clusters is mainly composed of *M. hyssopifolia* samples from the southern coast from subdesert environments, while another cluster is mainly composed of individuals form the wet northern coast. Moreover, samples of *M. hyssopifolia* from Teno share the same cluster with samples of *M. varia* from this same region. The best K for the analysis including only the central species was also K = 3 and the suboptimum is K = 5 (Fig. 5). Although *M. lasiophylla* and *M. lachnophylla* do not separate from each other in these runs, with higher values ok K they do. Results were the same for correlated and not correlated allele frequencies, so only analysis with correlated frequencies is shown.

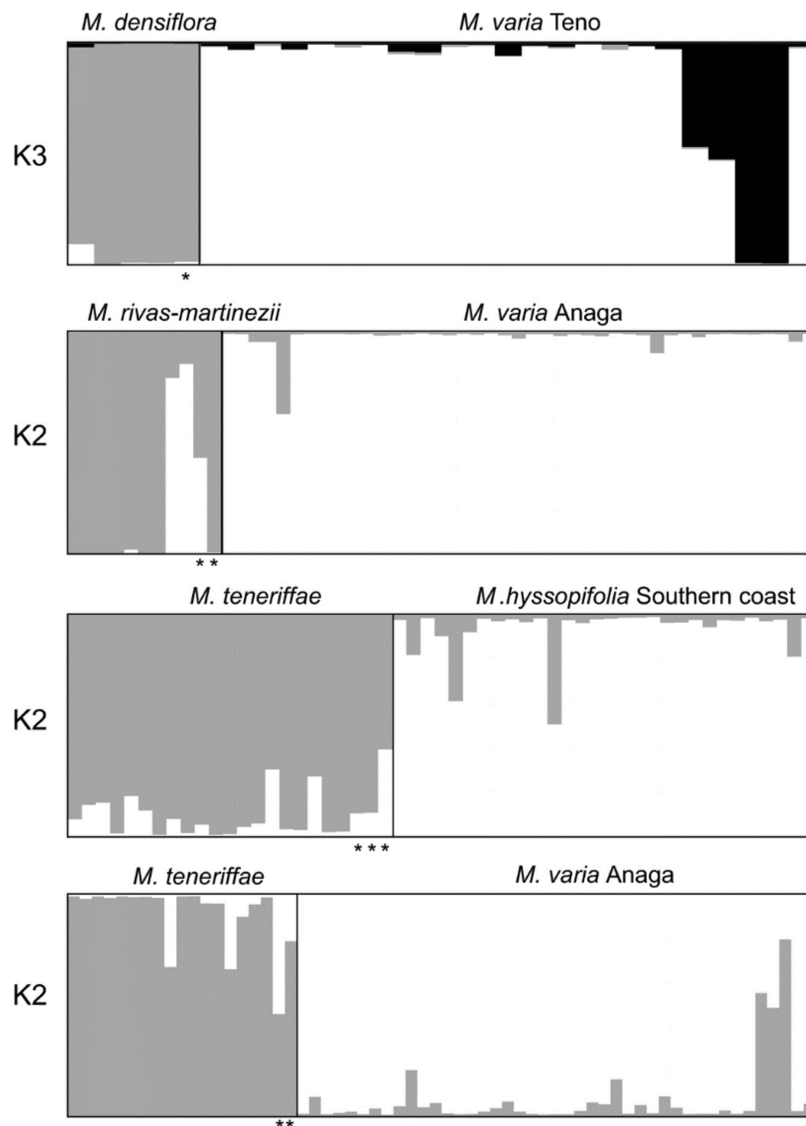
Gene flow and hybridization

Several individuals had been determined as hybrids because they present morphologically intermediate characteristics from two species. In a STRUCTURE analysis together with the putative parental species, the hybrid status of most of these individuals were confirmed. According to the DK method, the best K was K = 3 for the *M. densiflora* and *M. varia* dataset and K = 2 for the remaining species pairs (Fig. 6). From the two morphological hybrids between *M. rivas-martinezii* and *M. varia*, one showed an almost equal assignment to both clusters (44% assignment to *M. varia* cluster), while the other was assigned to the *M. rivas-martinezii* cluster so it is likely a backcross. In addition to these hybrids, two individuals that were morphologically identified as *M. rivas-martinezii* showed an almost complete assignment to *M. varia* evidencing introgression between both species. In the analysis between *M. teneriffae* and *M. hyssopifolia*, only one morphological hybrid could be confirmed with high assignment rates to both clusters (39%, of assignment to *M. hyssopifolia* cluster). Two *M. hyssopifolia* individuals showed mixed assignment (41% and 52% to the *M. teneriffae* cluster) suggesting them as hybrids or backcrosses. For the *M. varia* and *M. teneriffae* analysis, only one of the morphological hybrids was confirmed (54% assignment to *M. varia* cluster). Additionally, three *M. varia* individuals showed a high assignment to the *M. teneriffae* cluster (50% to 81%).

BAYSASS and MIGRATE were used to estimate contemporary and historical gene flow among species, respectively. Contemporary gene flow as indicated by BAYSASS was generally low showing migration rates below 10% of individuals originated from other populations for most of the comparisons (Table 3). The exceptions were migration rates from *M. densiflora* to *M. lasiophylla* (17%), *M. glomerata* to *M. rivas-*

martinezii (16%), *M. lachnophylla* to *M. hyssopifolia* (25%), and *M. varia* from Teno to *M. hyssopifolia* (26%).

Fig. 6. STRUCTURE analyses of potential hybrid individuals of *Micromeria* between four species pairs*: M. densiflora* and



M. varia; *M. rivas-martinezii* and *M. varia; M. teneriffae* and *M. hyssopifolia* from the South; and *M. teneriffae* and *M. varia* from Anaga. Only the optimal K values according to the Evanno method are presented. The individuals marked with * were identified as morphological hybrids.

The historical migration rates calculated by MIGRATE varied between 1.4 and 19.1 individuals per generation (Table 4). *Micromeria hyssopifolia* showed to be the main source of interspecific gene flow because it had the highest migration rate (to *M. lachnophylla*). And, from the eight comparisons, five showed migration rates above 10 individuals per generation. *Micromeria varia* was the second main source of migrants, with both *M. varia* from Teno and *M. varia* from Anaga showing three migration comparisons above 10 individuals per generation. *Micromeria rivas-martinezii* was the main sink population because it received more than 10 migrants per generation from six

other species. The species with lowest emigration and immigration, less than 10 individuals per generation, were *M. rivas-martinezii* and *M. glomerata*, respectively. Some loci showed higher values of migration rate than others. On average, the overall migration rate per locus varied from 4.66 individuals per generation for locus 5419 to 175.41 for locus 3963 (not shown).

Table 3. Contemporary migration rates calculated by BAYESASS between species of *Micromeria*. Results are presented in percentage (%) of individuals from species on top originated from the species in the left; values for migration within taxa are highlighted in gray; values above 10% are presented in bold; standard deviation values are in parentheses.

Source/Sink	*M. densiflora*	*M. glomerata*	*M. hyssopifolia*	*M. lachnophylla*	*M. lasiophylla*
M. densiflora	69.44 (2.39)	1.9 (1.8)	1.94 (1.81)	1.93 (1.81)	**16.98 (4.44)**
M. glomerata	1.93 (1.82)	69.46 (2.41)	1.93 (1.83)	1.94 (1.83)	1.93 (1.81)
M. hyssopifolia	0.21 (0.21)	0.21 (0.21)	97.78 (0.69)	0.21 (0.21)	0.24 (0.24)
M. lachnophylla	1.01 (0.96)	1.01 (0.96)	**24.92 (2.56)**	67.91 (1.17)	1.02 (0.98)
M. lasiophylla	1.84 (1.73)	1.85 (1.73)	1.86 (1.76)	1.85 (1.76)	84.91 (3.86)
M. rivas-martinezii	1.64 (1.56)	1.62 (1.54)	1.63 (1.56)	1.63 (1.56)	1.63 (1.56)
M. teneriffae	1.05 (1.01)	1.06 (1.03)	2.46 (1.8)	1.05 (1.01)	1.06 (1.02)
M. varia Anaga	0.64 (0.62)	0.64 (0.63)	0.87 (0.83)	0.65 (0.64)	0.91 (0.81)
M. varia Teno	0.9 (0.87)	0.9 (0.86)	**25.96 (2.3)**	0.9 (0.87)	0.9 (0.86)

Table 3. Continued

Source/Sink	*M. rivas-martinezii*	*M. teneriffae*	*M. varia* Anaga	*M. varia* Teno
M. densiflora	1.91 (1.81)	2.03 (2.01)	1.94 (1.83)	1.92 (1.81)
M. glomerata	**15.59 (6.06)**	3.37 (4.98)	1.92 (1.81)	1.93 (1.8)
M. hyssopifolia	0.39 (0.3)	0.29 (0.29)	0.45 (0.33)	0.22 (0.21)
M. lachnophylla	1.01 (0.97)	1.09 (1.05)	1.01 (0.97)	1.02 (0.97)
M. lasiophylla	2.04 (1.91)	1.95 (1.85)	1.85 (1.75)	1.86 (1.76)
M. rivas-martinezii	81.43 (5.88)	7.14 (6.1)	1.64 (1.56)	1.64 (1.57)
M. teneriffae	1.06 (1.02)	89.06 (2.96)	2.18 (1.61)	1.04 (1.01)
M. varia Anaga	0.72 (0.7)	0.78 (0.77)	94.15 (1.82)	0.65 (0.64)
M. varia Teno	0.89 (0.86)	0.91 (0.88)	0.9 (0.87)	67.75 (1.03)

Table 4. Historical migration rates calculated by MIGRATE between species of Micromeria. Results are presented in average number of individuals per generation; values correspond to migration rates from the species in the left to the species on top; values for migration within taxa are highlighted in gray; values above 10 migrants per generation are presented in bold; values for 95% confidence intervals are in parentheses

Source/Sink	*M. densiflora*	*M. glomerata*	*M. hyssopifolia*	*M. lachnophylla*	*M. lasiophylla*
M. densiflora	–	6.56 (0–23.33)	4.56 (0–21.33)	7.17 (0–24)	8.97 (0–25.33)
M. glomerata	5.36 (0–22)	–	4.06 (0–20.67)	2.57 (0–18.67)	6.27 (0–23.33)
M. hyssopifolia	**10.74 (0–26.67)**	1.41 (0–18)	–	**19.06 (1.33–36.67)**	8.88 (0–27.33)
M. lachnophylla	4.53 (0–21.33)	5.79 (0–22.67)	6.09 (0–22.67)	–	**14.91 (0–30.67)**
M. lasiophylla	8.17 (0–24.67)	5.23 (0–22)	5.85 (0–22)	5.97 (0–22.67)	–
M. rivas-martinezii	6.07 (0–22.67)	3.13 (0–19.33)	4.75 (0–21.33)	8.75 (0–24.67)	4.78 (0–21.33)
M. teneriffae	7.48 (0–24)	3.82 (0–20.67)	9.36 (0–25.33)	3.86 (0–20)	**10.19 (0–26.67)**
M. varia Anaga	7.19 (0–24)	3.65 (0–20)	**10.72 (0–26.67)**	9.14 (0–25.33)	7.94 (0–24.67)
M. varia Teno	**11.75 (0–27.33)**	3.8 (0–20)	**12.64 (0–28)**	6.49 (0–23.33)	4.45 (0–21.33)

Table 4. Continued

Source/Sink	*M. rivas-martinezii*	*M. teneriffae*	*M. varia* Anaga	*M. varia* Teno
M. densiflora	**13.55 (0–29.33)**	3.42 (0–20)	4.21 (0–20.67)	3.95 (0–20.67)
M. glomerata	**10.61 (0–26.67)**	**10.46 (0–26.67)**	4.07 (0–20.67)	9.42 (0–26.67)
M. hyssopifolia	**14.9 (0–30.67)**	3.62 (0–21.33)	**17.66 (0.67–34.67)**	**13.11 (0–29.33)**
M. lachnophylla	5.86 (0–22)	**10.89 (0–26.67)**	7.84 (0–24)	6.21 (0–22.67)
M. lasiophylla	9.14 (0–25.33)	5.9 (0–22.67)	**10.96 (0–26.67)**	4.89 (0–21.33)
M. rivas-martinezii	–	5.8 (0–22.67)	5.67 (0–22)	6.96 (0–23.33)
M. teneriffae	**10.08 (0–26.67)**	–	9.99 (0–26)	4.4 (0–21.33)
M. varia Anaga	**10.38 (0–26.67)**	**10.5 (0–26.67)**	–	9.55 (0–26)
M. varia Teno	**13.77 (0–29.33)**	9.23 (0–25.33)	8.25 (0–24.67)	–

Discussion

Geomorphological impact on genetic structure

In geologically complex islands such as Tenerife, secondary connection of previously isolated parts, successive volcanic activity, caldera collapses, landslides, etc, could have produced a strong impact on the diversification of its species (Whittaker et al. 2007, 2008; Fernández-Palacios et al. 2011). Several molecular studies in different organisms have found diversification patterns coinciding with the different geological events in Tenerife (e.g., Juan et al. 2000; Carine et al. 2004; Moya et al. 2004; Trusty et al. 2005; Mairal et al. 2015). In *Micromeria*, Puppo et al. (2014) showed that species restricted to the paleo-islands are early diverging lineages and are older than the central area species. Hereby the restricted ranges of *M. densiflora* from Teno, *M. glomerata* and *M. rivas-martinezii* from Anaga can be interpreted as contracted ranges, remnants of an earlier, wider distribution, while the range of *M. teneriffae* can be regarded as a shift from

Anaga to the surrounding areas after the uprising of the Teide (Puppo et al. 2014). In the present analysis, we found that the highest differentiation is between these four species restricted to the paleo-islands, which is in accordance to Puppo et al. (2014) phylogenetic hypothesis. The AMOVA results also support this previous study since higher variation was found among paleo-endemic species than among central species. Since these species are older, they had more time to accumulate genetic differences and are more reproductively isolated. The low differentiation between paleo-island and non-paleo-island species might be explained by the fact that high genetic variation found among paleo-island species is increasing the variation within groups.

The distance analysis of pairwise R_{ST}, is highly congruent with the previous phylogenetic inferences. In both analyses, the paleo-island species are clustering independently from the central species group. The difference is mainly in the most widespread species: using the microsatellite dataset, they are positioned more pronouncedly according to geography. For example, species from Teno are always clustering together while *M. varia* from Anaga appears together with geographically proximate *M. hyssopifolia* populations. The same is observed for *M. lasiophylla* and *M. lachnophylla* that occur on high altitude in the Teide Mountain. This might be a result of gene flow between the respective populations and is further discussed below.

Genetic diversity of the restricted species was lower than the common species, indicating the possibility that their ranges are contracted. This is supported also by the bottleneck analysis for the populations of *M. densiflora* and *M. glomerata*.

Our study shows that the two most widespread species on the island, *M. varia* and *M. hyssopifolia,* present a genetic structure that is highly correlated to the geological composition of Tenerife. In *M. varia*, samples from Teno and from Anaga are assigned to two different clusters. Samples of *M. hyssopifolia* from Teno cluster together with the *M. varia* samples from this region. This clustering is already indicated in the STRUCTURE analysis when $K = 2$ and is also evident in the PCoA. Additionally, the optimal division in STRUCTURE corresponds to the appearance of a unique cluster of *M. varia* from Anaga showing that this corresponds to a deep divergence. *Micromeria varia* is assumed to be distributed along the northern part of Tenerife from Teno to Anaga. However, samples from the central part of the northern coast have been identified as a different subspecies of *M. hyssopifolia*, subsp. *glabrescens* (sensu Pérez de Paz 1978). Therefore, *M. varia* might be restricted to the paleo-islands. Hence, the genetic structure observed might be either a consequence of the ancestral split of the two paleo-islands or a consequence of the reisolation of Anaga after the central shield was formed. For example, Anaga was reisolated by several events such as two massive landslides in the north of Tenerife: one occurred ca. 650–370 ka giving origin to La

Orotava valley, the second ca. 170 ka formed Las Cañadas Caldera (Ancochea et al. 1990; Watts and Masson 1995; Juan et al. 2000). The populations of *M. varia* from these two paleo-islands might have been isolated since then. In our previous work (Puppo et al. 2014), we found that *M. varia* from Anaga was grouped together with the older lineages resulting in a separation from Teno before these landslides. This was assumed to be a consequence of hybridization of *M. varia* populations with the other Anaga species. However, now more Anaga populations are included and all show the same pattern. Other events might have contributed to the isolation of both *M. varia* groups. As in *M. varia*, the divergence between Teno and Anaga populations has been observed in at least two other plant species, *Hypericum canariense* (Clusiaceae; Dlugosch and Parker 2007) and *Canarina canariensis* (Campanulaceae, Mairal et al.2015), and also in studies of mitochondrial haplotype diversity in several animal groups (e.g., Gübitz et al. 2000; Brown et al. 2006). It had been suggested that this difference stems from habitat discontinuities between and within paleo-islands that causes strong divergent selection and impedes migration (Gübitz et al. 2000; Moya et al. 2004).

Similar to the structure observed within *M. varia*, genetic divergence within *M. hyssopifolia* seems also related to the paleo-islands, in particular since these samples were assigned to multiple clusters in the STRUCTURE plot. Especially evident is the segregation of the individuals from Adeje which is the oldest paleo-island. Differently from Teno and Anaga which are forming rather independent shields, the remnant of Adeje is to a higher extent incorporated into the central massif. Our data show that even though secondary contact of Adeje and Teide central massif is supposed to be around 2 million years ago (Ancochea et al. 1990; Cantagrel et al. 1999), the imprint in genetic structure can still be observed. This is the case for the samples from *M. hyssopifolia* collected in Adeje which form a distinct cluster in the STRUCTURE analyses. This can be either explained by Adeje as origin of *M. hyssopifolia*, by different conditions that favors certain genotypes by selection, or recent volcanic events that kept these populations isolated.

Hybrid zones and potential ecological effects

Our analysis indicates a strong influence of historical and contemporary gene flow between the species on the genetic structure, most pronouncedly in *M. hyssopifolia*. Hybridization between different *Micromeria* species in Tenerife is well documented and hybrids between most of the species have been described: *M. varia × teneriffae, M. varia × rivas-martinezii, M. varia × densiflora, M. varia × M. glomerata, M. teneriffae × hyssopifolia* (Pérez de Paz 1978; Santos-Guerra et al. 2011). Some of these individuals were included in our dataset and their status as hybrids were confirmed:

M. varia × rivas-martinezii, *M. varia × teneriffae*, *M. teneriffae × hyssopifolia* because they show genotypes intermediate between the parent species. For the two first species pairs, these intermediate genotypes were found in individuals morphologically not classified as hybrids suggesting that they might be backcrosses. The respective two individuals were collected in the contact zone between *M. varia* and *M. rivas-martinezii* populations where both species grow together (Puppo pers. obs.). Three samples of *M. varia* growing in this contact zone were also assigned to *M. teneriffae*, which might be a consequence of introgression of ancestral alleles shared by *M. teneriffae* and *M. rivas-martinezii*.

A lower degree of reproductive isolation between island species, compared to continental ones, is generally assumed because of a potentially comparable lower effect of fitness decrease after hybridization resulting from the lower levels of interspecific competition in island systems (Herben et al. 2005). This context had been discussed in several studies and reviews (i.e., Thomas and Leggett 1974; Charmet et al. 1996; Herben et al. 2005; Silvertown et al. 2005).

In Tenerife, species of *Micromeria* have a pronounced allopatric distribution, i.e., species do not occur in sympatry but only come into contact in relatively small areas where ranges overlap (Fig. 1). Is in these contact zones where hybridization occurs. There are two possible explanations for this distributional pattern. Species might either have evolved in parapatry (Gavrilets et al. 2000) where edge populations differentiate from a larger central population, i.e., in populations of *M. lachnophylla*/*M. lasiophylla* and populations of *M. varia* from Teno. Or, species ranges might have developed after secondary contact of well differentiated species after merging of the paleo-islands. In any case, even in the presence of hybridization, species boundaries are maintained due to differential local selective pressures causing postzygotic isolation and preservation of morphological differences (Seehausen et al. 2014). This typically leads to a hybrid zone dynamic (Barton and Hewitt 1985). Via backcrossing alleles at neutral loci can pass the hybrid zone in both directions, while loci under strong selection cannot and form the base for species specific differences in morphology and ecology. This differential introgression pattern is very well studied and regarded as a typical expression of the contact zone between two species that are able to form fertile hybrids (i.e., Teeter et al. 2010; Nosil et al. 2012; Larson et al. 2014). An example of how selection favors certain genotypes in dependence of the ecological zone is the gecko *Tarentola delalandii* (Gübitz et al. 2000). Despite being the same species, three highly distinct mitochondrial haplotypes originated from the three paleo-islands. This means that, despite the current contact zone, and being the same species, gene flow between the corresponding groups might be low.

Besides the tests for migration, the existence of hybrid zones between the allopatric ranges of the species is supported in our study by three main findings: (1) we observed and verified the status of hybrids in the contact zones of four species pairs; (2) the two species with the largest contact zones, *M. hyssopifolia* and *M. varia*, show also the highest interspecific migration rates. (3) With exception of *M. densiflora* and *M. lasiophylla,* all other connections through gene flow were indicated between species that have contacting ranges; and (4) cluster arrangement in the structure analysis gives increased assignment probability for adjunct populations even though they belong to different species, e.g., for *M. hyssopifolia* and *M. varia* from Teno and *M. hyssopifolia* and *M. lachnophylla*. Hereby, some loci show higher values for migration than others indicating asymmetric introgression at some degree.

The formation of distinct hybrid zones could be observed directly between *M. varia* and *M. rivas-martinezii*. Here, in a very small spatial scale hybridization occurs at the transition from the range of *M. rivas-martinezii* to *M. varia*. *Micromeria rivas-martinezii* grows in a very restricted area in a small peninsula in the Anaga massif (Hernández-Pacheco et al. 1990) where *M. varia* does not occur. In a few 100 m wide zone at the main island adjacent to the peninsula, *M. varia* × *M. rivas-martinezii* hybrids occur in small individual numbers that are giving way to morphological *M. varia* populations (Puppo pers. obs.). This transition can also be seen in our SSR data, indicating a transition in the allele frequency content between *M. rivas-martinezii* and adjacent *M. varia* populations more gradual than expected if the species were reproductively isolated.

The formation of hybrid zones may have contributed to the increase of genetic variation of some taxa facilitating adaptation to changing conditions, shift of ecological niches, or range shift for the species after secondary contact of the paleo-islands. An example would be *M. hyssopifolia*, which is the species with the largest range. It participates in most of gene-flow exchanges found in the island and it has one of the highest genetic diversity. Environmental conditions across the range of *M. hyssopifolia* are highly heterogeneous. The northern part of Tenerife is wetter due to the fog brought by the trade winds with high levels of rainfall (ca. 1000 mm precipitation per year) in the mid altitudes. Contrary to this, the southern part of the island is dry (below 100 mm precipitation per year) due to the shade effect caused by the Teide. As described below, this genetic structure might reflect these environmental differences, such as structure found between the wet northern and dry southern slopes. Like outlined above, we see the population from Adeje slightly differentiated from the remaining *M. hyssopifolia* populations. Besides this, at optimal ($K = 3$) and higher K (up to $K = 9$), we observed genetic structure among: (1) Teno and west Tenerife; (2) north coast, and (3) southeast

and south coast, corresponding to a medium, high and very low precipitation regime. It seems likely that genotypes are locally adapted to these different habitats and genotypes from the southern part may not be able to establish in the northern part and vice versa. These different habitats correspond roughly to the subdivision of *M. hyssopifolia*. Three varieties are recognized within this species*: var. hyssopifolia*, *var. glabrescens,* and *var. kuegleri* (Pérez de Paz 1978) reflecting its morphological diversity. The typical *M. hyssopifolia* (*var. hyssopifolia*) presents a strigose indumentum which gives the plants a grayish aspect and is distributed in the pine forest between 400–2000 m. *Micromeria hyssopifolia var. glabrescens* is mostly distributed in the north of the island in degraded areas between 300–600 m while *var. kuegleri* is the coastal form that grows in the southeast from the sea level up to 400 m. Thus, our structure pattern differentiates mostly *var. glabrescens* and *var. kuegleri*. Because the environmental conditions are not independent from geography, further work is currently being conducted to confirm the hypotheses that: hybridization after secondary contact of former paleo-island species allowed the colonization of the whole island by one or a few species, and the genetic structure that can be observed in *M. hyssopifolia* is an expression of local adaptation patterns rather than geography.

Low genetic differentiation levels and microsatellites

The pattern of hybridization found in our study might also explain the apparent low genetic distance between the species with microsatellite datasets and with our earlier multigene analyses (Puppo et al. 2014, 2015a). In Puppo et al. (2015a), low genetic differentiation and low tree resolution were not only found for the central species of Tenerife but also for the most widespread species from Gran Canaria. Because they are usually neutral and have a high mutation rate, microsatellites are frequently used in population genetic studies to identify genetic diversity levels and population differentiation within species but they are rarely used in investigations that cover multiple species (Barbará et al. 2007). Recent examples are Bonatelli et al. (2014) and Turini et al. (2014), where SSR markers and Bayesian clustering had been used to test species boundaries or to establish a species concept.

Gene flow between the species would impact genetic distance. In the case of *Micromeria*, the age estimate especially for the paleo-island species would suggest that alleles are highly diverged, and the amount of private alleles within one species should be rather high. Even though we found private alleles for the different species, only few of them have within species frequencies above 10%, and most of them are rare alleles. In addition, pairwise F_{ST} between populations is only slightly higher between species than within species. Using a microsatellite dataset to investigate different species is likely

to underestimate genetic distances between species when hybridization occurs, not only because of shared alleles but also because of the choice of loci during the screen for markers (Turini et al. 2014). With hybridization between species at a contact zone, screen is likely to be biased toward markers that are not linked to loci that are highly structured but to neutral loci that can pass the hybrid zone. We assume therefore that the degree of genetic differentiation between species might be underestimated using a dataset like this. This is especially true when we consider the high morphological distinctness of the species under investigation (Puppo et al. 2014). However, considering introgression and selection for alleles that are exchanged between species, F_{ST} below 0.1 could be plausible also between these morphologically highly differentiated species. This effect may also lead to overestimation of migration rates. Nevertheless, this would affect all measures in the same way and not influence interpretations that are made comparatively.

Phylogeographic and taxonomic considerations

Currently, there are eight species of *Micromeria* recognized in Tenerife with different levels of morphological differentiation. Recent phylogenetic analyses (Puppo et al. 2014) suggest that the genus was probably present in Anaga around 6.7 Ma, before the central shield was formed, and had a first diversification event that gave origin to *M. teneriffae*, and afterwards to *M. glomerata* and *M. rivas-martinezii*. A second diversification event probably took place in Teno giving origin to *M. densiflora* ca. 4.5 Ma. These four species are also today clearly related to the paleo-islands. According to this phylogeny, Teno colonized the central part of Tenerife where the remaining four species originated. These analyses were inconclusive with regard to the central species however, since relations among the species were poorly resolved (Puppo et al. 2014). Nevertheless, phylogenetic reconstruction seems to support a scenario where progressive adaptation to higher altitudes of *M. varia* gave origin to *M. hyssopifolia,* and this to *M. lachnophylla* and *M. lasiophylla* (Pérez de Paz 1978; Puppo et al. 2014).

Microsatellite analysis conclusively supports all species when we consider the formation of distinct clusters in the structure analysis. As explained above, it seems likely that hybridization between species is decreasing pairwise differences between the species. In addition, the paleo-island species appear to a higher extent differentiated from the others and microsatellite analyses provide new insights into the genetic structure of the central species. Interestingly, *M. lasiophylla* is showing close affinities to *M. teneriffae* in an analysis of Nei genetic distances as well as cluster together with the paleo-islands species for $K = 2$. Even though *M. lasiophylla* and *M. lachnophylla* are not early diverging lineages as the paleo-island species, this indicates that diversification

might precede the secondary contact that occurred ca. 2 Ma ago (Ancochea et al. 1990; Cantagrel et al. 1999). Both *M. lasiophylla* and *M. lachnophylla* grow in old rocks of volcanic origin. It is possible that progenitors of these species colonized from the paleo-islands independently from the other species instead of being the high altitude forms of *M. varia* or *M. hyssopifolia* as suggested by morphology and phylogenetic analysis (Pérez de Paz 1978; Puppo et al. 2014). In fact, it has been observed in several groups (i.e., Thorpe et al. 1994; Dlugosch and Parker 2007; Cox et al. 2012; Macías-Hernández et al. 2013) that taxa from Tenerife's paleo-islands colonized the central, younger part of the island, as well as other younger nearby islands, following a stepping-stone model (Kimura and Weiss 1964).

The PCoA shows *M. lasiophylla* and *M. lachnophylla* as distinct when analyzed separately from the rest of species (Fig. 3D). When K is increased (i.e., $K = 10$), some admixture between *M. lachnophylla* and *M. hyssopifolia* is found. This is probably caused by hybridization with *M. hyssopifolia* since *M. lachnophylla* is distributed from the high desert in Las Cañadas down to the border of the pine forest where *M. hyssopifolia* grows. Furthermore, morphologically intermediate individuals have been reported in several localities that constitute contact zones where both species occur (Pérez de Paz 1978).

Acknowledgments

The authors thank the University of La Laguna in Tenerife for providing a vehicle for the field trips. P.L. Pérez de Paz was an invaluable help throughout this study. V. Garzón-Machado provided the map for Tenerife. F. Faure helped during collection of the plant material and during preparation of Figure 1. This research was supported by the Fundação para a Ciência e Tecnologia (FCT) with a research grant to H.M. (PTDC/BIA-BEC/108866/2008) cofinanced by the Fundo Europeu de Desenvolvimento Regional (FEDER) through COMPETE – Programa Operacional Fatores de Competitividade (POFC) (FCOMP-01-0124-FEDER-008988), and PhD fellowships to P.P. (SFRH/BD/74747/2010) and M.C. (SFRH/BD/79010/2011).

Data Accessibility

Microsatellite data matrix is deposited in Demiurge as: Puppo P, Curto M, Meimberg H (2015) D-NMICR-99 http://www.demiurge-project.org/matrix_digests/D-NMICR-99.

Table S3 can be acessed in:
http://onlinelibrary.wiley.com/doi/10.1002/ece3.2094/full

References

Ancochea E, Fuster JM, Ibarrola E, Cendredo A, Coello J, Hernan F, Cantagrel JM, Jamond C (1990) Volcanic evolution of the island of Tenerife (Canary Islands) in the light of new K-Ar data. *Journal of Volcanology and Geothermal Research*, **44**, 231–249.

Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology*, 11, 771–783.

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*, **16**, 3759–3767.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Bräuchler C, Ryding O, Heubl G (2008) The genus *Micromeria* (Lamiaceae), a synoptical update. *Willdenowia* **38**, 363–410.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Beerli P, Palczewski M (2010) Unified Framework to Evaluate Panmixia and Migration Direction Among Multiple Sampling Locations. *Genetics*, **185**, 313–326.

Bertrand JAM, Bourgeois YXC, Delahaie B, Duval T, Garcia-Jimenez R, Cornuault J, Heeb P, Mila B, Pujol B, Thebaud C (2014) Extremely reduced dispersal and gene flow in an island bird. *Heredity*, **112**, 190–196.

Bonatelli IAS, Perez MF, Peterson T, Taylor NP, Zappi DC, Machado MC, Koch I, Pires A, Moraes EM (2014) Interglacial microrefugia and diversification of a cactus species complex: phylogeography and palaeo distributional reconstructions for *Pilosocereus aurisetus* and allies. *Molecular Ecology*, **23**, 3044–3063.

Brown RP, Hoskisson PA, Welton J-H, Báez M (2006) Geological history and within-island diversity: a debris avalanche and the Tenerife lizard *Gallotia galloti*. *Molecular Ecology*, **15**, 3631–3640.

Cantagrel JM, Arnaud NO, Ancochea E, Fuster JM, Huertas MJ (1999) Repeated debris avalanches on Tenerife and genesis of Las Cañadas caldera wall (Canary Islands). *Geology*, **27**, 739–742.

Carine MA, Russell SJ, Santos-Guerra A, Francisco-Ortega J, (2004) Relationships of the Macaronesian and Mediterranean Floras: Molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Convolvulus* (Convolvulaceae). *American Journal of Botany*, **91**, 1070–1085.

Carracedo JC (1994) The Canary Islands: an example of structural control on the growth of large oceanic-island volcanoes. *Journal of Volcanology and Geothermal Research*, **60**, 225–241.

Caujapé-Castells J (2011) Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. In: Bramwell D & Caujapé-Castells J (Eds.) *The biology of island floras*. Cambridge University Press, Cambridge, 284–324.

Charmet G, Balfourier F, Chatard V (1996) Taxonomic relationships and interspecific hybridization in the genus *Lolium* (grasses). *Genetic Resources and Crop Evolution*, **43**, 319–327.

Conflitti IM, Shields GF, Murphy RW, Currie DC (2014) Molecular Phylogenetics and Evolution The speciation continuum : Population structure , gene flow , and maternal ancestry in the Simulium arcticum complex (Diptera : Simuliidae). *Molecular Phylogenetic and Evolution*, **78**, 43–55.

Cornuet JM, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Cox SC, Carranza S, Brown RP (2012) Divergence times and colonization of the Canary Islands by Gallotia lizards. *Molecular Phylogenetics and Evolution*, **56**, 747–757.

Curto MA, Tembrock LR, Puppo P, Nogueira M, Simmons MP, Meimberg H (2013) Evaluation of microsatellites of *Catha edulis* (qat; Celastraceae) identified using pyrosequencing. *Biochemical Systematics and Ecology*, **49**, 1–9.

Dlugosch KM, Parker IM (2007) Molecular and quantitative trait variation across the native range of the invasive species *Hypericum canariense*: evidence for ancient patterns of colonization via pre-adaptation? *Molecular Ecology*, **16**, 4269–4283.

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

Fernández-Palacios JM, de Nascimento L, Otto R, Delgado JD, García-del-Rey E, Arévalo JR, Whittaker RJ (2011) A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography*, **38**, 226–246.

Garcia-Verdugo C, Sajeva M, La Mantia T, Harrouni C , Msanda F, Caujapé-Castells J (2015) Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology*, **4**, 726–741

Gavrilets S, Li H, Vose M (2000) Patterns of parapatric speciation. *Evolution*, **54**, 1126–1134.

González-Pérez MA, Lledó MD, Lexer C, Fay M, Marrero M, Bañares A, Carqué E, Sosa PA (2009). Genetic diversity and differentiation in natural and reintroduced populations of *Bencomia exstipulata* and comparisons with *B. caudata* (Rosaceae) in the Canary Islands: an analysis using microsatellites. *Botanical Journal of the Linnean Society*, **160**, 429–441.

Gübitz T, Thorpe RS, Malhotra A (2000) Phylogeography and natural selection in the Tenerife gecko *Tarentola delalandii*: testing historical and adaptive hypotheses. *Molecular Ecology*, **9**, 1213–1221.

Guillou H, Carracedo JC, Paris R, Pérèz-Torrado FJ (2004) Implications for the early shield-stage evolution of Tenerife from K/Ar ages and magnetic statigraphy. *Earth and Planetary Science Letters*, **222**, 599-614.

Herben T, Suda J, Munclinger P (2005) The ghost of hybridization past: niche pre-emption is not the only explanation of apparently monophyly in island endemics. *Journal of Ecology*, **93**, 572–575.

Hernández-Pacheco A, de la Nuez J, Cubas CR, Hernán F, Fernández S (1990) Los Domos sálicos de Tenerife, Islas Canarias. *Estudios Geológicos*, **46**, 175–184.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

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

Jones KE, Reyes-Betancort JA, Hiscock SJ, Carine, MA (2014) Allopatric diversification, multiple habitat shifts, and hybridization in the evolution of *Pericallis* (Asteraceae), a Macaronesian endemic genus. *American Journal of Botany,* **101**, 637–651.

Juan C, Emerson BC, Oromí P, Hewitt GM (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology and Evolution*, **15**, 104–109.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.

Larson EL, White TA, Ross CL, Harrison RG (2014) Gene flow and the maintenance of species boundaries. *Molecular Ecology*, **23**, 1668–1678.

MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, NJ.

Macías-Hernandez N, Bidegaray-Batista L, Emerson C, Oromí P, Arnedo MA (2013) The imprint of Geologic History on Within-Island diversification of woodlouse-hunter spiders (Araneae, Dysderidae) in the Canary Islands. *Journal of Heredity*, **104**, 341–356.

Mairal M, Sanmartín I, Aldasoro JJ, Culshaw V, Manolopoulou I, Alarcón M (2015) Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic *Canaria canariensis* (Campanulaceae). *Molecular Ecology*, **24**, 3944–3963.

Martín JL, Vera MA, Arechavaleta M (1999) Biodiversidad taxonómica y análisis de prioridad para el establecimiento de áreas protegidas. *Vieraea*, **27**, 245–253

Meimberg H, Abele T, Bräuchler C, McKay JK, Pérez de Paz PL, Heubl G (2006) Molecular evidence for adaptive radiation of *Micromeria* Benth. (Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data. *Molecular Phylogenetics and Evolution*, **41**, 566–578.

Moya O, Contreras-Díaz HG, Oromí P, Juan C (2004) Genetic structure, phylogeography and demography of two ground-beetle species endemic to the Tenerife laurel forest (Canary Islands). *Molecular Ecology*, **13**, 3153-3167.

Nosil P, Parchman TL, Feder JL, Gompert Z (2012) Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in *Timema* stick insects. *BMC Evolutionary Biology*, **12**, 1–12.

Peacock E, Sonsthagen SA, Obbard ME, Boltunov A, Regehr EV, Ovsyanikov N, Aars J, Atkinson SN, Sage GK, Hope AG, Zeyl E, Bachmann L, Ehrich D, Scribner KT, Amstrup SC, Belikov S, Born EW, Derocher AE, Stirling I, Taylor MK, Paetkau D, Talbot SL (2015) Implications of the Circumpolar Genetic Structure of Polar Bears for Their Conservation in a Rapidly Warming Arctic. *PlosOne*, **10**, e112021.

Pérez de Paz PL (1978) Revisión del género *Micromeria* Bentham (Lamiaceae-Stachyoideae) en la Región Macaronésica. *Instituto de Estudios Canarios, Monografías*, **16**, 1–306.

de Paz JP, Caujapé-Castells J (2013) A review of the allozyme data set for the Canarian endemic flora: causes of the high genetic diversity levels and implications for conservation. Annals of botany, **111**, 1059-1073.

Piry S, Luikart G, Cornuet J-M (1999) Bottleneck: a program for detecting recent reductions in effective population size reductions from allele frequency data. *Journal of Heredity*, **86**, 502–503.

Puppo P, Curto M, Velo-Antón G, Pérez de Paz PL, Meimberg H (2014) The influence of geological history on diversification in insular species: genetic and morphological

patterns of *Micromeria* Benth. (Lamiaceae) in Tenerife (Canary archipelago). *Journal of Biogeography*, **41**, 1871–1882.

Puppo P, Curto M, Gusmão-Guedes J, Cochofel J, Pérez de Paz PL, Bräuchler C, Meimberg H (2015a) Molecular phylogenetics of *Micromeria* (Lamiaceae) in the Canary Islands, diversification and inter-island colonization patterns inferred from nuclear genes. *Molecular Phylogenetics and Evolution*, **89**, 160–170.

Puppo P, Curto M, Meimberg H (2015b). Development and characterization of 16 microsatellite markers for *Micromeria* (Lamiaceae) from Tenerife (Canary Islands, Spain) using 454 sequencing. *Conservation Genetics Resources*, **7**, 743–749.

Puppo P, Meimberg H (2015) New species and new combinations in *Micromeria* (Lamiaceae) from the Canary Islands and Madeira. *Phytotaxa,* **230**, 1–21.

Rambaut A, Drummond AJ (2007) TRACER v1.4. Available from: http://beast.bio.ed.ac.uk/Tracer

Rohlf FJ (1993) NTSYS-PC: numerical taxonomy and multivariate analysis system – version 2.0. New York: Exeter Software.

Roy D, Lucek K, Walter RP, Seehausen O (2015) Hybrid ‘superswarm’ leads to rapid divergence and establishment of populations during a biological invasion. *Molecular Ecology*, **24**, 5394–5411.

Rumeu B, Vargas P, Jaén-Molina R, Nogales M, Caujapé-Castells J (2014) Phylogeography and genetic structure of the threatened Canarian *Juniperus cedrus* (Cupressaceae). *Botanical Journal of the Linnean Society,* **175**, 376–394.

Santos-Guerra A, Acevedo-Rodriguez A, Reyes-Betancort JA (2011) Redescubrimiento del endemismo tinerfeño *Micromeria densiflora* Benth. (Labiatae). *Anales del Jardín Botánico de Madrid*, **68**, 155–159.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G, Bank C, Brännström Å *et al*. (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.

Silvertown J, Francisco-Ortega J, Carine M (2005) The monophyly of island radiations: an evaluation of niche preemption and some alternative explanations. *Journal of Ecology*, **93**, 653–657.

Sosa PA, González-González EA, González-Pérez MA, Pérez de Paz PL (2013) Contrasting patterns of genetic differentiation in Macaronesian lineages of *Ilex* (Aquifoliaceae). *Botanical Journal of the Linnean Society*, **173**, 258–268.

Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, **64**, 472–485.

Thomas H, Leggett JM (1974) Chromosome relationships between *Avena sativa* and the two diploid species *A. canariensis* and *A. prostrata*. *Canadian Journal of Genetics and Cytology*, **16**, 889–894.

Thorpe RS, McGregor DP, Cumming AM, Jordan WC (1994) DNA evolution and colonization sequence of island lizards in relation to geological history: mtDNA RFLP, cytochrome B, cytochrome oxidase, 12S ribosomal RNA sequence, and nuclear RAPD analysis. *Evolution*, **48**, 230–240.

Trusty J, Olmstead RG, Santos-Guerra A, Sá-Fontinha S, Francisco-Ortega J (2005) Molecular phylogenetics of the Macaronesian-endemic genus *Bystropogon* (Lamiaceae): palaeo-islands, ecological shifts and interisland colonizations. *Molecular Ecology*, **14**, 1177–1189.

Turini FG, Steinert C, Heubl G, Bringmann G, Lombe BK, Mudogo V, Meimberg H (2014) Microsatellites facilitate species delimitation in Congolese *Ancistrocladus* (Ancistrocladaceae), a genus with pharmacologically potent naphthylisoquinoline alkaloids. *Taxon*, **63**, 329–341.

Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

van Hengstum T, Lachmuth S, Oostermeijer JGB, den Nijs H(J)CM, Meirmans PG, van Tienderen PH (2012) Human-induced hybridization among congeneric endemic plants on Tenerife, Canary Islands. *Plant Systematics and Evolution*, **298**, 1119–1131.

Watts AB, Masson DG (1995) A giant landslide on the north flank of Tenerife, Canary Islands. *Journal of Geophysical Research*, **100**, 24487–24498.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Whittaker RJ, Ladle RJ, Araújo MB, Fernández-Palacios JM, Delgado JD, Arévalo JR (2007) The island immaturity –speciation pulse model of island evolution: an alternative to the “diversity begets diversity” model. *Ecography*, **30**, 321–327.

Whittaker RJ, Triantis KA, Ladle RJ (2008) A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, **35**, 977–994.

3.2. Genetic diversity and differentiation patterns of *Micromeria* from the Canary Islands are congruent with multiple colonization dynamics and the establishment of species syngameons

Manuel Curto[1,2], Pamela Puppo[2], and Harald Meimberg[2]

[1]Institute for Integrative Nature Conservation Research, University of Natural Resources and Life Sciences, A-1180 Vienna, Austria
[2]CIBIO, Research Center in Biodiversity and Genetic Resources / InBio Associated Laboratory, University of Porto, Campus Vairão, Vairão 4485-661, Portugal

Abstract

Background: Especially in islands closer to the mainland such as the Canary Islands, different lineages originated by multiple colonization events could merge by hybridization, which furthermore could promote radiation events [1, 2, 3]. This is an alternative to the scenario, where evolution is mostly driven by drift [4, 5]. In this case hybridization should be reflected in the genetic structure and diversity patters of island species. In the present work we investigate *Micromeria* from the Canary Islands by extensively studying their phylogeographic pattern based on 15 microsatellite loci and 945 samples. These results are interpreted according to the hypotheses outlined above.

Results: Genetic structure assessment allowed us to infer detailed phylogeographic patterns such as the role of the region of Teno (Tenerife) in the colonization of other western islands of the Canary Islands. Moreover, we were able to genetically differentiate most *Micromeria* species supporting its current classification. We found that populations in younger islands were significantly more diverse genetically and less differentiated than the ones in older islands. Moreover, we found evidences of introgression among species and islands.

Conclusions: These results are congruent with a scenario of multiple colonizations during the expansion into new islands. Hybridization contributes for the grouping of multiple lineages into highly diverse populations. Thus, in our case, islands receive several colonization events from different sources, which are combined into sink populations. This mechanism is in accordance to the surfing syngameon hypothesis. Contrary to its current form our results might reflect a slightly different effect: hybridization might always be related to colonization also within the archipelago so initial genetic diversity might be always high. Thus the emergence of new islands promotes multiple colonizations events contributing to the establishment of hybrid swarms that may enhance adaptive ability and thus radiation events.

Keywords: hybrid swarm, species syngameon, genetic diversity, hybridization, genetic structure, oceanic islands

Background

Oceanic islands always had a special focus in evolutionary biology [6]. The high availability of empty niches paired with the low migration rate from the mainland may, on the one hand, contribute to a high prevalence for ecological speciation [6, 7, 8]. Hybridization might facilitate this adaptive evolution by increasing genetic diversity [9]. The processes linking hybridization and ecological adaptation as hypothesized by Seehausen [10], Seehausen *et al.* [9], and others (e.g. [11]), might be especially common on islands and form the basis for the prevalence of adaptive radiation on oceanic archipelagos. In agreement with this idea, Herben *et al.* [1] and Saunders and Gibson [2] suggested that multiple colonization events followed by hybridization, occur in particular on archipelagos close to the mainland, promoting adaptive radiation due to the increase of genetic diversity. By comparing island taxa with their mainland relatives, it has been found that genetic variation in several insular populations was not significantly lower than their close relative in the mainland [12, 13]. In case of a single colonization event, the founder effect would have created a significant lower genetic diversity in islands taxa. These observations led to the formulation of the surfing syngameon hypothesis where islands would constitute allelic sinks [3]. Through multiple colonization events originated from different sources, previously separated genotypes would be combined in hybrid populations (hybrid swarms) thus increasing genetic diversity. These populations might then differentiate ecologically into species that are still connected by gene-flow, thus forming a syngameon, i.e. a group of hybridizing species that evolve as one unit [3, 14]. Islands closer to the mainland are more likely to receive colonizers making them genetically more variable. The other islands are less likely to receive migrants from the mainland but rather from members of the syngameon. This will lead to the loss of genetic diversity compared to the source due to founder effects, and might increase differentiation by genetic dynamics of expanding populations due to allele surfing [3, 15]. In addition, this difference in likelihood of receiving colonization events explains the existence of different evolutionary processes within an archipelago, where in some island sink events are more predominant, and in others differentiation events prevail. This is the surfing syngameon hypothesis.

The predictions of the surfing syngameon hypothesis are apparently in accordance with the distribution of genetic diversity across the Canary Islands, where genetic diversity is negatively- and differentiation positively-related with island distance

to the mainland when considering the overall genetic diversity per species per island [13]. The Canary Islands are a volcanic archipelago composed of seven islands located between 100 and 450 km off the Western Saharan coast. These islands originated from east to west from a hot spot movement being older near the mainland, varying in age between 20.6 Ma (Fuerteventura) and 1.1 Ma (El Hierro). Like all volcanic islands, the Canaries present high levels of endemism as a consequence of their complex geomorphological composition and the high diversity of ecological zones [12, 16]. Each island is in a different stage of the oceanic islands' life cycle [17], being some of them composed by older and younger parts [18]. For example, Tenerife resulted from the connection of three older palaeo-islands (Anaga, Teno, and Adeje) by a central volcano, and in Gran Canaria the SW part of the island date back to the Miocene (palaeo-canaria) and the NE to the Pliocene (neocanaria; [19]). This has been shown to have a high impact on the evolutionary history of the species (i.e. [20, 21, 22, 23, 24, 25]). Hereby, species occupying the younger part of the island can be genetically more diverse but less differentiated among each other. This had been shown in our study system, *Micromeria* (Lamiaceae) in Tenerife [24, 25].

Here, we expand our previous microsatellite study [25] to include all species of *Micromeria*, covering the whole Canarian archipelago, to investigate structure between species and to assess the distribution of within population diversity and between population differentiation. If multiple colonizations and hybridization do not occur with high frequency, we would expect a decrease of genetic diversity from the older islands to the most recently colonized islands. Since the older eastern islands are expected to have been colonized first, they would be significantly more diverse than the younger western islands. In addition, because the likelihood of gene-flow between populations is lower in younger islands, genetic differentiation should be higher. On an alternative scenario, hybrid swarm creation would be more recent in younger islands leading to a high diversity but higher homogeneity of populations, and genetic differentiation should decrease from the older towards the younger islands. In the latter case the formation of the syngameon would have a higher significance during the diversification within the archipelago.

Here, these expectations are tested by investigations of the genetic structure, gene-flow among species within and among islands, and patterns of genetic diversity and differentiation of *Micromeria* in the Canary Islands. We interpret these results in the light of the surfing syngameon hypothesis with special focus on patterns of within and among population differentiation across the archipelago. If single colonization events prevail we do not expect to find gene-flow between species in different islands and genetic diversity should drastically decrease from older to younger islands within and

between populations. These expectations outlined by Caujapé-Castells [3] are not completely met. We did find evidences of multiple colonization events providing high genetic diversity at the colonization front.

Material and Methods

Biological system

Micromeria Benth. (Nepetoideae, Lamiaceae) is a monophyletic genus with ca. 54 species [26] from which approximately 22 are present in the Canary Islands and Madeira [27, 28, 29]. As in many other Canarian taxa, molecular data show two main lineages of *Micromeria* in the Canary Islands archipelago [30, 31]. One includes the species found in the eastern islands of Gran Canaria, Lanzarote and Fuerteventura; the other contains taxa from the western islands of Tenerife, La Palma, and El Hierro. Taxa from La Gomera belong to both lineages: *M. lepida* and *M. gomerensis* are part of the eastern lineage, while *M. pedro-luisii* is part of the western lineage [31].

From all the species present in the Canary archipelago, six share similar morphological characters that lead to be previously classified as the same species (*M. varia*). These taxa occupy all islands except La Palma, and have been recently separated based on molecular phylogenetic analyses [28, 29, 31] (Fig. 1). La Palma and El Hierro where two species were identified in each island [27], and now only one species per island is recognized [28, 29].

Populations from the species previously classified as *M. varia* in Lanzarote, Gran Canaria and from one taxa in La Gomera form a monophyletic group while the remaining ones they do not [31]. Likewise, it was also shown that *M. varia* s.s. in Tenerife is genetically and morphologically highly similar to *M. hyssopifolia* [24]. This variety of morphological features and their complex evolution allowed us to evaluate genetic variation at different levels: by using only members of morphological and genetically similar species we can replicate genetic patterns among individuals with independent phylogenetic positions but similar ecological and morphological features; by using only individuals from one lineage we can compare the diversity among individuals belonging to the same monophyletic group; and by using the whole archipelago we can assess the diversity among all terminal branches of *Micromeria* phylogeny. In the following analysis the name *M. varia* s.l. is used to characterize this group of species.

Tenerife and Gran Canaria are the largest islands in the archipelago and contain the highest number of *Micromeria* species, eight and seven respectively. La Gomera presents three species and the remaining islands one species each. All *Micromeria* species are single island endemics. In Tenerife, it has been suggested that the composition of species of *Micromeria* is linked to the geological history of the island [24,

25, 31]. From the eight species occurring in this island, three are restricted to the palaeo-islands: *M. densiflora* to Teno and *M. rivas-martinezii* and *M. glomerata* to Anaga. These species are narrowly distributed in these regions growing in reduced areas (2500-6500 m^2) with few individuals (ca. 400-900; [32, 33]). A fourth species, *M. teneriffae*, is also growing in Anaga palaeo-island but its range extends towards the southeast of the island up to Fasnia. Molecular studies suggest that these four species are older lineages that evolved before the connection of the three palaeo-islands by the formation of the central shield [24]. *Micromeria varia* s.s. presents a disjoint distribution growing in the palaeo-islands of Anaga and Teno though molecular studies suggest it as a more recent lineage [24, 25]. The remaining three species are distributed in the central, younger part of Tenerife: *M. hyssopifolia* growing throughout the island from the coast up to the pine forest, *M. lachnophylla* from the pine forest up to the high desert, and *M. lasiophylla* in Las Cañadas cliffs, at the skirts of the Teide volcano. In Gran Canaria on the other hand, species distribution is not apparently correlated to the island's evolution [31]. Two species are narrowly distributed in the west of the island, *M. leucantha* restricted to an area of 7000 m^2 and a population of ca. 2400 individuals, and *M. pineolens* growing in an area of ca. 9500 m^2 and a population of ca. 4500 individuals [32]. The remaining five species are widely distributed throughout the island.

Samples used and DNA isolation

A total of 945 individuals were included in this study corresponding to all recognized taxa of *Micromeria* present in the Canary Islands (Fig. 1). These were collected during several excursions from 2010 to 2012 and some of these were already used in the studies of Curto *et al.* [34] and Puppo *et al.* [24, 25, 31, 35]. A total of 196 localities were sampled (Fig. 1). Each locality was considered to be an independent population and they were composed by one to 14 individuals (Supplementary Table S1).

DNA isolation was done using the Macherey-Nagel Plant DNA Extraction Kit (Macherey-Nagel, Düren, Germany) on 20 mg of dried leaf material according to Puppo et al. [25].

Genotyping and markers quality control

The 16 microsatellite markers described in Puppo *et al.* [35] were amplified using the same multiplex primer combinations, florescent dyes, and PCR conditions from Puppo *et al.* [25, 35]. Genotyping was done in an ABI3130xl automatic sequencer (Applied Biosystems, Inc., Foster City, CA; USA) using an internal size standard (Genescan-500 LIZ; Applied Biosystems, Inc.). GeneMapper ver. 4.0 was used for allele's scoring (Applied Biosystems, Inc.). A total of 96 individuals were genotyped two

times to evaluate scoring consistency. After scoring, only markers showing data for most of the samples were used for further analyses. This led to the exclusion of one marker (6493). Puppo *et al.* [25, 35] did not find any significant deviation from Hardy-Weinberg equilibrium, genotyping errors or high amount of null alleles for any of the markers so no further exclusion was necessary.

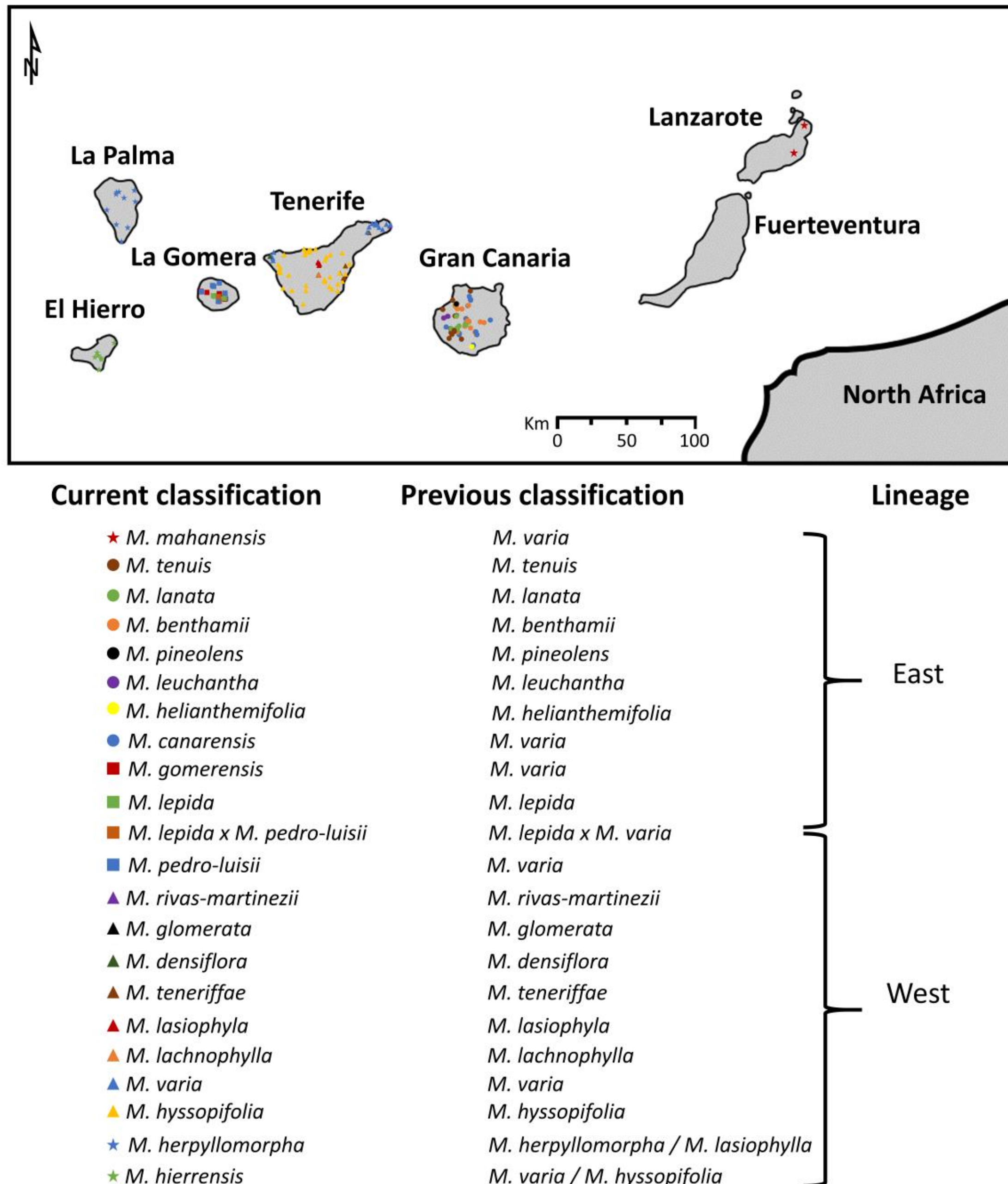


Fig. 1. Map of the Canary Islands showing sampled localities and recognized species per island and group (East/West) to which they belong.

Genetic structure assessment

Main genetic structure patterns were evaluated through distances and clustering analyses. Clustering analyses were conducted with the program STRUCTURE ver. 2.3.3

[36]. Considering there are 22 species and structure within species can be found, the optimal K was first tested by doing short runs (10,000 burnin and 10,000 additional generation) from K=2 to K=30. For each K, five replicates were performed and the best value was obtained using the Delta K method as implemented in STRUCTURE Harvester [37]. Longer runs of 1,000,000 generations followed by a burnin of 500,000 generations were performed from K=2 to the optimal values obtained in the first step. To evaluate possible further divisions we searched for the second best K. This was obtained by running STRUCTURE harvester with only results corresponding to K values above the best. The method only considers a continuous range of K values not allowing values smaller than the optimum. All tests were conducted using the admixture model since it is expected that individuals present multiple assignments to different clusters. Further divisions within the two major lineages (East / West) were evaluated by running two independent STRUCTURE analyses including only individuals from these groups.

To assess the relationships among species, an UPGMA dendrogram was constructed using uNeiD distance [38] among populations in PAUP v. 4.0 [39]. Principal coordinates analyses (PCoA) were calculated in Genalex v. 6.41 (http://biology-assets.anu.edu.au/GenAlEx/) considering each population as a sample point, and using three different datasets: 1) all populations, 2) Gran Canaria populations; and 3) populations assigned to the western group in analysis number 1. Gran Canaria was tested separately because it was clearly separated from the other islands in the whole archipelago analysis. These were calculated using pairwise genotypic distance among populations calculated in GenAlEx.

Gene-flow and hybridization

Migration rates were calculated as proxy estimates of gene-flow. Bidirectional contemporary and historical migration rates were calculated among species and island pairs using the program BayesAss v3.0 [40] and Migrate [41], respectively. Combination of both provides a perspective of gene-flow patterns in two time scales. Because calculating migration-rates among all possible pairwise population combinations was not possible computationally, this was done using both islands and species as groups. In case a clear division within species was observed in the STRUCTURE analyses, these were considered to be different groups. In La Gomera the two highly divergent lineages were treated separately in the island comparison. For Migrate, due to computational limitations migration among islands was only estimated using island groups. Species groups were just used to estimate gene-flow within islands. In BayesAss, the output values correspond to the portion of individuals originated from the population that it is being compared to, while for the Migrate they correspond to the average number of

migrants per generation. For BayesAss, only results with a minimum 95% confident interval value above 0.001 were considered to represent significant migration rates. In the case of Migrate, only migration rates significantly higher than 0 were considered. BayesAss parameters, such as mixing parameters, number of iterations, and burnin were optimized in several runs. An optimal run was considered to have migration, allele frequency, and inbreeding coefficient acceptance rates below 0.6 and an effective sample size above 100 for all parameters (calculated with TRACER v1.5.0, [42]). The final analyses ran for 200,000,000 generations, excluding the first 100,000,000, and sampling every 1,000th generation for both groupings. For species pair analyses the optimal admixing values were 0.5 for both allele frequency and inbreeding coefficient, while for island pairs no adjustment was necessary. Migrate was run as described in Puppo *et al.* [25]. We used a burnin of 5,000,000 generations and estimates were sampled every 100th generation until a total of 50,000 were recorded.

Hybridization between all lineages showing significant recent migration rates was tested. This was done with the program NewHybrids [43] by calculating the probability of individuals being F1 or F2 hybrids, or back crosses with their parent species. NewHybrids ran for 1,000,000 generations after a burnin period of 1,000,000.

Genetic diversity and quantitative genetic differentiation

To test how genetic diversity and differentiation varied through the archipelago, several statistics were calculated for three groupings: islands, species per island, and populations. We had a special focus on genetic diversity at the population level because it is not affected as much by sample size biases. Genetic diversity was assessed by calculating expected (H_E) and observed heterozygosity (H_O), and portion of private alleles. Two private allele measures were calculated: one comparing each group to the whole archipelago, the other comparing each group to the other groups from the same island. Genetic differentiation and quantification of genetic structure was done by estimating pairwise F_{ST} [44], R_{ST} [45] and unbiased Nei's distance (uNei) among groups. All measures were obtained with GenAlEx 6.41 (http://biology-assets.anu.edu.au/GenAlEx/) using the same matrix but containing only populations with at least four individuals; a total of 766 samples and 116 populations (Supplementary Table S1).

Significant variation of genetic diversity and differentiation between old and young islands, islands, and species per island were calculated by a variance analyses in JMP v. 12 (SAS Institute Inc., Cary, NC, USA). In this analysis, populations or species were respectively used as replicates. The Canary Islands are divided into two age groups: old areas with a geological age between 15.5 Ma to 5.8 Ma (Fuerteventura, Lanzarote, Gran

Canaria, La Gomera, Tenerife palaeo-islands) with the majority around 11 Ma; young areas varied between 2 to 1 Ma (central Tenerife, La Palma, and El Hierro) being mostly around 1.5 Ma. This resulted in the groups “young island species” and “old island species”. There is a large gap between the two islands age (at least 4 Ma) so if age has had any effect on genetic diversity, this will be detected by differences between these two groups. Tenerife has a special role because formerly independent islands got secondarily connected and the so called palaeo-islands have a more pronounced botanical singularity than the other islands. The species were grouped according to their position in the phylogenetic analysis to account for the possibility of recent range shift between old and young substrates. Thus the group of young island species were composed of species from the central area, *M. lachnophylla, M. lasiophylla* and *M. hyssopifolia,* and included the closely related species *M. varia* s.s., which expanded into the palaeo-islands from the central part. This was shown by being phylogenetically more closely related to the central area species [25]. *Micromeria teneriffae* distributed in both old and young areas because it is phylogenetically closely related with the species restricted to the palaeo-islands it probably underwent a range expansion into the central part of the islands. For that reason it was classified as old. When species were grouped according to their current main distributions, the results only changed slightly but were still significant. As a consequence of calculating pairwise genetic differentiation a third category was considered for this measure (differentiation between old and young areas). In summary, two categories for the genetic diversity measures were defined (old and young) and three for the genetic differentiation measures (among old, among young, between old and young).

Genetic differentiation results may reflect differences among lineages rather than among island’s age. This was overcome by doing these analyses including only members of the western lineage that occupy both old and young islands. Age-related biases were assessed by performing a test including only the recent diverged taxa from the *M. varia* / *M. hyssopifolia* species complex.

Results

Genetic Structure

Patterns of genetic distances (uNei among populations) were visualized using both dendrograms and PCoA analyses (Fig. 2). They were generally congruent with previous phylogenetic and similarity analyses [24, 25, 31]. Populations cluster in two main groups generally corresponding to the division between eastern and western islands, with populations from La Gomera assigned to both groups (*M. gomerensis* with the eastern islands; and *M. lepida* and *M. pedro-luisii* within the western islands). . Within

the eastern islands, *M. mahanensis* from Lanzarote clusters with *M. gomerensis* from La Gomera, while the position of Gran Canaria species was different between the PCoA and the dendrogram. In the dendrogram they are together with *M. mahanensis* and *M. gomerensis* while in the PCoA they are together with the remaining species from Gran Canaria. Within the western group, the position of Tenerife's species is congruent with earlier analyses Puppo et al. [25]. The other islands that are here included for the first time are separated from Tenerife with La Palma and La Gomera more similar. In the PCoA analyses, the population of *M. densiflora* from Teno is positioned very close to the La Palma / La Gomera cluster linking the Teno massif to these two islands. *Micromeria glomerata* is separated from all species from the western group in the PCoA and in the dendrogram. Interestingly, one *M. teneriffae* population (tetSC5) clusters within the other western islands which may indicate gene-flow among these islands.

In the STRUCTURE analysis (Fig. 3), the optimal K value according to Evanno's method was 13. In an analysis of K values above 13, the most likely K was 19. In Fig. 3 several runs are presented corresponding to major differentiation events in the archipelago. For K=2, the split between the eastern and western groups is indicated with the difference that *M. lepida* is clustered together with to the eastern islands species. At K=3, the eastern group representatives of *M. varia* s.l., *M. canariensis*, *M. gomerensis*, and *M. mahanensis*, form an independent group. Central Tenerife is separated from the remaining western species at K=5, and at K=8, *M. varia* from Tenerife is divided according to the two palaeo-islands where it grows. For the optimal K (K=13), most of the eastern lineage species are differentiated and most of the islands from the western lineage present their own cluster. Differentiation of most western lineage species is only observed at K=19. When the eastern and western groups are analyzed separately, the optimal K was K=2 and the suboptimal was K=11 and K=17 respectively (Fig. 3B). At these K, only few species were not assigned to individual clusters: *M. gomerensis* and *M. mahanensis* in the eastern group, and *M. rivas-martinezii* and *M. glomerata* in Tenerife. Subdivision within *M. canariensis* is congruent with Gran Canaria age division (palaeo-canaria and neocanaria; Fig. 3B). STRUCTURE analyses show several occurrences of gene-flow between species and populations. At all values of K, some individuals show high assignments probability to clusters mostly found in other species, which is an indication of introgression.

Gene flow and hybridization

Recent migration rates calculated with BayesAss varied between 0.07% and 24.44% between islands (all individuals per island are forming one group) and between 0.18% and 19.9% among species (Supplementary Table S2). All migration rates above

0.8% and 0.4 % were significant for islands and species analyses, respectively, and are shown on Fig. 4. Past migration rates calculated with Migrate varied from 2.8 to 32.3 among islands groups, and between 3.38 and 32.77 among species groups (Supplementary Table S3). For past migration rates analyses, only migration among island groups had a 95% confidence interval above zero. All significant migration rates are shown on Fig. 4.

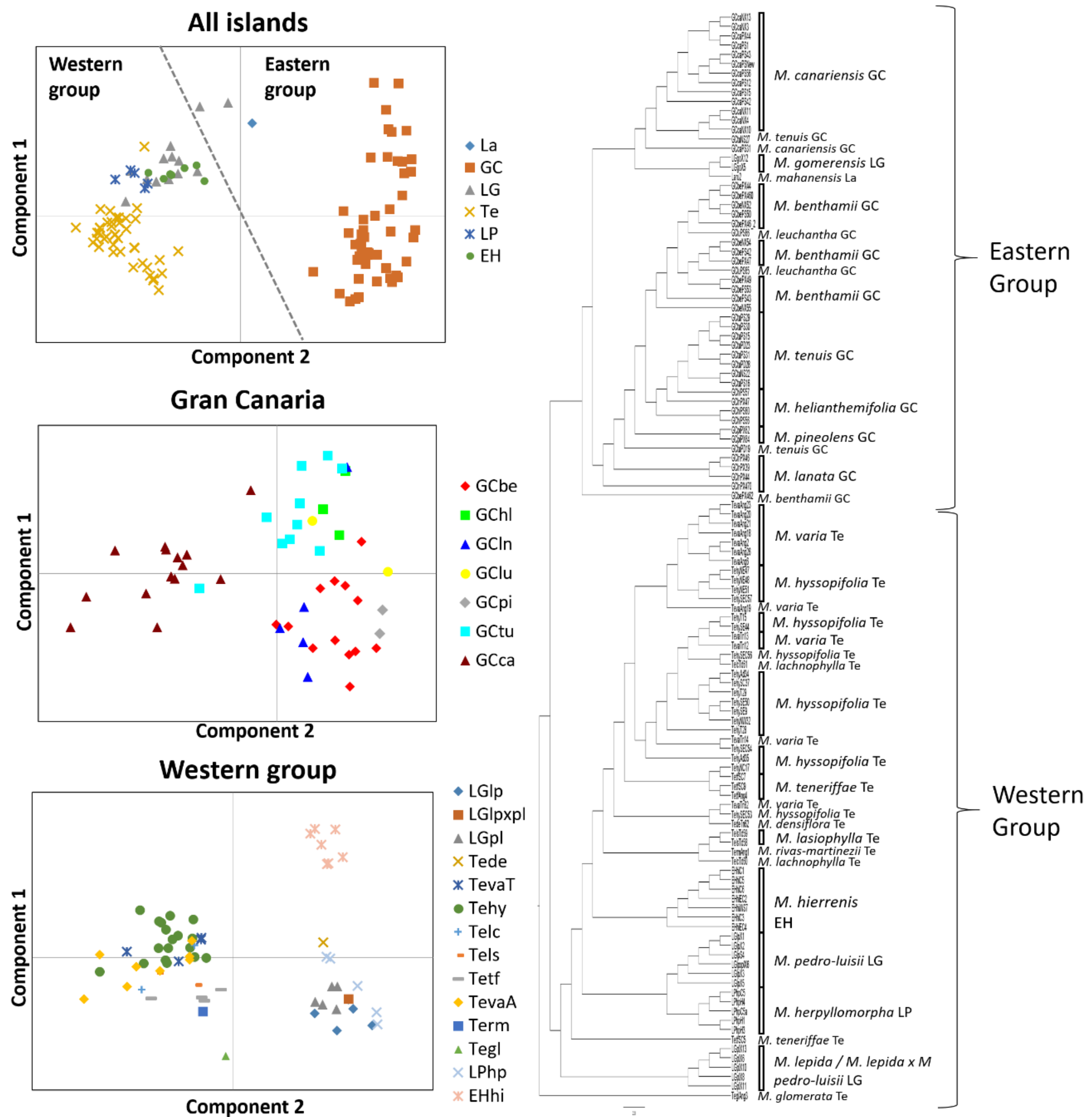


Fig. 2. Principal coordinates analyses and uNeiD UPGMA dendrogram including all localities with at least four individuals. For the PCoA, three analyses were performed: all populations; populations be-longing to the group composed by the eastern islands; and the ones belonging to the group composed by the western islands and *M. lepida* from La Gomera. The vertical and horizontal axis explained respectively: 33.1% and 8.3% of the variation for the all island analyses, 31.9% and 10.4% for the Gran Canaria analyses, and 17.6% and 13.9% for the western group analyses. Islands are represented by the following abbreviations: La (Lanzarote), GC (Gran Canaria), LG (La Gomera); LP (La Palma); EH (El Hierro).

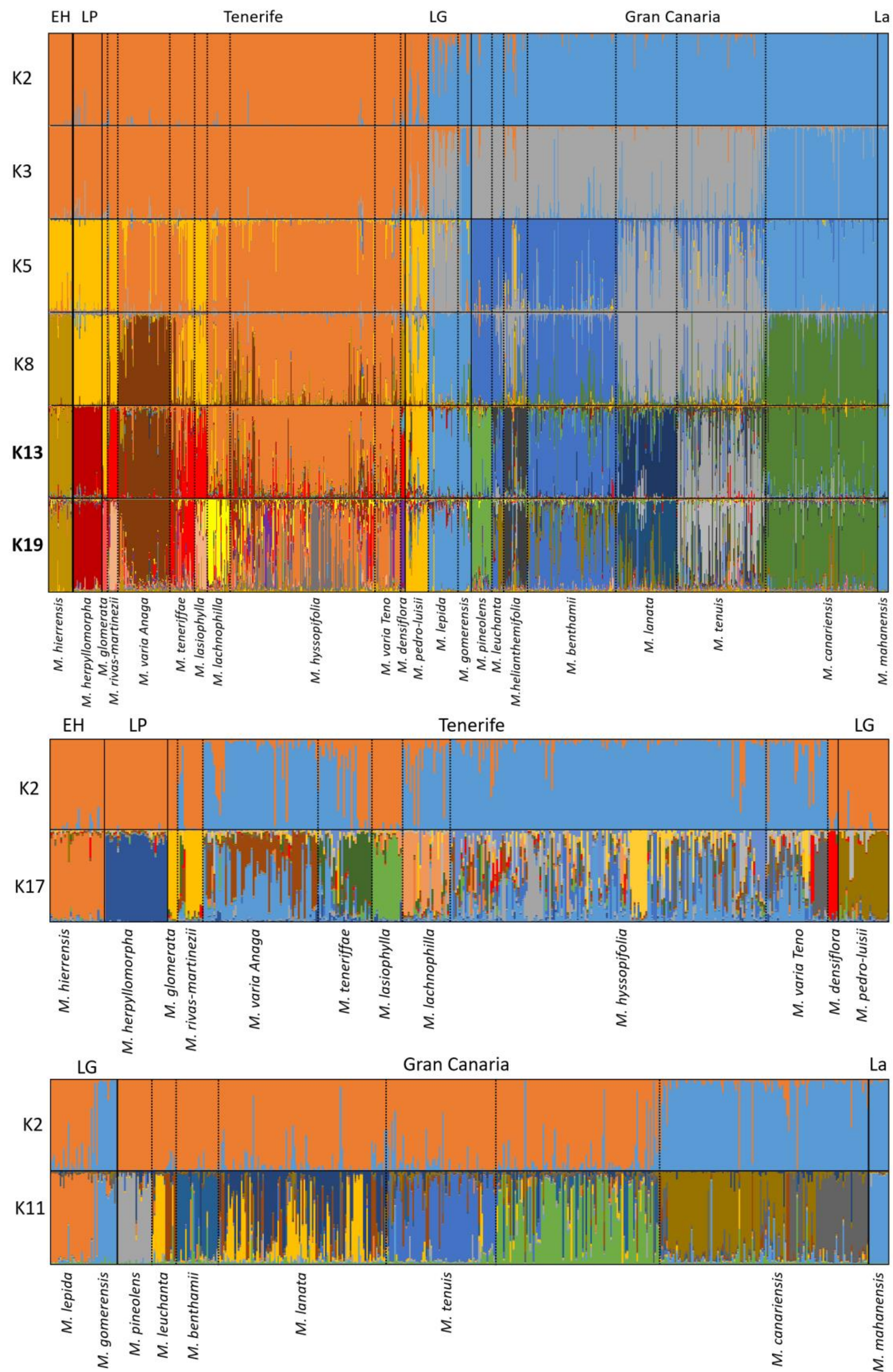


Fig. 3. Bayesian clustering analyses with STRUCTURE. A: Analyses including all samples; best value of K was K=13, second best was K=19. B: Analyses including only the western and eastern lineages; for both the optimum K value was K=2; the suboptimum was K=11 and K=17 for the eastern and western lineages, respectively.

Recent connectivity among islands was mostly found within each lineage (Supplementary Table S3; Fig. 4). When islands' groupings were considered, this resulted in high gene-flow to Lanzarote from members of the eastern lineage from La Gomera and from Tenerife to the western lineage from La Gomera and La Palma. When species groupings were considered, inter-island gene-flow was found from *M. canariensis* (Gran Canaria) to *M. mahanensis* (Lanzarote) and to *M. gomerensis* (La Gomera). Gene-flow among east and west lineages corresponded to migration between both lineages in La Gomera for island groupings and from El Hierro to Gran Canaria (*M. helianthemifolia*) and from *M. lepida* (La Gomera) to *M. densiflora* (Tenerife) in the species grouping. Past gene-flow was found from Tenerife to Gran Canaria, La Palma, and El Hierro; and from Gran Canaria to Tenerife.

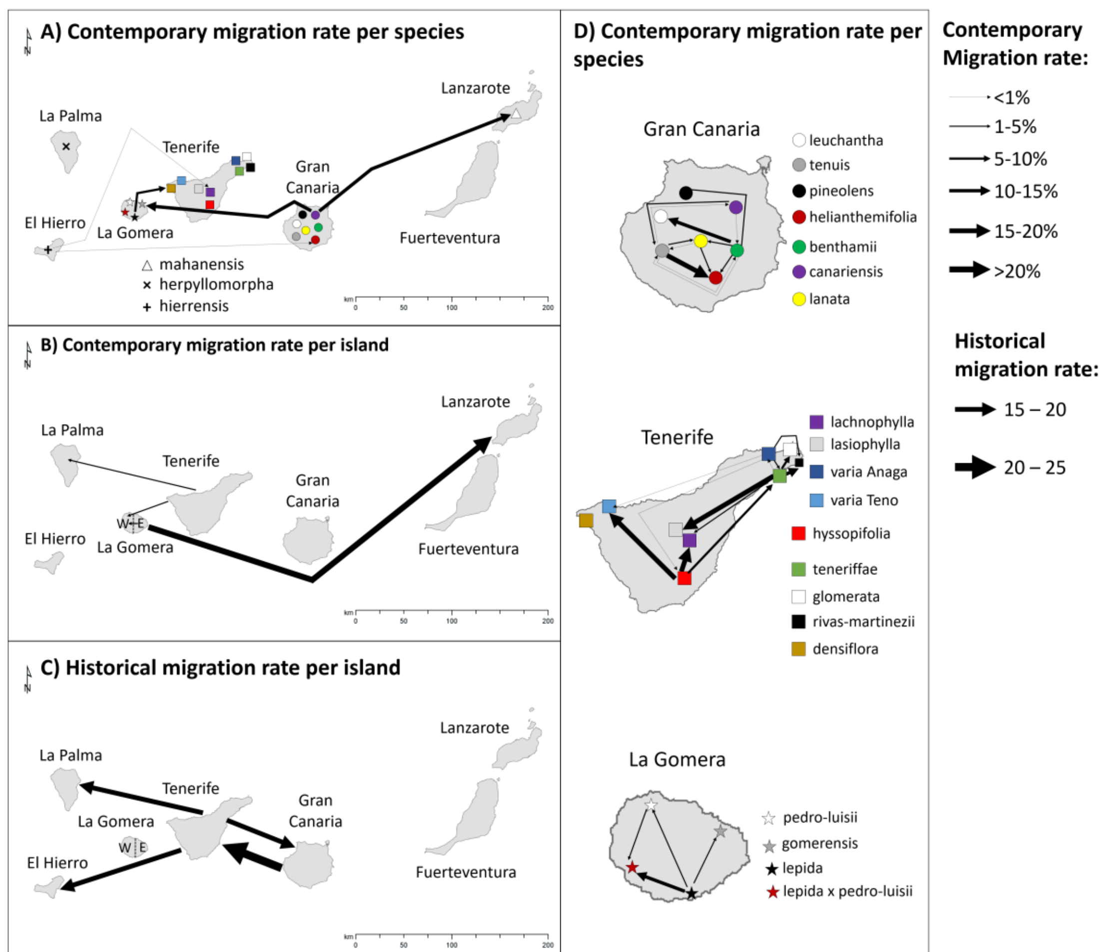


Fig. 4. Representation of recent and historical migration rates calculated with BayesAss and migrate, respectively. All arrows correspond to migration rates with 95% confidence intervals above 0.001 for BayesAss and 0 for Migrate. Arrow thickness is proportional to the migration rate. Panels A and B correspond to contemporary migration rates calculated between islands using species and islands as groups, respectively. Panel C shows historical migration rates between islands groups calculated with Migrate. These were the only significant values obtained from this program. In Panel D, contemporary migration rates between species within islands are shown.

Table 1. Summary of NewHybrid analyses showing the number of individuals with assignment above 50% of being F1 and F2 hybrids, and backcrosses with either species one or two.

Island 1	Island 2	Species 1	Species 2	nr. indiv.	nr. F1 >90%	nr. F2 >90%	nr. Back1 >90%	nr. Back2 >90%
El Hierro	Gran Canaria	*M. hierrensis*	*M. helianthemifolia*	53	0	0	0	0
El Hierro	Tenerife	*M. hierrensis*	*M. hyssopifolia*	191	0	5	0	0
Gran Canaria	Gran Canaria	*M. benthamii*	*M. canariensis*	225	0	39	0	0
Gran Canaria	Gran Canaria	*M. benthamii*	*M. helianthemifolia*	128	0	13	0	0
Gran Canaria	Gran Canaria	*M. benthamii*	*M. lanata*	170	0	17	0	0
Gran Canaria	Gran Canaria	*M. benthamii*	*M. leuchantha*	118	0	54	0	0
Gran Canaria	Gran Canaria	*M. benthamii*	*M. pineolens*	124	0	0	0	0
Gran Canaria	Gran Canaria	*M. benthamii*	*M. tenuis*	206	0	26	0	0
Gran Canaria	Gran Canaria	*M. canariensis*	*M. tenuis*	225	0	10	0	0
Gran Canaria	Gran Canaria	*M. helianthemifolia*	*M. tenuis*	128	0	52	0	0
Gran Canaria	Gran Canaria	*M. lanata*	*M. helianthemifolia*	92	0	5	0	0
Gran Canaria	Gran Canaria	*M. lanata*	*M. tenuis*	170	0	14	0	0
Lanzarote	Gran Canaria	*M. mahanensis*	*M. canariensis*	134	0	15	1	0
La Gomera	Gran Canaria	*M. gomerensis*	*M. canariensis*	136	0	18	0	0
La Gomera	La Gomera	*M. lepida*	*M. gomerensis*	41	0	1	0	0
La Gomera	Tenerife	*M. lepida*	*M. densiflora*	32	0	0	0	0
La Gomera	La Gomera	*M. pedro-luisii*	*M. lepida*	62	0	1	0	0
Tenerife	Tenerife	*M. glomerata*	*M. teneriffae*	33	0	1	0	0
Tenerife	Tenerife	*M. hyssopifolia*	*M. lachnophylla*	187	0	121	0	0
Tenerife	Tenerife	*M. hyssopifolia*	*M. teneriffae*	191	0	20	0	0
Tenerife	Tenerife	*M. hyssopifolia*	*M. varia Anaga*	221	0	36	0	0
Tenerife	Tenerife	*M. hyssopifolia*	*M. varia Teno*	193	0	154	0	0
Tenerife	Tenerife	*M. teneriffae*	*M. lachnophylla*	52	0	12	0	0
Tenerife	Tenerife	*M. teneriffae*	*M. lasiophylla*	43	0	4	0	0
Tenerife	Tenerife	*M. teneriffae*	*M. rivas-martinezii*	41	0	1	0	0
Tenerife	Tenerife	*M. teneriffae*	*M. varia Anaga*	86	0	14	0	0
Tenerife	Tenerife	*M. varia Anaga*	*M. rivas-martinezii*	71	0	3	0	0
Tenerife	Tenerife	*M. varia Anaga*	*M. varia Teno*	88	0	21	0	0

Within island, Tenerife and Gran Canaria showed recent gene-flow for 12 out of 56 possible connections and for 14 out of 42, respectively (Supplementary Table S3; Fig. 4). Although Gran Canaria showed more connections, intra-specific migration rate in Tenerife was higher, with an average of 7.39% compared to the 3.53% for Gran Canaria.

Within Tenerife, *M. hyssopifolia* and *M. teneriffae* were the source of interspecific migration. While *M. hyssopifolia* contributed with gene-flow to the central species, *M. teneriffae* contributed to the palaeo-endemic species. In Gran Canaria, all widespread species worked as source of migration. Within La Gomera, *M. lepida x pedro-luisii* hybrid received migration from both potential parental species. Additionally, *M. lepida* contributed with migration for the other two species in La Gomera.

We used NewHybrids to detect potential hybridization between species pairs exchanging recent gene-flow. We considered hybrids to exist if some of the individuals showed high probability of being hybrids or backcrosses (above 90 %). With the exception of the comparisons *M. lepida* and *M. densiflora*, *M. benthamii* and *M. pineolens,* and *M. hierrensis* and *M. helianthemifolia*; all other species pairs showed at least one individual with the assignment of being a hybrid or a back cross above 90% (Table 1). Potential hybridization between *M. pedro-luisii* and *M. lepida* in La Gomera would correspond to introgression between the two most divergent lineages however this was not verified since only F2 hybrids were found.

Genetic Diversity and differentiation

Genetic diversity and differentiation was estimated as heterozygosity, R_{ST}, F_{ST} and amount of private alleles (Supplementary Tables S4 and S5). Tenerife was the most diverse island showing the highest values of both H_E and H_O (0.53 and 0.80, respectively). The least diverse island was Lanzarote with H_E and H_O of 0.16 and 0.23, respectively. At the species level, taxa from Lanzarote were still the least diverse presenting the same values from the overall diversity per Island. *Micromeria hyssopifolia* presented the highest H_E (0.77) while *M. varia* from Teno the highest H_O (0.77). H_O varied between 0.14 and 0.68 while H_E between 0.22 and 0.78.

Pairwise differentiation among islands varied between 0.16 and 0.64 for R_{ST}, 0.06 and 0.37 for F_{ST} and 0.42 and 1.18 for uNei. None of the island pairs showed constantly either the highest or lowest values for these three measures. At the species level, differentiation among *M. varia* s.s. from Teno and *M. hyssopifolia* was consistently the lowest (R_{ST} = 0.01; F_{ST} = 0.02, uNei = 0.14) and differentiation between Lanzarote and *M. glomerata* the highest (R_{ST} = 0.86; F_{ST} = 0.62; uNeu = 3.09). At the population level, pairwise R_{ST} ranged from -0.21 to 0.91, pairwise F_{ST} to 0.05 and 0.66; and uNei from 0.13 to 5.16.

To determine to what extent island age influenced genetic diversity measures, analyses of variance were performed for differences among categories of: islands age, island, and species. These analyses were done using different measures of genetic diversity on the species and population levels (Table 2) for three datasets with different

sampling subsets: all samples, only *M. varia* s.l. and *M. hyssopifolia*, and within the western lineage. At the population level, both H_O and H_E were significantly higher on younger than older islands for all tests performed, which also corresponded to significant differences between island and species (Table 2 and Fig. 5). At the species level, this was only observed for H_O in the eastern island dataset, and for both H_E and H_O for the western island dataset. Overall, for both population and species level genetic diversity decreased in the direction of east to west (Fig. 5).

Table 2. Results of variance analyses of island age over genetic diversity for three datasets: whole archipelago, M. hyssopifolia and taxa belonging to the M. varia complex; and taxa belonging to the western lineage (Tenerife, La Palma and El Hierro and *M. pedro-luisii* from La Gomera). In the Tests column the letters correspond to the tests classes used. I: Islands; S: Species; O/N: old/new. * corresponds to p-values < 0.001.

Dataset	Level	Test	H_O				H_E			
			DF	SS	F	*p*	DF	SS	F	*p*
All samples	Populations	I	5	0.4	8.0	*	5	0.7	12.5	*
		S	20	1.0	8.8	*	20	1.3	12.0	*
		O/N	1	0.3	28.3	*	1	0.4	32.7	*
	Species	I	5	0.1	1.1	0.39	5	0.2	1.2	0.37
		S	1	0.1	8.3	0.01	1	0.1	4.2	0.05
M. varia* and *M. hyssopifolia	Populations	I	4	0.5	15.9	*	4	0.7	34.7	*
		S	6	0.6	15.7	*	6	0.8	26.5	*
		O/N	1	0.4	38.1	*	1	0.6	76.9	*
	Species	I	4	0.1	2.7	0.23	4	0.2	3.3	0.17
		S	1	0.1	6.4	0.04	1	0.1	4.0	0.09
Within West	Populations	I	5	0.4	18.8	*	5	0.7	29.8	*
		S	11	0.4	9.2	*	11	0.8	25.0	*
		O/N	1	0.1	7.9	0.01	1	0.2	18.6	*
	Species	I	3	0.0	0.4	0.73	3	0.0	0.2	0.89
		S	1	0.1	6.9	0.03	1	0.1	5.6	0.04

The same categories were used to evaluate the effect of island age on pairwise genetic differentiation among populations, species, and islands groups (Table 3; Fig. 5). The same three datasets were used (all samples, only *M. varia* and *M. hyssopifolia*, within the western lineage). At the population level, all differentiation measures were significantly different among classes for all tests in all datasets (Table 3). At the species level, differences among islands and species were not significant for uNei at the *M. varia* and *M. hyssopifolia* dataset. Also no differences among islands were observed for F_{ST} and uNei when only the western lineage was included. For all analyses, genetic differentiation among species inhabiting younger areas was lower than among older ones (not shown). Pairwise R_{ST} and uNeiD between younger and older islands was higher than the other categories when the datasets containing samples from the whole archipelago and only *M. varia* and *M. hyssopifolia* species were analyzed. For the remaining tests differentiation between younger and older islands was in an intermediate

position relatively to the other two classes. The patterns found for the comparisons among age classes were reflected on how genetic differentiation varied across the archipelago. While for the western islands dataset there was a gradual decrease of genetic differentiation from older to younger islands, for the other two datasets this pattern was just evident for F_{ST} (Fig. 5).

Table 3. Results of variance analyses of island age over genetic differentiation for three datasets: all archipelago, *M. hyssopifolia* and taxa belonging to the *M. varia* complex; and taxa belonging to the western lineage (Tenerife, La Palma and El Hierro and *M. pedro-luisii* from La Gomera). I: Islands; S: Species; O/N: old/new. * corresponds to p-values < 0.001.

Dataset	Level	Test	R_{ST}				F_{ST}				uNeiD			
			DF	SS	F	*p*	DF	SS	F	*p*	DF	SS	F	*p*
All samples	Populations	I	5	40,9	170,0	*	5	11,3	319,7	*	5	42,6	26,9	*
		S	21	93,5	100,6	*	21	25,8	204,2	*	21	246,5	38,9	*
		O/N	2	52,0	598,7	*	2	11,5	926,5	*	2	486,5	985,7	*
	Species	I	5	1,2	7,1	*	5	0,4	8,2	*	5	3,1	2,6	0,03
		S	21	3,7	6,1	*	21	2,2	15,5	*	21	18,9	4,3	*
		O/N	2	0,5	8,2	0	2	0,5	29,6	*	2	7,1	16,6	*
M. varia* and *M. hyssopifolia	Populations	I	4	46,5	217,3	*	4	7,2	272,2	*	4	84,2	69,1	*
		S	7	51,1	140,1	*	7	7,4	162,8	*	7	92,7	43,8	*
		O/N	2	39,3	446,7	*	2	8,5	1147,0	*	2	287,7	893,3	*
	Species	I	5	0,7	6,1	0	5	0,2	7,3	*	5	0,8	1,9	0,1
		S	9	0,9	4,2	0	9	0,2	5,4	*	9	1,0	1,3	0,26
		O/N	2	0,1	2,0	0,15	2	0,1	11,9	0	2	1,8	20,3	*
Within West	Populations	I	3	29,7	181,1	*	3	4,7	218,8	*	3	35,3	100,6	*
		S	11	55,2	106,8	*	11	10,6	182,7	*	11	97,6	90,3	*
		O/N	2	17,5	164,4	*	2	5,1	469,7	*	2	62,8	350,3	*
	Species	I	3	0,9	6,5	0	3	0,0	0,3	0,82	3	0,7	1,5	0,21
		S	11	2,6	7,0	*	11	0,7	8,7	*	11	7,1	6,4	*
		O/N	2	0,6	7,5	0	2	0,4	32,4	*	2	3,8	20,3	*

Portion of private alleles per islands varied between none in Lanzarote to 26.49% in Gran Canaria, and per species between zero for several taxa and 8.5% for *M. hyssopifolia*. When compared with the remaining species from the same island, *M. pedro-luisii* had the highest portion of private alleles (51.85 %). On the opposite end was *M. rivas-martinezii* where all alleles are shared with other Tenerife species. At the population level, the portion of private alleles varied between 0 % for several populations from several species to 6.45 % for one population from La Palma (*M. herpyllomorpha*). This population presented also the highest portion of within island private alleles (62.06 %). Private alleles were not as informative when assessing the variation among the defined categories. Nevertheless, the number of private alleles per island was significantly higher in younger islands when the dataset containing the whole archipelago

was considered. This measure was also significantly different among species and island groups.

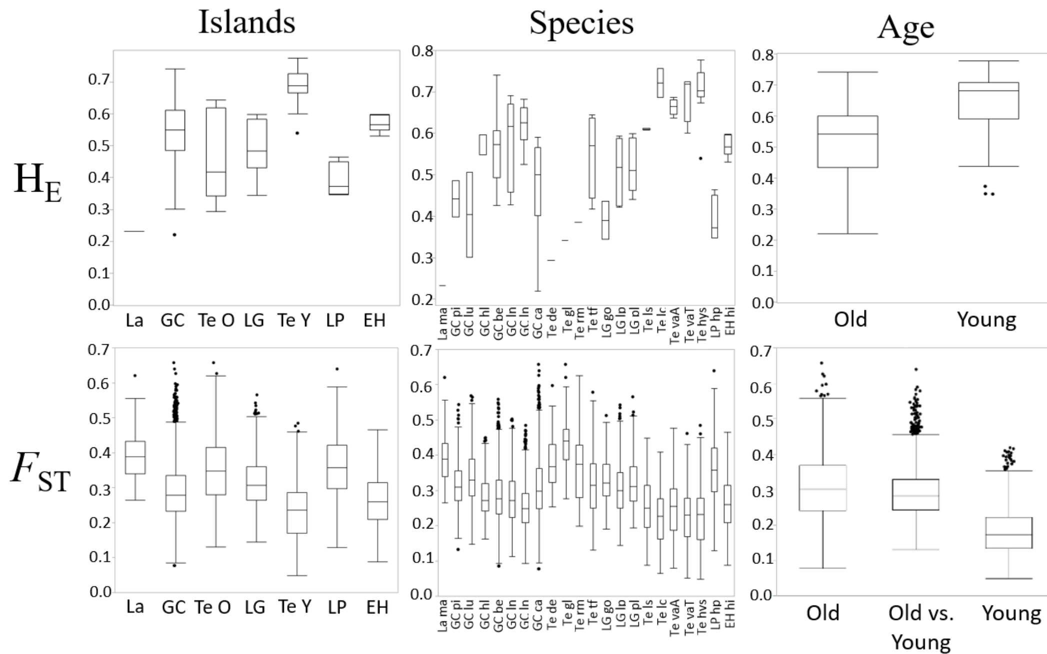


Fig. 5. Representation of expected heterozygosity (HE) and FST per population used as proxy of genetic diversity and differentiation. For each measure there is a graph per island, per species, and per age class. Here only the test including all samples is shown. Graphs for the other tests are shown in supplementary material.

Discussion

Factors contributing for island gene-flow

Codominant genotyping data are rarely applied to questions that cover multiple species. In a recent application, Puppo *et al.* [25] used a codominant dataset to determine the genetic structure at a fine scale for the species of *Micromeria* in Tenerife, and also provided evidence of gene-flow between populations and species. In the study at hand, we expanded this investigation to cover all the species of this genus present in the Canarian archipelago, including multiple populations for each island. We found that gene-flow through hybridization does not only exist within islands, but that it also seems to link different islands to each other. This is an important prerequisite that indicates that connectivity between islands might be higher than generally expected, with the subsequent impact on colonization ability and diversification.

Using the analysis of different programs we found evidence for contemporary and historical gene-flow between islands. This was indicated by recent connectivity within the eastern and western lineages and by historical connectivity among the central species of Tenerife, Gran Canaria, and La Gomera. Tenerife and Gran Canaria are the largest and most diverse islands of the archipelago, making them more likely to work as

migration sources. High historical migration rates are found between the two major lineages (east and west) but recent gene-flow between them is found to be sparse. Barriers to gene-flow might accumulate with time and can affect how genetic diversity is distributed. Nevertheless, introgression between these lineages is still found especially through hybridization in La Gomera. This indicates that although highly diverged, these lineages are not yet completely isolated, which can be a consequence of continuous genetic exchange during the divergence of these lineages [1].

In our previous study, we determined gene-flow between species within Tenerife, most pronouncedly between the species in the central part of Tenerife and members of the *M. varia* group. Gene-flow between these species is also indicated by earlier phylogenetic reconstruction using a multilocus analysis, where incongruences between markers prevent a clear phylogenetic assignment of the respective taxa [24, 31]. This can be either attributed to ancestral lineages not yet sorted out [46], or, at least occasionally, gene-flow creates a picture of reticulate evolution [47]. Nevertheless, analysis of contemporary gene-flow indicates that this is an ongoing process and that the species are connected by the occasional hybrids that can be observed. We show here that this is not only the case with Tenerife, but also with Gran Canaria and La Gomera, i.e. in all islands where more than one species can be found. Interestingly, Tenerife shows a larger amount of gene-flow than Gran Canaria when the total amount of migration is considered. This might be consequence of the different geological history of each island, resulting in a different age of the respective species groups, and thus different degree of reproductive isolation and different pattern of potential geographic isolation. Both islands are composed by old and young substrates from the Miocene and Pleistocene [19, 48]. Puppo et al. [24, 25] found genetic structure among species correlated with these regions in Tenerife, which was confirmed by our data. Differences in gene-flow between species, and age estimates of phylogenetic relationships between the two largest islands, might be related to differences in the geological formation of the islands. While both are of similar maximum age, Tenerife as we know it today was created by the unification of three older islands by a central area while Gran Canaria got its current shape by subsequent additions of land in the northeast direction [19, 48]. These different island formation processes resulted in different ways of land occupation. While the central, younger part of Tenerife was occupied by colonization from the older regions creating a contact zone, younger areas of Gran Canaria were occupied by expansion of pre-existing populations. Results from Puppo *et al.* [24, 25] support the idea that the central part of Tenerife was colonized after its emergence and only later it came into contact with the palaeo-island species.

Genetic differentiation, diversity patterns, and the surfing syngameon hypothesis

Hybridization can be a driver for adaptive evolution [11, 49] and the underlying mechanism could explain different patterns of evolutionary success, e. g. range expansion in the context of biological invasions [50, 51], or polyploid speciation [52, 53]. Hybridization combines genotypes that evolved independently resulting in the appearance of new genomic rearrangements. This can impact adaptive ability during colonization in two ways. First, new genotype combinations could result in novel characteristics so that sorting out processes can lead to the occupation of different ecological niches. This process of hybrid speciation where the hybrid has novel characteristics had been prominently shown for *Helianthus* [54, 55] or *Iris* [56, 57] where characteristics are exchanged between species. Secondly, new diversity might lead to higher phenotypic variability for selection to act upon, providing increased adaptability to facilitate range expansion into lower quality habitat facilitating long term adaptation [53, 58]. The creation of new diversity through genetic exchange has been used to explain the diversification processes on oceanic islands [1, 2, 3]. However, ecological speciation that occurs in islands could be particularly promoted by high levels of genetic diversity.

The most recent model that includes the effect of genetic diversity on adaptive evolution is the surfing syngameon hypothesis, which implies that islands close to the mainland should have a higher likelihood to receive multiple colonizers from different sources [3]. When these colonizers hybridize with each other and with new arriving colonizers, they are forming a syngameon, a group of hybridizing species that evolve like one joint group of organisms [3, 14]. The islands closer to the mainland constitute therefore recipient islands for colonists and genetic diversity (allelic sinks, [3]). Hereby, genetic diversity is created by recombination between these colonists. When the remaining archipelago is colonized, colonizers are likely to stem from these allelic sinks. The high genetic diversity resulting from the allelic sink effect might facilitate the establishment on the remaining islands by provision of suitable genotypes. Single lineages can then establish in different islands and form the base for future species diversity. Effectively, this allows to hypothesize a reduction of genetic diversity in a stepping stone model in the course of the range expansion and the colonization of other islands [59]. Therefore, the surfing syngameon hypothesis [3] allows that genetic diversity decreases with colonization steps, something that had been shown in several examples [13], and also after comparison between mainland and island taxa [60]. In this scenario, the syngameon is maintained in the islands by receiving gene-flow as an ongoing process from the mainland. Because this will prevent genetic differentiation, it can be expected that among populations, differentiation will increase during colonization. The hypothesis had been outlined for the Canary Islands where island age decreases

with the distance to the mainland (east to west). This should amplify this effect [3], and not only genetic diversity should be higher, but also genetic differentiation lower in older islands.

Our data supports rather a scenario where the younger islands receive a significant amount of genetic diversity and differentiation is lower than expected when considering the surfing syngameon hypothesis. The existence of gene-flow within and between islands may explain this pattern, where high levels of admixture increase genetic diversity and can ultimately have an adaptive effect. Here, we see that both distance to the source of colonization and time have an effect on diversity. Within populations, genetic diversity was higher in younger islands, while genetic differentiation was always significantly higher for older islands. This might be explained by the following mechanism: in younger islands the colonization process is still ongoing, so they contain a low number of taxa that may not have yet occupied all niches available, making them represented by a lower number of populations and individuals. Colonization might have here a higher impact on the genepool of already established populations. In older islands, taxa had time to expand and are now represented by higher number of populations, individuals, and species. Thus gene-flow into these islands might have a lower effect and does not contribute much to changes on the genepool of established populations. Therefore, colonizers should have a lower effect preventing population differentiation on the older islands, and also a lower effect of increasing within-population diversity. This scenario would correspond to models explaining biodiversity as species richness by Chen and He [61]. The model shows that, at some point, immigration would have less impact in the appearance of new species and speciation events become the main source of new biodiversity.

The surfing syngameon hypothesis implies that the sink populations in a source sink dynamics can have a genetic accumulative nature by maintaining their connection to their source through gene-flow. We can apply these assumptions to explain our results by presenting a scenario of inter-island colonization with continuous gene-flow. After emergence of an island, individuals that can hybridize, colonize it outgoing from different sources. In this case the syngameon expands with the appearance of new islands and prevents the loss of genetic diversity at the colonization front by buffering founder effects. In the Canary Islands this has an east-west direction. With the syngameon expansion populations in old islands might become isolated, speciate, and thus become disconnected with increasing reproductive isolation. This dynamics is similar to the model of species range shifts: in the expansion front, new populations have a high connectivity to the source, and thus they receive higher amounts of gene-flow, but between populations differentiation is low, while the rear relict populations become increasingly

isolated showing low within and high between population differentiation [62]. In the present study, both genetic diversity and gene-flow results support the syngameon expansion. Within populations, genetic diversity is increased on young islands while between-population diversity decreases independently of the dataset used. The central islands (Tenerife and Gran Canaria) are the only sources of historical migration while contemporary migration is originated from western islands like El Hierro. In *Micromeria*, the syngameon is therefore not only expanding and shifting, but it is continuously being recreated at the colonization edge, which might hereby increase the potential for ecologic speciation at the expansion front.

Phylogeographic implications and conclusions

We chose to use microsatellites, which are mostly used at the intraspecific level. Nevertheless, these have successfully been used to evaluate species boundaries and test hypothesis related with species concept [63, 64] supporting our approach.

The genetic structure found in the present study was mostly congruent with previous phylogenetic studies. All species described in Tenerife had already been supported with microsatellite data [25]. At the present study, we were able to do the same for most of the remaining *Micromeria* species, supporting the last *Micromeria* species delimitation from Puppo & Meimberg [28, 29]. STRUCTURE was able to define unique clusters for most species at K=19. With exception of *M. gomerensis* and *M. mahanensis,* all remaining species were differentiated in the analyses containing only individuals from the eastern and western groups. *Micromeria gomerensis* and *M. mahanensis* were separated when EPIC nuclear markers were used [31] so this pattern may be a consequence of characteristics intrinsic to microsatellites' nature. These markers are mostly neutral [65] making them more affected by introgression [66]. In fact, we found a high migration rate between these two species supporting this explanation.

Structure patterns were congruent with the colonization paths described for *Micromeria* in previous work [31] (Puppo *et al.* 2015a). We found a first division between the species from the eastern and western groups. Within these groups, species from the other islands share clusters with Tenerife and Gran Canaria. This supports the findings of Puppo *et al.* [25] where these central islands were reported to play an important role in the diversification and dispersal of *Micromeria*. Further divisions in the structure analyses together with the PCoA and distance dendrogram results allowed us to access more detailed diversification and colonization patterns. The data supported a colonization of the remaining western islands by the palaeo-islands, in case of La Palma and La Gomera from the Teno region. In both STRUCTURE and distance dendrogram the eastern species previously classified as *M. varia* grouped together. In Puppo *et al.*

[25] a similar pattern was found which indicates a common origin for these species.

The results support the existence of three lineages with independent origins in La Gomera: *M. pedro-luisii* originated from Tenerife; *M. gomerensis* originated from an eastern *M. varia* s.l. ancestor; and *M. lepida* with uncertain origin. *Micromeria lepida* showed high assignment to both eastern and western groups depending on the method used. The STRUCTURE analyses clustered it with Gran Canaria for some values of K. In the PCoA it grouped together with *M. pedro-luisii* (La Gomera), *M. herpyllomorpha* (La Palma), and *M. densiflora* (Tenerife). We found evidences of gene-flow between *M. lepida* and *M. densiflora* and in lower amount with *M. pedro-luisii*. In the previous phylogenetic analyses, this species shared a clade with Gran Canaria species supporting an origin from this island [31]. So the incongruences found can be a consequence of high introgression with species from the western group.

Like described before, we found multiple origins for the species previously classified as *M. varia*. This result raises questions related with how this morphology is maintained and if this is a consequence of multiple colonization dynamics described here. Further research is currently being developed in our lab where hypothesis concerning this observation are outlined and tested.

This was the first time neutral genetic variation was studied for *Micromeria* across the Canarian archipelago. All our results are congruent with the existence of syngameons that may facilitate colonization and speciation in oceanic islands. These seem to expand and shift in accordance with a range expansion scenario. Island colonizations have many parallelisms with other evolutionary processes, and theoretically the model proposed here could be applied to scenarios where a range expansion is accompanied by adaptive speciation. We showed how genetic diversity is affected by gene-flow and hybridization, but we do not assess its adaptive implications. To do so, genomic approaches where both coding and non-coding variation is compared with ecological and morphological features might successfully be implemented.

References

1. Herben T, Suda J, Munclinger P. The ghost of hybridization past: niche pre-emption is not the only explanation of apparent monophyly in island endemics. J Ecol. 2005; 93: 572–575.
2. Saunders NE, Gibson DJ. Breeding system, branching processes, hybrid swarm theory, and the humped-back diversity relationship as additional explanations for apparent monophyly in the Macaronesian island flora. J Ecol. 2005; 93: 649–652.

3. Caujapé-Castells J. Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. In Author/Editor, The biology of island floras, company, 2011. p. 284-324.
4. Silvertown J. The ghost of competition past in the phylogeny of island endemic plants. J Ecology. 2004; 92: 168-173.
5. Silvertown J, Francisco-Ortega J. Carine M. The monophyly of island radiations: an evaluation of niche pre-emption and some alternative explanations. J Ecology. 2005; 93: 653-657.
6. Losos JB, Ricklefs, RE. Adaptation and diversification on islands. Nature. 2009;457: 830–6.
7. Heaney LR. Dynamic disequilibrium: a long-term , large-scale perspective on the equilibrium model of island biogeography. Globa Ecol Biogeogr. 2000; 9: 59–74.
8. Whittaker RJ, Triantis KA, Ladle RJ. A general dynamic theory of oceanic island biogeography. Journal Biogeogr. 2008; 35: 977–994.
9. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G, Bank C, Brännström Å et al. Genomics and the origin of species. Nature Rev Genet. 2014; 15: 176–192.
10. Seehausen O. Conditions when hybridization might predispose populations for adaptive radiation. J Evol Biol. 2013; 26: 279-281.
11. Arnold ML. Evolution through genetic exchange. Oxford University Press, Oxford; 2006. p. 34-150.
12. Francisco-Ortega J, Santos-Guerra A, Kim SC, Crawford DJ. Plant genetic diversity in the Canary Islands: a conservation perspective. Am J Bot. 2000; 87:909-919.
13. Pérez de Paz JP, Caujapé-Castells J. A review of the allozyme data set for the Canarian endemic flora: causes of the high genetic diversity levels and implications for conservation. Ann Bot-London. 2013; 111: 1059-1073.
14. Grant V. Plant speciation. Columbia University Press, New York. 1981.
15. Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol. 23, 347-351.
16. Juan I, Emerson B, Orom I, Hewitt G. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends Ecol Evol. 2000;15:104–109.
17. Fernández-Palacios JM, de Nascimento L, Otto R, Delgado JD, García-del-Rey E, Arévalo JR, Whittaker RJ. A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. Journal of Biogeography. 2011; 38: 226-246.

18. Otto R, Whittaker RJ, Gaisberg M, Stierstorfer C, Naranjo-Cigala A, Steinbauer MJ, Fernández-Palacios JM. Transferring and implementing the general dynamic model of oceanic island biogeography at the scale of island fragments: the roles of geological age and topography in plant diversification in the Canaries. Journal Biogeogr. 2016; 43: 911–922.
19. del-Arco M, Salas M, Acebes JR, Marrero MC, Reyes-Betancort JA, Pérez de Paz PL. Bioclimatology and climatophilous vegetation of Gran Canaria (Canary Islands). In Ann Bot Fenn. 2002; 39:15-41.
20. Macías-Hernandez N, Bidegaray-Batista L, Emerson C, Oromí P, Arnedo MA. The imprint of Geologic History on Within-Island diversification of woodlouse-hunter spiders (Araneae, Dysderidae) in the Canary Islands. J Hered. 2013; 104:341–356.
21. Rumeu B, Vargas P, Jaén-Molina R, Nogales M, Caujapé-Castells J. Phylogeography and genetic structure of the threatened Canarian Juniperus cedrus (Cupressaceae). Botanical Journal of the Linnean Society, 2014; 175: 376–394.
22. Jones KE, Reyes-Betancort JA, Hiscock SJ, Carine, MA. Allopatric diversification, multiple habitat shifts, and hybridization in the evolution of Pericallis (Asteraceae), a Macaronesian endemic genus. Am J Bot. 2014; 101:637–651.
22. Mairal M, Sanmartín I, Aldasoro JJ, Culshaw V, Manolopoulou I, Alarcón M. Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic Canaria canariensis (Campanulaceae). Mol Ecol. 2015; 24: 3944–3963.
24. Puppo P, Curto M, Velo-Antón G, Pérez de Paz PL, Meimberg H. The influence of geological history on diversification in insular species: genetic and morphological patterns of Micromeria Benth. (Lamiaceae) in Tenerife (Canary archipelago). Journal Biogeogr. 2014; 41: 1871-1882.
25. Puppo P, Curto M, Meimberg H. Genetic structure of Micromeria (Lamiaceae) in Tenerife, the imprint of geological history and hybridization on within-island diversification. Ecol Evol. 2016; 6: 3443-3460.
26. Bräuchler C, Ryding O, Heubl G. The genus Micromeria (Lamiaceae), a synoptical update. Willdenowia. 2008; 38:363-410.
27. Pérez de Paz PL. Revisión del género Micromeria Bentham (Lamiaceae-Stachyoideae) en la Región Macaronésica. Instituto de Estudios Canarios, Monografías, 1978; 16: 1–306.
28. Puppo P, Meimberg H. New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa. 2015a; 230: 1-21.

29. Puppo P, Meimberg H. Corrections to Phytotaxa 230: New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa. 2015b; 239: 197-198.
30. Meimberg H, Abele T, Bräuchler C, Mckay JK, Pérez PL, Paz D, Heubl G. Molecular evidence for adaptive radiation of Micromeria Benth . (Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data. Mol Phylogenet Evol. 2006; 41: 566-578.
31. Puppo P, Curto M, Gusmão-Guedes J, Cochofel J, de Paz PLP, Bräuchler C, Meimberg H. Molecular phylogenetics of Micromeria (Lamiaceae) in the Canary Islands, diversification and inter-island colonization patterns inferred from nuclear genes. Mol Phylogenet Evol. 2015a; 89: 160-170.
32. Bañares Á, Blanca G, Güemes J, Moreno JC, Ortiz S. Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Dirección General de Conservación de la Naturaleza, Madrid. 2004.
33. Bañares Á, Blanca G, Güemes J, Moreno JC, Ortiz S. Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Adenda 2010. Dirección General de Medio Natural y Política Forestal (Ministerio de Medio Ambiente, y Medio Rural y Marino)-Sociedad Española de Biología de la Conservación de Plantas, Madrid. 2010.
34. Curto MA, Puppo P, Ferreira D, Nogueira M, Meimberg H. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). Mol Phylogenet Evol. 2012; 63:758-767.
35. Puppo P, Curto M, Meimberg H. Development and characterization of 16 microsatellite markers for Micromeria (Lamiaceae) from Tenerife (Canary Islands, Spain) using 454 sequencing. Conservation Genetics Resources. 2015b; 7: 743-749.
36. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Mol Ecol Res. 2009; 9:1322–1332.
37. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005; 14:2611-2620.
38. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 1978; 89: 583-590.
39. Swofford DL. PAUP Phylogenetic analysis using parsimony (and other methods), Version 4.0. Sinauer Associates, Sunderland, Massachusetts. 2002.

40. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. Genetics. 2003; 163: 1177–1191.
41. Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. P Natl Acad Sci USA. 2001; 98:4563–4568.
42. Rambaut A, Drummond AJ. TRACER v1.4. 2007. Available from: http://beast.bio.ed.ac.uk/Tracer
43. Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. Genetics, 2002; 160:1217-1229.
44. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 1358-1370.
45. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 1995; 139: 457-462.
46. Avise JC. Phylogeography: the History and Formations of Species. Harvard University Press, Cambridge, MA. 2000.
47. Linder CR, Rieseberg LH. Reconstructing patterns of reticulate evolution in plants. Am J Bot. 2004; 91: 1700-1708.
48. del-Arco M, Pérez de Paz PL, Acebes JR, González-Mancebo JM, Reyes-Betancort JA, Bermejo, JA, González-González R. Bioclimatology and climatophilous vegetation of Tenerife (Canary Islands). In Ann Bot Fenn. 2006; 43:167-192.
49. Arnold, ML. Divergence with genetic exchange. Oxford University Press. Oxford, 2015. p. 34-129.
50. Kolbe JJ, Glor RE, Schettino LR, Lara AC, Larson A, Losos JB. Genetic variation increases during biological invasion by a Cuban lizard. Nature. 2004; 431:177-181.
51. Meimberg H, Milan NF, Karatassiou M, Espeland EK, McKay J, Rice KJ. Patterns of introduction and adaptation during the invasion of Aegilops triuncialis (Poaceae) into Californian serpentine soils. Mol Ecol. 2010; 19: 5308-5319.
52. Soltis DE, Soltis PS. Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol. 1999; 14: 348-352.
53. Meimberg H, Rice KJ, Milan NF, Njoku CC, McKay JK. Multiple origins promote the ecological amplitude of allopolyploid Aegilops (Poaceae). Am J Bot. 2009; 96: 1262-1273.
54. Rieseberg LH, Carter R, Zona S. Molecular tests of the hypothesized hybrid origin of two diploid Helianthus species (Asteraceae). Evolution. 1990; 17: 1498-1511.

55. Lexer C, Welch ME, Raymond O, Rieseberg LH. The origin of ecological divergence in Helianthus paradoxus (Asteraceae): selection on transgressive characters in a novel hybrid habitat. Evolution. 2003; 57: 1989-2000.
56. Arnold ML, Buckner CM, Robinson JJ. Pollen-mediated introgression and hybrid speciation in Louisiana irises. P Natl Acad Sci USA. 1991; 88, 1398-1402.
57. Brothers AN, Barb JG, Ballerini ES, Drury DW, Knapp SJ, Arnold ML. Genetic architecture of floral traits in Iris hexagona and Iris fulva. J Hered. 2013; est059.
58. Dyer AR, Rice KJ. Effects of competition on resource availability and growth of a California bunchgrass. Ecology, 1999; 80:2697-2710.
59. Clegg SM, Phillimore AB. The influence of gene flow and drift on genetic and phenotypic divergence in two species of Zosterops in Vanuatu. Philos Trans R Soc Lond B Biol Sci .2010; 365: 1077-1092.
60. García-Verdugo C, Sajeva M, La Mantia T, Harrouni C, Msanda F, Caujapé-Castells J. Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. Mol Ecol. 2015; 24:726-741.
61. Chen XY, He F. Speciation and endemism under the model of island biogeography. Ecology. 2009; 90:39-45.
62. Hampe A, Petit RJ. Conserving biodiversity under climate change: the rear edge matters. Ecol Lett. 2005; 8:461-467.
63. Bonatelli IAS, Perez MF, Peterson T, Taylor NP, Zappi DC, Machado MC, Koch I, Pires A, Moraes EM. Interglacial microrefugia and diversification of a cactus species complex: phylogeography and palaeo distributional reconstructions for Pilosocereus aurisetus and allies. Mol Ecol. 2014; 23:3044–3063.
64. Turini FG, Steinert C, Heubl G, Bringmann G, Lombe BK, Mudogo V, Meimberg H. Microsatellites facilitate species delimitation in Congolese Ancistrocladus (Ancistrocladaceae), a genus with pharmacologically potent naphthylisoquinoline alkaloids. Taxon. 2014; 63: 329–341.
65. Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett. 2006; 9: 615-629.
66. Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK. The variable genomic architecture of isolation between hybridizing species of house mice. Evolution. 2010; 64: 472–485.

Availability of data and material

Microsatellite data matrix is deposited in Demiurge as: Puppo P , Curto M , Meimberg H D-NMICR-105. 2016. http://www.demiurge-project.org/matrix_digests/D-NMICR-105

Funding

This research was supported by the Fundação para a Ciência e Tecnologia (FCT) with a research grant to H.M. (PTDC/BIA-BEC/108866/2008) co-financed by the Fundo Europeu de Desenvolvimento Regional (FEDER) through COMPETE – Programa Operacional Fatores de Competitividade (POFC) (FCOMP-01-0124-FEDER-008988), and PhD fellowships to P.P. (SFRH/BD/74747/2010) and M.C. (SFRH/BD/79010/2011).

Author Contributions

All authors were involved in writing the manuscript. H.M. developed the research idea with support of the other authors, and obtained funding for the research. Laboratory work was performed by M.C. and P.P. and analyses mostly by M.C. Field work was performed by P.P. and H.M.

Acknowledgements

We would like to thank Christian Bräuchler and Pedro Luis Pérez de Paz for providing samples and helping with field work. We also thank Susana Lopes for helping with microsatellites scoring.

Chapter 4: Distribution of adaptive traits during adaptive radiation.

Paper I: Curto M., Schachtler C., Puppo P., Meimberg H. (submitted) Phylogenetic analyses of RAD data to infer the evolution of *Micromeria* in the Canary Islands.

4.1. Phylogenetic analyses of RAD data to infer the evolution of *Micromeria* in the Canary Islands.

Manuel Curto[1,2], Christina Schachtler[3,4], Pamela Puppo[2], Harald Meimberg[1]

[1]Institute for Integrative Nature Conservation Research, University of Natural Resources and Life Sciences, A-1180 Vienna, Austria
[2]CIBIO, Research Center in Biodiversity and Genetic Resources / InBio Associated Laboratory, University of Porto, Campus Vairão, Vairão 4485-661, Portugal
[3]Chair of Restoration Ecology, Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Technische Universität München, D-85354 Freising, Germany.
[4]Cell Biology, Department of Biology, University of Erlangen-Nuremberg, D-91058 Erlangen, Germany

Abstract

Some of the endemics of *Micromeria* from the Canary Islands are morphologically very similar, so they had been previously recognized as one species, *M. varia* s.l., with different subspecies described in each island. Ecological requirements of these units are very diverse, making them a good model to investigate niche shifts and adaptation. Previous molecular studies have failed to reconstruct their phylogenetic relationship presumably due to introgression or incomplete lineage sorting. In the present study, we develop a RAD-sequencing variant with which we recovered 3571 loci genotyped for a total of 45 individuals from *Micromeria* in order to improve phylogenetic resolution and better understand the relationship among these taxa.

RAD-sequencing data produced a highly resolved phylogenetic tree where all *M. varia* species corresponded to independent units for each species confirming the latest species reclassification. This allowed to uncover patterns among some of the species that were missed in previous studies. More specifically, we found a close phylogenetic relationship between taxa occupying similar ecological conditions in different islands, exemplified by the taxa from the laurel forest from La Gomera and Gran Canaria. We hypothesize that either the laurel forest worked as a filter and only allowed the establishment of colonizers already pre-adapted to these conditions, or the genes that facilitated the adaptation to laurel forest were exchanged through introgression from Gran Canaria to La Gomera or vice versa. The observations obtained in this study can serve as bases for research where potential adaptive related genetic variation is compared to phylogeographic patterns, which will ultimately result in a more comprehensive view of the radiation processes in oceanic islands. We also present a

brief discussion concerning possible biases associated with RAD-sequencing methods when applied to phylogenetic studies.

Keywords: Phylogenomics, oceanic islands, *Micromeria,* niche conservatism, adaptive radiation.

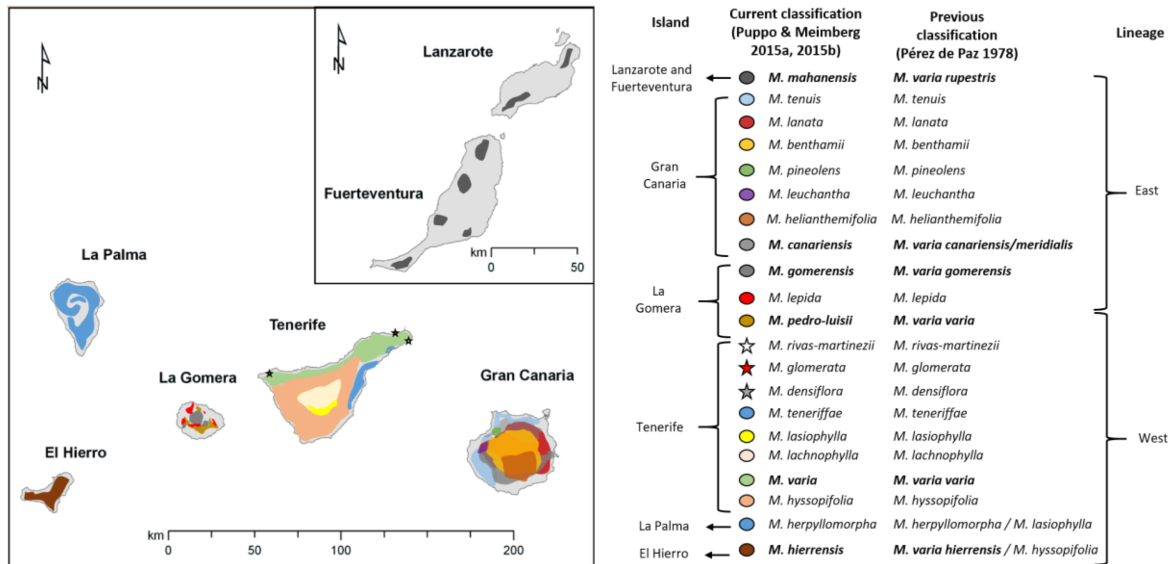
Introduction

Oceanic islands typically harbor a high number of endemic species which can occur either as single island endemics or as archipelago endemics occurring in multiple islands. The latter usually present some degree of niche conservatism, meaning that the taxa occupy the same ecological conditions in different islands (Francisco-Ortega et al., 1996) and may or not be connected by continuous genetic exchange. However, in some cases, such species can occupy different islands' niches and thus constitute examples of niche shift, i.e. the change of the ecological niche during differentiation and colonization (Kim et al., 1996; Francisco-Ortega et al. 1996). In the genus *Micromeria*, *M. varia sensu* Pérez de Paz 1978 (hereafter M. varia *s. l.*), represents a taxon, with an archipelago-wide distribution on the Canary Islands and associated to a wide variety of habitats (Pérez de Paz 1978). Based on recent phylogenetic results (Puppo et al., 2015) and morphological variation *M. varia* was separated in several species, all but one (*M. mahanensis*), single island endemics (Puppo and Meimberg 2015 a, b; Fig. 1); however, the morphological similarities might be a consequence of adaptation to similar ecological zones in the different islands. In the present study we reconstruct in detail the phylogenetic relationships of *M. varia s. l.* using a RAD-sequencing variant in order to further support this hypothesis as well as to explore the mechanisms behind this process.

The genus *Micromeria* Benth. (Lamiaceae) distributed in Asia, all Mediterranean basin, Macaronesia, and some regions of Africa and Asia with approximately 54 species (Bräuchler et al., 2008). In the Canary Island it has a center of diversity with 21 single island endemic species (Pérez de Paz 1978; Puppo and Meimberg 2015a, b). *Micromeria varia s.l.* is hereby treated as 6 several independent species, The central islands, Gran Canaria and Tenerife, are the most diverse with eight and seven species respectively, while the remaining islands have three (La Gomera) or one species each (Fig. 1). The Canary archipelago was formed by a hotspot movement from east to west, resulting on islands with an age gradient in the same direction. The islands have different altitudinal gradients depending on their age (Fernández-Palacios et al., 2011), and harbor a range of ecological zones from costal desert, passing through laurel forest, to sub-alpine desert (Juan et al., 2000). Most *Micromeria* species distribution is associated with these ecological zones indicating that most of the diversity found in the Canary

Islands arose through adaptive radiation (Meimberg et al., 2006; Puppo et al., 2014, 2015a). Similar morphological traits might occur in taxa in different islands in consequence to similar ecological zones.

Fig. 1. Up: Map of the Canary Islands showing the distribution of the species of *Micromeria* present in the Canary Islands. Down: species classification according to both Pérez de Paz 1978 (1978) and Puppo and Meimberg (2015 a, 2015b) indicating to which lineage (east or west) each species is assigned to according to molecular studies. Species showing the *M. varia* phenotype are indicated in bold.



Micromeria in the Canary Islands is considered to be monophyletic and divided in two major lineages (Puppo et al., 2015a; Fig. 1): one composed by the eastern islands of Lanzarote, Fuerteventura, and Gran Canaria, and the other by the western islands of Tenerife, La Palma and El Hierro. Two taxa from La Gomera (*M. gomerensis* and *M. lepida*) belong to the eastern clade while the third taxon, *M. pedro-luisii,* belongs to the western lineage. Gran Canaria and Tenerife are paraphyletic relatively to the remaining islands and, with exception of La Gomera, all taxa from the other islands are monophyletic. Puppo et al. (2014, 2015a, 2016) showed that *Micromeria* diversification pattern was highly congruent with the Canary Island's geological history. Puppo et al. (2016) and Curto et al. (sub.) found that interspecific gene-flow is frequent in *Micromeria*, especially among species from the same island. In these studies, the authors suggested that, when a new island emerges, each island works as an allelic sink due to the combination, through hybridization, of lineages coming from different sources. This process results in the maintenance of high genetic diversity during colonization, increasing the likelihood of adaptive radiation events and providing opportunities for adaptive introgression (Caujapé-Castells 2011). Furthermore, the high frequency of

hybridization events might have also had an impact in the phylogenetic signal of previous studies, suggesting the use of phylogenomic approaches.

No genomic information is available for *Micromeria* so parallel sequencing of the complete genome is not feasible to date. Instead, a reduced representation sequencing technique was used allowing the screening of thousands of loci across the genome without the need of previous genetic information (Cronn et al., 2012). The most common approach uses restriction enzymes to reduce genome complexity (i.e. Baird et al., 2008; Elshire et al., 2011; Peterson et al., 2012; Poland et al., 2012). This results in a new class of markers where the regions associated with enzymatic cleavage are sequenced. Baird et al. (2008) were the first to introduce this principle calling the marker type RAD (Restriction Associated DNA). Variations of this method were presented, all of them having in common: the use of a restriction enzyme, selection of fragments with the appropriate size, and the ligation of adapters specific to a sequencing platform (Andrews et al., 2016). In the present study we developed a new RAD-sequencing variant where only one restriction enzyme is used and a size selection is made through agarose gel electrophoresis, reducing the number of fragments in a way that the Illumina MiSeq could be used for sequencing and also decreasing costs. In this variant, individuals are identified by adding both barcodes and indexes through adapter ligation and PCR, respectively. Because the Illumina platform is being used, it is necessary to add different adaptors in opposite ends of each fragment. For this, an adapter containing partly complementary oligonucleotides, commonly called Y-adapter, was added.

In the present study, we use a RAD-sequencing variant in order to recover a better resolved phylogeny of *Micromeria* in the Canary Islands. Since RAD-sequencing allows assessing thousands of loci across the genome from both coding and non-coding regions, we expect to obtain a more comprehensive view of the evolutionary history of this group. Particularly, we aim to solve some of the ambiguous relationships among the species of the *M. varia* species complex and test our hypothesis that the morphological similarities found among its taxa result from adaptation to similar ecological zones. The RAD-sequencing variant method here developed is also described in detail below.

Materials and methods

Samples used and DNA isolation

Most samples used belong to species previously classified as *M. varia* (Supplementary Table S1). If present, at least one population per species per island was included. Some individuals were used in previous studies (Curto et al., 2012; Puppo et al., 2014, 2015a, b, 2016) while the remaining ones were sampled in the same excursions described in these publications. DNA isolation was done using the

NucleoSpin® Plant kit from Macherey-Nagel's (Düren, Germany) applying the manufacturer's recommendations for dried leaf material. DNA quality and quantity was evaluated through agarose gel electrophoresis in a comparative way. Samples were classified according to their DNA quality as: High (high amount of intact DNA); Medium (medium amount of intact DNA); Low (low amount of intact DNA); and smeared (fragmented DNA). Only samples with some visible DNA content were used in the next steps resulting in a total of 93 initial samples representing 25 populations and 11 species (Supplementary Table S1).

Method overview

With our approach we intended to reduce genome's complexity allowing sequencing a higher number of samples. This was obtained by using a rare cutter as restriction enzyme (EcoRI) and by selecting fragments of a certain size range through gel electrophoreses. We reduced the initial investment in primers and adaptors by using a combination of barcodes and indexes as molecular identifiers. Barcodes appear in the beginning of each read while Indexes are read directly by the Illumina machine. The addition of Illumina's P5 and P7 was possible by the ligation of a Y-adaptor. This is an adaptor where only part of its length is overlapping, and in our case contained the barcode information as well. One of the oligonucleotides present in this adapter serves as template for index primers containing the P5 and P7 flow cell binding motifs. These primers also contain the Index information. This method is divided in four steps and is summarized in Fig. 2: 1) DNA digestion with EcoRI and ligation of Y adaptor containing the barcode; 2) Multiplex of samples with different barcodes and first size selection in gel electrophoreses; 3) Two step PCR where Indexes are added; and 4) Multiplex of samples with different Indexes and second size selection electrophoreses to select the target size range and exclude excess primers. Details about each of these steps are described below. All adapters and primers information is available in the Supplementary Table S2.

Step 1: DNA digestion and Y-adaptor ligation

The Y adapter was obtained by combining 2 nmol of each oligonucleotide in 50 µL of a Tris-HCl 8 µM solution and incubated in the following temperature scheme: initial incubation at 97 °C for 2 minutes followed by a gradual decrease of temperature of 1 °C every 30 second until 25 °C where it incubated for an additional 5 minutes. As explained before the non-overlapping part is used to add the flow cell specific elements to the final fragment. The overlapping part is composed, from 5' to 3' by Illumina sequencing primer annealing motif, a 5 bp barcode and a sequence overhang (TTAA) in the P7

oligonucleotide that allows the ligation to the fragments digested with EcoRI. Restriction enzyme digestion and adaptor ligation were done in the same reaction tube containing 10 µL of DNA, 20 unities of EcoRI (New England Biolabs, Ipswich, Massachusetts, USA), 3.5 unities of T4 DNA (Promega, Fitchburg, Wisconsin, United States), 2 µg of BSA (Promega), 20 pmol of ATP, 0.4 pmol of Y-adapter, 2 µL of each 10X T4 and EcoRI buffers, and water for a total volume of 40 µL. The reaction was performed at 37 °C for 8 hours followed by an enzymatic inactivation period of 65 °C for 20 minutes.

Step 2: First multiplex and size selection

Nine samples with similar DNA quality category (see “Samples used and DNA isolation”) were combined to prevent the over representation of any individual in one multiplex library. To make gel loading possible, a total of 40 µL from each multiplex was concentrated into final volumes of 15 µL using the NucleoSpin Gel and PCR Clean-up Kit from Macherey-Nagel. Size selection was performed by running the samples in a 1.5% TAE agarose gel and regions ranging from 200 bp to 650 bp were extracted. DNA was purified from the gel using the same Macherey-Nagel kit following manufacturer’s recommendations.

Step 3: Two-step PCR for Index addition

This step was divided in two PCR reactions using a total of three primer pairs. The first PCR was done with the sequencing primers to create motifs complementary to the Index primers and to increase the amount of template. In the second PCR, indexes and flow-cell ligation motifs were added by using the Index and Amplification primers. The latter primers were used to enrich fragments containing index information and were composed by the last 25 and 24 bases of the P5 and P7 Index primers. Two independent PCRs with different index combinations were performed for each sample to randomize the effects of index sequence content (Supplementary Table S1). The first PCR was conducted in a 12.5 µL reaction containing 0.8 pmol of each primer, 3.5 µL of Qiagen’s Multiplex PCR kit (Qiagen, Venlo, Netherlands), and 1 µL of DNA template. The second PCR was performed in a 25 µL solution with 0.08 pmol of each Index-primer, 0.8 pmol of each amplification primer, 8 µL of Qiagen’s Multiplex PCR mastermix, and 1 µL of the product of the previous PCR diluted 1:10. Both PCR reactions were made with the following temperature conditions: Initial denaturation at 95 °C for 15 minutes; 35 cycles of denaturation at 95 °C for 30 seconds, annealing at 57 °C for 30 seconds, and extension at 72 °C for one minute followed by a final extension at 72 °C for 10 minutes.

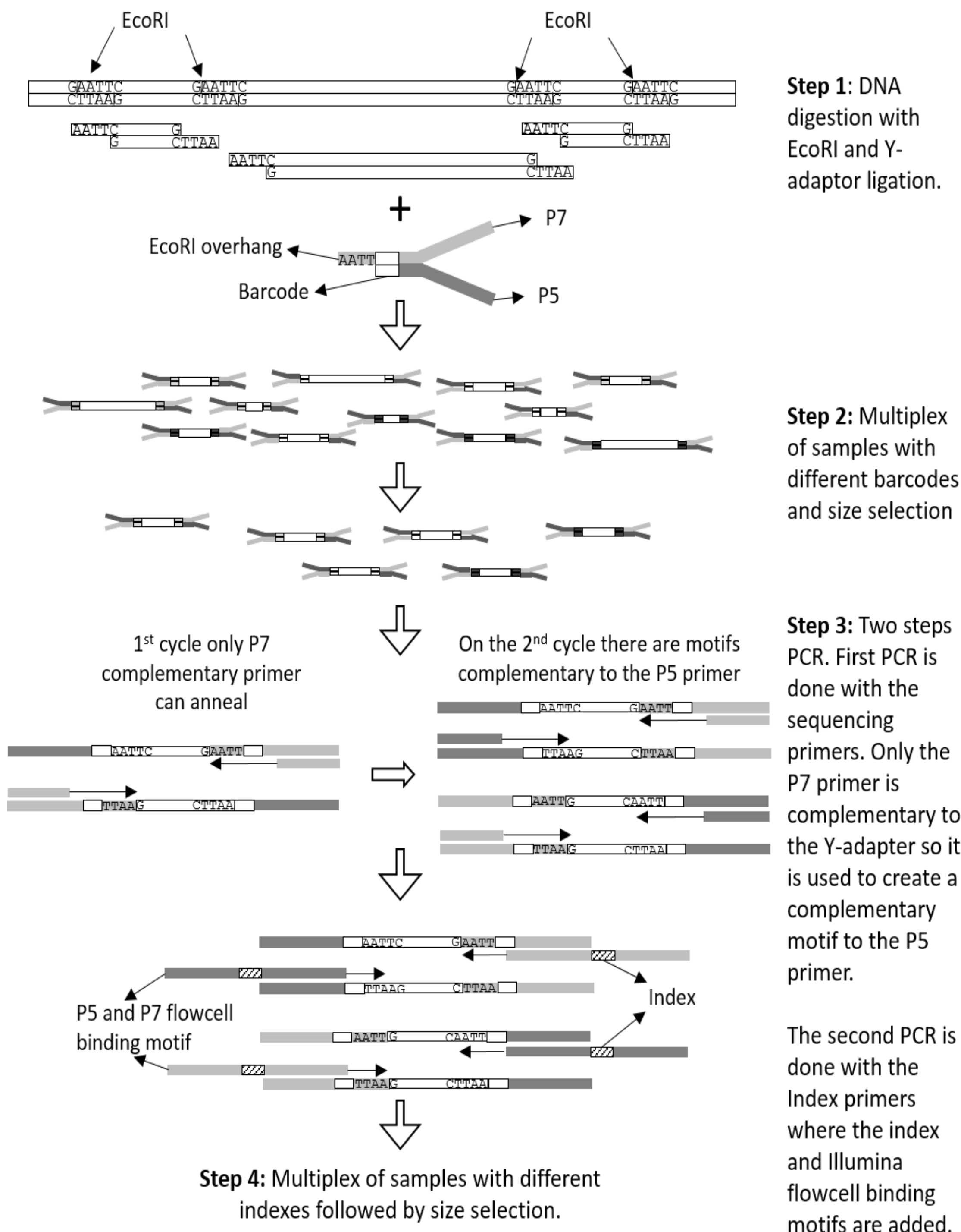


Fig. 2. Schematic summary of the RAD-sequencing library preparation method developed applied in the current study. This was divided into four steps, further details of which can be found in the text beside each figure section.

Step 4: Second multiplex and size selection

Index multiplex solution was obtained by mixing 10 µL of each Index PCR reaction. This solution was then run in a 1.5% agarose TAE gel and regions between 300 bp and 650 bp were collected for gel extraction. This last DNA purification step was

performed by the NucleoSpin Gel and PCR Clean-up Kit from Macherey-Nagel following the manufactures recommendations.

Library quality assessment and sequencing

Library quality was evaluated by measuring its DNA concentration in a Qubit® DNA assay (Life Technologies, Carlsbad, USA); the length profile was assessed in a Agilent 2100 Bioanalyzer machine (Santa Clara, USA). Illumina paired-end sequencing was performed in a 600 cycles (300 in each direction) Illumina MiSeq run (Illumina, San Diego, California, USA). Library quality control and sequencing was done as a service in the Genomics Service Unit from Ludwig-Maximilians-University (Munich, Germany).

Sequence analyses

Sequence quality was evaluated with the program FastQC v. 0.10.1 (Andrews 2010). Regions containing low quality bases and adaptor's sequences in the 3' end were removed by the program Cutadapt v. 1.5 (Martin 2011). Cutadapt trimmed low quality regions with an average Phred quality below 20. Adapter trimming was done by looking for matches with the Illumina adaptors and removing it together with the 3' down streaming sequence. Only matches of at least 20 bp with a maximum mismatch of 15% were considered. In addition, only paired reads larger than 70 bp were kept.

We merged paired reads using the program PEAR v. 0.9.4 (Zhang et al., 2014). Only minimum overlaps of 15 bp with a p-value below 0.01 for the highest observed expected alignment scores (OESs) were taken. Reads that did not overlap were put together in the same sequence with four "n" characters separating them. We used a custom python script to separate the reads according to their barcode (Supplementary Table S2). The script searches for barcode and restriction site motifs in both sequence ends only keeping sequences with these motifs in the expected position. The barcodes were cut off so the output sequences start and end with the restriction enzyme binding motif. A maximum allowed mismatch of two was applied for both restriction motifs and barcodes. Because the barcodes AGCAT and ATCAT only differ in one bp we did not allow any mismatch while searching them.

PyRAD v. 3.0.63 (Eaton 2014) was used to assemble RAD loci and find homology among samples. This was done in two clustering steps. In the first one reads were clustered within the same individual with a similarity above 88 % into potential loci. In the second, the loci defined for each individual were compared among samples and put together into alignments if they had a similarity above 85 %. With our approach each paired read is sequenced from both directions making it necessary to consider its reverse complement while clustering. This was done by implementing the GBS clustering method

available in PyRAD. Only loci with a minimum depth of 10 reads and with heterozygosity below 0.5 were considered.

The resulting loci were blasted against themselves to look for the presence of paralogs. This was done with Blastn v. 2.2.28 (Camacho et al., 2009); only hits with at least 50% overlap of the total length of the respective sequences, and a minimum similarity of 85% (the threshold used for clustering among samples) were considered. Loci showing significant hits with other loci were excluded from further analyses. The data obtained was then filtered in two steps. First, we excluded individuals with missing data above 95%, and second, all loci with missing data above a certain threshold were excluded. Because we wanted to test the influence of missing data in our final results, two threshold values were used, 90% and 50%, resulting in two final datasets.

Phylogenetic analyses

Phylogenetic relationships were assessed using a Bayesian analyses with the program Mr.Bayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003). Since we genotyped a high number of loci from all over the genome with distinct evolutionary histories, we implemented the most complex model in the Bayesian analyses (GTR+G+I). Analyses were done with two independent runs containing four chains each and saving trees every 1000th generation. We ran Mr.Bayes until all parameters reached an effective sample size of 100. This was assessed in Tracer v. 1.6 (http://tree.bio.ed.ac.uk/software/tracer/) as was the appropriate burnin value. The program ran for 2,000,000 generations from which the first 25% were excluded.

We assessed tree resolution by estimating the percentage of node with posterior probability above 0.7 when compared with the maximum number of possible nodes. A tree with N terminal branches can have a maximum number of internal nodes equal to N-1.

Results

Sequence analysis

The MiSeq run resulted in a total of 15,923,850 paired reads from which 8,892,248 passed the quality control step. From those 8,673,421 (97%) overlapped creating sequences with a size range from 70 bp to 592 bp. A total of 4,915,911 sequences had barcodes and restriction enzymes cutting motifs in the correct position. The number of sequence per individuals ranged from 462 for GCle_501 (*M. leucantha* from Gran Canaria) to 250,173 for GCv_457 (*M. canariensis* from Gran Canaria) (Supplementary Table S1). PyRAD recovered 6403 loci from which 242 were excluded

because they showed several hits on the blast search. These corresponded to 154 loci that may have several copies in the genome. The number of loci per individual varied between 1 and 3021, while the individual with the least number of loci from the 45 samples kept had 200 loci (Supplementary Table S1). All taxa showing the *M. varia* phenotype from both eastern and western lineages were kept. After filtering loci according to missing data, 3571 loci were kept in the 90% missing data matrix and 104 on the 50% missing data matrix. As shown in Table 1 the matrix containing more missing data was larger and it contained more variable and informative positions.

Table 1. Alignment information content from the two missing data extractions including: number of loci used in the alignment; number of total, variable and informative characters, and portion of branches with a posterior probability above 0.7.

Missing data	Nr. loci	Nr. characters	Variable characters	Informative characters	Portion of supported branches		
					Complete tree	Eastern lineage	Western lineage
90%	3571	653083	18199	7449	87%	97%	60%
50%	104	20671	558	286	44%	47%	40%

Phylogenetic analysis

We tested the effect of missing data and lineage content on tree resolution. This was measured by assessing which percentage of the maximum number of possible nodes had a support above 70%. Higher missing data resulted in more supported nodes for all lineages (Table 1). The eastern clade was more resolved than the western one in both datasets although this difference is not as pronounced in the 50% missing data matrix (Table 1). Both trees showed a division between the eastern and western *Micromeria* lineages with the same species content described in previous phylogenetic analyses (Puppo et al., 2014, 2015a). Given its higher degree of resolution, only the 90% missing data tree is being included (Fig. 3). The other can be found in supplementary material (Fig. S1).

In the eastern lineage, for both analyses, all taxa of *M. varia* s. l. created a monophyletic group. However, only the 90% missing data matrix was resolved enough to recover informative phylogenetic relationships within this group. The eastern species of *M. varia* s. l. were divided into four lineages (Fig. 3): *M. mahanensis* from Lanzarote, which is in a neighbor group position relatively to the remaining taxa; *M. gomerensis* from La Gomera and *M. canariensis* from Gran Canaria, which is paraphyletic since the individuals from the laurel forest in Gran Canaria are closer to *M. gomerensis* than to the remaining *M. canariensis* and form a group with *M. gomerensis* (Fig. 3). The relationships of the remaining species from the eastern group are as follows: *M. benthamii* is sister to the *M. varia* clade followed by a clade composed of one *M. gomerensis* individual and

one sample of hybrid morphology (*M. lepida* x *M. pedro-luisii*). *Micromeria helianthemifolia* appears at the base of the eastern lineage.

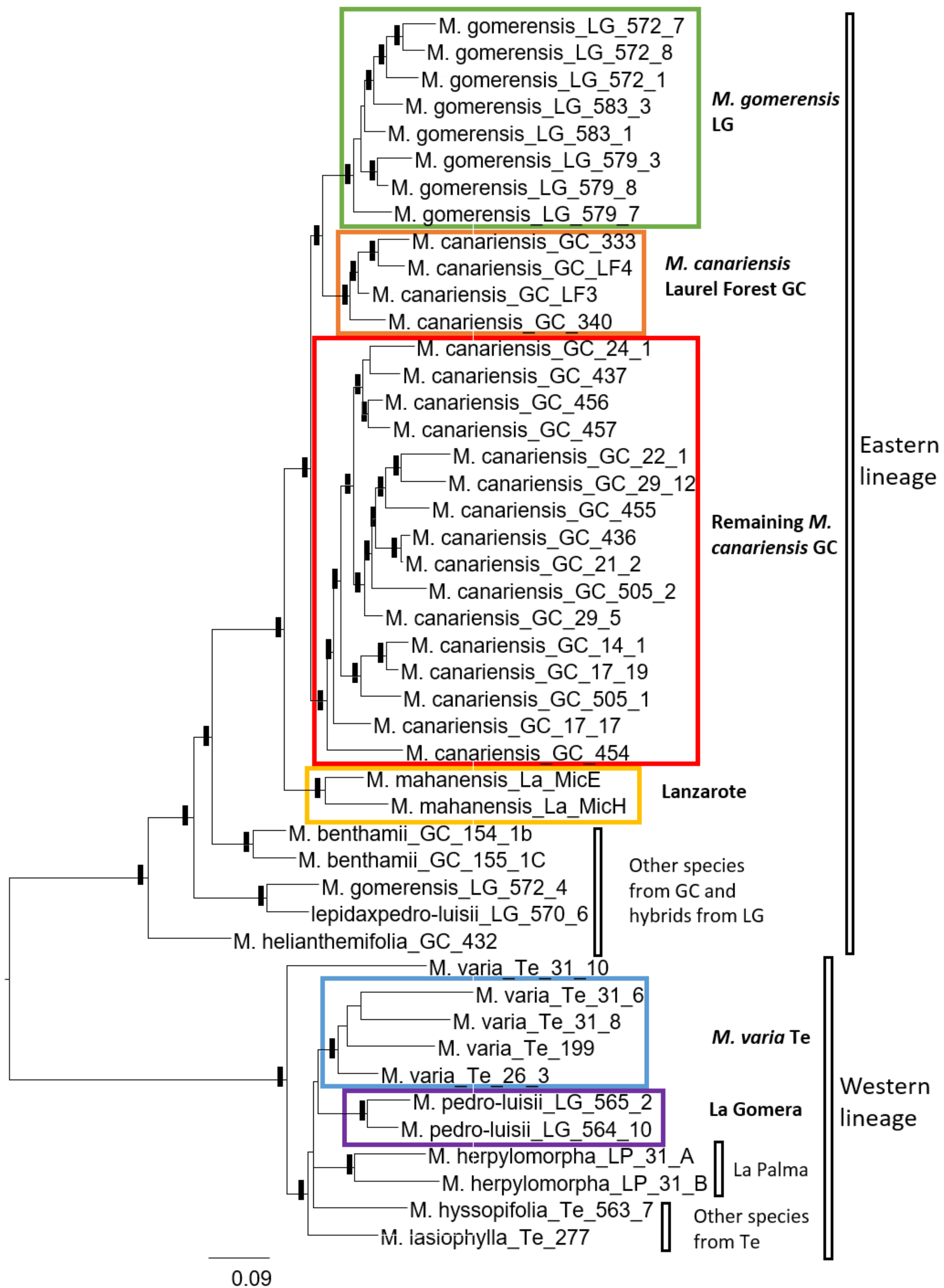


Fig. 3. Mr.Bayes tree calculated with the dataset including only loci with a maximum missing data of 90%. Bars in the nodes correspond to posterior probability values >90.

In the western group, La Palma and La Gomera samples form single clades. Most *M. varia* s. s., the species from Tenerife, form a group sister to La Gomera, except for one individual which appears unresolved to western lineage. *Micromeria lasiophylla* is the neighbor group to the remaining taxa with *M. hyssopifolia* forming a polytomy with La Palma and La Gomera-*M. varia* s.s. clades. The maximum 50% missing data matrix showed a similar pattern but *M. hyssopifolia* grouped with *M. lasiophylla.*

Discussion

Improvement of Micromeria phylogeny

The phylogeny obtained in the present study was mostly congruent with previous studies for *Micromeria* from the Canary Islands (Puppo et al., 2015a), and as expected, we were able to obtain a more detailed phylogenetic pattern with our RAD approach. Particularly, each one of the species previously classified as M. varia constituted independent entities. This result is in agreement with the most recent classification of these species made by Puppo & Meimberg (2015a, b). Solely based on morphological data, Pérez de Paz (1978) recognized different subspecies within *M. varia*, roughly one per island. The reevaluation of these morphological differences led to its reclassification and was indicated by microsatellite data. The results of the present phylogeny further support the recognition of one lineage per island.

The higher resolution obtained in the present phylogeny is probably related with the number and type of loci analyzed. Previous studies used codominant information from two (Meimberg et al., 2006) to 15 loci (Puppo et al., 2016; Curto et al., sub.) thus using only a small portion of the genome for phylogenetic inference. Molecular phylogenetic analyses using a low amount of loci are more likely to be affected by incomplete lineage sorting and hybridization (Rokas et al., 2003) resulting in patterns that do not reflect species divergence history. Lineage sorting signal is random, so when a genomic dataset is used (thousands of loci spread throughout the genome) its signal should be canceled out (Green et al., 2010; Durand et al., 2011). Hybridization is not random so its signal persists (Linder and Reiseberg 2004) and depending on how strong it is, it can still affect the overall phylogeny even when a high number of loci are used.

Increasing the number of loci alone may not be always the solution for ambiguous phylogenetic relationships. Philippe et al. (2011) stressed that in phylogenomic datasets erroneous phylogenetic signal such as the one originated from incorrect identification of orthologues, wrong alignments, and incorrect variant call can be more predominant. These cannot be overcame by increasing the number of loci but by improving the methodological approaches. In our case we used the PyRAD pipeline because it was developed with the purpose of processing RAD data for phylogenic analyses of closely

related species (Eaton 2014), avoiding many of these errors. This pipeline is able to deal with insertions and deletions more efficiently than other methods which allows the recovery of more trustworthy comparisons among species (Eaton 2014). By combining this approach with paired-end 300 bp reads we analyzed longer RAD loci (up to 592 bp) which contain more phylogenetic information than using a single SNP per restriction site.

Micromeria varia evolution

As in all previous phylogenies, our results show *Micromeria* is divided into two lineages, East and West. Also in agreement with previous studies (Puppo et al., 2015a) species previously classified as *M. varia* from the eastern lineage show a common origin. With exception of one individual from M. varia s. s. (Tenerife) both species previously classified as *M. varia* (*M. pedro-luisii* from La Gomera and M. varia s. s. from Tenerife) group together. It is the first time this pattern is recovered. It was not possible to include a comprehensive sampling for the western lineage so it is not possible to draw further conclusions on those species at this time.

Also, and in contrast to what was previously described, *M. varia* individuals from the two main Tenerife populations (Anaga and Teno) create a monophyletic group. Both microsatellites and nuclear sequencing data showed two independent lineages for these regions (Puppo et al., 2014, 2016). In these previous studies the authors suggested this was a consequence of introgression with other species present in the island. Since we are using a multilocus dataset, the patterns obtained here are a consequence of the sum of the phylogenetic signal from each individual locus. The comparison of the microsatellite and RAD dataset indicate therefore, that even if many loci can show a deviating pattern probably because of gene-flow with other species (Puppo et al., 2016), the majority support a common origin of both *M. varia* s.s. populations.

Niche conservatism between islands

Our results showed M. canariensis, endemic to Gran Canaria, as paraphyletic to *M. gomerensis* from La Gomera. This results from a lineage in *M. canariensis* consisting of the individuals from the Laurel forest remain at the northern part of Gran Canaria, which created an independent clade that is neighbor group to *M. gomerensis*. This result suggests that *M. canariensis* from Gran Canaria contributed for the colonization of La Gomera resulting in the divergence of *M. gomerensis* (Puppo et al., 2015a). However, the role of the laurel forest populations in this event was not yet explored. Based on our results three possible scenarios can be outlined. First, *M. gomerensis* originated from *M. canariensis* from the laurel forest that colonized La Gomera. Second, the close relatedness of *M. gomerensis* and *M. canariensis* from the laurel forest is a consequence

of high introgression between them. Finally, this introgression may have an adaptive value and it introduced characters that contributed for the adaptation of *M. gomerensis* to the laurel forest.

The scenario of a unique origin of M. gomerensis from Gran Canaria laurel forest individuals would be congruent with the existence of strong ecological filters during the colonization of an island. The laurel forest is very common in La Gomera and most of *M. gomerensis* is distributed in this habitat. So, the close relatedness of these taxa may be related with adaptations to the laurel forest conditions resulting in niche conservation between two related species from different islands. A similar pattern has been observed for other plant taxa from the Canary Islands (Francisco-Ortega et al., 1996; Francisco-Ortega et al., 2000; Barber et al., 2000). This would imply that *M. gomerensis* only differentiated after *M. canariensis* became adapted to the laurel forest, which would make this lineage one of the youngest in the archipelago. In addition this implies that adaptation to laurel forest conditions took place in Gran Canaria where *M. canariensis* grows in a wide range of habitats from semi-arid areas in the south to very humid areas in the north. In congruence with this hypothesis is the observation that in all the phylogenies obtained so far the species of *M. varia* s. l. are relatively recent lineages (Puppo et al., 2014, 2015a).

It is possible that introgression between Gran Canaria and La Gomera would have contributed to the intermediate position of the laurel forest population relatively to these two islands. As mentioned above, in a parallel study we found that multiple colonization events contributed to the establishment of taxa connected by gene-flow, being this observed even between islands (Curto et al., sub.). This would create an accumulative effect where a large part of the *M. gomerensis* genome would contain alleles of Gran Canaria origin. Since we are using a multi-locus approach the phylogeny represents the average signal across the genome and if introgression is high enough this signal may be more prominent than the one originated from cladogenesis events. This idea uses the same premises from Philippe et al. (2011) where despite using whole genome data contradictory signal originated from artifacts can still affect the final outcome. The same would be expected for introgression.

Gene-flow from Gran Canaria may have contributed to the inclusion of genes that facilitated the adaptation of M. gomerensis populations into the laurel forest. This same mechanism may be behind ecological shifts or expansion of other Micromeria species. In our focus group it can explain the distribution of *M. canariensis*, which undergone an ecological expansion either from arid to humid parts areas or vice-versa. In this case introgression with other taxa from Gran Canaria may have promoted this process. Alternatively, it is possible that introgression did not have an adaptive nature and

introgression is only allowed among similar ecological zones. The transmission of advantageous traits through introgression has already been described in several taxa (Lexer et al., 2003; Brothers et al., 2013; Arnold 2006, 2015). In oceanic islands hybridization was already suggested to be an important process in increasing taxa adaptive ability (Saunders and Gibson 2005; Caujapé-Castells 2011). In the case of *Micromeria*, a parallel study showed that hybridization and multiple colonization events contributed to the maintenance of a high genetic diversity during *Micromeria* inter-island colonization (Curto et al., sub.). The transmission of adaptive traits may be another consequence of this process. Further testing estimating the degree of introgression between species is underway and will be necessary to confirm this hypothesis.

Biases related to the use of RAD-sequencing for phylogenetic inference

Different molecular marker systems have been used to study evolutionary processes in *Micromeria* (Curto et al., 2012; Puppo et al., 2015b), and applied, both archipelago-wide (Puppo et al., 2015a; Curto et al., sub.) and island-specific (Puppo et al., 2014; 2016). This allows to compare the RAD method with other methods. Each method has its characteristics that result in different degrees of suitability to phylogenetic or population genetics analyses. Unlike the classical methods RAD characteristics are still not fully understood (Hodel et al., 2016). This is especially related with: how recent the method is, the requirement of bioinformatics skills to analyses this type of data, and the comparable high amount of missing data in a RAD dataset compared to other methods.

Our results showed that the exclusion of loci with high missing data results in the loss of phylogenetic information. By using a more stringent approach we are selecting for loci that are more similar among all lineages and thus more conserved. Indirectly, this will also result in the selection of regions that are more likely to contribute for introgression. Hence the difference between both analyses. This was also the explanation given to justify the lower genetic differentiation detected among species of *Ancistrocladus* and *Micromeria* when microsatellites were used (Turini et al., 2014; Puppo et al., 2016).

An uneven distribution of loci across individuals can create biases when they are excluded according to missing data. Huang and Knowles (2014) showed that, in divergent lineages, mutations in restriction enzyme recognition site result in a non-random representation of loci across lineages. Which, by excluding loci according to missing data, results in loci only present in one or the other lineage. In our case, we have more individuals belonging to the eastern lineage resulting in higher probability of excluding loci only present in the western lineage. Simulation analyses showed that high

missing data has little effect in phylogenetic accuracy. (Rubin et al., 2012; Huang and Knowles 2014). Rubin et al. (2012), showed that an uneven distribution of informative characters across taxa is more likely to result in a misleading phylogeny than missing data per locus. So an approach as implemented in this study where data is filtered according to missing data per sample rather per loci is more likely to result in better resolved and accurate phylogenetic analyses.

Conclusions and outlook

In the present study we demonstrated that our RAD method is an additional option for phylogenetic analyses using genomic data. RAD-sequencing allows the screening of both coding and non-coding regions throughout the genome recovering a more comprehensive phylogenetic signal. However, there are some biases associated with this method by excluding data according to missing data thus complementary approaches can be used. Amplicon sequencing using NGS platforms is widely used for barcoding and metagenomics projects (i.e. Fierer et al., 2012; Shokralla et al., 2015; Mata et al., 2016) but it can also be applied for phylogenetic analyses. An example of this was the sequencing of the complete mitochondria for eastern Africa white eye birds (*Zostesterops*; Meimberg et al., 2016), where a highly resolved phylogenetic signal was obtained.

By evaluating the degree and nature of introgression among and within islands we could test all the hypothesis outlined above in order to find if local adaptation of *M. varia* to different island environments was influenced by introgression from other taxa. More specifically in La Gomera we will know if the adaptation to the laurel forest is a consequence of introgression or if this taxon diverged uniquely from laurel forest populations from Gran Canaria. This will be possible by partitioning the phylogenic signal of the genome into regions showing different phylogenetic patterns and defining which can be involved in introgression or being a consequence of selection. For this purpose, a whole genome resequencing approach is more appropriate because we can define linkage groups behaving differently and then infer which genes may be responsible for this pattern.

To further investigate our hypothesis, it will be necessary to perform an additional phylogeographic study using a comprehensive sampling from *M. canariensis*. We found a division in *M. canariensis* in the microsatellite data which is concordant with substrate age (Curto et al., sub.). The laurel forest population was assigned to the younger substrate. Since we did not include a non-laurel forest population from the younger substrates in this work, we cannot conclude that this division has an ecological meaning.

Acknowledgements

The authors would like to thank Sarah Sturm for technical assistance and Andreas Brachmann for guidance during the library preparation and sequencing process. We also thank Saurabh Dilip and Peter Sykacek for support regarding data analyses. This research was supported by the Fundação para a Ciência e Tecnologia (FCT) with two research grants to H.M.: (PTDC/BIA-BEC/108866/2008) co-financed by the Fundo Europeu de Desenvolvimento Regional (FEDER) through COMPETE – Programa Operacional Fatores de Competitividade (POFC) (FCOMP-01-0124-FEDER-008988); and (PTDC/BIA-BEC/098596/2008) co-financed by the Fundo Europeu de Desenvolvimento Regional (FEDER) through COMPETE – Programa Operacional Fatores de Competitividade (POFC) (FCOMP-01-0124-FEDER-008900). FCT also contributed by financing M.C. and P.P. PhD scholarships (SFRH/BD/79010/2011 and SFRH/BD/74747/2010, respectively)

References

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

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Publ. Gr. 17, 81-92.

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

Arnold, M. L., 2015. Divergence with genetic exchange. Oxford University Press.

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

Barber, J. C., Ortega, J. F., Santos-Guerra, A., Marrero, A., Jansen, R. K. 2000. Evolution of endemic *Sideritis* (Lamiaceae) in Macaronesia: insights from a chloroplast DNA restriction site analysis. Syst. Botany 25, 633-647.

Bräuchler, C., Ryding, O., & Heubl, G., 2008. The genus Micromeria (Lamiaceae), a synoptical update. Willdenowia 363-410.

Brothers, A. N., Barb, J. G., Ballerini, E. S., Drury, D. W., Knapp, S. J., & Arnold, M. L., 2013. Genetic architecture of floral traits in Iris hexagona and Iris fulva. J. Hered. est059.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST + : architecture and applications BMC bioinformatics, 10, 421.

Caujapé-Castells, J., 2011. Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. In: The biology of island floras, 284-324.

Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., Udall, J., 2012. Targeted enrichment strategies for next-generation plant biology. Am. J. Bot. 99, 291-311.

Curto, M.A., Puppo, P., Ferreira, D., Nogueira, M., Meimberg, H., 2012. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae) Mol. Phylogenet. Evol. 63, 758–767.

Curto, M., Puppo, P., & Meimberg, H., sub. Genetic diversity and differentiation patterns in oceanic islands are congruent with multiple colonization dynamics and the establishment of species syngameons. Submitted to BMC Evolutionary Biology.

Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for Ancient Admixture between Closely Related Populations. Mol. Biol. Evol. 28, 2239–2252.

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844-1849.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6, e19379.

Fernández-Palacios, J. M., de Nascimento, L., Otto, R., Delgado, J. D., García-del-Rey, E., Arévalo, J. R., Whittaker, R. J., 2011. A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. J. Biogeogr. 38, 226-246.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., et. al., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc. Natl. Acad. Sci. U.S.A. 109, 21390-21395.

Francisco-Ortega, J., Jansen, R. K., & Santos-Guerra, A. 1996. Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. Proc. Natl. Acad. Sci. U.S.A. 93, 4085-4090.

Francisco-Ortega, J., Santos-Guerra, A., Kim, S. C., Crawford, D. J. 2000. Plant genetic diversity in the Canary Islands: a conservation perspective. Am. J. Bot. 87, 909-919.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et. al., 2010. A draft sequence of the Neandertal genome. Science 328, 710-722.

Hodel, R. G., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., et al., 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. Appl. Plant Sci. 4, 1600025 .

Huang, H., & Knowles, L. L., 2014. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Syst. Biol. 65, 353-356.

Juan, I., Emerson, B., Orom, I., Hewitt, G., 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends Ecol. Evol. 15, 104–109.

Kim, S. C., Crawford, D. J., Francisco-Ortega, J., & Santos-Guerra, A. (1996). A common origin for woody *Sonchus* and five related genera in the Macaronesian islands: molecular evidence for extensive radiation. Proc. Natl. Acad. Sci. U.S.A, 93, 7743-7748.

Lexer, C., Randell, R. A., & Rieseberg, L. H., 2003. Experimental hybridization as a tool for studying selection in the wild. Ecology 84, 1688-1699.

Linder, C. R., & Rieseberg, L. H., 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Bot. 91, 1700-1708.

Martin, M., 2011) Cutadapt removes adapter sequences from high-throughput sequencing EMBnet. journal 17, 10-12

Mata, V. A., Amorim, F., Corley, M. F., McCracken, G. F., Rebelo, H., & Beja, P., 2016. Female dietary bias towards large migratory moths in the European free-tailed bat (Tadarida teniotis). Biol. Lett.12, 20150988.

Meimberg, H., Abele, T., Bräuchler, C., Mckay, J.K., Pérez, P.L., Paz, D., Heubl, G., 2006. Molecular evidence for adaptive radiation of Micromeria Benth. (Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data. Mol. Phylogenet. Evol. 41, 566–578.

Meimberg, H., Schachtler, C., Curto, M., Husemann, M., & Habel, J. C., 2016. A new amplicon based approach of whole mitogenome sequencing for phylogenetic and phylogeographic analysis: An example of East African white-eyes (Aves, Zosteropidae). Mol. Phylogenet. Evol. 102, 74-85.

Pérez de Paz, P. L., 1978. Revisión del género Micromeria Bentham (Lamiacea-Stachyoideae) en la Región Macaronésica. Instituto de Estudios Canarios, Monografías 16, 1-306.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double Digest RADseq : An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PloS One 7, e37135.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol 9, e1000602.

Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J., 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PloS One 7, e32253.

Puppo, P., Curto, M., Velo-Antón, G., Pérez de Paz, P.L., Meimberg, H. 2014. The influence of geological history on diversification in insular species: genetic and morphological patterns of Micromeria Benth. (Lamiaceae) in Tenerife (Canary archipelago) J. Biogeogr. 41, 1871–1882.

Puppo, P., & Meimberg, H., 2015a. New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa, 230, 1-21.

Puppo, P., & Meimberg, H., 2015b. Corrections to Phytotaxa 230: New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa 239, 197-198.

Puppo, P., Curto, M., Gusmão-Guedes, J., Cochofel, J., Pérez de Paz, PL., Bräuchler, C., Meimberg, H., 2015a. Molecular phylogenetics of Micromeria (Lamiaceae) in the Canary Islands , diversification and inter-island colonization patterns inferred from nuclear genes q. Mol. Phylogenet. Evol. 89, 160–170.

Puppo, P., Curto, M., Meimberg, H., 2015b. Development and characterization of 16 microsatellite markers for Micromeria (Lamiaceae) from Tenerife (Canary Islands, Spain) using 454 sequencing. Conserv. Genet. Resour. 7, 743–749.

Puppo, P., Curto, M., Meimberg, H., 2016. Genetic structure of Micromeria (Lamiaceae) in Tenerife, the imprint of geological history and hybridization on within-island diversification. Ecol. Evol. 6, 3443-3460.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rubin, B.E., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS One, 7, e33394.

Saunders, N. E., Gibson, D. J. 2005. Breeding system, branching processes, hybrid swarm theory, and the humped-back diversity relationship as additional explanations for apparent monophyly in the Macaronesian island flora. J. Ecol. 93, 649-652.

Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., et al., 2015. Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Sci. Rep. 5, 9687.

Turini, F. G., Steinert, C., Heubl, G., Bringmann, G., Lombe, B. K., Mudogo, V., Meimberg, H. 2014. Microsatellites facilitate species delimitation in Congolese

Ancistrocladus (Ancistrocladaceae), a genus with pharmacologically potent naphthylisoquinoline alkaloids. Taxon 63, 329-341.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614-620.

.

Chapter 5: General discussion

5.1 New insights into evolution on oceanic islands

Until recently, limitations associated with classical molecular methods only allowed the genotyping of a few loci, thereby preventing the detection of reticulate evolution patterns (Herben 2005). This influenced the way evolutionary processes were viewed, especially concerning the role of hybridization. As a consequence, the colonization process on oceanic islands was described as a stepping stone mechanism with single introduction events on each island (Sanmartin 2008). These observations biased the conclusions made by the authors, exemplified by the niche preemption hypothesis proposed by Silvertown et al. (2004, 2005). This hypothesis explains the prevalence of single colonization events on islands close to the mainland. It posits that island niches are filled relatively fast during the colonization process, which prevents secondary colonizers from establishing through competition. Multi-loci genotyping approaches showed that reticulated patterns exist (i.e. Kim 2008, Barber et al. 2007) and these patterns suggest that hybridization may be common, which contradicts the niche preemption hypothesis. Archipelago-wide genetic diversity studies showed a low impact of the founder effect, providing some further evidence for the existence of multiple colonization events (Francisco-Ortega et al. 2000; Caujapé-Castells 2011; Pérez de Paz and Caujapé-Castells 2013). The work gathered in this thesis provides additional examples of reticular patterns, thereby supporting the idea that islands are colonized multiple times and that gene-exchange between species is rather common.

Evidence of evolution with gene-flow and its impact on genetic variation

This thesis presents several studies focusing on *Micromeria* evolution, which show evidence of gene-flow between species. These studies employed different sampling schemes and marker systems but showed similar results, which provides extra confidence in this assumption.

In chapter 2, Curto et al. (2012) developed 19 EPIC markers and applied them in eight *Micromeria* individuals. This was done by testing their ability in recovering resolved phylogenetic trees and their concordance with the expected phylogenetic hypothesis through topology comparison. Incongruence was found in both tests, thus supporting a reticulate pattern. Using eight of these markers to assess the phylogenetic relationships of *Micromeria* on *Tenerife* and the entire archipelago, Puppo et al. (2014, 2015) recovered some degree of incongruence between these loci as well. These three studies provided the first indications that introgression may exist among *Micromeria* species. Similar reticulated patterns were described when chloroplast and ITS loci were

sequenced to recover phylogenetic relationships of Canarian flora (i.e. Gruenstaeudl et al. 2012; Jones et al. 2014). An early study from Barber et al. (2007) provides evidence of multiple colonization events and hybridization in Canarian *Sideritis*. Discordance between gene trees can be caused by lineage sorting as well (Degnan and Rosenberg 2009), and due to their low number loci these approaches are not able to differentiate between lineage sorting and hybridization. With these shortcomings in mind, alternative approaches able to detect gene-flow among species directly were investigated.

Microsatellites are mostly neutral (Selkoe and Toonen 2006), and thus more permeable to genetic exchange between hybridizing species under differential selective conditions (Teeter et al. 2010). For this reason detecting gene-flow between species should be easier using these markers. Microsatellite markers had already been developed for *Micromeria* (Puppo et al. 2015b) and were applied with the objective of detecting evidences of gene-flow and hybridization between *Micromeria* species. In chapter 3, significant historical and recent migration rates among most species in Tenerife were determined, and it was found that the most widespread species were the largest contributors for genetic exchange (Puppo et al. 2016). Hybridization was detected for most species pairs showing morphological hybrids. When the entire archipelago was considered, data was able to recover some migration rate between islands (Chapter 3: Curto et al. sub. a). Additionally, hybrids, which contribute to gene flow, were detected between most species. As described above in the case of Tenerife, widespread species in Gran Canaria contributed to most of the gene-flow between species. These publications provide direct evidence of gene-flow among Canarian species using microsatellites and show for the first time introgression between two islands. Previous studies on other plant taxa from the Canaries only detected introgression between a few species pairs and did not consider the entire archipelago (González-Pérez and Sosa 2009).

Although hybridization and gene-flow exists, they do not necessarily have an impact on genetic variation. According to both the hybrid swarm theory from Seehausen (2004) and the surfing syngameon hypothesis from Caujapé-Castells (2011), introgression contributes to an increase in genetic diversity, which will boost populations' adaptive ability. The microsatellite data support this assumption. Both on Tenerife and Gran Canaria the species contributing the most genetic-exchange are also the most genetically diverse. Incidentally they are also the most widespread species, which can be an indication of higher adaptive ability. These factors might not be independent, so this needs more testing. One approach would be to evaluate which genes are more prone to exchange during the colonization process, and whether they are in regions with higher genetic diversity and if their function is related with species adaptation, which is

not possible with only microsatellites. By doing so, it would be possible to test the impact of introgression on the diversity of regions associated with species fitness.

So far allozymes have been used to assess genetic diversity in multiple species on archipelagoes (Francisco-Ortega et al. 2000; Pérez de Paz and Caujapé-Castells 2013). Despite being able to associate these patterns with potential multiple colonization events, this data was not used to estimate gene-flow directly. In the present thesis a similar approach was employed using neutral variation. By combining evidence of gene-flow together with genetic diversity results, it was possible to extend the surfing syngameon hypothesis to the colonization process within the archipelago (Chapter 3: Curto et al. sub. a). Data suggests that multiple colonization events do not only occur during the colonization of the archipelago but also during the occupation of each individual island (Fig. 1). Consequently, the resulting species syngameon expands in the direction of the colonization front instead of remaining static. Because some taxa become more reproductively isolated, they are excluded from the syngameon, which causes the syngameon to vary in composition as well. I expect that with more studies directly testing the surfing syngameon hypothesis new features of this mechanism will be described. As yet, only Riley et al. (2016) tested this hypothesis outside of the Canary Islands. In their work the authors did not find evidence of gene-flow and multiple colonization on the Californian Channel Islands. These islands are continental fragments, so this process may be promoted by the creation of new land dynamics characteristic to oceanic islands.

Island geological dynamics has an impact on hybridization and diversification patterns

Phylogeographic and phylogenetic studies on islands show high congruence between species evolution and island geological history (i.e. Gómez et al. 2003; Moya et al. 2004), and *Micromeria* was no exception (Puppo et al. 2014, 2015). Phylogenetic analyses for species both on a single island and between multiple islands supported a colonization direction from older to younger islands (Puppo et al. 2014, 2015). Therefore the way new land emerges through volcanic activity is one of the major factors shaping *Micromeria* diversification. Other geological events such as mega-landslides seem to have influenced *Micromeria* genetic structure as well (Chapter 3: Puppo et al. 2016). In this section I will discuss how some mechanisms of island morphogenesis influenced *Micromeria* evolution.

Results indicate that different forces shaped the diversification process in Tenerife and Gran Canaria. In Tenerife genetic differentiation was congruent with island age while in Gran Canaria this was not the case (Puppo et al. 2014, 2015; Chapter 3: Puppo et al. 2016; Curto et al. sub. a). This might be related to the islands' different

morphogenesis processes. When the central part of Tenerife emerged it was not connected to the remaining palaeo-islands, so *Micromeria* arrived there through colonization. *Micromeria* encountered new niches during colonization, thus triggering speciation. In Gran Canaria land was always subsequently added to the pre-exiting island so populations expanded to the new area. Species expansion occurred alongside an expansion of established niches, so speciation was less likely to occur.

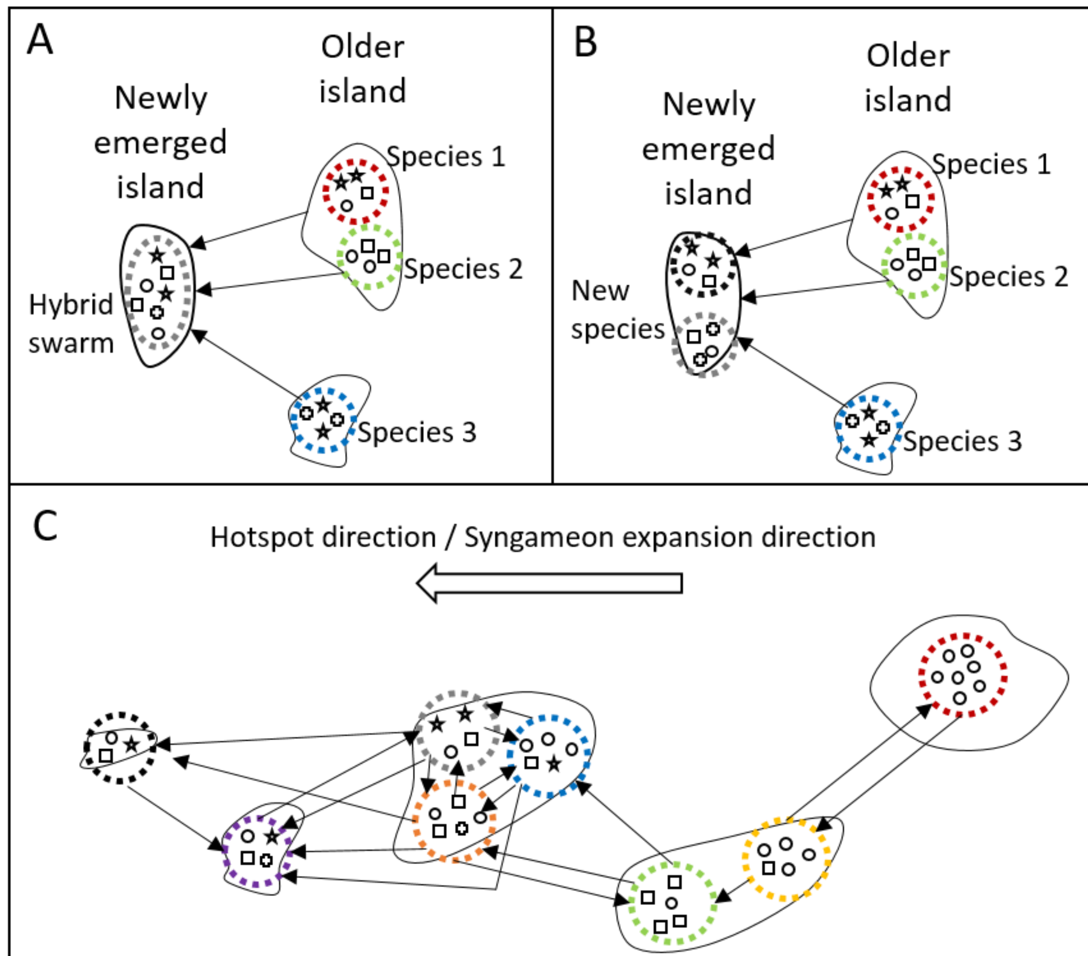


Fig. 1. Schematic summary of how the surfing syngameon hypothesis can be applied to single-island colonization based on the data presented in this thesis. Unless indicated otherwise, dashed circles represent species. Symbols represent alleles for a specific locus. A: Multiple colonization on a newly emerged island resulting in the creation of a hybrid swarm. Arrows represent colonization events. B: The hybrid swarm differentiates into two species through an ecological speciation process. Alleles are sorted according to local selective pressures. C: Connectivity through gene-flow of the species syngameon resulted from the multiple colonization process. Species occurring on younger islands are more connected than the ones occurring in older ones. This process results in an expansion of the syngameon in the same direction of the addition of new land.

So it seems that the appearance of new land was a major driver of diversification on oceanic islands (Fig. 1). In chapter 3, taxa occupying younger islands are genetically more diverse and less differentiated, which is congruent with multiple colonization of each island and subsequent establishment of hybrid swarms (Curto et al. sub. a; Fig.

1A). These populations would have a higher adaptive ability and would be more likely to result in radiation events. Because populations in younger islands remain connected by gene-flow when they differentiate into multiple species, they result in the expansion of the species syngameon (Fig. 1B, C). Gene-flow does not affect the diversity of populations in older regions, making them more isolated from the syngameon. Overall this results in a shift, similar to a species range shift process, in both content and size of the species syngameon. In this case the new islands work as the satellite populations in the expansion front and are highly connected to the syngameon, while the older islands correspond to the rear populations and are less connected to the syngameon (Fig. 1C).

The appearance of new detached land may trigger the process described in the previous paragraph. The emergence of new land provides a high amount of free niches, contributing to a decrease in intra-specific competition and making the selection against hybrids less severe. This may be the reason why hybridization on islands is so frequent (Francisco-Ortega et al. 1996). Following this reasoning, increased likelihood of hybrid swarms on new islands also results in increased likelihood of adaptive radiation events. Similar mechanisms were responsible for the adaptive radiation in other island-like systems. For example, the radiation that gave origin to around 800 cichlid species in Malawi lake was shown to originate from a hybrid swarm created by the admixture of riverine and paleo lakes species (Seehausen et al. 2003; Joyce et al. 2005; Schwarzer et al. 2012; Nichols et al. 2015). In this case the creation of a lake followed by colonization from other lakes and rivers is an equivalent process to the emergence of an island and its colonization from mainland or other islands.

Erosion and the occurrence of land-slides can also have a role in shaping genetic diversity in oceanic islands. Using microsatellites in populations spread throughout the archipelago it was possible to genotype using denser and more evenly distributed sampling, and thus resulting in the detection of new structure patterns. For example, in Tenerife, intra-specific structure patterns were found (Chapter 3: Puppo et al. 2016): *M. hyssopifolia* populations from Adeje were separated from the remaining island, and *M. varia* was divided into two groups (Anaga and Teno). Although the *M. hyssopifolia* example resulted as a consequence of island age, the separation of the two *M. varia* populations was probably caused by multiple landslides that occurred during Tenerife's morphogenesis, sectioning off populations from one another (Ancochea et al. 1990; Watts and Masson 1995; Juan et al. 2000). These geological events seem to have contributed toward population structures of both plant and animal species (Gübitz et al. 2000; Brown et al. 2006; Dlugosch and Parker 2007; Mairal et al. 2015). In Gran Canaria genetic structure and differentiation according to island geology (subtract age) was only found among *M. canariensis* populations (Chapter 3: Curto et al. sub. a). Other authors

found similar patterns and suggested that this was a consequence of restrictions to gene-flow caused by lava flows during volcanic events occurring around 2.8–3.5 Ma in the northern part of the island (Pestano and Brown 1999; Contreras-Díaz et al. 2003).

Species inhabiting older areas are less diverse and show higher pairwise population differentiation, which is a consequence of the decreased effect of gene-flow relative to the increased population size on each island (Chapter 3: Curto et al. sub. a). This ultimately results in the reproductive isolation of these species and their exclusion from the syngameon. Considering the general dynamic theory island biogeography (GDM) from Whittaker et al. (2008), erosion contributes to a decrease in island carrying capacity, thus increasing the extinction rate. This same process promotes the exclusion of species from the syngameon and their loss of genetic diversity. As shown above, catastrophic events create barriers to gene-flow and habitat loss, both of which contribute to the extinction of populations, and thus their loss of genetic variability. Moreover, with the decrease in an island's carrying capacity, colonizers are less likely to establish successfully, which promotes reproductive isolation relative to populations found on other islands.

In summary, there is a complex set of factors shaping the evolution of *Micromeria* in the Canary Islands. The increase of speciation rate and species syngameon size are positively related with the appearance of new islands and their increase in topological complexity. In the other hand, the detachment of species from the syngameon and consequent accumulation of genetic differences is promoted by erosion and catastrophic events such as mega-landslides.

5.2. Effect of syngameon dynamics on species definition and morphological traits

Maintenance of species boundaries despite gene-exchange

If strong selective forces that set certain phenotypes do not exist, populations belonging to hybrid swarms are not able to differentiate (Seehausen 2004, Arnold 2006, 2015). As described in the previous section, selection may be the first trigger of differentiation but with time geological events contributing for reproductive isolation can further promote this process. *Micromeria* in the Canary Islands is a great system for studying the factors that influence morphological differentiation: there are islands with more than one species, allowing the correlation of morphological traits and their diversification patterns with an island's ecological features; and it is composed of species

that diverged at different times (Puppo et al. 2014), providing examples of different stages of the differentiation process.

Micromeria species distribution is mostly concordant with island ecological zones, indicating that selective forces played a role in the diversification of these species. In fact, Puppo et al. (2016) found evidence of hybrid zones in Tenerife, which is a strong indicator that selection plays a role in *Micromeria* morphological differentiation (Chapter 3). In hybrid zone dynamics, strong stabilizing selection contributes to the fixation of alleles responsible for local adaptation (Nosil and Feder 2012). If the divergence of these species resulted from an adaptive radiation event, then the genomic regions under selection should be responsible for maintaining species morphological integrity despite hybridization (Feder et al. 2012).

Other mechanisms besides selection may have contributed to the morphological diversity observed in *Micromeria*. As mentioned above, older lineages tend to be excluded from the species syngameon and become more reproductive isolated, which may lead to morphological divergence. In Puppo et al. (2014), *Micromeria* species restricted to older parts of the island were found to be older as well as more genetically and morphologically differentiated. In chapter 3, these species did not contributed as much to inter-specific gene-flow as the other species (Puppo et al. 2016). A similar pattern was found in Gran Canaria where the most morphologically differentiated species (*M. leucantha* and *M. pineolens*) were the ones showing less connectivity (Chapter 3: Curto et al. sub. a). All of these examples serve as evidence that reproductive isolation also contributes to morphological differentiation. However, this is contradicted by the existence of species that are morphologically similar but genetically divergent. This is the case for the *M. varia* s.l. species complex. The mechanism involved in the maintenance of this phenotype despite the high divergence observed in some lineages sharing this morphology is discussed in the next section.

The appearance and maintenance of morphologically similar species with independent origins

The fourth chapter of this thesis focuses on the phylogenetic relationships between the species previously classified as *M. varia* s.l. (Curto et al. sub. b). These species are morphologically similar and were divided into six independent taxa based on molecular data (Puppo and Meimberg 2015a, 2015b). In the study presented in chapter 4, we hypothesized that adaptations may have been transferred between two of these species from different islands through introgression. This was based on the close phylogenetic relationship between *M. gomerensis* from La Gomera and a population of *M. canariensis* from Gran Canaria that grows in the laurel forest. Although we were not

able to show that introgression was the main factor contributing for this pattern the high gene-flow found between these species, with microsatellite data, make us more confident in this hypothesis. If this is the case, these same mechanisms may play an important role in the evolution of the remaining *M. varia* s.l. species.

In previous work using eight nuclear EPIC sequencing markers *M. varia* s.l. had several different origins (Puppo et al. 2015): one in the eastern lineage and two in the western lineage (*M. varia* from Tenerife and *M. pedro-luisii* from La Gomera belong to different lineages). When the microsatellites were considered the same pattern was found. Since a high degree of gene-flow among species from the same island was found, introgression may be hiding the phylogenetic signal of a common origin. This effect would be more pronounced when just a few loci were used. To overcome this limitation we used a phylogenomic approach. The results did not contradict previous phylogenetic analyses and confirmed that *M. varia* evolved several times. The most likely explanation for this observation is convergent evolution, and in this section I discuss the mechanism that might be behind this process.

There are three genetic processes contributing to convergent evolution (Stern 2013): mutations of independent genomic regions resulting in the same phenotype; the maintenance of an ancestral allele; and the introduction of common alleles through introgression. The patterns gathered in this thesis suggest that a combination of the last two processes might be responsible for the convergent evolution of *M. varia* s.l. morphology.

Introgression was detected in all datasets, so its possible influence in the evolution of *M. varia* s.l. phenotype must not be ignored. Additionally, the mechanisms behind this apparent convergent evolution must be inferred while taking into account the multiple colonization dynamics described in chapter 3 (Curto et al. sub. a). A consequence of this process would be the spread of similar genetic content on all islands, allowing for species to show similar adaptations and thus similar phenotypic traits. This might be the reason why the *M. varia* s.l. phenotype could have appeared several times during *Micromeria* evolution. The alleles responsible for this morphology would have come to each island during the radiation process and were integrated into the gene-pool of the hybrid swarm (Fig. 2). Through selection these would have been fixed in the populations that gave origin to the *M. varia* s.l. species on each island. The transmission of traits showing some adaptive value through introgression has been described for several systems (Lexer et al. 2003; Brothers et al. 2013; Arnold 2006, 2015).

The process described above is congruent with a single origin of the *M. varia* s.l. phenotype. In this case a common origin would only be detected in the genes responsible

for this morphology, while the remaining parts of the ancestral *M. varia* s.l. genome would have been replaced with alleles from other species through introgression. If this were the case, this phenotype would have appeared early in the divergence process since it is shared among highly divergent lineages. A single origin of the *M. varia* s.l. phenotype would be confirmed by the detection of genomic regions common to all individuals showing this phenotype.

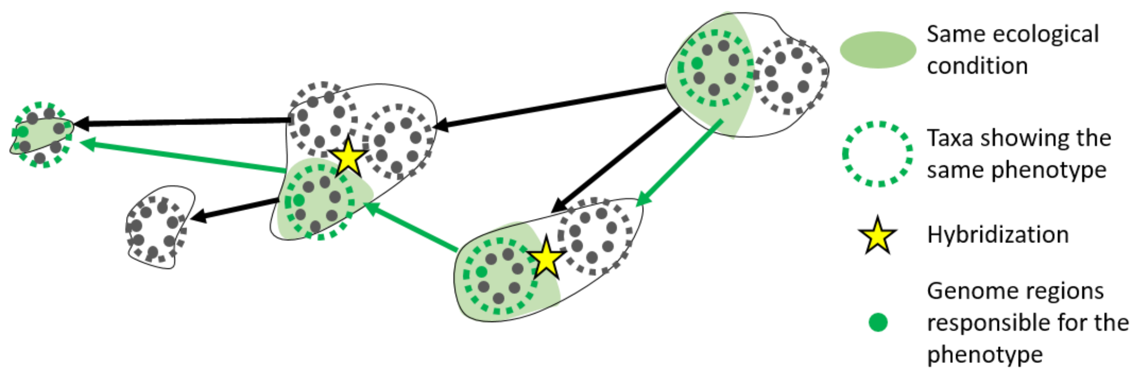


Fig. 2. Representation of how a single genetic origin of a phenotype is congruent with a convergent evolution scenario due to the multiple colonization dynamics. Dashed circles represent different species. Green dashed circles show the convergent phenotype (ex. *M. varia* morphology). Arrows represent colonization events and green arrows correspond to the establishment of a colonizer containing target phenotype.

The fact that there are other *Micromeria* species groups showing patterns of convergent evolution indicate that this process may occur frequently. *Micromeria lasiophylla* from Tenerife and *M. herpyllomorpha* from La Palma show similar morphologies and have independent origins. In fact, the population from *M. herpyllomorpha* found in high altitudes was once classified as *M. lasiophylla* (Fig. 3). In both the phylogeny from Puppo et al. (2015) and the microsatellite data (Chapter 3: Curto et al. sub. a) *M. herpyllomorpha* is more closely related to *M. densiflora* than *M. lasiophylla*. Both *M. herpyllomorpha* high-altitude populations and *M. lasiophylla* occupy similar environments on different islands, indicating that they have the same adaptations. Therefore it is possible that some colonizers from the *M. lasiophylla* ancestry contributed to the colonization of La Palma, and that the regions related to high-altitude adaptations fostered/supported the occupation of the mountain regions.

Evidence of convergent evolution and the transmission of adaptive traits through hybridization have been described in *Heliconius* butterflies (wing color; Heliconius Genome Consortium 2012) and in the house mouse (Warfarin resistance; Song et al. 2011). Both studies used genome sequence information, which provides some confidence that, if a similar approach is used, the genetic basis of *M. varia* s.l. morphology can be revealed.

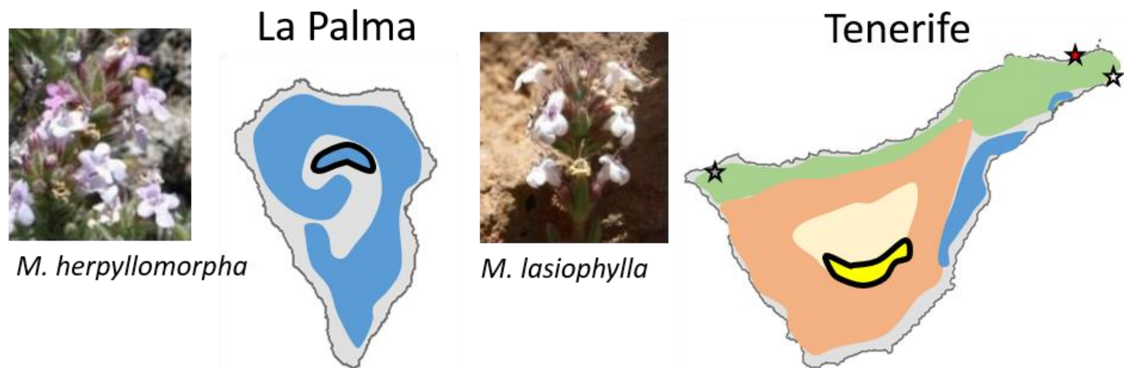


Fig. 3. Distribution and pictures of *M. herpyllomorpha* from La Palma (blue) *and M. lasiophylla* (yellow).from Tenerife. The regions surrounded by the black line correspond to the high-altitude populations. Species distributions are based in Perez-de-Paz (1976). Photos were taken by Pamela Puppo.

5.3. Some remarks about molecular markers

Comparison of information obtained among marker systems

By using three types of markers in this research it was possible to access *Micromeria* evolutionary patterns from different perspectives. Nevertheless, the main phylogeographic/phylogenetic patterns were concordant among all marker systems in the following ways: all showed a division between the eastern and western islands groups; all revealed that all species previously classified as *M. varia* from the eastern lineage belong together in the same clade/cluster; and all makers support the idea that Tenerife and Gran Canaria contributed to most of the diversification and colonization of *Micromeria* in the archipelago. Each of these patterns corresponded to evolutionary events occurring in the early stages of the *Micromeria* diversification process. In these stages, *Micromeria* genetic variation was probably more homogeneous and only a few lineages existed. As a consequence, incongruences among different genomic regions resulting from lineage sorting and introgression were less likely to occur. With this reasoning the incongruent patterns among marker systems should correspond to more recent events.

By comparing EPIC (Puppo et al. 2015) and microsatellite (Chapter 3: Curto et al. sub. a) markers a few incongruences were found. In the phylogenetic analyses constructed based on EPIC markers, *M. lepida* belongs to the eastern lineage. With microsatellites this pattern was not as clear because it is clustered together with either eastern or western lineages depending on the analyses method used. This was considered to be a consequence of gene-flow with members of the western group. Another incongruence was the inability of the microsatellites to differentiate between *M. mahanensis* (Lanzarote) and *M. gomerensis* (La Gomera), which was not the case with

EPIC markers. This is somewhat unexpected since microsatellites should be able to detect more detailed structure patterns (Ellegren 2004). The lack of structure between these two species was explained by characteristics inherent to this marker. EPIC markers should be less permeable to genetic exchange and thus the recovered pattern should be more concordant with species delimitations (Thomson et al. 2010). Microsatellites, on the other hand, due to their neutral nature are more likely to contribute to introgression (Selkoe and Toonen 2006). A high migration rate was found between *M. mahanensis* and *M. gomerensis*, which could be causing a lack of structure in the neutral regions. The marker selection process may have had an impact on this result. Because microsatellites are known for being highly variable (Ellegren 2004), they are less likely to contain flanking regions preserved among several species, therefore making them suitable for primer design (Barbará et al. 2007). By selecting markers that amplify in most of the archipelago, we are selecting for markers that are automatically more homogenous among species. In doing so we are also increasing the likelihood of selecting loci that are contributing to introgression among islands. As it was shown in both microsatellite development studies presented in chapter 2, this effect is already detected in the population level where the markers showing highest genetic structure were also the ones with higher missing data (Curto et al. 2013; 2015).

In the publication developing microsatellites for *Catha edulis* we found that markers containing "AT" rich repeats were more variable as thus could be serve as a good alternative for obtaining highly informative data with lower dropout. The microsatellite primers used on *Micromeria* were designed base on 454 libraries previously enriched with probes which are not very efficient in targeting these types of motifs.

With RAD-sequencing we are accessing around 100 times more loci than in the traditional methods, so more detailed relationships should be recovered. This was the case when a high amount of missing data was allowed; but when loci missing more than 50% of their data were excluded, much of the resolution was lost. This was mostly reflected in the relationships among the species previously classified as *M. varia* in the eastern clade. In Puppo et al. (EPIC marker phylogeny; 2015a), a monophyletic group containing these species had already been recovered, but a differentiation pattern among them was not clear. Rubin et al. (2012) and Huang and Knowles (2014), based on simulated RAD data, suggested that excluding loci according to missing data contributes to the loss of phylogenetic information. The same mechanism contributing to the lack of structure between some species in microsatellites can also be applied in RAD-sequencing. In the RAD method presented in this thesis, only fragments with two EcoRI cutting sites within a distance of 650 bp are sequenced. Not all restriction sites will be

shared among all lineages, creating a bias (Huang and Knowles 2014). By selecting loci according to missing data, only more conserved regions and regions highly permeable to introgression might be kept. Hence the lack of phylogenetic information.

These conclusions show that while making inferences, the intrinsic characteristics of the method must be considered. Moreover it shows the importance of studies which focus on their biases and caveats.

Is it still worthy to invest in classical methods?

With the appearance of second generation sequencing platforms, there was a shift in the technology used to access genetic information (van Dijk e al. 2014). With them it was possible to analyze genetic variation on a genomic level, and when combined with methods like RAD-sequencing this also became possible for non-model species (Davey et al. 2013). Nevertheless, classical methods are still widely used. For example, by searching the key words “microsatellite” and “RAD sequencing” in Web of Science for the year 2016, until present time, microsatellites had 5,022 entries compared to 30 from RAD. Now we face the question of whether these classical markers systems should be discarded completely and replaced with new methods to accessing genomic information.

As pointed out by Hodel et al. (2016), the decision to use microsatellites as opposed to genomic approaches such as RAD-sequencing hinges on a tradeoff between the investment that a project requires and the amount and type of information needed. For short-term projects with small budgets, the investment in a genomic dataset it is still not viable. This is particularly true when the researchers still need to acquire the bioinformatics skills necessary to analyze NGS data. Many projects require a high number of individuals, but not a high number of loci. In this case, the use of markers such as microsatellites is still advantageous. For example, according to calculations by Hodel et al. (2016), the use of RAD-sequencing in 96 individuals has a similar cost as the use of 12-15 microsatellites for the same sampling. However, RAD-sequencing becomes more expensive than microsatellites with the addition of more samples. Another advantage of using classical marker systems is the fact that they have been used for a long time and most of their caveats have already been described. Moreover, by combining them with genomic resources it is possible to have a better understanding of the patterns obtained from previous publications. For example, by combining microsatellites with transcriptomic data it will be possible to access whether markers were under selection and to which genes they are connected. Finally, we still do not know enough about genomic approaches and the biases associated with them to have complete trust in the data obtained. Using RAD-sequencing as an example, several problems have been found: inclusion of paralogs due to erroneous assemblies (Etter et

al. 2011; Xu et al. 2014); wrong genotypes as a result of sequencing and PCR errors (Arnold et al. 2013); biases related to library preparation (Arnold et al. 2013); overestimation of heterozygosity (Arnold et al. 2013; Gautier et al.2013); limitation related to DNA quality (Andrews et al. 2016); and non-random cleavage of the restriction enzyme (Arnold et al. 2013; Huang and Knowles 2014). Although many of the problems have been solved by improving bioinformatics analyses algorithms, the restriction-enzyme-associated biases have not. This ultimately results in a non-random distribution of missing data, which may have affected the RAD-sequencing dataset as presented in this thesis.

With time, a better understanding of these new methods and their limitations will be achieved and they will eventually replace the classic markers. In the meantime, some efforts have been made to genotype markers such as microsatellites with NGS platforms (De Barba et al. 2016; Vartia et al. 2016). In fact, amplicon sequencing using Illumina is commonly used both for DNA barcoding (Caporaso et al. 2012; Shokralla et al. 2015; Mata et al. 2016) and phylogenetic applications (Meimberg et al. 2016, Uribe-Convers et al. 2016). These approaches are good alternatives to more traditional methods, mostly because they are cheaper and less labor intensive.

5.4. Future directions

A few questions remain unanswered and, in the process of the work presented in this thesis, new ones were raised.

In the phylogenetic analyses from Puppo et al. (2014, 2015) only three species were used as outgroups: *M. graeca* from southern mainland Spain; *M. hochreutineri* from the high Atlas in Morocco; and *M. inodora* from the Baleares. Of these *M. inodora* was the closest to *Micromeria* from the Canaries. However, this does not mean that this is the most closely related *Micromeria* mainland species. Knowing the most closely related mainland species will allow a better understanding of the diversification process in the archipelago.

Several authors have suggested that colonization of the Canary Islands occurs through multiple colonization events (Herben 2005, Sauders and Gibson 2005, Caujape-Castels 2011). And several studies have supported this idea with empirical data (i.e. Percy and Cronk 2002; Galbany-Casals et al. 2009; García-Verdugo et al. 2015). A better understanding of *Micromeria's* colonization and adaptation process would be obtained by having the most closely related species for comparison. More specifically, it would be possible to detect if there is still some degree of connectivity with the mainland

populations. This would expand our knowledge about the syngameon hypothesis and help us determine which *Micromeria* adaptations are ancestral or derived, and what their genetic bases are.

The biggest advantage of knowing the most closely related mainland species is having a more reliable outgroup for the *Micromeria* phylogeny. The directionality of current phylogenetic relationships may not be correct. Ancestral lineages in Tenerife were defined because they were congruent with the island's geological history, and Gran Canaria was differentiated enough to serve as an outgroup (Puppo et al. 2014, 2015). However, this is not possible when the whole archipelago is considered. This is particularly true if the outgroups are too divergent to share enough ancestral states with insular lineages. Directionality will allow us to assess where colonization started, and thus have a better understanding of the *Micromeria* diversification process. We are currently in the process of genotyping chloroplast haplotypes for several *Micromeria* species from the mainland, which will allow new insights into this topic.

In this thesis, mostly neutral variation was assessed and it was not possible to make conclusions concerning the genetic bases of *Micromeria* phenotypes. When the RAD-sequencing approach was used, which includes both coding and non-coding regions, the data was not powerful enough to identify candidate genes responsible for *M. varia* morphology. Since *Micromeria* species are a result of an adaptive radiation event, the assessment of the genes responsible for morphological features will provide a better understanding of the adaptation process. This can be studied archipelago-wide by looking for regions that are unique in species showing the same phenotype, which can be particularly interesting in the *M. varia* and *M. lasiophylla* morphological species complexes. By accessing regions that are unique to these groups, it will be possible to assess the genetic basis of these phenotypes and consequently confirm whether they appear just once or several times during *Micromeria* evolution. If the same candidate genes are found for all taxa, it is possible to conclude that the morphology only appeared once and it spread across the islands during the multiple colonization process.

On single islands, the adaptation process can be studied by focusing on hybrid zone dynamics. In a hybrid zone resulting from the interaction of two species with different ecological requirements only neutral regions will be permeable to gene-exchange (Harrison and Larson 2014). The regions that are different between both species should be under strong purifying selection, and thus are involved in the maintenance of species integrity (Larson et al. 2014). This is usually done by looking for F_{ST} outliers or genomic clines. The first approach considers that regions associated with genes under stabilizing selection should be highly differentiated, and thus show an F_{ST} significantly higher than the genome average (Gosset and Bierne 2013). In research

such as Nosil et al. (2012), this method was deemed prone to false positives since drift could create similar patterns. If a hybrid zone is present, the using of genomic clines across the contact zone can be a good alternative in the determination of regions under selection. (i.e. Payseur 2010; Nosil et al. 2012; Carneiro et al. 2013; Larson et al. 2014). These clines result in the comparison of the allele frequency of loci that are fixed in pure species and individual hybrid index. If a loci is fixed for different alleles in pure populations (hybrid index of 0 or 1), it is a candidate gene. If it is under selection, the alleles responsible for a certain habitat adaptation should be kept in high frequency independently if the populations contain hybrids. This will result in clines where the frequency of an allele changes drastically in the contact zone.

For most questions outlined above, the application of an entire genome sequence approach is ideal. This requires the construction of a good reference genome where most genes can be completely annotated. We started this process already and once it is complete, resequencing approaches can be employed.

All of the hypotheses developed here are solely based on empirical data produced using *Micromeria* from the Canary Islands as a model, and are therefore not necessarily applicable in other systems. Further investigation is necessary to explore whether the conclusions made here can be generalized. Meta-analyses approaches similar to the one employed by Pérez de Paz and Caujapé-Castells (2013) are good alternatives. Other archipelagos also need be studied to ensure that the patterns observed are not unique to the Canary Islands.

5.5. Summary of the major conclusions

This thesis was a continuation of the work of a previous PhD thesis (Puppo 2015). There, the first detailed phylogeographic patterns of *Micromeria* diversification on the Canary Islands archipelago and on the island of Tenerife were described. In this thesis a more detailed view of how *Micromeria* differentiate within each lineage was presented. This allowed the confirmation of some of the hypotheses outlined by Puppo et al. (2014, 2015) concerning the importance of Tenerife and Gran Canaria in the dispersal of *Micromeria.*

By using multiple markers with a neutral nature it was possible to evaluate the role of multiple colonization and hybridization in the species diversification process. As a consequence a scenario was confirmed in which multiple colonizations exist and have an impact on genetic variation. This led to the confirmation that a mechanism similar to the surfing syngameon hypothesis can be applied to single-island colonization. As

suggested in Caujapé-Castells (2011) the data support the establishment of species syngameon; however, instead of being geographically fixed, they expand by adding new species in the colonization front.

In the present work I discussed how the establishment of species syngameon can impact the distribution of morphological traits across the archipelago. By using a phylogenomic approach, it was possible to evaluate how the *M. varia* s.l. phenotype could have appeared through this process. RAD data supports multiple origins, however the concordance of phylogenetic signal with ecological zones for the first time in this species group raises questions regarding the role of introgression in the origin of these taxa phenotype. This led to the creation of the hypothesis that during the multiple colonization process the genes responsible for this trait are introduced on each island through introgression and then fixed through selection.

During this research several marker systems were established for *Micromeria,* resulting in the creation of many molecular resources for this genus and closely related taxa. Additionally, several new hypotheses were outlined, which may lead to the creation of new research projects that will contribute to a better understanding of evolutionary processes in island systems.

5.6. References

Ancochea, E., Fuster, J., Ibarrola, E., Cendrero, A., Coello, J., Hernan, F., ... & Jamond, C. (1990). Volcanic evolution of the island of Tenerife (Canary Islands) in the light of new K-Ar data. Journal of Volcanology and Geothermal Research, 44(3), 231-249.

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

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

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

Arnold, M. L. (2015). Divergence with genetic exchange. Oxford University Press.

Barbará, T., Palma-Silva, C., Paggi, G. M., Bered, F., Fay, M. F., Lexer, C. (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. Molecular Ecology, 16, 3759–3767.

Barber, J. C., Finch, C. C., Francisco-Ortega, J., Santos-Guerra, A., & Jansen, R. K. (2007). Hybridization in Macaronesian Sideritis (Lamiaceae): evidence from

incongruence of multiple independent nuclear and chloroplast sequence datasets. Taxon, 56(1), 74-88.

Brothers, A. N., Barb, J. G., Ballerini, E. S., Drury, D. W., Knapp, S. J., & Arnold, M. L. (2013). Genetic architecture of floral traits in Iris hexagona and Iris fulva. Journal of Heredity, est059.

Brown, R. P., Hoskisson, P. A., Welton, J. H., & Baez, M. (2006). Geological history and within-island diversity: a debris avalanche and the Tenerife lizard Gallotia galloti. Molecular Ecology, 15(12), 3631-3640.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., ... & Gormley, N. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal, 6(8), 1621-1624.

Carneiro, M., Baird, S. J., Afonso, S., Ramirez, E., Tarroso, P., Teotónio, H., ... & Ferrand, N. (2013). Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. Molecular ecology, 22(9), 2511-2525.

Caujapé-Castells, J. (2011). Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. The biology of island floras, 284-324.

Contreras-Díaz, H. G., Moya, O., Oromí, P., & Juan, C. (2003). Phylogeography of the endangered darkling beetle species of Pimelia endemic to Gran Canaria (Canary Islands). Molecular Ecology, 12(8), 2131-2143.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD Sequencing data: implications for genotyping. Molecular Ecology, 22(11), 3151-3164.

de Barba, M., Miquel, C., Lobreaux, S., Quenette, P. Y., Swenson, J., Taberlet, P. (2016). New prospects for standardized microsatellite genotyping in ecology using high-throughput sequencing (oral presentation), Concervation Genomics 2016, Vairao, Portugal

de Paz, J. P., & Caujapé-Castells, J. (2013). A review of the allozyme data set for the Canarian endemic flora: causes of the high genetic diversity levels and implications for conservation. Annals of botany, 111(6), 1059-1073.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in ecology & evolution, 24(6), 332-340.

Dlugosch, K. M., & Parker, I. M. (2007). Molecular and quantitative trait variation across the native range of the invasive species Hypericum canariense: evidence for

ancient patterns of colonization via pre-adaptation?. Molecular ecology, 16(20), 4269-4283.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nature reviews genetics, 5(6), 435-445.

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Molecular methods for evolutionary genetics, 157-178.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. Trends in Genetics, 28(7), 342-350.

Francisco-Ortega, J., Santos-Guerra, A., Kim, S. C., & Crawford, D. J. (2000). Plant genetic diversity in the Canary Islands: a conservation perspective. American Journal of Botany, 87(7), 909-919.

Galbany-Casals, M., Garcia-Jacas, N., Sáez, L., Benedí, C., & Susanna, A. (2009). Phylogeny, biogeography, and character evolution in Mediterranean, Asiatic, and Macaronesian Helichrysum (Asteraceae, Gnaphalieae) inferred from nuclear phylogenetic analyses. International journal of plant sciences, 170(3), 365-380.

García-Verdugo, C., Sajeva, M., La Mantia, T., Harrouni, C., Msanda, F., & Caujapé-Castells, J. (2015). Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. Molecular ecology, 24(4), 726-741.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., ... & Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology, 22(11), 3165-3178.

Gómez, A., González-Martínez, S. C., Collada, C., Climent, J., & Gil, L. (2003). Complex population genetic structure in the endemic Canary Island pine revealed using chloroplast microsatellite markers. Theoretical and Applied Genetics, 107(6), 1123-1131.

González-Pérez, M. A., & Sosa, P. A. (2009). Hybridization and introgression between the endemic Phoenix canariensis and the introduced P. dactylifera in the Canary Islands. Open For Sci J, 2, 78-85.

Gosset, C. C., & Bierne, N. (2013). Differential introgression from a sister species explains high FST outlier loci within a mussel species. Journal of Evolutionary Biology, 26(1), 14-26.

Gruenstaeudl, M., Santos-Guerra, A., & Jansen, R. K. (2012). Phylogenetic analyses of Tolpis Adans.(Asteraceae) reveal patterns of adaptive radiation, multiple colonization and interspecific hybridization. Cladistics, 1, 1-19.

Gübitz, T., Thorpe, R. S., & Malhotra, A. (2000). Phylogeography and natural selection in the Tenerife gecko Tarentola delalandii: testing historical and adaptive hypotheses. Molecular Ecology, 9(9), 1213-1221.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. Journal of Heredity, 105(S1), 795-809.

Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature, 487(7405), 94-98.

Herben, T., Suda, J., & Munclinger, P. (2005). The ghost of hybridization past: niche pre-emption is not the only explanation of apparent monophyly in island endemics. Journal of Ecology, 93(3), 572-575.

Hodel, R. G., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., ... & Soltis, D. E. (2016). The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. Applications in Plant Sciences, 4(6).

Huang, H., & Knowles, L. L. (2014). Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Systematic biology, syu046.

Jones, K. E., Reyes-Betancort, J. A., Hiscock, S. J., & Carine, M. A. (2014). Allopatric diversification, multiple habitat shifts, and hybridization in the evolution of Pericallis (Asteraceae), a Macaronesian endemic genus. American journal of botany, 101(4), 637-651.

Joyce, D. A., Lunt, D. H., Bills, R., Turner, G. F., Katongo, C., Duftner, N., ... & Seehausen, O. (2005). An extant cichlid fish radiation emerged in an extinct Pleistocene lake. Nature, 435(7038), 90-95.

Juan, C., Emerson, B. C., Oromí, P., & Hewitt, G. M. (2000). Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends in Ecology & Evolution, 15(3), 104-109.

Kim, S. C., McGowen, M. R., Lubinsky, P., Barber, J. C., Mort, M. E., & Santos-Guerra, A. (2008). Timing and tempo of early and successive adaptive radiations in Macaronesia. PLoS One, 3(5), e2139.

Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. Molecular ecology, 23(7), 1668-1678.

Lexer, C., Randell, R. A., & Rieseberg, L. H. (2003). Experimental hybridization as a tool for studying selection in the wild. Ecology, 84(7), 1688-1699.

Mairal, M., Pokorny, L., Aldasoro, J. J., Alarcón, M., & Sanmartín, I. (2015). Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in

Africa: the case of the Rand Flora genus Canarina (Campanulaceae). Molecular ecology, 24(6), 1335-1354.

Mata, V. A., Amorim, F., Corley, M. F., McCracken, G. F., Rebelo, H., & Beja, P. (2016). Female dietary bias towards large migratory moths in the European free-tailed bat (Tadarida teniotis). Biology letters, 12(3), 20150988.

Meimberg, H., Schachtler, C., Curto, M., Husemann, M., & Habel, J. C. (2016). A new amplicon based approach of whole mitogenome sequencing for phylogenetic and phylogeographic analysis: An example of East African white-eyes (Aves, Zosteropidae). Molecular phylogenetics and evolution, 102, 74-85.

Moya, Ó., Contreras-Diaz, H. G., Oromí, P., & Juan, C. (2004). Genetic structure, phylogeography and demography of two ground-beetle species endemic to the Tenerife laurel forest (Canary Islands). Molecular Ecology, 13(10), 3153-3167.

Nichols, P., Genner, M. J., Van Oosterhout, C., Smith, A., Parsons, P., Sungani, H., ... & Joyce, D. A. (2015). Secondary contact seeds phenotypic novelty in cichlid fishes. Proceedings of the Royal Society of London B: Biological Sciences, 282(1798), 20142272.

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. Phil. Trans. R. Soc. B, 367(1587), 332-342.

Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Molecular Ecology Resources, 10(5), 806-820.

Percy, D. M., & Cronk, Q. C. (2002). Different fates of island brooms: contrasting evolution in Adenocarpus, Genista, and Teline (Genisteae, Fabaceae) in the Canary Islands and Madeira. American Journal of Botany, 89(5), 854-864.

Pestano, J., & Brown, R. P. (1999). Geographical structuring of mitochondrial DNA in Chalcides sexlineatus within the island of Gran Canaria. Proceedings of the Royal Society of London B: Biological Sciences, 266(1421), 805-812.

Puppo, P. (2015). Diversification patterns of *Micromeria* Benth. (Lamiaceae) in the Canary Islands: genetic and morphological implications of inter-island colonization. PhD Thesis, University of Porto, Porto, Portugal.

Puppo, P., Curto, M., Velo-Antón, G., Pérez de Paz, P. L., & Meimberg, H. (2014). The influence of geological history on diversification in insular species: genetic and morphological patterns of Micromeria Benth. (Lamiaceae) in Tenerife (Canary archipelago). Journal of Biogeography, 41(10), 1871-1882.

Puppo, P., Curto, M., Gusmão-Guedes, J., Cochofel, J., de Paz, P. L. P., Bräuchler, C., & Meimberg, H. (2015a). Molecular phylogenetics of Micromeria (Lamiaceae) in the Canary Islands, diversification and inter-island colonization patterns inferred from nuclear genes. Molecular phylogenetics and evolution, 89, 160-170.

Puppo, P., Curto, M., & Meimberg, H. (2015b). Development and characterization of 16 microsatellite markers for Micromeria (Lamiaceae) from Tenerife (Canary Islands, Spain) using 454 sequencing. Conservation Genetics Resources, 7(3), 743-749.

Puppo, P., & Meimberg, H. (2015a). New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa, 230(1), 1-21.

Puppo, P., & Meimberg, H. (2015b). Corrections to Phytotaxa 230: New species and new combinations in Micromeria (Lamiaceae) from the Canary Islands and Madeira. Phytotaxa, 239(2), 197-198.

Riley, L., McGlaughlin, M. E., & Helenurm, K. (2016). Narrow water barriers prevent multiple colonizations and limit gene flow among California Channel Island wild buckwheats (Eriogonum: Polygonaceae). Botanical Journal of the Linnean Society, 181(2), 246-268.

Rubin, B. E., Ree, R. H., & Moreau, C. S. (2012). Inferring phylogenies from RAD sequence data. PLoS One, 7(4), e33394.

Sanmartín, I., Van Der Mark, P., & Ronquist, F. (2008). Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. Journal of Biogeography, 35(3), 428-449.

Schwarzer, J., Swartz, E. R., Vreven, E., Snoeks, J., Cotterill, F. P. D., Misof, B., & Schliewen, U. K. (2012). Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. Proceedings of the Royal Society of London B: Biological Sciences, 279(1746), 4389-4398.

Seehausen, O. (2004). Hybridization and adaptive radiation. Trends in Ecology & Evolution, 19(4), 198-207.

Seehausen, O., Koetsier, E., Schneider, M. V., Chapman, L. J., Chapman, C. A., Knight, M. E., ... & Bills, R. (2003). Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. Proceedings of the Royal Society of London B: Biological Sciences, 270(1511), 129-137.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology letters, 9(5), 615-629.

Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., ... & Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Scientific reports, 5, 9687.

Silvertown, J. (2004). The ghost of competition past in the phylogeny of island endemic plants. Journal of Ecology, 92(1), 168-173.

Silvertown, J., Francisco-Ortega, J., & Carine, M. (2005). The monophyly of island radiations: an evaluation of niche pre-emption and some alternative explanations. Journal of Ecology, 93(4), 653-657.

Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F. R., Shih, C. H., ... & Kohn, M. H. (2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Current Biology, 21(15), 1296-1301.

Stern, D. L. (2013). The genetic causes of convergent evolution. Nature Reviews Genetics, 14(11), 751-764.

Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W., & Tucker, P. K. (2010). The variable genomic architecture of isolation between hybridizing species of house mice. Evolution, 64(2), 472-485.

Thomson, R. C., Wang, I. J., & Johnson, J. R. (2010). Genome-enabled development of DNA markers for ecology, evolution and conservation. Molecular Ecology, 19(11), 2184-2195.

Uribe-Convers, S., Settles, M. L., & Tank, D. C. (2016). A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the South American species of Bartsia L.(Orobanchaceae). PloS one, 11(2), e0148203.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends in genetics, 30(9), 418-426.

Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., ... & FitzGerald, R. D. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. Royal Society open science, 3(1), 150565.

Watts, A. B., & Masson, D. G. (1995). A giant landslide on the north flank of Tenerife, Canary Islands. Journal of Geophysical Research: Solid Earth, 100(B12), 24487-24498.

Whittaker, R. J., Triantis, K. A., & Ladle, R. J. (2008). A general dynamic theory of oceanic island biogeography. Journal of Biogeography, 35(6), 977-994.

Xu, P., Xu, S., Wu, X., Tao, Y., Wang, B., Wang, S., ... & Li, G. (2014). Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. The Plant Journal, 77(3), 430-442.

Appendix

Appendix I: Supplementary material from section 2.3.

Curto, M.A., Tembrock, L.R., Puppo, P., Nogueira, M., Simmons, M.P., Meimberg H. (2013). Evaluation of microsatellites of *Catha edulis* (qat; Celastraceae) identified using pyrosequencing. Biochemical Systematics and Ecology, 49, 1–9.

Table S1. Primers designed with their respective SSR sequence and expected product length.

Locus	Genbank accession #	SSR	Expected product size (bp)	Forward primer	Reverse primer	# of samples amplified
CE1	JX406758	(AAG)$_7$	223	GAGGAGGAAGAGAGTTCAGCA	GCTTCTTGTTGGTCCGACTC	23[a]
CE2	JX406759	(CCT)$_7$	189	CATGTCAGGCAGAATCAACA	GCCTGCCATATTTAGGGATT	11
CE3	JX406760	(CT)$_{12}$	156	CCTTCTATCACCCTCCCACA	CCCTCTGTATTGCACGGTTT	23
CE4	JX406761	(AAG)$_8$	185	GCAATCCCAATTGAGAGCA	CAAGTCCACCAGCATTAGCA	23
CE5	JX406762	(AT)$_{10}$	238	ACATGTCAACATCCCACCAT	TGGATTAAGGCCTGGTTGTT	0
CE6	JX406763	(CT)$_{13}$	163	CAAGCCTCACCACACATCTC	AAGAGTGCAAACTTGGTGAGC	23[a]
CE7	JX406764	(AT)$_9$	195	GGCTTACCCTCAACCTTTCC	CGCGTACATGAAATCTGTGA	23[a]
CE8	JX406765	(ATT)$_8$	188	ATTGTTGAAAGGGCCAAGAG	CACCCAATCATGCTTTCAGA	21
CE9	JX406766	(AAT)$_7$	222	AGCAGGCCCAAAGTTCAATA	CGGAGACCGGTTAAAATGAC	0
CE10	JX406767	(AT)$_1$3	194	CGATGAGGTCCAATCTCTCC	CACATGTCCTCGTGATTTGG	10
CE11	JX406768	(ATGT)$_9$	223	ACGATTGCAGGTCCTCTCAT	TTGTCGATGCTGTTGAGACA	0
CE12	JX406769	(AT)$_{12}$	151	TCCTGGCAATCCCACATAA	TGCGACTTGCAACCATGTA	10
CE13	JX406770	(AAG)$_7$	213	CAGAAACAACACCTGCCTCA	ATCTCTCCCTCCTGACACCA	0
CE14	JX406771	(CT)$_{12}$	174	TCGTTTTCTCTCGTGGACTG	ACCCTGTCGCAGCTATGTTT	3
CE15	JX406772	(AG)$_9$	162	CCTCAATCGGACAACATCAA	GCGTACCTGAATCCCTCTTG	23
CE16	JX406773	(AAG)$_8$	238	CCATTGTTGTGGTGGTTGAG	ACCATCAACCATGCCTTCTC	23
CE17	JX406774	(AAT)$_7$	154	AGTGTTTCCATTGTGGATGG	GCATCGATTCCCTGATTGTT	0
CE18	JX406775	(AGG)$_8$	226	CTTGTGCCAATCCAAATCCT	TTCCACCCTGGTTTCTTCC	0
CE19	JX406776	(AC)$_{12}$	123	CAAGTGGTGGCCGTAAAAG	AGTCACAAGATCCAACTGTCG	10
CE20	JX406777	(ATT)$_9$	275	AACTTCATCGAGGCACTCAAG	GGGGTCACCAAGATACGAAG	0
CE21	JX406778	(AT)$_{13}$	132	AACATGTATTCCGGTGGTAGC	CAACAACAACCCACAACAAC	16
CE22	JX406779	(AG)$_{10}$	225	GCTGCAAGAGGTAGGTGGAT	TTCTCTCTCTCGCTTGGCTA	23
CE23	JX406780	(ATC)$_{13}$	167	AGCAGCAGCAACAACAAGAA	CAGCAAGGGAGGCCTTATTA	23

Table S1. Continued

Locus	Genbank accession #	SSR	Expected product size (bp)	Forward primer	Reverse primer	# of samples amplified
CE24	JX406781	$(AC)_{11}$	204	TCTTGCTCCTTCAACCTCAA	ATCTTGCCAGCTTCCGTCTA	23
CE25	JX406782	$(AAT)_{14}$	146	AACCTGCAATCTGTTTGACC	TGATGGGCATCAACTGGTAT	14
CE26	JX406783	$(AT)_{10}$	239	GATCCAGACCCAACCCATT	TGGTGCATGTGTCCTTTTTC	0
CE27	JX406784	$(AG)_{21}$	300	AATGCATGAGCATGACAAGG	AGGAAGAGGCGATTGTTTTG	0
CE28	JX406785	$(CTT)_{9}$	223	CGACCACACAAATCTTCCAG	GGAAGTTCCTCTTCTGGGGTA	1
CE29	JX406786	$(AAT)_{7}$	232	GCCAACCTCTTGTTCTGGAG	TAGGTTTGGCCATTCGATTC	23
CE30	JX406787	$(AGG)_{7}$	224	TGAAGAGTCCAAAGCAGCAG	AGACCCATGCACTACCCAAC	23
CE31	JX406788	$(AC)_{9}$	173	TTCCCAAAAGTGTTGCTGAG	CTTTACTAGGGCCCGTCCTT	23
CE32	JX406789	$(AT)_{11}$	234	TGAACTTGTGGGTTGTTGGA	GGGACCTGGTTGTGTTTGAT	21[a]
CE33	JX406790	$(CTT)_{7}$	213	AGGTTGAGCTGGAACGAGAA	TTCTCATTTGCCTTGTGTGC	23
CE34	JX406791	$(CT)_{12}$	173	CGGATGCCAAAACACTATCA	ATCCAAGAGGTTTTGGTTGC	22
CE35	JX406792	$(AG)_{12}$	178	GCAAATGTCGGTCAAACCTT	CTTCTCCAGTGGGCTTCTGT	0
CE36	JX406793	$(ATT)_{15}$	320	GATTGGTGGCCACTTCTTTC	CATGCATGCAGGAGACTTGT	7
CE37	JX406794	$(ATCT)_{11}$	215	ACTCGAAAAACATGGTGCAG	TGAGCCTCAATCTGGAGACA	23
CE38	JX406795	$(ATT)_{11}$	196	GGATGACCAAGTCGATTCAA	GTTCTGCACAGCCCTAAACC	9
CE39	JX406796	$(AAC)_{9}$	163	GTGGTTCGAGTCCAGTCCTT	CGTTGGAATACACGTGTTGG	23
CE40	JX406817	$(GT)_{10}$	151	TGGTATAGCCCATATCGTCAG	CACACTACGCTTCACGCTTC	23
CE41	JX406797	$(AG)_{15}$	219	GGACAGAATTCCCAAAACGA	ATTGCCAGCTCGATCACTCT	23
CE42	JX406798	$(GT)_{9}$	230	AAGGGGAAGGAGAGAGATGC	CCTCATCCTGATGTGGCTAA	23
CE43	JX406799	$(ATT)_{10}$	156	CAGATCCTCCTCCTCCTTCA	GGATGCCACAAATGCTGAT	23
CE44	JX406800	$(AGT)_{8}$	226	CGACTGCAGATGGTGAGAAA	CCAGTCCAAAGCACCCTAAC	0
CE45	JX406801	$(AG)_{11}$	247	GGTTTAGCCTCTCCTGCAAGT	CAAGTGCGGACTCAACAAGT	23
CE46	JX406802	$(ATT)_{10}$	259	GGGGTTTGTGGCTTGTTTT	AGGGATCCACCCCTGATAGT	0
CE47	JX406803	$(GTT)_{8}$	238	GTGATGATGGGTTGGAATTG	ATCCTCATCATCCCCATCAG	23

Table S1. Continued

Locus	Genbank accession #	SSR	Expected product size (bp)	Forward primer	Reverse primer	# of samples amplified
CE48	JX406804	$(AG)_{12}$	226	GATGTGGATCTCCCACCTGT	ACCCCCTACCTCAATCCACT	13
CE49	JX406805	$(AC)_{9}$	166	GGAAAGAGGGAAGTTGGAATG	CCACTCACCATTCTCTGCAA	20[a]
CE50	JX406806	$(CT)_{12}$	158	AACTACCCGCCATTTCGAC	AGCTGGGCGATTGACTAAAG	23
CE51	JX406807	$(TC)_{11}$	167	GAAATTCCCACGACTGTGAAG	TTTAACGAATTCTAATGACGGAG	0
CE52	JX406771	$(CT)_{11}$	222	AACTACACCGGAGCACAGCA	CAGCTATGTTTGCAGAATCGC	0
CE53	JX406819	$(GA)_{15}$	163	TTAGGAATCGAAAGTGAGGC	TCTTACGTCCTCATATGTACAGG	0
CE54	JX406812	$(CT)_{12}$	153	TTTGGTCTTTAATGGCGAAGTC	GGAAGGATCTATGACAAGTTGC	0
CE56	JX406813	$(GA)_{10}$	199	ACAACCAATGTCGTACAGAGA	CAGTATTTGTATCTGCAGTACAG	23
CE57	JX406814	$(TTA)_{13}$	187	TCCTGTCCTGATAATATCCTG	AGTCCCACTGATTGTTATGAC	23
CE58	JX406815	$(TTA)_{8}$	163	TTCCATTGTGGAAATTTGGTG	CTTAATGGAACCTATGATCAGC	23
CE59	JX406816	$(TTA)_{15}$	192	GAATTCTGATCAAAACTACCAG	GGCTTCAGAAAGAATTGGATG	22
CE60	JX406808	$(TTAA)_{6}$	269	CTACTTGTCATAACCTTATCCAC	TGGGTTGAATTCAATAGAGTAGC	19
CE61	JX406818	$(AT)_{19}$	145	TATAATCAGTGTGGGATAGA	ACCTAATAATTAGTTACTTGTAC	0
CE62	JX406809	$(TA)_{13}$	196	TTGAACTGTTGCCCTCTTGC	CTACTGCAACTTATCATTATACC	17
CE63	JX406810	$(CT)_{19}$	196	AGAACTCCCTGTAAACTTCCAC	AATACTAACCGCTTAATCGACTG	14
CE64	JX406811	$(CT)_{11}$	232	CATCCTCAGCCATCTGAGCA	AACTGTCTACATGAACTTGTATTC	23

[a] Locus with more than one allele per sample or fixed heterozygosity.

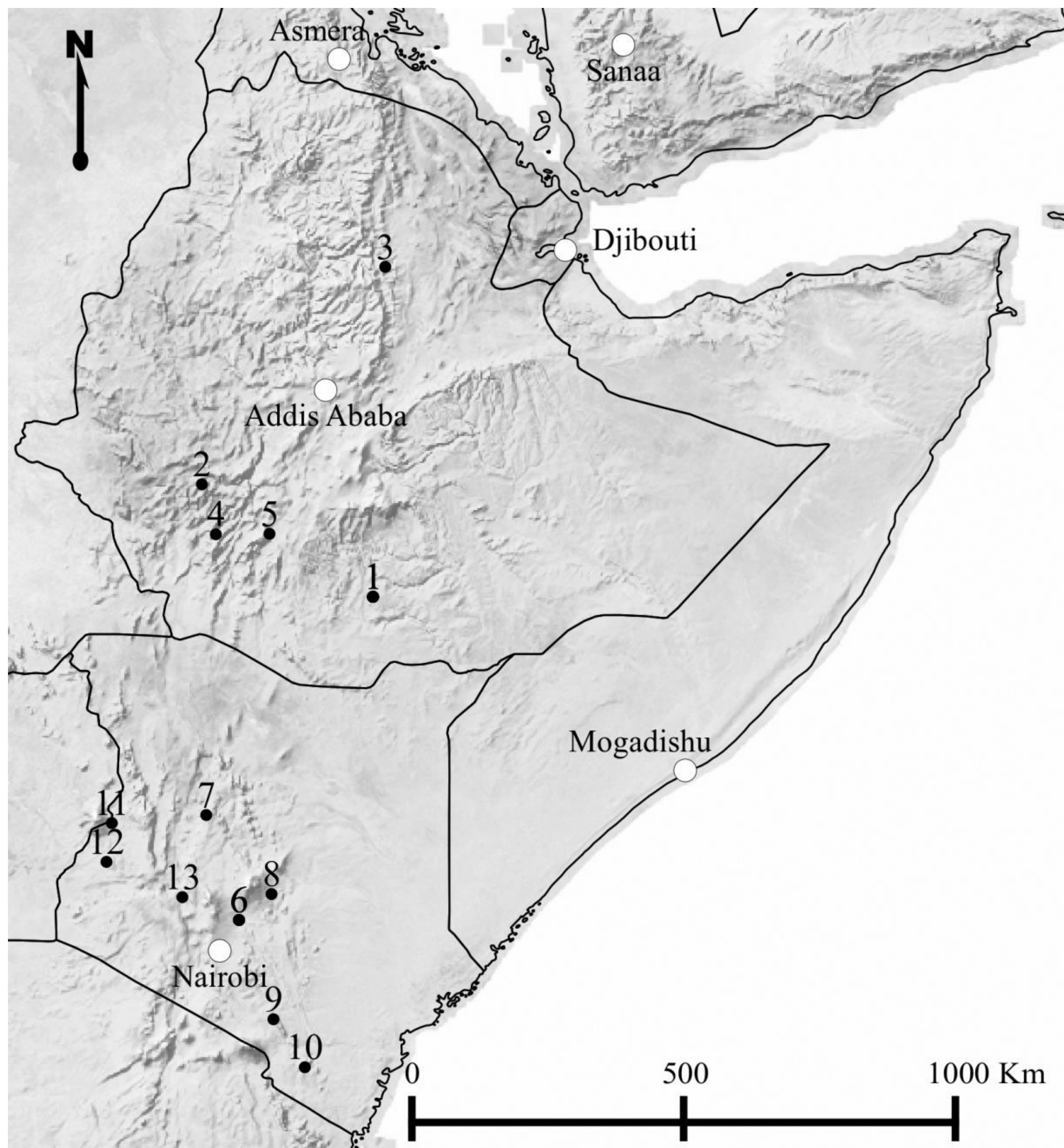


Fig. S1. Localities of the qat samples used in this study. The numbers correspond to the population number on Table 1.

Appendix II: Supplementary material from section 2.4.

Curto, M., Nogueira, M., Beja, P., Amorim, F., Schümann, M., & Meimberg, H. (2015). Influence of past agricultural fragmentation to the genetic structure of Juniperus oxycedrus in a Mediterranean landscape. Tree Genetics & Genomes, 11(2), 1-13.

Table S1. List of samples used with information about the a priori population they were assigned, individual ID, coordinates, altitude and trunk perimeter measures. Information about unmeasured individuals visually larger than 100 cm is also included ("> 100"). NA corresponds to data that was not possible to be collected.

Population	Individual ID	Latitude	Longitude	Altitude (m)	Trunks perimeter in cm (TP)
I	12_1	41.47779	-6.79929	534	50
I	12_1.2	41.478026	-6.798474	NA	NA
I	12_2	41.47817	-6.79933	595	NA
I	12_4	41.47789	-6.79926	595	35
I	12_5	41.47807	-6.79806	602	25
I	12_6	41.47821	-6.79643	604	NA
II	13_1	41.48153	-6.6891	556	50
II	13_11	41.46907	-6.68252	466	50
II	13_12	41.46593	-6.67987	447	90
II	13_14	41.462142	-6.67823	417	114
II	13_15	41.45816	-6.67923	385	70
II	13_16	41.453175	-6.675993	302	NA
II	13_16.2	41.454993	-6.677665	339	NA
II	13_17	41.45657	-6.68044	396	NA
II	13_18	41.44901	-6.6865	392	NA
II	13_19.2	41.44764	-6.68763	378	35
II	13_20	41.44453	-6.69224	313	> 100
II	13_3	41.48051	-6.69189	449	75
II	13_4	41.47807	-6.68971	555	132
II	13_5	41.47704	-6.68771	534	50
II	13_7	41.47639	-6.68564	522	50
II	13_8	41.475795	-6.686857	546	60
II	13_9	41.47122	-6.68363	499	NA
II	14_1	41.44095	-6.69122	260	NA
II	14_1.2	41.44095	-6.69122	NA	NA
II	14_2	41.43901	-6.69235	295	NA
II	14_3	41.43704	-6.69788	310	NA
II	14_3.2	41.4378	-6.69707	300	67
II	14_5	41.43391	-6.70344	293	NA
II	14_6	41.43107	-6.70692	264	NA
II	14_7	41.43126	-6.70961	287	NA
II	14_8	41.43052	-6.70607	300	NA
II	14_9	41.43134	-6.70833	263	75
III	15_1	41.42517	-6.676283	394	> 100
III	15_1.2	41.42517	-6.676283	NA	NA
III	15_11	41.40684	-6.80341	NA	NA
III	15_2	41.42413	-6.76438	385	> 100
III	15_3	41.42186	-6.76716	365	NA
III	15_4	41.42103	-6.76602	361	NA

Table S1. Continued.

Population	Individual ID	Latitude	Longitude	Altitude (m)	Trunks perimeter in cm (TP)
III	15_5	41.41669	-6.77273	298	NA
III	15_6	41.41278	-6.78473	236	NA
III	15_7	41.40871	-6.78794	229	NA
III	15_8	41.40665	-6.79607	226	NA
III	15_9	41.40636	-6.80252	226	NA
IV	1_10	41.31004	-6.88167	NA	150
IV	1_12	41.31057	-6.88263	NA	75
IV	1_3	41.30853	-6.87978	NA	NA
IV	1_4	41.30861	-6.87969	NA	100
IV	1_5	41.30855	-6.87941	NA	50
IV	1_7	41.30811	-6.87958	NA	50
IV	1_8	41.3083	-6.87931	NA	100
IV	1_9	41.30873	-6.87943	NA	100
IV	16_1	41.33322	-6.86083	468	NA
IV	16_1.2	41.33322	-6.86083	NA	NA
IV	16_11	41.33296	-6.82593	204	93
IV	16_12	41.32613	-6.82534	194	NA
IV	16_13	41.33231	-6.82323	322	NA
IV	16_14	41.32948	-6.8225	394	NA
IV	16_15	41.32918	-6.81318	516	133
IV	16_15	41.32918	-6.81318	NA	NA
IV	16_16	41.32908	-6.81306	474	37
IV	16_2	41.33863	-6.85192	537	75
IV	16_3	41.33908	-6.84353	506	NA
IV	16_4	41.33878	-6.83808	465	NA
IV	16_5	41.33857	-6.83111	407	NA
IV	16_6	41.34734	-6.82166	206	NA
IV	16_7	41.3415	-6.82802	309	110
IV	16_8	41.34513	-6.8222	253	50
IV	2_1	41.32343	-6.86595	450	NA
IV	2_1.2	41.32343	-6.86595	NA	NA
IV	2_2	41.32336	-6.86626	NA	25
IV	2_3	41.32318	-6.86628	NA	25
IV	2_4	41.32306	-6.86597	NA	50
IV	2_5	41.3233	-6.8657	NA	75
IV	3_1	41.32086	-6.86477	NA	75
IV	3_2	41.3069	-6.865	NA	75
IV	3_3	41.32018	-6.86575	NA	50
IV	3_4	41.32004	-6.86521	NA	50
IV	4_1	41.31957	-6.86504	NA	100

Table S1. Continued.

Population	Individual ID	Latitude	Longitude	Altitude (m)	Trunks perimeter in cm (TP)
IV	4_2	41.31834	-6.8647	NA	100
IV	4_3	41.31755	-6.86398	NA	50
IV	4_4	41.31551	-6.86331	NA	NA
IV	4_6	41.31466	-6.86365	NA	75
IV	5_2	41.30899	-6.85683	186	50
IV	5_2	41.30899	-6.85683	NA	NA
IV	5_4	41.30956	-6.85591	187	37,1
IV	5_7	41.31041	-6.85822	NA	51,2
IV	6_2	41.3153	-6.8614	443	NA
IV	6_2.2	41.3153	-6.8614	NA	NA
IV	6_3	41.3186	-6.86286	420	> 100
IV	6_5	41.31717	-6.86249	404	NA
IV	6_6	41.318	-6.86269	410	NA
V	10_1	41.23927	-6.96116	165	NA
V	10_2	41.2389	-6.96083	186	NA
V	10_3	41.2397	-6.9563	267	> 100
V	10_5	41.23853	-6.95402	188	NA
V	10_6	41.23683	-6.95108	187	NA
V	10_8	41.22591	-6.94637	406	NA
V	11_2	41.24107	-6.92163	420	> 100
V	11_3	41.24083	-6.92174	420	> 100
V	11_4	41.23916	-6.92389	420	> 100
V	11_5	41.24061	-6.92221	NA	> 100
V	11_7	41.24052	-6.92124	NA	> 100
V	11_8	41.24044	-6.922142	NA	NA
V	4_5	41.28670	-6.9224	511	35
V	7_1	41.2808	-6.9208	519	25
V	7_11	41.40684	-6.80341	NA	NA
V	7_2	41.27983	-6.91936	445	15
V	7_2	41.2806	-6.9211	516	50
V	7_4	41.27896	-6.91922	422	20
V	7_6	41.2803	-6.9227	507	50
V	7_7	41.2798	-6.9222	NA	NA
V	7_8	41.2801	-6.9206	485	100
V	7_9	41.2803	-6.92122	NA	NA
V	8_1	41.247	-6.9244	375	NA
V	8_2	41.2484	-6.928	429	25
V	9_1	41.24697	-6.938	340	15
V	9_11	41.24663	-6.93846	NA	NA
V	9_2	41.24725	-6.93811	343	30

Table S1. Continued.

Population	Individual ID	Latitude	Longitude	Altitude (m)	Trunks perimeter in cm (TP)
V	9_2.2	41.2464	-6.938	399	50
V	9_3	41.2471	-6.9384	349	NA
V	9_4	41.24678	-6.93859	351	NA
V	9_5	41.24651	-6.9391	357	NA
V	9_6	41.24656	-6.9393	370	NA
V	9_7	41.2463	-6.9388	404	> 100
V	9_9	41.2462	-6.9393	437	> 100

Table S2. Markers showing positive amplification results for most of the samples. The information displayed corresponds Genebank reference for the sequence used to design the primers, sequences used for primer construction, microsatellite motif, repetition number, and size range from the alleles observed.

Marker name	Sequence Genebank reference	Repeat Number	Motif	Alleles Size	Tail primer (M13 oligo)	Primer Forward	(GTTT) + Primer Reverse
Joxy1	KM013332	11	ga	277-285	TAATACGACTCACTATAGGG	TTGGTGTACCGATCAACACAG	(GTTT)GCTAGGGGAGTTTGATACAAGG
Joxy3	KM013320	9	aca	121-133	TGTAAAACGACGGCCAGT	CTTATCCCAAAGGCAACCAG	(GTTT)ATTGGAGCCACTACCACCAC
Joxy4	KM013322	10	tc	129-143	TAATACGACTCACTATAGGG	AATTGGTCATAACCCAAAAGG	(GTTT)AACAATGAAAGAGAATAGCACCA
Joxy8	KM013323	13	ca	130-152	TGTAAAACGACGGCCAGT	TCGCTAGTCCAAATCAACCTG	(GTTT)TCCACAAGGTGTTTCATATTTCT
Joxy10	KM013318	7	gt	117-124	TTTCCCAGTCACGACGTTG	CCTGCAGCTTCTAAAAGATTGT	(GTTT)TCCCACATTGGTGGGTAATC
Joxy11	KM013328	7	tct	191-191	TGTAAAACGACGGCCAGT	TGCTTTATTGGCCTGGTCTC	(GTTT)GAAGGAGAAATTTAAGGAGGTGG
Joxy12	KM013317	8	ttc	111-120	TAATACGACTCACTATAGGG	CTCAAGCTCTTCAAGCTTTGTT	(GTTT)GAGCTTGTGGTAATCTTGGAGA
Joxy13	KM013333	8	ct	91-97	TGTAAAACGACGGCCAGT	GGAGATCCACATTCATCCATC	(GTTT)CCTCATGGAATTTATTGTCGTG
Joxy14	KM013316	8	ac	110-114	TGTAAAACGACGGCCAGT	TCCACTTTATACATATAGCTTGTGGG	(GTTT)GACAAGTCCAACCACAAAGGA
Joxy17	KM013319	8	agg	120-126	TAATACGACTCACTATAGGG	CCTTTTGGGAAGGGAAAGAG	(GTTT)TCCTTAATCACCTTTCACACCA
Joxy20	KM013321	8	agg	127-136	GATAACAATTTCACACAGG	ACGAGCCACTAGGAAGGAGG	(GTTT)CTCCCCTGGTAGGCTTCTTT
Joxy22	KM013325	7	ttc	149-152	TAATACGACTCACTATAGGG	TAGGCTTGGAAACAACTGGC	(GTTT)GGCATACTTAGCAAAGGACCA
Joxy23	KM013324	9	caacat	146-188	TTTCCCAGTCACGACGTTG	GTTGGCAACTTCATGACTGG	(GTTT)GTGTAATGTTGGTGCGGATG
Joxy28	KM013326	9	ag	162-168	GATAACAATTTCACACAGG	CACATGGGGAAAATAAGAGCA	(GTTT)TTAATCATATAATTCATGGTTCACATT
Joxy31	KM013327	7	ag	173-177	TGTAAAACGACGGCCAGT	TTGGCGACTTTGTACTAGCC	(GTTT)GGGAAACAATCTTTGGCAATGA
Joxy35	KM013329	13	tg	196-204	TAATACGACTCACTATAGGG	CTCCCATAACCTTAGACAATGAGAA	(GTTT)CCCTGAAGATTCTCCTCTAGCA
Joxy37	KM013330	7	tg	222-224	GATAACAATTTCACACAGG	TCTTGGTAGTGGCATGTGGA	(GTTT)CAACTTTCGTGACAATGAGAATC
Joxy42	KM013331	9	ca	257-263	TTTCCCAGTCACGACGTTG	TGCTTCCTGTCTTCATCTCTTG	(GTTT)CATGTTGATCTCACTCAGGCA

Table S3. Family structure analyses result at the individual level including to which population they belong. The probability of that relationship is also included.

17 Markers

Potential offspring	Offspring population	Potential parent	parent population	Probability
13_5	II	11_2	V	0.990
1_5	IV	16_11	IV	1.000
4_3	IV	6_3	IV	1.000
1_7	IV	11_2	V	0.990
3_3	IV	11_7	V	1.000
7_6	V	16_15	IV	0.578
7_4	V	6_3	IV	1.000
7_1	V	11_2	V	0.784

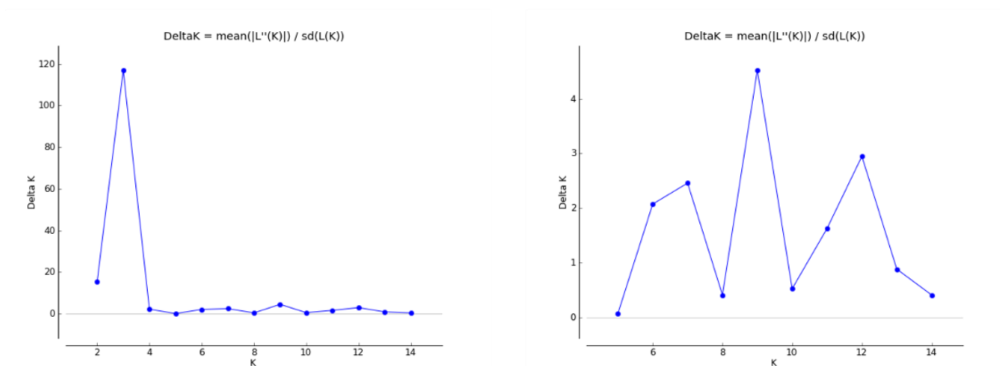
11 Markers

Potential offspring	Offspring population	Potential parent	parent population	Probability
12_5	I	6_3	IV	1.000
13_5	II	11_2	V	0.994
13_11	II	16_15	IV	0.766
13_1	II	4_2	IV	1.000
1_7	IV	11_2	V	0.994
3_3	IV	11_2	V	0.961
16_16	IV	16_15	IV	0.764
3_4	IV	4_2	IV	1.000
4_3	IV	6_3	IV	1.000
8_2	V	16_15	IV	0.766
7_2	V	4_2	IV	1.000
7_1	V	6_3	IV	1.000
7_4	V	6_3	IV	1.000

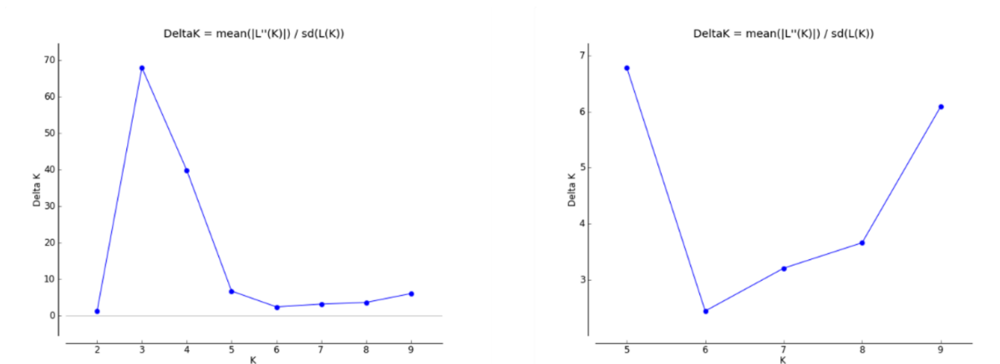
Appendix III: Supplementary material from section 3.2.

Puppo, P., Curto, M., Meimberg, H (2016) Genetic structure of *Micromeria* (Lamiaceae) in Tenerife, the imprint of geological history and hybridization on within-island diversification. Ecology and Evolution. 2045-7758

All species



Central species



Hybrids Test

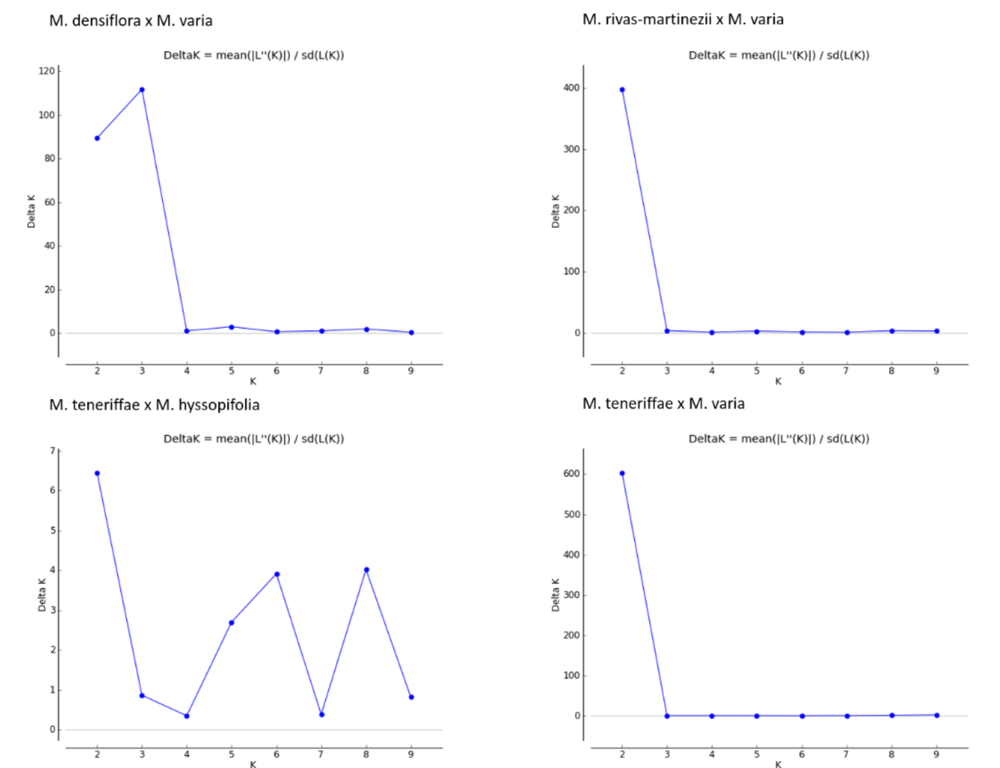


Fig. S1. Delta K plots obtained by STRUCTURE Harvester for all STRUCTURE tests performed.

Table S1. List of *Micromeria* samples used in the present study including region, locality name and number, geographical coordinates (Latitude, Longitude), number of samples per locality (N), and collection information. TFC, Herbarium of the Universidad de la Laguna in Tenerife. Numbers of localities correspond to those in Fig. 1. * No geographical coordinates are provided for restricted species.

Species	Region	Locality	Loc. Nr.	Latitude	Longitude	N	Collection
M. teneriffae var. cordifolia	Southern Coast	Fasnia	1	28,2189	-16,4153	4	Puppo 299-301, 303 (TFC)
M. teneriffae var. cordifolia	Southern Coast	Fasnia	2	28,22351	-16,41321	3	Meimberg 25
M. teneriffae var. teneriffae	Southern Coast	Güímar	4	28,2944	-16,403	3	Puppo 151, 154, 156 (TFC)
M. teneriffae var. teneriffae	Southern Coast	Btw Fasnia-Güímar	5	28,2497	-16,4382	4	Puppo 160-161, 163-164 (TFC)
M. teneriffae var. teneriffae	Southern Coast	Bco Eras	6	28,2129	-16,4532	2	Puppo 166, 168 (TFC)
M. teneriffae var. teneriffae	Anaga	Bco Tahodio, Anaga	3	28,4982	-16,2589	7	Puppo 185-188, 190-192 (TFC)
*M. glomerata**	Anaga	Taganana, Anaga	7			5	Puppo 200-203 (TFC)
*M. rivas-martinezii**	Anaga	Roque Juan Bay, Anaga	8			11	Puppo 208-215, 218, 227-228 (TFC)
*M. densiflora**	Teno	Bujame, Teno	9			5	Puppo 255-257, 259-260 (TFC)
*M. lasiophylla**	Teide	Teide Nat. Park	10			9	Puppo 274-276, 279-281, 283, 286, 288 (TFC)
M. varia	Anaga	Anaga	11	28,54328	-16,20538	7	Meimberg 26
M. varia	Anaga	Anaga	12	28,54757	-16,21141	5	Meimberg 27
M. varia	Anaga	Anaga	13	28,54597	-16,21688	7	Meimberg 28
M. varia	Anaga	Anaga	14	28,53219	-16,25843	7	Meimberg 29
M. varia	Anaga	S. Andrés, Anaga	15	28,5162	-16,1748	1	Puppo 184 (TFC)
M. varia	Anaga	Bco Tahodio, Anaga	16	28,504	-16,2644	3	Puppo 193-195 (TFC)
M. varia	Anaga	Parque Rural, Anaga	17	28,5296	-16,1941	2	Puppo 196, 197 (TFC)
M. varia	Anaga	Parque Rural, Anaga	18	28,5626	-16,2094	2	Puppo 198, 199 (TFC)
M. varia	Anaga	Afur, Anaga	19	28,5526	-16,2378	2	Puppo 204, 206 (TFC)
M. varia	Anaga	Antequera, Anaga	20	28,5425	-16,1305	1	Puppo 223 (TFC)
M. varia	Anaga	Antequera, Anaga	21	28,5482	-16,1486	5	Puppo 230, 232, 236, 237-238 (TFC)
M. varia	Anaga	Antequera, Anaga	22	28,5434	-16,1379	1	Puppo 224 (TFC)
M. varia	Teno	Teno	23	28,32792	-16,85619	10	Meimberg 31
M. varia	Teno	Teno	24	28,37647	-16,85258	5	Meimberg 32
M. varia	Teno	Teno	25	28,3421	-16,8615	6	Puppo 247-248, 250-253 (TFC)
M. varia	Teno	Bujame, Teno	26	28,3518	-16,8717	2	Puppo 261-262 (TFC)
M. lachnophylla	Teide	El Portillo	27	28,3091	-16,5672	3	Puppo 290, 294, 296 (TFC)
M. lachnophylla	Teide	ca. El Portillo	28	28,31343	-16,57074	2	Meimberg 12
M. lachnophylla	Teide	ca. El Portillo	29	28,31593	-16,5743	2	Meimberg 13
M. lachnophylla	Teide	ca. El Portillo	30	28,30353	-16,56701	12	Meimberg 11
M. hyssopifolia var. glabrescens	Northen Coast	Guancha	31	28,38088889	-16,65402778	2	Puppo 561 (TFC)

Table S1. Continued

Species	Region	Locality	Loc. Nr.	Latitude	Longitude	N	Collection
M. hyssopifolia var. glabrescens	Northen Coast	Guancha	32	28,36280556	-16,65727778	2	Puppo 589 (TFC)
M. hyssopifolia var. glabrescens	Northen Coast	Guancha	33	28,4	-16,6647	1	Puppo 563 (TFC)
M. hyssopifolia var. glabrescens	Northen Coast	Mirador de Mazapé	35	28,3895	-16,63627778	2	Puppo 588 (TFC)
M. hyssopifolia var. glabrescens	Northen Coast	Rambla de Castro	36	28,39525	-16,62472222	1	Puppo 560 (TFC)
M. hyssopifolia var. glabrescens	Northen Coast	Rambla de Castro	37	28,39592	-16,58981	3	Puppo 239, 241, 243 (TFC)
M. hyssopifolia var. glabrescens	Northen Coast	Realejos	38	28,37808333	-16,6295	2	Puppo 587 (TFC)
M. hyssopifolia var. kuegleri	Southern Coast	Los Abades	39	28,14114	-16,45456	3	Puppo 175-176, 178 (TFC)
M. hyssopifolia var. kuegleri	Southern Coast	Acantilado de la Hondura	40	28,200861	-16,424861	3	Puppo 179-181 (TFC)
M. hyssopifolia var. hyssopifolia	Southern Coast	Arico	42	28,18306	-16,45317	3	Puppo 169-171 (TFC)
M. hyssopifolia var. hyssopifolia	Southern Coast	Arico	43	28,15231	-16,49158	1	Puppo 172 (TFC)
M. hyssopifolia var. hyssopifolia	Southern Coast	Btw Fasnia y Güímar	44	28,2497	-16,4382	2	Puppo 162, 165 (TFC)
M. hyssopifolia var. hyssopifolia	Southern Coast	East coast	45	28,304	-16,38094	6	Meimberg 7
M. hyssopifolia var. hyssopifolia	Southern Coast	Güímar	46	28,2944	-16,403	1	Puppo 153 (TFC)
M. hyssopifolia var. hyssopifolia	Southern Coast	Southeast, nr Fasnia	59	28,24405	-16,40731	5	Meimberg 6
M. hyssopifolia var. hyssopifolia	Southern Coast	Southeast, nr Fasnia	60	28,20048	-16,42502	6	Meimberg 8
M. hyssopifolia var. hyssopifolia	Northeast	Arafo	41	28,37269	-16,42764	3	Puppo 268-269, 272 (TFC)
M. hyssopifolia var. hyssopifolia	Northeast	Northeast	47	28,33034	-16,53259	5	Meimberg 14
M. hyssopifolia var. hyssopifolia	Northeast	Northeast	48	28,34888	-16,52472	9	Meimberg 15
M. hyssopifolia var. hyssopifolia	Northeast	Northeast	49	28,3614	-16,50061	4	Meimberg 16
M. hyssopifolia var. hyssopifolia	North	Northwest	50	28,23585	-16,7598	6	Meimberg 3, 10
M. hyssopifolia var. hyssopifolia	North	West coast	66	28,15732	-16,79511	3	Meimberg 18
M. hyssopifolia var. hyssopifolia	Adeje	South, nr Adeje	51	28,06075	-16,66895	8	Meimberg 20
M. hyssopifolia var. hyssopifolia	Adeje	Adeje	52	28,14551	-16,74021	11	Meimberg 19
M. hyssopifolia var. hyssopifolia	South	Southcenter	53	28,17351	-16,65225	3	Meimberg 4
M. hyssopifolia var. hyssopifolia	South	Southcenter	54	28,14278	-16,65101	3	Meimberg 5
M. hyssopifolia var. hyssopifolia	Southeast	Southeast	55	28,20875	-16,5392	10	Meimberg 22
M. hyssopifolia var. hyssopifolia	Southeast	Southeast	56	28,20578	-16,53791	3	Meimberg 23
M. hyssopifolia var. hyssopifolia	Southeast	Southeast	57	28,19777	-16,53239	2	Meimberg 24
M. hyssopifolia var. hyssopifolia	Southeast	Southeast	58	28,16728	-16,50658	9	Meimberg 21
M. hyssopifolia var. hyssopifolia	Southeast	Southeast	65	28,23511	-16,48109	5	Meimberg 9
M. hyssopifolia var. glabrescens	Teno	Lomo Morin, Teno	34	28,35981	-16,78911	2	Puppo 562 (TFC)
M. hyssopifolia var. hyssopifolia	Teno	Teno	61	28,30052	-16,82512	10	Meimberg 30
M. hyssopifolia var. hyssopifolia	Teno	Teno	62	28,26191	-16,82088	3	Meimberg 1
M. hyssopifolia var. hyssopifolia	Teno	Teno	63	28,28123	-16,81356	5	Meimberg 17
M. hyssopifolia var. hyssopifolia	Teno	nr Teno	64	28,2516	-16,81083	4	Meimberg 2
Total						289	

Table S2. Results for HWE and Bottleneck test per population. Here we present the number of loci deviating from HWE and the p-value for deviations the from mutation-drift equilibrium (Bottleneck).

Population	Species	Region	nr. Loci deviating from HWE	Bottleneck p-value
3	*M. teneriffae*	Anaga	1	0.330
5	*M. teneriffae*	Southern Coast	1	0.593
7	*M. glomerata*	Anaga	3	0.039
8	*M. rivas-martninezii*	Anaga	5	0.219
9	*M. densiflora*	Teno	0	0.031
10	*M. lasiophylla*	Teide	0	0.352
11	*M. varia*	Anaga	2	0.389
12	*M. varia*	Anaga	5	0.580
13	*M. varia*	Anaga	0	0.365
14	*M. varia*	Anaga	2	0.572
21	*M. varia*	Anaga	0	0.068
23	*M. varia*	Teno	1	0.195
24	*M. varia*	Teno	0	0.014
25	*M. varia*	Teno	1	0.190
30	*M. lachnophylla*	Teide	3	0.225
45	*M. hyssopifolia*	Southern Coast	0	0.439
47	*M. hyssopifolia*	North east	2	0.461
48	*M. hyssopifolia*	North east	2	0.109
50	*M. hyssopifolia*	North west	2	0.499
51	*M. hyssopifolia*	Adeje	2	0.407
52	*M. hyssopifolia*	Adeje	5	0.573
55	*M. hyssopifolia*	Southeast	4	0.573
58	*M. hyssopifolia*	Southeast	3	0.039
59	*M. hyssopifolia*	Southern Coast	2	0.377
60	*M. hyssopifolia*	Southern Coast	1	0.318
61	*M. hyssopifolia*	Teno	3	0.390
63	*M. hyssopifolia*	Teno	4	0.532
65	*M. hyssopifolia*	Southeast	3	0.144

Appendix IV: Supplementary material from section 3.3.

Curto M., Puppo P., Meimberg H. (submitted) Genetic diversity and differentiation patterns of *Micromeria* from the Canary Islands are congruent with multiple colonization dynamics and the establishment of species syngameons.

Table S1. List of all localities with information regarding: the number of individuals sampled; species classification according to Puppo and Meimberg (2015); coordinates.

Pop	Nr. Samples	Species	Island	UTM Zone 28N Y	UTM Zone 28N X
Laru1	3	*M. mahanensis*	Lanzarote	*641246*	*3215140*
Laru2	8	*M. mahanensis*	Lanzarote	*647652*	*3232688*
Laru3	1	*M. mahanensis*	Lanzarote		
GCbe42	11	*M. benthamii*	Gran Canaria	*443984*	*3087082*
GCbe43	4	*M. benthamii*	Gran Canaria	*444738*	*3085125*
GCbe44	8	*M. benthamii*	Gran Canaria	*438613*	*3095957*
GCbe46	16	*M. benthamii*	Gran Canaria	*437917*	*3091772*
GCbe47	10	*M. benthamii*	Gran Canaria	*434909*	*3100327*
GCbe48	3	*M. benthamii*	Gran Canaria	*432428*	*3086161*
GCbe49	6	*M. benthamii*	Gran Canaria	*432543*	*3100595*
GCbe50	7	*M. benthamii*	Gran Canaria	*432653*	*3085908*
GCbe51	3	*M. benthamii*	Gran Canaria	*437537*	*3090884*
GCbe52	1	*M. benthamii*	Gran Canaria	*440143*	*3103983*
GCbe53	9	*M. benthamii*	Gran Canaria	*443891*	*3087012*
GCbe54	10	*M. benthamii*	Gran Canaria	*447591*	*3092941*
GCbe55	6	*M. benthamii*	Gran Canaria	*449898*	*3092547*
GCbeNX52	9	*M. benthamii*	Gran Canaria	*440143*	*3103983*
GCca1	11	*M. canariensis*	Gran Canaria	*426219*	*3089404*
GCca10	6	*M. canariensis*	Gran Canaria	*441064*	*3109656*
GCca11	4	*M. canariensis*	Gran Canaria	*441624*	*3107616*
GCca12	10	*M. canariensis*	Gran Canaria	*433302*	*3084706*
GCca13	6	*M. canariensis*	Gran Canaria	*453948*	*3094118*
GCca14	3	*M. canariensis*	Gran Canaria	*430545*	*3087138*
GCca15	6	*M. canariensis*	Gran Canaria	*430547*	*3087138*
GCca2	3	*M. canariensis*	Gran Canaria	*433440*	*3085286*
GCca3	11	*M. canariensis*	Gran Canaria	*439861*	*3103791*
GCca31	8	*M. canariensis*	Gran Canaria	*428604*	*3088307*
GCca4	4	*M. canariensis*	Gran Canaria	*441607*	*3107648*
GCca42	7	*M. canariensis*	Gran Canaria	*443984*	*3087082*
GCca43	7	*M. canariensis*	Gran Canaria	*444738*	*3085125*
GCca44	5	*M. canariensis*	Gran Canaria	*438613*	*3095955*
GCca5	1	*M. canariensis*	Gran Canaria	*441732*	*3107161*
GCca56	12	*M. canariensis*	Gran Canaria	*442558*	*3076212*
GCca57	1	*M. canariensis*	Gran Canaria	*442847*	*3077129*
GCca58	2	*M. canariensis*	Gran Canaria	*443111*	*3077399*
GCca6	1	*M. canariensis*	Gran Canaria	*442868*	*3077191*
GCca67	1	*M. canariensis*	Gran Canaria	*443042*	*3077314*
GCca7	3	*M. canariensis*	Gran Canaria	*443147*	*3077353*
GCca8	2	*M. canariensis*	Gran Canaria	*443168*	*3077306*
GCca9	2	*M. canariensis*	Gran Canaria	*443294*	*3077235*
GCcaNew	5	*M. canariensis*	Gran Canaria	*443042*	*3077314*
GChl56	8	*M. helianthemifolia*	Gran Canaria	*442558*	*3076212*
GChl57	6	*M. helianthemifolia*	Gran Canaria	*443042*	*3077314*
GChl58	1	*M. helianthemifolia*	Gran Canaria	*443111*	*3077399*
GChl59	1	*M. helianthemifolia*	Gran Canaria	*443072*	*3077385*
GChl60	8	*M. helianthemifolia*	Gran Canaria	*443114*	*3077477*
GChl61	1	*M. helianthemifolia*	Gran Canaria	*443126*	*3077400*
GCln29	3	*M. lanata*	Gran Canaria	*431175*	*3087367*
GCln31	1	*M. lanata*	Gran Canaria	*428604*	*3088307*
GCln32	1	*M. lanata*	Gran Canaria	*431709*	*3103135*
GCln33	2	*M. lanata*	Gran Canaria	*432162*	*3088094*

Table S1. Continued

Pop	Nr. Samples	Species	Island	UTM Zone 28N Y	UTM Zone 28N X
GCln34	1	*M. lanata*	Gran Canaria	*432227*	*3088126*
GCln35	3	*M. lanata*	Gran Canaria	*432320*	*3087923*
GCln36	3	*M. lanata*	Gran Canaria	*432371*	*3087366*
GCln37	3	*M. lanata*	Gran Canaria	*433443*	*3089414*
GCln38	2	*M. lanata*	Gran Canaria	*437890*	*3092001*
GCln39	8	*M. lanata*	Gran Canaria	*438415*	*3092818*
GCln40	1	*M. lanata*	Gran Canaria	*440461*	*3093480*
GCln41	3	*M. lanata*	Gran Canaria	*441247*	*3089747*
GCln44	9	*M. lanata*	Gran Canaria	*438613*	*3095957*
GCln46	13	*M. lanata*	Gran Canaria	*437917*	*3091772*
GCln47	13	*M. lanata*	Gran Canaria	*434909*	*3100327*
GClnxbe46	1	*M. lanataxbenthamii*	Gran Canaria	*431603*	*3102671*
GClu65	8	*M. leucantha*	Gran Canaria	*424920*	*3095052*
GClu66	7	*M. leucantha*	Gran Canaria	*427243*	*3096018*
GCpi62	7	*M. pineolens*	Gran Canaria	*432335*	*3103782*
GCpi63	2	*M. pineolens*	Gran Canaria	*432182*	*3103371*
GCpi64	12	*M. pineolens*	Gran Canaria	*432337*	*3103763*
GCtu14	3	*M. tenuis*	Gran Canaria	*430545*	*3087138*
GCtu15	14	*M. tenuis*	Gran Canaria	*430547*	*3087138*
GCtu16	9	*M. tenuis*	Gran Canaria	*424743*	*3100857*
GCtu17	3	*M. tenuis*	Gran Canaria	*424929*	*3095129*
GCtu18	1	*M. tenuis*	Gran Canaria	*427243*	*3096018*
GCtu19	4	*M. tenuis*	Gran Canaria	*427540*	*3082909*
GCtu20	3	*M. tenuis*	Gran Canaria	*430567*	*3086681*
GCtu21	1	*M. tenuis*	Gran Canaria	*430583*	*3086663*
GCtu22	10	*M. tenuis*	Gran Canaria	*431131*	*3107660*
GCtu23	9	*M. tenuis*	Gran Canaria	*434508*	*3082204*
GCtu24	1	*M. tenuis*	Gran Canaria	*432071*	*3096244*
GCtu25	1	*M. tenuis*	Gran Canaria	*432078*	*3096250*
GCtu26	2	*M. tenuis*	Gran Canaria	*429302*	*3085088*
GCtu27	9	*M. tenuis*	Gran Canaria	*441956*	*3113102*
GCtu28	5	*M. tenuis*	Gran Canaria	*428854*	*3086184*
GCtu29	10	*M. tenuis*	Gran Canaria	*431332*	*3096190*
GCtu30	5	*M. tenuis*	Gran Canaria	*432043*	*3096062*
GCtu31	13	*M. tenuis*	Gran Canaria	*428604*	*3088307*
GCtuxca14	1	*M. tenuis x canariensis*	Gran Canaria	*430545*	*3087138*
LGgo11	2	*M. gomerensis*	La Gomera	*272067*	*3114002*
LGgo12	4	*M. gomerensis*	La Gomera	*275672*	*3113273*
LGgo5	4	*M. gomerensis*	La Gomera	*285344*	*3110162*
LGgo6	1	*M. gomerensis*	La Gomera	*286523*	*3110078*
LGgo9	3	*M. gomerensis*	La Gomera	*283049*	*3113035*
LGlp1	4	*M. lepida*	La Gomera	*279844*	*3111428*
LGlp2	8	*M. lepida*	La Gomera	*280172*	*3111082*
LGlp3	4	*M. lepida*	La Gomera	*282195*	*3111442*
LGlp4	4	*M. lepida*	La Gomera	*282923*	*3108046*
LGlp5	4	*M. lepida*	La Gomera	*285344*	*3110162*
LGlp6	3	*M. lepida*	La Gomera	*286523*	*3110078*
LGlpxpl3	1	*M. lepida x pedro-luisii*	La Gomera	*282195*	*3111442*
LGlpxpl4	2	*M. lepida x pedro-luisii*	La Gomera	*282923*	*3108046*
LGlpxpl6	6	*M. lepida x pedro-luisii*	La Gomera	*286523*	*3110078*
LGpl10	4	*M. pedro-luisii*	La Gomera	*279413*	*3117631*

Table S1. Continued

Pop	Nr. Samples	Species	Island	UTM Zone 28N Y	UTM Zone 28N X
LGpl11	4	*M. pedro-luisii*	La Gomera	*279413*	*3114002*
LGpl13	4	*M. pedro-luisii*	La Gomera	*279102*	*3118023*
LGpl4	1	*M. pedro-luisii*	La Gomera	*282923*	*3108046*
LGpl6	4	*M. pedro-luisii*	La Gomera	*286523*	*3110078*
LGpl7	2	*M. pedro-luisii*	La Gomera	*282972*	*3119896*
LGpl8	4	*M. pedro-luisii*	La Gomera	*286742*	*3113200*
LGpl9	3	*M. pedro-luisii*	La Gomera	*283049*	*3113035*
Tede62	4	*M. densiflora*	Tenerife	*316554*	*3137590*
Tedexva62	1	*M. varia x densiflora*	Tenerife	*316554*	*3137590*
Tegl3	5	*M. glomerata*	Tenerife	*381771*	*3160144*
Tehy10	1	*M. hyssopifolia*	Tenerife	*362295*	*3130619*
Tehy15	10	*M. hyssopifolia*	Tenerife	*321033*	*3131845*
Tehy16	2	*M. hyssopifolia*	Tenerife	*336907*	*3142444*
Tehy17	6	*M. hyssopifolia*	Tenerife	*344288*	*3142006*
Tehy27	3	*M. hyssopifolia*	Tenerife	*321385*	*3127560*
Tehy28	5	*M. hyssopifolia*	Tenerife	*322135*	*3129690*
Tehy29	4	*M. hyssopifolia*	Tenerife	*322353*	*3126403*
Tehy30	3	*M. hyssopifolia*	Tenerife	*323741*	*3115933*
Tehy31	2	*M. hyssopifolia*	Tenerife	*324644*	*3138283*
Tehy32	5	*M. hyssopifolia*	Tenerife	*327335*	*3124584*
Tehy33	1	*M. hyssopifolia*	Tenerife	*327338*	*3124500*
Tehy34	12	*M. hyssopifolia*	Tenerife	*329114*	*3114546*
Tehy35	8	*M. hyssopifolia*	Tenerife	*335984*	*3105056*
Tehy36	2	*M. hyssopifolia*	Tenerife	*337619*	*3138495*
Tehy37	6	*M. hyssopifolia*	Tenerife	*337795*	*3117528*
Tehy38	2	*M. hyssopifolia*	Tenerife	*337931*	*3140511*
Tehy39	2	*M. hyssopifolia*	Tenerife	*339687*	*3141439*
Tehy40	1	*M. hyssopifolia*	Tenerife	*340823*	*3142100*
Tehy41	2	*M. hyssopifolia*	Tenerife	*340981*	*3140280*
Tehy44	10	*M. hyssopifolia*	Tenerife	*348944*	*3121286*
Tehy45	3	*M. hyssopifolia*	Tenerife	*349067*	*3120956*
Tehy46	2	*M. hyssopifolia*	Tenerife	*349598*	*3120061*
Tehy47	5	*M. hyssopifolia*	Tenerife	*349764*	*3134750*
Tehy48	11	*M. hyssopifolia*	Tenerife	*350562*	*3136795*
Tehy49	3	*M. hyssopifolia*	Tenerife	*351094*	*3118406*
Tehy50	10	*M. hyssopifolia*	Tenerife	*352089*	*3116651*
Tehy51	4	*M. hyssopifolia*	Tenerife	*352942*	*3138152*
Tehy52	6	*M. hyssopifolia*	Tenerife	*353541*	*3114975*
Tehy53	5	*M. hyssopifolia*	Tenerife	*360114*	*3139313*
Tehy54	6	*M. hyssopifolia*	Tenerife	*360141*	*3120233*
Tehy55	3	*M. hyssopifolia*	Tenerife	*360169*	*3120299*
Tehy56	5	*M. hyssopifolia*	Tenerife	*361935*	*3125040*
Tehy57	6	*M. hyssopifolia*	Tenerife	*364598*	*3131653*
Tehy7	2	*M. hyssopifolia*	Tenerife	*358912*	*3125710*
Tehy9	5	*M. hyssopifolia*	Tenerife	*361330*	*3122771*
Telc42	2	*M. lachnophylla*	Tenerife	*345654*	*3133206*
Telc43	2	*M. lachnophylla*	Tenerife	*345999*	*3132925*
Telc60	8	*M. lachnophylla*	Tenerife	*346345*	*3132424*
Telc61	12	*M. lachnophylla*	Tenerife	*346351*	*3131823*
Tels58	9	*M. lasiophylla*	Tenerife	*345354*	*3124417*
Tels59	6	*M. lasiophylla*	Tenerife	*348456*	*3124501*

Table S1. Continued

Pop	Nr. Samples	Species	Island	UTM Zone 28N Y	UTM Zone 28N X
Term x va1	2	*M. rivas-martinezii x varia***	Tenerife	*389591*	*3157861*
Term1	9	*M. rivas-martinezii*	Tenerife	*389591*	*3157861*
Term2	2	*M. rivas-martinezii*	Tenerife	*388725*	*3157862*
Tetf10	3	*M. teneriffae*	Tenerife	*362295*	*3130619*
Tetf4	4	*M. teneriffae*	Tenerife	*376797*	*3153040*
Tetf5	4	*M. teneriffae*	Tenerife	*361115*	*3122266*
Tetf6	1	*M. teneriffae*	Tenerife	*376257*	*3153686*
Tetf7	5	*M. teneriffae*	Tenerife	*358912*	*3125710*
Tetf9	4	*M. teneriffae*	Tenerife	*361330*	*3122771*
Tetfxhy10	3	*M. teneriffae x hyssopifolia*	Tenerife	*362295*	*3130619*
Tetfxhy8	2	*M. teneriffae x hyssopifolia*	Tenerife	*358912*	*3125710*
Tetfxva6	2	*M. teneriffae x varia*	Tenerife	*376257*	*3153686*
Teva11	3	*M. varia*	Tenerife	*316573*	*3137268*
Teva12	8	*M. varia*	Tenerife	*317536*	*3136504*
Teva13	10	*M. varia*	Tenerife	*318033*	*3134928*
Teva14	5	*M. varia*	Tenerife	*318469*	*3140302*
Teva18	7	*M. varia*	Tenerife	*376876*	*3156804*
Teva19	4	*M. varia*	Tenerife	*378916*	*3159058*
Teva2	8	*M. varia*	Tenerife	*388725*	*3157862*
Teva20	7	*M. varia*	Tenerife	*380957*	*3158289*
Teva21	5	*M. varia*	Tenerife	*381494*	*3158461*
Teva22	2	*M. varia*	Tenerife	*381711*	*3160119*
Teva23	7	*M. varia*	Tenerife	*382079*	*3157979*
Teva24	2	*M. varia*	Tenerife	*383166*	*3156454*
Teva26	10	*M. varia*	Tenerife	*387643*	*3158473*
Teva6	6	*M. varia*	Tenerife	*376257*	*3153686*
Teva62	4	*M. varia*	Tenerife	*316554*	*3137590*
LPhp1	4	*M. herpyllomorpha*	La Palma	*218227*	*3183546*
LPhp2	2	*M. herpyllomorpha*	La Palma	*218506*	*3183953*
LPhp3	4	*M. herpyllomorpha*	La Palma	*220098*	*3184898*
LPhp4	4	*M. herpyllomorpha*	La Palma	*223206*	*3181300*
LPhp5	4	*M. herpyllomorpha*	La Palma	*218919*	*3164257*
LPhp5a	4	*M. herpyllomorpha*	La Palma	*213170*	*3173527*
LPhp6	2	*M. herpyllomorpha*	La Palma	*222272*	*3152723*
LPhp7	2	*M. herpyllomorpha*	La Palma	*226161*	*3162465*
LPhp8	3	*M. herpyllomorpha*	La Palma	*229510*	*3185748*
LPhp9	3	*M. herpyllomorpha*	La Palma	*230702*	*3178927*
EHhi1	4	*M. hierrensis*	El Hierro	*204428*	*3071541*
EHhi2	4	*M. hierrensis*	El Hierro	*205015*	*3061895*
EHhi3	4	*M. hierrensis*	El Hierro	*206275*	*3069905*
EHhi4	4	*M. hierrensis*	El Hierro	*214503*	*3079961*
EHhi5	4	*M. hierrensis*	El Hierro	*206417*	*3069310*
EHhi6	4	*M. hierrensis*	El Hierro	*202356*	*3070599*
EHhi7	4	*M. hierrensis*	El Hierro	*204034*	*3073369*

Table S2.1. Migration rates among species groups calculated with BaysAss. Values correspond to percentage of individuals originated from species in the left column present in species in the upper line. The second value corresponds to the 95% confidence interval. Migration rates significantly higher than zero are marked in bold.

Source/Sink		Lanzarote	Gran Canaria							La Gomera			
		M. mahanensis	***M. benthamii***	***M. canariensis***	***M. helianthemifolia***	***M. lanata***	***M. leucantha***	***M. pineolens***	***M. tenuis***	***M. gomerensis***	***M. lepida***	***M. lepida x M. pedro-luisii***	***M. pedro-luisii***
Lanzarote	***M. mahanensis***	67,74 ± 1,02	0,27 ± 0,26	0,23 ± 0,23	0,64 ± 0,63	0,37 ± 0,37	0,81 ± 0,79	0,75 ± 0,74	0,27 ± 0,26	0,84 ± 0,81	0,67 ± 0,65	0,98 ± 0,94	0,68 ± 0,66
Gran Canaria	***M. benthamii***	0,89 ± 0,86	93,12 ± 1,23	0,41 ± 0,35	**1,33 ± 1,09**	**1,51 ± 0,81**	**14,91 ± 2,81**	**2,56 ± 1,5**	**0,97 ± 0,55**	0,85 ± 0,82	0,67 ± 0,65	0,98 ± 0,94	0,68 ± 0,66
	M. canariensis	**13,57 ± 2,77**	**0,83 ± 0,5**	94,12 ± 1,07	0,66 ± 0,64	0,43 ± 0,43	0,91 ± 0,86	0,76 ± 0,74	**2,73 ± 0,95**	**13,04 ± 2,67**	0,68 ± 0,67	0,98 ± 0,94	0,68 ± 0,66
	M. helianthemifolia	0,89 ± 0,86	0,27 ± 0,26	0,23 ± 0,22	67,41 ± 0,72	0,37 ± 0,37	0,81 ± 0,78	0,76 ± 0,75	0,27 ± 0,26	0,85 ± 0,82	0,66 ± 0,66	0,98 ± 0,94	0,68 ± 0,67
	M. lanata	0,89 ± 0,86	0,3 ± 0,29	0,23 ± 0,23	**1,23 ± 0,91**	89,9 ± 1,67	0,99 ± 0,93	0,76 ± 0,75	**0,71 ± 0,52**	0,85 ± 0,82	0,67 ± 0,66	0,98 ± 0,94	0,7 ± 0,68
	M. leucantha	0,89 ± 0,86	0,27 ± 0,27	0,22 ± 0,22	0,65 ± 0,63	0,37 ± 0,37	67,63 ± 0,92	0,76 ± 0,74	0,26 ± 0,26	0,84 ± 0,81	0,67 ± 0,66	0,97 ± 0,93	0,67 ± 0,67
	M. pineolens	0,89 ± 0,86	0,31 ± 0,3	0,23 ± 0,23	0,65 ± 0,63	0,37 ± 0,37	0,82 ± 0,79	81,51 ± 2,57	0,27 ± 0,27	0,85 ± 0,82	0,67 ± 0,67	0,97 ± 0,94	0,68 ± 0,66
	M. tenuis	0,89 ± 0,86	**0,64 ± 0,44**	**0,87 ± 0,47**	**16,51 ± 2,57**	**1,07 ± 0,67**	0,94 ± 0,9	0,78 ± 0,76	90,49 ± 1,47	0,84 ± 0,81	0,66 ± 0,65	0,97 ± 0,93	0,68 ± 0,66
La Gomera	***M. gomerensis***	0,89 ± 0,85	0,26 ± 0,26	0,23 ± 0,22	0,64 ± 0,63	0,37 ± 0,37	0,81 ± 0,79	0,76 ± 0,74	0,27 ± 0,26	67,67 ± 0,96	0,66 ± 0,64	0,97 ± 0,93	0,68 ± 0,67
	M. lepida	0,89 ± 0,85	0,28 ± 0,28	0,23 ± 0,22	0,86 ± 0,78	0,4 ± 0,4	0,82 ± 0,79	0,75 ± 0,74	0,28 ± 0,28	**2,36 ± 1,37**	85,32 ± 2,32	**10,7 ± 2,86**	**1,21 ± 0,94**
	M. lepida x M. pedro-luisii	0,9 ± 0,86	0,27 ± 0,26	0,23 ± 0,23	0,65 ± 0,63	0,37 ± 0,36	0,81 ± 0,79	0,76 ± 0,74	0,27 ± 0,27	0,84 ± 0,81	0,66 ± 0,65	67,87 ± 1,13	0,67 ± 0,66
	M. pedro-luisii	0,89 ± 0,86	0,27 ± 0,26	0,23 ± 0,22	0,65 ± 0,64	0,38 ± 0,38	0,81 ± 0,78	0,75 ± 0,74	0,27 ± 0,27	0,84 ± 0,81	0,66 ± 0,65	**1,91 ± 1,29**	84,15 ± 2,41
Tenerife	***M. densiflora***	0,89 ± 0,85	0,26 ± 0,26	0,23 ± 0,23	0,65 ± 0,63	0,37 ± 0,37	0,82 ± 0,79	0,76 ± 0,74	0,27 ± 0,26	0,84 ± 0,81	0,67 ± 0,65	0,98 ± 0,94	0,68 ± 0,66
	M. glomerata	0,89 ± 0,86	0,26 ± 0,26	0,23 ± 0,23	0,65 ± 0,63	0,37 ± 0,37	0,81 ± 0,79	0,76 ± 0,74	0,26 ± 0,26	0,84 ± 0,81	0,67 ± 0,65	0,99 ± 0,95	0,67 ± 0,66
	M. hyssopifolia	0,89 ± 0,86	0,26 ± 0,27	0,23 ± 0,23	0,65 ± 0,63	0,37 ± 0,36	0,81 ± 0,78	0,75 ± 0,73	0,27 ± 0,26	0,85 ± 0,82	0,66 ± 0,65	0,97 ± 0,94	0,75 ± 0,73
	M. lachnophylla	0,89 ± 0,86	0,27 ± 0,26	0,23 ± 0,23	0,65 ± 0,63	0,37 ± 0,37	0,81 ± 0,79	0,75 ± 0,73	0,26 ± 0,26	0,84 ± 0,81	0,66 ± 0,65	0,97 ± 0,93	0,68 ± 0,66
	M. lasiophylla	0,89 ± 0,86	0,27 ± 0,26	0,23 ± 0,23	0,64 ± 0,63	0,36 ± 0,36	0,81 ± 0,79	0,76 ± 0,74	0,27 ± 0,27	0,85 ± 0,82	0,66 ± 0,65	0,97 ± 0,93	0,69 ± 0,67
	M. rivas-martinezii	0,89 ± 0,86	0,26 ± 0,26	0,23 ± 0,23	0,65 ± 0,63	0,38 ± 0,37	0,81 ± 0,79	0,76 ± 0,73	0,27 ± 0,27	0,84 ± 0,81	0,67 ± 0,65	0,97 ± 0,93	0,68 ± 0,66
	M. teneriffae	0,89 ± 0,86	0,27 ± 0,26	0,23 ± 0,23	0,64 ± 0,62	0,37 ± 0,37	0,81 ± 0,78	0,76 ± 0,75	0,27 ± 0,27	0,85 ± 0,82	0,67 ± 0,65	0,97 ± 0,93	0,77 ± 0,74
	M. varia Anaga	0,89 ± 0,86	0,27 ± 0,26	0,23 ± 0,23	0,65 ± 0,64	0,36 ± 0,36	0,81 ± 0,78	0,76 ± 0,74	0,27 ± 0,26	0,87 ± 0,84	0,67 ± 0,65	0,98 ± 0,94	0,75 ± 0,73
	M. varia Teno	0,89 ± 0,86	0,26 ± 0,26	0,23 ± 0,23	0,65 ± 0,63	0,37 ± 0,36	0,81 ± 0,78	0,76 ± 0,74	0,26 ± 0,26	0,84 ± 0,81	0,66 ± 0,65	0,98 ± 0,93	0,68 ± 0,67
La Palma	***M. herpyllomorpha***	0,89 ± 0,86	0,27 ± 0,27	0,23 ± 0,23	0,65 ± 0,63	0,38 ± 0,38	0,81 ± 0,78	0,77 ± 0,75	0,27 ± 0,27	0,84 ± 0,82	0,68 ± 0,66	0,98 ± 0,94	0,74 ± 0,71
El Hierro	***M. hierrensis***	0,89 ± 0,86	0,27 ± 0,27	0,23 ± 0,23	**1,64 ± 1,01**	0,37 ± 0,36	0,81 ± 0,79	0,75 ± 0,73	0,29 ± 0,29	0,85 ± 0,82	0,67 ± 0,66	0,98 ± 0,94	0,75 ± 0,73

Table S2.1. Continued

	Source/Sink	Tenerife									La Palma	El Hierro
		M. densiflora	***M. glomerata***	***M. hyssopifolia***	***M. lachnophylla***	***M. lasiophylla***	***M. rivas-martinezii***	***M. teneriffae***	***M. varia Anaga***	***M. varia Teno***	***M. herpyllomorpha***	***M. hierrensis***
Lanzarote	***M. mahanensis***	1,13 ± 1,09	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,64	0,81 ± 0,79	0,86 ± 0,83	0,64 ± 0,63	0,41 ± 0,41	0,58 ± 0,57	0,61 ± 0,59	0,66 ± 0,65
Gran Canaria	***M. benthamii***	1,14 ± 1,09	1,14 ± 1,08	0,18 ± 0,18	0,66 ± 0,64	0,82 ± 0,79	0,86 ± 0,84	0,65 ± 0,65	0,41 ± 0,41	0,59 ± 0,58	0,6 ± 0,59	0,65 ± 0,63
	M. canariensis	1,13 ± 1,09	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,64	0,81 ± 0,79	0,87 ± 0,83	0,65 ± 0,64	0,41 ± 0,41	0,58 ± 0,57	0,6 ± 0,59	0,65 ± 0,64
	M. helianthemifolia	1,13 ± 1,08	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,65	0,8 ± 0,78	0,86 ± 0,83	0,66 ± 0,64	0,41 ± 0,41	0,59 ± 0,57	0,61 ± 0,6	0,65 ± 0,64
	M. lanata	1,13 ± 1,08	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,65	0,81 ± 0,78	0,87 ± 0,83	0,65 ± 0,64	0,41 ± 0,4	0,59 ± 0,58	0,61 ± 0,6	0,65 ± 0,64
	M. leucantha	1,13 ± 1,08	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,64	0,81 ± 0,79	0,86 ± 0,83	0,65 ± 0,64	0,41 ± 0,41	0,59 ± 0,57	0,61 ± 0,59	0,65 ± 0,64
	M. pineolens	1,14 ± 1,09	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,65	0,82 ± 0,79	0,85 ± 0,82	0,66 ± 0,64	0,41 ± 0,4	0,59 ± 0,58	0,6 ± 0,59	0,65 ± 0,64
	M. tenuis	1,13 ± 1,08	1,13 ± 1,08	0,18 ± 0,17	0,66 ± 0,64	0,81 ± 0,79	0,86 ± 0,83	0,64 ± 0,63	0,41 ± 0,41	0,65 ± 0,63	0,6 ± 0,59	0,66 ± 0,64
La Gomera	***M. gomerensis***	1,13 ± 1,08	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,64	0,81 ± 0,78	0,86 ± 0,83	0,66 ± 0,65	0,41 ± 0,4	0,59 ± 0,58	0,61 ± 0,59	0,66 ± 0,64
	M. lepida	**7,85 ± 3**	1,13 ± 1,07	0,18 ± 0,18	0,67 ± 0,64	0,81 ± 0,78	0,85 ± 0,82	0,65 ± 0,64	0,42 ± 0,41	0,59 ± 0,58	0,61 ± 0,6	0,65 ± 0,63
	M. lepida x M. pedro-luisii	1,14 ± 1,09	1,13 ± 1,08	0,18 ± 0,17	0,66 ± 0,64	0,81 ± 0,79	0,86 ± 0,83	0,65 ± 0,64	0,42 ± 0,41	0,59 ± 0,57	0,61 ± 0,6	0,65 ± 0,64
	M. pedro-luisii	1,14 ± 1,09	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,64	0,82 ± 0,79	0,86 ± 0,83	0,65 ± 0,64	0,41 ± 0,4	0,6 ± 0,58	0,61 ± 0,59	0,66 ± 0,65
Tenerife	***M. densiflora***	68,1 ± 1,33	1,13 ± 1,07	0,18 ± 0,18	0,66 ± 0,64	0,81 ± 0,79	0,85 ± 0,82	0,65 ± 0,64	0,41 ± 0,41	0,59 ± 0,57	0,61 ± 0,6	0,65 ± 0,64
	M. glomerata	1,13 ± 1,07	68,1 ± 1,33	0,18 ± 0,18	0,66 ± 0,64	0,81 ± 0,78	0,86 ± 0,83	0,65 ± 0,64	0,41 ± 0,4	0,59 ± 0,57	0,61 ± 0,6	0,65 ± 0,64
	M. hyssopifolia	1,36 ± 1,27	1,13 ± 1,08	95,73 ± 0,84	**17,38 ± 2,39**	0,81 ± 0,79	0,86 ± 0,83	**7,37 ± 2,7**	0,64 ± 0,6	**19,9 ± 2,25**	0,65 ± 0,64	0,67 ± 0,66
	M. lachnophylla	1,13 ± 1,09	1,13 ± 1,07	0,18 ± 0,18	67,42 ± 0,73	0,82 ± 0,79	0,86 ± 0,83	0,66 ± 0,66	0,41 ± 0,41	0,59 ± 0,57	0,61 ± 0,59	0,66 ± 0,64
	M. lasiophylla	1,14 ± 1,08	1,12 ± 1,07	0,18 ± 0,18	0,66 ± 0,64	67,63 ± 0,92	0,86 ± 0,83	0,66 ± 0,64	0,41 ± 0,41	0,58 ± 0,57	0,61 ± 0,6	0,65 ± 0,63
	M. rivas-martinezii	1,14 ± 1,08	1,13 ± 1,08	0,18 ± 0,18	0,66 ± 0,64	0,81 ± 0,78	67,69 ± 0,98	0,65 ± 0,64	0,41 ± 0,41	0,59 ± 0,57	0,61 ± 0,59	0,66 ± 0,64
	M. teneriffae	1,16 ± 1,12	**8,18 ± 2,97**	0,25 ± 0,24	**1,38 ± 0,93**	**15,3 ± 2,66**	**13,73 ± 2,81**	77,59 ± 3,08	**0,8 ± 0,61**	0,69 ± 0,65	0,65 ± 0,64	0,66 ± 0,64
	M. varia Anaga	1,14 ± 1,09	1,13 ± 1,07	**0,41 ± 0,31**	0,68 ± 0,66	0,81 ± 0,78	**1,43 ± 1,11**	**1,95 ± 1,35**	90,3 ± 1,73	**0,85 ± 0,74**	0,61 ± 0,6	0,66 ± 0,65
	M. varia Teno	1,13 ± 1,08	1,14 ± 1,08	0,18 ± 0,18	0,66 ± 0,63	0,81 ± 0,79	0,85 ± 0,83	0,65 ± 0,64	0,41 ± 0,41	67,33 ± 0,65	0,6 ± 0,59	0,65 ± 0,64
La Palma	***M. herpyllomorpha***	1,13 ± 1,08	1,13 ± 1,07	0,19 ± 0,19	0,66 ± 0,64	0,81 ± 0,78	0,85 ± 0,83	0,69 ± 0,68	0,42 ± 0,41	0,59 ± 0,57	86,53 ± 2,19	0,66 ± 0,65
El Hierro	***M. hierrensis***	1,13 ± 1,08	1,13 ± 1,08	0,2 ± 0,2	**1,25 ± 0,89**	0,82 ± 0,79	0,85 ± 0,82	0,65 ± 0,64	0,42 ± 0,42	0,58 ± 0,57	0,63 ± 0,62	85,6 ± 2,29

Table S2.2. Migration rates among island groups calculated with BaysAss. Values correspond to percentage of individuals originated from islands in the left column present in islands in the upper line. The second value corresponds to the 95% confidence interval. Migration rates significantly higher than zero are marked in bold.

Source/Sink	Lanzarote	Gran Canaria	La Gomera	Tenerife	La Palma	El Hierro
Lanzarote	68,63 ± 1,79	0,07 ± 0,07	0,41 ± 0,4	0,1 ± 0,1	0,88 ± 0,86	0,98 ± 0,95
Gran Canaria	1,45 ± 1,39	99,37 ± 0,25	0,42 ± 0,41	0,1 ± 0,1	0,88 ± 0,86	0,98 ± 0,96
La Gomera	25,55 ± 3,09	0,13 ± 0,13	97,26 ± 1,08	0,1 ± 0,1	0,9 ± 0,88	0,98 ± 0,94
Tenerife	1,46 ± 1,4	0,08 ± 0,08	**0,81 ± 0,64**	99,44 ± 0,24	1,2 ± 1,1	1,14 ± 1,11
La Palma	1,45 ± 1,38	0,08 ± 0,08	0,46 ± 0,45	0,14 ± 0,13	95,24 ± 1,94	0,98 ± 0,95
El Hierro	1,45 ± 1,39	0,26 ± 0,18	0,64 ± 0,6	0,13 ± 0,12	0,91 ± 0,89	94,93 ± 2,06

Table S3.1. Migration rate calculated with Migrate among islands. Values correspond to average number of individuals migrating, per generation, from islands in the left column to islands in the upper line. Values with 95% intervals above zero (under brackets) are marked as bold. NA: Not available.

Source/Sink	Lanzarote	Gran Canaria	La Gomera	Tenerife	La Palma	El Hierro
Lanzarote	NA	2.80 (0.00 - 19.33)	6.26 (0.00 - 22.67)	3.65 (0.00 - 20.00)	6.29 (0.00 - 22.67)	3.13 (0.00 - 19.33)
Gran Canaria	16.16 (0.00 - 32.00)	NA	**20.50 (3.33 - 36.67)**	13.00 (0.00 - 28.67)	**32.03 (14.67 - 48.67)**	14.36 (0.00 - 30.00)
La Gomera	5.97 (0.00 - 22.67)	11.22 (0.00 - 27.33)	NA	14.43 (0.00 - 30.00)	9.69 (0.00 - 26.67)	12.26 (0.00 - 28.00)
Tenerife	6.64 (0.00 - 23.33)	4.56 (0.00 - 21.33)	3.75 (0.00 - 20.00)	NA	7.15 (0.00 - 23.33)	7.35 (0.00 - 23.33)
La Palma	10.39 (0.00 - 26.00)	**28.41 (11.33 - 44.67)**	**18.21 (1.33 - 34.67)**	15.85 (0.00 - 31.33)	NA	16.94 (0.00 - 32.67)
El Hierro	4.15 (0.00 - 20.67)	5.73 (0.00 - 22.00)	8.26 (0.00 - 24.00)	8.83 (0.00 - 24.67)	9.58 (0.00 - 25.33)	NA

Table S3.2. Migration rate calculated with Migrate among Gran Canaria species. Values correspond to average number of individuals migrating, per generation, from species in the left column to species in the upper line. Values with 95% intervals above zero (under brackets) are marked as bold. NA: Not available.

Source/Sink	*M. benthamii*	*M. canariensis*	*M. helianthemifolia*	*M. lanata*	*M. leucantha*	*M. pineolens*	*M. tenuis*
M. benthamii	NA	16.12 (0.00 - 48.00)	19.61 (0.00 - 52.00)	26.85 (0.00 - 57.33)	20.00 (0.00 - 52.00)	11.89 (0.00 - 45.33)	25.27 (0.00 - 56.00)
M. canariensis	16.76 (0.00 - 49.33)	NA	9.10 (0.00 - 42.67)	29.87 (0.00 - 60.00)	9.30 (0.00 - 42.67)	10.09 (0.00 - 44.00)	31.30 (0.00 - 62.67)
M. helianthemifolia	5.78 (0.00 - 38.67)	3.38 (0.00 - 36.00)	NA	20.84 (0.00 - 53.33)	13.91 (0.00 - 46.67)	22.46 (0.00 - 53.33)	4.94 (0.00 - 37.33)
M. lanata	17.92 (0.00 - 49.33)	10.87 (0.00 - 44.00)	12.87 (0.00 - 45.33)	NA	26.03 (0.00 - 57.33)	13.27 (0.00 - 45.33)	6.45 (0.00 - 40.00)
M. leucantha	7.83 (0.00 - 41.33)	5.97 (0.00 - 40.00)	16.92 (0.00 - 49.33)	21.05 (0.00 - 54.67)	NA	6.05 (0.00 - 38.67)	11.20 (0.00 - 44.00)
M. pineolens	6.78 (0.00 - 40.00)	4.90 (0.00 - 37.33)	7.94 (0.00 - 41.33)	7.96 (0.00 - 42.67)	5.83 (0.00 - 38.67)	NA	4.05 (0.00 - 37.33)
M. tenuis	29.22 (0.00 - 60.00)	12.77 (0.00 - 46.67)	13.24 (0.00 - 46.67)	19.74 (0.00 - 57.33)	9.32 (0.00 - 42.67)	13.24 (0.00 - 45.33)	NA

Table S3.3. Migration rate calculated with Migrate among La Gomera species. Values correspond to average number of individuals migrating, per generation, from species in the left column to species in the upper line. Values with 95% intervals above zero (under brackets) are marked as bold. NA: Not available.

Source/Sink	*M. gomerensis*	*M. lepida*	*M. pedro-luisii x M. lepida*	*M. pedro-luisii*
M. gomerensis	NA	12.33 (0.00 - 45.33)	20.94 (0.00 - 53.33)	5.00 (0.00 - 37.33)
M. lepida	13.82 (0.00 - 46.67)	NA	23.86 (0.00 - 56.00)	5.41 (0.00 - 38.67)
M. pedro-luisii x M. lepida	4.10 (0.00 - 37.33)	16.51 (0.00 - 49.33)	NA	4.30 (0.00 - 37.33)
M. pedro-luisii	5.11 (0.00 - 38.67)	4.27 (0.00 - 37.33)	15.52 (0.00 - 48.00)	NA

Table S3.4. Migration rate calculated with Migrate among Tenerife species. Values correspond to average number of individuals migrating, per generation, from species in the left column to species in the upper line. Values with 95% intervals above zero (under brackets) are marked as bold. NA: not valieble.

Source/Sink	*M. densiflora*	*M. glomerata*	*M. hyssopifolia*	*M. lachnophylla*	*M. lasiophylla*	*M. rivas-martinezii*	*M. teneriffae*	*M. varia Anaga*	*M. varia Teno*
M. densiflora	NA	13.65 (0.00 - 49.33)	6.18 (0.00 - 38.67)	3.87 (0.00 - 37.33)	4.18 (0.00 - 37.33)	5.09 (0.00 - 37.33)	8.72 (0.00 - 41.33)	4.21 (0.00 - 37.33)	14.22 (0.00 - 48.00)
M. glomerata	18.20 (0.00 - 49.33)	NA	10.70 (0.00 - 42.67)	13.94 (0.00 - 45.33)	4.50 (0.00 - 37.33)	5.31 (0.00 - 38.67)	5.42 (0.00 - 38.67)	7.40 (0.00 - 40.00)	12.32 (0.00 - 44.00)
M. hyssopifolia	13.49 (0.00 - 45.33)	26.55 (0.00 - 57.33)	NA	26.52 (0.00 - 58.67)	13.93 (0.00 - 46.67)	17.07 (0.00 - 49.33)	32.77 (0.00 - 64.00)	17.65 (0.00 - 49.33)	31.67 (0.00 - 62.67)
M. lachnophylla	4.55 (0.00 - 37.33)	7.13 (0.00 - 40.00)	18.17 (0.00 - 50.67)	NA	8.87 (0.00 - 41.33)	19.88 (0.00 - 50.67)	20.07 (0.00 - 52.00)	5.39 (0.00 - 38.67)	16.28 (0.00 - 50.67)
M. lasiophylla	8.72 (0.00 - 41.33)	5.68 (0.00 - 38.67)	15.79 (0.00 - 48.00)	4.90 (0.00 - 37.33)	NA	4.68 (0.00 - 37.33)	8.38 (0.00 - 41.33)	6.55 (0.00 - 40.00)	17.89 (0.00 - 49.33)
M. rivas-martinezii	7.22 (0.00 - 40.00)	5.16 (0.00 - 38.67)	5.34 (0.00 - 38.67)	7.03 (0.00 - 40.00)	5.01 (0.00 - 37.33)	NA	17.63 (0.00 - 49.33)	5.32 (0.00 - 38.67)	14.68 (0.00 - 46.67)
M. teneriffae	15.12 (0.00 - 46.67)	5.48 (0.00 - 38.67)	27.57 (0.00 - 58.67)	24.29 (0.00 - 54.67)	6.64 (0.00 - 40.00)	21.15 (0.00 - 52.00)	NA	14.57 (0.00 - 46.67)	14.37 (0.00 - 46.67)
M. varia Anaga	6.74 (0.00 - 40.00)	5.42 (0.00 - 38.67)	11.06 (0.00 - 44.00)	8.64 (0.00 - 41.33)	5.70 (0.00 - 38.67)	9.00 (0.00 - 41.33)	12.63 (0.00 - 45.33)	NA	8.87 (0.00 - 41.33)
M. varia Teno	15.53 (0.00 - 48.00)	22.18 (0.00 - 56.00)	9.21 (0.00 - 42.67)	14.04 (0.00 - 46.67)	19.72 (0.00 - 56.00)	12.60 (0.00 - 45.33)	9.92 (0.00 - 42.67)	6.53 (0.00 - 40.00)	NA

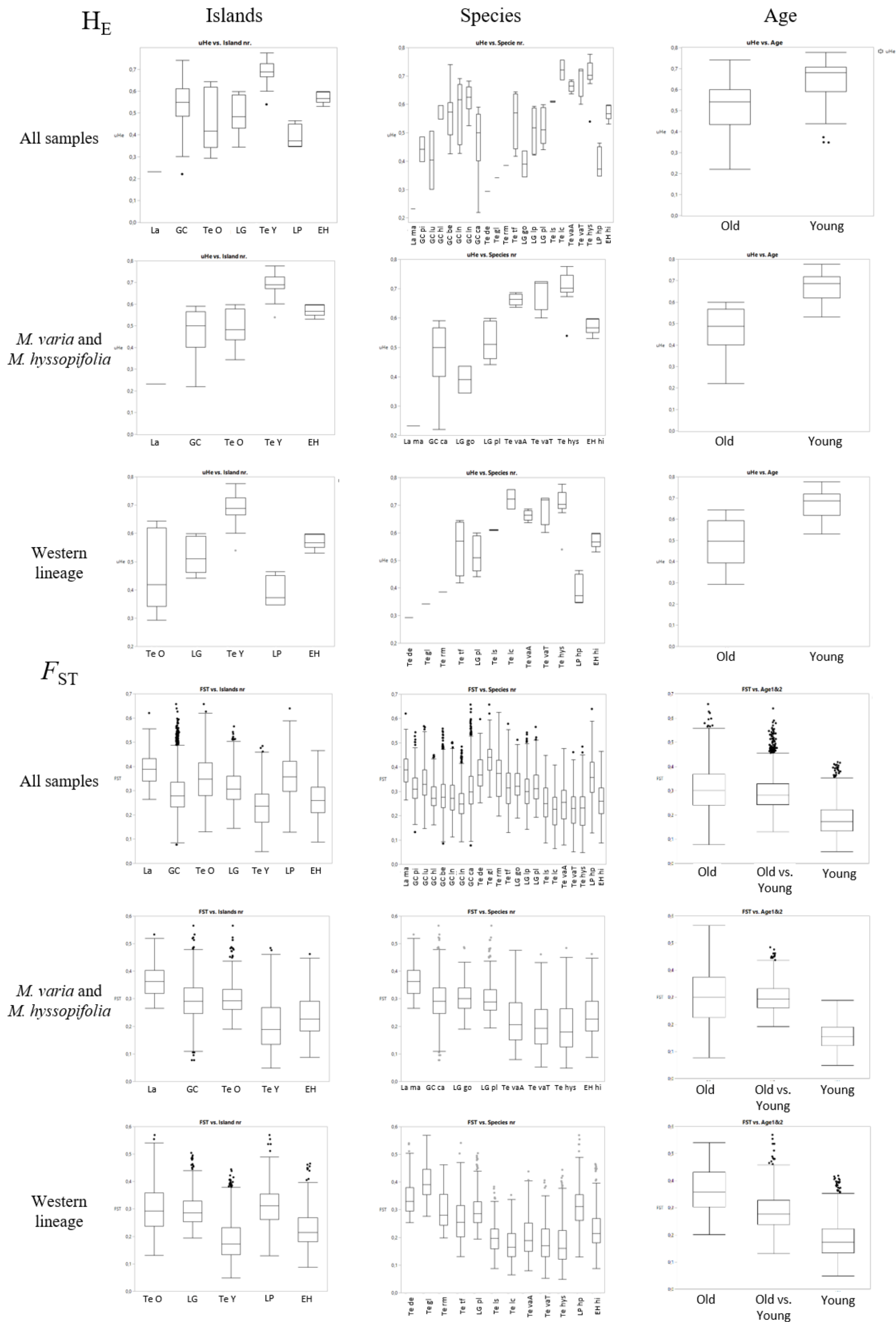


Fig. S1. Representation of expected heterozygosity (He) and F_{ST} per population used as proxy of genetic diversity and differentiation. For each measure there is a graph summarizing it per island, per species, and per age class. From top to bottom the box-plots correspond to the different datasets used: including all samples, .only M. varia and M. hyssopifolia, and Western lineage (Tenerife, EH, LP, and part of LG).

Appendix V: Supplementary material from section 4.2.

Manuel Curto, M., Christina Schachteler, C., Puppo, P., Meimberg, H., (submitted.) Phylogenetic analyses of RAD data to infer the evolution of *Micromeria* in the Canary Islands.

Table S1. Samples used in this study including individual code, islands of origin, species, population code, evaluation of DNA quality (High: strong high molecular weight band; Medium: less strong molecular weight band; Low: faint molecular weight band; Smeared: fragmented DNA), number of reads retained per sample after quality filtering (Nr of final reads), and number of loci produced by PyRAD (Nr of loci). The column "Included in phylogeny" contains information regarding the inclusion of each sample in the final phylogenetic analyses. The last four columns include information regarding the barcode and index used for each sample.

Individual	Island	Species	Population	Sample quality	Nr of final reads	Nr of loci	Included in phylogeny	Barcode P5	Barcode P7	Index P5	Index P7
GCbe_154_1b	Gran Canaria	*M. benthamii*	Gcbe 1	High	108447	850	yes	ATCAT	ATGAT	TATAGCCT / TATAGCCT	GAGATTCC / ATTACTCG
GCbe_155_1C	Gran Canaria	*M. benthamii*	Gcbe 2	High	86483	724	yes	AACGT	ACGTT	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
GChl_432	Gran Canaria	*M. helianthemifolia*	GChl 1	High	23202	200	yes	TCAGA	TCTGA	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
GCv_14_1	Gran Canaria	*M. canariensis*	GCca 1	Medium	53691	444	yes	AGCAT	ATGCT	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
GCv_17_17	Gran Canaria	*M. canariensis*	GCca 2	High	124911	1726	yes	ATCAT	ATGAT	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
GCv_17_19	Gran Canaria	*M. canariensis*	GCca 2	High	21186	231	yes	TACGA	TCGTA	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
GCv_21_2	Gran Canaria	*M. canariensis*	GCca 3	High	25753	200	yes	TCGAA	TTCGA	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
GCv_22_1	Gran Canaria	*M. canariensis*	GCca 3	Medium	172394	921	yes	AACGT	ACGTT	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
GCv_24_1	Gran Canaria	*M. canariensis*	GCca 3	High	41766	329	yes	AGCAT	ATGCT	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
GCv_29_12	Gran Canaria	*M. canariensis*	GCca 4	Medium	57211	519	yes	TCGAA	TTCGA	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
GCv_29_5	Gran Canaria	*M. canariensis*	GCca 4	High	109398	1518	yes	AACGT	ACGTT	TATAGCCT / TATAGCCT	GAGATTCC / ATTACTCG
GCv_333	Gran Canaria	*M. canariensis*	GCca 5	High	76831	700	yes	ACAGT	ACTGT	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG
GCv_340	Gran Canaria	*M. canariensis*	GCca 5	High	57195	762	yes	ACGTT	AACGT	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
GCv_436	Gran Canaria	*M. canariensis*	GCca 6	High	41413	417	yes	AGCAT	ATGCT	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
GCv_437	Gran Canaria	*M. canariensis*	GCca 6	High	103883	1188	yes	ATCAT	ATGAT	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
GCv_454	Gran Canaria	*M. canariensis*	GCca 7	Medium	57539	375	yes	TCAGA	TCTGA	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
GCv_455	Gran Canaria	*M. canariensis*	GCca 7	Medium	77482	635	yes	TCAGA	TCTGA	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
GCv_456	Gran Canaria	*M. canariensis*	GCca 7	High	250095	2924	yes	ACAGT	ACTGT	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
GCv_457	Gran Canaria	*M. canariensis*	GCca 7	High	250173	2792	yes	ACGTT	AACGT	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
GCv_505_1	Gran Canaria	*M. canariensis*	GCca 8	Low	150589	1022	yes	AACGT	ACGTT	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
GCv_505_2	Gran Canaria	*M. canariensis*	GCca 8	Low	36782	292	yes	AGCAT	ATGCT	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
GCv_LF_4	Gran Canaria	*M. canariensis*	GCca 10	High	86084	1069	yes	AACGT	ACGTT	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
GCv_LF3	Gran Canaria	*M. canariensis*	GCca 10	High	124683	1881	yes	AGCAT	ATGCT	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
Lav_MicE	Lanzarote	*M. mahanensis*	Lama 2	Low	32795	246	yes	ACGTT	AACGT	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
Lav_MicH	Lanzarote	*M. mahanensis*	Lama 2	Low	37542	329	yes	ACAGT	ACTGT	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
LGvg_572_1	La Gomera	*M. gomerensis*	LGgo 2	High	163107	1639	yes	AACGT	ACGTT	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG

Table S1. Continued

Individual	Island	Species	Population	Sample quality	Nr of reads	Nr of loci	Included in phylogeny	Barcode P5	Barcode P7	Index P5	Index P7
LGvg_572_4	La Gomera	*M. gomerensis*	LGgo2	High	38484	380	yes	TCAGA	TCTGA	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
LGvg_572_7	La Gomera	*M. gomerensis*	LGgo2	High	22761	210	yes	TTGCA	TGCAA	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
LGvg_572_8	La Gomera	*M. gomerensis*	LGgo2	High	155534	1481	yes	TCAGA	TCTGA	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG
LGvg_579_3	La Gomera	*M. gomerensis*	LGgo3	Medium	66112	495	yes	ATCAT	ATGAT	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
LGvg_579_7	La Gomera	*M. gomerensis*	LGgo3	High	48721	503	yes	TCGAA	TTCGA	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
LGvg_579_8	La Gomera	*M. gomerensis*	LGgo3	High	232869	2278	yes	ATCAT	ATGAT	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG
LGvv_564_10	La Gomera	*M. pedro-luisii*	LGpl1	High	68669	424	yes	ACGTT	AACGT	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
LGvv_565_2	La Gomera	*M. pedro-luisii*	LGpl2	High	117891	970	yes	ATCAT	ATGAT	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
LGvv_583_1	La Gomera	*M. gomerensis*	LGgo4	High	25240	247	yes	TCAGA	TCTGA	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
LGvv_583_3	La Gomera	*M. gomerensis*	LGgo4	High	247869	3021	yes	AGCAT	ATGCT	TATAGCCT / TATAGCCT	GAGATTCC / ATTACTCG
LGvvxlp_570_6	La Gomera	*M. lepida x M. pedro-luisii*	LGplxlp1	High	221429	2450	yes	ACAGT	ACTGT	TATAGCCT / TATAGCCT	GAGATTCC / ATTACTCG
LPla_31_A	La Gomera	*M. herpyllomorpha*	LPhe1	High	66990	589	yes	TACGA	TCGTA	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG
LPla_31_B	La Gomera	*M. herpyllomorpha*	LPhe1	Medium	69519	362	yes	AACGT	ACGTT	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
Tela_277	La Gomera	*M. lasiophylla*	Tela1	High	160899	1230	yes	AGCAT	ATGCT	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG
Tev_26_3	La Gomera	*M. varia*	Tev5	High	99753	713	yes	ACAGT	ACTGT	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
Tev_31_10	La Gomera	*M. varia*	Tev7	Medium	58188	292	yes	TCAGA	TCTGA	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
Tev_31_6	La Gomera	*M. varia*	Tev7	Medium	72130	310	yes	ACAGT	ACTGT	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
Tev_31_8	La Gomera	*M. varia*	Tev7	Medium	54849	274	yes	AACGT	ACGTT	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
Tev_563_7	La Gomera	*M. hyssopifolia*	Tehy2	Medium	128497	406	yes	ACGTT	AACGT	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
GCv_453	Gran Canaria	*M. canariensis*	GCca7	High	18190	138	d3	TTGCA	TGCAA	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
GCv_L509	Gran Canaria	*M. canariensis*	GCca9	High	17000	102	d3	ACGTT	AACGT	TATAGCCT / TATAGCCT	GAGATTCC / ATTACTCG
GCv_LF1	Gran Canaria	*M. canariensis*	GCca10	Smeared	15359	43	d3	ATCAT	ATGAT	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
LGlp_581_1	La Gomera	*M. gomerensis*	LGgo1	High	21247	158	d3	TCGAA	TTCGA	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
LGvg_572_2	La Gomera	*M. gomerensis*	LGgo2	High	14541	105	d3	TCAGA	TCTGA	TATAGCCT / TATAGCCT	GAGATTCC / ATTACTCG
LGvg_577_9	La Gomera	*M. lepida*	LGlp2	High	13640	82	d3	TACGA	TCGTA	TATAGCCT / TATAGCCT	GAGATTCC / ATTACTCG
LGvvxlp_570_4	La Gomera	*M. lepida x M. pedro-luisii*	LGplxlp1	High	17195	109	d3	TTGCA	TGCAA	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
Tev_197	La Gomera	*M. varia*	Tev1	Smeared	959	5	d3	TACGA	TCGTA	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
EHv_H5_E	El Hierro	*M. hierrensis*	EHh1	High	5300	29	No	TCGAA	TTCGA	GGCTCTGA / TATAGCCT	CTGAAGCT / ATTACTCG
EHv_H5_F	El Hierro	*M. hierrensis*	EHh1	High	4741	41	No	ACGTT	AACGT	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG
GCle_501	Gran Canaria	*M. leucantha*	GCle1	Medium	463	5	No	TTGCA	TGCAA	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG

Table S1. Continued.

Individual	Island	Species	Population	Sample quality	Nr of reads	Nr of loci	Included in phylogeny	Barcode P5	Barcode P7	Index P5	Index P7
GCpi_390	Gran Canaria	*M. pineolens*	GCpi1	Medium	2145	19	No	ACGTT	AACGT	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
GCpi_393	Gran Canaria	*M. pineolens*	GCpi1	Medium	2696	27	No	ACGTT	AACGT	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
GCpi_394	Gran Canaria	*M. pineolens*	GCpi1	Medium	11771	70	No	ACAGT	ACTGT	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
GCte_321	Gran Canaria	*M. tenuis*	GCte1	Medium	732	10	No	TACGA	TCGTA	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
GCv_14_2	Gran Canaria	*M. canariensis*	GCca2	Smeared	5160	22	No	AGCAT	ATGCT	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
GCv_14_5	Gran Canaria	*M. canariensis*	GCca3	Smeared	568	1	No	TTGCA	TGCAA	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
GCv_17_15	Gran Canaria	*M. canariensis*	GCca2	Medium	17569	128	No	TACGA	TCGTA	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
GCv_335	Gran Canaria	*M. canariensis*	GCca5	Medium	668	8	No	TCGAA	TTCGA	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
GCv_338	Gran Canaria	*M. canariensis*	GCca5	High	3917	21	No	ACAGT	ACTGT	CCTATCCT / TATAGCCT	GAATTCGT / ATTACTCG
Lav_1	Lanzarote	*M. mahanensis*	Lama1	Smeared	10633	32	No	ACGTT	AACGT	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
Lav_2	Lanzarote	*M. mahanensis*	Lama1	Low	4791	40	No	TCGAA	TTCGA	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
Lav_MicG	Lanzarote	*M. mahanensis*	Lama2	Low	10189	75	No	TCGAA	TTCGA	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
LGlp_571_1	La Gomera	*M. lepida*	LGlp1	Medium	17215	137	No	ATCAT	ATGAT	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
LGvg_572_6	La Gomera	*M. gomerensis*	LGgo2	Medium	15056	111	No	TCGAA	TTCGA	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
LGvg_579_5	La Gomera	*M. gomerensis*	LGgo3	Low	4009	42	No	TTGCA	TGCAA	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
LGvv_564_11	La Gomera	*M. pedro-luisii*	LGpl1	Low	1011	6	No	AGCAT	ATGCT	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
LGvv_564_3	La Gomera	*M. pedro-luisii*	LGpl1	Low	4944	33	No	ACAGT	ACTGT	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
LGvv_564_7	La Gomera	*M. pedro-luisii*	LGpl1	Medium	12273	88	No	ATCAT	ATGAT	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
LGvv_564_9	La Gomera	*M. pedro-luisii*	LGpl1	Low	15163	93	No	AACGT	ACGTT	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
LGvv_565_10	La Gomera	*M. pedro-luisii*	LGpl2	Medium	7516	53	No	AGCAT	ATGCT	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
LGvv_568_4	La Gomera	*M. pedro-luisii*	LGpl3	High	1463	9	No	TACGA	TCGTA	ATAGAGGC / TATAGCCT	ATTCAGAA / ATTACTCG
LGvv_568_5	La Gomera	*M. pedro-luisii*	LGpl3	Medium	11929	77	No	ACAGT	ACTGT	TCATCTTA / TATAGCCT	CGGCTATG / ATTACTCG
LGvvxlp_570_5	La Gomera	*M. lepida x M. pedro-luisii*	LGplxlp1	Low	8733	77	No	TACGA	TCGTA	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
LPhe_32_A	La Gomera	*M. herpyllomorpha*	LPhe2	Low	15731	70	No	TCAGA	TCTGA	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
Mav_6	La Gomera	*M. maderensis*	Mam1	Smeared	1886	7	No	TCGAA	TTCGA	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
Tehy_2_3	La Gomera	*M. hyssopifolia*	Tehy1	Low	2764	24	No	TTGCA	TGCAA	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
Tela_289	La Gomera	*M. lasiophylla*	Tela2	Medium	4748	42	No	TTGCA	TGCAA	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG
Tev_199	La Gomera	*M. varia*	Tev1	Medium	84677	393	No	ATCAT	ATGAT	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
Tev_260	La Gomera	*M. varia*	Tev3	Smeared	8030	24	No	AACGT	ACGTT	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
Tev_261	La Gomera	*M. varia*	Tev3	Smeared	1050	5	No	TCAGA	TCTGA	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA

Table S1. Continued

Individual	Island	Species	Population	Sample quality	Nr of reads	Nr of loci	Included in phylogeny	Barcode P5	Barcode P7	Index P5	Index P7
Tev_28_2	La Gomera	*M. varia*	Tev6	Medium	12635	84	No	TACGA	TCGTA	CAGGACGT / TATAGCCT	TGCGCGAA / ATTACTCG
Tev_28_4	La Gomera	*M. varia*	Tev6	Smeared	11397	28	No	ACAGT	ACTGT	TATAGCCT / ATAGAGGC	CTGAAGCT / TCCGGAGA
Tev_31_2	La Gomera	*M. varia*	Tev7	Low	3420	28	No	TACGA	TCGTA	ATACACCG / ATAGAGGC	GAATTCGT / ATTACTCG
Tev_319	La Gomera	*M. varia*	Tev4	Medium	5423	40	No	TTGCA	TGCAA	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
Tev_32_3	La Gomera	*M. varia*	Tev8	Low	19821	116	No	TCAGA	TCTGA	TATACGGT / ATAGAGGC	ATTCAGAA / ATTACTCG
Tev_563_10	La Gomera	*M. hyssopifolia*	Tehy2	High	9334	48	No	TTGCA	TGCAA	AGGCGAAG / TATAGCCT	TTAATGCGC / ATTACTCG
Tev_563_9	La Gomera	*M. hyssopifolia*	Tehy2	Medium	33607	124	No	AGCAT	ATGCT	GTACTGAC / ATAGAGGC	GAGATTCC / ATTACTCG

Table S2. Oligonucleotides used in our RAD-sequencing approach. Barcodes and Index information are available in Table S1.

Name	Sequence
Y-adapter P5	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT[Barcode]-3
Y-adapter P7	[Phos] 5'-AATT[Barcode]AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'
P5seq	5'-ACACTCTTTCCCTACACGACGCTCTT-3'
P7 seq	5'-GTGACTGGAGTTCAGACGTGTGCTCTT-3'
P5amp	5'-AATGATACGGCGACCACCGAGATCT-3'
P7amp	5'-CAAGCAGAAGACGGCATACGAGAT-3'
P5_Index_Primer	5'-AATGATACGGCGACCACCGAGATCTACAC[IndexP5]ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
P7_Index_Primer	5'-CAAGCAGAAGACGGCATACGAGAT[IndexP7]GTGACTGGAGTTCAGACGTGTGCTCTT-3'

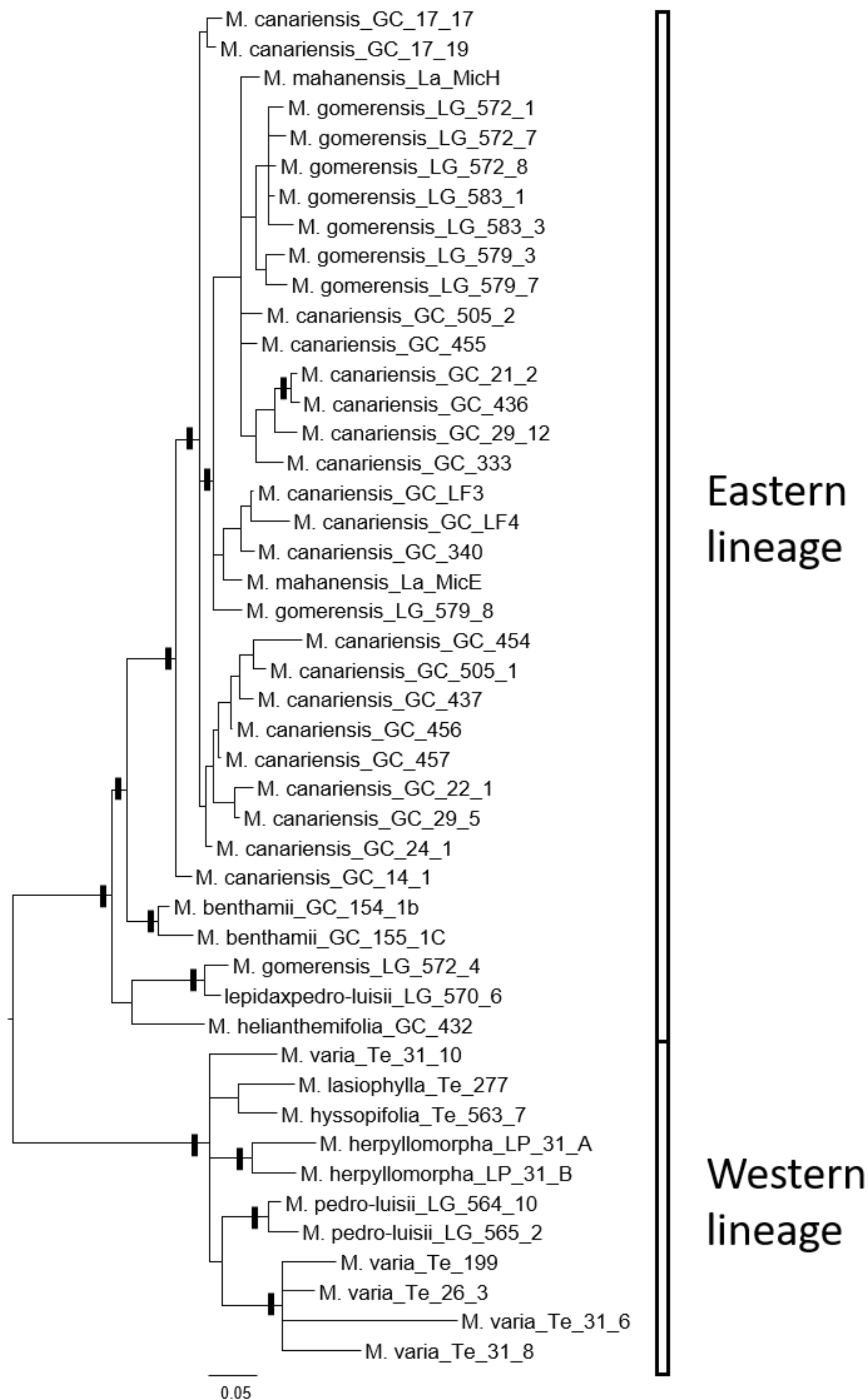


Fig. S1. Mr.Bayes tree calculated with the dataset including loci with a maximum missing data of 50%. Bars in the nodes correspond to posterior probability values >90%. The last two letters after species name correspond to island information as in Table S3.