UNIVERSITÁ DEGLI STUDI DI VERONA

FACOLTÁ DI SCIENZE MM.FF.NN.

DOTTORATO DI RICERCA IN BIOTECNOLOGIE MOLECOLARI INDUSTRIALI E AMBIENTALI

XVIII CICLO

"Genetic variability in Italian populations of *Cupressus sempervirens* L. assessed by SSR and RAPD markers"

Coordinatore: Ch.mo Prof. Hugo Luis Monaco

Relatore: Ch.ma Prof. Antonella Furini

Tutore: Dott. Nicola La Porta

Dottoranda: Maria Chiara Valgimigli

ANNO ACCADEMICO 2004-2005

To Matteo

&

To my Family

INDEX

INTRODUCTION

1.1 Biodiversity and genetic variation

Biodiversity can be regarded as the richness and variety of genetic information present in an ecosystem, in a species or in a population. Biodiversity at specific level is described as genetic variability existing among individuals, and genetic variability of a species can be defined as the number of different alleles per locus (allelic variability) or as the number of combinations of different alleles per genome (genotypic variability).

Allelic diversity mainly arises from mutational events on DNA sequence. In eukaryotic organisms genotypic variability can be increased by all the processes associated to reproduction and generally termed as recombination events: crossingover, independent segregation of non-homologous chromosomes, and, above all, syngamy. The major benefit of recombination is the generation of diversity, rare genotypes, and well-adapted lineages in the face of environmental changing.

Species with high levels of biodiversity include individuals with different genetic information. Biodiversity is narrowly correlated to the adaptive potential of populations: the more individuals are different the more probability at least some of them have to face successfully changing environmental conditions.

On the contrary, species with low levels of biodiversity include very similar individuals that may probably react to environmental request in a very similar way: new parasite appearance or climate changes could have disastrous consequences if all the individuals of a population are homogeneous and lack genetic tools necessary to confer resistance or tolerance against adversity.

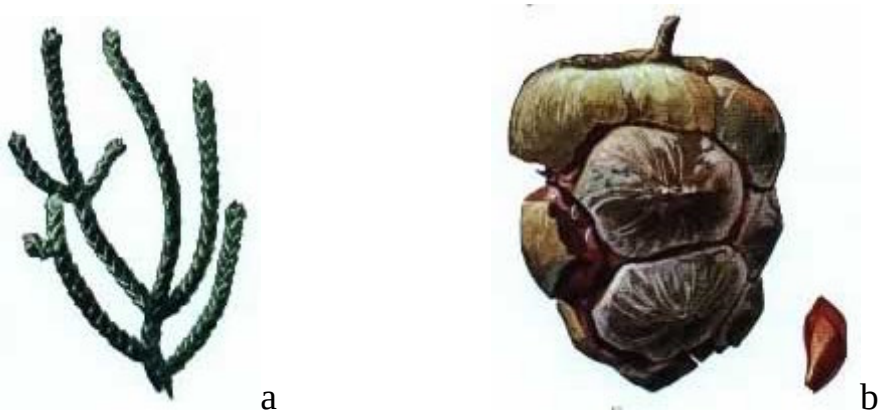
Quantifying genetic variability and evaluating its distribution among and within the populations of a species permit to infer the best ways to preserve diversity of populations, to find propagation material of good quality, to find mother plants for seeds collections.

Pathogen diseases, commercial management and breeding practice often lead to reductions in genetic diversity (Herlan, 1975), and loss of genetic diversity has long been recognized as a potentially serious problem in human managed forest tree species (Libby, 1969; Jasso, 1970; Richardson, 1970; Ledig, 1992; Rogers, 1996). In fact, conservation of genetic diversity may be one of the most important issues influencing forestry practices (Namkoong, 1992) and survival of a species (Boyle, 1992).

1.2 *Cupressus sempervirens* L.

1.2.1 Botanical description and productive cycle

C. sempervirens is a conifer tree about 20-30 m height with a straight trunk. The bark is thin, smooth and gray for quite a long time, later it becomes gray-brown and longitudinally furrowed. Shoots grow radiating in all directions; they are about 1 mm in diameter, round or quadrangular. Leaves are scale-like, decussate, small, ovate, and obtuse with a dark green color and a dorsal gland in the shape of longitudinal furrow (Figure 1 a). Flowers appear early in spring (Figure 2 a). Cones are 2-3 cm long and pendulous, they have short stalks and look glossy, brown to gray; the shape is from globose to elliptic (Figure 1 b). Seeds are brown, flattened, minute, without resin blisters and narrowly winged (Figure 1 b).



a b

Figure 1: a) Leaves, b) Cones and seeds of the cypress

The Cypress is a diploid monoecious outcrossing wind-pollinated species. Male and female cones (strobili) are separated along the branch. First flowering appears after 3-4 years and reproductive cycle develops over 3 years. The first year flower buds initiate from late-spring to early summer. Male cones are differentiated in spring while female cones in autumn and develop faster. The second year pollination drops appear in January - February on erect and exposed ovules, indicating pollen receptivity of female cones. Pollen germinates immediately on the pollination drop. The female ovule matures until July when fertilization takes place. Cone seed and embryo start developing and stop in late summer. The third year maturation starts again and it is complete in autumn.

Pollen dispersal, fertilization and embryo development are stages where disturbances caused by climatic, biotic or genetic factors play an important role in seed quality. Flowering time and male-female synchronism, for example, may affect the percentage of empty seeds in natural populations and seed orchards. The species can live up to 1000 years of age (Figure 2 b).

a

b

Figure 2: a) cypress flowers b) an ancient cypress (800 years old), Villaverucchio (RN).

The Cypress has two principal varieties according to the brunch type, which is either erect or horizontal (Figure 3). Var. *horizontalis* has spreading branches, with a broad conical crown. This is the most common shape found in natural distribution areas. Var. *pyramidalis* (or *stricata*, or *fastigiata*) has erect branches parallel to the trunk or very short horizontal branches forming a pyramid or columnar shape. This was the variety described by Linnaeus. It is the popular ornamental variety and the most widely planted of all cypress. The two varieties are interfertile and can give progenies that, apart from the parental forms, present different intermediate types of crown structure.

Figure 3: C. sempervirens *var. pyramidalis (left) and var. horizontalis (right).*

1.2.2 Origin and present diffusion area

The genus *Cupressus* is native to warm temperature climates in the Northern Hemisphere. It can be found around the Mediterranean, in North America and in Asia. Mediterranean species appear to have a common ancestor with a single species whose range extended from Morocco to the Caspian Sea. The gradual climate warming and increased aridity separated this species into the present-day ones: *C. sempervirens*, *C. dupreziana* and *C. atlantica*.

C. sempervirens L., (Cupressaceae) is native to the Iran, as well as Syria, Turkey, Cyprus and several Greek Islands (Crete, Samos, Rhodes ecc..), and it was introduced in most countries around Mediterranean. Probably during the Tertiary it occupied larger areas (Axelrod, 1958) reduced during millennia mainly by strong human pressure (Boscherini, 1994; Vendramin, 1995), intensive and unregulated forest utilisation, burning, grazing and cypress canker disease (Kayacik, 1979; Sumer, 1987; Graniti, 1998) leaving only small areas of forest.

At present its natural geographic distribution is characterised by disjoint and often relic populations growing in Iran, Syria, Jordan, Lebanon, Libya, the Aegean Islands, Crete, Turkey, Cyprus (Zohary, 1973). It grows also in Italy, France, Spain, Portugal and former Yugoslavia (Ducrey, 1999) where it was introduced presumably during the Roman era or even before, since the Phoenicians and Etruscans started to sail along the Mediterranean (Giannelli, 2002; Santini, 2000). Since historical times, the cypress has been extensively cultivated far beyond its natural geographic range, in earlier times through its association with religious rites and later for aesthetic reasons. This tree species can actually live under various Mediterranean climates, from sea level up to 2000 m or more on a variety of soil types and in a variety of plant association (Zohary, 1973). Such a spread of the cypress is still in act, not only in Mediterranean countries, but in every climatically similar area too, where the cypress is able to fit to the local environmental conditions (Santini, 2000).

Nowadays the common cypress has an important role in the characterization of Mediterranean landscape mainly for its aesthetic function. The cypress does not have any specific soil requirements, even if, as most forest species, is best adapted to rich, deep, moist, well aerated soil with neutral pH. However, its rustic nature makes possible for it to grow on poor, dry soils. It has actually demonstrated to be an excellent pioneer species for reforestation of rocky, argillaceous, limestone, barren and superficial lands. It prevents the hydrogeological erosion and constitutes a source of yield for the very good quality of its wood. It is also used like windbreak plant.

The cypress acclimatized to the Italian environment very well, as to be considered a naturalized species (Ducrey, 1999) although there are no natural forests of *C. sempervirens* in Italy (Raddi, personal communication). Cypress groves are present in coastal hills from Liguria to Calabria and in Sicily. In the

central part of Italy, especially in Tuscany near Florence, Siena and Pisa, cypress woods are more present and productive. In the north of Italy, cypress stands and groves can be mainly found around lakes where climatic conditions are favourable (Xenopoulos, 1990).

1.2.3 Cypress canker

In the last fifty years the cypress has been attacked by a parasitic fungus, *Seiridium cardinale* (canker of cypress), which is seriously threatening the survival of this plant in Italy and in other Mediterranean countries (Graniti, 1998). Several cypress improvement programs for resistance were set up with the attempt to cultivate resistant clones throughout wide-reaching territories and areas with highly diverse pedoclimatic conditions. Several resistant clones were actually produced. The strong effect of environment and of environment by genotype interaction on cypress clones has been noted. The cypress is a very plastic species: clones growing in completely different habitats take very different shapes in accordance with variations in environmental conditions, ecological factors and soil characteristics (Santini, 1994 b).

Similar conclusions are also being reached in works involving stability in the resistance to cypress canker disease (Santini, 1994 a). Clones to use should perhaps be tested locally before spread on a big area, instead of aiming the entire research effort at finding a universal clone adaptable to all environments (Santini, 1994 b).

The low temperatures that the cypress has often to stand in the Italian northern regions act indirectly to increase the strength of penetration of *S. cardinale* spores by means of microlesions created by frost. In this context plants resistant to *S. cardinale* and adapted to cold northern regions guarantee a better protection against pathogen.

Up to now, only little genetic information about some *C. sempervirens* stands is available: only few populations have been studied with allozymes. In fact there is little genetic information about existing stands where the cypress has been more recently introduced and we have no information about cypress population structure in Italy.

Preliminary studies to increase knowledge in our country can be helpful and interesting for several reasons. Due to the importance of *C. sempervirens* L. for reforestation, ornamental and windbreak plantings all around the Mediterranean region and Italy, there is considerable interest in the geographic patterns of genetic variation (Korol, 1997). Quantifying genetic variability and evaluating its distribution among and within populations permit to infer the best ways to preserve diversity of populations, to find propagation material of good quality, to find plants for seeds. Moreover, the knowledge of the population dynamics of a species can be helpful in predicting the effect of the biotic and abiotic stresses it may encounter. The possibility that future climate changes may render the North of Italy more favourable to cypress cultivation could aid the increasing

spread of this arboreal plant on the territory and suggested to us that we might study this conifer in the Italian northern regions. But conservation of gene resources is required to ensure adaptation and survival in environmental changes. And more information about cypress gene resources is necessary. In fact, the conservation of genetic diversity may be one of the most important issues influencing future forestry practices (Namkoong, 1992).

In the case of the cypress we have very poor knowledge about population structure, genetic diversity, frequency and entity of genetic flow between populations, compared, for example, with other temperate trees such as oaks (Petit, 2002) beeches (Demesure, 1996) or chestnuts (Fineschi, 2000).

1.3 Genetic markers

Biodiversity can be detected with genetic markers. Any physical or molecular characteristic that differs among individuals and is easily detectable in the field or laboratory is a potential genetic marker. Markers can be morphological characteristics, isozymes, actual genes or DNA segments that have no known function but whose inheritance pattern can be followed. In order for a marker to be useful, it must be polymorphic, i.e. alternative forms must exist so that they are detectable among individuals in family studies (Cruzan, 1998).

An ideal genetic marker should be stable, polymorphic, easy to detect or to observe, heritable in a simple manner, codominant, reproducible within and among different laboratories and detectable with methodology applicable to many distinct species.

The main classes of markers are: morphological (the characteristics that can be easily observed) cytogenetic, biochemical and molecular (the markers for which laboratory techniques are required to observe the polymorphisms). Traditional approaches to taxonomy and evolution have typically used morphological traits, sexual crossability, and cytogenetic considerations. These approaches are now complemented by biochemical and molecular methods to more directly trace genetic diversity. Biochemical techniques used to measure differences within species include protein sequencing, immunological responses to analogous proteins between different species, and isozyme analysis.

The polymerase chain reaction (PCR) method of DNA amplification has provided geneticists and ecologists with new ways of generating molecular markers (Avise, 1994). Each marker system is characterized by its own set of advantages and disadvantages, so the choice of a marker system is determined, to a certain extent, by its desired application.

In this section we will introduce only markers used in this study (RAPD, SSR) or previously used for studies on *C. sempervirens* genetic variability (morphological markers, isozymes).

1.3.1 Morphological markers

Morphological traits may be (1) qualitative - falling into discrete classes, e.g. purple or white as in flower colour; or (2) quantitative - showing a continuous range of phenotypes. Single genes most often control qualitative traits, whereas several genes control quantitative traits. For morphological traits, levels of variability can be estimated, and their response to selection and their genetic background can be determined.

Lower heritability of traits and evident presence of interactions between genotype and the environment require the application of special experimental designs to be used to distinguish genotypic from phenotypic variation (Greenwood, 1992). Morphological analyses require extensive observations of

mature plants, are labour-intensive, and in many cases lack definition and objectivity (Kearsey, 1998).

1.3.2 Isozymes

Gel electrophoresis of proteins is a standard research tool in a number of biological disciplines. One form, called isozyme analysis, has become prominent in systematic and evolutionary biology as well as in agronomy and forestry. Isozymes were first widely employed as molecular markers since the 1950's (Hamrick, 1989). They are multiple molecular forms of an enzyme that share a common substrate but differ in electrophoretic mobility. They provide a series of readily scorable, single-gene markers (Hamrick, 1989).

For genetic studies, allozyme is the more correct term, referring specifically to enzyme forms that are the products of different alleles and not the product of different genes with similar enzymatic activity. Allozymes have been extensively used as genetic markers for: identifying cultivars, species, and hybrids, confirming hybridity, measuring genetic diversity of plant populations, verifying taxonomic relationships, constructing linkage maps, determining distance within and between species (Jasieniuk, 2001; Loveless, 1984; Brown 1979).

The utility of allozymes as genetic markers is generally attributed to their frequent polymorphism, codominance, single gene-Mendelian inheritance, and their ubiquity in plant tissue. The assay is relatively rapid, inexpensive, and relatively simple (Brown 1979; Hamrick 1989).

The analysis begins with electrophoretic separation of proteins on a gel matrix. The gel is then soaked in a reaction solution specific for a given enzyme. Enzymatic activity identifies enzyme location by a local change in colour. The genetic analyses depend on differences in gel mobility among different forms of the enzyme (Bergmann, 1991).

The advantages of this technique of analysis are: (1) individuals can be scored for several allozymes at the same time; and (2) allozyme markers are co-dominant, i.e. both alleles in a diploid organism are identifiable and so heterozygotes can be distinguished from homozygotes. This is a prerequisite for estimation of allele frequencies in population genetic studies (Bergmann, 1991; Hamrick, 1989).

Though the typical allozyme locus encodes codominant alleles with no genotype x environment interactions, null, dominant, and epistatic allozymes have been identified (Bergmann, 1991).

However, there are limitations to allozyme analyses. A new allele will only be detected as a polymorphism if a nucleotide substitution has led to an aminoacid substitution, which thus affects the mobility of the enzyme in question. But because of the redundancy of the genetic code and because not every aminoacid replacement leads to detectable electrophoretic mobility

difference, only 30% of all nucleotide substitutions leads to polymorphic fragment patterns.

Thus allozymes underestimate genetic variability. In addition, allozymes study only those parts of the DNA that encode stainable enzymes, which is not necessarily a random sample of the genome (Bergmann, 1991).

Allozyme analysis cannot be applied to distinguish between closely related genotypes because of limited polymorphisms. Plant tissue intended for allozyme studies has to be processed shortly after harvest because proteins usually are quite unstable.

In contrast, DNA-based methods allow for storage at ultralow temperatures thus providing a longer time between harvest and processing. Also, environment, tissue type, and developmental stage may affect allozymes. The spectrum of allozymes present may vary under these conditions and may show changes in intensity, and the appearance or disappearance of bands (Avise, 1994). Allozyme markers are limited in number.

1.3.3 DNA-based markers

The development of DNA markers has facilitated detection, monitoring, and manipulation of genetic variation (reviewed in Parker, 1998). The advantages that make DNA markers more attractive than allozymes and morphological traits are:

1) the genotype of the organism is examined directly;
2) environmental and developmental influences on the phenotype do not interfere with the result;
3) an almost unlimited number of detectable polymorphisms exist;
4) a variety of different techniques has been developed to generate suitable markers for several specific application (Parker, 1998).

Polymorphic DNA is thought to generate ideal genetic markers because:

1) nucleotide sequence variation is presumably selectively neutral, at least for non-coding sequences;
2) certain complications, which reduce heritability with protein analysis (mutation and modifications) may be minimized; and
3) plant cells have three distinct genomes (nuclear, chloroplast, and mitochondria) that are known to evolve separately and in different ways.

Most DNA marker applications focus on nuclear DNA because it generally evolves more rapidly than plant organelle DNA and thus is expected to be more polymorphic (Newton, 1999).

Some of these new PCR-based marker systems are Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeat (SSR), Restriction Fragment Length Polymorphism PCR (RFLP-PCR).

Random Amplified Polymorphic DNA

Random Amplified Polymorphic DNA (RAPD), also termed Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) markers are well suited for efficient, non-radioactive DNA fingerprinting of organisms, especially for determining intra- and interspecific genetic relationships.

The RAPD technology, developed concurrently by Williams (1990) and Welsch (1990), is based on the amplification of genomic DNA with single 10-mer primers of arbitrary nucleotide sequence to generate a unique pattern of fragments in low stringency PCR reactions.

Since the primers are so short, they often anneal to the template DNA at multiple sites. Some primers anneal in the proper orientation and at suitable distance from each other (i.e. max 3 kb, the approximate maximum size of a PCR fragment) to support chain amplification of the unknown sequence between them (Figure 4). Among the set of fragments there are some that can be amplified from certain genomic DNA samples but not from others.

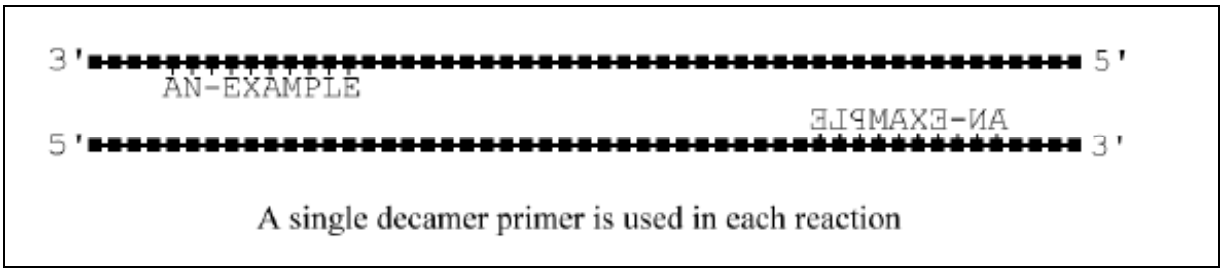


Figure 4: Source: Gandolfi A. "Genetic variability and life history traits of three freshwater ostracod species". PhD thesis.

The identity and sequence of a particular amplification product is unknown, but its presence or absence in different organisms or individuals can be used as a highly informative character for the evaluation of genetic diversity and relatedness. Conditions are normally chosen whereby the number of fragments generated is between 1 and 20.

The existence of a particular product may be a good Mendelian character, which is typically, but not always, dominant (Williams, 1990; Wu, 1999). Consequently, RAPDs will generally segregate in simple 1:1 and 3:1 ratios (Grattapaglia, 1994).

The major advantage of the RAPD assay is that no previous knowledge of any gene in the target organism is required (Williams, 1990; Parker, 1998).

RAPDs sidestep the problems associated with primer design by using large sets of short random oligonucleotides, with an arbitrary not palindromic sequence and high G-C content, thus some useful products are likely to be amplified in virtually any species (Bucci, 1993; Williams, 1990). The ability of RAPDs to produce multiple bands using a single primer means that a relatively small number of primers can be used to generate a very large number of

fragments. Besides these fragments are usually generated from different regions of the genome and hence multiple loci may be examined very quickly. Another great benefit of this technique, common to many PCR-based techniques, is that only small amounts of DNA are required for analysis (Williams, 1990).

The technical simplicity of RAPDs is also an advantage to geneticists and ecologists. RAPDs offer the advantage of being simpler to use, less expensive and less time-consuming than RFLPs (Powell, 1996) or AFLPs (Lanham, 1999).

This technique also has its handicaps. One of these is that it is difficult to distinguish many of the polymorphisms generated from PCR artefacts. As some of the results from RAPD analysis come from negative evidence (lack of a band), precise control of the amplification conditions is critical (Penner, 1993). In the experience of several workers, template DNA reagent concentrations, primer sequence and length, and experimental conditions must all be strictly controlled during DNA amplification, because they may strongly affect banding patterns (Williams, 1992). However, the poor reproducibility in early RAPD analyses can now be avoided through improved laboratory techniques (Skrok, 1995) and with a rigorous set up of reactions, checking pattern reproducibility with different amplifications, and using more stringent PCR conditions (i.e. high annealing temperature).

Another flaw is the dominant nature of RAPDs, which makes distinguishing between homozygotes and heterozygotes theoretically impossible (Parker, 1998). Though, pairs of RAPD markers tightly linked in repulsion phase would allow for the identification of heterozygotes (Grattapaglia, 1996). RAPD analyses generally detect the occurrence of a single allele, while other DNA techniques can distinguish among many alleles at specific loci. Thus RAPD markers may be inferior to codominant markers, though the frequency of alleles coding for fragment occurrence or absence may be estimated from RAPD data via relevant statistical analysis.

Another potential problem is that true homology between bands of the same length obtained with the same primer should be verified. Numerous studies have demonstrated that (usually) comigrating fragments are identical by descent, at least at the intraspecific level (Riesberg, 1996; Wu, 1999). Consequently this method is especially suited to situations where little or no molecular genetics research has been conducted previously.

Microsatellites

Microsatellites or Simple Sequence Repeats (SSRs) are molecular markers widely used for DNA fingerprinting, paternity testing, linkage map construction and population genetic studies. Based on tandem repeats (from 5 to 100 units) of short (1-6 bp) DNA sequences, these markers are highly polymorphic due to variation in the number of repeat units (Tautz, 1986). The repeat length at

specific SSR loci is easily assayed by PCR using primers specific to conserved regions flanking the repeat array. They are widely dispersed throughout eukaryotic genomes (Tautz, 1986).

SSR can be classified as perfect (without interruption in the repeated sequence (TG)n or (AAT)n), imperfect (with one or more interruption in the repeated sequence), or composed by two or more perfect or imperfect repeated units ((GT)n(TC)i).

They have been well characterized in mammalian genomes and in a number of plant genomes (Akkaya, 1992; Morgante, 1993; Zhao, 1993). The most frequent dinucleotide repeats in plants are (AT)n, (AG)n and (AC)n (Morgante; 1993). The hypervariability and co-dominance of SSRs, their dispersion throughout genomes and suitability for automation are the principal reasons for their wide utility (Powell, 1996; Jarne, 1996; Gupta, 1996). Microsatellites evolve faster than the surrounding DNA and are thus very polymorphic. Besides, they are potentially multiallelic having the potential for more information per marker. SSRs are inherited in a codominant Mendelian fashion and are generally stable (Rafalski, 1993).

PCR amplification protocols used for microsatellites are generally standard and can be carried out in a total volume of 10 µl. Depending on which of the possible strategies for electrophoresis and subsequent scoring of alleles is used, PCR amplifications employs either unlabelled primer pairs or primer pairs with one of the primers being fluorolabelled.

Several PCR products can be pooled (i.e. multiple loading) or several loci can be co-amplified during PCR (i.e. multiplexing). Multiplexing allows the rapid genotyping of large sample sizes across several loci. However, considerable time has to be spent in designing reliable multiplex systems.

A major advantage of automated systems is the availability of dyes that fluoresce at different wavelengths (e.g. FAM, HEX, TAMRA) enabling highly efficient, simultaneous electrophoresis of several loci with overlapping allele size ranges. Loci that have non-overlapping allele size range can be separated simultaneously in fluorescent approach, too. The accurate sizing of alleles is achieved running an internal size marker unique fluorescent labelled. Automated electrophoresis system, usually have high resolution.

Mono and dinucleotide repeat unit microsatellites present stutter bands (or peaks) due to replication slippage during amplification process. These slippage products are present as less intense bands of usually one to five repeat units smaller and (occasionally greater) than the actual allele. The slippage bands become relatively less intense the more they deviate in size from the native allele. A major limitation of SSR is the time and cost required to isolate and characterize each locus when pre-existing DNA sequence is not available.

Typically, this process requires the construction and screening of a genomic library of size-selected DNA fragments with SSR-specific probes, followed by DNA sequencing of isolated positive clones, PCR primer synthesis and testing

(Hayden, 2001; Edwards, 1996). Only after this procedure, informativeness of SSR loci can be determined. Significant cost and time-savings could be made if library screening is eliminated and sequence information for more than one SSR locus is obtained from each plasmid clone.

One potentially very large advantage of microsatellites in conservation genetics, especially for future studies, is the fact that primers developed for a particular species have now been increasingly shown to be applicable across a wide range of related taxa in animals. With the increasing numbers of microsatellites being produced in a large range of animal and plant species it is conceivable that in a few years cloning microsatellites will be unnecessary for many species. However when attempting to apply aspecific primers in new species, it is necessary to try a range of amplification protocols and changing annealing temperature and template DNA concentration is usually sufficient to explore fully the possible applicability of the system.

It is possible that the frequency of “null alleles” may increase with the degeneracy of the primers used to a greater tendency for sensitivity to mis-priming. However, there is little evidence for this phenomenon to date. Successful cross-species amplification normally involves the reduction of specificity of PCR reactions in order to allow annealing of primers with target sequences displaying less similarity compared with the species for which primers were designed. This, however, may interfere with non-target sequences that may interfere with scoring of the allele size.

In general, the use of non-specific primers offers an exciting prospect for laboratories unable to undertake the laborious and time-consuming process of cloning microsatellites in new species.

1.3.4 Molecular DNA techniques

Random Amplified Polymorphic DNA (RAPD; Williams, 1990) and microsatellites (SSR; Tautz, 1989) are two widely used markers for estimating genetic diversity in many species. Deciding on which technique would be the most appropriate for any given investigation is not obvious and depends on a number of factors including the purpose of the research, the biology of the species and the resources available. When this study began, little was known about Cypress genetic variability. There were no specific markers for this species.

RAPD markers have been extensively used in assessing genetic variation of both wild and cultivated trees: *Pilgerodendron uviferum* (Allnutt, 2003), *Camellia sinensis* (Kaundun, 2002), *Pinus oocarpa* (Diaz, 2001), *Quercus petraea* (Le Corre, 1997), *Taxus baccata* (Hilfiker, 2004), *Chamaecyparis* (Hwnag, 2001), *Picea abies* (Jeandroz , 2004), *Castanea sativa* (Fornari, 1999), *Olea europea* L. (Belaj, 2002), *Acacia raddiana* (Shrestha, 2002), particularly as DNA sequence information is not required prior to investigating a previously unstudied species. Consequently this method is especially suited to situations where little or no molecular genetic research has been conducted previously. RAPDs offer the advantage of being simpler to use, less expensive and less time-consuming than RFLPs (Powell, 1996) or AFLPs (Lanham, 1999).

In order to obtain preliminary information about genetic variability of *C. sempervirens* in the north of Italy, we initially sampled a restricted set of populations (seven) from Trentino-South Tyrol. Later we included three Tuscan populations in the analysis, in order to compare northern samples with other Italian cypresses, and one Turkish population as “control” to check the technique used.

Recently, Sebastiani (2005) developed nine couples of microsatellite primers characterized by high polymorphism nuclear marker specific for *C. sempervirens*. Simple-sequence repeat (SSR) markers have been largely used for variability studies in genetic populations, as they are transferable, highly polymorphic, multiallelic polymerase chain reaction (PCR)-based codominant markers, relatively simple to interpret (Rafalski, 1993). Obtaining SSR markers require great initial effort because sequence information is needed (Morgante, 1993). Microsatellites allow to study distribution of diversity in natural and artificial stands and describe the distribution of diversity at different geographical scales. Eight new SSR specific markers were used to amplify already sampled *C. sempervirens* populations in order to confirm or reject results obtained from RAPD and to compare the techniques used.

Comparisons of molecular markers for measuring genetic diversity have been carried out in several plant species (Powell, 1996; Milbourne, 1997; Russell, 1997; Pejic, 1998; Crouch, 2000; Garcia-Mas, 2000; Staub, 2000) but, to our knowledge, no such studies have been reported in cypresses yet. We were

interested in comparing and evaluating the utility of the RAPD and SSR techniques in terms of the study of population genetics of the cypress. The results obtained are presented in the first section “RAPD and SSR” of this thesis.

We decided to amplify sampled areas in the north of Italy. We chose SSR markers to continue the work. We included 15 more populations in the analysis in order to obtain more information about *C. sempervirens* genetic structure. The results are described in the second section “SSR” of this thesis.

1.4. Aims of the work

This study is a part of a wider project called ECOCYPRE, financed by the Autonomous Province of Trento and dealing with “Ecological assessment and sustainable management of the cypress in the landscape of Trentino”.

The aim of ECOCYPRE project was the selection of some ornamental cypress clones resistant to *S. cardinale* and adapted to the climatic condition (cold tolerance) of the Italian northern regions.

Considering that the desiderable future clones of *C. sempervirens* resistant to canker and cold, that may arise from the project ECOCYPRE, could have negative impact on genetic structure of cypress populations and more in general on the biodiversity richness of the species, we want preliminary to assess the grade and the structure of the cypress genetic biodiversity resources in North Italy where presumably the new selected clones will be employed. The assessment was evaluated at two different geographical scales.

In the first stage of our work the objectives (described in “RAPD and SSR” sections) were

- to investigate the level of genetic diversity of *C. sempervirens* in stands located in Trentino-South Tyrol using RAPD markers,
- to amplify the study including some other Tuscan populations and one foreign population as “control” for the technique employed,
- to compare the discriminating capacity and informativeness of the PCR based molecular markers RAPD, and SSR for genotype identification and genetic diversity analyses;

The purpose of the second stage (described in “SSR” sections) was to study

- distribution of diversity in different Italian stands along the biggest glacial lakes in the Southern slopes of the Alps,
- cypress genetic structure at a wider geographical range.

MATERIALS AND METHODS

2.1 Sample collection

Leave samples were collected from 598 cypresses in twenty-seven different sites (Table 1). Twenty significant sites are situated in the North of Italy and give a good representation of *C. sempervirens* distribution in the area (Figure 5). Samples of five populations come from Tuscany where the cypress is largely spread, and samples of the last population come from Turkey, one of the regions of origin *C. sempervirens*. Individuals were chosen at random with a log diameter larger than 30 cm to evaluate genetic variability of mature plants. Eleven populations were used to obtain preliminary data with RAPD markers, while all the collected populations were analyzed with microsatellites (Table 1). Collected samples were frozen and stored at –80°C until liophylization.

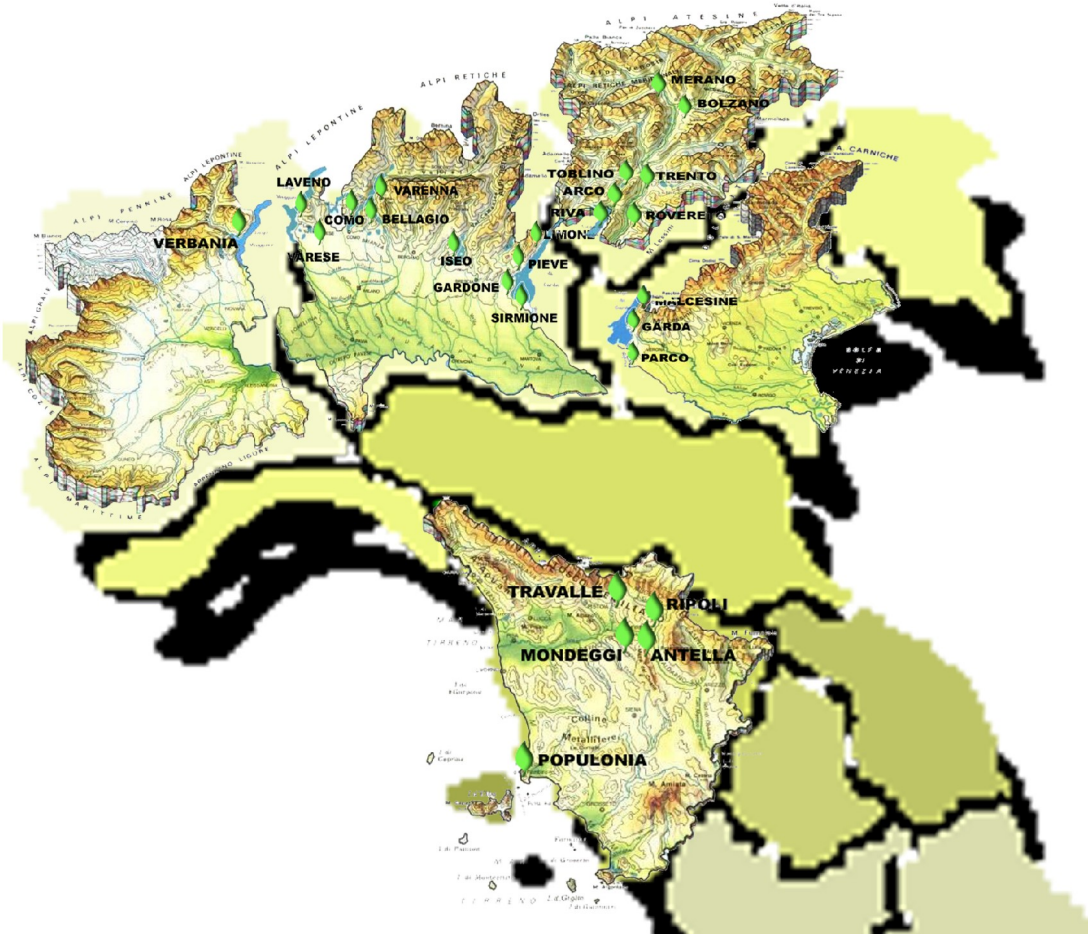


Figure 5: Location of Cupressus sempervirens *populations sampled in Italy.*

Group	Provenance	Population	N. Sample	Long. East	Lat. Nord	Markers
TUSCAN	TUSCANY	BAGNO A RIPOLI	21	11°19”	43°45”	RAPD & SSR
		MONDEGGI	33	11°19”	43°42”	RAPD & SSR
		ANTELLA	10	11°19”	43°43”	RAPD & SSR
		POPULONIA	18	10°31”	42°59”	SSR
		TRAVALLE	20	11°09”	43°52”	SSR
NORTH EAST	TRENTINO-SOUTH TYROL	BOLZANO	30	11°20”	46°29”	RAPD & SSR
		TRENTO	30	11°07”	46°04”	RAPD & SSR
		MERANO	27	11°09”	46°40”	RAPD & SSR
		ROVERETO	30	11°02”	45°53”	RAPD & SSR
		LAGO DI TOBLINO	30	10°57”	46°03”	RAPD & SSR
		ARCO	29	10°53”	45°55”	RAPD & SSR
		RIVA DEL GARDA	29	10°50”	45°53”	RAPD & SSR
	LAKE GARDA	PIEVE	20	10°45”	45°46”	SSR
		LIMONE	19	10°45”	45°48”	SSR
		MALECESINE	19	10°48”	45°45”	SSR
		GARDONE	20	10°33”	45°37”	SSR
		GARDA	20	10°42”	45°34”	SSR
		SIRMIONE	20	10°36”	45°29”	SSR
		PARCO SIGURTA’	18	10°44”	45.21”	SSR
NORTH WEST	LAKE ISEO	ISEO	27	10°03”	45°44”	SSR
	LAKE VARESE	VARESE	18	8°49”	45°48”	SSR
	LAKE MAGGIORE	LAVENO	20	8°36”	45°55”	SSR
		VARBANIA	18	8°33”	45°55”	SSR
	LAKE COMO	BELLAGIO	20	9°15”	45°59”	SSR
		COMO	13	9°04”	45°50”	SSR
		VARENNA	20	9°17”	46°00”	SSR
TURKISH	TURKEY	TURKEY	20	28°58”	41°01”	RAPD & SSR

Table 1: C. sempervirens *populations: group, region, population name, sample number in each population, geographic coordinates, marker used for analysis.*

2.2 DNA extraction

2.2.1 Sample lyophilization

Place approximately 100 mg of fresh leaves of each plant into a 2 ml Safe-Lock microtube. Let caps open, close the microtubes with parafilm and make two little holes on it. In this way samples can't go out but air and liquid nitrogen can go through parafilm plug. Froze the samples in liquid nitrogen and store at -80 °C until lyophilization.

Frozen samples are lyophilized in a vacuum pump in approximately 36 hours. Ensure that the material is dry otherwise grinding leaves with TissueLyser (Qiagen) can lead to clumps. Remove parafilm from each microtube and close caps. Samples can be stored until extractions.

2.2.2 Disruption using the TissueLyser

The TissueLyser (Qiagen) provides rapid and efficient disruption of 2 x 24 samples in parallel using 2 ml Safe-Lock microtubes and 5 mm stainless steel beads. Disruption is performed in three or four 1–2 minute high-speed (20–30 Hz) shaking steps. Using lyophilized tissue of *C. sempervirens* with TissueLyser (Qiagen) facilitates leaves disruption, improving the final yield.

Place a 5 mm stainless steel bead into the 2 ml Safe-Lock microtube, together with the sample material. Freeze the tubes in liquid nitrogen for 30 s. Place the tubes into the TissueLyser Adapter Set 2 x 24 and fix into the clamps of the TissueLyser. Immediately grind the samples for 1 min at 30 Hz. Disassemble the Adaptor Set, remove the microtubes, and re-freeze the samples in liquid nitrogen for 2 minutes.

Repeat last two steps until all samples have been ground to a fine powder (typically three or four times), reversing the position of microtubes within the Adaptor Set (in this way all the samples are thoroughly and equally disrupted).

2.2.2 DNA extraction

Dna was extracted using slight modified protocol of DNeasy Plant Mini Kit (Qiagen).

1) cells lyses: add 400 µl of Buffer AP1 and 4 µl of RNase A stock solution (100 mg/ml) to the frozen tissue and vortex vigorously. No tissue clumps should be visible. Vortex further to remove any clumps. Clumped tissue will not lyse properly and will therefore result in a lower yield of DNA. Incubate the mixture for 10 min at 65°C. Mix 2-3 times during incubation by inverting tube.

2) Precipitation of detergent, proteins, and polysaccharides: add 130 µl of Buffer AP2 to the lysate, mix, and incubate for 5 min on ice.

Cypress leaves generate very viscous lysates and large amounts of precipitates during this step resulting in shearing of the DNA in the next step. In this case, optimal results are obtained if the majority of these precipitates are removed by centrifugation for 5 min at 20,000 x g (14,000 rpm).

3) After centrifugation, apply supernatant to QIAshredder Mini Spin Column placed in a 2 ml collection tube and centrifuge for 2 min at 20,000 x g (14,000 rpm). The QIAshredder Mini Column removes most precipitates and cell debris.

Transfer flow-through fraction to a new tube without disturbing the cell-debris pellet. Typically 300-350 µl of lysate is recovered for cypress leaves.

5) Add 1.5 volumes of Buffer AP3/E ethanol-added to the cleared lysate and mix by pipetting. It is important to pipet Buffer AP3/E directly onto the cleared lysate and to mix immediately.

6) Apply 650 µl of the mixture from the previous step, including any precipitate that may have formed, to the DNeasy Mini Spin Column sitting in a 2 ml collection tube (supplied).

7) Centrifuge for 1 min at ≥6000 x *g* (corresponds to ≥8000 rpm for most microcentrifuges) and discard flow-through. Repeat the step with remaining sample. Discard flow-through and collection tube.

8) Place DNeasy Mini Spin Column in a new 2 ml collection tube , add 500 µl Buffer AW ethanol-added to the DNeasy Mini Spin Column and centrifuge for 1 min at 6000 x *g* (8000 rpm). Discard flow-through and reuse the collection tube.

9) Add 500 µl Buffer AW to the DNeasy Mini Spin Column and centrifuge for 2 min at 20,000 x *g* (14,000 rpm).

10) Discard flow-through and reuse the collection and centrifuge for 5 min at 20,000 x *g* (14,000 rpm) to dry the membrane.

It is important to dry the membrane of the DNeasy Mini Spin Column since residual ethanol may interfere with subsequent reactions. This spin ensures that no residual ethanol will be carried over during elution. Discard flow-through and collection tube.

11) Transfer the DNeasy Mini Spin Column to a 1.5 ml or 2 ml microcentrifuge tube and pipet 100 µl of Buffer AE preheated at 65°C directly onto the DNeasy membrane. Incubate for 5 min at room temperature and then centrifuge for 1 minute 6000 x *g* (8000 rpm) to elute. Repeat the step in the same microcentrifuge tube.

12) A new microcentrifuge tube must be used for a third elution step (in 200 µl of Buffer AE) to prevent dilution of the first eluate. Besides more than 200 µl should not be eluted into a 1.5 ml microcentrifuge tube because the DNeasy Mini Spin Column will come into contact with the eluate.

2.2.3 DNA quantitation

Double-stranded DNA concentration can be determined, by comparison with a known concentration standard, mainly by three methods: spectrophotometric determination, fluorometric determination and ethidium bromide fluorescent quantitation, generally by DNA electrophoresis on an agarose gel. The main difference among these techniques is in the minimum amount of DNA needed to obtain a measure. The exact lowest detection limit depends on the

instruments and tools used: in the first two cases, it can for example depend on using cuvettes or capillaries and on their volume; with the last method, even in standardised conditions detection limit can change depending on the teeth dimensions of the comb that is used. Anyway, lowest DNA detection limit can be estimated in about 250 ng for spectrophotometry, 15 ng for fluorimetry and 1-5 ng for ethidium bromide fluorescence on a gel. DNA fluorimeters are very useful to measure the concentration of digested DNA, but they are much less when looking at whole genomic DNA.

Apart from the lowest detection limit, another possible advantage of the gel method is the qualitative information on the extract condition, that is the DNA fragmentation state. In fact, when DNA is cut into small pieces it is not visualised on the gel as a clear band but as a “smear”. On the other hand, the major limit with this technique is the subjectivity introduced in measures by reading bands intensity and comparing with standards “by eye”.

The cypress DNA concentration was estimated by running samples on a 1% agarose gel in 0,5X TBE buffer. After staining with ethidium bromide gels were photographed in UV light and the brightness of each sample bands were compared with λ DNA standard of known concentration. Typically, about 50-100 ng/µl were recovered. When necessary, DNA aliquots were diluted by sterile water before RAPD analysis. RAPD patterns seem to be most affected by very low DNA concentrations, i.e., in the picogram range (Williams, 1993). Very high DNA concentrations and fragmented DNA, however, can also affect banding repeatability (Munthali, 1992) probably by inhibiting the reaction due to the increased presence of plant-derived contaminants (Vroh Bi, 1997).

Since the HotStarTaq polymerase seems to be insensitive to DNA concentration variation (within the tested range from 10 to 80 ng), it is believed that the simple method employed for the calculation of DNA concentration is accurate enough.

2.3 Setting up of RAPD-PCR amplification protocols

To set up the amplification conditions a long preliminary analysis was applied by considering a matrix of different combined conditions, including annealing temperatures, number of cycles and $MgCl_2$ concentration and time of elongation. Sixty-four 10mer primers RAPD (Operon Primer Technology: kit OPB, kit OPP and kit OPE and OPF3, OPH7, OPL6 e OPL19) were initially screened at standard conditions in order to see which primers amplify useful band patterns on *C. sempervirens*.

The PCR reactions were run in a T Gradient thermal cycler (Biometra) according to the following program: one cycle of 95 °C for 15 min, 35 cycles of 95 °C for 60 sec, 40°C for 60 sec, 72 °C for 120 sec, and final elongation cycle of 10 min at 72 °C. PCR reactions were carried out in a 20 μl reaction mix containing about 50 ng cypress DNA, 1X Buffer (provided by the manufacturer of the Taq enzyme), 200 μM each dNTP, 0,6 μM primer (Operon Primer Technology), 2,5 mM $MgCl_2$ (see table 2), 0,75 U HotStarTaq DNA Polymerase (Qiagen). Twenty-three primers were chosen for further analysis because of their good pattern in this preliminary screening. For each selected primer a PCR gradient was performed varying the annealing temperature from 36 to 48°C, keeping all the other conditions constant. The best annealing temperature for each primer was chosen for further analysis. Then reaction was set up for time of elongation, $MgCl_2$ concentration and number of cycles in order to obtain the best results for each primer. Variability and reproducibility of patterns were tested for each primer at the selected conditions.

Nine primers from the screening process that exhibited a good level of polymorphism, reproducibility of patterns and showed the best readability were chosen for further study of the individual genotypes.

2.3.1 RAPD amplification

Final PCR reactions were carried out in a 20 μl reaction mix containing 50 ng template DNA, 1X Buffer (provided by the manufacturer of the Taq enzyme), 200 μM each dNTP, 0,6 μM primer (Operon Primer Technology), 2,5 – 4 mM $MgCl_2$ (see table 2), 0,75 U HotStarTaq DNA Polymerase (Qiagen). PCR reactions were run in a T Gradient thermal cycler (Biometra) according to the following program: one cycle of 95 °C for 15 min, 40-45 cycles (Table 2) of 95 °C for 60 sec, annealing temperature of each primer (Table 2) for 60 sec, 72 °C for 90 sec, and final elongation cycle of 10 min at 72 °C.

Reproducibility of amplification profiles was tested for each primer. Only those bands consistently reproduced in different analyses were considered. Poor amplifications occurred systematically with individuals from different populations; these were excluded from the analysis and they mainly account for the different sample sizes of this study.

Primer	Sequence	n. cicles	T annealing	[MgCl2]
OPB 1	5'-GTTTCGCTCC-3'	40	44°C	2.5mM
OPB 5	5'-TGCGCCCTTC-3'	40	45°C	2.5mM
OPB 6	5'-TGCTCTGCCC-3'	40	45°C	2.5mM
OPB 7	5'-GGTGACGCAG-3'	45	47°C	2.5mM
OPB 8	5'-GTCCACACGG-3'	45	47°C	2.5mM
OPB 10	5'-CTGCTGGGAC-3'	40	38°C	2.5mM
OPB 11	5'-GTAGACCCGT-3'	40	44°C	2.5mM
OPP 16	5'-CCAAGCTGCC-3'	40	44°C	2.5mM
OPP 17	5'-TGACCCGCCT-3'	45	47°C	4.0mM

Table 2: Selected PCR conditions for the amplifications of each primer used.

PCR amplification RAPD products were separated by size on 1.5% agarose gel run in 0.5X TBE buffer, post-stained with ethidium bromide, and photographed in UV light. Pictures of the gels were taken using a digital camera and saved to computer as image files. To aid interpretation of band identity between different gels, each contained Hyperladder II (Bioline) migration standard.

2.4 SSR analysis

2.4.1 SSR amplification

Eight of nine primers couples developed by Sebastiani (2005) were used for genetic analysis on *C. sempervirens* (Table 3). Cyp 250 was excluded from analysis owing to its low polymorphism and because it repeatedly gave poor amplifications.

PCRs were performed in 15 µl containing 50 ng of DNA, 1x PCR reaction buffer (Bioline), 200 µM of each dNTP, 0.75 U Taq polymerase (BioTaq, Bioline), 1.5 mM MgCl2, 0.4 µM of each primer. The PCR cycle (except for cyp258) was: denaturation at 94ºC for 5 min followed by 30 cycles of 94ºC for 30s, 50ºC for 30s, 72ºC for 30s and final extension at 72ºC for 7 min.

A "touchdown" PCR protocol was used to amplify Cyp 258: 3 min at 94°C, 10 touchdown cycles of 94ºC for 30s, 60ºC for 30s (-1ºC/cycle), 72ºC for 60s; 25 cycles of 94ºC for 30s, 50ºC for 30s, 72ºC for 30s and final extension at 72ºC 8 min.

Locus name (Accession no.)	**Primer sequences 5'->3'**	**Label**	**Repeated motif**	**Clone size (bp) (Alleles size range)**	**N. of Alleles**	**H_O**	**H_E**
Cyp52	F-CATCCACTGCCAATACTTTT	FAM	$(GT)_{21}$	169	6	0.524	0.787
AY854181	R-AGCATCTTCCCATTACTTGA			146-174			
Cyp84*	F-CATTTCAATTTGCATAAGTTCT	TAMRA	$(GT)_{13}(TC)_{23}$	243	6	0.909	0.785
AY854182	R-GCAATGGGATGACTACAAAA			226-248			
Cyp101	F-AGGCCACACTCAAACTTATG	HEX	$(GT)_{12}$	175	13	0.773	0.862
AY854183	R-ATGACAATGGGTGAAGTCAT			175-209			
Cyp139	F-ACAACTAGAGAGGGAGTGAAAA	FAM	$(GA)_{19}$	234	8	0.733	0.864
AY854184	R-TGGTTGAAACAATAAAGGAGA			220-270			
Cyp174	F-CAACCCTTCTCTTCGATAGT	FAM	$(GA)_{21}$	210	8	0.800	0.856
AY854185	R-ACCTTCTTTGTCATCGTCAC			192-214			
Cyp250	F-ATGGATGCAAGAGATTTTGT	HEX	$(GT)_{16}$	138	2	0.444	0.366
AY854186	R-TGGTCCGATAGAAGTACTCG			136-138			
Cyp257	F- AACTTGCACATTTAGGGATG	HEX	$(GT)_{10}(TA)_4$	159	5	0.545	0.690
AY854187	R-TGATGGAATAACATGGACAG			159-183			
Cyp258**	F-AATTTGGGCTCATGAAATTA	FAM	$(GT)_{12}$	222	10	0.650	0.779
AY854188	R-TCTAGACCGATTCTATGGTCA			224-256			
Cyp293	F-GGCAAGTAATGAAACTCCAC	HEX	$(GT)_{14}$	182	8	0.455	0.525
AY854189	R-TACAAACATGCATGGCTAAC			167-195			

Table 3: Characteristics of dinuclotide microsatellite loci developed for C. sempervirens *(Sebatiani; 2005): H_O observed heterozigosity and H_E expected heterozygosity found on 24 cypresses from Italy, Turkey, France and Greece (6 samples each provenance).*

Five µL of each PCR product were loaded on 1,5% agarose gel, run in 0.5X TBE buffer, stained with ethidium bromide and photographed under UV light to control the success of the amplifications and to quantify fragments.

2.4.2 Electrophoresis analysis

Forward primers were labelled with a fluorescent dye (FAM HEX or TAMRA; Applied Biosystems) at 5' end (Table 3). This allowed the pooling of three PCR products together with an internal size standard in each well after adjustment of their relative concentrations.

Based on the intensity of the resultant band, the rest of the product (10 μl) was diluted with sterile water; from these dilutions, 0,5 μl of a Fam-labelled product, 0,5 μl of a Hex-labelled product and 0,5 μl of a Tamra-labelled product (or another Fam-labelled product when size ranges did not overlap) were mixed with 9,92 μl of formamide and 0,08 μl of GeneScan-500 ROX^{TM} Size standard (Applied Biosystems). Since the size ranges of loci were overlapping and primers had the same label, it was not possible to load more than three PCR products in one well. Samples were than heated to 95°C for 5 min and chilled on ice for 2 min.

Electrophoresis and detection of PCR products were carried on using the ABI 3100 Genetic Analyzer (Applied Biosystems).

The resulting electropherograms were analysed with the Genotyper® 3.7 software (Applied Biosystems). The standard included with each sample provided a set of DNA markers of suitable size to resolve the dimension of the analyzed PCR products. GENESCAN 3.7 (Applied Biosystems) software was then used for further analysis. All the genotypes were scored by eye, to attribute the exact allele size when ambiguous.

2.5 RAPD and SSR Statistical analysis

2.5.1 Intrapopulation analysis

Analyses were performed on eleven populations (Table 4).

Since RAPD were dominant markers, we assumed that each band represented the dominant homozygote-heterozygote phenotype at a single biallelic locus. Amplified fragments, named by the primer used and the size in base pairs (bp), were scored as presence (1) or absence (0) of homologous bands, and a matrix of the different RAPD phenotypes was assembled. Variations in intensity between bands of the same molecular weight across samples were not considered to be polymorphism.

Our analysis was based on those RAPD bands that fulfilled Lynch and Milligan's assumptions for the analysis of dominant markers: bands with observed frequency < 3/*N* (where N is the total number of studied individuals) were included in the analysis (Lynch, 1994).

The POPGENE 32 freeware (Yeh, 1997) was used for RAPD markers to calculate the percentage of polymorphic loci (%P), Nei's gene diversity (*He*; Nei, 1973) and Shannon's Information index (*I*; Shannon, 1949) to provide a relative estimate of the degree of variation within each population. The same software was used to estimate the total genetic diversity (H_T), the mean gene diversity (H_S) and the proportion of variation distributed among populations (G_{ST}) averaged over all the polymorphic loci (Nei, 1978), and over all the Italian and regional populations. Furthermore POPGENE 32 was used to output Nei's standard genetic distance matrix.

GenAlEx 6.0 (Peakall, 2005) freeware was used for SSR dataset to calculate the number of alleles per locus, the mean number of allele per locus for each population (Na), the percentage of polymorphic loci (%P), the Shannon information index (I), the observed heterozygosity (H_o), the expected heterozygosity (H_E) and the fixation index (*F*) and Nei's standard genetic distance matrix.

GenAlEx 6.0 was also employed to analyze relationship between 240 RAPD phenotypes using the Euclidean square distance (pairwise difference). The obtained matrix was used to perform principal coordinate analysis (PCA) and results were visualized with "SPSS 13.0 for Windows" statistical analysis program to provide a 3D graphical representation of the RAPD relationships between individuals.

The analysis of molecular variance (AMOVA) was used to describe population structure for RAPDs and SSRs, respectively. GenAlEx provided AMOVA both for RAPDs and SSRs. A pairwise, individual-by-individual (*N x N*) genetic distance matrix was generated for binary data. This calculation of pairwise genetic distances for binary data follows the method of Huff (2003), in which any comparison with the same state yields a value of 0 (both 0 vs 0

comparisons and 1 vs 1 comparisons), while any comparison of different states (0 vs 1 or 1 vs 0) yields a value of 1. When calculated across multiple loci for a given pair of samples, this is equivalent to the tally of differences between the two genetic profiles. This distance option is used to calculate Φ_{PT} via AMOVA, a measure of population genetic differentiation for binary data that is analogous to *Fst* (Excoffier, 1992; Huff, 1993; Peakall, 1995). A pairwise, individual-by-individual (*N x N*) genetic distance matrix is calculated for SSR data. For a single-locus analysis, with *i*-th, *j*-th, *k*-th and *l*-th different alleles, a 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 $d_2(ii, jj) = 4$ (Peakall, 1995; Smouse, 1999). This distance option was used to calculate Φ_{PT} via AMOVA, a measure of population genetic differentiation that suppresses intra-individual variation and is therefore ideal for comparisons between codominant and binary data (Maguire, 2002), where no intra-individual variation (heterozygosity) is available.

SAMOVA analysis (Spatial Analysis of Molecular Variance) (Dupanloup, 2002) was also performed on the whole dataset. SAMOVA is a freeware program that uses a new approach for defining groups of populations that are geographically homogeneous and maximally differentiated from each other. The method is based on a simulated annealing procedure that aims to maximize the proportion of total genetic variance due to differences between groups of populations.

2.5.2 Interpopulation analysis

Interpopulation divergence was calculated for both markers using Nei's genetic distance coefficient (Nei, 1978) and Φ_{PT} pairwise genetic distance coefficient (with statistical testing by random permutation; Excoffier, 1992). The two distance matrices were then used for the following statistical analyses.

Mantel test for Matrix Correspondence (Mantel, 1967; Smouse, 1992), allows tests for a statistical relationship between the elements of two-distance matrices with matching entries with the option for statistical testing by random permutation.

First we tested for the correlation of two genetic distance matrices calculated from the different genetic markers (SSRs and RAPDs). Than we tested isolation-by-distance, comparing genetic distance matrices with the geographic distance matrix for the respective populations.

In addition, population relationships were inferred both using the UPGMA clustering method on the basis of Nei's genetic distance (the tree was subsequently visualized with Mega 3.1) and using PCA analysis (performed with GenAlEx).

2.6 SSR statistical analysis

2.6.1 Allelic variation and intrapopulation analysis

Analyses were performed on twenty-seven populations (Table 1).

The following statistics of genetic variation were computed with the software GenAlEx 6.0:

Na is the number of alleles per locus calculated over all the entire dataset and for each population. The arithmetic mean across loci was also provided.

Na >5% is the number of alleles with frequency greater than 5%

Allele frequency was calculated on population and locus. To visualize allele distribution, microsatellite allele frequency by locus was graphically represented.

No. of private alleles is the number of alleles unique to a single population in the data set.

$Ne = \frac{1}{1 - He}$ is the effective number of alleles. It was calculated for single loci and then the arithmetic mean across loci was made. This measure enables meaningful comparisons of allelic diversity across loci with diverse allele frequency distributions. It provides an estimate of the number of equally frequent alleles in an ideal population with homozygosity equivalent to the actual population;

$He = 1 - \sum p_i^2$ is the expected Heterozygosity or Genetic Diversity (Nei, 1973) calculated on a single locus basis where *pi* is the frequency of the *i*-th allele.

$\overline{He} = \frac{\sum_{i=1}^{k} H_e(i)}{k}$ *Mean He* is the average per-population genetic diversity, used in the calculation of *F*-statistics. *He* is the expected heterozygosity in population *i*, *k* is the number of populations.

$Ho = \frac{No.\_of\_Hets}{N}$ is the Observed Heterozygosity (calculated on a per locus basis) where the number of heterozygotes is determined by direct count and *N* is the number of samples.

$\overline{Ho} = \frac{\sum_{i=1}^{k} H_o(i)}{k}$ *Mean Ho* is the average observed heterozygosity of a collection of populations, used in the calculation of *F*-statistics. Here, *Ho*(*i*) is the observed heterozygosity in population *I* and *k* is the number of populations.

$F = \frac{He - Ho}{He}$ is the fixation index calculated on a per locus basis. The arithmetic mean across loci GenAlEx was also provided. Values close to zero

are expected under random mating, while substantial positive values indicate inbreeding or undetected null alleles. Negative values indicate excess of heterozygosity, due to assortative mating, or heterotic selection.

$F_{IS} = \frac{\overline{H_e} - \overline{H_o}}{\overline{H_e}}$ Fis: is the inbreeding coefficient within individuals, relative to the population. *Fis* measures the reduction in heterozygosity of an individual, due to non random mating within each population.

$F_{IT} = \frac{H_T - H_o}{H_T}$ *Fit* is the inbreeding coefficient within individuals, relative to the total. This statistic takes into account the effects of both non random mating within and genetic differentiation among populations.

$F_{ST} = \frac{H_T - \overline{H_E}}{H_T}$ *Fst* is the inbreeding coefficient within subpopulations, relative to the total. *Fst* provides a measure of the genetic differentiation among populations, that is the proportion of the total genetic diversity (~ heterozygosity) that is distributed among the populations. *Fst* is typically greater than or equal to zero. If all subpopulations are in Hardy-Weinberg equilibrium with the same allele frequencies, *Fst* = 0. *Fst* is more or less equivalent to *Gst*.

$H_T = 1 - \sum_{i=1}^{k} \overline{p_i}^2$ *Ht* is the total expected heterozygosity, a partition of the total genetic diversity among populations, used in the calculation of *F*-statistics. Here, *pi* is the frequency of the *i*-th allele, averaged over populations. *Ht* is the expected heterozygosity if all the populations were pooled (i.e., if there were no population subdivision).

PI, Probability of identity, is an estimate of the average probability that two unrelated individuals, drawn from the same randomly mating population, will by chance have the same multilocus genotype. It is used as an indication of the statistical power of a specific set of marker loci. Probability of identity to increase locus combinations (PI), expected number of individuals with the same multilocus genotype to increase locus combination (Exp n. PI), PI by single locus, expected number of individuals with the same genotype at a given locus were calculated on the Italian dataset.

A *Multilocus Matches Analysis* was performed in order to identify the presence of putative clones between sampled populations.

Correlation tests between several populations' parameters (Na, He, Ho, F) and their geographical coordinate (lat and long) were performed to verify correspondences between genetic variations along geographic distances.

GENEPOP 3.4 (Raymond, 1995) was used for testing:

genotypic linkage disequilibrium: the null hypothesis was that genotypes at one locus were independent from genotypes at the other locus. GENEPOP

creates contingency tables for all pairs of loci in each sample and performs a probability test (or Fisher exact test) for each table using a Markov chain. The estimation input parameters used for the Markow chain were 10 000 dememorization steps, 100 batches, and 5000 iterations per batch.

Hardy –Weinberg equilibrium. Excesses or deficits in genetic diversity were measured for each locus by Weir and Cockerman (1984) fixation index. As the samples included a large number of individuals and many alleles were found, that produced many low-frequency genotypes, goodness-of-fit tests such as χ^2 were poorly suited for testing departures from Hardy–Weinberg proportions (HWP; Guo, 1992 a). Instead, tests for departures from HWP were conducted using GENEPOP 3.4 program. In this program, probability values were enumerated completely for loci with fewer than five alleles, while for loci with five or more alleles the *P* -values were estimated using the Markov chain methods of Guo and Thompson (1992 b); estimation input parameters were 1000 dememorization steps, 100 batches, and 5000 iterations per batch.

2.6.2 Interpopulation analysis

GENEPOP 3.4 was used for testing:

Excat test of Populatoin differentiation is concerned with the allelic distribution of alleles in the various samples. For each locus, an unbiased estimate of the *P*-value of the probability test (or Fisher exact test) was performed, (Raymond, 1995). This procedure tests each locus for significant differences in allele frequencies among the populations, using a Markov Chain to generate an exact probability distribution under the null hypothesis that there is no differentiation among populations and is not biased by rare alleles or small sample sizes. The estimation input parameters were 10000 dememorization steps, 100 batches, and 5000 iterations per batch.

GenAlEx 6.0 was used to perform:

$Nei\_I = \frac{J_{xy}}{\sqrt{(J_x J_y)}}$ *Nei standard genetic identity*; $J_{xy} = \sum_{i=1}^{k} p_{ix} p_{iy}$, $J_x = \sum_{i=1}^{k} p_{ix}^2$ and $J_y = \sum_{i=1}^{k} p_{iy}^2$. Here p_{ix} and p_{iy} are the frequencies of the *i*-th allele in populations *x* and *y*. For multiple loci, *Jxy*, *Jx* and *Jy* were calculated by summing over all loci and alleles and dividing by the number of loci. These average values were then used to calculate *I*.

$Nei\_D = -\ln(I)$ Nei standard genetic distance: Where I is Nei's Genetic Identity (see below) (Nei, 1972 and 1978). The calculation of Nei's standard genetic distance was computed between pairs of populations. This measure is one of the most widely used for estimating genetic distance between populations. Nei' standard genetic distance was used to show relationship between populations by using the unweighted pair group method using

arithmetic averages (UPGMA) method. The tree was constructed and visualized with MEGA 3.1.

PCA (principal coordinate analysis) was performed on the basis of Nei's genetic distance matrix and results were visualized with "SPSS 13.0 for Windows" statistical analysis program to provide a 3D graphical representation of the relationships between populations.

AMOVA was also used to measure population differentiation (Excoffier, 1992), including a permutation method for testing significance of differentiation. Hierarchical AMOVA tests were conducted with the populations divided into three groups based on their geographical locations.

$D = \sqrt{(xi - xj)^2 + (yi - yj)^2}$ *Geographic distance* xi and yi are the coordinates for the i-th sample and xj and yj are the coordinates for the j-th sample. A pairwise, linear geographic distance matrix is generated from X and Y coordinates of populations.

Isolation by distance Isolation by distance was tested by a Mantel test between the matrix of population pairwise estimates of Nei's genetic distance and the matrix of their geographic distance.

Spatial genetic structure. The 'global' spatial autocorrelation method of Smouse and Peakall (Smouse, 1999) employs a multivariate approach to simultaneously assess the spatial signal generated by multiple genetic loci. Both pairwise geographic and pairwise genetic distance matrices are required as inputs for such analysis. The autocorrelation coefficient generated (*r*) provides a measure of the genetic similarity between pairs of populations whose geographic separation falls within the specified distance class. The results are summarised by a correlogram. Following Peakall, (2003), GenAlEx offers tests for statistical significance, From 1000 random permutations, the values of the 25th and 975th ranked *rp* values are taken to define the upper and lower bounds of the 95% confidence interval. If the calculated *r*-value falls outside this confidence belt, then significant spatial genetic structure is inferred. If positive spatial genetic structure is found, the first *x*-intercept has often been interpreted in the literature as an estimate of the extent of non-random (positive) genetic structure.

RESULTS

3.1 RAPD and SSR results

3.1.1 Plant material

Eleven populations were used to obtain preliminary data using RAPD markers (Table 4). Poor RAPD amplifications occurred systematically with individuals from different populations, these were excluded from the analysis. In order to compare the discriminating capacity of the two markers, the same samples were analyzed with SSR and RAPD. The results obtained in this section using SSRs can be slightly different from data obtained in the other section with the same markers on the same stands due to difference in the analyzed populations size.

Group	Provenance	Population	N. Sample	Long. Est	Lat.Nord
TUSCAN	TUSCANY	BAGNO_A_RIPOLI	21	11°19"	43°45"
		MONDEGGI	33	11°19"	43°42"
		ANTELLA	10	11°19"	43°43"
NORTH EAST	TRENTINO-SOUTH TYROL	BOLZANO	21	11°20"	46°29"
		TRENTO	20	11°07"	46°04"
		MERANO	19	11°09"	46°40"
		ROVERETO	24	11°02"	45°53"
		LAGO_DI_TOBLINO	23	10°57"	46°03"
		ARCO	27	10°53"	45°55"
		RIVA_DEL_GARDA	22	10°50"	45°53"
TURKISH	TURKEY	TURCHIA	20	28°58"	41°01"

Table 4: C. sempervirens populations included in RAPD and SSR analysis: group, region, populations' names, number of analyzed individuals per populations, geographic coordinate.

3.1.2 Setting up of RAPD-PCR amplification protocols

To set up the amplification conditions a long preliminary analysis (described in "Material and Methods") was applied by considering different combined settings including annealing temperatures, number of cycles, $MgCl_2$ concentration, time of elongation. Than variability and reproducibility of patterns were tested. Table 5 resumes selected conditions for each primer.

Primer	Sequence	n. cicles	T annealing	[MgCl2]
OPB 1	5'-GTTTCGCTCC-3'	40	44°C	2.5mM
OPB 5	5'-TGCGCCCTTC-3'	40	45°C	2.5mM
OPB 6	5'-TGCTCTGCCC-3'	40	45°C	2.5mM
OPB 7	5'-GGTGACGCAG-3'	45	47°C	2.5mM
OPB 8	5'-GTCCACACGG-3'	45	47°C	2.5mM
OPB 10	5'-CTGCTGGGAC-3'	40	38°C	2.5mM
OPB 11	5'-GTAGACCCGT-3'	40	44°C	2.5mM
OPP 16	5'-CCAAGCTGCC-3'	40	44°C	2.5mM
OPP 17	5'-TGACCCGCCT-3'	45	47°C	4.0mM

Table 5: 10-mer primers used for RAPD analysis, primer sequences, number of PCR cycles, optimal annealing temperatures, optimal $MgCl_2$ concentration.

3.1.3 Fingerprinting of *C. sempervirens* populations

The nine RAPD selected primers generated a total of 55 consistently well-amplified bands (with an average of 6,1 bands scored per primer), ranging in size from 350 to 1420 base pairs (Table 6). Fifty of these bands (90,9%) were polymorphic among the 11 populations, while five were monomorphic.

It's to be noted that a higher number of bands appeared polymorphic but were rejected because the assignment of the presence or absence was ambiguous for several individuals. By using this approach, we were fully aware of the possibility of losing useful information, but our aim was to obtain reproducible and clear data.

All the 50 polymorphic bands selected fulfil Linch and Milligan criterion over all the populations. No individuals were characterized by the same RAPD profile and population-specific bands were not found.

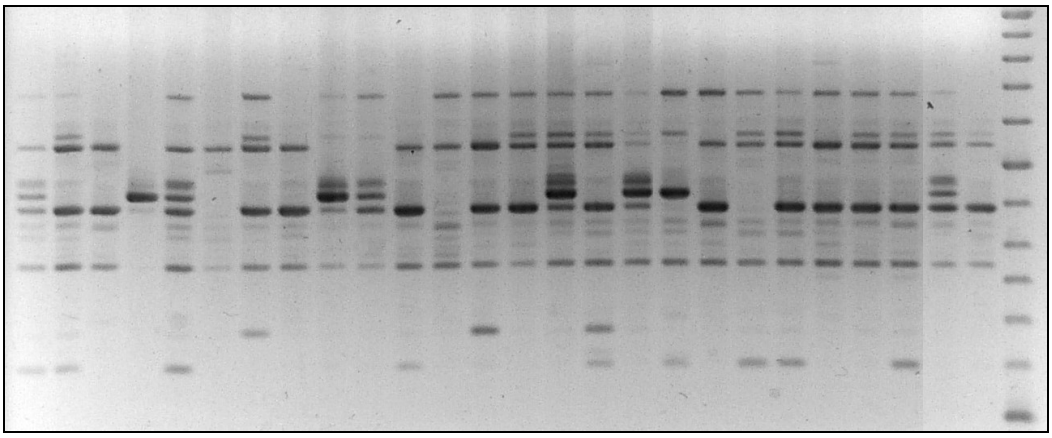


Figure 6: illustrates a typical example of the band pattern generated with the OPB16 primer.

Amplification of the SSR markers was performed with 8 primer pairs that produced in total 87 fragments: 20 of them are private Turkish allele, while 67 are present in Italian populations. Primer pair Cyp 84 produced the maximum number of alleles (18), while Cyp 257 and 293 the minimum (8).

SSR locus Cyp 174 had an observed heterozygosity similar to the one expected under Hardy-Weinberg equilibrium (*Fis*=0.005). All the other loci showed a deficiency of heterozygotes (*Fis* ranged from 0.135 to 0.688; Table 6) with an overall *Fis* 0.382.

RAPD				SSR			
Primer	Bp max-min	polimorfic	Total	Primer	Bp Max-Min	N. alleles	Fis
OPB 1	1120-490	4	4	**Cyp 52**	145-186	10 (6)	0.514
OPB 5	1050-800	2	2	**Cyp 84**	226-288	18 (17)	0.443
OPB 6	970-450	3	4	**Cyp 101**	174-208	11 (8)	0.579
OPB 7	800-530	8	8	**Cyp 139**	217-270	11 (7)	0.688
OPB 8	700-350	10	10	**Cyp 174**	192-231	12 (9)	0.005
OPB 10	1420-530	4	7	**Cyp 257**	145-183	8 (5)	0.368
OPB 11	900-600	3	4	**Cyp 259**	216-256	9 (8)	0.135
OPP 16	1380-400	9	9	**Cyp 293**	175-194	8 (7)	0.322
OPP 17	1050-350	7	7				

Table 6: 10-mer primers used for RAPD analysis, size range of useful bands, number of polymorphic bands, number of total amplified bands for each primer. Primers used for SSR analysis, size range of alleles, number of alleles over all the populations and in Italian populations (in brackets), Fis over all populations.

3.1.4 Intrapopulation variation

The percentage of polymorphic markers in each population ranged from 68% to 80% for RAPD and from 88 to 100% for SSR.

Diversity measures were calculated with Nei's genetic diversity index (*He*: expected Heterozygosity) and Shannon Information Index (*I*) both for RAPD and SSR.

There was a significant difference in expected heterozygosity among markers types (Gl=1; P<0,001). Expected heterozygosity as measured by SSRs was about 1,7 times the one obtained using RAPD markers. There were slight differences in *He* among the Italian populations and RAPD loci showed relatively little variation in *He* (0.208-0.265) as compared to SSR loci (0.410-0.529) (Table 7). *He* and *I* values measured by SSR were not significantly correlated with *He* and *I* values found with RAPD.

The total diversity (H_T) within all the populations was 0.286 while the average diversity (H_S) was 0.238. The mean level of genetic differentiation between all the populations (G_{ST}) over all the loci was 0.168. This indicates that a low proportion of diversity is observed between populations as compared with diversity within populations (Table 8). When considering only Italian plants the total diversity within populations (H_T) was 0.269, the average diversity (H_S) was 0.237. The level of genetic differentiation between Italian populations (G_{ST}) decreases to 0.12, so demonstrating that the Turkish population was quite different from Italian ones.

Trentino-South Tyrol populations showed H_T = 0.253, H_S = 0.231 and a significant value of G_{ST} = 0,088 while Tuscan populations had H_T = 0.267, H_S = 0.252 and were very similar (G_{ST} = 0.043) (table 8).

			SSR						RAPD		
Group	Population	N. samples	Na	Ne	%P	I	He	Ho	%P	I	He
Trentino-South Tyrol		155	3.804	2.305	95%	0.868	0.464	0.281	98%	0.392	0.253
	Bolzano	21	4.375	2.356	100%	0.992	0.529	0.330	74%	0.330	0.213
	Trento	20	4.250	2.450	100%	0.957	0.509	0.317	76%	0.381	0.252
	Merano	19	3.500	2.182	100%	0.835	0.463	0.258	76%	0.360	0.237
	Rovereto	24	3.750	2.432	88%	0.868	0.455	0.296	70%	0.329	0.216
	Lago_di_Toblino	23	3.625	2.294	88%	0.846	0.452	0.238	72%	0.319	0.208
	Riva_del_Garda	22	3.500	2.233	100%	0.784	0.415	0.241	72%	0.359	0.240
	Arco	27	3.625	2.186	88%	0.793	0.428	0.290	70%	0.369	0.246
Tuscany			3.458	1.997	100%	0.791	0.442	0.303	90%	0.405	0.267
	Bagno_a_Ripoli	65	3.625	2.036	100%	0.814	0.453	0.305	80%	0.380	0.253
	Mondeggi	21	3.375	1.892	100%	0.719	0.410	0.300	82%	0.400	0.265
	Antella	33	3.375	2.064	100%	0.840	0.463	0.306	68%	0.354	0.237
Italy		220	3.700	2.213	96%	0.845	0.457	0.288	74%	0.415	0.269
Turkey	Turchia	20	7.125	3.950	100%	1.563	0.720	0.400	78%	0.379	0.250
	Total population	240	3.945	2.330	97%	0.896	0.481	0.298	74%	0.437	0.286

Table 7: Groups, populations in each group, number of samples per popualtion. SSR mean number of alleles (Na), *effective number of alleles, percentage of polimorphic loci (P%), Shannon inforamtion index* (I), *expected heterozygosity* (He), *obeserved heterozygosity* (Ho). *RAPD percentage of polimorphic loci* (P%), *Shannon information index* (I), *expected heterozigosity* (He).

	Trentino-A A pops	Tuscany	Italiy	All population
RAPD HT	0.253	0.267	0.269	0.286
RAPD HS	0.231	0.252	0.237	0.238
RAPD GST	0.088	0.043	0.120	0.168
SSR F_{ST}	0.053	0.040	0.072	0.141

Table 8: RAPD H_T H_S G_{ST} and SSR Fst for all the principal region analyzed.

3.1.5 Relationship between individuals

The 240 RAPD phenotypes were analyzed using the Euclidean square distance (pairwise difference). The obtained matrix was used to perform principal coordinate analysis (PCA) and to provide a graphical representation of the RAPD relationships between individuals (Figure 7). The first three coordinate axes accounted for 14,67%, 8,54%, and 6,05% of the total variance respectively, and identified two main groups corresponding to the Italian and Turkish populations. Italian plants were approximately separated into two groups that partially reflected their regional provenance while there was no evidence for population specific clustering of individuals. These data proved the

reliability of the differences between Italian and Turkish populations and the prevalent grouping of Italian individuals within their own provenance region.

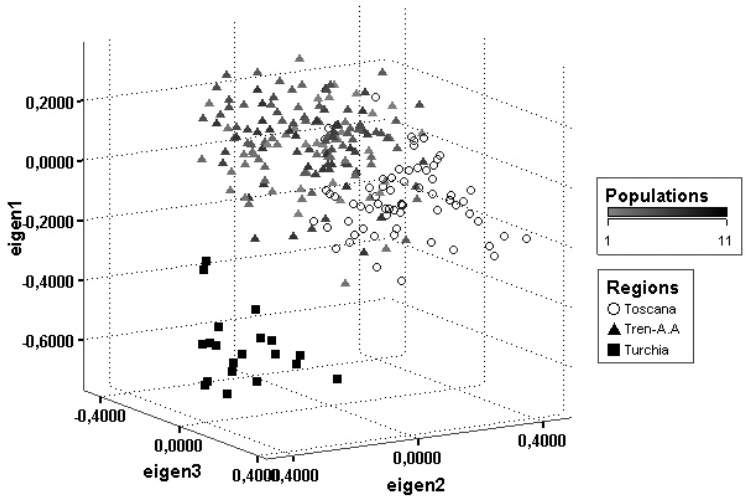


Figure 7: PCA analysis between individuals: Turkish plants; Trentino Alto-Adige plants; Tuscan plants

PCA was performed also based on SSR Euclidean square distance matrix. The first three coordinate axes accounted for 20,52%, 12,59%, and 8,39% of the total variance respectively, and substantially reflected the results obtained with RAPD markers, even if the Turkish population was more scattered (data not shown).

3.1.6 AMOVA analysis

AMOVA analysis on RAPD and SSR markers for all the populations allowed a partitioning of the overall variation into three levels (Table 9 a). The results indicated that the division between the Italian populations and the Turkish one was very pronounced (variance among groups: 30,03% RAPDs; 32,5% SSRs). Both marker types showed that the majority (64,7%, RAPDs; 63,5%, SSRs; based on Φ_{PT}) of genetic diversity was contained within populations (Table 9). The remaining diversity was distributed between populations within groups (6,3% RAPDs; 4% SSRs). SAMOVA analysis (Dupanloup, 2002), based both on SSRs and RAPD, confirmed the partition of the 11 populations into the same two groups: the Italian populations on one side and the Turkish population on the other.

A different partitioning was found taking into account only the Italian populations (Table 9 b): the majority of variance was found within populations, the residual diversity was distributed among groups (Φ_{RT} = 0.10, P < 0.001 RAPDs; Φ_{RT} 0.068, P < 0.001 SSRs) and between populations within groups (Φ_{PR} = 0.045, P < 0.001 RAPDs; Φ_{PR} = 0,033 SSRs). The overall differentiation

among the Italian populations was Φ_{PT} = 0.14 (P < 0.001 RAPDs) and Φ_{PT} 0,099 (P < 0,001 SSRs).

SAMOVA analysis based on RAPD confirmed this structure, while SAMOVA based on SSRs divided the Italian populations into three groups: the first included Tuscan populations, the second included Bolzano, and the third all the other populations.

When the studies were restricted to populations within a particular region, a low but significant structure was evident in Trentino-South Tyrol found both with RAPD (Φ_{PT} = 0.06, P < 0,001), and SSR (Φ_{PT} = 0.033, P < 0,001; Table 9 c). RAPDs found that Tuscan populations were not significantly different (Table 9 d). SSR attributed 2.9% of the total variance to differentiation among populations but the value is not statistically significant (P = 0.079).

	AMOVA	**a) 2 Groups**	**b) 2 Groups**	**c) 1 group**	**d) 1 group**
	1° group	Turkey pop	Tuscan pops	Trentino-ST pops	Tuscan pops
	2° group	Italian pops	Trentino-ST pops		
RAPD	Among groups	30.0%	10.1%		
	Among pops within groups	6.5%	4.1%	6.1%	0%
	Within pops	64.7%	85.8%	93.9%	100%
	Φ_{RT}	0.30***	0.10 ***		
	Φ_{PR}	0.09 ***	0.045 **		
	Φ_{PT}	0.36 ***	0.14 ***	0.06***	ns
SSR	Among groups	32.5%	6.8%		
	Among pops within groups	4.0%	3.1%	3.3%	2.9%
	Within pops	63.5%	90.1%	96.7%	97.1%
	Φ_{RT}	0.32***	0.068***		
	Φ_{PR}	0.06***	0.033***		
	Φ_{PT}	0.36***	0.099***	0.033***	0.0286 ns

Note:; * P< 0,05; ** P < 0.01; *** P < 0,001; ns not significant

Table 9: Hierarchical AMOVA analysis based on RAPD and SSR markers. a) The eleven populations analyzed were divided into two groups: the first group includes the Turkish population and the second all the Italian populations; b) Italian populations were divided into two groups: the first group includes all the Trentino-South Tyrol populations and the second all the Tuscan populations; c) Trentino-South Tyrol populations; d) Tuscan populations.

3.1.7 Interpopulation variation

Interpopulation divergence was calculated for both markers using Φ_{PT} pairwise genetic distance coefficient (Table 10) with statistical testing by random permutations and the Nei' s standard genetic distance. All pairwise RAPD Φ_{PT} values between populations were significant, except comparisons between Tuscan populations. Instead several SSR Φ_{PT} values between populations were not significant.

	Bol	Tre	Mer	Rov	Tob	Gar	Arc	Mon	Rip	Ant	Tur
Bol	0.000	0.044	0.049	0.072	0.073	0.092	0.061	0.086	0.111	0.107	0.394
Tre	0.064	0.000	0.040	0.064	0.076	0.088	0.048	0.135	0.126	0.154	0.315
Mer	0.031	0.032	0.000	0.065	0.058	0.051	0.051	0.126	0.111	0.141	0.356
Rov	0.053	0.029	0.017	0.000	0.026	0.036	0.047	0.157	0.143	0.181	0.414
Tob	0.083	0.036	0.014	0.019	0.000	0.072	0.064	0.167	0.156	0.209	0.434
Gar	0.096	0.033	0.007	0.005	0.000	0.000	0.065	0.120	0.093	0.136	0.374
Arc	0.064	0.033	0.002	0.037	0.024	0.006	0.000	0.147	0.133	0.138	0.360
Rip	0.028	0.063	0.049	0.048	0.077	0.074	0.083	0.000	0.000	0.000	0.346
Mon	0.051	0.157	0.119	0.125	0.136	0.179	0.166	0.038	0.000	0.006	0.316
Ant	0.000	0.108	0.050	0.066	0.098	0.119	0.100	0.028	0.008	0.000	0.334
Tur	0.278	0.284	0.292	0.307	0.337	0.327	0.317	0.304	0.386	0.246	0.000

Table 10: Below diagonal SSR pairwise Φ_{PT} between Italian populations; above diagonal RAPD pairwise Φ_{PT} between Italian populations. Underlined values are not significant (P>0,05).

A Mantel test was performed to discover correlation between pairs of matrices. The Turkish population was excluded from these analyses to avoid its over-influence. The comparison of RAPD and SSR markers showed a good agreement between pairwise Φ_{PT} estimates (Figure 8). The correlation found between Φ_{PT} matrices calculated with RAPD and SSR markers was fairly high (r = 0.594; P = 0.007). A higher correlation was found when Nei's genetic distances were considered (r = 0.708; P = 0.004; Figure 9).

The relative power of SSR markers with respect to RAPD was assessed by considering the slope of the regressed function between RAPDs and allozymes distances (or divergences). The sensitivity of SSR markers to population differentiation turned out to be 1.57 and 2.80 (1/0.638 and 1/0.358) times greater than the RAPDs's.

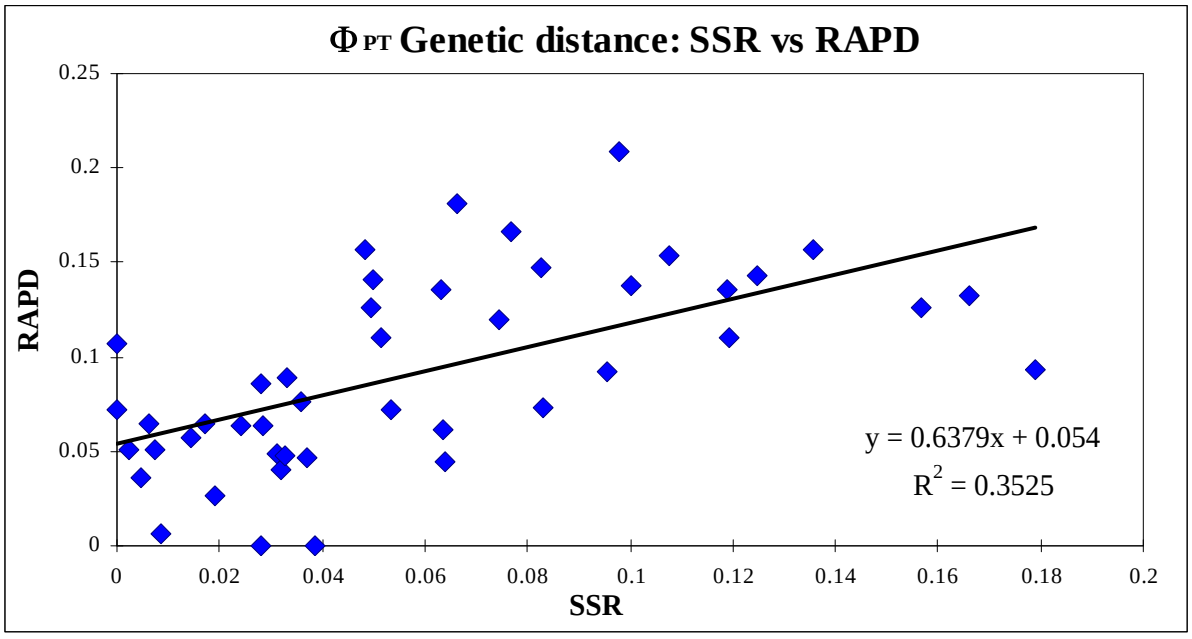


Figure 8: Mantel test between Φ_{PT} pairwise genetic distance matrix obtained from SSR and from RAPD markers; r =0.594, r^2=0.352, P=0.007.

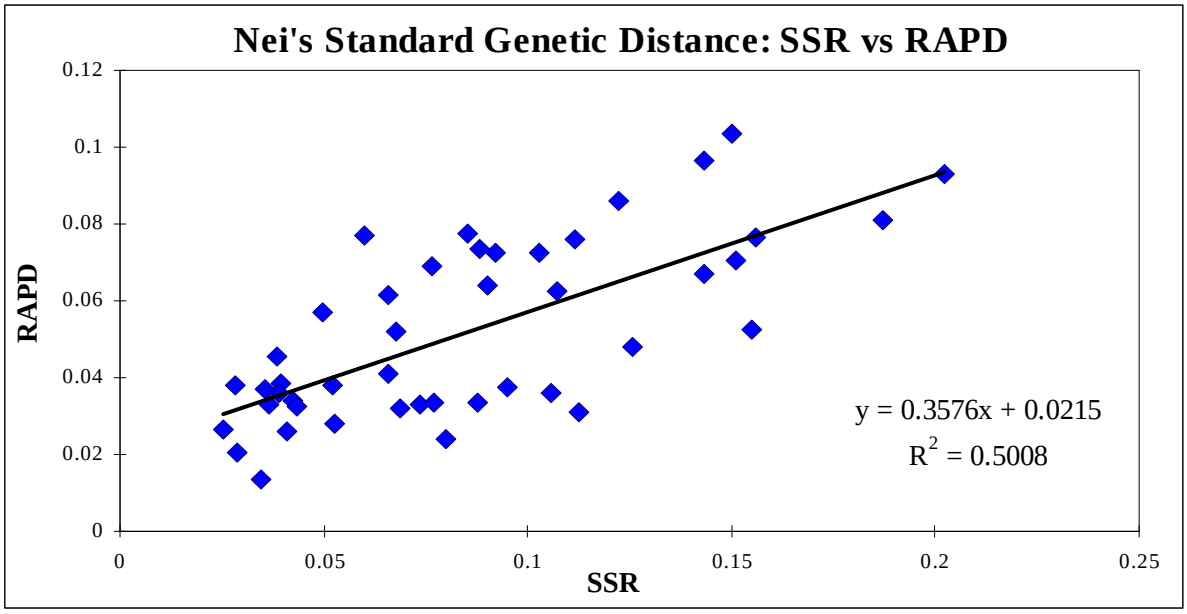


Figure 9: Mantel test between Nei standard genetic distance matrix Obtained from SSR and from RAPD markers;, r =0,708 r^2=0,501, P=0,001.

Two dendrograms were constructed to express the results of the cluster analyses based, respectively, on the RAPD (Figure 10) and SSR fragments (Figure 11). The UPGMA dendrograms obtained from the Nei's genetic distance matrices revealed a good grouping of the Cypress populations into three main clusters, each one corresponding to their respective regional location (Trentino-South Tyrol, Tuscany, and Turkey; Figure 10 and Figure 11). SSR dendrogram shows that Bolzano clusters with Tuscan populations (Figure 11). Also PCA analysis was performed on the same matrices (excluding Turkish population; Figure 12 a and b).

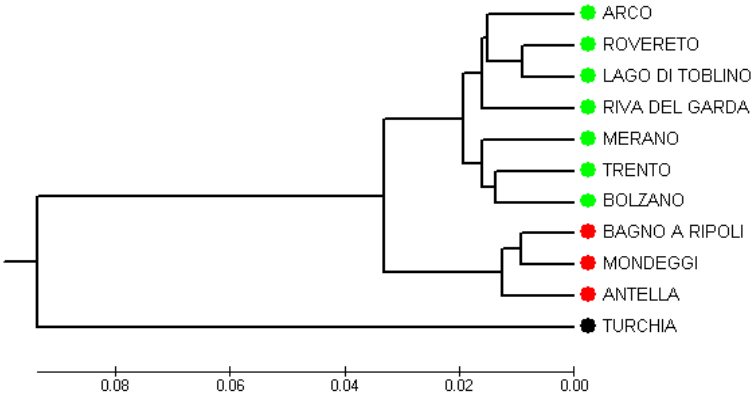


Figure 10: UPGMA dendrogram based on RAPD Nei's genetic distance matrix. ●Red: populations from Tuscany; ●green: populations from Trentino-South Tyrol; ●black: population from Turkey.

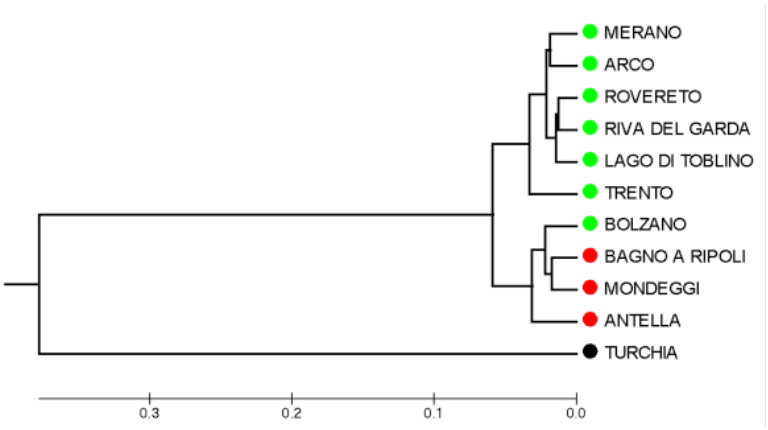
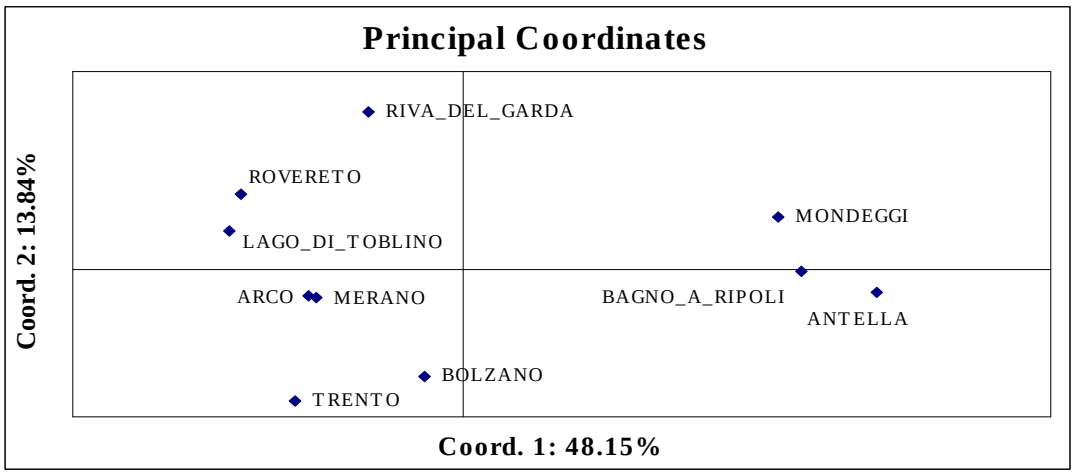
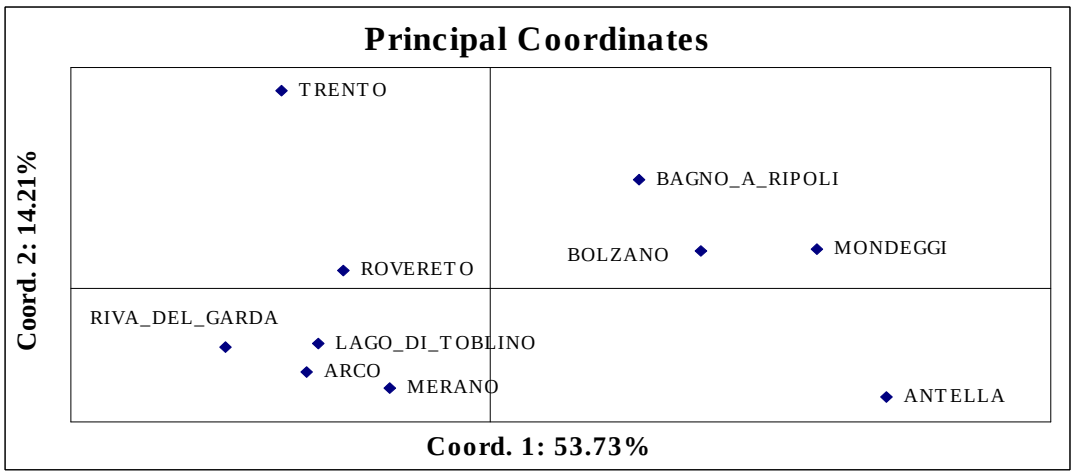


Figure 11: UPGMA dendrogram based on SSR Nei's genetic distance matrix. ●Red: populations from Tuscany; ●green: populations from Trentino-South Tyrol; ●black: population from Turkey.



a



b

Figure 12: PCA based on Nei's genetic distance matrices for RAPD (a) and SSR (b). The first three axes explain respectively 70,83% for RAPD and 78,84% for SSR of the total variance.

A mantel test was used also to discover correlation between genetic and geographic distance matrices (excluding the Turkish population to avoid its over-influence). In all cases there is a positive correlations between genetic and geographic distance (Table 11), mainly due to difference between Tuscan and Trentino-South Tyrol populations.

Marker	Genetic distance vs Gographic distance	r^2	r	P
RAPD	Nei's genetic distance matrix vs Geogrphic distance matrix	0.707	0.841	0.003
	Fst pairwise distance matrix vs Geogrphic distance matrix	0.6806	0.825	0.005
SSR	Nei's genetic distance matrix vs Geogrphic distance matrix	0.3201	0.55	0.015
	PhiPT pairwise distance matrix vs Geogrphic distance matrix	0.335	0.579	0.016

Table 11: Mantel tests between genetic distance matrices obtained with RAPD or SSR and Geographic distance matrix.

3.2 SSR results

3.2.1 Allelic variation

PCR amplifications produced results for all the 8 loci in more than 95% of individuals. One to two alleles were detected in each sample by using 8 SSR.

All the 8 loci were polymorphic, having a total of 104 alleles among the 598 diploid individuals. Turkey showed 15 private alleles, every one of them rare (freq. <5%). The number of observed alleles (Na) in all the populations per locus varied from 8 to 22 (Figure 13) with a mean of 13 alleles per locus and their size ranges were broader than the ones found by Sebastiani (2005) (Table 12).

As we were interested mainly in Italian population structure, the Turkish population was excluded from further analysis and used as an “out-group” only when necessary or interesting.

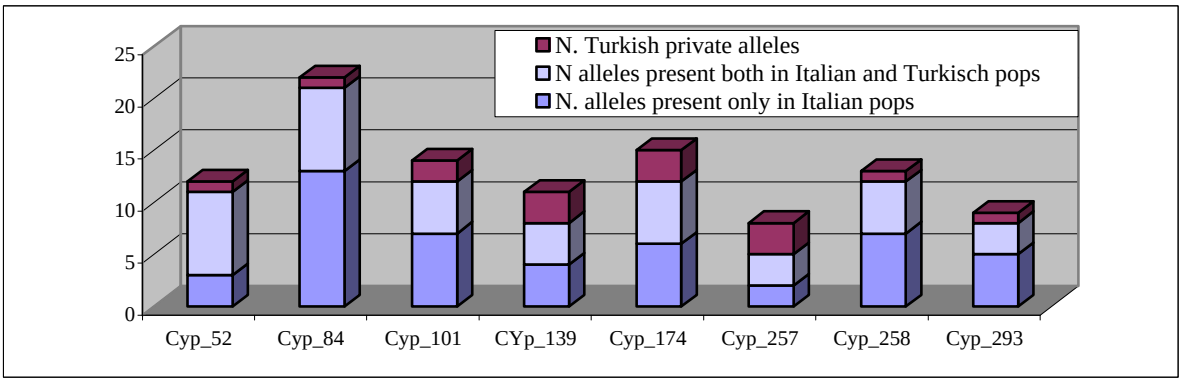
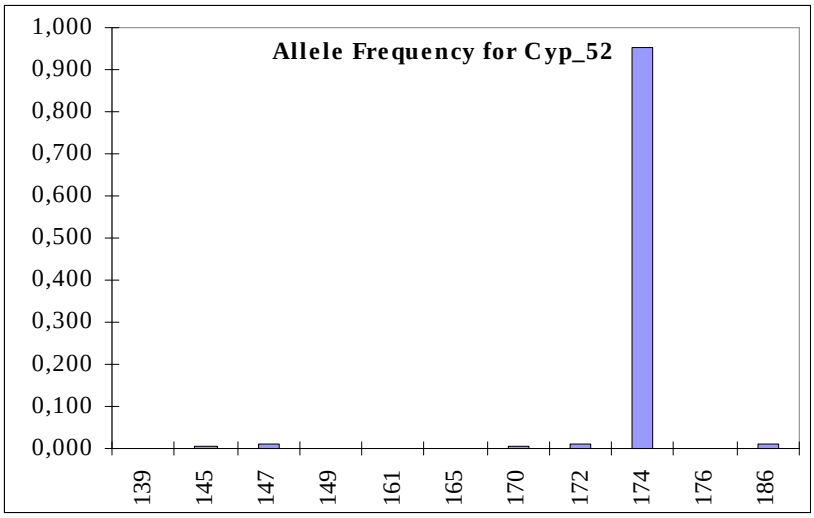


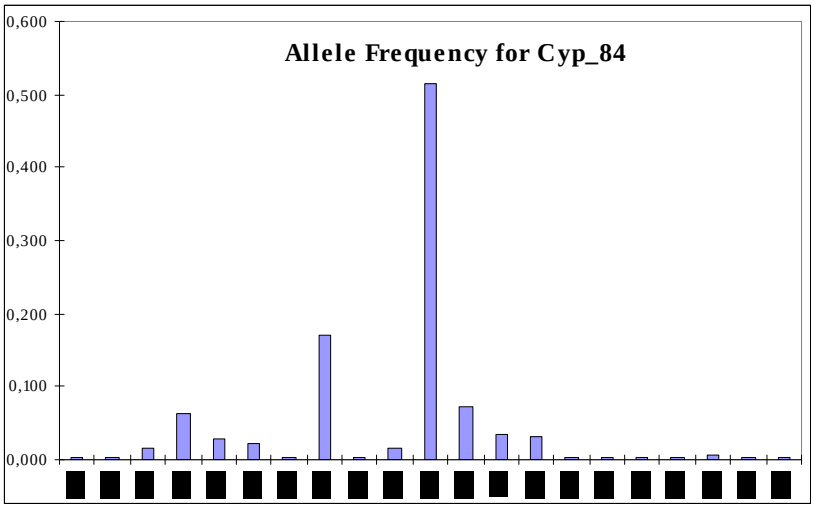
Figure 13: Total number of alleles detected for each primer. ¦ Fraction of the total alleles found only in Turkish population, ¦ fraction of the total allele found both in Turkish and in Italian populations, ¦ and fraction of total alleles found only in Italian populations but not in the Turkish one.

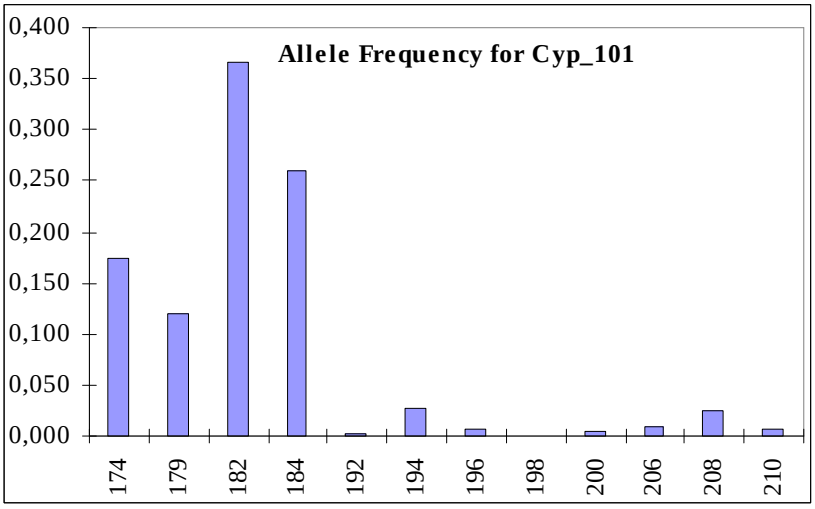
Of the 89 Italian alleles found, most (62) were rare (had frequencies < 0.05), 16 were common (frequencies between 0.05 and 0.5), and only 6 were abundant (frequencies > 0.5). Figure illustrates the distribution frequency of the alleles produced from eight SSR markers over the Italian populations. Most alleles were separated by single or multiple unit sizes (e.g. 2 bp for dinucleotide repeat loci), but some alleles did not fit these standard series: Cyp 84 and Cyp 293 showed low frequency alleles that differed by single base pair, while Cyp 52 showed two alleles that differed for 3 bp. In case of such strange series, the actual number of nucleotides in the allele needs to be established by sequencing, since the mobility of Fam- or Hex-labeled products could be slightly different from the Rox-labeled standards used for size estimations.

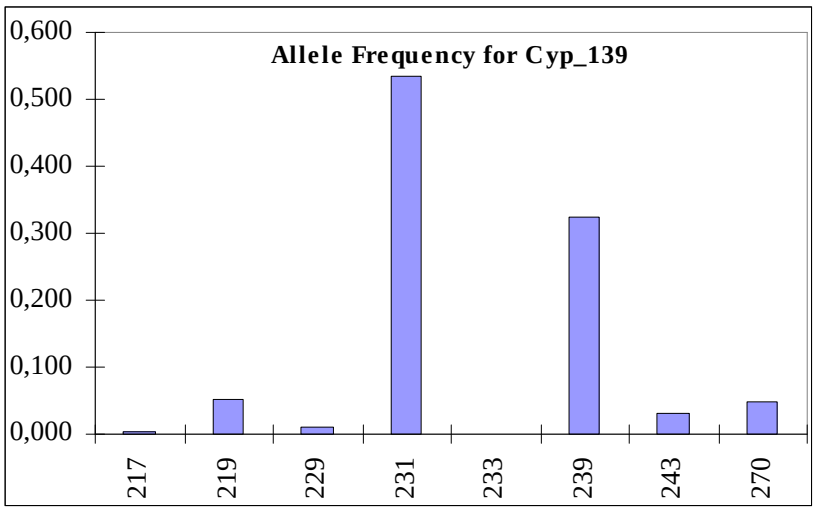
The microsatellite allele frequency distributions of the eight markers was very different from a distribution “under stepwise mutation model”, and seemed

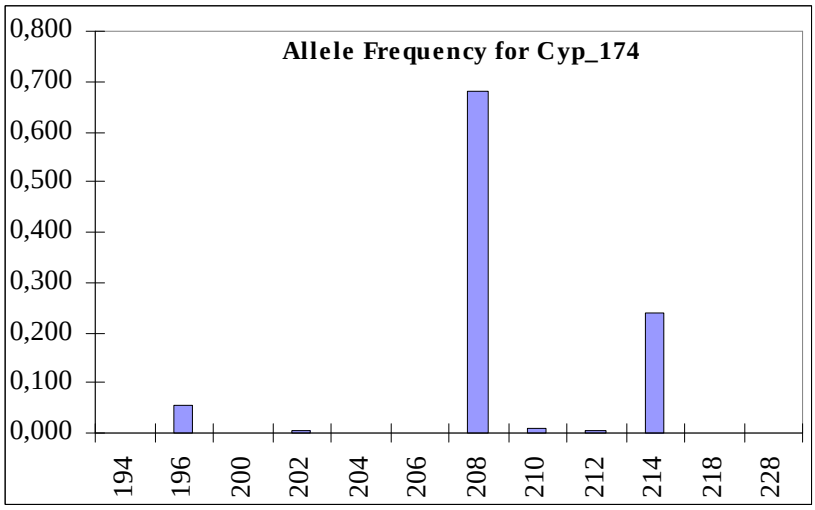
affected by drift and selection that had removed some allele sizes while randomly selecting others to be greatly over-represented.

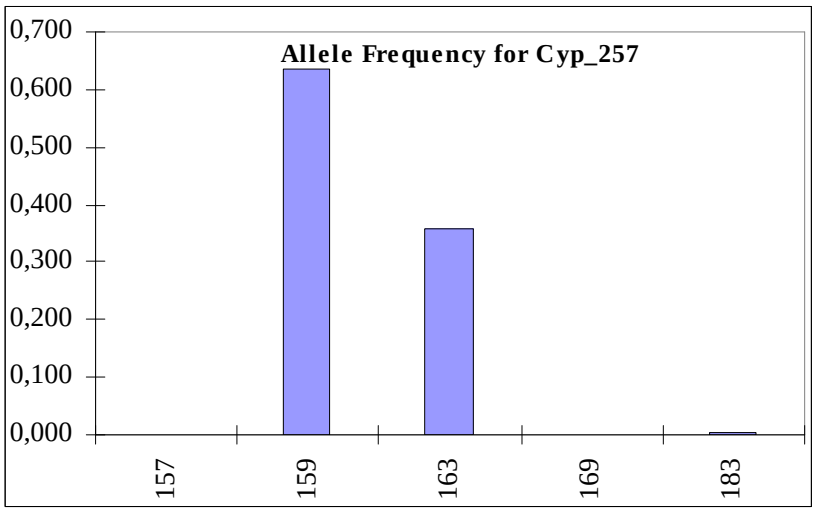







Allele Frequency for Cyp_139
0,600
0,500
0,400
0,300
0,200
0,100
0,000
217
219
229
231
233
239
243
270


Allele Frequency for Cyp_174
0,800
0,700
0,600
0,500
0,400
0,300
0,200
0,100
0,000
194
196
200
202
204
206
208
210
212
214
218
228


Allele Frequency for Cyp_257
0,700
0,600
0,500
0,400
0,300
0,200
0,100
0,000
157
159
163
169
183

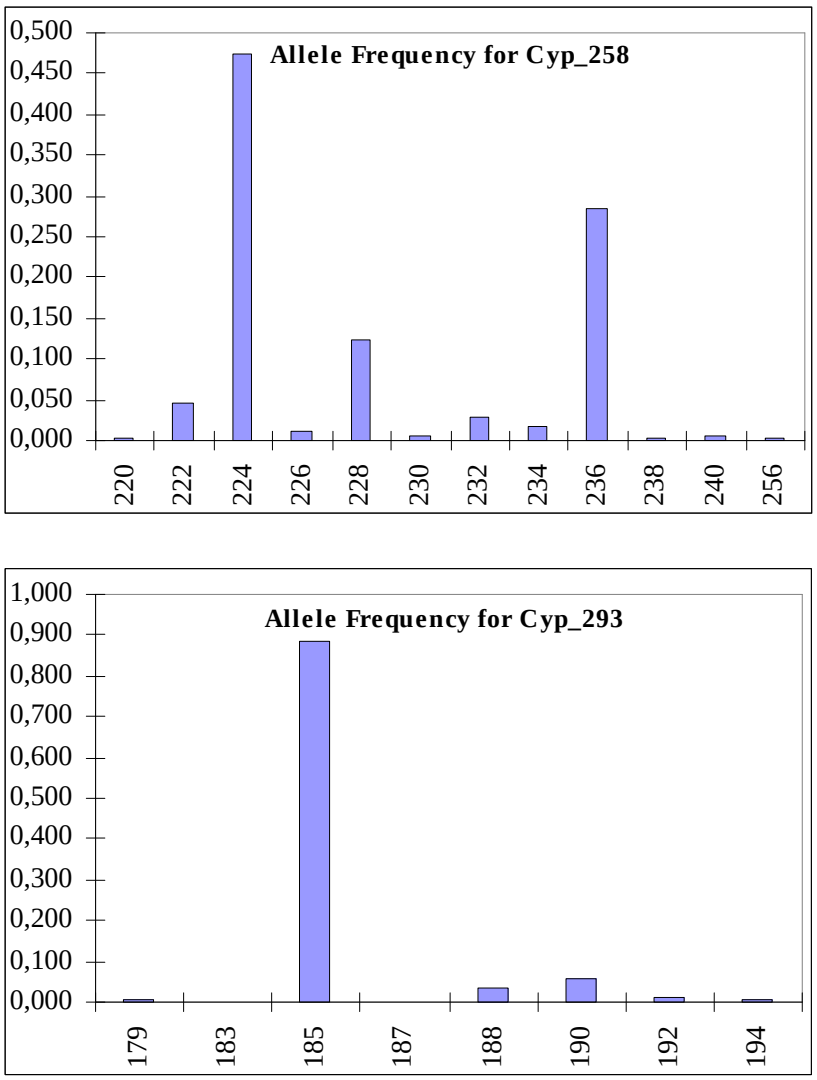


Figure 14: Allele distribution frequency generated from eight SSR markers over the Italian population.

Ho and *He* overall Italian samples (calculated in the populations "Italy") analyzed together range respectively from 0,043 to 0,515 and from 0,090 to 0,753 (Table 12). Every locus showed high significant excess of homozygotes.

Locus	Repeated unit	Range bp	Na	Ne	Ho	He	F		Ht	MeanHo	MeanHe	Fis	Fit	*Fst*
Cyp_52	$(GT)_{21}$	139-186	11	1.099	0.043	0.090	0.527	***	0.089	0.047	0.083	0.430	0.466	0.064
Cyp_84	$(GT)_{13}(TC)_{23}$	226-288	21	3.262	0.477	0.693	0.312	***	0.679	0.497	0.631	0.213	0.269	0.071
Cyp_101	$(GT)_{12}$	174-210	12	4.041	0.284	0.753	0.623	***	0.750	0.292	0.669	0.564	0.611	0.108
CYp_139	$(GA)_{19}$	217-270	8	2.521	0.200	0.603	0.669	***	0.595	0.192	0.507	0.621	0.677	0.148
Cyp_174	$(GA)_{21}$	194-228	12	1.922	0.444	0.480	0.074	***	0.482	0.446	0.463	0.035	0.075	0.041
Cyp_257	$(GT)_{10}(TA)_4$	157-183	5	1.878	0.242	0.468	0.482	***	0.473	0.228	0.436	0.477	0.518	0.078
Cyp_258	$(GT)_{12}$	220-256	12	3.094	0.515	0.677	0.239	***	0.683	0.511	0.641	0.204	0.252	0.061
Cyp_293	$(GT)_{14}$	179-194	8	1.276	0.094	0.216	0.565	***	0.238	0.097	0.215	0.549	0.593	0.096

Table 12: Information for each locus: Number of alleles (Na), number of effective alleles (Ne), observed heterozygosity over all Italian samples (Ho), expected heterozygosity over all Italian samples (He), average expected heterozygosity among population (mean He), average observed heterozygosity among populations (mean Ho), inbreeding coefficient within individuals relative to the population (Fis), inbreeding coefficient within individuals relative to the total (Fit), inbreeding coefficient within subpopulations, relative to the total (Fst)

The Probability of Identity (PI) test provide an estimate of the average probability that two unrelated individuals drawn from the population can have the same multilocus genotype. The PI for Increasing Locus combinations over all the dataset with 8 loci was 0.00. The estimated expected number of individuals with the same multilocus genotype for increasing locus combination (Exp n. PI) was 0,015. This test demonstrated that the eight selected loci were useful for reliable genetic tagging (Table 13)

Pop	Cyp_52	+Cyp_84	+Cyp_101	+CYp_139	+Cyp_174	+Cyp_257	+Cyp_258	+Cyp_293
PI	0.828	0.097	0.010	0.002	0.001	0.000	0.000	0.000
Exp. No PI	478.634	56.149	5.550	1.231	0.401	0.155	0.024	0.015
PI by Locus	0.828	0.117	0.099	0.222	0.326	0.386	0.152	0.621
Exp No. PI by Locus	478.634	67.806	57.133	128.146	188.297	223.324	87.914	359.189

Table 13: Probability of identity for increasing locus combinations (PI), expected number of individuals with the same multilocus genotype for increasing locus combination (Exp n. PI), Probability of identity by single locus (PI by Locus) and expected number of individuals with the same genotype at a given locus (Exp No. PI by Locus) calculated over all Italian samples.

A Multilocus Matches Analysis was performed in order to identify clones. Thirty-three samples identified 13 putative clones in the dataset: most of the matching genotypes belonged to the same populations or to one of the nearest neighboring populations with some exceptions: one cypress from Varenna matched with one cypress of Garda, one sample from Varese matched with one sample from Lake Toblino and five samples from Bolzano matched with one plant from Merano and one from Bagno a Ripoli.

Linkage disequilibrium was also assessed. Eight loci afforded 28 pairwise comparison of independence, 23 of which were consistent with hypothesis of

independence ($P >> 0.05$). Genotypes at Cyp 101 showed to be in significant linkage disequilibrium with Cyp 139 ($P < 0.01$), Cyp 257 ($P < 0.01$), Cyp 258 ($P < 0.001$) and Cyp 293 ($P < 0.01$). Genotypes at Cyp 84 and Cyp 258 were also not independent ($P < 0.05$).

3.2.2 Intrapopulation variation

The total number of alleles detected in a single population varied from 23 for Bellagio to 40 for Trento and Bolzano with an overall average of 31,5.

In general populations from the northwest (Lake Como, Lake Maggiore, Lake Varese and Lake Iseo) produced a lower number of allele than the ones from the northeast (Trentino-South Tytrol, Lake Garda): there is a positive correlation ($r^2 = 0.337$, $P < 0.01$) between latitude and number of alleles per population in the north of Italy.

Of the 19 private alleles detected (Table 14), 13 belonged to the north-east populations, four to the Tuscan populations and two to the North-west populations; however none of these provided a distinctive, constant character because the majority were found at low frequency (freq. < 0.05 for 74% of private alleles). Eight of the nineteen private alleles detected among the Italian populations, were common among Turkish samples. This result suggests that some individuals from Turkey were occasionally introduced and mixed with Italian populations.

Pop	Locus	Private Allele	Freq
MONDEGGI	Cyp_293	*183*	0,016
POPULONIA	Cyp_101	*198*	0,029
POPULONIA	Cyp_174	228	0,056
TRAVALLE	Cyp_52	*176*	0,025
BOLZANO	CYp_139	217	0,040
TRENTO	Cyp_101	200	0,093
TRENTO	Cyp_257	*157*	0,033
ARCO	Cyp_174	*194*	0,034
ARCO	Cyp_174	*206*	0,017
RIVA_DEL_GARDA	Cyp_257	169	0,034
MALCESINE	Cyp_52	165	0,026
GARDONE	Cyp_52	139	0,028
GARDONE	Cyp_52	161	0,028
GARDA	Cyp_101	210	0,158
GARDA	Cyp_174	200	0,025
GARDA	Cyp_174	218	0,025
PARCO SIGURTA	Cyp_258	220	0,059
LAVENO	Cyp_52	*149*	0,053
BELLAGIO	Cyp 139	*233*	0,029

Table 14: Summary of Private alleles by populations: population, locus, allele dimension, frequency in the population. Alleles in Italics were found also in Turkish population.

Heterozygosities, both observed and expected, differed among populations (Table 15). Average expected heterozygosity (*He*) was 0,456, ranging from 0,352 to 0,551. *Ho* varied from 0,191 of Merano and 0,384 of Travalle and was generally higher in southern populations than in northern ones: there is a negative correlation between longitude and mean *Ho* ($r^2 = 0,269$, P=0,0066). *He* was always higher than *Ho* and all the populations have a significant heterozygote deficit. No significant correlations were found between populations *He* and their geographical coordinates. A positive correlation exists between *F* and longitude ($r^2 = 0.2934$; P= 0.004).

Of the 208 locus-by-locus Italian population comparisons, 95 departed significantly from hardy-Weinberg equilibrium (HWE) and all of them but one showed a significant excess of homozygotes. Cyp 52 was fixed in 13 populations and Cyp 293 was fixed in 7 populations. Cyp 101 was out of equilibrium in all the populations except Travalle, Laveno and Verbania, Cyp 174 was in equilibrium in 22 populations out of 26 and Cyp 84 was out of equilibrium in all Trentino-South Tyrol populations; and they followed HWE for all the other populations but three. The locus/population combinations that were outside HWE were not concentrated in a single population or at any definite locus. In the Turkish population Cyp 174, Cyp 257 and Cyp 258 followed HWE.

Linkage disequilibrium was also assessed. Eight loci afforded 28 pairwise comparison of independence, 23 of which were consistent with hypothesis of independence (P>>0,05). Genotypes at Cyp 101 showed to be in significant linkage disequilibrium with Cyp 139 (P< 0,01), Cyp 257 (P<0,01), Cyp 258 (P<0,001) and Cyp 293 (P<0,01). Genotypes at Cyp 84 and Cyp 258 were also not independent (P<0,05).

				Mean			
Population	N	N allele	No. Private Alleles	Na	Ne	He	Ho
BAGNO_A_RIPOLI	21	29	0	3,625	2,036	0,453	0,305
MONDEGGI	33	27	1	3,375	1,892	0,410	0,300
ANTELLA	10	27	0	3,375	2,064	0,463	0,306
POPULONIA	18	31	2	3,875	1,898	0,422	0,326
TRAVALLE	20	31	1	3,875	2,049	0,456	0,384
BOLZANO	30	40	1	5,000	2,472	0,539	0,282
MERANO	27	32	0	4,000	2,130	0,444	0,191
TRENTO	30	40	2	5,000	2,570	0,508	0,284
ROVERETO	30	32	0	4,000	2,496	0,469	0,296
LAGO_DI_TOBLINO	30	34	0	4,250	2,404	0,479	0,271
ARCO	29	29	1	3,625	2,218	0,437	0,290
RIVA_DEL_GARDA	29	30	1	3,750	2,197	0,425	0,244
PIEVE	20	34	0	4,250	2,417	0,470	0,296
LIMONE	19	27	0	3,375	1,749	*0,352*	0,287
MALCESINE	19	32	1	4,000	2,114	0,433	0,276
GARDONE	19	36	2	4,500	2,653	0,455	0,281
GARDA	20	39	3	4,875	2,583	0,444	0,297
SIRMIONE	20	32	0	4,000	2,448	0,462	0,316
PARCO SIGURTA'	18	36	1	4,500	2,099	0,450	0,283
ISEO	27	36	0	4,500	2,217	0,475	0,340
LAVENO	20	35	1	4,375	2,247	0,514	0,354
VERBANIA	18	25	0	3,125	2,046	0,468	0,267
VARESE	18	25	0	3,125	1,828	0,381	*0,195*
BELLAGIO	20	23	0	2,875	1,897	0,429	0,257
COMO	13	30	0	3,750	2,506	0,551	0,341
VARENNA	20	26	0	3,250	2,058	0,460	0,240
Mean	22,231	31,462	0,654	3,933	2,203	0,456	0,289
Total	578	89	19	11,125	2,387	0,498	0,287

Table 15: Italian analyzed populations, numeber of smaple per popualtion, number of alleles per populetion, number of private alleles per population, Na *average number of allele,* Ne *effective number of allele,* He *expected heterozygosity,* Ho *observed heterozygosity.*

3.2.3 Interpopulation variation

The exact test of population differentiation (Raymond, 1995) indicated highly significant differentiation among the populations overall ($p<<0{,}001$). All loci showed significant differentiation among the 26 Italian populations. To elucidate the genetic relationship in cypress populations, a dendrogram was constructed based on Nei's genetic distance between populations by the UPGMA method (Figure 15). Populations were divided into 2 clusters that roughly reflect geographic distribution: the first (cluster 1) contains all the populations from Lake Garda and from Trentino-South Tyrol except Bolzano. The second cluster (cluster 2) contains all the other populations and Bolzano. This indicates that exchange of genetic material exists; probably the five clones detected in Bolzano came from Tuscany. When these clones were excluded from analysis, Bolzano formed an independent clade separated from the other two groups that remained substantially unchanged (data not shown).

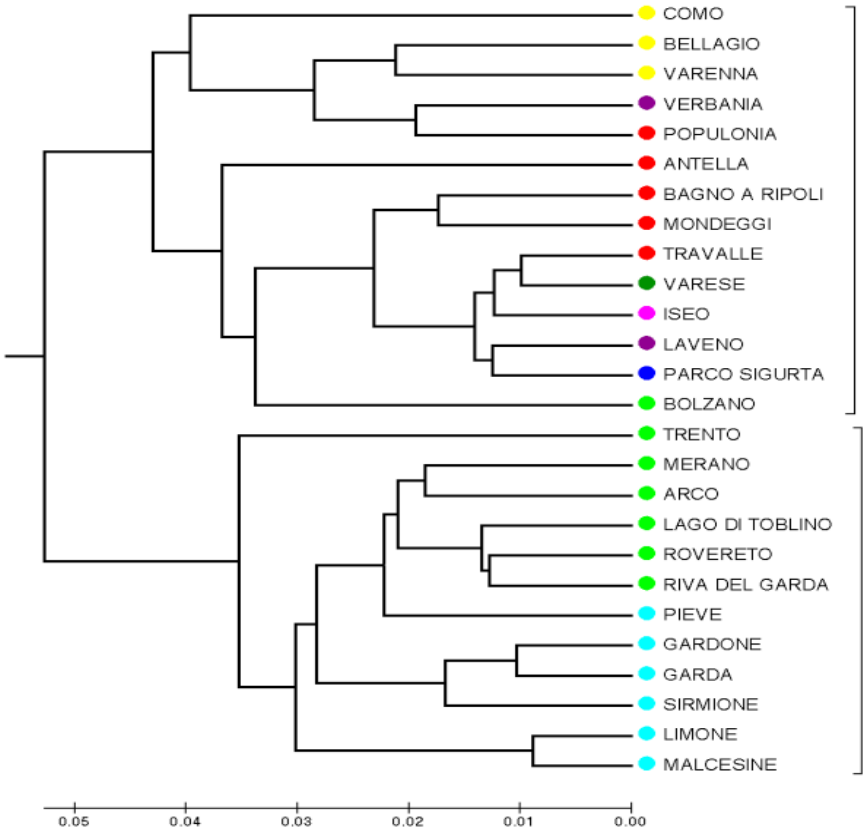


Figure 15: ●yellow: populations from Lake Como (Como, Bellagio, Varenna); ●red: populations from Tuscany (Populonia, Antella, Bagno a Ripoli, Mondeggi, Travalle); ●violet: population from Lake Maggiore (Verbania, Laveno); ●fuchsia: popuation from Lake Iseo; ●dark-green: population from Lake Varese; ●blue: population from Parco Sigurtà; ●Green: populations from Trentino-South Tyrol (Bolzano, Trento, Merano, Rovereto, Lago di Toblino, Riva del Garda Arco); ●sky-blue: populations from Lake Garda (Pieve, Limone, Malcesine, Sirmione, Garda, Gardone).

Using PCA analysis, based on the same matrix, we had a better tri-dimensional representation of genetic relationship between demes (Figure 16). Cluster 1 and cluster 2 were easily discriminated by the first three PCA axes (49,45%, 19,77% and 12,20% of the total variance explained respectively, in total was 81,41%). Cluster 2 turned out to be widely scattered and groups of populations are not strictly defined on the basis of geographic distance, for example: populations from Lake Maggiore and Varese are more similar to the Tuscan populations, while Parco Sigurtà and Iseo stand between Tuscan populations and Trentino- Lake Garda cluster, only partially reflecting their geographic position. Lake Como populations cluster on one side of the graphic while Laveno and Verbania are fairly distant.

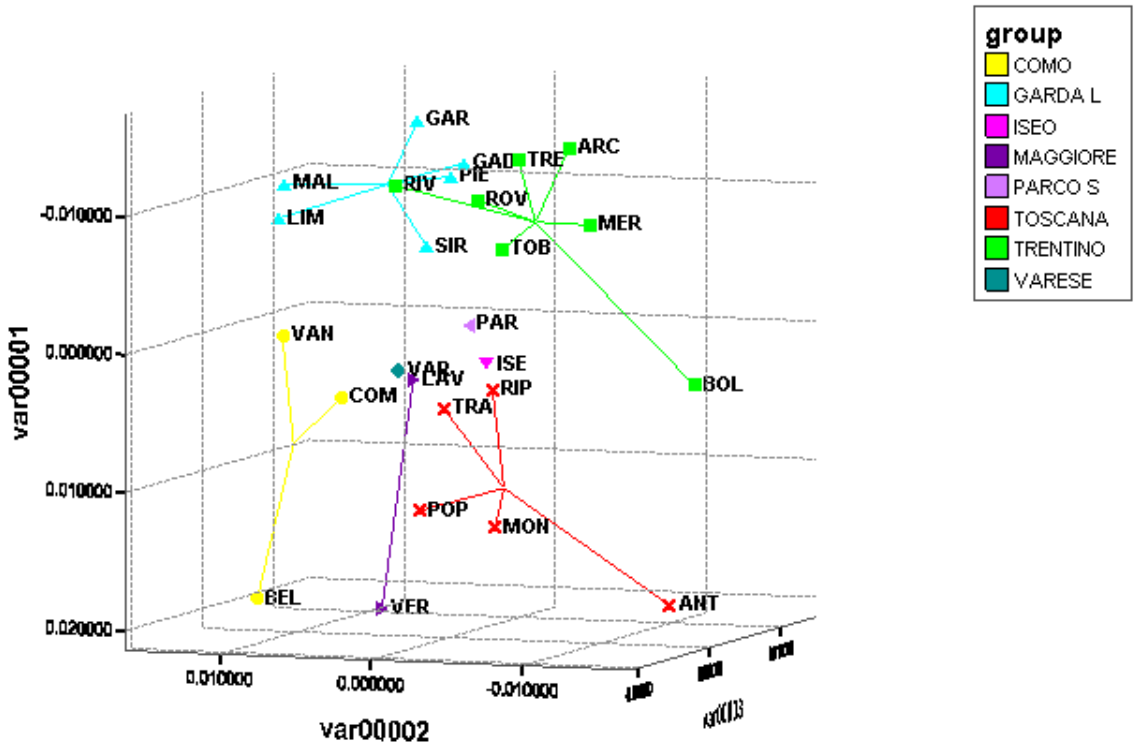


Figure 16: The percentage of variation explained was 49,45% from the first axis, 19,77% from the second axis, 12.20% from the third axis. Populations from the same group are linked with spikes.

Based on both geographical population distribution and PCA results, the populations were divided into three distinct groups that combined genetic and geographic distance: the first, called “North-east” group, included populations from Trentino-South Tyrol, Lake Garda, and Parco Sigurtà; the second, called

"Tuscan" group, included all the populations coming from Tuscany, and the third, called "North-west" group, included all the other populations (from Lake Maggiore, Lake Varese, Lake Como and Lake Iseo; (see Table 1 in Materials and Methods).

Hierarchial AMOVA analyses were undertaken on different combinations of the three distinct groups just defined (Table 16). The first analysis revealed that 3% of total variation was distributed among groups (*Fst* = 0,031; p = 0,001), 4% among populations within groups (*Fsr* = 0,042; P = 0,001) and 93% among individuals within populations. *Fst* was 0,072, (P = 0,001). Other results are summarized in table Table 16.

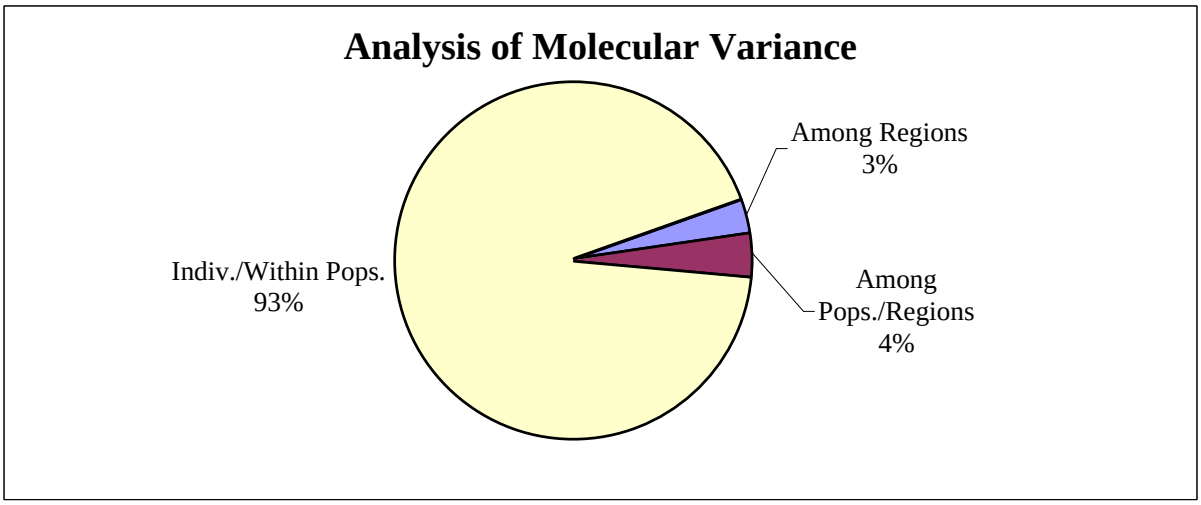


Figure 17: Hierarchical AMOVA analysis between three groups of populations.

AMOVA	a) 3 Groups	b) 2 Groups	c) 2 group	d) 2 group	e)	f)	g)
1° group	North east	North east	North east	Tuscany	North east	North west	Tuscany
2° group	North west	North west	Tuscany	North west			
3° group	Tuscany						
Among groups	3,1%	2,9%	4,2%	0,8%			
Among pops within groups	4,1%	4,2%	4,0%	3,8%	4,4%	4,2%	3,1%
Within pops	92,8%	92,9%	91,8%	95,4%	95,6%	95,8%	96,9%

Table 16: Hierarchical AMOVA: a) between populations divided into three groups: "north east", "north west" and "Tuscany" b) between populations from two groups: "north east" and "north west"; c) between populations from two groups: "north east", and "Tuscany" d) between populations from two groups: "north west" and "Tuscany". AMOVA e) between "north east" populations f) between "north west" populations g) between "Tuscany" populations.

A Mantel test between genetic and geographic distance matrices was performed to test for isolation by distance. The genetic distance values between pairs of populations increased linearly with the natural logarithm of the geographic distance (Figure 18; Mantel test: $r^2 = 0.177$, $P < 0.001$), showing a typical pattern of isolation by distance (Rousset, 1997), explaining only 18% of the total variance.

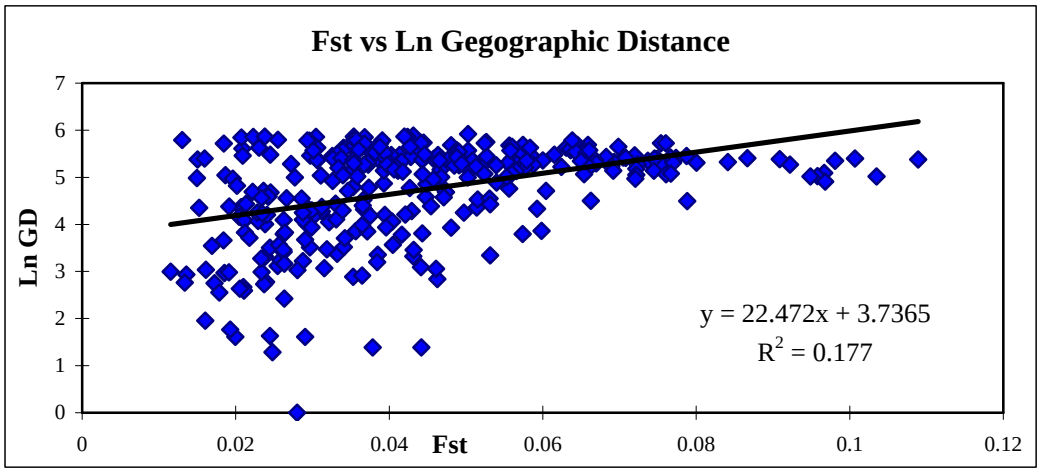


Figure 18: Mantel test between Fst matrix and Ln of the geographic distance matrix. Under isolation by distance, values of the ratios are expected to increase linearly with the logarithm of distance (Rousset, 1997).

Isolation by distance was tested within each group of populations already defined and between couples of groups. 38% of variability inside "north east" group was explained by distance (P<0,05), while the other two groups showed no significant correlations between genetic and geographic distance matrix. There was no isolation by distance between "north east" and Tuscan populations. A more powerful test for assessing spatial vs genetic structure is the spatial autocorrelation analyses (Figure 19). The test showed a positive and significant correlation up to 80 Km and after 330 Km (even if data in the two last distance classes were based on few comparisons), while a negative correlation was present between 170 and 260 Km. On the intermediate class, r was not significant. X intercept (inversion of correlation) was about 130 Km. So, although there are anomalous genetic similarities between certain pairs of populations, some of the differentiations among populations can be explained on the basis of geographic distance.

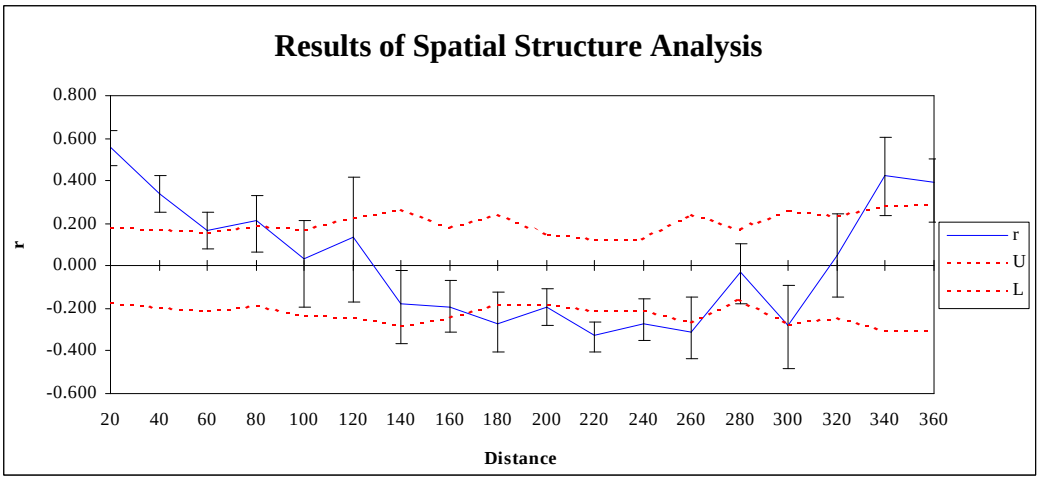


Figure 19: Correlogram obtained for all the population with all loci.

DISCUSSION

4.1 RAPD and SSR analysis

In the first part of this study, a preliminary data on the genetic structure of 11 populations of *C. sempervirens* were investigated with RAPD and SSR markers. The discriminating capacity and informativeness of these two methodologies were compared for genotype identification and genetic diversity analyses.

4.1.1 Within populations analysis

The higher level of expected heterozygosity within Italian populations detected in the cypress by SSR markers rather than with RAPDs highlights the stronger discriminating capacity of the former marker. This result is in accordance with previous studies where SSRs were compared to RAPD (Belaj, 2003; Thomas, 1999; Powell, 1996; Russell, 1997; Pejic, 1998). SSR-derived estimates of within-population variation are usually considerably higher discriminant than dominant marker based estimates (Turpeinen, 2003; Maguire, 2002; Diaz, 2001; Mariette, 2001).

Since RAPDs are a dominant marker type, the presence or absence of a band is defined as representing two alleles at a locus and, therefore, calculation of genetic diversity values should assume Hardy-Weinberg equilibrium. Consequently, these data do not allow the determination of allelic richness, the effective number of alleles or the calculation of fixation indices.

In those cases where Hardy-Weinberg Equilibrium (HWE) is violated, statistical methods using allele frequencies may not be valid and methods that use genotype frequencies should be preferred (Xu, 2002). Estimation of allele frequencies indirectly from RAPD phenotypes observed in diploid tissue could cause download bias in *Hs* and *Ht* while *Gst* could be upwardly biased (Aagard, 1998; Isabel, 1995) when compared with codominant markers.

He and *Ht* obtained on the Italian cypresses using RAPD were in accord with values obtained in others long-perennial woody plants with the same marker (some examples in Table 17), while higher values were detected in *Pinus oocarpa* (Diaz, 2001) and *Pinus Contorta* (Thomas, 1999).

Species	Hs	Ht	I	Ref.
Psedotsuga menizesii	0.22	0.32		Aagard, 1998
Quercus petraea	0.233	0.239		Le Corre, 1997
Picea Abies	0.25			Jeandroz , 2004
Chamaecyparis formosiensis			0.448	Hwnag, 2001
C. taiwanensis			0,409	Hwnag, 2001
Pinus massoniana	0.2451			Peng, 2003

Table 17: RAPD Hs Ht and I found with RAPD in other studied species.

In this study no significant correlation was found between intrapopulation expected heterozygosities originated by RAPD and SSR. RAPD also failed to detect high levels of expected heterozygosity in the Turkish population. In previous studies, correlations between within-diversity estimates calculated within a set of different populations using SSR and AFLP (Mariette, 2001) or SSR and RAPD (Sun, 1998) have not proved significant. Moreover, in a compilation of 307 studies using nuclear DNA markers for evaluating among and within-population diversity in wild angiosperms and gymnosperms, Nymbom (2004) established that RAPD-based data showed no association at all with geographical range (endemic, narrow, regional or widespread populations), whereas both SSR and isozymes produced a significant result, with the lowest values showed for endemic species, followed by narrow, regional and widespread for allozymes or narrow, widespread and regional for SSR. Nevertheless, it is striking and somewhat puzzling that endemics population and species contain only average levels of RAPD variability in spite of showing significantly reduced levels of allozyme and SSR variability.

Gst values, obtained with RAPD in this study, were congruent with the differentiation obtained with SSR even if *Fst* was a bit lower than *Gst*. It also fits with the general observation that woody perennial and outbreeding species maintain most of their variation within populations (Hamrick, 1992; Nimbom, 2000).

4.1.2 Relationship between individuals and AMOVA analysis

PCA analysis showed that both markers were able to distinguish very well between Turkish and Italian populations and revealed a good relationship between populations and their geographical location.

The same difference was detected with AMOVA. Both marker types showed that the majority of differentiation was caused by variation between the Italian and Turkish populations (30% rapd and 32% SSR), which was approximately fivefold and eightfold higher than differentiation among the Italian population. A discrete differentiation was also found between Tuscan and Trentino-Alto Adige groups (Φ_{PT} RAPD = 10,1% and Φ_{PT} SSR = 6,8%). RAPD found a moderate differentiation among Trentino-Alto Adige populations (6,1%), while SSR found a lower but significant value (3,3%). Both markers didn’t found

significant differentiation among Tuscan populations. Previous studies found that estimates of among population variation were almost identical between SSR and the dominant markers (Sun, 1999; Thomas, 1999; Mariette, 2001; Fahima, 2002; Turpeinen, 2003). Large discrepancies have occasionally been reported (Maguire, 2002; Lowe, 2003) but appear to involve a very low number of either samples or loci.

4.1.3 Divergence at the population level

The comparison of pairwise genetic divergence between populations using SSR and RAPD markers showed a good agreement. The Turkish population was excluded from the analysis to avoid its overinfluence. Correlation coefficients between estimates obtained with the two types of markers were fairly high for both genetic distance (Nei GD, r = 0.708) and genetic divergence (Φ_{PT}, r = 0.594). The greater sensitivity of SSRs to population divergence may reflect a greater power of these markers for detecting differences among populations or a grater number of mutations of SSR than RAPD.

Geographically more distant populations were more dissimilar than populations from the same region: correlation coefficients between geographic and genetic distance matrices obtained with the two types of markers were positive and fairly high.

The two markers showed a high degree of similarity in dendrogram topologies, though with one difference in the positioning of Bolzano population in the main groups. The dendrograms reflect relationships among most of the populations, depending upon their area of provenance.

Both RAPDs and SSR, were highly efficient in detecting genetic spatial structure in the cypress, but SSR was more powerful to detect intrapopulation and interpopulation differentiation and genetic distance. To conclude, the codominant nature of SSRs, their better resolutive power and their simple detection have made them the marker of choice for genetic population studies also in the Cypress.

4.2 SSR analysis

Among plants, conifers stand out as wind-pollinated, long-lived, highly fecund taxa. All of these traits were found to be associated with high levels of genetic variation in two different multivariate studies of many species of plants using protein-based markers (Hamrick, 1989) and RAPD markers (Nymbom, 2000). Wind pollination increases gene flow and neighbourhood size. Studies of geographic variation in conifers with genetic markers have revealed that most have high levels of genetic variation within populations and little differentiation among populations. Typically, 90% or more of the total genetic variation is found within populations (Bucci, 1995; Hamrick, 1989).

Only isoenzymes have been employed previously to quantify the genetic diversity within *C. sempervirens* germplasm collections coming from Mediterranean eastern region (Turkey, Israel and Greece; Raddi, 1999; Schiller, 1997; Sumer, 1987). These markers revealed a large genetic diversity, allele richness, enzyme polymorphism and a relatively low interpopulation diversity. The high genetic diversity observed in these populations seems to exclude past bottleneck or cycles of extinction and recolonization; rather it suggest that the extant natural populations in eastern Mediterranean are remnants of once widespread populations (Raddi, 1999). The phenomenon of relatively high heterozygosity in *C. sempervirens* L. might be due also to a special type of serotinous cone which stay alive within trees canopy for several years (Korol, 1997; Lev-Yadun, 1995): this phenomenon might possibly be one of the ways of transfer, through out the millennia and in spite of catastrophic events, of the high genetic diversity that has been maintained in the species notwithstanding of the very disjoint areas of distribution and relatively small population (Korol, 1997).

4.2.1 Allelic variation

The 8 loci analyzed differed greatly in the level of variability, from loci with few variants for allele to others with abundant polymorphism for alleles widely differing in size. Most alleles were separated by single or multiple unit sizes, but some alleles differed by a single base or three bases, rather than by two bases as expected for this dinucleotide repeat. This may have arisen from an ancestral insertion/deletion of one or three bases somewhere between the two priming sites for this locus, with subsequent evolution by the addition or deletion of dinucleotide units.

The microsatellites allele frequency distributions of the eight SSR was very different from a distribution “under stepwise mutation model”, and seemed affected by drift and selection that have removed some alleles sizes while randomly selecting others to be greatly over-represented. Less common alleles could represent introduction of foreign plants or more recent mutations.

4.2.2 Analysis of genetic population diversity at each locus

Comparing *He* and *Ho* calculated on loci developed for *C. sempervirens* L. by Sebastiani (2005) with the results obtained in this study, several differences were revealed. The heterozygosity expected and observed in the Italian populations was much lower than the one previously estimated by Sebastiani: especially Cyp 52 and Cyp 293 showed very low values. Instead the Turkish population had values more similar to the ones obtained in the primer notes (Sebastiani, 2005). This may indicate that the Italian populations are less variable than the "potentiality" of this species.

F estimates at each locus are large enough to suggest not only high inbreeding but also the existence of null alleles, especially for cyp 101 and cyp 139 locus. Also Sebastiani, (2005) found significative deviations from HWE. When they estimated the frequency of null alleles with the Software package CERVUS 2.0 (Marshall, 1998) found that it was higher than 5% for cyp 52, cyp 139, cyp 258 and cyp 293 (0.19, 0.06, 0.09 and 0.07, respectively). However, null alleles seem to be actually quite frequent in conifer species. This has already been reported in a study on *Pinus radiata* (Fisher, 1998), which pointed out the high frequency of null alleles in microsatellites of this species. It seems also to be the case in other species such as *Picea abies* (Derory, 2002). The use of microsatellites may therefore lead to underestimation of heterozygosity and allelic richness in conifer species.

The Probability of Identity for increasing locus combinations over all the dataset was zero, demonstrating that selected loci are reliable for genetic tagging and putative clone identification. In a previous study on *Populus* species, nine SSR loci with number of alleles ranging from two to nine were used for clonal identification (Leena, 2005). This should be very useful for future study on clonal identification. Indeed, several clones were found not only within populations or between neighbouring populations, but also in very distant populations, suggesting the enormous influence of human action on this species. Human activities have repercussions on woods gene diversity. Such transfer of seeds or plants on different local stands can also result in an increased genetic diversity in the next generations, but frequently with negative consequences for local adaptation (Ledig, 1992).

4.2.2 Intrapopulation differentiation

Several private alleles were detected among the Italian populations, while other rare alleles were in common with the Turkish population. Some of these are very different in size from the most frequent Italian alleles and are not probably originated by a direct mutation, indicating that probably some individuals from Turkey were occasionally introduced and mixed with Italian populations.

The mean expected genetic diversity within Italian populations found with SSR were on the same order than the one found with isoenzymes polymorphic

loci in Turkish, Greek and Israeli populations of *C. sempervirens* where interpopulation variability averaged respectively H_{exp} = 0.350, 0.400 and 0.479. (Raddi, 1999; Schiller, 1997; Sumer, 1987).

All the Italian populations showed an excess of homozygosity. Deviation from Hardy-Weinberg equilibrium may be due to several reasons: presence of null alleles, inbreeding, small population size that results in random sampling errors and unpredictable genotype frequencies, selection of one or a combination of genotypes. Selection could be explained by the fact that *C. sempervirens* was introduced in Italy. The erect form (var. pyramidalis) was most probably selected for ornamental and short rotation planning (Korol, 1997), and then spread all over the Mediterranean in ancient times. Also the common seed collection from few mother plants to ensure the nursery germination before planting instead of the natural regeneration could create some bottleneck effect and so higher homozygosity.

A significant negative correlation was found between observed heterozygosisty and longitude indicating that in general southern populations had higher *Ho* values than northern populations while a positive correlation was found between *F* and longitude, indicating that southern populations seem to have more heterozygotes than northern ones.

Heterozygote deficiencies can be caused by inbreeding (assortative mating), selection against heterozygotes, the Wahlund effect (population subdivision into separate breeding units) or selection induced microscale differentiation (Sproule, 1996; Epperson, 1990; Brown, 1979).

4.2.3 Interpopulation differentiation

There are two major ways to quantify the degree of population structure: with or without underlying biological models. Distance methods with no biological assumptions, also known as geometric distances, analyze the data as a set of numbers without making any biological assumptions. Approaches could include principal components analyses (PCA), Euclidean distances or somewhat more complex geometric distances. Many of these will allow us to create a sort of abstract "map" of the populations in one, two, three or more dimensions.

Several distance measures incorporate assumptions about the importance of drift and mutation as forces of change: measures that use a stepwise mutation model (SMM; Figure 20 a), specifically developed for microsatellites, assumes that alleles more similar in size will presumably be more closely related (*Rst;* Slatkin 1995). Other measures use the Infinite Allele Model (IAM: Figure 20 b) where every new mutation is assumed to give rise to a new distinguishable allele with the same probability (e.g. Nei's distance -Nei 1972, 1978- and *Fst* -Wheir, 1984). IAM, with its neglected importance of mutation, seems to work better for small population size, high potential of drift and "missing steps" or allele size (Balloux, 2002). For the above reasons our AMOVA analysis and our pairwise population difference were based on IAM and *Fst*.

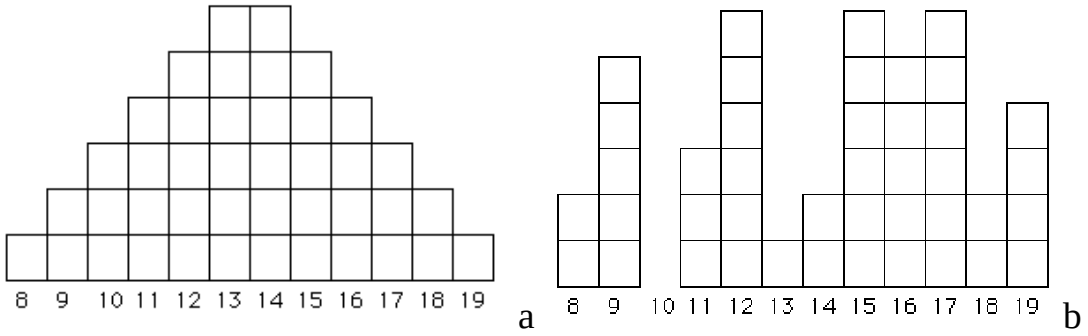


Figure 20: a) A microsatellite allele frequency distribution under a strict ***stepwise mutation model*** *(SMM). The* X*-axis shows the number of repeat units (e.g., AC_8 to AC_{19}), while the* Y*-axis shows the number of alleles. Starting with either a 13 or 14 repeat chain as the ancestor, we have tended to accumulate more alleles at sizes close to the starting point because of equal likelihood of additions or subtractions. b) An allele frequency distribution that has been greatly affected by drift and may better fit an* ***infinite-alleles model*** *(IAM). Even if the mutations that generated the original variation did occur in stepwise fashion, drift has removed some allele sizes (e.g., the 10-repeat category) while randomly selecting others to be greatly over-represented (e.g., 12, 15 and 17).*

Nei's genetic distance matrix and a pairwise *Fst* genetic distance matrix were performed on our dataset. While *Rst* was not used in this study. A dendrogram was constructed based on Nei's genetic distance between populations with the UPGMA method and a PCA was performed on the same matrix to have better three-dimensional representation of relationship between populations.

Populations were divided into two main clusters that roughly reflect geographic distribution: the first contains all the populations from Lake Garda and from Trentino-South Tyrol except Bolzano. The second cluster contains all the other populations and Bolzano. This indicates that exchange of genetic material exists, probably the five clones detected in Bolzano came from Tuscany. When these clones were excluded from analysis, Bolzano formed an independent clade separated from the other two groups that remained substantially unchanged. It is possible that some collected samples from Bolzano were originated from plants coming from other regions of the Austro-Hungarian Empire such as Istria or Dalmatia, where cypresses grow abundantly.

The second cluster is more scattered than the first one and not always genetic population distances reflect their geographic positions. For example: the populations from Lake Maggiore and Varese are more similar to the Tuscan populations, while Parco Sigurtà and Iseo stand between the Tuscan populations and Trentino-South Tyrol - Lake Garda cluster, only partially reflecting their geographic position. Probably a mix of old "autoctonous" plants and new imported materials composes population of Parco Sigurtà.

Lake Como populations cluster on one side of the graphic while Laveno and Verbania (Both coming from Lake Maggiore) are fairly distant.

Amova

Highly outcrossing and long-lived perennial species, such as conifers and *C. sempervirens*, typically show very high levels of within-population genetic variation and relatively less differentiation among populations (Nybom, 2004; Nybom, 2000; Hamrick, 1996). This value is higher when observed in species where gene flow occurs freely. For example in a study of the genetic structure of Oak (Quercus spp.) trees in the Mediterranean region (Petit, 1993) was found that only 7% of the genetic diversity could be explained by between population differences, and this was accounted by the contiguous nature and frequent gene flow among the populations. Nevertheless species with extremely small and disjoint geographic ranges are expected to diverge by a larger extent due to a lowered interpopulation gene flow, larger influence of drift and influence of selection.

Considering both genetic distance and geographic locations, the Italian populations were divided into three distinct groups: “North-east” “Tuscan” and “North-west” group. The highest among-group differentiation was found between “Tuscan” and “North-east”, as indicated by *Fst v*alues: 0.042. Group “North-east” was significantly differentiated from the group “North west” of populations (*Fst* = 0.029), whereas the differentiation between “North west” and Tuscan group of the populations (*Fst* = 0.008) had a much lower value. Differentiations were highly significant for all pairs of groups (all cases $P < 0.001$).

May be it is not irrelevant that historical data actually show that the cypress was probably present in Trentino-South Tyrol starting from XIV century (Turri, 2004). There are no information before that century about the presence of the cypress in Trentino and Lake Garda, even tough it cannot be excluded. Moreover, the presence of cypresses in the north of Italy is documented todate back at least to a thousand years ago: plants of 1000 years of age were still living until the beginning of the past century (Cormio, 1935; Rossi, 1994). This period of time could have been sufficient to create a lowly but significantly differentiated group of local populations even if genetic differentiation and drift among groups was probably inflated by plantations of genetic material from different stands.

Spatial analysis

The *Fst* for pairs of populations increased linearly with the natural logarithm of the geographical distance. Under isolation by distance, values of the ratios are expected to increase linearly with the logarithm of distance (Rousset, 1997). Similar significant values were found in other studies with SSR between *Fraxinus excelsior* populations ($r^2 = 0.15$; Heuertz, 2001).

Autocorrelations analyses clarified the distribution of this correlation: for populations distant under 80 Km there was a good significant correlation between genetic and geographic distance, while between 130 and 260 this correlation was negative. It is probably due to genetic similarities found between distant populations.

4.3 Conclusions

The results of this study have demonstrated that RAPD technique can be applied to measure the degree of variability between *C. sempervirens* populations while it is not so useful to find within population variation. RAPD and SSR show similar values for estimation of between-population variation and spatial structure in the populations analyzed.

SSR is a very good tool to investigate genetic variability of *C. sempervirens*.

The set of used markers could be used to identify putative clones.

Italian cypresses have reduced genetic variability compared with the one expected from this species. However the genetic variability detected was similar to the one found for both natural and artificial populations of conifer species.

Trentino-South Tyrol and Lake Garda show a good clustering spatial structure in spite of the presence of clone planted material.

To increase knowledge about the cypress genetic structure we also recommend further analysis using co-dominant molecular markers to have a better resolution power.

REFERENCES

Aagard JE, Krusovskii KV, Strauss SH. 1998. RAPDs and allozyme exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir. Heredity 81, 69-78.

Akkaya MS, Bhagwat AA, Cregan PB. 1992. Length polymorphism of simple sequence repeats DNA in soybean. Genetics 132, 1131-1139.

Allnutt TR, Newton AC, Premoli A, Lara A. 2003. Genetic variation in the threatened South American conifer *Philgerodendrum uviferum* (Cupressaceae), detected using RAPD markers. Biological Conservation 114, 245-253.

Avise JC. 1994. Molecular markers, natural history, and evolution. Chapman and Hall, New York, NY.

Axelrod DI. 1958. Evolution of the Madro Tertiary geoflora. Botanical Review 24, 433-509.

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

Belaj A, Satovic Z, Rallo L, Trujillo I. 2002. Genetic diversity and relationship in Olive (Olea europea L.) germplasm collections as detemined by Randomly Amplified Polymorphyc DNA. Theoretical and Applied Genetics 105, 648-644.

Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I. 2003. Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. Theoretical and Applied Genetics 107, 736–744.

Bergmann F. 1991. Isozyme gene markers. In: Muller-Starck G, Ziehe M (eds) Genetic variation in European populations of forest trees. Sauerlander's Verlag, Frankfurt, 67-78.

Boscherini G, Morgante M, Rossi P, Vendramin, GG. 1994. Allozyme and chloroplast DNA variation in Italian and Greek populations of *Pinus leucodermis*. Heredity 73, 284-290.

Boyle TJB. 1992. Biodiversity in Canadian forest: current status and future challenges. Forestry Chronicle 68, 444-453.

Brown AHD. 1979. Enzyme polymorphism in plant populations. Theoretical Population Biology 15, 1-41.

Bucci G, Menozzi P. 1993. Segregation analysis of Random Amplified Polymorphic DNA (RAPDs) in *Picea abies* Karst. Molecular Ecology 2 (4), 227-232.

Bucci G, Menozzi P. 1995. Genetic Variation of Rapd Markers in a *Picea Abies* Karst. Population. Heredity 75, 188-197.

Cormio R. 1935. Il cipresso di Somma Lombardo. Tipografia del “Popolo d’Italia”. Milano.

Crouch JH, Ortiz R, Crouch HK, Jarret RL, Ford-Lloyd BV, Howel, EC, Newbury HJ. 2000. Utilization of molecular genetic techniques in support of plantain and banana improvement. Acta Horticulturae 540, 110-116.

Cruzan MB, 1998. Genetic markers in plant evolutionary ecology. Ecology 79, 400-412.

Demesure B, Comps B, Petit RJ. 1996. Chloroplast DNA phylogeography of the common beech (*Fagus Sylvatica* L.) in Europe. Evolution 50, 2515-2520.

Derory J, Mariette S, Gonzaléz-Martínez SC, Chagné D, Madur D, Gerber S, Brach J, Persynd F, Ribeiroe MM, Plomion C. 2002. What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenance certification strategies? Annals of Forest Science 59, 699-708.

Diaz V, Muniz LM, Ferrer E. 2001. Random Amplified Polymorphic DNA an Amplified Fragment Length Polymorphism assessment of genetic variation in Nicaraguan populations of *Pinus oocarpa*. Molecular Ecology 10, 2593-2603.

Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. Molecular Ecology 11, 2571-2581.

Double MC. 2005 Dispersal, philopatry and infidelity: dissecting local genetic structure in superb fairy-wrens (*Malurus cyaneus*). Evolution 59, 625-635.

Ducrey M, Brofas G, Andreoli C, Raddi P. 1999. Cypress. A practical handbook. Ed. Du Cross T. Studio Leonardo, Florence, Italy. 9-24.

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A. 1996. Microsatellite libraries enriched for several microsatellite sequences. Proceedings of the National Academy of Science of the United States of America. 89, 3419-3223.

Epperson BK. 1990. Spatial patterns of genetic variation within plant populations. Plant Population Genetics, Breeding, and Genetic Resources, 229-253.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 179-191.

Fahima TR, Röder MS, Wendehake K, Kirzhner VM, Nevo E. 2002. Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theoretical and Applied Genetics 104, 17–29.

Fineschi S, Taurchini D, Villani F, Vendramin, GG. 2000. Chloroplast DNA polymorphism reveals little geographical structure in Castanea sativa Mill. (Fagaceae) throughout southern European countries. Molecular Ecology, 9, 1495-1503.

Fisher PJ, Richardson TE, Gardner RC. 1998. Characteristics of single- and multi-copy microsatellites from *Pinus radiata*. Theoretical and Applied Genetics 96, 969-979.

Fornari B, Taurchini D, Villani F. 1999. Genetic structure and diversity of two Turkish *Castanea sativa* Mill. populations investigated with isozyme and RAPD polymorphisms. Journal of Genetics and Breeding 53, 315-325.

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

Garcia-Mas J, Oliver M, Gomez-Paniagua H, Vicente MC. 2000. Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. Theoretical and Applied Genetics 10, 860-864.

Giannelli L, Bezzini L. 2002. Il cipresso: storie e miti di terre toscane. Firenze, Ed. Scramasax. pp. 151.

Graniti A. 1998. Cypress Canker: A Pandemic in Progress. Annual Review of Phytopathology 36, 91-114

Grattapaglia D, Sederoff R. 1994. Genetic linkage map of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137, 1121-1137.

Greenwood MS, Volkaert HA. 1992. Morphophysiological traits as markers for the early selection of conifer genetic families. Canadian Journal of Forest Research 22, 1001-1008.

Guo SW, Thompson EA. 1992 a. Performing the exact test of Hardy-Weinberg proprtion for multiple alleles. Biometrics 48, 361-372.

Guo SW, Thompson EA. 1992 b. A Monte Carlo method for combines segregation and linkage analysis. America Journal of Human Genetics 51, 111-1126.

Gupta PK, Balyan HS, Sharma PC. 1996. Microsatellites in plants: a new class of molecular markers. Current Science 70, 45-54.

Hamrick JL, Godt MJW. Allozyme diversity in plant species. 1989. In: Plant Population Genetics, Breeding and Genetic Resources (eds Brown AHD Clegg MT Kahler AL Weir BS), pp. 43–63. Sinauer, Sunderland MA.

Hamrick JL, Godt M.J.W, Sherman-Broyles SL. 1992. Factors influencing levels of genetic diversity in woody plant species. New Forests 6, 95-124.

Hamrick JL, Godt MJW. 1996. Effects of life history traits on genetic diversity in plant species. Philosophical transactions of the Royal Society in London. Series B, 351, 1291-1298.

Hayden MJ and Sharp PJ. 2002. Sequence tagged microsatellite profiling (STMP): improved isolation of DNA sequence flanking targhet SSRs. Nucleic Acid Research 30, 1-5.

Herlan JR. 1975. Our vanishing genetic resources. Science 188, 618-621.

Heuertz M, Hausman JF, Tsvetkov I, Frascaria-Lacoste N, Vekemans X. 2001. Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.). Molecular Ecology 10, 1615–1623.

Hilfiker K, Holderegger R, Rotach P, Gugerli F. 2004. Dynamics of genetic variation in *Taxus baccata*: local versus regional perspectives. Canadian Journal of Botany 82, 219-227.

Huff DR, Peakall R, Smouse PE. 1993. RAPD variation within and among natural populations of outcrossing buffalograss *Buchloe dactyloides* (Nutt) Engelm. Theoretical and Applied Genetics 86, 927-934.

Huxel GR, Hastings A. 1999. Habitat loss, fragmentation and restoration. Restoration Ecology 7, 309-315.

Hwnag SY, Lin HW, Kuo YS, Lin TP. 2001. RAPD variation in relation to population differentiation of *Chamecyparis formosensis* and *Chamaecyparis taiwanensis*. Botanical Bulletin of Academia Sinica 42, 173-179.

Isabel N, Beaulieu J, Bousquet J. 1995. Complete congruence between gene diversity estimates derived from genotypic data at enzyme and Random Amplyfied Polymorfic DNA loci in black spruce. Proceedings National Academic Science U.S.A. 81, 69-78.

Isik K, Yaltirik F, Akesen A. 1997. The interrelationship of forests, biological diversity and the maintenance of natural resources. S.A. Dembner. An International Journal of Forestry and Forest Industries. 48, 19-29.

Jarne P, Lagoda PJG. 1996. Microsatellite from molecules to populations and back. Trends in Ecology & Evolution 11,424-429.

Jasieniuk M, Maxwell BD. 2001. Plant diversity: new insights from molecular biology and genomics technologies. Weed Science. 49, 257-265.

Jasso JM. 1970. Impact of selviculture on forest gene resources. Unasylva 24, 70-75.

Jeandroz S, Collignon AM, Favre JM. 2004. RAPD and mt DNA variation among autochthonous and planted populations of *Picea abies* from the Vosges Mountains (France) in reference to other French populations. Forest and Ecology Management. 197, 225-229.

Kaundun S, Park YG. 2002. Genetic structure of six Korean tea populations as revealed by RAPD PCR markers. Crop Science 42, 594-601.

Kayacik H, Yaltrik F, Elicin G. 1979. The floristic composition of the Italian cypress (*C. sempervirens* L.) forest within the Antalyas region in Turkey. Webbia 34, 145-153.

Kearsey MJ, Farquhar AGL. 1998. QTL analysis in plants; where are we now?. Heredity 80, 137-142.

Korol L, Kara N, Isik K, Schiller G. 1997. Genetic differentiation among and within natural and planted *Cupressus sempervirens* L. eastern Mediterranean populations. Silvae Genetica 46, 151-155.

Lanham PG, Brennan RM. 1999. Genetic Characterization of gooseberry (*Ribes grossularia* subgenus grossularia) germplasm using RAPD, ISSR and AFLP markers. Journal of Horticultural Science and Biotechnology 74, 361-366.

Le Corre V, Dumolin-Lapegue S, Kremer A. 1997. Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: the role of history and geography. Molecular Ecology 6, 519-529.

Ledig FT. 1992. Human impacts on genetic diversity in forest ecosystem. Oikos 63, 87-108.

Leena IS, Latva-Karjanmaa TB. 2005. Clone identification and clonal structure of the European aspen (*Populus tremula*). Molecular Ecology 14, 2851–2860

Lev-Yadun, S. 1995. Living serotinous cones in *Cupressus sempervirens*. International Journal of Plant Sciences 156, 50-54.

Libby WJ, Stettler RF, Seitz FW. 1969. Forest genetics and forestree breeding. Annual Review of Genetics 3, 469-494.

Loveless MD, Hamrick JL, 1984. Ecological determinants of genetic structure in plant populations. Annual Review of Ecology Evolution and Systematics 15, 65-95.

Lowe AJ, Jourde B, Breyne P. 2003. Fine-scale genetic structure and gene flow within Costa Rican populations of mahogany (*Swietinia macrophylla*). Heredity 90, 268-275.

Lynch M, Milligan BG. 1994. Analysis of population genetic structure with RAPD markers. Molecular Ecology 3, 91-99.

Maguire TL, Peakall R, Saenger P. 2002. Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. Theoretical and Applied Genetics 104, 388-398.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27, 209-220.

Mariette S, Chagné D, Lézier C. 2001. Genetic diversity within and among *Pinus Pinaster* populations: comparisons between AFLP and microsatellite markers. Heredity 86, 469-479.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7, 639-655.

Michalakis Y, Excoffier L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142, 1061-1064.

Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, Waugh R. 1997. Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Molecular Breeding 3, 127-136.

Morgante M, Olivieri AM. 1993. PCR-amplification microsatellites as markers in plant genetics. Plant Journal 3, 175-182.

Munthali M, Ford-Lloyd BV, Newbury HJ. 1992. The Random Amplification of Polymorphic DNA for fingerprinting plants. Pcr Methods and Applications 1, 274-276.

Namkoong G. 1992. Biodiversity - issues in genetics, forestry and ethics. Forestry Chronicle 68, 438-443.

Nei M. 1972. Genetic distance between populations. American Naturalist 106, 283-392.

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proceedings National Academic Science U.S.A. 70, 3321-3323.

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

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA. 1999. Molecular phylogeography, intraspecific variation and the conservation of tree species. Tree 14, 140-145.

Nybom H, Bartisch I. 2000. Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. Perspectives in Plant Ecology, Evolution and Systematics 3, 93-114.

Nybom H. 2004. Comparison of different nuclear DNA markers for estimating inraspecific genetic diversit in plants. Molecular Ecology 13, 1143-1155.

Pancoresi A, Raddi P, Andreoli C, Ramos P, Xenopoulos S, Caetano F, Pinto-Ganhao J. 1999. Cypress. A practical handbook. Ed. Du Cross T. Studio Leonardo Florence, Italy. 55-74.

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA, 1998. What molecules can tell us about populations: choosing and using a molecular marker. Ecology 79, 361-382.

Peakall R, Smouse PE, Huff DR. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Molecular Ecology 4, 135-147.

Peakall R, Ruibal M, Lindenmayer DB. 2003. Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. Evolution 57, 1182-1195.

Peakall R, Smouse PE. 2005. GenAlEx: Genetic Analysis in Excel. Population genetic software for teaching and research.

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplik V, Castiglioni P, Taramino G, Motto M. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. Theoretical and Applied Genetics 97, 1248-1255.

Peng SL, Li QF, Wang ZF, Wang DP. 2003. Genetic diversity of *Pinus massoniana* revealed by RAPD markers. Silvae Genetica 52, 60-63.

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Laroche A, Scoles G, Molnar SL, Fedak G, 1993. Reproducibility of Random Amplified DNA, RAPD analysis among laboratories. PCR Methods and Applications 2, 341-345.

Petit RL, Kremer A, Wagner DB. 1993. Geographic structure of chloroplast DNA polymorphism in European oaks. Theoretical and applied genetics 87, 122-188.

Petit RJ, Csaikl UM, Bordacs S, Burg K, Coart E, Cottrell J, Dam B, Deans JD, Dumolin-Lapegue S, Fineschi S, Finkeldey R, Gillies A, Glaz I, Goicoechea PG, Jensen JS, Konig AO, Lowe A. J, Madsen SF, Matyas G, Munro RC, Olalde M, Pemonge, MH, Popescu F, Slade D, Tabbener H, Taurchini D. (Et Al). 2002. Chloroplast DNA variations in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. Forest Ecology and Management 156, 103-114.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2, 225-238.

Raddi S, Sümer S. 1999. Genetic diversity in natural *C. sempervirens* L. populations in Turkey. Biochemical Systematics and Ecology 27, 799-814.

Rafalski A, Gidzinska M, Wisniewska I. 1996. PCR-based systems for evaluation of relationships among maize inbred. Proceedings of the XVIIth conference on genetics, biotechnology and breeding of maize and sorghum held at Thessaloniki, Greece, 20-25 October 1996. 106-111.

Rafalski A, Tingey SV. 1993. Genetic diagnostics in Plant Breeding: RAPDs, Microsatellites and Machines. Trends in Genetics 9, 275-280.

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

Richardson SD. 1970. Gene pools in forestry. Genetic resources in plants - their exploitation and conservation. Oxford, Eds. Frankel, Bennett. 353-365.

Riesenberg LH, 1996. Homology among RAPD fragments in interspecific comparisons. Molecular Ecology 5, 99-105.

Rogers DL, Ledig FT. 1996. The status of temperate North American forest genetic resources. University of California. Genetic Resources Conservation Program Report, 16, xvi + pp. 85.

Rossi A. 1994. Antiche località di Somma Lombardo. Antonio Ferrario. Industria grafica Gallarate.

Rousset F. 1997. Genetic differentiation and estimation of gene flow from F- statistics under isolation by distance. Genetics 145, 1219-1228.

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theoretical and Applied Genetics 95, 714-722.

Sant Santini A, Casini N, Panconesi A, Lonardo V. 1994a. Environmental effect on plant morphology and growth of some cypress clones and possible relation to *Seiridium cardinale* infection. Monti e Boschi 45, 42-48.

Santini A, Casini N, Panconesi A, Lonardo V, Nembi V. 1994b. Responses to *Seiridium cardinale* inoculation of some common cypress clones at two locations in Italy. Italia Forestale e Montana 49, 389-400.ini A, Lonardo V. 2000. Genetic variability of the 'bark Canker Resistance' character in several natural provenances of *Cupressus sempervirens*. Forest Pathology 30, 87-96.

Schiller G, Leonid K. 1997. Electrophoretic analysis of diversity within *C. sempervirens* L. growing in Israel. Israel Journal of Plant Science 45, 1-8.

Schneider S, Roessli D, Excoffier L. 2000. Arelequin ver. 2.000: a software for population genetic data analysis.

Sebastiani F, Buonamici A, Fineschi S, Racchi ML, Raddi P, Vendramin GG. 2005. Novel polymorphic nuclear microsatellites in *C. sempervirens* L. Molecular Ecology Notes 5, 393-395.

Shannon CE, Weavwe W. 1949. The matematica theory of communication. Urbana, University of Illinois press.

Shrestha MK, Golan-Goldhirsh A, Ward D. 2002. Population genetic structure and the conservation of isolated populations of *Acacia raddiana* in the Nagev Desert. Biological Conservation, 108, 119-127.

Skroch P, Nienhuis J. 1995. Impact of scoring error and reproducibility of RAPD data on RAPD based estimates of genetic distance. Theoretical and Applied Genetics 91, 1086-1091.

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

Smouse PE, Long, JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology 35, 627-632.

Smouse PE, Long JC. 1992. Matrix correlation analysis in anthropology and genetics. Yearbook of Physical Anthropology 35, 187-213.

Smouse PE, Peakall R. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82, 561-573.

Sproule AT, Dancik BP. 1996. The mating system of black spruce in north-central Alberta, Canada. Silvae Genetica 45, 159-164.

Staub, JE, Danin-Poleg Y, Fazio G, Horejsi T, Reis N, Katzir N. 2000. Comparative analysis of Cultivated melon groups (*Cucumis melo* L.) using Random Amplified Polymorphic DNA and Simple Sequence Repeat markers. Euphytica 115, 225-241.

Sumer S. 1987. The Distribution of Cypress (*Cupressus* L.) In Turkey and the current status in its pests and diseases, especially cypress canker disease. Istanbul Universities Orman Fakultesi Dergisi. Seri a 37, 46-66.

Sun G-L, Díaz O, Salomon B, von Bothmer R. 1999. Genetic diversity in *Elymus caninus* as revealed by isozyme, RAPD, and microsatellite markers. Genome 42, 420-431.

Szmidt AE, Wang X, Lu M. 1996. Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* L. using aploid tissue analysis. Heredity 76, 412-420.

Tautz D, Trick M, Dover GA. 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature 322, 652-656.

Thomas BR, Macdonald SE, Hicks M, Adams DL, Hodgetts RB. 1999. Effects of reforestation methods on genetic diversity of lodgepole pine: as assessment using microsatellite and Randomly Amplified Polymorphic DNA markers. Theoretical and Applied Genetics, 98, 793–801.

Turpeinen T, Vanhala T, Nevo E, Nissilä E. 2003. AFLP genetic polymorphism in wild barley (*Hordeum spontaneum*) populations in Israel. Theoretical and Applied Genetics 106, 1333-1339.

Turri D, La Porta N. 2004. History and landscape in Trentino: the case of *C. sempervirens* L. History and sustainability. Napoli, Stamperia M Armano srl. pp. 131-140.

Vendramin GG, Michelozzi M, Lelli L, Tognetti R. 1995. Genetic Variation in *Abies nebrodensis*: a case study for a highly endangered species. Forest Genetics 2, 171-175.

Vitousek PM, Walker LR, Whiteaker LD, Mueller-Dombois D, Matson PA. 1987. Biological invasion by *Myrica faya* alters ecosystem development in Hawaii. Science 238, 802-804.

Vroh Bi I, Jardin PD, Mergeai G, Baudoin JP. 1997. Optimization and application of RAPD (Random Amplified Polymorphic DNA) in a recurrent selection programme of cotton (Gossypium Spp.). Biotechnologie, Agronomie, Societe et Environnement 1, 142-150.

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

Welsh J, McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acid Research 18, 7213-7218.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA plymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18, 6531-6535.

Williams CG, Neale DB, 1992. Conifer wood quality and marker-aided selection, a case study. Canadian Journal of Forest Research 22, 1009-1017.

Wu RL, O'malley DM, Mckeand SE. 1999. Understanding the genetic architecture of a quantitative trait in gymnosperms by genotyping haploid megagametophytes. Theoretical and Applied Genetics 99, 1031-1038.

Xenopoulos S, Andreoli C, Pancoresi A, Pinto Ghnhao J, Tuset JJ. 1990. Importance of cypress. In: Ponchet Juornal (Ed). Progress in EEC research on cypress disease. AGRIMED, Antiber, France. 1-13.

Xu J, Turner A, Little J, Bleecker FR, Mayers DA. 2002. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? Human Genetics 111, 573-4.

Yeh FC, Yang R, Boyle TBJ, Ye ZH, Mao JX. 1997. POPGENE, the user-friendly shareware for population genetic analysis.

Zhao X, Kochert G, 1993. Phylogenetic distribution and genetic mapping of a (GGC)n microsatellite from rice (*Oryza sativa* L.). Plant Molecular Biology 21, 607-614.

Zohary, M. 1973. Geobotanical Foundations of the Middle East. Journal of Ecology 62, 349-350.

Ringraziamenti:

Numerose persone mi hanno permesso di raggiungere questo obiettivo.

Desidero ringraziare:

la Professoressa Antonella Furini per la sua disponibilità,

il Dott. Nicola la Porta, responsabile del progetto ECOCYPRE, per avermi dato l'oppotunità di svolgere questa ricerca,

il Dott. "Beppe" Vendramin, la Dott. ssa Silvia Fineschi e il Dott. Federico Sebastiani per le collaborazioni, i consigli e gli insegnamenti,

il Dott. Andrea Gandolfi per le "correzioni di bozze", per tutti i consigli e per la sua disponibilità,

tutti i colleghi e amici con cui ho condiviso questi anni e in particolare i miei compagni di ufficio e di laboratorio

la mia famiglia

Matteo.