Università degli Studi di Cagliari

PHD DEGREE

Earth and environmental sciences and technologies

Cycle XXX

TITLE OF THE PHD THESIS

Characterisation of Sardinian *Malus* and *Pyrus* varieties, through comparative seed image and genetic analyses

Scientific Disciplinary Sector(s)

BIO/03

PhD Student	Silvia Sau
Coordinator of the PhD Programme	Aldo Muntoni
Supervisor	Prof. Gianluigi Bacchetta PhD. Guy D'hallewin PhD. Oscar Grillo PhD. Mariano Ucchesu

Final exam. Academic Year 2016 – 2017

Thesis defence: February-March 2018 Session

Summary

Abstract

History of fruit tree cultivation in Sardinia seems to be dated from the Bronze Age and during the Phoenician-Punic period (Agabbio et al., 2015; Ucchesu et al., 2015). Many fruits were collected for human consumption as a fig, plum, grape, olive, almond and hazelnuts (Condit, 1947; Ucchesu et al., 2014; Sarigu et al., 2016). The long tradition of the cultivation of fruit trees in Italy has created over time a genetic biodiversity that currently is disappearing because of their substitution with new, more productive varieties. However, to avoid the loss of genetic diversity these they must be are properly maintained (Choen et al., 1991).

For Sardinia, the protection of biodiversity of wild and domesticated fruit species with limited diffusion is even more justified in that they represent a gene pool for the long geographical isolation and genetic shows great interest for research not only in Italy (Agabbio et al., 2015). The old varieties have been almost excluded from the orchards because of their low productivity, which in many cases did not have met the standards of modern varieties.

The fruit germplasm of Italian varieties is an important resource of genetic diversity that can be used in addition characterizations for the germplasm collections of European, to optimize the efficiency of the association studies within the genome and to identify genomic regions control the main horticultural characteristics (Liang et al., 2015).

Presently for the preservation of ancient varieties in Italy, there are different catalogue fields where the characteristics of these ancient varieties are preserved and studied both from the genotype and the phenotypic point of view.

References

Agabbio, M., Ansaldi, N., Beccaro, G.L., Berra, L., Botta, R., Brunu, A., et al., 2015. FRUTTI DIMENTICATI E BIODIVERSITA' RECUPERATA. Il germoplasma frutticolo e viticolo delle agricolture tradizionali italiane. Casi di studio: Piemonte e Sardegna. Quaderni NATURA E BIODIVERSITA'. Vol.7.

Condit, I., J., 1947. The Fig. Chronica Botanica, Waltham, MA.

Cohen, J., Williams, J., Plucknett, D., and Shands, H., 1991. Ex Situ Conservation of Plant Genetic Resources: Global Development and Environmental Concerns. Science. 253, 866-872.

Liang, W., Dondini, L., De Franceschi, P., Paris, R., Sansavini, S., Tartarini, S., 2015. Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. Plant Mol Biol Report. 33, 458-473.

Ucchesu, M., Peña-Chocarro, L., Sabato, D., Tanda, G., 2014.Bronze Age subsistence in Sardinia, Italy: cultivated plants and wild resources. Veg Hist Archaeobot. 24, 1-13.

Sarigu, M., Ucchesu, M., Loi, M.C., Grillo, O., Venore, G., D'hallewin, G., Bacchetta, G., 2016. La caratterizzazione delle varietà autoctone di susino (*Prunus domestica* L.) della Sardegna attraverso l'analisi morfo-colorimetrica degli endocarpi. EDUCAZIONE ALLA BIODIVERSITA' ANTICHE VARIETÀ DI FRUTTA DELLA SARDEGNA IL SUSINO, Editors: ISPA, LAORE. 61-67.

PhD structure

My PhD research was focused on the study of characterisation and comparison different Sardinian varieties of *Malus* and *Pyrus* throught morphometric and genetic analyses.
This research is divided in four chapter:

Chapter 1: we present the results of a study conducted on 25 Sardinian varieties of *Malus domestica* for investigating the possible existence a relationship between seed morphology and apple skin colour.
The main goals of this work are to:
(1) build a descriptive morphometric database of Sardinian apple seeds, useful for classifying the phenotypic characteristics;
(2) investigated the relationship between seed morphology and fruit skin colour.

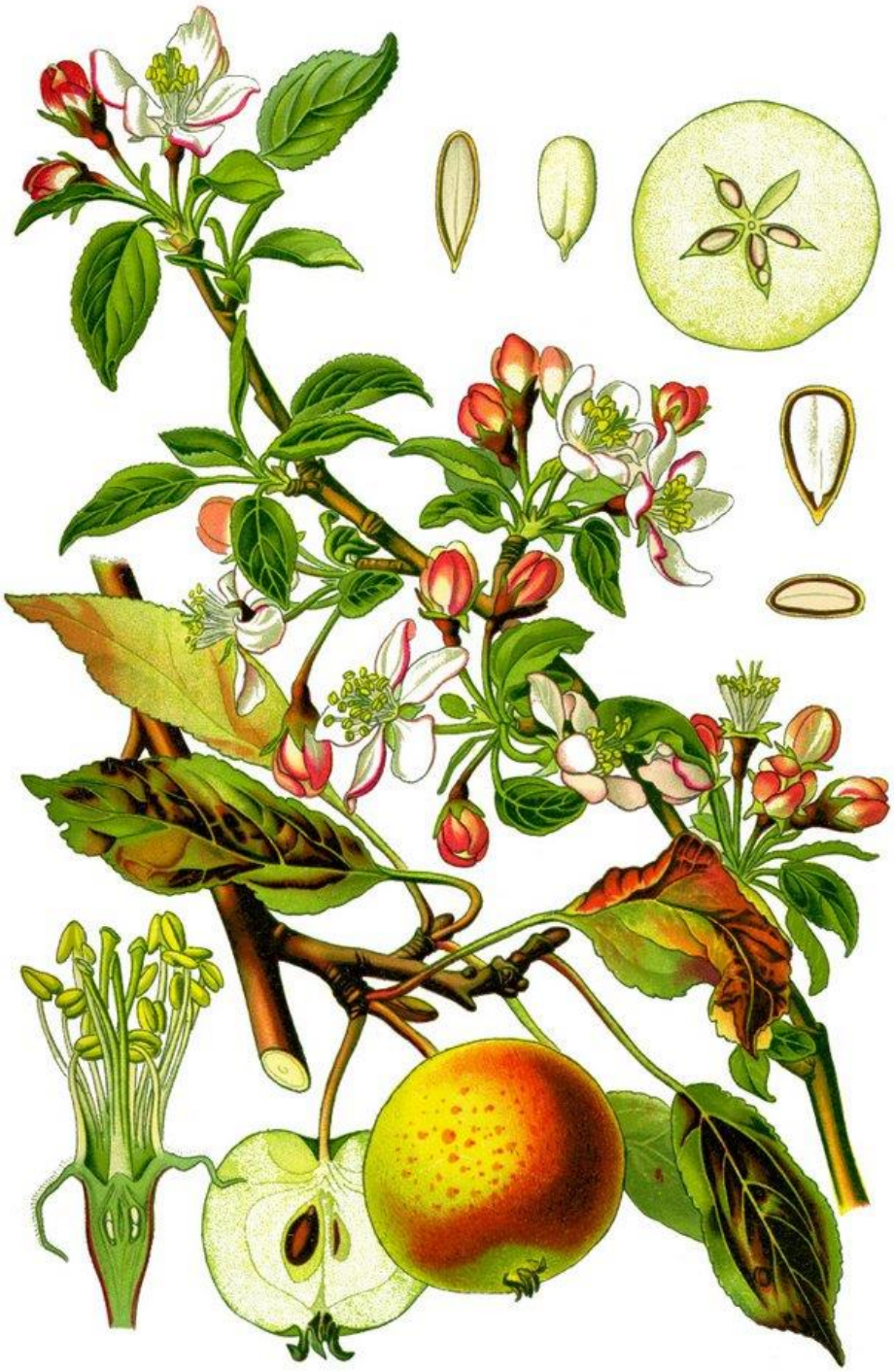
Chapter 2: we evaluate the trustworthiness of seed image analysis as an approach to discriminate apple germplasm accessions.
The main goals of this work were to:
(1) build a database of seed morphological variables of apple cultivars, suitable for variety characterization;
(2) assess the phenotypic diversity of apples by morphological seed image analysis techniques and by LDA;
(3) compare our seed image analysis data with a genetic study previously conducted on the same varieties (Liang et al., 2015).

Chapter 3: we present the result of study for investigating the relationship between *Pyrus communis* and *P.spinosa* and the phenomenon of over-ripening in Sardinian *P. communis* varieties. In this study, we compared 65 Sardinian varieties of *P. communis* with 44 international varieties and 7 Sardinian varieties of *P. spinosa*.
The main goals of this study are to investigate by seed image analysis the phenotype diversity of pears, in particular:
(1) the relationship between local cultivars and wild populations of Sardinian pear and national and international varieties;
(2) the phenomenon of pear over-ripening in Sardinian local cultivars.

Chapter 4: we present a **c**haracterization of microsatellite loci in Sardinian pears (*Pyrus communis* L. and *Pyrus spinosa* Forssk.).

The aim of the present study was to estimate the genetic relationship among wild and local pear varieties from Sardinia and national and international ones, using SSR markers. This molecular characterization will: identify the genetic diversity; investigate cases of homonymous and/or synonymous genotypes that are difficult to distinguish using standard morphological descriptors and increase the *Pyrus* molecular marker datasets.

The species *Malus domestica* Borkh.



1. Botanical description

Domestic apple (*Malus domestica* Borkh.) belongs to the great family of Rosaceae, which includes different forms with closely related fruits such as the *Pyrus* and *Cydonia* genera and ornamental genres such as *Amelanchier*, *Aronia*, *Chaenomeles*, *Cotoneaster*, *Crataegus*, *Pyracantha* and *Sorbus* (Challice, 1974).

The difficulty in delimiting species within the genus has been widely investigated (Korban, 1986; Rohrer et al., 1994; Robison et al., 2001). The morphological characters used to delineate *Malus* species and subspecies change continuously and in some cases overlap, contradicting each other (Harris et al., 2002; Hummer and Janick, 2009).

The genus *Malus* possesses pentamers flowers, with tubular cavities, do not open at the top and completely included carpels, which are a number of 3-5 and connected with the hypanthium, with cartilaginous walls in the fruit, each of which contains two or more ovules. From the carpels and the hypanthium, develop a fake fruit (pomo) more or less globally (Pignatti, 1982). The baseline chromosomal number is n = 17, while ploidy levels may be variable (from haploid to triploid). It includes about 55 species, distributed throughout the boreal hemisphere. Of them, only some have crop interest as fruit plants. Others are cultivated for ornamental purposes, for flowers and abundant and vibrantly colorful fruits.

2. Origin

Domestic apple is one of the world's most important fruit cultivars in terms of production level (FAOSAT 2014) and occupies a central position in culture, folklore and art of many populations.

The culture of apple is known to the Greeks end italics already in the early stages of their social development, at least from 800 BC Greek and Latin texts talk about it widely.

Moreover, the study conducted by Vavilov suggests that Turkestan wild apple and close relatives represented the progeny of domesticated apple, and see it as the centre of domestication origin of this Almaty in Kazakhstan (Forsline, 1995). Vavilov believed that the wild apple had similar fruits to the home apple could have had the same progenitor. Moreover, Janick and Moree's work (1996) confirmed that the area considered by Vavilov represents the area of greater biodiversity and the centre of origin of the domestic apple. With the passage of time humans began to occupy the area about 5000-8000 years ago, the first evolution of the apple was almost complete, and its migration was skillfully assisted by the use of domesticated horses.

Over several thousand years, within this migration flow, the many thousands of apple cultivars have come notes to date, as a result of unconscious and conscious selection.

Moreover, based on archaeological data combined with the molecular ones, it seems likely that in the late Neolithic or early Bronze Age, travelers on the major commercial routes from central China to the Danube transported the seed of Western Asia's wild apple, in handbags or bowels of horses (Crosby, 2007).

At present we know that the technique of grafting described as taming by Teofrasto (323 BC) was indispensable to optimize production, and how direct sowing normally yielded inferior quality fruits

Two stages seem to have been important in the domestication of apples; the initial introduction of apples in Western Europe and subsequently the hybridizations between cultivar and between cultivars and wild species (Zohari, 1991).

Morphology, biochemical, and molecular variation within wild apples indicate that the first selection of tamed apples could originate directly from their natural environment without the involvement of other species (Crosby, 2007). However, subsequent hybridization could have been important in creating new cultivars that carry economically important features. These analyses would require better research and sampling of cultivar varieties in Central Asia.

Central Asia currently represents a very important centre of diversity of wild apples that are distinguished by colour, taste and shape.

In the 1920s, Vavilov (1930) travelled through central Asia and reported that *M.sieversii*'s great wild specimens existed in specific locations and suggested the region as the centre of origin for domesticated apple (Janick, 2003).

Dzhangaliev (1977), confirmed the existence of wild apple forests at the time, also noted that they were under pressure in some areas due to urbanization, agriculture, fodder and wood harvesting.

The man has exploited this fruit for centuries by selecting it and giving birth to several thousands of documented cultivars. Numerous genetic studies have been carried out over the last decades in order to evaluate the origin of cultivars (Harris et al., 2002).

Studies on the variability of genetically informative traits were dealt with using different molecular biology techniques that were designed not only to determine the characteristic traits of the genome of each cultivar but also to identify progenitors.

Among the most widely used molecular biology techniques are those based on the study of restriction fragments (RFLPs), amplified fragment lengths (AFLPs), satellite DNA (SSR) or sequencing of genome traits, also used for the analysis of fruit trees (Coart et al., 2006).

The use of these methodologies implies a comparative study in order to choose the genome marker (nuclear or plastid) that best suits the needs of the operator.

All of these studies have been dealt with in the program of conservation of the genome of apple tree varieties in order to provide information on abandoned old varieties with good organoleptic characteristics, resistance to disease and therefore useful to provide native genetic resources to exploit also commercial purposes.

Most authors concur that the apple cultivars derive from *Malus pumila* (Heywood and Zohary, 1995) or from crosses between the latter and other species, such as *Malus sieversii* (Mill.), *M. orientalis* (Uglitzk.) and *M. sylvestris* (Mill.), all from Europe or Asia (Forsline et al., 2003). Clearly, there is still much to discover about the origin of apple domestication, the processes that led to its domestication and the origins of both desserts and cider apples.

3. Economic importance

M. domestica is the most important species from the economic point of view of the Rosaceae family, with world production of 84,630,275 tons (www.fao.org/faostat/en/#compare).

The popularity of apple is due to the fact that the fruit has many uses. It can be consumed fresh or preserved for months, while a large percentage of the crop is converted into sauces, juices, cider, pectin vinegar and baked goods. Winter apples, harvested in late autumn and stored at higher temperatures than freezing, are considered an important food in Europe, Asia, Argentina and the United States. The fruit has a low-calorie content and contains vitamin C. Some studies suggest that apples can help slow down the development of cancer, manage diabetes, and help patients prepare for surgery (Gossè et al., 2004).

In addition, other studies have shown that antioxidants contained in apples can protect nerve cells from diseases such as Alzheimer's and Parkinson's (Boyer et al., 2004).

However, further research is needed to confirm these results.

References

Angelini, R., 2009. Il melo. 1st Ed. Bayer Crop Science, Italy.

Boyer, J., Liu, R. H., 2004. Apple phytochemicals and their health benefits. Nutrition journal. 3, 5.

Challice, J., 1974. Rosaceae chemotaxonomy and the origins of the Pomoideae. Botanical Journal of the Linnean Society. 69, 239-259.

Crosby, M. R., 2007. The Story of the Apple.

Forsline, P. L., 1995. Adding diversity to the national apple germplasm collection: collecting wild apples in Kazakstan. New York Fruit Quarterly. 3, 3-6.

Gossé, F., Guyot, S., Roussi, S., Lobstein, A., Fischer, B., Seiler, N., and Raul, F., 2005. Chemopreventive properties of apple procyanidins on human colon cancer-derived metastatic SW620 cells and in a rat model of colon carcinogenesis. Carcinogenesis. 26, 1291-1295.

Harris, S. A., Robinson, J. P., and Juniper, B. E., 2002. Genetic clues to the origin of the apple. TRENDS in Genetics. 18, 426-430.

Heywood V. H., Zohary D., 1995. A catalogue of the wild relatives of cultivated plants native to Europe. Flora Mediterranea. 5, 375-415.

Hummer, K. E. and J. Janick. 2009. Rosaceae: taxonomy, economic importance, genomics. Genetics and Genomics of Rosaceae. 1-17.

http://www.fao.org/faostat/en/#data/QC

Janick, J., 2003. *Horticultural reviews*. John Wiley & Sons. Vol. 30.

Janick, J., 2005. The origins of fruits, fruit growing, and fruit breeding. Plant breeding reviews. 25, 5-320.

Juniper, B. E., and Mabberley, D. J., 2006. The story of the apple. Timber Press (OR).

Korban, S.S., 1986. Interspecific hybridization in Malus. HortScience. 21, 41-48.

Pignatti, S., 1982. Flora d'Italia. Edagricole. 1, 604-605.

Robinson, J. P., Harris, S. A., and Juniper, B. E. 2001. Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant. Syst. Evol. 226, 35-58.

Rohrer, J.R., K.R., Robertson, and J.B., Phipps. 1994. Floral morphology of Maloideae (Rosaceae) and its systematic relevance. Am. J. Bot. 81, 574-581.

Vavilov, N. I., and Freier, F., 1951. Studies on the origin of cultivated plants.

Zohary, D., 1991. Conservation of wild progenitors of cultivated plants in the Mediterranean Basin. Bot. Chron. 10, 467-474.

CHAPTER 1: Relationship between seeds morphology and skin colour of Sardinian apple cultivars.

1. Introduction

The genus *Malus* (Rosaceae) includes about 55 species mainly distributed in temperate regions (Cornille et al., 2012). The domesticated apple (*Malus domestica* Borkh.) is one of the most important fruit crops in terms of production levels (http://faostat.fao.org/).

The origin of apple domestication is still remained unknown due to his out-crossing, self-incompatibility and also because of the genus *Malus* capacity to hybridize itself and to generate highly variable progenies (Zohary et al., 2012; Cornille et al., 2012; Velasco et al., 2010).

In addition, morphological characters used to delimit species and subspecies in genus *Malus* are continuous and overlapping, making it difficult to effectively understand the exact number of this genus (Robison et al., 2001). Recent, genetic studies have allowed us to hypothesize that domestic apple is derived from *M. sieversii* and *M. roem* (Velasco et al., 2010; Cornille et al., 2012).

Today, there are over 10,000 varieties of apples with different variations in fruit skin colour and taste of the pulp (Harris et al., 2002; Cornille et al., 2012, 2014).

In Italy, the germplasm fruit collections is an important genetic resource that can be used for breeding program and for investigating genomic regions that control the main horticultural characteristics (Liang et al., 2015). In this way, the use of the molecular markers plays an important role determining apple biodiversity relationships, which can clear away from the frequent synonyms and homonyms within specific germplasm collections (Velasco et al., 2010; Cornille et al., 2012; Liang et al., 2015; Urrestarazu et al., 2012, 2016).

In the case of Sardinia, although the cultivation of apple is not very widespread, there is a considerable number of old varieties with excellent organoleptic characteristics. Currently, these old apple varieties have been almost excluded from the orchards because of their low productivity and stored in several catalogues of fields (Agabbio et al., 2015).

Among the different phenotypic traits for commercial interest, apple skin colour plays a very crucial role on the way to capture the attention of the consumers giving a significantly influence the market value (Kim et al., 2003). Apple skin colour is an important quality indicator that influence consumer acceptance, with a strong increase for red skin cultivars. This trend influences directly the growers and is becoming a very important goal in breeding programs (Velasco et al., 2010; Bi et al., 2014).

The class of flavonoids represents a fundamental role in the coloring of apple skin, in particular from anthocyanins that are responsible for the red colour but also for other variants of coloring

(Lancaster and Dougall, 1992). In fact, anthocyanins with chlorophyll and carotenoids determine the colour variation in apple skin (Honda et al., 2002). Several studies have established that in some apple cultivars high levels of anthocyanins are accumulated especially in red apples (White and Lespinasse, 1986; Honda et al., 2002; Kim et al., 2003; Ubi, 2004; Jaakola, 2013).

In apple, an R2R3 MYB transcription factor has been shown to control the pigmentation of the pulp and leaf anthocyanins (MYB10) and (MYB1) the fruit skin colour (Chagnè et al., 2007; Honda et al., 2002).

Over the last 20 years, the use of seed image analysis through Linear Discriminant Analysis (LDA) application has allowed characterising different cultivars such as *Vitis vinifera* subsp. *vinifera* L., *Olea europaea* L., *Prunus domestica* L., *Cucumis melo* L. (Orrù et al., 2012, 2013; Sabato et al., 2015; Piras et al., 2016).

In order to identify multiple variables measured on multiple samples multivariate statistical analysis could be very useful. Multivariate data contain much more information than univariate analysis. Multivariate data analysis techniques can be used to model factors and responses and find the relationship that exists between them. Information resulted from multivariate data are usually very helpful to understand the characteristics of different systems and processes (Zude et al. 2003; Perk et al., 2010; Caboni et al., 2017; Murgia et al., 2016).

In this paper, the relationship between the seeds morphology and the apple skin colour be analysed by seed image analysis and data processed by Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA).

The main goals of this work are to:

(1) build a descriptive morphometric database of Sardinian apple seeds, useful for classifying the phenotypic characteristics;

(2) investigated the relationship between seed morphology and fruit skin colour.

2. Material and Methods

2.1. Apple varieties

In this work 25 traditional Sardinia, apple varieties were investigated using seed image analysis. The germplasm varieties were collected in the field catalogue of CNR-ISPA (Nuraxinieddu, Oristano, Sardinia) (Table 1). The fruits were harvested at full maturity, after removing the pulp, the seeds were cleaned and washed, and dried at room temperature, following the manual of standard procedures adopted in the Germplasm Bank of Sardinia BG-SAR (Atzeri et al., 2012).

Table 1. *M. domestica* varieties investigated in this study. Keys skin colour classification UPOV 35: YY= yellow; GW= green-white; YG= yellow-green; BO= brown-ocher; YO= yellow-orange; YW: yellow-white. UPOV 37: OR= orange-red; PR= pink-red; NT= not detected; RR= red; RP= red-purple; PO= pink-orange. IBPGR 6.2.12: YY= yellow; GY= green-yellow; WC= white-cream; OO: orange; GG= green. IBPGR 12.6.13: OO= orange; RR= red; NT= not detected; PP: pink; BB: brown.
* = varieties used to LDA and multivariate analysis.

Code	Variety name	Origin	N° Seeds	UPOV 35	UPOV 37	UPOV 39	IBPGR 6.2.12	IBPGR 6.12.13
C1*	APPICCADORZA	BONARCADO	101	YY	OR	Uniform and mottled (Elstar)	YY	OO
C2	APPIO ROSSEGGIANTE	OLBIA-TEMPIO	60	GW	RP	Uniform (Red Jonaprince)	GY	RR
C3*	APIONE	LACONI	48	GW	NT	NT	GY	NT
C4	BACCALARISCA	BONARCADO	71	YY	OR	Uniform and mottled (Elstar)	YY	OO
C5	BIANCA DI ARITZO	ARITZO	60	YG	RR	Uniform (Red Jonaprince)	GY	RR
C6*	BIANCA DI USSASAI	USSASAI	60	YY	PR	Uniform and mottled (Elstar)	YY	RR
C7	BONARCADO 'A'	BONARCADO	60	YG	PO	Uniform and mottled just mentioned	GY	PP
C8	CADDINA	NUCHIS	39	GW	RP	Slight and uniform with well delimited streaks (Gravensteriner) up to 70%	GY	RR
C9	DAMA	NUCHIS	60	YG	RR	Uniform, streaked and mottled (Jonagold)	GY	RR
C10*	DE FERRU	LACONI	48	BO	NT	Uniform	WC	BB
C11	DE JERRU DE ARITZO	ARITZO	60	GW	RR	Uniform with streaks well delimitate (Jonagored)	GY	RR
C12*	DI CUGLIERI	CUGLIERI	60	YG	PR	uniform and mottled (Elstar)	GY	PP
C13*	DI LUGLIO	OLBIA-TEMPIO	60	YG	PR	Slight and uniform with well delimited streaks (Gravensteriner) up to 70%	YY	RR
C14	LACONI B	LACONI	49	YW	PR	Uniform with streaks well delimitate (Jonagored)	GY	RR
C15	LACONI D	LACONI	34	YO	RR	Uniform, streaked and mottled (80% of the area)	OO	RR
C16	MELA DI CUGLIERI	CUGLIERI	60	YG	PR	Uniform and mottled (Elstar)	GY	PP
C17	MIALI	SASSARI	72	YG	RR	Uniform, streaked and mottled (90% of the area) "	YY	RR
C18	OXIU	SASSARI	40	YG	OR	uniform and mottled (Elstar)	GY	OO
C19	OZZU	SASSARI	60	YG	NT	NT	GY	NT
C20	RANETTA	SASSARI	77	YG	RR	Uniform with streaks well delimitate (Jonagored)	GY	RR
C21	RENETTA	UNKNOW	60	YG	RR	Uniform and mottled (Elstar) 40-60%	YY	RR
C22	ROSA	LACONI	77	GW	RR	Uniform and mottled (Elstar) 40-60%	GY	RR
C23	SAN GIOVANNI ARRUBIA	LACONI	58	YG	PR	Uniform, streaked and mottled (80% of the area) "	GY	PP
C24*	SONADORE	BONARCADO	60	GW	RR	Uniform, streaked and mottled (80% of the area) "	GY	RR
C25	ZAZZARI	SASSARI	77	GW	NT	NT	GG	NT

Table 2. List of the 26 morphometric seed features measured on endocarps and calculated by Particles8 plugins from ImageJ v. 1.49.

Parameter	Description
Perim	Perimeter, calculated from the centres of the boundary pixels
Area	Area inside the polygon defined by the perimeter
Pixels	Number of pixels forming the seed image
MinR	Radius of the inscribed circle centred at the middle of mass
MaxR	Radius of the enclosing circle centred at the middle of mass
Feret	Larges taxis length
Breadth	Largest axis perpendicular to the Feret
CHull	Convex hull or convex polygon calculated from pixel centres
CArea	Area of the convex hull polygon
MBCRadius	Radius of the minimal bounding circle
AspRatio	Aspect ratio = Feret/Breadth
Circ	Circularity = $4 \cdot \pi \cdot \text{Area}/\text{Perimeter}^2$
Roundness	Roundness = $4 \cdot \text{Area}/(\pi \cdot \text{Feret}^2)$
ArEquivD	Area equivalent diameter = $\sqrt{((4/\pi) \cdot \text{Area})}$
PerEquivD	Perimeter equivalent diameter = Area/π
EquivEllAr	Equivalent ellipse area = $(\pi \cdot \text{Feret} \cdot \text{Breadth})/4$
Compactness	Compactness = $\sqrt{((4/\pi) \cdot \text{Area})}/\text{Feret}$
Solidity	Solidity = Area/Convex_Area
Concavity	Concavity = Convex_Area-Area
Convexity	Convexity = Convex_hull/Perimeter
Shape	Shape = $\text{Perimeter}^2/\text{Area}$
RFactor	RFactor = Convex_Hull /$(\text{Feret} \cdot \pi)$
ModRatio	Modification ratio = $(2 \cdot \text{MinR})/\text{Feret}$
Sphericity	Sphericity = MinR/MaxR
ArBBox	Area of the bounding box along the feret diameter = Feret·Breadth
Rectang	Rectangularity = Area/ArBBox

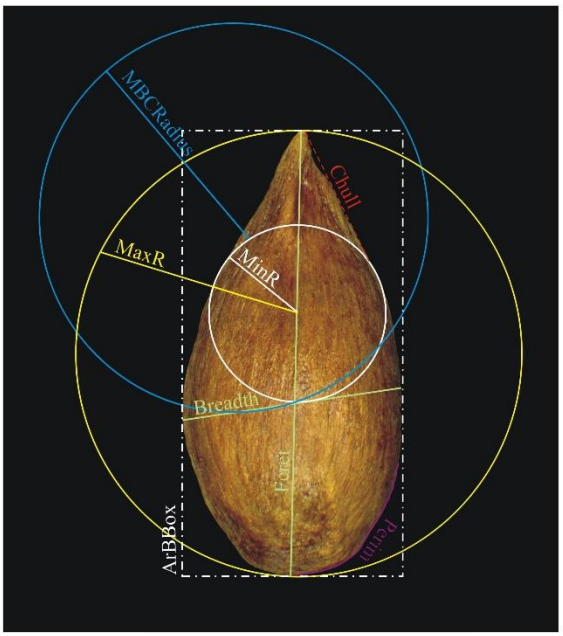


Fig 1. Graphical representation of principal morphometric parameters measured on each seed. Sample details are reported in Table 2.

2.2.2. Seed image analysis

The digital images of seeds were acquired using a flatbed scanner (Epson Perfection V550), with a digital resolution of 800 dpi for a scanning area not exceeding 1024×1024 pixels (Bacchetta et al., 2008). Digital images of seeds were processed and analyzed using the software package ImageJ v. 1.49 (http://rsb.info.nih.gov/ij). A plugin, Particles8, freely available on the official website (http://www.mecourse.com/landinig/ software/software.html) was used to measure 26 seed morphometric features (Table 2).

2.2.3. Linear Discriminant analysis

The morphometric parameters were used to build a database of features descriptive of seed size and shape. Statistical analyses were executed using the IBM SPSS software (Statistical Package for Social Science) release 16.0 (SPSS Inc. for Windows, Chicago, Illinois, USA).

The data were statistically processed by linear discriminant analysis (LDA) to identify and discriminate among the investigated apple varieties.

LDA is a method generally used to classify/identify unknown groups through the quantitative and qualitative variables (Sugiyama, 2007). Therefore, LDA is able to select and change the different predictor variables entered into the database minimizing the classification of short distance achieving and maximum the discrimination between the different classes (Hastie et al., 2002; Holden et al., 2011; Rencher et al., 2012; Kuhn and Johnson, 2013).

To verify the existence of some correlation between the seeds morphology and skin colour. The 25 apple varieties were characterized according to the descriptors of the International Union for the Protection of New Varieties of Plants (UPOV) (UPOV TG / 23/06, 2013) (Table 3) and those reported by the International Board for Plant Genetic Resources (IBPGR), (IBPGR 2015) (Table 4).

2.2.4. Multivariate statistical data analysis

From image analysis of apple varieties, an X matrix was created and subsequently was elaborated through chemometric techniques using SIMCA Software (14.0, Umetrics, Umeå, Sweden). In a multivariate analysis, data set has to be pre-treated and converted in a more eligible form to the analysis. In this work, the unit variance scaling was used. To show the complexity of a phenome constituted by a large number of variables using a small number of summary indicators (latent variables) principal component analysis was achieved (PCA). The PCA may be very useful to investigate variables measured on a large number of samples and to provide further support to understand the complex connection between the morphology of the seeds and characteristics of the fruit. In fact, multivariate data analysis techniques can be used to model factors and responses and understand the characteristics of different systems and process (Zude et al., 2003; Perk et al., 2010; Caboni et al., 2017; Murgia et al., 2016). This explorative method identifies the distribution of a data set and highlights similarities and differences no suspected among the data. PCA results are reported into different plots. The scores plot reported the samples projection into the model space, calculated by the use of principal components, while the loadings plot reports the projection of variables, using the same rules. These two plot can be reported overlaid as a Biplot. The quality of the model can be evaluated by the R^2 and Q^2 parameters. The R^2 is the fraction of the variation of the variables explained by the model. Q^2 is an estimate of the predictive ability of the model calculated by cross-validation. To better appreciate the presence of outliers Hotelling's T^2 and DmodX analyses were achieved. With the aim to classify the different classes of apples, a discriminant analysis (PLS-DA) was performed. The generated R^2Y and Q^2 values describe the reliability and the predictive ability of the fitting. Q^2 is made on the basis of a cross-validation analysis. R^2Y described the classificatory power of the model. Permutation test, a non-parametric test, was also performed to highlight the classificatory power of the model. A variable importance in the project test (VIP) gave the measure of a variables importance in the PLS-DA model.

2.2.5. Univariate statistical analysis.

GraphPad Prism software (version 7.01, GraphPad Software, Inc., CA, USA) was used to verify the significance of discriminant variables obtained by the supervised analysis. A one-way ANOVA test was performed and $P = 0.05$ was used as a limit of significance.

3. Results

3.1. Linear Discriminant Analysis

A first comparative analysis of the 25 apple varieties from Sardinia was performed, achieving 27% cross-validated grouped cases (SM 1; SM 2).

The highest discrimination was achieved by 'Appione' (70.8%), 'De Ferru' (60.4%) and 'Laconi B' (67.3%) (SM 1). For the same varieties, a second analysis was performed considering the apple group division according to the colour classifications UPOV 35, 37, 39 and IBPGR 6.2.12 and 6.12.13.

Considering the UPOV 35 parameter (ground colour), the 25 varieties were divided into six groups (Table 3a; SM 2). The LDA analysis conducted on these six groups achieved cross-validated grouped cases of 54.9% (Table 3a; SM 2). The highest classification was obtained for the yellow apple group (YG) with a 91.5% of classification (Table 3a, SM 2). Furthermore, considering the UPOV 37 parameter (over colour), the 25 varieties were divided into another six groups, the LDA achieving a cross-validated of 41.4% (Table 3b,). In this case, the highest classification was obtained for the red apple group (RR) with a correct percentage of classification of 84% (Table 3b). Finally, considering the UPOV 39 parameter (colour distribution), the 25 varieties were divided into six groups and in this case, the LDA achieving a cross-validated grouped cases of 36.9% (Table 3c). The group with uniform and the mottled colour (1) was classified with a high level of discrimination 81% (Table 3c).

Additional analyses were conducted considering the ground colour IBPGR 6.2.12 and IBPGR 6.12.13 colorimetric parameter (Table 1). Considering the IBPGR 6.2.12, the 25 varieties were divided into four groups, the LDA achieved cross-validated grouped cases of 66.8% (Table 4a; Fig. 2a). The green/yellow apple group (GY) obtained the highest percentage of discrimination with a 98.8% of classification (Table 4a). In this case, all other groups tend to be confused with the yellow/green apple group (Table 4a).

Finally, LDA conducted on five apple groups considering the parameter IBPGR 6.12.13 (over colour) were divided in to five groups and LDA achieved cross validated grouped cases of 59.9%. The high level of discrimination has been shown for the red apple group (RR) with a 94.4% (Table 4b; Fig. 2b). Moreover, the other groups tend to be confused with the group with the highest percentage of discrimination (Table 4b).

Considering that the highest classification percentages were obtained by the ground colour colorimetric parameters UPOV35 and IPBGR 6.2.12, we selected seven apples representative of all colorimetric parameters and performed the LDA and PCA considering only the UPOV35 and IPBGR 6.2.12 parameters (Table 1).

A first comparative analysis of seven varieties was performed by LDA, obtaining cross-validated grouped cases of 56.1% (Table 5a; Fig. 3a). Almost all varieties obtained a good percentage of discrimination except for the varieties 'Di Luglio' that tends to be confused with 'Appicadorza' (21,7%), 'Bianca di Ussassai' (16,7%) and 'De Ferru' (15%), respectively (Table 5a).

Considering the UPOV 35, the seven varieties were divided into four groups (Table 5b; Fig 3b), in this case the LDA showed, a cross-validated grouped cases 66.8%, were the green/white (GW) group was the best discriminated with a percentage of 81.5% (Table 5b). Moreover, LDA analysis showed that the yellow/green (YG) group was confused with the yellow (YY) apple group in 45% of the cases (Table 5b).

Finally, the LDA analysis performed on three groups of apples classified according to the IBPGR 6.2.12 parameter showed that almost all groups were correctly discriminated, with a cross-validated grouped cases of 69.9% (Table 5c; Fig 3c). In this case the yellow (YY) group was the best discriminated with a percentage of 83,7% (Table 5c).

Table 3(a). Classification percentage considering UPOV 35 parameters (ground colour), for 25 varieties.

Classification Results									
UPOV 35			Predicted Group Membership						Total
			YY	GW	YG	BO	YW	YO	
Cross-validated	Number seeds	YY	0	38	176	6	1	0	221
		GW	3	179	293	8	8	0	491
		YG	2	46	638	10	1	0	697
		BO	0	0	29	19	0	0	48
		YW	0	13	26	1	9	0	49
		YO	0	4	30	0	0	0	34
	%	YY	0	17,2	79,6	2,7	0,5	0	100,0
		GW	0,6	36,5	59,7	1,6	1,6	0	100,0
		YG	0,3	6,6	91,5	1,4	0,1	0	100,0
		BO	0	0	60,4	39,6	0,0	0	100,0
		YW	0	26,5	53,1	2,0	18,4	0	100,0
		YO	0	11,8	88,2	0	0	0	100,0
54,9% of cross-validated grouped cases correctly classified.									

Table 3 (b). Classification percentage considering UPOV 37 parameters (over colour) for 25 varieties.

Classification Results									
UPOV 37			Predicted Group Membership						Total
			OR	RP	NT	RR	PR	PO	
Cross-validated	Number seeds	OR	3	1	18	124	25	0	171
		RP	1	8	52	63	34	1	159
		NT	3	10	96	142	53	0	304
		RR	0	1	11	426	69	0	507
		PR	2	10	32	190	105	0	339
		PO	0	1	6	25	28	0	60
	%	OR	1,8	0,6	10,5	72,5	14,6	0	100,0
		RP	0,6	5,0	32,7	39,6	21,4	0,6	100,0
		NT	1,0	3,3	31,6	46,7	17,4	0	100,0
		RR	0	0,2	2,2	84,0	13,6	0	100,0
		PR	0,6	2,9	9,4	56,0	31,0	0	100,0
		PO	0	1,7	10,0	41,7	46,7	0	100,0
41,4% of cross-validated grouped cases correctly classified.									

Table 3 (c). Classification percentage considering UPOV 39 parameters (distribution colour) for 25 varieties. 1: Uniform and mottled (Elstar); 2: Uniform (red Jonaprince); 3: NT (absent); 4: Uniform and mottled just mentioned; 5: Slight and uniform with well delimited streaks (Gravensteriner) up to 70%; 6: Uniform, streaked and mottled (Jonagold); 7: Uniform with streaks well delimitate (Jonagored).

Classification Results										
UPOV 39			Predicted Group Membership							Total
			1	2	3	4	5	6	7	
Cross-validated	Number seeds	1	405	1	17	0	24	53	0	500
		2	105	4	2	0	17	40	0	168
		3	108	0	47	0	3	26	0	184
		4	35	0	7	0	5	13	0	60
		5	53	1	4	0	33	7	0	98
		6	240	1	23	0	1	79	0	344
		7	129	0	4	0	12	41	0	186
	%	1	81,0	0,2	3,4	0	4,8	10,6	0	100,0
		2	62,5	2,4	1,2	0	10,1	23,8	0	100,0
		3	58,7	0	25,5	0	1,6	14,1	0	100,0
		4	58,3	0	11,7	0	8,3	21,7	0	100,0
		5	54,1	1,0	4,1	0	33,7	7,1	0	100,0
		6	69,8	0,3	6,7	0	0,3	23,0	0	100,0
		7	69,4	0	2,2	0	6,5	22,0	0	100,0
36,9% of cross-validated grouped cases correctly classified.										

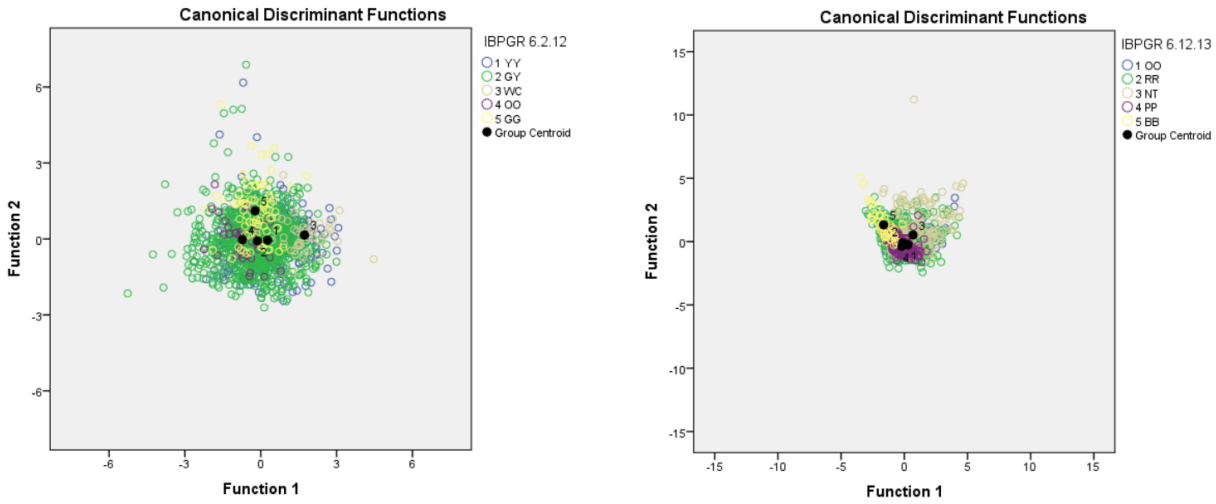


Fig. 2. Scatter plot graphs (A, B) based on LDA analysis discrimination conducted considering the colorimetric parameters on 25 varieties of *M. domestica*: (A) IBPGR 6.2.12 (ground colour); (B) IBPGR 6.12.13 (over colour).

Table 4 (a). Classification percentage considering IBPGR 6.2.12 (ground colour) parameters for 25 varieties.

Classification Results								
IBPGR 6.2.12			Predicted Group Membership					Total
			YY	GY	WC	OO	GG	
Cross-validated	Number seeds	YY	0	342	8	0	3	353
		GY	0	1016	4	0	8	1028
		WC	0	40	8	0	0	48
		OO	0	34	0	0	0	34
		GG	0	72	0	0	5	77
	%	YY	0	96,9	2,3	0	0,8	100,0
		GY	0	98,8	0,4	0	0,8	100,0
		WC	0	83,3	16,7	0	0	100,0
		OO	0	100,0	0	0	0	100,0
		GG	0	93,5	0	0	6,5	100,0
66,8% of cross-validated grouped cases correctly classified.								

Table 4 (b). Classification percentage considering IBPGR 6.12.13 (over colour) parameters for 25 varieties.

Classification Results								
IBPGR 6.12.13			Predicted Group Membership					Total
			OO	RR	NT	PP	BB	
Cross-validated	Number seeds	OO	0	153	18	0	0	171
		RR	1	837	35	0	14	887
		NT	0	183	70	0	3	256
		PP	0	177	1	0	0	178
		BB	0	32	0	0	16	48
	%	OO	0	89,5	10,5	0	0	100,0
		RR	0,1	94,4	3,9	0	1,6	100,0
		NT	0	71,5	27,3	0	1,2	100,0
		PP	0	99,4	0,6	0	0	100,0
		BB	0	66,7	0	0	33,3	100,0
59,9% of cross-validated grouped cases correctly classified.								

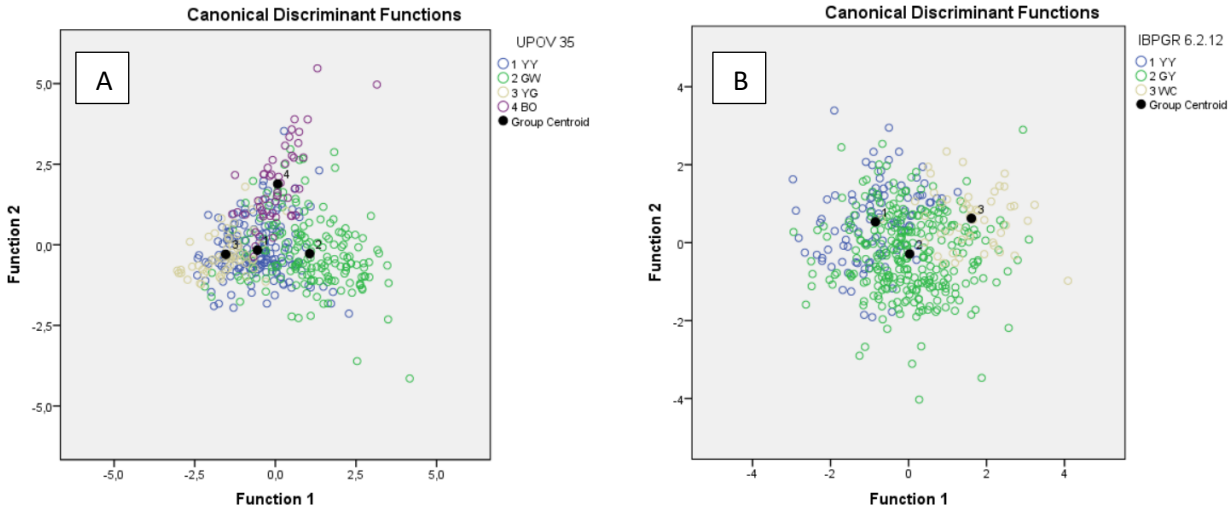


Fig. 3. Scatter plot graphs based on LDA analysis discrimination conducted considering 7 varieties of *M. domestica* for their colorimetric parameters: (A) UPOV 35 (ground colour), (B) IBPGR 6.2.12. (ground colour).

Tab 5(a). Classification percentage considering preliminary analysis for 7 varieties.

Classification Results										
7 ACCESSIONS			Predicted Group Membership							Total
			Appicadorza	Appiona	Bianca di Ussassai	De Ferru	Di Cuglieri	Di luglio	Sonadore	
Cross-validated	Number Seeds	Appicadorza	62	0	7	1	22	2	7	101
		Appiona	2	33	0	0	0	0	13	48
		Bianca di Ussassai	8	0	29	7	7	9	0	60
		De Ferru	4	0	10	32	0	2	0	48
		Di Cuglieri	22	0	5	3	30	0	0	60
		Di luglio	13	1	10	9	2	23	2	60
		Sonadore	9	10	2	2	0	1	36	60
	%	Appicadorza	61,4	0	6,9	1,0	21,8	2,0	6,9	100,0
		Appiona	4,2	68,8	0	0	0	0	27,1	100,0
		Bianca di Ussassai	13,3	0	48,3	11,7	11,7	15,0	0	100,0
		De Ferru	8,3	0	20,8	66,7	0	4,2	0	100,0
		Di Cuglieri	36,7	0	8,3	5,0	50,0	0	0,0	100,0
		Di luglio	21,7	1,7	16,7	15,0	3,3	38,3	3,3	100,0
		Sonadore	15,0	16,7	3,3	3,3	0	1,7	60,0	100,0
56,1% of cross-validated grouped cases correctly classified.										

Tab.5(b). Classification percentage considering UPOV 35 (ground colour) parameters for 7 varieties.

Classification Results							
UPOV 35			Predicted Group Membership				Total
			YY	GW	YG	BO	
Cross-validated	Number seeds	YY	121	6	28	6	161
		GW	17	88	1	2	108
		YG	54	1	55	10	120
		BO	12	0	8	28	48
	%	YY	75,2	3,7	17,4	3,7	100,0
		GW	15,7	81,5	0,9	1,9	100,0
		YG	45,0	0,8	45,8	8,3	100,0
		BO	25,0	0	16,7	58,3	100,0
66,8% of cross-validated grouped cases correctly classified.							

Tab.5 (c). Classification percentage considering IBPGR 6.2.12 (ground colour) parameters for 7 varieties.

Classification Results						
IBPGR 6.2.12			Predicted Group Membership			Total
			YY	GY	WC	
Cross-validated	Count	YY	185	24	12	221
		GY	70	93	5	168
		WC	18	4	26	48
	%	YY	83,7	10,9	5,4	100,0
		GY	41,7	55,4	3,0	100,0
		WC	37,5	8,3	54,2	100,0
69,6% of cross-validated grouped cases correctly classified.						

3.2 Multivariate statistical data analysis

Based on the morphometric analysis of the seven selected apple varieties based on the ground colour UPOV35 and IPBGR 6.2.12 (Table 5a) a 437 X 26 matrix was shaped and analyzed by a multivariate data statistical analysis using the SIMCA software. The matrix was set up with 437 seed samples (Table 5a) and 26 morphometric parameters (Table 2). The Hotelling's T^2 and DmodX analysis showed 24 outliers that were excluded from the pool of samples. A new PCA was than performed showing the following validation parameters: $R^2X=0.93$ and $Q^2=0.88$. The results, reported as a Biplot in figure 4a, display different clusters along both of the two principal components. No clusters were observed based on the geographical location or varieties classification. Considering the ground skin colour (IBPGR 6.2.12), the White-Creamy (WC) samples clustered along the second principal component. This cauterization is due to the minR and the Breadth parameters. On the other hand, the yellow (YY) and yellow/green (YG) samples were more scattered indicating a strong infraclass variability (Fig. 4a).

Based on the UPOV 35 classification, the PCA showed different clusters among the first principal component. In fact, in the scores plot of figure 4b it is remarkable a clusterisation of class GY and BO classes on the left side of the plot, while the GW class clustered on the right side. On the other hand, the YY class appeared more scattered in both of the two first principal components.

Excluding the YY class from the model the resulted PLS-DA of the remained three classes BO *vs* GY *vs* GW showed an R^2Y of 0.51 and a Q^2 of 0.49 (Fig. 4c). Classificatory values are reported in table 6 from the classification test. From the loadings and the VIP analyses the resulted discriminant morphometric parameters were: Feeret, MaxR, MBCRadius, CHull, Perim, Pixels, Area, AspRatio and ArEquivD for the GW class, while Breadth, RFactor, Compactness, Roundness, MinR, Rectang Mod Ratio and Circ for the GY and BO classes (SM 3). All the variables were subjected to an ANOVA non parametric test to confirm the significance of the variance taking $P=0.05$ as the limit of significance.

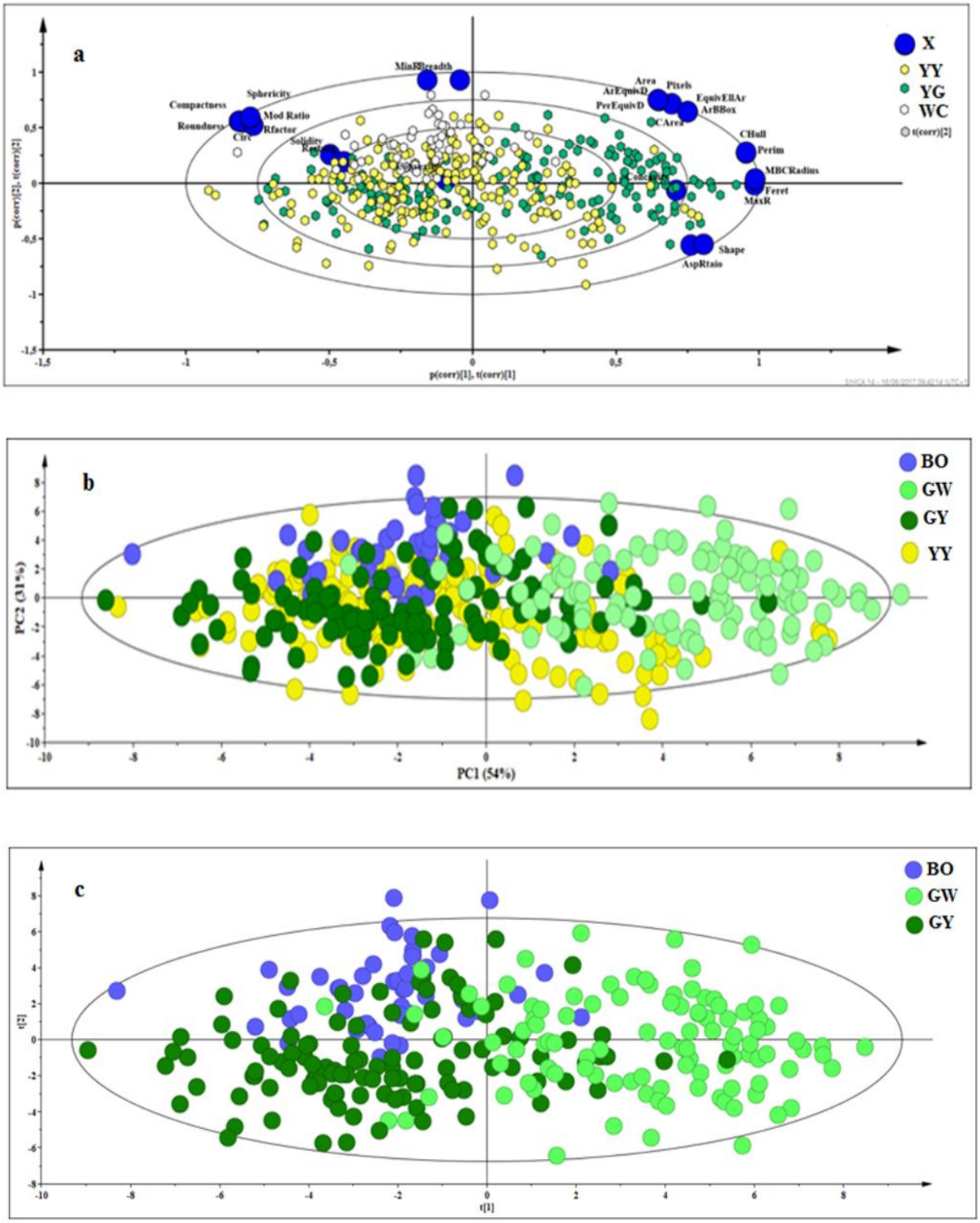


Fig 4. PCA Biplot of apple samples analysis PC1 vs PC2: $R^2X=0.93$ $Q^2=0.88$.
a: Scores plot colored by IBPGR 6.2.12 classification. Yellow class (YY); green/yellow class (GY); white/creamy class (WC); blue circles variables (X).
b: PCA scores plot of apple samples colored by UPOV 35 classification. Yellow class (YY); green/white (GW); yellow/green (YG); brown/orange (BO).
c: PLS-DA score plot of apple samples classified by UPOV 35 parameters: green/white (GW) *vs* yellow/green (YG) *vs* brown/orange (BO). $R^2X=0.51$ $Q^2=0.49$.

Table 6. Classification list related to the PLS-DA made based on UPOV 35 (ground colour) parameter for 3 classes GW (green/white); YG (yellow/green) and BO (brown/orange). The analysis classified correctly an average of 81% of the samples.

	Members	Correct	G-W	Y-G	B-O
G-W	104	92,31%	96	6	2
Y-G	111	84,68%	10	94	7
B-O	48	52,08%	2	21	25
No class	174	-	34	134	6
Total	437	81,75%	142	255	40

4. Discussion

Considering the importance of the apple skin colour in breeding program, we have investigated this phenotypic characteristic by seed image analysis and using two different types of statistical analysis: LDA and PCA.

The first LDA analysis conducted on 25 apple varieties showed a high degree of discrimination for the varieties ‘Apione’, ‘De Ferru’ and ‘Laconi B’. These varieties come from the Laconi territory (placed in the Sarcidano region), that was considered for a long time one great biodiversity area of wild and cultivated plant species.

In fact, the literature reported that already in 1800 there have been more than 65 thousand fruit trees such as almonds, walnuts, chestnuts, figs, pears, apricots, cherries and plum trees (Agabbio et al., 1994; Angius, 2006). This testifies that in this area several fruit species have been cultivated for a long time and that probably many of them have been preserved to present day. In fact, ‘Apione’, ‘De Ferru’ and ‘Laconi B’ are considered the ancient local Sardinian varieties (Agabbio et al., 1994) and this could justify the high level of diversity when compared to the other apple varieties studies in this work.

In addition, the LDA analysis, considering the different UPOV and IBPGR classifications, allowed studying the relationship between the seed morphology and the apple skin colour.

LDA showed a high discrimination between the apple groups with yellow/green ground colour and red over colour. Even considering the other colorimetric groups it was observed that these tend to be confused with ground colour yellow/green and the red over colour.

Some considerations about this result can be attributed to the ability of some apples to develop the red skin colour. As it is well known, the colour of the fruit skin is due to anthocyanin pigments that produce colors ranging from red, blue and purple (Pomar et al., 2005; Torres et al., 2011; Jaakola, 2013). Red fruit skin colour is an important characteristic for consumer preference and marketability (Allan et al., 2008). Thus, several genetically studies have been conducted to understand which mechanisms involved in the red coloring of apples (Honda et

al., 2002; Kim et al., 2003; Ubi, 2004; Allan et al., 2008; Liu et al., 2013). These studies have established that the red coloration is induced by specific alleles (Zhang et al., 2014; Chagné et al., 2007, 2013; Ban et al., 2007; Lin-Wang et al., 2010). In addition, other studies have established that apples with yellow/green colour easily developed red colour respect to yellow apples after being properly stimulated (Honda et al., 2002; Liu et al., 2013). Thus, the correct classification of yellow/green and red apple groups by LDA could be related to the capacity of these apples to develop red coloring in their skin.

When the UPOV 39 colorimetric parameter was considered, the results obtained with the LDA showed a low discriminating power, in fact, all groups tend to be confused with the uniform and mottled colorimetric parameter.

The result obtained by comparing the seven varieties of apples through LDA shows how image analysis is able to discriminate the varieties studied considering them as individual varieties and as considering descriptive colorimetric parameters. In fact, the results achieved considering these small sample of apples and analysed by LDA and the multivariate statistical data analysis, suggest that the apple groups with ground colour green/white in the case of UPOV 35 and yellow for IBPGR 6.2.12 were the groups that obtained the best classification.

Considering the UPOV 35 colorimetric classification, the groups of apples that tend to confused among themselves were the group with ground colour yellow and yellow/green respectively. These results were confirmed with LDA and PCA analysis. The same result also occurred with IBPGR 6.2.12 classification. In this case, the yellow apple group were confused with the green/yellow group.

Moreover, based on the UPOV35 skin colour classification from the multivariate statistical analysis, several morphometric parameters were found statistically discriminant for the different colour classes. These results highlight a connection between the variability of seeds shape and the apple skin colour. Taking in consideration the good statistical results obtained from both analysis, the multivariate approach, used together with the morphometric analysis, can be used in future as original tool for the investigation of skin colour phenotype in apple varieties.

The most important result of this work is the demonstration that there is a direct correlation between seed morphology and apple skin colour. This relationship has never been described until now and it is an important starting point to investigate phenotype characters in modern varieties.

References

Agabbio, M., 1994. Patrimonio genetico di specie arboree da frutto le vecchie varietà della Sardegna. Carlo Delfino Editore. pp.1-424.

Allan, A.C., Hellens, R.P., Laing, W.A., 2008. MYB transcription factors that colour our fruit. Trends Plant Sci. 13, 99-102.

Angius, V., and Carta, L., 2006. Città e villaggi della Sardegna dell'Ottocento. Ilisso.

Atzeri, P., Bacchetta, G., Casti, M., Fenu, G., Mattana, E., Meloni, F., et al., 2012. La Banca del Germoplasma della Sardegna (BG-SAR). Studi Trent. Sci. Nat. 88, 69-75.

Bacchetta, G., Grillo, O., Mattana, E., Venora, G., 2008. Morpho-colourimetric characterization by image analysis to identify diaspores of wild plant species. Flora. 203, 669-682.

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T., 2007. Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. Plant and cell physiology. 48, 958-970.

Bi, X., Zhang, J., Chen, C., Zhang, D., Li, P., and Ma, F., 2014. Anthocyanin contributes more to hydrogen peroxide scavenging than other phenolics in apple peel. Food chemistry. 152, 205-209.

Caboni, P., Manis, C., Ibba, I., Contu, M., Coroneo, V., and Scano, P., 2017. Compositional profile of ovine milk with a high somatic cell count: A metabolomics approach. International Dairy Journal. 69, 33-39.

Chagné, D., Carlisle, C. M., Blond, C., Volz, R. K., Whitworth, C. J., Oraguzie, N. C., and Gardiner, S. E., 2007. Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. Bmc Genomics. 8, 1-1.

Cornille, A., Gladieux, P., Smulders, M. J. M., Roldan-Ruiz, I., Laurens, F., Le Cam, B., et al., 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genetics 8: e1002703.

Hastie, T., Tibshirani, R., and Friedman, J., 2002. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Biometrics.

Harris, S. A., Robinson, J. P., Juniper, B. E., 2002. Genetic clues to the origin of the apple. Trends Genet. 18, 426-430.

Holden, J. E., Finch, W. H., Kelley, K., 2011. A comparison of two-group classification methods. Educ Psychol Meas. 71, 870-90.

Honda, C., Kotoda, N., Wada, M., Kondo, S., Kobayashi, S., Soejima, J., et al., 2002. Anthocyanin biosynthetic genes are coordinately expressed during red colouration in apple skin. Plant Physiol Biochem. 40, 955-962.

Jaakola, L., 2013. New insights into the regulation of anthocyanin biosynthesis in fruits. Trends Plant Sci. 18, 477-483.

Kim, S. H., Lee, J. R., Hong, S. T., Yoo, Y. K., An, G., Kim, S. R., 2003. Molecular cloning and analysis of anthocyanin biosynthesis genes preferentially expressed in apple skin. Plant Science. 165, 403-413.

Kuhn, M., and Johnson, K., 2013. Discriminant analysis and other linear classification models. Applied Predictive Modeling. Springer. pp. 275-328.

Lancaster, J. E., and Dougall, D. K., 1992. Regulation of skin colour in apples. Crit. Rev. Plant Sci. 10, 487-502.

Ling-Wang, K. U. I., Micheletti, D., Palmer, J., Volz, R., Lozano, L., Espley, R., et al., 2011. High temperature reduces apple fruit colour via modulation of the anthocyanin regulatory complex. Plant, Cell & Environment. 7, 1176-1190.

Liu, Y., Che, F., Wang, L., Meng, R., Zhang, X., & Zhao, Z., 2013. Fruit colouration and anthocyanin biosynthesis after bag removal in non-red and red apples (Malus× domestica Borkh.). Molecules, 18, 1549-1563.

Murgia, A., Scano, P., Contu, M., Ibba, I., Altea, M., Bussu, M., et al., 2016. Characterization of donkey milk and metabolite profile comparison with human milk and formula milk. LWT-Food Science and Technology. 74, 427-433.

Orru, M., Grillo, O., Venora, G., & Bacchetta, G., 2012. Computer vision as a method complementary to molecular analysis: Grapevine cultivar seeds case study. Comptes rendus biologies. 335, 602-615.

Orrù, M., Grillo, O., Lovicu, G., Venora, G., Bacchetta, G., 2013. Morphological characterisation of *Vitis vinifera* L. seeds by image analysis and comparison with archaeological remains. Veg Hist Archaeobot. 22, 231-242.

Perk, S., Cinar, A. and Zude, M., 2010. Multivariate analysis and statistical monitoring of apple fruit development on the tree. Acta Horticulturae. 858, 439-442.

Piras, F., Grillo, O., Venora, G., Lovicu, G., Campus, M., Bacchetta, G., 2016. Effectiveness of a computer vision technique in the characterization of wild and farmed olives. Comput. Electron. Agric. 122, 86-93.

Pomar, F., Novo, M. and Masa, A., 2005. Varietal differences among the anthocyanin profiles of 50 red table grape cultivars studied by high performance liquid chromatography. Journal of Chromatography a. 1094, 34-41.

Rencher, A. C., Christensen, W. F., 2012. Methods of multivariate analysis. New York: John Wiley. pp. 800.

Robinson, J. P., Harris, S. A. and Juniper, B. E., 2001. Taxonomy of the genus *Malus* Mill (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant Systematics and Evolution. 226, 35-58.

Sabato, D., Esteres, C., Grillo, O., Bacchetta, G., 2015. Seeds morpho-colourimetric analysis as complementary method to molecular characterization of melon diversity. Sci Hort. 192, 441-452.

Sugiyama, M., 2007. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. J Mach Learn Res. pp. 27-106.

Torres, B., Tiwari, B. K., Patras, A., Cullen, P. J., Brunton, N., and O'donnell, C. P., 2011. Stability of anthocyanins and ascorbic acid of high pressure processed blood orange juice during storage. Innovative Food Science & Emerging Technologies. 12, 93-97.

Ubi, B.E., 2004. The genetics of anthocyanin reddening in apple fruit skin Intl. J. Food Agricult. Environ. 1, 163-165.

Urrestarazu, J., Miranda, C., Santesteban, L. G., Royo, J. B., 2012. Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers. Tree Genet Genomes. 8, 1163-80.

Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R, Feugey, L, Poncet, C., 2016. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. BMC Plant Biology. pp. 16-130.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., et al., 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat. Genet. 42, 833-839.

Williams, K., Munkvold, J., & Sorrells, M. (2013). Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* L.). Euphytica. 190, 99-116.

Zhang, X. J., Wang, L. X., Chen, X. X., Liu, Y. L., Meng, R., Wang, Y. J., and Zhao, Z. Y., 2014. A and MdMYB1 allele-specific markers controlling apple (Malus x domestica Borkh.) skin colour and suitability for marker-assisted selection. Genet Mol Res. 13, 9103-14.

Zohary, D., Hopf, M., and Weiss, E., 2012. Domestication of plants in the old world. The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford: Oxford University Press.

Zude, M., 2003. Comparison of indices and multivariate models to non-destructively predict the fruit chlorophyll by means of visible spectrometry in apple fruit. Analytica Chimica Acta. 481, 119-126.

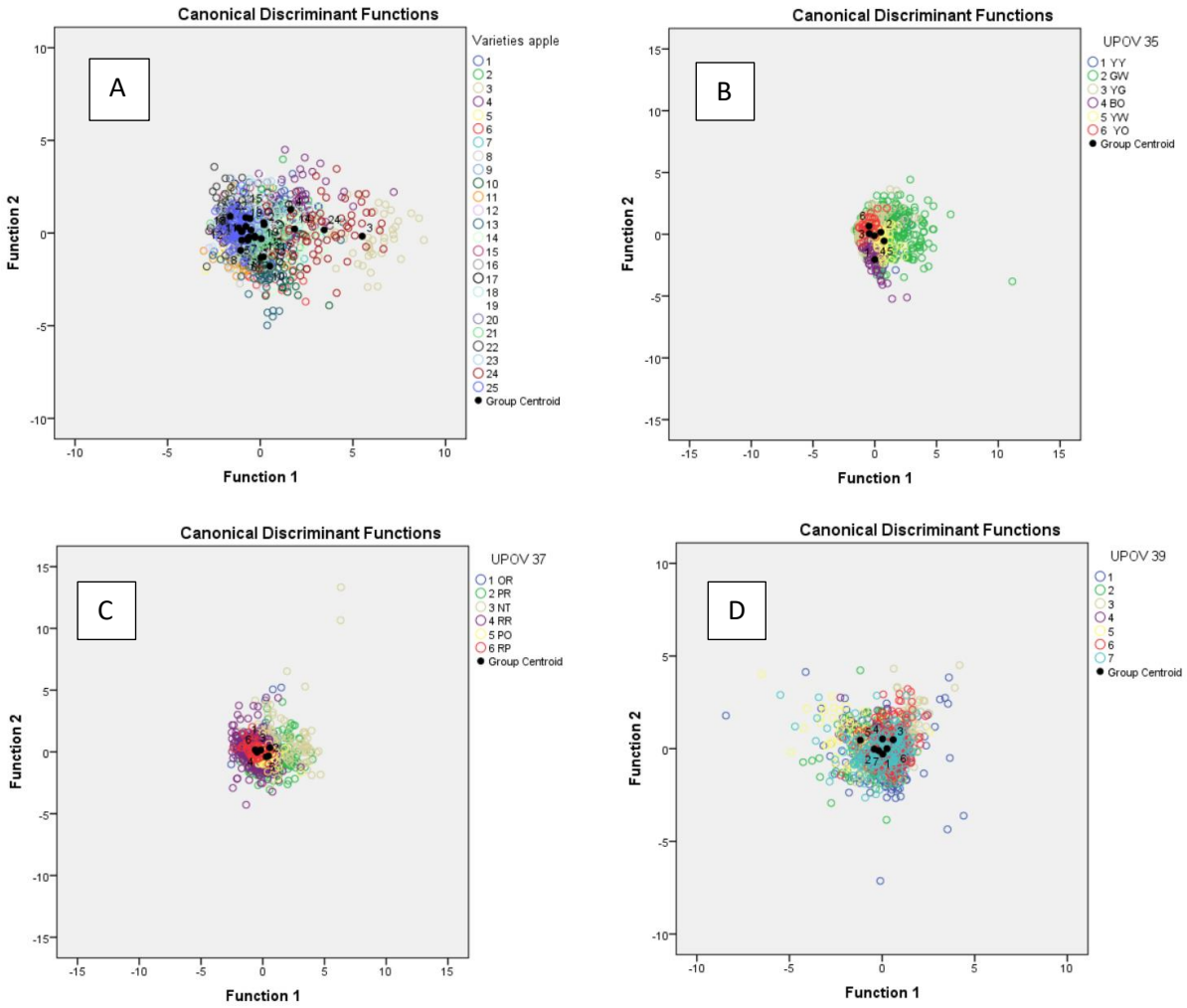
Supplementary material: Chapter 1 SM 1

Classification Results

	25 apple varieties	Predicted Group Membership																									Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
Cross-validated	ALPENLADREA	12,9	5,0	1,0	16,8	3,0	2,0	1,0	2,0	2,0	0,0	0,0	0,0	0,0	5,9	0,0	4,0	8,9	1,0	2,0	8,9	1,0	5,0	6,9	4,0	6,9	100,0
	ALPENLADRENGRANTE	13,3	21,7	0,0	5,0	1,7	6,7	8,3	1,7	3,3	1,7	1,7	1,7	3,3	8,3	0,0	0,0	3,3	3,3	1,7	3,3	0,0	5,0	5,0	0,0	0,0	100,0
	ARPONE	0,0	0,0	70,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,1	0,0	0,0	0,0	0,0	2,1	25,0	0,0	100,0
	BIACCALARRICA	19,7	2,8	1,4	47,9	0,0	1,4	2,8	0,0	0,0	0,0	0,0	0,0	4,2	8,5	0,0	0,0	2,8	0,0	0,0	0,0	0,0	1,4	1,4	5,6	0,0	100,0
	BIANCA DI ARTIZIO	3,3	1,7	0,0	0,0	35,0	5,0	1,7	1,7	5,0	0,0	10,0	5,0	5,0	1,7	0,0	5,0	1,7	1,7	0,0	5,0	1,7	3,3	3,3	0,0	3,3	100,0
	BIANCA DI ORGNANI	0,0	1,7	0,0	0,0	6,7	36,7	3,3	1,7	3,3	10,0	1,7	1,7	11,7	1,7	0,0	6,7	5,0	0,0	0,0	6,7	0,0	0,0	0,0	0,0	1,7	100,0
	BIONALOCALDO A'	13,3	5,0	0,0	0,0	0,0	0,0	11,7	0,0	3,3	11,7	0,0	0,0	8,3	1,7	0,0	1,7	6,7	5,0	1,7	5,0	5,0	0,0	10,0	1,7	8,3	100,0
	CALDDINA	2,6	0,0	0,0	0,0	0,0	13,2	0,0	28,9	0,0	5,3	10,5	2,6	5,3	0,0	0,0	10,5	0,0	0,0	0,0	7,9	2,6	2,6	5,3	0,0	2,6	100,0
	DALAIA	1,7	10,0	0,0	0,0	1,7	1,7	1,7	1,7	31,7	3,3	1,7	6,7	0,0	1,7	0,0	11,7	6,7	0,0	1,7	0,0	0,0	3,3	0,0	0,0	13,3	100,0
	DE FERRAU	0,0	0,0	0,0	0,0	0,0	12,5	4,2	2,1	0,0	60,4	0,0	0,0	4,2	0,0	0,0	6,3	6,3	0,0	0,0	0,0	2,1	0,0	0,0	0,0	2,1	100,0
	DE FERRU DE ARTIZZO	3,3	3,3	0,0	0,0	18,3	5,0	0,0	13,3	8,3	0,0	13,3	1,7	8,3	0,0	0,0	3,3	0,0	3,3	0,0	1,7	0,0	3,3	1,7	0,0	11,7	100,0
	DI CICCILIERI	5,0	0,0	0,0	0,0	1,7	0,0	0,0	5,0	10,0	1,7	1,7	8,3	0,0	1,7	0,0	3,3	3,3	3,3	5,0	21,7	3,3	15,0	6,7	0,0	3,3	100,0
	DI LUCILLO	3,3	1,7	1,7	0,0	8,3	13,3	1,7	3,3	1,7	10,0	0,0	0,0	31,7	3,3	0,0	0,0	5,0	0,0	1,7	1,7	0,0	3,3	6,7	1,7	0,0	100,0
	LACONE B	10,2	2,0	0,0	0,0	0,0	0,0	4,1	0,0	0,0	4,1	2,0	0,0	4,1	67,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	6,1	0,0	100,0
	LACONI D	17,6	2,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	5,9	2,9	2,9	0,0	0,0	0,0	8,8	5,9	0,0	14,7	0,0	11,8	17,6	0,0	8,8	100,0
	OMELA DI CICCILIERI	10,0	0,0	0,0	0,0	5,0	3,3	5,0	5,0	15,0	0,0	0,0	3,3	1,7	1,7	0,0	5,0	13,3	1,7	3,3	13,3	0,0	3,3	5,0	0,0	5,0	100,0
	OMLALI	13,9	0,0	0,0	0,0	1,4	4,2	4,2	0,0	11,1	1,4	0,0	1,4	1,4	1,4	0,0	2,8	22,2	8,3	2,8	8,3	9,7	0,0	0,0	0,0	5,6	100,0
	OXXU	11,4	0,0	0,0	1,4	5,7	2,9	5,7	1,4	2,9	2,9	2,9	0,0	8,6	0,0	1,4	2,9	8,6	8,6	0,0	11,4	2,9	5,7	11,0	0,0	2,9	100,0
	OZZU	5,0	1,7	0,0	1,7	3,3	0,0	5,0	0,0	3,3	1,7	1,7	1,7	1,7	5,0	0,0	5,0	1,7	0,0	6,7	13,3	1,7	8,3	18,3	0,0	13,3	100,0
	RANETTA	3,9	0,0	0,0	0,0	0,0	2,6	3,9	1,3	3,9	0,0	2,6	5,2	0,0	0,0	0,0	3,9	10,4	10,4	0,0	24,7	1,3	14,3	2,6	0,0	9,1	100,0
	RENETTA	11,7	1,7	0,0	0,0	0,0	1,7	6,7	0,0	5,0	5,0	0,0	3,3	3,3	0,0	0,0	3,3	26,7	3,3	0,0	10,0	6,7	3,3	1,7	0,0	6,7	100,0
	ROSA	0,0	0,0	0,0	0,0	1,3	1,3	0,0	2,6	2,6	0,0	2,6	2,6	0,0	0,0	1,3	0,0	1,3	0,0	3,9	14,3	0,0	46,8	6,5	0,0	13,0	100,0
	SANGIOVANNI ARRUBIA	12,1	3,4	0,0	1,7	0,0	0,0	5,2	0,0	5,2	0,0	0,0	0,0	1,7	1,7	1,7	1,7	5,2	1,7	13,8	8,6	0,0	8,6	27,6	0,0	0,0	100,0
	SIOVLADORE	3,3	1,7	31,7	11,7	0,0	5,0	0,0	0,0	0,0	6,7	0,0	0,0	0,0	13,3	0,0	0,0	1,7	0,0	0,0	0,0	0,0	0,0	0,0	25,0	0,0	100,0
	ZAZZARI	0,0	1,3	0,0	0,0	2,6	1,3	1,3	6,5	7,8	0,0	1,3	3,9	0,0	0,0	1,3	1,3	0,0	0,0	3,9	18,2	1,3	6,5	3,9	0,0	37,7	100,0

27,0% of cross-validated grouped cases correctly classified.

Supplementary material: Chapter 1 SM 2



SM 2. Scatter plot graph based on LDA analysis discrimination conducted considering the 25 varieties of *M. domestica* for their colorimetric parameters: (A) preliminary analysis, (B) UPOV 35 (ground colour), (C) UPOV 37 (over colour), (D) UPOV 39 (distribution colour).

Supplementary material: Chapter 1 SM 3

SM 3. Most discriminant list related to the VIP of the PLS-DA made based on UPOV 35 (ground colour) parameters for 3 classes GW (green/white); YG (yellow/green) and BO (brown/orange). The statistical significance was confirmed by a one way ANOVA t test: all the p value were found < to 0.005.

Discriminant variables	VIP
AspRatio	1,1479
Feret	1,13856
MBCRadius	1,13851
RFactor	1,10694
CHull	1,0875
Perim	1,07637
Compactness	1,05651
Roundness	1,04767
MaxR	1,03715
ArEquivD	1,019
Pixels	1,01566
Area	1,01491
PerEquivD	1,01491
Rectang	1,00985
ModRatio	1,00806
Circ	1,00723

CHAPTER 2: Seed morphometry is suitable for apple germplasm diversity-analyses.

1. Introduction

The Rosaceae family comprises about 3,000 taxa that include many genera of great importance for human nutritional and ornamental use (Hancock, 2008). Among these, the genus *Malus* comprises about 55 species, including the domestic apple (*Malus domestica* Borkh.), one of the most economically important fruit crops grown in temperate zones (Zohary et al., 2012).

M. domestica Borkh. domestication likely began in the Tian Shan Mountains in Central Asia (Harris et al., 2002; Cornille et al., 2012). This area contains multiple crop wild relatives (CWR) of domestic apples, such as *M. sieversii* (Ledeb.) M. Roem., which is fully interfertile with *M. domestica* (Zohary et al., 2012). Other important species of wild apples have genetically contributed to domestic apple, including *M. orientalis* Uglitzk. ex Juz., with a distribution range identified in the Caucasus, and *M. sylvestris* (L.) Mill., distributed primarily in Europe (Cornille et al., 2014).

Molecular marker studies of wild and domesticated apples have confirmed the diffusion of apple across the silk road from Central Asia, passing through Turkey towards Europe (Cornille et al., 2013a, 2013b, 2012; Velasco et al., 2010; Harris et al., 2002). Nevertheless, the origin of apple domestication remains partially unclear because of interfertility and self-incompatibility of *Malus* species, which can hybridize, thereby generating highly variable progenies (Zohary et al., 2012; Cornille et al., 2012; Velasco et al., 2010). There are over 10,000 varieties of apples worldwide, showing huge variability in their traits, especially pomological features such as the fruit size, skin colour and taste (Cornille et al., 2014, 2012; Harris et al., 2002).

Today, the number of modern commercial varieties has been reduced because of clonal selection and breeding programs, which used a small number of genotypes (Hokanson et al., 2001; Noiton and Alspach, 1996). For this reason, to maintain the greatest variation of alleles that can be exploited in breeding programs, several researchers have recommended protecting and preserve CWRs and old apple varieties (Liang et al., 2015; Way et al., 1990; Nnadozie et al., 2003).

Several genetic studies were conducted to investigate the origins of apple domestication and genetic diversity within the species or within the local germplasm (Urrestarazu et al., 2016, 2012; Cornille et al., 2012; Liang et al., 2015; Velasco et al., 2010).

Liang et al. (2015), by simple sequence repeats (SSRs) analysis, described the genetic diversity within a large number of apple varieties (belonging mainly to the Italian peninsula), with the

goal of identifying synonymy and homonymy (which are extremely difficult to detect through phenotypic traits) and exploring the genetic structure detectable in this large asset of accessions. Other authors, by seed image analysis techniques and the linear discriminant analysis (LDA), have investigated the diversity in cultivated species such as *Vitis vinifera* L. spp. *vinifera*, *Olea europaea* L., *Cucumis melo* L. and *Prunus domestica* L., (Orrù et al., 2015, 2013, 2012; Ucchesu et al., 2017, 2016, 2015; Sabato et al., 2015; Piras et al., 2016; Sarigu et al., 2017).

The main objectives of this work were to:

(1) build a database of seed morphological variables of apple cultivars, suitable for variety characterization;

(2) assess the phenotypic diversity of apples by morphological seed image analysis techniques and by LDA;

(3) compare our seed image analysis data with a genetic study previously conducted on the same varieties (Liang et al., 2015).

2. Materials and Methods

2.1. Apple germplasm varieties

In this work, we have investigated 42 apple varieties previously subjected to genetic analysis by Liang et al. (2015) (Table 1).

The fruits were harvested in the apple collection field of the Cadriano Experimental Station of the Department of Agricultural Sciences (University of Bologna) at full ripening. After removing the flesh, seeds were cleaned, washed and naturally air-dried in the laboratory of the Sardinian Germplasm Bank (BG-SAR), University of Cagliari.

To ensure the highest morphological variability among seeds and to compare morphometric results with genetic data, 10 fruits of each cultivar were harvested from the same tree previously used for the genetic analysis. To facilitate the presentation of results and sample grouping, each cultivar has been coded (Table 1).

2.2 Seeds image analysis

Digital image analysis is an innovative method of recent use that allows a high number of morphometric features of the seeds (Keefe and Draper, 1988).

This method gives several advantages such as low-cost analysis, non-destruction of the sample, analytical speed compared to conventional methods, even in the presence of a large amount of data and the ability to standardize the process making it interactive and easy to use (Chitra et al., 2016; Sandeep et al., 2013; Nikam and Kakatkar, 2013).

Moreover, this methodology can be applied to a large field of investigations such as the agronomic one, for example for identifying new cultivars or identifying possible synonyms and homonyms groups (Orrù et al., 2015, 2013, 2012; Ucchesu et al., 2017, 2016, 2015; Sabato et al., 2015; Piras et al., 2016; Sarigu et al., 2017).

The digital images of all seeds were acquired using a flatbed scanner (Epson Perfection V550), with a digital resolution of 800 dpi for a scanning area not exceeding 1024 × 1024 pixels (Bacchetta et al., 2008). The images were processed and analysed using the open source software ImageJ v. 1.49. The Particles8 plugin (Landini, 2006) was used to measure 26 seed morphometric variables (Table 2, Fig. 1).

To increase the number of discriminant parameters, 80 elliptic Fourier descriptors (EFDs) of the seed's contour shape were computed using the open source SHAPE software (Iwata and Ukai, 2002). This method reports the boundary of the seed projection as an array of complex numbers, which correspond to the pixel positions on the edge of the seed. Therefore, from the seed apex, defined as the starting point in a Cartesian system, chain codes are generated (Kuhl and Giardina, 1982). In total, 26 morphometric variables were measured.

2.3. Statistical analysis

Morphometric data were processed and statistically analysed by applying a stepwise LDA method using IBM SPSS (Statistical Package for Social Science, release 16.0). The LDA method is commonly used to classify or identify unknown groups characterized by quantitative and qualitative variables (Sugiyama, 2007), finding the combination of predictor variables with the aim of simultaneously minimizing the within-class distance and maximizing the between-class distance, thus achieving maximum class discrimination (Hastie et al., 2002; Holden et al., 2011; Rencher and Christensen, 2012; Kuhn and Johnson, 2013).

This method allowed us to reduce the dataset size without losing important information and to classify statistical cases into groups (Fukunaga, 2013; Duda et al., 2012; Hastie et al., 2002; Holden et al., 2011; Sugiyama, 2007). Using three statistical variables (Tolerance, *F*-to-enter and *F*-to-remove), the stepwise method identifies and selects the most statistically significant variables among the 26 measured on each seed. The tolerance value, *F*-to-enter and *F*-to-remove values define the power of each variable and their role in the model. At each step, the variable with the largest *F*-to-enter value that exceeds the entry criterion chosen ($F \geq 3.84$) was added to the model. All variables with a value below 3.84 were excluded from the analysis (Venora et al., 2009). Finally, a cross-validation procedure was applied to verify the performance of the identification system, testing individual unknown cases and classifying them based on all others. This procedure, also referred to as rotation estimation (Picard and Cook, 1984; Kohavi, 1995), was used to evaluate the performance and to validate any

implemented classifier. Here we applied the leave one out cross validation (LOOCV) procedure (SPSS 2006).

Table 1. Varieties of *M. domestica* analysed in this study.

Code	Origin	Variety name	N° Seeds	Code	Origin	Variety name	N° Seeds
C1	ITA	ABBONDANZA	81	C22	ITA	MELA GIALLA 1	105
C2	TURK	AMASYA	118	C23	ITA	MELA ROZZA	112
C3	ITA	ANNURCA	80	C24	ITA	MELA TOSTA	60
C4	ITA	APPIA (RT)	93	C25	ITA	MELO FERRO (PD)	34
C5	ITA	APPIONA	80	C26	ITA	OXIU	94
C6	NLD	BELLA DI BOSKOOP	17	C27	ITA	PAOLUCCIA (VT)	100
C7	ITA	BELLA DEL GIARDINO	100	C28	ITA	PARADISA	69
C8	ITA	BOURAS	86	C29	ITA	PUMA TENERELLA	97
C9	ITA	CADDINA	54	C30	FRA	RAMBOUR FRANK	96
C10	ITA	CAVICCHIO	72	C31	USA	RED CHIEF	86
C11	GER	CLIVIA	99	C32	NLD	RENETTA ANANAS	91
C12	ITA	DURELLO	97	C33	ITA	RENETTA DI CHAMPAGNE	99
C13	ITA	EPPIA	57	C34	FRA	REINETTE FRANCHE (M.REGINA)	64
C14	ITA	FIOR DI CASSIA	77	C35	ITA	ROSA D'OSTA	35
C15	ITA	FRANCESCA (MI)	121	C36	ITA	RUNSE'	133
C16	ITA	GELATA	92	C37	ITA	SANT'AGOSTINO	76
C17	ITA/FRA	GRENOBLE (TO)	86	C38	ITA	SEL IDICE 3	51
C18	ITA	LIMONCELLA	113	C39	ITA	SEL IDICE 4	100
C19	ITA	LIMONCELLA URIDDU	89	C40	ITA	VERGINELLA	95
C20	ITA	LOSA D'GIAVENO	93	C41	ITA	VIGNONE	75
C21	ITA	MARCON (TN)	89	C42	ITA	PUMA OLIO	95

Fig. 1. Variety *M. domestica* utilised in this study.

Table 2. List of the 26 morphometric seed features measured on endocarps and calculated by Particles8 plugins from ImageJ v. 1.49.

Parameter	Description
Perim	Perimeter, calculated from the centres of the boundary pixels
Area	Area inside the polygon defined by the perimeter
Pixels	Number of pixels forming the seed image
MinR	Radius of the inscribed circle centred at the middle of mass
MaxR	Radius of the enclosing circle centred at the middle of mass
Feret	Larges taxis length
Breadth	Largest axis perpendicular to the Feret
CHull	Convex hull or convex polygon calculated from pixel centres
CArea	Area of the convex hull polygon
MBCRadius	Radius of the minimal bounding circle
AspRatio	Aspect ratio = Feret/Breadth
Circ	Circularity = $4 \cdot \pi \cdot \text{Area}/\text{Perimeter}^2$
Roundness	Roundness = $4 \cdot \text{Area}/(\pi \cdot \text{Feret}^2)$
ArEquivD	Area equivalent diameter = $\sqrt{((4/\pi) \cdot \text{Area})}$
PerEquivD	Perimeter equivalent diameter = Area/π
EquivEllAr	Equivalent ellipse area = $(\pi \cdot \text{Feret} \cdot \text{Breadth})/4$
Compactness	Compactness = $\sqrt{((4/\pi) \cdot \text{Area})}/\text{Feret}$
Solidity	Solidity = Area/Convex_Area
Concavity	Concavity = Convex_Area-Area
Convexity	Convexity = Convex_hull/Perimeter
Shape	Shape = $\text{Perimeter}^2/\text{Area}$
RFactor	RFactor = Convex_Hull /$(\text{Feret} \cdot \pi)$
ModRatio	Modification ratio = $(2 \cdot \text{MinR})/\text{Feret}$
Sphericity	Sphericity = MinR/MaxR
ArBBox	Area of the bounding box along the feret diameter = Feret·Breadth
Rectang	Rectangularity = Area/ArBBox

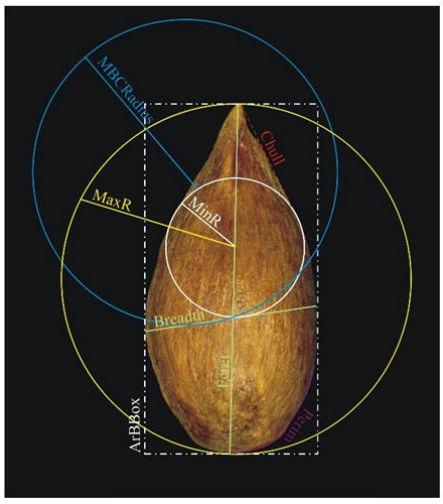


Fig. 2. Graphical representation of principal morphometric variables measured on each apple seed (refer to Table 2).

3. Results

A first comparison among the 42 apple varieties was executed, achieving a 38.8% cross-validation for correct identifications. This first analysis showed a high classification for some varieties, including ‘Bella del Giardino’ (70%), ‘Amasya’ (60.1%), ‘Paradisa’ (59.4%) and ‘Durello di Forlì’ (54.6%) (Table 3).

Seeds from the variety ‘Eppia’ showed the lowest discriminant characteristics, with a value of only 14% (Table 3), showing a degree of similarity with the ‘Francesca’ variety of 19.2%. Other varieties that showed a low level of discrimination were ‘Red Chief’ (16.2%) and ‘Cavicchio di Levizzano’ (18%) (Table 3).

By LDA analysis, the 42 apple varieties could be subdivided into two large groups labelled ‘A’ and ‘B’ (Fig. 3). The comparison between these two groups showed a classification of 100% and 91.6%, respectively (Table 4). No subgroups were identified in group ‘A’ and ‘B’. Convexity, solidity and concavity were identified as the most discriminant variables (Table 5).

With the aim of identifying possible similarities, we compared groups ‘A’ and ‘B’ with the groups ‘G1’ and ‘G2’ reported by Liang et al. (2015) (Table 3). In a preliminary analysis, the ‘G1’ and ‘G2’ groups were individually included in the database as unknown groups and compared by LDA with group ‘A’ and ‘B’. In the LDA analysis in which the ‘G1’ group was inserted as an unknown, we detected a correlation with groups ‘A’ (69.63%) and ‘B’ (30.36%) (Table 6). In addition, the same analysis was conducted by inserting group ‘G2’ as an unknown. In this case, the LDA showed a correlation with group ‘A’ of 85.26% and with group ‘B’ of 14.73% (Table 6).

Moreover, a comparison between the groups reported by Liang et al. (2015) using LDA analysis discriminated the two groups with a percentage of 27.6% for ‘G1’ and 91.8% for ‘G2’ (Table 7). A further analysis was conducted to verify the existence of subgroups within the ‘G1’ and ‘G2’ groups, as had been identified by Liang et al. (2015). As a result of this analysis, subgroups ‘C’ and ‘D’ were identified within group 'G1’ (Fig. 3), and subgroups ‘E’ and ‘F’ within the group ‘G2’ (Fig. 4).

In agreement with the genetic analysis (Liang et al., 2015), that included ‘Eppia’ and ‘Francesca’ varieties in ‘G2’ (‘G2.U’ subgroup), the LDA method also classified these two varieties within ‘G2’ (Table 8). However, the ‘Paradisa’ variety, which was assigned to ‘G1’ by the genetic analysis (‘G1.U’ subgroup), was assigned to ‘G2’ by the LDA method (Table 8).

Table 3. Apple varieties grouping and discriminant percentages determined by LDA using either our seed image data or the genetic data of Liang et al. (2015).

Code	Origin	Variety name	LDA analysis		SSR markers (Liang et al., 2015)	
			Group	Discriminant percentage (%)	Group	Soubgroub
C1	ITA	ABBONDANZA	A	49,3	G1	G1.1
C2	TURK	AMASYA	A	60,1	G2	G2.1
C3	ITA	ANNURCA	A	50	G2	G2.1
C4	ITA	APPIA (RT)	A	41,9	G2	G2.1
C5	ITA	APPIONA	A	27,5	G2	G2.1
C6	NLD	BELLA DI BOSKOOP	A	11,7	-	-
C7	ITA	BELLA DEL GIARDINO	A	70	G1	G1.1
C8	ITA	BOURAS	A	39,5	-	-
C9	ITA	CADDINA	A	40,7	-	-
C10	ITA	CAVICCHIO DI LEVIZZANO	A	18	G1	G1.1
C11	ITA	CLIVIA	A	23,2	G1	G1.2
C12	ITA	DURELLO DI FORLI'	A	54,6	G2	G2.2
C13	ITA	EPPIA	A	14	G2	G2.U
C14	ITA	FIOR DI CASSIA	A	42,8	G2	G2.2
C15	ITA	FRANCESCA (MI)	A	19,8	G2	G2.U
C16	ITA	GELATA	A	46,7	G2	G2.1
C17	ITA/FRA	GRENOBLE (TO)	A	26,7	G1	G1.2
C18	ITA	LIMONCELLA	A	25,6	G2	G2.1
C19	ITA	LIMONCELLA URIDDU	A	39,6	G2	G2.1
C20	ITA	LOSA D' GIAVENO	A	35,4	G1	G1.2
C21	ITA	MARCON (TN)	A	42,6	-	-
C22	ITA	MELA GIALLA 1	A	43,3	G2	G2.1
C23	ITA	MELA ROZZA	A	33	G2	G2.1
C24	ITA	MELA TOSTA	A	23,3	G2	G2.2
C25	ITA	MELO FERRO (PD)	A	50	G2	G2.1
C26	ITA	OXIU	A	37,2	G2	G2.2
C27	ITA	PAOLUCCIA (VT)	A	33	G2	G2.2
C28	ITA	PARADISA	A	59,4	G1	G1.U
C29	ITA	PUMA TENERELLA	A	19,5	G2	G2.2
C30	FRA	RAMBOUR FRANK	A	26	G2	G2.1
C31	USA	RED CHIEF	A	16,2	G1	G1.1
C32	NLD	RENETTA ANANAS	B	34	G1	G1.1
C33	ITA	RENETTA DI CHAMPAGNE	B	41,4	G1	G1.1
C34	FRA	REINETTE FRANCHE (M.REGINA)	B	42,1	-	-
C35	ITA	ROSA D'OSTA	B	20	-	-
C36	ITA	RUNSE'	B	54,1	G1	G1.2
C37	ITA	SANT'AGOSTINO	B	53,9	G1	G1.1
C38	ITA	SEL IDICE 3	B	41,1	-	-
C39	ITA	SEL IDICE 4	B	54	G2	G2.1
C40	ITA	VERGINELLA	A	47,3	G2	G2.1
C41	ITA	VIGNONE	B	38,6	G2	G2.2
C42	ITA	PUMA OLIO	B	40	G2	G2.1

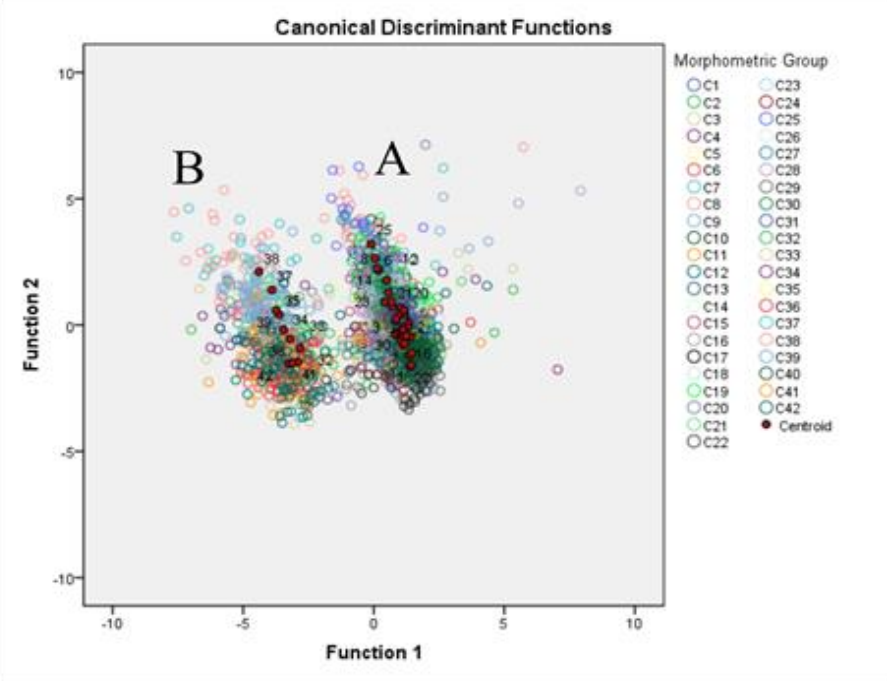


Fig. 3. Stepwise Linear Discriminant Analysis (LDA) of the 42 studied *M. domestica* varieties. Small differently coloured points represent data of a single seed. Red points indicate the average value (centroid). The variance of function 1 is 100%, and that of function 2 is 91.6%, with the remaining 8.4% distributed amongst the third function (not shown). The two groups generated by the LDA are labelled 'A' and 'B'.

Table 4. Correct classification percentages among group 'A' and 'B'. The number of seeds analysed is indicated in brackets.

Groups	A	B	Total
A	100 (2801)	0,03 (1)	100 (2802)
B	8,4 (69)	91,6 (750)	100 (819)
Cross-validated (%)			**98,10%**

Table 5. Ranking of the three best discriminant morphometric variables selected and used by the Stepwise Linear Discriminant Analysis for groups 'A' and 'B'.

Parameters	Tolerance	*F* to Remove	Wilks' Lambda
Convexity	0,091	2790,688	0,328
Solidity	0,011	1042,116	0,238
Concavity	0,175	793,025	0,225

Table 6. Correct classification percentages among group 'A' and 'B'. The number of seeds analysed is indicate in brackets.

Groups	A	B	Total
A	99,88 (2704)	0,11 (3)	100 (2705)
B	8,53 (78)	91,46 (836)	100 (914)
G1	69,63 (555)	30,36 (242)	100 (797)
G2	85,26 (1435)	14,73 (248)	100 (1638)
Cross-validated (%)			97,80%

Table 7. Correct classification percentages among group ‘G1’ and ‘G2’. The number of seeds analysed is indicated in brackets.

Groups	G1	G2	Total
G1	27,6 (259)	72,4 (681)	100 (940)
G2	8,2 (189)	91,8 (2109)	100 (2109)
Cross-validated (%)			**73,1% (3049)**

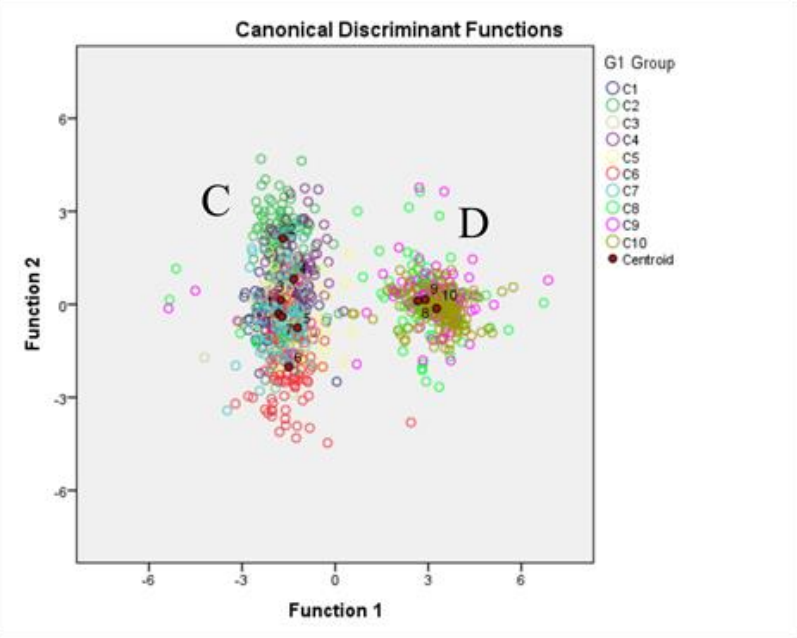


Fig. 3. Stepwise Linear Discriminant Analysis of group ‘G1’of *M. domestica.* Small differently coloured points represent data of a single seed. Red points indicate the average value (centroid). The two subgroups generated by the LDA are labelled ‘C’ and ‘D’.

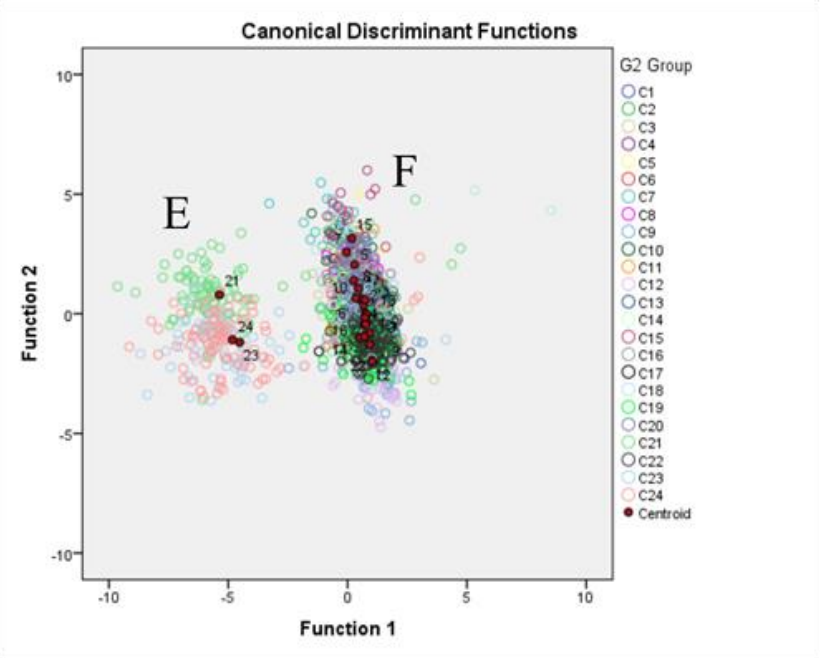


Fig. 4. Scatter plot graph based on LDA analysis discrimination group 'G2' of *M. domestica.* Small differently coloured points represent data of a single seed. Red points indicate the average value (centroid). The two subgroups generated by the LDA are labelled 'E' and 'F'.

Table 8. Classification percentage between the apple seed lot from subgroup 'G2. U', considered as an unknown group, and group 'G1' and 'G2'. The number of seeds analysed is indicated in brackets.

Groups	G1	G2	Total
G1	27,8 (261)	72,2 (679)	100 (940)
G2	8,5 (185)	91,5 (1,993)	100 (2,178)
Eppia	3,5 (2)	96,5 (55)	100 (57)
Francesca	12,5 (15)	87,5 (105)	100 (120)
Paradisa	5,8 (4)	94,2 (65)	100 (69)
Cross-validated (%)			**72,20%**

4. Discussion

In this work, we report the findings of an LDA based on seed morphometric parameters from 42 apple varieties. Our aim was to investigate the phenotypic diversity among these varieties and to compare this data with available genetic data (Liang et al., 2015). The first comparative analysis identified some varieties with high percentages of discrimination (e.g. 'Bella del Giardino'), as well as others with low separation (e.g., 'Eppia'). This high variability between *Malus* varieties can be explained by the origin of its genotypes; in fact, the currently cultivated varieties of apples are the result of a complex and long genetic history, in which natural evolution and human selection, presumably since prehistoric times, have contributed to shaping the extant species (Angelini, 2009; Cornille et al., 2012).

The variable degree of seed discrimination might be explained by the fact that apple fruit and seed development are strictly related to a complex network of endogenous and exogenous factors (Eccher et al., 2014), such as hormones (Gillaspy et al., 1993) and gene expression regulation by MADS-box genes (Sung et al., 2000; Yao et al., 2001). This developmental cross-talk involves a large number of genes and biosynthetic pathways, whose effects on the respective phenotypes are largely unknown (Eccher et al., 2014).

The phenotypic diversity among the tested seeds is presumably the result of the activation or repression of a large number of genes (and related allelic variants), with potential contributions from environmental factors.

In addition, in this case, the LDA analysis proved to be a valid method for discriminating groups and subgroups, as previously demonstrated by the genetic analysis conducted by Liang et al. (2015).

The comparison between the 'A' and 'B' groups show a clear differentiation, with a higher classification for group 'B'. This might be because group 'B' is composed of a greater number of varieties sharing common phenotypic traits.

By LDA, the 'Gelata', 'Verginella', 'Limoncella' and 'Limoncella Uriddu' varieties have been classified into the same group ('A'), confirming the results obtained by SSR analysis in which these varieties presented the same electrophoretic pattern (Liang et al., 2015). These results, therefore, suggest that the morphological and molecular approaches produce similar findings.

Moreover, by comparing the principal groups 'G1' and 'G2' by LDA, we detected a low classification for group 'G1', which tends to be confused with 'G2'. This is probably because 'G2' is larger than 'G1' and is composed mainly of varieties classified with a high percentage. In addition, as evidenced by genetic analysis, the 'G2' group is composed mostly of Italian varieties, whereas the international standard varieties are assigned to 'G1'.

Both approaches evidenced the existence of two principal groups, which, in turn, are divided into two additional sub-groups. The composition of the sub-groups, identified through the LDA approach, suggests a partial correlation of varieties with those identified by genetic analysis. Finally, the comparison of the "Francesca", "Eppia" and "Paradisa" varieties, which were assigned by genetic analysis as ‘G1.U’ and ‘G2.U’, have obtained a total allocation in the ‘G2’ group. This is because these three varieties belong to group ‘A’, which is mostly composed of varieties included in the ‘G2’ group.

These differences can be partly explained by the fact that the molecular markers (SSRs) used by Liang et al. (2015) identified a large number of polymorphisms in specific loci, whereas the morphometric analysis highlights a phenotypic variability that cannot be directly compared with those detected by the SSR analysis (Orrù et al., 2012). In fact, genes that are responsible for this variability are located in genomic positions that are not necessarily covered by the analysed SSRs. These observations could be used to understand the relationships between the morphological results and the molecular analyses reported by Liang et al. (2015). The different basis of the discrimination in groups made by genetic and morphological approaches could explain the few discrepancies observed in the present study.

Liang et al. (2015) identified the existence of the two groups, although these were little differentiated. The additional division into subgroups was determined by a nested approach, as described by Urrestarazu et al. (2012). Another recent paper that analysed the genetic diversity at the European level identified the existence of three main groups: varieties coming from the North East of Europe, those from the Western countries and mainly Mediterranean varieties collected in Italy and Spain (Urrestarazu et al., 2016). In addition, in this case, the three groups were linked with the geographical regions of origin, even if they were little differentiated. This would reflect a situation whereby the varieties from a given region were more frequently derived from crosses between parental varieties from the same region, rather than from varieties found elsewhere. Most of the samples analysed in this study belong to the Mediterranean group, and would, therefore, be expected to group together if compared to cultivars of a different origin; however, the number of non-Mediterranean genotypes included in our study was probably insufficient to allow the detection of an external group based on morphologic traits. Moreover, the migration of plant material associated with human movement, together with hundreds of years of empirical selection, have likely caused a significant gene flow across Europe (Urrestarazu et al., 2016), making it difficult to identify well-defined phenotypical differences between the main groups and subgroups of extant varieties.

5. Conclusion

In this work, the use of digital seed image analysis applied to investigate phenotypic characteristics of the apple samples has allowed to obtain valuable information that can be compared with genetic analysis.

Overall, this work has allowed the creation of a morphometric seeds database of Italian apple varieties, suitable for the characterization and conservation of germplasm. Moreover, a global correlation has been shown especially in the subdivision of two main groups and respective subgroups by both phenotypic and genotypic approaches. Furthermore, we provided evidence that seed morphometric and genetic analyses support the characterization of biodiversity among varieties of fruit tree species. These findings will be useful for efforts to preserve the genetic diversity of cultivated plant species and to enlarge the genetic component of existing fruit breeding programs.

References

Angelini, R., 2009. Il melo. 1st Ed. Bayer CropScience, Italy.

Bacchetta, G., Grillo, O., Mattana, E., Venora, G., 2008. Morpho-colorimetric characterization by image analysis to identify diaspores of wild plant species. Flora. 203, 669-682.

Chitra, H.S.H., Suguna, S., and Sujatha, S.N., 2016. A survey on image analysis techniques in agricultural product. Indian Journal of Science and Technology. 9, 13.

Cornille, A., Gladieux, P., Smulders, M.J.M., Roldan-Ruiz, I., Laurens, F., Le Cam, B., Nersesyan, A., et al., 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genetics 8:e1002703.

Cornille, A., Giraud, T., Bellard, C., Tellier, A., Le Cam, B., Smulders, M.J.M., Kleinschmit, J. et al., 2013a. Post-glacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. Mol. Ecol. 22, 2249-2263.

Cornille, A., Gladieux, P., Giraud, T., 2013b. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. Evol. Appl. 6, 737-748.

Cornille, A., Giraud, T., Smulders, M.J., Rolan-Ruiz, I., Gladieux, P., 2014. The domestication and evolutionary ecology of apples. Trends Genet. 30, 57-65.

Duda, R.O., Hart, P.E., & Stork, D.G., 2012. Pattern classification. John Wiley and Sons.

Eccher, G., Ferrero, S., Populin, F., Colombo, L., and Botton, A., 2014. Apple (*Malus domestica* L. Borkh) as an emerging model for fruit development. Plant Biosyst. 148, 157-168.

Fukunaga, K., 2013. Introduction to statistical pattern recognition. Academic press.

Gillaspy, G., Ben-David, H., and Gruissem, W., 1993. Fruits: a developmental perspective. The Plant Cell. 5, 1439-1451.

Hancock, J. F., 2008. Temperate fruit crop breeding: germplasm to genomics. Springer Science and Business Media.

Harris, S.A., Robinson, J.P., Juniper, B.E., 2002. Genetic clues to the origin of the apple. Trends Genet. 18, 426-430.

Hastie, T., Tibshirani, R., and Friedman, J., 2002. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Biometrics.

Holden, J.E., Finch, W.H., Kelley, K., 2011. A comparison of two-group classification methods. Educ. Psychol. Meas.71, 870-90.

Hokanson, S.C., Lamboy, W.F., Szewc-McFadden, A.K., McFerson, J.R., 2001. Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids. Euphytica. 118, 281-294.

Iwata, H, Ukai, Y., 2002. SHAPE: A computer program package for quantitative evaluation of biological shapes based on Elliptic Fourier Descriptors. J. Hered. 93, 384-385.

Keefe, P.D., and S.R., Draper, 1988. An automated machine vision system for the morphometry of new cultivars and genebank accessions. Plant Var. Seeds. 1,1-11.

Korban, S., 1986. Interspecific hybridization in *Malus*. HortScience. 21, 41-48.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai.14, 1137-1145.

Kuhn, M., Johnson, K., 2013. Discriminant analysis and other linear classification models. Applied Predictive Modeling New York: Springer. pp. 275-328.

Kuhl, F.P., Giardina, C.R., 1982. Elliptic Fourier features of a closed contour. Comput. Graph. 18, 259-278.

Landini, G., 2006. Quantitative analysis of the epithelial lining architecture in radicular cysts and odontogenic keratocysts. Head Face Med. 2, 1-9.

Liang, W., Dondini, L., De Franceschi, P., Paris, R., Sansavini, S., Tartarini, S., 2015. Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. Plant Mol. Biol. Report. 33, 458-473.

Nikam, M.S.V., and Kakatkar, M.N., 2013. SEED PROPERTY MEASUREMENT WITH IMAGE ANALYSIS. 4, 18-22.

Nnadozie, K., Littington, R., Bruch, C., Bass, S., King, S., 2003. African perspective on genetic resources, environmental law institute. A Handbook on Laws, Policies and Institutions. African Union/ Enviromental Law Institute.

Noiton, D.A.M., Alspach, P.A., 1996. Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J. Am. Soc. Hortic. Sci. 121, 773-782.

Orrù, M., Grillo, O., Venora, G., Bacchetta, G., 2012. Computer vision as a method complementary to molecular analysis: Grapevine cultivar seeds case study. C. R. Biol. 335, 602-615.

Orrù, M., Grillo, O., Lovicu, G., Venora, G., Bacchetta, G., 2013. Morphological characterisation of *Vitis vinifera* L. seeds by image analysis and comparison with archaeological remains. Veg. Hist. Archaeobot. 22, 231-242.

Orrù, M., Grillo, O., Venora, G., Bacchetta, G., 2015. Seed morpho-colorimetric analysis by computer vision: a helpful tool to identify grapevine (*Vitis vinifera* L.) cultivars. Aust. J. Grape Wine Res. 21, 508-519.

Picard, R.R., and Cook, R.D., 1984. Cross-validation of regression models. J. Am. Stat. Assoc. 79, 575-583.

Pinna, M.S., Grillo, O., Mattana, E., Cañadas, E.M., Bacchetta, G., 2014. Inter-and intraspecific morphometric variability in *Juniperus* L. seeds (Cupressaceae). Syst. Biodivers. 12, 211-223.

Piras, F., Grillo, O., Venora, G., Lovicu, G., Campus, M., Bacchetta, G., 2016. Effectiveness of a computer vision technique in the characterization of wild and farmed olives. Comput. Electron. Agric. 122, 86-93.

Rencher, A.C., Christensen, W.F., 2012. Methods of multivariate analysis. New York: John Wiley. p. 800.

Robinson, J.P., Harris, S.A., and Juniper, B.E., 2001. Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant. Syst. Evol. 226, 35-58.

Rohrer, G.A., Alexander, L.J., and Beattie, C.W., 1994. Mapping the β subunit of follicle stimulating hormone (FSHB) in the porcine genome. Mamm. Genome. 5, 315-317.

Sabato, D., Esteres, C., Grillo, O., Bacchetta, G., 2015. Seeds morpho-colourimetric analysis as complementary method to molecular characterization of melon diversity. Sci. Hort. 192, 441-452.

Sandeep, V.V., Kanaka, D.K., and Keshavulu, K., 2013. Seed image analysis: its applications in seed science research. Int Res J Agr Sci. 1, 30-36.

Sarigu, M., Grillo, O., Lo Bianco, M., Ucchesu, M., D'hallewin, G., Venora, G., Bacchetta, G., 2017. Phenotypic identification of plum varieties (*Prunus domestica* L.) by endocarps morpho-colorimetric and textural descriptors. Comput. Electron. Agric. 136, 25-30.

Sugiyama, M., 2007. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. J. Mach. Learn. Res. 8, 1027-1061.

Sung, S.K., Yu, G.H., Nam, J., Jeong, D.H., and An, G., 2000. Developmentally regulated expression of two MADS-box genes, MdMADS3 and MdMADS4, in the morphogenesis of flower buds and fruits in apple. Planta. 210, 519-528.

Ucchesu, M., Orrù, M., Grillo, O., Venora, G., Usai, A., Serreli, P.F., Bacchetta, G., 2015. Earliest evidence of a primitive cultivar of *Vitis vinifera* L. during the Bronze Age in Sardinia (Italy). Veg. Hist. Archaeobot. 24, 587-600.

Ucchesu, M., Orrù, M., Grillo, O., Venora, G., Paglietti, G., Ardu, A., Bacchetta, G., 2016. Predictive Method for Correct Identification of Archaeological Charred Grape Seeds: Support for Advances in Knowledge of Grape Domestication Process. PloS One. 11, 2 -e0149814.

Ucchesu, M., Sarigu, M., Del Vais, C., Sanna, I., d'Hallewin, G., Grillo, O., Bacchetta, G., 2017. First finds of *Prunus domestica* L. in Italy from the Phoenician and Punic periods (6th–2nd centuries BC). Veg. Hist. Archaeobot. 26, 539-549.

Urrestarazu, J., Miranda, C., Santesteban, L.G., Royo, J.B., 2012. Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers. Tree Genet Genomes. 8, 1163-1180.

Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R, Feugey, L, Poncet, C., 2016. Analysis of the genetic diversity and structure across a wide range of germplasm r eveals prominent gene flow in apple at the European level. Plant Biology. 16, 130.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., et al., 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat. Genet. 42, 833-839.

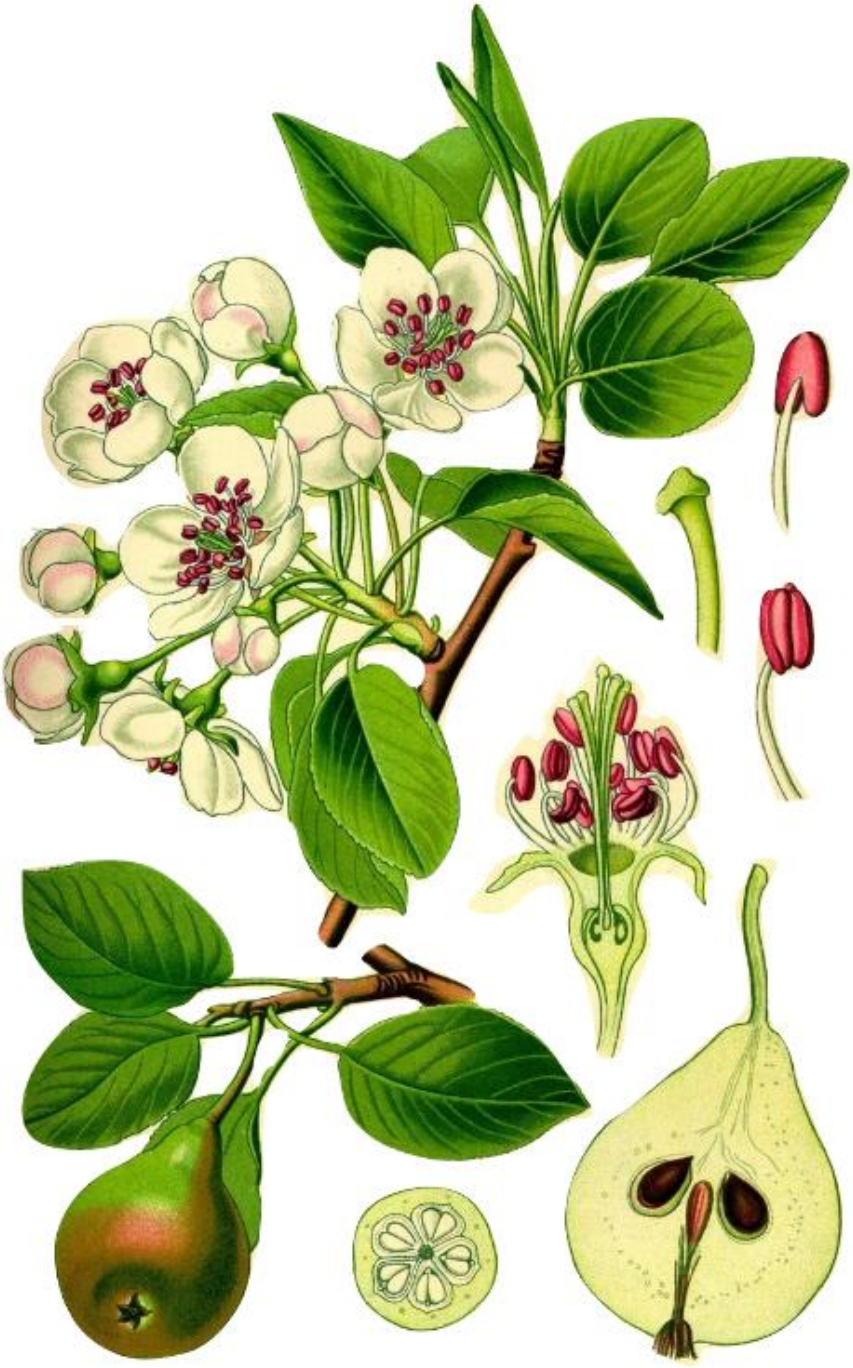
Venora, G., Grillo, O., Ravalli, C., Cremonini, R., 2009. Identification of Italian landraces of bean (*Phaseolus vulgaris* L.) using an image analysis system. Sci. Hortic. 121, 410-418.

Way, R.D., Aldwinckle, H.S., Lamb, R.C., Rejman, A., Sansavini, S., Shen, T., Watkins, R., Westwood, M.N., Yoshida, Y., 1990. Apples (*Malus*). In: J.N. Moore and J.R. Ballington (eds.). Genetic resources of temperate fruit and nut crops. Intl. Soc. Hort. Sci. 1, 5-62.

Yao, J.L., Dong, Y.H., and Morris, B.A., 2001. Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. Proc. Natl. Acad. Sci. 98, 130-1311.

Zohary, D., Hopf, M., and Weiss E., 2012. Domestication of plants in the old world. The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford: Oxford University Press.

The species *Pyrus communis* L.:



1. Botanical description

The *Pyrus* genus belongs to the family of the Rosaceae subfamily Pomidae.

The flower, gathered in a bunch of corymbiform of 5-15 flowers, is hermaphrodite with about 20 stamens, 5 petals of generally white color, sometimes rosé; The chalice is persistent in Western species, while it is generally fallen in the eastern species; The peduncle, which is generally elongated and curved, may be more or less fleshy in western species, not fleshy in the eastern ones.

The *Pyrus* genus is characterized by widespread autoincompatibility and factorial incompatibility. The fruit is a boss, also called a fake fruit because the part of the building is made up of the flower recipe. The pulp is predominantly white-cream, but it can also be of red colour more or less clear.

The simple leaves have varying shapes and sizes, are alternate with 2/5, fillets, crunchy, with crinkled edges in Eurasian species, clenched at a sharp angle or crenated silk in the oriental species, sometimes wavy. The tree, depending on the species, is a thorny bush, or as in cultivated varieties, a tall stem, pyramidal (up to 20 m and upwards), (Pignatti, 1982).

2. Origin

Pear belongs to rosaceous family, being a close "cousin" of the apple, but with some particularities that make this fruit special with a delicate flavour (Silva et al., 2014).

Common pears are thought to have developed from plants growing natively in Europe and Northern Asia. They were popular among the ancient Greeks and Romans, although they were used in drinks and cooked dishes rather than eaten raw, since the ancient varieties of pear were less palatable raw. However, through breeding, pears were made safe to eat raw and became a popular dessert fruit. They were brought to America in 1797, and continue to be popular worldwide today (Ferreira et al., 2002).

Most botanists believe that the *Pyrus* genus has differentiated in the Tertiary period, in the mountainous territory of present-day western China; from there it would have been dispersed both east and west, adapting to the different conditions of climate and territory, differentiating the species currently known (Fideghelli, 2007).

There are two domestication centers and primary origin of the genus *Pyrus*: the first is located in China, the second located in Asia Minor to the Middle East, in the Caucasus mountains, and a third secondary centre located in Central Asia (Zukovski, 1962).

Vavilov has identified two centers of primary origin:

A) China, where the species *P. pyrifolia*, *P. ussuriensis* and *P. calleryana* are grown;

B) Middle East (Caucasus, Asia Minor), the center of primary origin of the *P. communis*.

A third secondary centre (C) was also reported in Central Asia (northwestern India, Afghanistan, Tajikistan, Uzbekistan and the Tianshan province of China) (Vavilov, 1992), where *P.communis* is present hybridized with *P.* x *bretschneideri*.

In its evolution, *P. communis* hybridized with several species in Asia Minor and Europe:

P. elaeagrifolia, P. salicifolia, P. syriaca, P. nivalis, P. caucasia.

Of the species *P. communis* there is the pyramidal botanic form or variety present also in Southern Mediterranean Italy, which Linneo considered the wild form from which the cultivated varieties would have evolved.

In addition, cultivated varieties of the Eastern Peruvian are probably derived from natural hybridization among the various species, in particular, *P. pyrifolia*, *P. ussuriensis* and *P. betulaefolia*. *Pyrus* x bretschneideri is, among the various natural hybrids, the most important one from the point of view of cultivated varieties.

All *Pyrus* species have a 2n = 34 chromosomal number and only a few *P. communis* cultivars are known to be polyploidy (Angelini, 2007).

3. Economic importance

Pear is an important ingredient in many cuisines and is thus an important economic crop (Jackson et al. 2003).

Three-quarters of all cultivated pears are grown in Asia, and the world's largest producer is China (Layne et al., 1975); it produced a massive 18 million tons in 2014 (FAOSAT). The United States and Italy are also large producers of the fruit. There are three economically important species, *P. communis* (European pear), *P. pyrifolia*, (Japanese pear or Nashi), and *P. usuriensis* (Chinese pear). Japanese and Korean cultivars are complex hybrids of *P. pyrifolia* and *P. ussuriensis*. Pear has similar uses to apple, although its popularity may be somewhat lower because the best quality is ephemeral in European pear. In European winter pears this eating quality is achieved by ripening after harvest. Pear cider is usually made from cultivars of *P. nivalis* and is called perry. The pear tree is also an important ornamental and is beloved in Asia where pear is considered a sign of good luck. In the United States, the most popular ornamental pear trees were selections of *Pyrus calleryana,* the Callery pear. These street trees can be found from Oregon to Ohio to New York, and south to Alabama and Georgia. *Pyrus koehnii*, an evergreen species native to Taiwan, is planted in California and Florida.

Pears are consumed more for their taste than for nutritional value and their medicinal benefits. It has a high nutritional value with reasonable amounts of vitamins A, B1, B2, B3 and C and minerals such as sodium, potassium, phosphorus, calcium, magnesium and iron (Gonsalves et al., 2002; USDA 2012). Research has shown, however, that they have been helpful in lowering

cholesterol, decreasing blood lipid levels and improving the health of the stomach and intestine by protecting the ulcers and improving the production of useful bacteria to our body.

In addition, pears have been useful in diets of people suffering from obesity, kidney dysfunction, hypertension and cystitis, stimulating the elimination of urine and the proper functioning of the kidneys (Ferreira, 2002).

References

Ferreira, D., Guyot, S., Marnet, N., Delgadillo, I., Renard, C. M., and Coimbra, M. A., 2002. Composition of phenolic compounds in a Portuguese pear (*Pyrus communis* L. var. S. Bartolomeu) and changes after sun-drying. Journal of Agricultural and Food Chemistry. 50, 4537-4544.

Fideghelli, C., 2007. "Origine ed evoluzione," in Il Pero, R. Angelini, Ed., vol. 1, Bayer/CropScience, Milano, Italy.

Gonsalves, P. E., 2002. As frutas e seus benef´ıcios, in Frutas Que Curam. 1, 131-166.

http://ndb.nal.usda.gov/

http://faostat.fao.org/site/342/default.aspx.

Jackson J.E., 2003. The growing of apples and pears in Biology of Apples and Pears. Cambridge University Press, Cambridge, UK. 1, 4-20.

Layne, R. E. C., and Quamme, H. A., Pears in Advances in Fruit Breeding, J. Janick and J. N. Moore, 1975. EdsPurdue University Press, West lafayette, Ind, USA. 38-70.

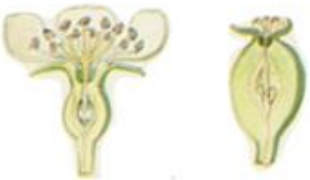
Pignatti, S.,1982. Flora d'Italia.

Silva, G. J., Souza, T. M., Barbieri, R. L., and Costa de Oliveira, A. 2014. Origin, domestication, and dispersing of pear (*Pyrus* spp.). Advances in Agriculture.

Vavilov N. I., 1992. Origin and Geography of Cultivated Plants. Cambridge University Press, Cambridge, UK. Vol.2.

Zukovskij, P. M., 1962. Cultivated Plants and Their Wild Relatives. Farnham Royal, London, UK, Translated by P.S. Hudson.

The species *Pyrus spinosa* Forssk. :



1. Botanical description

It is a shrub or small tree with slow growth, which can add 4-6 meters with stems and main branches erect, light drizzle. The trunk has a bark of grey colour, frozen in ripe specimens. It has erect-patented branches with lentils almost null. Younger jets are often spineless, with a reddish bark often dense or covered with a dense, non-persistent dense characteristic. These young branches carry oval-shaped gems, dull, enclosed in 6-9 brunet bristles, hairy or glabrous eyelashes.

The leaves are very variable, generally oblong-lanceolate, (1-2,5 - 3-7 cm), with rounded or acute apex and rounded base. The bottom page is white-torment in the juvenile phase and almost disperses later. The upper part of the adult leaves is shiny and has bluish shades.

The petiole is 10-20 mm long and the margins are wholly or finely pruned.

The flowers (8-12) are collected in umbrellatory peaks. Petals, white and elliptical (5-6 - 7-8 mm), are wholly or more often slightly apical, with short nails at the base.

The buds on the end have a reddish colour similar to the apple tree. Flowering takes place from March to May depending on altitudes and exposure. Fruits, spherical piridion, 2-4 cm in diameter, with a sturdy peduncle and as long as the fruit or more. The colour varies from green to brown to yellowish dirt due to its presence on almost all of its superficial formations (freckles), which often come together to cover the whole fruit. The pulp is woody for the presence of chlorides and harsh. The fruits remain on the trees until the beginning of the winter season (Zancheri, 1976; Pignatti, 1982;).

The fruits, though rough and hardly harvested, can be consumed with the aid of a plummeting, ancient practice by which maturation is induced by depositing the fruit in straw beds.

Small pears harvested at the end of summer or autumn can be stored for long periods of time.

2. Habitat

P.spinosa is a Steno-Mediterranean species spread from Spain to Turkey with particular reference to the following regions: Catalonia, Provence, Central and Southern Italy, Istria, Dalmatia, South Serbia, Peloponnese, Crete, Rhodes, Bitinia and Thrace (Zohary, 1997; Giardina et al., 2007).

Within the limits of its area and the highest altitude of survival, *P. spinosa* proves to possess a great ability to adapt to almost every type of soil: It can then be encountered in heavily drained back-soil, in rocky semi-arid areas, or on limestone or basal soil rich in metal oxides. It may occasionally grow along the slopes of seasonal waterways or on areas regularly occupied by temporary pools. It is common along the margins of roads, wooded areas or discontinuous cultivated land.

3. Systematic description

The complexity of the *Pyus* genus is such that to date one or more homogeneous classification criteria cannot be indicated. Nevertheless, while awaiting the identification of a classification system that includes exhaustively all known species of the *Pyrus* genus, the method adopted and proposed by Koehne (1980) remains the only one recognized as valid by most such specialists. According to this method, *P. spinosa* Forssk. It is located within the *Pyrus* group that exhibits fruit with persistent residual goblet; this group also belongs to *P. communis* L.

P. spinosa is species among the most polymorphic and phenotypically variable. This variability is the source of a particularly large number of descriptions by various authors between the 19th and 20th centuries of new species with related combinations of names, many of which are now invalidated or reduced to the synonym or variety. *P.spinosa* is currently considered a Crop Wild Relative (CWR) of the *P.communis* species, presenting a primary gene pool that allows hybridizing easily with it, (Vincent et al., 2013).

References

Giardina, G., Raimondo, F. M., and Spadaro, V., 2007. A catalogue of plants growing in Sicily. Bocconea. 20, 5-582.

Koehne, E., 1890. Die Gattung der Pomaceen.Wissenshaftliche Beitrage zum Programm des Falk-Realgymnasium zu Berlin. 95, 1-33.

Pignatti, S.,1982. Flora d'Italia.

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., and Maxted, N., 2013. A prioritized crop wild relative inventory to help underpin global food security. Biological conservation. 167, 265-275.

Zangheri, P., 1976. Flora Italica. Ed. Cedam, Padova. 2, 178-224.

Zohary, D., 1997. Wild apples and wild pears. In Proceedings of the workshop on conservation of the wild relatives of European cultivated plants. Bocconea. 7, 409-416.

CHAPTER 3: Seed image analysis to characterise Sardinian *P. communis* and *P. spinosa* and comparison with international varieties.

1. Introduction

Pear (*Pyrus communis* L.) after apple (*Malus domestica* Borkh.) and grape (*Vitis vinifera* ssp. vinifera) are one of the most important fruit trees cultivated in all temperate regions of the world (Wolko et al., 2010). Worldwide are produced 25,798,644 tons of pears where, Asia is the largest producer with 20,071,402 tones, followed by Europe with 2,976,540 tones, America with 1,869,017 tones, Africa with 758,522 tones and Oceania with 123,163 tones (FAOSAT 2014).

In Italy, the cultivation of the pear tree is strongly concentrated in Emilia-Romagna mainly in the triangle between Ferrara, Modena and Bologna, where in 2015 the production was 515,000 tons overall. In addition, Veneto and Sicily also boast considerable pear production (http://www.freshplaza.it/article/79363/Produzione-2015-di-pere-in-Italia-e-prime-indicazioni-commerciali).

Fossil evidence has shown that the genus *Pyrus* has differentiated in the Tertiary period (65 to 55 million years ago) and its origin has been placed in the mountainous areas of western and sub-western China (Kumar et al., 2017). Afterward, it would have expanded to adapt to the different climatic and environmental conditions, thus giving origin to present species (Rubtzov, 1944; Zielinski and Thompson, 1967). This location was also suggested by Nicolai Vavilov that in the 1930s identified two primary centres of domestication originated in regions extending from Asia west to the Caucasus (Vavilov, 1951). A third centre was also reported in Central Asia (northwestern India, Afghanistan, Tajikistan, Uzbekistan and the Tianshan province of China) (Zukovskij, 1962; Vavilov, 1992).

Domesticated cultivars of *P. communis* spp. *communis* hybridized with ancestors of *P. communis* subspecies and *P. nivalis* Jacquin. Also *P. communis* spp. *pyraster* and *P. communis* spp. *caucasian* are probably the ancestors of the cultivated European pear (Challice and Westwood, 1973).

Regarding pear tree cultivation, many species may have escaped the control of farmers, generating wild population that hybridizing with cultivars could have originated new pears variety (Dosdek, 1991; Aldasoro et al., 1996; Paganova, 2003).

The *P. communis* is practically the only one cultivated species in Europe, Africa, America, Oceania and West Asia, with few exceptions of local oriental productions. Wild populations are found in Europe, North Africa, West, Central and Eastern Asia and Japan (Terpò, 1992).

Cultivated pears were also known by the Greeks and Romans from about 2500 years ago (Hedrick, 1924).

During the Roman period, many pear varieties were introduced in Sardinia and some of these were mentioned by Elder Pliny in the "Naturalis Historia", at present day some of these still appear in the Sardinia germplasm (Agabbio et al., 1986). Moreover, during the Judicial period between the ninth and fifteenth centuries, pear had a great spread throughout Sardinia by grafting on the wild species (*P. spinosa* Forssk.), so called "pirastru" are widespread throughout the island, this maybe a sign of their ancient origin (Camarda and Valsecchi, 1983). Subsequently, in the Aragonese period (1323-1326 AD), the cultivation of pomaceae was intensely practised to support family needs. (Agabbio et al., 1994).

At present day, the local Sardinian varieties are progressively replaced with more productive modern ones, leading to the disappearance of local cultivars. Many of these ancient varieties are preserved from the CNR-ISPA catalogue fields (Nuraxinieddu, Oristano, Sardinia). The harvesting time of these varieties starts in June and ends late in November. Some of the late harvested varieties can only be consumed after a long storage period needed to ripen fruit and decrease its astringency. A characteristic core browning, known as 'ammezzimento' occurs in several ancient varieties and in the early harvested ones it takes place before or quickly after harvest while, it is a slow process in late-season harvested fruit. Moreover, this characteristic allows the autumn harvested pears to be preserved for a long time and consumed them throughout the New Year.

The loss of local varieties typical of a territory represents not only a serious biological, ecological and cultural defeat for the community but also the loss of great opportunities for development. Furthermore, for these local varieties, no morphological or genetic studies have been carried out for their preservation.

In order to counteract the phenomenon of genetic erosion of the old pear varieties in Sardinia, this work has been carried out to preserve germplasm diversity, and characterize pear varieties. The main objectives of this study are to investigate by seed image analysis the phenotype diversity of pears, in particular:

(1) the relationship between local cultivars and wild populations of Sardinian pear and national and international varieties;

(2) the phenomenon of pear over-ripening in Sardinian local cultivars.

2. Materials and Methods

2.1. Plant material

In this work, we have studied 65 *P. communis* Sardinian varieties from the CNR-ISPA catalogue field (Nuraxinieddu, Oristano, Sardinia) and seven varieties of *P. spinosa* collected in Sardinia (Table 1; Fig. 1). Moreover, additional 44 national and international varieties of *P. communis* were collected in the field catalogue of the experimental station of Cadriano of the Department of Agricultural Sciences Bologna, (Italy), and added in this study (Table 1; Fig. 2).

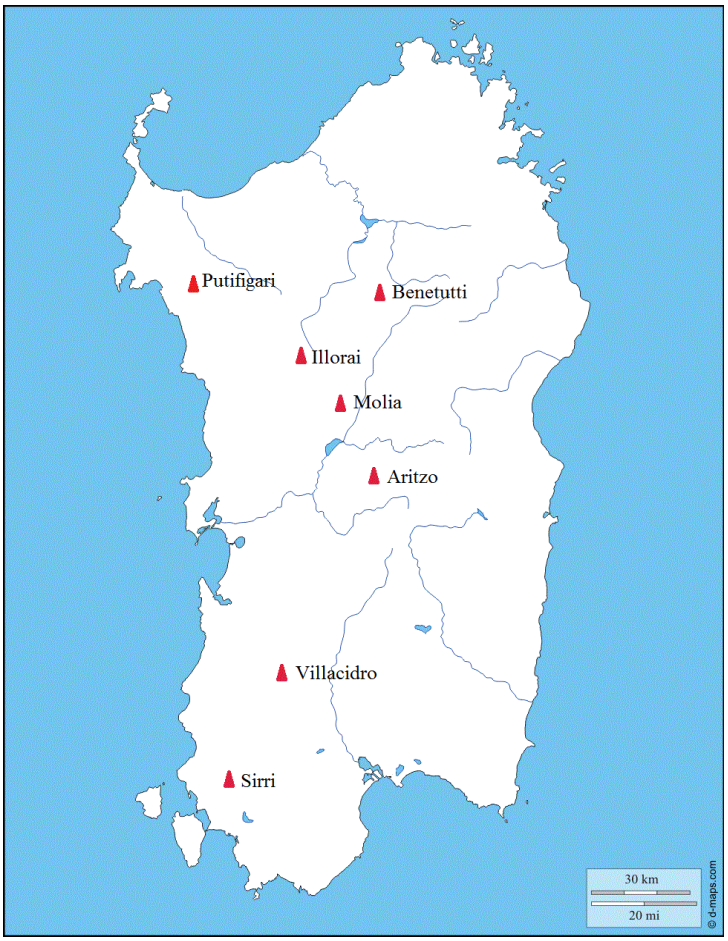


Fig.1 Distribution of *P. spinosa*, samples selected for this study.

Table 1. *P. communis* and *P. spinosa* seeds analysed in this study.
SA *P.c* = *P. communis* varieties of Sardinia; BO *P.c* = *P. communis* national and international varieties of Bologna; SA *P.s* = *P.spinosa* collected in Sardinia.

Code	Origin	Variety name	N° Seeds	Code	Origin	Variety name	N° Seeds	Code	Origin	Variety name	N° Seeds
C1	SA *P.c*	Belgamotto	42	C41	SA *P.c*	Piringino	77	C81	BO *P.c*	Grossa di castagnito	72
C2	SA *P.c*	Camusina	60	C42	SA *P.c*	Parena	44	C82	BO *P.c*	Idice b	91
C3	SA *P.c*	Orotelli	24	C43	SA *P.c*	Reale	15	C83	BO *P.c*	Kieffer	96
C4	SA *P.c*	Pera Mela	40	C44	SA *P.c*	Regina	53	C84	BO *P.c*	Leopardo	60
C5	SA *P.c*	Camusina precoce	60	C45	SA *P.c*	Roma	37	C85	BO *P.c*	Martin del bosc	59
C6	SA *P.c*	Alveghina	60	C46	SA *P.c*	Ruspu Nuchis	42	C86	BO *P.c*	Martin sec	55
C7	SA *P.c*	Appicadorza	10	C47	SA *P.c*	Buttiru de Jerru	60	C87	BO *P.c*	Mercedes	61
C8	SA *P.c*	Arangiu	41	C48	SA *P.c*	Collone de Manu	60	C88	BO *P.c*	Mora	56
C9	SA *P.c*	Armungesa	66	C49	SA *P.c*	Sant Andrea	39	C89	BO *P.c*	Mora di pirovano	61
C10	SA *P.c*	Arriabi	14	C50	SA *P.c*	San Giovanni Migliorata	51	C90	BO *P.c*	Pac hams	93
C11	SA *P.c*	Bau	60	C51	SA *P.c*	San Giovanni	60	C91	BO *P.c*	Passadel	58
C12	SA *P.c*	Bianca di Gonnos	64	C52	SA *P.c*	Sa Maddalena	60	C92	BO *P.c*	Per duche	17
C13	SA *P.c*	Bottida e Austu a	60	C53	SA *P.c*	Sarmentina	70	C93	BO *P.c*	Pera angelica	52
C14	SA *P.c*	Bragamota	39	C54	SA *P.c*	San Domenico	108	C94	BO *P.c*	Pera bianca	70
C15	SA *P.c*	Buttiu De Austu b	35	C55	SA *P.c*	San Giovanni orrubia	36	C95	BO *P.c*	Pera limone	95
C16	SA *P.c*	Cabudraxia	60	C56	SA *P.c*	Vacchesa	96	C96	BO *P.c*	Pera martini	85
C17	SA *P.c*	Camusina di Sassari	58	C57	SA *P.c*	Funtana Sones	58	C97	BO *P.c*	Pera proni	80
C18	SA *P.c*	Camusina grande	60	C58	SA *P.c*	Santa Maria	54	C98	BO *P.c*	Pero fiorenza	83
C19	SA *P.c*	Cauli	13	C59	SA *P.c*	Camusina di Bonarcado	13	C99	BO *P.c*	Pero rossellini	97
C20	SA *P.c*	Cozzone ainu	60	C60	SA *P.c*	Camusina di Precoce	72	C100	BO *P.c*	Pero truvela	35
C21	SA *P.c*	Cracchera	46	C61	SA *P.c*	Meba	32	C101	BO *P.c*	Pierre corneille	95
C22	SA *P.c*	De Puleu	15	C62	SA *P.c*	Piringino di Giugno	100	C102	BO *P.c*	Pignano	83
C23	SA *P.c*	De su Duca	57	C63	SA *P.c*	Santa Barbara	100	C103	BO *P.c*	Santa Lucia	104
C24	SA *P.c*	Di Luglio	60	C64	SA *P.c*	Sitzia	38	C104	BO *P.c*	Santa Maria	480
C25	SA *P.c*	E Donna	84	C65	SA *P.c*	Limoni b	8	C105	BO *P.c*	Spadona d'inverno	486
C26	SA *P.c*	Funtana Sones	26	C66	BO *P.c*	Buona Luisa D'avreances	90	C106	BO *P.c*	Spina carpi	265
C27	SA *P.c*	Laconi 2	59	C67	BO *P.c*	Butirra hardy red	68	C107	BO *P.c*	Suittore	189
C28	SA *P.c*	Laconi 1	57	C68	BO *P.c*	Campigna	78	C108	BO *P.c*	Tarda	100
C29	SA *P.c*	Laconi 4	81	C69	BO *P.c*	Campigna a	103	C109	BO *P.c*	Villa maria	100
C30	SA *P.c*	Lida	60	C70	BO *P.c*	Campigna b	42	C110	BO *P.c*	Avalle	104
C31	SA *P.c*	Limoni a	4	C71	BO *P.c*	Carola	77	C111	SA *P.s*	Benetutti	25
C32	SA *P.c*	Muscadeddu de Jerru	53	C72	BO *P.c*	Cedrata romana	80	C112	SA *P.s*	Molia	41
C33	SA *P.c*	Meana	36	C73	BO *P.c*	Curato	14	C113	SA *P.s*	Putifigari	9
C34	SA *P.c*	Meli	48	C74	BO *P.c*	Dell'Auzzana	75	C114	SA *P.s*	Sirri	64
C35	SA *P.c*	Muscadeddu	59	C75	BO *P.c*	Dirce	89	C115	SA *P.s*	Aritzo	101
C36	SA *P.c*	Natalina	66	C76	BO *P.c*	Dr Guyot	53	C116	SA *P.s*	Illorai	52
C37	SA *P.c*	Natauna	60	C77	BO *P.c*	Duchessa D'anguleme	39	C117	SA *P.s*	Villacidro	52
C38	SA *P.c*	Oddinesa	60	C78	BO *P.c*	Garzon	95				
C39	SA *P.c*	Oliena 1	26	C79	BO *P.c*	Gioma	83				
C40	SA *P.c*	Pira-Mamoi	22	C80	BO *P.c*	Goriziana rossa	35				

Fig. 2 *Pyrus* varieties provided by the Department of Agricultural Sciences Bologna used in this study.

The fruits were harvested at full maturity after removed the pulp, the seeds were cleaned, washed, and subsequently air dried according to the standard protocol adopted in the Germplasm Bank of Sardinia BG-SAR (Atzeri et al., 2012) (Fig.3).

Fig. 3 The protocol of the Germplasm Bank of Sardinia adopted to collect, clean and store pear seeds. (from upper left to right: seed removal, soaking and air-drying. Lower sequence from left to right: seed selection, packaging and storage).

2.2. Seeds image analysis

A total of 8,119 seeds were analysed using the seed image and Linear Discriminant Analysis (LDA) method.

Digital images of seeds were acquired using a flatbed scanner (Epson Perfection V550), with a digital resolution of 800 dpi for a scanning area not exceeding 1024×1024 pixels (Bacchetta et al., 2008) (Fig. 4). The images were processed and analysed using the open source software ImageJ v. 1.49 (http://rsb.info.nih.gov/ij). A plugin, Particles8 (Landini 2006), freely available on the official website (http://www.mecourse.com/landinig/software/software.html) was used to measure 26 seed morphometric features (Table 2; Fig.5).

The morphometric parameters were used to build a database of features descriptive of seed size and shape.

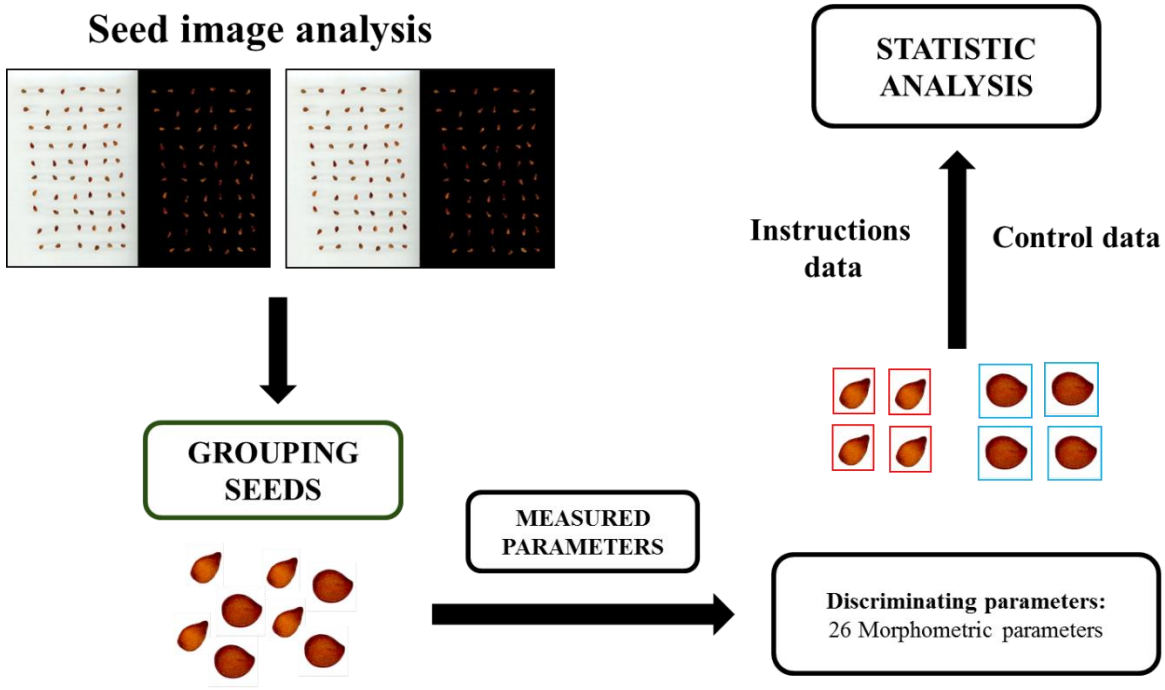


Fig. 4 Flow-chart of the seed image analysis.

2.3. Statistical analysis

Morphometric data have been processed and statistically analysed by applying the stepwise LDA method and by using the IBM SPSS software (Statistical Package for Social Science) release 16.0 (SPSS Inc. for Windows, Chicago, Illinois, USA). LDA method is commonly used to classify or identify unknown groups characterized by quantitative and qualitative variables (Sugiyama, 2007), finding the combination of predictor variables with the aim of minimizing the within-class distance and maximizing the between-class distance simultaneously, thus achieving maximum class discrimination (Hastie et al., 2001; Holden et al., 2011; Kuhn and Johnson, 2013).

This method allows reducing the dataset size without losing important information and classifying statistical cases into groups (Fukunaga, 2013; Duda et al., 2012; Hastie et al., 2001; Holden et al., 2011; Sugiyama, 2007). LDA achieve the highest-class discrimination combining the all predictor variables by minimizing the distance within the class (Holden et al., 2011; Rencher and Christensen 2012; Kuhn and Johnson, 2013). The stepwise method identifies and selects the most statistically significant features, among the 26 measured on each seed, using three statistical variables: Tolerance, *F*-to-enter and *F*-to-remove. The variable variance indicates by the tolerance value, *F*-to-enter and *F*-to-remove values define the power of each variable in the model and are useful in describing what happens if a variable is inserted and removed, from the current model. At each step, the largest F-to-enter value respect to the chosen value ($F \geq 3.84$) is added to the model. Therefore, all variables with values smaller than 3.84

will not be added to the analysis at the last step (Venora et al., 2009). Finally, a cross-validation procedure is applied to verify the performance of the identification system, testing individual unknown cases and classifying them according to all others. This procedure also called rotation estimation (Picard and Cook, 1984; Kohavi, 1995) was applied, both to evaluate the performance and to validate any implemented classifier. The validation procedure used here is the leave-one-out cross-validation (LOOCV) (SPSS 2006).

Table 2. Morphometric features measured on seeds and calculated by Particles8 plugins from ImageJ v. 1.49.

Parameter	Description
Perim	Perimeter, calculated from the centres of the boundary pixels
Area	Area inside the polygon defined by the perimeter
Pixels	Number of pixels forming the seed image
MinR	Radius of the inscribed circle centred at the middle of mass
MaxR	Radius of the enclosing circle centred at the middle of mass
Feret	Larges taxis length
Breadth	Largest axis perpendicular to the Feret
CHull	Convex hull or convex polygon calculated from pixel centres
CArea	Area of the convex hull polygon
MBCRadius	Radius of the minimal bounding circle
AspRatio	Aspect ratio = Feret/Breadth
Circ	Circularity = $4 \cdot \pi \cdot \text{Area}/\text{Perimeter}^2$
Roundness	Roundness = $4 \cdot \text{Area}/(\pi \cdot \text{Feret}^2)$
ArEquivD	Area equivalent diameter = $\sqrt{((4/\pi) \cdot \text{Area})}$
PerEquivD	Perimeter equivalent diameter = Area/π
EquivEllAr	Equivalent ellipse area = $(\pi \cdot \text{Feret} \cdot \text{Breadth})/4$
Compactness	Compactness = $\sqrt{((4/\pi) \cdot \text{Area})}/\text{Feret}$
Solidity	Solidity = Area/Convex_Area
Concavity	Concavity = Convex_Area-Area
Convexity	Convexity = Convex_hull/Perimeter
Shape	Shape = $\text{Perimeter}^2/\text{Area}$
RFactor	RFactor = Convex_Hull /$(\text{Feret} \cdot \pi)$
ModRatio	Modification ratio = $(2 \cdot \text{MinR})/\text{Feret}$
Sphericity	Sphericity = MinR/MaxR
ArBBox	Area of the bounding box along the feret diameter = $\text{Feret} \cdot \text{Breadth}$
Rectang	Rectangularity = Area/ArBBox

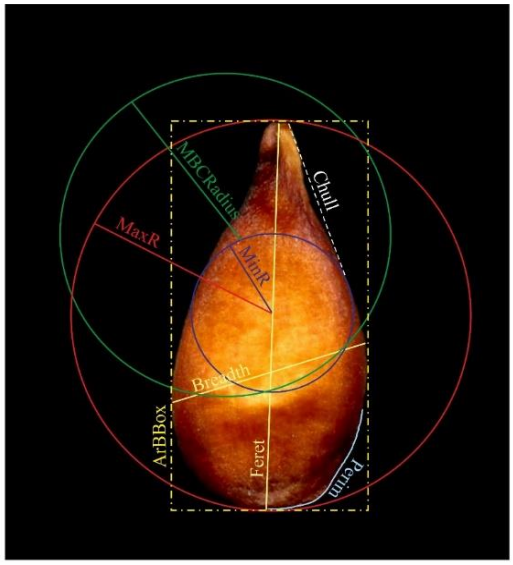


Fig. 5 Graphical representation of principal morphometric parameters measured on each seed (see Tab. 2).

3. Results

The first comparison among the 65 ancient *P. communis* local varieties from Sardinia was executed, achieving a 38.8% cross-validation (Fig. 6).

The first analysis showed a high classification within each varieties: Belgamotta (100%): Meli (72,9%); Sa Maddalena (71,7%); Piringino di Giugno (68,0%); Santa Barbara (66,0%); Roma (64,9%); Alveghina (63,3%); Limoni b (62,5%); Lida (61,7%). Other varieties that showed a discriminatory power of more than 50% were: Camusina di Paegle (58,3%); Pera Mela (55%); San Giovanni Migliorata (54,9%); Vacchesa (51,0%); Laconi2 (50,8%) San Domenico (50,9%). Moreover varieties that showed a low percentage of discrimination below 10% were: Appicadorza (0%); Funtana Sones (0%); Limoni a (0%); Parena (0%); Camusina di Bonarcado (0%); Meba (0%); Buttiu de Austu (5,7%); Arriabi (7,1%) e Pira Mamoi (9,1%).

The LDA analysis, conducted to differentiate pear varieties from the catalog field of Cadriano (Bologna) showed a cross-validation of 38.1% (Fig.6). Noteworthy, of the 45 varieties analysed 12 showed a high discriminant percentage: Pero rossellini (95,9%); Martin del bosc (78,0%); Idice B (72,5%); Campigna (73,1%); Pec hams (62,4%) e Suittore (64,1%), Pera fiorenza (57,8%); Pera bianca (60%); Avalle (59.6); Per Duche (52,9%); Martin Sec (52,7%), Spadona d'inverno (51,2%); e Compigna b (50%). Moreover, varieties that showed a percentage of discrimination below 10% were: Grossa di castagnito (2,8%); Pero truvela (8,6%); Tarda (8,9%) e Dr. Guyet (9,4%).

In addition, the LDA analysis conducted on the varieties of *P. spinosa* from Sardinia showed a general cross-validated by 74.4% (Table 3; Fig. 8). Varieties with the highest rates of discrimination were:

Benetutti (95,6%) e Illorai (90%). Moreover, the variety with the lowest percentage of discriminant was Sirri (21,7%) (Table 3).

Furthermore, the LDA conducted on the two *P. communis* groups (SA and BO *P.c.*) and *P. spinosa* of Sardinia showed that these three groups are able to differentiate from each other, showing, in fact, a percentage of discrimination of 62,7% for Sardinian varieties, 68,8% for national and international varieties of Bologna catalogue field and 76,5% for the varieties of *P. spinosa* of Sardinia (Table 4; Fig. 9).

Moreover, considering *P.spinosa* which unknown group, LDA analysis showed similarities with the cultivars of Sardinia with 56,6 % cases (Table 5).

Considering the different degree of ripeness and over-ripening in Sardinia pears, LDA showed a cross-validated of 77.6% (Table 6; Fig. 10). Early-ripening pears were discriminated with a percentage of discrimination of 90.7%, intermediate varieties with of 26.4% and late varieties with a percentage of 58.2% (Table 6).

Finally, the analysis conducted for investigating the characteristic of over-ripening showed an optimal discrimination with a cross-validated 98.2% (Table 7; Fig. 11). In this case, the varieties with high over-ripening were discriminated to 96.8% and with low over-ripening to 99,2%.

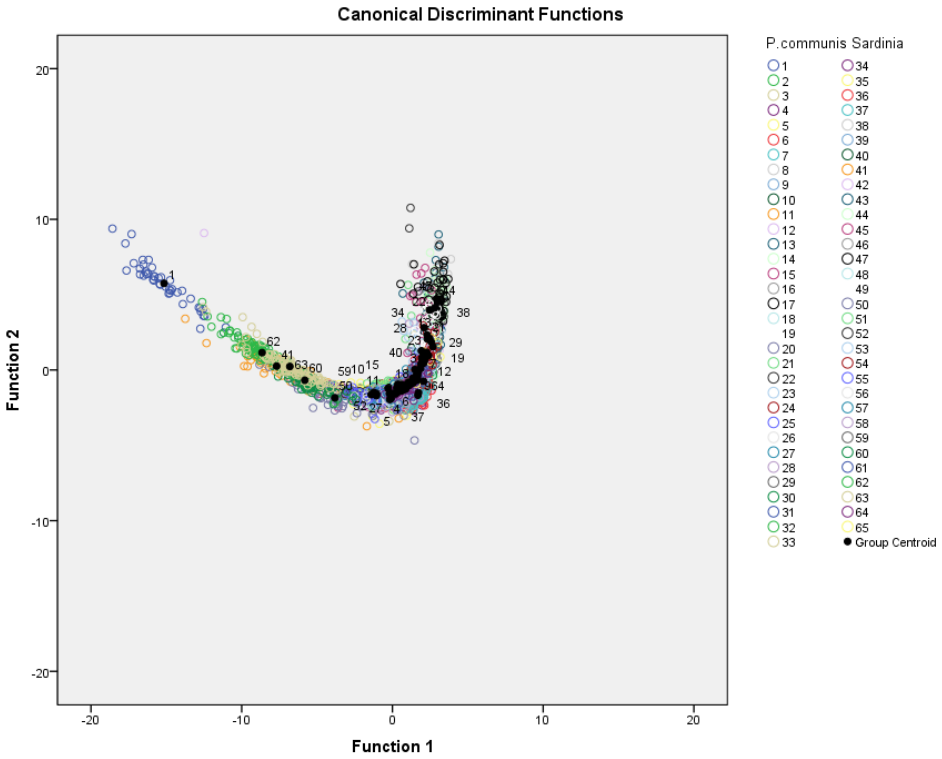


Fig. 6 Scatter plot graph based on LDA analysis discrimination of the 65 *P. communis* Sardinian varieties. Small points variously colored represent single seed data, black ones represent their average (centroid).

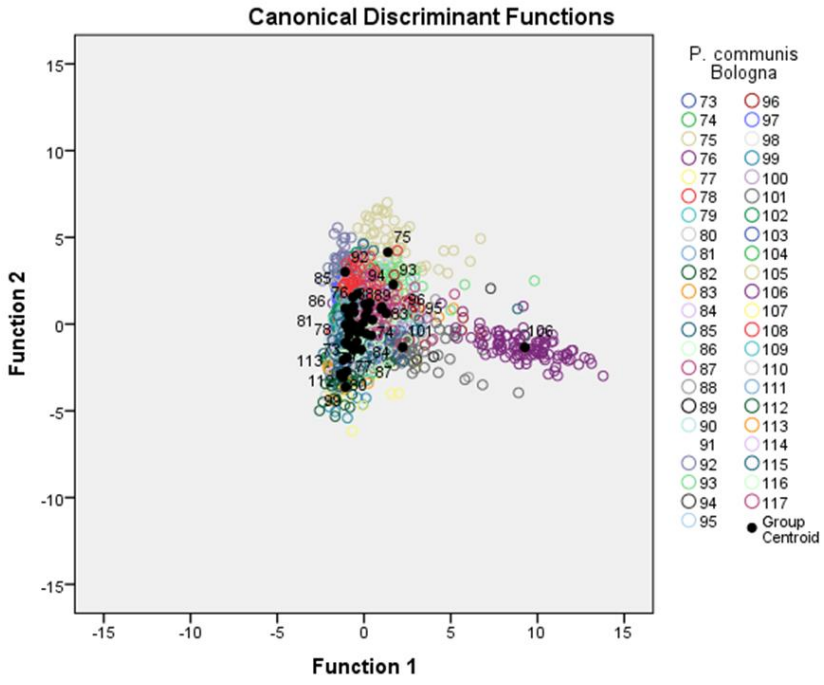


Fig. 7 Scatter plot graph based on LDA analysis discrimination of the 45 international varieties of catalogue field of Bologna of *P. communis*. Small points variously coloured represent single seed data, black ones represent their average (centroid).

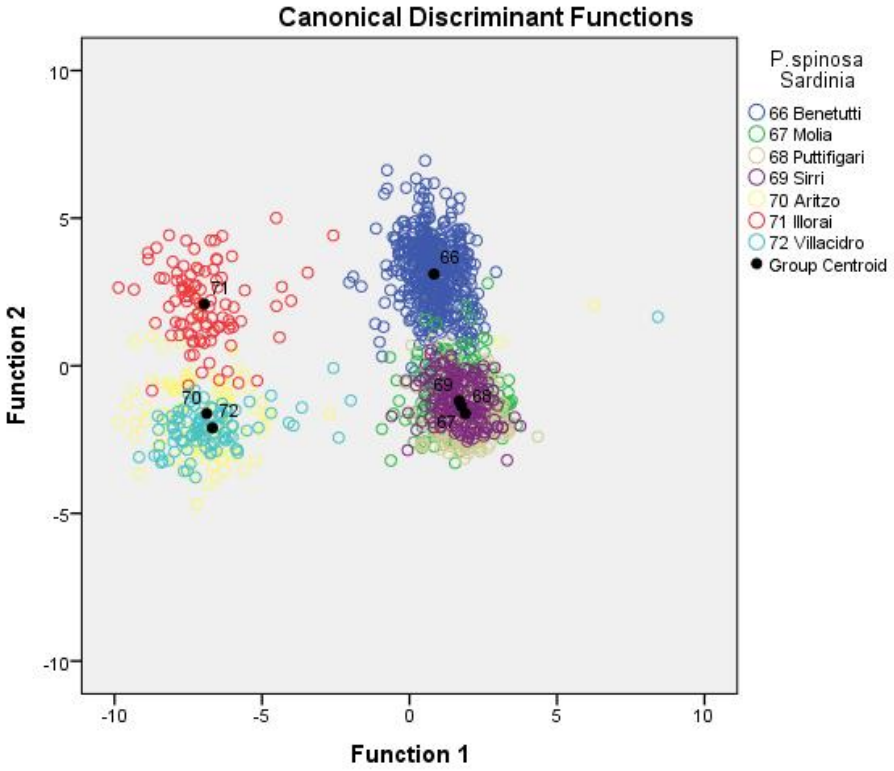


Fig. 8 Scatter plot graph based on LDA analysis discrimination of seven *P. spinosa* Sardinian varieties. Small points variously coloured represent single seed data, black ones represent their average (centroid).

Table 3. Correct classification percentages among *P. spinosa* Sardinian varieties.

Classification Results									
P. spinosa			Predicted Group Membership						Total
			Benetutti	Molia	Putificagari	Sirri	Aritzo	Illorai	Villacidro	
Cross-validated	Number seeds	Benetutti	459,0	19,0	0,0	2,0	0,0	0,0	0,0	480,0
		Molia	5,0	362,0	86,0	33,0	0,0	0,0	0,0	486,0
		Putifigari	2,0	78,0	173,0	12,0	0,0	0,0	0,0	265,0
		Sirri	0,0	115,0	33,0	41,0	0,0	0,0	0,0	189,0
		Aritzo	1,0	0,0	0,0	1,0	81,0	5,0	12,0	100,0
		Illorai	2,0	0,0	0,0	0,0	7,0	90,0	1,0	100,0
		Villacidro	0,0	1,0	0,0	1,0	16,0	0,0	86,0	104,0
	%	Benetutti	**95,6**	4,0	0,0	0,4	0,0	0,0	0,0	100,0
		Molia	1,0	**74,5**	17,7	6,8	0,0	0,0	0,0	100,0
		Putifigari	0,8	29,4	**65,3**	4,5	0,0	0,0	0,0	100,0
		Sirri	0,0	60,8	17,5	**21,7**	0,0	0,0	0,0	100,0
		Aritzo	1,0	0,0	0,0	1,0	**81,0**	5,0	12,0	100,0
		Illorai	2,0	0,0	0,0	0,0	7,0	**90,0**	1,0	100,0
		Villacidro	0,0	1,0	0,0	1,0	15,4	0,0	**82,7**	100,0
74,9% of cross-validated grouped cases correctly classified.										

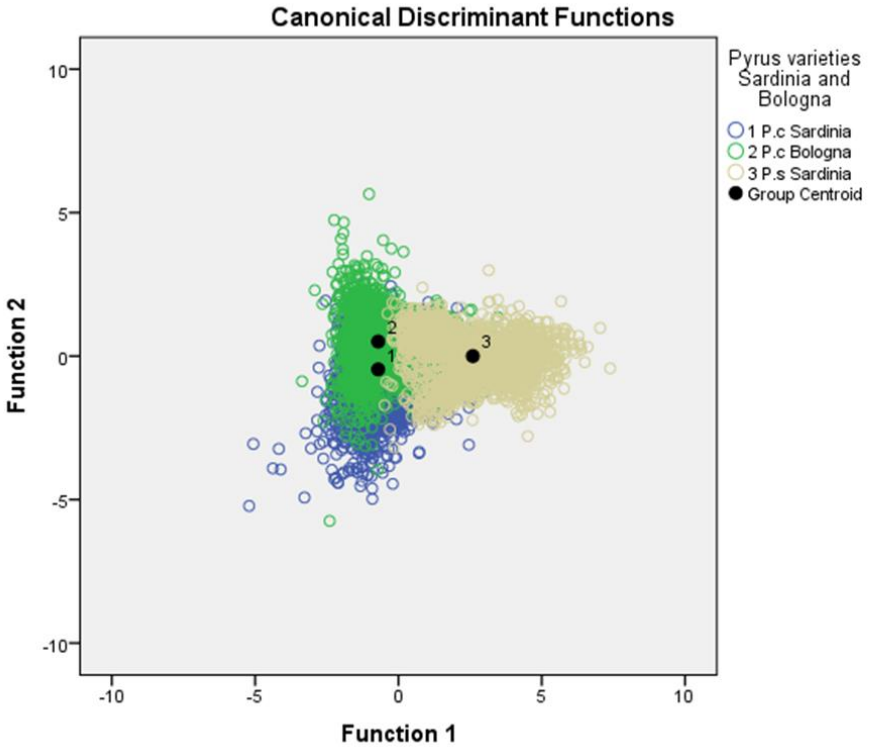


Fig. 9 Scatter plot graph based on LDA analysis discrimination of *P. communis* varieties from: Sardinia (1) Bologna (2) and *P. spinosa* ones from Sardinia (3). Small points variously coloured represent single seed data, black ones represent their average (centroid).

Table 4. Correct classification percentages among *P. communis* and *P. spinosa* varieties from Sardinia and *P. communis* ones from Bologna.

Classification Results						
P.communis and *P.spinosa*			Predicted Group Membership			Total
			P.c Sardinia	P.s Sardinia	P.c Bologna	
Cross-validated	Number seeds	P.c Sardinia	2061	69	1158	3288
		P.s Sardinia	129	1319	276	1724
		P.c Bologna	826	116	2081	3023
	%	P.c Sardinia	62,7	2,1	35,2	100,0
		P.s Sardinia	7,5	76,5	16,0	100,0
		P.c Bologna	27,3	3,8	68,8	100,0
68,0% of cross-validated grouped cases correctly classified.						

Table 5. Correct classification percentage between the *P. spinosa* from Sardinia, considered as unknown group, and *P. communis* from Sardinia and Bologna.

Classification Results					
P. communis and *P.spinosa*			Predicted Group Membership		Total
			P.c Sardinia	P.c Bologna	
Cross-validated	Number seeds	P.c Sardinia	2150	1138	3288
		P.c Bologna	955	2068	3023
		P.s Sardinia	976	748	1724
	%	P.c Sardinia	65,4	34,6	100,0
		P.c Bologna	31,6	68,4	100,0
		P.s Sardinia	56,6	43,4	100,0
66,7% of cross-validated grouped cases correctly classified.					

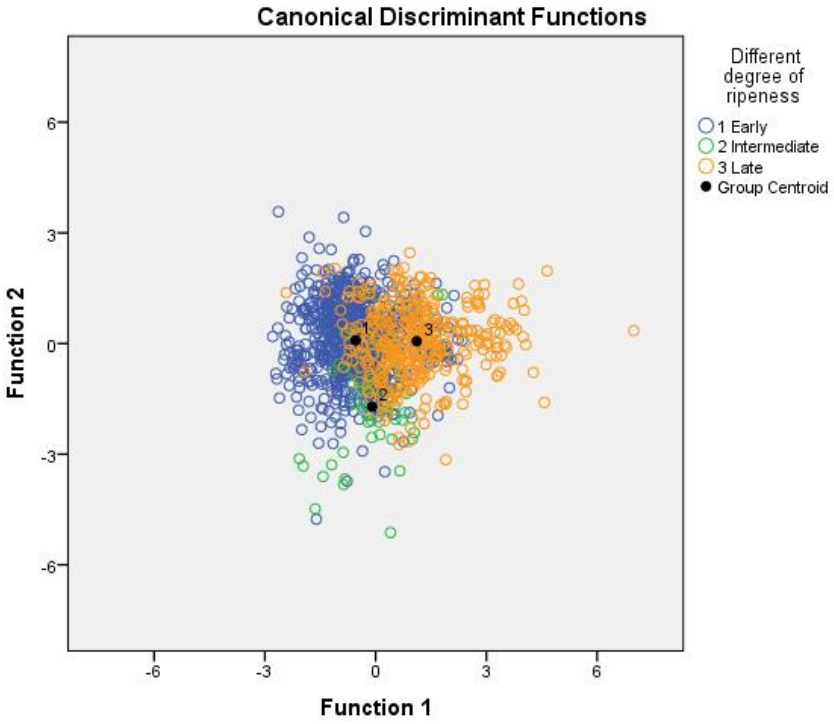


Fig. 10 Scatter plot graph based on LDA analysis discrimination of *P. communis* varieties from Sardinia for different degree of over-ripening. Small points variously coloured represent single seed data, black ones represent their average (Centroid).

Table 6. Correct classification percentages among *P. communis* varieties from Sardinia selected by fruit over-ripening.

Classification Results						
Ripeness			Predicted Group Membership			Total
			Early	Intermediate	Late	
Cross-validated	Number seeds	1	719	8	66	793
		2	33	14	6	53
		3	158	6	228	392
	%	1	**90,7**	1,0	8,3	100,0
		2	62,3	**26,4**	11,3	100,0
		3	40,3	1,5	**58,2**	100,0
77,6% of cross-validated grouped cases correctly classified.						

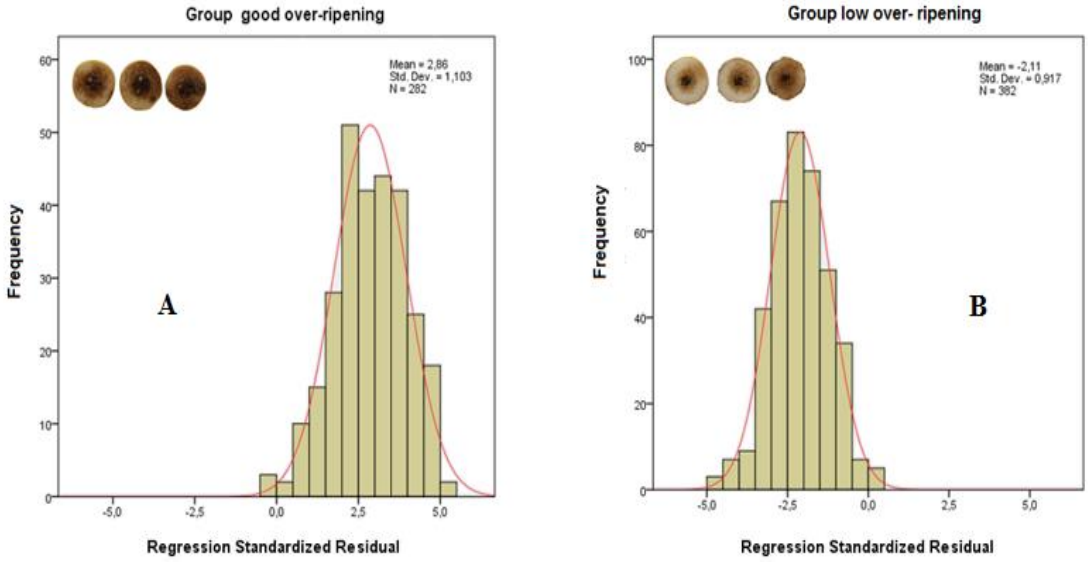


Fig. 11 Histogram of the standardised residuals based on the LDA analysis discrimination of *P. communis* varieties from Sardinia according to over-ripening degree: (A) high and (B) low.

Table 7. Correct classification percentages among *P. communis* varieties to Sardinia selected by over-ripening.

Classification Results					
Group over-ripening			Predicted Group		Total
			High	Low	
Cross-validated	Number seeds	High	273	9	282
		Low	3	379	382
	%	High	96,8	3,2	100,0
		Low	0,79	99,2	100,0
98,2% of cross-validated grouped cases correctly classified.					

4. Discussion

The LDA conducted on *P. communis* varieties from Sardinia and Bologna used in this study showed a low percentage of discrimination. This may be due to the fact that *Pyrus* has a genetic history similar to that of the *Malus* genus even in the case of the *Malus* genus, the LDA shows a low diversification of varieties within the studied collections (Kole, 2011).

Regarding the Sardinian varieties, there was good discrimination only for some ones, such as the "Belgamotta" variety. Unfortunately, the scarcity or total missing of bibliographic information does not allow us to discuss these results. Further morphological and genetic analyses could help in understanding the results obtained.

Moreover, the LDA carried out on the Bologna varieties showed a high degree of diversification for some individual cultivars, such as Pero Rossellini. In fact, in the work of Martinelli et al. 2008, "Pero Rossellini" is mentioned among the old varieties of Trentino Alto Adige. Genetic and morphometric studies based on agronomic data of "Pero Rossellini" have shown that this variety is both genetically and morphologically different from the others, supporting our results.

The LDA conducted on *P. spinosa* varieties showed a good percentage of discrimination. This may be explained by the fact that *P. spinosa* was collected in different territories of Sardinia and environmental conditions, *e.g.* tree age and growth phases, may influence phenotypic characteristics (Rotondi et al., 2003).

The results obtained through LDA, considering varieties of *P. communis* from Sardinia and Bologna and *P. spinosa* showed that *P. spinosa* ones are more closely related to Sardinian *P. communis* ancient varieties, in comparison to Bologna varieties grouping separately.

This result can be explained by the fact that some varieties of Sardinian *P.communis* have certainly originated from the populations of *P.spinosa* present in Sardinia, while the varieties of national and international *P.communis* came from different populations of *P.spinosa*.

In fact, as known *P.spinosa* represents a Wild Crop Relative (CWR), with primary gene pool, of *P.communis* and hence interfertile with it (Vincent et al., 2013).

LDA conducted considering the different degree of over-ripening of Sardinian *P. communis* varieties, showed that early and late harvested varieties were those with the highest percentage of discrimination, showing an effective correlation between the shape of the seed and the different degrees of over-ripening. In addition, the same analysis conducted considering the high maturity varieties, showed that it was possible to perfectly discriminate all varieties that present this characteristic.

This work is the first study conducted to characterise pear varieties in Sardinia based on seed morphometric data and the results obtained show that seed morphology reflects phenotypic and genotype characteristics of the varieties and how these data can be compared with genetic

analyses to shed light on the relationships between seeds morphology and characteristics of fruit.

References

Agabbio, M., 1994. Le vecchie varietà della Sardegna. Carlo Delfino editore.1-424.

Agabbio, M., D' Aquino, S., Piga, A., 1996. Risposta alla frigoconservazione di alcune cultivar di pere estive del germoplasma autoctono. Atti del 2° Convegno Nazionale Biodiversità e Produzioni Biologiche. Matera.

Agabbio, M., Suelzu, R., Mulas, M., Mannoni, G., 1986. Patrimonio genetico delle pomacee in Sardegna: il pero: cenni storici e cultivar individuate. Studi Sassaresi. 32, 119-138.

Aldasoro, J. J., Aedo, C., Munoz Garmendia, F., 1996. The genus *Pyrus* L. (Rosaceae) in southwest Europe and North Africa. Bot. J. Linnean Soc. 121,143-158.

Atzeri, P., Bacchetta, G., Casti, M., Fenu, G., Mattana, E., Meloni, F., and Sarigu, R., 2012. La Banca del Germoplasma della Sardegna (BG-SAR). Studi Trentini di Scienze Naturali Acta Biologica. 88, 69-75.

Bacchetta, G., Grillo, O., Mattana, E., Venora, G., 2008. Morpho-colorimetric characterization by image analysis to identify diaspores of wild plant species. Flora-Morphology, Distribution, Functional Ecology of Plants. 203, 669-682.

Boni, C., 1994. Flora Sarda Alberi e Arbusti. Edisar.

Camarda, I., and Valsecchi, F., 1983. Alberi ed arbusti spontanei della Sardegna. Gallizzi Editore.

Challice, J. S., and Westwood, M. N., 1973. Numerical taxonomic studies of the genus *Pyrus* using both chemical and botanical characters. Botanical Journal of the Linnean Society. 67, 121-148.

Dosdek, J., 1991. Qnu *pratln* in Czechoslovakia. Preslia. 63, 113-118.

Duda, R. O., Hart, P. E., Stork, D. G., 2012. Pattern classification. John Wiley and Sons.

Fukunaga, K., 2013. Introduction to statistical pattern recognition. Academic press.

Gonsalves, P. E., 2002. “As frutas e seus benef´ıcios,” in Frutas Que Curam. 1: 131-166.

Grillo O., Draper D., Verona G., Martínez-Laborde J. B. Seed image analysis and taxonomy of *Diplotaxis*. (Brassicaceae Brassiceae) Systemat. Biodivers. 10, 57-70.

Hastie, T., Tibshirani, R., Friedman, J., 2013. The elements of statistical learning: data mining, inference and prediction. Second Edition. New York: Springer Ser. Statist. p. 536.

Hedrick, U. P., Howe, G. M., Taylor, O. M., Francis, E. H., and Tukey, H. B., 1924. The pears of New York. Journal of Pomology and Horticultural Science. 3,153-155.

Holden, J. E., Finch, W. H., Kelley, K., 2011. A comparison of two-group classification methods. Educ Psychol Meas.71, 870-90.

Kole C., 2011. Wild crop relatives: genomic breeding resource. Temperate fruits.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai.14, 1137-1145.

Kuhn, M., Johnson, K., 2013. Discriminant analysis and other linear classification models. Applied Predictive Modeling New York: Springer. pp. 275-328.

Kumar, S., Kirk, C., Deng, C., Wiedow, C., Knaebel, M., and Brewer, L., 2017. Genotyping-by-sequencing of pear (*Pyrus* spp.) accessions unravels novel patterns of genetic diversity and selection footprints. Horticulture Research, 4:17015.

Landini, G., 2006. Quantitative analysis of the epithelial lining architecture in radicular cysts and odontogenic keratocysts. Head Face Med. 2, 1-9.

Lombard, P., Hull Jr, J., and Westwood, M. N., 1980. Pear cultivars of North America. Fruit Varieties Journal, 34: 74-83.

Paganova, V., 2003. Taxonomic reliability of leaf and fruit morphological characteristics of the

Pyrus L. taxa in Slovakia. Zahradnictvi. Horticultural Science. 30, 98-107.
Picard, R. R., and Cook, R. D., 1984. Cross-validation of regression models. J Am Stat Assoc. 79, 575-583.
Rencher, A.C., Christensen, W. F., 2012. Methods of multivariate analysis. New York: John Wiley. p. 800.
Rubzov, G. A., 1944. "Geographical distribution of the genus *Pyrus* and trends and factors in its evolution," American Naturalist, 78, 358-366.
Silva, G. J., Souza, T. M., Barbieri, R. L., & Costa de Oliveira, A. (2014). Origin, domestication, and dispersing of pear (*Pyrus* spp.). Advances in Agriculture.
Sugiyama, M., 2007. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. J Mach Learn Res. pp. 27-106.
Vavilov, N. I., 1930. Wild progenitors of the fruit trees of Turkestan and the Caucasus and the problem of the origin of fruit trees. in Proceedings of the 9th International Horticultural Congress.The Royal Horticultural Society. pp. 271-286.
Vavilov N.I., 1951. The origin, variation, immunity and breeding of cultivated plants. Selected writings of N.I. Vavilov, translated from Russian by K. Starr Chester. Chronika Botanika, 13, 1-366.
Vavilov, N. I., 1992. Origin and Geography of Cultivated Plants. Cambridge University Press, Cambridge, UK. Vol. 15.
Venora, G., Grillo, O., Ravalli, C., Cremonini, R., 2009. Identification of Italian landraces of bean (*Phaseolus vulgaris* L.) using an image analysis system. Sci. Hortic. 121, 410-418.
Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., and Maxted, N., 2013. A prioritized crop wild relative inventory to help underpin global food security. Biological conservation. 167, 265-275.
Terpò, A., 1992. Qw taxa in Hungav, and their practical importance. Zhaiszia. KosiGe. 2: 41-57.
Volk, G. M., Richards, C. M., Henk, A. D., Reilley, A. A., Bassil, N. V., and Postman, J. D., 2006. Diversity of wild *Pyrus communis* based on microsatellite analyses. Journal of the American Society for Horticultural Science. 131, 408-417.
Wolko, L., Antkowiak, W., Lenartowicz, E., and Bocianowski, J., 2010. Genetic diversity of European pear cultivars (*Pyrus communis* L.) and wild pear (*Pyrus pyraster* (L.) Burgsd.) inferred from microsatellite markers analysis. Genetic resources and crop evolution. 57, 801-806.
Zielinski, Q. B., and Thompson, M. M., 1967. "Speciation in *Pyrus*: chromosome number and meiotic behavior," Botanical Gazette. 128, 109-112.
Zukovskij, P. M., 1962. Cultivated Plants and Their Wild Relatives, Farnham Royal, London, UK, Translated by P.S. Hudson.

CHAPTER 4: Characterisation of microsatellite loci in Sardinian pears (*Pyrus communis* L. and *P. spinosa* Forssk.).

1. Introduction

The cultivation of the pear trees in Sardinia has a long history originating from its spread in ancient times which is extensively documented by the historical bibliography. Pliny reminds him of 39 varieties. Aristotle exalts fruit from Sardinia, as well as other authors of antiquity (Pausania, Polibio, Diodoro) (Agabbio, 1994). In the Roman world, Virgilio, Celsus, Columella and Pliny evidence the splendour on the cultivation of pear trees in Italy and providing us with valuable indications on the cultivar then widespread and known. More than a hundred pear varieties in Sardinia are now recognized, of ancient origin with morphological and qualitative characteristics that have allowed the differentiation of a rich heritage of varieties (Agabbio et al., 1986).

However, in the centuries in Sardinia, a rich local germplasm has been differentiated (Agabbio, et al., 1986, 2015) which presents varieties with agronomic and qualitative characteristics that can be found in modern fruit groves, increasingly conditioned by competition and the difficulty of ensuring comparable income with other economic activities (Muresu et al., 1997). The importance of local cultivars lies in the high genetic variability that characterizes them and has allowed them to adapt to adverse environmental conditions over the years, in which they have however, managed to have a good yield, thus protecting populations from complete crop loss. (Hammer et al., 2003).

Despite the low yields and unsatisfactory quality of the old local pear varieties, they represent a precious source of genetic variability for many traits that are not yet exploited by current breeding programs as already done for apple trees (Liang et al., 2015). In addition, an efficient molecular characterisation of pears varieties, would be required to manage this genetic heritage by identifying clonal relationships, synonyms and homonyms, their propagation, and correct taxonomy. Microsatellite or simple sequence repetition (SSR) is omnipresent in eukaryotic genomes, thus becoming a marker of choice for both animal and plant species, due to their reproducibility, multi-allelic nature, codominant inheritance, relative abundance, and good genome coverage (Weber and Mat, 1989; Powell et al., 1996).

In recent years, the work of SSR primers focused on the varietal heritage of *Pyrus pyrophilia* Japanese, leading to a small study on the European *Pyrus* SSR varieties (Bao et al., 2007, Fernàndez-Fernàndez et al., 2006, Yamamoto et al., 2002a, b, c). Some SSRs have been standardized for pear varieties in Europe (Evans et al., 2009) in order to compare germplasm

collections. The low morphological diversity of fruits, the lack of distinctive features among species, and the high crossing potential have made it very difficult to estimate the genetic diversity of the different *Pyrus* genres.

The aim of the present study was to estimate the genetic relationship among wild and local pear varieties from Sardinia and national and international ones, using SSR markers. This molecular characterization will: identify the genetic diversity; investigate cases of homonymous and/or synonymous genotypes that are difficult to distinguish using standard morphological descriptors and increase the *Pyrus* molecular marker datasets.

2. Materials and Methods

2.1. Germplasm material, DNA extraction and SSR genotyping

A list of the 109 varieties analysed is included in Table 1. The samples including 78 local varieties from the CNR-ISPA catalogue field (Nuraxinieddu, Oristano, Sardinia) and 5 varieties from the Fo.Re.S.T.A.S catalogue field (Monte Pisanu, Bono, Sardinia). Moreover, 24 *P. spinosa* collected in Sardinia and 3 international reference cultivars ('Williams', 'Abate', 'Keiser') (Table 1).

For each variety, genomic DNA was extracted from 50mg of young freeze-dried leaves following the standard CTAB protocol (Maguire et al., 1994). Genomic DNA was quantified by Nanodrop™ ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and diluted to 10ng/µl.

A set of nine SSRs was chosen from the HiDRAS (High-quality Disease Resistant Apples for Sustainable Agriculture) website (http://users.unimi.it/hidras/), mainly based on their distribution across the pears genome. Forward SSR primers were labelled whit 2'-fluorescent dyes: VIC and FAM (Table 2).

The amplification of markers was performed using PCR protocol for apple cultivar as described by Liang et al. (2015). PCR was performed in a 10µl reaction mixture containing 1× reaction buffer (Applied Biosystems, Foster City, CA, USA), 1.5 mM $MgCl_2$, 0.2 mM dNTPs (Fermentas, Lithuania), 5n Meach SSR locus-specific primer,76 nM dye-labeled tag-F and unlabeled tag-R primers, 10ng genomic DNA and 0.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems). The PCR reactions were carried out in a 2720 thermal cycler (Applied Biosystems) with the following amplification protocol: an initial denaturation step of 10 min at 95°C, followed by 20 cycles of 30 s at 92 °C, 90 s at 60 °C, and 60 s at 72 °C, and then 40 cycles of 15 s at 92 °C, 30 s at 45 °C, and 60 s at 72 °C, with a final extension step of 10 min at 72 °C (Fig.1).

Multi-pooling groups (MPG) of SSRs labelled with the four different fluorescent dyes were designed for SSR genotyping on an ABI PRISM 3730 DNA analyzer. SSRs were pooled by mixing PCR products labelled with different dyes in a ratio of 1:1:1:2 for VIC: FAM; 3µl of the PCR products mixture was added to 7µl of formamide containing 0.2µl of GeneScan500 LIZ size standard (Applied Biosystems). Fragments were visually analysed and scored by using Peak Scanner v.1.0 (Applied Biosystems). To monitor the reproducibility in different amplifications, three reference cultivars, ‘Abate’, ‘Kaiser’, ‘William’, were included in each single run.

Table 1. In uppercase varieties of *P. communis* (*P.c*) and (*P.c**) and in lowercase varieties of *P.synosa* (*P.s*) analysed in this study. In bold, reference varieties.

Variety	Code	Species	Variety	Code	Species
COZZONE AINU	1	*P.c*	DE JERRU NURAXINIEDDU	56	*P.c*
SANGUIGNA	2	*P.c*	SPADONE	57	*P.c*
CODIS	3	*P.c*	CAMUSINA CAGLIARI	58	*P.c*
APPICADORZA	4	*P.c*	BRUTTA E BONA	59	*P.c*
OLIENA 3	5	*P.c*	LIMONI	60	*P.c*
REGINA	6	*P.c*	ARANGINU	61	*P.c*
REUSPU NUCHIS	7	*P.c*	MUSCADEDDU	62	*P.c*
SANTA BARBARA	8	*P.c*	COIMBINU	63	*P.c*
CAULI	9	*P.c*	SAMENTINA	64	*P.c*
SA MADDALENA	10	*P.c*	DI MUCARGIA	65	*P.c*
ALVEGHINA	11	*P.c*	VACCHESA	66	*P.c*
LACONI 1	12	*P.c*	CAMUSINA SASSARI	67	*P.c*
REALE	13	*P.c*	BAU	68	*P.c*
ENOSA	14	*P.c*	FUNTANA SONES	69	*P.c*
DE PULEU	15	*P.c*	SITZIA	70	*P.c*
SAN GIOVANNI MIGLIORATA	16	*P.c*	MEANA	71	*P.c*
CAMUSINA GRANDE	17	*P.c*	BRAGAMOTTA	72	*P.c*
ROMA	18	*P.c*	ARMUNGIA	73	*P.c*
BUTIDU DE AUSTU	19	*P.c*	MUSCADE DI LACONI	74	*P.c*
MELA LACONI	20	*P.c*	COIMBINU	75	*P.c*
E' DONNA	21	*P.c*	ABDRO	76	*P.c*
OLIENA 1	22	*P.c*	ANTONI SALE	77	*P.c*
SAN GIOVANNI	23	*P.c*	PIRA CONTU E DOM	81	*P.c**
CAMUSINA BONARCADO	24	*P.c*	PIRA DI URZULE'	82	*P.c**
PREIDERINA	25	*P.c*	PIRA MODDE ARDA	83	*P.c**
BIANCA	26	*P.c*	PIRA DI ORTINE URZU	84	*P.c**
PARENA	27	*P.c*	PINA CONA ARRUBIA	85	*P.c**
MELA LACONI	28	*P.c*	**Abate**	78	*P.c*
SANTA MARIA	29	*P.c*	**Kaiser**	79	*P.c*
LACONI 4	30	*P.c*	**William**	80	*P.c*
SANT'ANNA	31	*P.c*	Meana	86	*P.s*
SAN DOMENICO	32	*P.c*	Santadi	87	*P.s*
CAMPANA	33	*P.c*	Villacidro	88	*P.s*
OLIENA 2	34	*P.c*	Orroli	89	*P.s*
PIBIRI	35	*P.c*	Escalaplano	91	*P.s*
OZALE	36	*P.c*	Su Carropu	92	*P.s*
MAMOI	37	*P.c*	Villasalto	93	*P.s*
DI OROTELLI	38	*P.c*	Ardara	94	*P.s*
AMMIABI	39	*P.c*	Berchidda	95	*P.s*
PIRINGINU	40	*P.c*	Tuaredda	96	*P.s*
CARBUDRAXIA	41	*P.c*	Domusnovas	99	*P.s*
CRACCHERA	42	*P.c*	Alà dei Sardi	100	*P.s*
MELI	43	*P.c*	Perfugas	101	*P.s*
BERGAMOTTA	44	*P.c*	Tonara	102	*P.s*
MUSCADELLU	45	*P.c*	Narcao	103	*P.s*
LACONI 2	46	*P.c*	Oliena 5	104	*P.s*
DE SU DUCA	47	*P.c*	Monte Nai	105	*P.s*
DI PERFUGAS	48	*P.c*	Feraxi	106	*P.s*
CAMUSINA PRECOCE	49	*P.c*	Laconi	107	*P.s*
BUTIRRA DE JERRU	50	*P.c*	Tempio	108	*P.s*
MEBA	51	*P.c*	Samugheo	110	*P.s*
ARMUNGESA	52	*P.c*	Quirra	111	*P.s*
ARBANISCO	53	*P.c*	Nuoro	112	*P.s*
MUSCADEDDU DE JERRU	54	*P.c*	Monte Arci	113	*P.s*
IL DE NOA	55	*P.c*			

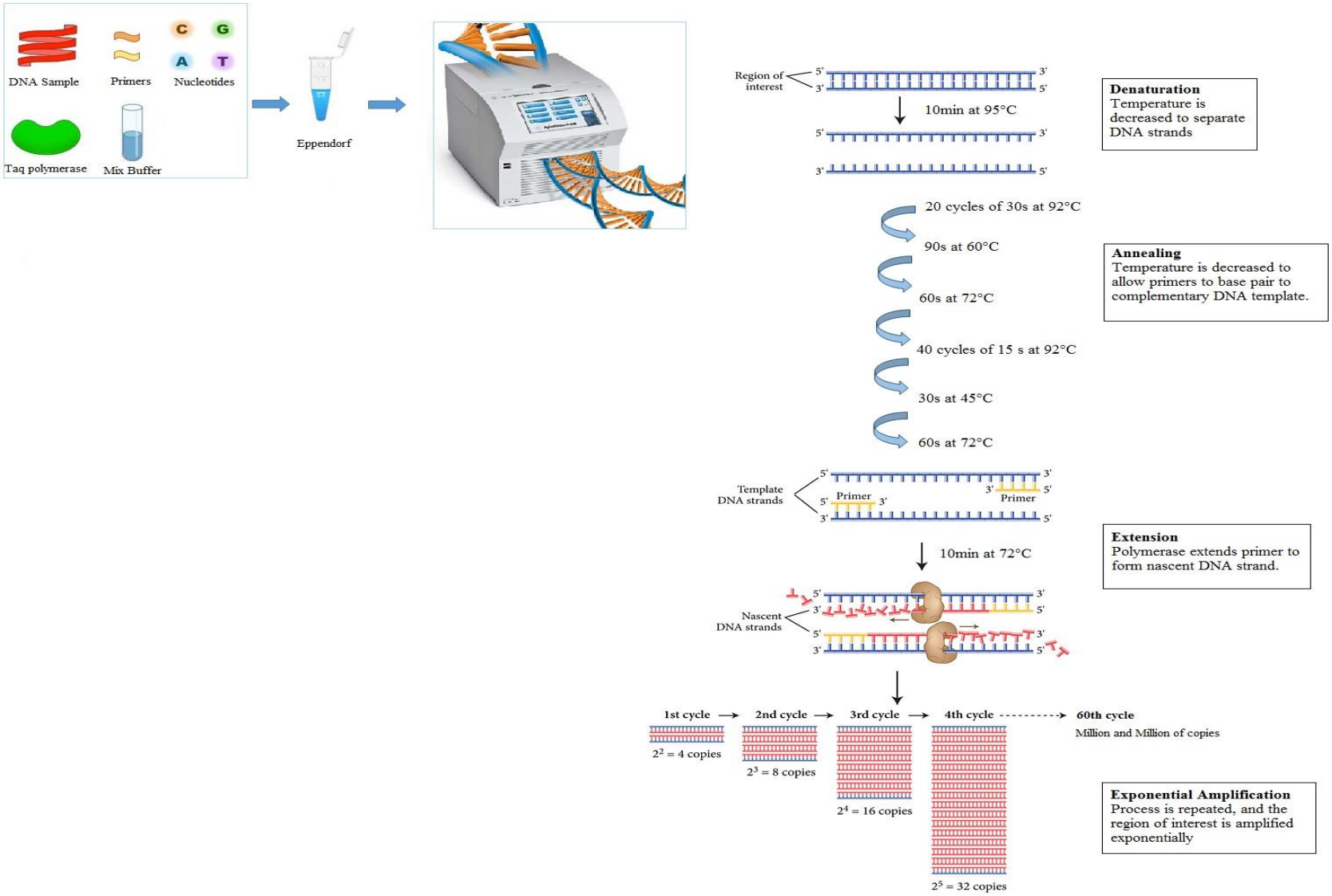


Fig.1 PCR protocol used in this study.

2.2. SSRs Marker

The PCR products were amplified using the ABI 3730 DNA analyser made available by Sant'Orsola Hospital in Bologna. To facilitate a highly parallel SSR genotyping, the products of nine SSRs were assigned to two multi-pooling groups (MPGs). The procedure consists of making two different SSR amplification products with two fluorescent dyes (VIC and FAM), then combining them with single MPG based on the range of their allele length. If on the markers panel, the two SSR alleles will be separated, without overlapping each other, the two SSRs will be strongly recommended as a multi-pooling group.

The protocol used to prepare SSR products for electrophoresis it was as follows: PCR products labelled with different fluorescent dyes were pooled in a total volume of 25 µl at a ratio of 1 : 1 : 1 : 2 for VIC: FAM, according to differences in the relative fluorescence of each fluorophore; 3 µl of PCR products mixture was added 7 µl of deionized formamide containing 0.2 µl of GeneScan500 LIZ size standard (Applied Biosystems). This multi-pooling mixture containing two different SSR amplification products was separated on a capillary electrophoresis ABI 3730 DNA sequences. Raw fragment size data were analysed using Peak Scanner Software v1.0 (Applied Biosystems) (Fig. 2), and all automated results were manually reviewed.

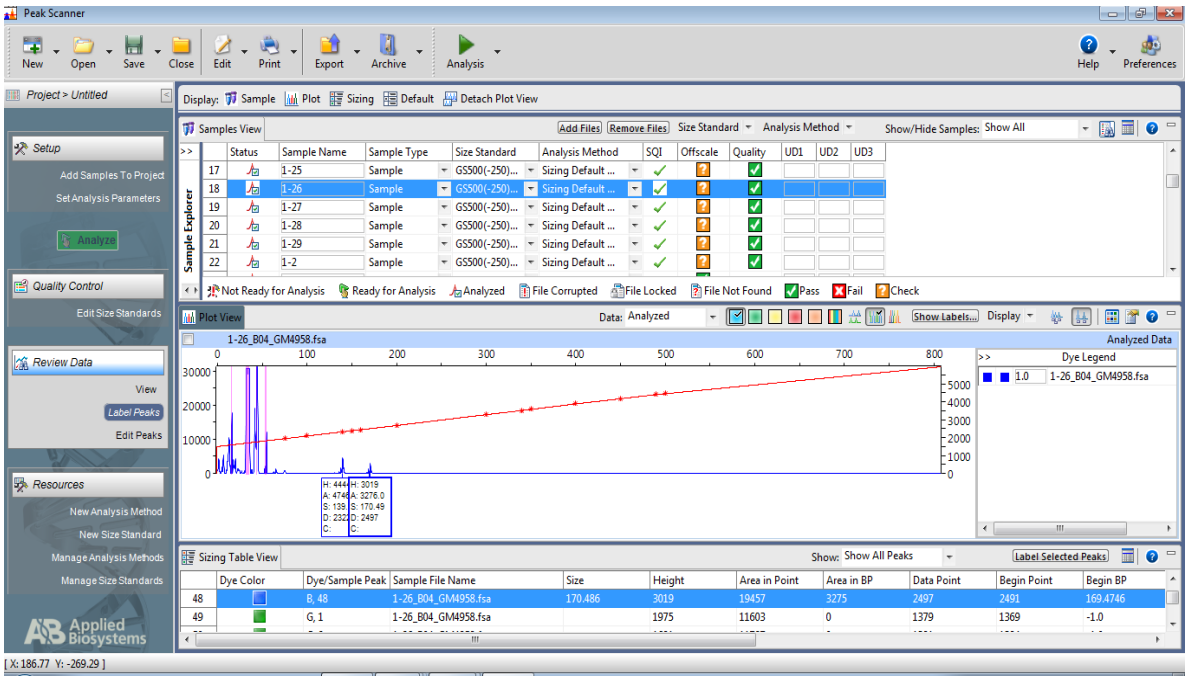


Fig. 2 Peak Scanner Software v. 1.0.

2.3. Data analysis

To investigate the different levels of polyploidy the software STRUCTURE (Pritchard et al., 2000),(https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4 /html/structre.html) was used in a dataset formatted by 454 unique pear accesses created by Fernandez-Fernandez in 2008 and by 22 unique Sardinian pear genotype.

Moreover, through this software it was possible to detect different genetic statistical information including: the number of alleles per locus (A); the expected heterozygosity (He) and Wright's fixation index (Fst). To determine the possible presence of null alleles, the genetic uniqueness of each varieties, and to quantify redundancy, the multilocus DNA profile of all the varieties was compared pairwise using the program Cervus ver. 3.0 (Kalinowski et al., 2007) under the identity analysis with the setting of minimum number of matching loci is 15 and allow 0 mismatch. The collected data were organized in a square matrix in which the code '0' was used for allele absence and the code '1' for presence of allele (code for missing data was 9).

The genetic distance between cultivars was then calculated through the DICE coefficient (Dice 1945) by the SimQual procedure of NTSYSpc 2.0 (Rohlf 1994). The dendrogram was constructed using the unweighted pair group method average (UPGMA) clustering and drawn with NTSYSpc 2.0 program (Rohlf, 1994). To investigate the population structure using the genotype data of our dataset, a Bayesian clustering method (Pritchard et al. 2000a; Falush et al. 2003) was applied using the STRUCTURE v.2.3.4 free software.

Through this method it was possible to detect the underlying genetic population among a set of individuals genotyped at multiple markers and computes the proportion of the genome of an individual originating from each inferred population (quantitative clustering method). For the structure analysis, putative triploid varieties were removed from the dataset. Previous information was not used to define clusters. Independent runs were done by setting the number of clusters (k) from 2 to10. Each run comprised a burn-in length of 10000 followed by 100000 MCMC (Monte Carlo Markov Chain) replicates. To estimate the most likely number of clusters (k), Structure Harvester software was used (taylor0.biology.ucla.edu/structureHarvester/).

For each cluster, the varieties have been divided for Q> 0.8 compared to the others that are to be assigned to a cluster only for a greater Q, but that are not unequivocally.

Subsequently, the statistic Δk was calculated to identify the number of clusters (k), using the Structure Harvester software (taylor0.biology.ucla.edu/structureHarvester/), using a zip file analysis with all the simulation results for a structure hypothesis from two to ten populations. In addition, subsequently, each of the k groups was analysed individually by STRUCTURE

v.2.3.4. (Pritchard et al., 2000). Finally, for each of the clusters, the fixation index was calculated.

3. Results

We created a dataset for Sardinian pear varieties composed of 83 local varieties of *P. communis*, 24 *P. spinosa* and three international reference cultivars ("William", "Kaiser" and "Abate"), (Table 1). The nine SSR markers showed a clear and easy readable peak amplification. Peaks evaluation of the varieties used as references has allowed ensuring that the SSRs used have been reliable markers to assess the diversity of pear germplasm as confirmed by Seich et al. (2012). Thus, a total of nine polymorphic loci were statistically analysed, giving a total of 194 alleles for 109 varieties, with a minimum of 14 (CH04e03) to a maximum of 26 different alleles (CH01d09) (Table 2). In addition, the analysis of allelic frequencies within the locus examined showed uneven distribution (low or high) with values between 0.006 (CH01d09 and Ch01f07a) and 0.651 (CH04e03) (Table 2). For all the SSR loci studied, 15 specific rare alleles were identified with a minimum of two (EMPc11, EMPc117) to maximum 17 (CH01f07a). Were identified 78 varieties with three alleles in at least one locus, of which 20 varieties with three alleles in two loci, six varieties with three alleles in three loci, 10 varieties with three alleles in four loci, five varieties with three alleles in five loci and two varieties with three alleles in six loci. The remaining 36 varieties have presented only one locus with three alleles (Table 3). After a careful examination of our dataset, we found three varieties with four alleles in a single locus: "De Puleu", "Buttidu de Austu" and "Mela". In our work, the number of alleles present in each locus and the range of values in which they are found differs from that found in the literature (Table 4). The overall allelic diversity have showed by the set of 9 SSRs used revealed a high genetic variation in the estimated germplasm. Respect to other large-scale studies of the genetic diversity of pear, the average number of alleles per locus (21.5) was different (9.4) than that reported in the literature (Table 3). Comparison of SSR profiles have showed four groups of diploid varieties that had the same SSR profile, the size of these groups consisting of two varieties with different names have showed the presence of homonymous in our dataset. In addition, one of these groups was constituted of a variety used as reference "William" and the variety called "Reale" in our dataset. Cluster analysis carried out on the Sardinian dataset has allowed identifying different homonymous varieties: "Cazzone Ainu" with "Bragamotta"; "Enosa" with "E' Donna"; "Reale" with "William" and "Acueghina" with "Bau" it is confirming the comparison of SSR profiles (Fig. 3). Furthermore, the varieties "Vacchesa" and "Pira

Ortine"; "Laconi2" and "Camusina precoce"; "White" and "Mamoi"; "Armungesa" and "Di Mulargia" have discriminated for less than 10 alleles placed in different loci (Fig. 3, Table 5).

In addition, the analysis showed that the *P.communis* group is phylogenetically separated from that of *P. spinosa* although , the varieties of *P.communis* "Pira cona arrubia" and "Pira di urzulè" were resulted phylogenetically closer to varieties of *P. spinosa* respect to other *P.communis* in the data set.

The datasets of international varieties has presented numerous cases of homonymies among the varieties (Fig. 4). Moreover, it has also been shown that there are no great relationships between genotypes of Sardinian and international varieties.

The variety called "San Domenico" in Sardinian dataset showed a single difference allele (110) located on the same locus (CH3d12) with "Beur-Giffard" by Fernàndez-Fernàndez (106), showing a 100% likeness. An important result of the alignment of the two datasets confirmed the homonymies present in Sardinian varieties and the phylogenetical position of *P.spinosa*.

Another, showed a correlation between a small part of Sardinian and Japanese varieties (Fig. 4). The analysis of the allelic frequencies of the entire dataset, including Sardinian and international varieties, showed how many alleles of the Sardinian are in common with the European varieties. Structure Harvester software was used to conducted structure analysis on 449 unique pear genotypes, (including 22 Sardinian single diploid genotypes and a total of 454 international genotypes) showed the formation of four distinct genetic groups. Sardinian varieties were included into a single group, showing a 0.95 Q value (Fig. 5). A second group enclosed the Japanese *P. pyrifolia* varieties (Q = 0.01), a third group including, late varieties (Q= 0.02) and a fourth group composed of the all most famous varieties used in breeding (Q = 0.02) (Fig. 5).

Although the differences between the groups did not show very large values, they were still statistically valid, the level of differentiation between these four groups has been very evident for Sardinian varieties, which unequivocally were distinguished from all others.

The F_{st} was calculated for each single groups: group one had 0.0015; group two 0.3981; group three 0.1232; and group four 0.2322. The value of the fixing index (F_{st}) for each gene groups was found to be very low.

As far as, *He*, got higher values with group one values of 0.834; for group two of 0.6305; for group three of 0.7069 and for group four of 0.6308. Evanno's statistical analysis Δk has shown unequivocally k = 4 ($\Delta k = 150$) as the most likely stratification level of the population (Fig. 6). Finally, the Baterplot has charted clearly shows the presence of four distinct gene groups (Fig. 7), confirming the hypothesis of k = 4.

Table 2. Allele frequency of 9 SSRs in Sardinian *Pyrus* germplasm with 454 varieties. Rare alleles are indicated in bold, unique alleles are evidence by underlining.

SSRs Marker		alleles																																	Total
CH01d09	alleles	122	126	128	130	132	134	136	138	140	143	**146**	147	149	151	153	155	157	160	162	165	**167**	168	169	170	172	175								26
	frequency	0.006	0.014	0.109	0.035	0.078	0.027	0.102	0.057	0.062	0.015	0.007	0.034	0.079	0.133	0.043	0.036	0.036	0.053	0.007	0.013	0.007	0.009	0.009	0.010	0.012	0.007								
CH02b10	alleles	116	118	120	**121**	122	124	126	128	130	132	134	136	138	141	143	145	147	150	154	156	160													21
	frequency	0.008	0.051	0.109	0.008	0.121	0.028	0.151	0.124	0.096	0.135	0.020	0.033	0.007	0.020	0.014	0.013	0.011	0.016	0.017	0.007	0.010													
CH03d12	alleles	84	88	90	93	97	**100**	103	106	107	108	110	111	112	113	114	118	121	123	125	126	130	**133**	147	153	155	157	159							27
	frequency	0.007	0.022	0.085	0.007	0.007	0.010	0.010	0.089	0.013	0.129	0.123	0.018	0.060	0.014	0.012	0.044	0.019	0.182	0.012	0.027	0.013	0.007	0.008	0.007	0.058	0.008	0.010						
CH04e03	alleles	167	180	**182**	184	188	190	194	196	197	198	200	203	205	207																				14
	frequency	0.007	0.651	0.008	0.013	0.009	0.023	0.011	0.011	0.050	0.030	0.017	0.011	0.121	0.037																				
CH01f07a	alleles	**91**	**111**	175	177	179	180	181	183	185	187	188	189	191	192	193	194	195	197	198	199	200	201	**202**	203	205	206	207	208	209	214				30
	frequency	0.007	0.007	0.126	0.006	0.007	0.171	0.007	0.082	0.040	0.055	0.056	0.087	0.031	0.021	0.038	0.016	0.023	0.047	0.015	0.010	0.010	0.007	0.009	0.012	0.011	0.060	0.010	0.015	0.007	0.009				
CH05c06	alleles	83	87	91	93	95	97	99	100	101	103	104	105	107	109	111	119	**142**	**149**																18
	frequency	0.011	0.258	0.311	0.041	0.024	0.081	0.042	0.007	0.022	0.027	0.007	0.014	0.065	0.028	0.036	0.013	0.007	0.007																
EMPc11	alleles	**113**	**115**	123	136	138	140	142	143	144	**145**	146	**147**	149	151	153	155	157	171																18
	frequency	0.007	0.007	0.007	0.080	0.078	0.060	0.170	0.008	0.072	0.010	0.016	0.007	0.197	0.035	0.133	0.062	0.035	0.007																
EMPc117	alleles	82	88	91	92	95	97	99	101	103	105	107	109	111	113	115	117	118	119	121	123	125	133	135											23
	frequency	0.007	0.064	0.037	0.027	0.067	0.033	0.013	0.010	0.013	0.010	0.092	0.033	0.078	0.177	0.042	0.203	0.009	0.023	0.014	0.013	0.010	0.010	0.018											
GD147	alleles	112	121	123	125	127	129	131	133	135	137	139	141	148	150	152	158	164																	17
	frequency	0.007	0.476	0.142	0.042	0.068	0.040	0.081	0.007	0.031	0.011	0.007	0.016	0.007	0.038	0.007	0.007	0.013																	
Total																																			194

Table 3. Triploids and tetraploids found in the Sardinian set of varieties.

Accession name	SSRs marker with three, four and five alleles									Total	
	EMPc117	EMPc11	CH01d09	CH03d12	CH04e03	CH01f07a	CH05c06	GD147	CH02b10	Tri (X)	Tetra (XX)
CODIS			X							1	
APPICADORZA					X					1	
REGINA	X	X			X	X				4	
RUSPU NUCHIS	X									1	
SANTA BARBARA		X	X		X					3	
CAULI		X								1	
SA MADALLENA		X	X		X					3	
LACONI 1		X					X			2	
DE PULEU	XX										1
SAN GIOVANNI MIGLIORATA	X									1	
BUTIDU DE AUSTU		X			XX		X			2	1
MELA LACONI			X		X					2	
PREIDERINA			X							1	
BIANCA		X	X	X			X			4	
PARENA			X	X						2	
MELA		X				XX				1	1
LACONI4									X	1	
OLIENA2	X		X		X		X	X		5	
PIBIRI	X									1	
MAMOI		X	X	X			X	X		4	
DI OROTELLI			X		X					2	
PIRINGINU		X		X						2	
CARBUDRAXIA	X									1	
MELI				X			X			2	
BERGAMOTTA								X		1	
MUSCADELLU	X		X	X			X	X		5	
LACONI2	X			X						2	
DE SU DUCA						X	X			2	
CAMUSINA PRECOCE						X				1	
MEBA		X	X	X		X	X	X		6	
ARMUNGESA		X	X					X		3	
ARBANISCO				X						1	
MUSCADEDDU DE JERRU						X		X		2	
IL DE NOA		X								1	
DE JERRU NURAXINIEDDU	X		X				X	X		4	
SPADONE						X				1	
CAMUSINA CAGLIARI	X		X	X		X	X			5	
BRUTTA E BONA				X						1	
LIMONI	X		X	X					X	4	
MUSCADEDDU							X			1	
SANT'ANDREA			X	X		X	X			4	
SAMENTINA				X						1	
DI MUCARGIA	X	X	X				X	X		5	
VACCHESA	X		X			X	X			4	
CAMUSINA SASSARI	X	X	X				X	X	X	6	
BAU	X									1	
FUNTANA SONAS	X			X		X				3	
SITZIA	X	X	X	X			X			5	
MEANA	X									1	
BRAGAMOTTA	X					X				2	
MUSCADE DI LACONI	X						X			2	
ANTONI SALE	X									1	
COIMBINU							X			1	
PIRA DI ORTINE URZU	X		X	X			X			4	
PIRA CONTU E DOM		X								1	
PIRA CONA ARUBIA	X									1	
PIRA DI URZULE'			X							1	
Meana	X									1	
Santadi	X									1	
Villacidro	X									1	
Orroli									X	1	
Feraxi	X			X		X	X			4	
Laconi	X			X						2	
Samugheo	X									1	
Nuoro	X									1	
Escalaplano		X				X			X	3	
Su carroppu	X	X	X							2	
Villasalto						X				1	
Ardara		X	X						X	3	
Berchidda			X						X	2	
Tuaredda			X							1	
Domusnovas	X									1	
Alà dei sardi						X		X		2	
Perfugas	X		X							2	
Tonara	X									1	
Narcao							X			1	
Armungia	X					X				2	
Oliena 5	X					X				2	

Table 4. Characteristics of the 9 SSRs studied.

SSR	Florescent dye	LG	Primer (5'-3')		Allele size renge (bp)	Allele size renge in this work (bp)	Total N° of alleles	Genotype specific allele	Rare allele	Reference
			Fwd	RVs						
CH01d09	FAM	12	GCCATCTGAACAGAATGTGC	CCCTTCATTCACATTTCCAG	198-384	122-175	26 (6)	5	2	Liebhard et al. 2002
CH02b10	VIC	2	CAAGGAAATCATCAAAGATTCAAG	CAAGTGGCTTCGGATAGTTG	124-142	116-160	21 (12)	5	1	Gianfranceschi et al. 1998
CH03d12	VIC	6	GCCCAGAAGCAATAAGTAAACC	ATTGCTCCATGCATAAAGGG	108-154	84-159	27 (7)	7	2	Liebhard et al. 2002
CH04e03	VIC	5	TTGAAGATGTTTGGCTGTGC	TGCATGTCTGTCTCCTCCAT	194-198	167-207	14 (11)	3	1	Liebhard et al. 2002
CH01f07a	FAM	10	CCCTACACAGTTTCTCAACCC	CGTTTTTGGAGCGTAGGAAC	174-206	91-214	30 (8)	17	3	Liebhard et al. 2002
CH05c06	VIC	5	ATTGGAACTCTCCGTATTGTGC	ATCAACAGTAGTGGTAGCCGGT	104-126	83-149	18 (8)	3	2	Liebhard et al. 2002
EMPc11*	FAM	11	GCGATTAAAGATCAATAAACCCATA	AAGCAGCTGGTTGGTGAAAT	121-161	113-171	18 (9)	2	4	Fernández-Fernández et al. 2006
EMPc117*	FAM	7	GTTCTATCTACCAAGCCACGCT	CGTTTGTGTGTTTTACGTGTTG	76-119	82-135	23 (9)	2	0	Fernández-Fernández et al. 2006
GD147	VIC	13	TCCCGCCATTTCTCTGC	GTTTAAACCGCTGCTGCTGAAC	124-156	112-164	17 (15)	9	0	Hokanson et al. 1998
Total							194 (85)	53	15	

Table 5. Variety with similar genotypes. In bold, allele differentiates the genotype.

Cultivar Name	CH01d09		CH02b10	CH03d12		CH04e03	CH01f07a		CH05c06		EMPc11		EMPc11		GD147	
VACCHESA	130	153	126	93		180	180	**202**	91	107	142		**87**	111	**121**	
VACCHESA	138	0	143	113		205	188	0	101	0	0		107		**131**	
PIRA ORTINE OZU	130	153	126	93	113	180	-1		91	107	**138**		107	**113**	**123**	
PIRA ORTINE OZU	138	**157**	143	**107**	0	205	-1		101	0	142		111	0	**127**	
LACONI2	134		126	**107**	111	180	180		89		**142**		107	113	123	
LACONI2	140		0	**110**	0	207	184		95		144		**111**	0	131	
CAMUSINA PRECOCE	134		126	111		180	180	**194**	89		**140**		107		123	
CAMUSINA PRECOCE	140		0	**126**		207	185	0	93		144		113		131	
BIANCA	132	165	-1	90	113	180	-1		87	107	**138**	144	99		129	133
BIANCA	140	167	-1	111	0	0	-1		95	0	142	0	**107**		131	0
MAMOI	132	165	-1	90	113	180	**175**		87	107	**136**	145	**97**		129	133
MAMOI	140	167	-1	111	0	0	**182**		95	0	142	0	99		131	0
ARMUNGESA	128	157	120	110		194	177		87		136	153	107		122	131
ARMUNGESA	136	0	130	**123**		0	183		107		140	0	113		129	0
DI MUCARGIA	128	157	120	110		**180**	177		87	107	136	153	107	113	**123**	131
DI MUCARGIA	136	0	130	**125**		**196**	183		**105**	0	140	0	**109**	0	129	0

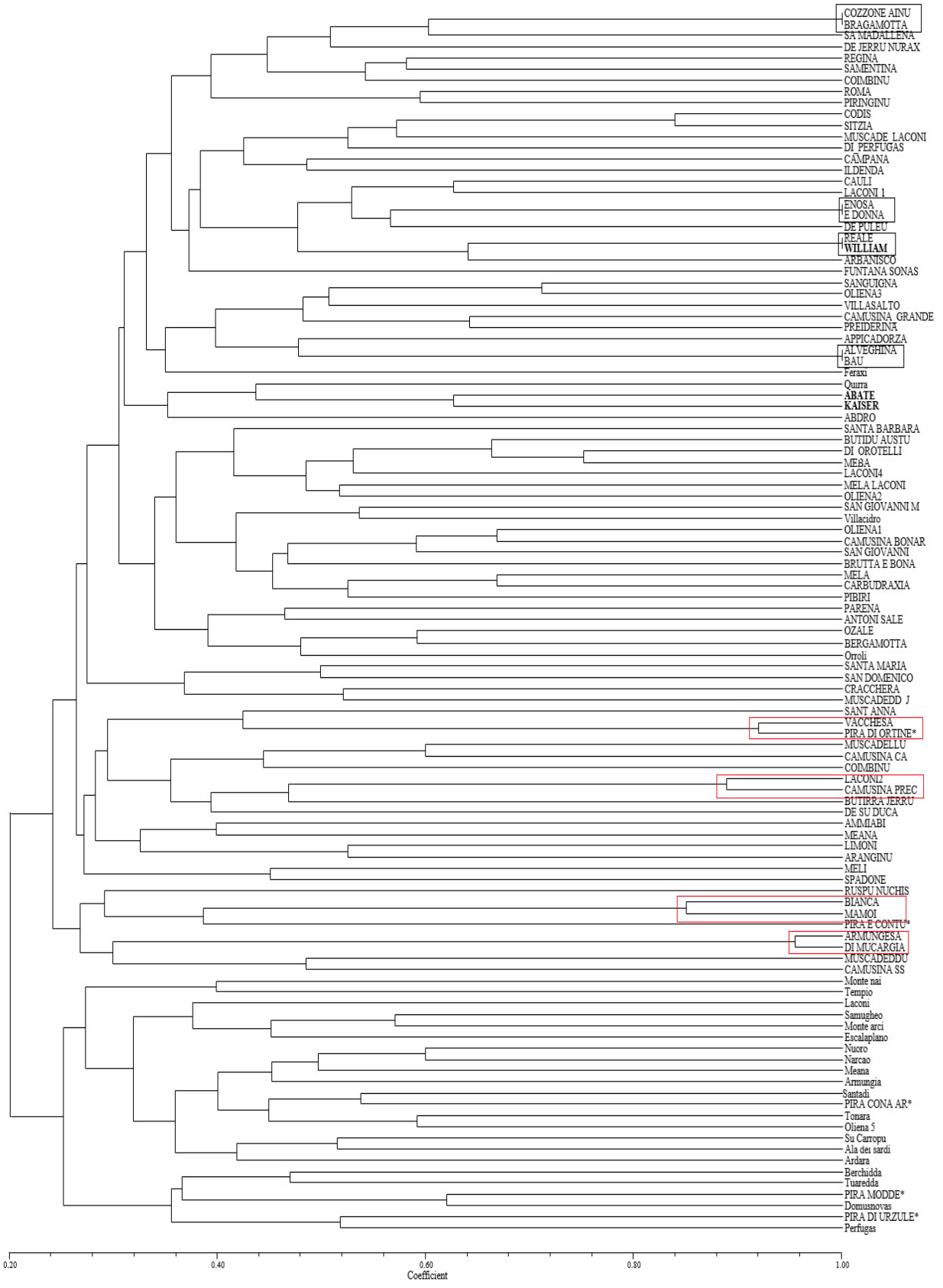
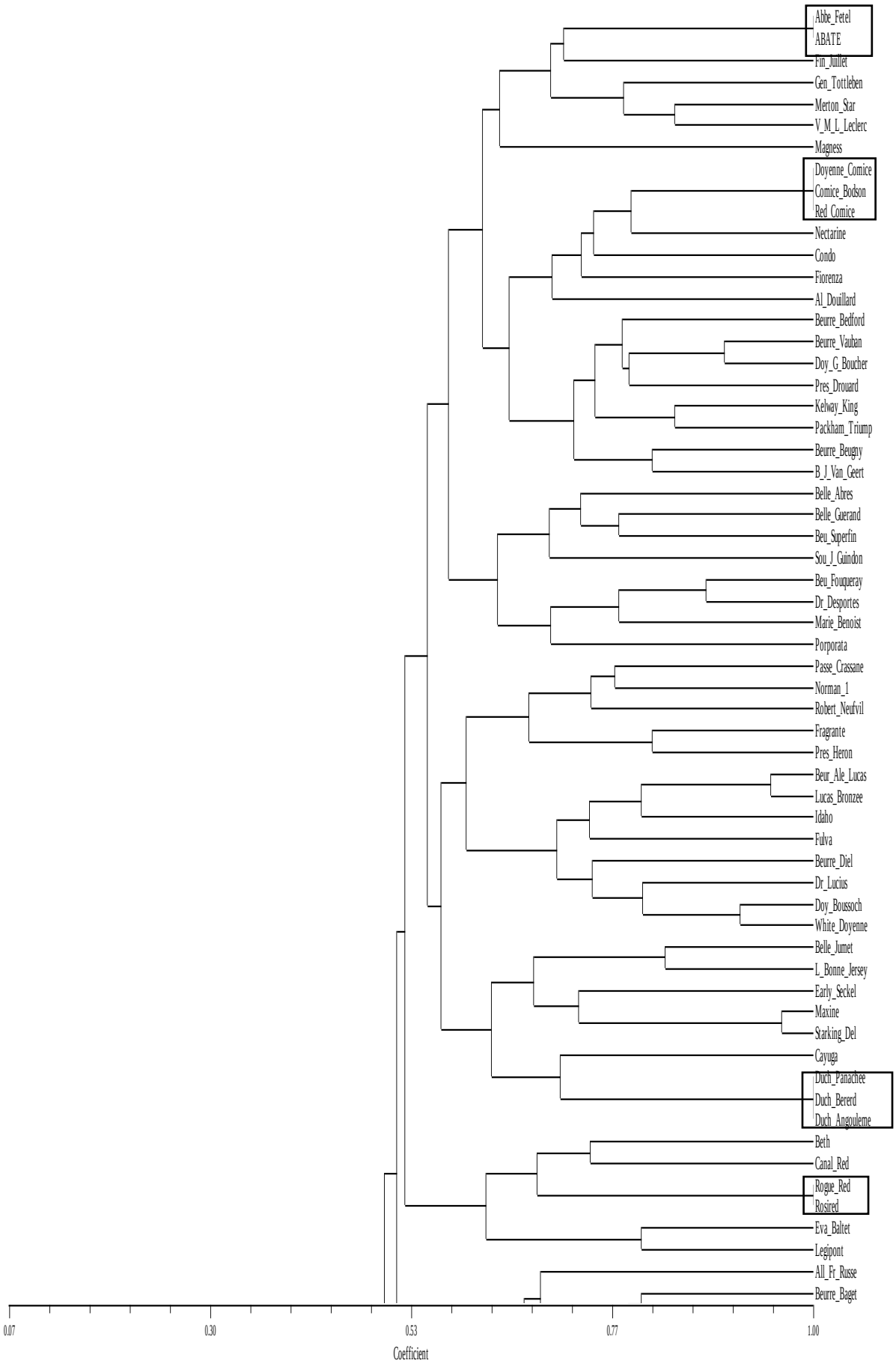
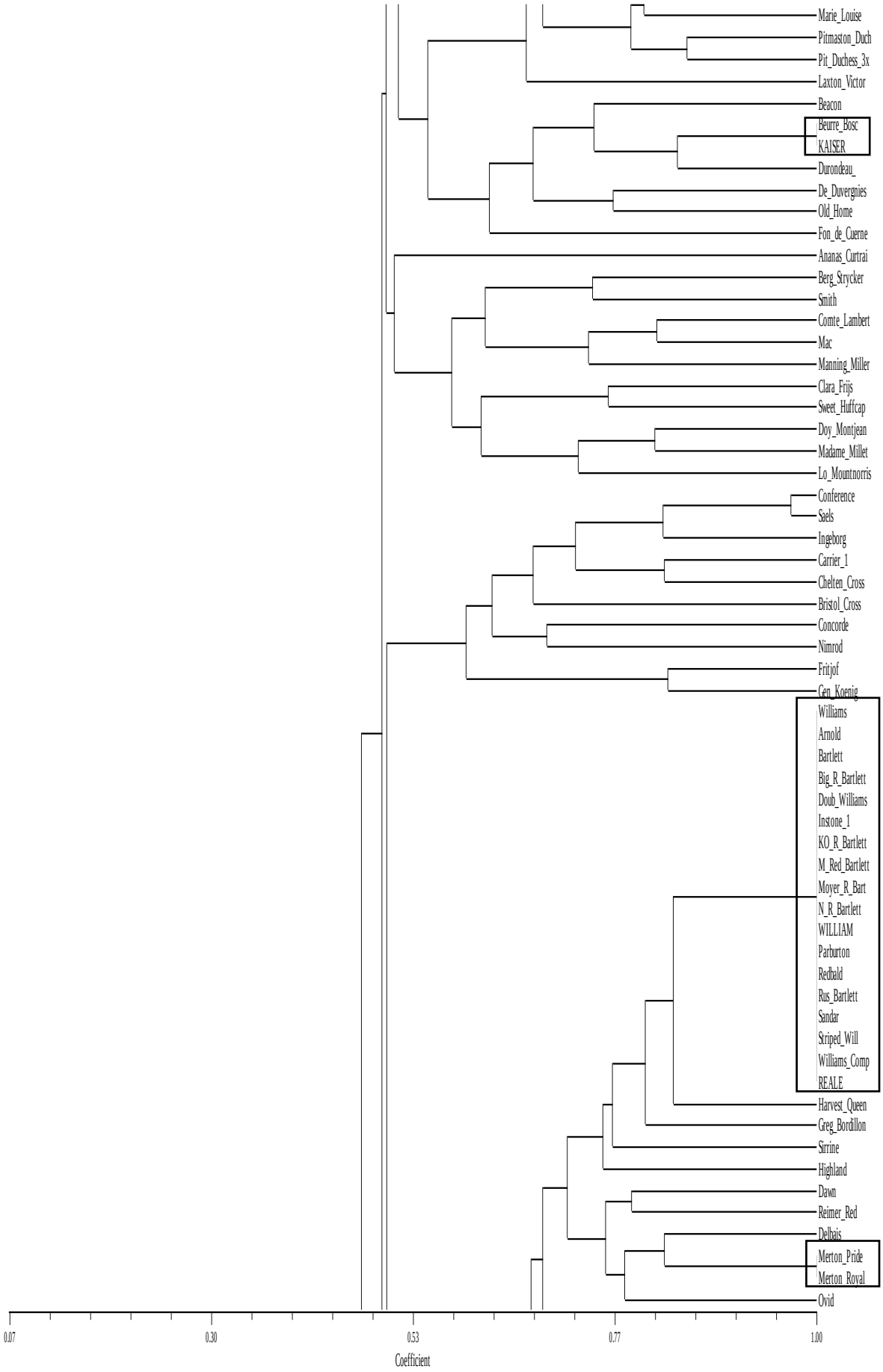
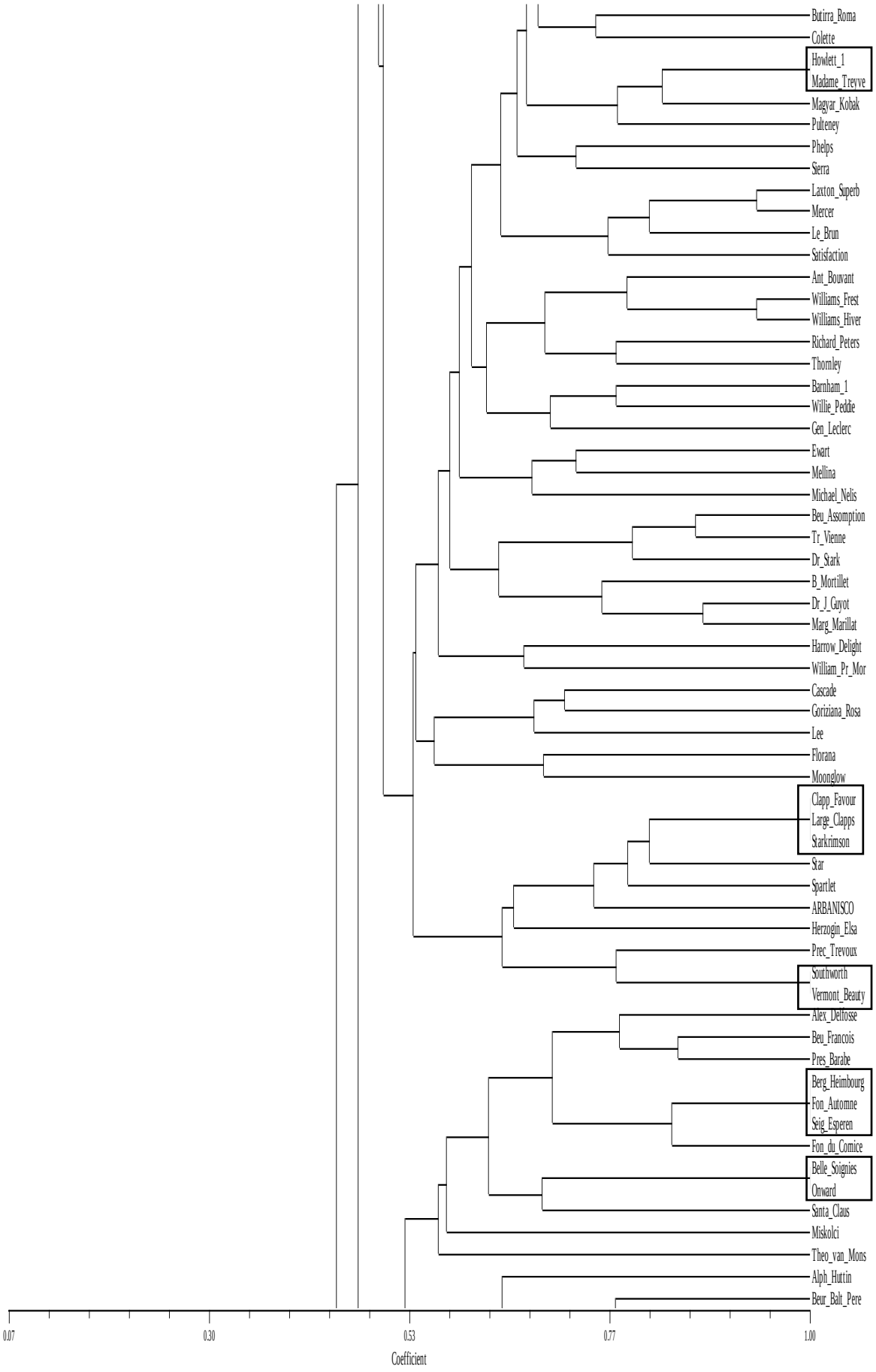
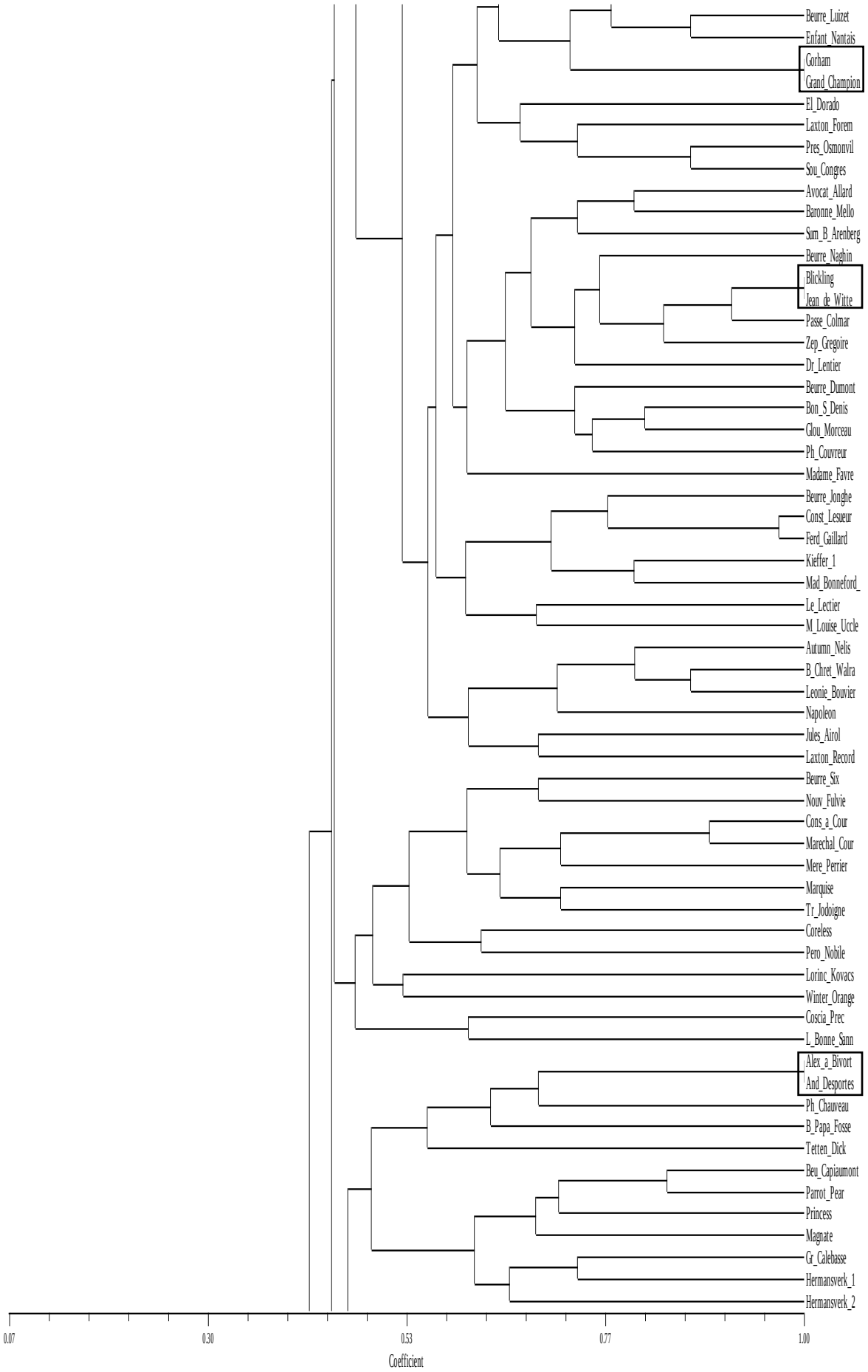


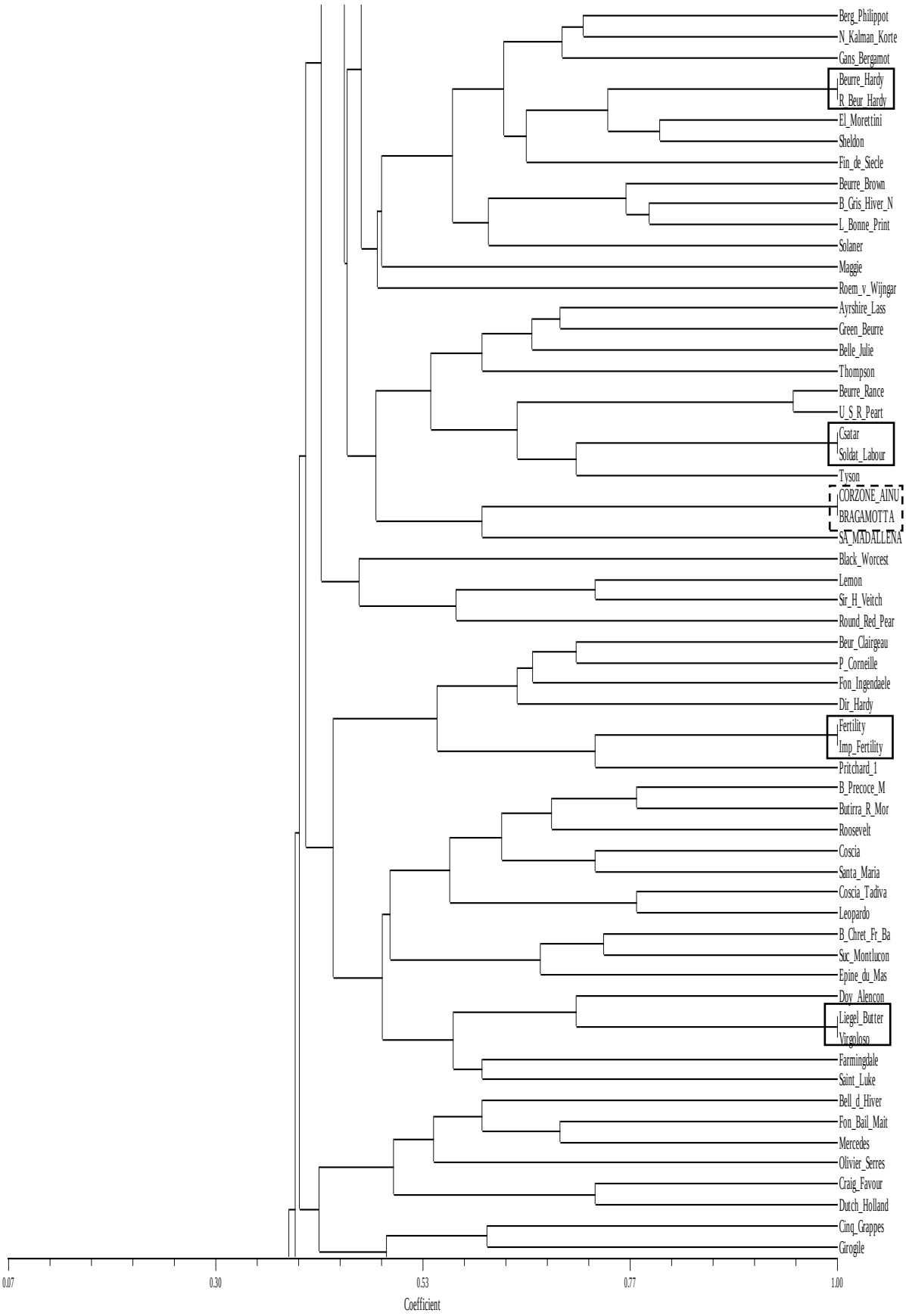
Fig. 3 Dendrogram of 109 pear varieties from Sardinia germplasm based on Cophenetic Correlation coefficient calculated from the allele frequencies of 194 alleles found in 9 SSR loci. In uppercase letters the *P. communis*, in lowercase ones the *P. spinosa* varieties. In the black rectangles, the homonymies found in the Sardinian *Pyrus* accessions. In the red rectangles with the dotted edge, the varieties that are differentiated for less than 10 alleles.

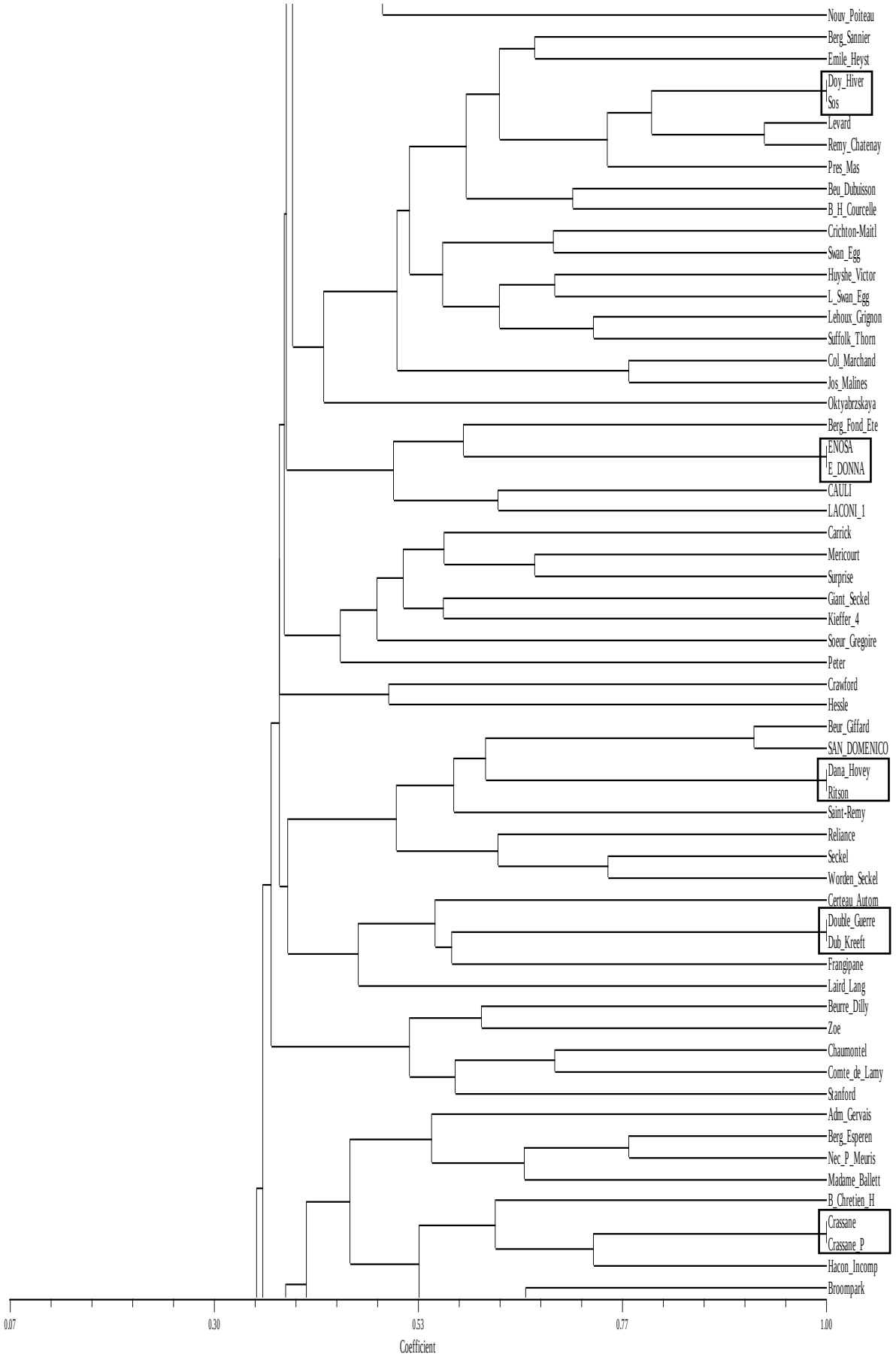

Abbe_Fetel
ABATE
Fin_Juillet
Gen_Tottleben
Merton_Star
V_M_L_Leclerc
Magness
Doyenne_Comice
Comice_Bodson
Red_Comice
Nectarine
Condo
Fiorenza
Al_Douillard
Beurre_Bedford
Beurre_Vauban
Doy_G_Boucher
Pres_Drouard
Kelway_King
Packham_Triump
Beurre_Beugny
B_J_Van_Geert
Belle_Abres
Belle_Guerand
Beu_Superfin
Sou_J_Guindon
Beu_Fouqueray
Dr_Desportes
Marie_Benoist
Porporata
Passe_Crassane
Norman_1
Robert_Neufvil
Fragrante
Pres_Heron
Beur_Ale_Lucas
Lucas_Bronzee
Idaho
Fulva
Beurre_Diel
Dr_Lucius
Doy_Boussoch
White_Doyenne
Belle_Jumet
L_Bonne_Jersey
Early_Seckel
Maxine
Starking_Del
Cayuga
Duch_Panachee
Duch_Bererd
Duch_Angouleme
Beth
Canal_Red
Rogue_Red
Rosired
Eva_Baltet
Legipont
All_Fr_Russe
Beurre_Baget
0.07
0.30
0.53
0.77
1.00
Coefficient

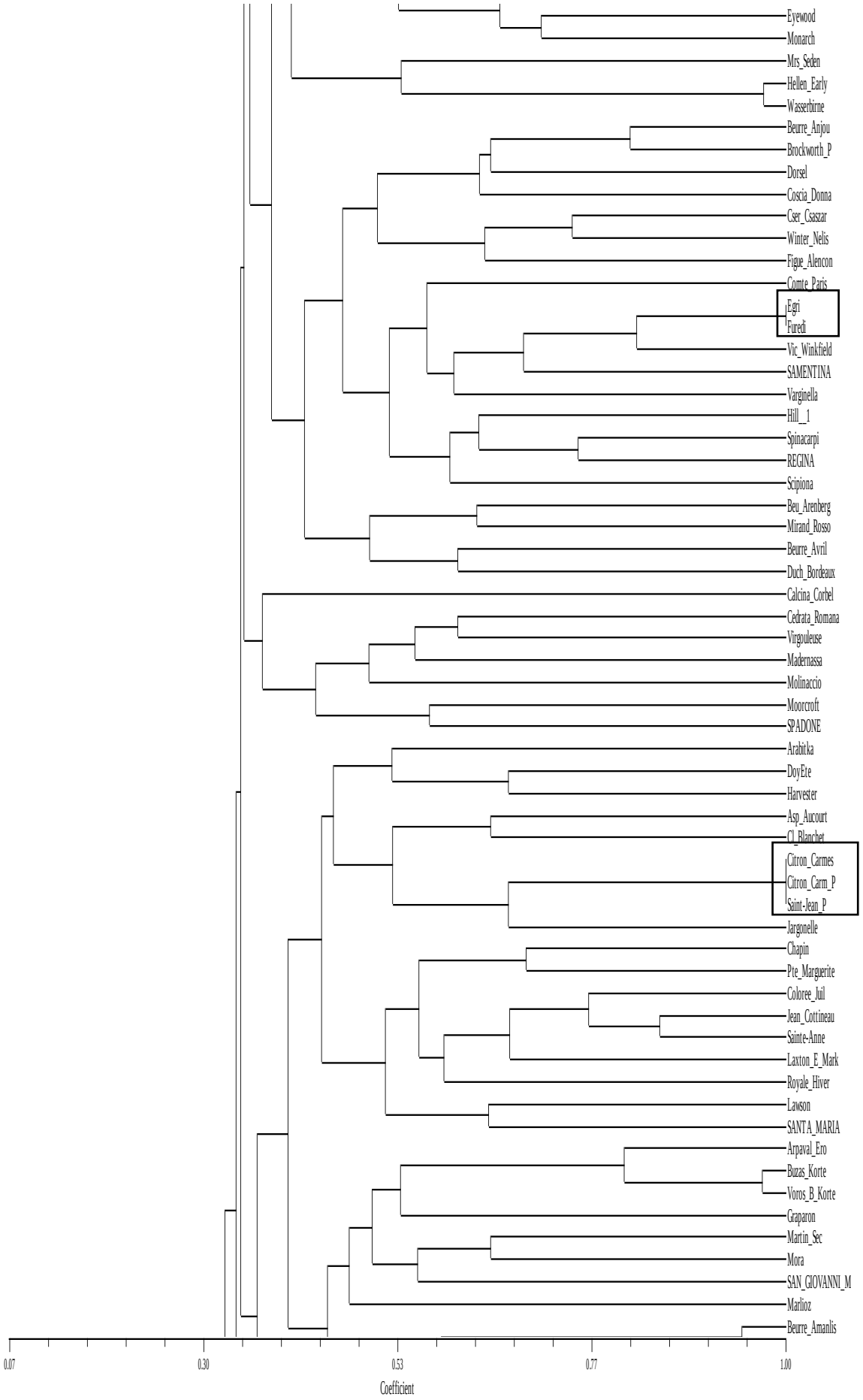

Marie_Louise
Pitmaston_Duch
Pit_Duchess_3x
Laxton_Victor
Beacon
Beurre_Bosc
KAISER
Durondeau_
De_Duvergnies
Old_Home
Fon_de_Cuerne
Ananas_Curtrai
Berg_Strycker
Smith
Comte_Lambert
Mac
Manning_Miller
Clara_Frijs
Sweet_Huffcap
Doy_Montjean
Madame_Millet
Lo_Mountnorris
Conference
Saels
Ingeborg
Carrier_1
Chelten_Cross
Bristol_Cross
Concorde
Nimrod
Fritjof
Gen_Koenig
Williams
Arnold
Bartlett
Big_R_Bartlett
Doub_Williams
Instone_1
KO_R_Bartlett
M_Red_Bartlett
Moyer_R_Bart
N_R_Bartlett
WILLIAM
Parburton
Redbald
Rus_Bartlett
Sandar
Striped_Will
Williams_Comp
REALE
Harvest_Queen
Greg_Bordillon
Sirrine
Highland
Dawn
Reimer_Red
Delbais
Merton_Pride
Merton_Royal
Ovid
0.07
0.30
0.53
0.77
1.00
Coefficient

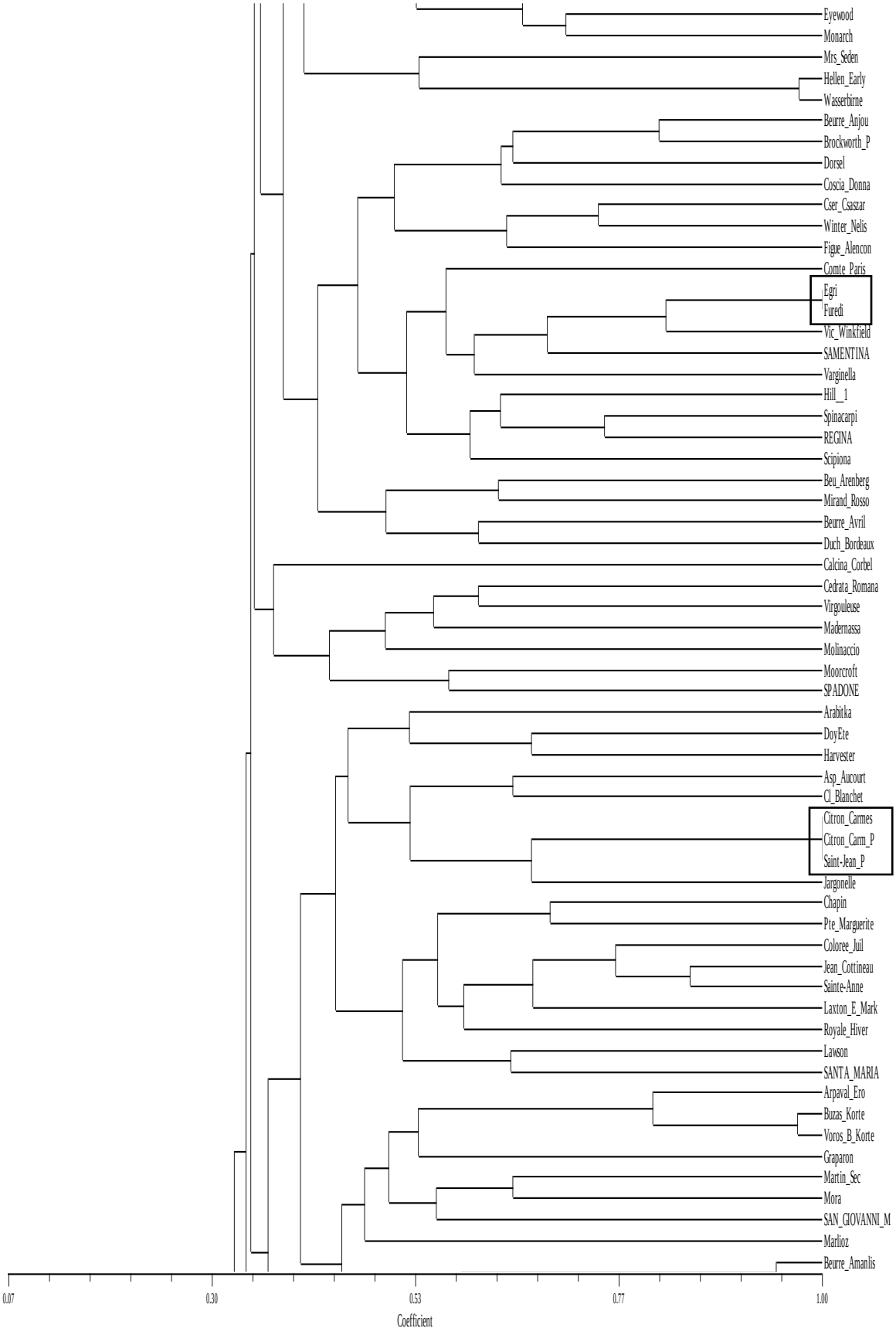

Butirra_Roma
Colette
Howlett_1
Madame_Treyve
Magyar_Kobak
Pulteney
Phelps
Sierra
Laxton_Superb
Mercer
Le_Brun
Satisfaction
Ant_Bouvant
Williams_Frest
Williams_Hiver
Richard_Peters
Thornley
Barnham_1
Willie_Peddie
Gen_Leclerc
Ewart
Mellina
Michael_Nelis
Beu_Assomption
Tr_Vienne
Dr_Stark
B_Mortillet
Dr_J_Guyot
Marg_Marillat
Harrow_Delight
William_Pr_Mor
Cascade
Goriziana_Rosa
Lee
Florana
Moonglow
Clapp_Favour
Large_Clapps
Starkrimson
Star
Spartlet
ARBANISCO
Herzogin_Elsa
Prec_Trevoux
Southworth
Vermont_Beauty
Alex_Delfosse
Beu_Francois
Pres_Barabe
Berg_Heimbourg
Fon_Automne
Seig_Esperen
Fon_du_Comice
Belle_Soignies
Onward
Santa_Claus
Miskolci
Theo_van_Mons
Alph_Huttin
Beur_Balt_Pere
0.07
0.30
0.53
0.77
1.00
Coefficient

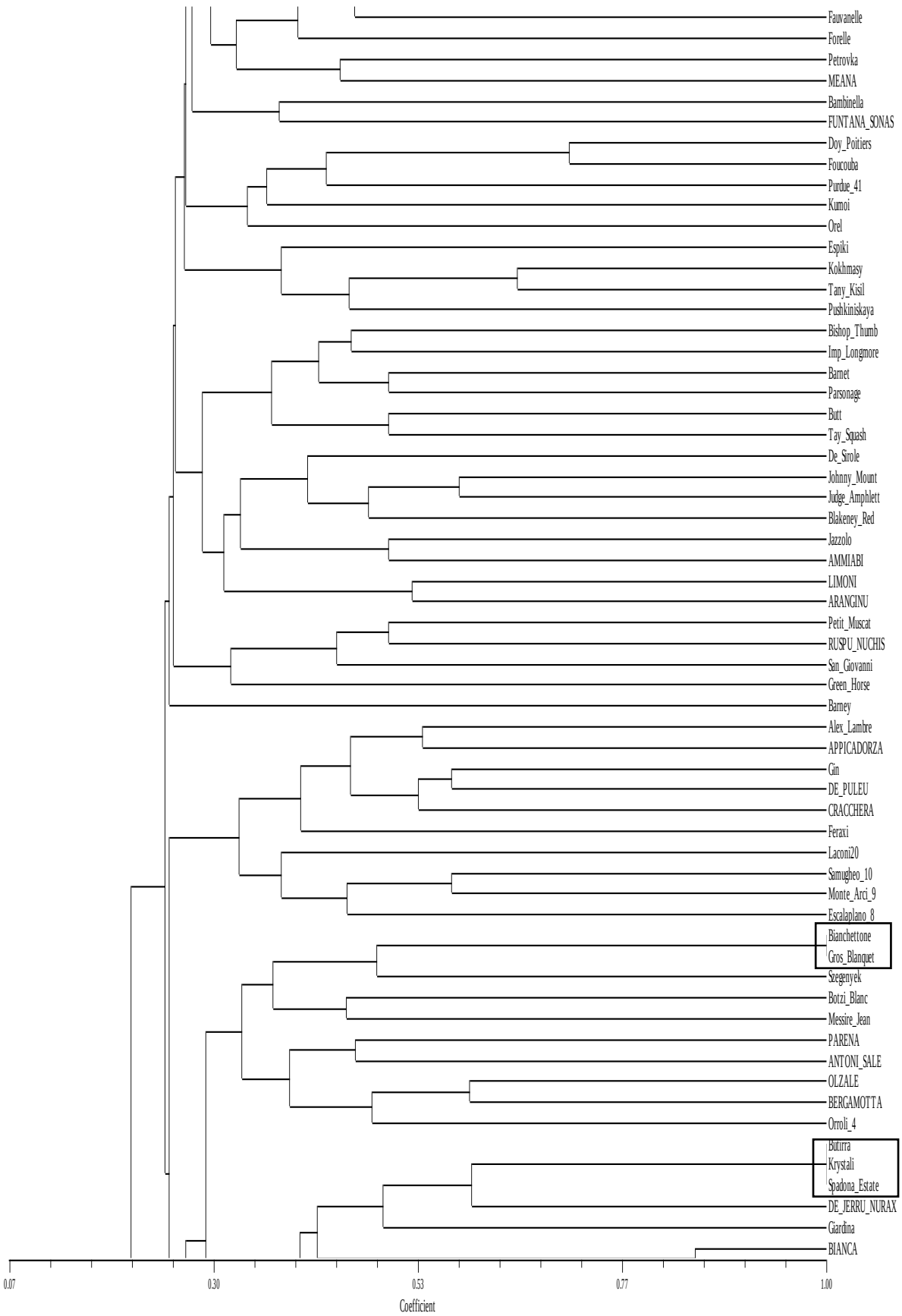

Beurre_Luizet
Enfant_Nantais
Gorham
Grand_Champion
El_Dorado
Laxton_Forem
Pres_Osmonvil
Sou_Congres
Avocat_Allard
Baronne_Mello
Sum_B_Arenberg
Beurre_Naghin
Blickling
Jean_de_Witte
Passe_Colmar
Zep_Gregoire
Dr_Lentier
Beurre_Dumont
Bon_S_Denis
Glou_Morceau
Ph_Couvreur
Madame_Favre
Beurre_Jonghe
Const_Lesueur
Ferd_Gaillard
Kieffer_1
Mad_Bonneford_
Le_Lectier
M_Louise_Uccle
Autumn_Nelis
B_Chret_Walra
Leonie_Bouvier
Napoleon
Jules_Airol
Laxton_Record
Beurre_Six
Nouv_Fulvie
Cons_a_Cour
Marechal_Cour
Mere_Perrier
Marquise
Tr_Jodoigne
Coreless
Pero_Nobile
Lorinc_Kovacs
Winter_Orange
Coscia_Prec
L_Bonne_Sinn
Alex_a_Bivort
And_Desportes
Ph_Chauveau
B_Papa_Fosse
Tetten_Dick
Beu_Capiaumont
Parrot_Pear
Princess
Magnate
Gr_Calebasse
Hermansverk_1
Hermansverk_2
0.07
0.30
0.53
0.77
1.00
Coefficient

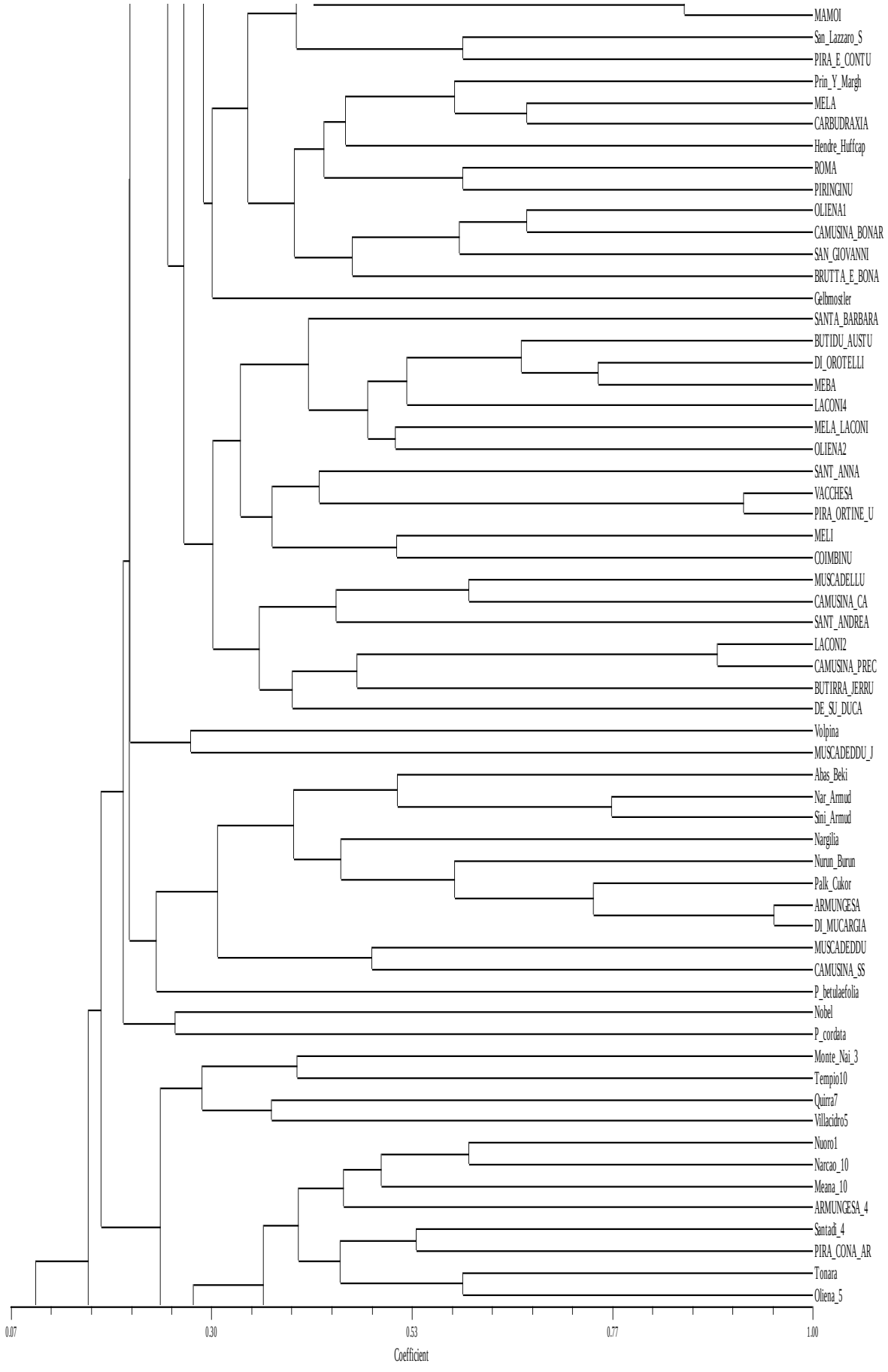

Berg_Philippot
N_Kalman_Korte
Gans_Bergamot
Beurre_Hardy
R_Beur_Hardy
El_Morettini
Sheldon
Fin_de_Siecle
Beurre_Brown
B_Gris_Hiver_N
L_Bonne_Print
Solaner
Maggie
Roem_v_Wijngar
Ayrshire_Lass
Green_Beurre
Belle_Julie
Thompson
Beurre_Rance
U_S_R_Peart
Csatar
Soldat_Labour
Tyson
CORZONE_AINU
BRAGAMOTTA
SA_MADALLENA
Black_Worcest
Lemon
Sir_H_Veitch
Round_Red_Pear
Beur_Clairgeau
P_Corneille
Fon_Ingendaele
Dir_Hardy
Fertility
Imp_Fertility
Pritchard_1
B_Precoce_M
Butirra_R_Mor
Roosevelt
Coscia
Santa_Maria
Coscia_Tadiva
Leopardo
B_Chret_Fr_Ba
Suc_Montlucon
Epine_du_Mas
Dov_Alencon
Liegel_Butter
Virgoloso
Farmingdale
Saint_Luke
Bell_d_Hiver
Fon_Bail_Mait
Mercedes
Olivier_Serres
Craig_Favour
Dutch_Holland
Cinq_Grappes
Girogile
0.07
0.30
0.53
0.77
1.00
Coefficient


Nouv_Poiteau
Berg_Sannier
Emile_Heyst
Doy_Hiver
Sos
Levard
Remy_Chatenay
Pres_Mas
Beu_Dubuisson
B_H_Courcelle
Crichton-Maitl
Swan_Egg
Huyshe_Victor
L_Swan_Egg
Leboux_Grignon
Suffolk_Thorn
Col_Marchand
Jos_Malines
Oktyabrzskaya
Berg_Fond_Ete
ENOSA
E_DONNA
CAULI
LACONI_1
Carrick
Mericourt
Surprise
Giant_Seckel
Kieffer_4
Soeur_Gregoire
Peter
Crawford
Hessle
Beur_Giffard
SAN_DOMENICO
Dana_Hovey
Ritson
Saint-Remy
Reliance
Seckel
Worden_Seckel
Certeau_Autom
Double_Guerre
Dub_Kreeft
Frangipane
Laird_Lang
Beurre_Dilly
Zoe
Chaumontel
Comte_de_Lamy
Stanford
Adm_Gervais
Berg_Esperen
Nec_P_Meuris
Madame_Ballett
B_Chretien_H
Crassane
Crassane_P
Hacon_Incomp
Broompark
0.07
0.30
0.53
0.77
1.00
Coefficient


Eyewood
Monarch
Mrs_Seden
Hellen_Early
Wasserbirne
Beurre_Anjou
Brockworth_P
Dorsel
Coscia_Donna
Cser_Csaszar
Winter_Nelis
Figue_Alencon
Comte_Paris
Egri
Furedi
Vic_Winkfield
SAMENTINA
Varginella
Hill__1
Spinacarpi
REGINA
Scipiona
Beu_Arenberg
Mirand_Rosso
Beurre_Avril
Duch_Bordeaux
Calcina_Corbel
Cedrata_Romana
Virgouleuse
Madernassa
Molinaccio
Moorcroft
SPADONE
Arabitka
DoyEte
Harvester
Asp_Aucourt
CL_Blanchet
Citron_Carmes
Citron_Carm_P
Saint-Jean_P
Jargonelle
Chapin
Pte_Marguerite
Coloree_Jul
Jean_Cottineau
Sainte-Anne
Laxton_E_Mark
Royale_Hiver
Lawson
SANTA_MARIA
Arpaval_Ero
Buzas_Korte
Voros_B_Korte
Graparon
Martin_Sec
Mora
SAN_GIOVANNI_M
Marlioz
Beurre_Amanlis
0.07
0.30
0.53
0.77
1.00
Coefficient


Eyewood
Monarch
Mrs_Seden
Hellen_Early
Wasserbirne
Beurre_Anjou
Brockworth_P
Dorsel
Coscia_Donna
Cser_Csaszar
Winter_Nelis
Figue_Alencon
Comte_Paris
Egri
Furedi
Vic_Winkfield
SAMENTINA
Varginella
Hill__1
Spinacarpi
REGINA
Scipiona
Beu_Arenberg
Mirand_Rosso
Beurre_Avril
Duch_Bordeaux
Calcina_Corbel
Cedrata_Romana
Virgouleuse
Madernassa
Molinaccio
Moorcroft
SPADONE
Arabitka
DoyEte
Harvester
Asp_Aucourt
Cl_Blanchet
Citron_Carmes
Citron_Carm_P
Saint-Jean_P
Jargonelle
Chapin
Pte_Marguerite
Coloree_Juil
Jean_Cottineau
Sainte-Anne
Laxton_E_Mark
Royale_Hiver
Lawson
SANTA_MARIA
Arpaval_Ero
Buzas_Korte
Voros_B_Korte
Graparon
Martin_Sec
Mora
SAN_GIOVANNI_M
Marlioz
Beurre_Amanlis
0.07
0.30
0.53
0.77
1.00
Coefficient


Fauvanelle
Forelle
Petrovka
MEANA
Bambinella
FUNTANA_SONAS
Doy_Poitiers
Foucouba
Purdue_41
Kunoi
Orel
Espiki
Kokhmasy
Tany_Kisil
Pushkiniskaya
Bishop_Thumb
Imp_Longmore
Barnet
Parsonage
Butt
Tay_Squash
De_Sirole
Johnny_Mount
Judge_Amphlett
Blakeney_Red
Jazzolo
AMMIABI
LIMONI
ARANGINU
Petit_Muscat
RUSPU_NUCHIS
San_Giovanni
Green_Horse
Barney
Alex_Lambre
APPICADORZA
Gin
DE_PULEU
CRACCHERA
Feraxi
Laconi20
Samugheo_10
Monte_Arci_9
Escalaplano_8
Bianchettone
Gros_Blanquet
Szegenyek
Botzi_Blanc
Messire_Jean
PARENA
ANTONI_SALE
OLZALE
BERGAMOTTA
Orroli_4
Butirra
Krystali
Spadona_Estate
DE_JERRU_NURAX
Giardina
BIANCA
0.07
0.30
0.53
0.77
1.00
Coefficient


MAMOI
San_Lazzaro_S
PIRA_E_CONTU
Prin_Y_Margh
MELA
CARBUDRAXIA
Hendre_Huffcap
ROMA
PIRINGINU
OLIENA1
CAMUSINA_BONAR
SAN_GIOVANNI
BRUTTA_E_BONA
Gelbmostler
SANTA_BARBARA
BUTIDU_AUSTU
DI_OROTELLI
MEBA
LACONI4
MELA_LACONI
OLIENA2
SANT_ANNA
VACCHESA
PIRA_ORTINE_U
MELI
COIMBINU
MUSCADELLU
CAMUSINA_CA
SANT_ANDREA
LACONI2
CAMUSINA_PREC
BUTIRRA_JERRU
DE_SU_DUCA
Volpina
MUSCADEDDU_J
Abas_Beki
Nar_Armud
Sni_Armud
Nargilia
Nurun_Burun
Palk_Cukor
ARMUNGESA
DI_MUCARGIA
MUSCADEDDU
CAMUSINA_SS
P_betulaefolia
Nobel
P_cordata
Monte_Nai_3
Tempio10
Quirra7
Villacidro5
Nuoro1
Narcao_10
Meana_10
ARMUNGESA_4
Santadi_4
PIRA_CONA_AR
Tonara
Oliena_5
0.07
0.30
0.53
0.77
1.00
Coefficient

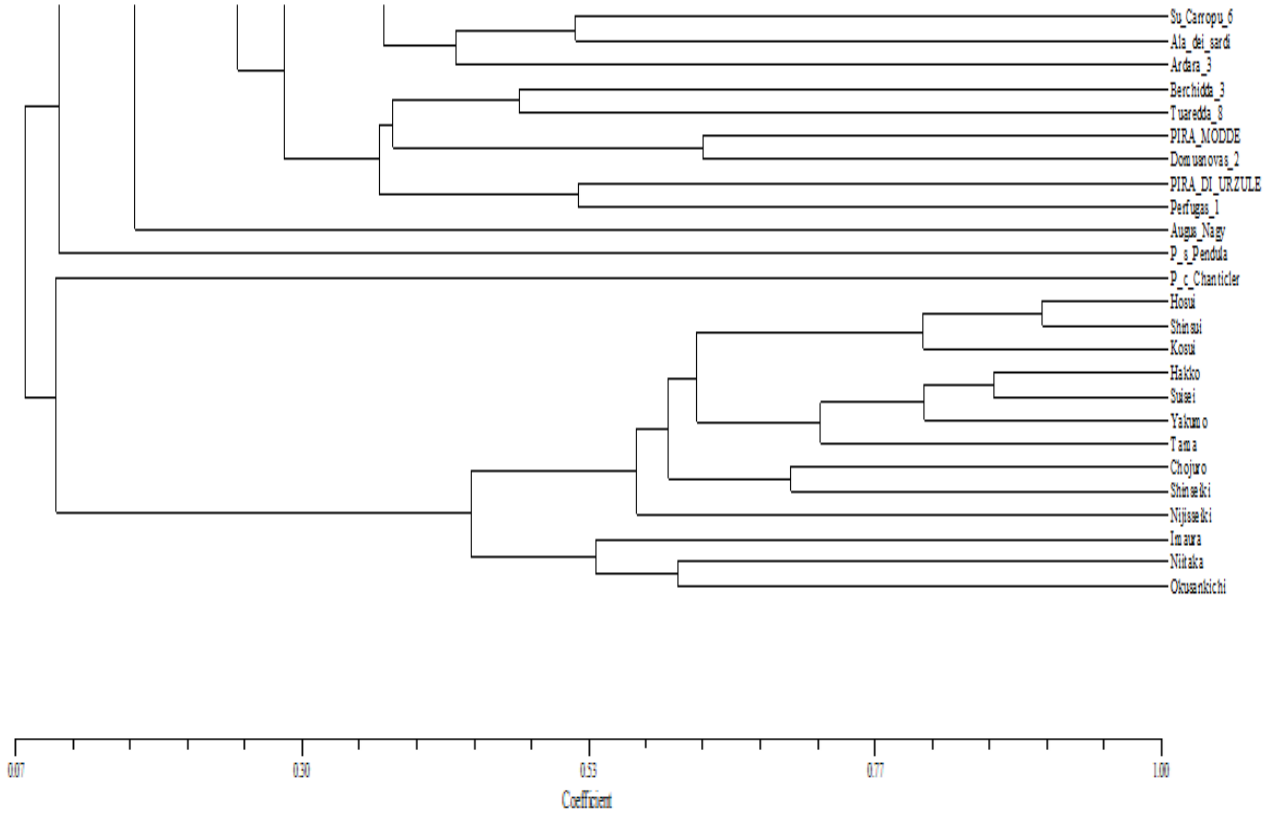


Fig. 4 Dendrogram of 625 pear varieties from international germplasm based on Cophenetic Correlation Coefficient calculated from the allele frequencies of 194 alleles found in 9 SSR loci. In the rectangles, the homonymies found in the international dataset of *Pyrus* accessions.

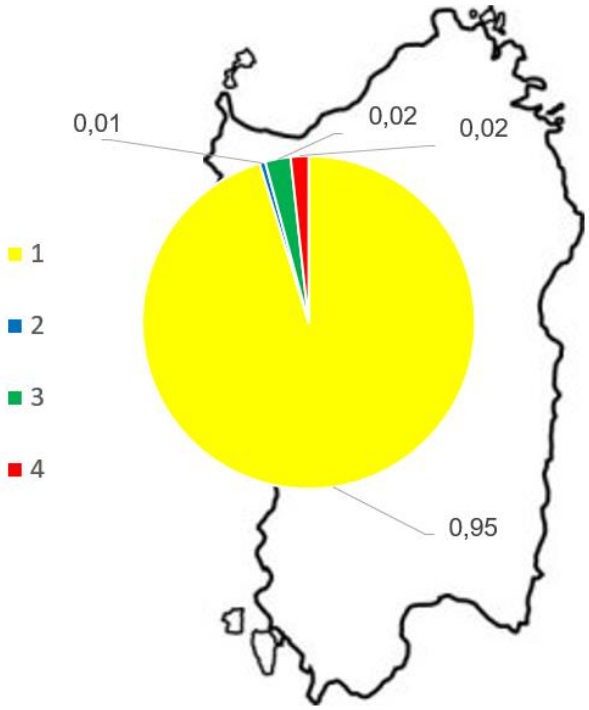


Fig. 5. Distribution of Q values related to the Cluster of the varieties of Sardinia. (1): Sardinian varieties; (2): Japanese varieties; (3): late varieties; (4): the most famous varieties used in breeding**.**

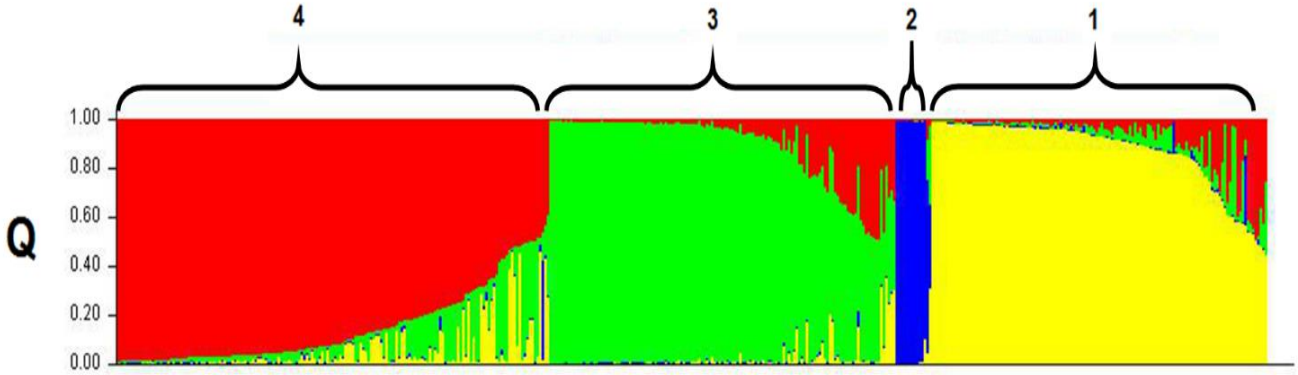


Fig. 6. Groups identified in the first round structure analysis. 1,2,3,4 corresponding a number of cluster. In yellow Sardinian varieties; in blue Japanese varieties; in green late varieties; in red the most famous varieties used in breeding**.**

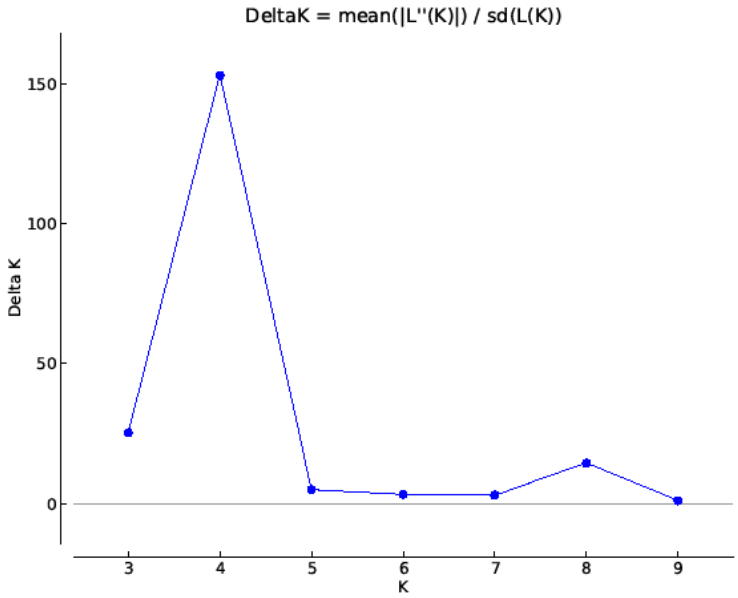


Fig. 7 Slope rate change estimates of the log probability curve (Δk) calculated according to Evanno et al. (2005) plotted against number of cluster (k).

4. Discussion.

The nine SSR markers used for our the analysis were selected on the basis of the work carried out by Seih et al. (2012), in which eight markers for apple and two for pear were used. The transferability of SSR information from a species to a related second species has been demonstrated abundantly in the work of Pierantoni et al., (2004). Reproducibility and coherence in amplified peak readings in the references "William", "Keizer" and "Abate" have allowed ensuring that SSRs used in this work are reliable markers of the diversity of pear germplasm.

The 15 alleles recognized as unique were present in a single genotype, indicating a substantial level of allelic diversity ever used in breeding programs. The presence of a few rare alleles is reflected in the structure, in fact there is a single large group that encloses all diploids.

The use of reliable SSRs is essential for effective differentiation of cultivars and for establishing genetic relationships such as in the case of identification of homonymies and synonymies. In fact, the case of synonymies between Sardinian varieties called "Reale" and the "William" reference, is definitely not due to an error in reading the peaks but may be explained by the fact the variety called in Sardinia "Reale" is surely the same variety called "William" on international data set.

The overall allelic diversity shown by the set of nine SSRs used revealed genetic variability in pear germplasm, especially as regards the varieties of *P. communis* in comparison to *P. spinosa*. Moreover, the phylogenetic proximity between the varieties of *P.communis* "Pira di Urzulè" and "Pira cona arrubia" with the population of *P.spinosa* can be explained by the fact that these varieties may originate from hybridization with the populations of *P. spinosa* present in the territory.

In fact, the hybridizations between these two species are known (Vincent et al., 2013).

The high number of triploids in Sardinian varieties can be explained by the fact that *Pyrus* is an allotetraploid species, so the marker might have amplified another locus. Additionally, the identification of four homonymies in the Sardinian varieties emphasized the importance of testing germplasm referee's collections with powerful tools such as molecular markers to avoid real repetitions and eliminate any duplications that may be present.

Thanks to the existence of a large dataset created by Fernàdez-Fernàndez et al. (2006), that comprised *Pyrus*'s international varieties, it has been possible to compare these varieties with our dataset, thus increasing the availability of usable data for future analysis.

The alignment of the two datasets was possible because the same markers were used. In addition, the use of the same Dice index has allowed us to certify the synonymies present in the cluster analysis. In fact, this index allows establishing the synonymies based on the coexistence of the bands and not for the co-absence (Dice, 1945).

Fernàndes-Fernàndez dataset's reliability, on which we have made no modification, is confirmed by the presence of the same synonymies found in Sardinia's varieties, once the two datasets have been merged.

The only homonymy found among the varieties of Sardinia and the international varieties is represented by the variety “San Domenico” and “Beur-Giffard” for international varieties. In future a pomological comparison would be an excellent methodology for the confirmation of molecular data.

The confusion generated between few Japanese and Sardinian varieties could be related to a play of weights, elongation, and shortening of the alleles, which could have the same molecular weight.

Moreover, the fact that many of the alleles present in Sardinia's varieties are also present in European varieties confirms the hypothesis of germplasm exchange between Italy and France that probably occurred during the Sardinian-Piedmont reign (1720-1861).

Regarding Structural Analysis, the value of Q found in the group of Sardinian varieties indicates that there is virtually no genetic flow from the other groups.

The very low F_{st} observed in all four groups, indicates that genetic variability is not massively linked to the structure of the population analysed. In fact, although there is a difference between the statistically valid, groups we are in a fairly homogeneous pool gene.

The high *He*, found in all four groups was predictable, considering the two species compatibility (pear and apple).

Moreover, the ΔK chart clearly shows the validity of the hypothesis that there are four distinct groups within the analyzed population, a hypothesis that is clearly confirmed by the Baterplot graphic.

5. Conclusion

The main consideration in building a data set from a very large germplasm collection is to develop reliable classification criteria. The use of the same SSR markers and the same DICE index. In addition to the confirmation of the same synonymies found in the two datasets that were compared, allowed to obtain reliable data that could be used for future work on characterization of germplasm *Pyrus.* In addition, the study of Sardinian varieties has made it possible the know a genetic diversity that is not yet investigated as an important resource that could provide the answer to the main goals of genetic improvement.

References

Agabbio, M. C. S., Suelzu, R., Mulas, M., and Mannoni, G., 1986. Patrimonio genetico delle pomacee in Sardegna: il pero: cenni storici e cultivar individuate. Studi sassaresi: organo ufficiale della Società sassarese di Scienze mediche e naturali. Sez. 3: Annali della Facoltà di Agraria dell'Università di Sassari, 32, 120-139.

Agabbio, M., 1994. Le vecchie varieta'della Sardegna: patrimonio genetico di specie arboree da frutto. C. Delfino.

Agabbio, M., Ansaldi, N., Beccaro, G.L., Berra, L., Botta, R., Brunu, A., et al., 2015. Frutti dimenticati e biodiversita' recuperata. Il germoplasma frutticolo e viticolo delle agricolture tradizionali italiane. Casi di studio: Piemonte e Sardegna. Quaderni NATURA E BIODIVERSITA'. Vol.7.

Cohen, J., Williams, J., Plucknett, D., and Shands, H., 1991. Ex Situ Conservation of Plant Genetic Resources: Global Development and Environmental Concerns. Science. 253, 866-872.

Dice, L. R., 1945. Measures of the amount of ecologic association between species. Ecology. 26, 297-302.

Earl, D. A., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources. 4, 359-361.

Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troggio, M., and Grando, M. S., 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC plant biology. 13, pp. 39.

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

Fernández-Fernández, F., Harvey, N. G., and James, C. M., 2006. Isolation and characterization of polymorphic microsatellite markers from European pear (*Pyrus communis* L.). Molecular Ecology Resources. 6, 1039-1041.

Gianfranceschi, L., Seglias, N., Tarchini, R., Komjanc, M., and Gessler, C., 1998. Simple sequence repeats for the genetic analysis of apple. Theoretical and Applied Genetics. 96, 1069-1076.

Hammer, K., Arrowsmith, N., Gladis, T., 2003. Agrobiodiversity with emphasis on plant genetic resources. Naturwissenschaften. 90, 241-250.

Kalinowski, S. T, Taper, M. L, Marshall, T. C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. 16,1099-1106.

Hokanson, S. C., Szewc-McFadden, A. K., Lamboy, W. F., and McFerson, J. R., 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a Malus× domestica Borkh. core subset collection. TAG Theoretical and Applied Genetics. 97, 671-683.

Lia, V. V, Poggio, L., Confalonieri, V. A., 2009. Microsatellite variation inmaize landraces from Northwestern Argentina: genetic diversity, population structure and racial affiliations. Theor. Appl. Genet. 119, 1053-1067.

Liang, W., Dondini, L., De Franceschi, P., Paris, R., Sansavini, S., and Tartarini, S., 2015. Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. Plant Molecular Biology Reporter. 33, 458-473.

Liebhard, R., Gianfranceschi, L., Koller, B., Ryder, C. D., Tarchini, R., Van de Weg, E., and Gessler, C., 2002. Development and characterisation of 140 new microsatellites in apple (Malus x domestica Borkh.). Molecular breeding. 10, 217-241.

Maguire, T. L., Collins, G. G, Sedgley, M., 1994 A modified CTAB DNA extraction procedure for plants belonging to the family Proteaceae. Plant Mol. Biol. Rep. 12, 106-109.

Martinelli, F., M., Busconi, M., Camangi, F., Fogher, C., Stefani, A., and Sebastiani, L., 2008. Ancient Pomoideae (*Malus domestica* Borkh. and *Pyrus communis* L.) cultivars in "Appenino Toscano" (Tuscany, Italy): molecular (SSR) and morphological characterization. Caryologia, 61, 320-331.

Muresu, E., Castiglia, P., Piana, A., Masia, M., and Ciappeddu, P. L., 1997. Caratterizzazione di prodotti alimentari dell'area sardo-corsa: analisi chimico-fisica e batteriologica. pp. 5-18.

Pierantoni, L., Cho, K. H., Shin, I. S., Chiodini, R., Tartarini, S., Dondini, L., and Sansavini, S., 2004. Characterisation and transferability of apple SSRs to two European pear F1 populations. Theoretical and Applied Genetics. 109, 1519-1524.

Powell, W., Machray, G. C., & Provan, J., 1996. Polymorphism revealed by simple sequence repeats. Trends in plant science. 1, 215-222.

Pritchard, J. K., Stephens, M., and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics. 155, 945-959.

Rohlf, F.J., 1994. NTSYS-pc Numerical taxonomy and multivariate analysis system, version 1.80. State University of New York.

Sehic, J., Garkava-Gustavsson, L., Fernández-Fernández, F., and Nybom, H., 2012. Genetic diversity in a collection of European pear (*Pyrus communis*) cultivars determined with SSR markers chosen by ECPGR. Scientia horticulturae. 145, 39-45.

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., and Maxted, N., 2013. A prioritized crop wild relative inventory to help underpin global food security. Biological conservation. 167, 265-275.

Weber, J. L., and May, P. E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American journal of human genetics. pp. 44, 388.

General Conclusion

Germplasm Banks' collections of local fruit varieties are still incomplete, and the gaps in these gatherings may create limitations on the options available for researchers to study and introduce new taxa into protection programs.

At the same time, many local varieties are threatened in nature by changes to their habitats, by the modernization of agricultural areas, invasive species, and climate change that may aggravate their vulnerability. The loss of old varieties typical of a territory represents for local communities not only a serious defeat from the biological, ecological and cultural point of view but also the loss of great opportunities for development.

The main goals achieved through this doctoral research were different and can be summarized in the following four points:

1. Through the morphometric analyses performed on the ancient Sardinian apples varieties it was possible to investigate the bonds between the morphology of the seeds and the color of the skin, showing the existence of a relationship that had never been described before and which represent an important starting point for future phenotypic studies on new varieties.
2. Comparison of existing genetic analysis and morphometric analysis of national and international apple varieties has made it possible to evaluate phenotypic and genotypic diversity and compare the two types of analysis, demonstrating how morphometric analysis was an important method to support molecular analyses.
3. Through the morphometric analysis carried out on ancient Sardinian pome varieties and on national and international pear varieties, it has been possible to characterize the different varieties and to demonstrate how Sardinian *P.spinosa* varieties are more closely related to local varieties of *P.communis* than to the ones provided by the University of Bologna, which clustered in a distinct group. Moreover, through the analyses carried out on pear fruit core browning, it was possible to establish that early and late harvested varieties better describe the correlation between fruit maturation and seed morphology.
4. Finally, through the genetic analysis of Sardinian local pear varieties and national and international ones, it was possible to investigate a genes pool that had never been subjected to analysis. Moreover, it was possible to shed light on cases of synonymy and homonymy discovering also new alleles useful for future breeding programs.

In conclusion, the research carried out during the PhD program allowed to build-up a morphometric and genetic data set of *Malus* and *Pyrus* varieties, increasing the knowledge on these varieties and laying the foundations for future studies.

Acknowledgements

In priimis un sentito ringraziamento al Tutor di questa tesi Prof Gianluigi Bacchetta per la capacità che ha dimostrato di stimolare il mio interesse per l'argomento qui discusso, per avermi mostrato aspetti relativi allo stesso che senza di lui probabilmente avrei ignorato, e infine, per la sua pazienza, e per la fiducia in me e l'entusiasmo che mi ha trasmesso per questo lavoro.

Un sentito ringraziamento ai Cotutor Oscar e Guy ed in particolare a Mariano per il lavoro svolto in questi anni, per i confronti costruttivi, gli approfondimenti, gli stimoli, gli spunti forniti alla tesi, oltre che per la sua disponibilità.

Ringrazio inoltre tutti i colleghi del CCB. Un singero Grazie a Giuseppe, per l'esempio che mi ha dato e che considero molto importante per la mia formazione, di come sia possibile amare e far amare lo studio e la conoscenza di una materia non sempre facile, come la Botanica, attraverso dialoghi e confronti mai banali ma anzi sempre costruttivi.

È un bagaglio importante, sia culturale che di impostazione mentale, che porterò sempre con me.

Un grazie di cuore alla collega Silvia, indiscutibile ascoltatrice e compagna di merende, che mi ha sempre saputa rasserenare e sostenere nei momenti più faticosi e nei quali non nego di aver pensato di abbandonare questo percorso.

Un profondo ringraziamento al Prof Luca Dondini e a tutto il suo team, siete stati un pilastro importante della mia formazione in questi anni di Dottorato, Grazie per non avermi fatto sentire la nostalgia di casa e per avermi insegnato come si lavora seriamente per ottenere buoni risultati.

Sentiti ringraziamenti indubiamente vanno ai miei genitori, che hanno permesso di raggiungere questo traguardo dandomi sempre opportuni consigli e non facendomi mai mancare il loro supporto, unitamente a tutti i miei fratelli, sorelle e nipoti.

Un sentito Grazie ai miei suoceri e a mio cognato Paolo voi sapete il perchè, Grazie!

Un Grazie infinito al mio compagno di vita Daniele, finalmente il traguardo è stato raggiunto e tu che mi sei sempre stato vicino, sei certamente la persona che meglio può capire come mi senta in questo momento, voglio dirti che il merito è anche tuo e voglio ringraziarti, per la pazienza che hai avuto nel sostenermi, rasserenarmi e per tutte le volte che con il tuo sorriso hai saputo far sorridere anche me.

Infine a conclusione di questo percorso di Dottorato sento il dovere e il piacere di ringraziare tutti coloro che mi hanno sostenuto, supportata e sopportata in questi tre anni a chi mi sostiene ancora e a chi purtroppo è venuto a mancare...

Grazie mille a tutti Voi!!!