

Genetic diversity and population structure of wild Sardinian Myrtle (*Myrtus communis* L.) genotypes from different microclimatic areas

Sara Melito^{1*}, Leonarda Dessena², Luana Sale², Maurizio Mulas²

¹Department of Agriculture, University of Sassari, Sassari, Italy

²Department of Nature and Land Sciences, University of Sassari, Sassari, Italy

*Corresponding author: smelito@uniss.it

Abstract

The wild Sardinian Myrtle (*Myrtus communis* L.) is an aromatic shrub widely spread in the Mediterranean region. Because of the large use of the plant in food and cosmetic industry, a domestication process started from the last decades. Today, the myrtle may be considered as an aromatic crop. In this artificial evolution, the genetic characterization of wild populations is essential to understand the bases of the phenotype variability useful for cultivar selection. The aim of the present study was to explore the genetic diversity and structure of wild myrtle populations within eco-geographical zone of Sardinia, where the specie is still widely abundant. Eighteen populations were sampled randomly by picking 20 individuals from every site. A total of 360 samples were analysed by AFLP markers, by which 150 polymorphic fragments were detected. In total, two well-differentiated genetic groups were identified. Agrometeorological and geographical variables influenced the distribution of the two clusters. A significant different partition of the two groups was observed from South West to North East direction. Moreover, altitude plays a significant role in the genetic distribution: one cluster is predominant at low altitude and the other at medium ($P = 0.0014$). AMOVA was performed to explore the effect of different environmental factors to the genetic variance. The highest F_{ST} value (0.44) was associated to the origin of the sampling site followed by the genetic partition at $K = 2$ (0.34). A map of myrtle genetic diversity as result of the collection site was described. These findings are relevant for planning collection for future domestication programs.

Keywords: AFLP; Genotyping; Micrometeorological areas; Myrtle; Population clustering.

Abbreviations: AFLP amplified fragment length polymorphism; UPGMA unweighted pair group method using arithmetic averages; F_{ST} fixation index.

Introduction

In the last years, the interest for aromatic and medicinal plants is growing because of new uses for food and industrial no food products (Lange, 2006). The following speciality materials can be obtained from plants as industrial crops: essential oils, pharmaceuticals, herbal health products, dyes and colorants, cosmetics and personal care products, plant protection products and intermediates, from which the cited products can be produced (Lubbe and Verpoorte, 2011).

Because of the relative plant biodiversity abundance in the Mediterranean area since old times, many autochthonous or newly introduced plants represented the basic resources of folk medicine and of any other activity related to the human development (Cowling et al., 1996; Mulas, 2006).

Myrtle (*Myrtus communis* L.) represents a well-known resource of the folk medicine. Recently, the importance of domestication and cultivation of this species was determined due to wide application of this plant. Beside its use as ornamental specie of the Mediterranean "maquis" (Bruna et al., 2007), myrtle is intensively exploited for the aromatic proprieties of berries and leaves (Mulas et al., 1998; Messaoud et al., 2006). Essential oil extraction, antioxidants and liquor production are only some of the food and pharmaceutical industries utilization of the myrtle leaves and fruits (Gardeli et al., 2008; Mulas, 2012; Mulas and Melis, 2011). Considering the economic importance of myrtle and the success of some food products such as the liquor, the request of raw material in processing industries is constantly

growing. Overall most of the myrtle biomass comes from wild plants, causing a serious risk of genetic diversity reduction. This phenomenon is even more evident in Sardinia, where the liquor production is more than 4 ML (Mulas, 2012). In this case, the uncontrolled harvest of fruit and in general myrtle biomass (Mulas et al., 1998), associated to the natural fires, livestock and wild herbivores grazing and its cutting for firewood, is increasing the pressure of genetic erosion.

As a consequence, a conservation program is required to supply the liquor and pharmaceutical industries plant material request and to preserve and contain the reduction of natural population's diffusion (Melito et al., 2013a; Mulas and Melis, 2011). The preservation of myrtle population's natural diversity represents a long term strategy for domestication and future cultivars selection.

Particularly in Sardinia land, this strategy could evolve in the selection of cultivars with adaptability to growing conditions (Melito et al., 2013a).

The natural genetic resources contain higher variability than the pool of cultivars selected and their evaluation may be useful to support and improve the breeding program of myrtle (Mulas and Cani, 1999). Microclimatic diversity of spreading areas of plant natural populations is a factor determining their genetic variability (Melito et al., 2014).

Genetic characterization based on molecular markers, has been successfully performed to explore the biodiversity status

to evaluate the population genetic structure and to plan suitable conservation management strategies (Agrimonti et al., 2007).

Nowadays, several molecular markers have been used on natural plant species to assess the genetic variation of endangered species (Juan et al., 2004). RAPD, SSR, SRAP, CAPS-SNP and AFLP are some of the possible genotyping techniques applied to assess the germplasm diversity in several plant species such as *Eucaliptus*, Almond, *Citrus* and *Helichrisum* (Osman et al., 2012; Rigoldi et al., 2015; Amar et al., 2011; Melito et al., 2013b).

Previous studies on Italian (Melito et al., 2016; Agrimonti et al., 2007; Melito et al., 2013a; Melito et al., 2014) and Mediterranean myrtle populations (Bruna et al., 2007; Nora et al., 2015; Messaud et al., 2011; Serce et al., 2008; Albaladejo et al., 2009) showed that dominant and codominant molecular markers could associate genotype to a specific geographic area. In particular, AFLPs are extremely useful to assess the genetic diversity of myrtle, screening a large number of loci simultaneously without any preliminary knowledge on the genotype sequence.

The present research was aimed to evaluate the diversity of myrtle natural populations of Sardinia using AFLP molecular markers. Furthermore, the influence of historical agrometeorological characteristics of some explored areas on diversity-distribution of relative genotypes was studied with reference to the regional macroarea. The genetic fingerprinting and the population genetic structure data were applied to explore the relationship between genetic variation and environmental factors such as altitude, geographical origin and landscape feature.

Results and Discussion

AFLP analysis and population genetic structure

Several studies conducted on Italian myrtle populations have shown that molecular markers can be successfully used to characterize natural populations (Agrimonti et al., 2007; Bruna et al., 2007; Melito et al., 2013a; Melito et al., 2016). More recently, AFLP analysis has been performed to characterize domesticated (Melito et al., 2014) and wild myrtle populations (Melito et al., 2016), respectively, in Sardinia and Sicily. Natural populations of myrtle are subjected to ecological risk. As consequence, the assessment of genetic diversity state represents an essential step to prevent loss of genetic variability.

AFLP fingerprinting showed to be useful to explore the influence of specific environmental factors (such as altitude, soil type, climatic condition, etc.) in the selection of ecotypes adapted to the associated ecological niches (Melito et al., 2014; Melito et al., 2013b; Rapposelli et al., 2015a, b).

The exploration of the genetic diversity in natural population of myrtle represents an important step to find interesting genotypes and to plan new breeding programs. Myrtle selection could be performed for several purposes such as fruit and biomass yield, identification of interesting agronomical traits or for morphological characters. Mulas (2012) reported the selection and use as ornamental plant of myrtle accessions characterized by re-flowering recurrence. Molecular markers could be helpful to select targeted genotypes of myrtle with interesting combination of characters for domestication programs.

The AFLP analysis on the 18 wild myrtle populations, collected from 6 climatic zones in Sardinia, produced 150 polymorphic fragments ranging from 50bp to 600bp. Overall,

the AFLP reaction produced 30-40 fragments for the primer set used. The *EcoRI*+1T/*MseI*+1C and *EcoRI*+1C/*MseI*+1A revealed the highest number of bands (40 bands for combination). Faint polymorphic bands were excluded from the AFLP analysis to minimize the AFLP scoring errors. Control genotypes were used to align different run gels in the same primer combination.

AFLP profiling was elaborated with STRUCTURE tool to evaluate the presence of genetic cluster. As previously reported by Melito et al. (2014) in candidate cultivar selection, we also detected two principal clusters in this collection of wild myrtle populations, referred as Cluster A (CA) and Cluster B (CB) (Fig 1). In general, more than 78% of genotypes had $Q \geq 0.7$ and were consequently assigned to one of the two genetic clusters. The two groups are not equally distributed among the identified climatic areas of Sardinia. Overall, 62% and 16% of individuals belonged to CA and CB, respectively (data not shown). The remaining 22% of individuals showed a low Q ($Q < 0.7$), and were classified as "admixed" (data not shown). These myrtle accessions were not assigned to any specific genetic cluster and were not considered in the successive investigations. The admixed individuals are probably the consequence of gene flow, generated by crossing events, among the two clusters identified at $K = 2$. Similar trend was observed in wild and domesticated myrtle populations from different area of Sardinia (Melito et al., 2013a; Melito et al., 2014), suggesting that different environmental factors could be involved in myrtle germplasm diversification.

In Fig 2, the genetic distance among the 18 myrtle populations based on the UPGMA approach is shown. The dendrogram confirms the general feature displayed by the Bayesian clustering model (Fig 1). The pie chart indicates the mean coefficient of membership of each population at $K = 2$. Two principal clusters can be identified in the dendrogram: Cluster I and Cluster II: NE2 and CW1, which belonged mostly to CB, clustered together in a separated branch II; NW5, NE3 and SW2 with a lower CB contribution, are grouped in a sub-branch of the cluster II.

To evaluate the role of meteorological micro area on the population genetic diversity the average Q was calculated for each population/locality (Fig 3). As expected, CA and CB are specifically distributed on each sampling site as reported in Fig 1 and 2. The S, SE and NW presented the higher percentage of CA (>90%), while in the other three meteorological areas (SW, CW and NE), the CB showed a higher percentage (>30%). Based on these findings, the contingency analysis was conducted. Significant different distribution of the two genetic clusters were found in the 6 meteorological areas ($P < 0.001$) (Supplementary Fig 2).

Genetic diversity and habitat

Heterozygosity (H_e) was used as indicator of genetic diversity of wild myrtle population of Sardinia. The H_e ranged from 0.066 (CW3) to 0.253 (CW1) with an average of 0.142, suggesting a general high genetic diversity of the species in Sardinia. The H_e value observed in this study was higher than one reported by Melito et al. (2014) in candidate cultivars collection and in wild Sardinian populations (Agrimonti et al., 2007). The apparent discrepancy found with previous studies could be the results of different sampling size and collection strategies. In fact, Melito et al. (2014) explored the genetic diversity in myrtle candidate cultivars selected for plant fruit and biomass

Table 1. Sampling sites features (geographical coordinates, altitude and climatic zones) of 18 myrtle populations.

Population's Code	Coordinates			Altitude (m)	Climatic zones
	Zone	N	E		
NW1	32T	4462812	456221	94	North West (NW)
NW2	32T	4459887	458358	101	
NW3	32T	4501303	433373	39	
NW4	32T	4502896	445301	53	
NW5	32T	4444035	459720	236	
NE1	32T	4340345	489951	481	North East (NE)
NE2	32T	4515988	529566	356	
NE3	32T	4522211	541729	80	
SE1	32S	4346876	543020	68	South East (SE)
SE2	32S	4358510	543006	63	
CW1	32T	4440703	458885	118	Central West (CW)
CW2	32S	4411400	484623	152	
CW3	32S	4411011	508840	709	
SW1	32S	4435751	478509	270	South West (SW)
SW2	32S	4373605	461082	50	
S1	32S	4366476	449519	68	South (S)
S2	32S	4346524	484089	151	
S3	32S	4358431	543126	411	

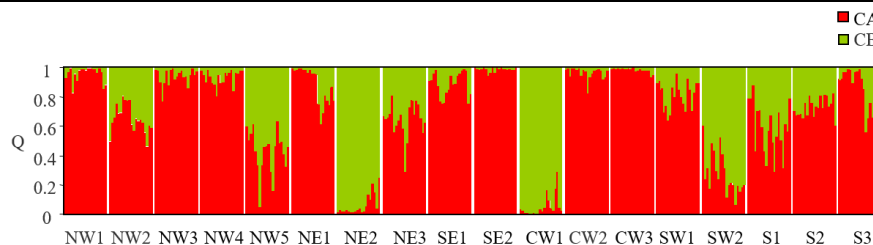


Fig 1. Population genetic structure of 360 wild *M. communis* plants from 7 climatic zones in Sardinia. Each zone presented at least 2 populations (see Table 1 for details). The figure shows the Bayesian clustering model analysis results implemented in STRUCTURE program (Pritchard et al., 2000) at $K = 2$. Each individual was assigned to a specific genetic clusters (A, B) based on the coefficient of membership ($Q \geq 0.7$). Plants with a $Q < 0.7$ were classified as admixed and were not assigned to any specific genetic group.

Table 2. Pairwise genetic distance matrix (FST) among the 6 climatic zones with the 18 myrtle populations. the highest and lowest FST values are indicated in bold. All comparisons were significant after 1000 random permutation tests ($P < 0.05$).

	NW	NE	SE	CW	SW	S
NW	0.000					
NE	0.101	0.000				
SE	0.147	0.172	0.000			
CW	0.099	0.059	0.138	0.000		
SW	0.155	0.115	0.253	0.157	0.000	
S	0.123	0.133	0.172	0.127	0.143	0.000

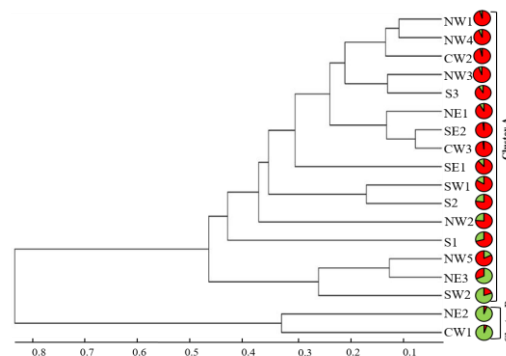


Fig. 2

Fig 2. Dendrogram of 18 myrtle populations based on UPGMA clustering obtained by AFLP analysis. The distribution of myrtle population's genetic clusters as calculated by STRUCTURE software is shown on the right part of the image. Each individual pie chart are based on the average coefficient of membership (Q) for each population at $K = 2$.

Table 3. Partition of genetic diversity determined by AMOVA analysis. The overall, in selection and neutral alleles, the climatic zone, the altitude classes, the geographic classes, the partition at K = 2, were considered sources of molecular variance.

Population	Source of variation	d.f.	% Variation	FST
Overall	Among populations	17	43.84	0.44
	Within populations	342	56.16	
Alleles in selection	Among populations	17	66.85	0.67
	Within populations	342	33.15	
Alleles neutral	Among populations	17	40.47	0.40
	Within populations	342	59.53	
K = 2	Among cluster	1	36.97	0.34
	Within cluster	279	63.03	
Climatic zone	Among climatic zone	5	12.93	0.13
	Within climatic zone	354	87.07	
Altitude classes	Among altitude classes	2	6.58	0.07
	Within altitude classes	357	93.42	

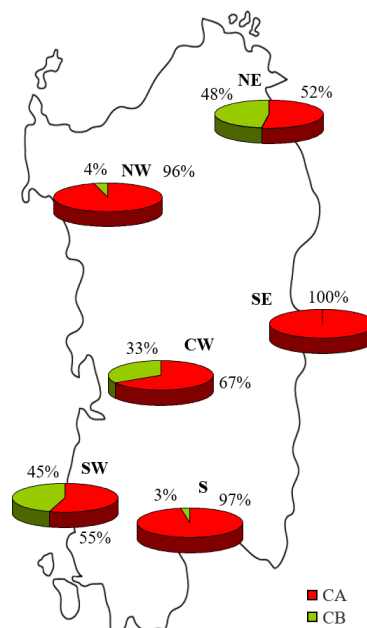


Fig. 3

Fig 3. Distribution of the myrtle population's genetic clusters in Sardinia, as evaluated by STRUCTURE, based on 7 climatic zones. Each individual pie chart shows the average coefficient of membership (at K = 2) for all the populations sampled in each climatic zone.

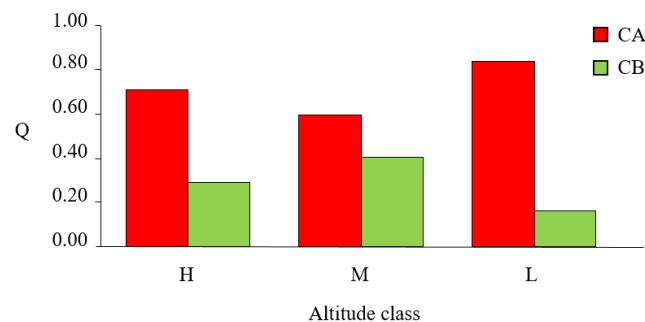


Fig. 4

Fig 4. Distribution of genetic clusters (CA and CB) in wild myrtle populations in Sardinia. CA and CB, identified by STRUCTURE were studied according to three altitude classes: >200 m asl (high altitude, H), < 100m asl (low altitude, L), 101-200 m asl (medium altitude, M).

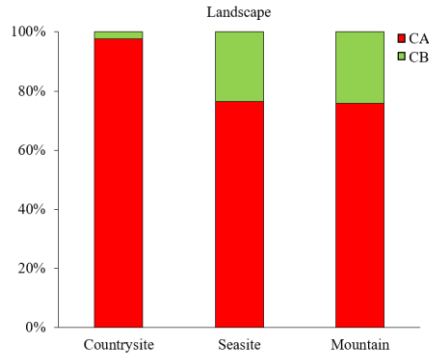


Fig. 5

Fig 5. Partition of the two genetic cluster CA and CB in the three principal landscapes studied: Countryside, Seaside and Mountain.

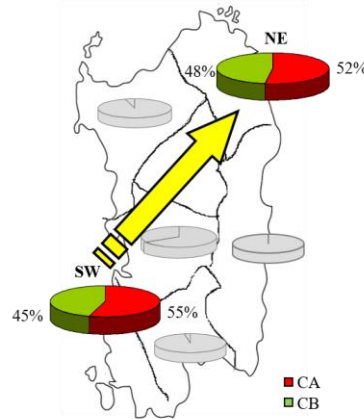


Fig.6

Fig 6. Geographic distribution of the two main clusters identified at K = 2 based on South West (SW) to North East (NE) orientation.

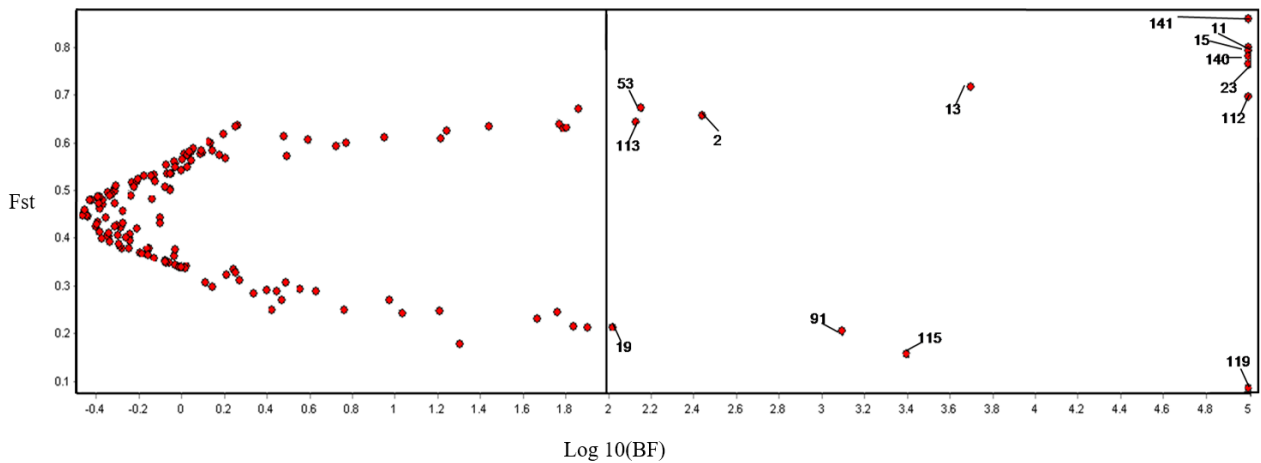


Fig 7. AFLP loci putatively under selection based on Bayescan 1.0 software (Foll and Gaggiotti, 2006). Each point corresponds to a single AFLP band. In figure the FST is plotted against the log10 of posterior odds. The vertical dashed line indicates the chosen threshold value of 2.0.

productivity, while Agrimonti et al. (2007) compared 10 Sardinian populations with 4 myrtle populations from Calabria. Pairwise distance matrix of FST was used to explore the genetic diversity of myrtle from different micro climatic areas (Table 2). Overall the average FST is 0.130 indicating a lower level of divergence compared to other myrtle populations in Mediterranean area such as Sicily (Melito et al., 2016) and Tunisia (Messaoud et al., 2006). These differences might be related to the specific localization

of the sampling sites. In the first case, Melito et al. (2016) explored the divergence of a reduce number of populations mainly from the coast area of Sicily. In the second case, the aim of the research was to study the genetic diversity of *M. communis* samples from upper-semiarid and sub-humid bioclimatic regions in Tunisia.

SE and SW plants presented the highest FST value (0.253), while NE and CW had the lowest FST value (0.059) as

shown in Table 2. These populations have different distribution of the two genetic clusters CA and CB (Fig 3).

Genetic clusters and environmental conditions

Difference in gene flow could have caused a differential distribution of the two genetic clusters identified. The microclimatic condition of the 6 areas as well as geographical distances and physical barriers could have affected the pollen and seed dispersion and the distribution of the two clusters in specific micro-areas.

The climatic parameters are reported to be relevant for plant biomass production, plant development, flowering and genetic diversity distribution. In Sicily, for instance cluster assignment and winter conditions (rainfalls and temperatures) are reported to be significantly correlated (Melito et al., 2016). The principal meteorological data for the climatic zones identified in Sardinia are reported in Supplementary Table 1. In Supplementary Table 2 the pairwise correlations results between climatic parameters and genetic parameters has been reported. The average minimum temperature of the winter months of January, February and March are negatively correlated to the He, while the average minimum temperatures of August are negatively correlated to the coefficient of membership of the CA.

The particular shape of Sardinia with hill and mountains localized in specific geographical areas rise interesting observations on the relationship between habitat and genetic diversity distribution. The altitude level is reported as an environmental factor influencing the population genetic diversity of aromatic plants (Melito et al., 2013ab; Melito et al., 2014; Melito et al., 2016).

Relationship between altitude and genetic differentiation has been reported in cultivar selections and in wild myrtle populations from Sardinia and Sicily, respectively (Melito et al., 2014, 2016). As reported by Melito et al., (2014), the sampling sites were divided into three main altitude classes: <100 m asl (low altitude; L), 101-200 m asl (medium altitude; M), and >201 m asl (high altitude; H). Significant correlation between the two genetic clusters and the altitude levels of each sampling site was found (Pearson, $P = 0.0014$). As shown in Fig 4, CA and CB followed a specific trend. CA is predominant a low altitude level, while CB is more representative between 101-200 m asl. In addition, where one of the two clusters is predominant, the other one was followed an opposite trend. Beside the genetic cluster, also He significantly varied based on the altitude. The highest He was found in plant collected above 200 m (0.260), while the lowest was recorded at <100 m asl (0.218) (Supplementary Fig 3). Beside the altitude levels, other environmental conditions could have influenced the genetic cluster distribution in Sardinia. The island is characterized mostly by hill (67.9%), followed by lowland (18.5%) and then by 13.6% of mountains (data not shown). In addition, being an island, the proximity with the sea has affected the microclimatic conditions of the coast. Based on these considerations, the relationship between genetic cluster assignment and micro area were evaluated. Seaside, mountain zone and countryside were considered the three principal habitats explored. The contingency test (Pearson 0.05; $P = 0.0077$) displayed a significant different distribution of CA and CB in the considered landscapes (Fig 5). CA is predominant in the countryside and presented at the same level in mountain and sea landscapes, while CB is almost absent in countryside and equally represented in seaside and mountain zones.

A similar approach was inferred to explore, if the geographic localization of the studied sampling sites influence the genetic clusters distribution. The genetic groups did not display a significant difference distribution from North to South, and from West to East. However, the contingency test between clusters assignment and geographic origin showed a geographical trend from South-West to North-East (Fig 6) (Pearson 0.05; $P \leq 0.0001$). The CB increased in representativeness from South-West to North-East, while CA showed an opposite trend. Similar trend was found in myrtle candidate cultivars selected for biomass and berry production by Melito et al. (2014).

Several factors could have induced such feature in the population genetic structure and among them the distribution of mountain and hill could represent one of the predominant one. The specific distribution Sardinian's mountains (from West to East) associated to the geographical distance North-South could have act as barrier for the natural gene flow of myrtle populations.

AMOVA

Analysis of molecular variance was performed to explore the genetic variation distribution and to evaluate the contribution of different sources of variance. Considering all AFLP scored alleles, most of the genetic variance was found within populations (56.16), while a lower contribute was associated to the among wild myrtle populations (43.84) (Table 3). The associated F_{ST} based on the 18 populations was quite high (0.44), suggesting a general deep divergence among the considered *M. communis* populations.

BAYESCAN 1.0 software (Foll and Gaggiotti, 2006) identified 14 loci putatively under selection (threshold \log_{10} value of 2.0, posterior odds probability 0.99) (Fig 7). The AMOVA was then calculated considering the 14 loci under selection and the remaining neutral alleles (Table 3). The variance base on the "under selection alleles" was partitioned into 66.85% "among populations" and 33.15% "within populations", while the apportioning of the variance based on neutral loci produced a picture similar to that obtaining using the whole AFLP markers.

To evaluate the contribution of the population genetic structure to the genetic variance, AMOVA was conducted using the partition identified at $K = 2$. As shown for the full allele dataset of the 18 myrtle populations, also in this case a similar distribution of the genetic variance was revealed. Most of the variance was found within (63.03%) CA and CB and a minor contribute was associated to the among clusters partition (36.97%).

Beside the genetic structure the influence of environmental factors on the molecular variance was evaluated. Considering the 6 climatic areas identified in Sardinia and the three altitude classes (see materials and methods) again the same partition of the genetic variance was observed. Climatic area and altitude presented a highest percentage of variance associated to the "within populations" (83.07 and 93.42 respectively) and the lowest one related to the "among population" (12.93 and 6.58, respectively).

Overall considering all source of molecular variance the F_{ST} showed a high value when *M. communis* was divided in the 18 populations. As expected the F_{ST} using only under selection alleles increased to 0.67, indicating that these alleles are responsible of differentiation between the all myrtle populations. Furthermore the partition identified at $K = 2$ has an important role in the genetic differentiation, as shown by the F_{ST} value (0.34), climatic zone and altitude division load

a minor contribution to the variance with an F_{ST} of 0.13 and 0.07, respectively.

The AMOVA results underlined that, despite a significant part of variation belonged to the difference among populations from different climatic areas and among two genetic groups identified by Bayesian clustering model, the principal source of variance is due to “within population’s partition”. The high level of differentiation at the intrapopulation level could be associated to the myrtle reproduction system. As reported by several authors (Lughadha and Proenca, 1996; Mulas and Fadda, 2004) *M. communis* is an outcrossing species that partially use a self-pollination system. The principal pollination vectors in myrtle are insect and in particular coleopterans. The AMOVA and the F_{ST} results pointed out a prevalent source of variance at the intrapopulation level. These data is compatible with the limited mobility coleopterans that consequently cause a deep genetic exchange among plants of the same populations and climatic area (Agrimonti et al., 2007).

Materials and methods

Sampling and extraction of DNA

Leaf tissues of 360 plants were collected from 18 wild populations growing in 6 climate zones of the Sardinian region according to ARPAS (Regional Agency for Environmental Protection of Sardinia)(Supplementary Table 1). For each sampling site the geographic position and altitude was reported (Table 1). Global Positioning System (GPS) was also used to record the coordinates of each locality. The twenty samples of the same locality were selected from an area of 40 m of diameter.

For each climate zone the historical meteorological data from weather stations located within each zone, were acquired (Supplementary Table 1). Meteorological data (average monthly temperature, average minimum and maximum monthly temperatures) were obtained from historical series (1991-2011). The data were supplied by the Hydrological Service of the Region of Sardinia.

Analysis AFLP

Total genomic DNA was extracted from 100 mg of young leaf tissue using an electric homogenizer TissueLyser (Qiagen, Italy) and the DNeasy Plant Mini Kit (Qiagen, Italy) following the manufacturer’s protocol.

AFLP analysis was conducted following Vos et al., (1995) protocol with minor modifications. For each individual, 250 ng of genomic DNA were digested with 5U of *EcoRI* and *MseI* enzymes, in a 10X restriction/ligation buffer (100 mM of Tris base, 100 mM of MgAc, 500 mM of Kac, 50 mM of DTT, 100 ng / μ l of BSA). The restricted mix was then treated adding 10 μ l of the ligation mix (5 μ M *EcoRI* + 1A adapters, 50 μ M of *MseI* + 1C adapters, 10 mM of ATP and 1 U of T4 ligase). The digested/ligated (D/L) mix was then incubated at 37 °C for 3 h. Subsequently, the D/L DNA was diluted with sterile water (1: 9) and 5 μ l of mix were then used for the preamplification reaction.

The preamplification was performed in a total of 20 μ l volume containing 5 μ l of D/L DNA, 1.5 mM of MgCl₂, 10X Buffer, 10 mM dNTPs, 2.75 μ M *EcoRI* primer, 2.75 μ M of primer *MseI* and 1 unit of Taq polymerase (Invitrogen). The preamplification products were diluted with sterile MilliQ water (1: 4), and 5 μ l were used for the selective PCR according to Vos et al., (1995) protocol. The selection of the number of preamplification and the selective amplification

primer combinations and sequences was performed based on previous studies conducted on Sardinian and Sicilian myrtle genotypes, with few modifications (Melito et al 2014; Melito et al 2016; Agrimonti et al 2006).

Two primers combinations in the pre-amplification reaction (*EcoRI*+T/*MseI*+C; *EcoRI*+C/*MseI*+A) and 4 primer combinations in the selective amplification reaction (*EcoRI*+TGG/ *MseI*+CGG; *EcoRI*+TTT/ *MseI*+CCT; *EcoRI*+TAG/ *MseI*+CGG; *EcoRI*+CCT/ *MseI*+ATA) were assessed.

The AFLP amplicons were separated in polyacrylamide gel (6% w/v) and the band were visualized by silver staining procedure (Bassam et al., 1991). Alleles score was manually performed based on the presence/absence of selected specific bands in each sample. A binary matrix 0 (absence), 1 (presence), based on the fragments scoring was elaborated. Forty/fifty bands for each primer was recorded. The molecular weight of bands has a range between 50 bp to 500 bp. To ensure the AFLP band detection and to minimize the scoring errors, faint polymorphic fragments were excluded. Control genotypes were loaded in different run gels for each primer combination, in order to easily follow polymorphic bands in different gels. Finally to guarantee the reproducibility of the scoring 10% of the samples were twice scored.

BAYESCAN 1.0 software (Foll and Gaggiotti, 2006) was conducted to identify loci putatively neutral or under selection (99% confidence interval, pilot run length 5,000). The loci identified were used for subsequent AMOVA.

Population genetic structure

Population genetic structure was explored using the Bayesian clustering model inferred in STRUCTURE 2.3.4 tool (Pritchard et al., 2000). The range in K considered was 1–12. The number of groups (K) was estimated by 20 independent runs for each K, with 20,000 Markov Chain Monte Carlo (MCMC) repetitions after a burn-in period of 20,000 interactions for each group number K. The most likely number of genetic cluster (K) was evaluated based on Evanno et al. (2005) method through the Structure Harvester software (Earl et al., 2012). Individuals were assigned to each cluster if their coefficient of membership (Q) was ≥ 0.70 .

Population divergence

Analysis of molecular variance (AMOVA), evaluation of genetic distance between populations (F_{ST}), percentage of polymorphic loci (P) and estimation of expected heterozygosity (H_e) were calculated using Arlequin 3.5.2.2. software (Excoffier and Lischer, 2010). The input file for Arlequin was generated using the PGD Spider 2.0.9.2 tool (Lischer and Excoffier, 2012). Population genetic distance matrix (Nei, 1973) was computed with Arlequin to evaluate the pairwise genetic distance (F_{ST}) among the *M. communis* population studied.

AMOVA and the F_{ST} values were also evaluated using different source of variance, such us climatic zone, altitude classes and geographic location to evaluate the factor that contribute mostly to the genetic variance.

A phylogenetic tree was constructed using Popgene v 3.2 software (Yeh et al., 1997) based on UPGMA (Sokal and Michener, 1958) clustering.

Statistical analysis

Pearson’s χ^2 test for contingency tables was conducted for categorical variables to evaluate the correlations with the

genetic clades identified by STRUCTURE. Altitude, climatic zone and geographic feature were considered as categorical variables. In particular, according to Melito et al. 2014 three altitude classes were considered: high (H) (> 200 m), medium (M) (between 100 and 200 m), and low (L) (<100 m); climatic zone were classified in North West (NW), North East (NE), South East (SE), Center West (CW), South West (SW) and South (S) based on ARPAS data. Finally sampling sites were assigned to specific geography classes (Sea, Mountain, Countryside) (Supplementary Figure 1). Pearson correlation between genetic cluster assignment (Q) and climatic (meteorological data) and environmental variable (altitude levels) were assessed with JMP 7 software (SAS Institute 2007) (significance at $P < 0.05$).

Conclusion

In this study, the evaluation of myrtle genetic diversity was associated with several environmental factors. Altitude, climatic conditions, landscape and geographic location of sampling site influenced the genetic partition of wild myrtle population in Sardinia. These findings could be useful to select candidate clone for future domestication programs.

Acknowledgments

Research was supported by the Regione Autonoma della Sardegna (Special Grant L.R. 7/2007-2012: “The myrtle: genetics and metabolomics, an integrated approach for industry development”). The authors would like to thank Dr. Emma Rapposelli for the useful suggestions in AFLP analysis.

References

- Agrimonti C, Bianchi R, Bianchi A, Ballero M; Poli F, Marmiroli N (2007) Understanding biological conservation strategies: A molecular-genetic approach to the case of myrtle (*Myrtus communis* L.) in two Italian Regions: Sardinia and Calabria. *Conserv Genet.* 8 (2): 385-396.
- Albaladejo RG, Fernandez-Carrillo L, Aparicio A, Fernandez-Manjarres J, Gonzalez-Varo JP (2009) Population genetic structure in *Myrtus communis* L. across a chronically fragmented landscape in the Mediterranean: can gene flow counteract habitat perturbation? *Plant Biol.* 11: 442-453.
- Amar MH, Biswas MK, Zhang Z, Guo WW (2011) Exploitation of SSR, SRAP and CAPS SNP markers for genetic diversity of citrus germplasm collection. *Sci Hortic-Amsterdam.* 128: 220-227.
- Bassam BJ, Caetano-Anolle's G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem.* 196: 80-83.
- Bruna S, Portis E, Cervelli C, De Benedetti L, Schiva T, Mercuri A (2007) AFLP-based genetic relationships in the Mediterranean myrtle (*Myrtus communis* L.). *Sci Hortic Amsterdam.* 113 (4): 370-375.
- Cowlin RM, Rundel PW, Lamont BB, Arroyo MK, Arianoutsou M (1996) Plant diversity in Mediterranean-climate regions. *Trends Ecol Evol.* 11(9): 362-366.
- Earl DA, von Holdt BM (2012) STRUCTURE HARVESTERS: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 4: 359-361.
- Evanno G, Reganut 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, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Eco Resour.* 10: 564-567.
- Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics.* 174:875-891.
- Gardeli C, Vassiliki P, Athanasios M, Kibouris T, Komaitis M (2008) Essential oil composition of *Pistacia lentiscus* L.: and *Myrtus communis* L.: evaluation of antioxidant capacity of methanolic extracts. *Food Chemistry.* 107: 1120-1130.
- Juan A, Crespo MB, Cowan RS, Lexer C, Fay MF (2004) Patterns of variability and gene flow in *Medicago citrina*, an endangered endemic of islands in the western Mediterranean, as revealed by amplified fragment length polymorphism (AFLP). *Mol Ecol.* 13: 2679-2690.
- Lange D (2006) International trade in medicinal and aromatic plants. In: Bogers, R.J., Craker, L.E., Lange, D. (Eds.), *Medicinal and Aromatic Plants: Agricultural, Commercial, Ecological, Legal, Pharmacological and Social Aspects.* Springer, Dordrecht, pp. 155-170.
- Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics.* 28: 298-299.
- Lubbe A, and Verpoorte R (2011) Cultivation of medicinal and aromatic plants for specialty industrial materials. *Ind Crops Prod.* 34(1): 785-801.
- Lughadha EN, Proenca C (1996) A survey of the reproductive biology of the Myrtoideae (Myrtaceae). *Ann Missouri Bot.* 480-503.
- Melito S, Chessa I, Erre P, Podani J, Mulas M (2013a) The genetic diversity of Sardinian myrtle (*Myrtus communis* L.) populations. *Electron J Biotechnol.* 16:1-14.
- Melito S, Sias S, Petretto GL, Chessa M, Pintore G, Porceddu A (2013b) Genetic and metabolite diversity of Sardinian populations of *Helichrysum italicum*. *PLoS ONE.* 8 (11): e79043
- Melito S, Fadda A, Rapposelli E, Mulas M (2014) Genetic diversity and population structure of Sardinian myrtle (*Myrtus communis* L.) selections as obtained by AFLP markers. *HortScience.* 49: 531-537.
- Melito S, La Bella S, Martinelli F, Cammalleri I, Tuttolomondo T, Leto C, Fadda A, Molinu MG, Mulas M (2016) Morphological, chemical and genetic diversity of wild myrtle (*Myrtus communis* L.) populations in Sicily. *Turk J Agric For.* 40: 249-261.
- Messaoud C, Khoudja ML, Boussaid M (2006) Genetic diversity and structure of wild Tunisian *Myrtus communis* L. (Myrtaceae) Populations. *Genet Resour Crop Evol.* 53(2): 407-417.
- Messaoud C, Bejaoui A, Boussaid M (2011) Fruit color, chemical and genetic diversity and structure of *Myrtus communis* L. var. *italica* Mill. morph populations. *Biochem Syst Ecol.* 39: 570-580.
- Mulas M (2006) Traditional uses of labiatae in the Mediterranean area. *Acta Hortic.* 723:25-32.
- Mulas M, and Melis RAM (2011) Essential oil composition of myrtle (*Myrtus communis*) leaves. *J Herbs Spices Med Plants.* 17(1): 21-34.
- Mulas M (2012) The myrtle (*Myrtus communis* L.) case, from a wild shrub to a new fruit crop. *Acta Hortic.* 948: 235-242.
- Mulas M, Cani MR, Brigaglia N (1998) Characters useful to cultivation in spontaneous populations of *Myrtus communis* L. *Acta Hortic.* 457: 271-278.

- Mulas M, Cani MR (1999) Germplasm evaluation of spontaneous myrtle (*Myrtus communis* L.) for cultivar selection and crop development. *J Herbs Spices Med Plants*. 6(3): 31-49.
- Mulas M, Fadda A (2004) First observation on biology and organ morphology of myrtle (*Myrtus communis* L.) flower. *Agricoltura Med*. 134: 223–235.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Nat Acad Sci USA*. 70 (12): 3321-3323.
- Nora S, Albaladejo RG, Aparicio A (2015) Genetic variation and structure in the Mediterranean shrubs *Myrtus communis* and *Pistacia lentiscus* in different landscape contexts. *Plant Biol J*. 17: 311–319.
- Osman AK, Abd El-Mageed AA, Tawfik AQ, Mohammed A (2012) Genetic diversity among four *Eucalyptus* species (Myrtaceae) based on random amplified polymorphic DNA (RAPD) analysis. *Afr J Biotechnol*. 11: 4729–4739.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*. 155: 945–959.
- Rapposelli E, Melito S, Barmina GG, Foddai M, Azara E, Scarpa GM (2015a) AFLP fingerprinting and essential oil profiling of cultivated and wild populations of Sardinian *Salvia desoleana*. *Genet Resour Crop Evol*. 62: 959-970.
- Rapposelli E, Melito S, Barmina GG, Foddai M, Azara E, Scarpa GM (2015b) Relationship between soil and essential oil profiles in *Salvia desoleana* populations: preliminary results. *Nat Prod Commun*. 38: 1615-1618.
- Rigoldi MP, Rapposelli E, De Giorgio D, Resta P, Porceddu A (2015) Genetic diversity in two Italian almond collections. *Electron J Biotechn*. 18(1): 40-45.
- SAS Institute Inc (2007) JMP. 7.0 Ed Cary, NC.
- Serce S, Şimşek O, Gunduz K, Aka-Kacar Y, Ercişli S (2008) Relationships among myrtle accessions from Turkey as revealed by fruit characteristics and RAPD. *Rom Biotech Lett*. 13: 4054-4065.
- Sokal R and Michener C (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*. 38: 1409–1438.
- Yeh FC, Yang RC, Boyle T, Ye ZH, Mao JX (1997) POPGENE, the user friendly shareware for population genetic analysis. *Molecular Biology and Biotechnology Center*. University of Alberta, Edmonton.
- Vos P, Hogers R, Bleeker M, Reijans M, Lee TVD, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zebau M (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res*. 23: 4407–4414.