UNIVERSITÀ DELLA CALABRIA

Dipartimento di CHIMICA

Submitted in the partial fulfilment of the requirements of the

Degree of Doctor of Philosophy

(PhD) in

Methodologies for the development of molecules of pharmacological interest

XIX cycle (CHIM/06)

Biochemistry of olive tree. From genomic to proteomic application on different olive tissues and oil

(PhD) Thesis

Department of Chemistry

University of Calabria

Italy

Supervisor
Prof. Dr. Giovanni SINDONA

Candidate
Ashif Sajjad

Coordinator
Prof. Dr. Giovanni SINDONA

Academic Year 2005/2006

Declaration

I certify that this thesis submitted for the degree of Ph.D. is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed: .....
(Ashif Sajjad)

Date: ……30/11/2006…..

Table of Contents

ACKNOWLEDGMENT

It is perhaps not possible to express in words my gratitude to my research supervisor Prof. Dr. Giovanni Sindona for his guidance and support during the past years. My interaction with him has had an enormous influence on my scientific development Prof. Sindona has given me invaluable scientific knowledge and has encouraged me to develop independent thinking.

I would also send a special thanks to Dr. Enzo Perri Director, C.R.A., Experimental Institute for Olive Growing. The time I spent in your Institute was a very pleasant time and it gave me invaluable experience.

Thanks, Prof. Dr. Anna Napoli, for your years of supervision, suggestions, guidance, teachings and your willingness to listen when I needed to talk about everything technical. Thanks, Dr. Innocenenzo Muzzalup, you always give me very helpful suggestions for my thesis writing. Thanks, Prof. Dr. Antonio De Nino, for your encouragement and support. My thanks also go out to all the Department of Ecology. Thanks, Prof. Dr. Silvia Mazzuca for your help to work in you laboratory, your guidance and useful discussions.

I wish to thank all the people in Pakistan who have helped me in several ways. I am most indebted to the University of Balochistan, Quetta Pakistan the professors and teachers. Everyone has been a part of my experience here in Italy and the assistance I've received and friendships I've made will never be forgotten.

I thank Italian Embassy Islamabad, Pakistan as well as the Italian Ministry of Foreign Affairs for giving me Italian Government Scholarship for PhD.

Finally, I owe all of my accomplishments to my parents. My father and mother, they sacrificed much to provide me with the quality of life from which I could choose what I want to be and to do. This thesis is dedicated to them both, with love and thanks for all they have done for me throughout my life.

SUMMARY

Olive (*Olea europaea* L.) belongs to a sclerophyll tree species. A complete understanding of the olive tree's biochemistry is of paramount importance. Genomic to proteomic, these technologies moving towards the high-throughput. Proteome, the end product of the genome. The hunt for new biomarkers has been augmented by the advent of genomics and proteomics. The Genome has been defined as GENe and chromosOME the genome of an organism is the whole hereditary information of an organism that is encoded in the DNA, genomics referred to the study of an organism's genomes. Proteome as the PROTEin complement expressed by a genOME or tissue. Proteomics is the field that involves the identification, characterization, and quantification of proteins in tissues or whole cells. Advances in DNA and protein sequence analysis has resulted in increasing input in the number of sequences deposited in database. The sequence databases has become an increasingly popular method for the protein characterization and cellular localization, plus identification of post-translational modifications. During recent years, polymerase chain reaction (PCR) based DNA markers and mass spectrometric methods have become an essential tool for plant genomics and proteomics research. Olive Genomics and Proteomics research has stood at the forefront. However, the genomics & proteomics are providing new insights about the olive tree biodiversity.

This dissertation presents (i) the utility of microsatellite DNA markers for management of olive germplasms and points out the high level of polymorphisms in microsatellite repeats when coupled with DNA sequence analysis. The establishment of genetic relationships among cultivars in the Abruzzo, Molise, Sicily and Apulia germplasm collection allows for the construction of a molecular database that can be used to establish the genetic relationships between known and unknown cultivars, (ii) the antigenic profile of *Olea europaea* pollen from different Mediterranean cultivars obtained by MALDI mass spectrometry. A simple procedure of chemical fractionation of the whole antigen extract was developed. The proposed experimental procedure, can supply valuable information on the antigens' micro heterogeneity, (iii) devised Phenol based method in order to extract high quality protein from olive (*Olea europea* L.) fruit which is required for proteomics and could be useful for the characterisation of some

enzymes and (iv) developed a quick, reliable & reproducible method for olive (*Olea europea* L.) oil protein extraction for getting different protein sequence in order to correlate their presence in the oil, which could be helpful to find the biological and geographical origin.

1. OVERVIEW.

The olive tree (*Olea europaea* L.) can reach a variable height from few meters to 15 m. The trunk is irregular and the tortuous branches bear evergreen, leathery, elliptical and/or lanceolate leaves with whole margin, green and silvery in the upper and lower part, respectively. The bloom happens in April-June and the pollination is prevalently anemophilous. The olive fruit is a drupe and its pulp, characterized by a bitter taste due to the presence of glycoside oleuropein, shows very high oil content. Owing to the beneficial properties of olive oil for the human health and to the presence of edible fruits, *Olea europaea* L. has been cultivated for millennia in the Mediterranean basin, becoming one of the most important crop in this area. Olive plants have been propagated vegetatively by local populations, from the Grecian age to the present time, to take advantage of the best genetic complement derived from genomic mutations and random crosses. Therefore, to look for alternative sources of material useful for increasing production and/or quality and to preserve minor cultivars as well as wild oleaster (*O. europaea* ssp. *sylvestris* Hoffgg et Link) and feral olive trees (originate either from varieties or from hybridization between a variety and an oleaster), which are a natural reserve of genetic diversity, efforts should be made to collect and characterize the olive germplasm (Angiolillo et al., 1999). Despite the long cultivation history and the great socio-economic interest of the olive tree, its germplasm is today poorly characterized on the whole. Indeed, elaiographical and biometrical studies are surprisingly insufficient to well address its management and preservation. In addition, reliable molecular standardized methods in order to elucidate the potential occurrence of homonyms (one denomination for several genotypes) or synonyms (one genotype with several denominations) are needed to eliminate ambiguities in variety identification. Although in the Mediterranean basin olive crop is usually cultivated by means of traditional methods on a large surface (about 7,000,000 ha), at present the risk of a genetic erosion of its yet not-sufficiently known germplasm exists (Khadari et al., 2003). In fact, there is an even more diffuse tendency to substitute the traditional orchards with new plantation made by only few varieties chosen on the basis of the agro-alimentary characteristics required by the market. Consequently, in most cultivation areas the rarefaction of the minor varieties occur, which now could be

considered as a threatened kind of biodiversity. Therefore, to prevent this rarefaction and to look for alternative source of material useful to increase production and/or quality of food we need to make efforts in order to preserve minor varieties as well as wild *oleasters* and feral olives, which are a natural genetic diversity reserve.

At the same time, it would be important to improve the *ex-situ* plant germplasm collection and utilise them in order to adequately characterise all accessions and develop future breeding programs. In this respect, several Mediterranean countries have promoted olive germplasm collections, like that of Cordoba (Spain), Porquerolles (France) and Marrakech (Morocco) which host most of the Mediterranean varieties.

Italian germplasm is large and more variegated at regional scale, because, in the years, every region selected some varieties which, by a progressive adaptation, took on characteristics of typical ness, at present recognized by the Council Regulation (EEC) n.2081/92 on the Protected Designation of Origin (PDO) and Protected Geographical Indication (PGI). This reinforces the need to build up a molecular data base for the Italian olive germplasm considering that it, at present in Italian olive germplasm is estimated to include over 500 cultivars and over 1300 synonyms (Bartolini et al., 1998), most of which are landraces vegetatively propagated at the farm level since ancient times. The number is probably underestimated because of inadequate information on minor local varieties widespread in the different olive growing areas., according to the FAO Olive Germplasm (Bartolini et al., 1998) and the Italian Olive Card-index[‡]. The study of less-common or minor cultivars on a regional scale is important since they may have traits not considered important in the past, but that might be important to meet the challenges of modern olive growing. Low vigor, resistance to low temperatures, salinity tolerance (Gucci and Tattini, 1997), adaptability to low pruning systems (Cantini et al., 1998), late ripening (Montedoro et al., 1990), phenol content (Perri et al., 1999) and fatty-acid content (Lanteri et al., 2002) are important traits for olive or olive oil quality.
This represents an important genetic diversity reserve for the whole Mediterranean basin. Which include more possible integrating all the different data collected by several research groups working in Italy.

[‡] "Elenco delle varietà di olive ufficialmente iscritte nello Schedario Oleicolo Italiano" Suppl. Ordinario G.U. 05-01-1994.

1.1 GENOMICS

Genomics can also be defined as the identification, analysis and utilization of DNA sequence information. After the completion of human genome sequence the genome of the model plant *Arabidopsis thaliana* has been sequenced. The information derived from plant-based genomics studies and their application on other plant species, including crop plants helps to understand how genes enable a plant to function, how the confusing diversity of functions that distinguish plant species is related to simple changes in their genomes. In genomics analysis there are different markers have been used, like morphological markers, protein based markers and DNA based markers. Microsatellites which provide a good genome coverage and hence are suitable as genetic markers in plant genome research.

1.2 Microsatellite DNA

Microsatellites are simple sequence repeats (SSRs) or short tandem repeats (STRs) of genomic sequences, discovered by Hamada and colleagues (1982). Tandemly repeated sequences with a very short nucleotide motif (1-6 bases long) came to be called microsatellites. Attempts have been made to standardise the nomenclature of microsatellites, no absolute consensus has been agreed upon (Tautz, 1993; Jarne and Lagoda, 1996; Chambers and MacAvoy, 2000). The microsatellites locus might be a streach of DNA with base sequence CA repeated 17 times in succession [$(CA)_{17}$] or AAT repeated 10 times [$(CA)_{10}$] (Queller et al., 1993). The repeated sequence (motif) is often simple, consisting of mononucleotide SSR $(A)_{11}$ AAAAAAAAAAA, dinucleotide SSR $(GT)_6$ GTGTGTGTGTGT, Trinucleotide SSR $(CTG)_4$ CTGCTGCTGCTG and tetranucleotide SSR $(ACTC)_4$ ACTCACTCACTCACTC. Microsatellites occur in many places (loci) throughout the genome, but the majority are in non-coding regions (Wren et al., 2000). For example, AC_{22} might occur in thousands of places in the genome. The trick is to find and sequence the flanking regions at a particular place (locus)-the DNA on each side of the repeat. The flanking region is a presumably random (unpatterned) stretch of nucleotides. It should occur at only one locus, because the probability that a longish unpatterned sequence arises more than once is vanishingly small -- *P* approximately equal to 4^{-60}, if the total length of the flanking

regions is 60 base pairs. Because the 'beads' are all the same, microsatellite mutation can occur relatively easily, by a process called strand slippage during DNA replication (Levinson and Gutman, 1987; Schlötterer and Tautz, 1992; Strand et al., 1993). Even if a few 'beads' in the complementary strands of DNA don't match up perfectly, the rest do, and strand 1 can add some base pairs to repair the DNA match, or strand 2 can cut out a few. Mutation may also occur during recombination or during meiosis (Blouin et al., 1996). The mutation rate is relatively high (between 10^{-2} and 10^{-6} mutations per locus per generation, and on average 5×10^{-4}) that generate the high levels of allelic diversity necessary for genetic studies (Schlötterer, 2000; Selkoe et al., 2006). Thus, a given population might contain variants of AC_n, where n is the number of 'beads' (the motif AC repeated, head to tail, n times), with n ranging from, for example, 10 to 28. If the adjacent sequence information for the flanking regions are known, the forward and reverse primers for the polymerase chain reaction (PCR) can be designed. The PCR greatly amplifies the target microsatellite, yielding billions of times more copies than the started one. After that the PCR-amplified product run on an electrophoretic gel, through which fragments of different size travel at different rates. Amplified fragments of the same size (e.g., AC_{26}, AC_{27}) will form bands in the gel that can be visualized by a variety of techniques like the use of a highly efficient system of fluorescent labeling that is analyzed by a laser beam in an automated DNA sequencer. The size variants are alleles-variants of the genetic material. At any particular locus, an individual has one band/allele (if it is a homozygote) or two bands (if it's a heterozygote). In the population as a whole, however, there may be many alleles.. Most of the observed changes in microsatellite allele sizes are attributable to deletion or addition of a few repetitive units (Straub et al., 1993; Deka et al., 1995).Typical vertebrate populations can have as many as 5 to 15 alleles at a given microsatellite locus. The analyses of several or many loci, it is possible to perform a genetic analyses on the resulting data sets. Potential analyses of microsatellites range from individual identification, to parentage exclusion, to relatedness calculations, to diagnosis of genetically mediated diseases, and on up to analysis of genetic differentiation between populations or species. Thus microsatellites markers are the preferred markers for wide and varied applications like genetic diversity analysis (Yang et al., 1994), cultivar identification (Thomas and Scott, 1993),

determination of hybridity (Provan et al., 1996) and diagnosis of important traits in plant breeding programs (Yu et al., 1994).

1.2.1 Microsatellites amplification

Microsatellites can be amplified for identification using Polymerase Chain Reaction (PCR), using templates of flanking regions (primers). DNA is denatured at a high temperature, separating the double strand, allowing annealing of primers and the extension of nucleotide sequences along opposite strands at lower temperatures. This process results in production of enough DNA to be visible on agarose or acrylamide gels; only small amounts of DNA are needed for amplification as thermocycling in this manner creates an exponential increase in the replicated segment (Griffiths et al., 1996) with the abundance of PCR technology, primers that flank microsatellite loci are simple and quick to use, but the development of such primers is often a tedious and costly process.

1.2.2 Microsatellite primers

Microsatellite primers are developed by cloning random segments of DNA from the focal species. These are inserted into a plasmid or phage vector, which is in turn implanted into *Escherichia coli* bacteria. Colonies are then developed, and screened with fluorescently–labelled oligonucleotide sequences that will hybridise to a microsatellite repeat, if present on the DNA segment. If positive clones can be obtained from this procedure, the DNA is sequenced and PCR primers are chosen from sequences flanking such regions to determine a specific locus. This process involves significant trial and error on the part of researchers, as microsatellite repeat sequences must be predicted and primers that are randomly isolated may not display significant polymorphism (Queller et al., 1993; Jarne and Lagoda, 1996). Microsatellite loci are widely distributed throughout the genome and can be isolated from semi-degraded DNA of older specimens, as all that is needed is a suitable substrate for amplification through PCR.

1.3 Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) a revolutionary method of gene analysis was developed by Kary Mullis in mid-1980's. The polymerase chain reaction is a test tube system for DNA replication that allows a "target" DNA sequence to be selectively amplified, or enriched, several million-fold in just a few hours **(Fig.1)** without the reliance to cloning. Within a dividing cell, DNA replication involves a series of enzyme-mediated reactions, whose end result is a faithful copy of the entire genome. Within a test tube, PCR product is amplified from the DNA template using a heat-stable DNA polymerase from Thermus aquaticus (Taq DNA polymerase) and using an automated thermal cycler to put the reaction through 30 or more cycles of denaturing, annealing of primers, and polymerization.

To perform a PCR reaction, a small quantity of the target DNA is added to a test tube with a buffered solution containing DNA polymerase, oligonucleotide primers that flank the sequence, the four deoxynucleotide building blocks of DNA, and the cofactor $MgCl_2$. The PCR mixture is taken through 6 replication cycles consisting of:

i. Denature the DNA at 94˚C.

ii. Rapidly cool the DNA (37-65˚C) and anneal primers to complementary single stranded sequences flanking the target DNA.

iii. Extend primers at 70-75˚C using a heat-resistant DNA polymerase (e.g., Taq polymerase derived from Thermus aquaticus).

iv. Repeat the cycle of denaturing, annealing, and extension 20-45 times to produce 1 million (220)to 35 trillion copies (245) of the target DNA.

v. Extend the primers at 70-75˚C once more to allow incomplete extension products in the reaction mixture to extend completely.

vi. Cool to 4˚C and store or use amplified PCR product for analysis.

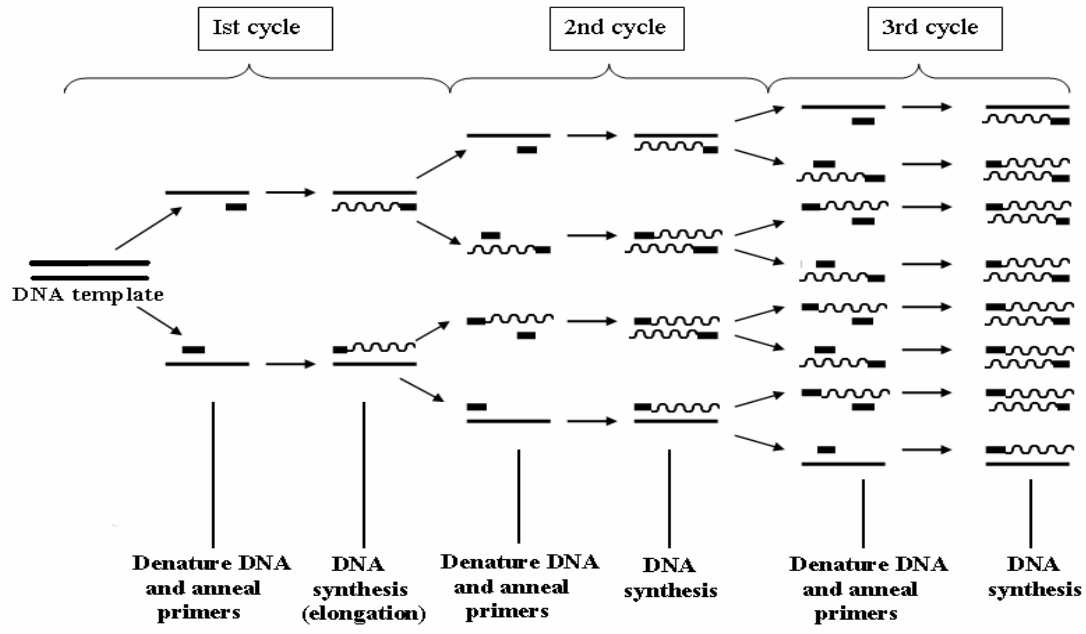


Fig. 1. Polymerase Chain Reaction (PCR)

1.3.1 Resolving PCR products.

After amplification by PCR, the products are separated by Agarose or polyacrylamide gel electrophoresis and are directly visualized after staining with ethidium bromide. Another alternate method, which is commercially available called chip-based nucleic acid separation system can replace the earlier method.

1.3.2 Agarose Gel Electrophoresis.

The standard method used to separate, identify, and purify DNA fragments is electrophoresis through agarose gels. The technique is simple, rapid to perform, and capable of resolving mixtures of DNA fragments that cannot be separated adequately by other sizing procedures. Furthermore, the location of DNA within the gel can be determined directly: Bands of DNA in the gel are stained with the intercalating dye ethidium bromide ; as little as 1 ng of DNA can be detected by direct examination of the gel in ultraviolet light (Sharp et al., 1973).

Agarose, which is extracted from seaweed, is a linear polymer whose basic structure is shown in **(**Fig. 2).

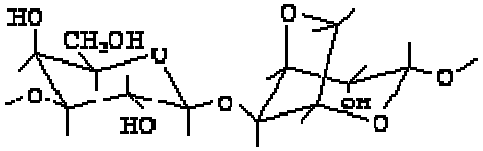


Fig. 2. Agarose structure unit

Agarose gels are cast by melting the agarose in the presence of the desired buffer until a clear, transparent solution is achieved. The melted solution is then poured into a mold and allowed to harden. Upon hardening, the agarose forms a matrix, the density of which is determined by the concentration of the agarose. When an electric field is applied across the gel, DNA, which is negatively charged at neutral pH, migrates toward the anode. The electrophoretic migration rate of DNA through agarose gels is dependent upon four main parameters, which are as below:

The molecular size of the DNA. Molecules of linear, duplex DNA, which are believed to migrate in an end-on position (Fisher and Dingman, 1971; Aaij and Borst, 1972) travel through gel matrices at rates that are inversely proportional to the logarithm of their molecular weights (Helling et al., 1974).

The agarose concentration. A DNA fragment of given size migrates at different rates through gels containing different concentrations of agarose. There is a linear relationship between the logarithm of the electrophoretic mobility of DNA (μ) and gel concentration (τ), which is described by the equation:

$$\log \mu = \log \mu_{0-K} \tau$$

where μ_0 is the free electrophoretic mobility and κ is the retardation coefficient, a constant that is related to the properties of the gel and the size of the migrating molecules. Thus, by using gels of different concentrations, it is possible to resolve a wide-range of DNA fragments.

The conformation of the DNA. Closed circular, nicked circular and linear DNA of the same molecular weight migrate through agarose gels at different rates. The relative mobilities of the three forms are dependent primarily on the agarose concentration in the

gel but are also influenced by the strength of the applied current, the ionic strength of the buffer, and the density of superhelical twists in the DNA.

The applied current. At low voltages, the rate of migration of linear DNA fragments is proportional to the voltage applied. However, as the electric field strength is raised, the mobility of high-molecular-weight fragments of DNA is increased differentially. Thus, the effective range of separation of agarose gels decreases as the voltage is increased. Gels should be run at no more than 5 V/cm.

Base composition and temperature. The electrophoretic behavior of DNA in agarose gels (by contrast to polyacrylamide gels (Allett et al., 1973) is not significantly affected either by the base composition of the DNA (Thomas and Davis, 1975) or the temperature at which the gel is run. Thus, in agarose gels the relative electrophoretic mobilities of DNA fragments of different sizes do not change between 4psy176 C and 30psy176 C.

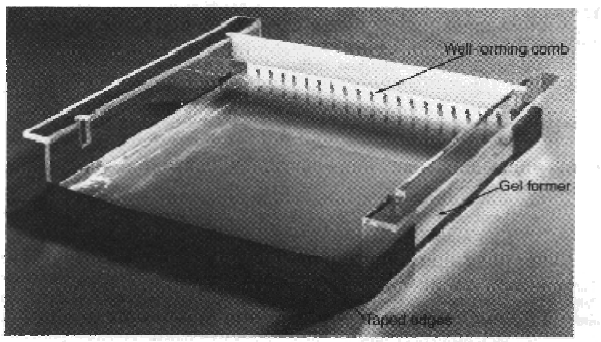


Fig. 3. Photograph of agarose gel, showing taped edges and well-forming comb in position.

The agarose concentration is varied for different fragment ranges. For analyzing the complete 2ar codon (1239 bp), a 1% agarose gel is made by dissolving agarose in 1x TAE buffer by heating in a microwave oven. After cooling to about 60 psy 176 C, ethidium bromide is added to a final concentration of 0.5 μg/ml. The agarose solution is poured into a taped gel former mold to make the gel (Fig. 3). A well-forming comb (12 slots for minigels) is placed near one edge of the gel. The gel is cooled to harden until it becomes milky and opaque (approximately one hour). The gel mold is placed

horizontally into the electrophoresis tank, which is filled with 1xTAE (0.5 μg/ml ethidium bromide).

The gel loading buffer is applied to the samples and they are carefully added to individual wells. The electrophoresis is run by 70-100 V/20-80 mA for about an hour or at 20 to 30 V overnight. The size of fragments can be determined by calibrating the gel, using known standards (e.g., λDNA *Eco*RI / *Hind*III digest, Boehringer Mannheim, or 100bp ladder, BioVentures, Inc.), and comparing the distance the unknown fragment has migrated.

The most convenient method of visualizing DNA in agarose gels is by use of the fluorescent dye ethidium bromide (Sharp et al., 1973) (2,7-Diamino-10-ethyl-9-phenyl-phenanthridinium bromide). This substance (Fig. 4) contains a planar group that intercalates between stacked bases of DNA. The fixed position of this group and its close proximity to the bases causes dye bound to DNA to display an increased fluorescent yield compared to dye in free solution. UV-irradiation absorbed by the DNA at 260 nm and transmitted to the dye, or irradiation absorbed at 300 nm and 360 nm by the bound dye itself, is emitted at 590 nm in the red-orange region of the visible spectrum.

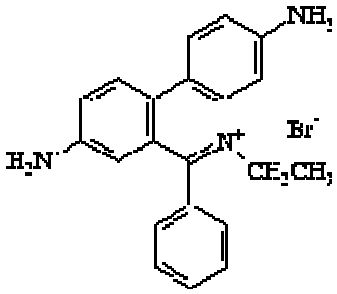


Fig. 4. Ethidium bromide

Ethidium bromide can be used to detect both single- and double-stranded nucleic acids. However, the affinity of the dye for single-stranded nucleic acid is relatively low and the fluorescent yield is poor.

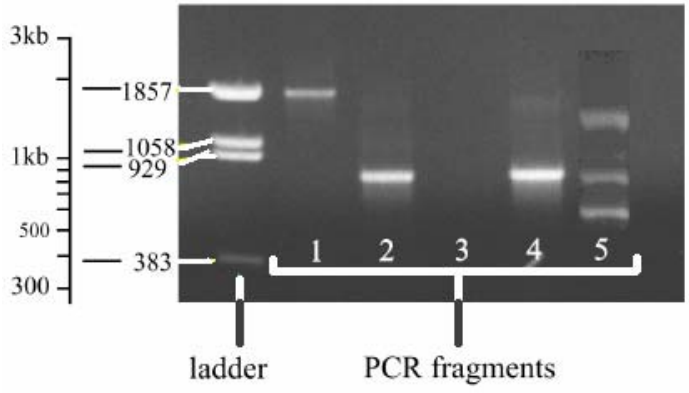


Fig. 5. Verification of PCR product on Agarose gel.

1.3.3 Chip-based nucleic acid separation.

Principles of Nucleic Acid and Protein Analysis on a Chip[2]

The electrophoretic assays are based on traditional gel electrophoresis principles that have been transferred to a chip format. The chip format dramatically reduces separation time and sample consumption. The system provides automated sizing and quantitation information in a digital format. On-chip gel electrophoresis is performed for the analysis of DNA, RNA and proteins.

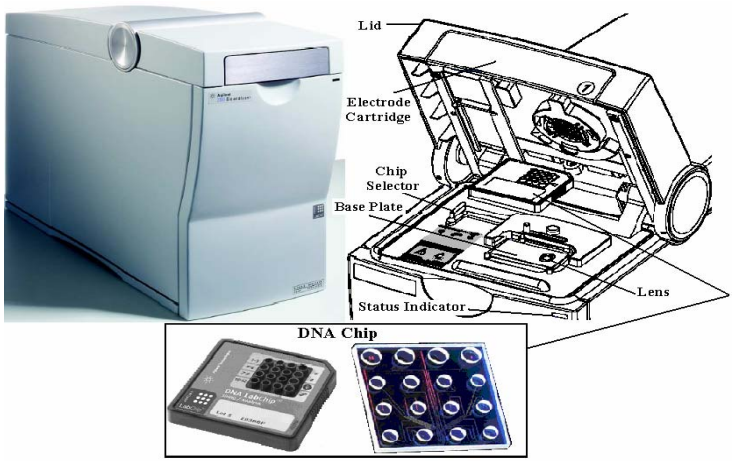


Fig. 6. Major components and DNA chip of Agilent 2100 Bioanalyzer

[2] (From Agilent 2100 expert software online help)

The chip accommodates sample wells, gel wells and a well for an external standard (ladder). Micro-channels are fabricated in glass to create interconnected networks among these wells (Fig. 6). During chip preparation, the micro-channels are filled with a sieving polymer and fluorescence dye. Once the wells and channels are filled, the chip becomes an integrated electrical circuit. The 16-pin electrodes of the cartridge are arranged so that they fit into the wells of the chip. Each electrode is connected to an independent power supply that provides maximum control and flexibility. Charged biomolecules like DNA or RNA are electrophoretically driven by a voltage gradient-similar to slab gel electrophoresis. Because of a constant mass-to-charge ratio and the presence of a sieving polymer matrix, the molecules are separated by size. Smaller fragments are migrating faster than larger ones. Dye molecules intercalate into DNA or RNA strands or Protein-SDS micells. These complexes are detected by laser-induced fluorescence. Data is translated into gel-like images (bands) and electropherograms (peaks) (Fig. 7). With the help of a ladder that contains fragments of known sizes and concentrations, a standard curve of migration time versus fragments size is plotted. From the migration times measured for each fragment in the sample, the size is calculated.

For DNA and protein assays, quantitation is done with the help of the upper marker. The area under the upper marker peak is compared with the sample peak areas. Because the concentration of the upper marker is known, the concentration for each sample can be calculated. Besides this relative quantitation, an absolute quantitation is available for protein assays, using external standard proteins.

The 2100 expert software plots fluorescence intensity versus migration time and produces an electropherogram for each sample.

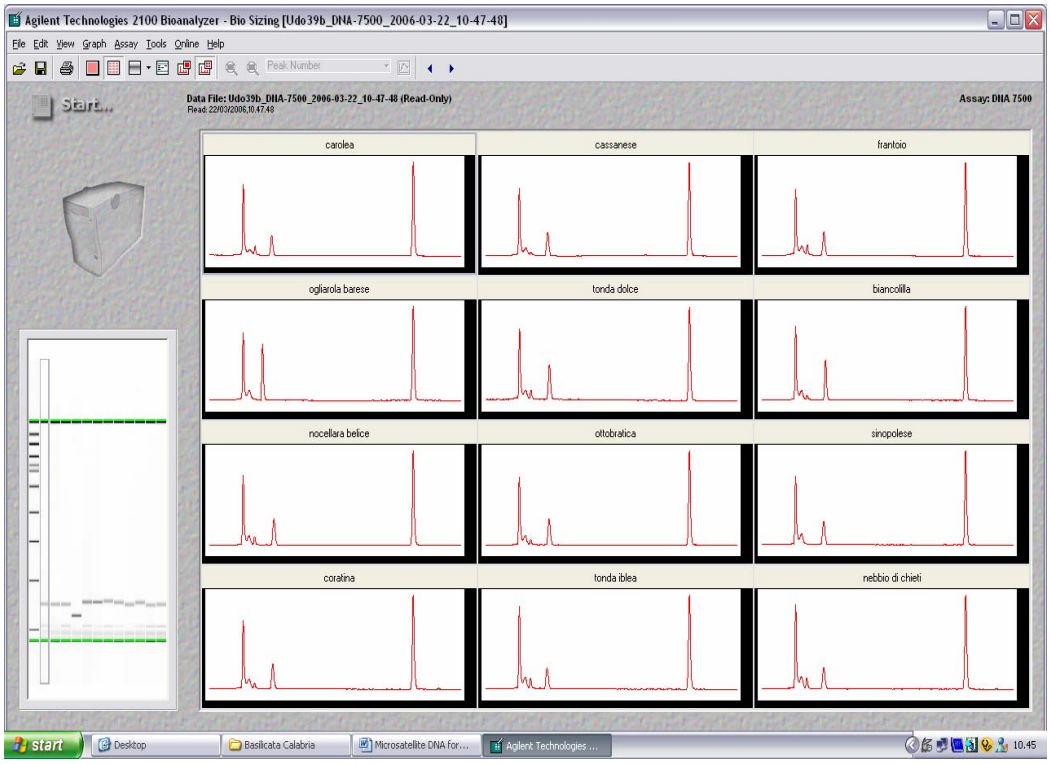


Fig. 7. A DNA 7500 assay as displayed in the Aglent 2100 biosyzing software. Data is presented both as electrophograms and a slab gel-like image.

1.3.4 Polyacrylamide gel electrophoresis[3]

Separation of DNA molecules differing in length by just one nucleotide.

Polyacrylamide gel electrophoresis is used to examine the families of chain-terminated DNA molecules resulting from a sequencing experiment. Agarose gel electrophoresis cannot be used for this purpose because it does not have the resolving power needed to separate single-stranded DNA molecules that differ in length by just one nucleotide. Polyacrylamide gels have smaller pore sizes than agarose gels and allow precise separations of molecules from 10–1500 bp. As well as DNA sequencing, polyacrylamide gels are also used for other applications where fine-scale DNA separations are required, for instance in the examination of amplification products from PCRs directed at microsatellite loci, where the products of different alleles might differ in size by just two or three base pairs. A polyacrylamide gel consists of chains of

[3] (http://www.ncbi.nlm.nih.gov)

acrylamide monomers (CH_2=CH-CO-NH_2) cross-linked with *N, N'*-methylenebisacrylamide units (CH_2=CH-CO-NH-CH_2-NH-CO-CH-CH_2), the latter commonly called 'bis'. The pore size of the gel is determined by both the total concentration of monomers (acrylamide + bis) and the ratio of acrylamide to bis. For DNA sequencing, a 6% gel with an acrylamide : bis ratio of 19 : 1 is normally used because this allows resolution of single-stranded DNA molecules between 100 and 750 nucleotides in length. About 650 nucleotides of sequence can therefore be read from a single gel. The gel concentration can be increased to 8% in order to read the sequence closer to the primer (resolving molecules 50–400 nucleotides in length) or decreased to 4% to read a more distant sequence (500–1500 nucleotides from the primer). Polymerization of the acrylamide : bis solution is initiated by ammonium persulfate and catalyzed by TEMED (*N, N, N', N'*-tetramethylethylenediamine). Sequencing gels also contain urea, which is a denaturant that prevents intra-strand base pairs from forming in the chain-terminated molecules. This is important because the change in conformation resulting from base-pairing alters the migration rate of a single-stranded molecule, so the strict equivalence between the length of a molecule and its band position, critical for reading the DNA sequence is lost.

Polyacrylamide gels are prepared between two glass plates held apart by spacers. This arrangement serves two purposes. First, it enables a very thin (<1 mm) gel to be made, which facilitates sequence reading by improving the sharpness of the bands. Second, it ensures that polymerization, which is inhibited by oxygen, occurs evenly throughout the gel. With such a thin gel the amount of DNA per band is small and the banding pattern is only barely visible after ethidium bromide staining. For this reason, a radioactively labeled nucleotide is usually included in the sequencing reactions so that the banding pattern can be visualized by autoradiography

1.4 DNA Sequencing.

Rapid and efficient methods for determining the nucleotide sequence of DNA were first devised in the mid-1970s. Two different procedures were published at almost the same time:

1.4.1 Principles of DNA Sequencing[4]

The process of determining the order of the nucleotide bases along a DNA strand is called sequencing. In 1977, twenty-four years after the discovery of the structure of DNA, two separate methods for sequencing DNA were developed: the chain termination method (Sanger et al., 1977) and the chemical degradation method (Maxam and Gilbert, 1977). Both methods were equally popular to begin with, but, for many reasons, the chain termination method is the method more commonly used today (Fig. 8). This method is based on the principle that single-stranded DNA molecules that differ in length by just a single nucleotide can be separated from one another using polyacrylamide gel electrophoresis, described earlier. All the steps in a sequencing reaction (like in PCR), which are repeated for 30 or 40 cycles. The DNA to be sequenced, called the template DNA, is first prepared as a single-stranded DNA. Next, a short oligonucleotide is annealed, or joined, to the same position on each template strand. The oligonucleotide acts as a primer for the synthesis of a new DNA strand that will be complimentary to the template DNA. This technique requires that four nucleotide-specific reactions--one each for G, A, C, and T--be performed on four identical samples of DNA. The four sequencing reactions require the addition of all the components necessary to synthesize and label new DNA, including:

• A DNA template;
• A primer tagged with a mildly radioactive molecule or a light-emitting chemical;
• DNA polymerase--an enzyme that drives the synthesis of DNA;
• Four deoxynucleotides (G, A, C, T); and
• One dideoxynucleotide, either ddG, ddA, ddC, or ddT.

After the first deoxynucleotide is added to the growing complementary sequence, DNA polymerase moves along the template and continues to add base after base. The strand synthesis reaction continues until a dideoxynucleotide is added, blocking further elongation. This is because dideoxynucleotides are missing a special group of molecules, called a 3'-hydroxyl group, needed to form a connection with the next nucleotide. Only a small amount of a dideoxynucleotide is added to each reaction,

[4] (http://depts.washington.edu/pceut/pceut_services/DNA-Sequencing-NCBI)

allowing different reactions to proceed for various lengths of time, unti, by chance, DNA polymerase inserts a dideoxynucleotide , terminating the reaction. Therefore, the result is a set of new chains, all of different lengths. To read the newly generated sequence, the four reactions are run side-by-side on a polyacrylamide sequencing gel. The family of molecules generated in the presence of ddATP are loaded into one lane of the gel and the other three families, generated with ddCTP, ddGTP, and ddTTP, are loaded into three adjacent lanes. After electrophoresis, the DNA sequence can be read directly from the positions of the bands in the gel.

1.4.2 Separation of the molecules.

After the sequencing reactions, the mixture of strands, all of different length and all ending on a fluorescently labeled ddNTP have to be separated; This is done on an acrylamide gel, which is capable of separating a molecule of 30 bases from one of 31 bases, but also a molecule of 750 bases from one of 751 bases. All this is done with gel electrophoresis. DNA has a negative charge and migrates to the positive side. Smaller fragments migrate faster, so the DNA molecules are separated on their size.

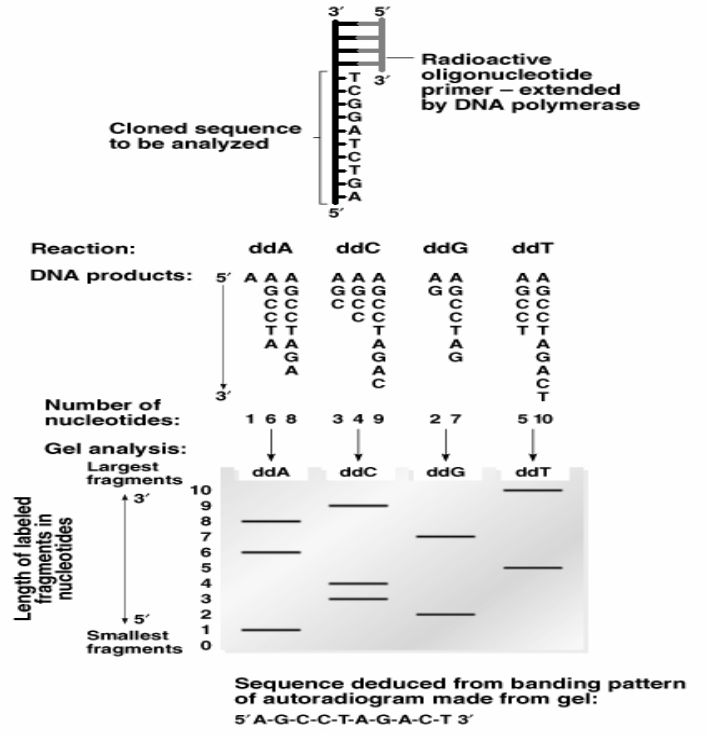


Fig. 8. DNA Sequencing by the chain termination (Sanger) method.

1.4.3 Automated DNA sequencing with fluorescently labeled dideoxynucleotides.

Variations of this method have been developed for automated sequencing machines. In one method, called cycle sequencing, the dideoxynucleotides--not the primers--are tagged with different colored fluorescent dyes, thus all four reactions occur in the same tube and are separated in the same lane on the gel. As each labeled DNA fragment passes a detector at the bottom of the gel, the color is recorded and the sequence is reconstructed from the pattern of colors representing each nucleotide in the sequence (Fig. 9).

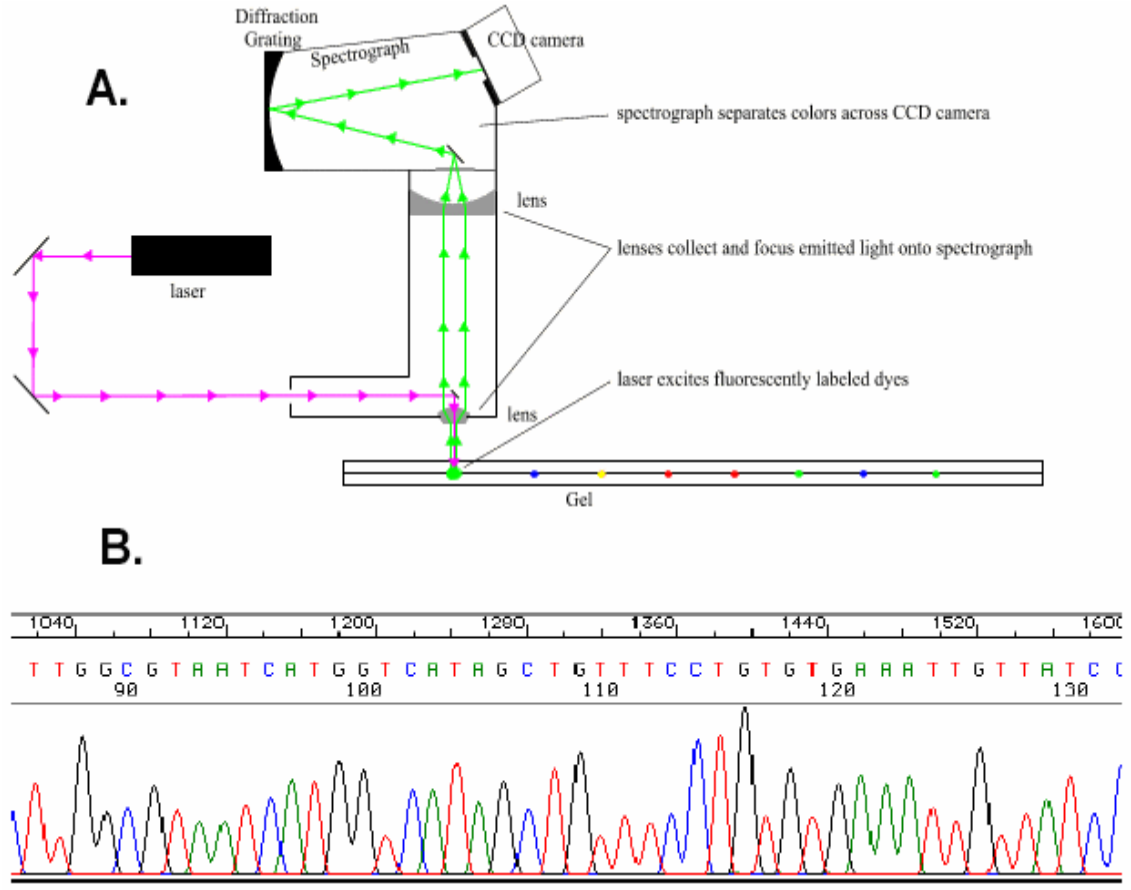


Fig. 9. **A**. The basic set up of a capillary system. The fluorescently labelled fragments that migrate trough the gel, are passing a laser beam at the bottom of the gel. The laser exites the fluorescent molecule, which sends out light of a distinct color. That light is collected and focused by lenses into a spectrograph. Based on the wavelength, the spectrograph separates the light across a CCD camera (**c**harge **c**oupled **d**evice). Each base has its own color, so the sequencer can detect the order of the bases. **B**. The final sequence electropherogram output.

1.5 Phylogenetics.[5]

The word derived form (Greek: phylon = tribe, race and genetikos = relative to birth, from genesis = birth) is the study of evolutionary relatedness among various groups of organisms (e.g., species, populations).

1.5.1 Purpose of phylogenetics.

With the aid of sequences, it should be possible to find the genealogical ties between organisms. Experience learns that closely related organisms have similar sequences, more distantly related organisms have more dissimilar sequences. One objective is to reconstruct the evolutionary relationship between species.
An other objective is to estimate the time of divergence between two organisms since they last shared a common ancestor.

1.5.2 Disclaimers.

The theory and practical applications of the different models are not universally accepted.
With one dataset, different software packages can give different results. Changes in the dataset can also give different results. Therefore, it is important to have a good alignment to start with.
Trees based on an alignment of a gene represent the relationship between genes and this is not necessarily the same relationship as between the whole organisms. If trees are calculated based on different genes from organisms, it is possible that these trees result in different relationships.

[5] (http://users.ugent.be/~avierstr/principles/phylogeny.html)

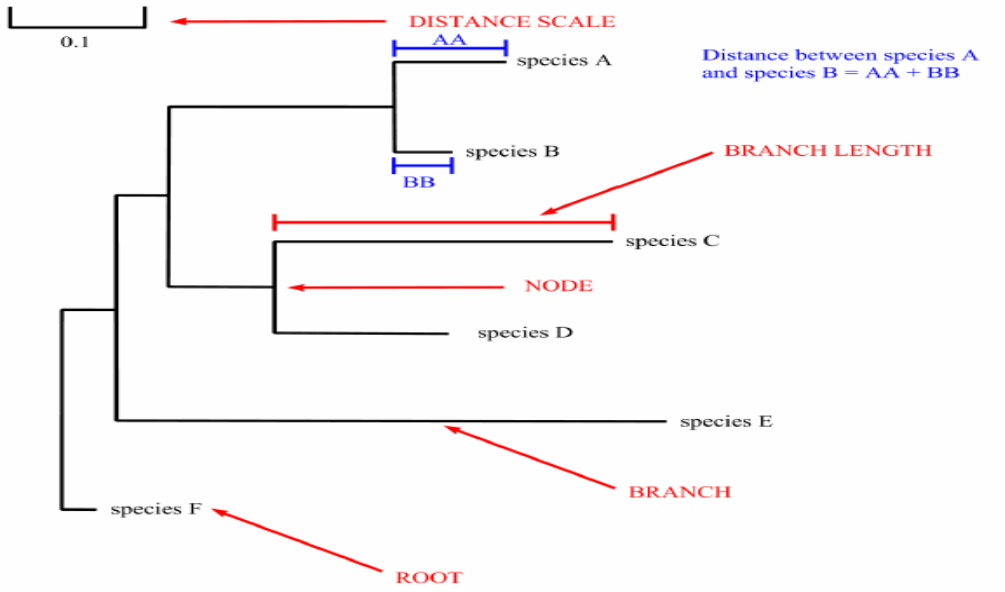


Fig. 10. The tree terminology. **node :** a node represents a taxonomic unit. This can be a taxon (an existing species) or an ancestor (unknown species : represents the ancestor of 2 or more species). **branch :** defines the relationship between the taxa in terms of descent and ancestry. **topology :** is the branching pattern. **branch length :** often represents the number of changes that have occurred in that branch. **root :** is the common ancestor of all taxa. **distance scale :** scale which represents the number of differences between sequences (e.g. 0.1 means 10 % differences between two sequences)

1.5.3 Methods of phylogenetic analysis

There are two major groups of analyses to examine phylogenetic relationships between sequences:

i. Phenetic methods: trees are calculated by *similarities of sequences* and are based on **distance** methods. The resulting tree is called a **dendrogram** and does not necessarily reflect evolutionary relationships. Distance methods compress all of the individual differences between pairs of sequences into a single number.

ii. Cladistic methods : trees are calculated by considering the *various possible pathways of evolution* and are based on **parsimony** or **likelihood** methods. The resulting tree is called a **cladogram**. Cladistic methods use each alignment position as evolutionary information to build a tree.

1.5.4 Phenetic methods based on distances.

i. Starting from an alignment, **pairwise distances** are calculated between DNA sequences as the sum of all base pair differences between two sequences (the most similar sequences are assumed to be closely related). This creates a **distance matrix**.
All base changes can be considered equally or a matrix of the possible replacements can be used.
Insertions and deletions are given a larger weight than replacements. Insertions or deletions of multiple bases at one position are given less weight than multiple independent insertions or deletions. it is possible to correct for multiple substitutions at a single site.

ii. From the obtained **distance matrix**, a phylogenetic tree is calculated with **clustering algorithms**. These cluster methods construct a tree by linking the least distant pair of taxa, followed by successively more distant taxa.

UPGMA clustering (Unweighted Pair Group Method using Arithmetic averages) :
this is the simplest method.
Neighbor Joining : this method tries to correct the UPGMA method for its assumption that the rate of evolution is the same in all taxa.

2. PROTEOMICS.

Proteomics is the study of the function of all expressed proteins (Tyers and Mann, 2003). The term proteome was first coined to describe the set of proteins encoded by the genome (Wilkins et al., 1996). The study of the proteome, called proteomics, which deal with all the proteins in a cell, all protein isoforms, post-translational modifications. Proteomics is much more diverse and complicated than genomics. Most importantly, while the genome is a rather constant entity, the proteome differs from cell to cell and is constantly changing through its biochemical interactions with the genome and the environment. One organism has radically different protein expression in different parts of its body, in different stages of its life cycle and in different environmental conditions.

Proteomics is a quite new field of research that study and analyse the proteome of an individual. This new discipline can be defined as "the qualitative and quantitative comparison between the proteome in different conditions with the objective of understanding the cellular processes which subject the biological one (Anderson and Anderson, 1996).

Furthermore, proteomics is a complex methodology because it is made by the combination of more than one techniques, mixed together to obtain complete information about all the proteins inside of a cell. There are many possible applications: identification and characterization of single or mixture proteins; analysis of proteins profile; finding of new proteins, post-traditional modifications and of new markers; characterization of glycosilation and phosphorilation sites; assays of peptides and proteins. Many research groups use also this discipline for the characterization of protein mutations by the comparison between healthy and diseased cells (David et al., 2005). Proteomics can detect and assay with accuracy all the changes induced by a specific alteration as for examples, a genetic mutation and/or pharmacological and pathological matters (Gygi and Ruedi, 2000)

Nowadays, the proteomics research can be divided into two principal branches expression proteomics and functional proteomics (Tyers and Mann, 2003). Expression

proteomics target the qualitative and quantitative definition of the growth and reduction of the protein expression levels induced by variations of cellular conditions. The main targets of the functional proteomics are, instead, the recognition of biological function of proteins, which nowadays unknown, and the identification of *in vivo* protein-protein interactions, with the purpose of description cellular processes at molecular level.

The great versatility of the proteomic approach is substantially legacy to the multiple techniques employed for the study of proteome. Generally it is carried out by an efficient extraction/isolation and purification of protein from complex mixtures. The separation of simple protein mixtures by using one- dimension (1-D) sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). While the complex protein mixtures, such as total cell lysates, require the use of the highly resolving two-dimensional (2-D) SDS-PAGE. And the identification by using the mass spectrometry, MALDI-MS (Matrix Assisted Laser Desorption Ionization-Mass Spectrometry) and/or MSMS followed by extensive database searches. In some cases if more structural information is required form the separated protein mixture the liquid chromatography (LC)-electrospray ionization (LC-ESI- MS and LC-ESI-MS/MS) is often employed (Fenn et al., 1989).

2.1 Proteomics Strategy.

2.1.1 Protein extraction and sample preparation.

The extraction and preparation of samples is a critical step and has a great impact on any proteomics study. Since proteins are not a homogeneous entity, having different biophysical and chemical properties, it is important to characterize the correct conditions for solubilization of biological samples (Herbert, 1999). Protein solubilization is the process of breaking interactions involved in protein aggregation (Rabilloud, 1996). Sample solubilization solutions typically contain a number of compounds including detergents, reducing agents, phosphatase and protease inhibitors and also some time denaturing solvents are used for the "linearization" of proteins which facilitate the solubilization. The successive step follows after the sample

preparation is the separation of proteins. The method more common in order to separate the protein mixture by electrophoresis mono or bidimensional (1-D or 2-D SDS-PAGE) (Rabilloud, 1996).

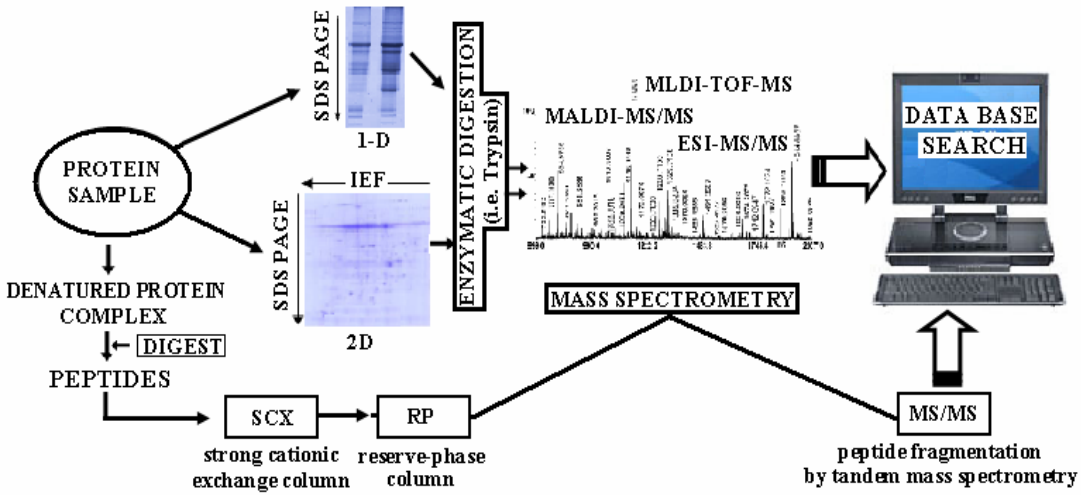


Fig. 11. Proteomics strategy

2.1.2 Protein separation.

Various approaches are possible for protein separation (Fig. 11) including 1-D SDS-PAGE. However, for quantitative and comparative proteomic analysis, there is still no widely available technology that surpasses 2-D gel electrophoresis (2-DE), which has been the method of choice for separation of complex protein mixtures for several decades (Rose et al., 2004). In the 2-D-PAGE the protein mixtures resolved first in one dimension isoelectrofocusing (IEF), that separates proteins on the basis of isoelectric point (*pI*). These proteins are then resolved on a second dimension SDS- PAGE, which separate according to molecular weight (MW). The proteins, that run on the gel, that have been separated during electrophoresis, can be visualized using different staining techniques, most common are Coomassie Brilliant Blue staining is relatively easy, cost-effective and compatible with subsequent protein identification by mass spectrometry (MS), but it is only moderately sensitive, with a limit of approximately 10 ng protein. The other alternative has been silver-staining, which is more sensitive, detecting as little as 0.5 ng protein (Heazlewood and Millar, 2003) but not particularly quantitative and less suitable for MS identification (Lopez, 2000; Patton, 2000). Moreover, although the

two dimension gel electrophoresis (2-DE) is the method of choice for separation of complex protein mixtures. The basic proteins and/or the proteins of low molecular weight are difficult to separate by electrophoresis, an alternative approach is the protein fractionation is pre-fractionation at the peptide level. Then the protein are digested in solution and resulting peptides are separated using 2-D chromatography: in the first dimension according to their charge (typically SCX chromatography) and in the second dimension according to hydrophobicity by reversed-phase chromatography. (Stasyk et al., 2004).

2.1.3 Protein identification

The mass spectrometry is widely used and is the method of choice for both protein identification and characterization. The first step toward protein identification is typically excision of 1-D or 2-D gel containing the protein bands or spots of interest, in-gel digestion with a site-specific protease (commonly trypsin), and finally MS analysis of the resultant eluted peptides (Figure 11). Two MS platforms in particular represent powerful tools for proteomic studies. The first, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS is typically used to measure the masses of the peptides derived from the trypsinized parent protein spot, generating a 'peptide mass fingerprint' (PMF). A common alternative to the PMF approach is de novo sequencing by electrospray ionization tandem mass spectrometry (ESI) MS/MS, which yields amino acid sequences of selected tryptic peptides (Figure 11). The first step of tandem MS involves ionization of a sample and separation based upon the mass-to-charge ratio (*m/z*) of the primary ions. An ion with a specific *m/z* value is then selected, fragmented, and the fragment ions detected after passing through the second mass spectrometer. This process produces a series of fragment ions that can differ by single amino acids allowing a portion of the peptide sequence, termed an 'amino acid sequence tag', to be determined and used for database searching (Rose et al., 2004).

2.2 Electrophoresis of Proteins.

2.2.1 One-Dimensional SDS-PAGE.

The single most widely used analytical separation in all of protein chemistry, is reasonably useful for proteomic analysis. SDS-PAGE, officially sodium dodecyl sulfate polyacrylamide gel electrophoresis, uses a discontinuous polyacrylamide gel as a support medium and sodium dodecyl sulfate (SDS) to denature the proteins a very common method for separating proteins according to their size (length of polypeptide chain or molecular weight). One-dimensional-SDS-PAGE is done on gels in which the extent of cross-linking (i.e., polymerization of the acrylamide). There are many different methods of performing SDS-PAGE based on whether analysis is required in a native or denatured state and the size of the molecules being analysed.The first ever published paper by Laemmli (1970) using SDS-PAGE is the method of choice, while the method of (Schägger et al., 1987) is popular for polypeptides of 10 kDa or less.

SDS (also called lauryl sulfate) is an anionic detergent that binds to proteins according to a constant weight ratio (1.4 g SDS per gram of protein) independent of the nature of the protein (Reynolds and Tanford, 1970). The intrinsic charges of polypeptides are negligible compared to the negative charges provided by SDS, so that SDS-polypeptide complexes have essentially identical charge densities. Under this condition, proteins migrate in polyacrylamide gels strictly according to their size (Weber and Osborn, 1969). Separation of Proteins under Denaturing conditions, it is usually necessary to reduce disulphide bridges in proteins before they adopt the random-coil configuration necessary for separation by size: this is done with β-mercaptoethanol or dithiothreitol. In denaturing SDS-PAGE separations therefore, migration is determined not by intrinsic electrical charge of the polypeptide, but by molecular weight.

2.2.2 Determination of Molecular Weight.

This is done by SDS-PAGE of proteins or PAGE or agarose gel electrophoresis of nucleic acids of known molecular weight along with the protein or nucleic acid to be characterised. A linear relationship exists between the logarithm of the molecular weight

of an SDS-denatured polypeptide, or native nucleic acid, and its *Rf*. The Rf is calculated as the ratio of the distance migrated by the molecule to that migrated by a marker dye-front. A simple way of determining relative molecular weight by electrophoresis (*Mr*) is to plot a standard curve of distance migrated vs. log10MW for known samples, and read off the log*Mr* of the sample after measuring distance migrated on the same gel (Fig. 12).

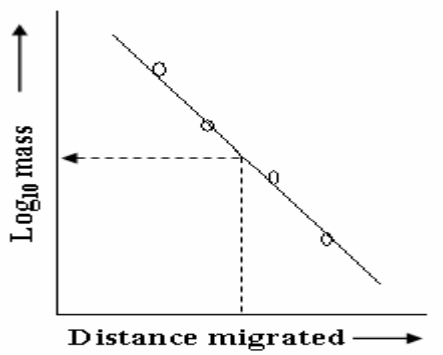


Fig. 12. Determining relative molecular weight of an unknown protein by comparison of its electrophoretic mobility with markers of known mass

2.2.3 Buffer Systems.

There are two types of buffer systems in electrophoresis, continuous and discontinuous. Most protein separations are performed using a "discontinuous" buffer system that significantly enhances the sharpness of the bands within the gel. During electrophoresis in a discontinuous gel system, an ion gradient is formed in the early stage of electrophoresis that causes all of the proteins to focus into a single sharp band. This occurs in a region of the gel that has larger pores so that the gel matrix does not retard the migration during the focusing or "stacking" event. Negative ions from the buffer in the tank then "outrun" the SDS-covered protein "stack" and eliminate the ion gradient so that the proteins subsequently separate by the sieving action in the lower, "resolving" region of the gel.

Many labs use a tris-glycine or "Laemmli" buffering system that stacks and resolves at a pH of ~8.3-9.0. These pHs promote disulfide bond formation between cysteine residues in the proteins, especially when they are present at high concentrations because the pKa of cysteine ranges from 8-9 and because reducing agent present in the loading buffer doesn't co-migrate with the proteins. Recent advances in buffering technology alleviate

this problem by resolving the proteins at a pH well below the pKa of cysteine (e.g., bis-Tris, pH 6.5) and include reducing agents (e.g. sodium bisulfite) that move into the gel ahead of the proteins to maintain a reducing environment. An additional benefit of using buffers with lower pHs is that the acrylamide gel is more stable so the gels can be stored for long periods of time before use.

2.2.4 Two-Dimensional SDS-PAGE.

O'Farrell in 1975 described Two-dimensional (2-D) gel electrophoresis, separates proteins in the first dimension according to their isoelectric point (*pI*), and in the second dimension according to their molecular weight (Fig. 13). It is possible to separate several hundred proteins from a total cellular extract. 2-D gel electrophoresis proved a very useful tool for proteomic studies when combined with microsequencing, amino acid composition and mass spectrometry.

The two different ways of separating proteins in the first dimension on the basis of their isoelectric point. The first proteins are separated in a pH gradient generated by applying an electric field to a gel containing a mixture of free carrier ampholytes (An der Lan and Chrambach, 1985). Carrier amphoiytes are low molecular mass components with both amino and carboxyl groups. According to the second way, the pH gradient is generated by a different type of chemicals, the immobilines (Bjellqvist et al., 1982). The immobilines are acrylamide derivatives carrying amino or carboxyl groups. These immobilines are copolymerized with the acrylamide gel matrix such that an immobilized pH gradient is generated.

2.2.5 Principles of Two-Dimensional Gel Electrophoresis.

Two-Dimensional gel electrophoresis of proteins is carried out under denaturing conditions. In order to separate proteins in the first dimension, proteins are solubilized in the presence of urea which essentially works by disrupting hydrogen bonds. This denaturant has the advantage that it does not affect the intrinsic charge of proteins so that it allows us to separate proteins only on the basis of their charge. When loaded on a

pH gradient of adequate porosity, proteins will migrate until they have no net charge, i.e when they reach the pH of the gradient corresponding to their isoelectric point (*PI*). After separation of proteins according to their charge, proteins are separated in a second dimension in the presence of sodium dodecylsulfate (SDS).

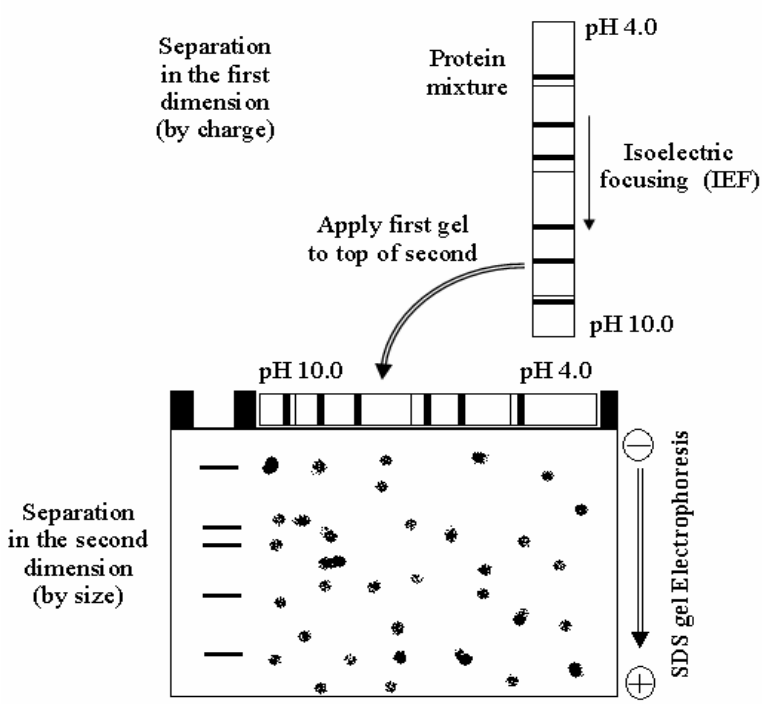


Fig. 13. Two-dimensional gel electrophoresis. The protein mixture is first subjected to isoelectric focusing (IEF) in one dimension and then to SDS-PAGE in the second dimension.

2.2.6 Isoelectric focusing (IEF)[6].

This technique is similar to the first step in 2D-SDS-PAGE. Isoelectric focusing (IEF) is an electrophoretic method that separates proteins according to their isoelectric points (*pI*). Proteins are amphoteric molecules; they carry either positive, negative, or zero net charge, depend-ing on their amino acid composition and the pH of their surroundings. The net charge of a protein is the sum of all the negative and positive charges of its amino acid side chains and amino- and carboxyl-termini. The isoelectric point is the specific pH at which the net charge of the protein is zero. Proteins are positively charged at pH values below their pI and negatively charged at pH values above their pI (Fig. 14).

[6] (www5.amershambiosciences.com)

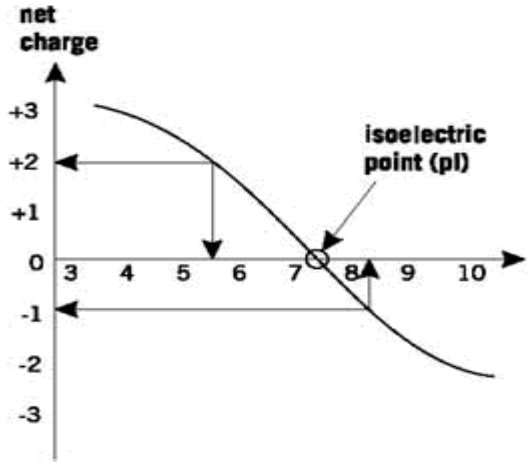


Fig.14. Net charge on a protein as a function of pH. In this example the protein has a net charge of +2 at pH 5.5, 0 at pH 7.5 (the isoelectric point), and -1 at pH 8.5.

The presence of a pH gradient is critical to the IEF technique. In a pH gradient, under the influence of an electric field a protein will move to the position in the gradient where its net charge is zero. A protein with a positive net charge will migrate toward the cathode, becoming progressively less positively charged as it moves through the pH gradient until it reaches its pI. A protein with a negative net charge will migrate toward the anode, becoming less negatively charged until it also reaches zero net charge. If a protein should diffuse away from its pI, it immediately gains charge and migrates back to its isoelectric position. This is the focusing effect of IEF, which concentrates proteins at their pIs and separates proteins with very small charge differences. Because the degree of resolution is determined by electric field strength, IEF is performed at high voltages (typically in excess of 1000 V). When the proteins have reached their final positions in the pH gradient, there is very little ionic movement in the system, resulting in a very low final current (typically below 1 mA).

One method for generating pH gradients in IEF gels relies on carrier ampholytes. Carrier ampholytes are small, soluble, amphoteric molecules with a high buffering capacity near their *pI*. Commercial carrier ampholyte mixtures comprise hundreds of individual polymeric species with pIs spanning a specific pH range. When a voltage is applied across a carrier ampholyte mixture (Fig 15) , the carrier ampholytes with the lowest *pI* (and the most negative charge) move toward the anode, and the carrier ampholytes with the highest pI (and the most positive charge) move toward the cathode.

The other carrier ampholytes align themselves between the extremes, according to their pIs, and buffer their environment to the corresponding pH. The result is a continuous pH gradient.

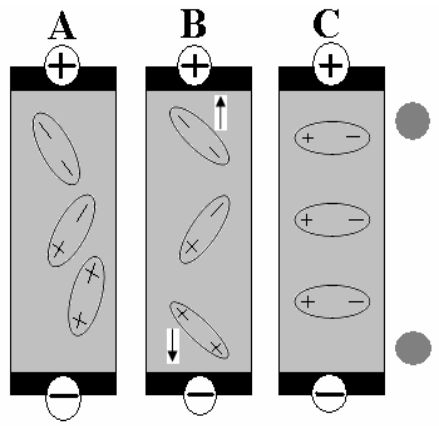


Fig 15. Creating a carrier ampholyte pH gradient. (A) No voltage applied. (B) Ampholytes and proteins move by electrophoresis when charged. (C) At isoelectric pH, ampholytes and proteins are focused.

IEF can be run in either a native or a denaturing mode. Native IEF is the more convenient option, as precast native IEF gels are available in a variety of pH gradient ranges. This method is also preferred when native protein is required, as when activity staining is to be employed. The use of native IEF, however, is often limited by the fact that many proteins are not soluble at low ionic strength or have low solubility close to their isoelectric point. In these cases, denaturing IEF is employed. Urea is the denaturant of choice, as this uncharged compound can solubilize many proteins not otherwise soluble under IEF conditions. Detergents and reducing agents are often used in conjunction with urea for more-complete unfolding and solubilization. Urea is not stable in aqueous solution, so precast IEF gels are not manufactured with urea. Dried precast gels are a convenient alternative; they have been cast, rinsed, and dried and can be rehydrated with urea, carrier ampholytes, and other additives before use. Although useful, carrier ampholytes have some limitations. Because the carrier ampholyte–generated gradient is dependent on the electric field, it breaks down when the field is removed. The pH gradients are also susceptible to gradient drift (or cathodic drift), a phenomenon in which there is a gradual decrease in pH at the cathodic (–) end of the gel and a flattening out of the pH at the anodic (+) end. For this reason it is important to not over focus the protein, because cathodic drift will increase over time. There can be

significant batch-to-batch and company-to-company variations in the properties of carrier ampholytes, which limits the reproducibility of focusing experiments. Another problem encountered with carrier ampholytes is their tendency to bind to the sample proteins, which may alter the migration of the protein and render the separation of carrier ampholytes from the focused protein difficult.

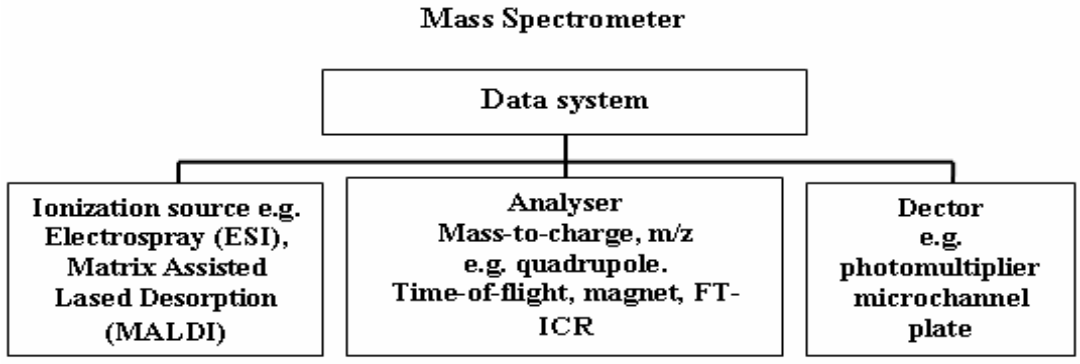
Acrylamido buffers are an alternative means to form pH gradients that circumvent most of the limitations of carrier ampholytes. Chemically, they are acrylamide derivatives of simple buffers and do not exhibit amphoteric behavior. The acrylic function of an acrylamido buffer co-polymerizes with the gel matrix and, by pouring a gel that incorporates an appropriate gradient of acrylamido buffers, an immobilized pH gradient (IPG) is formed. The protein sample can be applied immediately (no prefocusing is needed). The pH gradient is stable and does not drift in an electric field. Additionally, the gels are not susceptible to cathodic drift, because the buffers that form the pH gradient are immobilized within the gel matrix. Acrylamido buffers are available commercially with a specific pK value (or optimum pH buffering range), suitable for casting gradients from pH 3–10.

Because reproducible linear gradients with a slope as low as 0.01 pH units/cm can separate proteins with pI differences of 0.001 pH units, the resolution possible with immobilized pH gradient gels is 10–100 times greater than that obtained with carrier ampholyte–based IEF. IEF is best performed in a flatbed electrophoresis apparatus. This type of apparatus allows very effective cooling, which is necessary due to the high voltages employed for IEF. Manufactures offers a variety of precast gels for IEF, including ready to use carrier ampholyte gels, dried IPG gels, and dried acrylamide gels ready for reswelling in a mixture of carrier ampholytes and any other additives desired, such as detergent and denaturants.

2.3 Mass spectrometry[7].

Mass spectrometry is an analytical tool used for measuring the molecular mass of a sample. For large samples such as biomolecules, molecular masses can be measured to within an accuracy of 0.01% of the total molecular mass of the sample *i.e.* within a 4 Daltons (Da) or atomic mass units (amu) error for a sample of 40,000 Da. This is sufficient to allow minor mass changes to be detected, *e.g.* the substitution of one amino acid for another, or a post-translational modification. For small organic molecules the molecular mass can be measured to within an accuracy of 5 ppm or less, which is often sufficient to confirm the molecular formula of a compound, and is also a standard requirement for publication in a chemical journal. Structural information can be generated using certain types of mass spectrometers, usually those with multiple analysers which are known as tandem mass spectrometers. This is achieved by fragmenting the sample inside the instrument and analysing the products generated. This procedure is useful for the structural elucidation of organic compounds and for peptide or oligonucleotide sequencing.

Mass spectrometers can be divided into three fundamental parts, namely the ionisation source , the analyser , and the detector.



Simplified schematic of a mass spectrometer

[7] http://www.astbury.leeds.ac.uk

The sample has to be introduced into the ionisation source of the instrument. Once inside the ionisation source, the sample molecules are ionised, because ions are easier to manipulate than neutral molecules. These ions are extracted into the analyser region of the mass spectrometer where they are separated according to their mass (*m*) -to-charge (*z*) ratios (*m/z*) . The separated ions are detected and this signal sent to a data system where the *m/z* ratios are stored together with their relative abundance for presentation in the format of a *m/z* spectrum .

The analyser and detector of the mass spectrometer, and often the ionisation source too, are maintained under high vacuum to give the ions a reasonable chance of travelling from one end of the instrument to the other without any hindrance from air molecules. The entire operation of the mass spectrometer, and often the sample introduction process also, is under complete data system control on modern mass spectrometers.

2.3.1 Sample introduction.

The method of sample introduction to the ionisation source often depends on the ionisation method being used, as well as the type and complexity of the sample.

The sample can be inserted directly into the ionisation source, or can undergo some type of chromatography *en route* to the ionisation source. This latter method of sample introduction usually involves the mass spectrometer being coupled directly to a high pressure liquid chromatography (HPLC), gas chromatography (GC) or capillary electrophoresis (CE) separation column, and hence the sample is separated into a series of components which then enter the mass spectrometer sequentially for individual analysis.

2.3.2 Methods of sample ionisation

Many ionisation methods are available and each has its own advantages and disadvantages The ionisation method to be used depend on the type of sample under investigation and the mass spectrometer available. The ionisation methods used for the majority of biochemical analyses are Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI*)*.

2.3.3 Analysis and Separation of Sample Ions.

The main function of the mass analyser is to separate , or resolve , the ions formed in the ionisation source of the mass spectrometer according to their mass-to-charge (*m/z*) ratios. There are a number of mass analysers currently available, the better known of which include quadrupoles **,** time-of-flight (TOF) analysers, magnetic sectors , and both Fourier transform and quadrupole ion traps .

These mass analysers have different features, including the *m/z* range that can be covered, the mass accuracy, and the achievable resolution. The compatibility of different analysers with different ionisation methods varies. For example, all of the analysers listed above can be used in conjunction with electrospray ionisation, whereas MALDI is not usually coupled to a quadrupole analyser.

Tandem (MS-MS) mass spectrometers are instruments that have more than one analyser and so can be used for structural and sequencing studies. Two, three and four analysers have all been incorporated into commercially available tandem instruments, and the analysers do not necessarily have to be of the same type, in which case the instrument is a hybrid one. More popular tandem mass spectrometers include those of the quadrupole-quadrupole, magnetic sector-quadrupole , and more recently, the quadrupole-time-of-flight geometries.

2.3.4 Detection and recording of sample ions.

The detector monitors the ion current, amplifies it and the signal is then transmitted to the data system where it is recorded in the form of mass spectra . The *m/z* values of the ions are plotted against their intensities to show the number of components in the sample, the molecular mass of each component, and the relative abundance of the various components in the sample.

The type of detector is supplied to suit the type of analyser; the more common ones are the photomultiplier , the electron multiplier and the micro-channel plate detectors.

2.4 Electrospray.

Electrospray mass spectrometry (ESMS) has been developed for use in biological mass spectrometry by (Fenn et al., 1989).

2.4.1 Electrospray ionisation.

Electrospray Ionisation (ESI) is one of the Atmospheric Pressure Ionisation (API) techniques and is well-suited to the analysis of polar molecules ranging from less than 100 Da to more than 1,000,000 Da in molecular mass (Fig. 16).

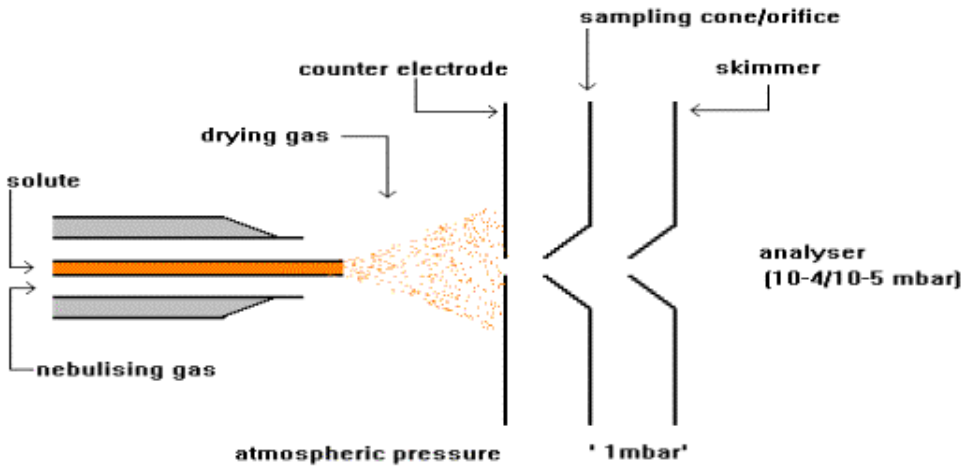


Fig. 16. Standard electrospray ionisation source (Platform II)

During standard electrospray ionisation (Yamashita and Fenn, 1984), the sample is dissolved in a polar, volatile solvent and pumped through a narrow, stainless steel capillary (75 - 150 micrometers i.d.) at a flow rate of between 1 µL/min and 1 mL/min. A high voltage of 3 or 4 kV is applied to the tip of the capillary, which is situated within the ionisation source of the mass spectrometer, and as a consequence of this strong electric field, the sample emerging from the tip is dispersed into an aerosol of highly charged droplets, a process that is aided by a co-axially introduced nebulising gas flowing around the outside of the capillary (Fig. 17). This gas, usually nitrogen, helps to direct the spray emerging from the capillary tip towards the mass spectrometer. The

charged droplets diminish in size by solvent evaporation, assisted by a warm flow of nitrogen known as the drying gas which passes across the front of the ionisation source. Eventually charged sample ions, free from solvent, are released from the droplets, some of which pass through a sampling cone or orifice into an intermediate vacuum region, and from there through a small aperture into the analyser of the mass spectrometer, which is held under high vacuum. The lens voltages are optimised individually for each sample.

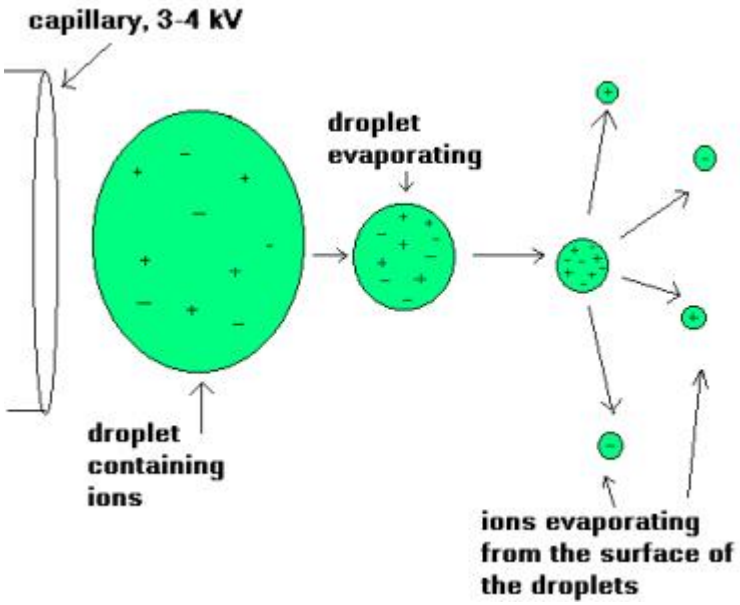


Fig. 17. The electrospray ionisation process

2.4.2 Nanospray ionisation.

Nanospray ionisation (Wilm and Mann, 1996) is a low flow rate version of electrospray ionisation. A small volume (1-4 microL) of the sample dissolved in a suitable volatile solvent, at a concentration of ca. 1 - 10 pmol/microL, is transferred into a miniature sample vial. A reasonably high voltage (ca. 700 - 2000 V) is applied to the specially manufactured gold-plated vial resulting in sample ionisation and spraying. The flow rate of solute and solvent using this procedure is very low, 30 - 1000 nL/min, and so not only is far less sample consumed than with the standard electrospray ionisation technique, but also a small volume of sample lasts for several minutes, thus enabling

multiple experiments to be performed. A common application of this technique is for a protein digest mixture to be analysed to generate a list of molecular masses for the components present, and then each component to be analysed further by tandem mass spectrometric (MS-MS) amino acid sequencing techniques.

ESI and nanospray ionisation are very sensitive analytical techniques but the sensitivity deteriorates with the presence of non-volatile buffers and other additives, which should be avoided as far as possible.

In positive ionisation mode, a trace of formic acid is often added to aid protonation of the sample molecules; in negative ionisation mode a trace of ammonia solution or a volatile amine is added to aid deprotonation of the sample molecules. Proteins and peptides are usually analysed under positive ionisation conditions and saccharides and oligonucleotides under negative ionisation conditions. In all cases, the *m/z* scale must be calibrated by analysing a standard sample of a similar type to the sample being analysed (e.g. a protein calibrant for a protein sample), and then applying a mass correction.

2.5 Matrix assisted laser desorption ionisation (MALDI).

Matrix Assisted Laser Desorption Ionisation (MALDI) (Hillenkamp et al., 1991) (Fig. 18) deals well with thermolabile, non-volatile organic compounds especially those of high molecular mass and is used successfully in biochemical areas for the analysis of proteins, peptides, glycoproteins, oligosaccharides, and oligonucleotides. It is relatively straightforward to use and reasonably tolerant to buffers and other additives. The mass accuracy depends on the type and performance of the analyser of the mass spectrometer, but most modern instruments should be capable of measuring masses to within 0.01% of the molecular mass of the sample, at least up to ca. 40,000 Da.

MALDI is based on the bombardment of sample molecules with a laser light to bring about sample ionisation. The sample is pre-mixed with a highly absorbing matrix compound for the most consistent and reliable results, and a low concentration of sample to matrix works best. The matrix transforms the laser energy into excitation energy for the sample, which leads to sputtering of analyte and matrix ions from the surface of the mixture. In this way energy transfer is efficient and also the analyte

molecules are spared excessive direct energy that may otherwise cause decomposition. Most commercially available MALDI mass spectrometers now have a pulsed nitrogen laser of wavelength 337 nm.

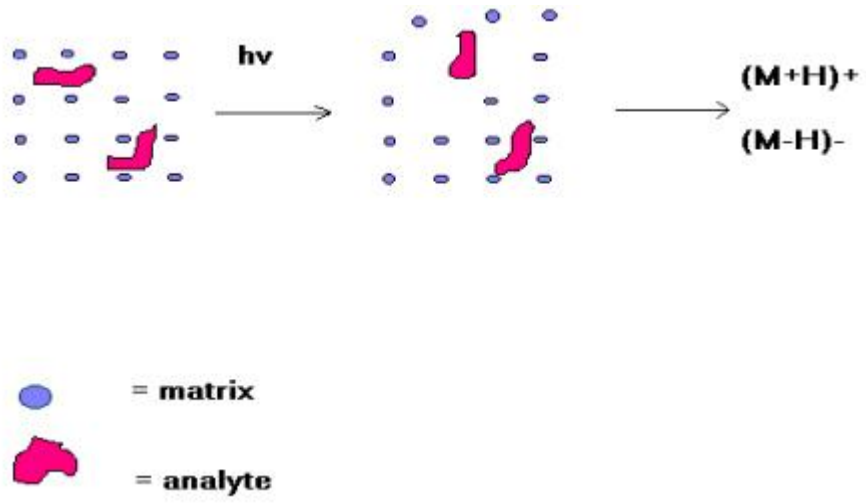


Fig. 18. Matrix assisted laser desorption ionisation (MALDI)

The sample to be analysed is dissolved in an appropriate volatile solvent, usually with a trace of trifluoroacetic acid if positive ionisation is being used, at a concentration of ca. 10 pmol/μL and an aliquot **(**1-2 μL**)** of this removed and mixed with an equal volume of a solution containing a vast excess of a matrix. A range of compounds is suitable for use as matrices: sinapinic acid is a common one for protein analysis while alpha-cyano-4-hydroxycinnamic acid is often used for peptide analysis. An aliquot (1-2 μL) of the final solution is applied to the sample target which is allowed to dry prior to insertion into the high vacuum of the mass spectrometer. The laser is fired, the energy arriving at the sample/matrix surface optimised, and data accumulated until a *m/z* spectrum of reasonable intensity has been amassed (Fig. 19). The time-of-flight analyser separates ions according to their mass(m)-to-charge(z) (*m/z*) ratios by measuring the time it takes for ions to travel through a field free region known as the flight, or drift, tube. The heavier ions are slower than the lighter ones.

The *m/z* scale of the mass spectrometer is calibrated with a known sample that can either be analysed independently (external calibration) or pre-mixed with the sample and matrix (internal calibration).

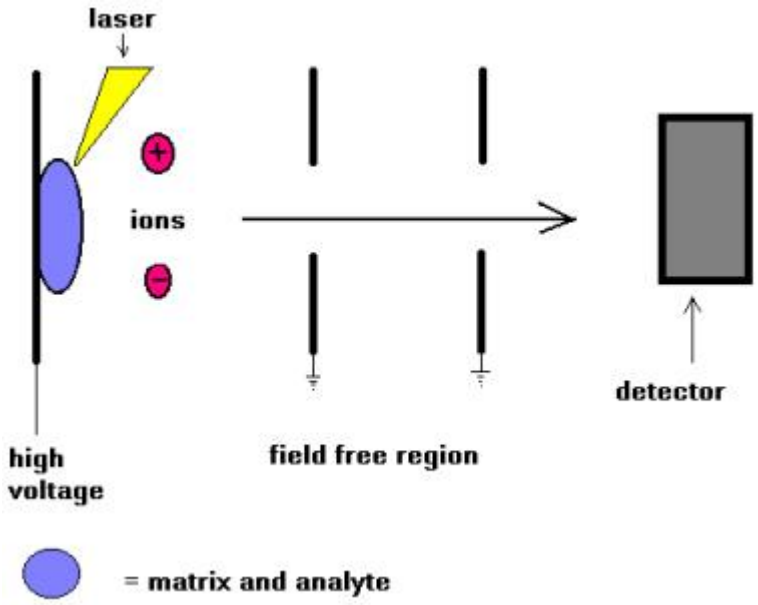


Fig. 19. Simplified schematic of MALDI-TOF mass spectrometry (linear mode)

MALDI is also a "soft" ionisation method and so results predominantly in the generation of singly charged molecular-related ions regardless of the molecular mass, hence the spectra are relatively easy to interpret. Fragmentation of the sample ions does not usually occur.

In positive ionisation mode the protonated molecular ions (M^+H^+) are usually the dominant species, although they can be accompanied by salt adducts, a trace of the doubly charged molecular ion at approximately half the *m/z* value, and/or a trace of a dimeric species at approximately twice the *m/z* value. Positive ionisation is used in general for protein and peptide analyses (Fig. 20).

In negative ionisation mode the deprotonated molecular ions (M-H-) are usually the most abundant species, accompanied by some salt adducts and possibly traces of dimeric or doubly charged materials. Negative ionisation can be used for the analysis of oligonucleotides and oligosaccharides.

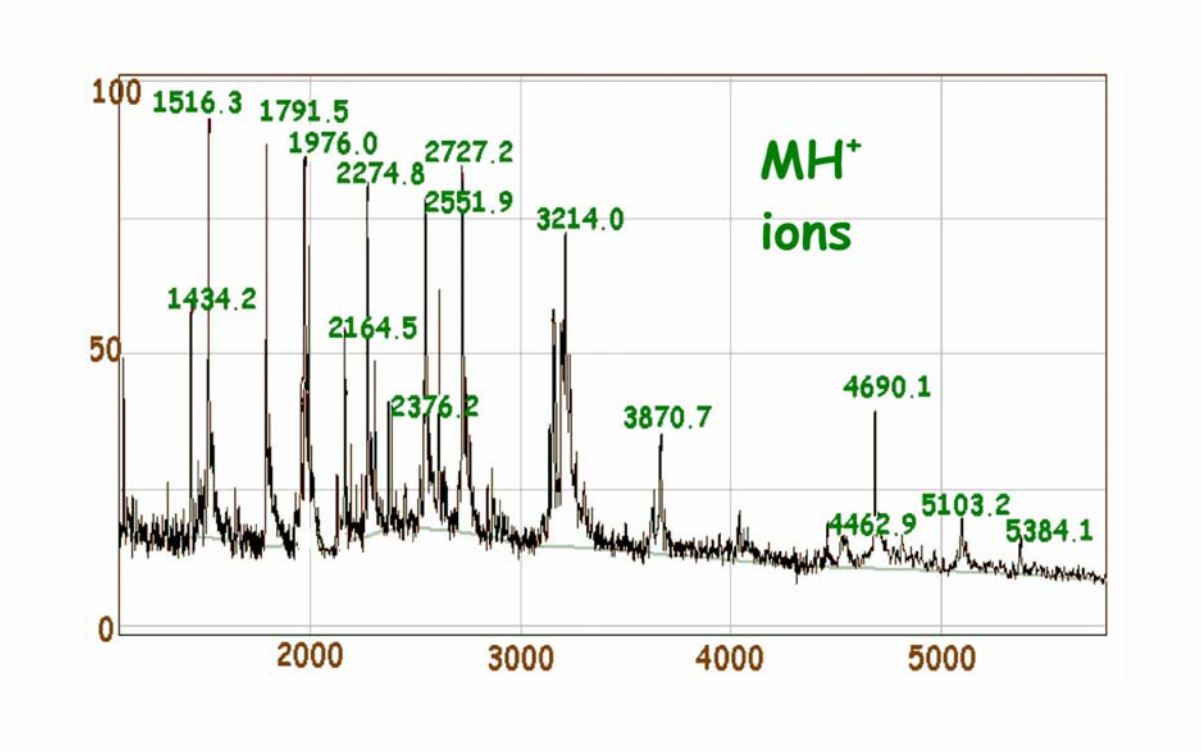


Fig. 20. Positive ionisation MALDI m/z spectrum of a peptide mixture using alpha-cyano-4-hydroxycinnamic acid as matrix

2.5.1 Positive or negative ionisation?

If the sample has functional groups that readily accept a proton (H^+) then positive ion detection is used e.g. amines $R\text{-}NH_2 + H+ = R\text{-}NH_3+$ as in proteins or peptides.

If the sample has functional groups that readily lose a proton then negative ion detection is used e.g. carboxylic acids $R\text{-}CO_2H = R\text{-}CO_2-$ and alcohols R-OH = R-O- as in saccharides or oligonucleotides

2.6 Tandem mass spectrometry (MS-MS).

Structural and sequence information from mass spectrometry.

2.6.1 Tandem mass spectrometry.

Tandem mass spectrometry (MS-MS) is used to produce structural information about a compound by fragmenting specific sample ions inside the mass spectrometer and identifying the resulting fragment ions. This information can then be pieced together to generate structural information regarding the intact molecule. Tandem mass

spectrometry also enables specific compounds to be detected in complex mixtures on account of their specific and characteristic fragmentation patterns.

A tandem mass spectrometer is a mass spectrometer that has more than one analyser, in practice usually two. The two analysers are separated by a collision cell into which an inert gas (e.g. argon, xenon) is admitted to collide with the selected sample ions and bring about their fragmentation. The analysers can be of the same or of different types, the most common combinations being:

- quadrupole - quadrupole
- magnetic sector - quadrupole
- magnetic sector - magnetic sector
- quadrupole - time-of-flight.

Fragmentation experiments can also be performed on certain single analyser mass spectrometers such as ion trap and time-of-flight instruments, the latter type using a post-source decay experiment to effect the fragmentation of sample ions.

2.6.2 Tandem mass spectrometry analyses.

The basic modes of data acquisition for tandem mass spectrometry experiments are as follows:

2.6.3 Product or daughter ion scanning:

The first analyser is used to select user-specified sample ions arising from a particular component; usually the molecular-related (i.e. $(M+H)^+$ or $(M-H)^-$) ions. These chosen ions pass into the collision cell, are bombarded by the gas molecules which cause fragment ions to be formed, and these fragment ions are analysed i.e. separated according to their mass to charge ratios, by the second analyser. All the fragment ions arise directly from the precursor ions specified in the experiment, and thus produce a fingerprint pattern specific to the compound under investigation.

This type of experiment is particularly useful for providing structural information concerning small organic molecules and for generating peptide sequence information.

2.6.4 Precursor or parent ion scanning.

The first analyser allows the transmission of all sample ions, whilst the second analyser is set to monitor specific fragment ions, which are generated by bombardment of the sample ions with the collision gas in the collision cell. This type of experiment is particularly useful for monitoring groups of compounds contained within a mixture which fragment to produce common fragment ions, e.g. glycosylated peptides in a tryptic digest mixture, aliphatic hydrocarbons in an oil sample, or glucuronide conjugates in urine.

2.6.5 Constant neutral loss scanning.

This involves both analysers scanning, or collecting data, across the whole *m/z* range, but the two are off-set so that the second analyser allows only those ions which differ by a certain number of mass units (equivalent to a neutral fragment) from the ions transmitted through the first analyser. e.g. This type of experiment could be used to monitor all of the carboxylic acids in a mixture. Carboxylic acids tend to fragment by losing a (neutral) molecule of carbon dioxide, CO_2, which is equivalent to a loss of 44 Da or atomic mass units. All ions pass through the first analyser into the collision cell. The ions detected from the collision cell are those from which 44 Da have been lost.

2.6.6 Selected/multiple reaction monitoring

Both of the analysers are static in this case as user-selected specific ions are transmitted through the first analyser and user-selected specific fragments arising from these ions are measured by the second analyser. The compound under scrutiny must be known and have been well-characterised previously before this type of experiment is undertaken. This methodology is used to confirm unambiguously the presence of a compound in a matrix e.g. drug testing with blood or urine samples. It is not only a highly specific method but also has very high sensitivity.

2.6.7 Peptide Sequencing by Tandem Mass Spectrometry.

The most common usage of MS-MS in biochemical areas is the product or daughter ion scanning experiment which is particularly successful for peptide and nucleotide sequencing.

2.6.8 Peptide sequencing: H_2N-CH(R')-CO-NH-CH(R")-CO_2H

Peptides fragment in a reasonably well-documented manner (Roepstorff and Fohlman, 1984; Johnson and Biemann, 1989). The protonated molecules fragment along the peptide backbone and also show some side-chain fragmentation with certain instruments

There are three different types of bonds that can fragment along the amino acid backbone: the NH-CH**,** CH-CO, and CO-NH bonds. Each bond breakage gives rise to two species, one neutral and the other one charged, and only the charged species is monitored by the mass spectrometer. The charge can stay on either of the two fragments depending on the chemistry and relative proton affinity of the two species. Hence there are six possible fragment ions for each amino acid residue and these are labelled as in the diagram, with the a, b, and c" ions having the charge retained on the N-terminal fragment, and the x, y", and z ions having the charge retained on the C-terminal fragment. The most common cleavage sites are at the CO-NH bonds which give rise to the b and/or the y" ions. The mass difference between two adjacent b ions, or y"; ions, is indicative of a particular amino acid residue (Fig. 21).

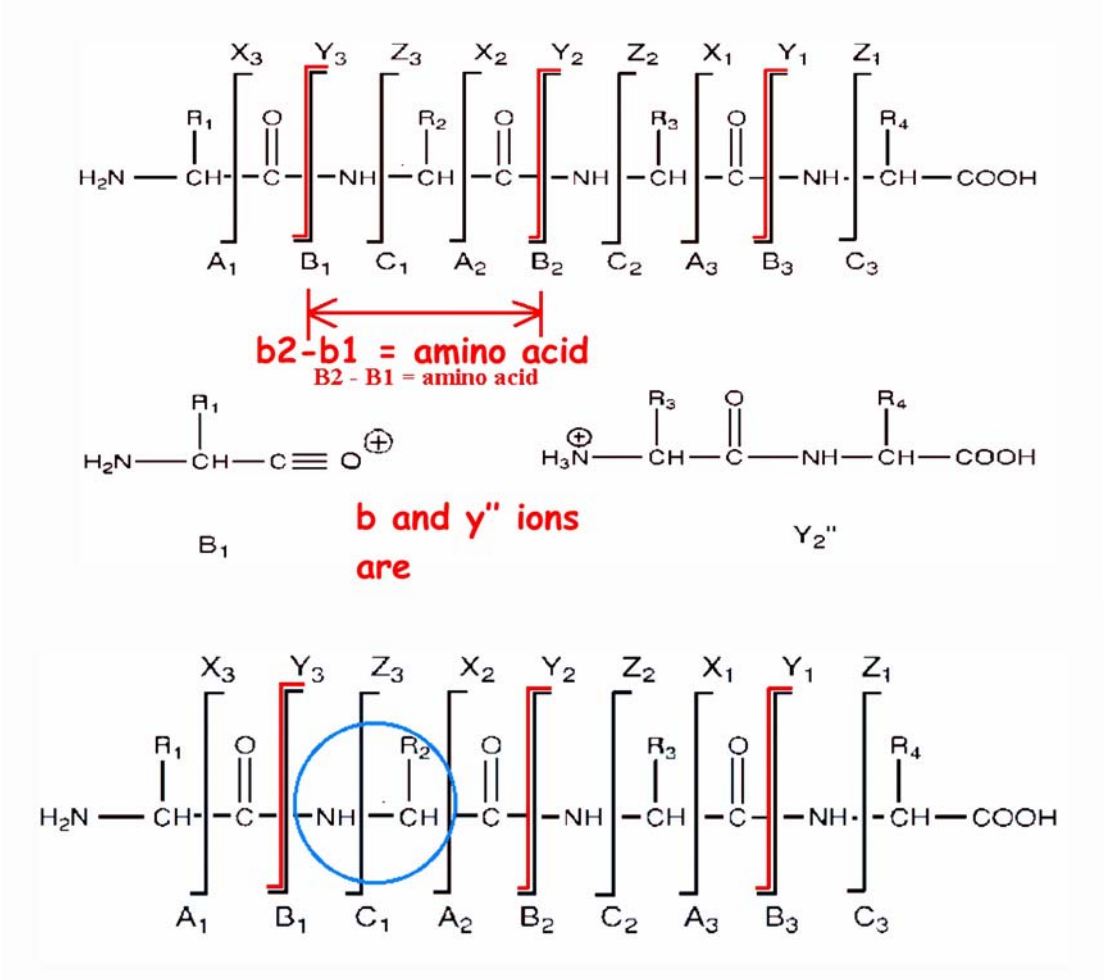


Fig. 21. Peptide sequencing by tandem mass spectrometry - backbone cleavages

The extent of side-chain fragmentation detected depends on the type of analysers used in the mass spectrometer. A magnetic sector - magnetic sector instrument will give rise to high energy collisions resulting in many different types of side-chain cleavages. Quadrupole - quadrupole and quadrupole - time-of-flight mass spectrometers generate low energy fragmentations with fewer types of side-chain fragmentations.

Immonium ions (labelled "i") appear in the very low *m/z* range of the MS-MS spectrum. Each amino acid residue leads to a diagnostic immonium ion, with the exception of the two pairs leucine (L) and iso-leucine (I), and lysine (K) and glutamine (Q), which produce immonium ions with the same m/z ratio, i.e. *m/z* 86 for I and L, *m/z* 101 for K and Q. The immonium ions are useful for detecting and confirming many of the amino acid residues in a peptide, although no information regarding the position of

these amino acid residues in the peptide sequence can be ascertained from the immonium ions.

2.6.9 A protein identification study would proceed as follows:

- a. The **protein** under investigation would be analysed by mass spectrometry to generate a molecular mass to within an accuracy of 0.01%.
- b. The protein would then be **digested** with a suitable enzyme. **Trypsin** is useful for mass spectrometric studies because each proteolytic fragment contains a basic **arginine (R)** or **lysine (K)** amino acid residue, and thus is eminently suitable for positive ionisation mass spectrometric analysis. The digest mixture is analysed - without prior separation or clean-up - by mass spectrometry to produce a rather complex spectrum from which the molecular weights of all of the proteolytic fragments can be read. This spectrum, with its molecular weight information, is called a **peptide map**. (If the protein already exists on a **database**, then the peptide map is often sufficient to confirm the protein.) For these experiments the mass spectrometer would be operated in the **"MS" mode**, whereby the sample is sprayed and ionised from the nanospray needle and the ions pass through the sampling cone, skimmer lenses, *Rf* hexapole focusing system, and the first (quadrupole) analyser. The quadrupole in this instance is not used as an analyser, merely as a lens to focus the ion beam into the second (time-of-flight) analyser which separates the ions according to their mass-to-charge ratio.

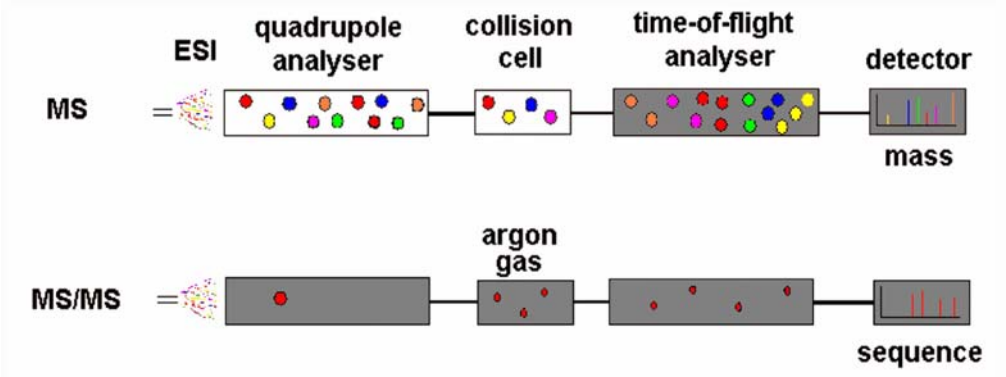


Fig. 22. Q-TOF mass spectrometer operating in MS (upper) and MS/MS mode (lower) modes.

- c. With the digest mixture still spraying into the mass spectrometer, the Q-Tof mass spectrometer is switched into **"MS/MS" mode** (Fig. 22) The protonated molecular ions of each of the digest fragments can be independently selected and transmitted through the quadrupole analyser, which is now used as an analyser to transmit solely the ions of interest into the **collision cell** which lies inbetween the first and second analysers. An inert gas such as argon is introduced into the collision cell and the sample ions are bombarded by the collision gas molecules which cause them to fragment. The optimum collision cell conditions vary from peptide to peptide and must be optimised for each one. The **fragment (**or **daughter or product**) **ions** are then analysed by the second (time-of-flight) analyser. In this way an **MS/MS spectrum** is produced showing all the **fragment ions** that arise directly from the chosen **parent** or **precursor ions** for a given peptide component.

  An **MS/MS daughter** (or **fragment**, or **product**) ion spectrum is produced for each of the components identified in the proteolytic digest. Varying amounts of sequence information can be gleaned from each fragmentation spectrum, and the spectra need to be interpreted carefully. Some of the processing can be automated, but in general the **processing** and **interpretation** of spectra will take longer than the data acquisition if accurate and reliable data are to be generated.

The amount of sequence information generated will vary from one peptide to another, Some peptide sequences will be confirmed totally, other may produce a partial sequence of, say, 4 or 5 amino acid residues. Often sequence "tag" of 4 or 5 residues is sufficient to search a protein database and confirm the identity of the protein.

2.7 Peptide sequencing in summary.

Peptides fragment along the amino acid backbone to give sequence information.

Peptides ca. 2500 Da or less produce the most useful data.

The amount of sequence information varies from one peptide to another. Some peptides can generate sufficient information for a full sequence to be determined; others may generate a partial sequence of 4 or 5 amino acids.

A protein digest can be analysed as an entire reaction mix, without any separation of the products, from which individual peptides are selected and analysed by the mass spectrometer to generate sequence information.

About 4 μL of solution is required for the analysis of the digest mixture, with a concentration based on the original protein of ca. 1-10 pmol/μL. MS/MS sequencing is a sensitive technique consuming little sample.

Sometimes the full protein sequence can be verified; some proteins generate sufficient information to cover only part of the sequence. 70 - 80% coverage is reasonable.

Often a sequence "tag" of 4/5 amino acids from a single proteolytic peptide is sufficient to identify the protein from a database.

The final point in this summary means that mass spectrometers have been found to be extremely useful for **proteomic** studies, as illustrated below.

The **proteomics procedure** usually involves excising individual spots from a **2-D gel** and independently **enzymatically digesting** the protein(s) contained within each spot and/or band, before analysing the digest mixture by mass spectrometer in the manner outlined above. Electrospray ionisation or MALDI could be used at this step.

The initial **MS spectrum** determining the **molecular masses** of all of the components in the digest mixture can often provide sufficient information to search a **database** using just several of the molecular weights from this **peptide map.**

If the database search is not fruitful, either because the protein has not been catalogued, is previously uncharacterised, or the data are not accurate or comprehensive enough to distinguish between several entries in the database, then further information is required.

This can be achieved by sample clean-up and then MS/MS studies to determine the amino acid sequences of the individual proteolytic peptides contained in the digest mixture, with which further database searching can be carried out.

2.8 Peptide Mass Fingerprinting (PMF).

PMF is an analytical technique for protein identification that was developed by John Yates and colleagues (Griffin et al., 1995). In this method, a "mass fingerprint " is obtained of a protein enzymatically degraded with a sequence-specific protease such as trypsin. This set of masses, typically obtained by MALDI-TOF, is then compared to the theoretically expected tryptic peptide masses for each entry in the database. The proteins can be ranked according to the number of peptide matches (Fig. 23).

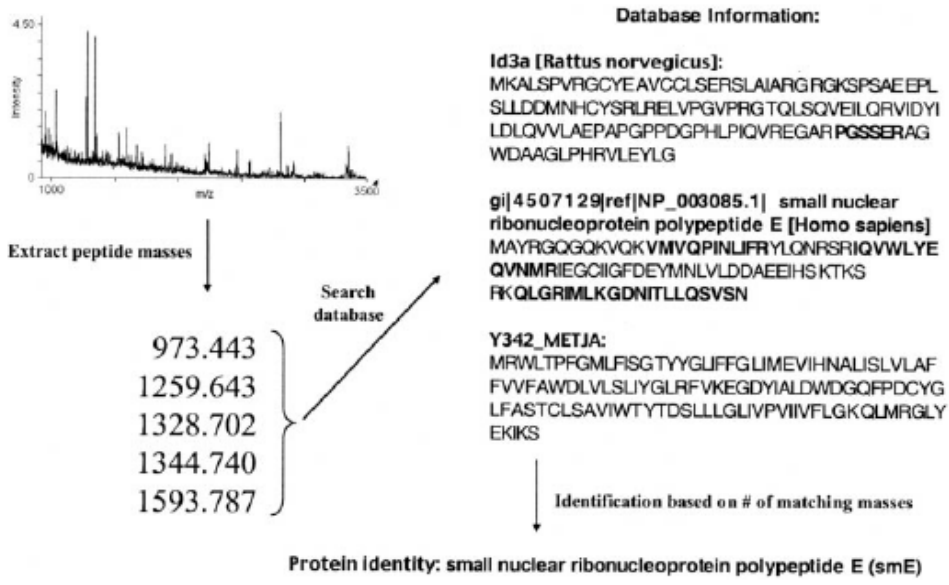


Fig. 23. Peptide mass fingerprinting

More sophisticated scoring algorithms take the mass accuracy and the percentage of the protein sequence covered into account and attempt to calculate a level of confidence for the match (Berndt et al., 1999; Perkins et al., 1999; Eriksson et al., 2000). Other factors can also be included, such as the fact that larger peptides are less frequent in the database and should therefore count more when matched. The accuracy obtained in the measurement of peptide mass strongly influences the specificity of the search (Jensen . et al., 1996; Clauser et al., 1999)**.** When high mass accuracy (10 to 50 ppm) is achieved,

as a rule at least five peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered for an unambiguous identification. After a match has been found, a second-pass search is performed to correlate remaining peptides with the database sequence of the match, taking into account possible modifications. Mass fingerprinting can also resolve simple protein mixtures, consisting of several proteins within a roughly comparable amount. For example, databases can be searched iteratively by removing the peptides associated with an unambiguous match (Jensen et al., 1997). Generally, peptide mass fingerprinting is used for the rapid identification of a single protein component. Protein sequences need to be in the database in substantially full length. Isoforms can be differentiated from each other, if peptides covering the sequence differences appear in the peptide map. If proteins from organisms with fully sequenced genomes can be identified with a 50–90% success rate when at least a few hundred femtomoles of gel-separated protein are present (Jensen et al., 1997).

3. MICROSATELLITE MARKERS FOR DNA SEQUENCE ANALYSIS OF OLIVE CULTIVERS[8].

In recent times, a great effort has been made in order to molecularly characterize olive germplasm by means of different classes of biochemical markers. To discriminate and to evaluate relationships of numerous olive varieties several studies have been made by use of isozymes (Lumaret et al., 2004; Perri et al., 1995), RAPDs (Belaj et al., 2002; Belaj et al., 2003a; Fodale et al., 2005; Lombardo et al., 2003; Lombardo et al., 2004; Perri et al., 1995), AFLPs (Angiolillo et al., 1999), ISSR and sequence analysis of nuclear ribosomal internal transcribed spacer 1 (Hesse et al., 2000), mitochondrial RFLPs (Besnard and Bervillé 2000) and chloroplast RFLP (Besnard and Berville 2002; Besnard et al., 2002).

More recently, several microsatellites *loci* have been isolated from olive (Carriero et al., 2002; Cipriani et al., 2002; De La Rosa et al., 2002, Rallo et al., 2000; Sefc et al., 2000) and, at present, it is very common utilizing these markers associated with other molecular markers (Belaj et al. 2003b; Khadari et al. 2003; La Rosa et al. 2003; Lombardo et al. 2005; Wu et al. 2004).

The microsatellites or simple sequence repeats (SSRs) shows several advantages because they are easily amplified by PCR and thus can be used on non-invasively sampled material and they are also highly-reproducible among different labs. The SSRs are co-dominant markers, showing a large number of polymorphisms per primer set and often multiple alleles in a variety, which can be highly informative.

In this thesis we presents the results of an experiment whose aim is to evaluate the extent to which SSR markers analysis was able to obtain the molecular characterisation and identification of traditional olive varieties from Abruzzo, Molise, Sicily and Apulia Regions (Italy), which is a major area for olive cultivation in Italy, with a strategic geographical location in the Mediterranean basin. We also show, by DNA sequence analysis of microsatellite alleles, that a very high degree of polymorphism is present and

[8] Part of this research work published (No. 1,3,4,7 & 8 Appendix)

also that alleles with an identical length may have a different repeated motif. Further analysis on the genetic significance of these polymorphisms is needed.

3.1 Olive SSR loci and genetic diversity.

Initial screening based on microsatellite loci and the genetic distance measures were used to draw inference on genetic associations among the samples belongs to Abruzzo, Molise, Sicily and Apulia Regions of Italy.

3.1.1 ABRUZZO AND MOLISE.

3.1.1 Characterization of olive SSR loci and genetic diversity.

The 9 primer pairs were selected based on the number of polymorphisms amplified and for their high reproducibility. An average of 4.2 alleles per locus were amplified, ranging from 2.0 at UDO01 to 6.0 at UDO39. A total of 38 alleles over 9 loci were observed (Table 2), which was comparable to the observed number of alleles among olive varieties reported by Rallo et al., (2000) and somewhat lower than that obtained by Khadari et al., (2003). It was also lower than that reported among different species of the olive family (Lefort et al., 1999). The effective number of alleles (Kimura and Crow 1964) ranged from 1.5 at the GAPU71A locus to 4.0 at GAPU103A, showing a mean value of 2.9 (Table 2). The shortest allele among the 9 loci was 125 base pairs, allele *A* at GAPU71B, while the longest was 243 base pairs, namely allele *F* at UDO39.

The lowest allele frequency (0.02) was observed in allele *C* of the GAPU59 locus in Gnagnaro and Posolella and (0.03) in allele *A* at Gapu103A locus in Gentile nera di Colletorto, Olivoce and in Noccioluta. One allele *B* of the low polymorphic locus Gapu71A showed the highest frequency (0.79).

The observed heterozygosity (*Ho*) for the 44 varieties ranged from 0.00 at UDO01 to 0.82 at GAPU71B with a mean value of 0.50. The expected heterozygosity (*He*) ranged from 0.34 at GAPU71A to 0.76 at GAPU103A, with a mean value of 0.64 (Table 3). As expected for vegetatively propagated plants, we found a moderately high overall loci *Fst*

(Hartl and Clark 1989) mean value of 0.60.

Table 2. SSR loci in 44 olive accessions of the Abruzzo and Molise region. For each locus, the number of alleles detected (*Nd*) and effective alleles (*Ne*) are reported.

Locus	*Size range (base pairs)*	*Nd*	*Ne*
GAPU59	208-222	5.0	3.2
GAPU71A	210-224	3.0	1.5
GAPU71B	125-145	5.0	3.3
GAPU103A	136-186	5.0	4.0
UDO01	140-143	2.0	2.0
UDO03	135-202	5.0	3.1
UDO12	166-193	3.0	2.9
UDO28	154-210	4.0	3.5
UDO39	170-243	6.0	3.0
Total	125-243	38.0	26.5
Mean	-	4.2	2.9
Std. Dev.	-	1.3	0.7

Table 3. Observed heterozygosity (*Ho*) and expected heterozygosity (*He*) values for olive SSR loci.

Locus	*Ho*	*He*
GAPU59	0.75	0.69
GAPU71A	0.41	0.34
GAPU71B	0.82	0.70
GAPU103A	0.50	0.76
UDO01	0.00	0.50
UDO03	0.18	0.68
UDO12	0.79	0.66
UDO28	0.79	0.72
UDO39	0.27	0.67
Mean	0.50	0.64
Std. Dev.	0.31	0.13

3.1.2 Genetic relationships between olive varieties from the Abruzzo region.

The polymorphism observed among accessions was appropriate to differentiate varieties. The fingerprinting patterns obtained allowed unequivocal identification of each variety (Table 4 Annex-I). Therefore, the highest values of genetic identity (data not shown) were observed between the accessions Caprina Casalanguida - Caprina Vastese (0.89) and between the varieties Nebbio di Chieti - Puntella (0.79). The lowest values of genetic identity were obtained Gentile de L'Aquila – Tortiglione (0.15) and Posolella – Rustica (0.15).

Nei's genetic distances were utilized to obtain a phenogram based on UPGMA algorithm of clustering (Fig. 24). Two distinct clusters of olive varieties were clearly recognizable. The first cluster included 52% of analyzed varieties and the second one 48%.

The first cluster (*1*) contained 12 olive accessions: Caprina Casalanguida, Caprina Vastese, Toccolana, Nebbio di Chieti, Puntella, Posola, Posolella, Carpinetana, Ghiandaro, Castiglionese, Peperella and Tortiglione. This cluster included most of the olive accessions situated along the hillside facing the east, towards the coast of the Adriatic Sea, in provinces of Chieti, Pescara and Teramo. The second cluster (*2*) containing 11 accessions: Carbonchio, Olivoce, Dritta, Gentile de L'Aquila, Gentile di Chieti, Intosso, Nebbio di Pescara, Olivastro di Bucchianico, Olivastro Frentano, Precoce and Rustica. This cluster included most of the olive accessions, resistance to low temperatures, situated in the central areas of the Abruzzo region.

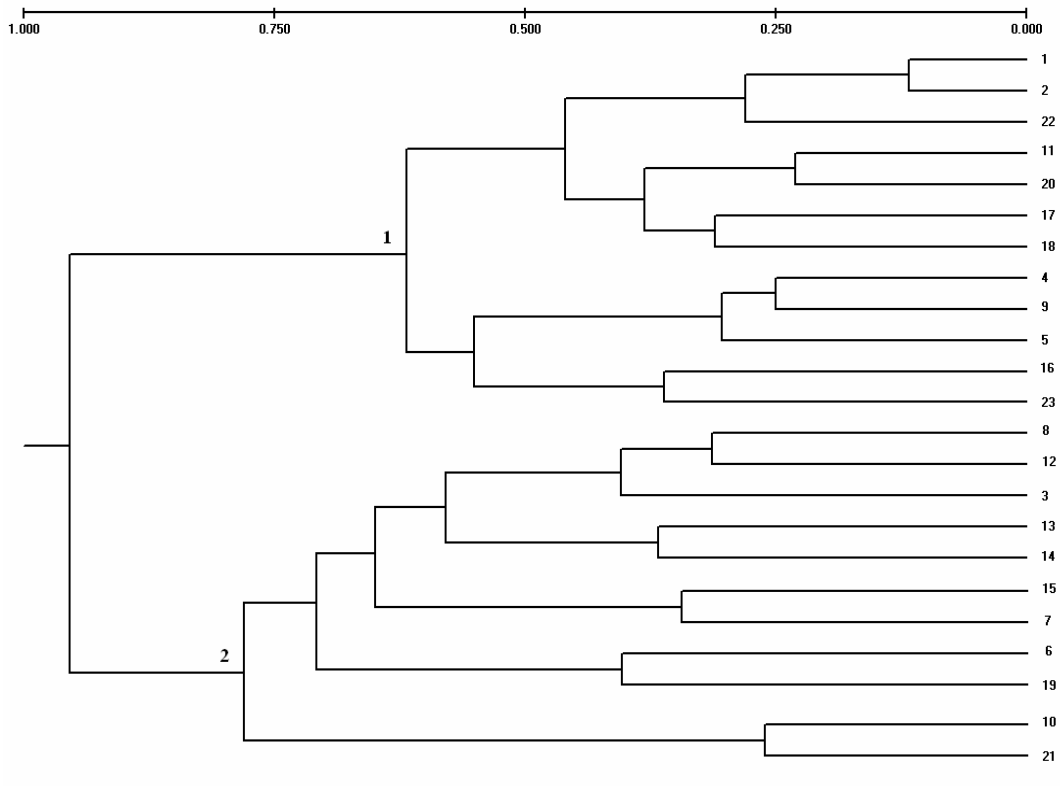


Fig. 24. Dendrogram of 23 olive genotypes from the Abruzzo region, generated by UPGMA cluster analysis based on Nei's (1972) genetic original distances. Numbers in the dendrogram locate the different clusters. The numbers correspond to the ordering number of accessions in (Table 4 Annex- I).

3.1.3 Genetic relationships between olive varieties from Molise region.

SSR marker technology was revealed to be a powerful tool for studying variation among accessions of *Olea europaea* L. In fact, the fingerprinting patterns obtained allowed unequivocal identification of each accession (Table 4 Annex-I). The highest values of genetic identity (data not shown) were observed between the accessions Paesana bianca - Rosciola di Rotello (0.96), Grossa di Venafro - Paesana nera (0.96), Cazzarella – Sperone di Gallo (0.88), Grossa di Venafro - Olivetta nera (0.86), Oliva nera di Colletorto - Noccioluta (0.85), and Olivetta nera - Paesana nera (0.82). These extremely high values of genetic identity can be explain how a few accessions can be considered as synonyms, e.g. the accessions Oliva nera di Colletorto – Noccioluta (Cicoria, et al., 2000). The lowest values of genetic identity were obtained between Rossuola and Sperone di Gallo (0.16).

Nei's genetic distances were utilized to obtain a phenogram based on a UPGMA algorithm of clustering (Fig. 25). Two distinct clusters of olive varieties were clearly recognizable. The first included 62% of varieties and the second one contained 38%. Evidence of relationships for most of the cultivars according to their geographic origin was found.

The first cluster (*1*) could be subdivided in three sub-clusters: (*1a*) contained 3 accessions: Olivastro Dritto, Paesana bianca and Rosciola di Rotello. The second (*1b*) containing 6 accessions: Aurina, Bottoni di Gallo, Olivastro di Montenero, Grossa di Venefro, Paesana nera, and Olivetta nera. Finally, the subcluster (*1c)* containing two accessions: Gnagnaro and Olivastro d'Aprile. All accessions belonging to the first cluster were present in the Isernia cultivation area, except for Rosciola di Rotello and Olivastro di Montenero. The second cluster (*2*) was subdivided in 3 sub-clusters: (*2a*) containing the 2 accessions Cellina di Rotello and Rossuola; (*2b*) comprised 4 accessions: Cazzarella, Sperone di Gallo, Gentile di Larino, Rumignana and (*2c*) comprised 4 accessions: Oliva nera di Colletorto, Noccioluta, Romanella molisana and Salegna di Larino. All accessions belonging to cluster *2* were present in the Campobasso cultivation area.

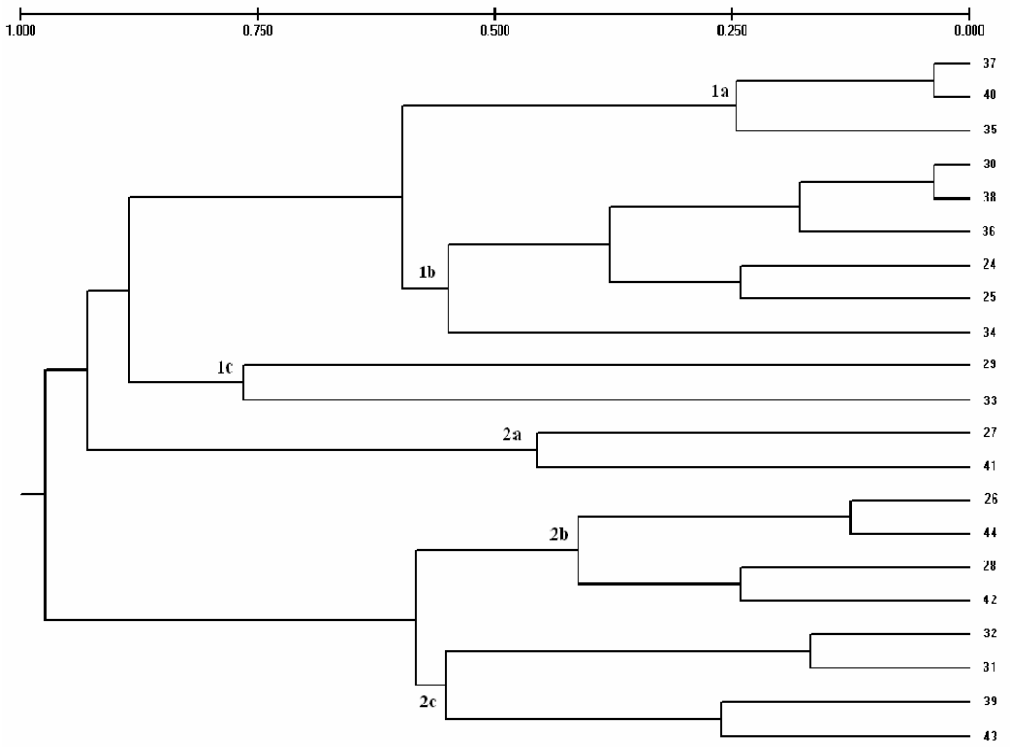


Fig. 25. Dendrogram of 21 olive genotypes from the Molise region, generated by UPGMA cluster analysis based on Nei's (1972) genetic original distances. Numbers in the dendrogram locate the different clusters and subclusters. The numbers correspond to the ordering number of accessions in (Table 4. Annex-I)

3.1.4 Genetic relationships between olive varieties from the Abruzzo and Molise regions.

Forty-four genotype profiles were obtained using 9 loci SSR and were able to identify all the varieties analyzed. All olive accessions were readily separated from each other (Table 4. Annex - I). The greatest values of genetic identity were observed between the accessions of the Molise Olivetta nera and Abruzzi accession Toccolana (0.96), Caprina Vastese (Abruzzo) – Grossa di Venafro (Molise) (0.93), Caprina Vastese (Abruzzo) - Paesana nera (Molise) (0.92), and Ghiandaro (Abruzzo) - Olivastro dritto (Molise) (0.85). The lowest values of genetic identity were obtained between Gentile di Chieti (Abruzzo) – Rosciola di Rotello (Molise) (0.15) and Gentile di Chieti (Abruzzo) – Paesana bianca (Molise) (0.15).

Genetic distances (Nei, 1972) were utilized to obtain a phenogram based on the UPGMA algorithm of clustering (Fig. 26). Two distinct clusters of olive varieties were

clearly recognizable. The first cluster included 55% of olive varieties examined (12 Abruzzi accessions and 12 accessions of the Molise region), while the second cluster contained 45% of the varieties (11 Abruzzo accessions and 9 accessions of the Molise region). This first cluster included most of the accessions situated along the hillside facing east towards the coast of the Adriatic Sea. The second cluster included most of the olive accessions situated towards the interior of a country, probably because it is more resistant to cold.

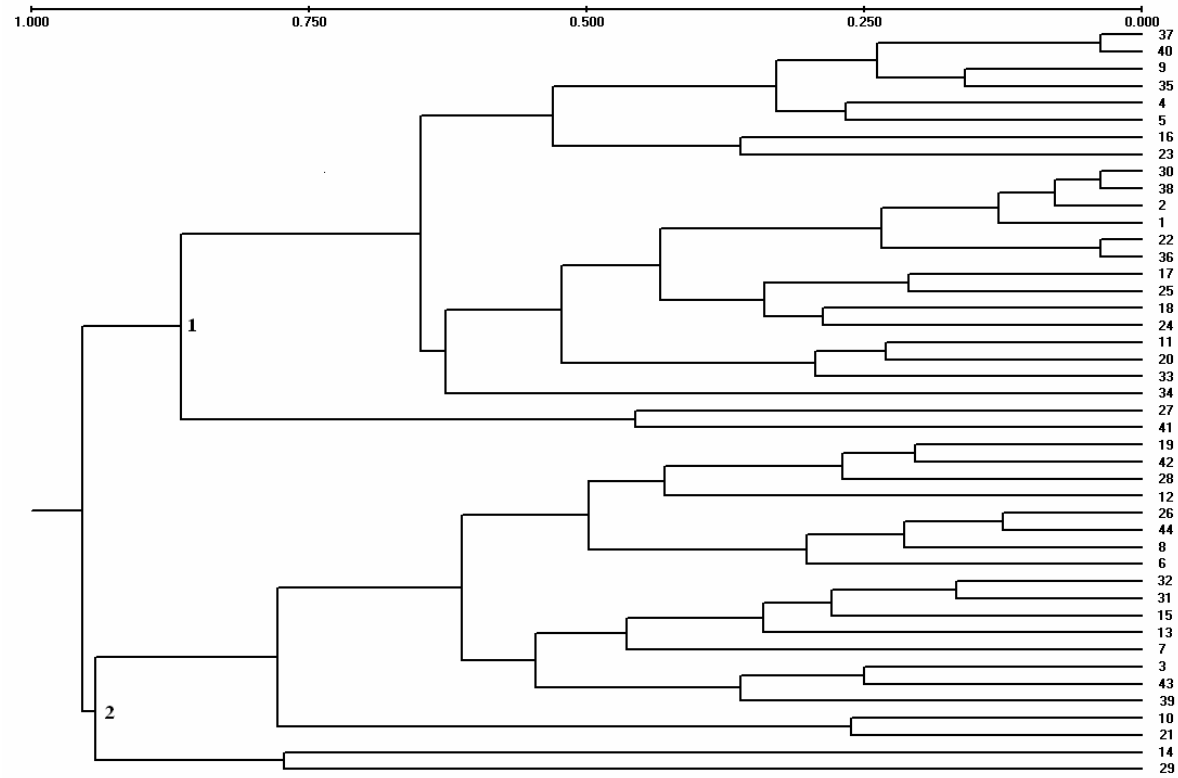


Fig. 26. Dendrogram of 44 olive genotypes, generated by UPGMA cluster analysis based on Nei's (1972) genetic original distances. Numbers in the dendrogram locate the different clusters. The numbers correspond to the ordering number of accessions in (Table 4. Annex-I)

3.2 SICILY

In this work 55 olive cultivars of CRA-ISOl germplasm, were studied, corresponding to the major part of the autochthon germplasm of Sicily region.

The eight SSRs primers produced polymorphic amplification products in the cultivars studied. The total number of alleles amplified was 36 with an average of 4,5 alleles per

locus, ranging from a minimum of 2 for UDO01 to a maximum of 8 for UDO39 (Table 5). This is comparable to the number of alleles among olive cultivars reported by Rallo et al., (2000), but somewhat lower than that published by La Mantia et al., (2005), probably because it used the major number of primer pairs and also because it included cultivars of other areas of origin.

Table 5: SSR loci in 55 olive accessions of Sicilian germplasm. For each locus, the number of alleles, size range of alleles, observed (Ho) and expected heterozygosity (He) and estimated frequency of null alleles (r) are reported.

Locus	N alleles	Size range (bp)	Ho	He	r
GAPU059	3	208-222	0.636	0.560	-0.05
GAPU071A	4	210-259	0.436	0.480	0.03
GAPU071B	5	125-145	0.855	0.660	-0.12
GAPU103A	6	136-186	0.891	0.783	-0.06
UDO001	2	140-143	0.000	0.323	**0.24**
UDO012	3	166-193	0.709	0.660	-0.03
UDO028	5	143-210	0.382	0.609	0.14
UDO039	8	108-232	0.327	0.787	**0.26**

The observed heterozygosity was higher the expected values at loci GAPU59, GAPU71B, GAPU103A and UDO12, and lower at the UDO01 and UDO39 locus. A possible explanation of such a deficit is the occurrence of null alleles at these loci since the corresponding probability is highly significant (Table 5). The frequency for each alleles ranged from 0,01 to 0,78 (data not shown). The lowest allelic frequency (0.01) was observed in alleles 259 bp of GAPU71A in Bottone di gallo, and 250 bp of GAPU71A in Erbano and 128 bp at GAPU71B in Bottone di gallo. This three alleles were present only once in all the cultivars analyzed. Whereas allele 144 bp of the less polymorphic locus UDO01 showed the highest frequency (0.78).

Fifty-one different genotype profiles were obtained with this combination of eight loci, being able to identify 93% of the varieties analyzed showing unique profiles. The remaining 7% is composed by four different pairs in which the varieties are genetically indistinguishable from one another. The first two possible case of synonymy comprises Nerba catanese/Olivo di Castiglione and Giarraffa/Pizzo di corvo 1 and were in

agreement to La Mantia et al., (2005). The third pairs of varieties comprises the Iacona and the Ogliarola messinese. Finally, the last pairs of varieties comprises the Aitana and the Cacaridduni. This last possible case of synonymy was in contrast with the survey carried out by La Mantia et al., (2005) probably because the source of plant material present in the CRA-ISOl collection was different from the Castelvetrano collection. In fact, olive trees under the Pizzo di corvo (1 and 2) denomination were classified into two molecular profiles which were differentiated by two SSR alleles (similarity value = 0,917). Apart from that, the greatest similarity was observed between the cultivars Monaca - Vaddara (s.v. = 0.966), and between the cultivars Nocellara Nissena – Tonda Iblea (s.v. = 0.963). The lowest similarity values were obtained between Nebba – Castricianella rapparina (s.v. = 0.167).

In the dendrogram (Fig. 27), most of the cultivars could be classified into three groups. Of the 16 varieties included in group I: seven varieties were from Messina province (Castricianella rapparina, Mantonica, Minuta, Monaca, Santagatese, unknown 1 and Vaddara), five were from Agrigento province (Abunara, Bottone di gallo, Erbano, Murtiddara and Pirunara), two from Catania province (Calatina and Cavalieri) and two from Palermo province (Sammartinara and Crastu). Group II was formed by 37 Sicilian varieties that neither their likely area of origin, nor their phenotypic traits can explain. Group III was formed by Pizzutella and Verdello (from Messina province) what stands alone, with a low level of similarity to all remaining cultivars.

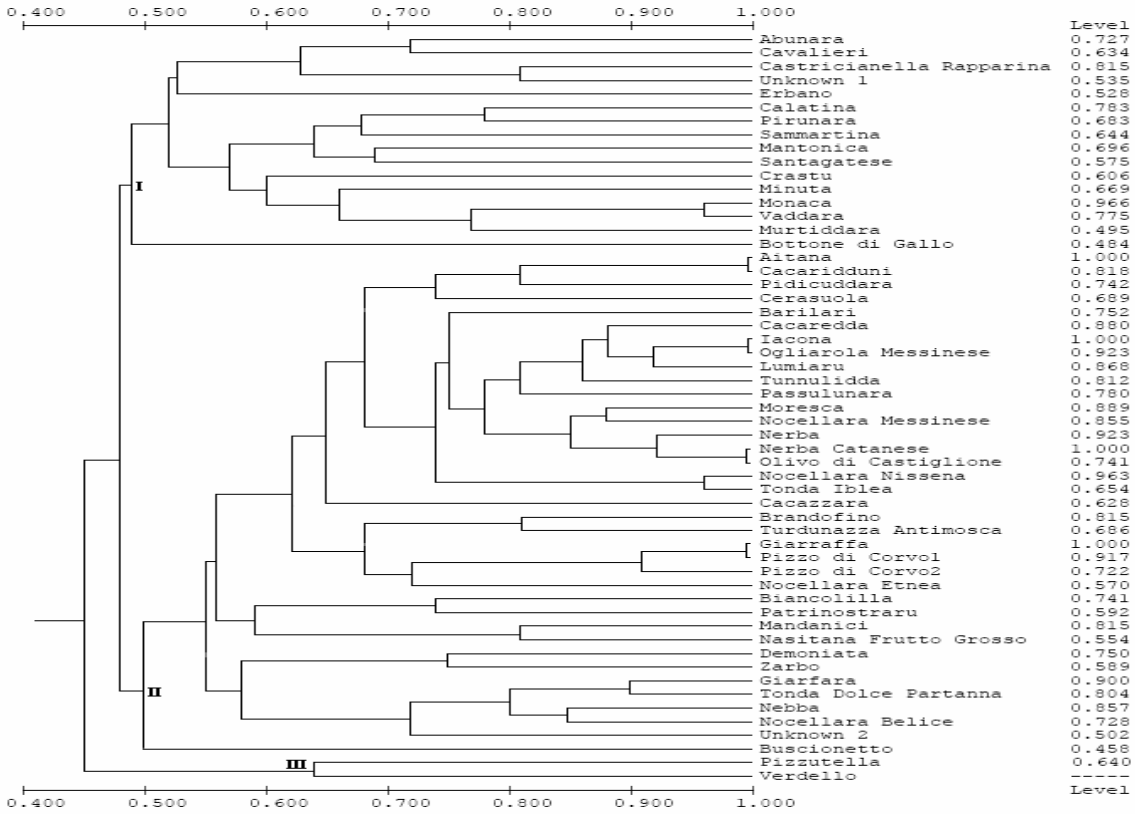


Fig. 27. UPGMA dendrogram showing genetic similarity between 55 Sicilian olive cultivars.

The unknown accessions (1 and 2) were placed in the dendrogram in the group I and II respectively, but none was identical to any other accessions analyzed. It was not possible to identify these accessions with any known cultivar in Sicily.

3.3 APULIA.

After an initial screening, five primer pairs were chosen for further analysis of Apulian cultivers based on their polymorphic index and reproducibility of amplification.

The DNA sequence of all amplicons was determined and the exact number of repeat units was established for each sample. In fact, we observed that some alleles showed the same length but different sequences due to a sequence variant that interrupted the tandem repeat unit motif. A total of 28 alleles was found in the five loci (Table 7) with an average of 5.6 alleles per locus, ranging from 2 at UDO01 to 9 at UDO39. This is comparable to the number of alleles among olive cultivars reported by Rallo et al.,

(2000), but somewhat lower than that published by Khadari et al., (2003), probably because it included a large number of foreign cultivars. Interestingly, only 24 of the 28 alleles were distinguishable by electrophoretic analysis, with an average of 4.8 alleles per locus (Table 7), since these had the same length but minor nucleotide changes.

Table 7. SSR loci in 39 olive accessions of the Apulian germplasm. For each locus, the number of detected (Nde), effective (Nef) and exclusive (Nex) alleles was obtained with and without sequencing. Nde, Nef and Nex were calculated using POPGENE 32 software.

Locus	With sequencing			Without sequencing		
	Nde	Nef	Nex	Nde	Nef	Nex
GAPU45	5.00	4.16	0.00	4.00	2.44	0.00
GAPU59	6.00	3.68	1.00	4.00	2.13	0.00
GAPU71A	6.00	2.66	1.00	5.00	2.43	1.00
UDO01	2.00	1.62	0.00	2.00	1.62	0.00
UDO39	9.00	5.11	1.00	9.00	5.11	1.00
Total	28.00	17.23	3.00	24.00	13.73	2.00
Mean	5.60	3.45	-	4.80	2.75	-
SD	2.51	1.35	-	2.59	1.36	-

The size range and the sequenced repeat motifs for all loci are reported in Table 8. The shortest allele among these five loci was 108 bp in length in UDO39, while the longest allele was 228 bp in GAPU71A. As mentioned above, four alleles were not distinguishable by electrophoresis and showed different DNA sequences (Table 8). These included allele 183A/183B bp at GAPU45, alleles 208A/208B bp and 222E/222F bp at GAPU59 and alleles 210A/210B bp at GAPU71A (Table 8).

Table 8. Repeat motif and sequence size of the SSR amplification products employed in the olive accessions characterization. For every locus only the variable sequence region is reported. The underlined nucleotides represent sequence variants detected by DNA sequence analysis.

Locus	Size of allel (bp)	Sequence motif
GAPU45	183 A	$C(AG)_7$ CTTC<u>AA</u>G
	183 B	$C(AG)_8$ CTTCG
	185	$C(AG)_8$CTTCAAG
	182	$GTG(AG)_9$
	196	$GTG(AG)_{12}TG(AG)_3$
GAPU59	208 A	<u>AA</u>$(CT)_{10}$
	208 B	$(CT)_{11}$
	212	$(CT)_{13}$
	218	$AA(CT)_{15}$
	222 E	$(CT)_{16}$<u>T</u>TCT
	222 F	$(CT)_{18}$
GAPU71A	210 A	$(AG)_6A$<u>A</u>AG
	210 B	$(AG)_8$
	212	$(AG)_9$
	214	$(AG)_{10}$
	224	$(AG)_{15}$
	228	$(AG)_{17}$
UDO01	140	$(CA)_8AA$
	144	$(CA)_{11}$
UDO39	108	$AA(ATAC)_2(AT)_2$
	142	$(AT)_7(GT)_5(GC)_6A(CGTG)_2TTG$
	146	$(AT)_7(GT)_8(GC)_4A(GTGC)_2ATGTTG$
	164	$(AT)_5(GT)_{12}(GC)_3(GT)_2GCAT(GC)_2(GT)_3GCGTGCATGTTG$
	170	$(AT)_5(GT)_{11}(GC)_3(GT)_2GC(GT)_3GCGT(GC)_2(GT)_3GCGTGCATGTTG$
	173	$(AT)_8(GT)_{20}(GC)_2(GTGC)_4GTG$
	175	$(AT)_8(GT)_{21}(GC)_2(GTGC)_4GTG$
	184	$(AT)_9(GT)_{16}(GC)_2(GTGC)2(GC)_5(GTGC)_3GTGTTG$
	188	$(AT)_5(GT)_{13}(GC)_3(GT)_2GC(GT)_3GCGT(GC)_2(GT)_3GCAT(GC)_2(GT)_3GCGTGCAT$ GTTG

The lowest allelic frequency (0.013) was observed in alleles 222F bp of GAPU59 and 228 bp GAPU71A in 'Cellina di Nardó' and 142 bp at UDO39 in 'Butirra di Melpignano', whereas allele 144 bp of the less polymorphic locus UDO01 showed the highest frequency (0.744). Three alleles were present only once in all the cultivars analyzed (Table 9. Annex-I).

The highest genotypic frequency (0.74) was observed at UDO01 for the 144-144 genotype, while the lowest frequencies (0.03) were detected for the 208A-208A, 208A-222F, and 218-218 genotypes at GAPU59, obtained for 'Frangivento', 'Cellina di Nardó' and 'Toscanina' respectively, for the 210A-224, 210B-210B, 212-214 and 214-228 genotypes at the locus GAPU71A observed in 'Mora', 'Simona', 'Peranzana' and 'Cellina di Nardó' respectively, and for the 108-142, 108-164, 108-173, 108-175, 108-188, 146-146, 164-164, 170-173, 173-173, 173-184, 184-184 and 188-188 genotypes at the UDO39 locus in 'Butirra di Melpignano', 'Simona', 'Peranzana', 'Dolce di Cassano', 'Cellina di Nardó', 'Oliva Dolce di Barbarano', 'Toscanina', 'Nociara', 'San Benedetto', 'Frangivento', 'Carmelitana' and 'Rotondella', respectively (Table 9. Annex-I).

The observed heterozygosity for the 39 cultivars ranged from 0.00 at UDO01 to 0.77 at the GAPU45 locus, which was lower than the expected value. (Table 10) shows the observed and expected heterozygosities obtained by either electrophoretic or DNA sequence analysis.

Table 10. Analysis of molecular data of 39 olive accessions of the Apulian germplasm by mean of five SSR loci. Observed (Ho) and expected heterozygosity (He) values were obtained from the combination of with and without sequencing SSR loci. Ho and He were calculated using POPGENE 32 software.

Locus	With sequencing		Without sequencing	
	Ho	He	Ho	He
GAPU45	0.77	0.76	0.61	0.60
GAPU59	0.67	0.73	0.38	0.54
GAPU71A	0.59	0.62	0.59	0.60
UDO01	0.00	0.38	0.00	0.39
UDO39	0.49	0.80	0.49	0.81
Mean	0.50	0.66	0.41	0.59
SD	0.30	0.17	0.25	0.15

3.3.1 Genetic relationships between olive cultivars.

28 unique genotype profiles were obtained using only five loci (Table 9 Anne-I), which identified 54% of the cultivars analyzed (genotypes 1÷21). The remaining 46% probably comprises seven different accession groups in which the cultivars are genetically indistinguishable from one another, potentially representing cases of synonymy (Table 9 Annex-I). The first group comprises ‘Bella di Spagna’ and ‘Bella di Cerignola’ (genotype 22), the second ‘Cerasella’, ‘Nolca’ and ‘Olivo Mele’ (genotype 23), the third ‘Cima di Mola’, ‘Ogliarola di Lecce’ and ‘Pizzuta’ (genotype 24), the fourth ‘Coratina’ and ‘Racemo’ (genotype 25), the fifth ‘Dolce d’Andria’ and ‘Dolce Mele’ (genotype 26), the sixth ‘Lezze’, ‘Oliastro’ and ‘Oliva Rossa’ (genotype 27) and the last ‘Marinese’, ‘Ogliarola Barese’ and ‘Ogliarola Garganica’ (genotype 28). These possible cases of synonymy were in agreement with previous data using RAPDs (39 primers) and morphological (36 parameters) analysis (Lombardo et al., 2004) obtained using the same set of olive trees.

Genotype 3 showed the highest percentage of shared alleles (90%) with genotype 8, as was the case for genotype 12 with 28, 17 with 22, and 19 with 26, while genotype 1 showed the lowest percentage (0%) with genotype 27. On the other hand, genotype 4 showed the highest percentage of shared alleles with genotypes 25 and 27, while genotype 6 showed the highest percentage of shared alleles with genotype 25. Therefore, the highest values of genetic identity (Nei, 1987) were observed between genotypes 17 and 22 (0.94), genotypes 3 and 8 (0.93), genotypes 12 and 28 (0.93), and genotypes 19 and 26 (0.93).

Genetic distances (Nei, 1972) were utilized to obtain a phenogram based on the UPGMA algorithm of clustering (Fig. 28). Two distinct clusters of olive cultivars were clearly recognizable. The first cluster included 74% of cultivars examined, while the second cluster contained 26% of cultivars.

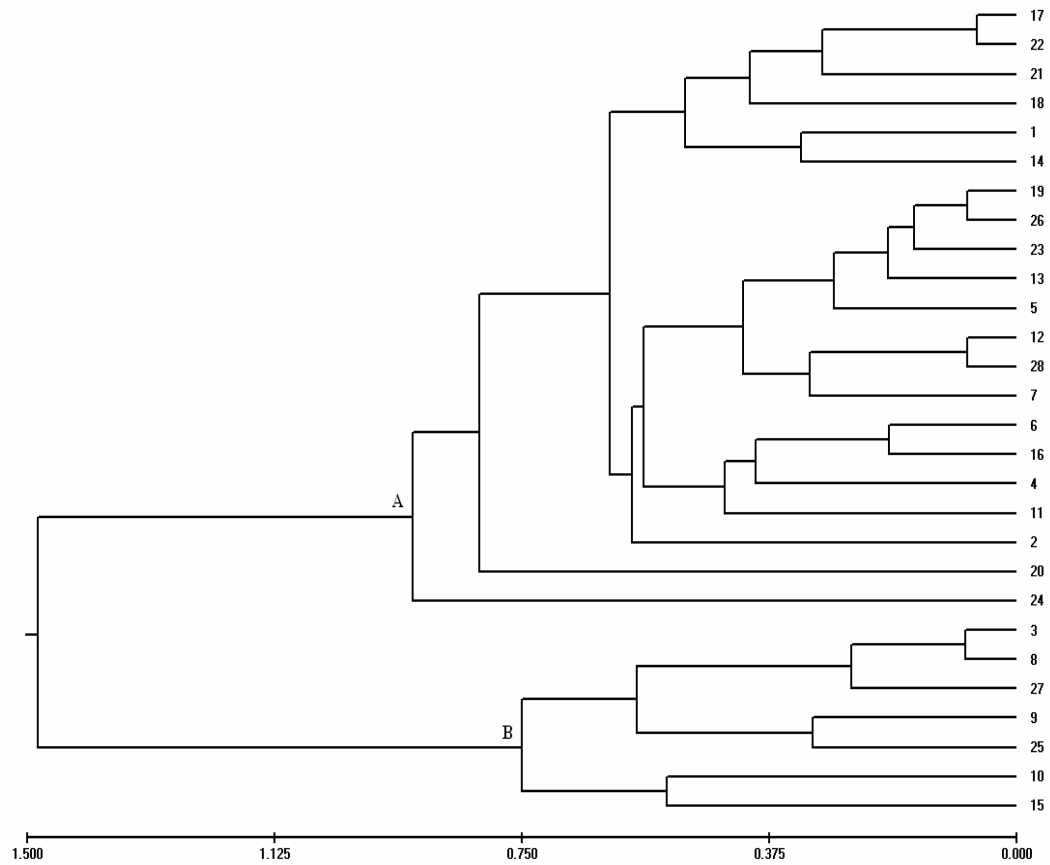


Fig. 28. Dendrogram of 28 olive genotypes generated by UPGMA cluster analysis based on Nei's (1972) genetic distances obtained using TFPGA software version 1.3. Genotype number corresponds to those reported in (Table 9. Annex-I). Capital letters in the dendrogram correspond to different clusters.

3.4 DISCUSSION.

Forty-four genotype profiles were obtained using 9 SSR and were able to identify all the analyzed varieties of the Abruzzo and Molise regions. Despite the small area of the two regions, the genetic diversity within olive germplasms from Abruzzo and Molise is quite high and can probably be ascribed to several introductions and evolutions of genetic material from other regions and countries.

By the identification of 51 different genotypes and the characterization of 47 cultivars with unique genotype. The high values of cultivars with unique genotype obtained from olive germplasm in Sicily is probably ascribable to several introductions of genetic material from other regions and countries. In fact, from a historical point of view Sicily has been dominated by many different cultures (Phoenicians, Greeks, Romans, Arabs, Normans) that undoubtedly contributed to the present-day germplasm in Sicily (Bartolini and Petruccelli, 2002).

One cases of homonymy were identified: olive trees under the Pizzo di corvo denomination were classified into two molecular profiles. Besides the following four presumable synonyms were detected: Aitana/Cacaridduni, Iacona/Ogliarola messinese, Nerba catanese/Olivo di Castiglione and Giarraffa/Pizzo di corvo. Therefore, homonymy and synonyms characterisation is very important to avoid genotype redundancy in order to maximise genetic diversity in Sicilian olive germplasm collection.

The identification of 28 different genotypes and the detection of 21 cultivars from Apulia, with unique genotypes Among the cultivar groups showing shared genotypes, we can conclude that as already inferred they are in fact synonyms. By comparison of the molecular profiles obtained from the five SSR loci, we observed that the same genotype is shared among two or more cultivars in genotypes ranging from 22 and 28 (Table 9 Annex-I). Genotype 22 consisted of two accessions, ‘Bella di Spagna’ and ‘Bella di Cerignola’, which was in accordance with morphological and molecular analysis data (Lombardo et al., 2004), but in contrast with the survey carried out by

Bartolini et al., (1998). Genotype 23 included three accessions, 'Cerasella', 'Mele' and 'Nolca', which is in agreement with morphological and molecular data (Lombardo et al., 2004), but in contrast to that reported by Bartolini et al., (1998). Genotype 24 comprises three accessions, namely 'Ogliarola di Lecce', 'Pizzuta' and 'Cima di Mola'. According to Bartolini et al., (1998) the 'Ogliarola di Lecce' and 'Pizzuta' accessions are synonyms. The results with the 'Cima di Mola' are in contrast with the data of Bartolini et al., (1998), but in agreement with morphological observations and the results of RAPDs (Lombardo et al., 2004). The 'Coratina' and 'Racemo' accessions, belonging to genotype 25, are synonyms according to literature data (Bartolini et al., 1998; Ferrara et al., 1995; Lombardo et al., 2004). Genotype 26 includes two accessions: 'Dolce di Andria' and 'Dolce Mele', which were identical according to morphological and RAPDs data (Lombardo et al., 2004), but in contrast to the survey carried out by Bartolini et al., (1998). Genotype 27 comprises three accessions, 'Lezze', 'Oliastro' and 'Oliva Rossa', which is in agreement with morphological observations and RAPDs analysis (Lombardo et al., 2004). Genotype 28 included the 'Marinese', 'Ogliarola Barese' and 'Ogliarola Garganica' accessions, which are synonyms according to available data (Bartolini et al., 1998; Lombardo et al., 2004). In fact, these names are local synonyms for the 'Cima di Bitonto' (Bartolini et al., 1998). It is very common to attribute different names to the same cultivar. Therefore, characterization of synonyms is very important in order to avoid genotype redundancy and to maximize genetic diversity in olive germplasm collections.

The results obtained during this study from Abruzzo, Molise, Sicily and Apulia regions confirms the utility of molecular analysis of olive germplasms and shows the high variability of these markers. This variability can be used to point out genetic distances among the different examined entities and to affirm with sufficient certainty if these are homogeneous genetic entities and/or belong to the same variety.

4. PROTEOMICS OF (*OLEA EUROPAEA* L.) TISSUES (POLLEN, FRUIT AND OIL).

The availability of genomic sequences made much easier the proteomic analyses of plants. In the field of plant proteomics, there have been some changes in the analytical aims. While early projects aimed at comparing different genotypes and plant lines to measure phylogenetic distances (Zivy et al., 1984; Bahrman et al., 1985; Damerval et al., 1986), later projects made use of N-terminal Edman protein sequencing (Tsugita et al., 1994; Kamo et al., 1995) or used amino acid analysis for protein determination (Touzet et al., 1996). Nowadays, with mass spectrometric methods the identification of proteins goes hand in hand with their separation (Porubleva et al., 2001).

Proteomic analyses of olive tissues (pollen, fruit and oil) were applied to determine the protein profile of the extracted samples for the purpose of their identification and characterization.

4.1 Allergenomics.

The term *allergenomics* has been proposed, recently, for a rapid and comprehensive analysis of protein allergen candidates by applying a proteomic strategy. *Allergenomics* requires not only the detection and the assignment of the putative allergen in a short time (Beyer et al., 2002; Yu et al., 2003), but also the correlation of the quantitative and qualitative change of the antigens with the environmental conditions. This procedure is based on the extraction from olive pollen of the protein pool containing the allergen candidates (Yagami et al., 2004). Pollens are important triggers for allergic asthma and seasonal rhinitis. Some trees, such as birch and olive, are able to induce dramatic airborne allergies because of the seasonal appearance of high density pollen levels. Allergy to olive trees pollen (*olea europaea*) affects more than 30 % of the population in Mediterranean areas during the pollination season. The structure which surround the pollen (intine and exine) contains numerous proteins which vary with the different plant species. Cultivars, and probably local varieties or sub-varieties of olive trees, present special features which depend upon their adaptation to the environment or ecotype. The ecoenvironment and crop management are determining factors to induce qualitative changes, allergologically fundamental in a particular variety or cultivar (Conde et al., 2002; Benincasa et al., 2003).

4.1.1 Extraction and characterization of olive (*Olea europaea* L.) pollen hydrophilic proteins from Italian cultivars by MALDI TOF MASS Spectrometry[9].

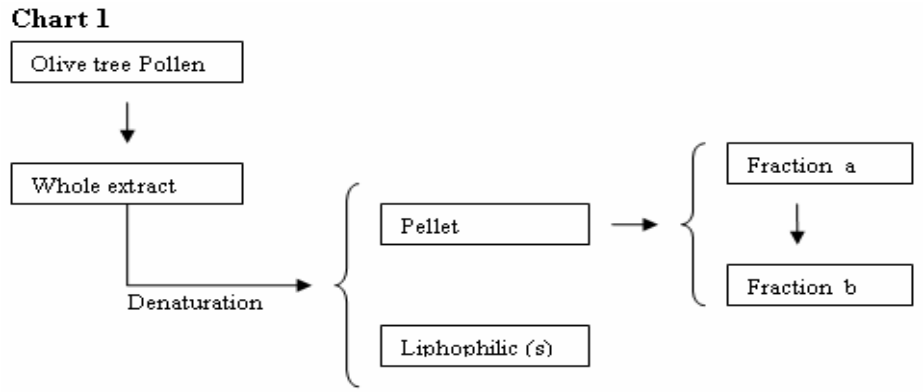
An important feature of allergenic molecules is the solubility in aqueous medium. The allergenicity of proteins is more linked to their concentration and rapid elution from airborne particles, than to other intrinsic properties (Vrtala et al., 1995). The extraction solution could help to determine the antigenic profile of olive pollen extracts (Carnés et al., 2002). Chemical fractionation of whole protein extracts, and the antigen contents of each fraction can be varied according to the selected experimental conditions. Specificity and better resolution of MALDI mass spectrometry with respect to a conventional 2-D (two-dimensional gel electrophoresis) (Conde Hernandez et al., 2002) chromatographic approach, represents the methodology of choice to get reliable results in the profiling of olive pollen.

The limitations in MALDI MS analysis lies within the sample preparation, like the use of solution containing sodium chloride, phosphate buffer, and borate buffer either to prepare the whole pollen extract or for its chemical fractionation. For that reason we choose the ammonium bicarbonate solutions to prepare the whole antigenic extract from pollen grain of cultivars 1-8. Ammonium bicarbonate suppose to help the formation of carboxylate/ammonium ion pairs, which as a result support the solubility of the proteins to be extracted in moderately polar solvents, for instance acetonitrile/water mixtures. Furthermore, ammonium counter ions have been often used to improve the desorption of high molecular weight protein (Rajnarayanan and Wang, 2004) and does not interfere with the mass spectrometric analysis. The olive (*Olea europaea*) pollens of the Mediterranean cultivars, Ottobratica (1), Carolea (2), Dolce di Rossano (3), Cassanese (4), Coratina (5), Nocellara del Belice (6), Villacidro (7), and Sinopolese (8), were selected as case studies to determine a protein profile of the whole extract and to identify and characterize specific proteins without any previous chromatographic or two-dimensional gel separations. we developed a simple procedure of chemical fractionation of the whole antigen extract, whereby less complex, or pure, fractions of antigen candidate were obtained. All fractions were directly analyzed by MALDI TOF

[9] Published (No. 2. Appendix)

in the linear mode. The devised a sample preparation protocol distribute the amount of information that are stored in the proteome of each olive pollen in a set of three MALDI spectra that could be independently evaluated and matched in order to retrieve data for their comparison. The data set displayed by the three spectra provides the entire profiling of a given entity. One lypophilic (s) and two hydrophilic fractions (a-b), respectively, can be obtained from each sample by following the procedure reported in the Experimental Section (Chart 1).

Chart 1



(Bold lower-case letters and bold numbers are used to identify the fraction and the cultivar, respectively; for example, sample 4a represents the fraction a of Cassanese).

Accordingly the antigenic profile of Cassanese 4 displayed four allergens: Ole e 7, Ole e 6, Ole e 2, and Ole e 1 in the 5-20-kDa mass range (Fig. 29). The four ion peaks at *m*/*z* 9791-10 041 mass range (Fig. 29) can be ascribed to Ole e 7. In fact, the apparent SDS-PAGE molecular mass of this allergen is ~9-11 kDa (Table 11, column 3, row 8). The predicted molecular mass of the Ole e 7 fragment from the peptide sequence is 2199 Da (Tejera et al., 1999) (Table 11, column 5, row 8). The only experimental values available are 9905-10 302 Da obtained by low-resolution MALDI mass spectrometry (Tejera et al., 1999) (Table 11, column 5, row 8). Therefore, it can be confidently suggested that peaks in the range 9791-10 041 Da (Fig. 29) should correspond to the expected Ole e 7.

The ion peak at *m*/*z* 5821 (Fig. 29) was attributed to the well known olive pollen allergen Ole e 6. This allergen has been isolated, purified and biochemically characterized, (Batanero et al., 1997) and its specific cDNA was cloned and sequenced (Barral et al., 2004; Trevino et al., 2004) (Table 11, row 7). Considering that the value of 5833 Da (Table 11, column 5, row 7) corresponds to the predicted molecular weight

from cDNA and that there are no other known allergens in this range, the observed ion peak at *m*/*z* 5821 can be ascribed to Ole e 6.

The doubly charged ion at *m*/*z* 7396 probably corresponds to Ole e 2, since it is known that this allergen consists of 134 amino acids, (Martinez et al., 2002; Asturias et al., 1997) and the predicted average molecular mass is 14.4 kDa (Table 11, row 3).

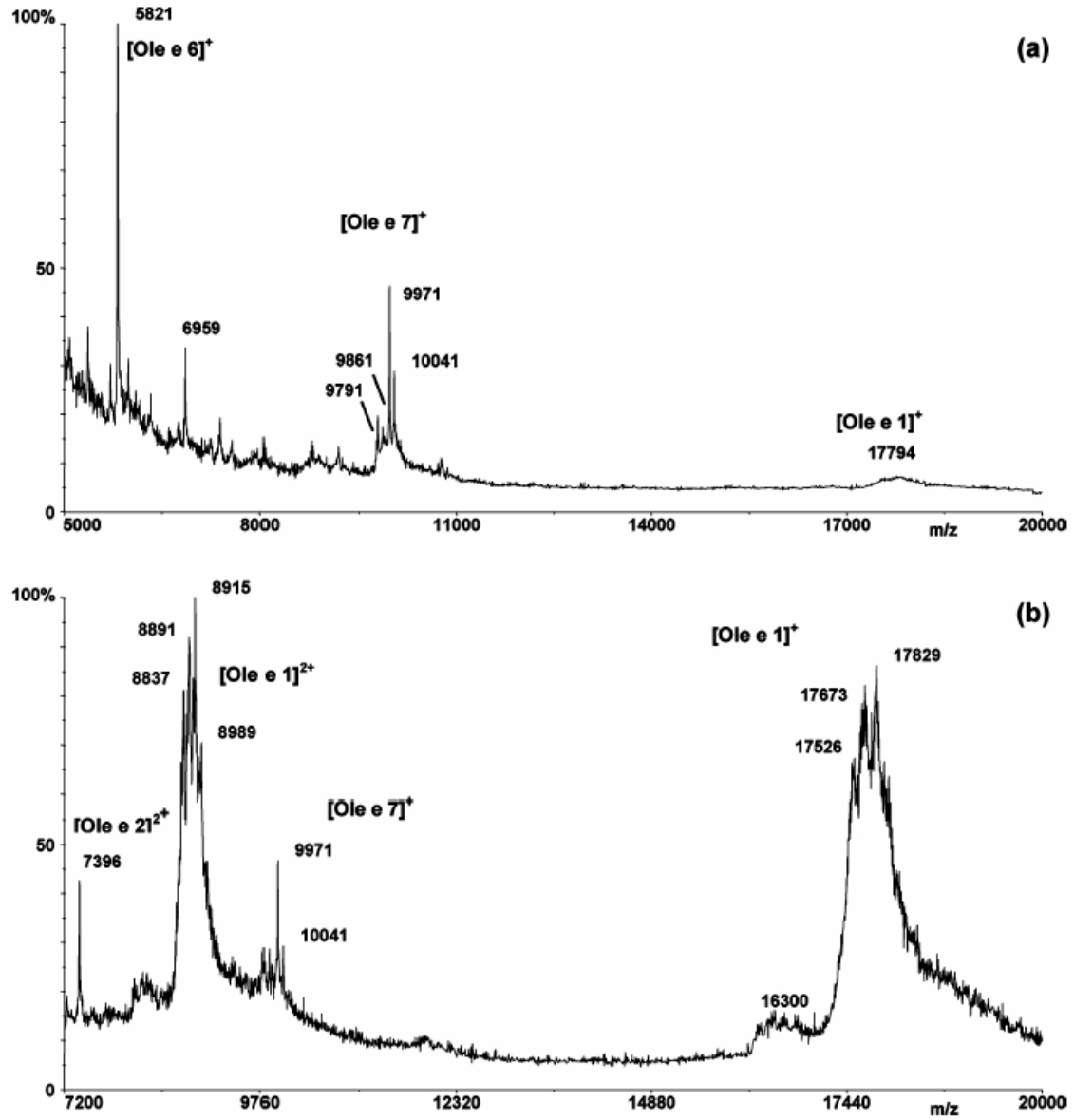


Fig. 29. MALDI spectra of fractions (a) **4a** and (b) **4b**. Both fractions represent a mixture of allergens. Part (a) shows the typical pattern of Ole e 7, Ole e 6, and some polypeptides between 5 and 7 kDa; whereas part (b) shows the ion species [Ole e 2]2+, [Ole e 1]2+, [Ole e 7]+, and [Ole e 1]+.

Table 11. Olive Pollen Allergens with Clinical Relevance Developed and Maintained by Allergen Nomenclature Subcommittee of the IUIS (www.allergen.org), Including Allergens Whose IgE Reactivity Has a Prevalence of >5%

allergen name	obsolete name	MW, kDa, SDS-PAGE[a]	variants	MW (Da)[b]	p.I.	sequence[c]	accession number[d]
Ole e 1	major pollen allergen	18-21	3-3, I → V (in Ole e 1.0102, Ole e 1.0103, & Ole e 1.0105) 23-23, R → P 24-25, AG → SR (in Ole e 1.0105) 39-39, L → V (in Ole e 1.0102, Ole e 1.0103, & Ole e 1.0105) 44-44, K →R (in Ole e 1.0105) 45-45, D f E (in Ole e 1.0105) 46-46, K → I 47-47, E → K 48-48, N → K 50-50, D → S 51-51, V → I (in Ole e 1.0105) 56-56, V → I (in Ole e 1.0102, 1.0103, & 1.0105) 69-69, V → I 87-87, R → S 91-91, N →D (in Ole e 1.0105) 95-95, T → I (in Ole e 1.0105) 106-106, K → I (in Ole e 1.0102, 1.0103, & 1.0105) 108-108, N → S (in Ole e 1.0103) 111-111, N → D 121-121, G → R (in Ole e 1.0102) 123-123, F → Y (in Ole e 1.0102 & 1.0103)	16 330	6.18	C	P19963
Ole e 2	profilin-1	15-18		14 489	5.06	C	O24169
	profilin-2	15-18		14 427	5.21		O24170
	profilin-3	15-18		14 399	5.21		O24171
Ole e 3	polcalcin Ole e 3	9.02		9356	4.49		O81092
	Ole e 3 allergen (fragment)			5798			Q5DTB7
Ole e 4	major pollen allergen Ole e 4 (fragments)	32		2711	3,77	P	P80741
Ole e 5	superoxide dismutase [Cu-Zn] (fragment)	16		2973	4,65	P	P80740
Ole e 6	pollen allergen Ole e 6	10		5833	4,96	C	O24172
	pollen allergen Ole e 6 *fragment)			4986	5,43		Q5DTB8
	Ole e 6 allergen (fragment)			4400	7,8	N	P81430
Ole e 7	pollen allergen Ole e 7 (fragment)	9 - 11	5-5, S f G (in type B) 10-10, L → K (in type B) 18-18, I → K (in type B)	9905-10 302[e] 2199	3.56	P	P81430
Ole e 8	Ca2+-binding protein	21	43-44, GV → CA 58-58, G → A	18 907	4.51	C	Q9M7R0
Ole e 9	β-1,3-glucanase	46	60-60, I → M	48 838	5.21	C	Q94G86

[a] Apparent molecular mass in SDS-PAGE. [b] Theoretical molecular weight. [c] Sequence information obtained by C, cDNA; P, peptide sequence; N, nucleotide sequence. [d] Swissprot database. [e] Mass spectrometry determination.

The mono and doubly charged ions at *m*/*z* 16 300-17 829 and at *m*/*z* 8837-8989, respectively, were ascribed to Ole e 1 (Table 11, row 2).

The two a and b fractions (Fig. 29) show a significantly different antigenic pattern; a complete pool of Ole e 1 isoforms is, in fact, predominant in hydrophilic fraction **4b**.

The MALDI spectrum of fraction **4s** is characterized by the presence of ion peaks corresponding to low molecular weight proteins as a consequence of the fractionation procedure which lowers the solubility of lower molecular weight proteins in aqueous ammonium bicarbonate.

The MALDI spectra of the first hydro-soluble fractions of 1, 3, and 5-8 showed similar protein expression. In particular, the MS spectrum of fraction **5a** (Fig. 30b) shows four peaks in the mass range 9.7-10 kDa, whereas that of 3a (Fig. 30a) displays one additional peak at *m*/*z* 9186. A closer inspection of the main four peaks present in that mass range shows a difference of 70 and 180 mass units between two adjacent peaks and couples, respectively.

The allergens in the 9-10-kDa range (Table 11, column 3) that have already been found in olive tree pollen are Ole e 3 (Ledesma et al., 1998) and Ole e 7.

Ole e 7 consists of one polymorphic polypeptide chain of ~10 kDa, which accounts for 88 amino acid residues, as deduced from the elution profile on HPLC and determined by mass spectrometry (Tejera et al., 1999) (Table 11, column 5, row 8). Data obtained from Edman degradation of NH_2-terminal amino acids (Tejera et al., 1999) indicate that the number of isoforms of Ole e 7 is four and differs by the exchange of three amino acids at the positions 5 (S-G), 10 (L-K), and 18 (I-K), leading to a difference of -30, +15, and +15 mass units, respectively (Table 11, column 4, row 8). Therefore, the observed difference of 70 mass units (Fig. 30a) could be associated with the combined effect of a different peptide sequence, to the presence of calcium adducts, or both, whereas the difference of 180 should be associated with a hexose unit. The four ion peaks in the *m*/*z*

9.7-10-kDa range can be attributed to Ole e 7 because of its known polymorphism and its very high solubility in saline solution.

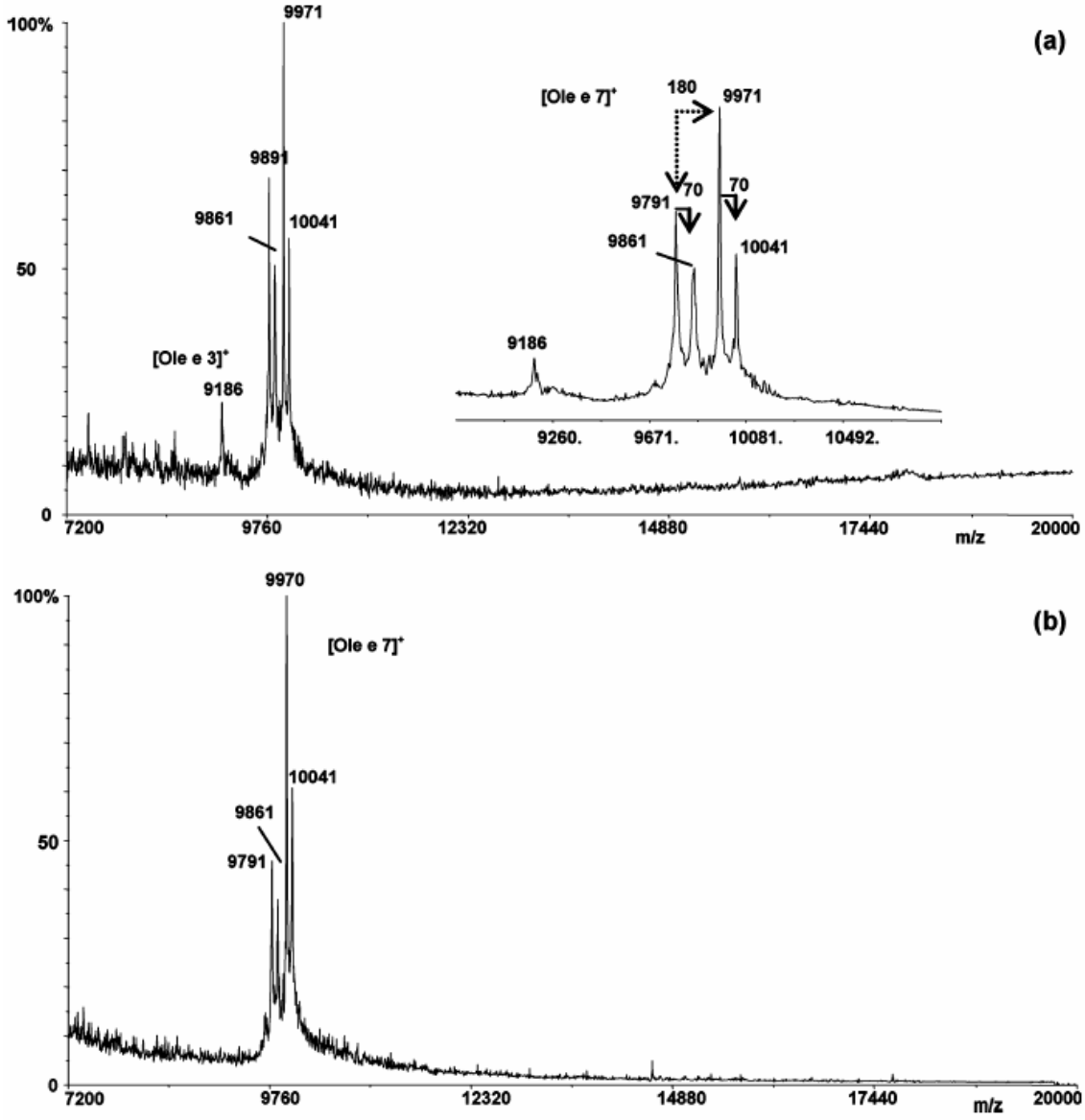


Fig. 30. MALDI spectra of fractions (a) **3a** and (b) **5a**.

Additional information was obtained for cultivar 3, whose pollen was consecutively extracted at room temperature with ammonium bicarbonate for 20, 60, and 120 min. The whole extracts **3I**, **3II**, and **3III** thus obtained were subsequently partitioned into three fractions (as shown in Chart 1) generically identified as **I-s**, **I-a**, and **I-b**, respectively.

The MALDI-MS spectrum of lipophilic fraction 3I-s, (Fig. 31a) shows ion peaks at *m*/*z* 5821 and 7396, likely corresponding to the species $[Ole\ e\ 6]^{+}$ and $[Ole\ e\ 2]^{2+}$, respectively; (Batanero, 1997; Barral et al., 2004; Trevino et al., 2004; Martinez et al., 2002; Asturias et al., 1997) those at *m*/*z* 9186 and 9256 (Δm) 70 Da), which can be attributed to $[Ole\ e\ 3]^{+}$; and that at *m*/*z* 9.7-10 kDa, which reasonably corresponds to $[Ole\ e\ 7]^{+}$. Two other different polypeptides at *m*/*z* 9686 and 9756, observed in fractions **3II-s** and **3III-s**, respectively (Fig. 31b and c), belongs to the isoform pool of Ole e 7.

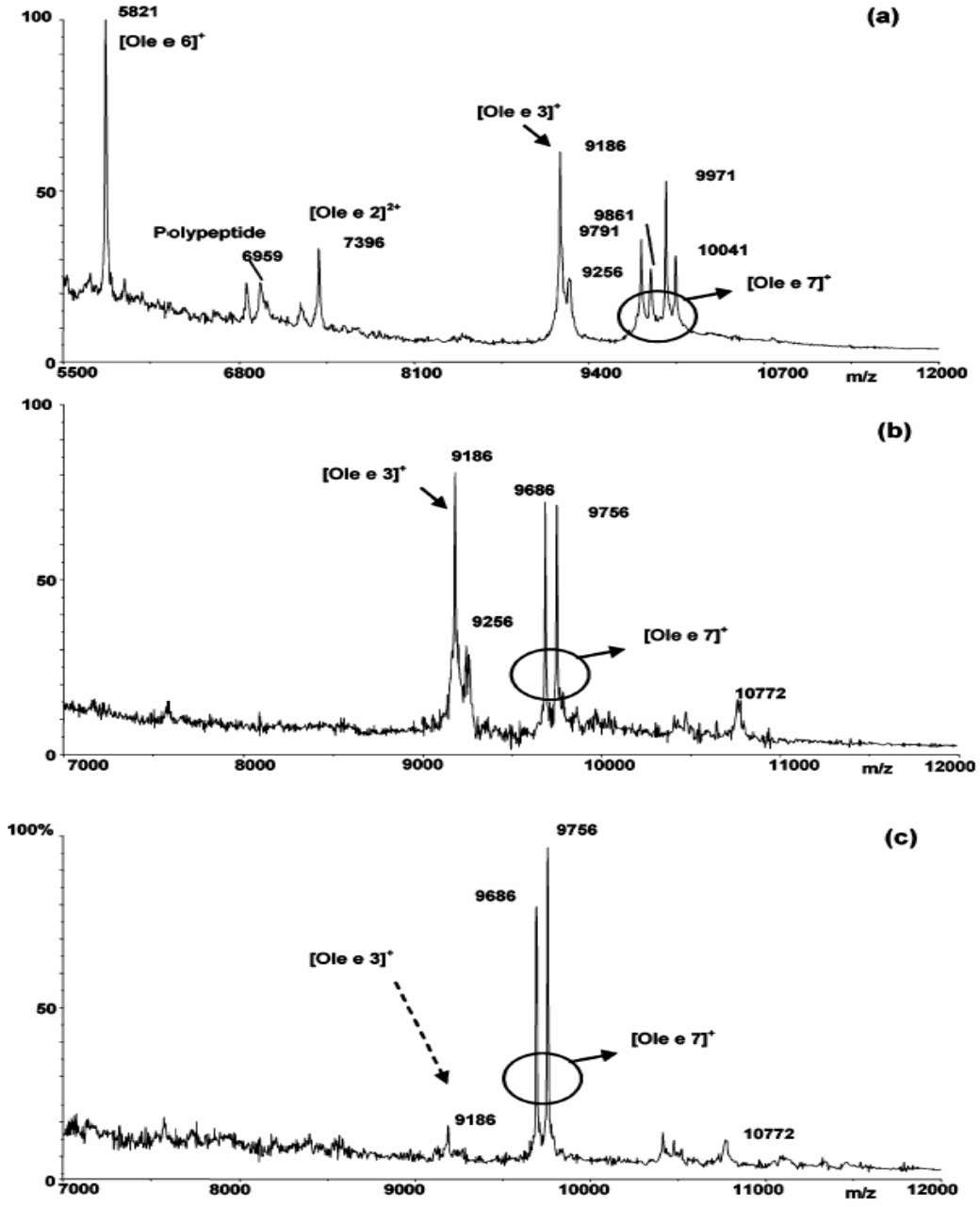


Fig. 31. MALDI spectra of fractions (a) **3I-s**, (b) **3II-s**, and (c) **3III-s**.

The antigen Ole e 3 (Table 11, row 4), known also as Poc3_Oleeu, could exist in the two cDNA-derived isoforms (Ledesma et al., 1998) whose 30-u difference is due to the replacement, at the positions 43 and 80, of the amino acid P with L and of V with I, respectively. The MALDI spectra of the two fractions **3I** and **3IIs** displayed two peaks, at *m*/*z* 9186 and 9256, differing by 70 u (Fig. 31a and b). Taking into account the calcium-binding properties (Ledesma et al., 1998) of Ole e 3, the observed difference likely proves the identification in the natural matrix of both the expected isoforms, as calcium ion adducts.

Ole e 3 and Ole e 7 are the only proteins identified in the whole extract of cultivar Ottobratica (1). MALDI spectra of fractions **1a** (Fig. 32a) show in the mass range 9.7-10 kDa the typical pattern of Ole e 7, previously discussed, whereas that of fraction **1b**

The presence of Ole e 3 in fraction 1b has been proved by a peptide mass fingerprinting (PMF) approach by tryptic digestion following deglycosylation by PNGase F. Database searching of the MALDI reflectron spectrum of the tryptic digest (Fig. 33) confirms that one component of the fraction **1b** is Polcalcin Ole e 3 (calcium-binding pollen allergen Ole e 3) (sptjO81092) with sequence coverage of 50%. The Ole e 3 identified sequences, detected by MALDI TOF, are shown in Table 12.

Table 12.

1 MADDPQEVAE HERIFKRFDA NGDGKISSSE LGETLKTLGS VTPEEIQRMM
51 AEIDTDGDGF ISFEEFTVFA RANRGLVKDV AKIF

start-end	Obs	*M*r (exptl)	*M*r (calcd)	Δ*M*r	Miss	sequence
1-16	1914.89	1913.88	1913.89	0.01	1	MADDPQEVAEHERIFK
18-25	823.37	822.35	822.35	0.01	0	FDANGDGK
18-36	1968.09	1967.08	1966.95	0.13	1	FDANGDGKISSSELGETLK
37-48	1329.71	1328.69	1328.69	-0.01	0	TLGSVTPEEIQR
72-78	757.50	756.49	756.46	0.03	1	ANRGLVK

A particular case is represented by the cultivar Carolea (2). Its fraction **2a** is characterized, in fact, by only one abundant doubly charged ion peak at *m*/*z* 7230, probably corresponding to Ole e 2.

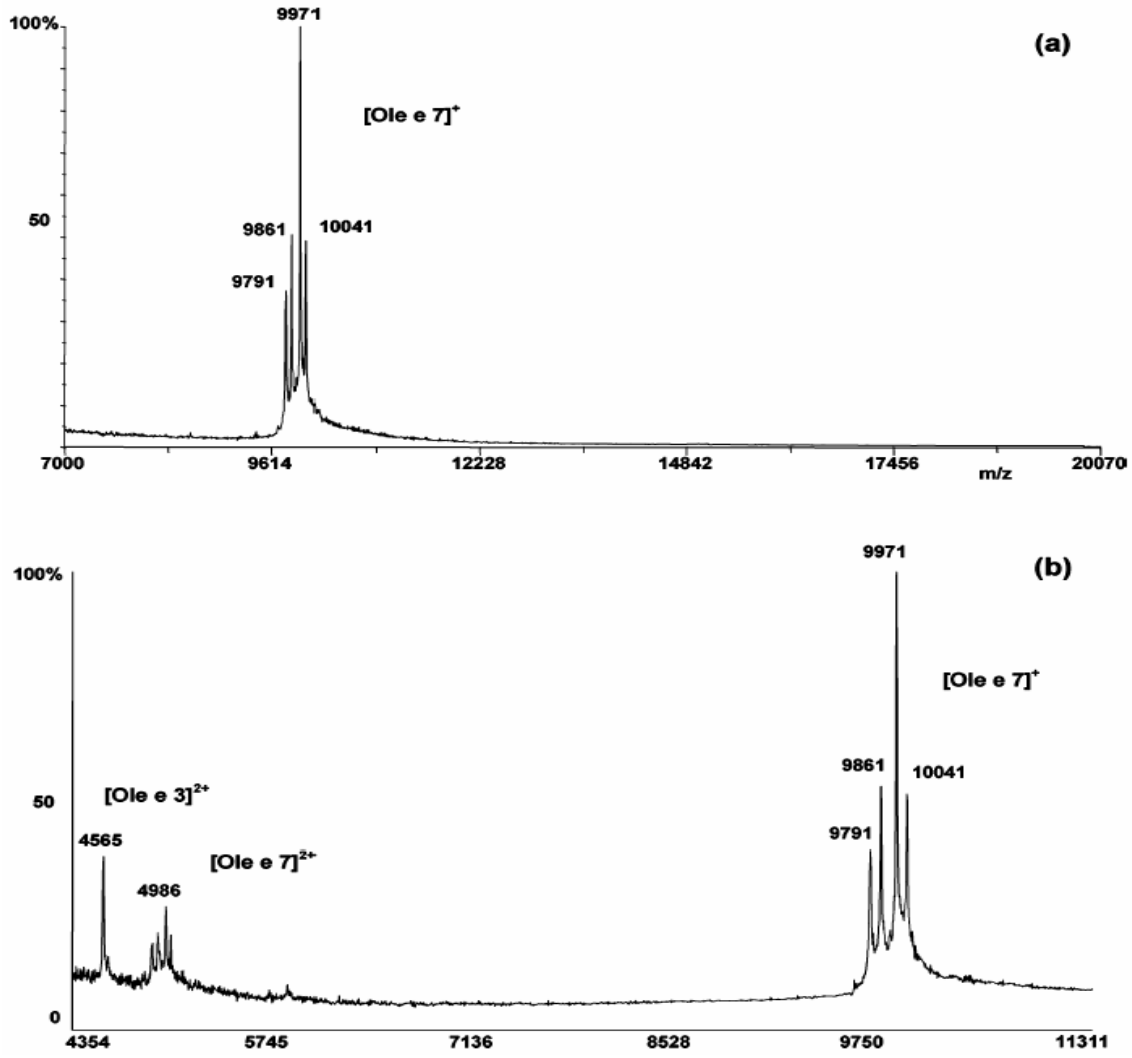


Fig. 32. MALDI spectra of fractions (a) **1a** and (b) **1b**.

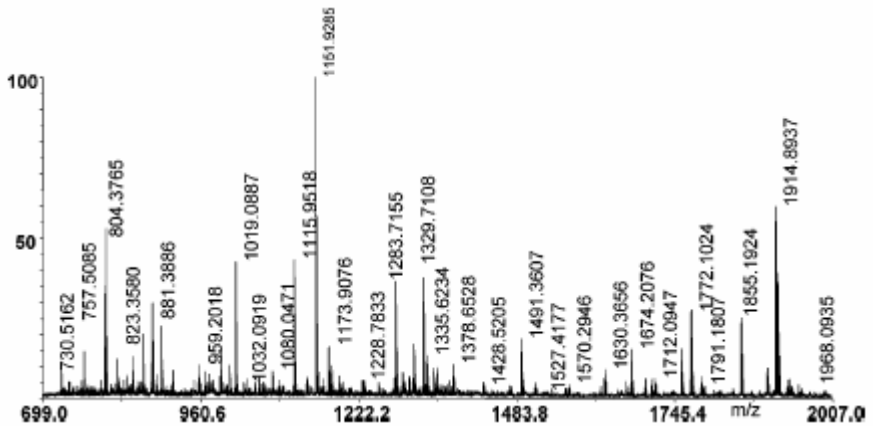


Fig. 33. MALDI-MS spectra of peptide mixture of fraction **1b**.

The antigenic profile of Coratina and Villacidro hydrosoluble **b** fractions, **5b** and **7b**, respectively, (Fig. 34a,b) shows a similar overenrichment of Ole e 1 allergens. The

fraction **7b** (Fig. 34b) contains the ionic species at *m*/*z* 8896, 17810, and 35635 only, corresponding to the doubly and monocharged and dimeric forms of Ole e 1, respectively. The peaks at *m*/*z* 8277 and 16 447 represent probably the doubly and the monocharged deglycosilated forms of the same allergen.

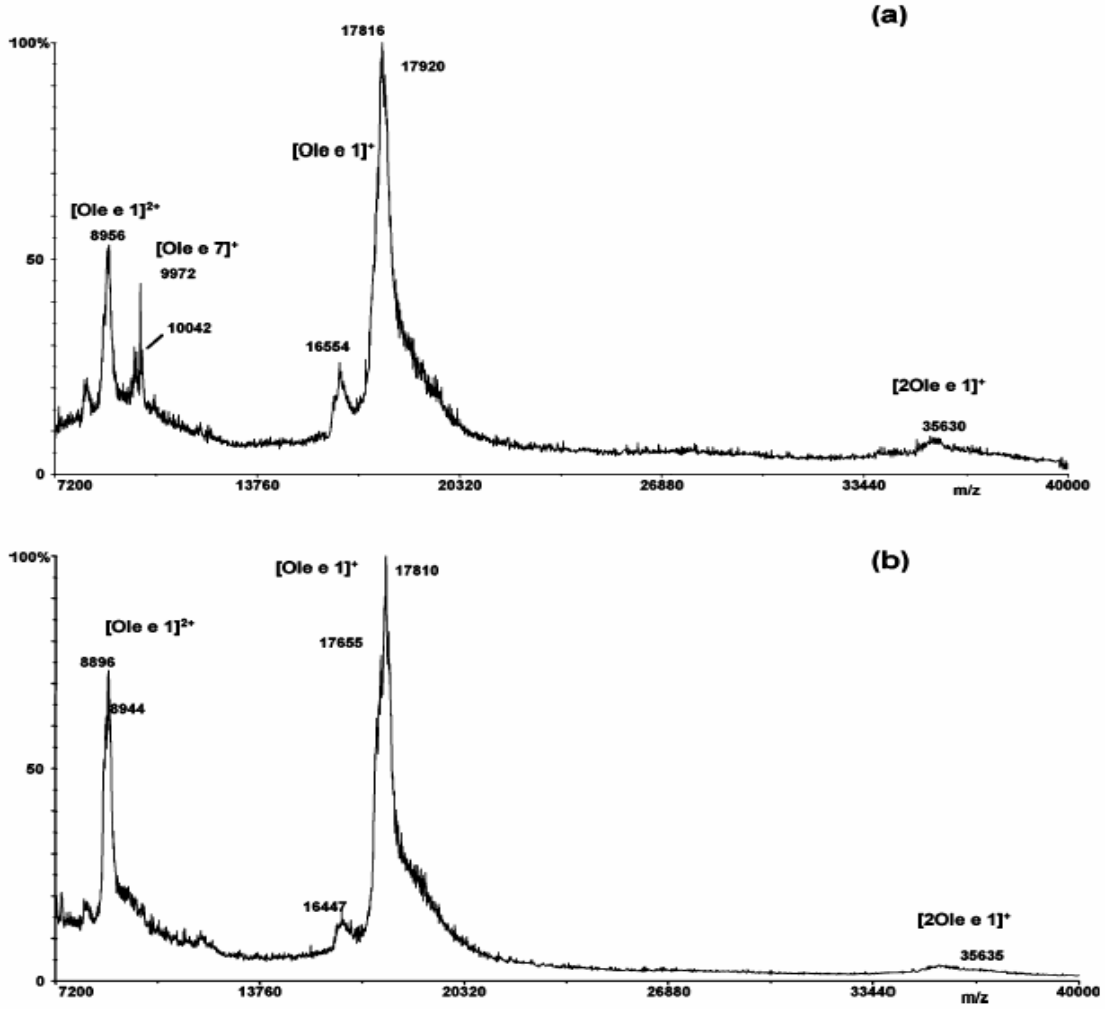


Fig. 34. MALDI spectra of fractions (a) **5b** and (b) **7b**.

Ole e 1 is the major allergen of *Olea europaea* (Table 11, row 2) and one of the best characterized allergens of the Oleaceae family; it has been isolated, purified, and biochemically characterized, (Villalba et al., 1990; Villalba et al., 1993) and its specific cDNA was cloned and sequenced (Lombardero et al., 1994; Villalba et al., 1994) (GenBank/ EMBL Data Bank accession numbers S75766 and X76395, respectively). Ole e 1 shows 85-95% identity with Lol p 11, Lig v 1, Syr v 1, Fra e 1 and Pla l 1 allergens and 36-38% identity with the deduced amino acids sequences from *LAT*52,

*Zmc*13, and *OSPSG* genes from tomato anthers and maize and rice pollens, respectively. (Lombardero et al., 2002).

Fraction **7b** was used for the identification of Ole e 1 by peptide mass fingerprinting after tryptic digest. The MALDI reflectron spectrum (Fig. 35) of tryptic peptides of intact Ole e 1 displays ion peaks corresponding to protonated tryptic peptides of the Ole e 1 family, with mass errors of 50 ppm.

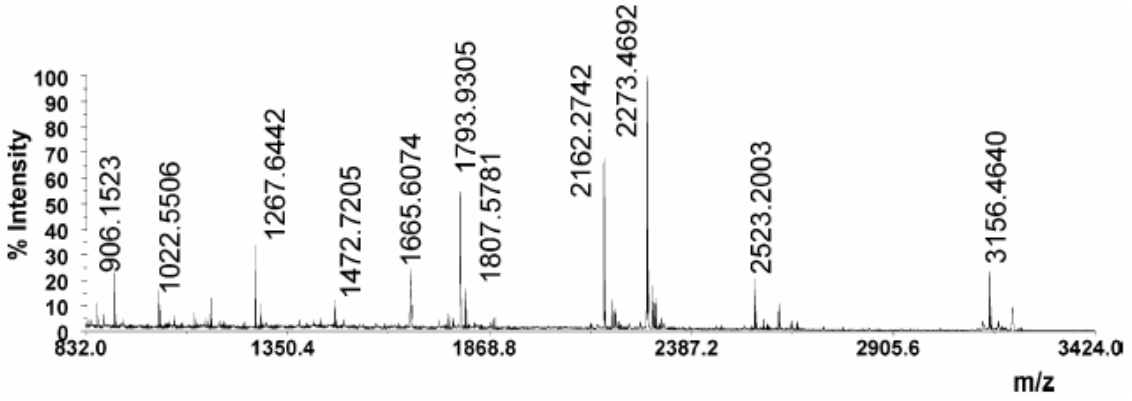


Fig. 35. MALDI-MS spectra of peptide mixture of fraction **7b**.

Database searching indicates that the major components of the mixture are (i) main olive allergen (*O. europaea*), (ii) major allergen OLE26, and (iii) major pollen allergen (allergen Ole e 1) with sequence coverages of 83, 62, and 50%, respectively (Table 13). The tryptic fragments of main olive allergen detected by MALDI-TOF are shown in (Table 14).

Table 13.

accession	description	mass	p.I.	sequence coverage, %	no. mass values matched
gi|13195753	main olive allergen (*Olea europaea*)	14 575	7.74	83	10
gi|1362133	major allergen OLE26, common olive (fragment)	15 288	7.74	62	6
gi|14424429	major pollen allergen (allergen Ole e 1) (Ole e I)	16 319	6.18	50	6

Table 14.

start-end	obs	*M*r (exptl)	*M*r (calcd)	Δ*M*r	miss	sequence
1-29	3234.45	3233.45	3233.57	-0.12	2	QVYCDTCRAGFITELSEFIPGASVRLQCK
9-2 5	1793.94	1792.93	1792.94	-0.01	0	AGFITELSEFIPGASVR
9-31	2523.20	2522.20	2522.32	-0.12	2	AGFITELSEFIPGASVRLQCKEK
32-45	1600.80	1599.80	1599.79	0.01	1	KNGDITFTEVGYTR
33-45	1472.72	1471.72	1471.69	0.02	0	NGDITFTEVGYTR
46-56	1267.64	1266.64	1266.63	0.01	0	AEGLYSMLVER
46-73	3156.46	3155.46	3155.53	0.07	2	AEGLYSMLVERDHKNEFCEITLISSGSK
74-89	1787.91	1786.90	1786.86	0.05	0	DCNEIPTEGWAKPSLK
90-100	1235.67	1234.66	1234.67	-0.00	0	FILNTVNGTTR
101-109	1022.55	1021.54	1021.56	-0.02	0	TVNPLGFFKK

The fractions **2b**, **3b**, and **8b** did not show any enrichment in Ole e 1, as previously observed for **5** and **7**, whereas they contain the antigenic mixture of Ole e 1, Ole e 2, Ole e 7 and one unidentified polypeptide at *m*/*z* 6959 (Fig. 36a-c).

The significant differences observed in the MALDI-MS spectra of **1b 8b** fractions suggest that the antigenic profile of the analyzed pollens can be divided into three distinct groups: (i) those with low Ole e 1 content, represented by Carolea (2), Dolce di Rossano (3), and Sinopolese (8); (ii) those characterized by the overenrichment of Ole e 1, corresponding to Cassanese (4), Coratina (5), Nocellara del Belice (6), and Villacidro (7); and (iii) that containing Ole e 3 and Ole e 7 only, which is represented by the cultivar Ottobratica (1).

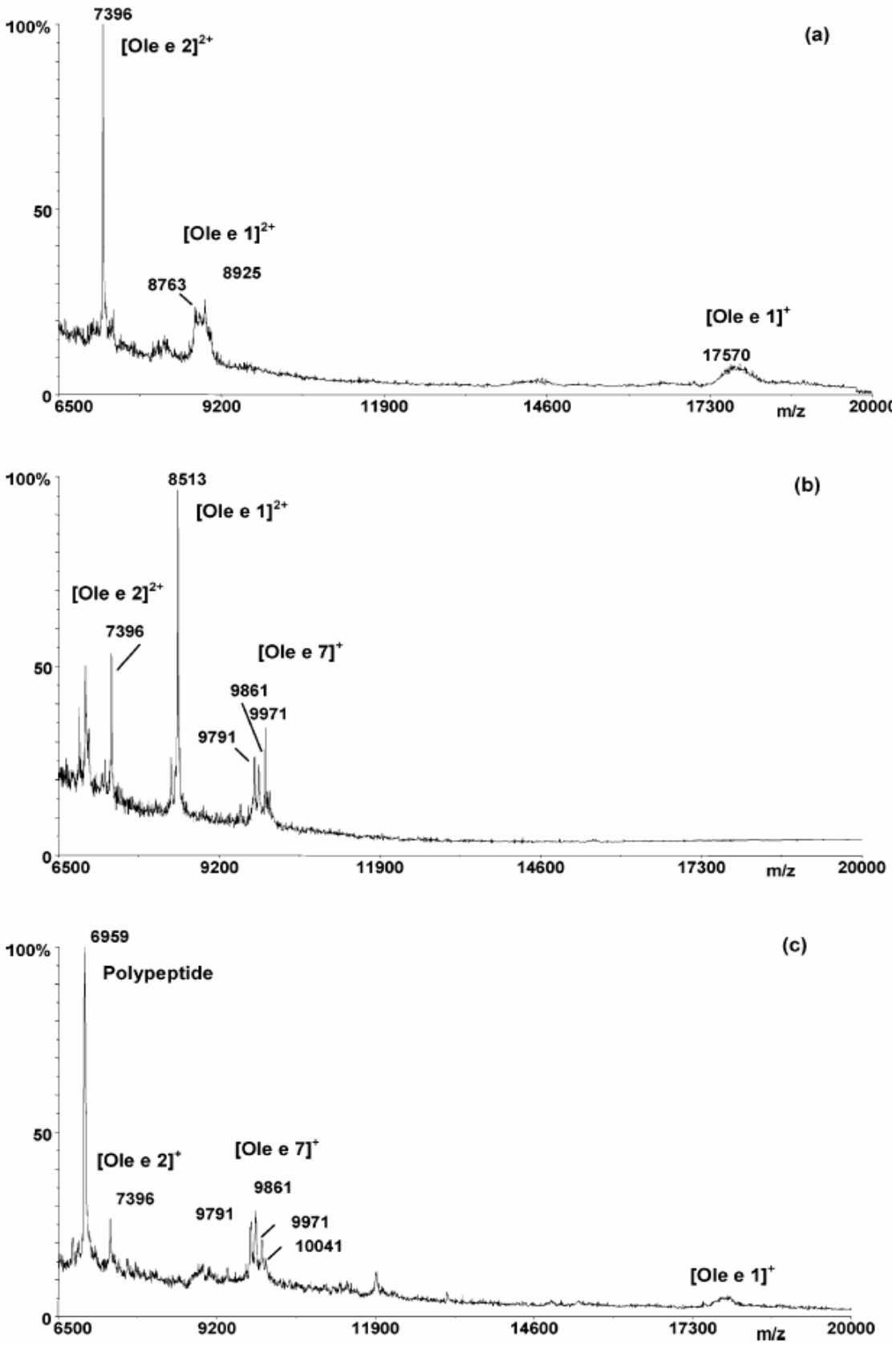


Fig. 36. MALDI spectra of fractions (a) **2b**, (b) **3b**, and (c) **8b**.

4.2 Extraction and charaterization of Olive (*Olea europaea* L.) Fruit and Oil Proteins from Italian Ccultivars by Gel Electrophoresis and MALDI TOF/TOF Mass Spectrometry[10].

In Mediterrranean basin *Olea europaea* has large economic interest for both fruits and oil production The proteic extract of olive fruit and oil could be useful for the characterization of some enzymes in order to determine their role in the biogenesis of the volatile compounds that influence the (aroma) and polyphenols (antioxidents) of extravirgin olive oil (Briante et al., 2002; Ridolfi et al., 2002). The most critical step in any proteomics study is protein extraction and sample preparation. The analysis of plant tissues is a difficult one for reproducible protein extraction, olive fruit contain high level of lipids. In this study we performed Phenol based method (Saravanan et al., 2004; Hurkman at al., 1986) with minor modifications to extract olive fruit protein from three Italian cultivars, the extraction procedure gave a high protein yield for electrophoretic separation. In-gel tryptic digestion of 1-D gel separated bends and MALDI TOF-TOF analysis was performed for the purpose of getting sequencing information by using online data base search.

The extravirgin olive oil is one of the antique and known vegetable oils. The ripe fruits of *Olea europea* tree are pressed to extract the oil, and is the only among vegetable oils because it can be consumed without any refining treatment. In the last years its popularity has grown due to health benefits like reduction of cholesterol LDL and the action of protection against some types of tumour (Bruckner, 2000; Simonsen et al., 1998). In this context it has been described the presence of various enzymatic activities like lipoxygenase in extravirgin olive oil, that takes part in the formation of volatile compounds, and is considered responsible for the aroma of the oil (Hatanaka, 1993), while β-glucosidase and esterase involve in hydrolysis of antioxidant molecule like oleuropein (Perri et al., 1999; Di Donna et al., 2001).

These considerations induce us to think about the extravirgin olive oil that the presence of protein as minor constituent, drive from fruit and transfer to the crude oil in the

[10] Parts of this research work presented (No. 5 & 6 Appendix)

crushing stage (Koidis and Boskou, 2006). While the polypeptides (4.6kDa) derived from the mesocarp of olive drupes are reported to persist even in refined oils (Hidalgo et al., 2001). It is problematic to extract protein due to the presence of phenolic compounds therefore, we compared a range of protein extraction protocols. These included methods (Koidis and Boskou, 2006; Hidalgo et al., 2001). In the present study we devised a simple system of protein extraction from the oil, that gives good electrophoretic separation of proteins for further identification through MALDI-TOF/TOF mass spectrometry.

4.2.1 Electrophoretic identification of fruit proteins.

The protocol presents a high resolution of olive fruit protein separation. SDS-polyacrylamide gel electrophoresis (PAGE) was performed according to the method of Laemmli (1970). In order to get reproducible and good protein separation we used different gel concentrations among them, by using 12.5% acrlamide gel (Fig. 37a & b Annex-I), protein bands were resolved but the bands belongs to low molecular mass were not well resolved. The best separation was achieved with 10% acrlamide gels (Fig. 37c Annex-I) and (Fig. 38a, b). Numerous protein bands revealed by the use of 1-D SDS-PAGE analysis, gels were stained with Coomassie brilliant blue. Electrophoratic pattern of olive fruit extract reveals well-resolved protein bands throughout the gel. The extensively stained major protein bands having molecular masses of approximately 55 and 65 kDa were observed. Additional lower-molecular-mass minor bands, ranging from 15 to 45 kDa, and higher-molecular mass less intense bands, ranging from 80 to116 kDa. The presence of minor/less intense bands could be due to proteolysis or loss during protein extraction procedure. While two band at 36 kDa probably could be the same proteins with different glycosylation degree. The presence of protein band of approximately 15 to 36 as well as 55 kDa, are also reported in olive leaves, among the major protein band of 55 kDa present in fruit protein extract is characteristic of leaf protein fractions of many plant species and correspond to the large subunit of ribulose-1,5-biphosphate carboxylase (Rubisco) (Wang et al., 2003; Garcia et al., 2000; Meyer et al., 1988; Hubbs and Roy, 1992). The second major protein band of 66 kDa could be the β-Glucosidase, which is reported as its belongs to the olive fruit tissues and is a key

enzyme in fruit ripening and defence response (Konno et al., 1999; Goupy et al., 1999; Brenes Balbuena, et al., 1992). The use of membrane technology to fractionate and purify protein extract from *Olea europaea* fruit containing β-glucosidase, and further separated by electrophoresis confirm the presence of protein band having molecular mass approximately 65 kDa (Mazzei et al., 2006). Another research study which supports our result in which it has been reported for the first time *in situ* detection of β-glucosidase in the mesocarp of *O. europaea* fruit during development and ripening stages (Mazzuca et al., 2006). Protein band of approximately 15, 25 and 36 kDa could probably belongs to Oleosins, it also has been reported that the Oleosins are alkaline proteins having molecular mass of 15–30 kDa (Gitte et al., 2001). No obvious differences in the number of major or minor protein bands were observed among the three cultivars used in this study except lane 2 (Nocellara del Belice) approximately 26 kDa (Fig. 37a & b). In order to evaluate the protein glycosylation, the sample Carolea was incubated with 5 µL of PNGase F (0.6 unit/ µL) for 3 days at 37°C. The electrophoretic separation shows no significant difference, that could be due to the proteins are not glycosylated and/or possibly *N* or *O*-linked glycoslyted, except the disappearance of a band of approximately 26 kDa (Fig. 38b lane 1).

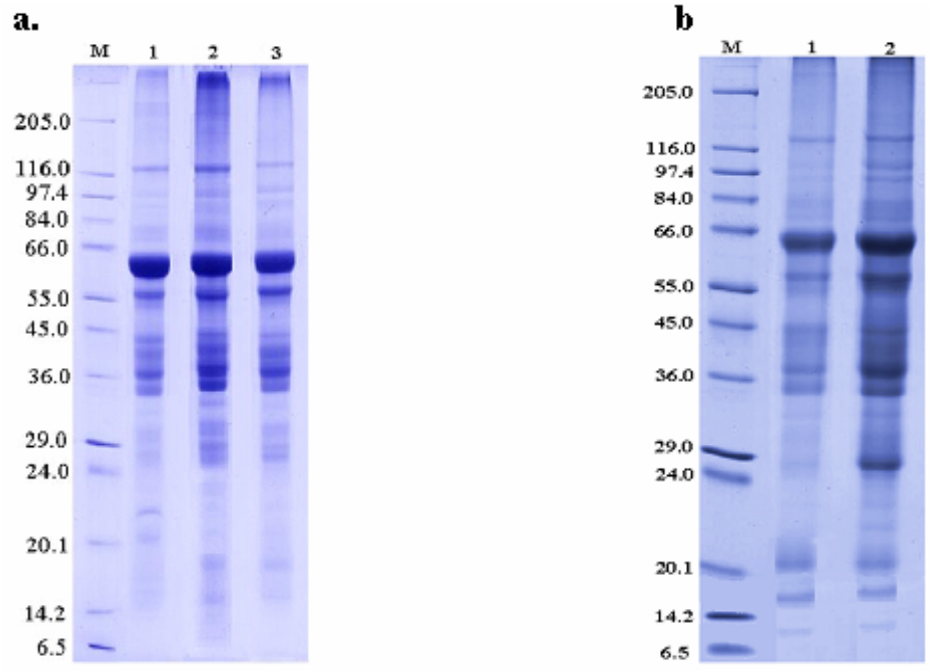


Fig. 38. SDS-PAGE separation of Olive (*Olea europaea.* L) fruit proteins. Gel **a.** 1. Carolea 2. Nocellara del Belice and 3. Coratina. Gel **b**. 1. Carolea 2. Nocellara del Belice . Samples were resolved using 10.0% plyacrylamide gel and visualized with Coomassie blue. Molecular weights (in kilodaltons) are shown to the left.

4.2.2 Electrophoretic identification of oil proteins.

Various methods for protein extraction were applied, which rely on different chemical principles. This study compares the precipitants for protein extraction of olive oil, after phenole extraction and precipitation with different concentration and ratio of Ammonium bicarbonate, Acetone, Methanol and trichloroacetic acid. Among the precipitants 20% trichloroacetic acid (TCA) yielded enough amount of protein pellet. The samples were analyzed by one-dimensional gel electrophoresis. All precipitation methods showed a smaller correlation with each other. However, the most intense bands were resolved for TCA-precipitated samples (Fig. 39. Annex-I). The protein was extracted from six cultivars after the oil extraction form fruits the samples were maintained at room temperature and at 4 °C.

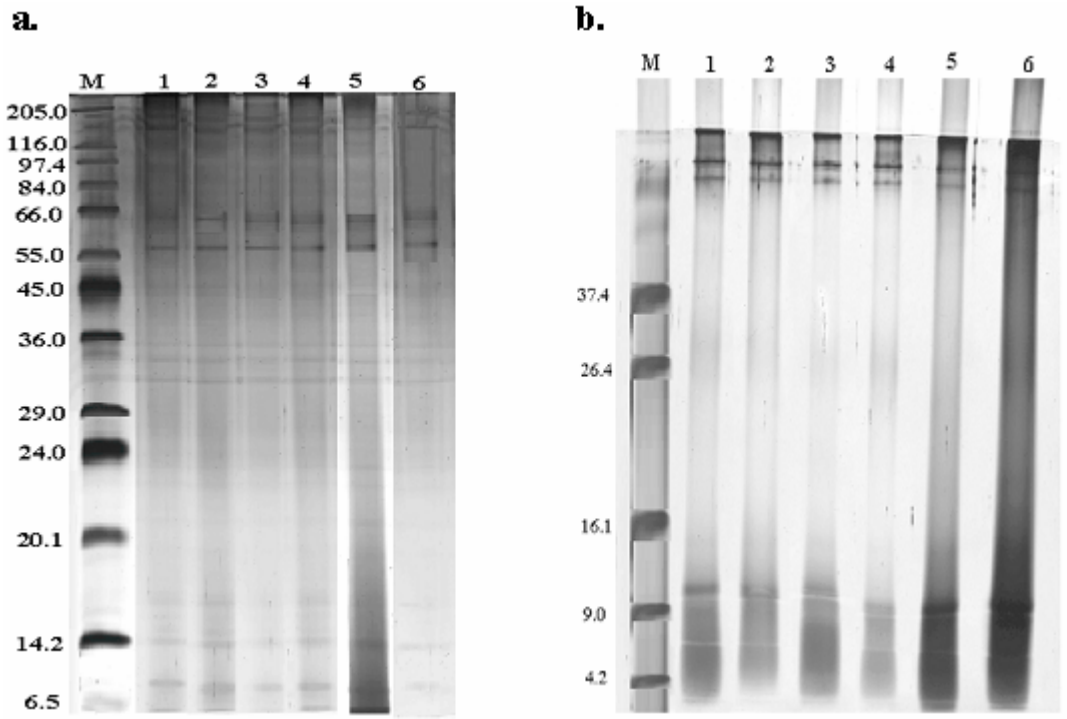


Fig. 42. SDS-PAGE eparation of olive oil proteins, gel a. & b. (1) Carolea (2) Frantoio (3) Cassanese (4) Coratina (5) Dolce di Rossano and (6) Ogliorola Barese, proteins were resolved using **a.** SDS-PAGE, 12.5.0% **b.** Tricine-DS-PAGE 16.5% plyacrylamide gel and visualized with silver stain. Molecular weights (in kilodaltons) are shown to the left.

Proteins were separated by SDS-PAGE using a discontinuous buffer system (Laemmli, 1970), with 0.75-mm and/or 1.5-mm thick 10 and 12.5% acrylamide separating gels, 4 and/or 6 % stacking gels. For separation of proteins of low molecular mass Tricine-

sodium dodecyl sulfate-polyacrylamide gel electrophoresis (tricine-SDS-PAGE) (Fig. 42b) was performed with 1.5-mm thick 16.5% total acrylamide gels containing 3% cross-linker (Schägger and Jagow, 1987). Gels were silver stained. The gels with the above mentioned two methods with different acrylamide concentrations are presented in (Fig. 39,40 & 41a,b,c. Annex-I). The best electrophoratic pattern was achieved with 12.5% acrylamide (Fig. 42a), shows intense bands having molecular masses of approximately from 55 and 66 kDa with dark background, to our knowledge the presence of these proteins bands in oil never been reported, the same band were also present in fruit as mentioned earlier. The protein band having molecular masses of approximately 55 kDa also present in olive leaves (Wang et al., 2003) and fruits. The presence of above bands supports the idea about the presence of enzymes in olive oils that pass from fruit during oil extraction. The lower-molecular-mass minor/less intense bands, ranging approximately from 6.5 to 14.2 kDa are also present. While no significance differences found among the six cultivars used in this study with regard to the relative intensity of bands, except lane 5 Dolce di Rossano (Fig. 42a) shows the presence of a polypeptide band less then 6.5 kDa. The presence of polypeptide band less then 6.5 kDa (Fig. 42b) with high background could be the olive oil protein fraction consists of a 4.6 kDa polypeptide (Koidis and Boskou, 2006; Hidalgo et al., 2001).

4.2.3 MALDI TOF-TOF analysis of in-gel trypsin digested bands.

After electrophoresis of fruit samples number of selected protein bands (Fig. 43) were excised and in-gel digested with trypsin and peptides mixture were analyzed by MALDI TOF/TOF. MS data obtained form all tryptic digested bands were matched through Mascot programs (www.matrixscience.com).

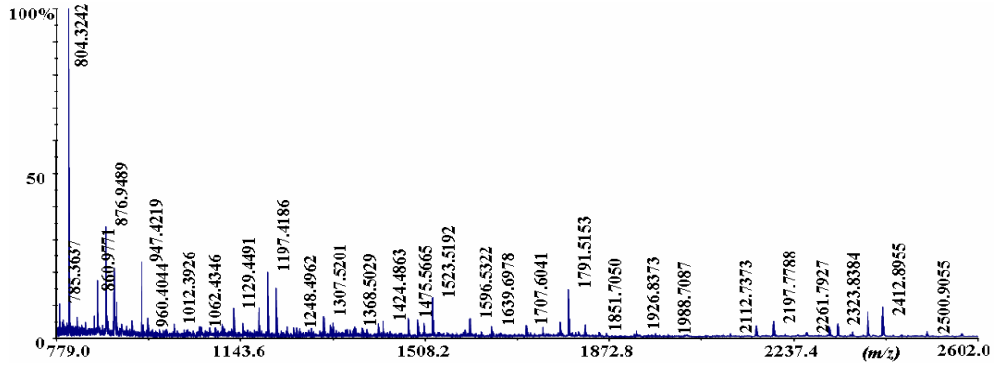


Fig. 43. MALDI MS: peptide mixture from band 66 kDa (Fig. 38b)

MS analysis and database search reveal the presence of Ribulose 1,5-bisphosphate carboxylase-oxygenase (Rubisco) in the band at 66kDa (Fig. 38b). The combination of two different approaches (protein identification based on cross-species sequence homology among *Viridiplantae* and protein identification based on an EST collection of a closely related species contributed to these results, together with the spectra and database researching. The database search and identification by homology search, the most intense peaks matched (NCBI BLAST) to Q5K3Y9_9ROSI (Ribulose 1,5-bisphosphate carboxylase-oxygenase) with Nominal mass (*Mr*): 52419; Calculated pI value: 6.01 and Taxonomy: (*Lachanea rupestris*) having Sequence Coverage: 20%. This protein containing 50% identity sequence with Ribulose 1,5-bisphosphate of *Olea europaea* (O20269), (Table 15). MSMS analysis (Fig. 44a & b) and MSMS data search confirm the presence of Ribulose 1,5-bisphosphate carboxylase-oxygenase from vegetables of the same taxonomy of *Olea europaea*, but different class (Table 16 and 17). In particular the peptides at m/z 1523.53 and 804.32 is matched in 50 isoforms of Ribulose 1,5-bisphosphate carboxylase-oxygenase belonging to different plants species (Garcia et al., 2000; Meyer et al., 1988; Hubbs and Roy 1992).

Table 15. Displays the identity of sequence

LACHNAEA Q5K3Y9	1	ETKASVGFKA	GVKEYKLTYY	TPEYETKETD	ILAAFRVTPQ	PGVPPEEAGA	AVAAESSTGT
OLEEU O20269		SVXFKAGVKE	YKLTYYTPEY	ETKDTDILAA	FRVTPQPGVP	PEEAGAAVAA	ESSTGTWTTV
	61	WTTVWTDGLT	NLDRYKGRCY	DIEPVPGEEN	QYICYVAYPL	DLFEEGSVTN	MFTSIVGNVF
		WTDGLTSLDR	YKGRCYHIEP	VPGEADQYIC	YVAYPLDLFE	EGSVTNMFTS	IVGNVFGFKA
	121	GFKALRALRL	EDLRIPTAYI	KTFQGPPHGI	QVERDKLNKY	GRPLLGCTIK	PKLGLSAKNY
		LRALRLEDLR	IPTAYVKTFQ	GPPHGIQVER	DKLNKYGRPL	LGCTIKPKLG	LSAKNYGRAV
	181	GRAVYECLRG	GLDFTKDDEN	VNSQPFMRWR	DRFLFCAEAI	YKAQAETGEI	KGHYLNATAG
		YECLRGGLDF	TKDDENVNSQ	PFMRWRDRFL	FCAEALYKAQ	AETGEIKGHY	LNATAGTCEE
	241	TCEEMMKRAV	FARELGVPIV	MHDYLTGGFT	ANTSLSHYCR	DNGLLLHIHR	AMHAVIDRQK
		MMKRAVFARE	LGVPIVMHDY	LTGGFTANTS	LAHYCRDNGL	LLHIHRAMHA	VIDRQKNHGI
	301	NHGMHFRVLA	KALRMSGGDH	IHSGTVVGKL	EGERDITLGF	VDLLRDDFIE	KDRSRGIYFT
		HFRVLAKALR	MSGGDHIHSG	TVVGKLEGER	DITLGFVDLL	RDDFIEKDRS	RGIYFTQDWV
	361	QDWVSLPGVL	PVASGGIHVW	HMPALTEIFG	DDSVLQFGGG	TLGHPWGNAP	GAVANRVALE
		SLPGVIPVAS	GGIHVWHMPA	LTEIFGDDSV	LQFGGGTLGH	PWGNAPGAVA	NRVALEACVK
	421	ACVQARNEGR	DLAREGNEII	RRACKWSPEL	AAACEVWKEI	KFEFEAVDIL	D
		ARNEGRDLAS	EGNVIIREAS	KWSPELAAAC	EVWKEIKFEF	AAMDTL	

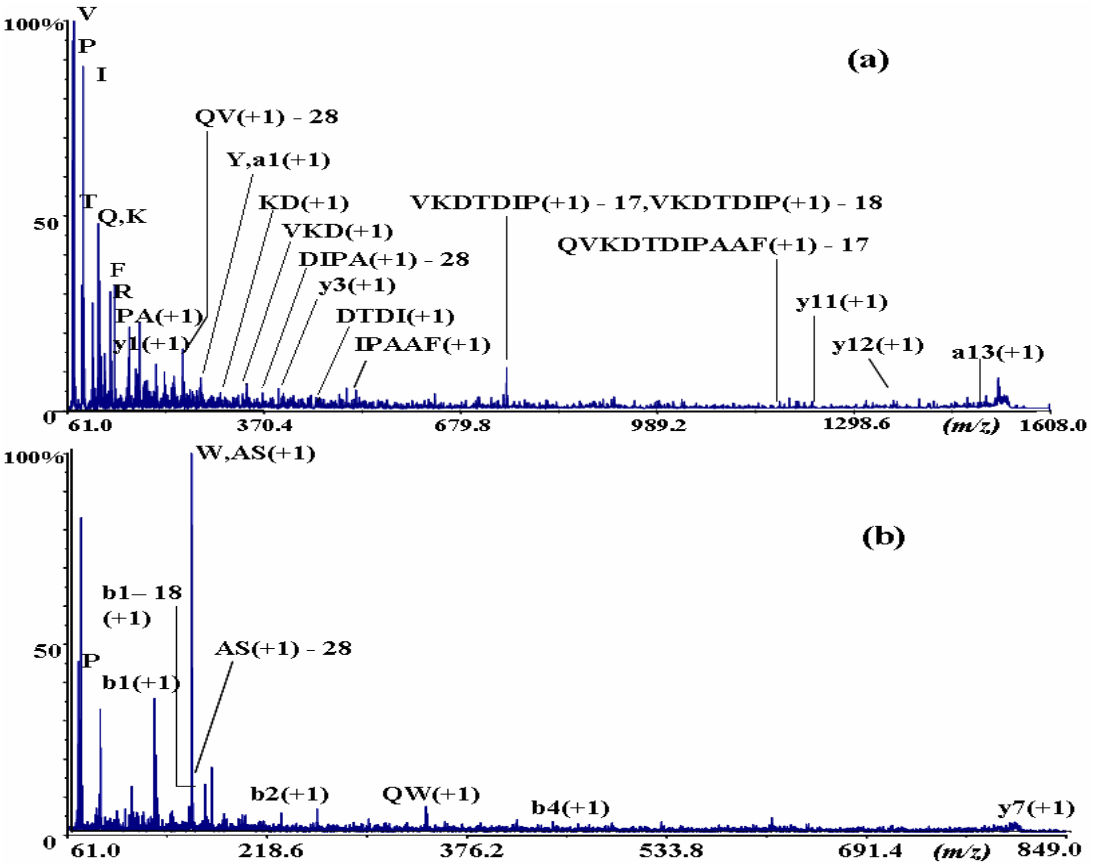


Fig. 44. MS/MS: peptide mixture from the peaks **a.** 1523.53 m/z and **b.** 804.34 m/z..

Table 16.

accession	Taxonomy	Mass	Peptides matched
gi|22859501	Other green plants (ribulose-1,5-bisphosphate carboxylase/oxygenase)	47149	1
gi|12274885	Other green plants (ribulose-1,5-bisphosphate carboxylase/oxygenase)	49077	1

Table 17.

Obs	*M*r (exptl)	*M*r (calcd)	Δ*M*ρ	Miss	sequence
1523.53.00	1522.52.00	1522.78	-0.2	1	YQVKDTDIPAAFR
804.32.00	803.31	803.34	-0.03	0	EASQWSP

MS analysis and database search also reveal the presence of two isoforms of **Oleosin** in the band at less than 36kDa.

Accession	**Mass**	**Description**
spt|P29530	23487 P24	oleosin isoform A (P89)
trm|Q9SS98	19742	Putative oleosin

The first one (spt|P29530) is from to the Glycine max., which belongs to the same taxonomy of *Olea europaea,* while the second one is from a very different plant family. The MW of matched proteins differs about 10 kDa from that observed by SDS-PAGE analysis. Proteins belong to vegetable world are characterized by a high homology. It can be suggested that the band observed at 36kDa probably corresponds to *Olea europaea* oleosin, which is reported as (*Olea europaea*) oil-body protein (Jonne et al., 2001). The *oleosin* cDNA gene have been isolated from young embryonic total RNA of *Olea europaea* with 165 amino acid residues and a predicted molecular mass 17.2 kDa and a pI at 9.91. Olive oleosin showed the highest homology (64%) or identity (51%) to oleosin from *Sesamum indicum*, followed by oleosin of *Daucus carota* (62% homology or 51% identity) and oleosin from *Gossypium hirsutum* (60% homology or 52% identity). Olive oleosin shows relatively low homology with glycine-rich proteins or oleosin-like proteins from *Arabidopsis thaliana* although the central hydrophobic domain is highly homologous (Polydefkis et al., 2002).

5. CONCLUSION

The origin of *Olea europaea* L. in the Mediterranean basin is unclear and was probably associated with a multilocal domestication of cultivated forms, as recent research with molecular markers has shown (Besnard and Bervillé, 2000; Besnard et al., 2001). Since olive cultivation occurred for millennium in all Mediterranean basin countries, the presence of a large number of synonymy and homonymy cases is thus probable. Therefore, the bona fide identity of olive germplasms and genetic variability present either in Italy or in other olive growing countries is still uncertain.

The results obtained in this work show that SSR markers can suitably characterize the olive collection such as the autochthon germplasm of the Abruzzo, Molise, Sicily and Apulia regions. Using existing olive SSRs primers we demonstrated that they can be successfully employed in cultivar characterization without the expensive development of new markers. In fact, the SSR loci used in this work were previously selected based on their high polymorphic index (Carriero et al., 2002; Cipriani et al., 2002). It should also be useful to distinguish other accessions that are introduced into the *CRA*. This study confirmed the utility of molecular analysis of olive germplasms and demonstrated the high level of polymorphisms in microsatellites, which is further augmented by DNA sequence analysis by revealing polymorphisms present in the microsatellite repeat. This is an important consideration since the additional costs of DNA sequencing must be weighed against the additional benefits in terms of cultivar identification and classification. In fact, the observed variability may also be used to measure genetic distances among different entities and to affirm with reasonable certainty if homogeneous genetic entities are present.

Almost all reports of cultivar identification using SSRs employed a greater number of primers with simple determination of the length of the amplified product. In the present study, we improved the specificity of this analysis by sequencing the amplicons from different SSR loci. In fact, alleles with similar or even identical lengths may have different sequences. The presence of sequence variants within microsatellites, including single nucleotide polymorphisms (SNPs), could be particularly useful in the

development of molecular markers for characterization of germplasms. The identification of different genotypes and the detection of cultivars 44 from Abruzzo & Molise, 41 from Sicily and 21 from Apulia the cultivars with unique genotypes demonstrated by analysis of SSRs show that it is an efficient tool to genotype the Italian collection of olive germplasms and may be useful to characterize additional accessions

Great efforts have been made in characterizing olive germplasms using different types of biochemical and molecular markers. Analysis of SSRs in cultivars in the studies of four regions germplasm collection allowed us to construct a molecular catalog that can be used to compare the molecular pattern of the various cultivars as well as to other samples of unknown origin, avoiding the collection of redundant genetic entities. The use of molecular markers like SSRs, in addition to other information, is imperative in order to build a database for cultivar analysis and for appropriate management of olive germplasm collections.

The mass spectral profiling of the examined samples (1-8) has to take into account that different allergenic and antigenic pattern can be found in the pollen of olive trees of the same species (Barber et al., 1990) and, also of different cultivars (Waisel et al., 1996) of the 260 existing varietal species. (Carnés et al., 2002) and/or 500 cultivars (Bartolini et al., 1998). Moreover, environmental adaptation and factor management could significantly affect qualitatively and quantitatively the protein profile of a given pollen. The experimental approach used so far is based on the SDS-PAGE or 2-D chromatographic separation followed by an immunoassay test (Tejera et al., 1999). The lack of sensitivity of those methods prevents the identification of cDNA-derived variants. The approach here proposed aims at characterizing olive cultivars from the profiling of the allergens of olive pollen by matching MALDI-TOF spectra of specific fractions containing chemically homogeneous proteins. The separation of proteins of different hydrophilicity from the lipophilic ones prevents the suppression effect typical of desorption methodologies (Rajnarayanan and Wang, 2004) and makes more effective the matching of MALDI spectra of different pollen fractions. The olive tree pollens (1-8) are distinguishable by the over enrichment of Ole e 1. An exception is represented by Ottobratica (1) which contains only Ole e 3 and Ole e 7. The data here presented

indicate clearly that two isoforms of Ole e 3 are present in the natural matrix and that Ole e 7 consists of four isoforms characterized by a different glycosilation degree. These results demonstrate that the proposed experimental procedure can supply valuable information on the antigens' micro heterogeneity.

In this study we devised a Phenol based method for olive fruit protein extraction for the application of proteomic analysis. Olive fruit is also like leaf tissue, notorious recalcitrant to common protein extraction methods due to the presence of high level interfering compounds. The different methods we applied for olive fruit proteic/enzymatic extraction failed to give electrophoratic separation, for *e.g* previously published method for protein extraction form olive leave (Wang et al., 2003), preparation of olive enzyme extract (Marta et al., 2002; Hidalgo et al., 2001). Our developed method is based on previously published methods (Hurkman at al., 1998; Saravanan and Rose., 2004) with minor modifications**,** after the Phe extraction the pellet was precipitated with acetone, and all the steps were carried out at room temperature. The average protein yield found to have 1.3 mg/g fresh weight. The extraction method gives optimal and reproducible 1-D SDS PAGE protein separation and could also provide well-resolved 2-D pattern.

The olive proteome is not completely sequenced, for that reason SDS-PAGE combined with MS and MSMS analysis are necessary to identify and characterized protein content in olive drupes. By SDS-PAGE profile it is possible to discriminate olive drupes belonging to different cultivars. Partially sequence of ribulose-1,5-bisphosphate carboxylase/oxygenase was identified by MS and MSMS experiments, conform the presence of (Rubisco) which is also present in olive leaves (Wang W. et al., 2003). This protein is the key enzyme in the Calvin Cycle and accounts for 16% of the protein content of the chloroplast and is likely the most abundant protein on earth. The possible function and presence of Rubisco have also been reported in oil synthesis in *Brassica napus* L. seeds, where Rubisco acts in a novel metabolic context, bypassing steps of glycolysis and increasing the efficiency of carbon use during conversion of sugars to oil (Schwender, 2006). The presence of two isoforms of Oleosin in olive fruit. It has been suggested that Oleosin play roles in the stability of oil bodies, and in their synthesis and

metabolism (Gitte et al., 2001). While another study it has been reported that the oil body fraction of olive mesoscarp tissue contained very little detectable protein, while oil bodies of mature olive seed tissue contained the Oleosin (Ross et al., 1993).

The method we developed during this study for olive oil protein extraction, gave reproducible electophoratic pattern of the analysed samples. The average protein yield found to have 0.8 to 2.0 mg/kg. Thus, in accordance with other studies in which the level of prtein was calculated from amino acid composition (Koidis and Boskou, 2006; Hidalgo et al., 2001) The oil sample were stored at room temperature or at 4˚C for further protein extraction showed no affect due to storage time period and temperature. After electrophoratic separation, number of protein bands were selected, excised and in-gel digested with trypsin and peptides mixture were analyzed by MALDI TOF/TOF. MS data obtained form all tryptic digested bands were matched. Due to unavailability of sequence information we were not able to get considerable peptide match. Therefore, it is concluded that MALDI-TOF/TOF mass spectrometry techniques could also identify peptides from olive oil. These peptides may prove to be useful markers of authentication or adulteration. Transformations of these proteins may occur as a result of refining, long-term storage or processing in complex food matrixes thereby providing further clues to sample history. Proteins should enhance oil stability due to their anti-oxidant properties (Pokorny et al., 2001), but (Hidalgo et al., 2002) observed a negative correlation between stability and protein content. Lipoxygenase and polyphenol oxidase catalytic activity has been reported (Georgalaki et al., 1998), the former are in part responsible for differences in the aroma of in different cultivars of extra virgin olive oil (Ridolfi et al., 2002).

6. EXPERIMENTAL SECTION

6.1 Plant material

All samples mentioned below were collected from plants growing in the olive germplasm collection of the C.R.A.- Experimental Institute for Olive Growing of Rende, Cosenza, located along the Ionian Coasts near Mirto-Crosia (Calabria region, southern Italy).

6.1.1 Leave samples.

A panel of 138 samples of olive plants were used, corresponding to the majority of the regional autochthon Italian germplasm. To evaluate intra-cultivar variability, we analyzed the accessions by verifying that the individuals belonging to the same cultivar showed identical genotypes (Lombardo et al., 2004). Among the studied samples, most of the cultivars are cultivated in the main production areas over a large surface (several thousand hectares), and are the most relevant for the regional agricultural economy. While some of the cultivars cover several thousand hectares, whereas the cultivars that are located in small areas and represent an important genetic reserve for the regional autochthon were also studied (Lombardo et al., 2004).

6.1.2 Pollen samples.

The olive (*Olea europea* L.) pollens of the Mediterranean cultivars, Ottobratica (1), Carolea (2), Dolce di Rossano (3), Cassanese (4), Coratina (5), Nocellara del Belice (6), Villacidro (7), and Sinopolese (8). From olive germplasm collection as mentioned above.

6.1.3 Fruit samples.

Samples of olive (*Olea europea* L.) fruit from cultivars (1) Carolea (2) Nocellara del Belice and (3) Coratina, were harvested by hand at the end of December 2005 from

plant growing in the olive germplasm collection site as mentioned above. Olive fruit were destoned by hand using a knife, stored at -80°C until protein extraction.

6.1.4 Oil samples.

Virgin olive oil samples were produced immediately after harvest at the end of December 2004 and 2005. Olive drupes (10 Kg) of each cultivars (1) Carolea (2) Frantoio (3) Cassanese (4) Coratina (5) Dolce di Rossano and (6) Ogliorola Barese, were used. Fruits were washed with distilled water, destoned by hand using knife. Monovarietal olive oils were produced from destoned olive fruits in an hammer mill. Malaxation was conducted at 25 °C for 20 min. No water was added for oil separation. Olive oil was obtained by centrifugation at 5000 x g for 20 min. Oil samples were stored at room temperature and at 4 °C until protein extraction.

6.2 DNA extraction.

Total genomic DNA was extracted from fresh leaves using the CTAB method according to Muzzalupo and Perri (2002). After extraction, samples were treated with RNase A (Sigma Chemical Co, St. Louis, Miss.) for 30 min at 37 °C and run on 0.8% (w/v) agarose gels (FMC BioProducts, Rockland, Maine) in 1X Tris-borate-EDTA (TBE) buffer in the presence of ethidium bromide (1 $\mu g \cdot mL^{-1}$). A DNA weight molecular marker (Sigma Chemical Co.) was used to estimate the concentration of samples.

6.2.1 DNA amplification.

Several primer pairs from previously published studies were evaluated. Among the primer sets designed by Carriero et al., (2002), GAPU45, GAPU59, GAPU71A, GAPU71B and GAPU 103A, were assessed due to the relatively large number of different alleles amplified. For the same reason, the UDO01, UDO-003, UDO-012, UDO-028 and UDO39 loci described by Cipriani et al., (2002) were also amplified. PCR reactions were carried out in 50 µL containing 20 ng of DNA, 10 mM Tris-HCl pH 8.0, 1.5 mM MgCl2, 50 mM KCl, 0.2 mM dNTPs, 0.25 µM forward and reverse

primers, and 0.05 units of Taq DNA polymerase (Roche Diagnostics, Manheim, Germany).

PCR conditions comprised an initial denaturing step at 94 °C for 3 min, followed by 40 cycles of 94 °C for 30 s, 60 °C for 40 s and 72 °C for 30 s plus a final extension at 72 °C for 5 min in a GeneAmp PCR 9600 (PE Applied Biosystems, Foster City, Calif.). PCR products were analyzed using 2100 Bioanalyzer (Agilent Tecnologies, Waldbronn, Germany) by using the DNA 500 LabChip Kit (Burns et al., 2003), which provides an estimate base pair of the length of any amplified product.

6.2.2 DNA sequencing.

All the SSR amplification products were subjected to direct DNA sequence analysis since different alleles may have similar or even identical lengths, but different repeating units. Consequently, the exact number of repeat units was established for every sample. PCR products were run in 3% (w/v) agarose gels (FMC BioProducts) in TBE 1X buffer in the presence of ethidium bromide (1 $\mu g \cdot mL^{-1}$) at 100V for 4 h. Amplified bands were excised from the gel and purified using the QIAquick Gel Extraction kit (Qiagen Spa, Milan, Italy). The purified products were analyzed using a 2100 Bioanalyzer (Agilent Tecnologies) on a DNA 500 Chip (Burns et al., 2003) and the yield of template for any allele was estimated for sequencing reactions.

Sequence analysis was performed utilizing Sequencing Analysis 3.7 software (PE Applied Biosystems). Alleles of the same length were isolated from 3% (w/v) agarose gels. The PCR products were cloned using a PCR-ScriptTMAMP Cloning Kit (Stratagene, La Jolla, Calif.) following the manufacturer's instructions. Plasmid DNA was isolated using a NucleoSpin Plasmid kit (Macherey-Nagel AG, Oensingen, Switzerland) and the inserts were sequenced.

The sequencing reactions were performed in a GeneAmp PCR 9600 (PE Applied Biosystems) using a BigDye Terminator v.1.1 Cycle Sequencing Kit (PE Applied

Biosystems) utilizing 25 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 4 min and then loaded on a ABI Prism 310 Sequencer (PE Applied Biosystems).

6.2.3 Data analysis.

Data were processed using POPGENE 32 software (Yeh et al., 1997). The software allowed calculation of the number of alleles, their frequency, and their observed and expected heterozygosity (*Ho* and *He*, respectively; Nei, 1973). The number of alleles detected (*Nde*, counts the number of alleles with nonzero frequency) and exclusive (*Nex*, counts the number of alleles present) with and without sequencing SSR loci was calculated using the same software (Tables 2,3,5, 7 and 10). The probability of null alleles was estimated according to the formula of Brookfield (1996) $r = (He - Ho)/(1 + He)$. The results of alleles identification were then used to create a qualitative data matrix of presence (1) and absence (0) that was processed using NTSYS-PC software (version 1.80; Rohlf, 1994). Dice's coefficient (Dice, 1945) was used and the cultivars were grouped by cluster analysis using the unweighted pair-group method (UPGMA) (Figures. 24-28). The cophenetic correlation coefficient was calculated and Mantel's test (Mantel, 1967) was performed to check the goodness of fit of a cluster analysis to matrix on which it was based.

6.3 Allergen extraction.

Allergen extracts. Portions (50 mg) of pollen grains (1-8) were extracted with 1 mL of aqueous 50 mM NH_4HCO_3 for 20 min at room temperature, followed by centrifugation at 14 000 rpm for 2 min. The supernatant portion was separated and stored at -20 °C. A 200-µL portion of whole extract was precipitated with 400 µL of $CHCl_3/CH_3OH$ 1:3 (v/v), and the precipitated protein pellet was partitioned consecutively, under magnetic stirring and at room temperature, for 10 min with (a) 150 µL of 50 Mm NH_4HCO_3/CH_3OH 1:1 (v/v) and (b) 150 µL of 50 mM NH_4HCO_3/ CH_3CN 1:1 (v/v). The MALDI spectra of the residue of the previous extractions reveals traces of the allergens found in fraction b. Each step (a, b) was followed by centrifugation at 14 000 rpm for 1 min.

A 50-µL portion of fraction 1b (pH 8) was fully reduced by treatment with 10 µL of DTT (50 mM) under magnetic stirring for 6 h at room temperature. The sample was then incubated with 4 µL of PNGase F (0.5 unit/µL) for 3 days at 37 °C. After 78 h, the resulting mixture was digested overnight with 1 µL (10 pmol/µL) of trypsin.

A 50-µL portion of fraction 7b (pH 8) was fully reduced by treatment with 10 µL of DTT (50 mM) under magnetic stirring for 6 h at room temperature. The sample was then digested overnight with 1µL (10 pmol/µL) of trypsin.

6.3.1 Mass Spectrometry.

A 1-µL portion of each fraction of olive pollen and in-gel digested with trypsin and peptides mixture were analyzed were analyzed by linear MALDI using R-cyano-4-hydroxy-*trans*cynnamic acid (R-CHCA, 0.3% in TFA) as matrix.

MALDI-TOF analyses were performed using a 4700 Proteomics Analyzer mass spectrometer from Applied Biosystems (Foster City, CA) equipped with a 200-Hz Nd:YAG laser at 355-nm wavelength. The MS spectra were acquired in reflectron mode (20-keV accelerating voltage), with 400-ns delayed extraction, averaging 2000 laser shots with a mass accuracy of 50 ppm. A 1-µL portion of a premixed solution of each fraction for pollen samples, in-gel tryptic digested (samples from fruit & oil) peptides mixture and R-CHCA (0.3% in TFA) was spotted on the matrix target, dried at room temperature, and analyzed in the mass spectrometer. Proteins were identified by searching a comprehensive protein database using Mascot programs (www.matrixscience.com). One/two missed cleavages per peptide was allowed, and an initial mass tolerance of 50 ppm was used in all searches.

Linear MALDI MS spectra were acquired averaging 2500 laser shots with a mass accuracy of 500 ppm in default calibration mode that was performed using the following set of standards: insulin (bovine, $[M + H]^{+}_{avg}$) 5734.59), apomyoglobin (horse, $[M + H]^{2+}_{avg}$) 8476.78, $[M + H]^{+}_{avg}$) 16 952.56), and thioredoxin (*Escherichia coli*, $[M + H]^{+}_{avg}$) 11 674.48).

6.4 Protein extraction from olive fruits.

Protein from olive fruits were extracted according to the previously described procedure (Saravanan et al., 2004; Hurkman at al., 1986) with some modifications. One grams of frozen plant tissue was finely powdered in liquid nitrogen using a pestle and mortar and resuspended in 5 mL of extraction buffer 1 (3% polyvinylpyrrolidone (PVP), 0.7 M sucrose, 0.1 M KCl, 0.5 M Tris-HCl pH 7.5, 500 mM EDTA, 1mM *p*-Toluenesulfonyl fluoride, 1% β-mercaptoethanol). The mixture was extensively homogenized on ice for 30 min using Ultra-Turrax T8 IKA. Werke (GMBH & Co. Germany), then the homogenate was centrifuged at 13,000 rpm for 15 min, the supernatant an equal volume 1:1 of Tris-HCl pH 7.5-saturated Phenol was added and the mixture was homogenized, than the mixture was centrifuged at 13,000 rpm for 30 min. The upper Phe phase was removed and precipitated with five volumes of saturated ammonium acetate in methanol overnight at -20ºC. Finally pelleted by centrifugation at 12,000 rpm for 30 min and washed with cold saturated ammonium acetate in methanol once and cold 80% acetone twice. The pellet was dried and dissolved in a buffer of choice.

6.5 Protein extraction from olive oil.

Oil 4ml + 8ml 20 %TCA (1:2), vortexed and left at 4 °C for half hour. Centrifuged at 12000 r.p.m for 5 min (4 °C) the upper oil phase was discarded. To the aqueous part added ~8ml cold acetone 80%, vortexed for 1 min and centrifuged at 12,000 r.p.m for 5 min (4 °C). The pellet was washed with once with cold 80% Acetone & one wash with Hexane. The pellet was air dried at room temp and dissolved in a buffer of choice. The extracted oil samples were maintained for different time period before protein extraction at room temperature and 4 °C in order to monitor if the incubation /storage time or temperature affect the presence of protein in oil.

6.6 Determination of total protein concentration.

Total protein was determined by the micro-Bradford method (Bradford reagent, Sigma Chemical Co), which utilises the method of Bradford (1976); involving the shift in

A595nm caused by the binding of Coomassie Brilliant Blue G-250 to protein. The kit was used as per the manufacturer's instructions. A standard curve was prepared with BSA. The average protein yield from fruit samples found to have 1.3 mg/g fresh weight, the oil samples the average protein yield found to have from 0.8 to 2.0 mg/kg.

Table.18 (concentration of protein the average of two extraction per sample)

Olive fruits	Protein yield mg/g	Olive oils	Protein yield mg/kg
Carolea	1.42	Carolea	0.81
Nocellara del Belice	1.01	Frantoio	0.87
Coratina	1.58	Cassanese	1.75
		Coratina	1.23
		Dolce di Rossano	2.0
		Ogliorola Barese	1.68

6.7 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) .

Samples were prepared as individual sections 6.4 and 6.5 specified. Proteins were separated by one dimensional SDS-PAGE using a discontinuous buffer system (Laemmli, 1970), with 0.75 and 1.5-mm thick 10 and 12.5% (w/v) acrylamide (29:1 acrylamide-bisacrylamide) separating gels with or without glycerole, and 4 and 6 % (w/v) stacking gels. Samples diluted 1:1 in the gel loading buffer were boiled in a water bath for 5 min, and after they had been cooled, they were applied onto the polyacrylamide gel. The 10 and 15 -well form comb was used depending upon the apparatus used. Sigma and Bio-Red molecular weight standards, broad and small range were used in the marker lane. Electrophoresis was carried out at 30 mA constant current (Power supply Bio-Rad Laboratories 1000/500 and/or Amersham Biosciences SPS 601) until the tracking dye front reached the bottom of the gel (approximately 4 hr). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis (tricine-SDS-PAGE) was performed with 1.0-mm thick 16.5% total acrylamide gels containing 3% cross-linker (Schägger and Jagow, 1987).

6.8 Polyacrylamide gel staining.

***Coomassie blue staining*:** After electrophoresis the gel were stained with 0.025% coomassie brilliant blue R250, 40 % (v/v) methanol, for over night. After staining, immersed the gels in destaining solution (40% methanol, 10% acetic acid) for 3 hours

on a shaking platform. Changed the destaining solution 3-4 times until the clear background was achieved.

***Silver staining*:** The method used is based to the method of Heukeshoven and Dernick (1985). Processed the gel(s) according to Table 18. Used 250 ml of each solution per gel for standard (14 x 16 cm) gels. Shaked slowly on a shaking platform.

Table. 19

Solution	Time for 1.5-mm thick gel
Silver staining fixing solution (40% (v/v) ethanol, 10% (v/v) acetic acid)	30 min
Sensitizing solution (30% (v/v) ethanol, 6.8% (w/v) sodium acetate, 0.2% (w/v) sodium thiosulphate, 0.125% (v/v) glutaraldehyde)	30 min
Distilled or deionized water	3 x 10 min
Silver solution (0.25% (w/v) silver nitrate, 0.015% (v/v) formaldehyde)	30 min
Distilled or deionized water	2 x 1 min
Developing solution (2.5% (w/v) sodium carbonate, 0.0074% (v/v) formaldehyde)	5–10 min
Stopping solution (1.5% (w/v) Na_2EDTA)	10 min
Distilled or deionized water	3 x 5 min
Preserving solution (30% (v/v) ethanol, 4% (v/v) glycerol)	30 min

6.9 In-Gel Tryptic Digestion.

In-gel trypsin digestion of the proteins and further analysis of the resulting peptide fragments by mass spectrometry (MS). This method for the protein identification and characterization based on (Aulak et al., 2001; Hilton et al*.,* 2001; Lim et al., 2002; Tani et al., 2000). Coomassie or silver stain was used to visualize proteins that were separated in 1-D gels.

Prepared, 1 μg/0.1 ml solution of high-quality trypsin in ultrapure water.

6.9.1 Gels Stained with Coomassie blue.

Destaining Solution: 25 mM ammonium bicarbonate in 50% acetonitrile.
Digestion Buffer: 25 mM ammonium bicarbonate in water.

After excising the band of interest by scalpel from 1-D gel pieces, added 200 µl destaining solution to gel pieces. Incubated the samples at 37°C for 30 minutes with shaking. Removed and discarded destaining solution from the tube and repeated the above steps twice. Shrieked the gel pieces by adding 50 µl of acetonitrile. Incubated the samples for 15 minutes at room temperature. Carefully removed acetonitrile and allowed the gel pieces to air-dry for 5-10 minutes. Added 10 µl (or more in order to cover gel piece) of prepared trypsin solution to the tube containing the shrunken gel pieces; incubated at room temperature for 15 minutes to allow gel pieces to swell and absorb the trypsin solution. Added 25 µl of Digestion Buffer to the tube. Incubated the samples at 37°C for overnight with shaking. Removed the digestion mixture and placed in a clean tube.

6.9.2 Gels Stained with Silver.

Destaining Solution: Prepared separate 100 mM sodium thiosulfate and 30 mM potassium ferricyanide solutions, then mixed them in a 1:1 (v:v) ratio.

Wash Solution: 25 mM ammonium bicarbonate in 50% acetonitrile.
Digestion Buffer: 25 mM ammonium bicarbonate in water.

Added 0.1 ml of destaining solution to excised gel piece. (a sufficient volume to completely cover gel piece.) Incubated samples at room temperature for 15 minutes with shaking. Decanted and discarded destain solution. Added 0.2 ml of wash solution to the sample and incubated with gentle shaking for 10 minutes. Repeated the above steps twice. Added 10 µl (or little more in order to cover gel piece) of prepared trypsin solution to the tube containing the shrunken gel pieces; incubated at room temperature for 15 minutes to allow gel pieces to swell and absorb the trypsin solution. Added 25 µl of Digestion Buffer to the tube. Incubated the samples at 37°C for overnight with shaking. Removed the digestion mixture and placed in a clean tube.

REFERENCES

Aaij, C., Borst, P. Biochim. *Biophys. Acta*. **1972,** 269:192.

Allett, B., Jeppesen, P. G.N., Katagiri, K. J., Delius, H *Nature*. . **1973,** 241:120.

An der Lan, B., Chrambach, A. *IRL Oxford Press*. **1985.**

Anderson, N.G., Anderson, N.L. *Electrophoresis*. **1996,** *17*, 443-453.

Angiolillo, A., Mencuccini, M., Baldoni, L. *Theor Appl Genet.* **1999,** *98*:411-421.

Asturias, J.A., Arilla, M. C., Gomez-Bayon, N., Martinez, J., Martinez, A., Palacios, R. *J. Allergy Clin. Immunol.* **1997,** *100*, 365-372.

Aulak, K. S., Miyagi, M., Yan, L., West, K. A., Massillon, D., Crabb, J. W., Stuehr, D.J. *PNAS* 98. **2001**,12056-61.

Bahrman, N., de Vienne, D., Thiellement., H. Hofmann., J.P. *Biochem. Genet*. **1985,** *23,* 247–255.

Barber, D., Carpizo, J., Garcý`a-Rumbao, M.C. Polo, F., Juan, F. *Ann. Allergy*. **1990**, *64*, 43-46.

Barral, P. Tejera, M.L., Trevino, M. A., Batanero E., Villalba, M., Bruix, M., Rodriguez, R. *Protein Expr Purif*. **2004,** *37*(2), 336-43.

Bartolini, G., Prevost, G., Messeri, C., Carignani, G. FAO *Plant Production and Protection Div*. **1998,** Rome, Italy.

Bartolini, G and Petruccelli, R. *Food and Agricolture organization of the United Nations* **1998** Rome, Italy.

Batanero, E., Ledesma, A., Villalba, M., Rodríguez, R. *FEBS Lett*. **1997,** *57*, 293–296.

Belaj, A., Satovic, Z., Ismaili, H., Panatoti, D., Rallo, L., Trujillo, I. *Euphytica*. **2003a**, 130, 387-395.

Belaj, A., Satovic, Z., Cipriani, G., Baldoni, L., Testolin, R., Rallo, L., Trujillo, I. *Theor Appl Genet.* **2003b,** *107*,736-744.

Belaj, A., Satovic, Z., Rallo, L., Trujillo, I. *Theor Appl Genet*. **2002**, *105*, 638-644.

Benincasa, C., De Nino, A., Lombardo, N., Perri, E., Sindona, G., Tagarelli, A. *J. Agric Food Chem*. **2003**, *51*, 733-741.

Berndt, P., Hobohm, U., Langen, H. *Electrophoresis*. **1999**, *20*, 3521–3526.

Besnard, G., Bervillé, A. *Life Sci*. **2000,** *323*,173-181.

Besnard, G., Bervillé, A. *Theor Appl Genet*. **2002,** *104*, 1157-1163.

Besnard, G., Khadari, B., Baradat, P., Bervillé, A. *Theor Appl Genet.* **2002,** *104*,1353-1361.

Besnard, G., Baradat, P. Bervillé, A. *Theor. Appl. Genet.* **2001,** *102*, 251-258.

Beyer, K., Bardina, L., Grihina, G., Sampson, H.A. *J Allergy Clin Immunol*. **2001,** *110*,154-159.

Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Görg, A., Westermeier, R., Postel W. *J. Biochem Bio-phy Methods*. **1982,** *6*, 317-339.

Blouin, M.S., Parsons, M., Lacaille, V., Lotz, S. *Molecular Ecology*. **1996**, *5*, 393 - 401.

Bradford, M.M., A .*Analyt. Biochem*. **1976**,*72*, 248-54.

Brenes Balbuena, M., Garcia Garcia, P., Garrido Fernandez, A. *J. Agric. Food Chem*. **1992,** *40*, 1192–1196.

Briante, R., Patumi, M., Limongelli, S., Febbraio, F., Vaccaro, C., Di Salle, A., La Cara, F., Nucci, R. *Plant Science*. **2002,** (8)*162*, 791-798.

Brookfield JFY. *Molecular Ecology*. **1996**, *5*, 453–455.

Bruckner, G. *Ed. M. Dekker New York*. **2000,** 843-846.

Burns, M., Shanahan, D., Valdivia, H., Harris, N. *Eur. Food Res. Technol*. **2003,** *216,* 428-433.

Cantini, C., Gucci, R. Sillari, B. *Hort. Technology*. **1998**, *8*, 409–412.

Carnés, J., Fernández-Caldas, E., Boluda, L., Casanovas, M., Sastre, J., Lluch Bernal, M., Blanca, M. *Allergy*. **2002**, *57*,9, 798-804.

Carriero, F., Fontanazza, G., Cellini, F., Giorio, G. *Theor Appl Genet*. **2002,***104,*301-307.

Chambers, G.K., and MacAvoy, E.S. *Comp. Biochem. Physiol.B*. **2000**, *126*, 455-476.

Cicoria, M., Corbo, M., D'Uva, T. Ruggiero, A. Ed. *E.R.S.A. Molise*. **2000.**

Cipriani, G., Marrazzo, M.T., Marconi, R., Cimato, A., Testolin, R. *Theor Appl Genet. 2002, 104,* 223-228.

Cipriani, G., Marrazzo, M.T., Marconi, R., Cimato, A., Testolin, R. *Theor Appl Genet.* **2002**,*104*, 223-228.

Clauser, K.R., Baker, P., Burlingame, A.L. *Anal. Chem*, **1999**, *71*, 2871-2882.

Conde, J., Conde Hernàndez, H.P., Gònzales Quevedo Tejrina, M.T., Conde Alcaniz, M.A., Conde Alcaniz, E.M., Crespo Moreira, P., Cabanillas Platero, M. *Allergy*. **2002** *57* (Suppl. 71), 60-65.

Damerval, C., de Vienne, D., Zivy, M., Thiellement, H. *Electrophoresis*. **1986,** *7,* 52–54.

David, A., Colantonio, D.l., Chan, W. *Clinica Chimica Acta*. **2005**, *357*: 151-158.

De La Rosa, R., James, C.M., Tobutt, K.R. *Mol Ecol Not*. **2002**, *2*, 265-267.

Deka, R., L. Jin, M.D., Shriver, L.M., Yu, S., DeCroo, J., Hundrieser, C.H., Bunker, R.E., Ferrel., and Chakraborty, R. Am. *J. Hum. Genet*. **1995**, *56*, 461-474.

Di Donna, L., De Nino, A., Mazzotti, F., Perri, E., Raffaelli, A., Sindona, G. *Cambridge: Royal society of chemistry.* **2001,** 131-132.

Dice, L.R. *Ecology.* **1945**, *26,* 297-302.

Eriksson, J., Chait, B.T., and Fenyo, D. *Anal., Chem*. **2000**, *72*, 999–1005.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. *Science*. **1989**, *246*, 64–71.

Ferrara, E. and Lamparelli, F. *Atti del Convegno L'Olivicoltura. Istituto Sperimentale per l'Olivicoltura, Cosenza, Italy*. **1995**, 133-141.

Fisher, M.P., Dingman, C.W. *Biochemistry*. **1971**, *10,* 895.

Fodale, A.S., Mulé, R., Muzzalupo, I., Pellegrino, M., Perri, E. *Atti del VII Convegno Nazionale sulla Biodiversità, Catania, Italy.* **2005,** (in press).

Garcia, J. L., Avidan, N., Troncoso, A., Sarmento, R., Lavee, S. *Sci. Hort*. **2000**, *85*, 271–284.

Georgalaki, M.D., Sotiroudis, T.G., Xenakis, A. *Journal of the American Oil Chemists' Society*. **1998**, *75*, 155–159.

Gitte, I., Frandsena, J. M., Jason, T.C.T. *Physiologia Plantarum*. **2001**, *112*, 301–307.

Goupy, P., Fleuriet, A., Amiot, M.J., Macheix, J.J. *J. Agric. Food Chem*. **1991**, *39*, 92–95.

Griffin, P.R., MacCoss, MJ., Eng, J.K., Blevins, R.A., Aaronson, J.S., Yates, J,R. *Rapid Commun Mass Spectrom*. **1995**, *9(15)*,1546-51.

Griffiths, A.J.F., Miller, J.F., Suzuki, D.T., Lewontin, R.C., Gelbart, W.M. *Introduction to Genetic Analysis, 5th Edition. W.H. Freeman, New York*. 1996.

Gucci, R. and M. Tattini. *Hort. Rev*. **1997**, *21*,177–213.

Gygi, S. P.and Ruedi, A. Proteomics: A Trends Guide. 2004, 31-36.

Hamada, H., Petrino, M.G., Kakunaga, T. *Proc. Natl. Acad. Sci. USA*. **1982**,*79*, 6465-6469.

Hartl, D.L. and Clark, A.G. *Principles of population genetics, 2nd Ed. Sinauer Associates Inc., Sunderland, MA*. **1989**.

Hatanaka, A. *Phytochemistry*. **1993**, *34* 1201- 1218.

Heazlewood, J.L. and Millar, A.H. *Funct. Plant Biol*. **2003**, *30*, 471–482.

Helling, R.B., Goodman, H.M., Boyer, H.W. *J. Virol*. **1974**, 14,1235.

Herbert, B. *Eletrophoresis*. **1999**, *20* (4-5),660-3.

Hesse, J., Kadereit, J.W., Vargas, P. *Mol Ecol*. **2000**, *9*, 857-868.

Heukeshoven, J. and Dernick, R. *Electrophoresis*. **1985**, *6*, 103–112.

Hidalgo, F. J., Alaiz, M., Zamora, R. *Anal. Chem*. **2001**, *73*, 698-702.

Hidalgo, F. J., Alaiz, M., Zamora, R. *Journal of the American Oil Chemists' Society*. **2002**, *79*, 685–689.

Hillenkamp, F. M., Karas, R.C., Beavis and Chait, B. *Anal. Chem*. **1991**, *63*, 1193-1203.

Hilton, J.M., Plomann, M., Ritter, B., Modregger, J., Freeman, H.N., Falck, J.R., Krishna, U. M. and Tobin, A. B. *J. Biol Chem*. **2001**, *276*, 16341-7.

Hubbs, A. and Roy, H. *Plant Physiol*. **1992**, *100*, 272-. 281.

Hurkman, W.J, Tanaka, C.K. *Plant Physiol*. **1986**, *81*, 802-806.

Jarne, P. & Lagoda, P.J.L. *Trends in Ecology and Evolution*. **1996**, *1*,424 – 429.

Jensen, O.N., Podtelejnikov, A.V. and Mann, M. *Anal Chem*. **1997**, *69,* 4741-4750.

Jensen, O.N., Podtelejnikov, A.V. and Mann, M. *Rapid Commun. Mass Spectrom*. **1996**, *10*, 1371 –1378.

Johnson, R.J., Biemann, K. Biomed. *Environ. Mass Spectrom*. **1989**, *18*, 945-957.

Jonne, H.E.R., Juan, S., Francisco, M., Denis, J.M. *Plant Science*. **2001**, *93*; 203-210.

Kamo, M., Kawakami, T., Miyatake, N. and Tsugita, A. *Electrophoresis*. **1995**, *16,* 423–300.

Khadari, B., Breton, C., Moutier, N., Roger, J.P., Besnard, G., Bervillé, A., Dosba, F. *Theor Appl Genet, Rome*. **2003**, *106,* 521-529.

Kimura, M. and J.F. Crow. *Genetics*. **1964**, *49*, 725-38.

Koidis, A., Boskou, D. *Eur. J. Lipid Sci. Technol*. **2006**, *108,* 323-328.

Konno, K., Hirayama, C., Yasui, H. and Nakamura, M. *Proc. Natl. Acad. Sci. Unit. States Am*. **1996**, *96*; 9159–9164.

La Mantia, M., Lain, O., Caruso, T. and Testolin, R. J. *Hort. Sci. Biot*. **2005**, *80*, 628-632.

La Rosa, R., Angiolillo, A., Guerrero, C., Pellegrini, M., Rallo, L., Besnard, G., Bervillé, A., Martin, A., Baldoni, L. *Theor Appl Genet*. **2003**,*106,* 1273-1282.

Laemmli, U. K. *Nature*. **1970**, *227*, 680-685.

Lanteri, S., Armanino, C., Perri, E. and Palopoli, A. *Food Chemistry*. **2002**, *76*, 501-507.

Ledesma, A., Villalba, M., Batanero, E., Rodrý`guez, R. *Eur. J. Biochem*. **1998**, *258,* 454-459.

Lefort, F., Brachet, S., Frascaria-Lacoste, N., Edwards, K.J. and Douglas, G.C. *Mol. Ecol*. **1999**, *8,* 1075-1092.

Levinson, G. and Gutman, G.A. *Mol. Biol. Evol*. **1987**, *4*, 203-221.

Lim, J., Bustin, M., Ogryzko, V.V. and Postnikov, Y.V. J. *Biol Chem*. **2002**, *277*, 20774-82.

Lombardero, M., Barbas, J.A., Moscoso del Prado, J. and Carreira, J. *Clin. Exp. Allergy*. **1994**, *24*, 765-770.

Lombardero, M., Obispo, T., Calabozo, B., Lezau` an, A., Polo, F., Barber, D. *Allergy*. **2002**, *57 (Suppl. 71)*, 29-34.

Lombardo, N., Godino, G., Alessandrino, M., Belfiore, T., Muzzalupo, I., Perri, E. *Ed. I.S.Ol., Cosenza*. **2004**.

Lombardo, N., Perri, E., Muzzalupo, I., Madeo, A., Godino, G., Pellegrino, M. *Ed. Co.R.Ass.Ol., Regione Calabria*. **2003**

Lombardo, N., Muzzalupo, I., Perri, E. and Sajiad, A. *XII Simposium Científico-Técnico Expoliva, Jaén, Spain. Available:http://www.expoliva.com/expoliva.***2005**.

Lopez, M.F. *Electrophoresis*. **2000**, *21*, 1082–1093.

Lumaret, R., Ouazzani, N., Michaud, H., Vivier, G., Deguilloux, M.F., Di Giusto, F. *Heredity*. **2004,** *92*:343-351.

Macrì, T., Picone, G. and La Porta, G. *Terra e Sole*. **1997**, *3*, 146-149.

Mantel, N. *Cancer Res*. **1967**, *27,* 209-220.

Marta, R., Stefana, T., Maurizio, P., Giuseppe, F. *J. Agric. Food Chem*. **2002**, *50*, 835-839.

Martinez, A., Asturias, J.A., Monteseirin, J., Moreno, V., Garcia-Cubillana, A., Hernandez, M., de la Calle, A., Sanchez-Hernandez, C., Perez-Formoso, J. L., Conde. *Allergy*. **2002**, *57,s71*, 17-23.

Maxam, A.M. and Gilbert, W. *Proc Natl Acad Sci USA*. **1977**, *74* 560-564.

Mazzei, R., Giorno, L., Mazzuca, S., Spadafora, A., Drioli, E. *Desalination*. **2006**, *200,* 483–484.

Mazzuca, S., Spadafora, A., Innocenti, A.M. *Plant Science*. **2006**, *171*; 726–733.

Meyer, Y., Grosser, J., Chartier, Y., Cleyet-Marel, J.V.C. *Electrophoresis*. **1988**, *9*, 704–712.

Montedoro, G.F., Garofalo, L., Bertuccioli, M. and Pannelli, G. *Proc. of the 6th Int. Flavor Conference. Develop Food Sci*. *Elsevier, Amsterdam*. **1990**, *24*, 881–891.

Muzzalupo, I. and Perri, E. *European Foof Res. Technol*. **2002**, *214*, 528-531.

Nei, M. *Amer. Nat*. **1972**, *106,* 283-292.

Nei, M. *Proc. Natl. Acad. Sci. USA*. **1973**, *70*, 3321-3323.

O'Farrell, P.H. *The Journal of Biological Chemistry*. **1975**, *250*, 4007–4021.

Patton, W.F. *Electrophoresis*. **2000**, *21*, 1123–1144.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. *Electrophoresis*. **1999**, *20*, 3551-67.

Perri, E., Parlati, M.V., Palopoli, A., Rizzati, B. *Italus Hortus*. **1995**, *2,* 31-36.

Perri E., Raffaelli, A., Sindona, G. *Journal of Agricultural and Food Chemistry*. **1999**, *47*, 4156 4160.

Pietrangeli, F., Russo, A. *Ed. Regione Abruzzo*. **1997**.

Pokorny, J., Yanishlieva, N., Gordon, M. *Cambridge, UK: Woodhead Publishing*. 2001.

Polydefkis, H., Georgios, B., Katerina, G., Fotis, G., Nikos, N., Dimitra, M., Kosmas, H. *Eur. J. Lipid Sci. Technol*. **2002**, *104,* 574–586.

Porubleva, L., Vander Velden, K., Kothari, S., Livier, D.J. and Chitnis, P.R. *Electrophoresis*. **2001,** *22*: 1724–1738.

Provan, J., Shepherd, L., Powell, W., Waugh, R. *Plant Cell Rep*. **1996**, *16,* 196-199.

Queller, D.C., Strassman, ‚J.E. & Hughes, C.R. *Trends in Ecology and Evolution*. **1993**, *8*, 285 – 288.

Rabilloud, T. *Electrophoresis*. **1996**, *17*, 813–829.

Rajnarayanan, R.V., Wang, K. *J. Mass Spectrom*. **2004**, *39*, 79-85.

Rallo, P., Dorado, G., Martin, A. *Theor Appl Genet*. **2000**, *101*, 984-989

Reynolds, J.A., Tanford, C. *Proc. Natl. Acad. Sci. USA*. **1970**, *66*, 1002-1007.

Ridolfi, M., Terenzani, S., Patumi, M., Fontanazza, G. *J. Agric. Food Chem*. **2002**, *50*, 835-839.

Roepstorff, P., Fohlman, J. *Biomed Mass Spectrom*. **1984**, (*11*),601.

Rohlf, F.J. *NTSYS-pc. Exeter Software, Setauket, N.Y*. **1994**.

Rose, J.K.C., Bashir, S., Giovannoni, J.J., Jahn, M.M. and Saravanan, R.S. *Plant J.* **2004**, *39,* 715-733.

Ross, J.H.E., Sanchez, J., Millan, F., Murphy, D.J. *Plant Sci.* **1993**, *93,* 247–280.

Sanger, F., Nicklen, S., Coulson, A.R. *Proc Natl Acad Sci U S A*. **1977**, *74(12)*, 5463–5467.

Saravanan, R.S., Rose, J.K.C. *Proteomics*. **2004**, *4*, 2522-2532.

Schägger, H., von Jagow, G. *Anal. Biochem*. 1987, *166*, 368-379.

Schlötterer, C. *Chromosoma*. **2000**, *109*, 365–371.

Schlötterer, C. and Tautz, D. *Nucleic Acids Res*. **1992**, *20,* 211-215.

Schwender, J. *17th International Symposium on Plant Lipids. Michigan State University, East Lansing, Michigan July 16 − 21*, **2006**.

Sefc, K.M., Lopes, M.S., Mendonca, D., Rodrigues dos Santos, M., Laimer da Camara, M.A. *Mol Ecol*. **2000**, *9*,1171-1173.

Selkoe, K.A. & Toonen, R.J. *Ecology Letters*. **2006**, *9,* 615-629.

Sharp, P. A., Sugden, B., Sambrook, J. *Biochemistry*. **1973**, *12*, 3055.

Simonsen, N.R., Fernandez-Crehuet, N.J., Martin-Moreno, J.M. *Am J Clin Nutr*. **1998**, *68*,134-141.

Stasyk T., Huber L.A. *Proteomics*. **2004**, *12(4)*, 3704-16.

Strand, M., Prolla, T.A., Liskay, R.M. and Petes, T.D. *Nature*. **1993**, *365,* 274-276.

Straub, R.E., Speer, M.C., Luo, Y., Rojas, K., Overhauser, J. Ott, J. and Gilliam, T.C. *Genomics*. **1993**, *15*, 48-56.

Tani, M., Okino, N., Mitsutake, S., Tanigawa, T., Izu, H. and Makoto, I. *J. Biol Chem*. **2000**, *275,* 3462-8.

Tautz, D. *Exs*. **1993**, *67*, 21-28.

Tejera, M.L. *J. Allergy Clin. Immunol*. **1993**, *104*, 797-802.

Thomas, M.R, Scott, N.S. *Theor Appl Genet*. **1993**, *86*, 985-990

Thomas, M., Davis, R.W. *J. Mol. Biol*. **1975**, *91*,315.

Touzet, P., de Vienne, D., Huet, J.C., Ouali, C., Bouet, F. and Zivy, M. *Electrophoresis*. **1996**, *17* 1393–1401.

Trevino, M. A., Garcia-Mayoral, M.F., Barral, P., Villalba, M., Santoro, J., Rico, M., Rodrýguez, R., Bruix, M. *J. Biol. Chem*. **2004**, *279*, 39035-39041.

Tsugita, A., Kawakami, T., Uchiyama, Y., Kamo, M., Miyatake, N. and Nozu, Y. *Electrophoresis*. **1994**, *15*, 708–720.

Tyers, M. and Mann, M. *Nature*. **2003**, *422*, 193-7.

Villalba, M., Batanero, E., Monsalve, R.I., Gonzalez de la Pena, M.A., Lahoz, C. and Rodriguez, R. *J. Biol. Chem*. **1994**, *269*, 15217-15222.

Villalba, M., Batanero, E., Lc¸pez-Otý`n, C., Sa`nchez, L.M., Monsalve, R.I., Gonza`lez de la Pena, M.A., Lahoz, C., Rodrý`guez, R. *Eur J Biochem*. **1993**, *216*, 863-869.

Villalba, M., Lope´z-Otý`, C., Martin-Orozco, E., Monsalve, R. I., Palomino, P., Lahoz, C., Rodrihuez, R. Biochem. *Biophys. Res. Commun*. **1990**, *172*, 523-528.

Vrtala, S., Grote, M., Duchene, M. Int. *Arch. Allergy Immunol*. **1995**, *107*, 160-169.

Waisel Y., Geller-Berstein C., Keynan N., Arad G. *Allergy*. **1996**, *51*, 819-825.

Wang, W., Scali, M., Vignani, R., Spadafora, A., Sensi, E., Mazzuca, S., Cresti, M. *Electrophoresis*. **2003**, *24*, 2369-2375.

Weber, K., Osborn, M. *J Biol Chem*. **1969**, *244*, 4406-4412.

Wilkins, M.R., Pasquali, C., Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, Hochstrasser DF. *Biotechnology*. **1996**, *14*, 61–65.

Wilm, M., Mann, M. *Anal Chem*. **1996**, *1;68*, 1–8.

Wren, J.D., Forgacs, E., Fondon, J.W. 3rd, Pertsemlidis, A., Cheng, S.Y., Gallardo, T., Williams, R.S., Shohet, R.V., Minna, J.D. and Garner, H.R. *Am. J. Hum. Genet*. **2000**, *67* 345-356.

Wu, S.B, Collins, G., Sedgley, M. *Genome*. **2004**, *47*, 26-35

Yagami,T., Haishima, Y., Tsuchiya, T., Tomitaka-Yagami, A., Kano, H. and Matsunaga, K. *Int. Arch. Allergy Immunol*. **2004**, *135*, 3-11.

Yamashita, M., Fenn, J.B. *J. Phys. Chem*. **1984**, *88*, 4451-4459.

Yang, G.P., Saghai Maroof, M.A., Xu, C.G., Zhang, Q., Biyashev, R.M. *Mol Gen Genet*. **1994**, *245*, 187-194.

Yeh, F.C., Yang, R.C., Boyle, T.B.J., Ye, Z.H. and Mao, J.X. *POPGENE: Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Alberta, Canada*. **1997**.

Yu, Y.G., Saghai Maroof, M.A., Buss, G.R., Tolin, S.A. *Phytopathol*. **1994**, *84*, 60-64.

Yu, C.J., Lin, Y.F., Chiang, B.L. and Chow, L.P. *J. Immunol*. **2003**,*170*, 445-453.

Zivy, M., Thiellement, H., de Vienne, D. and Hofmann, J.P. *Theor. Appl. Genetics*. **1984**, *68*, 335–345.

ANNEX-I

Table 1. List of the olive accessions from Abruzzo & Molise regions Olive Germplasm Collection of the C.R.A.-ISOl (Italy).

Code	Cultivar	Local synonyms*	Cultivation area
B 184 I	Aurina	Liciniana, Aurea	Isernia (Molise region)
A 174 III	Bottoni di Gallo	-	Isernia (Molise region)
B 260 II	Caprina di Casalanguida	Caprigna	Chieti (Abruzzo region)
B 261 I	Caprina Vastese	Caprigna	Chieti (Abruzzo region)
B 216 I	Carbonchio	Carbogna,Carbonchia, Imbrattasacchi, etc	Chieti, Pescara, Teramo (Abruzzo region)
B 204 I	Carpinetana	Femmina, Posolella, etc	Pescara, Chieti (Abruzzo region)
A1 I	Castiglionese	Dritta, San Francesco	Pescara, Teramo (Abruzzo region)
C_2 369 I	Cazzarella	Tortarella	Campobasso (Molise region)
B 185 I	Cellina di Rotello	-	Campobasso (Molise region)
A4 I	Dritta	Dritta di Loreto, Moscufese, etc	Chieti, Pescara, Teramo (Abruzzo region)
B217 II	Gentile de L'Aquila	Monachella	L'Aquila (Abruzzo region)
A5 I	Gentile di Chieti	Gentile, Ordinaria, etc	Chieti, Pescara, Teramo(Abruzzo region)
A 166 I	Gentile di Larino	Olivacchione	Campobasso (Molise region)
A2 I	Ghiandaro	Crognalegna/o, Circelluta, etc	L'Aquila, Pescara, Chieti (Abruzzo region)
B 187 II	Gnagnaro	Gnagnaro lungo	Isernia (Molise region)
B 188 I	Grossa di Venafro	-	Isernia (Molise region)
A6 I	Intosso	Indosso, Oliva Pana, etc	Chieti, L'Aquila, Pescara, Teramo (Abruzzo region)
A102 I	Nebbio di Chieti	Comune, Nebbia	Chieti (Abruzzo region)
B176 I	Nebbio di Pescara	Neja	Pescara (Abruzzo region)
C_2 371 I	Noccioluta	Rosciola di Rotello	Campobasso (Molise region)
B 190 I	Oliva nera di Colletorto	Gentile di Colletorto, Noccioluta	Campobasso (Molise region)
C_2 372 I	Olivastro d'Aprile	Olivastro breve	Isernia (Molise region)
B218 I	Olivastro di Bucchianico	-	Chieti (Abruzzo region)
B219 I	Olivastro Frentano	-	Chieti (Abruzzo region)

C_2 374 II	Olivastro di Montenero	Olivastra di Montenero, Livastrello	Campobasso (Molise region)
C_2 373 II	Olivastro dritto	Olivastro, Olivastrello	Isernia (Molise region)
C_2 375 II	Olivetta near	Olivetta	Isernia (Molise region)
A3 I	Olivoce	Cucco, Chietina, Olivone, etc	Chieti, L'Aquila, Pescara, Teramo (Abruzzo region)
B 233 I	Paesana bianca	Oliva bianca	Isernia (Molise region)
B 234 III	Paesana near	Ceciarola, Oliva nera	Isernia (Molise region)
A7 I	Peperella	Pescarese, Paparella, etc	Pescara, Teramo (Abruzzo region)
B220 II	Posola	Pennese, Pizzutella	Pescara, Chieti (Abruzzo region)
B221 I	Posolella	Precoce, Carpinetana	Pescara (Abruzzo region)
A146 I	Precoce	Gentile	Pescara (Abruzzo region)
B177 I	Puntella	-	Pescara (Abruzzo region)
A 167 I	Romanella molisana	-	Campobasso (Molise region)
B 186 I	Rosciola di Rotello	Noccioluta, Rossuola	Campobasso (Molise region)
B 189 I	Rossuola	Resciola or Risciola di Venafro	Isernia (Molise region)
B 191 I	Rumignana	Remugnana,Ramignana,Rimignanna	Campobasso (Molise region)
B222 I	Rustica	-	L'Aquila (Abruzzo region)
B 235 II	Salegna di Larino	Saligna	Campobasso (Molise region)
A 175 I	Sperone di Gallo	Zampa di gallo, Oliva torta	Campobasso (Molise region)
A8 I	Toccolana	Police	Pescara (Abruzzo region)
A9 I	Tortiglione	-	Pescara, Teramo (Abruzzo region)

*According to: Cicoria et al., 2000; Bartolini et al., 1998; Macrì et al., 1997; Pietrangeli et al., 1997.

Table 4. Genotypic profiles obtained from the combination of 9 SSR loci.

	N*	Accessions	GAPU 59	GAPU 71A	GAPU 71B	GAPU 103A	UDO 01	UDO 03	UDO 12	UDO 28	UDO 39
Abruzzo Region	1	Caprina di Casalanguida	AB	BB	AA	BB	BB	CD	AC	AC	CC
	2	Caprina Vastese	AB	BB	AD	BB	BB	CE	BC	AC	CC
	3	Carbonchio	BD	BB	AA	BD	AA	CC	BC	AC	CC
	4	Carpinetana	BE	BB	BE	EE	BB	CC	BC	AD	BB
	5	Castiglionese	AB	AB	BC	CE	BB	CC	AC	AD	BE
	6	Dritta	AB	BB	AD	CD	AA	BB	AC	BB	DD
	7	Gentile de L'Aquila	BB	BB	EE	BE	AA	DD	AC	BC	CC
	8	Gentile di Chieti	AB	AB	AD	DD	AA	BB	AC	AC	CC
	9	Ghiandaro	AB	BB	AE	CD	BB	CC	BC	AC	BB
	10	Intosso	BE	AB	AD	CE	AA	CC	AA	AC	BB
	11	Nebbio di Chieti	BE	BB	AD	CC	BB	AA	BC	AC	CC
	12	Nebbio di Pescara	AA	AB	AD	DD	AA	CC	AC	AD	CE
	13	Olivastro di Bucchianico	AE	AB	AA	CD	AA	DD	BC	AD	CC
	14	Olivastro Frentano	AE	AB	CE	DD	AA	DD	BB	AD	BE
	15	Olivoce	BD	BB	AB	AC	AA	DD	AC	AC	CE
	16	Peperella	AB	BC	AC	CD	BB	BB	BC	AC	DF
	17	Posola	BD	BB	AE	DE	BB	AC	AC	AC	CC
	18	Posolella	CD	BB	CE	CC	BB	CC	BC	AC	CC
	19	Precoce	AB	BB	BC	CD	AA	CC	AC	BB	BE
	20	Puntella	EE	BB	AD	CC	BB	EE	AC	AC	CC
	21	Rustica	BE	AB	AB	BE	AA	AA	AA	AC	BB
	22	Toccolana	BE	BB	AD	BB	BB	CD	BC	AC	BB
	23	Tortiglione	AB	BC	AB	CD	BB	CC	BB	AD	DF
Molise Region	24	Aurina	EE	AB	AE	CD	BB	CC	BB	AC	CC
	25	Bottoni di Gallo	BB	AB	AE	DD	BB	CC	BC	AC	CC
	26	Cazzarella	BB	BB	AA	DD	AA	BB	AC	BB	CC
	27	Cellina di Rotello	EE	AB	AB	BE	AA	EE	BC	AD	AC
	28	Gentile di Larino	AB	BB	AD	BE	AA	CC	AC	BB	CC

	29	Gnagnaro	BC	AB	AE	DD	BB	CD	AA	AB	DF
	30	Grossa di Venafro	BE	BB	AD	BB	BB	CC	BC	AC	CC
	31	noccoluta	AB	BB	AB	AD	AA	DD	BC	AC	CC
	32	Oliva nera di Colletorto	AB	BB	AB	AD	AA	AA	BC	AC	CC
	33	Olivastro d'Aprile	EE	AB	AB	CC	BB	AD	AA	CC	CC
	34	Olivastro di Montenero	DE	AB	AA	DD	BB	AC	BC	DD	AC
	35	Olivastro dritto	BB	AB	AE	CC	BB	CC	BC	AD	BB
	36	Olivetta nera	BE	BB	AD	BB	BB	CC	BC	AC	BB
	37	Paesana bianca	AB	BC	AE	CE	BB	CC	BB	BD	BB
	38	Paesana nera	BE	BB	AD	BB	BB	CD	BC	AC	CC
	39	Romanella molisana	BE	BB	DB	BB	AA	BB	BC	AC	CF
	40	Rosciola di Rotello	AB	BC	AE	CE	BB	CC	BB	BB	BB
	41	Rossola	EE	BB	AA	BE	BB	EE	BC	AC	FF
	42	Rumignana	AB	BB	AC	BD	AA	CC	AC	BB	DD
	43	Salegna di Larino	BE	BB	AD	BB	AA	CC	BC	AC	DF
	44	Sperone di Gallo	BB	BB	DD	DD	AA	BB	AC	BB	CC

∗ **The numbers used here coincides with the numbers used to illustrate the dendrogram of the olive accessions (Fig. 24,25 and 26).**

Table 6. List of the 39 olive accessions analyzed from the olive germplasm collection of the Apulia region, including their local synonyms and the main production areas.

Accessions	Local synonyms[z]	Cultivation area
'Bella di Cerignola'[z]	'Grossa di Spagna', 'Oliva a Prugna', etc	Bari, Foggia, Lecce, Taranto
'Bella di Spagna'[z]	'Belle d'Espagne', 'Belle of Spain'	Foggia
'Butirra Melpignano'[z]	-	Lecce
'Carmelitana'[z]	-	Bari, Foggia
'Cazzinicchio'[z]	-	Bari
'Cellina di Nardò'[z]	'Cafarella', 'Cellina Inchiastra', etc	Bari, Brindisi, Lecce, Taranto
'Cerasella'[z]	'Bicarrese', 'Riccia', 'Rotondella'	Bari, Foggia, Taranto
'Cima di Mola'[z]	'Cima di Monopoli', 'Molese', etc	Bari, Brindisi
'Coratina'[z]	'Cima di Corato', 'Racemo', etc	Apulia Region
'Dolce di Cassano'[y]	-	Bari
'Dolce d'Andria'[z]	'Dolce di Sannicandro', 'Dolce di Puglia'	Bari, Foggia, Lecce
'Dolce Mele'[z]	'Mele'	Bari
'Frangivento'[z]	'Cipressino', 'Olivo di Pietrafitta', etc	Lecce, Taranto
'Lezze'[y]	'Oliastro', 'Oliastro di Conversano'	Bari, Brindisi, Taranto
'Marinese'[z]	'Cima di Bitonto', 'Ogliarola Barese', 'Ogliarola Garganica', etc	Foggia
'Mele'[z]	'Amele', 'Melo Dolce', etc	Bari, Brindisi, Taranto
'Mora'[w]	-	Bari
'Morellona di Grecia'[z]	'Nera di Grecia'	Lecce
'Nociara'[z]	-	Bari, Brindisi, Taranto
'Nolca'[z]	'Annolca', 'Dolce di Barbarano', etc	Apulia region
'Nzimbibolo'[z]	-	Foggia
'Ogliarola Barese'[z]	'Cima di Bitonto'	Bari, Brindisi, Foggia, Lecce

'Ogliarola Garganica'[z]	'Cima di Bitonto'	Foggia
'Ogliarola di Lecce'[z]	'Ogliarola Salentina', 'Pizzuta', 'Chiarita', etc	Apulia Region
'Oliastro'[y]	'Olivastro di Conversano', 'Lezze'	Bari, Brindisi, Taranto
'Oliva Dolce di Barbarano'[z]	'Nolca'	Lecce
'Oliva Rossa'[x]	-	Bari
'Pasola'[z]	'Fasola', 'Passula', 'Calabrese', 'Frasola', etc	Apulia region
'Pasola d'Andria'[y]	-	Bari
'Peppino Leo'[z]	-	Bari
'Peranzana'[z]	'Provenzale', 'Francese', 'Tondina', etc	Bari, Foggia, Lecce, Taranto
'Pizzuta'[z]	'Ogliarola di Lecce'	Bari, Brindisi, Lecce, Taranto
'Racemo'[z]	'Coratina'	Bari
'Rotondella'[z]	-	Foggia
'San Benedetto'[z]	-	Lecce
'Sant'Agostino'[z]	'Grossa Andriesana', 'Oliva Grossa', etc	Apulia Region
'Simona'[z]	-	Bari
'Termite di Bitetto'[z]	'Cima di Bitetto', 'Mele di Bitetto', etc	Bari, Brindisi, Taranto
'Toscanina'[z]	'Oliva a Grappa'	Brindisi

Apulian germplasm according to: [z]Bartolini et al., 1998, [y]Lombardo et al., 2004, [x]State of Italy, 1994, [w]Ferrara and Lamparelli, 1995

Table 9. Molecular characterization of 39 olive accessions by sequencing of the amplicons from five SSR loci (GAPU45, GAPU59, GAPU71A, UDO01 and UDO39). Unique Genotype (UG) profiles and the exclusive alleles (in bold) obtained from the combination of sequencing SSR loci are reported. Seven different accession groups (genotypes from 22 to 28) are genetically indistinguishable from one another.

Accessions	GAPU45	GAPU59	GAPU71A	UDO01	UDO39	UG[z]
'Butirra di Melpignano'	182-196	222E	214	144	108-**142**	*1*
'Carmelitana'	183A-196	208A-208B	214-224	144	184	*2*
'Cazzinicchio'	183A-183B	208B	210A-214	140	108-184	*3*
'Cellina di Nardó'	196	208A-**222F**	214-**228**	144	108-188	*4*
'Dolce di Cassano'	183B-185	208A-208B	214-224	144	108-175	*5*
'Frangivento'	196	208A	214	144	173-184	*6*
'Mora'	183A-196	208A-208B	210A-224	144	108-170	*7*
'Morellona di Grecia'	183A-183B	208B-222E	210A-214	140	108-184	*8*
'Nociara'	183A	208B-222E	210B-214	140	170-173	*9*
'Oliva Dolce di Barbarano'	183B	222E	214-224	140	146	*10*
'Pasola'	182-196	208A-218	214	144	175	*11*
'Pasola d'Andria'	183A-183B	208A-208B	214-224	144	170	*12*
'Peppino Leo'	183B-182	208A-218	214	144	108	*13*
'Peranzana'	183B-182	208B-222E	212-214	144	108-173	*14*
'Rotondella'	183B-196	208B-222E	214	140	188	*15*
'San Benedetto'	183A-196	208A-208B	214	144	173	*16*
'Sant'Agostino'	183A-196	208B	210B-214	144	108	*17*
'Simona'	182-196	208B-222E	210B	144	108-164	*18*
'Termite di Bitetto'	183B-185	208B-218	214	144	108-170	*19*
'Toscanina'	183A-182	218	210B-214	144	164	*20*
'Nzimbibolo'	182-196	208B-218	214-224	144	108	*21*
'Bella di Spagna'	183A	208B	210B-214	144	108	*22*
'Bella di Cerignola'	183A	208B	210B-214	144	108	*22*
'Cerasella'	183B-196	208A-208B	214	144	108-170	*23*
'Mele'	183B-196	208A-208B	214	144	108-170	*23*
'Nolca'	183B-196	208A-208B	214	144	108-170	*23*
'Cima di Mola'	183B	212	210B-214	144	175	*24*
'Ogliarola di Lecce'	183B	212	210B-214	144	175	*24*

‘Pizzuta’	183B	212	210B-214	144	175	*24*
‘Coratina’	183A-182	208B	210B-224	140	170	*25*
‘Racemo’	183A-182	208B	210B-224	140	170	*25*
‘Dolce d’Andria’	183B-182	208A-218	214	144	108-170	*26*
‘Dolce Mele’	183B-182	208A-218	214	144	108-170	*26*
‘Lezze’	183A-183B	208B-218	210A	140	146-184	*27*
‘Oliastro’	183A-183B	208B-218	210A	140	146-184	*27*
‘Oliva Rossa’	183A-183B	208B-218	210A	140	146-184	*27*
‘Marinese’	183B-182	208A-208B	214-224	144	170	*28*
‘Ogliarola Barese’	183B-182	208A-208B	214-224	144	170	*28*
‘Ogliarola Garganica’	183B-182	208A-208B	214-224	144	170	*28*

[z] The numbers used coincide with the numbers used in the dendrogram of the Apulian germplasm

Olive fruit protein SDS-PAGE.

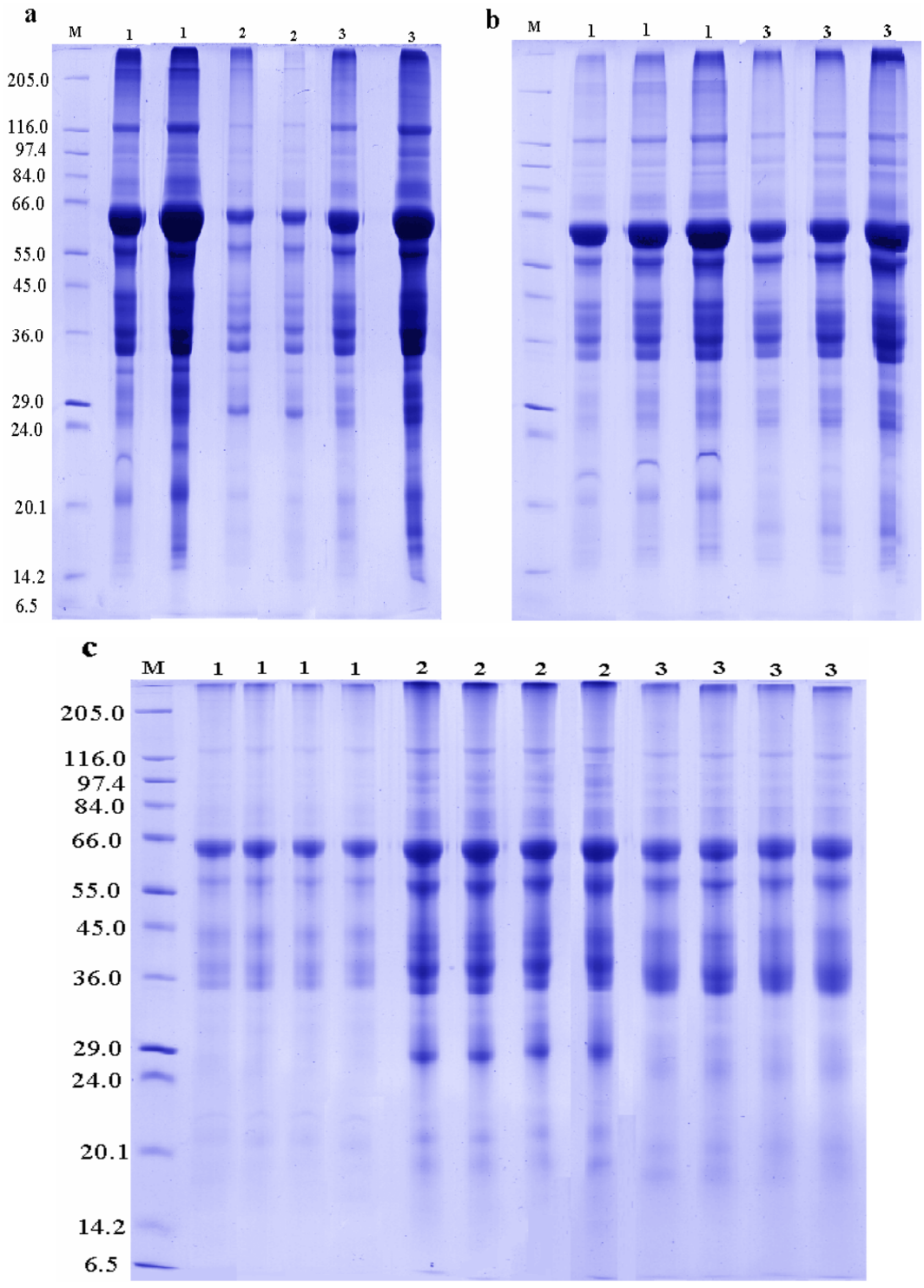


Fig. 37. Electrophoratic pattern of olive fruit proteins, (1). Carolea (2). Nocellara del Belice and (3). Coratina. Proteins were resolved using SDS-PAGE. **a**. 4% separating & 12.5% resolving gel. **b.** 6% separating & 12.5% resolving gel. **c.** 6% separating & 10% resolving gel, and visualized with Coomassie blue. Molecular weights (in kilodaltons) are shown to the left.

Olive oil protein extraction and SDS-PAGE.

A range of methods for protein extraction were applied, *e.g.* (Saravanan and Rose., 2004), which rely on different chemical principles. The diverse methods/solutions used:

- NH_4HCO_3 25 mM in H_2O
- NH_4HCO_3 25 mM in 2:1 (H_2O:MeOH)
- TCA 20% (Aqueous)
- NH_4HCO_3 50 mM in H_2O
- NH_4HCO_3 50 mM in 2:1 (H_2O : MeOH)
- Phenol/SDS buffer (containing 30% sucorose, 2% SDS, Tris HCl 0.1 M pH 8 & 5% β-mercaptoethanole), ratio 1:1.

After extraction of protein SDS-PAGE was performed according to Laemmli, (1970) method. 12.5% polyacrylamide separating gels and visualized with silver stain.

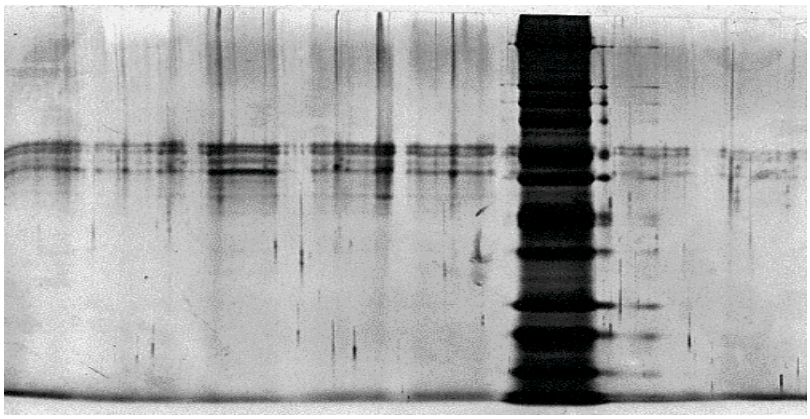


Fig. 39. Electrophoratic pattern of oil (Carolea) 12.5% resolving gel.

Lane-1. Oil 1 part and extraction with 1 part of NH_4HCO_3 25 mM in H_2O

Lane-2. Oil 1 part and extraction with 1 part of NH_4HCO_3 25 mM in H_2O + MeOH (2:1)

Lane-3. Oil 1 part and extraction with 1 part of TCA 20% Aqueous

Lane-4. Oil 1 part and extraction with 1 part of NH_4HCO_3 50 mM in H_2O

Lane-5. Oil 1 part and extraction with 1 part of NH4HCO3 50 mM in 2 H_2O+ 1 MeOH

Lane-6. Marker Sigma (6.5 - 205 kDa)

Lane-7. Oil 1 part and extraction with 1 part of Phenol + SDS buffer (ratio 1:1) and successive precipitation of phenolic phase with NH_4HCO_3 25 mM in H_2O (ratio 1:5).

Lane -8. Oil 1 part and extraction with 1 part of Phenol + SDS buffer (ratio 1:1) and successive precipitation of phenolic phase with NH_4HCO_3 25 mM in H_2O + MeOH (ratio 1:5).

The above Fig. 39 lane 3 shows the intense bands having molecular masses of approximately 55 and 66 kDa were resolved, the protein was extracted with TCA 20%.

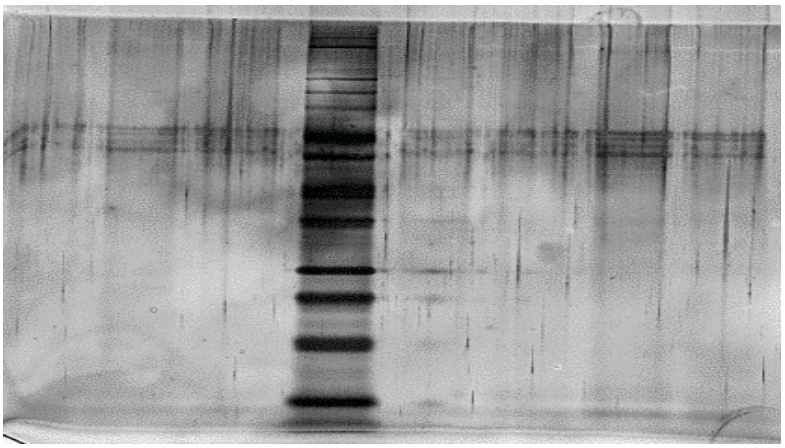


Fig. 40. Electrophoratic pattern and comparison between filtered/unfiltered and with/without centrifugation of fresh oil sample (Carolea), 12.5% resolving gel.

Lane-1: Filtered oil before protein extraction with TCA 20% (ratio oil/TCA 1:2).

Lane-2: Centriguged oil at 6500 rpm for 3 min before protein extraction with TCA 20% (ratio oil/TCA 1:2).

Lane-3: Centriguged oil at 13000 rpm for 1 min before protein extraction with TCA 20% (ratio oil/TCA 1:2).

Lane-4: Marker Sigma (6.5 - 205 kDa)

Lane-5: Extraction with NH_4HCO_3 50 mM (ratio oil/ NH_4HCO_3 1:2).

Lane-6: Extraction with TCA (ratio oil/TCA 1:3)

Lane-7: Phenol + SDS buffer (ratio oil/phe-SDS-buffer 1:1) and successive precipitation of phenolic phase with NH_4HCO_3 50 mM.

Lane-8: Filtered oil and treated with (ratio oil/phe-SDS-buffer 1:1) and successive precipitation of phenolic phase with NH_4HCO_3 50 mM.

The above Fig. 40, shows the same bands having molecular masses of approximately 55 and 66 kDa. It is observed that there is no significant difference in electrophoratic pattern if the sample is filtered/unfiltered and with/without centrifugation.

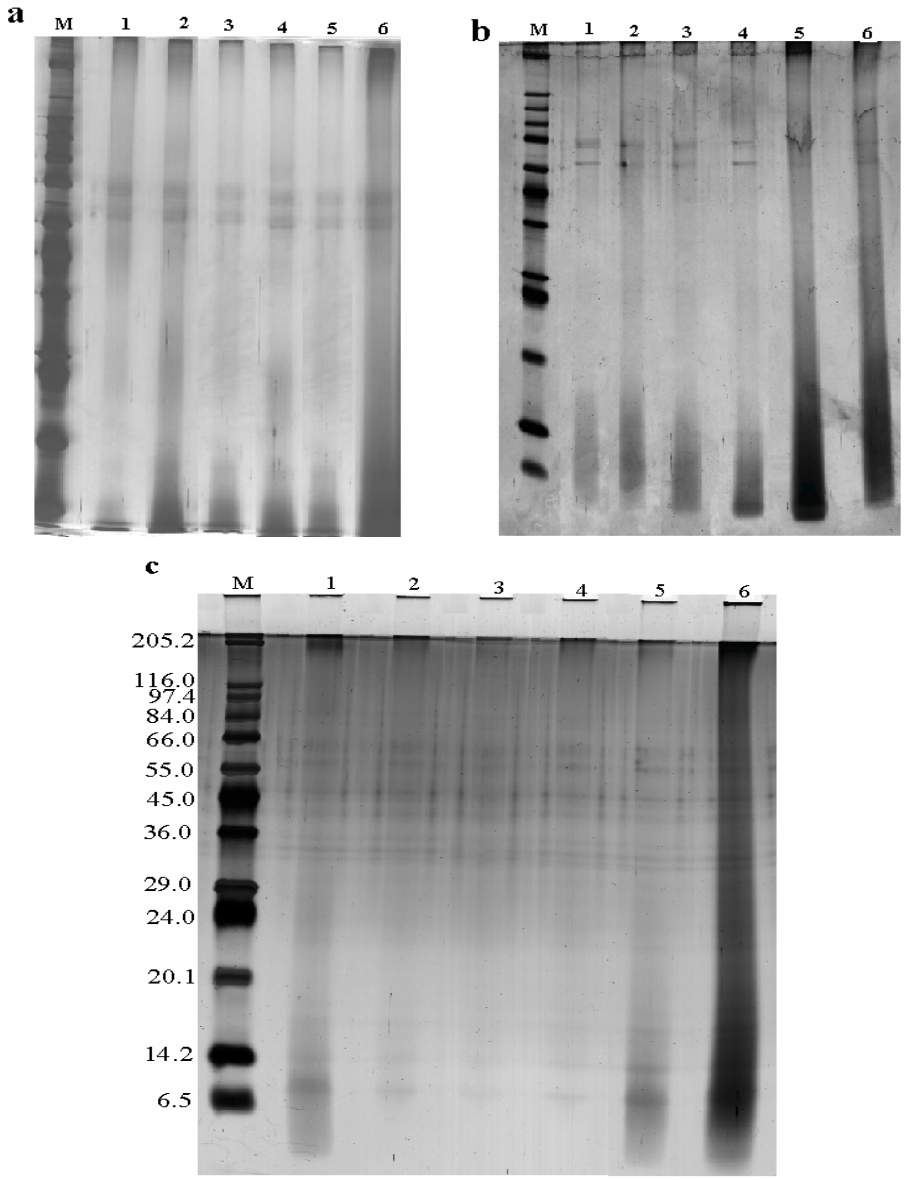


Fig. 41. Electrophoratic pattern of olive oil proteins, (1) Carolea (2) Frantoio (3) Cassanese (4) Coratina (5) Dolce di Rossano and (6) Ogliorola Barese, proteins were resolved using SDS-PAGE. **a**. 4% separating & 10% resolving gel. **b.** 6% separating & 10% resolving gel. **c.** 6% separating & 12.5% resolving gel, and visualized with silver stain. Molecular weights (in kilodaltons) are shown to the left.

APPENDIX

List of Scientific publications:

1. Innocenzo Muzzalupo, Nicola Lombardo, Aldo Musacchio, Maria Elena Noce, Giuseppe Pellegrino, Enzo Perri, and ***Ashif Sajjad***., "DNA Sequence Analysis of Microsatellite Markers Enhances Their Efficiency for Germplasm Management in an Italian Olive Collection".Journal of the American Society for Horticultural Science, 2006, Vol.131, No.3, pp. 352-359.

2. Anna Napoli, Donatella Aiello, Leonardo Di Donna, ***Ashif Sajjad***, Enzo Perri, and Giovanni Sindona., "Profiling of Hydrophilic Proteins from *Olea europaea* Olive Pollen by MALDI TOF Mass Spectrometry". Analytical chemistry, 2006, Vol 78, n, 10, pp 3434 – 3443.

3. Lombardo N., Muzzalupo I., Perri E., ***Sajjad A***. *"Microsatellite markers for Olive (Olea Europea.* L.*) cultivar identification"*. International Olive oil exhibition, 12-13 May 2005 Jaén Spain. pp 1-8. Available at: http://www.expoliva.com/expoliva2005/

4. ***Sajjad A***.,Muzzalupo I.,Perri E., Sindona G. "Characterisation of olive germplasm from Abruzzo, Apulia, Molise, Tuscany and Sicily regions by microsatellite markers". Proceedings of 3rd Workshop on The Contribution of Chemistry, Molecular Biology and Medicinal Chemistry to Life Sciences. 2006. pp. 36-50. ISBN: 88-548-0736-2.

ABSTRACTS/POSTERS:

5. Donatella Aiello, Antonio De Nino, Anna Napoli, Enzo Perri, ***Ashif Sajjad***, Giovanni Sindona " Extraction and characterization of olive (*Olea europaea* L.) fruit proteins from Italian cultivars by Gel Electrophoresis and MALDI TOF/TOF Mass Spectrometry*"*. XXII Congresso Nazionale della Societa Chimica Italiana, 10-15 Settembre 2006, Firenze, Italy.

6. Dalpozzoa R., De Nino A., Di Donnaa L., Mazzucab S., Napolia A., ***Sajjad A***., Sindona G., Spadaforab A. "*Determinazione e caratterizzazione di proteine e peptici in olio extravergine d'oliva mediante elettroforesi su gel e spettrometria di massa MALDI-TOF/TO"F*. XIX Congresso di Chimica Analitica, 11-15 Settembre 2005, Pula (CA), Italy.

7. Mzzalupo I., Lombardo N., Pellegrino M., Perri E., ***Sajjad A***. *"Characterisation of Olive germplasm from Abruzzo region"*. ISBN 88-900622-6-6. XLIX Italian Society of Agricultural Genetics Annual Congress 12-15 September, 2005, Potenza, Italy.

8. Lombardo, N., Muzzalupo, I., Perri E., ***Sajjad A***. *"Characterisation of olive germplasm from Molise region by microsatellite markers"*. XVII International Botanical Congress,Vienna, Austria Center , 17 - 23 July 2005.

9. Aiello D., Matrici D., ***Sajjad A***., Di Donna L., Mazzotti F., Perri E., Napoli A., Sindona G. "*Proteic profile of the Olive Pollen by MALDI TOF/TOF. 53rd ASMS Conference*", 5-9 June 2005, San Antonio, USA.

10. Napoli A., Matrici D., ***Sajjad A***., Perri E., Sindona G. *"Profiling of the Olive Pollen by MALDI-TOF/TOF*". MASSA 2004 An International symposium on Mass Spectrometry, 26-30 September 2004, Bari, Italy.