UNIVERSITÀ DEGLI STUDI DELLA TUSCIA

Dipartimento di Scienze dell'Ambiente Forestale e delle sue Risorse
DISAFRI

Dottorato di ricerca in ecologia forestale
Ciclo XXIII

Development of a novel set of microsatellite markers and genetic characterization of Italian natural *Tamarix* populations

(AGR/05)

Coordinatore del corso:
Prof. Paolo De Angelis

Tutori:

Dott.ssa Elena Kuzminsky

Dott. Maurizio Sabatti

Correlatore:
Dott. Isacco Beritognolo

La dottoranda:
Serena Terzoli

There is grandeur in this view of life, with its several powers, having been originally breathed by the Creator into a few forms or into one; and that, whilst this planet has gone circling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved.

Charles Darwin

Non omnes arbusta iuvant humilesque Myricae;
si canimus silvas, silvae sint consule dignae.

Virgilio

Table of contents

Abstract

Tamarix plants are characterized by tolerance to extreme environmental conditions and represent an alternative resource for the recovery of marginal areas. However, their taxonomy is troublesome, and few molecular markers are available to enable species identification. Transcriptome sequencing projects offer a potential source for the development of new markers, named EST-SSRs.

Thirteen polymorphic simple sequence repeat (SSRs) markers derived from Expressed Sequence Tags (ESTs) from *Tamarix hispida*, *T. androssowii*, *T. ramosissima*, and *T. albiflonum* were identified and screened on 24 samples of *T. africana* to detect polymorphism. The number of alleles per locus ranged from two to eight, with an average of 4.3 alleles per locus, and the mean expected heterozygosity was 0.453. Amplification products of these 13 loci were also generated for *T. gallica*.

Tamarix plants were collected in seven sites from Italian islands, Central and Southern Italy: Imera, Simeto and Alcantara from Sicily, Crati from Calabria, Basento from Basilicata, Baratz from Sardinia, and Marangone from Lazio. It was performed a blind sampling, thus individuals were collected without any regard for species identity. Indeed, despite during our germplasm collection flowers for species identification were collected; a large set of samples remained unidentified. The number of individuals surveyed for site ranges from 24 to 84, with a total of 316 plants. The identities of 85 plants were determined with Baum's morphological keys, while all the rest (72% of the total) remained unidentified. An assignment method based on the use of molecular markers with a Bayesian statistical approach allowed the identification of individuals among species. It was found a clear assignment of *T. africana* individuals. Otherwise, it was not the same for *T. gallica* and *T. canariensis*, whose individuals were assembled together in a unique genetically homogeneous group (*T. gallica*-like), consistent with the hypothesis of introgression between species. Starting from a total of 221 unidentified plants the Bayesian approach assigned 142 individuals to *T. africana*, 78 to the *T. gallica*-like group, while 11 individuals (3.49%) were considered admixed and discarded for further analysis. The advantage concerning the use of molecular markers for classification purposes is that they can be used to assess differentiation across a wide range of taxonomic levels and address questions of species status by comparing inter-intra taxa differentiation. Thus, a high assignment power panel could be employed not only in the species studied in the present work, but could represent a new, fast and cheap method for species identification in all the taxa belonging to the genus *Tamarix*. It is noteworthy that only two loci, T1B8 and Ta1350, were required to achieve assignment to the species.

Three sites resulted monospecific stands of *T. africana* (Alcantara, Baratz, and Marangone), four stands resulted mixed with *T. africana* and *T. gallica*-like group contemporarily present in the same site (Crati, Imera Basento, and Simeto), while no any monospecific composition of *T. gallica*-like was found. The analyses of genetic structure of these populations pointed out the existence of a unique gene pool in Southern Italy for both *T. africana* and *T. gallica*-like, with populations characterized by low variability. On the other hand, *T. africana*

populations from Central Italy and Sardinia resulted more differentiated, despite the Mantel test for isolation by distance was not significant. It means that probably the genetic variability is affected by environmental factors, but it is not clear yet which features are involved.
At the best of our knowledge the present work is the first one regarding the characterization of genetic resources in Italian tamarisks.

Key-words: *Tamarix*, microsatellite, population genetics, taxonomy, EST-SSRs

Riassunto

Le tamerici sono caratterizzate da una elevate tolleranza a condizioni ambientali estreme e rappresentano una potenziale risorsa per il recupero delle aree marginali. Tuttavia, la tassonomia del genere *Tamarix* è problematica ed esistono pochi marcatori molecolari che possano essere utilizzati per l'identificazione delle specie. I progetti di sequenziamento del trascrittoma offrono, tuttavia, una risorsa per lo sviluppo di nuovi marcatori.
Sono stati sviluppati 13 nuovi marcatori microsatellite derivati da sequenze espresse (EST) di *Tamarix hispida*, *T. androssowii*, *T. ramosissima* e *T. albiflonum*. Questi marcatori sono stati testati su 24 individui di *T. africana* per evidenziarne il polimorfismo. Il numero di alleli per locus varia da due a otto con una media di 4.3 alleli per locus, mentre l' eterozigosi attesa media è pari a 0.453. L'amplificazione di questi 13 loci è stata inoltre verificata in *T. gallica*.
Le tamerici sono state raccolte in sette siti localizzati nelle isole e nell'Italia centrale e meridionale: Imera, Simeto e Alcantara dalla Sicilia, Crati dalla Calabria, Basento dalla Basilicata, Baratz dalla Sardegna e Marangone dal Lazio. È stato condotto un campionamento alla cieca in cui le piante sono state raccolte senza nessuno riguardo per l'identità specifica. Infatti, sebbene durante la collezione del germoplasma Italiano siano stati raccolti i fiori per l'identificazione specifica, non si è pervenuti all'identificazione di molti campioni. Il numero di individui collezionati per sito varia da 24 a 84 per un totale di 316 tamerici. È stata determinata l'identità specifica di 85 piante utilizzando la chiave morfologica di Baum, mentre le rimanenti piante (72%) sono state considerate indeterminate. L'identificazione specifica degli individui indeterminati è stata realizzata mediante un metodo di assegnazione basato sull'utilizzo di marcatori microsatelliti ed un approccio statistico di tipo Bayesiano. Gli individui della specie *T. africana* hanno presentato un chiara assegnazione. Mentre, non è stato ottenuto lo stesso risultato per le specie *T. gallica* e *T. canariensis* i cui individui hanno formato un unico gruppo omogeneo dal punto di vista genetico (*T. gallica*-like), coerente con l'ipotesi di introgressione tra le specie. Complessivamente gli individui indeterminati erano 221, il metodo Bayesiano ha asseganto 142 individui alla specie *T. africana*, 78 al gruppo ascrivibile a *T. gallica*, mentre 11 individui (3.49% del totale) non sono stati assegnati e quindi sono stati scartati dalle analisi successive. Il vantaggio dell'utilizzo dei marcatori molecolari negli studi di classificazione che questi possono essere utilizzati per evidenziare la differenziazione a diversi

livelli tassonomici, e possono chiarire lo status delle specie comparando la differenziazione all'interno e tra taxa. Un gruppo di marcatori caratterizzati da un alto livello di confidenza nell'identificazione specifica potrebbe essere utilizzato non soltanto nelle specie studiate nel presente lavoro, ma potrebbe rappresentare un nuovo metodo, poco costoso, riproducibile e veloce per l'identificazione delle specie appartenenti al genere *Tamarix*. È da sottolineare che i nostri risultati hanno evidenziato che per ottenere l'identificazione specifica sono sufficienti due loci, T1B8 e Ta1350.

Tre siti sono risultati essere formazioni monospecifiche di *T. africana* (Alcantara, Baratz, e Marangone), quattro popolamenti sono misti con la contemporanea presenza di *T. africana* e del gruppo ascrivibile a *T. gallica* (Crati, Imera Basento, e Simeto), mentre non sono state riscontrate formazioni monospecifiche del gruppo ascrivibile a *T. gallica.* L'analisi della struttura genetica delle popolazioni ha evidenziato l'esistenza di un unico pool genico presente nelle regioni del meridione d'Italia per entrambe le specie, con popolazioni caratterizzate da una bassa variabilità genetica. D'altro canto, le popolazioni di *T. africana* provenienti dall'Italia centrale e dalla Sardegna appaiono differenziate rispetto alle altre, nonostante il test di Mantel per l'isolamento per distanza non sia risultato significativo. Ciò significa che presumibilmente la variabilità genetica è affetta da parametri ambientali che tuttavia ancora non sono stati definiti.

Al meglio della nostra conoscenza, il presente lavoro è il primo riguardante la caratterizzazione delle risorse genetiche del genere *Tamarix* in Italia.

Parole-chiave: *Tamarix*, microsatellite, genetica di popolazione, tassonomia, EST-SSR

Introduction

The Mediterranean Basin is located in the transition zone between the northern latitudes and the desert belt. Due to this specific location, the present global warming is expected to result in lower precipitations and higher temperature. This may cause a general degeneration of Mediterranean forests, as water availability and quality are determinant for productivity and sustainability of plant, crop, and agro-forestry plantations. Conservation strategies represent crucial issue in this area which is one of the world's major centers of plant diversity (Médail and Quézel 1999). Nevertheless, some species of the Mediterranean Basin natural vegetation, such as *Tamarix* plants, thrive in zones where extreme climate conditions are regular phenomena; so, during their evolution, they have developed stress-adapted mechanisms. For these reasons this genus is likely to be well adapted to future conditions caused by the global warming effects. Moreover, these plants live in habitats that could become extremely fragile as subjected to flooding by the increasing sea level due to the melting of polar ices. In this scenario tamarisks represent an additional resources that should be protected and appreciate.

Since the last century, some specimens of tamarisks were introduced in North America and became invasive in riparian habitats with negative ecological and environmental impacts (Gaskin and Schaal 2002). Conversely, in Europe the genus *Tamarix* is native and does not show invasiveness, moreover it is considered interesting for its tolerance of abiotic stresses and for recovering of marginal areas.

At the best of our knowledge our work is the first contribution about Italian germplasm with molecular approach. Our work is a part of a wide collection of *Tamarix spp.* germplasm in Italy, where the main two species are *T. africana* and *T. gallica*, anyway even *T. canariensis* was found. The genetic resources for these species are very poor, and there is no any information about the structure of Italian natural populations.

The present work is included in a broader international project in collaboration with the Tel Aviv University (Israel), and other Ph.D. students were involved in the same project. In particular, Renée Abou Jaoudé was involved in germplasm collection and physiological characterization of tamarisks in response to salt and flooding stress, and Grazia Abbruzzese's

work was concerned the characterization of *Tamarix* micro-morphology of leaf and flower features in relation to environmental condition and taxonomy.

As described in literature, the taxonomy of this genus is troublesome, as the species cannot be distinguished at vegetative status and the floral traits used in the species identification could often be misleading, varying from season to season on the same individual (Gaskin 2003). For this reason, most of the *Tamarix* plants sampled during our germplasm collection in Southern and Central Italy remained unidentified. Unfortunately, only few molecular markers have been developed in *Tamarix* (Gaskin et al. 2006). For this reason, in the present work a novel set of gene based microsatellite markers has been developed. These markers are named EST-SSRs. Nonetheless, the increasing number of available sequences from large-scale transcriptome sequencing of *Tamarix* species offers a potential resource for a rapid and cheap development of these markers. Expressed sequence tag (EST) from four Asian species (*T. hispida*, *T. androssowii*, *T. ramosissima*, and *T. albiflonum*) were used for mining of EST-SSRs, and a cross-species amplification on Italian specimens was conduct.

These newly developed EST-SSRs represent additional resources for genetic characterization in this genus, and, with their large transferability, they could be used as new tool for species identification or to track the spread of invasive genotypes where tamarisks represent a threat.

Species identity represents the fundamental unit in ecology and evolution, although biological species definition is still matter of debate (Goldstein et al. 2000; De Queiroz 2007). Thus, it was realized a new method for species identification based on the use of molecular markers and a Bayesian statistical approach, instead of the classical method relied on morphological traits. It was found a clear assignment of *T. africana* individuals. Otherwise, it was not same for *T. gallica* and *T. canariensis*, whose individuals were assembled together in a unique molecularly homogeneous group (*T. gallica*-like). This result suggested the existence of a unique taxon or introgression between species, and it is consistent with a recent speciation scenario.

Microsatellite markers at the whole had a tremendous impact on population genetics and have became the genetic markers most commonly employed (Balloux and Goudet 2002). Population genetics describe the genetic structure of populations more or less connected by gene flow. Beyond, it is a starting point to suggest which evolutionary processes are actually involved in shaping population variability. Understanding gene flow and its effects is crucial for many fields of research including population genetics, populations ecology, conservation

biology and epidemiology (Balloux and Lugon-Moulin 2002). In particular, in this work it was found that Italian populations of *T. africana* are characterized by a great genetic differentiation between Southern Italy, Central Italy and Sardinia populations. Whereas, *T. gallica*-like showed a higher genetic diversity within populations and a low differentiation among populations.

The set of markers described in this work was tested even in T. *jordanis, T. tetragyna,* and *T. aphylla* from the desert of Negev (Israel), showing transferability and polymorphism. Even if only one genotype for species was available, we could demonstrate the transferability of our loci among species geographically distant, which, for this reason, are genetically different.

Chapter 1

Genomic resources in *Tamarix spp*.

1.1. The genus *Tamarix*

The family Tamaricaceae comprises three genera (*Tamarix*, *Myricaria*, and *Reaumuria*) and contains about 80 species. The genus *Tamarix* (the largest one) is composed by 54 species of shrubs or trees that can outcross, self pollinate, and propagate clonally from woody fragment (Gaskin and Schaal 2002). *Tamarix* plants, also called tamarisks or saltcedars, are broadly spread across many habitats with a great variability of environmental conditions. In fact, it native in Europe, Asia and North Africa and thrives in several riparian habitats like dunes, banks of rivers, and alluvial planes (Figure 1.1).

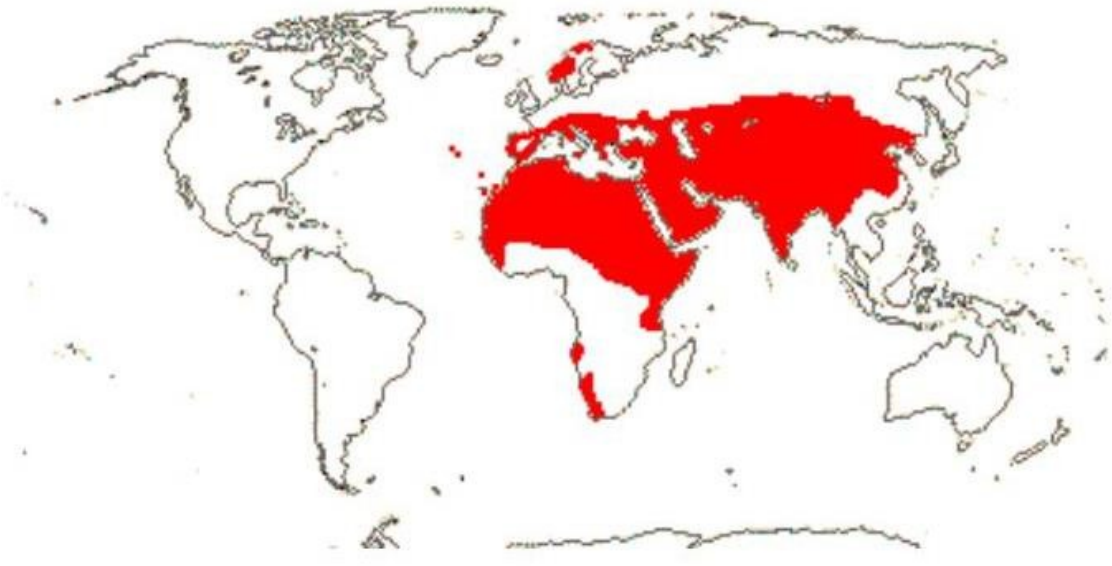

Figure 1.1: Native range distribution of the genus Tamarix. The centre of differentiation is located in central Asia.

The latest comprehensive revision of the genus by Baum (1978) contains three distinct sections, separated primarily by stamen number, petal length, androecial disk shape, and

position of filament insertion on the androecial disk. These sections are further divided into nine series based on morphological characters. In parenthesis are reported the species that will be studied in this work.

- Sections I Tamarix: series Gallicae (*T. gallica*, *T. ramosissima*), Leptostachyae (*T. arborea*, *T. canariensis*, *T. hispida*), Vaginates (*T. aphylla*),
- Sections II Oligadenia: series Laxae (*T. chinensis*), Anisandrae (*T. africana, T. tetragyna*), Arbusculae (*T. androssowii*), Fasciculatae,
- Sections III Polyadenia: series Arabiacae, Pleiandrae.

The taxonomy of the genus *Tamarix* is one of the most difficult among Angiosperms as its members offer few distinctive external features. The floral morphology include some characteristics which can be used for species identification, but they can be discerned after the dissection of the tiny flowers (Baum 1978). The inflorescence consists of solitary or compound racemes that occur on the branches, which are composed by clustered tetra or pentamerous small flowers. The width of the racemes show inter-specific variation, while the length is more variable even within the same species (Figure 1.2). Leaves are usually small, scale-like, and adpressed to the branches along their major axes, moreover, on the surface the presence of deep salt-secreting glands is common.

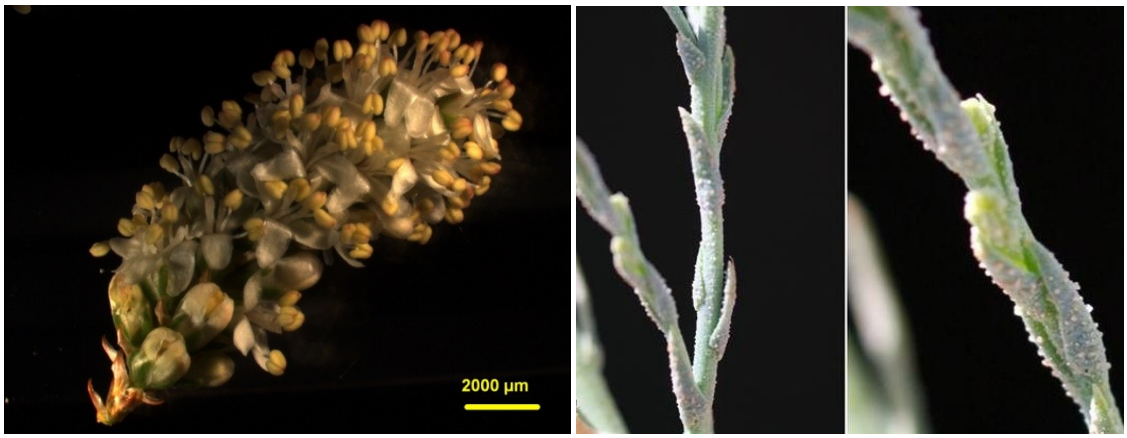

Figure 1.2: *T .africana* raceme (courtesy Grazia Abbruzzese) and salt glands in *T. tetragyna* leaves.

All the species belonging to this genus are deciduous with the except of *T. aphylla* which is the only one evergreen (Baum 1978).

1.2. The genus *Tamarix* in Italy

In 1984 De Martis and co-workers found 11 species in Sardinia, however the species *T. parviflora* was considered cultivated and become naturalized, while the species *T. nilotica* was found for the first time in Europe, but it was not reported in successive works. In fact, a later paper reports the existence of 10 species in Italy (Conti et al. 2005) that thrive in several natural habitats like dunes, banks of rivers, and alluvial planes. The main two species are *T. gallica* and *T. africana*, anyway even *T. canariensis* (Figure 1.3), *T. dalmatica*, *T. parviflora*, *T. arborea*, *T. hampeana*, *T. passerinoides*, and *T. tetragyna* were observed. In a recent work, Venturella and co-workers (2007) added two specimens usually used as ornamental *T. rosea* and *T. chinensis* that did not have been found before in the wild, moreover, they did not find *T. dalmatica* during their investigation in Sicily.

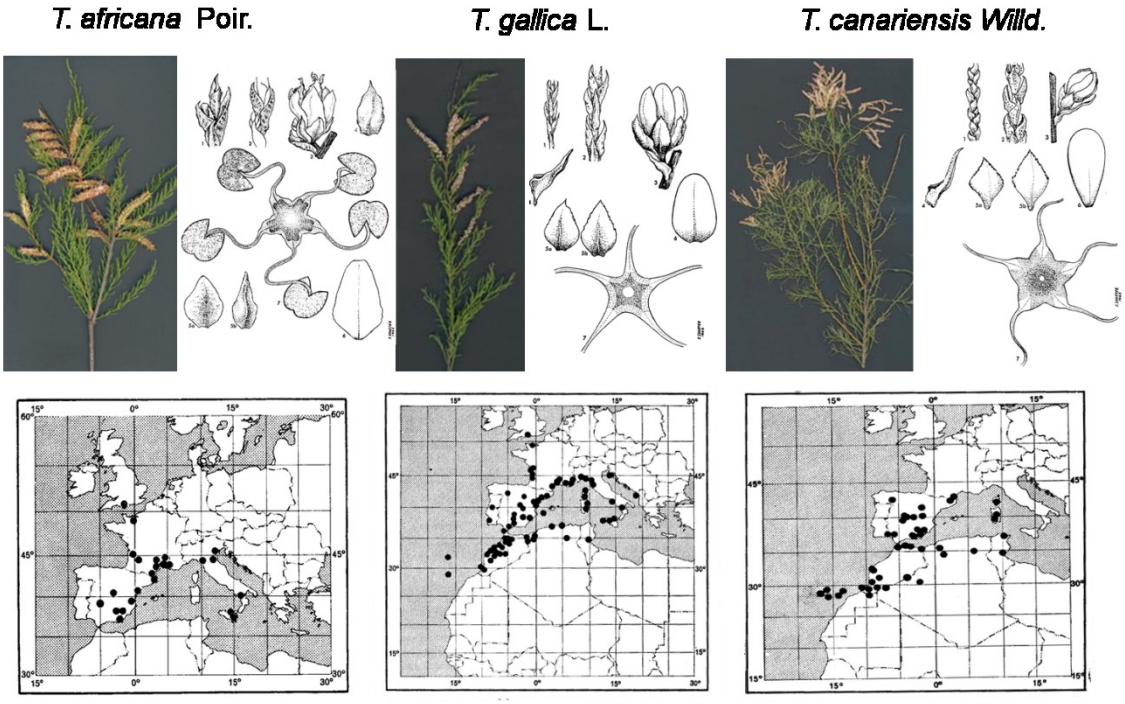

Figure 1.3: Flowering branch, floral features and native range distribution of *T. africana*, *T. gallica*, and *T. canariensis* which are the species more deeply investigated in this study (modified from Baum, 1978).

Despite sometimes the genus *Tamarix* is considered a treat as it became invasive in North America, in the native range it does not show any trait of invasiveness. On the other hand, it is considered a very interesting genus for its tolerance to abiotic stress and for the recovery of degraded lands. Anyway, no information is available about the genetic structure of the Italian natural germplasm of *Tamarix*.

1.3. Ecological features

Tamarix species are pheatophytes, characterized by deep roots that reach the water table for their water supply. However, under particular conditions they can grow where the groundwater is not accessible, thus *Tamarix* are defined facultative phreatophyte rather than obligate phreatophyte (Zhang et al. 2002). Temperature has strong effect on plant growth and survival limiting the geographical distribution of the species (Sexton et al. 2002), in fact, for instance *T. aphylla*, *T. nilotica*, *T. dioica*, *T. macrocarpa* show high temperature requirement, otherwise other species inhabit temperate regions withstanding freezing temperatures of the European winters. The species also differ in their degree salt tolerance, in fact, some species are considered extreme halophytes (e.g., *T. chinensis*, *T. hispida*) while others are not (Baum 1978).

1.4. Uses

Tamarix species were used as wind-breaks, hedges in desert conditions or dunes fixers along costal zones as some species are tolerant of salt spray, moreover the deep root allow to use *Tamarix* plants for the fixation of river banks to prevent erosion (Baum 1978) (Figure 1.4).

Figure 1.4: Tamarisks are able to survive in extreme environmental conditions

Recently this adaptation to an extreme range of environmental condition pointed out the suitability of tamarisks to be employed for the reforestation of degraded lands that might be irrigated by saline water or by reused urban and industrial water waste.

1.5. *Tamarix* invasiveness

Since the last century around ten species of *Tamarix* were introduced in north America to be use for shade and erosion control, and became invasive in riparian habitats with negative ecological and environmental impacts. The ability of invasive plants to compete and proliferate can be caused by intrinsic factor such as physiological or reproductive capacities, and to extrinsic factor such as a loss of competitors, herbivores, or pathogens upon introduction. An additional influence may be due to the creation of new genotypes; in fact, the introduction of species in a new region can provide opportunities for hybridization events, bringing into contact related species that previously were isolated (Gaskin and Schaal 2002). Once naturalized recombination of genetic variation can provide a range of hereditable phenotype to respond to local selection pressures and produce offspring with a greater fitness (Sexton et al. 2002).

Many species have become noxious due to their invasive habit and competition with neighbouring plants, and even because they impoverish water sources thought transpiration, lowering the water table and altering stream and river dynamics (Gaskin 2003). Salt secretion may led to the desalinization of deeper soil layers, while increasing the salinity of upper soil layers, often the soil surface being covered year by year by a layers of twiggy salty litter (Baum 1978). In fact, once established, *Tamarix* can tolerate drought by utilizing deep ground water sources. Moreover they can exclude excess salt from salinized water sources from glands in their scale-like leaves, which seasonally are dropped, forming a saline duff on the soil surface that inhibits the germination of other plants. Thus *Tamarix* have also a negative role being a strong competitor which can form monospecific stand that decrease biodiversity (Zhang et al. 2002).

1.6. Molecular systematics

Recently studies about the molecular systematic of *Tamarix* (Gaskin and Schaal 2002; Gaskin 2003) pointed out that the molecular analyses did not support the three taxonomic sections identified by Baum, and morphological traits are often misleading as a means of identifying specimens. Gaskin and Schaal (2003) used both nuclear and chloroplast markers to compare the evolutionary dynamics of the maternally and biparentally inherited genome, allowing the investigation of putative hybridization within the genus *Tamarix*. In that study

they analysed several *Tamarix* species collected both across the native range (Europe, Asia, southern Africa) and in north America. The molecular phylogenetic analysis was performed using both nuclear ribosomal ITS and chloroplast *trn*S-*trn*G intergenic spacer sequence data, but the results were incongruent with earlier partitioning of the genus. For example, some species that in Baum's classification belong to different sections, have identical phylogeny both in nuclear and chloroplast genome. Moreover, they found incongruence between nuclear and chloroplast evolutionary histories and not all species could be distinguished with these molecular markers. In particular, in this work *T. gallica* and *T. canariensis* had the same chloroplastic DNA sequences (Figure 1.5), otherwise following the nuclear sequences data these species fell into two clades, the first composed by *T. gallica* and *T. canariensis* alone, and the other one include *T. africana* as well. Gaskin and Schaal (2003) suggested that the inability of their molecular markers to distinguish between *T. canariensis* and *T. gallica* may be due to these species being the same taxon or the species are introgressive.

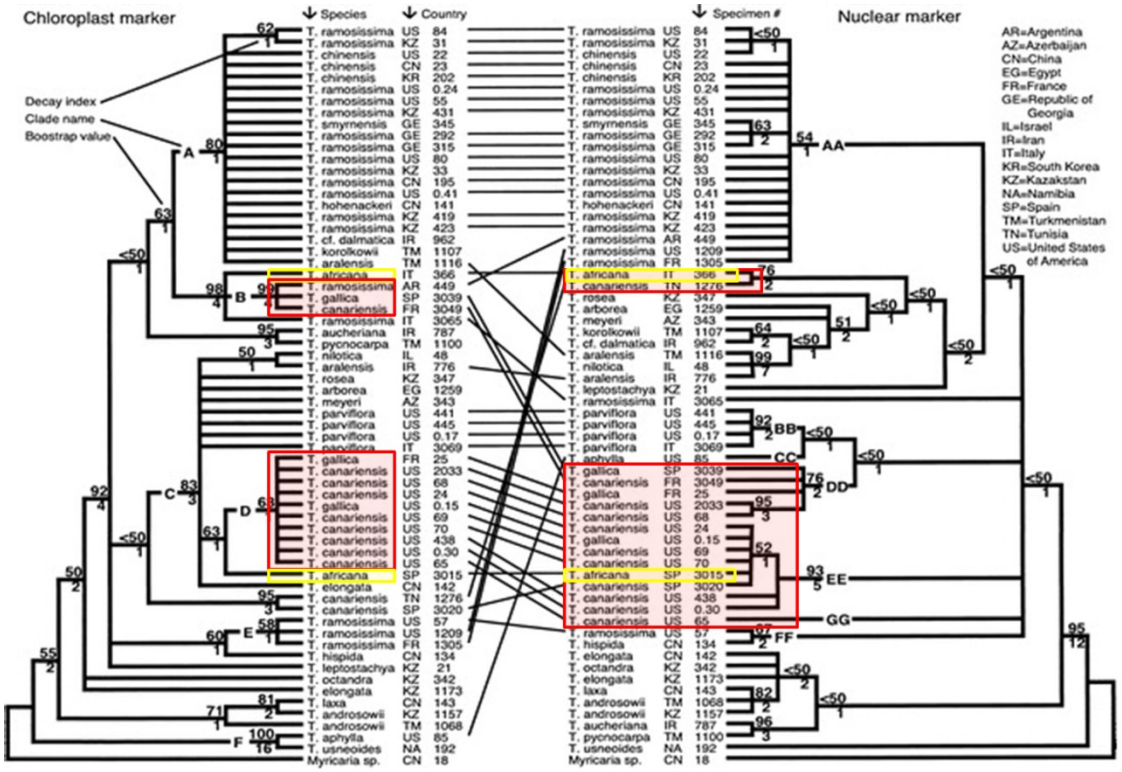

Figure 1.5: Chloroplast and nuclear marker gene trees reported in Gaskin and Schaal (2003).

1.7. Hybridization

Hybrids formation is a common phenomenon in plants both in the wild and under cultivation, it is caused by the formation of gametes from two different entities (species, subspecies, etc) (Weising et al. 2005). Gaskin and Schaal (2002) conducted a study on the spread of invasive population of *Tamarix* in USA. In that study nuclear sequence form an intron of phosphoenolpyruvate carboxylase (PepC) gene was used and vouchered *Tamarix* species both from native and invasive range were collected. This study showed that within the USA invasion many of the plants were novel hybrids between *T. chinensis* and *T. ramosissima*. In fact, it was found that the most common genotype is a hybrid combination between two species-specific haplotype that were geographically isolated in *Tamarix* native range in Asia. Moreover, they observed a lower genetic diversity than the native rage, suggesting the hypothesis of founder events or genetic bottleneck. In a successive work Gaskin and Krazmer (2006) tried to establish a relationship between ornamental and wild saltcedars by means of chloroplast *trn*S-*trn*G intergenic spacer sequence and the fourth intron of phosphoenolpyruvate carboxylase (PepC) gene. This second study confirmed the hybrid origin of invasive *Tamarix* in USA and indicate that the wild genotype originated from wild plants, as chloroplast and nuclear genotypes found in ornamental plants were dissimilar from genotypes found in nearby wild stands of *Tamarix*. The same authors studied the introgression of invasive saltcedars in USA with AFLP markers including both plant material from Asia and USA (Gaskin and Krazmer 2009). They found that *T. ramosissima* and *T. chinensis* plants from Asia were genetically distinct, and that the USA plants were genetically intermediate the parental species, forming a continuum between the parental genotypes. A recent study on *Tamarix* cold hardiness was conducted on native *T. chinensis* and *T. ramosissima* and invasive population from USA with microsatellites markers (Friedman et al. 2008); the authors pinpointed the lack of a strong genetic isolation among invasive populations, with a gradual trend from individuals that resemble to *T. chinensis* in the southern to individual that resemble to *T. ramosissima* in the northern latitude. In that study the absence of genetic isolation and the similarity with the parental species were observed, supporting the hypothesis of a post-introduction hybridization, which may allowed the creation of a greater genetic variability with respect to cold hardiness than any of the originally introduced population.

1.8. Expressed Sequences Tag analysis in *Tamarix*

The knowledge of the distribution of genetic variability between and within natural plant populations is essential to adopt competent strategies for *ex situ* and *in situ* germplasm conservation, for molecular breeding or for detecting the spread of invasive specimens. Anyway, only 10 neutral microsatellite markers have been developed for *T. chinensis*, *T. ramosissima* and their hybrids, but it is worth to note that the transferability of microsatellite loci across different species is not complete, since, usually, the primer sequences are species-specific (see Chapter 2). Nonetheless, the increasing number of available sequences from large-scale transcriptome sequencing of *Tamarix* species offers a potential resource for rapid and cheap development of new markers, EST-SSRs. As mentioned above tamarisks are one of the most remarkable salt-tolerance woody plant species, however, only recently studies on their stress resistance mechanisms have been conducted. Two species from China were investigated, *T. androssowii* and *T. hispida*, for the construction of cDNA libraries which lead to a better understanding of the gene expression profiles manifested in response to stress. In these studies an expressed sequence tags (ESTs) analysis was conducted. It is an effective method in discovering novel genes that is a hard topic especially in *Tamarix*, as in this genus, the genomic data are poor.

In the first work, Wang and co-workers (2006) constructed a cDNA library of *T. androssowii* stressed by $NaHCO_3$, and obtained 2455 EST sequences. The authors found about 400 differential expressed genes in response to $NaHCO_3$ stress, which were analyzed using BLASTX, and the ESTs of known putative function were further grouped into 12 functional categories. The most abundant ESTs were involved in defence and photosynthesis and a high portion of them were identified as salt stress related genes in *T. androssowii.* The authors compared the gene expression pattern of *T. androssowii* with those of other species to allow a better understanding of transcripts profile under salinity stress, and they found that metallothionein-like (MTL) proteins, lipid transfer protein (LTP), and Germin-like proteins (GLPs) were significantly expressed suggesting a role in the salinity tolerance processes in plants. It is worth noting that even Late embryogenesis abundant protein (LEA), a well known osmotic stress related gene, was found in the EST collection (Wang et al. 2006). This gene was cloned and introduced in tobacco, where, the over expression of LEA gene conferred high tolerance to salt, dehydration and cold (Zhao et al. 2011). Moreover the differential expression of genes involved in defence from reactive oxygen species (ROS) were observed during the construction of *T. androssowii* cDNA library (Wang et al. 2006). Recently, these

observation were employed for the transformation of transgenic poplars with a manganese superoxide dismutase (MnSOD); the overexpression of this gene conferred enhanced salt tolerance to a hybrid poplar (*Populus davidiana* X *P. bolleana*) (Wang et al. 2009).

A second work by Gao and co-workers (2008) produced a cDNA library in response to $NaHCO_3$ and NaCl in *T. hispida*. They obtained 9447 expressed sequences from leaf tissue after increasing time of exposure to $NaHCO_3$ stress. Once identified the genes responsive to $NaHCO_3$, nine of them were further investigated and the expression patterns in response to $NaHCO_3$ and NaCl were compared by real time RT-PCR. The authors found that the expression patterns of *T. androssowii* and *T. hispida* were very similar, in fact, even in this later work an increased expression level of LEA genes, LTP and ROS scavenging (MTLs) genes were observed. The authors observed that in *T. hispida* the short-term response in $NaHCO_3$ and NaCl stresses were similar, but prolonged stress triggers different mechanisms in saline or saline-alkali stress.

Moreover, in a recent work (Dong et al. 2007) the gene encoding the plasma aquaporin of *T. albiflonum* was cloned an sequenced. Aquaporins are water selected channels involved in seed germination, cell elongation, stoma movement and also play a role in drought stress response. The authors obtained the sequence of the *T. albiflonum* aquaporin (AQP) from a subtractive hybridization library constructed under drought stress and conducted a BlastX homology search to assess the homology with kwon genes. The comparative molecular analysis of the nucleotide sequences of *T. albiflonum* AQP showed 95% homology with the gene of *Arabidopsis thaliana*.

Each of these works provided a large number of expressed sequences that were deposited in publicly available gene bank, that enriched the scarce genetic resources in *Tamarix* and could allow a better knowledge of stress tolerance in woody plants. Even if these sequences belong the species that are not present in Italy, the expressed sequences promise for the development of molecular markers (Ellis and Burke 2007).

Chapter 2

Microsatellites or SSRs (Simple Sequence Repeats)

2.1. Definitions and applications

Microsatellites, also known as simple sequence repeats (SSRs), consist of short motifs of 1 to 6 nucleotide tandem repeats stretches of DNA, that are widely spread across the genome occurring both in coding and non-coding regions (Pashley et al. 2006). SSRs located in non-coding portion of the genome are defined neutral as they are not subjected to environmental selection, but it is clear now that microsatellites could be located also in expressed sequences within both the coding and the untranslated regions. (Ellis and Burke 2007). Moreover, microsatellites could be classified according to their type of repeat sequence as perfect, imperfect, interrupted or composite. In a perfect microsatellite the repeated sequence is not interrupted by any base not belonging to the motif, while in an imperfect microsatellite there is a pair of bases between the repeated motif that does not match with the motif sequence. In the case of an interrupted microsatellite there is a small sequence within the repeated sequence that does not match with the repeated sequence, while in a composite microsatellite the sequence contains two adjacent distinctive sequence-repeats (Figure 2.1) (Bhargava and Fuentes 2010).

Figure 2.1: A. Examples of perfect microsatellites composed by mono-, di-, tri-, tetra-, Penta-, and Hexanucleotide repeats, respectively. B. Examples of perfect, imperfect, interrupted and compound microsatellites (modified from Weising et al. 2005).

Microsatellites are highly polymorphic regions and the polymorphism between individuals arises from the different number of repeated units within the sequence (Figure 2.2).

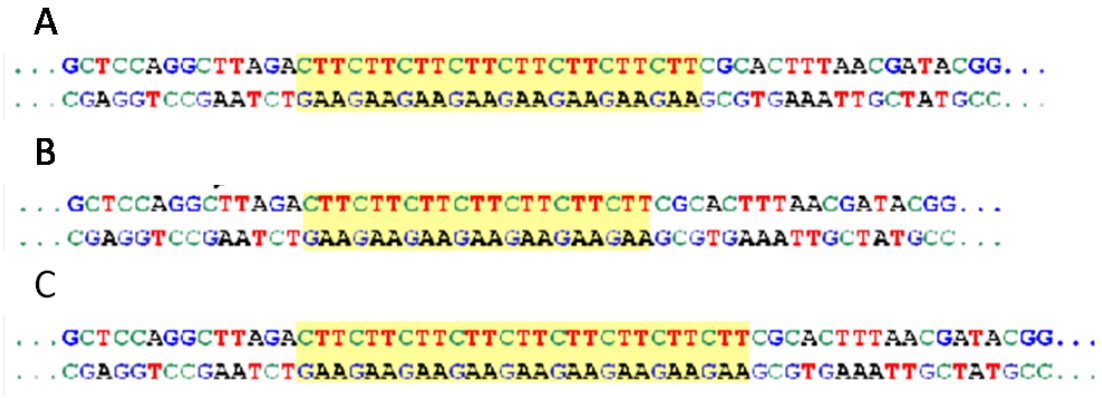

Figure 2.2: Three different alleles of a tri-nucleotide microsatellite where the allele A has eight repeated units, B seven repeated units, and the longest C has nine repeated units.

So, an homozygous microsatellite locus has the same number of repeats on both homologous chromosomes, whereas a heterozygous microsatellite locus has different number of repeats for each allele.

Otherwise, at the same locus a population usually contains several alleles, each with a different number of repeats, which means that microsatellite markers could be very useful for discriminating different individuals. According to the information provided about heterozygous, molecular markers are classified in dominant and codominant markers. The advantage of codominant markers over dominant markers is the differentiation between homozygous and heterozygous individuals that makes the analysis and the interpretation of both kinds of markers very different. The employment of SSR markers shows different advantages correlated to their hypervariability, codominant inheritance, multiallelic nature, extensive genome coverage and simple detection by PCR reaction. Microsatellites are locus-specific and sequence specific markers, thus their development is more difficult with respect to the dominat markers. In fact, unfortunately the development of SSRs is expensive and time-consuming and frequently the PCR primers used to amplify these loci are often species specific, so the marker developed in one taxon could not be readily transferred to another one. Anyway, codominat markers are particularly precious since they are orthologus, meaning that they derive from a common ancestral locus that makes them suitable in comparative genomics analysis, population genetics, parentage analysis and estimate of gene flow.

2.2. Genomic distribution

Microsatellites are not regularly distributed within a single genome due to differences in their frequencies within coding and non-coding regions and the possible functional roles of different repeats. Despite microsatellites have ubiquitous occurrence both in eukaryotic and prokaryotic genomes, their density and distribution could vary markedly across the genome (Li et al. 2002; Sharma et al. 2007); and their frequency could vary per taxon in terms of absolute number of microsatellite loci and preferential repeats. Eukaryotic genomes are characterized by the prevalence of mononucleotide repeats over the other classes of microsatellites repeats (Sharma et al. 2007); while in prokaryotic genomes microsatellite frequency is relatively low and tri nucleotide microsatellites are the most common ones. In plants, a positive and linear relationship between microsatellite frequency and percentage of

single-copy DNA was observed (Morgante et al. 2002), which suggest that microsatellites frequently occur within and near genes (Bhargava and Fuentes 2010).

Microsatellites are broadly used as molecular markers since they suffer higher rates of mutation than the rest of the genome (Oliveira et al. 2006). High mutability at microsatellite loci has a role in genome evolution by creating genetic variation within a gene pool and may provide an evolutionary advantage of fast adaptation to new environments (Li et al. 2004). Although, all types of SSR repeats appear to be more common in non-coding regions, a large number of microsatellite are located in protein-coding regions. In general they are less frequent with exception of tri- and hexanucleotide microsatellite repeats that are found to be in excessive numbers over the others repeats unit size. Probably, the differences between the non-coding and the coding SSRs frequencies arise from specific negative selection against frameshift mutations in coding regions resulting from length changes in nontriplet repeats (Li et al. 2002), that could generate a mutation pressure or a positive selection for specific amino acid stretches. In fact, since the RNA bases are read as triplets, the selection against mutations, that change the reading frame of a gene, restrict the presence of microsatellite in coding regions. So, while microsatellites with repeats in multiples of three develop evenly in both coding and non-coding regions, the potential expansion of di- or tetra- nucleotide repeats microsatellite at the coding or the untranslated region could lead to disruption of the original protein and the formation of new genes by frameshift (Li et al. 2002). Changes in lengths of triplet or amino acid repeats could affect protein function, as frameshift within coding region caused by microsatellite expansion or contraction may cause gain or loss of function, gene silencing or induce pathogenesis. Anyway, as explained above the repeat elongation/shortening processes also lead to the increase of biological complexity, which is considered to be crucial for biological evolution.

2.3. Functional perspectives

Microsatellites can have either selectively neutral effect on the genome or affect crucial regulatory gene functions, but their actual role in plant genes is still poorly understood (Kalia et al. 2011). Moreover, there are some evidence that genomic distribution of microsatellites is non-random, presumably because of their effects on chromatin organization, regulation of gene activity, recombination, DNA replication, cell cycle, and mismatch repair (MMR) system (Figure 2.3). Interestingly, members of some gene groups have preferential

distribution of microsatellites in specific gene regions. For example, members of the "transcriptional factor activity" group have more microsatellites than expected in all regions except introns, otherwise genes involved in transport have more microsatellite than expected in introns and 3' regions (Sharopova 2008). Although, usually in literature microsatellites are considered evolutionary neutral markers, it is clear that they can have either a neutral effect on the genome or perform important functions (Oliveira et al. 2006), and the microsatellites location in the genome determines their functional role (Kalia et al. 2011).

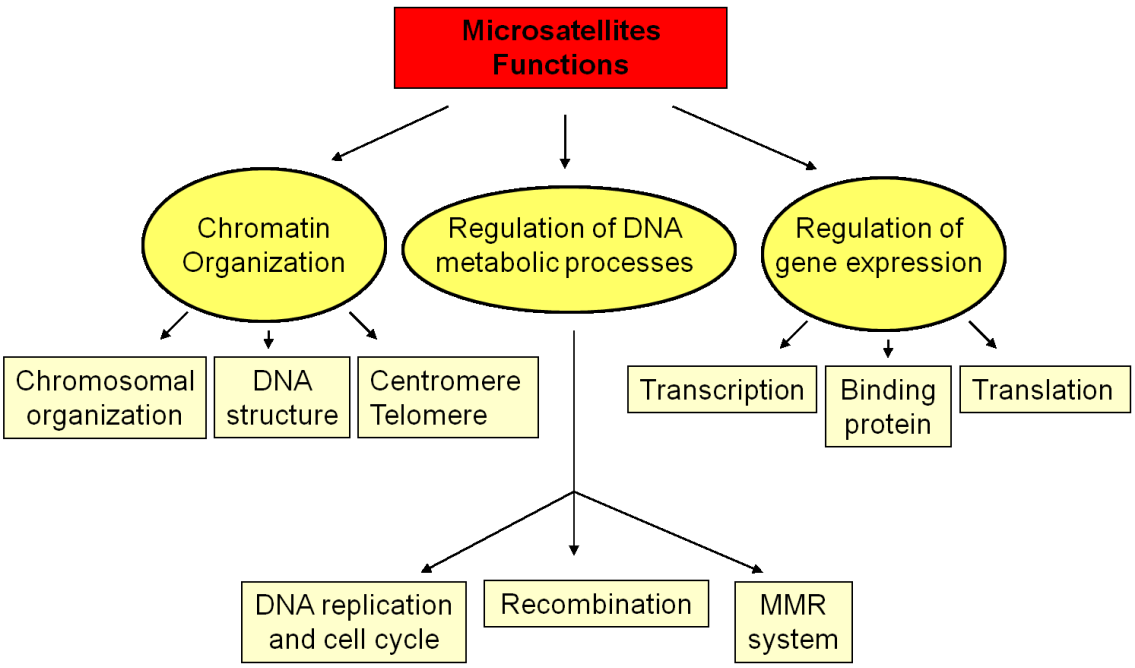

Figure 2.3: Microsatellites putative function (modified from Li et al. 2002).

2.3.1. Chromatin organization

Chromosomal organization. Several aspects of microsatellite distribution suggest a special role as possible ancient genomic component of taxon-specific chromosome structure. Moreover several evidences indicate that repetitive elements participate in the packaging or the genome, allowing the expression of coordinately regulated genes in a cell type-specific manner (Figure 2.4) (Kumar et al. 2010).

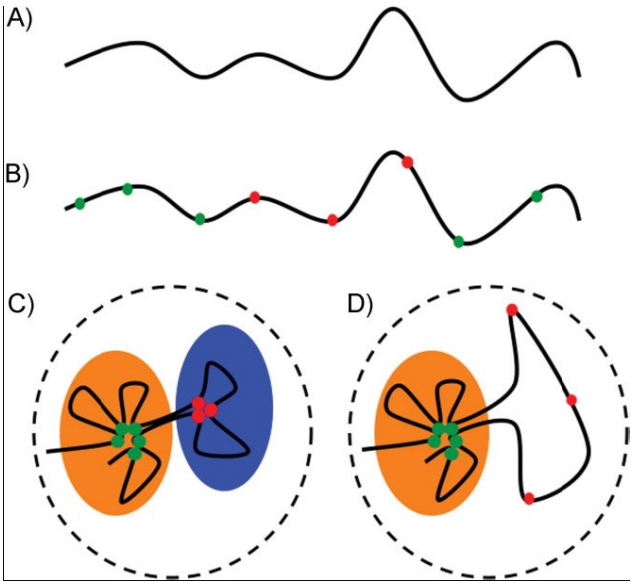

Figure 2.4: Repeat-dependent cell type-specific genomic packaging code. A: Genomic DNA. B: Genome marked with different repeat elements indicated as different coloured spheres. Repeat-specific DNA-binding proteins can allow chromatin loop formation by bringing the repeats together. C: Different kinds of repeat can interact with the help of different set of proteins and cluster the linked loci to different nuclear compartments. D: In a different cell type where a particular repeat-specific DNA-binding proteins is absent, the clustering can be altered, leading to a different conformation (Kumar et al. 2010).

DNA structure. Microsatellites are able to form a wide variety of unusual DNA structure with simple and complex loop-folding patterns. The formation of such stable structure offer an advantageous mechanism during transcription, providing unique protein recognition motifs. Moreover, the repeats number seems to be a critical parameter that determines the balance between the advantage gained and the disadvantage during replication posed by these structure.

Centromere and telomere. In many species the centromeric region of chromosomes is characterized by the presence of a large number of microsatellites. The assembly of divergent tandem sequence into chromosome specific higher order repeats appears to be a common organizational feature of many organism centromeres, and also suggest that evolutionary mechanisms that creates higher order repeats are conserved among their genomes. Telomeric repeats are located at the extreme ends of eukaryotic chromosomes and represent a special version of microsatellites (Weising et al. 2005).

2.3.2. Regulation of DNA metabolic processes

Recombination. Recombination events are not evenly distributed along the chromosomes rather occur nonrandomly, often clustered in regions called "recombination hot spots" (Ellegren 2004). Several authors have suggested that both the repeat motif and the repetition number of microsatellites play a role as recombination hot spots. In particular dinucleotide repeated microsatellite are proposed as preferentially site for recombination since their affinity with recombination enzymes (Oliveira et al. 2006). Thus, microsatellites can act as evolutionary switches that modulate the mutation rate under condition require rapid evolution and allow the population to respond quickly to changing environmental conditions (Li et al. 2004). Moreover the association between microsatellite and recombination hot spots is stronger for longer repeats, while it is absent or weak for repeats with less then six repeats.

DNA replication and cell cycle. Microsatellites could affect enzymes controlling cell cycles. In fact, during DNA replication as it was observed that some tracts act as arrest site; while in other loci a loss of cell cycle control was observed if mutation occurs.

Mismatch repair system. DNA MMR proteins correct replication errors and actively inhibit recombination between divergent sequences, thus controlling mutation rate and evolutionary adaptation. Anyway the effectiveness of MMR system is strongly influenced by the genomic position and the DNA surrounding the mismatch, moreover the MMR genes are vulnerable to spontaneous indel mutations of a mononucleotide microsatellite within their coding region resulting in MMR deficiency. It was observed that if MMR genes mutate, microsatellite instability increase (Ellegren 2004).

2.3.3. Regulation of gene expression

Microsatellite and transcription. The occurrence of microsatellites within the genes is correlated with changes in the expression of many genes; in particular, the effect of microsatellite on the average transcript level depend on its position within the transcribed region. Microsatellites in introns and 3' region are often associated with low expression levels, while if the repeats are located in the 5' region genes have higher than average transcript levels (Shaporova 2008). Moreover, there are several evidences showing that microsatellites located in the promoter regions could affect gene activity, as expansion or deletion of the repeats alters the transcriptional activity of the promoter itself. The number of

repeats appears to be a key factor for gene expression and expression level. Moreover, some genes can be expressed within a narrow range of SSR repeats number, and out of this range gene activity would be turned off; while another group of genes show adjusted expression levels by changing their regulatory microsatellite repeat numbers in a more wide range (Li et al. 2002).

Protein binding. Some microsatellites found in upstream activation sequences, serve as binding sites for different regulatory proteins. For instance, the regulation of several genes depends on the binding of the GAGA transcription factor to a small fragment of a CT microsatellite presents in the promoter region of the gene. Moreover, repeat-specific DNA-binding proteins contribute to coordinately regulation of genes as allow chromatin loop formation by bringing the repeats together and clustering the linked loci into different nuclear compartments. Thus microsatellites can create a specific "microlandscape" for a particular group of genes by changing DNA structure locally (Kumar et al. 2010). Different kinds of repeats can interact with a large set of DNA-binding protein leading to different chromatin organization and to different expression pattern (Figure 2.5).

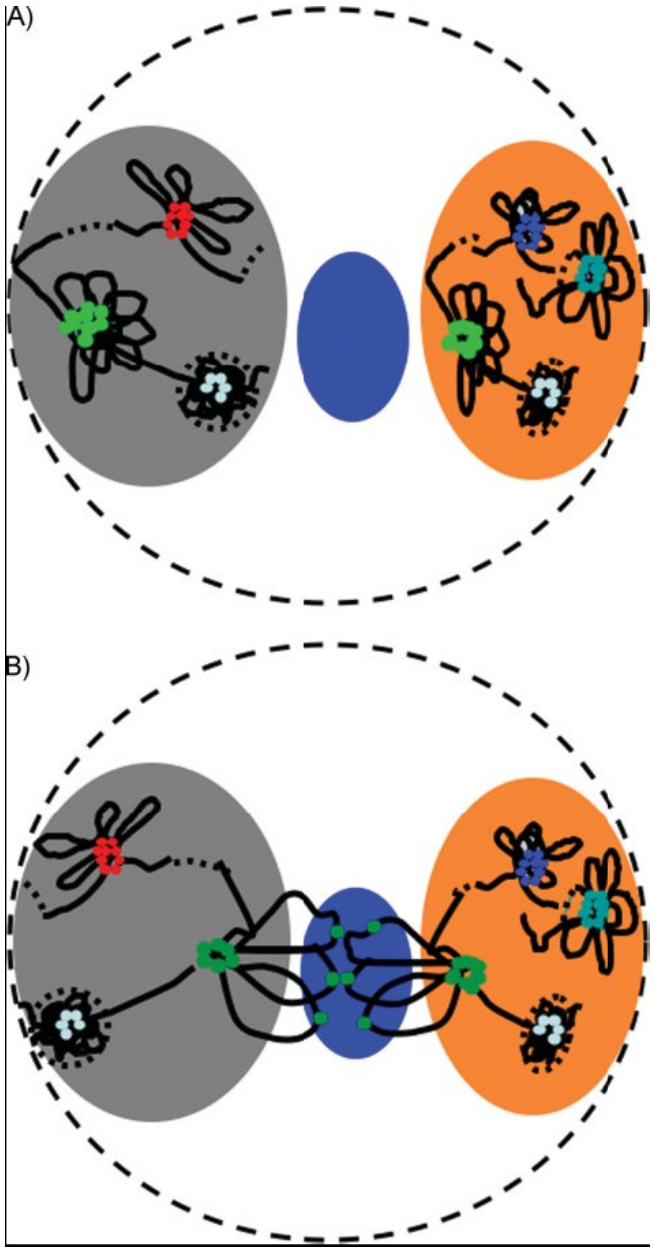

Figure 2.5: Long-range interactions mediated by repeats. Repeats can organize chromatin into specific compartments. A: Similar kinds of repeat-associated loci are shown in same colour. B: Associated loci, due to differential expression of corresponding repeat-binding protein (e.g., repeat-interacting protein, green) can allow movement of the associated loci to a common compartment. Such chromatin movement can have regulatory consequences for the associated genes (Kumar et al. 2010).

Translation. Microsatellites are involved in gene translation as expanded repeats may form altered DNA secondary structures that confer genetic instability and most likely contribute to transcriptional silencing (Li et al. 2004). Microsatellites can form a several types

of DNA structure and their ability is proportional to the length and the composition of the repeated region (Shaporova 2008).

2.4. Mutation rates and mechanisms

Despite microsatellites have been extensively studied, the mutational dynamics of these regions is not well understood; in fact, their mutation rate is much higher than other part of the genome, ranging from 10^{-2} and 10^{-6} nucleotides per locus per generation (Oliveira et al. 2006). There is no uniform microsatellite mutation rate; the rates tend to differ among loci, alleles, and even among species (Ellegren 2004). The mutation rate at a microsatellite depends in part on its intrinsic features and the repeated motif. In general microsatellites with a large number of repeats are more mutable probably due to increased probability of slippage, since larger number of repeats provides more opportunities for misalignment during the reannealing of the nascent strand. Moreover, longer stretches of repeated units pose more problems to polymerase than shorter ones, making longer alleles more prone to slipped strand mispairing. Moreover, mutation rate of microsatellites equal in length have been found to be inversely proportional to their motif size (length of the repeated unit in base pair). A threshold of minimum repeats seems to be required before a microsatellite locus could become hypervariable, but the association between mutation rate and size of repeated region still remained a matter of debate (Bhargava and Fuentes 2010). Even if a threshold size of eight nucleotide irrespective of different motifs was estimated, many workers disagreed and suggested that no critical point exist for microsatellite extension. Mutational rates may not only vary among repeated types, base composition of the repeat and microsatellite type (perfect, imperfect or compound), but also among taxonomic group. Additional factors like repeated motifs, allele size, chromosome position, cell division, sex and the GC content in flanking DNA have been found to affect the rate of mutation at SSR loci. Another important factor is the heterozygosity; in fact, it was observed that microsatellite alleles of any given length are more likely to mutate when their homologue is unusually different in length. Due to the high mutation rate it is to be expected that coding regions have low microsatellite density, otherwise these regions would be significantly altered, possibly leading of loss of functionality SSRs located within genes also show a higher mutation rate with respect of non-repetitive regions, even if they are less variable with respect to the microsatellites in non-

coding regions. Several hypothesis about the evolutionary dynamics of microsatellites have been suggested to explain their high mutation rate:

- *Point mutation* consist in errors during recombination that anyway did not like to be the predominant mechanism in the generation of microsatellite variability.
- *Unequal Crossing-over* that can lead to drastic changes in the number of repeats.

This mechanism occurs since the presence of a microsatellite could generate a hairpin during synapsis, meaning that each chromosome homologous received a different number of repeats (Figure 2.6). A possible explanation of high levels of microsatellite variability in recombination hot spots is that recombination is mutagenic to microsatellite, promoting the generating and maintenance of polymorphic repeats in tracts of high recombination.

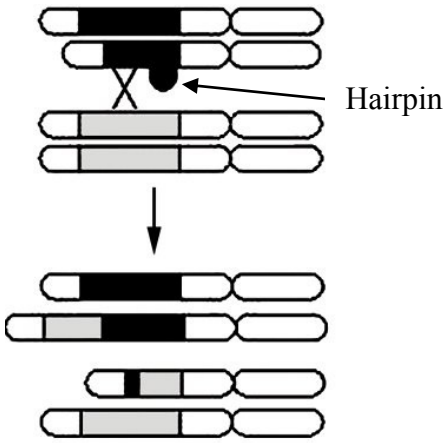

Figure 2.6: unequal crossing-over between homologous chromosomes, where black and gray regions correspond to repeated sequences (Oliveira et al. 2006).

- *Replication Slippage* could occur during replication or repair and occurs when one DNA strand is temporary dissociated from the other and rapidly rebinds in a different position, leading to base-pair errors and continued lengthening of the new strand.

This may result in an increase in the number of repeats in the allele if the error occurs on the complementary strand or a decreased number of repeats if the error occurs on the parental strand (Figure 2.7). Despite the majority of these slippage events involve the addition/deletion of a single repeat unit, high incidence of multiple repeat units has been observed.

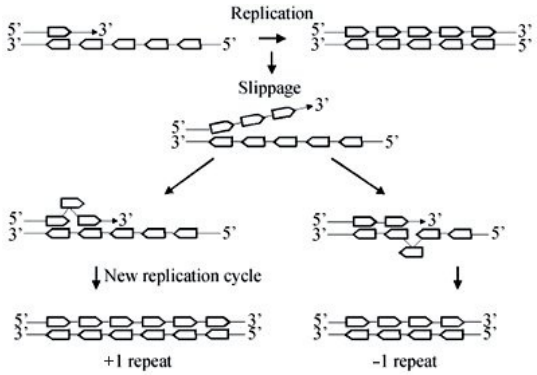

Figure 2.7: Slippage during replication, the repeats are represented by box. Slippage leads to the formation of new alleles and the gain or floss of repeats depends on which strain contain the polymerase error (Oliveira et al. 2006).

Some recent studies have suggested that equilibrium distributions of SSR repeat length are a result of balance between slippage events and point mutation (Ellegren 2004). Whereas replication slippage favours growth, point mutation break down a long array into a shorter ones. In fact, mutation within the repeated regions causes an interruption that slip the original repeat into two shorter ones, which would increase locus stability by reducing the substrate for polymer slippage and recombination. The relative rates of point mutation and slippage might be altered by differences in genome structure or organization between species as well as by changes in the efficiency of MMR and proofreading during DNA replication. Although MMR has been found to have a greater effect on SSR stability with respect to proofreading, it should be noted that the effectiveness of MMR proteins is influenced by the presence of a microsatellite in their coding region (see paragraph 2.2.3).

2.5. Origin

As described previously, the genesis of microsatellites in genomes appear to be non-random, with an imbalance between mechanisms that prevent and those that promote the initiation of microsatellites. Current data suggest two alternative, but not mutually exclusive, hypothesis to explain microsatellites genesis:

De novo microsatellites which arise spontaneously from/within unique sequences

Adopted microsatellites that are brought about in a primal form into a repetitive genomic location by mobile elements

De novo microsatellites hypothesis assumes that a minimum of repeats (proto-microsatellite) is required before DNA polymerase slippage can extend the number of repeats. A proto-microsatellite is a short intermediate stage with as few as 3-4 repeat units, within cryptically simple sequence, which are defined as a scramble of repetitive motifs lacking a clear tandem rearrangement. These proto-microsatellites were initially thought to originate from base substitutions and served as substrate for further expansion. However, subsequent studies showed that proto-microsatellites are often formed without preceding nucleotide substitution, but rather through indel events. A study on insertions in human genes demonstrated that the processes leading to the expansion of microsatellite loci occur even with few repeats. In fact, it was observed that often nucleotide insertions resulted in new repeats, and most of them are not extensions of pre-existing repeats but new microsatellites originating from random sequences (Li et al. 2002).

The alternative model is that microsatellites sequences are adopted from other genomic regions via a number of transposable elements found in abundance in eukaryotic genomes. In mammals association between microsatellites and retrotransposons were observed, in particular it was found that microsatellites rich in A-base were generated by the extension of terminal 3' of retrotranscripts, similarly to the mRNA polyadenylation mechanism. Anyway, in plants retrotransposition does not seem to be the generalized mechanism of microsatellites genesis, in fact the high density of retrotransposons does not always fit with the density of microsatellites (Morgante et al. 2002).

2.6. Microsatellite transferability

The development of microsatellite markers via traditional methods is expensive and time-consuming, and the PCR primer needed for the amplification of these loci are often species-specific. However many study showed that primer pairs designed for one species can be used for other species from the same genus or even from different genus from the same family (Ellis and Burke 2007). This characteristic is called transferability or cross-species amplification. Transferability can be a very important factor in facilitating the use of microsatellites since it reduces costs when working with species that lack a clear economic importance. Microsatellite transferability amongst related species is allowed by the

homologous nature of the DNA sequences in microsatellite flanking regions. Anyway, the successful amplification rate declines as genetic divergence between species increases. Moreover the degree of microsatellite polymorphism is not transferable, thus high levels of polymorphism detected in one species may not be found in another one after primers have been transferred (Oliveira et al. 2006; Barbará et al. 2007).

As described above, mutation rates for length variation in microsatellites have been found to be higher than point mutations rates. In order to explain this difference, two kinds of mutational mechanism have been proposed: replication slippage, and unequal crossing-over. Both processes result in changes in the number of repeat units which is compatible with the observed size polymorphism of microsatellites. One consequence of these mutational mechanisms is that the same genetic state (i.e. number of repeats) may evolve in two different microsatellite lineages through independent mutational events, a phenomenon known as homoplasy. Homoplasic microsatellite alleles are alleles similar in length, but different in descent, thus homoplastic loci are not orthologous. Microsatellite homoplasy can be divided into two types:

- microsatellite alleles identical in length, but not in sequence (indistinguishable by fragment length analyses)
- alleles identical in both length and sequences, but with different evolutionary history (only detectable through mutations documented in known pedigrees)

The likelihood of orthology vs. paralogy of cross-amplified loci will have to be evaluated on a case by case basis (Barbará et al. 2007).

Moreover the generation of non-amplifying alleles (so-called null alleles) could occur when mutation in one or both primer binding site prevents PCR amplification. Homozygous individuals for a null allele do not display any band at all, whereas heterozygotes show only one band therefore appear to be homozygotes on a gel. Undetected null alleles can give the erroneous impression of an apparent homozygote excess in population studies.

2.7. Microsatellite markers in taxonomic studies

There has been much debate about the use of microsatellite markers for phylogenetic studies. Despite multilocus dominant markers (AFLP, ISSR, etc) are broadly used to characterize taxonomic relationship between specimen, microsatellites DNA analyses

provides a higher chance that specific alleles rather than unrelated bands are compared with each other. However there are several evidence that could affect the usefulness of microsatellites such as the unpredictable mutation rate, the possible constrained range of alleles size and the frequent occurrence of indel events in the flanking sequences. Anyway, one important determinant of the extend of markers transferability across species is the source and characteristic of the library of origin. Thus primers binding site are expected to be more conserved when the microsatellite flanking sequences are maintained under selective constraints. Consequently microsatellites within the transcribed regions provide good chances to design primer pairs that are broadly transferable among taxa and even within plant families (Ellis and Burke 2007). EST-derived microsatellite markers usually are less polymorphic than genomic microsatellites, but often show an increased level of conservation between taxa.

2.8. Potential applications

Compared with other class of molecular markers microsatellites are highly polymorphic, because of which they have been used not only to answer several questions related to plant population genetics, such as gene flow and paternity analysis but also for the study of natural population. The knowledge of the distribution of genetic variability between and within natural plant populations is essential to adopt competent strategies for ex situ and in situ germplasm conservation and microsatellites are extremely useful for estimating genetic population parameters as population structure, parentage and paternity analysis, and gene flow.

Chapter 3

Population genetics in Plants

3.1. Population genetics: goals and applications

Population genetics study frequencies of occurrence of alleles and genotypes within and between populations. It is also the study of changes in gene frequencies and, therefore, is closely related to evolutionary genetics, as evolution depends heavily on changes in gene frequencies. The nature, timing and geographic context of historical events and population processes shape the spatial distribution of genetic diversity. This issue is, therefore, critical for addressing questions relating to speciation, selection, and applied conservation management (Garrick et al. 2010). Genetic variation is usually distributed in a hierarchical way: within and between individuals, within or among populations, within regions of origin or all regions inhabited by a species (Weising et al. 2005). Following the analyses of variation and how it is partitioned over these hierarchical levels, important conclusions about the biology of a species can be drawn. The amount of genetic variation in a species and its distribution among populations is determined by a large number of factors including the mating system, environmental barriers, the demographic history, the effective population size, and the extend of gene flow regarding migration or seed dispersal between populations. Sometimes in plant it is possible to observe gene flow between different species, hybridization between crop and wild relatives are well documented, and introgression between sympatric species are often reported for forest trees (Gugerli et al. 2008). Moreover, as species geographical distribution is usually more extended than an individual's dispersal capacity, populations are often genetically differentiated through isolation by distance (Balloux and Lugon-Moulin 2002). Hence, populations in close proximity are genetically more similar than more distant populations. Long lived, outcrossing and late successional taxa retain most of their variation within population, whereas annual, selfing, and early successional taxa allocate more variation among populations (Belaj et al. 2007). Within-population diversity, in general, is negatively correlated with the level of population differentiation (Weising et al. 2005).

Genetic structuring reflects the number of alleles exchanged between populations, so this has essential consequences on the genetic composition of individuals themselves. Understanding gene flow and its effects is crucial for many fields of research including population genetics, populations ecology, conservation biology and epidemiology (Balloux and Lugon-Moulin 2002). The exchange of genes between populations homogenizes allele frequencies and determines the relative effects of selection and genetic drift. High gene flow precludes local adaptation through the fixation of alleles which are favoured under local conditions, and will also impede the process of speciation (Siol et al. 2010). On the other hand, gene flow generates new polymorphisms in populations, and increases the ability to resist random changes in allele frequencies. Thus it has the opposite effect to random genetic drift, since it generates new gene combinations on which selection could act (Lugon-Moulin et al. 1999). Reliable estimates of population differentiation are also crucial in conservation biology, where it is necessary to understand whether populations are genetically isolated form each other, and if so, to what extend. Small isolated populations are subjected to genetic drift, which will affect their evolutionary potential, through fixation of deleterious mutations. For these reasons, the knowledge of population structuring may therefore provide valuable guidelines for conservation strategies and management.

3.2. The Hardy-Weinberg principle

In 1908 G. Hardy and W. Weinberg independently developed a mathematical model that predicts genotype frequencies when a population is not being affected by evolutionary forces (Barcaccia and Falcinelli 2005). The model is known as the Hardy-Weinberg Equilibrium Model. The Hardy-Weinberg Equilibrium is a baseline by means of the evolution of populations can be measured, and is the foundation for the genetic theory of evolution. The Hardy-Weinberg Equilibrium theory states “in a large randomly breeding population, allelic frequencies will remain the same from generation to generation assuming no evolutionary forces are acting within the population”.

Populations will conform to the Hardy-Weinberg Equilibrium assertions only if no evolutionary forces or mechanisms influence the loci under consideration; the assumption that population must satisfy to hold the Hardy-Weinberg Equilibrium include:

Large population size. Random chance can alter allele frequencies through mating processes and death within small population.

The random sampling among gametes in small population could result in fluctuation in allele frequencies that occur by chance; a phenomenon also known as genetic drift.

Random mating. The choice of mates by individuals in the population is determined by chance, and not influenced by the genotype.
When non-random mating occurs individuals that are more or less closely related mate more often that would be expected by chance for the population. It increases the homozygosity of a population and its effect is generalized for all alleles.

No difference in *mutation rates* between alleles of the same locus. Mutation lead the occurrence of novel alleles, which may be favourable or deleterious to the individual's ability to survive. If changes are advantageous then the new alleles will tend to prevail by being selected in the population.

No migration or gene flow. Reproductive isolation from other population or migration implies changes in allele frequencies through the movement of individuals and new alleles into a population. It causes an increased genetic variability with more copies of an allele that was already present or new allele arrives. Gene flow is the passage and establishment of genes typical of one population in the gene pool of another one by natural or artificial hybridization and backcrossing.

No Selection. No differential survival or reproduction among phenotypes. The effect of selection could be directional (decrease diversity), balancing (it increase diversity as heterozygotes have the highest fitness thus selection favours the maintenance of multiple alleles), and frequency dependent (it increase diversity, and fitness is a function of allele or genotype frequency and changes over time).

When a population meets all of these conditions it is said to be in Hardy-Weinberg Equilibrium, thus:

Allele and genotype frequencies should remain the same from one generation to the next (Russell 2002). For instance when at a certain locus there are only two alleles, the sum of the frequency of the first allele (A) plus the frequency of the other (a) equals one (3.1).

$$p+q=1 \qquad (3.1)$$

Where p is the frequency of the allele A and q is the frequency of the allele a.

Given a certain set of allele frequencies, genotype frequencies should conform to those calculated using basic probability. Within a population, the frequencies of all possible combinations of pair of alleles at one locus can be mathematically expressed based on a simple binomial or multinomial distribution. The case of a biallelic locus (A, a) is reported in equation 3.2.

$$p^2+2pq+q^2=1 \quad (3.2)$$

Where p^2 is the frequency of the homozygotes AA, q^2 is the frequency of the homozygotes aa, and 2pq is the frequency of heterozygotes Aa (Barcaccia and Falcinelli 2005).

If the genotype frequencies obtained from a real population do not agree with those predicted by Hardy-Weinberg Equilibrium, then some evolutionary mechanisms must operate.

Populations in their natural environment usually do not meet the conditions required to archive Hardy-Weinberg Equilibrium, thus their alleles frequencies will change from one generation to the next and the population will evolve (Russell 2002). Anyway, the comparison of observed versus expected outcomes offers an estimation of how far the population deviates from Hardy-Weinberg Equilibrium, and it is an indication of the effect of external factors. In population genetics it is recommended that the first test to be done is to establish if a population is in Hardy-Weinberg Equilibrium or not, and whether there is linkage of markers (Linkage Disequilibrium). Deviations from Hardy-Weinberg Equilibrium could occur from some mistakes during the sampling procedure, and from self-fertilization or selection. In particular, selection could act at a locus itself or at gene linked to it. On the other hand, another possibility is the occurrence of null alleles that alters the detection of heterozygous individuals (see Chapter 2) and cause deviation from Hardy-Weinberg Equilibrium. Thus, a locus that exhibit strong deviation from Hardy-Weinberg Equilibrium should be omitted from further analyses. Linkage Disequilibrium is a deviation from the random association of alleles in a population that may be caused by population substructuring or high levels of inbreeding. Consequently alleles are not transmitted independently but instead as a haplotype, thus, the occurrence of linkage disequilibrium is problematic for population genetics studies.

3.3. The theoretical models

Another important factor that should be taken into account is the choice of the most appropriate mutational model, as many parameters, such as number of migrants, population structure, and effective population size are highly dependent on this assumption. The mutation models are used to derive the expected number of alleles in a population from the observed heterozygosity, and even in the statistical analyses of genetic variation. In general four models can be used (Balloux and Lugon-Moulin 2002).

Infinite alleles model (IAM). Each mutation randomly creates a new allele at a given rate. This model does not allow for homoplasy; consequently, identical alleles share the same ancestry and are identical-by-descent. Applying this model to microsatellite loci, mutations alter the number of repeats and the proximity in terms of number of repeats does not indicate a greater phylogenetic relationship.

Stepwise mutation model (SMM) each mutation creates a novel allele either by adding or deleting a repeat of the microsatellite, with an equal probability in both directions. It implies that two alleles differing by only one motif are more related than alleles differing by several repeats. This model is preferred to assess for relations between individual and population structure.

Two phase Model is an extension of the SMM, developed to account for a proportion of larger mutation events. It states that most mutational events result in an increase or decrease of one repeat unit with probability p, and large variation in number of repeats occurs with probability (1-p).

K- alleles model assumes that if there are exactly k possible alleles in a given locus then the probability of a given allele mutating into any other is $\mu/k-1$, where μ is the mutation rate. This model allows for homoplasy, it means that alleles identical-by-state could not be identical-by-descent.

However, no single model of evolution could play a role in the mutation-driven variability of microsatellites (Bhargava and Fuentes 2010), thus all models present disadvantages when applied to microsatellite markers due to their high mutation rate and to the constrained allele size (Weising et al. 2005).

3.4. Measures of genetic variation

The characterization of plant genetic resources at DNA level will determine the identity of each individual in a population or in a germplasm collection by a band pattern or fingerprint. The simplest quantitative measure is the number of polymorphic markers (P) often expressed as the percentage of all markers scored in a set of samples. Another simple measure is the Average number of allele per locus A, which is usually reported over all loci tested (3.3).

$$A= (1/K)\Sigma n_i \quad (3.3)$$

Where n_i is the number of alleles per locus and K is the number of loci. This measure is, however, sensitive to sample size as a larger sample size increase the chance to observe rare alleles. In fact, for instance, a population where are present three alleles with similar frequency is more polymorphic than one that harbour one very frequent allele and two rare alleles. Thus, a better measure of variation is the effective number of alleles (A_e), that corrects the absolute number of alleles by taking in account the allele frequency (3.4).

$$A_e= 1/\Sigma p_i^2 \quad (3.4)$$

Where p_i is the frequency of the i^{th} allele in a locus, even if this parameter is often averaged over all loci studied. When allelic frequencies are similar, value of effective number of alleles is close to the observed number of alleles in a locus. Therefore, large differences between these parameters indicate low frequencies of some allele, thus the effective number of alleles could provide evidences indicating rare alleles (Laurentin 2009).

Another measure of genetic variation is the observed frequency of heterozygous (H_o) samples averaged over loci that can provide evidence of inbreeding or could be due to the occurrence of null alleles. One of the most commonly employed methods to estimate within-population diversity is the expected heterozygosity, which is the frequency of heterozygous individuals expected under Hardy-Weinberg equilibrium. There are two main formulas to calculate the expected heterozygosity. In 1973 Nei proposed an index of genic variation called heterozygosity H, that is also named Nei's gene diversity (Nei 1973). H is defined as the probabilities of non-identity of two random chosen genes (correlation of uniting gametes), and is calculated from the allele frequencies following the equation 3.5.

$$H= 1-\Sigma_k p_i^2 \quad (3.5)$$

The gene diversity H is summed from i=1 to k, where p_i is the frequency of the i^{th} allele and k is the number of alleles. Gene diversity is averaged across all loci. Later, Nei (1978) proposed a second formula called unbiased heterozygosity H_e that reduces the bias caused by a small sample size (3.6-3.7)

$$h_k = 2n(1-\Sigma p_i^2)/(2n-1) \quad (3.6)$$

Where h_k is the heterozygosity or gene diversity for the locus k, p_i is the frequency of the i^{th} allele, and n is the number of individuals.

$$H_e = \Sigma h_k/r \quad (3.7)$$

While H_e is the expected heterogyzosity over all loci where r is the number of loci.

3.5. F-statistics in genetic differentiation

Separate populations of a species differ in the relative allele frequencies of genes and markers. Random drift, selection, founder effects, and bottlenecks can cause populations to differentiate, whereas high migration rate between populations will prevent or slow down differentiation. The partitioning of genetic variation, within or between populations, can provide evidences about the factors involved in populations evolution. Differentiation of populations causes a reduction in the proportion of observed heterozygotes compared with the number of expected heterozygotes. The extend of this reduction can be used to obtain a measure of population structure employing Wright's F statistics. Originally a fixation index was developed by Wright (1921) to account for the effect of inbreeding within samples. He defined this quantity in terms of a correlation coefficient. Later Wight (1951) expanded this concept to a population subdivided into a set of subpopulations, leading to the traditional hierarchical F-statistics F_{IS}, F_{ST}, and F_{IT} introducing the following formula (3.8):

$$1-F_{IT} = (1-F_{IS})(1-F_{ST}) \quad (3.8)$$

Wright defined F_{IT} and F_{IS} as the correlations between the two uniting gametes to produce the individuals relative to the total population and relative to the subpopulation, respectively, whereas F_{ST} is the correlation between two alleles chosen at random within subpopulation relative to alleles sampled at random from the total population (Wright 1951). Therefore F_{ST} measures inbreeding due to the correlation among alleles because they are

found in the same subpopulation. F_{IS} and F_{IT} may become negative, while F_{ST} is always non-negative and ranges from zero to one. Theoretically, when considering two subpopulations and a biallelic locus, this quantity will reach a value of one when the two subpopulations are totally homozygous and fixed for the alternative allele, and a value of zero when the frequencies in the two populations are identical. F_{ST} represents a measure of heterozygote deficiency due to population subdivision, hence it measures the heterozygote deficit relative to its expectation under Hardy-Weinberg equilibrium.

Later in 1973, Nei redefined the fixation indices for multiple allele. He showed that the gene diversity of the total population can be partitioned into its components (intra-subpopulation and inter-subpopulation), when the gene diversity is defined as the heterozygosity expected under Hardy-Weinberg equilibrium.

Nei also defined an analogue of F_{ST} among infinite number of subpopulations, called coefficient of gene differentiation (G_{ST}) as being the ratio 3.9:

$$G_{ST}= D_{ST}/Ht= (Ht-Hs)/Ht \quad (3.9)$$

Where D_{ST} is the average gene diversity between sub population, including the comparisons of subpopulation themselves with (3.10):

$$D_{ST}= (Ht-Hs) \quad (3.10)$$

Where H_S and H_T are defined in terms of gene diversity, but, in random mating subpopulations, they correspond to the expected heterogyzosity under Hardy-Weinberg equilibrium averaged among subpopulations (Hs) and of the total population (Ht).
In 1977, Nei extended this concept and defined the F-statistics as a function of observed and expected heterozygosities and all of them take similar mathematical form (3.11).

$$F_{ST}= H_T-H_S/H_T \quad (3.11)$$

In which H_T is the expected heterozygosity in the total population and H_S is the mean expected heterozygosity in the subpopulations.

Since F_{IS}, F_{ST}, and F_{IT} are defined in terms of the present gene and genotype frequencies they can be applied to any situation, whether there is selection or not and maintain the same mathematical properties. As in the case of Wright's definition F_{IS} and F_{IT} measure the deviations of genotype frequencies from Hardy-Weinberg proportions in the subpopulation and in the total population, respectively, whereas F_{ST} measures the degree of genetic differentiation of subpopulation.

The high mutation rate that characterize microsatellite markers can invalidate many assumption used in conventional population structure analysis, as different populations could share homoplastic alleles at frequencies that depend on both the rate and the mutation process in detail (Oliveira et al. 2006). Since microsatellites are often assumed to follow a stepwise mutation model (SMM); it was derived a statistic parameter explicitly based on this mutation model R_{ST} (3.12).

$$R_{ST} = (S - S_W)/S \quad (3.12)$$

Where S is the average squared difference in allele size between all pairs of alleles, and S_W is the average sum of squares of the differences in allele size within each subpopulation. R_{ST} is an F_{ST} analogue that assumes a mutational model thought to reflect more accurately the mutation pattern of microsatellites. R_{ST} is calculated from the variances of allele size within each subpopulation, whereas F_{ST} is derived from the variances of allele frequencies. The high mutation rate of microsatellites causes a decreased probability of identity that deflated F_{ST} values. In fact, F_{ST} is a decreasing function of the product of local population size and the sum of migration and mutation. Thus the magnitude of the ratio of mutation over migration affect deeply F_{ST} estimator for its sensitivity to the mutation rate when migration is low (Balloux and Lugon-Moulin 2002). On the other hand, R_{ST} is independent of the mutation rate, in fact, assuming this model, allele size differences are relevant for the calculation of distances between loci and individuals, but this estimator is affected by high variance. Under a SMM R_{ST} could be more accurate than F_{ST} reducing the sampling variance by increasing the number of populations sampled, the number of individuals per population or the number of loci scored. R_{ST} is a better predictor of interspecific divergence as it better detect longer historical separations, whereas F_{ST} appear to be more sensitive to detect intraspecific differentiation.

As described in Chapter 2, the mutation rates of microsatellites loci show variations among different repeat types, base composition and microsatellites type. Similarity the mutation patterns appears to involve both the addition/deletion of single repeats and nonstepwise mutation, thus, none of the models appear to perfectly fit all microsatellites loci. Consequently, both F_{ST} and R_{ST} are commonly reported in studies using microsatellites markers.

Since G_{ST} (3.9), F_{ST} (3.11) and their relatives fail to detect differentiation when within-subpopulation heterozygosity is very high, Jost (2008) pointed out the opportunity to derive new self-consistent descriptive measures of diversity and differentiation. Thus, this author proposed the effective number of alleles (described above in 3.4) to describe the diversity, and

a new parameter, called true differentiation (D), to estimate differentiation. This measure can be directly related to the migration and mutation rate of the finite-island model, whereas it is independent of the within-population diversity H_S and of the sample size. For these features D should be more appropriate for estimation of subpopulation structuring, for correctly ranking populations in terms of their differentiation. It is worth to note that D was strongly criticized, as it is affected by mutation and heterozygosity such as G_{ST}, and it is not useful in estimation of migration rate, on the other hand this is a robust tool in situation were interest is focused on a single locus (Ryman and Leimar 2009). On the other hand, R_{ST} is not affected by mutation and it is a good estimator under a perfect stepwise mutation (SMM) scenario, but the applicability of this observation may be limited as no groups of molecular maker seems to mutate exclusively according to this model (Balloux and Lugon-Moulin 2002). Anyway, as pinpointed by Jost (2009), the choice of the most fitting estimation measure depends exclusively on the questions that being asked, in fact, G_{ST} and D measure different aspects of population structure. D measures the actual relative degree of differentiation of allele frequencies among subpopulations, whereas G_{ST} is a useful tool for estimating the amount of migration between subpopulations (Jost 2009). Finally, it is clear that the perfect estimator does not exist, but there are several indices of genetic differentiation that could describe different variables and have to be chosen according to the question being asked.

3.6. Computer programs for population genetics

The analysis of genetic diversity within species is vital for understanding evolutionary processes at the population level and at the genomic level. Recent population genetics methods can provide accurate information on the past demography of a population, which is a parameter necessary to provide a correct interpretation of linkage disequilibrium, recognize regions of the genome that are under selection or help to develop good conservation strategies and priorities. The advent of cheap genotyping techniques has broadly facilitated the assessment of genetic diversity within populations, thus, powerful novel methods have been developed to analyse these data, sometimes relying on massive computations. These methods are implemented in different software packages and programs, which has grown in a huge number in the few past years (Excoffier and Heckel 2006). Usually the main programs are freely available on line and can be grouped in three categories:

Multi-purpose packages that compute basic statistics that describe the genetic diversity within and between populations as well as a few more elaborated analyses;

Individual-centred programs which focus on the analysis of individuals and the very recent history of a population employing population assignment tests, relatedness, and parentage analyses;

Specialized programs that intended to infer some population parameters under a particular evolutionary scenario.

3.7. Bayesian inference in phylogeny estimation

In application of population genetics, it is often useful to infer populations structure or classify individuals in a sample into populations. The standard approach include three basic methods that have been used to estimate phylogeny: genetic distance, maximum parsimony, and maximum likelihood. In 1996, three groups independently proposed to use Bayesian inference of phylogeny, which utilizes a simulation technique in combination with the chosen model and the data, to produce a posterior probability distribution of trees. The method chosen for approximating the posterior probability is a numerical method called Markov Chain Monte Carlo (MCMC) (Huelsenbeck et al. 2002). This method is required for taking valid, albeit dependent, samples from the probability distribution of interest, in the case of Bayesian inference of phylogeny it is the posterior probabilities of phylogenetic trees (Huelsenbeck et al. 2001). Essentially in Bayesian approach there is no logical distinction between model parameters and data, that both are random variables with a joint probability distribution that is specified by a probabilistic model. The joint distribution is a product of the likelihood and the prior, which incorporates information about the values of a parameter before examining the data in form of a probability distribution. The likelihood is a conditional distribution that specifies the probability of the observed data given any particular values for the parameters and it is based on a model of the underlying process, combining all available information about the parameters. Thus, Bayesian statistics implies the manipulation this joint distribution in various ways to make inferences about the parameters, or the probability model, given the data. The main aim of Bayesian inference is to calculate the posterior distribution of the parameters, which is the conditional distribution of parameters given the data (Beaumont and Rannala 2004). Similar to the maximum likelihood method, Bayesian

inference is based on the likelihood function and it focuses on a quantity known as the posterior probability which is the probability that a tree is correct based on the prior beliefs and the likelihood (Alfaro and Holder 2006). Thus, unlike maximum likelihood, Bayesian inference of phylogeny can incorporate background information into the specification of the model. While the prior probability distribution describes the probability of different trees given the prior knowledge, the posterior probability of trees describes the probability of trees considering the prior distribution, the model and, most importantly, the data. Thus, for example, the values observed on a Bayesian phylogeny are not simply the posterior probability that a clade is true, but rather then the probability that it is true given the model and parameters used, the priors, and the data (Archibald et al. 2003).

This likelihood approach was not applicable to population genetic until the development of coalescent theory. This theory describes the statistical distribution of branch lengths in genealogy of chromosomes or genes, under many life history schemes, and taking predefined limits. The coalescent theory allows to calculate the expected values of statistics, and also provide a parametric bootstrapping for generating a simulated sample data set, which permit a more sophisticated calculation of confidence intervals and hypothesis testing with respect to the traditional frequentist methods. Despite it is not possible the application of coalescent theory in all areas of population genetics, it is the basin for likelihood calculation that allowed the use of Bayesian approaches to infer demographic histories from genetic data and assign individuals to populations. In this last case the assignment method calculate the probability of an individual's multilocus genotype given the allele frequencies at different populations.

Chapter 4

Materials and Methods

4.1. Plant material

In 2008, a *Tamarix* germplasm collection was conducted and three species were found: *T. africana*, *T. gallica* and *T. canariensis*. Plants were collected by Renée Abou Jaoudé in six different populations from Central and Southern Italy: Imera, Simeto and Alcantara Rivers from Sicily, Crati River from Calabria, Basento River from Basilicata, and Baratz lake from Sardinia. Later, in 2009, a further population from central Italy were surveyed and included in this study: Marangone Creek from Lazio. It was performed a blind sampling, thus individuals were collected without any regard for species identity. The number of individuals surveyed for site ranges from 24 to 84, with a total of 316 plants. Later in laboratory, the identities of 85 plants were determined by Grazia Abbruzzese with Baum's morphological keys (Baum 1978) while all the rest (72% of the total) remained unidentified as shown in Table 4.1.

Sites	Individuals	*T. gallica*	*T. africana*	*T. canariensis*	Unknown
Basento	84	13	12	2	57
Crati	36	3	8	1	24
Simeto	54	8	8	7	31
Imera	36	0	13	0	23
Alcantara	24	0	9	0	15
Baratz	32	0	4	0	28
Marangone	50	-	-	-	50

Table 4.1: Number of plants collected per population and number of individuals identified by morphological traits according to Baum's dichotomical key divided per population and species (species identification by Grazia Abbruzzese).

Leaf tissue was collected from individuals not in the field, but *ex situ* from the vegetative propagated material. Leaves were frozen at -20°C and conserved until the successive analysis. Moreover, although only few genotypes for each species were available, even T. *jordanis, T. tetragyna, T. aphylla* from the desert of Negev (Israel) were used to test transferability and polymorphism of microsatellites markers. DNA samples were provided by our project partner Aviah Zilberstain.

4.2. DNA extraction

Leaf tissue from the whole set of samples was collected, and total genomic DNA was extracted following the Doyle and Doyle protocol (1990).

The Doyle and Doyle procedure use a CTAB isolation buffer composed as following:

- 2% hexadecyltrimethylammonium bromide (CTAB)
- 1,4 M NaCl
- 2% β-mercaptoethanol
- 20 mM EDTA
- 100 mM Tris-HCl (pH 8)

The fresh leaf tissue (100-300 mg) was grind in a mortar cooled by liquid nitrogen to obtain a fine powder, the powdered plat material was transferred in a sterile 2 ml tube. The isolation buffer was heated at , and 1 ml was added to the tube allowing the powder to mix with the buffer. This suspension was incubated at for and occasionally swirled. Afterwards, 1 ml of chloroform-isoamyl alcohol (1:24) was added. This step produce two phases, an upper aqueous one that contain the DNA, and a lower chloroform phase which enclose the degraded protein. The interface between these two phases contain cell debris and degraded protein. The tube was spun at 9000 rpm at room temperature for 10 minutes to allow the separation of the aqueous phase, which after centrifugation was transferred to a new tube using a micropipette. This step was repeated once upon a time. To permit the removal of the aqueous phase of cold isopropanol was added to denature the nucleic acid. Usually this stage yields large strands of DNA that can easily seen by eyes. The tube was spun at 6000 rpm at for 10 minutes to concentrate the DNA pellet to the bottom of the tube and to share it from the

supernatant. A solution of 76% ethanol and sodium acetate was added to wash the pellet, after 20 minutes of washing the tube was spun and the supernatant was discarded. A second washing buffer composed by 76% ethanol and ammonium acetate was added to the pellet that, usually, became white or transparent. The nucleic acids were spun down to pour off the supernatant, and the pellet was dried at room temperature in a sterile hood. In the last stage the DNA pellet was resuspended in distilled water. Sometimes, the DNA needs further purification protocol, thus in our case we used a supplementary protocol for purification from polysaccharides. For this purpose 0,5 volumes of 7,5 M ammonium acetate were added to the distilled water and this solution was incubate for 30 minutes at . After this period the tube was centrifuged at for 10 minutes at to precipitate the polysaccharides at the bottom of the tube. The surnatant containing the DNA was transferred in a new tube and added of 2 volumes of 95% ethanol, and it was incubated for 30 minutes at . the presence of ethanol allow the denaturing of DNA that once spun at 4°C for 20 minutes at 13000 rpm, precipitated in a pellet at the bottom of the tube. The pellet was dried at room temperature in a sterile hood and resuspended in distilled water.

4.3. Quantification genomic DNA

After isolation of DNA, quantification and analysis of quality are necessary to ascertain the approximate quantity of DNA obtained and the suitability of DNA sample for further analysis. The quantification of genomic DNA was performed by detection of absorbance at 260nm in a spectrophotometer instrument. Analysis of UV absorption by the nucleotides provides a simple and accurate estimation of the concentration of nucleic acids in a sample. If the DNA sample is pure without contamination of proteins or organic solvents, purines and pyrmidines show a peak of absorption around 260nm. The genomic DNA extracted was diluted 1:100, this solution was of added to a clean cuvette, distilled water was used as blanck. The ratio of OD_{260}/OD_{280} should be determined to assess the purity of the sample. This method is however limited by the quantity of DNA and the purity of the preparation. Accurate analysis of the DNA preparation may be impeded by the presence of impurities in the sample or if the amount of DNA is too little. In the estimation of total genomic DNA, for example, the presence of RNA could interfere with the estimation of total genomic DNA. A OD_{260}/OD_{280} ratio between 1.8-2.0 denotes that the absorption in the UV range is due to nucleic acids. Otherwise, a ratio lower than 1.8 indicates the presence of proteins and/or other

UV absorbers, whereas a ratio higher than 2.0 indicates that the samples may be contaminated with chloroform or phenol.

4.4. Microsatellite markers detection and scoring

A huge number of works employ fluorescence-labelled PCR primers with a capillary automated sequencer to visualize PCR-generated fragment. This technique, combined with specific fragment analysis software allows an accurate fragment length determination and provides the discrimination of microsatellites, also in those alleles which differ by a single base pair.

Fluorescent PCR products are detected by real-time laser scanning during electrophoretic migration. The allelic information is immediately stored in a computer as the PCR fragment pass the detection window. Differential labelling with different fluorochromes allows the combination of several primer pairs (up to five) in a single run, but the PCR products should have non-overlapping size ranges.

The evaluation and comparison of marker data from different samples requires that individual bands within lane are assigned to particular positions that correspond to the fragment size expressed in base pair. The assessment of fragment size is often done by molecular weight marker-assisted sizing that provides standard landmarks to recognise the length of the fragment. Different lanes are screened for co-migrating bands contemporarily allowing the analyses of many samples and many markers.

4.5. In silico data mining

A total of 22713 ESTs, derived from the species *T. hispida*, *T. androssowii*, *T. ramosissima*, and *T. albiflonum,* were obtained by searching NCBI database and assembled separately to eliminate redundancy with CAP3 software with the criteria of 93% identity and 40bp overlap (Huang and Madan 1999). This program assembles short reads into long sequences, computing overlaps between reads. In a first phase CAP3 remove 5' and 3' poor regions, then it joins reads to form contigs in decreasing order of overlapping score. In the last phase, the software constructs a multiple sequence alignment of reads and computes a consensus sequence along with a quality value for each base and for each contig. The four unigene sets (one for each species) were obtained and screened for the presence of microsatellites using

Magellan software (Lim et al. 2004). This program looks for repeated sequence starting from the first base in the sequence data. If no motifs are present, the program search for repeats in the first two bases. The algorithm will continue it either found a microsatellite or found that there is not a microsatellite array. Magellan identifies all possible mono- to hexanucleotide repeats and permits the operator to choose the search settings. The criteria of minimum ten repeat units for mono-, nine for di-, and five repeats for tri- to esa- nucleotide were adopted. The software avoids the problem of complementarity grouping together all the possible combinations of a repeated motif, however it is not able to detect the redundancy in different non-annotated genome sequence reads.

4.6. Criteria for EST-SSRs markers development

The identification of repeated sequence is not respective of their polymorphism, thus the analyses were not performed on the whole set, but a sub-group of microsatellites was analysed in greater detail. Mononucleotide microsatellites were discarded as alleles scoring is easier in the other classes of microsatellites. Moreover, the detection of short fragments in agarose gel is easy, so were chosen microsatellites with amplicons size less than 300 bp. It is worth to note that it is not possible to design primers if the microsatellite is near both the ends on the sequence. It is well established in literature that the likelihood of polymorphism increase with the number of repeated units.

The resulting potential markers were named on the basis of their species of origin and the progressive number of the unigene sequence. Thus, EST-SSRs deriving from *T. hispida* were assigned the prefix Th, while for those deriving from *T. androssowii* the prefix was Ta.

4.7. Designing of primers

After the detection of the repeated regions, primers flanking the microsatellites were designed using the software Primer3 (Rozen and Skaletsky 2000). *T. hispida* and *T. androssowii* had the largest data sets, thus only these two unigenes were considered for primer designing. The software is freely available on line and the primers were chosen on the basis of the criteria of 18-27 nucleotide in length, optimum 55% GC content, optimum melting temperature of , and absence of secondary structures. The repeated sequence was labelled to

allow its inclusion within sequence amplified by the primer pair, and the size of amplicons was set between 100 bp and 300 bp.

4.8. SSRs and EST-SSRs amplification tests and screening of polymorphism

A subset of 35 SSRs was selected and the cross-species amplification was tested on *T. africana*, *T. gallica*, *T. jordanis*, *T. tetragyna*, and *T. aphylla* one individual for each species. It is worth to note that 10 neutral SSRs (Gaskin et al. 2006) have been developed in *T. chinensis* and *T. ramosissima* and their hybrids, thus even the cross-species amplification of these markers in the above mentioned species were tested. In the case of microsatellite markers in a cross species amplification, the sequences in the primer site often contain mismatch, due to mutations. Thus, the amplification efficiency could be compromised. For this reason, it was performed a test to assess the melting temperature in the species studied in this work by a gradient PCR reaction. The Polymerase chain reaction was performed in 12.5 µl, containing 1X reaction Buffer (GE healthcare), 2 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of forward and reverse primers, 0.25 U of Taq DNA Polymerase (GE healthcare), and 25 ng of template DNA. Samples were amplified following this thermal protocol: 3 min of denaturation at 94°C, followed by 30 cycles at 94°C for 30 s, annealing temperature varying from 53-60 °C for 30 s, extension at 72°C for 30 s; 8 cycles at 94°C for 30 s, 51°C for 30 s, 72°C for 30 s; and one step at 72°C for 10 min. Separation and detection of microsatellite marker was verified by separating 5µl of the amplicons on a 2.5% agarose gel in 0.5X TBE buffer, and staining the gel by ethidium bromide to evaluate the new melting temperature.

In order to evaluate polymorphism of the novel set of EST-SSR markers, it was chosen to analyse a sub-set of 24 individuals of *T. africana* and four of *T. gallica*. Individuals were collected from four populations (Basento, Imera, Crati, Alcantara, six individuals per population) and two populations (Basento and Crati, two individuals per population), respectively. A 19-bp M13 tail (5'-CACGACGTTGTAAAACGAC-3') was added to the 5' end of all the forward primers following Oetting et al. (1995). The Polymerase chain reaction was performed in 12.5 µl, containing 1X reaction Buffer (GE healthcare), 2 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of labeled 6-fam or Hex M13 and reverse primer, and 0.1 µM of the forward M13 tailed primer, 0.25 U of Taq DNA Polymerase (GE healthcare), and 25 ng of template DNA. Samples were amplified following this thermal protocol: 3 min of

denaturation at 94°C, followed by 30 cycles at 94°C for 30 s, annealing temperature specific to each primer pair for 30 s, extension at 72°C for 30 s; 8 cycles at 94°C for 30 s, 51°C for 30 s, 72°C for 30 s; and one step at 72°C for 10 min. Separation and detection of microsatellite marker can be achieved in several ways, in the first step, the amplification of the EST-SSRs was verified by separating 5µl of the amplicons on a 2.5% agarose gel in 0.5X TBE buffer, and staining the gel by ethidium bromide. While, to detect the presence of polymorphism a 2.5% high resolution agarose MethaPhor (Bioproducts controllare) in 0.5X TBE buffer was employed. Once established the presence of polymorphism by scoring the MetaPhor gel, the PCR products were diluted up to 1:20 in water and 1 µl of the diluted PCR product was mixed with 0.25 of a 500 bp internal-lane size standard (Gene Scan^{TM}- ROX 500, Applied Biosystems) and 9.75 µl of pure deionized-formammide. This solution was denatured at 95 °C for 5 minutes, and immediately chilled on ice. PCR amplification fragments were resolved by capillary electrophoresis with an ABI Prism 3700 (Applied Biosystems) to determine the exact size of the amplified microsatellite fragments (CNR IBAF- Porano institute facility). The amplified SSR fragment data were collected using Gene Scan Analysis version 3.7 Software and genotype profiles were assigned with Genotyper version 3.7 NT Software (Applied Biosystems).

4.9. Sequencing of EST-SSRs amplicons

Amplicons of homozygotes individuals for each EST-SSR were purified using Qiaquick PCR purification kit (Qiagen). The purified products were quantified in Qubit fluorometer (Invitrogen) following the manufactures instructions. Successively 30 ng of the purified product was added to two 1,5 ml tubes, in each tube was added alternatively 15 pM of both the forward and the reverse primers. The tube containing the solution with purified amplicons and the primers were shipped to Eurofins MWG Operon (M-Medical) for value read sequencing service.

4.10. EST-SSRs putative homology

Amplicons of polymorphic EST-SSRs of *T. africana* were sequenced to allow for actual identification of the fragments, sequences submission in GenBank, and assignment of putative homology to known genes. In *T. gallica* the successful marker amplification was determined

by comparison to the expected fragment size. The actual identification of the PCR fragments was performed searching the sequence obtained from PCR products in BLASTn (Altschul et al. 1997). The sequence obtained from PCR products was also aligned to the corresponding unigene sequence by ClustalW (Thompson et al. 1994). The sequences were searched against the GenBank non-redundant database using BLASTX (Altschul et al. 1997) for functional annotation with the expected value $<10^{-7}$. BLASTX finds regions of similarity between a nucleotide query sequence translated in all the possible reading frames and a protein sequence database.

4.11. Assessment of EST-SSRs characteristics

Population genetics parameters for the 13 polymorphic loci were estimated using GenAlEx 6.4 software (Peakal and Smouse 2006) including the number of observed alleles per locus (A), the observed heterozygosity (H_o) and expected heterozygosity (H_e). Tests for Hardy-Weinberg equilibrium and linkage disequilibrium were calculated using GenePop 4.0.10 (Raymond and Rousset 1995) using the following Markov chain parameters: 10000 dememorization, 100 batches and 10000 iterations per batch.

4.12. SSRs and EST-SSRs analyses in natural populations

Once established the cross species amplification and the presence of polymorphism, the whole set of 316 *Tamarix* individuals and 17 polymorphic microsatellite markers were analyzed using a multiplex PCR approach. Among these markers five were neutral microsatellite markers (Gaskin et al. 2006), and 12 were EST-SSR markers developed in this work (Table 4.2).

Locus	Repeats	Primer sequence	Ta (°C)
T1B8	(AC)(GC)(AC)	F: CGTTAGCAGGTTGGACATGA R: TTTGAGTGTCAGTCGATGGTG	60
T1C1	(AC)	F: GAGGCAAGCCTCTTGAAATG R: TGTGCTGCCGTCTATTTCTC	57
T1C10	(AC)	F:AACGAGGATCATGAAAAGGA R: GACACATGTCCCTACCATTGAA	60
T1E1	(GT)	F: ATTACGACCTGCAAGCATCC R: AATCGAATGCCTCGTGACTT	60
T1G9	(ATC)	F: CCATAAGTGCCCCATCAAAG R: AAAAGCTTTCCCCAAATACCA	58

Table 4.2: Characteristics of the five neutral SSR markers developed in *T. ramosissima* and *T. chinensis* used in this work (Gaskin et al. 2006). In the table are shown the locus name, the repeated motifs, forward (F) and reverse (R) sequences, and the annealing temperatures used during PCR.

Each reaction was performed with 15ng of DNA template. Polymerase chain reaction was performed by both multiplex and duplex PCR in 12.5 µl reaction volume under the conditions specified in the Type-it microsatellite multiplex kit (QIAGEN) using 1X reaction Buffer, $MgCl_2$, dNTPs, and HotStart Taq DNA Polymerase. In the duplex reaction 0.08 µM of the two forward M13 5'-tail-end primers, 0.1 µM of the two reverse primers, 0.1 µM of labeled 6-fam M13 primer were added to the mix. Samples were amplified with the following thermal protocol: 5 min of denaturation at 95°C, followed by 7 cycles at 94°C for 30 s, annealing temperature 1 min and 30 s, extension at 72°C for 30 s; 24 cycles at 94°C for 30 s, 51°C for1 min and 30 s, 72°C for 30 s; and one step at 60°C for 30 min. While in the multiplex reactions 0.2 µM of labeled Vic and Ned forward primer and 0.2 µM of reverse primers were added. Samples were amplified with the following thermal protocol: 5 min of denaturation at 95°C, followed by 28 cycles at 94°C for 30 s, annealing temperature for 1min and 30 s, extension at 72°C for 30 s; and one step at 60°C for 30 min. The PCR products from duplex and multiplex reaction were shuffled together avoiding the overlap of fragment size between different markers. The diluted solutions up to 1:40 were separated by capillary electrophoresis with a 500 bp size standard (ROX 500, Applied Biosystems) using an ABI Prism 3700 (Applied Biosystems) automatic sequencer as reported above.

4.13. Species assignment

The assignment of unidentified individuals was performed using a Bayesian model-based method implemented in STRUCTURE 2.3.3 (Pritchard et al. 2000; Falush et al. 2003), which determined genetically homogeneous clusters by means of 17 microsatellites markers. STRUCTURE identifies both the number of clusters (K) and individual's probability to belong the inferred clusters on the base the multilocus genotypes of all individuals considered in the study. The K clusters are inferred by minimizing Hardy-Weinberg and linkage disequilibrium within clusters and all sampled individuals are assigned probabilistically to clusters. Length burn-in period and number of Markov chain Monte Carlo were set to 100000 repetitions; and the option "correlated alleles frequencies model" for ancestry was chosen. Analyses were performed on the whole set of individuals without assuming predefined structure. The number of clusters K was tested in the range of 1 to 10 with 10 iterations for each value of K. The "admixture" model was chosen, as considered more powerful to detect subtle populations structure (Pritchard et al. 2000; Evanno et al. 2005). As highlighted by Evanno and co-workers (2005), the posterior probability of the data for a given K value, does not always show a clear mode for the true K. They recommend to use an ad hoc quantity based on the second order rate of change of the likelihood function with respect to K (ΔK). This approach was used to determine the appropriate number of clusters.

Following the admixture model of STRUCTURE, the assignment probability of an individual to each cluster (q) can be interpreted as the membership of its genome in each cluster. Individuals were considered assigned if possessing >90% of ancestry, otherwise they were considered admixed.

A second assignment test was performed with the software GENECLASS 2 (Piry et al. 2004) using the Paetkau frequency method (1995). The program is able to compare a reference sample constituted by individuals of *T. africana* and *T. gallica* identified by morphological traits, and a group of individuals to be assigned constituted by our unidentified samples. This approach computes the likelihood of the individual's multilocus genotypes occurring in each species and assigns the individual to the species with the highest likelihood. The statistical threshold was calculated simulating 1000 genotypes by the Monte Carlo resampling method (Paetkau et al. 2004) to obtain a confidence level for each individual assignment ($P>0.01$).

4.14. Selection of the most informative markers

The best combination of loci for individuals assignment to the species was searched with WHICHLOCI version 1.0 (Banks et al. 2003). The baseline data set comprehends all the individuals identified both by morphological trait and by the Bayesian approach. Individuals are divided in groups corresponding to their species. The program first places in the order of markers according to the markers information contents in individual assignment generating a rank. Then the program invokes loci from this rank increasing the number of loci one at time until the assignment score matches or exceed an accuracy threshold set by the user. Simulations were performed with 1000 new genotypes (population size N= 100) based on the allele frequencies of each species, using the standard 95% assignment accuracy threshold and LOD=4.0 of stringency option. Following the critical population setting, it is possible to focus on accuracy for assignment of a specific set of samples defined by the used and thus assess which loci are required for the identification of a specified species. Separate tests were performed considering both *T. africana* and *T. gallica* as "critical", where the program evaluated which loci are needed to distinguish the critical species from the other species without attempting to differentiate it. Finally the test was also performed to evaluate which loci are required to differentiate all species at the same time. The Locus Score for each marker is determined by applying the following formula:

Locus Score= % correctly assigned genotypes– (% incorrectly assigned genotypes* Score Multiplier)
where:
Score Multiplier= (100- Accuracy)- Inaccuracy

Accuracy and inaccuracy being the criteria of intent for assignment set by the user.

It is worth to note that usually in studies concerning individual identification, it is required the estimation of the statistical power of either the markers and the combinations of markers employed. In recent years, in DNA forensics, the Probability of Identity (PI) has been widely used for this purpose (Taberlet and Luikart 1999). PI provides an estimate of the average probability that two unrelated individuals drawn by chance from the same population have the same multilocus genotype (Waits et al. 2001). In highly sub-structured populations the theoretical PI could underestimate the probability of finding identical genotypes for the presence of relatives. Therefore it was also calculated the probability that two randomly selected fill-sibs would exhibit identical genotypes PI_{sib}. These calculations were performed

with GenAlEx 6.4 (Peakall and Smouse 2006) following the markers assignment power rank pointed out using WHICHLOCI.

4.15. Population genetic analysis

Population genetics parameters were estimated using the software GenAlEx 6.4 (Peakall and Smouse 2006), GenePop 4.0.10 (Raymond and Rousset 1995; Rousset 2008), and Arlequin 3.5 (Excoffier et al. 2005; Excoffier and Lischer 2010) in all the populations of the two species investigated in this study. The statistics of genetic diversity within populations were calculated per locus and per populations. The indices considered were the number of observed and effective alleles(A and A_e), the number of private alleles by population (A_p), the observed and expected heterozygosity (H_o and H_e). The departure from Hardy-Weinberg equilibrium was tested through the values of the fixation index F_{IS}, while the confidence intervals based on the Markov Chain method evaluated the statistical significance of the results adopting the following parameters: 10000 dememorization, 100 batches and 10000 iterations per batch. Moreover, linkage disequilibrium for each pair of loci across all populations was tested. Population differentiation was calculated by pairwise F_{ST} and R_{ST} using the software Arlequin 3.5. R_{ST} is an F_{ST} analogue that accounts for the stepwise mutation model typical of microsatellite markers (see Chapter 3). The comparison of F_{ST} and R_{ST} values can shed light on the relative importance of drift and mutation underpinning genetic differentiation. The null distribution of pairwise F_{ST} and R_{ST} values under the null hypothesis of no differentiation between populations were tested by a permutation test of 10000 replicates and 0.05 significance level. Moreover, SMOGD software (Crawford 2010) was used to measure the actual differentiation D_{est} among population according to Jost (2008) using 500 replicate permutations for bootstrapped values. SMOGD is freely available on line and report not only the diversity measured for each locus but also the intermediate values. A Bayesian analysis was performed to infer population structure and evaluate grouping. The parameters used were the same reported above for the inference of species identity.

It was performed the analysis of molecular variance AMOVA to examine differentiation between populations. The Phi_{PT} approach was used to partitioning the variation into within-population and among-population components adopting 999 permutations. Phi_{PT} is a F_{ST} analogue that is calculated as the proportion of the variance among populations, relative to the total variance; in other words it represents the correlation between individuals within a population, relative to the total (Excoffier et al. 1992).

Correlations between pairwise genotypic differentiation of Phi_{PT} and pairwise geographic distances were analysed though a Mantel test implemented in GenAlEx 6.4 adopting 999 permutations.

Multivariate principal coordinate analyses were conducted for the populations of the species studied in this work according to a codominant genotypic distance matrix. The pairwise individual-by-individual genetic distance matrix were constructed separately for each species. The program defines a set of squared distances for a single locus analysis. For instance, in the case of a four alleles locus (i-th, j-th, k-th and l-th), the set of squared distances is defined as $d^2(ii, ii) = 0$, $d^2(ij, ij) = 0$, $d^2(ii, ij) = 1$, $d^2(ij, ik) = 1$, $d^2(ij, kl) = 2$, $d^2(ii, jk) = 3$, and $d^2(ii, jj) = 4$ (Smouse and Peakall 1999). Genetic distances are summed across loci under the assumption of independence. Covariance genetic distances matrix were plotted within a multidimensional data set to detect genetically homogeneous grouping of populations.

Detection of outlier loci was performed following the method FDIST (Excoffier et al. 2009) implemented in Arlequin 3.5. It was performed a simulation setting 20000 permutation and 100 simulated demes. It is worth to note that selection can affect genetic diversity between populations, since loci under balancing selection show too even allele frequencies across populations, otherwise loci under local directional selection should show large differences between populations. This method obtain a joint distribution of F_{ST} across all loci as a function of heterozygosity between populations and identify outlier loci that with an unusually high or low F_{ST} value as potentially under selection.

4.16. Computer programs used in this work

A brief description of the different software employed in this work is reported in the following Table 4.3.

Name	Functionalities	Special features	Availability
Multipurpose packages			
Arlequin	Computes indices of genetic diversity, F-statistics and genetic distances between populations, performs tests to assess Hardy-Weinberg equilibrium and linkage disequilibrium	Hierarchical analysis of genetic structure based on the AMOVA framework, tests linkage disequilibrium without specifying gametic phase	http://cmpg.unibe.ch/software/arlequin3.
GenAlEx	Computes basic indices of genetic diversity, Mantel test, and probability of identity	Provides new tools for multilocus codominant data sets and individual-by-individual pairwise distances for interpolating missing data	http://www.anu.edu.au/BoZo/GenAlEx/
GenePop	Computes basic indices of genetic diversity and F-statistics, performs tests to assess Hardy-Weinberg equilibrium and linkage disequilibrium	Available web interface for remote computations, estimates the number of migrants exchanged between populations based on rare alleles	http://genepop.curtin.edu.au/
Individual centred programs			
Geneclass	Detects immigrants from multilocus genotypes, assignment of individuals to populations	Assesses whether a given genotype can be excluded from a population	http://www.montpellier.inra.fr/CBGP/softwares
Structure	Detects the underlying genetic structure among a set of individuals genotyped at multiple markers; can detects new immigrants or individuals whose ancestors were immigrants	Computes the proportion of the genome of an individual originating from the different inferred populations, reports genetic distances between inferred populations and the ancestral one	http://pritch.bsd.uchicago.edu/structure.html
Specialized programs			
Cap3	Assembles short reads into longer sequences on the basis of their overlaps	Constructs multiple sequence alignment in order to eliminates redundancy form non annotated sequences	mail to corresponding author
Magellan	Identifies all possible repeated sequences in a database information	Groups together the same repeated motifs to remove complementarity	http://www.medfac.usyd.edu.au/people/academics/profiles/dcarter.php
Whichloci	Determines the relative discriminatory power of and loci combinations for population assignment of individuals	Selects the best combination of loci required for population assignment through empiric analysis of data drawn from natural populations	http://www-bml.ucdavis.edu/whichloci.htm

Chapter 5

Results

5.1. DNA extraction and quantification

The genomic DNA was isolated from leaves tissue applying the Doyle and Doyle protocol (1989). Quantification was performed both using a spectrophotometer instruments, and the amount of DNA extracted varied between 1 µg to 117 µg per sample.

5.2. Analyses of EST sequences

The NCBI database contained 22731 expressed sequences of four Asian *Tamarix* species, which are *T. androssowii*, *T. hispida*, *T. ramosissima*, and *T. albiflonum*.

All the sequences were downloaded separately, and assembled into unigenes sets representing both contigs and singlets as shown in the Table 5.1.

Species	ESTs	Contigs	Singlets	Unigenes	GC%
T. androssowii	4756	1459	916	2375	42.77
T. hispida	17401	2035	5296	7331	45.33
T. albiflonum	208	5	197	202	44.44
T. ramosissima	347	25	227	252	45.11

Table 5.1: Number of expressed sequences, contigs and singlets reported for species. Unigenes represent the non redundant set of sequences.

T. hispida was characterized by the largest set of sequences that carried about 4Mb, whereas the smallest was the data set that belongs to *T. albiflonum*.

The non-redundant sequences were used also to calculate the GC content. The average GC content across all the four analyzed species was around 44%, ranging from 42.77% in *T. androssowii* and 45.33% in *T. hispida*.

5.3. Frequency and distribution of EST-SSRs

The unigenes set was screened for the presence of repeated regions resulting in a total of 637 microsatellites for the four species as reported in Table 5.2.

Species	EST-SSRs	Density	Average distance
T. androssowii	170	7.1%	5.98 kb
T. hispida	453	6.1%	8.63 kb
T. albiflonum	4	1.9%	19.54 kb
T. ramosissima	9	3.5%	10.26 kb

Table 5.2: Number of perfect microsatellites, density for unigenes set and average distance between microsatellites.

The density, reported as percentage of unigenes that contained at least one repeated sequence, ranged from 1.9 of *T. albiflonum* to 7.1 of *T. androssowii*. The average frequency of one SSR for EST-SSRs varied between about 6 kb for *T. androssowii* and 19.54 kb for *T. albiflonum*.

5.4. Distribution of EST-SSRs based on number of repeats and their motif

The different kinds of SSR motifs recorded in the unigenes sets ranged between 5 for *T. albiflonum* (the smallest data set) to 48 for *T. hispida* (the biggest one). Mononucleotide (68.8%-40% range of abundance of mononucleotide repeats in the four unigene sets) and trinucleotide (50%-18.8% range of abundance of trinucleotide repeats in the four unigene sets) repeats being the most common, followed by di-, tetra-, penta-, and esa-nucleotide repeats as shown in the Figure 5.1.

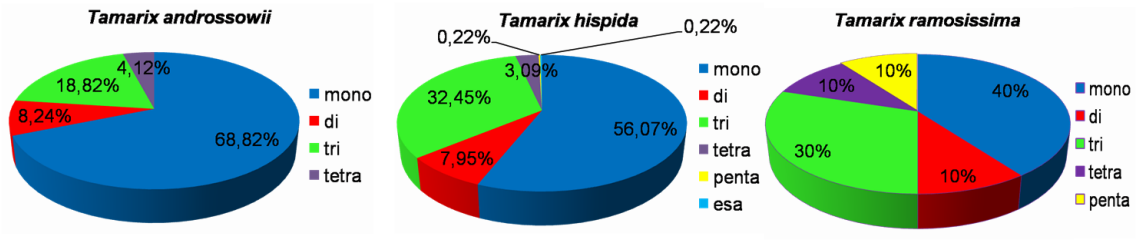

Figure 5.1: Abundance of repeated sequences divided per classes of motif length. *T. albiflonum* is not included in the figure for its small data set, that is composed by two mono- and two trinucleotide microsatellites.

The number of repeats varied from 5 to 33 but the most frequent number of repeats was five. An inverse relation between the frequency of putative EST-SSRs was found with increasing number of repeats as shown in the Table 5.3.

Tamarix hispida

Motif	numer of repeats				SSR
	5-9	10-14	15-19	>20	
A	-	167	26	21	214
C	-	39	1		40
AG	3	9		1	13
GA	2	6	1		9
AT	3	3			6
TA	5				5
CA		2			2
AC		1			1
AAT	14	1			15
CAG	14				14
TAA	8	3			11
AAG	8	2			10
CTC	9				9
GCA	7				7
ACC	6				6
ATG	6				6
CAC	6				6
GAA	6				6
AGC	5				5
AGG	5				5
ATC	5				5
CCA	5				5
CCG	5				5
AAC	4				4
AGA	4				4
ATA	4				4
GCC	4				4
GGA	4				4
CGA	3				3
GAC	2				2
TCA	2				2
ACA	1				1
ACG	1				1
ACT	1				1
CAA	1				1
CGC	1				1
AAGA	3				3
ATAC	2				2
GAAA	2				2
AATC	1				1
AGAA	1				1
AGAT	1				1
ATAA	1				1
CATC		1			1
CCTC	1				1
CTTC	1				1
AAGAA	1				1
CCTGCT	1				1

Tamarix androssowii

Motif	numer of repeats				SSR
	5-9	10-14	15-19	>20	
A	-	78	10	13	101
C	-	15	1		16
AT	2	4			6
TA		1			1
AG		3			3
GA	3	1			4
AAT	7				7
ATA	6				6
CAG	4				4
AGC	3				3
TAA	3				3
AGA	2				2
GAA	2				2
AAC	1				1
AAG	1				1
CTC	1				1
GAC	1				1
TCA	1				1
AAAT	2				2
AGAA	1				1
ATCC	1				1
ATTA	1				1
GCTA	1				1
TACA	1				1

Tamarix ramosissima

Motif	numer of repeats				SSR
	5-9	10-14	15-19	>20	
A		3			3
C		1			1
AG			1		1
AAG	1				1
CAA	1				1
CAC	1				1
ATTA	1				1
GAAAA	1				1

Table 5.3: Relative abundance of microsatellites divided for motifs and number of repeated motifs. For each class of motifs are reported the number of SSRs found by Magellan.

The A mononucleotide motif was the most abundant in all the species analyzed. Trinucleotide motif AAT was the most abundant both in *T. androssowii* and *T. hispida* with frequencies of 22% and 10%, respectively. Otherwise, between dinucleotide repeats the most common motifs were AT for *T. androssowii* and AG for *T. hispida* with frequencies of 43% and 36%, respectively.

5.5. Primer design

A total of 102 primers were designed. Primers could not be designed for the remaining putative EST-SSRs when the presence of the tandem repeats was too close to each end of the sequence, or the nature of the sequence did not allow for primer design using Primer3 selection criteria.

5.6. EST-SSRs amplification tests and detection of polymorphism

Since the microsatellites detected in this work were derived in ESTs of Asian species, the cross-species amplification of the microsatellites primers was tested on one individual for each species of *T. africana*, *T. gallica*, *T. jordanis*, *T. tetragyna*, and *T. aphylla.* Thus it was possible to demonstrate the transferability of EST-SSRs loci among species geographically distant, which, for this reason, are genetically different. Among the subset of 35 primer pairs selected, 25 SSRs (71.4%) showed amplification in *T. africana*, while 27 loci (77.1%) were transferable in *T. gallica* (Figure 5.2).

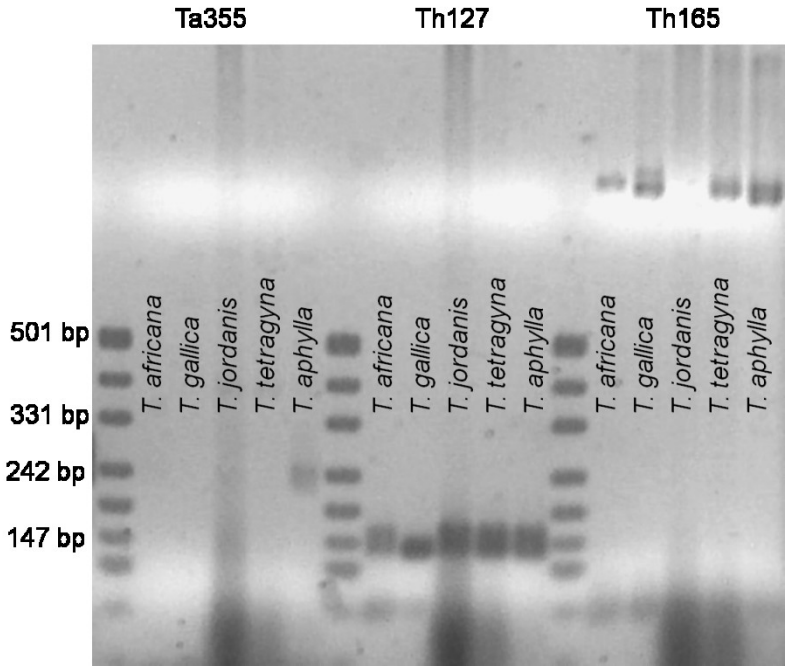

Figure 5.2: In the picture is shown an example of agarose gel with the results of a transferability test at three loci. The locus Ta355 did not amplify, the locus Th127 amplified and showed polymorphism among the species, while the locus Th165 amplified but the amplicon size was unexpected.

The amplifying loci with expected amplicons size were tested by a gradient PCR to assess the melting temperature to ensure the best amplification efficiency. The melting temperature observed varied from 58 °C (Th321) to 54 °C (Th6387).

Twenty-four *T. africana* and four *T. gallica* were used to evaluate polymorphism (Figure 5.3).

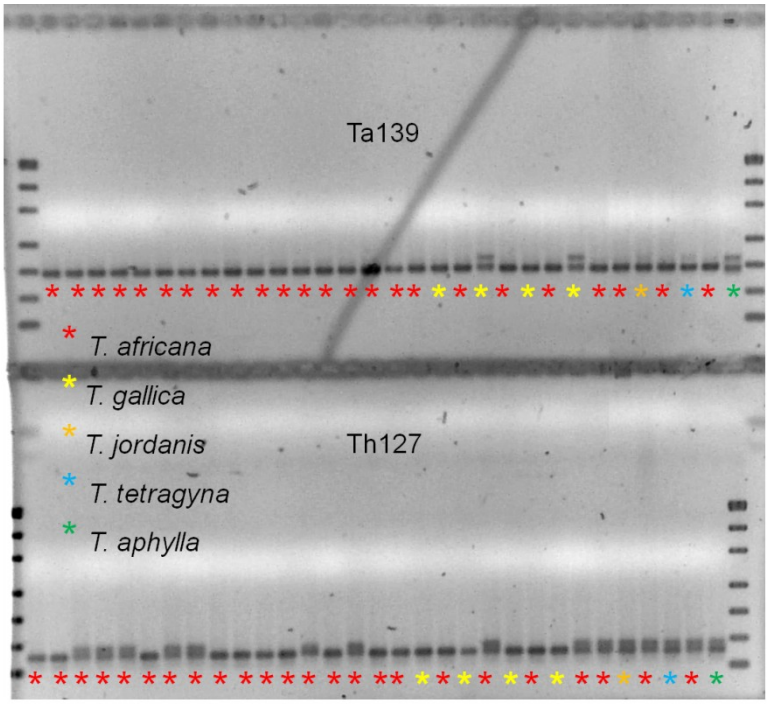

Figure 5.3: Results of the screening of 24 individuals of *T. africana*, four *T. gallica*, and *T. jordanis*, *T. tetragyna*, and *T. aphylla* (one individual for each species) in a high resolution agarose gel. The locus Ta139 was monomorphic in *T. africana* and monomorphic in *T. gallica*, whereas Th127 was polymorphic in *T. africana* but it was monomorphic in *T. gallica*.

It is worth to note that in the case of *T. jordanis*, *T. tetragyna*, and *T. aphylla* only few genotypes were available, thus our results about the polymorphism observed could change analysing a larger sample size. Results of tests for transferability and for polymorphism on the 35 EST-SSRs developed are summarized in Table 5.4.

Species	Transferability	Polymorphism
T. africana	25	13
T. gallica	27	21
T. canariensis	27	21
T. jordanis	26	21
T. tetragyna	27	16
T. aphylla	26	11

Table 5.4: Number of EST-SSR primers transferable and polymorphic in the five species studied.

Polymorphism among species was observed in terms of different allele size as shown in the Figure 5.3.

The ten neutral microsatellites developed in *T. ramosissima* and *T. chinensis* (Gaskin et al. 2006), were tested in the studied species, and nine of them amplified in *T. gallica* and *T. canariensis*, whereas only seven amplified in *T. africana* (Table 5.5). Even the neutral markers were tested by a gradient PCR to assess the melting temperature to ensure the best amplification efficiency. The melting temperature obtained ranged from 57 °C (T1C1) to 60 °C (T1B8).

Species	Transferability	Polymorphism
T. africana	7	6
T. gallica	9	9
T. canariensis	9	9
T. jordanis	10	10
T. tetragyna	9	9
T. aphylla	9	4

Table 5.5: Number of neutral SSR primers transferable and polymorphic in the five species studied.

The locus named T1D12 was not analysed for technical reasons, whereas the locus T1C7 was monomorphic in *T. africana*. Otherwise, between the 13 loci that were polymorphic in *T. africana* one locus (Th127) was monomorphic in *T. gallica* and *T. canariensis*. For these reasons the above mentioned loci were discarded for population genetics analyses, thus in this work five neutral and 12 EST-SSRs were used to detect species identity and genetic structuring of Italian populations of *Tamarix*.

5.7. EST-SSRs putative homology

The 13 polymorphic EST-SSRs developed in *T. africana* were sequenced to allow for assignment of putative homology to known genes and for submission in GenBank. Four loci showed significant similarities to known genes (Th321 with XM_002302479; Th715 with XM_002267690) or with proteins of unknown function (Th2620 with XM_002269093; Th6976 with XM_002270702). The GenBank accession numbers are listed in Table 5.5.

5.8. Characteristics of the novel set of EST-SSRs

The total number of observed alleles ranged from two (Th321 and Th127) to eight (Ta201) with an average of 4.3 alleles per locus. While, the size range of the PCR products varied from 121 bp to 252 bp including 19 bp of M13-tail. In Table 5.6 are reported characteristics of EST-SSR markers.

Locus	Accession	Repeat Motif	Primer Sequence	Ta (C°)	A	Size Range (bp)	Accession of putative omology
Th127	FN686792	$(AGG)_4$	F: TTGGCTGTTGAAGAAGATCG R:TCTCCAAACCTTGACCGACT	58	2	(129-132)	None
Th321	FN597589	$(CTC)_6$	F:TACCTTGCGAACACAACTGC R:TACACCGAGAGAGACGCTGA	58	2	(121-124)	XM_002302479
Ta201	FN686795	$(AT)_7$	F:AATTTGTCCGACTCCACTGT R:CGTCTCCTTTTCAGGCGTAG	56	8	(198-243)	None
Th412	FN686793	$(AG)_{11}$	F:CTGGCAAGTAGCAACACCTCT R: GGATGAACAACCCAACCATC	58	5	(232-250)	None
Th715	FN597590	$(AG)_{10}$	F:ACGTGGTTTGGTGAAAGGAG R:CCACCCTTAACCCACTCAGA	54	5	(126-136)	XM_002267690
Th1071	FN597591	$(TTTC)_4$	F:CGCTCTGTTGATCATCTTCG R:TGTCCCAATCCGTTACAAAA	58	4	(147-166)	None
Ta1350	FN597592	$(GA)_7$	F:CATGGCAGTGATGGATTCAG R:GGACAGTTCAGCCTCCACAT	57	3	(133-139)	None
Th2620	FN597593	$(AAG)_6$	F:GTTGAGCAGCAATCACATGC R:GAAGGGGCAGTGTTTTTCAA	58	3	(226-239)	XM_002269093
Th2876	FN597594	$(CCTGCT)_4$	F:CTGTAGCCAAGCATGGGACT R:AAGACACGTAAACCCGCAAC	58	3	(182-194)	None
Th3484	FN686794	$(CATC)_4$	F:TCAGATTTTGCAAACCACCA R:AAGCCTTTGCATACCACCAC	58	4	(186-202)	None
Th5990	FN597595	$(ATG)_{11}$	F:GCCGAATTTTGTTGTGGATT R:AATAAAAGGCACCCTCATCG	57	2	(177-184)	None
Th6387	FN597596	$(TTA)_6$	F:TCGGATTCTGGAAGGTGTTT R:TGCAACGAAAACATTATTACCC	54	7	(228-246)	None
Th6976	FN597597	$(ATG)_{11}$	F:CCGTGGACTAACCTTGCCTA R:CAAGCAAACGCAGGGTAGAT	58	6	(235-252)	XM_002270702

Table 5.6: Characteristics of 13 polymorphic EST-SSRs in developed *Tamarix africana*. Forward sequence (F), reverse sequence (R), annealing temperature (Ta), number of observed alleles (A), allele size range, GenBank accession no. and Accession no. of putative homology.

No locus deviated significantly from Hardy-Weinberg equilibrium ($P<0.05$), and no significant linkage disequilibrium ($P<0.05$) between locus pairs was observed (Table 5.7).

	Basento River			Imera River			Crati River			Alcantara River		
Locus	A	Ho	He	A	Ho	He	A	Ho	He	A	Ho	He
Th127	2	0.500	0.375	2	0.833	0.486	2	0.167	0.153	2	0.833	0.486
Th321	2	0.333	0.444	2	0.167	0.153	2	0.667	0.500	1	0	0
Ta201	3	0.333	0.542	5	0.500	0.611	5	0.667	0.736	5	0.833	0.722
Th412	3	0.833	0.569	4	0.500	0.681	3	0.750	0.594	1	0	0
Th715	3	0.333	0.486	1	0	0	3	0.333	0.486	3	0.333	0.292
Th1071	3	0.500	0.569	4	0.667	0.597	4	0.500	0.708	2	0.167	0.375
Ta1350	2	0	0.278	2	0.500	0.375	2	0.333	0.278	3	0.167	0.403
Th2620	2	0.333	0.278	2	0.333	0.278	3	0.167	0.292	3	0.500	0.403
Th2876	3	0.500	0.403	3	0.667	0.569	3	0.500	0.569	2	0.333	0.278
Th3484	4	0.667	0.597	3	0.667	0.500	4	0.667	0.681	3	0.667	0.569
Th5990	2	0.167	0.153	2	0.500	0.375	2	0.333	0.500	2	0.500	0.375
Th6387	2	0.167	0.486	4	0.667	0.583	3	0.333	0.653	4	0.667	0.625
Th6976	4	0.667	0.583	5	0.667	0.750	3	0.500	0.500	4	0.333	0.681

Table 5.7: Initial primer screening in populations of *T. africana*. Number of alleles (A) and Observed heterozygosity (H_o) and Expected heterozygosity (H_e) are shown for each population.

5.9. Assignment test

A new technique based on a Bayesian clustering methods was used to assign the unidentified individuals to clusters that correspond to species identity. Our samples are composed by three different species, thus it was expected to find three clusters, one for each species. Values of Ln P(D) began to plateau at K=2, but their values increased slightly for increasing K. The ad hoc quantity of Evanno and co-workers (2005) was investigated to infer the most likely number of clusters, which showed a clear mode at K=2 (Figure 5.4), instead of the expected number of clusters K=3 corresponding to the morphologically identified taxa.

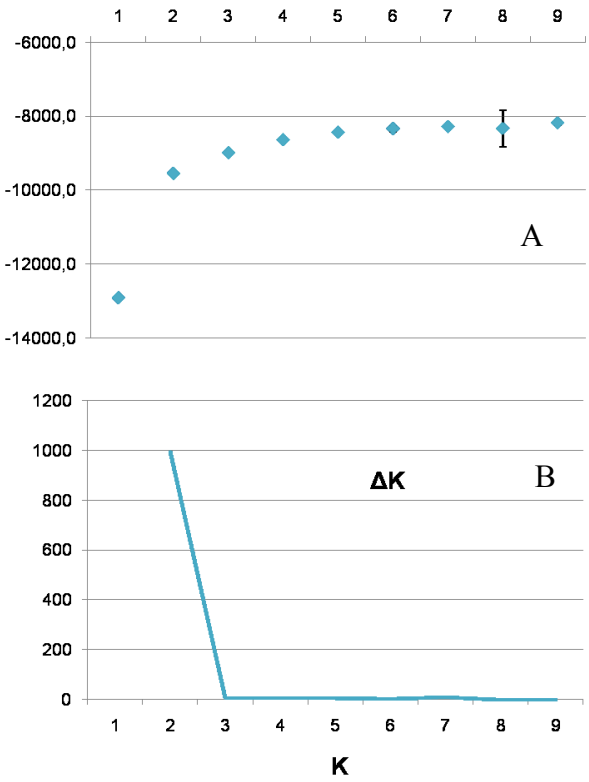

Figure 5.4: Identification of the optimal number of clusters. A. Increase of the likelihood function against the number of cluster K obtained for K=1-9. B. ΔK computed according to Evanno et al. (2005) against the number of cluster.

The mean standard error of the variance in Ln P(D) over all the 20 iterations for K=2 was 0.744. These clusters were considered representative of taxonomic groups, since the cluster one was composed by a group of unidentified samples and all the individuals identified as belonging to the species *T. africana* (194 individuals), while the cluster two included the remaining set of unidentified samples and all the individuals identified as *T. gallica* together with those classified as *T. canariensis* (111 individuals) (Figure 5.5).

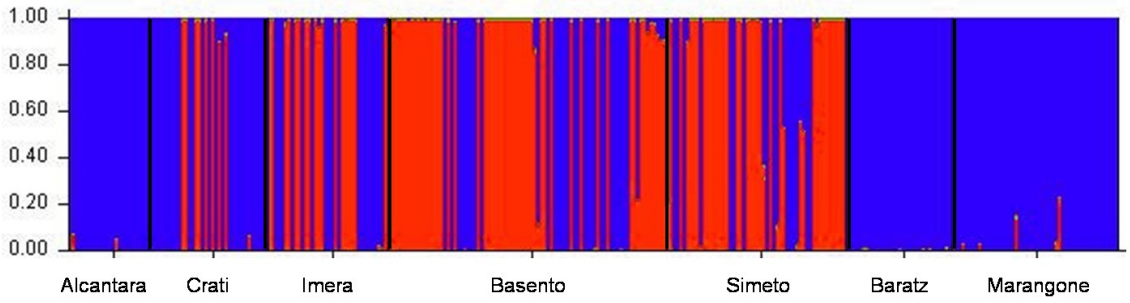

Figure 5.5: The identity of individuals is represented as bar plots partitioned in red or blue colour whose length is proportional to the individual's membership in the cluster (q). The identified plants were used to establish the

correspondence with the inferred clusters: bleu bars correspond to *T. africana*, and red bars correspond to both the species *T. gallica* and *T. canariensis* contemporarily.

It was found that the cluster one comprehends individuals that belong to the species *T. africana*, while the second one determine the existence of a unique group formed by the species *T. gallica* and *T. canariensis*, which hereafter will be called *T. gallica*-like group. The majority of individuals (96.51%) had a posterior probability >0.9 to belong either of the two clusters with and average of 0.97 for the cluster one and 0.96 for the cluster two. Starting from a total of 221 unidentified plants the Bayesian approach assigned 142 individuals to *T. africana*, 78 to the *T. gallica*-like group, while 11 showed an assignment threshold lower that 0.9 and were considered admixed. These 11 individuals resulted not assigned were originated, three in the Basento, six in the Simeto, and two in the Marangone creek population.

A second assignment test was performed following the frequency method of Paetkau (1995) that assigned 122 individuals to the species *T. africana*, 75 to the *T. gallica*-like group, but 34 remained not assigned. Both methods were in agreement in assigning plants to either species, bur Bayesian method performed better than Frequency based method.

Our results allowed to define the species composition in the seven sites considered for the germplasm collection (Table 5.8).

T. africana	Sites	*T. gallica*-like
24	Alcantara	0
18	Imera	18
15	Simeto	33
28	Crati	8
29	Basento	52
32	Baratz	0
48	Marangone	0

Table 5.8: *Tamarix* specific composition of each Italian site.

In *T. africana* the number of individuals collected per sites ranged from 24 in the Alcantara river and 48 in the Marangone creek, whereas in *T. gallica*-like group the number of individuals varied from 8 in the Crati river to 52 in the Basento river. It is worth to note that three sites resulted monospecific stands of *T. africana* (Alcantara, Baratz, and Marangone), four stands resulted mixed with *T. africana* and *T. gallica*-like group contemporarily present in the same site (Crati, Imera Basento, and Simeto), while no any monospecific composition of *T. gallica*-like was found.

5.10. Selection of best performing loci

An assignment test procedure implemented in the program WHICHLOCI (Banks et al. 2003) was used to assess locus-specific assignment power to evaluate what minimum number of high ranking loci are necessary in order to achieve desired assignment accuracy for species identification (Table 5.9).

Rank	Locus	Score	% (Relative Score)
1	T1B8	0.9810	8.1661
2	Ta1350	0.9643	8.0271
3	T1C1	0.9505	7.9121
4	Th715	0.9120	7.5918
5	Th1071	0.8984	7.4790
6	Th6387	0.8522	7.0946
7	Th2620	0.8168	6.7998
8	Th6976	0.8024	6.6798
9	Th412	0.7946	6.6150
10	Ta201	0.7546	6.2818
11	T1E1	0.7281	6.0609
12	T1C10	0.6042	5.0295
13	T1G9	0.5511	4.5875
14	Th3484	0.4508	3.7529
15	Th5990	0.4445	3.7004
16	Th2876	0.3669	3.0543
17	Th321	0.1402	1.1674

Table 5.9: Rank order, locus name and locus score are shown in the table.

This method indicated that the loci T1B8 and Ta1350 are the only required for high assignment success at the 95% stringency level. This finding could allow the creation of a panel of few SSRs that could be used to develop a new method for species identification in the genus *Tamarix*.

Moreover, it was investigated the statistical power of loci in distinguish genotypes within a population. For this reason the probability of identity (PI) was computed for each species. The probability of two unrelated individuals would share the same genotype (PI) and the probability of two full-sib have identical genotype (PI_{sib}) were very low at 17 loci for both *T. africana* and *T. gallica*-like group. In fact, in *T. africana* PI values ranged form 7.1^{-9} in the Baratz population and 2.0^{-11} in the Imera population, while PI_{sib} varied from 2.1^{-4} and 2.1^{-5} in

the Baratz population and in the Imera population, respectively. In *T. gallica*-like group PI values ranged form 6.2^{-11} in the Crati population and 1.3^{-12} in the Simeto population, whereas PI_{sib} varied from 2.5^{-5} and 1.3^{-5} in the Crati population and in the Simeto population, respectively. Moreover, according to the rank pointed out using WHICHLOCI, it was performed the estimation of the statistical power of the markers used in this work. It was determined the minimum number of loci necessary to identify individuals by choosing the number and the combination of loci for which no individuals share the same genotype. In the Figure 5.6 are reported the results obtained for both *T. africana* (graph A) and *T. gallica*-like group (graph B).

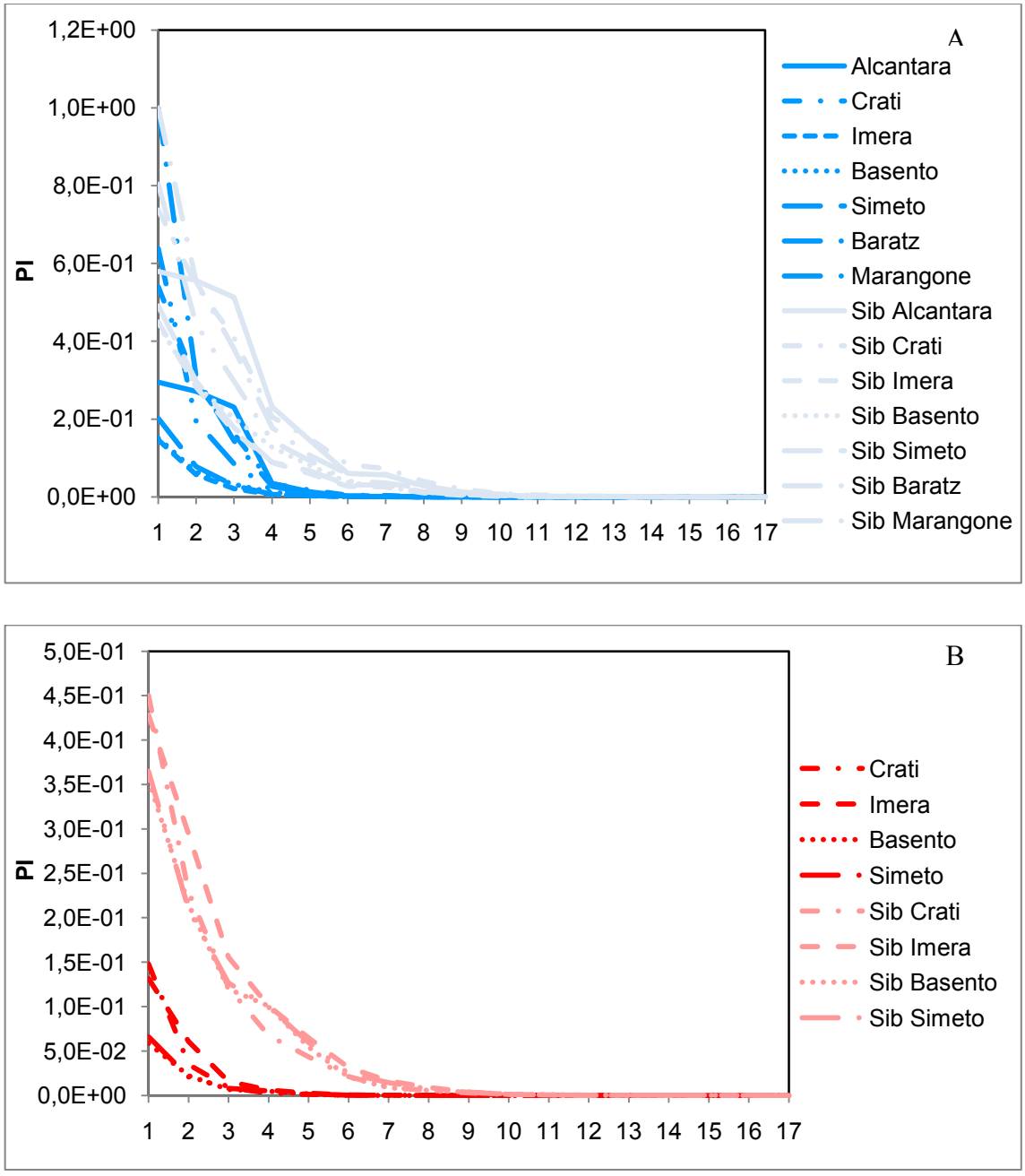

Figure 5.6: Relationship between PI and PI_{sib} for populations and the number of microsatellites used. A results for *T. africana*, B results for *T. gallica*-like. PI provides an estimate of the average probability that two unrelated individuals drawn by chance from the same population have the same multilocus genotype, while PI_{sib} is the probability that two randomly selected fill-sibs would exhibit identical genotypes.

In *T. africana* four and eight loci are required in the case of unrelated individuals and for the presence of relatives, respectively. Otherwise, in *T. gallica*-like group three and seven loci are required in the case of unrelated individuals and in presence of relatives, respectively.

5.11. Population Genetics

Unidentified plants were identified following their distribution within the clusters inferred by the Bayesian approach. Thus, it was possible to perform genetic diversity and population genetics analysis of the Italian natural populations of *T. africana* and *T. gallica*-like group by means of 17 microsatellite markers.

5.11.1. *T. africana*

5.11.1.1. Genetic variability within *T. africana* populations

The total number of alleles detected in the 194 individuals (seven populations combined) at the 17 loci analyzed was 93, with a mean across populations of 3.88 alleles per locus. The Imera population (Sicily) showed the highest mean number of alleles per locus and effective number of alleles (mean A=4.3 and A_e=2.6); whereas the lowest values were observed in the Baratz population (Sardinia, with mean A=3.5 and A_e=2.1). The genetic diversity observed and estimated in each population ranged from H_o=0.51 and H_e=0.55 in the Imera population and H_o=0.39 and H_e=0.44 in the Baratz population (Table 5.10).

	A	A_e	A_p	H_o	H_e
Alcantara	3.941	2.521	4	0.488	0.479
Imera	4.294	2.614	5	0.515	0.552
Simeto	3.588	2.368	0	0.462	0.528
Crati	3.824	2.116	4	0.409	0.478
Basento	3.765	2.220	2	0.380	0.513
Baratz	3.529	2.130	3	0.392	0.439
Marangone	4.235	2.096	8	0.371	0.457

Table 5.10: Genetic diversity parameter of seven population of *T. africana*. In the table are reported the mean number of observed alleles (A), the mean number of effective alleles (A_e), number of private alleles (A_p), the average observed (H_o) and expected heterozygosity (H_e) across loci.

Twenty-six private alleles (A_p) were detected across 12 loci. A_p are equivalent to the number of alleles unique to a single population in the data set. The frequencies of these alleles were always low ranging between 0.01 and 0.22. The distribution of unique alleles among

microsatellites varied from 1 (T1B8, T1C1, Ta1350, Th3484, Th5990, Th6387) to 4 (T1G9), and the microsatellites that present unique alleles were both neutral SSRs and EST-SSRs. In particular, in neutral SSRs 10 private alleles were detected in six populations, while in functional SSRs 12 private alleles were detected in five populations. The populations which displayed the highest number of private alleles is that of Marangone (eight private alleles).

After Bonferroni correction no loci exhibited Linkage Disequilibrium indicating all analyzed loci segregate independently to each other. Significant deviation from Hardy-Weinberg equilibrium (P<0.05) was found within some populations as reported in Table 5.11.

Locus	Alcantara	Imera	Simeto	Crati	Basento	Baratz	Marangone
SSR							
T1B8	-0.145	0.446	0.044	0.071	0.035	-	-0.109
T1C1	1	0.894	0.614	0.108	1*	-0.069	0.191
T1C10	0.153	0.201	-0.186	0.008	0.276	0.231	0.194
T1E1	-0.040	-0.286	-0.139	-0.089	0.087	0.207	0.428*
T1G9	0.013	-0.004	-0.122	0.250	0.260	0.132	-0.016
EST-SSR							
Th321	-	-0.076	0.607	0.376	0.252	-	-
Ta201	-0.029	0.379	-0.076	0.295	-0.477	0.168	-0.170
Th412	-0.163	0.053	0.713*	0.125	0.434*	0.170	0.620*
Th715	-	-	-	-0.050	0.791*	-	-0.034
Th1071	0.436	0.116	0.148	0.030	0.374	0.255	0.447
Ta1350	-	-0.108	0.471	0.286	0.918*	0.446	0.449
Th2620	-0.080	0.113	-0.157	-0.130	0.038	-0.158	-0.083
Th2876	-0.202	0.164	0.428	0.188	0.460	0.297	0.235*
Th3484	-0.098	-0.070	0.328	0.128	0.024	0.009	0.281
Th5990	-0.314	-0.223	0.103	0.372	0.529	-	-0.040
Th6387	-0.133	-0.014	0.263	0.296	0.508	0.003	0.272
Th6976	0.231	-0.167	-0.302	0.136	-0.075	-0.147	-0.064

Table 5.11: Inbreeding coefficient within population F_{IS} values per locus and population. * indicates significant deviation from Hardy-Weinberg expectation with P<0.05 after Bonferroni correction. - indicates monomorphic locus.

In particular, the locus Th412 revealed significant homozygotes excess in three populations (Simeto, Basento, and Marangone). The majority of departures from Hardy-Weinberg equilibrium were accompanied with deficiencies of heterozygotes, especially in the Basento and Marangone populations. In the Marangone population F_{IS} ranged from 0.235 at the locus Th2876 and 0.620 at the locus Th412. Whereas, in the Basento population F_{IS} varied

from 0.434 at the locus Th412 and 1 at the locus T1C1. The locus Th715 resulted polymorphic only in three populations (Basento, Crati, and Marangone) which were originated in Italian peninsula and was monomorphic in the populations derived in Italian islands (Alcantara, Imera, and Simeto from Sicily, and Baratz from Sardinia). No significant heterozygote excess were observed at any locus in any population.

5.11.1.2. Genetic differentiation among *T. africana* populations

Different coefficients of genetic differentiation among populations were estimate for both neutral and EST-SSR markers (Table 5.12). The average values of fixation index (F_{IS}) were positive with a mean of F_{IS}=0.166 for neutral SSRs and F_{IS}=0.189 for functional SSRs. Neutral SSRs showed higher levels of variation respect to the functional markers in all the evaluated parameters, but the differences were weak. In fact, both F_{ST} and R_{ST} values displayed similar values for both kind of markers. The mean F_{ST} value for neutral SSRs was 0.197 and 0.184 for EST-SSRs; whereas the mean R_{ST} value was 0.175 for neutral SSRs and 0.129 for EST-SSRs. Anyway, it is worth to note that the average multilocus R_{ST} values were always smaller than F_{ST} values for both kind of markers.

Locus	F_{IS}	F_{ST}	R_{ST}	D_{est}
Neutral SSRs				
T1B8	0.067	0.174	0.083	0.141
T1C1	0.438	0.367	0.334	0.271
T1C10	0.157	0.190	0.230	0.242
T1E1	0.087	0.113	0.165	0.201
T1G9	0.079	0.141	0.064	0.377
Mean	**0.166**	**0.197**	**0.175**	**0.246**
EST-SSRs				
Th321	0.281	0.194	0.015	0.067
Tan201	0.019	0.193	0.132	0.382
Th412	0.315	0.141	0.079	0.294
Th715	0.452	0.097	0.072	0.013
Th1071	0.295	0.168	0.194	0.301
Ta1350	0.444	0.138	0.135	0.125
Th2620	-0.069	0.248	0.283	0.143
Th2876	0.234	0.209	0.023	0.154
Th3484	0.085	0.123	0.052	0.172
Th5990	0.068	0.284	0.277	0.205
Th6387	0.184	0.253	0.048	0.432
Th6976	-0.046	0.157	0.240	0.270
Mean	**0.189**	**0.184**	**0.129**	**0.213**

Table 5.12: Genetic differentiation coefficients for neutral and functional markers. F_{IS} inbreeding coefficient within population, F_{ST} fixation index among populations according to Weir and Cockerham (1984), R_{ST} differentiation among populations according to Slatkin (1995), D_{est} estimator of actual differentiation according to Jost (2008).

These results were confirmed by the value of actual differentiation D_{est} across populations which presented similar values in both kind of markers. According to D_{est} values per locus, the most informative one is Th6387 (D_{est}=0.432) while the less informative is Th715 (D_{est}=0.013).

5.11.1.3. Population genetic structure in *T. africana*

The genetic divergence between populations was investigated computing pairwise F_{ST} (Table 5.13 and Figure 5.7), R_{ST} (Table 5. 14 and Figure 5.8), and D_{est} (Table 5.15 and Figure 5.9) matrix. Multilocus F_{ST} values varied between 0.027 (Simeto and Imera) and 0.315 (Crati and Marangone). All pairwise F_{ST} values were significantly greater than zero indicating a strong amount of population variability in Italian populations of *T. africana*. In particular Marangone and Baratz resulted the most differentiated populations.

Population	Alcantara	Crati	Imera	Basento	Simeto	Baratz	Marangone
Alcantara	0.0						
Crati	0.099*	0.0					
Imera	0.055*	0.085*	0.0				
Basento	0.183*	0.111*	0.086*	0.0			
Simeto	0.122*	0.117*	0.027*	0.066*	0.0		
Baratz	0.159*	0.234*	0.145*	0.258*	0.168*	0.0	
Marangone	0.265*	0.315*	0.209*	0.240*	0.187*	0.226*	0.0

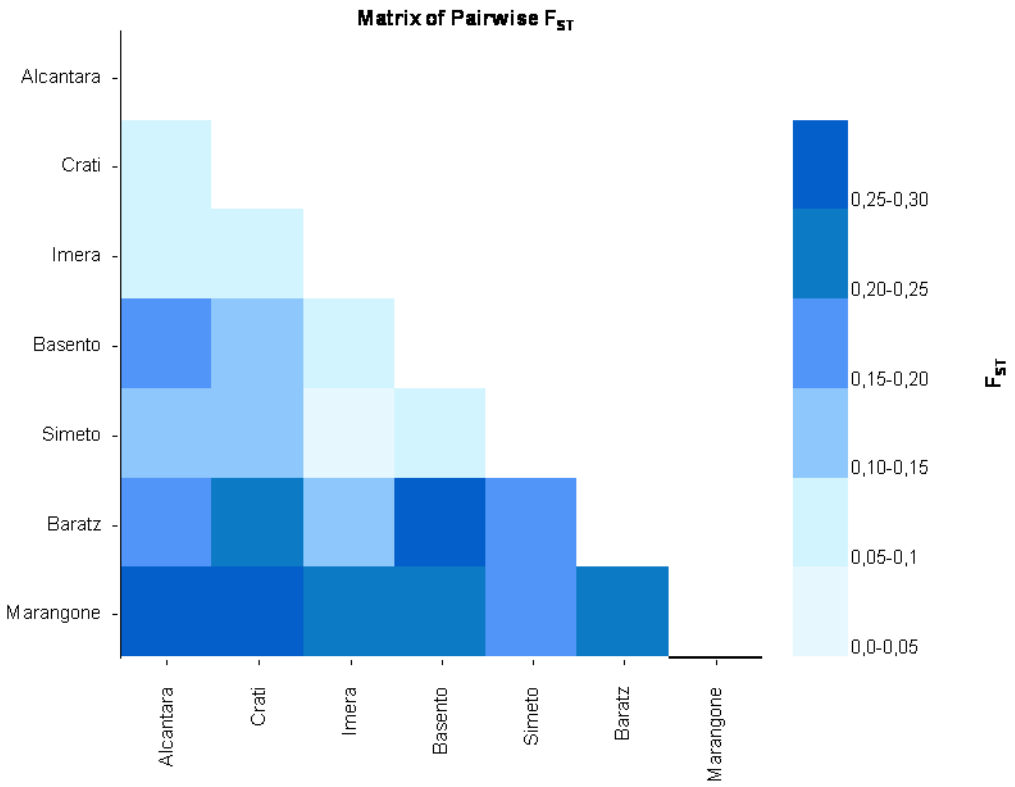

Table 5.13 and Figure 5.7: Pairwise F_{ST} values between seven *T. africana* populations. * indicates that all the values obtained were significant at P<0.05.

Multilocus R_{ST} values ranged between 0.018 (Simeto and Imera) and 0.294 (Basento and Alcantara). All the values obtained were significant with except of the Simeto-Imera comparison. Pairwise R_{ST} values resulted lower than $F_{ST,}$ moreover the genetic divergence of Marangone and Baratz pointed out by F_{ST} values was not observed according to R_{ST} values.

Population	Alcantara	Crati	Imera	Basento	Simeto	Baratz	Marangone
Alcantara	0.0						
Crati	0.118*	0.0					
Imera	0.092*	0.049*	0.0				
Basento	0.292*	0.134*	0.115*	0.0			
Simeto	0.234*	0.117*	0.018^{ns}	0.052*	0.0		
Baratz	0.132*	0.174*	0.159*	0.157*	0.140*	0.0	
Marangone	0.182*	0.127*	0.074*	0.116*	0.097*	0.144*	0.0

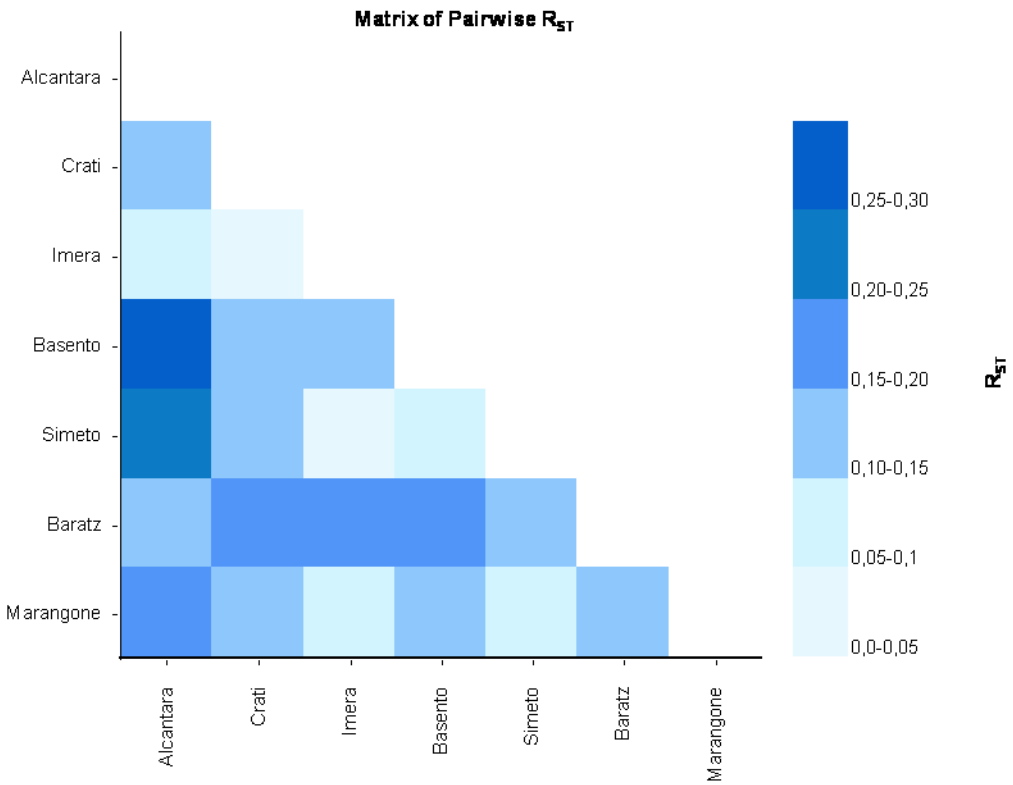

Table 5.14 and Figure 5.8: Pairwise R_{ST} values between seven *T. africana* populations. * marked values were significant at P<0.05; ns indicates not significant.

D_{est} values varied from 0.024 (Imera and Alcantara) and 0.293 (Crati and Marangone). Even pairwise D_{est} pinpointed a strong genetic differentiation between Southern Italy populations

(Alcantara, Crati, Basento, and Simeto) and Sardinia and Central Italy populations (Marangone and Baratz), as previously evidenced by pairwise F_{ST} values.

Population	Alcantara	Crati	Imera	Basento	Simeto	Baratz	Marangone
Alcantara	0.0						
Crati	0.052	0.0					
Imera	0.024	0.064	0.0				
Basento	0.088	0.035	0.044	0.0			
Simeto	0.036	0.063	0.016	0.034	0.0		
Baratz	0.131	0.187	0.127	0.206	0.171	0.0	
Marangone	0.223	0.293	0.198	0.233	0.164	0.149	0.0

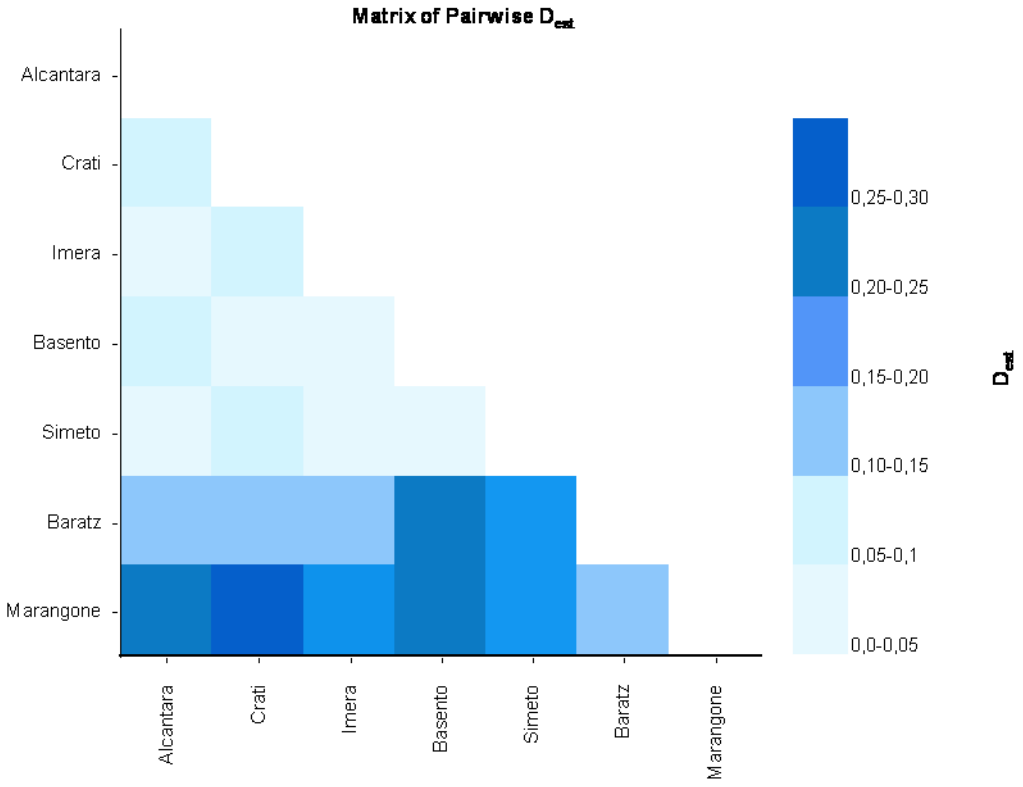

Table 5.15 and Figure 5.9: Pairwise D_{est} values between seven *T. africana* populations.

Genetic divergence among the seven *T. africana* populations was tested by AMOVA using estimator's parameters Phi_{PT} (Excoffier et al. 1992). The variance among population relative to the total variance was significant (Phi_{PT}= 0.246; P<0.001) and indicated a moderate genetic diversity among populations (Table 5.16).

Source	d.f.	Sum of Squares	Variance Component	Percentage of Variation (%)
Among Populations	6	700.303	3.867	25%
Within Populations	187	2221.192	11.878	75%
Total	193	2921.495	15.745	100%

Table 5.16: Analysis of molecular variance (AMOVA) calculated using Phi_{PT} in seven *T. africana* populations.

The AMOVA showed that most of the genetic variability was attributable to differences among individuals within the same population, in fact 75% of the diversity was expressed within population, while 25% of the variation was found among populations (Figure 5.10).

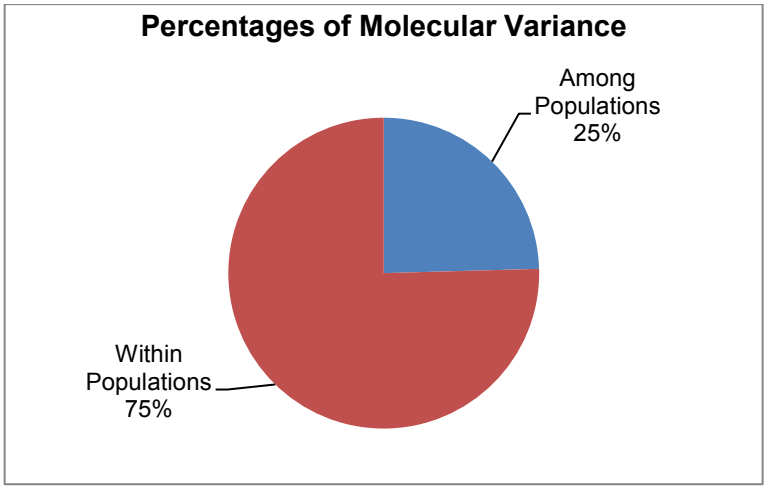

Figure 5.10: Partitioning of molecular variance according to AMOVA calculated using Phi_{PT}.

The analysis of the genetic structure of Italian *T. africana* populations was conducted using the software STRUCTURE. The posterior analysis of the likelihood function according to Evanno and co-workers (2005) pointed out the existence of two distinctive peaks at K=2 and K=5 (Figure 5.11). The peak at K=2 had a greater magnitude, indicating that it was hierarchically stronger.

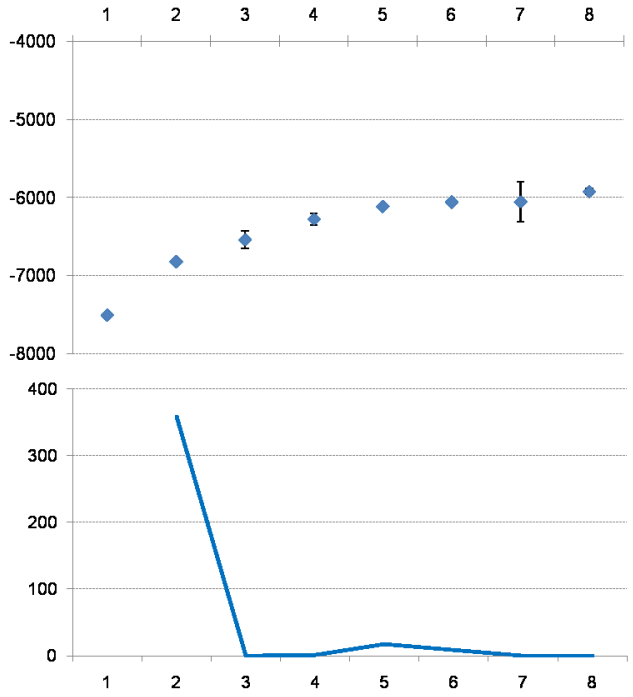

Figure 5.11: Identification of the optimal number of clusters according to ΔK variation (Evanno et al. 2005). ΔK showed two peaks with the peak at K=2 hierarchically stronger.

The model with two genetically distinct clusters delineated two groups that corresponded with Southern Italy (Crati, Basento) and Sicily (Alcantara, Imera, and Simeto) populations in the first cluster, and Central Italy (Marangone) and Sardinia (Baratz) populations in the second one (Figure 5.12).

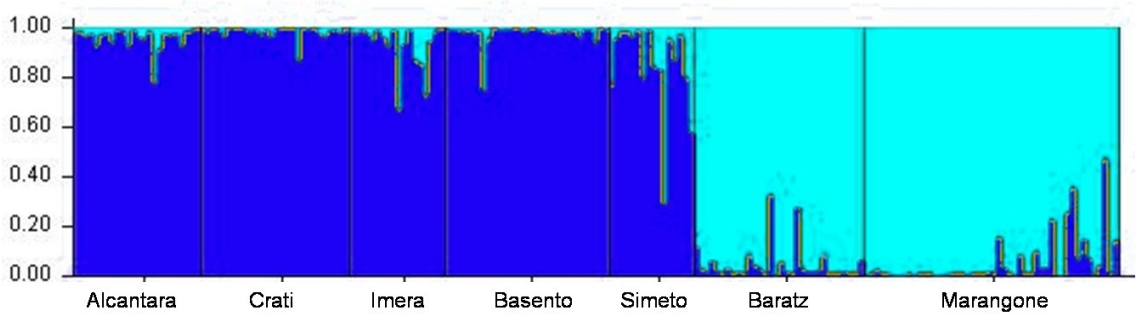

Figure 5.12: Representation of *T. africana* populations according to the model that assumed two clusters. Dark blue bars characterized the first cluster that comprehended individuals derived from Southern Italy (Crati, Basento) and Sicily (Alcantara, Imera, and Simeto) populations, and light-blue bars corresponded to the second cluster that comprehended individuals collected in Central Italy (Marangone) and Sardinia (Baratz) populations.

Individuals were assigned to either clusters with a mean of 93.76% for the first cluster and 95.1% for the second one. Simeto individuals showed the lowest values of assignment with an average assignment to the first cluster of 84.7%, and a mean assignment to the second one of 15.3%.

The alternative structure model with five homogeneous clusters showed that individuals from the Basento, Crati, Baratz and Marangone populations formed four clusters, while the fifth cluster was formed by individuals of the Alcantara and Imera populations. The Marangone and Baratz individuals showed a lower level of admixture with respect to the other clusters, in fact they were assigned to the corresponding cluster with the highest q values observed with 92.3% and 95.8% of assignment, respectively. Populations originated from Sicily showed a higher degree of admixture, in fact, Alcantara and Imera individuals formed a unique cluster (q=83.9% and q=73.6%, respectively) while Simeto population was completely admixed, as its individuals were assigned to two cluster contemporarily (q=42.4% and q=41.4%) as shown in the Figure 5.13.

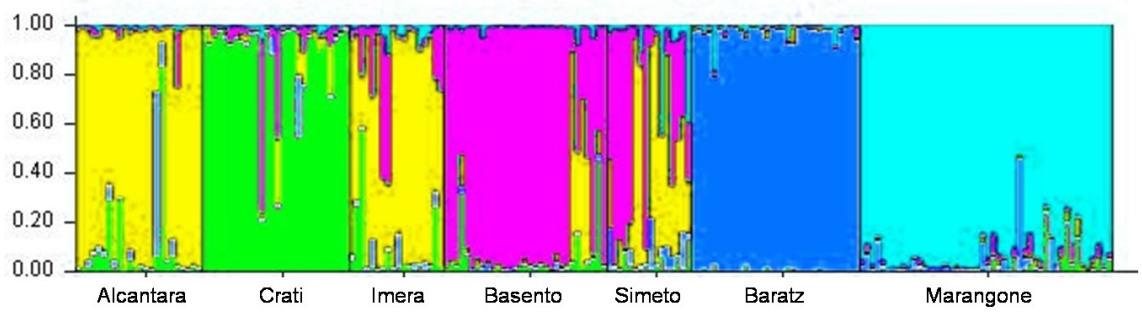

Figure 5.13: Representation of *T. africana* populations according to the model that assumed five clusters. Yellow bars represented individuals from Alcantara and Imera, green bars identified individuals from Crati, pink bars characterized individuals from Basento, dark blue bars represented individuals from Baratz and light-blue bars corresponded to those from Marangone population.

Analysis of principal component (PCoA) based on multilocus genotype data was conducted to provide an alternative graphical representation of genetic structuring of *T. africana* populations. The PCoA confirmed the results obtained by the Bayesian approach, and indicated a high level of genetic affinity in the populations originated in Southern Italy (Crati, Basento) and in Sicily (Alcantara, Imera, and Simeto) (Figure 5.14), while the populations from Central Italy (Marangone) and Sardinia (Baratz) appeared to be more differentiated.

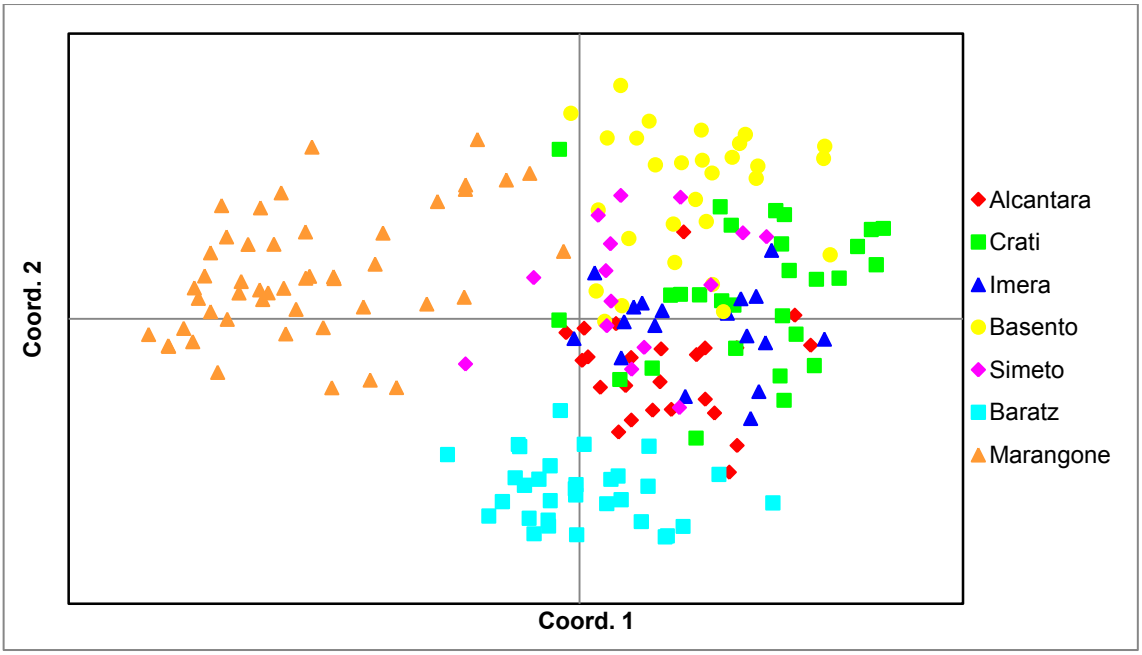

Figure 5.14: Principal component analysis of genotypic data using 17 microsatellite markers in seven Italian natural populations of *T. africana*.

In particular, the first coordinate separated Baratz population from the others, whereas the second coordinate divided Marangone population from the others. It is worth to note that 53.5% of variation was explained by the first two components (33.09% and 20.42%, respectively).

The correspondence between estimates of genetic distance and geographic distance was assessed by a Mantel test for matrix correlation as implemented in GenAlEx 6.4 (Smouse and Peakall 2006). Anyway the test was marginally not significant (P= 0.056) thus the *T. africana* populations were not affected by isolation-by-distance (Figure 5.15).

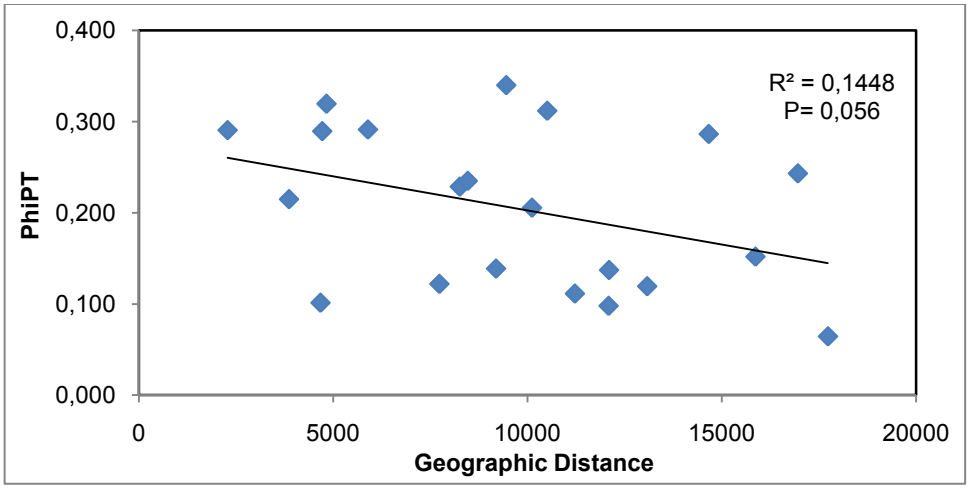

Figure 5.15: Mantel test for isolation-by-distance. Phi_{PT} matrix was plotted against a geographical distance matrix among populations. It was verified the null hypothesis of no correlation thus the *T. africana* populations were not affected by isolation-by-distance.

5.11.1.4. Detection of loci under selection

In this work most of the markers employed were EST-SSRs, which, due to their functional origin, could be subjected to selection in response to local adaptation to different environmental conditions. Thus, it was performed an analysis of outlier loci following the method FDIST2 as implemented in Arlequin 3.5. The neutral locus T1C1 showed F_{ST}= 0.366; this value was significantly above 95% confidence level. While the rest of the loci were included in the confidence level (Figure 5.16).

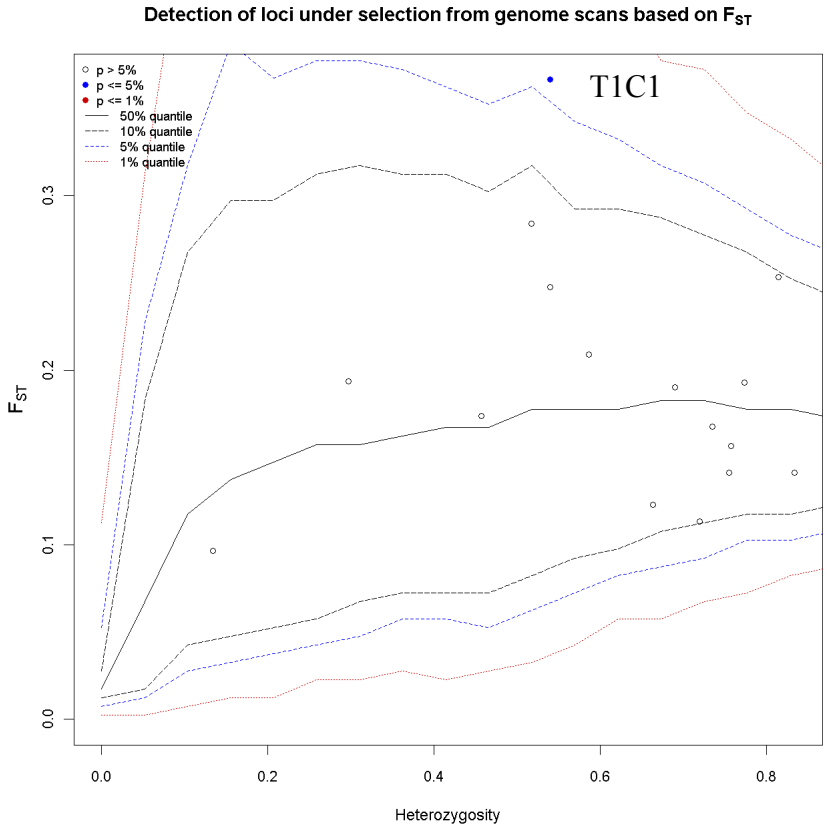

Figure 5.16: Distribution of F_{ST} values in function to the expected heterozygosity. T1C1 was identified as outlier locus above the 95% confidence level indicating selection.

The T1C1 nucleotide sequence was compared with the EST collection of the NCBI database performing a BlastX analysis to assess for a putative functional role of this locus. The translated sequence showed a homology of 63% (E value 3^{-12}) at the amino acidic level with the protein XP_002323308.1 from *Populus trichocarpa*. This protein was a 1-acylglycerol-3-phosphate O-acyltransferase which is an enzyme involved the biochemical pathway of lipids.

5.11.2. *T. gallica*-like group

5.11.2.1. Genetic variability within *T. gallica*-like group populations

The total number of alleles detected in the 111 individuals originated in four populations at the 17 loci analyzed was 89, with a mean across populations of 4.85 alleles per locus. The Simeto population (Sicily) showed the higher mean number of observed alleles per locus and effective number of alleles (mean A=5.23 and A_e=3.04); whereas the lowest values were observed in the Crati population (Calabria, with mean A= 3.41 and A_e= 2.40). The genetic diversity observed and estimated in each population ranged from H_o= 0.49 and H_e= 0.56 in the Imera population and H_o= 0.39 and H_e= 0.55 in the Crati population (Table 5.17).

	A	A_e	A_p	H_o	H_e
Imera	4.47	2.82	4	0.49	0.56
Simeto	5.24	3.05	16	0.46	0.56
Crati	3.41	2.40	2	0.39	0.55
Basento	5.24	3.05	7	0.39	0.55

Table 5.17: Genetic diversity parameter of four populations of *T. gallica*-like. In the table are reported the mean number of observed alleles (A), the mean number of effective alleles (A_e), number of private alleles (A_p), the average observed (H_o) and expected heterozygosity (H_e) across loci.

Twenty-nine private alleles (A_p) were detected across 11 loci. The frequencies of these alleles varied between 0.01 and 0.50. The distribution of unique alleles among microsatellites varied from one (Ta1350, Th2876, T1E1) to five (T1B8, Ta201, T1G9), and the unique alleles were detected in four neutral SSRs and seven EST-SSRs. In particular, in neutral SSRs 14 private alleles were observed, while in functional SSRs 15 private alleles were detected. The population which displayed the highest number of private alleles was Simeto with 16 private alleles.

After Bonferroni correction no loci showed Linkage Disequilibrium. Significant deviation from Hardy-Weinberg equilibrium ($P<0.05$) was observed in four loci within the Basento population, in particular significant deficiency of heterozygotes from Hardy-Weinberg expectation was observed following the inbreeding coefficient (F_{IS}) values reported in Table 5.18.

Locus	Crati	Imera	Basento	Simeto
SSR				
T1B8	0.225	0.117	-0.007	0.260
T1C1	-0.484	-0.038	0.220	0.324
T1C10	0.760	0.457	0.948*	0.338
T1E1	0.011	0.035	0.090	0.064
T1G9	0.307	0.131	0.274	0.235
EST-SSR				
Th321	1.000	-	-0.010	-
Ta201	0.681	-0.126	0.123	0.188
Th412	0.253	0.376	0.431*	0.419
Th715	0.800	0.385	0.896*	0.232
Th1071	1.000	0.575	0.173	0.318
Ta1350	0.815	0.093	-0.012	0.008
Th2620	0.259	0.316	0.253	-0.111
Th2876	-0.076	0.157	0.067	0.377
Th3484	-0.217	-0.054	0.232	0.177
Th5990	-0.176	-0.155	-0.142	0.105
Th6387	0.705	0.511	0.639*	0.482
Th6976	0.127	-0.091	0.322	-0.223

Table 5.18: Inbreeding coefficient within population F_{IS} values per locus and population. * indicates significant deviation from Hardy-Weinberg equilibrium with $P<0.05$ after Bonferroni correction. - indicates monomorphic locus.

In the Basento population significant F_{IS} values varied from 0.431 at the locus Th412 and 0.948 at the locus T1C10.

The locus Th321 resulted polymorphic only in the two populations which were originated in Italian peninsula (Basento and Crati), and it was monomorphic in both the population derived in Sicily (Imera and Simeto). No significant heterozygote excess were observed at any locus in any population.

5.11.2.2. Genetic differentiation among *T. gallica*-like group populations

The same coefficients of genetic differentiation estimated in *T. africana* populations were calculated for *T. gallica*-like populations, as well (Table 5.19). The average values of

inbreeding coefficient within population (F_{IS}) were positive with a mean of F_{IS}=0.248 for neutral SSRs and F_{IS}=0.259 for functional SSRs. The fixation index among population (F_{ST}) values displayed similar values for both kind of markers. In fact, the mean F_{ST} value for neutral SSRs was 0.068 and 0.066 for EST-SSRs. On the other hand, the inverse relationship was observed when R_{ST} was analyzed, since the mean R_{ST} value was 0.036 for neutral SSRs and 0.066 for EST-SSRs. Moreover, when only neutral markers were considered, R_{ST} displayed lower level of genetic differentiation with respect to F_{ST}. According to D_{est} values per locus, the most informative marker is Th412 (D_{est}=0.563) while the less informative is Th2876 (D_{est}=-0.007), with a mean of 0.275 for neutral markers and 0.140 for functional markers.

Locus	F_{IS}	F_{ST}	R_{ST}	D_{est}
Neutral SSRs				
T1B8	0.111	0.053	0.060	0.368
T1C1	0.149	0.034	0.041	0.051
T1C10	0.670	0.095	0.091	0.522
T1E1	0.066	0.056	0.002	0.083
T1G9	0.243	0.105	-0.016	0.353
Mean	**0.248**	**0.068**	**0.036**	**0.275**
EST-SSRs				
Th321	0.464	0.083	0.083	0.011
Tan201	0.139	0.073	0.014	0.525
Th412	0.409	0.160	0.314	0.563
Th715	0.617	0.097	-0.024	0.181
Th1071	0.450	0.049	0.013	0.039
Ta1350	0.086	0.007	-0.001	0.014
Th2620	0.161	0.039	0.024	0.034
Th2876	0.160	-0.012	-0.016	-0.007
Th3484	0.128	0.028	0.060	0.056
Th5990	-0.072	0.022	0.032	0.066
Th6387	0.560	0.189	0.288	0.148
Th6976	0.009	0.060	0.011	0.044
Mean	**0.259**	**0.066**	**0.066**	**0.140**

Table 5.19: Genetic differentiation coefficients for neutral and functional markers in *T. gallica*-like populations. F_{IS} inbreeding coefficient within population, F_{ST} fixation index among populations according to Weir and Cockerham (1984), R_{ST} differentiation among populations according to Slatkin (1995), D_{est} estimator of actual differentiation according to Jost (2008).

It is worth to note that neutral SSRs showed higher levels of variation estimated as average F_{ST} and D_{est} values. Anyway, in the case of F_{ST} the difference between the values obtained by neutral and by functional markers was weak, while D_{est} mean value provided by neutral markers was almost double than that obtained by EST-SSR markers. The average multilocus R_{ST} values for neutral of markers were always smaller than F_{ST} and D_{est} values, while EST-SSRs showed the same F_{ST} and R_{ST} values.

5.11.2.3. Population genetic structure in *T. gallica*-like group

As previously reported for *T. africana* populations, the genetic divergence between *T. gallica*-like group populations was investigated computing pairwise F_{ST} (Table 5.20 and Figure 5.18), R_{ST} (Table 5.21 and Figure 5.19), and D_{est} (Table 5.22 and Figure 5.20) matrix. Multilocus F_{ST} values varied between 0.079 (Basento and Simeto) and 0.103 (Crati and Simeto). Almost all pairwise F_{ST} values were significantly greater than zero indicating a slight, but significant, amount of population structuring in Italian populations of *T. gallica*-like. In particular Crati resulted the most differentiated population.

	Crati	Imera	Basento	Simeto
Crati	0.0			
Imera	0.100*	0.0		
Basento	0.083*	0.103*	0.0	
Simeto	0.103*	0.016^{ns}	0.079*	0.0

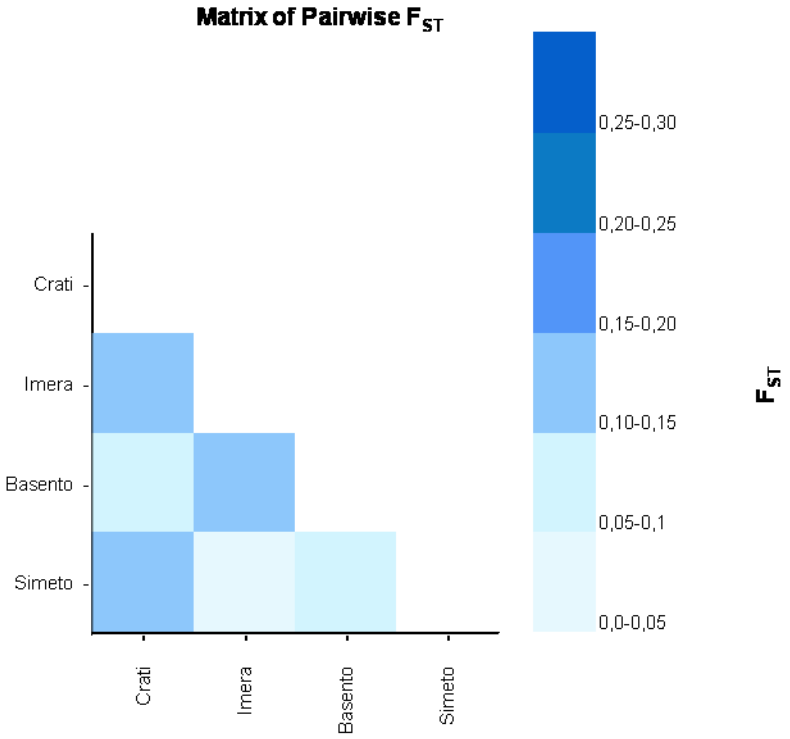

Table 5.20 and Figure 5.18: Pairwise F_{ST} values between four *T. gallica*-like populations. * indicates that all the values obtained were significant at P<0.05; ^{ns} indicates not significant value.

Multilocus pairwise R_{ST} values ranged between 0.016 (Simeto and Imera) and 0.342 (Crati and Basento). Pairwise R_{ST} values resulted generally higher than F_{ST} values, moreover the genetic divergence of Crati pointed out by F_{ST} values was confirmed by R_{ST} values.

	Crati	Imera	Basento	Simeto
Crati	0.000			
Imera	0.342*	0.000		
Basento	0.101^{ns}	0.097*	0.000	
Simeto	0.207*	0.016^{ns}	0.033*	0.000

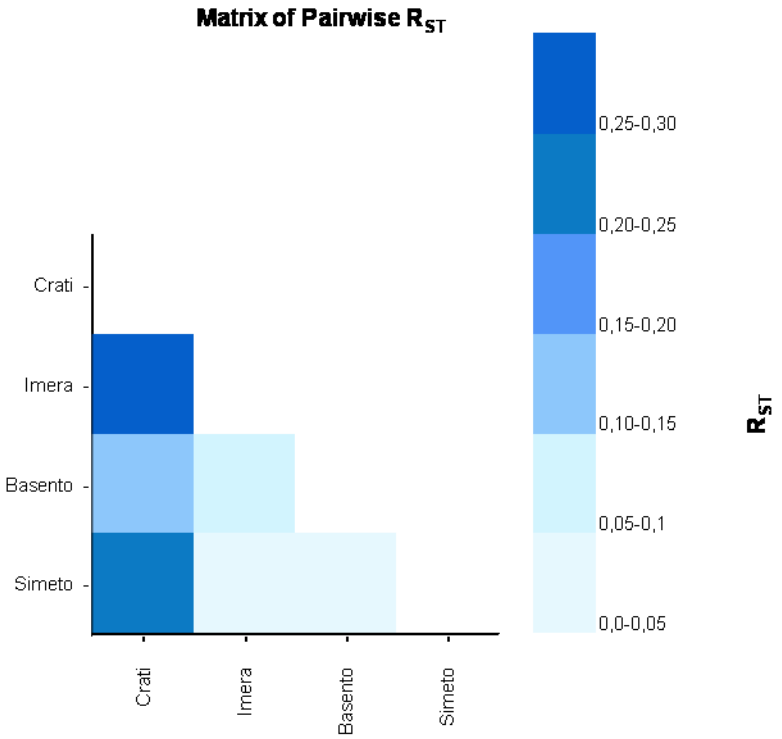

Table 5.21 and Figure 5.19: Pairwise R_{ST} values between four *T. gallica*-like populations. * marked values were significant at P<0.05; ^{ns} indicates not significant.

D_{est} values varied from 0.015 (Imera and Simeto) and 0.089 (Imera and Basento). Even pairwise D_{est} pinpointed a low genetic differentiation between the four *T. gallica*-like populations.

	Crati	Imera	Basento	Simeto
Crati	0.0			
Imera	0.056	0.0		
Basento	0.058	0.089	0.0	
Simeto	0.088	0.015	0.073	0.0

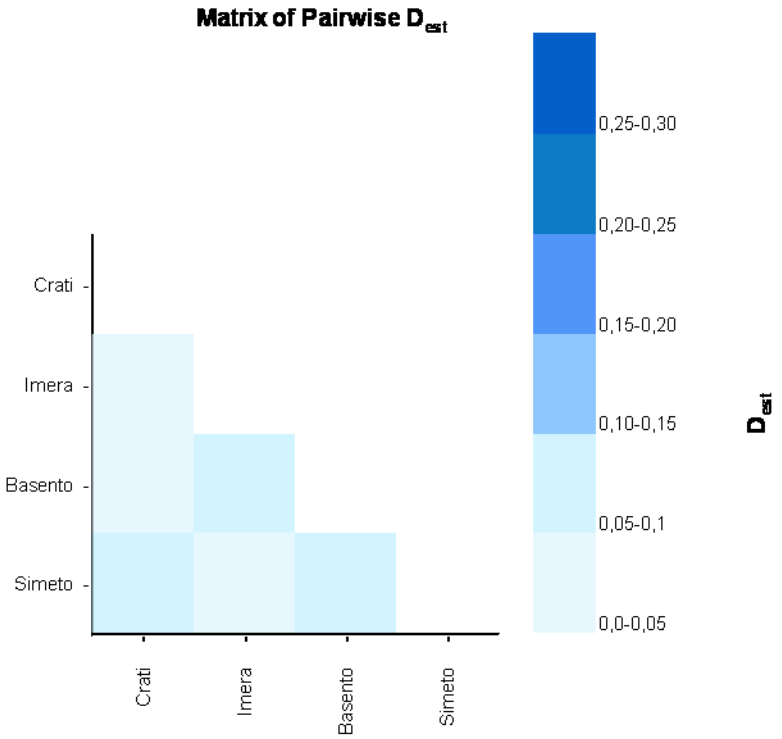

Table 5.22 and Figure 5.20: Pairwise D_{est} values between seven *T. gallica*-like populations.

Genetic divergence among the seven population was tested by AMOVA using estimator parameters Phi_{PT} (Table 5.23). The variance among populations relative to the total variance was significant (Phi_{PT}= 0.095; P<0.001) and indicated a slight genetic diversity among populations.

Source	d.f.	Sum of Squares	Variance Component	Percentage of Variation (%)
Among Pops	3	146.159	1.435	10%
Within Pops	107	1460.778	13.652	90%
Total	110	1606.937	15.087	100%

Table 5.23: Analysis of molecular variance (AMOVA) calculated using Phi_{PT} in *T.gallica*-like populations.

The AMOVA showed that the most of genetic variability was due to differences among individuals within the same population, in fact 90% of the diversity was retained within population, while 10% of the variation was found among population (Figure 5.21).

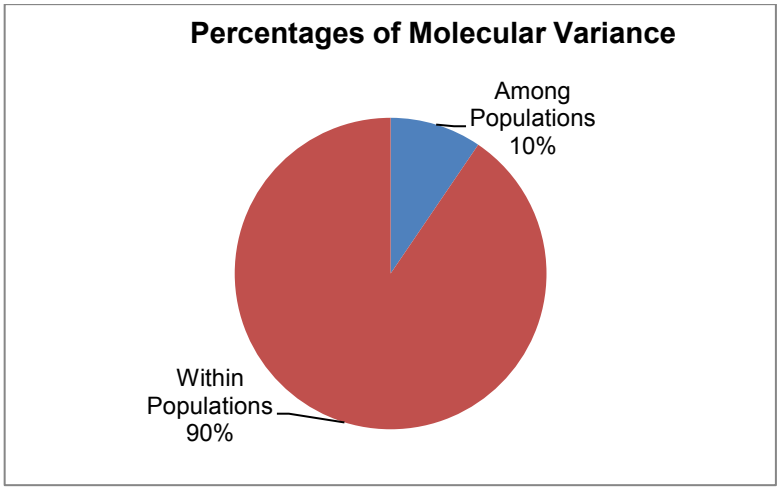

Figure 5.21: Partitioning of molecular variance according to AMOVA calculated using Phi_{PT}.

The analysis of the genetic structure of Italian *T. gallica*-like populations was conducted using the software STRUCTURE. The posterior analysis of the likelihood function according to Evanno and co-workers (2005) evidenced the existence of a peak at K=2 (Figure 5.22).

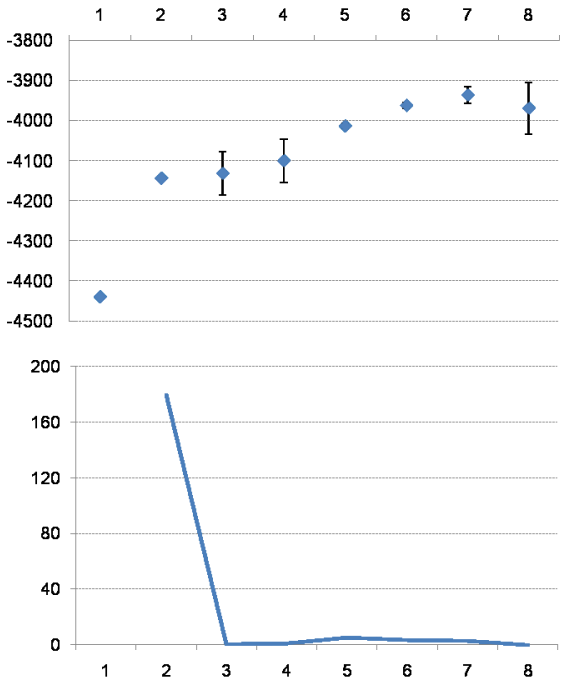

Figure 5.22: Identification of the optimal number of clusters according to ΔK variation (Evanno et al. 2005) that showed a peak at K=2.

The structure model with two homogeneous clusters showed that there was a separation of gene pool between the populations located in Italian peninsula (Basento and Crati) and those originated in Sicily (Imera and Simeto) (Figure 5.23). Individuals from the Simeto and Imera populations were assigned to the first cluster with a probability of 97.9% and 94.5% respectively. In contrast, individuals derived from the Basento and Crati populations resulted more admixed with lower probability of assignment. In fact, individuals were assigned to the corresponding cluster with a probability of 93.2% in the former population and 76.7% in the latter one.

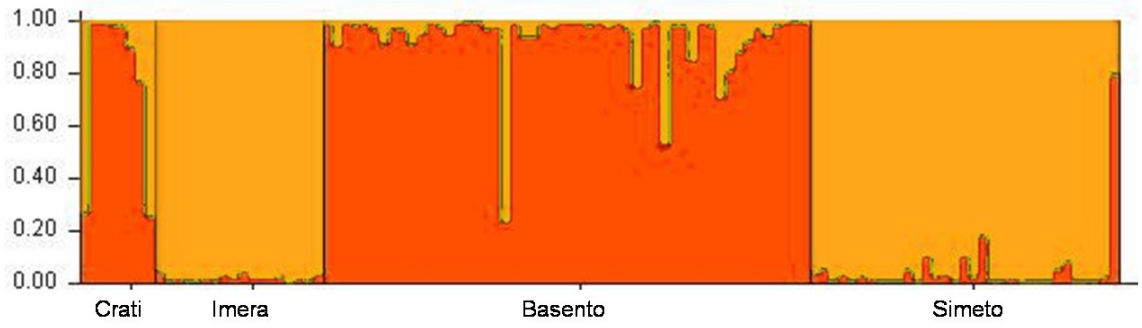

Figure 5.23: Representation of *T. gallica*-like populations according a two clusters model. Orange bars characterized the first cluster that comprehended the individuals from Sicilian populations (Imera and Simeto), and red bars corresponded to the second cluster that comprised the individuals from Italian Peninsula populations (Crati and Basento).

An alternative graphical representation of genetic structuring of *T. gallica*-like populations was obtained by the analysis of principal component (PCoA) based on multilocus genotype. The PCoA confirmed the results obtained by the Bayesian approach, and indicated a high level of genetic affinity in the populations originated in Sicily (Imera and Simeto) (Figure 5.24), while Basento population resulted more differentiated.

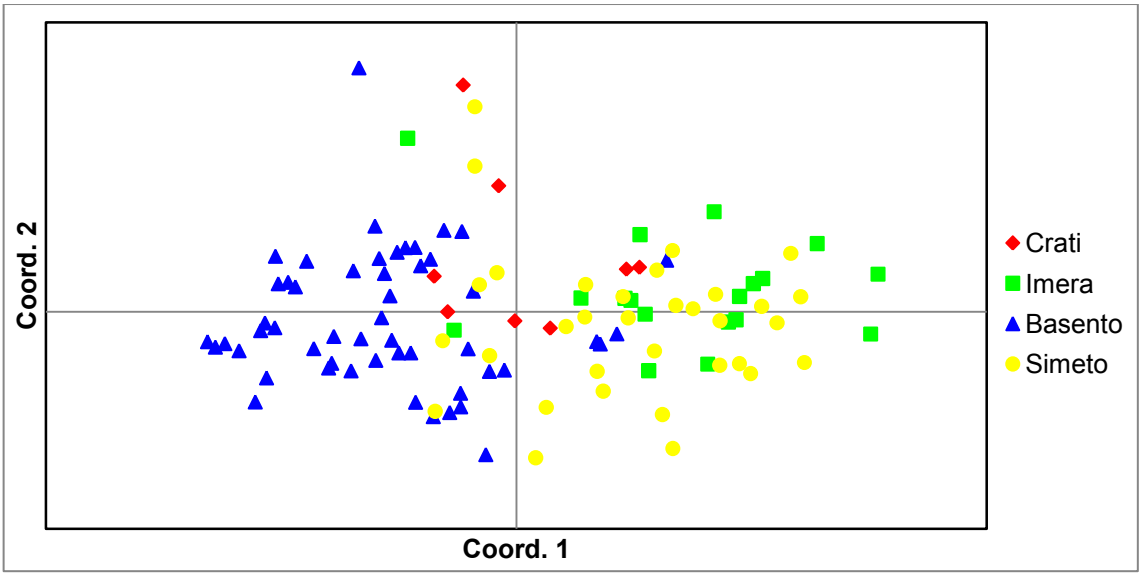

Figure 5.24: Principal component analysis of genotypic data using 17 microsatellite markers in four Italian natural populations of *T. gallica*-like.

In particular individuals from Basento population were divided from the others by the second coordinate, while individuals from Imera and Simeto overlapped and the Crati individuals formed a intermediate group. It is worth to note that 42.65% of variation was explained by the first two components (23.64% and 19.01%, respectively).

The correspondence between estimates of genetic distance and geographic distance was assessed by a Mantel test for matrix correlation as implemented in GenAlEx 6.4 (Smouse and Peakall 2006). Anyway, the test was not significant (P=0.376) thus the *T. gallica*-like populations were not affected by isolation-by-distance (Figure 5.25).

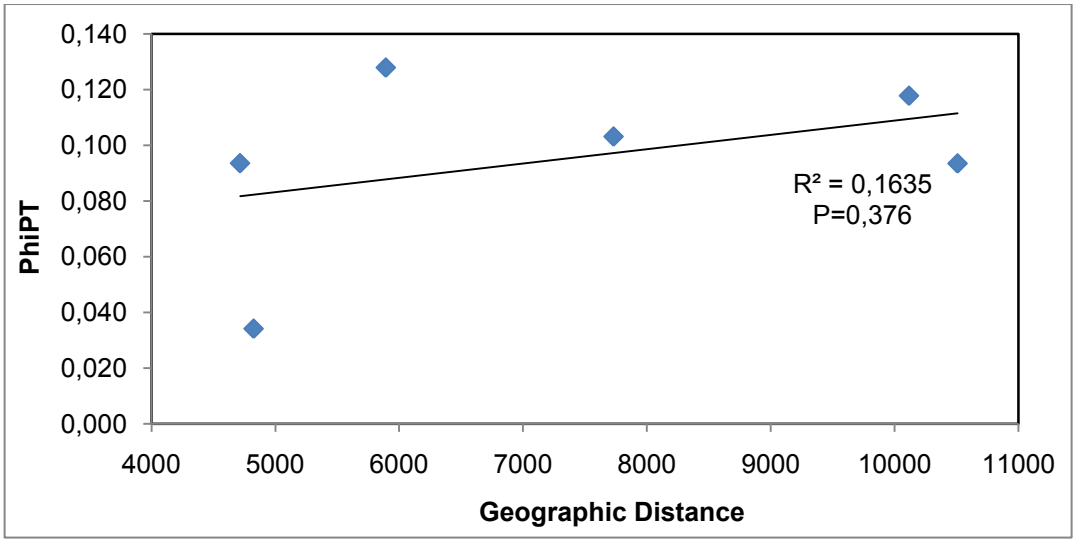

Figure 5.25: Mantel test for isolation-by-distance. Phi_{PT} matrix was plotted against a geographical distance matrix among populations. It was verified the null hypothesis of no correlation thus the *T. gallica*-like populations were not affected by isolation-by-distance.

5.11.2.4. Detection of loci under selection

The analysis of outlier loci was performed in *T. gallica*-like as well using the same method described for *T. africana*. Three EST-SSR loci (Th412, Th2876, and Th6387) showed F_{ST} values significantly above 95% confidence level. While the rest of the loci were included in the confidence level. The locus Th2876 exhibited the lowest F_{ST} as it was a negative value (F_{ST}=-0.011), while higher values were observed in Th6387 (F_{ST}=0.188) and Th412 (F_{ST}=0.160) (Figure 5.26).

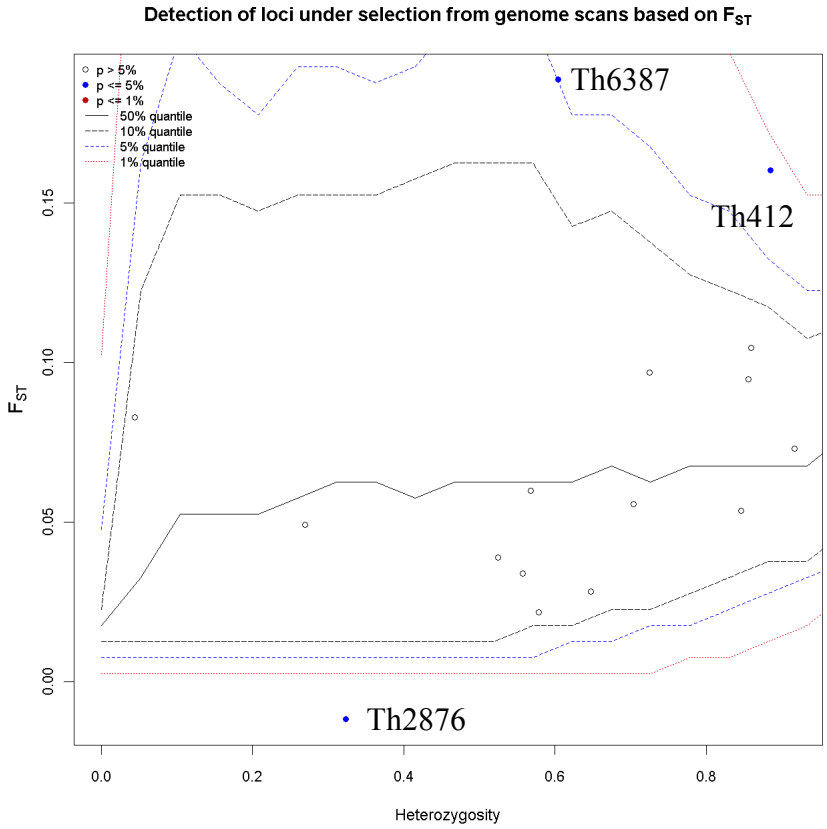

Figure 5.26: Distribution of F_{ST} values in function to the expected heterozygosity. Th412, Th2876, and Th6387 were identified as outlier locus above the 95% confidence level, indicating selection.

The sequences of Th412, Th2876, and Th6387 were previously analyzed in non-redundant EST database by BlastX, but no any homology was found.

Chapter 6

Discussion

6.1. Analyses of Expressed Sequences in *Tamarix*

The unique genetic resources available for the genus *Tamarix* were constituted by expressed sequences of four species (*T. androssowii*, *T. hispida*, *T. ramosissima*, and *T. albiflonum*), that are native in Central Asia. These sequences were processed and redundancy between them was eliminated. Thus, all the sequences were assembled separately for each species in a unigene set formed by both contigs and singlets. Only in *T. androssowii* the majority of EST fell into contigs, whereas in *T. hispida*, *T. ramosissima*, and *T. albiflonum* most of the EST remained as singleton. A possible reason for the presence of a large proportion of singlets for these species is a partial but not complete elimination of redundancy prior to the loading of sequences in GenBank (Kumpatla and Mukhopadhyay 2005).

6.2. EST-SSRs mining in *Tamarix*

The frequency of SSRs detected in this work in *T. androssowii* and in *T. hispida* was comparable with other dicotyledonous species (Morgante et al. 2002). The slight divergence in average distance between EST-SSRs could have resulted from the use of different setting of SSR search criteria, the small size of databases and the database mining tool used. On the other hand, *T. ramosissima* showed frequency somewhat lower than those observed in the former species, but even these values were similar with values previously reported for other dicotyledonous such as *Prunus armeniaca*, *Helianthus annuus*, and *Vitis vinifera* (Kumpatla and Mukhopadhyay 2005). Conversely, *T. albiflonum* displayed the lowest frequency observed, probably for the smaller size of its dataset.

In this work the most representative class of microsatellites was constituted by mononucleotide repeats, followed by trinucleotide microsatellites. Mononucleotide repeat class was reported to be the most common one in eukaryotic genomes (Sharma et al. 2007),

while, despite trinucleotide repeats are less frequent in genomic libraries, they represent the most common class in expressed sequence. A comparative study in five *Brassicaceae* species found the same distribution pattern (Bhati et al. 2010) obtained in our work, with mono and trinucleotide microsatellites overrepresented with respect to the other classes. The authors justified their observation on this pattern assessing that the theoretical assumption of finding mononucleotide repeats in a genome is higher than dinucleotide repeats, followed by tri- and tetranucleotide repeats. Whereas, a different and more practical explanation could be provided for overrepresentation of trinucleotide repeats. In fact, tri- and hexanucleotide SSRs do not cause frame shift mutations when are present in EST, because they are multiples of three which is the number of nucleotides in a codon. This is not the case for di, tetra and penta nucleotide motifs. A frame shift mutation could silence the gene, since it could result in a variation in amino acid residue number at the protein level that affect the protein activity (Cloutier et al. 2009).

6.3. Distribution of EST-SSRs based on their motif and number of repeats

The A/T motif was the most common among mononucleotide repeats in all the four species considered in this work; as previously reported in literature this is the most common motif among mononucleotide repeats (Morgante et al. 2002). Surprisingly, both *T. androssowii* and *T. hispida* displayed AAT as the most common motif between trinucleotide repeats, despite, AAG and CCG repeats were reported as the preferred motifs of coding regions in well investigated plants like *Arabidopsis thaliana*, cereals and legumes (Sharma et al. 2007). It is worth to note that usually AAT motif is least often present in coding regions, as it is associated with stop codons that have a direct effect on protein synthesis (Varshney et al. 2002).

Among the dinucleotide motifs AG was the most common one in *T. hispida*, while AT predominated in *T. androssowii*. It was observed in literature that often the motif AG predominates among the other classes of dinucleotide repeats, presumably reflecting the high frequency of Alanine (AGA) and Leucine (GAG) in polypeptides. Even AT is one of the most frequent motif in plant EST sequences (Stàgel et al. 2008; Cloutier et al. 2009) especially in conifers (Bérubé et al. 2007), but in other works this motif was found to be often associated with non-coding regions (Morgante et al. 2002).

In fungi it was observed a negative correlation between the presence of AAT motif and percentage GC content (Lim et al. 2004). In our work the percentage GC content values observed in the four species analysed ranged between 42,77% and 45,33%; these values were quite lower to those reported for dicotyledonous plants that usually attested around 50-55% (Bhati et al. 2010). Thus our finding confirmed this correlation between frequencies of repeats and percentage of GC content. However, it is worth to note that the percentage of CG content is not univocal determined across genomes, but a gradient along genes has been observed in either monocotyledonous and dicotyledonous plants (Lawson and Zhang 2006).

In this work it was observed a reduction of microsatellites frequencies with increasing repeat length with the five repeat category being the most frequent one. These results were consistent with other works (Varshney et al. 2002).

6.4. EST-SSRs amplification tests and detection of polymorphism

It is noteworthy that in the present work each microsatellite locus was subjected to a cross-species amplification, since the EST-SSRs derived from expressed sequences of four Asian species (*T. androssowii*, *T. hispida*, *T. ramosissima*, and *T. albiflonum*), and neutral SSRs were isolated in *T. ramosissima* and *T. chinensis*. Nevertheless, the aim of this work was the characterization of the natural Italian *Tamarix* populations. Thus, the transferability of both functional SSRs and neutral SSRs was verified in *T. africana*, *T. gallica* and *T. canariensis*, which are the most common species in Italy. *T. jordanis*, *T. tetragyna*, and *T. aphylla* were included in the analyses, as well, to test the transferability among species geographically distant, which, for this reason, should be genetically different. Surprisingly, the divergence of transferability between EST-SSRs and neutral SSRs was slight in the case of *T. africana* (71,4% and 70% respectively). On the other hand, in *T. gallica* and *T. canariensis* EST-SSRs reached 77% of transferability, and the most transferable primer pairs were those designed for neutral SSRs, which presented 90% of transferability. Usually in literature were reported greater value of transferability for EST-SSR, and lower for neutral SSR, since the functional constraints in transcribed regions could affect even primer site mutation that hence should be lower variable (Woodhead et al. 2005). Anyway, the range of transferability of EST-SSRs observed in this work was consistent with values reported in literature, that assessed a range of 60-80% of transferability for these kind of markers (Ellis and Burke 2007). Otherwise, it is well established in literature that for neutral markers the cross-species transferability is

unevenly distributed among different taxa, and a sort of taxonomic range effect was found for the successful markers amplification (Barbarà et al. 2007).

In particular, in the case of neutral SSRs, three primer pairs failed to amplify the *T. africana* genome, indicating divergence of the primer site sequences between species *T. africana* and *T. ramosissima - T. chinensis*; whereas the greater transferability of the same markers in *T. gallica* could be linked to a putative higher genetic affinity with the *T. ramosissima - T. chinensis* complex. Anyway the small number of neutral markers could bias these results that, for this reason, could be not representative of the true genomic divergence between the above mentioned species.

The differences of polymorphism upon cross-species transfer is dependent on the type of repeats but also on the features of the sequences located in the regions flanking the microsatellite (see Chapter 2). It is worth to note that even if a locus is conserved among different species, its polymorphism could not, thus the same SSR marker could be informative in a species but not in another one.

6.5. Characteristics of the novel set of EST-SSRs

It has been reported that EST-SSRs are usually less polymorphic with respect to those derived from genomic libraries. However, in the present work the existing genetic sequence information represented a unique opportunity to develop a novel set of molecular markers in *Tamarix*, EST-SSRs. In fact, the most important feature of EST-SSR is their transferability even among relatively distant species. Thus, the newly developed EST-SSRs characterized in this work represent an additional resource for the genetic characterization of all the species that belong to the genus *Tamarix*. Our EST-SSRs could be used as novel tools for species identification, for breeding programs and management of natural resources. It is noteworthy that from the past century the genus *Tamarix* has been invasive in North America, so a further employment of our markers could be to track the spread of invasive populations.

The novel set of EST-SSRs were characterized by a smaller number of observed alleles per locus with respect to the genomic SSRs isolated in *T. ramosissima* and *T. chinensis* by Gaskin and co-workers (2006). Our EST-SSRs did not exhibit significant Linkage Disequilibrium or deviation from Hardy-Weinberg equilibrium. On the contrary, six genomic SSRs displayed significant deviation from Hardy-Weinberg equilibrium, and in 47% locus pairs significant Linkage Disequilibrium was observed. But it should be underline that for the

screening of EST-SSRs polymorphism, it was employed 24 individuals while the genomic SSR were tested on a larger sample size (Gaskin et al. 2006). The authors justified their results with a possible close geographical relationship between the collected individuals that may have influenced these findings.

6.6. Species assignment by Bayesian approach

Species are considered the fundamental taxonomic unit of biology, and an accurate and efficient identification of species is usually required before any investigation in ecology, conservation and breeding (Millar et al. 2008). Anyway, the taxonomy of the genus *Tamarix* is a true challenge for botanists as species cannot be distinguished in vegetative status, and even if flowers are present their morphological features are often misleading for species identification (Gaskin 2003). Despite three taxa were identified among our samples (*T. africana*, *T. gallica*, and *T. canariensis*), unidentified plants were the largest group, representing 72% of the total plants collected. The species identity of subset of individuals were determined following morphological traits according to Baum's morphological key (Abbruzzese, personal communication), and the distribution of these individuals within genotypic clusters was assessed by a multilocus Bayesian clustering method. Thus, identified plants were used to establish the correspondence between taxa and genetic clusters, allowing the species identification of our unidentified samples not on the basis of their morphological traits, but on the use of microsatellite markers coupled with Bayesian assignment methods. The optimal number of cluster was two, instead of the three expected clusters that should correspond to the morphologically identified taxa. The partitioning of clusters showed a clear assignment of *T. africana* individuals, but it was not same for *T. gallica* and *T. canariensis*, whose individuals grouped together in the same cluster. These two species appeared completely genetically homogenised, even if the plants were collected in different site of origin. The existence of a unique group formed by *T. gallica* and *T. canariensis* was consistent with the hypothesis advanced by Gaskin and Schaal (2002) who suggested that these two species may be the same taxon or may be introgressive. An alternative hypothesis is that *T. gallica* and *T. canariensis* could be two species in early stage of divergence. In fact, species could share high levels of ancestral polymorphism before lineage sorting fixed alleles in different groups. On the other hand, gene flow between introgressive species could lead to shared polymorphism due to incomplete reproductive barriers (Drummond and Hamilton 2007).

Anyway, levels of differentiations reflects past and present barriers to reproduction that is the fundamental criterion in the biological species concept and for the delimitation of species (Avise and Walker 2000), although the definition of biological species is still matter of debate (Goldstein et al. 2000; de Queiroz 2007).

Another likely explanation of our results could be related to the fact that if the traits used to identify the species are under control of a few genes with categorical effects, microsatellite loci which are unlinked to these traits may fail to detect species divergence (Ochieng et al. 2010). Otherwise, natural selection could generate the same situation. For instance, it is possible that species identification traits consisted of genome regions that are under strongly disruptive selection, while the remaining part of the genome is neutral, thus it could generate phenotypical differences that are not supported by molecular data (Beaumont 2005). This suggest that morphological differences does not necessary reflect neutral genetic variation, and in some cases, detectable DNA differences between spatially distant population of the same species may be as great as or greater than the differences observed between species (Ochieng et al. 2010).

A second individual assignment test was applied following the frequency based method of Paetkau (1995). Both the frequency based method and the Bayesian assignment methods assumed markers being in Hardy-Weinberg equilibrium. Moreover, differently from the frequency method, the Bayesian clustering method does not necessary require a priori information about species partitioning, and it does not require loci at linkage equilibrium. In this first phase all individuals collected in all sites were analysed contemporarily, shuffling individuals from different populations. Thus Hardy-Weinberg expectation could be not satisfied. Anyway, the Bayesian clustering method and the frequency methods generally performed better that other assignment method, even if the assumption of Hardy-Weinberg equilibrium was not fulfilled (Cornuet et al. 1999). Thus both the assignment method used in this work could tolerate to some violation to such assumption. It was reported that Bayesian method could tolerate deviations from Hardy Weinberg equilibrium when a relatively large number of loci were employed, and a high genetic divergence between species was observed (Seikino and Hara 2007). Nevertheless, the Bayesian method displayed a higher performance with respect to the frequency based method, since it was possible to assign 305 individuals to the corresponding species, while only 11 individuals remained undetermined specimens. It is worth to note that individuals assigned to a species by the frequency based methods, were always assigned to the same species by the Bayesian one, but the number of undetermined individuals was higher when the frequency method of Paetkau (1995) was used.

It is worth to note that among the 11 unidentified individuals, seven of these showed a probability of assignment that narrowly missed the threshold of 90%. While three individuals collected from the Simeto population displayed lower probability of assignment with values around 50%. These three individuals should be further investigated, as their genome resulted assigned contemporarily to two clusters and could be interpreted as putative hybrids between *T. africana* and *T. gallica*-like. It was previously reported that invasive *Tamarix* populations in North America are mainly composed by hybrids between *T. chinensis* and *T. ramosissima*, two species that in the native range did not form hybrids since a species specific edaphic affinity (Gaskin and Schaal 2002). In the present work it was found a small number of putative hybrids even if the species *T. africana* and *T. gallica* are sympatric. The reason for this extremely reduced gene flow may be related to a different phenology that characterize these species. In particular, *T. africana* usually booms between March and May, whereas *T. gallica* blooms in April to September (Baum 1978).

6.7. Selection of best performing loci

An assignment test procedure implemented in the program WHICHLOCI (Banks et al. 2003) was used to assess locus-specific assignment power, to evaluate what minimum number of loci are necessary to achieve assignment accuracy for species identification. Although many studies reported a converse relationship between number of loci and number of individuals in order to achieve assignment, no simple predictor, such as number of alleles or heterozygosity of loci provided consistent prediction of individual based assignment performance (Cornuet et al. 1999). In fact, the assignment efficiency could be affected by several factors such as number and variability of markers, genetic divergence between species, and sample size (Seikino and Hara 2007). While more polymorphic and heterozygous loci generally showed higher rank, on occasion specific loci with as few as four or five alleles may rank high because of exclusive frequency and distribution of genotypes among populations under consideration (Banks and Jacobson 2004). The advantage concerning the use of molecular markers for classification purposes is that they can be used to assess differentiation across a wide range of taxonomic levels and address questions of species status by comparing inter-intra taxa differentiation. Thus, a high assignment power panel could be employed not only in the species studied in the present work, but could represent a new, fast and cheap method for species identification in all the taxa belonging to the genus *Tamarix*. It

is noteworthy that only two loci were required to achieve the assignment stringency threshold selected for this work. These two loci corresponded to the neutral marker T1B8 (Gaskin et al. 2006), and Ta1350, an EST-SSR developed in the present work.

WHICHLOCI indicated which markers were required for the correct assignment to the species. Thus, a second test was performed to detect the statistical power of our markers in discriminate individuals. For this reason, it was calculated the probability of identity (PI) to ensure the statistical confidence for individuals discrimination of the markers employed (Waits et al. 2001).

The probability of identity PI was very low at 17 loci for both *T. africana* and *T. gallica*-like group. This indicate that it is improbable at any pair, including siblings pair, to find two individuals with the same genotypic profile across all 17 loci. As previously reported in literature for other species, the PI and PI_{sib} estimators differed by approximately twofold in the number of loci required to achieve a $PI < 1^{-4}$ using codominant loci (Waits et al. 2001).

6.8. Population genetics

The aim of this work was to assess the level and the distribution of genetic variation of *T. africana* and *T. gallica*, the most common species of tamarisks in Italy, in order to get essential knowledge for planning activities in conservation and management of natural resources. As pointed out in Chapter 1, there is no information about *Tamarix* Italian natural germplasm. Moreover there is a lack of knowledge about the population genetics of the entire genus *Tamarix*, since no other works deal with this issue. Thus, it was very difficult to drawn our conclusions, since we could compare our results only with two works conducted on the invasive North American population by Gaskin and co-workers (2006) and Friedman and co-workers (2008).

6.8.1. *T. africana*

6.8.1.1. Diversity within population

Genetic diversity assessed by microsatellites markers in our seven Italian populations of *T. africana* was moderately low but comparable with the heterozygosity values observed in the invasive population in North America, but with fewer observed alleles per locus. It is noteworthy that these values are not consistent with high genetic diversity values usually reported for other forest trees (Belaj et al. 2007; Ferrazzini et al. 2007; Scalfi et al. 2009), but our results are comparable with values observed in other halophyte plant such as *Thellunginella salsuginea* (Gao et al. 2008), and *Populus euphratica* (Wu et al. 2009; Pascal et al. 2009). A low genetic diversity is usually associated with an erosion of the genetic diversity, bottleneck events or inbreeding. Although, extinction endangered species could present lower heterozygosity values than those observed in our work. Nevertheless, despite the narrow genetic diversity detected in USA invasive population (Gaskin et al. 2006), plasticity allowed tamarisks to spread and survive in a wide range of environments (Sexton et al. 2006). For these reasons, tamarisks plants cannot be considered endangered, anyway a wider work should be done, as the paucity of population genetic studies in *Tamarix* did not allow us to address concrete hypothesis, and the comparison with populations in a wider geographical range could be desirable

Almost all the populations studied hold exclusive alleles, but most of the private alleles detected were found in the Marangone Creek population. The presence of exclusive or private allele could be considered a measure of genetic distinctiveness and could be related to a low exchange of gene flow between *T. africana* populations.

The positive values of the inbreeding coefficient F_{IS} for most populations may be due to the presence of null alleles, although EST-SSRs were reported to be less prone to exhibit null alleles (Ellis and Burke 2007). Otherwise, an alternative hypothesis could be related to the sub-structuring of populations in subunits within which mating is more probable (Wahlund effect). Anyway in our populations plants were collected at a distance of about 50 m from one another, and this should exclude the presence of any significant sub-structure (Ferrazzini et al. 2007). A further explanation of excess of homozygotes is self-fertilization and the capability of *Tamarix* to vegetative reproduction from woody fragments. In fact, self-fertilization has been reported in *Tamarix* with smaller but viable seeds (Gaskin and Kazmer 2006).

6.8.1.2. Differentiation among populations

The analyses of genetic differentiation indices pinpointed a greater differentiation power of neutral markers, while EST-SSRs displayed a lower polymorphism. These results could be explained considering that EST-SSRs are designed on expressed sequences and therefore more conserved than sequences derived from non-coding regions.

A lower variability among populations was observed considering R_{ST} values with respect to the F_{ST} counterparts. This result was surprising, especially following the assertion that under stepwise mutation model, F_{ST} underestimates the degree of genetic differentiation among populations (Slatkin 1995, Jost 2008). Under this assumption, it was expected a smaller value of F_{ST} than that of R_{ST}. This discrepancy could be related to a deviation of our markers from the stepwise mutation model, to the large variance of this statistic, or to the relative effect of mutation and genetic drift (Lugon-Moulin et al. 1999). In particular, the low differentiation estimated by R_{ST} in this work is consistent with the hypothesis that drift and not mutation resulted in the differentiation between populations. In literature this observation was related to the presence of a founding event, where a new population share the same allele size range of the source population, but allele frequencies may diverge between populations due to genetic drift (Sefc et al. 2007). In this situation F_{ST} detects alleles frequency differences, on the contrary R_{ST} results in similar variance in alleles size. In general it has been reported a large variance in all studies that compare F_{ST} and R_{ST}. Although, R_{ST} provide a more accurate estimate of interspecific divergence that is better detected in a longer historical separation, gene flow is low, and the mutation rate is high. Whereas F_{ST} appear to be more sensitive to detect intra-specific differentiation (Balloux and Lugon-Moulin 2002; Balloux and Goudet 2002; Sefc et al. 2007). It is noteworthy that genetic drift is an evolutionary force which determines changes in alleles frequencies. Sometimes and under particular situations, such as in small populations for instance, it could leads to the fixation of few alleles. Our results suggest that genetic drift is not a synonymous of genetic erosion, but an evolutionary force that could have the main role in shaping the genetic diversity of the Italian populations of *T. africana*.

Moreover, these results were confirmed by a further estimator, D_{est}. In fact, F_{ST} estimators are often reported to be affected by their dependence of increasing polymorphism, especially when using highly polymorphic loci such as microsatellites (Jost 2008).

6.8.1.3. Population genetic structure in *T. africana*

Both the pairwise F_{ST} and D_{est} matrixes underpinned a moderate to great genetic divergence among the Southern Italy populations (Alcantara, Crati, Basento, and Simeto), and Sardinia and Central Italy populations (Marangone and Baratz). On the contrary R_{ST} did not show the same pattern, confirming the principal role of genetic drift in explaining the variability between populations. The analyses of molecular variance AMOVA showed a moderate intra-population diversity and a great genetic differentiation among *T. africana* populations; although, most of their variation was retained within populations as in most woody perennial outbreeding species (Belaj et al. 2007). It is worth to note that evidences of a large genetic variability between populations were found. Despite the partitioning of genetic variance pointed out that most of the variation was retained within populations, this percentage was lower than that observed by Friedman and co-authors (2008) in the invasive population in North America. In fact, the invasive *Tamarix* population in Central USA showed a lack of strong genetic isolation (Friedman et al. 2008), as the fixation index F_{ST} displayed a low value that indicated the absence of barriers to gene flow among invasive populations. Whereas the AMOVA indicated that the genetic variation within populations was large compared to variation between populations.

These observations were confirmed even by a Bayesian assignment procedure implemented in STRUCTURE, which allowed the detection of two main gene pools in *T. africana* populations, with a clear separation between geographical provenances. Individuals assignment showed a lower level of admixture in the Marangone and in the Baratz population, while the Southern Italy (Crati and Basento) and the Sicilian (Alcantara and Simeto) populations displayed greater values of admixture, that may be explained by the presence of a common genetic pool, as well as some extend of gene flow between these latter populations.

Then, a Mantel test to detect isolation-by distance was performed, but this test resulted marginally not significant. Thus, the genetic structure of Italian *T. africana* populations was not affected by isolation by distance, and other factors could be involved in shaping the genetic variability.

6.8.1.4. Detection of loci under selection

Finding genomic regions under selection is one of the first steps required to bridge the gap between the genotype and phenotype of adaptive traits, thus it is crucial to understand the process of adaptation (Siol et al. 2010). Loci under selection were identified by estimating the difference in F_{ST} of EST-SSRs and neutral SSRs among populations, and identifying outliers loci. Surprisingly, at 95% confidence level using Arlequin 3.5, the neutral locus T1C1 showed a possible outlier F_{ST} value indicative of selection. This locus was submitted to the non redundant protein database (NCBI) and resulted homologous of a *Populus trichocarpa* gene. This gene encode for an enzyme involved in the lipids metabolism that is over-expressed response to drought stress in almond (Campalans et al. 2001). This gene is an 1-acylglycerol-3-phosphate O-acyltransferase that encode a protein that was suggested to have a possible role in drought stress tolerance by protecting cells from water deficit by changing membrane composition (Campalans et al. 2001).

6.8.2. *T. gallica-like*

6.8.2.1. Diversity within population

Genetic diversity assessed by microsatellites markers in our four Italian populations of *T. gallica*-like displayed greater values of genetic diversity with respect to that observed in *T. africana* populations. Even in *T. gallica*-like populations private alleles were found in all the populations studied, with the Simeto population which harbour the highest number of exclusive alleles. Although, some populations were affected by a small sample size, and additional sampling might recover shared alleles that persist at low frequencies, the presence of private alleles suggests some degree of independence between gene pools. Despite the Crati population was characterized by a small sample size (8 individuals), deviations from Hardy-Weinberg expectation were detected only in the Basento population at three loci. As previously described for *T. africana* populations, all violations from Hardy-Weinberg expectation were associated to positive values of inbreeding coefficient F_{IS}, that was consistent with the presence of homozygotes excess.

6.8.2.2. Differentiation among populations

The analyses of genetic differentiation by F_{ST} and R_{ST} pinpointed an overall lower differentiation between *T. gallica*-like populations with respect to the great genetic differentiation found in *T. africana* populations. D_{est} values displayed higher differentiation between populations, with values similar to those obtained in *T. africana* at neutral markers. On the contrary, lower D_{est} values were observed at EST-SSR markers. This observation was consistent with the previous literature, in fact, F_{ST} often misses to detect the differentiation, as it declines with increasing polymorphism (Jost 2008; Jost 2009). Moreover, the discrepancy between F_{ST} and D_{est} estimates was higher at neutral loci, which were resulted the most polymorphic markers.

As observed in *T. africana* the higher value of F_{ST} was consistent with the hypothesis of a more important involvement of genetic drift for determining the genetic variability of *T. gallica*-like populations, even if with a slighter effect with respect to *T. africana*. Anyway, it is noteworthy that R_{ST} at EST-SSR loci was higher than the value obtained at neutral loci. This results demonstrated the usefulness of our EST-SSRs for analysing the genetic variability in *Tamarix* (Martin et al. 2010).

6.8.2.3. Population genetic structure in *T. gallica*-like group

As previously reported for *T. africana* populations, the genetic divergence between *T. gallica*-like group populations was investigated computing pairwise F_{ST}, R_{ST}, and D_{est} matrix. Each of these statistics pointed out a general low genetic variability among *T. gallica*-like populations. These results were confirmed by the AMOVA, which showed a slight but significant differentiation in *T. gallica*-like populations, with a high intra-population diversity and a low genetic differentiation among populations. In fact, most of their variation was retained within populations with a value similar to that observed in the invasive *Tamarix* populations in North America (Friedman et al. 2008). Our results suggest the existence of gene flow among populations.

The analysis of genetic structure of Italian *T. gallica*-like populations, conducted using a Bayesian approach implemented in STRUCTURE, pointed out the presence of two gene pools. In fact, individuals were assigned to the two genetic clusters following the geographical distribution of their provenance. Individuals collected in the populations located in Sicily (Imera and Simeto) were assigned to the cluster number one; while individuals originated in

the Italian peninsula populations (Basento and Crati) were assigned to the cluster number two. In particular, in Crati and Basento individuals, the probability of assignment to the second cluster was lower with respect to the first one, confirming the hypothesis of high rate of gene flow between *T. gallica*-like populations.

Moreover, as it was not found any evidence of divergence between *T. gallica* and *T. canariensis* it means that differentiation among the populations was greater than between species. The PCoA confirmed the results obtained by the Bayesian approach, and indicating a high level of genetic affinity in the populations originated in Sicily (Imera and Simeto). Moreover, while individuals from Imera and Simeto overlapped and Crati individuals formed a intermediate group, Basento population resulted more differentiated. As previously observed in *T. africana,* even in *T. gallica*-like populations there was a clear separation among the gene pool deriving from the Italian peninsula and that originated in Sicily.

The Mantel test for isolation-by distance resulted not significant. Thus the genetic structure of Italian *T. gallica*-like populations was not affected by isolation-by-distance. This finding was in agreement with the lack of a strong genetic structuring pointed out by our results.

6.8.2.4. Detection of loci under selection

The analysis of outlier loci was performed in *T. gallica*-like populations, with the same method described for *T. africana*. Three EST-SSR loci (Th412, Th2876, and Th6387) showed F_{ST} values that displayed too high or too low levels of differentiation between populations. The locus Th2876 exhibited a negative F_{ST} value, that indicated a lower level of differentiation than expected, consistent with selection acting in flavour of the same allele in different populations. On the contrary, positive and higher F_{ST} values were observed in Th6387 and Th412. These values were consistent with a great divergence in alleles frequency due to selection, which flavours different allele in different populations (Siol et al. 2010).

Conclusions

At the best of our knowledge the present work is the first one regarding the characterization of genetic resources in Italian tamarisks. The taxonomy of the genus *Tamarix* is considered a difficult task as it is one of the most troublesome among angiosperms. Indeed, morphological traits usually employed for species identification, such as flowering features, are often misleading. For this reason, most of the *Tamarix* plants sampled during our germplasm collection in Southern and Central Italy remained unidentified.

An helpful tool for species identification could be provided by molecular markers, but only 10 genomic microsatellite markers have been developed in *T. chinensis* and *T. ramosissima.* Furthermore, primer pairs are often species-specific, so these markers are generally less transferable between different taxa. During this work we characterized a novel set of gene based microsatellite markers. These new EST-SSR markers are characterized by fewer null alleles and higher transferability across related species with respect to genomic microsatellites. Our new EST-SSRs could be useful in genetic characterization of the genus *Tamarix*, as additional tools for taxonomic clarification and for studying invasive populations where they are a threat.

An individual assignment method allowed the identification of genetically homogeneous clusters, that corresponded to species boundaries. Thus, in this work a posterior identification method was provided instead of the classical method relied on morphological traits. The partitioning into clusters showed a clear assignment of *T. africana* individuals, but it was not same for *T. gallica* and *T. canariensis* whose individuals formed a unique group (*T. gallica*-like), consistent with the hypothesis advanced by Gaskin and Schaal (2002) who suggested that these two species may be the same taxon or may be introgressive. Three sites resulted monospecific stands of *T. africana* (Alcantara, Baratz, and Marangone), four stands resulted mixed with *T. africana* and *T. gallica*-like group contemporarily present in the same site (Crati, Imera Basento, and Simeto), while no monospecific formation of *T. gallica*-like was found. The analyses of genetic structure of these populations pointed out the existence of a unique gene pool in Southern Italy for both *T. africana* and *T. gallica*-like, with populations characterized by low variability. On the other hand, *T. africana* populations from Central Italy and Sardinia resulted more differentiated, despite the Mantel test for isolation by distance was

not significant. It means that probably the genetic variability is affected by environmental factors, but it is not clear yet which features are involved.

Tamarix plants are characterized by tolerance to extreme environmental conditions, thus their survival capacity could represents a resource for the recovery of marginal areas. Moreover tamarisks thrive in habitats that are becoming fragile, thus they should be protected and evaluated.

References

Alfaro ME, Holder MT (2006) The posterior and the prior in Bayesian phylogenetics. Annual Review of Ecology, Evolution, and Systematics 37:19-42.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Resources 25: 3389-3402.

Archibald JK, Mort ME, Crawford DJ (2003) Bayesian inference of phylogeny: a non-technical primer. Taxon 52: 187-191.

Avise JC, Walker D (2000) Abandon all species concepts? A response. Conservation Genetics 1: 77-80.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Molecular Ecology 11: 155-165.

Balloux F, Goudet J (2002) Stastistical properties of population differentiation estimators under stepwise mutation in a finite Island model. Molecular Ecology 11: 771-783.

Banks MA, Eichert W, Olson JB (2003) Which genetic loci have greater population assignment power? Bioinformatics 19(11): 1436-1438.

Banks MA, Jacobson DP (2004) Which Genetic Markers and GSI Methods are More Appropriate for Defining Marine Distribution and Migration of Salmon? NPAFC Technical Report No. 5.

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.

Barcaccia G, Falcinelli M (2005) Genetica e Genomica. Volume II Miglioramento genetico. Liguri Editore. Naples. 459-506.

Baum BR (1978) The genus Tamarix. Israel Accad. Sci. Hum., Jerusalem.

Beaumont MA (2005) Adaptation and speciation: what can FST tell us? Trends in Ecology and Evolution 20 (8): 435-440.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. Nature Review Genetics 5: 251-261.

Belaj A, Munõz-Diez C, Baldoni L, Porceddu A, Barranco D, Satovic Z (2007) Genetic Diversity and Population Structure of Wild Olives from the North-western Mediterranean Assessed by SSR Markers. Annals of Botany 100: 449-458.

Bérubé Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K (2007) Characterization of EST-SSRs in loblolly pine and spruce Trees Genetic and Genomes 3: 251-259.

Bhargava A, Fuentes FF (2010) Mutational Dynamics of Microsatellites. Molecular Biotechnology 44:250-266.

Bhati J, Sonah H, Jhang T, Kumar Singh N, Sharma TR (2010) Comparative Analysis and EST Mining Reveals High Degree of Conservation among Five Brassicaceae Species. Comparative and Functional Genomics 2010: 1-13.

Campalans A, Pagès M, Messeguer R (2001) Identification of differentially expressed genes by the cDNA-AFLP technique during dehydration of almond (Prunus amygdalus). Tree Physiology 21: 633-643.

Conti F, Abbate G, Alessandrini A, Blasi C (2005) An annotated checklist of the Italian vascular flora. Palombi Editore. Rome.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153: 1989-2000.

Cheng LS (2007) Tamarix Lineaus. Flora of China 13: 59-65.

Cloutier S, Niu Z, Datla R, Duguid S (2009) Development and analysis of EST-SSRs for flax (Linum usitatissimum L.). Theoretical Applied Genetics 119(1): 53-63.

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Molecular Ecology Resources 10: 556-557.

De Martis B, Loi MC, Polo MB (1984) Il genere Tamarix L. (Tamaricaceae) in Sardegna. Webbia 37 (2): 211-237.

De Queiroz K (2007) Species concepts and species delimitation. Systematic Biology 56 (6): 879-886.

Dong Y, Yang C, Zhang D, Wang Y (2007) Cloning and sequence analysis of gene encoding plasma aquaporin of Tamarix albiflonum. Frontiers of Forestry in China 2 (2): 217-221.

Doyle JJ, Doyle JL (1990) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Focus 12: 13-15.

Drummond CS, Hamilton MB (2007) Hierarchical components of genetic variation at species boundary: population structure in two sympatric varieties of Lupinus microcarpus (Leguminosae). Molecular Ecology 16: 753-769.

Duminil J, Caron H, Scotti I, Cazal SO, Petit RJ (2006) Blind population genetics survey of tropical rainforest trees. Molecular Ecology 15: 3505-3513.

Ellegren H (2004) Microsatellites: simple sequence with complex evolution. Nature Reviews Genetics 5: 435-445.

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99: 125-132.

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.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. Nature Reviews Genetics 7: 745-758.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity 103: 285-298.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformayics Online 1: 47-50.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to mitochondrial DNA restriction data. Genetics 131: 479-491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Ferrazzini D, Monteleone I, Bellettini P (2007) Genetic variability and divergence among Italian populations of common ash (Fraxinus excelsior L.). Annals of Forest Sciences 64: 159-168.

Friedman JM, Roelle JE, Gaskin JF, Pepper AE, Manhart JR (2008) Latitudinal variation in cold hardiness in introduced Tamarix and native Populus. Evolutionary Applications ISSN: 598-607.

Gao C, Wang Y, Liu G, Yang C, Jiang J, Li H (2008) Expression profiling of salinity-alkali stress responses by large-scale expressed sequence tag analysis in Tamarix hispid. Plant Molecular Biology 66: 245-258.

Gao D, Wang Q, Wu Y, Xu H, Yu Q, Liu J (2008) Microsatellite DNA loci from the typical halophyte *Thellungiella salsuginea* (Brassicaceae). Conservation Genetics 9 (4): 953-955.

Garrick RC, Caccone A, Sunnucks P (2010) Inference of Population History by Coupling Exploratory and Model-Driven Phylogeographic Analyses. International Journal of Molecular Sciences 2010 11 (4): 1190-1227.

Gaskin JF (2003) Molecular systematics and the control of invasive plants: A case study of Tamarix (Tamaricaceae) Annals Missouri Botanical Garden 90: 109-118.

Gaskin JF, Kazmer DJ (2006) Comparison of ornamental and wild saltcedar (Tamarix spp.) along eastern Montana, USA riverways using chloroplast and nuclear DNA sequence markers. Wetlands 29 (4): 939-950.

Gaskin JF, Kazmer DJ (2009) Introgression between invasive saltcedars (Tamarix chinensis and T. ramosissima) in the USA. Biological Invasions 11: 1121-1130.

Gaskin JF, Pepper AE, Manhart JR (2006) Isolation and characterization of 10 polymorphic microsatellites in saltcedars (Tamarix chinensis and Tamarix ramosissima). Molecular Ecology Notes 6: 1147-1149.

Gaskin JF, Schaal BA (2002) Hybrid Tamarix widespread in U.S. invasion and undetected in native Asian range. PNAS 99: 11256-11259.

Gaskin JF, Schaal BA (2003) Molecular phylogenetic investigation of U.S. invasive Tamarix. Systematic Botany 28: 86-95.

Goldstein PZ, DeSalle R, Amato G, Vogler AP (2000) Conservation genetics at the species boundaries. Conservation Biology 14: 120-131.

Gugerli F, Brodbeck S, Holderegger R (2008) Utility of multilocus genotypes for taxon assignment in stands of closely related European white oaks from Switzerland. Annals of Botany 102 (5): 855-863.

Jost L(2009) D vs. G_{ST}: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009) (2009) Molecular Ecology 18: 2088-2091.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Molecular Ecology 17: 4015-4026.

Kalia RK, Rai MK, Kaila S, Singh R, Dhawan AK (2011) Microsatellites markers: an overview of the recent progress in plants. Euphytica 177: 309-334.

Kumar RP, Senthilkumar R, Singh V, Mishra RK (2010) Repeat performance: how do genome packaging and regulation depend on simple sequence repeats? Biological Essays 32: 165-174.

Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48 (6): 985-998.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163: 1467-1482.

Huelsenbeck JP, Larget B, Miller RE, Ronquist F (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. Systematic Biology 51 (5): 673-688.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. Science 294: 2310-2314.

Laurentin H (2009) Data analysis for molecular characterization of plant genetic resources. Genetic Resources and Crop Evolution 56: 277-292.

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes. Genome biology 7 (2): R14.

Lim S, Notley-McRobb L, Lim M, Carter DA (2004) A comparison of the nature and abundance of microsatellites in 14 fungal genomes. Fungal Genetics and Biology 41: 1025-1036.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology 11: 2453-2465.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2004) Microsatellites within genes: structure, function, and evolution Molecular Biology Evolution 21 (6): 991-1007.

Lugon-Moulin N, Brünner H, Wyttenbach A, Hausser j, Goudet J (1999) Hierarchical analyses of genetic differentiation in a hybrid zone of Sorex araneus (Insectivora: Soricidae). Molecular Ecology 8: 419-431.

Martin MA, Mattioni C, Cherubini M, Taurchini D, Villani F (2010) Genetic diversity in European chestnut populations by means of genomic and genic microsatellite markers. Tree Genetics & Genomes 6: 735-744.

Médail F, Quézel P (1999) Biodiversity hotspots in the mediterranean basin: setting global conservation priorities. Conservation Biology 13: 1510-1513.

Millar MA, Byrne M, Nuberg I, Sedgle M (2008) A rapid PCR-based diagnostic test for the identification of subspecies of Acacia saligna. Tree Genetics & Genomes 4: 625-635

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Review Genetics: 194-200.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy Science USA 70 (12): 3321-3323.

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics 41: 225-233.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

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 30: 450-458.

Ochieng JW, Shepherd M, Baverstock PR, Nikles DG, Lee,DJ, Henry RJ (2010) Two sympatric spotted gum species are molecularly homogeneous. Conservation Genetics 11 (1): 45-56.

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Carneiro Vieira ML (2006) Origin, evolution and genome distribution of microsatellites. Genetics and Molecular Biology 29 (2): 294-307.

Pascal E, Steffen F, Martin S (2008) Development of two microsatellite multiplex PCR systems for high throughput genotyping in Populus euphratica. Journal of Forestry Research 20 (3): 195-198.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Molecular Ecology 4: 347-354.

Paetkau D, Slade R, Burden M, Estoup A (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Molecular Ecology 13: 55-65.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.

Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST Databases as a Source for Molecular Markers: Lessons from Helianthus. Journal of Heredity 97 (4): 381-388.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection. Journal of Heredity 95 (6): 536-539.

Rannala B, Mountain JL (1997) Detecting immigrants by using multilocus genotypes. Proceedings of the National Academy Science USA 94:9197-9201.

Raymond M, Rousset F (1995). Genepop (version-1.2) - population-genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248-249.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press. Totowa: 365-386.

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

Russell PJ (2002) iGenetica. EdiSES s.r.l. Naples: 661-709.

Ryman N, Leimar O (2009) GST is still a useful measure of genetic differentiation - a comment on Jost's D. Molecular Ecology 18: 2084-2087.

Scalfi M, Piotti A, Rossi M, Piovani P (2009) Genetic variability of Italian southern Scots pine (Pinus sylvestris L.) populations: the rear edge of the range. European journal of Forest Research 128 (4): 377-386.

Sekino M, Hara M (2007) Individual assignment tests proved genetic boundaries in a species complex of Pacific abalone (genus Haliotis). Conservation Genetics 8: 823-841.

Sexton JP, McKay JK, Sala A (2002) Plasticity and genetic diversity may allow saltcedar to invade cold climates in north America. Ecological Applications 12(6): 1652-1660.

Shaporova N (2008) Plant simple sequence repeats: distribution, variation, and effects on gene expression. Genome 51: 79-90.

Sharma P, Grover A, Kahl G (2007) Mining microsatellites in eukaryotic genomes. TRENDS in Biotechnology 25 (11): 490-498.

Siol M, Wright SI, Barrett SC (2010) The population genomics of plant adaptation. New Phytologist 188 (2): 313-332.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-462.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of multi-allele and multi-locus genetic microstructure. Heredity 82: 561-573.

Stàgel A, Portis E, Toppino L, Rotino GL, Lanteri S (2008) Gene-based microsatellite development for mapping and phylogeny studies in eggplant. BMC Genomics 9: 357.

Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. Biological Journal of the Linnean Society 68 (1-2): 41-55.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Resources 22: 4673-4680.

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cellular & Molecular Biology Letters 7: 537-546.

VenturellaG, Baum B, Mandracchia G (2007) The genus Tamarix (Tamaricaceae) in Sicily: first contribution. Flora Mediterranea 17: 25-46.

Wang YC, Gao C, Liang Y, Wang C, Yang CP, Liu GF (2010) A novel bZIP gene from Tamarix hispida mediates physiological responses to salt stress in tobacco plants. Journal of Plant Physiology 167: 222-230.

Wang YC, Qu GZ, Li HY, Wu JH, Wang C, Liu GF, Yang CP (2010) Enhanced salt tolerance of transgenic poplar plants ex pressing a manganese superoxide dismutase from Tamarix androssowii. Molecular Biology Reports 37:1119-1124.

Wang YC, Yang CP, Liu GF, Jiang J, Wu JH (2006) Generation and analysis of expressed sequence tags from a cDNA library of Tamarix androssowii. Plant Science 170: 28-36.

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology 10: 249-256.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38 6: 1358-1370.

Weising K, Nybom H, Wolff K, Kahl G (2005) DNA Finger printing in plants- Principles Methods, and Applications. CRC Press. Taylor and Francis Group.

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M, Powell W (2005). Comparative analysis of population genetic structure in Athyrium distentifolium (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. Molecular Ecology 14: 1681-1695.

Wright S (1921) Systems of mating I-V. Genetics 6: 111-178.

Wright S (1951) The genetical structure of populations. Annals of Eugenics 15: 323-354.

Wu Y, Wang J, Liu J (2008) Development and characterization of microsatellite markers in Populus euphratica (Populaceae). Molecular Ecology Resources 8 (5): 1142-1144.

Zhang D, Yin L, Pan B (2002) Biological and ecological characteristics of Tamarix L. and its effect on the ecological environment. Science in China 45: 18-22.

Zhao X, Zhan LP, Zou XZ (2011) Improvement of cold tolerance of the half-high bush Northland blueberry by transformation with the LEA gene from Tamarix androssowii. Plant Growth Regulation 63 (1): 13-22.

Acknowledgements

This work was supported by the Italian-Israeli Cooperation on Environmental Research and Development Project. I thank Riccardo Valentini, as project coordinator, Giovanbattista De Dato, Renée Abou Jaoudé, Grazia Abbruzzese for plant collection and morphological discrimination of the species studied; Claudia Mattioni, Matilde Tamantini and Isacco Beritognolo for laboratory support, Victoria Dawailibi and Critina Monteverdi for cooperation in the project work.

Ringraziamenti

Ringrazio la mia famiglia, babbo Pino e mamma Maria Letizia, per il supporto e il conforto che mi hanno offerto in questi tre anni, soprattutto nel momento in cui ho avuto problemi di salute. Ringrazio in mio fidanzato e compagno di vita Massimo per la pazienza che ha avuto ogni volta che lo stress mi faceva agitare inutilmente. Lui puntualmente mi riportava alla ragione, o come dice lui, faceva si che lo scienziato prendesse il sopravvento sulla donna. Mia nonna Maria Antonietta, tutti i miei zii e i miei cugini che mi hanno sempre sostenuto e incoraggiato, in particolare mio cugino Simone, sua moglie Alessandra e i piccoli pucciarotti Raoul e Christian.

Sono grata a tutti i miei colleghi dottorandi per le chiacchierate davanti alle macchinette del caffè, gli aperitivi-cena al Blitz e le varie pizzate consumate tutti insieme. Con particolare affetto vorrei menzionare le mie amiche e compagne di avventura Renée e Grazia. I nostri destini si sono incrociati facendo cominciare e finire insieme il percorso del dottorato di ricerca. Non posso tuttavia dimenticare quei colleghi che sono "emigrati" verso nuovi istituti di ricerca. Vorrei quindi ringraziare Daniela, Erika ed Olga per la loro amicizia e per la loro splendida compagnia.

Ringrazio tutti i miei colleghi della stanza 120 A. Federica per il suo supporto in laboratorio e per l'aiuto fornito (non dimenticherò mai la fatica dello sbrinamento del congelatore). Francesco per la sua memoria sulle battute dei film e per l'aiuto che mi ha dato sulle analisi con R. Giannetto per la sua disponibilità e per pazienza che dimostra sempre nei confronti dei suoi colleghi meno esperti.

Vorrei ringraziare tutto il personale (tecnici, ricercatori e professori) del CNR-IBAF Istituto di Porano. In particolare Claudia e Paola per avermi aiutato nelle analisi e per i preziosi consigli sulla elaborazione dei dati. Spero di poter mantenere la promessa fatta a Paola di poter cogliere l'occasione di passare il tempo in sua compagnia e assistere di nuovo a un congresso insieme.

Ringrazio tutti i miei amici, quelli che nel corso di questi tre anni si sono dimostrati tali. In particolare Margherita e Ivano, che più che essere considerati amici sono per me una sorella e un fratello. I miei ex colleghi studenti forestali Silvia, Daniela, Daniele, Pier Luigi, Ludovico, Mauro, Lando, Marco e Salvatore per le innumerevoli cene, le scampagnate e le rimpatriate che abbiamo fatto insieme.

In fine vorrei ringraziare in miei supervisori: il dott. Maurizio Sabatti e la dott.ssa Elena Kuzminsky per la fiducia che hanno sempre riposto in me e per i preziosi consigli che mi hanno sempre dato nei momenti di difficoltà. Sono grata a Isabella, Isacco e Muriel per tutto ciò che mi hanno trasmesso umanamente e professionalmente. Le mie competenze le ho acquisite grazie ai loro insegnamenti.

Grazia a tutti