University of Calabria

Ph.D. in Molecular Bio-pathology

(Disciplinary Field BIO18-Genetics)

Exploring new routes in genetic studies on human aging and longevity

Candidate
Alberto Montesanto

Supervisor
Prof. Giuseppe Passarino

Co-ordinator
Prof. Giovanna De Benedictis

2006

A Gigi e Giuseppe

Table of Contents

SOMMARIO

Negli ultimi decenni si è verificato un crescente interesse scientifico verso gli studi volti alla comprensione delle basi genetiche dell'invecchiamento e della longevità umana. Questo crescente interesse è giustificato dall'aumento della popolazione anziana verificatosi negli ultimi 50 anni nei Paesi sviluppati, causato da un generale miglioramento delle condizioni igienico-sanitarie. Il rapido incremento di questa fascia della popolazione rappresenta oggi un importante problema di carattere sia sanitario che sociale.

La difficoltà nello studio delle basi biologiche dell'invecchiamento e della longevità umana è principalmente dovuta (i) alla specificità dell'invecchiamento rispetto alla coorte ed alla popolazione analizzate, (ii) alla mancanza di una chiara ed oggettiva definizione del fenotipo da analizzare, (iii) alla natura poligenica della longevità umana. Sulla base di queste osservazioni, gli studi qui riportati rappresentano dei nuovi approcci per lo studio dell'invecchiamento e della longevità umana. Essi sono stati condotti su dati ottenuti dalla popolazione calabrese, caratterizzata da un'elevata omogeneità genetica e un basso tasso di immigrazione dovuti a motivi geografici, storici e sociali.

Il primo approccio che abbiamo utilizzato si basa sull'analisi spaziale degli individui longevi in Calabria. Mediante questo approccio abbiamo verificato che la distribuzione spaziale di tali individui in tale regione non è uniforme. Inoltre, mediante l'analisi dei cognomi calabresi, abbiamo verificato l'esistenza di una correlazione significativa fra inbreeding e longevità maschile in una particolare area di questa regione.

Il secondo approccio si basa sull'applicazione di una Cluster Analysis, con parametri geriatrici ben riconosciuti, per l'identificazione di "fenotipi di invecchiamento" nella popolazione calabrese. I risultati ottenuti dimostrano che i "fenotipi di invecchiamento"

identificati mediante Cluster Analysis hanno una solida base geriatrica e presentano una chiara componente genetica.

Infine, mediante un approccio multilocus, abbiamo analizzato l'influenza della variabilità di geni candidati sulla sopravvivenza in età avanzata. Sulla base di una curva di sopravvivenza sintetica ottenuta utilizzando i dati storici di mortalità della popolazione italiana e utilizzando modelli di regressione multivariati, abbiamo dimostrato che i fattori genetici influenzano la sopravvivenza in età avanzata in modo sesso ed età-specifici

I risultati qui riportati dimostrano che l'applicazione di questi nuovi approcci può risultare utile negli studi dell'invecchiamento e della longevità umana. D'altro canto, essi dimostrano come per lo studio di tali tratti complessi sia importante e necessario un approccio di tipo multidisciplinare.

SUMMARY

The past few decades has witnessed a growing scientific interest in genetic studies on human aging and longevity. This growing interest may be explained by the increasing number of elderly subjects in developed countries over the last 50 years due to the continued improvements in health care. Such a fast increase of these population segments represent a huge problem for the societies in terms of social care and welfare.

The difficulty in understanding the biological basis of human aging and longevity is mainly represented by (i) the cohort and the population-specificity of human aging and longevity, (ii) the lack of a clear and objective definition of the phenotype, (iii) the polygenic nature of the human longevity. On the basis of these observations, the studies reported here represent new approaches for the study of human aging and longevity. They were carried out on data obtained from the Calabrian population, characterized by high genetic homogeneity and a scarce level of immigration due to geographical, historical and social reasons.

The first approach we used was to analyze the spatial distribution of long-lived individuals in Calabria. Using this approach, we verified that the spatial distribution of such individuals in this region is not uniform. In addition, by using surname data, we verified a significant correlation between population inbreeding and male longevity in a particular area of this region.

The second approach is based on the application of a Cluster Analysis with well established geriatric parameters to identify aging phenotypes in the Calabrian population. The results obtained show that, the aging phenotypes recognized by Cluster Analysis are consistent from a geriatric point of view and have a clear genetic component.

Finally, using a multilocus approach, we analyzed the influence of genetic variability of candidate genes on survival at old age in good health. On the basis of a synthetic survival curve built using historic mortality data from the Italian population and using multiple regression models, we found that genetic factors influence survival at advanced ages in good health in a sex and age specific way.

The results reported here show that the application of these new approaches may be useful in human aging and longevity studies. Furthermore, they demonstrate that a multidisciplinary approach is necessary to analyze human aging and longevity.

List of abbreviations

ADL	Activity of Daily Living
APOA1	Apolipoprotein A1
APOA4	Apolipoprotein A4
APOB	Apolipoprotein B
APOE	Apolipoprotein E
BZ	Blue Zone
CA	Cluster Analysis
CR	Centenarian Rate
DZ	Dizygote
ECHA	European Challenge for Healthy Ageing
F/M	Female/Male
G1	Group 1
G2	Group 2
G3	Group 3
GD	Genetic Demographic model
GDS	Geriatric Depression Scale
GIS	Geographic Information System
HSP70-1	Heat Shock Protein 70-1
HSP90α	Heat Shock Protein 90 alpha
HSP90β	Heat Shock Protein 90 beta
IGF	Insulin-like Growth Factor
IGF-I	Insulin-like Growth Factor I
LD	Linkage Disequilibrium
MMSE	Mini Mental State Examination
mNR	Male Nonagenarian Rate
mtDNA	Mitochondrial DNA
MZ	Monozygote
NP	Non Parametric
NR	Nonagenarian Rate
OR	Odds Ratio
PR	Parametric

QI (I=1,…,4)	Municipalities belonging to the I-th quartile of the mNR distribution (Fig. 2, Chapter 2) or of theFisher's alpha distribution (Fig. 5, Chapter 2)
RR	Relative Risk
S(x)	Synthetic survival function
S''(x)	Second derivative of the synthetic survival function
S_1	Sample 1 (65-85 year old subjects)
S_2	Sample 2 (> 90 year old subjects)
SA	Spatial Analysis
SES	Socio Economic Status
SIR	Silent Information Regulator
SIRT3	Silent Information Regulator 3
SP	Semi Parametric
TH	Tyrosine Hydroxylase
VNTR	Variable Number of Tandem Repeats

1. INTRODUCTION

Aging is a complex process of progressive physical decline that characterizes every biological species and leads to a dramatic reduction of the individual survival probability and, ultimately, to death. Most of the characteristics of aging, such as lifespan, have been observed to be species specific, suggesting a genetic control on aging. On the other hand, within each species the quality and the rate of the aging process are characterized by a considerable variability, which implies noticeable differences in the individual aging phenotype and in life span. Such variability is due to the interplay of genetic, environmental and stochastic factors (Kirkwood, 2005; Herndon et al., 2002; Kirkwood and Finch, 2002). Almost every aspect of the individual phenotype undergoes modification with aging, and over the years this phenomenological complexity has led to a proliferation of ideas about specific cellular and molecular causes. Recent advances have resulted in significant progress in the theoretical underpinnings of aging research. The emerging idea is that, although aging is not a programmed process, a network of evolutionarily conserved cellular and molecular mechanisms modulates the aging process (the network theory of aging). According to this theory, aging is characterized by an age-related remodelling and adaptation of every cell and organ of the body in order to cope with the continuous attrition caused by internal and external stresses (Franceschi et al., 2000).

In the last few decades there has been a growing scientific interest regarding the basis of individual variability in aging, especially concerning humans. This growing interest may be explained by the increasing number of elderly subjects in developed countries over the last 50 years due to the continued improvements in health care (Kannisto, 1994). For instance, in 1961 the population aged 65 and older in Italy was 4.8 million (9.5% of the total population), while in 1981 this number increased up to 7.5 million (13.2% of the total population) and in 2001 it grew up to 10.6 million (18.7% of the total population). In addition, the population aged 90 and older is growing at a faster pace as it has quintuplicated in the last 10 years (Istituto Centrale di Statistica 1986; www.istat.it). Such a fast increase of these population segments represent a huge problem for societies in terms of social care and welfare. This problem is strengthened by the high prevalence among these subjects of chronic diseases, cognitive impairment and other disorders responsible for functional limitations that are often associated with aging. In order to cope with the challenges posed by the increase of the elderly population, both the European Union and

the United States of America have devoted notable resources to research on the biological basis of aging.

The studies aimed to understand the biological basis of aging in humans, as does all the research dealing with complex traits, have faced the difficulty of clearly defining the phenotypes associated with aging. In this context, particularly important is the concept of "successful aging". The concept of successful aging, put forward by Rowe and Kahn (1987), suggested that within the category of normal aging, a distinction can be made between usual and successful aging. Although today there are a number of possible definitions of successful aging, the one proposed by the same authors, and perhaps the most inclusive, incorporates several interactive components: absence of disease, maintenance of cognitive and physical functioning, engagement with life, and longevity (Rowe and Kahn, 1998). This definition takes into account the notion that there is a distinction to be made among primary (normal), secondary (impaired), and optimal (successful) aging. Consequently, most of the studies on aging have focused on the assessment of health in the elderly, while this is still poorly understood especially for very old subjects such as nonagenarians and centenarians, and on the biological basis of human longevity.

From a genetic point of view two most important points must be taken into account in studies of human aging and longevity: the polygenic nature of the trait and the effects of a population genetic structure on the same.

The polygenic nature of human longevity

In order to infer the effect of genetic factors on the aging process, it is important to keep in mind that aging is a complex trait such that no single gene or attribute can be considered as an independent predictor of it. Morever, common genetic variants with important effects on human longevity are unlikely to exist because of the lack of Mendelian patterns of inheritance for this trait. On the contrary, many interacting genes whose individual effects are rather small probably affect survival at old age (Hjelmborg et al., 2006). Thus, studies that consider marker genotypes at one locus in connection with survival may capture only a small proportion of the total combined effects of the susceptibility genes which affect the phenotype. Since the development of complex traits involves multiple genes and their interactions, multi-locus association analyses have been proposed to overcome this kind of problem (Hoh and Ott, 2003).

Population genetic structure and human longevity

Human populations are characterized by specific gene pools which are the result of their history, in terms of chance (genetic drift), migrations and adaptative selection to environment. Recently, two main lines of evidence have shown a possible role played by the population genetic structure on human longevity. The first is the significant correlation found between population genetic patterns and Female/Male (F/M) ratio among centenarians in the Italian population (Passarino et al., 2002). The second is the observation that in Sardinia centenarians are clustered in restricted areas characterized by a high level of geographic isolation and endogamy (Poulain et al., 2004). Here, the F/M ratio among centenarians varies according to these areas. On the basis of these observations, the impact of geography on human longevity may represent a direct consequence of the genetic structure of the population.

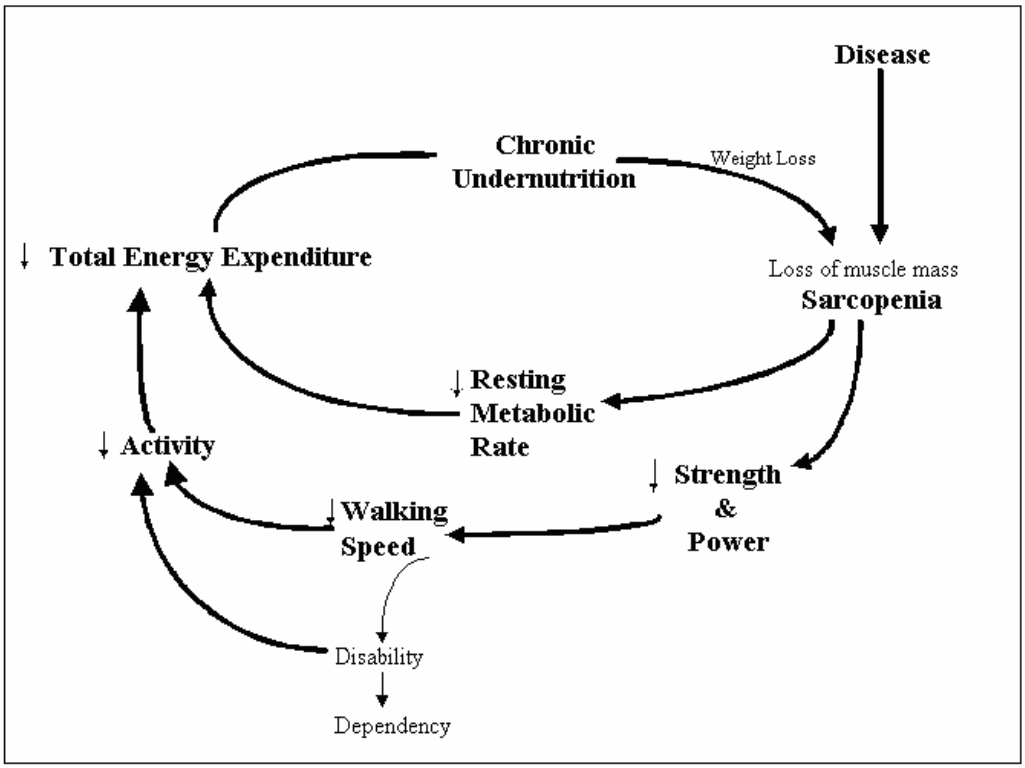
1.1 Aassessment of health status in the elderly

The choice of a phenotype is critical for the study of a complex genetic process, such as the aging process. In the last decade most of the studies aimed at explaining the biological basis of "healthy" or "successful" aging have based their phenotype definition on survival measures, primarily on age at death or longevity (Finkel et al., 1995; McGue et al., 1993). Although longevity should imply a quality of aging which is quite good, it is common experience that nonagenarians and centenarians constitute a very heterogeneous group, and that most suffer of disabilities or diseases (Jeune, 2002). In addition, the continuous improvement of medical care is leading to an increased number of old people afflicted with many chronic diseases but surviving for many years. Therefore, an ideal definition of the aging phenotype should take into account not only the age but also the health status of the subject. To date, a quantitative definition of the health status in old subjects is still lacking. This represents a serious obstacle in disentangling the complex interplay of genetic, environmental, and historical factors that modulate the aging process and the physical decline which is associated with this process.

The arbitrary definition of healthy aging has represented for many years the most serious obstacle for the definition of a criterion to assess the health status in the elderly. Nowdays there is agreement on the fact that a multidimensional assessment could measure the capacity to function well and cope with environmental challenges in domains assessing physical, mental and social well-being (Peel et al., 2004). The effort to define a phenotype for outlining the health status in the elderly took a step forward when the *frailty syndrome* was recognized. Practionners and geriatric clinicians have long recognized the existence of a subset of older adults who are frail, that is a subset of old subjects who are more vulnerable than others to a number of poor health outcomes (such as falls, hospitalization and death). The identification of these subjects may represent an important step for facing the challenge posed by the aging population. However, it has been extremely difficult to define frailty and to understand its biological basis. In fact, the diagnosis of frailty was mostly subjective, and physicians seldom thought of specific treatment. The first attempts to try to set an "objective" measure of frailty have led to consider frailty as synonymous of disability, co-morbidity or advanced old age (Hoffman et al., 1996; Wagner et al., 1996; Pope and Tarlov, 1991). More recently, clinicians and investigators have begun to recognize frailty as a distinct biological geriatric syndrome. Central to the clinical definition of frailty is the concept that no single altered system defines this state, but that multiple systems are involved (Rockwood, 2005). Patients classified as frail typically

exhibit loss of muscle strength, are physically inactive, and have a slow gait, with an increased risk (and fear) of falling. They are likely to have a poor appetite and to have undergone a recent, unintentional loss of weight. Frail individuals are more likely than the non frail to experience impaired cognition and depression. A very important step forward to further understand the biology of frailty was the study by Fried and Waltson (1998). These authors developed a hypothetical "cycle of frailty" that illustrated how disease and age-related changes may trigger frailty and how disability may evolve from this condition. Figure 1 illustrates the frailty cycle and how the loss of muscle mass, altered energy expenditure, and decline in nutritional intake may lead to this cycle of decline.

Figure 1. Cycle of frailty (modified from Fried et al., 2001).



Subsequently, Fried and colleagues (Fried et al., 2001) provided a potential standardized definition of the frailty phenotype. Based on a prior research and on clinical consensus, they proposed that frailty is fundamentally a wasting syndrome, characterized by weakness and poor nutritional status. They hypothesized that a series of characteristics compose the "phenotype of frailty" and then evaluated whether this phenotype identified a subset of individuals at high risk for adverse health outcomes clinically associated with frailty. In Fried's study, frailty was defined as a clinical syndrome of signs and symptoms in which at

least three of five possible criteria are present: (i) low grip strength, (ii) slow walking speed, (iii) low physical activity, (iv) self-reported unintentional weight loss, (v) self-reported exhaustion. Validation work has shown that patients meeting this definition of frailty were more likely to die, to be hospitalized, or to become disabled over 6 years of follow-up, independently of age and other risk factors for mortality (Fried et al., 2001).

After Fried and colleagues had outlined the basis to clinically define the frailty syndrome, the American Medical Association stated that frail subjects represent a group of patients who present the most complex and challenging problems to the physician and health care professionals (American Medical Association, white paper on elderly health). Moreover, in 2003, the National Institute on Aging recognised frailty as a priority for research and made $ 1.8 million available for studies on this topic.

Although much of the literature on frailty focuses mainly on biological and medical factors, recent work suggests that the social network of support and psychological components may influence the quality of life at old age and in turn influence the progression of frailty. Different markers of frailty have been proposed in the literature to monitor the progression of frailty and the onset of adverse outcomes and disability. A number of scores have been also developed for use in the research setting. Although the methodological approaches to identify markers of frailty differ in the specific criteria used, mounting evidence suggests that multiple interrelated physiological systems, including neurological and musculoskeletal functioning, inflammatory status, energy metabolism, influence the health and well-being of the entire organism and are modestly altered in frail older adults. A number of potential serum markers for frailty, disability, or adverse outcomes in old patients have also been identified. These include: peripheral blood markers of inflammation (IL-6, C-reactive Protein, Serum Albumin, Cholesterol) (Cohen et al., 1997; Ranieri et al., 1998; Corti et al., 1994), Coagulation Factors (Fibrinogen, Factor VIII, D-dimer) (Cohen et al., 2003; Walston et al., 2002), IGF-I levels, creatinine, haematological parameters (haemoglobin, haematocrit) (Leng et al., 2002). Moreover, in a recent work carried out by Bartali and co-workers (Bartali et al., 2006), evidence was provided that low intakes of energy and selected nutrients (protein, vitamins D, E and folate) are independently associated with frailty.

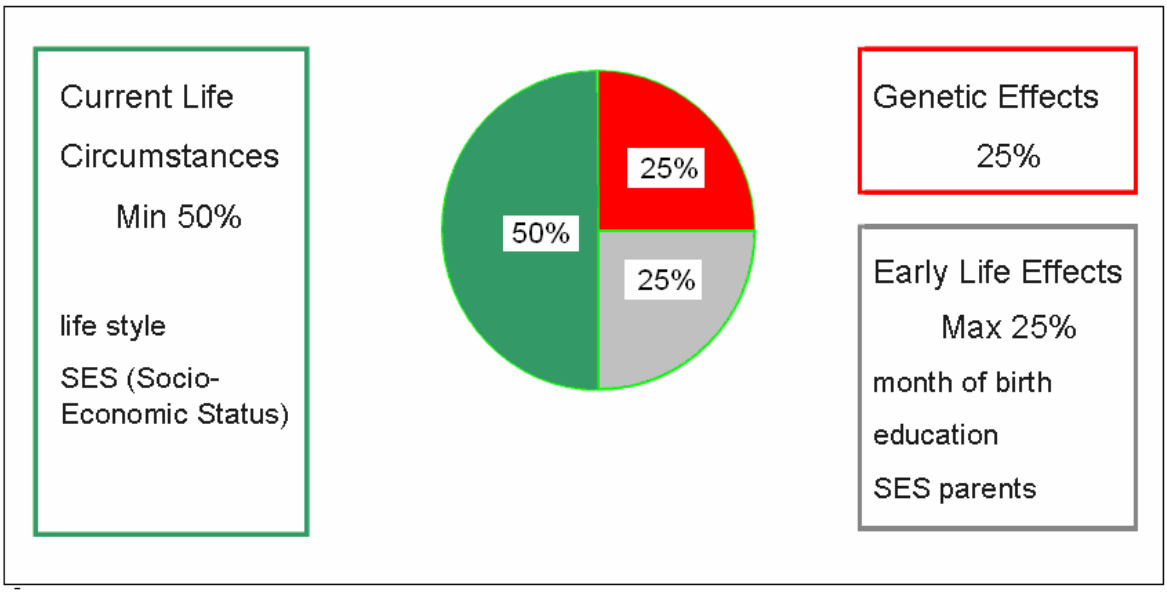
In the long run, testing these selected physiological parameters as potential correlates of frailty may provide clues for the development of more detailed etiologic studies.

1.2 The biological basis of human longevity

The factors determining human longevity have drawn the attention of researchers from a variety of disciplines including sociologists, biologists, gerontologists, psychologists and scientists. This great interest is chiefly due to the increase of the mean average life expectancy in developed countries. Reliable data from various European and North America countries show that the population of centenarians has doubled every decade since 1960, mostly as a result of the increase in survival past 80 years of age (Jeune and Vaupel, 1995). In several countries, as in Sweden, female death rates at older ages has fallen since 1950, with a large absolute reduction at advanced ages. The pattern is similar for males, although from conception to old age males suffer higher death rates than females, and progress in reducing male mortality has generally been slower than for females. Consequently, most old people in the world are women. With more and more people celebrating their 100-year birthdays, there is a great increase in studies aimed at understanding why these people survive where others failed, and what can help to explain the life span heterogeneity. From a demographical point of view, this growth of the oldest population can be attributed to two main factors: the increase in the probability of surviving to advanced age and the reduction of mortality at older ages, occurred since 1950 (Kannisto, 1996; Vaupel, 1997).

As a complex trait, life span is controlled by the same three components which modulate multifactorial phenotypes: genes, environment and chance. Vaupel et al. (1998) summarized the impact of different factors on human longevity and proposed that conditioning factors that arise in the first part of life (socio-economic state of parents, education and month of birth, which has been found to reflect the environmental conditions during the prenatal and early postnatal period) account for 25% of the interindividual variability in lifespan; life circumstances at adult and old age (including socio-economic status (SES) and medical assistance) may account for about 50%. As for the genetic influence on longevity, based on of a previous study on twins (Herskind et al, 1996), it should account for the remaining 25% (Fig. 2).

Figure 2. Components of the variation in lifespan in contemporary populations (modified from Vaupel et al., 1998).



Interestingly, much evidence suggests that genetic factors have a different impact on survival at different ages. In 1825 Benjamin Gompertz proposed that the mortality rate increased exponentially with age (Gompertz B, 1825). However, at very old age, mortality begins to decelerate in organisms such as medflies (Carey et al., 1992), nematode worms (Brooks et al., 1994; Johnson, 1990) and humans (Thatcher et al., 1990). Different hypotheses have been formulated to explain the mortality deceleration. Most likely this occurs because frailer individuals drop out of the population, leaving behind a more robust cohort who continues to survive. As a consequence, the intra-cohort distribution of specific genotypes and other survival-related attributes changes with older and older ages (Vaupel et al., 1979; Vaupel and Yashin, 1985; Vaupel, 1997). This selecting-out process is termed demographic selection (Vaupel et al., 1998). In addition, Hjelmborg et al. (2006) by studying a large population of Scandinavian monozygotic (MZ) and dizygotic (DZ) twins found that the genetic influence on lifespan is minimal prior to age 60, but increases thereafter.

The last decade has seen a surge of activity aimed at identifying genes controlling aging and longevity in model organisms such as the nematode worm and the fruitfly. Mutations in several genes have been found to increase lifespan by slowing the aging process (Lin K et al., 1997; Lin YJ et al., 1998; Guarente et al., 2000; Arantes-Oliveira et al., 2003; Bartke et al., 2003; Hsu et al., 2003). One class of genes identified in *C. elegans*, *Drosophila* and mice define a pathway of insulin/IGF signaling (Brown-Borg et al., 1996; Kimura et al.,

1997; Clancy et al., 2001). Mutations that downregulate this pathway, such as the *dwarf* mutations in mice, extend life span. The *dwarf* mutations reduce signalling by growth hormone, which, in turn, lowers the level of IGF-1. Moreover, experiments in model organisms have shown a correlation between this pathway and the gene silencing system controlled by SIR (Silent Information Regulator) proteins (Murphy at al., 2003). *Sir2* genes regulate the rate of aging in many species, modulating gene silencing and coordinating the pace of aging with the metabolic rate (Armstrong et al., 2002). Understanding the biological function of the product of such "longevity genes" turned out to be useful in studying human aging and longevity.

It is not obvious why "longevity genes" should persist in humans or any species, and to explain the persistence in different organisms of such "longevity genes" the antagonistic pleiotropic theory has been proposed. This theory was originally formulated by Williams (1957). It held that aging was due to the decline of the force of natural selection late in life, and that alleles with positive effects upon fitness early in life also have deleterious effects late in life. In this view, these later deleterious effects are the direct cause of aging (Leroi et al., 2005). Such theory remained largely untested until 1990 when different research groups observed a negative correlation between reproductive success and longevity in lines of fruit-flies (Rose, 1984; Sgrò and Partridge, 1999). Then, the antagonistic pleiotropic theory was reproposed and today it is quite popular in diverse areas of gerontology.

Taking advantage of the rapid development in molecular genetics, the past few decades have seen a significant increase in genetic studies on human longevity (De Benedictis et al., 2001). In these studies the model of centenarians emerged as crucial, as centenarians give information that no other experimental model can provide. In fact, unlike model organisms, centenarians personify the longevity phenotype naturally occurring in an outbreed species. In addition, their entire life was led in an environment that continuously pushed the organism to cope with intrinsic and extrinsic antigenic loads. Finally, no other biological category has experimented so many rapid changes as those which have occurred in the last century in all countries across the world, and chiefly in developed countries. Therefore, the model of centenarians is not simply an additional model with respect to common model organisms such as yeast, worm, *Drosophila* and mouse, but it provides unique insights on the complex network of biological and non biological factors which guide individual survival at old age (De Benedictis and Franceschi, 2006). The studies on centenarians have pointed out that human "longevity genes" could function in several important ways. They may slow the rate of age-related changes in cells and tissues,

improve the effectiveness of repair mechanisms, and increase resistance to environmental stresses like infection and injury. “Longevity genes” could also affect a wide spectrum of debilitating age-related conditions. These requirements are consistent with the observation that elderly children of centenarians have fewer incidence of diabetes and ischemic heart disease, and better self-rated health, than age-matched controls (Terry et al., 2003; Frederiksen et al., 2002). This suggests that the offspring of centenarians inherited a set of genes from their long-lived parent that protects them against these disorders.

In order to make appropriate inferences regarding the effect of observed genetic covariates on life span, one must bear in mind that life span is a complex trait such that no single gene or attribute can be considered to be an independent predictor of it. This is different from the situation where a single locus is responsible for a distinct dichotomous phenotype regardless of environment or genotypes at other loci (McClearn 1997). Because of complexity of the trait, efficient data analysis techniques are crucial in helping to interpret the results. In contrast with the rapid development in biological techniques involved, the statistical methods used in data analysis have remained mainly simple χ^2-tests. Various statistical approaches, borrowed from the genetic epidemiology and particularly developed in survival analyses, have been applied. In the following, I summarize the different analytical methods currently applied to identify genes affecting longevity.

1.2.1 The case-control approach

Most of the studies on the genetics of human longevity are cross-sectional association studies in unrelated people. In these studies, polymorphic variants for candidate genes are compared between a group of individuals selected for longevity, such as centenarians, and groups of individuals of younger ages. In this frame, centenarians are considered as cases and younger people as controls. Different applications of this approach can be found in the literature. Examples include the intensive study on apolipoprotein gene variations and their relationship to longevity (Kervinen et al., 1994; Schachter et al., 1994; De Benedictis et al., 1997, 1998; Pepe et al., 1998; Jian-Gang et al., 1998, Christensen et al., 2006). In this context, the χ^2-test is used to compare the frequency of a certain allele or genotype between cases and controls. An effect on aging/longevity by a certain genotype or allele can be detected when a significant difference is found. Although popular in use and powerful for revealing limited chromosomal regions encompassing the susceptibility locus, the case-control method has some disadvantages:

1. Life span is a continuous quantitative trait. Therefore, the case-control approach does not fully make use of the individual survival information available in the analysis, thus it presents a low efficiency for making inferences.

2. The case-control approach may be affected by confounding factors, such as heterogeneity in ethnicity, social environment, sex differential mortality. By stratifying the sample, it is possible to control some confounding factors, but this usually requires large sample sizes because the data have to be divided into smaller subsets and this consequently reduces the statistical power of the sample.

3. Similar to the above problem, the case-control approach is not an ideal method of dealing with interactions. Evidence of gene-environment and gene-sex interactions have been found in previous studies (De Benedictis et al., 1998, 1999; Ivanova et al., 1998). Interactions can be detected by making separate conclusions on different sexes or geographical regions when the sample is accordingly grouped. Again, inference has to be made on considerably smaller subsets of data.

4. As a continuous trait, life span is affected by factors that can be both biological and environmental. This results in individual differences in the aging process. The consideration of such differences is crucial for evaluating the influence of both genetic and environmental attributes in modulating life span. The case-control approach is not able to integrate unobserved heterogeneity.

5. In cross-sectional studies, participants are taken from different birth cohorts. They exhibit heterogeneous patterns of survival due to secular trends in mortality improvement (Vaupel et al., 1998). However, since the differences in individual survival probability are ignored using the case-control approach, the conclusions reached by such an approach could be biased.

1.2.2 The Logistic Regression Approach

As an extension of the case-control approach, the logistic regression model can be used to account for confounding factors or interactions. In the case-control approach, we look at the difference of gene frequencies in two age groups. A natural extension is to study the age trajectories of frequencies. Such a situation can be modeled by the logistic regression model (Tan et al., 2001). In the basic form of a logistic regression

$$\ln \frac{p(x)}{1-p(x)} = \beta_0 + \beta_1 x$$

the odd of the frequency *p(x)* of the allele or genotype of interest is modelled as a linear function of age *x*. When β_1 is significantly different from zero, the frequency of the allele or genotype increases ($\beta_1 > 0$) or decreases ($\beta_1 < 0$) when age increases. The odds ratio for frequency change between two adjacent ages can be measured as e^{β_1} . In the logistic regression approach it is not necessary to group the individuals into cases and controls, thus obtaining more power in the analysis. Moreover, assuming Hardy-Weinberg equilibrium, the logistic regression model with polytomous responses can be introduced to handle highly polymorphic genes (Tan et al., 2003). Genotype and allele-based parameterization can be used to investigate the modes of gene action and to reduce the number of parameters so that the power is increased while the number of multiple testing is minimized. Another very important feature of logistic regression is the possibility of modelling non-monotonous patterns of gene frequencies, which may arise from antagonistic pleiotropic effects in gene action during the aging process. (Leroi et al., 2005). This is achieved by modelling the allele or genotype frequency as a non-linear function of age *x* (Tan et al., 2003), for example by fractional polynomials (Royston et al., 1994). The partial likelihood ratio test can be applied to choose the best fitting model and to make inferences on the statistical significance of the age-dependent pattern as compared with a linear model.

1.2.3 Multi-locus approach

Several studies searched for associations between longevity and susceptibility genes by comparing gene pools of centenarians and younger subjects. However, studies that consider marker genotypes at one locus in connection with survival may capture only a small proportion of the total combined effects of the susceptibility genes which affect the phenotype. Since the development of complex traits involves multiple genes and their interactions, multi-locus association analysis is appealing (Hoh and Ott, 2003). The term multi-locus analysis includes two different procedures. The first is designed to find multiple risk loci, not necessarily on the same chromosome; the second is designed to analyse haplotypes using a set of ordered markers. In the former approach, if multiple risk loci are observed, a joint analysis should exhibit more power because it can capture the interactions among the loci, which are lost in single locus analysis. Statistical tools have been proposed for multi-locus analysis to uncover epistasis in human disease studies

(Cordell, 2002; Bohringer et al., 2003). However, more work is needed in developing and implementing multi-locus models for aging and longevity studies (Tan et al., 2006).

The study of haplotypes and linkage disequilibrium (LD) have proven fruitful in human population genetics. Because particular DNA variants may remain together on ancestral haplotypes for many generations, groups of neighbouring gene variants can form haplotypic diversity with distinctive patterns of LD that can be exploited in both genetic linkage and association studies (Schork et al., 2000). Haplotype analysis is more efficient than the single-locus association test because it makes use of LD information contained in the flanking markers (Akey et al., 2001). Haplotype approaches have been applied to detect allelic associations when parental genotypes are available for phase inference and for constructing the controls (Terwilliger and Ott, 1992; Clayton, 1999). Unfortunately, such methods are not applicable in longevity studies because parental genotype information is unavailable for old subjects. In order to reconstruct the missing phases in the multi-locus genotype data, different algorithms have been proposed among which the well-known EM algorithm (Excoffier and Slatkin, 1995). Recently, methods for haplotype-based multi-locus analysis of human survival have been proposed for cross-sectional (Tan et al., 2005) and for cohort (Tan et al., 2006) studies of unrelated individuals. In these models, a retrospective likelihood function is constructed upon the multinomial distribution of the explicitly observed multi-locus genotypes using haplotype-based parameterization. Moreover, these models are capable of capturing gene-sex and gene-environment interactions (Tan et al., 2005).

1.2.4 Demographic approach

In comparison with the case-control approach, the genetic-demographic (GD) approach uses information on survival functions and mortality rates for groups of individuals carrying the candidate gene or genotype (Yashin et al., 1999). Although more complex, GD remains really promising for its potential in the genetic research on aging and longevity. The complexity of the method is represented by the necessity of introducing:

- a survival model;
- a model that links gene frequencies and survival;
- a procedure for estimating model parameters starting from data;

The GD approach is based on two important assumptions. The first is that the initial gene frequencies in all birth cohorts represented in the cross-sectional sample under study are the same. This is true if there are no particular phenomena, such as bottleneck or genetic

drift, that rapidly change the genetic composition of the population. The second assumption is that the genotypic specific mortality does not depend on the birth year of the cohort. This assumption is not fulfilled because of the changes in socio-economic and sanitary conditions in the last century. To overcome this problem the use of a "synthetic" survival function has been proposed (Yashin et al., 1999; Dato et al., 2006).

The key point of the demographic approach consists in the possibility of using the representation of marginal survival function, *S(x)* - which can be taken from the cohort demographic life tables - as a discrete mixture of the respective survival functions for genotypes (or alleles). The GD approach requires that the whole sample population is divided into carriers and non-carriers of a given allele or genotype.

Let $S_A(x)$, $S_B(x)$ be the survival functions of carriers and non-carriers, respectively, evaluated at age *x*. The relation between the marginal survival functions for allele carriers/non-carriers and the survival function of the population is

$$S(x) = PS_A(x) + (1-P)S_B(x)$$

where *P* is the initial frequency of carriers in the population.

The relative frequency of carriers at age *x*, here denoted as $\pi_A(x)$, is given by:

$$\pi_A(x) = \frac{PS_A(x)}{S(x)} = \frac{PS_A(x)}{PS_A(x) + (1-P)S_B(x)}.$$

The frequency $\pi_B(x)$ of non-carriers at age x is given by:

$$\pi_B(x) = 1 - \pi_A(x) = \frac{(1-P)S_B(x)}{S(x)} = \frac{(1-P)S_B(x)}{PS_A(x) + (1-P)S_B(x)}.$$

The above formulas enable us to compute the age distribution of relative frequencies given the respective survival function and initial frequencies. The survival functions can be written in terms of the relative frequencies of carriers and non-carriers:

$$S_A(x) = \frac{\pi_A(x)S(x)}{P}; \qquad S_B(x) = \frac{\pi_B(x)S(x)}{1-P}.$$

Several approaches for the analysis of combined demographic and genetic data have been proposed (Yashin et al., 1999): the "non parametric method" (NP), the "relative risk method" (RR), the "parametric method" (PR), and the "semi-parametric method" (SP). All these methods are based on the formulas previously reported and share some similarities as, for example, the use of maximum-likelihood criterion for the estimation of the parameters and the existence of constraints such as the age-related decrease of the survival functions. The conceptual difference is that RR, PR and SP methods require that a model of mortality (a trend of the hazard function) is defined a-priori, while NP does not.

1.3 Aim of the study

Aim of the work carried out during my PhD appointment was to explore new routes in genetic studies on human aging and longevity, and in particular:

- to analyze longevity in Calabria in order to verify if the geographic distribution of long-lived subjects is random or not.
- to investigate the multidimensional nature of the quality of aging by trying to obtain an operative phenotypic classification able to measure the degree of vulnerability in elderly people.
- to combine demographic information and multi-locus analysis for estimating the influence by the genetic variability of ten candidate genes on survival at advanced ages.

In the following chapters the results of the above mentioned investigations are reported. In particular, the first part reports the results of a study focused on the spatial analysis of the distribution of the male longevity phenotype in Calabria (Italy).

The second part reports the integral version of a manuscript entitled "A cluster analysis to define human aging phenotypes", which is in press in Biogerontology.

Finally, the third part reports the integral version of a manuscript entitled "Sex-and-age specificity of susceptibility genes modulating survival at old age", which is in press in Human Heredity.

2. SPATIAL ANALYSIS OF THE DISTRIBUTION OF THE MALE LONGEVITY PHENOTYPE IN CALABRIA

(Montesanto et al, Ms. in preparation)

2.1 Background

Individual variability of human lifespan is influenced by an interplay of environmental, genetic and stochastic factors. The disentangling of such factors is complicated by cultural and genetic heterogeneity of human populations. Human societies undergo continuous changes with the improvement of the environment (for instance cleaner water and better food) and medical assistance. This leads to the increase of average and of maximum life span that are now far beyond where they were a few decades ago. Although the improvement of environmental conditions is occurring all over western societies, it has a different pace in different areas, and this reflects on the conditions of aging and of longevity (Jeune et al., 2006). In addition, it has been shown that genetic factors are able to predispose to longevity in certain populations but not in others, either because some variants are population specific or because the interaction of that variant with the environment is specific for the geographic area. On the whole it is emerging that most of the factors influencing longevity are heterogeneous and population specific. This calls for a close monitoring in different populations of the trajectories of lifespan and of their correlation with different factors that are likely to affect longevity in that population.

Recently, it has been reported an intriguing population-specific feature of longevity: the female/male (F/M) ratio among centenarians ranges from 2/1 in Sardinia and southern Italy to 5/1 in northern Italy and most of western European countries (Poulain et al., 2004; Passarino et al., 2002). Subsequently, Robine and co-workers (2006) analysed the F/M ratio among centenarians in four different Italian regions. They showed that the significant differences observed in F/M ratio among centenarians are mainly due to differences of male mortality over the age of 60 years. In particular, they observed that mortality of men over the age of 60 years is lower in Calabria (the most southern region of the Italian peninsula) than in other regions under study and in the whole of Italy. The identification of possible patterns of aggregation of areas characterized by high proportion of long-lived individuals may help to understand the effect of environmental factors on this trait, as well as the role played by the population genetic structure.

The recent observation that in Sardinia centenarians are clustered in restricted areas characterized by a high level of geographic isolation and endogamy (Poulain et al., 2004),

and recent findings indicating an increased homozigosity at loci involved in human longevity (Bonafé et al., 2001; Cardelli et al., 2006) prompted us to explore the effects of population inbreeding on human longevity. Therefore, we decided to study male longevity in Calabria with two objectives:

(i) to verify if the geographic distribution of long-lived subjects is random or not;

(ii) to explore the possible effects of inbreeding on the distribution of long-lived individuals.

The interest was on males because of their particular mortality pattern.

As a longevity index we used the Nonagenarian Rate (NR). It is a modified version of the index called Centenarian Rate (CR) proposed by Robine et al. (2006) which is an attractive tool to compare populations over time and across communities, controlling for the size of birth cohorts, infant mortality, past migrations, and policies of naturalisation (Robine and Paccaud, 2005).

To explore the possible effects of inbreeding on the distribution of long-lived individuals we used surname data. They represent a very important tool for population genetics. In fact, due to systems of surname attribution through the paternal line, the surname can be regarded as a single gene with many alleles transmitted only through the Y chromosome. Different studies have used surname analysis to estimate genetic parameters (e.g. drift, kinship and migrations) relevant in the study of the effects of the population structure on human evolution (Gottileb, 1983; Lasker, 1985; Brunet et al., 2001; Cavalli-Sforza et al., 2004). Recently, Cavalli-Sforza and co-workers (2004) have shown a significant correlation between the estimates of inbreeding obtained from consanguinity and those obtained from surname data.

2.2 Materials and Methods

2.2.1 Data

We used regional census data from 1971 and 2001 (Istituto Centrale di Statistica, 1974, 2002). In particular, we used the regional census data of 1971 to obtain the number of individuals resident in Calabria who in that year were 60-69 years old; we used the regional census data of 2001 to obtain the number of individuals resident in Calabria who in that year were 90-99 years old.

2.2.2 Nonagenarian Rate

The number of oldest-old people that we expect to observe today depends on the number of births in the corresponding cohorts: for example, the number of centenarians who we expect to observe today in a given region depends on the number of births a hundred years earlier in that region. However, a direct comparison of the observed number of oldest-old people with the number of births in the corresponding cohorts does not take into account important phenomena such as infant mortality, past migration etc, which are not related to aging. Recently, Robine et al. (2006), analyzing the evolution of the number of centenarians in Italy, proposed the *Centenarian Rate* (CR). Following this suggestion we have characterized longevity in different municipalities of Calabria according to the Nonagenarian Rate (NR) defined as the ratio between the number of survivors at age 90-99 years and the number of survivors at the age of 60-69 years, whitin the same cohort:

$$NR(i) = \frac{X_{90-99}(i)}{X_{60-69}(i)}$$

where

- $X_{90-99}(i)$ is the number of residents in the i-th municipality, 90-99 years old, registered by the 2001 national census;

- $X_{60-69}(i)$ is the number of residents in the i-th municipality, 60-69 years old, registered by the 1971 national census.

Obviously the Nonagenarian Rate can be defined for any area consisting of several municipalities:

$$NR_A = \frac{\sum_{i \in A} X_{90-99}}{\sum_{i \in A} X_{60-99}}$$

where the summation is over the municipalities enclosed in the area A. Note that NR_A is not the average of the NR values of the municipalities, but the weighted average.

2.2.3 Spatial Analysis

To test the hypothesis that the geographical distribution of oldest-old subjects in Calabria is non-random and that peculiar geographic areas could be identified where NR is consistently and significantly higher than in the whole of Calabria, we used Spatial Analysis (SA). SA is the process of extracting or creating new information about a set of geographic features. Methods of SA can be simple or very sophisticated, but usually they are applied by using a computer-based tool, the Geographic Information System (GIS).

2.2.4 Statistical analysis

If we hypothesize a uniform regional distribution of nonagenarians with respect to the size of the same cohort 30 years earlier in the different municipalities of the region, the expected number of nonagenarians in the i-th municipality is:

$$X'_{90-99}(i) = X_{60-69}(i) \times \frac{\sum_{i=1}^{N} X_{90-99}(i)}{\sum_{i=1}^{N} X_{60-69}(i)}$$

where N is the number of Calabrian municipalities and $NR_{Reg} = \frac{\sum_{i=1}^{N} X_{90-99}}{\sum_{i=1}^{N} X_{60-69}}$ is the regional value of the Nonagenarian Rate (NR).

Significant departures from a uniform distribution of the NR can be identified by using the *chi-square goodness-of-fit test* between observed and expected number of nonagenarians. For this test the statistic is

$$X^2_{Obs} = \sum_{i=1}^{N} \frac{\left[X_{90-99}(i) - X^{Ex}_{90-99}(i)\right]^2}{X^{Ex}_{90-99}(i)}.$$

Since several municipalities show a number of nonagenarians which are less than five, the assumption that the test statistic has a χ^2 distribution may be severely violated. Therefore we constructed the "empirical distribution" of X^2 by numerical simulation, on the null hypothesis that NR is uniform across Calabrian municipalities: $NR(i) = NR_{Reg}$.

Ten thousand patterns of NR(i) were generated, each characterized by a distribution of nonagenarians given as $X^{sim(k)}_{90-99}(i)$, i = 1,…, N. $X^{sim(k)}_{90-99}(i)$ is a pseudorandom integer from

the binomial distribution with parameters *(p,n)* given by $p = NR_{\mathrm{Re}\,g}, n = X_{60-69}(i)$. The rationale is that random departures from a uniform pattern of nonagenarians can come as number of "successes" in samples of different initial sizes ($X_{60-69}(i)$), but with the same "probability of success" ($NR_{\mathrm{Re}\,g}$). For each simulated pattern the statistic X^2 was computed as

$$X_S^2(k) = \sum_{i=1}^{N} \frac{\left[X_{90-99}^{sim(k)}(i) - X_{90-99}^{Ex}(i)\right]^2}{X_{90-99}^{Ex}(i)}, \; k=1,\ldots,10000.$$

The histogram of the values $X_S^2(k)$, k = 1,…, 10000 gives us an "empirical" probability distribution of the test statistic, on the null hypothesis of regional uniformity. Denoting by z the number of values in $\left\{X_S^2(k), k = 1,...,10000\right\}$ which exceed the observed value X_{Obs}^2, the empirical significance level of the test is $p = \frac{z}{10000}$.

With such test we are able to verify if the distribution of nonagenarians across the municipalities of Calabria is random or not.

If the hypothesis of uniformity is rejected, we verify if the observed distribution of nonagenarians in Calabria exhibits a clustering pattern. To this purpose, we repeated the above mentioned tests considering as reference unit the area consisting of one municipality with its neighbours. On the basis of these new reference units, we redefined NR values and tested if the pattern of $X_{90\text{-}99}$ in these new units departs from that corresponding to a uniform regional distribution of survival. Finally, the units (a given municipality with its neighbours) with a significant positive departure from the uniform distribution, identified on the basis of the relevant NR values, were heuristically overlapped to obtain wider areas with high NR values.

2.2.5 Surnames

In order to analyze the distribution of surnames we used the directory of Calabrian users registred by Italian telephone companies, edited by SEAT (CD-ROM 2005). Only persons listed as private telephone users who have given informed consent to use their personal data for direct marketing were used, for a total of 250000 subjects. Fisher's alpha, a

measure of species diversity in a sample (Fisher, 1943), was used as a measure of surname abundance to estimate inbreeding. Singletons surnames were excluded for reasons discussed by Zei et al., (1986), Du et al., (1992), Cavalli-Sforza (2001).

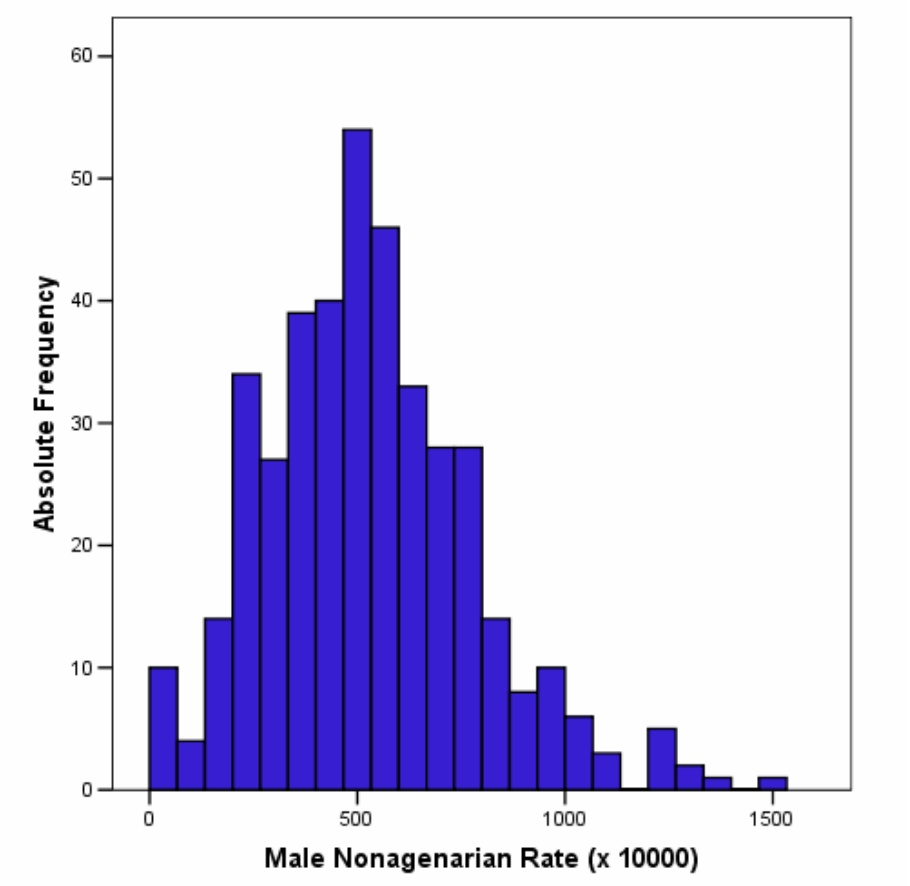
2.3 Results

2.3.1 Identification of Blue Zones

In this study we define as Blue Zone (BZ) an area characterized by a NR value significantly higher than the regional value.

For all municipalities of Calabria we computed the NR values of the male population (mNR). Figure 1 reports the distribution of the mNR computed index across the Calabrian municipalities.

Figure 1. Frequency distribution of Nonagenarian Rates in Calabrian municipalities computed on the basis of the male population (mNR).



Thus, we defined four quartiles of the mNR distribution. The four quartiles were used to define four groups of municipalities: we define as first group (Q1) the municipalities

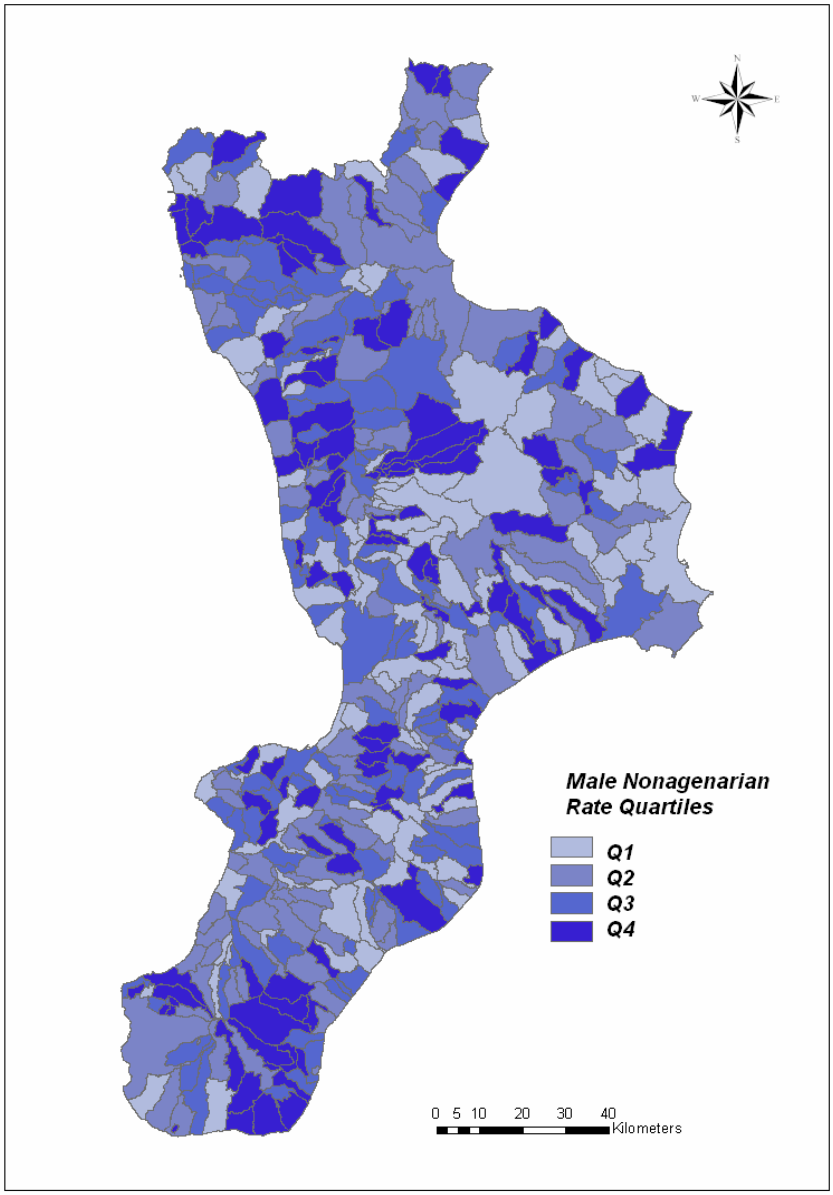
belonging to the quartile with the lowest values of mNR; as second group (Q2) those belonging to the second quartile; as third group (Q3) those belonging to the third quartile; as fourth group (Q4) those belonging to the fourth (Table 1).

Table 1. Municipality classification on the basis of the male Nonagenarian Rate (mNR).

	mNR ($x10^4$)
Group 1 (Q1)	<357
Group 2 (Q2)	357-505
Group 3 (Q3)	505-671
Group 4 (Q4)	>671

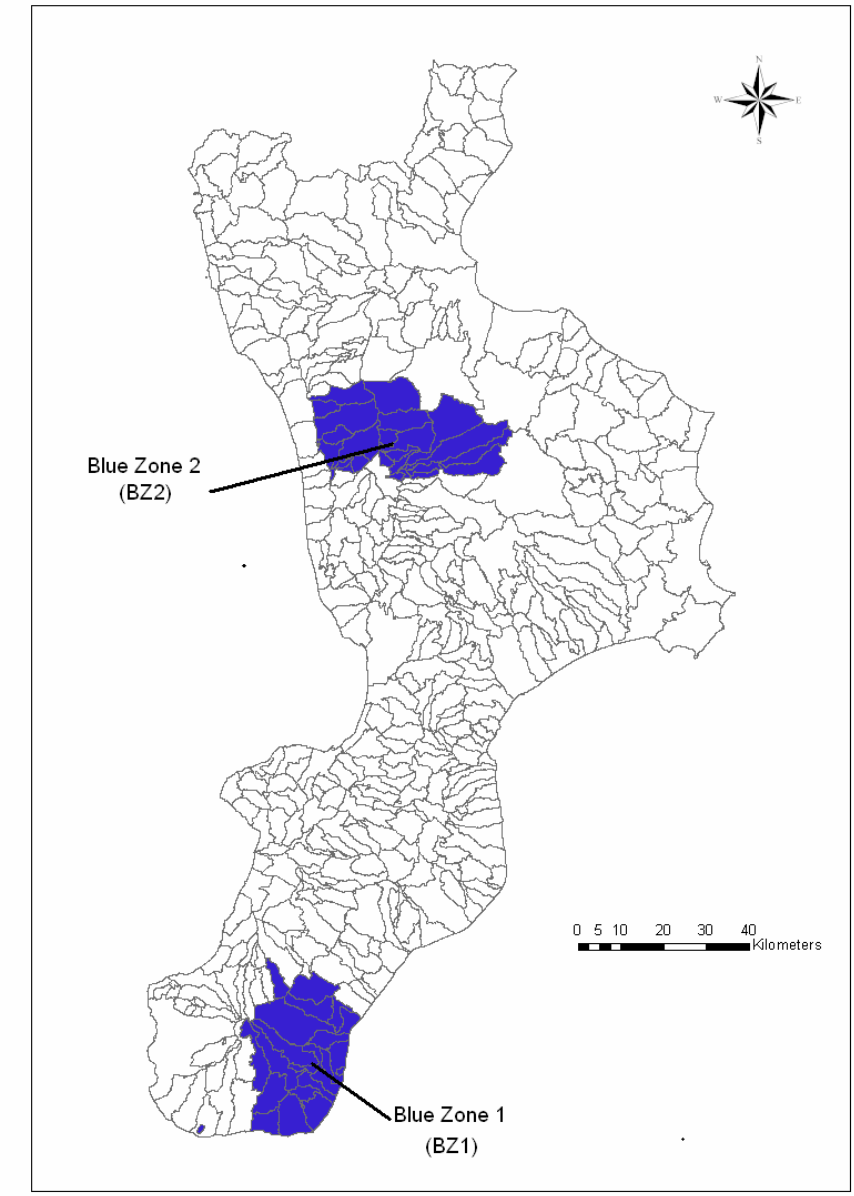
The following map represents the distribution of the mNR values across the Calabrian municipalities (Fig. 2):

Figure 2. Map of the Calabrian municipalities identified with different colours according to belonging to the different mNR groups (Q1, Q2, Q3 and Q4).



In order to verify the existence of areas characterized by very high NR values (see Materials and Methods), we evaluated the mNR for the areas consisting of municipalities and their neighbours. On the basis of these results we heuristically increased the level of aggregation, thus identifying two clusters of municipalities (including about 10% of the population of the respective provinces) characterized by a significant positive deviation of mNR values with respect to the average regional value. The first cluster is located in the province of Reggio Calabria; the second cluster is located in the province of Cosenza (Fig. 3).

Figure 3. Map of Blue Zones, that is of the areas characterized by a significant positive deviation of mNR values with respect to average regional value.



Following Poulain et al. (2004), we defined the areas in which these municipalities are located as Blue Zone 1 (BZ1) the cluster in the province of Reggio Calabria and Blue Zone 2 (BZ2) the cluster in the province of Cosenza. Table 2 reports the municipalities included in BZ1 and BZ2.

Table 2. List of municipalities included in the Blue Zones.

Municipality of BZ1 (Reggio Calabria province)	Municipality of BZ2 (Cosenza province)
Africo	Casole bruzio
Benestare	Castiglione Cosentino
Bianco	Castrolibero
Bova	Celico
Bova marina	Lappano
Bovalino	Lattarico
Brancaleone	Luzzi
Bruzzano Zeffirio	Marano Marchesato
Caraffa del Bianco	Marano Principato
Careri	Montalto Uffugo
Casignana	Rende
Ferruzzano	Rose
Palizzi	Rovito
Platì	San Benedetto Ullano
Roghudi	San Fili
Samo	San Pietro in Guarano
San Luca	San Vincenzo la Costa
Sant'Agata del Bianco	Serra Pedace
Santa Cristina d'Aspromonte	Spezzano della Sila
Staiti	Spezzano Piccolo
	Trenta
	Zumpano

The characteristic features of BZ1 and BZ2 are presented in Table 3, in comparison to the relevant Calabrian provinces and Italy.

Table 3. Characteristics of BZ1 and BZ2.

	BZ1	RC province	BZ2	CS province	Calabria	Italy
Number of municipalities	20	97	22	155	409	1092
Men 60-69 years old at 1971 census	2558	26723	2814	25908	77155	2420136
Men 90-99 years old at 2001 census	187	1295	195	1339	3798	101714
Total Male NR ($x10^4$)	731	485	693	517	492	420
F/M in nonagenarians at 2001 census	1.19	1.82	1.76	1.97	1.98	2.85

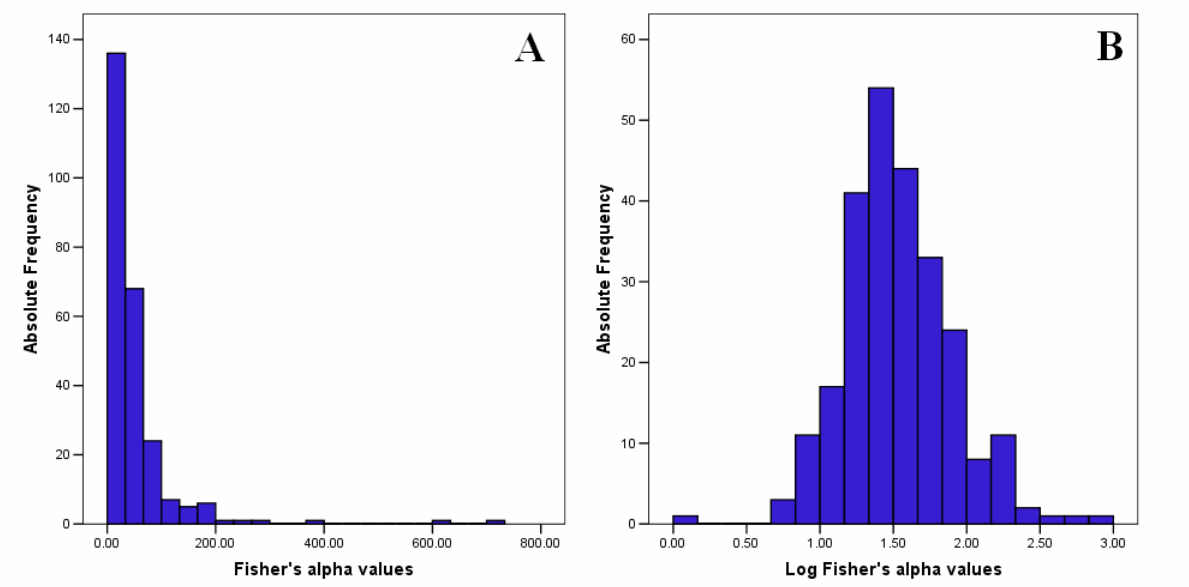
As expected, the mNR values in the BZ are higher than those in the respective provinces, in the entire region (Calabria) and nation (Italy). The significance of the departure of the mNR value in the BZ from provincial and regional values was ascertained by χ^2 test ($P<0.0001$ in all cases).

Interestingly, the F/M ratio in nonagenarians is extremely low in the BZ1, while it is not different from the regional value in the BZ2.

2.3.2 Surname Analysis

To verify a possible contribution of high inbreeding rate to the geographic pattern of male longevity shown in Figure 2, we carried out the surname analysis in Reggio Calabria and Cosenza provinces. The distribution of surname abundance, as measured by the Fisher's alpha values, is shown in Figure 4. The values of Fisher's alpha have been transformed into logarithms because the log-transformed values are approximately normally distributed.

Figure 4. Distribution of Fisher's alpha in Reggio Calabria and Cosenza provinces (A natural scale, B logarithmic scale).



In Figures 5a-d the maps of surname abundances in the two Calabrian provinces are compared to the correspondent BZ maps.

Figure 5. Comparison between the distribution of Fisher's alpha (A and C maps) and the Blue Zones (B and D maps) in Cosenza and Reggio Calabria provinces (Southern Italy). Fisher's alpha is a measure of surname abundance.

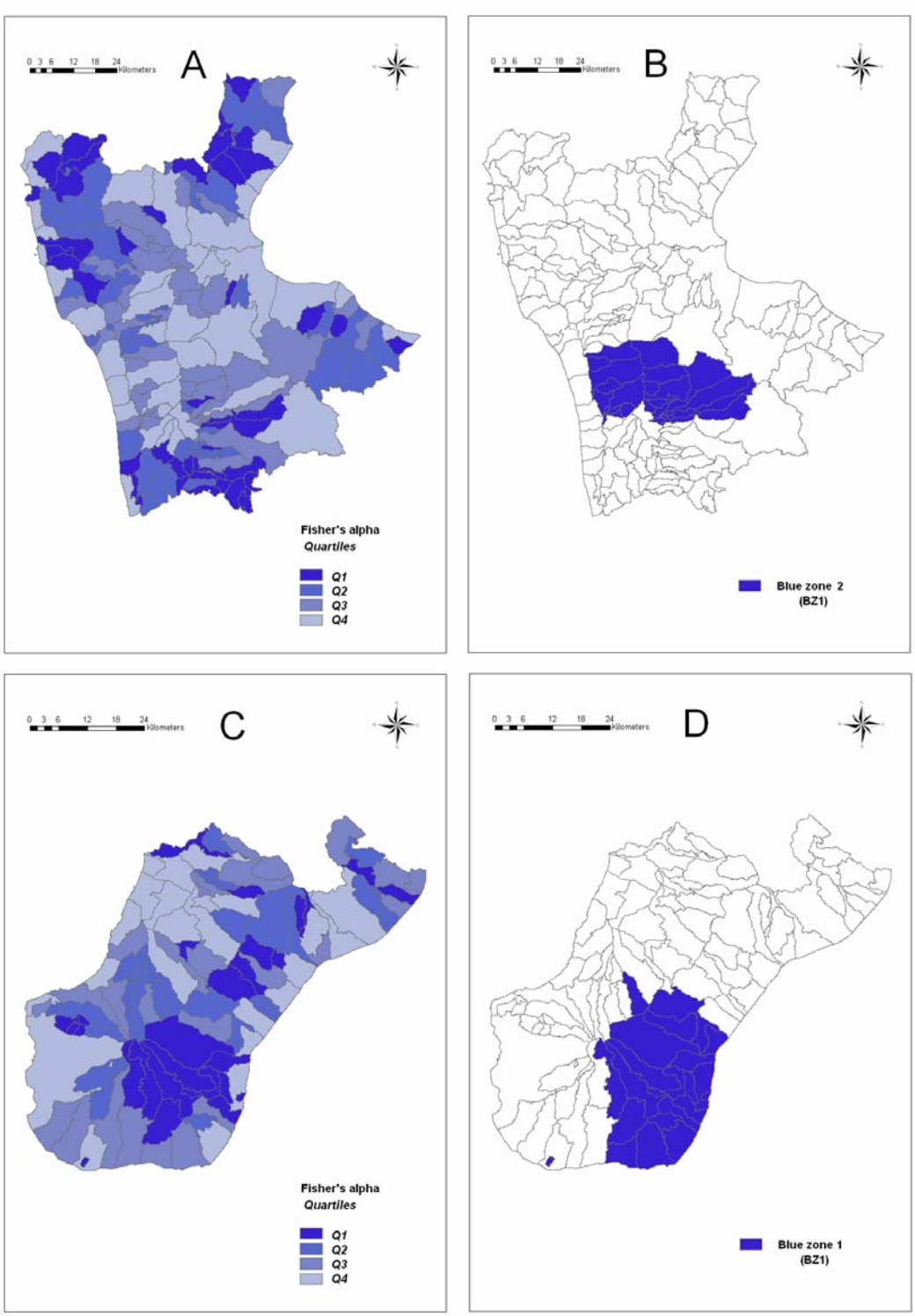


Figure 5 shows a significant overlapping between the BZ in the Reggio Calabria province and an area characterized by low surname abundance (high inbreeding level), whereas in

the Cosenza province the overlapping is not observed. In order to quantify the different relationships between surname abundance and mNR by which BZ were identified, we computed the corresponding correlation coefficients. As expected, we found a significant negative correlation between mNR and surname abundance values in Reggio Calabria province ($r=-0.254$; $P<0.05$), while in Cosenza province the correlation was not found ($r=-0.051$; $P=0.526$).

2.4 Discussion

Male mortality after the age of 60 has been found to be lower in Calabria than in the other Italian regions (Robine et al., 2006). This has led to a an extremely low F/M ratio among oldest olds when compared to other Italian and European areas. In order to better understand this phenomenon, we aimed to verify whether the geographic distribution of long-lived subjects is uniform throughout the Calabrian territory or it is concentred in specific areas. We found that the distribution of oldest old people across the region is not uniform. In addition, we observed a correlation between population inbreeding and male longevity.

By using a SA approach it was possible to identify two areas of male longevity: the BZ1 (located in the province of Reggio Calabria, around Platì territory) and the BZ2 (located in the province of Cosenza, along the middle course of river Crati and near the city of Cosenza). It is interesting to notice that BZ1 is characterized by a very low F/M ratio among nonagenarians, while in the BZ2 this ratio is not different from the regional value (Table 3), due to the female NR which is the highest in the region (data not shown). These data suggest that male longevity characterising the two Calabrian BZ may be related to different causes. In particular, male longevity in BZ1 might be related to some phenomena causing exclusively male longevity, while male longevity in BZ2 appears to be related to a more general phenomenon affecting also female longevity.

From an economic, cultural and geographic point of view BZ1 and BZ2 areas are extremely different. The BZ1 is located in a mountainous area which abruptly slopes down to the Ionian sea, with deep valleys (fiumare); its economy is essentially rural; communication with other parts of Calabria were very difficult up to a few years ago. By contrast, BZ2 is included in an economically, socially and culturally developed area, a node of communication between Ionian and Tirrenian coasts, and north and south of Calabria.

The recent observation that in Sardinia centenarians are clustered in restricted areas characterized by high level of geographic isolation and endogamy, suggested we explore the effects of population inbreeding on male longevity in Calabria. Surname analysis revealed that there was no correlation between surname abundance and mNR values in the province of Cosenza. On the other hand, the surname abundance analysis revealed a significant negative correlation between surname abundance and mNR values in the province of Reggio Calabria: high mNR values are associated with low surname abundance, and hence high inbreeding level.

On the whole these data suggest that the low male mortality in the BZ2 might be related to the improved economic and social conditions. As to BZ1, male longevity might be related to specific features of that area. In the present study, we show that population inbreeding may be one these features. These results, which are in line with those reported by Poulain et al (2004), might be interpreted by assuming that the high inbreeding level marked by the low surname abundance increases homozigosity at loci involved in male longevity (Bonafé et al., 2001; Cardelli et al., 2006).

3. A CLUSTER ANALYSIS TO DEFINE HUMAN AGING PHENOTYPES

Passarino G, Montesanto A et al., (2006) Biogerontology in press.

3.1 Introduction

A prerequisite for disentangling genetic, epigenetic and environmental factors which modulate rate and quality of human aging is the definition of the phenotype. Chronological age is not sufficient, because the variability of the aging trait is manifest within and between populations. Hence several studies searched for indicators of health and functional status in old and very old subjects by which objective phenotypes could be defined (Fried et al., 2001 and 2004; Bortz, 2002; Mitniski et al., 2002; Jones et al., 2004 and 2005). From these studies the concept of *frailty* emerged as a distinct clinical entity characterized by a state of vulnerability for adverse health outcomes, such as hospitalization or death, and therefore correlated to co-morbidity, disability and increased mortality hazard (Walston et al., 2006). Therefore the identification of a precise *frailty* phenotype could help to recognize homogeneous population groups enriched of genetic risk factors predisposing to a poor quality of aging. Cognitive, psychological and functional measures turned out to be the most effective to identify the *frailty* phenotype, since these parameters condense most of the *frailty cycle* that occurs in the elderly (Fried et al., 2004).

We wanted to verify if a Cluster Analysis (CA), that used well established geriatric parameters, was able to recognize the *frailty* phenotype. The term CA encompasses a number of different algorithms and methods for grouping cognate objects in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. By using Mini Mental State Examination (MMSE), Hand Grip strength and Geriatric Depression Scale (GDS) as variable parameters, we analyzed by CA two samples of old subjects recruited in Calabria (southern Italy). This region, as other European regions, is experiencing a great increase in the number of old subjects, probably also due to a reduced mortality at advanced ages consequent to improvement in medical care. For example, people older than 65 and 90 years of age represented 13% and 0.15% of the population, respectively, in 1991 (Italian Census data); at January 2004, these proportions were increased up to 18% and 0.6% respectively. However, despite the increased survival, the quality of aging in Calabria is still poor when compared with other European regions (Jeune et al., 2006).

We collected phenotypic and genetic data in two samples of old Calabrian subjects by using questionnaires, geriatric tests and laboratory analyses. Then we applied CA and

verified, from phenotypic and genetic perspectives, the reliability of the *frailty* phenotypes we obtained.

3.2 Materials and Methods

3.2.1 Samples

Two samples were analyzed. The first (S_1) included 65-85 years old subjects (252 subjects, 107 males and 145 females; median ages 73 and 75 years respectively); the second (S_2) included 117 subjects older than 90 years of age (54 males and 63 females; median ages 98 and 99 years respectively). Out of the 252 subjects in S_1, 80 had at least one living centenarian parent included in S_2. All the subjects were born in Calabria (southern Italy) and their ancestry in the region had been ascertained up to the grandparents generation. The samples had been recruited in the frame of the European research program *European Challenge for Healthy Aging* (ECHA project: http://biologia.unical.it/echa/) and of an independent study we carried out in unrelated people. In both the studies, we collected phenotypic information by using the ECHA questionnaires (http://biologia.unical.it/echa/results.htm). Vital status at 18 months after the visit was traced for 249 subjects in S_1 and for all the 117 subjects in S_2 through the population registers of the municipalities in which the respondents were domiciled. All the subjects had given informed consent for studies on aging carried out by our research group.

3.2.2 Anthropometric and geriatric measures

The physical examination included the record of height, weight, knee-to-floor height and waist and hip circumferences.

Cognitive function was assessed by Mini Mental State Examination (MMSE) test (Folstein et al., 1975). Since the test is affected by age and educational status, the scores were normalized for these variables.

Hand Grip strength was measured by using a handheld dynamometer (SMEDLEY's dynamometer TTM) while the subject was sitting with the arm close to his/her body. The test was repeated three times with the stronger hand. The maximum of these values was used in the CA analysis, after normalization for age, sex and height.

Depression was assessed by the short form (15 items) of the Geriatric Depression Scale (GDS) (Sheikh et al., 1986).

Functional activity was assessed by using a modification of the Katz'Index of ADL (Katz et al., 1963). The assessment was based on what the subject was able to do at the time of the visit.

Health status was ascertained by medical visit carried out by a geriatrician, who also conducted a structured interview including questions on common diseases occurred in the past.

3.2.3 DNA analysis

DNA was prepared from blood buffy-coats according to standard procedures and stored at –20°C until use. *APOE* genotyping (alleles ε2, ε3, ε4) was carried out according to the protocol described in Carrieri et al., 2001.

3.2.4 Statistical analysis

Statistical analyses were performed by using SPSS 10.0 (SPSS Inc., Chicago, IL).

Categorical variables were analyzed by contingency tables (chi-square test). A significance level of $\alpha = 0.05$ was chosen in all the tests. The Ward's method was used to realize hierarchical Cluster Analysis (Ward, 1963).

3.3 Results

Table 1 reports anthropometric characteristics in the two samples, together with information on the three geriatric parameters (MMSE, Hand Grip strength, GDS) utilized for the cluster analysis (CA).

Table 1. Mean values (standard deviation in parenthesis) of Mini Mental State Examination (MMSE), Hand Grip strength, Geriatric Depression Scale (GDS) and anthropometric characteristics of the surveyed subjects. Data are reported by age group and by gender.

	Sample 1 (n=252)		**Sample 2** (n=117)	
	Men (n=107) (65-85 years)	**Women** (n=145) (65-85 years)	**Men** (n=54) (>90 years)	**Women** (n=63) (>90 years)
Median Age (years)	73	75	98	99
MMSE	23.8 (4.72)	21.9 (4.66)	17.0 (5.01)	13.4 (5.56)
Hand Grip strength	28.9 (7.90)	17.6 (5.53)	15.9 (4.63)	10.6 (4.92)
GDS	3.4 (3.60)	6.5 (3.95)	4.8 (3.32)	6.0 (3.53)
Weight (kg)	72.0 (12.96)	64.7 (12.99)	57.0 (9.62)	49.7 (10.68)
Height (cm)	165.3 (6.80)	152.9 (6.00)	156.8 (6.32)	147.0 (7.14)
***BMI (kg/m^2)**	26.3 (4.00)	27.6 (5.03)	23.2 (3.77)	22.8 (4.40)
Waist (cm)	96.4 (10.63)	94.0 (12.22)	89.3 (8.56)	85.6 (11.76)
Hip (cm)	101.5 (9.36)	106.1 (12.06)	96.3 (7.60)	95.6 (11.32)
Knee height (cm)	49.7 (2.76)	44.9 (2.56)	47.31 (2.06)	44.0 (2.88)

*Body Mass Index

3.3.1 Phenotypic classification by CA

The values of MMSE, Hand Grip strength, GDS were used after normalization with respect to the non independent variables. In particular, MMSE scores were normalized for education level ($p=0.007$ in S_1 and $p=0.035$ in S_2) and age ($p<0.001$ in both S_1 and S_2); Hand Grip strength values were normalized for age, sex and height ($p<0.001$ in both S_1 and S_2 in all the cases). The cluster dendrogram plots suggested to stop the clustering process when three clusters were obtained in S_1 and two clusters in S_2. According to the average values of the classification variables (Table 2), we identified the three clusters in S_1 as *non frail, intermediate* and *frail*, phenotypes and the two clusters in S_2 as *frail* and *very frail* phenotypes (Fig.1).

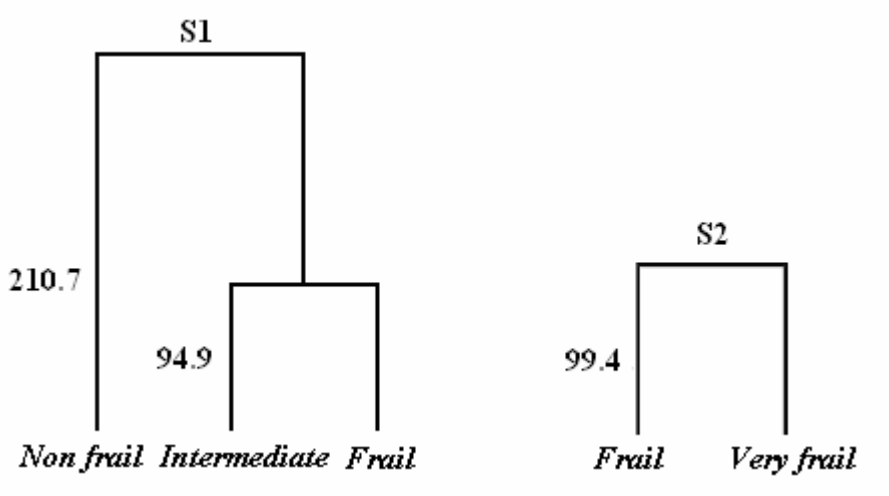
Table 2. Mean values (standard deviation in parenthesis) of MMSE, Hand Grip strength and GDS within the categories obtained by CA in the two samples. For MMSE and Hand Grip strength standardized residuals were used after appropriate adjustment.

Sample 1	**MMSE***	**Hand Grip****	**GDS**
Frail (n=34)	-1.47 (0.696)	-0.83 (1.031)	10.0 (2.42)
Intermediate (n=74)	0.39 (0.597)	-0.14 (0.725)	8.7 (2.69)
Non frail (n=144)	0.15 (0.921)	0.27 (0.986)	2.1 (1.79)
Sample 2	**MMSE***	**Hand Grip****	**GDS**
Very frail (n=72)	-0.58 (0.681)	-0.41 (0.905)	5.73 (3.810)
Frail (n=45)	0.94 (0.666)	0.66 (0.758)	5.02 (3.038)

* Adjusted for age and educational level by using mean differences in score estimated by multiple linear regression models adjusted for covariates (Kang et al., 2005).

** Adjusted for sex, age and height by using mean differences in score estimated by multiple linear regression models adjusted for covariates (Kang et al., 2005).

Figure 1. Schematic representation of the clusters obtained in S_1 and S_2 by applying hierarchical Cluster Analysis which used MMSE, GDS and Hand Grip strength data. We report the sum of square distances between the elements of different clusters (Ward, 1963).



3.3.2 Check of the clusters from a phenotypic perspective.

In order to verify if the classification obtained by CA was indeed able to identify the *frailty* phenotype we looked at the correlation of our classification with co-morbidity and disability. Tables 3 and 4 report physical performance (ADL) and co-morbidity, respectively, in S_1 and S_2 samples classified according to CA.

Table 3. Distribution of disabilities across the groups obtained by CA in the two samples.

Sample 1[#]	**No ADL Disabilities**	**At least 1 ADL disability**	**p (χ^2 test)**
Frail (n=34)	22 (64.7%)	12 (35.2%)	<0.001 (2df)
Intermediate (n=74)	45 (60.8%)	29 (39.2%)	
Non frail (n=143)	135 (94.4%)	8 (5.6%)	
Sample 2	**No ADL Disabilities**	**At least 1 ADL disability**	**p (χ^2 test)**
Very frail (n=72)	9 (12.5%)	63 (87.5%)	0.003 (1df)
Frail (n=45)	16 (35.6%)	29 (64.4%)	

[#] 1 missing data

Table 4. Distribution of co-morbidity across the groups obtained by CA in the two samples. The diseases included in co-morbidity were: diabetes; hypertension; angina pectoris; heart failure; irregular heart rhythm; asthma; chronic bronchitis; arthrosis; arthritis; migraine; osteoporosis; glaucoma; gastric ulcer; hip fracture; stroke; heart attack; cancer; pneumonia.

Sample 1[#]	**No disease**	**1-2 diseases**	**>2 diseases**	**p (χ^2 test)**
Frail (n=34)	-	6 (17.6%)	28 (82.4%)	<0.001 (4df)
Intermediate (n=73)	-	11 (15.1%)	62 (84.9%)	
Non frail (n=143)	11 (7.7%)	62 (43.4%)	70 (49.0%)	
Sample 2	**No disease**	**1-2 diseases**	**>2 diseases**	**p (χ^2 test)**
Very frail (n=72)	1 (1.4%)	20 (27.8%)	51 (70.8%)	0.727 (2df)
Frail (n=45)	-	13 (28.9%)	32 (71.1%)	

[#] 2 missing data

The ADL performances turned out to be significantly correlated to the CA classification in both S_1 and S_2 samples ($p<0.001$ and $p=0.003$, respectively). By contrast, co-morbidity showed to be correlated to the CA classification in S_1 ($p<0.001$) while not in S_2 ($p=0.727$) We further checked our CA classification by verifying how many subjects where still alive after 18 months of our visit (Table 5).

Table 5. Mortality within 18 months from the visit across the groups obtained by CA in the two samples.

Sample 1[#]	**Living subjects**	**Dead subjects**	**p (χ^2 test)**
Frail (n=34)	27 (79.4%)	7 (20.6%)	0.037 (*frail* vs. *intermediate* + *non frail*; 1df)
Intermediate (n=72)	64 (88.9%)	8 (11.1%)	
Non frail (n=143)	132 (92.3%)	11 (7.7%)	
Sample 2	**Living subjects**	**Dead subjects**	**p (χ^2 test)**
Very frail (n=72)	35 (50.0%)	35 (50.0%)	0.128 (1df)
Frail (n=45)	31 (65.9%)	16 (33.1%)	

[#] 3 missing data

It is evident that mortality increased from the least to the most frail groups both in S_1 and S_2, although the difference turned out to be significant in S_1 ($p=0.037$) while not in S_2 ($p=0.128$). Furthermore, in the S_1 sample, the *intermediate* cluster showed an intermediate result.

Then we compared the distribution across the clusters of 80 S_1 subjects having a living centenarian parent with that of 97 sex- and age-matched S_1 subjects whose parents were born in the same years as centenarians, but died at the average life expectancy in the population (Table 6). We found that the two sample groups were differently distributed across the clusters (p=0.009) and that the 61.2% of centenarians' offspring were represented in the *non frail* cluster.

Table 6. Distribution of S_1 subjects across the groups obtained by CA according to the presence/absence of at least one centenarian parent.

Sample 1	Subjects having one centenarian parent	Subjects* having no centenarian parent	p (χ^2 test)
Frail (n=25)	12 (15.0%)	13 (13.4%)	
Intermediate (n=63)	19 (23.8%)	44 (45.4%)	0.009 (2df)
Non frail (n=89)	49 (61.2%)	40 (41.2%)	

* individuals whose parents were born in the same years as the living centenarians, but died at the average life expectancy in the population.

We also tested the correlation of the *frailty* classification obtained by CA with socio-economic variables and found that *frailty* is inversely correlated with the level of economic and cultural status of the subjects in both S_1 and S_2 samples (data not shown).

3.3.3 Check of the clusters from a genetic perspective.

In order to check from a genetic perspective the classification obtained by CA, we verified if the CA phenotypes were correlated to the variability of the *APOE* gene (ε2, ε3, ε4 alleles) that has been widely described as related to *frailty* (Gerdes et al., 2000) and mortality (Schachter et al., 1994). The relatives present in the ECHA sample were not included; therefore 192 subjects in S_1 and 106 subjects in S_2 were analyzed (unrelated people only). In S_1 (while not in S_2) we found that the ε4 carriers were significantly more frequent in the *frail* cluster than in the others (p=0.034, see Table 7).

Table 7. Distribution of carriers and non carriers of ε4 allele across the groups obtained by CA in the two samples (unrelated people only).

Sample 1	**ε4 carriers**	**ε4 non carriers**	**p (χ^2 test)**
Frail (n=27)	6 (22.2%)	21 (77.8%)	0.034 (*frail* vs. *intermediate* + *non frail*; 1df)
Intermediate (n=61)	4 (6.6%)	57 (93.4%)	
Non frail (n=104)	11 (10.6%)	93 (89.4%)	
Sample 2	**ε4 carriers**	**ε4 non carriers**	**p (χ^2 test)**
Very frail (n=66)	10 (15.2%)	56 (84.8%)	0.244 (1df)
Frail (n=40)	3 (7.5%)	37 (92.5%)	

3.4 Discussion

The recent research in gerontology has pointed out the importance of identifying *frail* subjects among old adults, both for understanding the mechanisms which modulate the quality of aging and for obvious social reasons. Although many definitions have been proposed, there is a growing consensus that *frailty* represents a state of vulnerability for adverse health outcomes, including disability, dependency, falls, need for long-term care and mortality (Fried et al., 2004). What is more, the definition of a precise *frailty* phenotype is essential to disentangle the genetic components affecting the decline of the physical functioning at old age (McClearn et al., 1997; Frederiksen et al., 2002). In the present study, we used MMSE, Hand Grip strength and GDS to carry out a hierarchical cluster analysis by which groups of subjects with different degrees of cognitive, functional and psychological status could be classified. In such a way a precise *frailty* phenotype could be defined.

The CA provided a quite effective *frailty* classification as it regards 65-85 years old subjects. Indeed, the three clusters we obtained (Fig.1, S_1) showed to be correlated with disability (Table 3, S_1), co-morbidity (Table 4, S_1) and mortality within 18 months (Table 5, S_1). These results highlight the effectiveness of the geriatric parameters we selected and also show that CA is an appropriate tool for classifying the aging phenotype. It is important to notice that each of the single parameters was not able by itself to be predictive of the different outcomes. Thus, only the combination of the three parameters performed by the CA algorithm provided a consistent classification in three aging phenotypes correlated to ADL, co-morbidity and mortality in a different way. Moreover, the use of parameters

reflecting cognitive (MMSE), psychological (GDS) and physical (Hand Grip strength) functioning produced a classification which is more likely to be based on a real "*frailty*", that is an alteration of multiple systems, than on a specific disease or disorder. The CA classification turned out to be less effective in providing a *frailty* classification when it comes to very old people (S_2). This is likely due to the large prevalence of the *frail* status among this population segment. Indeed, out of 117 subjects who participated to the study, 51 (43.6%) died within 18 months from the visit. However, among the dead subjects, the percentage were 68.6% and 31.4% in *very frail* and *frail* clusters respectively (Table 5, S_2). On the other hand, although disabilities were significantly higher in the *very frail* than in the *frail* group (Table 3, S_2), co-morbidity was equally distributed in the two groups (Table 4, S_2). On the whole, the results obtained in S_2 suggested that among nonagenarians the CA approach could be useful to evaluate at least the quality of life.

An evidence that the classification we obtained by CA corresponds to different *frailty* phenotypes was provided by the analysis of children of centenarians. The use of offspring of centenarians to overcome cohort effects in population studies on human aging is emerging (Atzmon et al., 2006 and references therein). By this approach, it was recently shown that the offspring of centenarians is characterized by a significant lower prevalence of cardiovascular diseases and cardiovascular risk factors, including hypertension and diabetes mellitus, than age-matched controls whose parents died at average life expectancy (Terry et al., 2003). Furthermore, lower all-causes, cardiovascular and cancer mortality has been observed in the centenarians' offspring (Terry et al., 2004). On the basis of the definition of *frailty* (Fried et al., 2001), and on the results of the two cited studies, we can therefore say that the offspring of centenarians is less *frail* than the offspring of non centenarians. The data shown in Table 6 confirm this assumption and validate the phenotypic classification obtained by the CA.

A strong evidence that the CA *frailty* phenotypes correspond to true biological categories, and are not statistical artifacts, is provided by the results shown in Table 7. That allele ε4 is a *frailty* allele is a well established literature data (Gerdes et al., 2000; Lao et al., 2005; Christensen et al., 2006), and recent data show that this allele is associated with a substantial excess of mobility limitation (Melzer et al., 2005). Therefore, the finding that allele ε4 carriers are over-represented in the *frail* cluster ($p=0.034$) shows that this group includes the frailest subjects. The fact that this difference is not shown in very old subjects (S_2) probably depends on the absence of non frail subjects in this age category. In addition,

since the detrimental effect of ε4 allele is evident in the elderly more than in very old people (Dato et al., 2006), it is likely that other genetic components contribute to the phenotypic decline in this age group.

In conclusion, the data here presented show that, within the age range 65-85 years, the *frailty* phenotypes recognized by CA are consistent from a geriatric point of view and have a clear genetic component. However, we wish to underline two points. First, large longitudinal studies are required to validate the CA method as a predictive tool for estimating the quality of aging along the time. Second, the *frailty* classification obtained by means of our approach is highly population-specific and reflects the actual variability of the aging phenotypes within the elderly population. Thus, in order to use such a method in a new population it would be necessary to collect and analyze a suitable amount of data in that specific population. Then the CA classification of new cases/subjects can be correlated to the classification carried out in the relevant population. In any case, the method we propose may provide a further tool for identifying objective phenotypes suitable for the study of genetic variants affecting the quality of aging.

4. SEX-AND-AGE SPECIFICITY OF SUSCEPTIBILITY GENES MODULATING SURVIVAL AT OLD AGE

Passarino G., Montesanto A. et al., (2006) Human Heredity in press.

4.1 Introduction

The probability of reaching very advanced ages in good health depends on a complex interplay of genetic, environmental and stochastic factors. As for genes, it is difficult to disentangle the genetic network which affects the quality of aging and survival up to advanced ages, consequently several strategies have been adopted over the years which utilize familial or population data. In population studies the ideal design is based on longitudinal studies, but these studies are difficult to carry out and usually regard only advanced ages (Bathum et al., 2005; Murakami et al., 2005). On the other hand, most of the information regarding the genetics of human aging has been obtained by cross-sectional studies, under the hypothesis that unfavorable genotypes should be dropped out of the population by a sort of "demographic selection" (Perls et al., 2002) which finally results in an enrichment of favorable genotypes in the gene pool of very old people. However, cross-sectional studies may suffer from the lack of appropriate control groups, as cohort-specific effects may confound comparisons between very old people (for example centenarians) and younger cohorts (Christensen et al., 2006). The problem is hindered by the rapid changes which occur in human society that increase the level of population heterogeneity, thus introducing a further complicating factor. To cope with this problem, algorithms which integrate genetic and demographic data have been proposed (Yashin et al., 1999).

A further complication arises from evidence showing that the same gene variant may affect aging and survival according to age-specific patterns (De Benedictis and Franceschi, 2006). Consequently, the identification of well-defined age-related phenotypes is a key objective in the analysis of susceptibility genes affecting the quality of aging and survival at old age.

Last but not least, common genetic variants with important effects on human longevity are unlikely to exist because of the rather low genetic contribution to the trait. On the contrary, many interacting genes whose individual effects are rather small probably affect survival at old age (Hjelmborg et al., 2006). Thus, studies that consider marker genotypes at one locus in connection with survival may capture only a small proportion of the total combined effects of the susceptibility genes which affect the phenotype. In the genetic analysis of

complex traits multilocus approaches based on regression analyses have been proposed to overcome this kind of problem (Hoh and Ott, 2003). Multilocus approaches could be especially informative if candidate genes for which there is evidence of a biological role for the trait are included in the analysis.

In the present study we analyzed the hypothesis that genes exert a sex an age specific effect on the probability to reach advanced age in good health. We tackled the problems described above by adopting the following strategy. First, although we used cross-sectional data, these were collected in a population (Calabria, Southern Italy) which is characterized by high genetic homogeneity and a scarce level of immigration due to geographical, historical and social reasons. This choice should minimize cohort effects. Second, we used demographic information to categorize the sample into rationally defined age classes. Third, we applied a multilocus approach to a dataset including only genes for which there is documented evidence regarding their biological effects on aging and longevity.

The results presented here show that genetic risk factors act on the probability of reaching advanced ages in good health according to sex-and-age specific patterns.

4.2 Materials and Methods

4.2.1 Genetic data

The dataset comprised of genotype information from 972 unrelated subjects (451 males and 521 females) free of clinically overt pathologies (ascertained by medical visit), whose ages ranged from 18 to 106 years (see Supplementary Material, Fig.1_{SM}). All the subjects were born in Calabria (Southern Italy), and their ancestry in the area was ascertained (by interview) back to the grandparent's generation. The samples were collected during the period 1996-2004 and the buffy-coat DNA extracted and stored in our laboratory. The recruitment of centenarians (identified through the birth registers of the 409 Municipalities in Calabria) started on 1996, while the remaining subjects were recruited by an appropriate campaign launched in 1999 and concluded within two years. The recruitment campaign was focused on students and staff of the University of Calabria (18-60 year old subjects) and people visiting thermal baths in the area and the Academy of the Elderly (60-80 year old subjects). All subjects consented to their phenotypic and genetic data to be used anonymously for genetic studies on aging and longevity (informed consent).

The genetic polymorphisms considered in the present study (Table 1) were chosen because of functional effects that were in some way related to aging and longevity according to the current literature. The relevant experimental protocols are reported elsewhere (see note to

Table 1). For all the loci genotypic and allelic frequencies are reported in the Supplementary Material.

Table 1. Loci and polymorphisms analyzed*.

Locus	Polymorphism (Gene Bank Accession)	Biological effect
APOA1	628 A>G (J00098)	Plasma level of LDL-cholesterol (Garasto et al., 2003)
APOA4	1033 A>G (AY422950)	Plasma level of LDL-cholesterol (Garasto et al., 2003)
APOB	3' VNTR (AC115619)	Plasma level of LDL-cholesterol (Garasto et al., 2004)
APOE	21250 T>C and 21388 C>T (AF050154)	Antioxidant activity (Miyata and Smith, 1996)
HSP70-1	163 A>C (M11717)	Protein synthesis of HSP70.1 (Marini et al., 2004)
HSP90α	888 C>T (M27024)	Folding and maturation of proteins (Passarino et al., 2003 and references
HSP90β	5976 A>C (J04988)	Folding and maturation of proteins (Passarino et al., 2003 and references
SIRT3	5^{th} intron VNTR (AC136475)	Enhancer activity (Bellizzi et al., 2005)
TH	HUMTHO.1 STR (M23597)	Gene expression regulation (Meloni et al., 1998, 2002)
mtDNA	Haplogroups (www.mitomap.org)	OXPHOS efficiency (Ruiz Pesini et al., 2000)

Notes.
* Relevant experimental protocols are reported in: De Benedictis et al., 1997 (APOB); De Benedictis et al., 1998 (TH); De Benedictis et al., 1999 (mtDNA); Carrieri et al., 2001 (APOE); Altomare et al., 2003 (HSP70.1); Passarino et al., 2003 (HSP90-α and HSP90-β); Garasto et al., 2003 (APOA1 and APOA4); Bellizzi et al., 2005 (SIRT3).

4.2.2 Demographic curves

The samples were drawn from the population living in Calabria in the years 1996-2004 and contain individuals from 18 to 106 years of age. In these samples the cohorts born from the end of the XIX century to 1980 are represented. Each cohort is characterized by a specific mortality pattern which reflects the changes occurring in this period. If x denotes the variable representing age and t_B the birth year of the cohort, it follows that the age-related evolution of the population in that cohort is represented by the mortality function $\mu(x;t_B)$ and the related survival function $S(x;t_B)$. The latter function is defined as the solution of the differential equation $\frac{dS}{dx} = -\mu(x;t_B)S(x;t_B)$ with initial condition $S(0)=1$, and is given by $S(x;t_B) = exp\left(-\int_0^x \mu(\xi;t_B)d\xi\right)$. The survival function is used to obtain a partitioning of the samples into subgroups to which the logistic analysis is applied. However, in order to take into account the cross-sectional nature of our samples, we constructed a "synthetic" survival function, $S_S(x)$, as suggested by Yashin et al., 1999, who for the first time introduced a genetic-demographic model for the study of human longevity (see also Dato et al., 2006). The value of the "synthetic" survival curve at age x is taken as $S_S(x) = S(x;t_{BX})$ where t_{BX} is the calendar year of birth of the individuals aged x at the time of recruitment of the sample.

The survival functions for all cohorts from 1890 to 1980 were computed using death counts per calendar year and per year of birth, and cohort initial values as reported for the Italian population in the Human Mortality Database, available online at www.mortality.org. Individual survival functions were used to construct the synthetic survival function. However, since the resulting survival function shows unacceptable oscillations due to errors and inconsistencies in demographic data, a smoothed function is used. The smoothed function assumes a Gompertz-Makeham model for $\mu(x)$, with parameters obtained by least square fitting of the raw data, according to a procedure routinely used especially for datasets containing older ages (Horiuchi and Coale, 1990). In order to avoid cumbersome notations, the smoothed synthetic survival function of the Italian population will hereafter be denoted by $S(x)$.

4.2.3 Strategies for logistic model building

We set up a multinomial logistic regression model to evaluate the effects of genotypes (independent variables) on the probability of belonging to different age groups (dependent variable). In order to build the multinomial logistic regression models genotype data needed to be coded as binary independent variables. This is usually done by assigning code 1 for a genotype (or a group of genotypes encompassing a given allele) and using 0 for the remaining ones. In order to minimize the inevitable loss of information which is implied in this step, the coding of genetic variability needs to take into account the dominant/recessive effects of different alleles (Relton et al., 2004). Thus, for each locus we evaluated the effect of genotypes on the probability to be part of the next age-group (Tan et al., 2003) by comparing specific logistic models including the different genotypes. Those genotypes including the allele showing the highest impact on such probability, that is the logistic model with the highest significance, were then pooled together (see Table 2).

Table 2. Genotypes used for multivariate analysis. Allele nomenclature: APOA1, APOA4, HSP70-1, HSP90α, HSP90β: see Table 1; APOB: number of repeats <35=S, ≥35=L; APOE: ε2/ε3/ε 4; SIRT3 and TH: allele number refers to the repeat number; mtDNA: haplogroups.

Locus	Genotypes in Females	Genotypes in Males
APOA1	A/-	A/-
APOA4	A/-	A/-
APOB	S/S	S/-
APOE	ε4/-	ε4/-
HSP70-1	C/-	A/-
HSP90α	A/-	A/-
HSP90β	C/-	A/-
SIRT3	2/-	2/-
TH	9/-	6/6
mtDNA	H,V	U,K

Using this procedure, the result is a homozygous genotype *versus* all the others when an allele has a recessive effect; whereas many genotypes carrying a given allele *versus* all the others when the effect of that allele is dominant. Then the resulting genotypes were used as independent variables for building logistic multivariate regression models using a stepwise approach (Lachin, 2000). The stepwise procedure for the selection of variables is based on an algorithm that checks for the "importance of variables" (Hosmer and Lemeshow, 2000) and either includes or excludes them on the basis of their fit on the logistic model. The level of significance for the covariates (loci) to be added and remain in the stepwise procedure was set at 0.1 and 0.15 respectively (Hosmer and Lemeshow, 2000). The "importance of a variable" in the model is estimated by the deviance in the likelihood ratio (G^2) between the model excluding the variable and the model including it. The variable with the best fit is the first to be entered; then the second best is entered, and so on. When the addition of a new variable does not produce a significant increase in G^2, no further variables are added to the model.

Once the model with the best predictor variables was obtained, two-way interactions between all the available covariates were tested. Also in this case the estimation of the "importance" of the two-way interactions was assessed by likelihood ratio testing.

The goodness-of-fit statistic for the models was assessed by the Hosmer and Lemeshow test (Hosmer and Lemeshow, 2000).

In order to compare the models obtained for each sex we calculated the Nagerlkerke index (Nagerlkerke, 1991). The Nagerlkerke index is $\mathrm{R}^2 = \dfrac{\widetilde{\mathrm{R}}^2}{\max(\widetilde{\mathrm{R}}^2)}$ where $\widetilde{\mathrm{R}}^2 = 1 - \left(\dfrac{l(0)}{l(\boldsymbol{\beta})}\right)^{\frac{2}{\mathrm{n}}}$, $\max \widetilde{\mathrm{R}}^2 = 1 - [l(0)]^{\frac{2}{\mathrm{n}}}$, $l(0)$ is the log-likelihood for the model with only the constant, and $l(\boldsymbol{\beta})$ is the log-likelihood for the model under consideration.

4.3 Results

Identification of age classes according to demographic information

In order to categorize the samples into age classes by using a rational criterion, we set up a synthetic survival function (Fig.1). It should be noted that the curves of Fig.1 do not reflect the mortality experience of any of the cohorts we considered; nevertheless they provide a criterion for choosing the age thresholds for the subsequent logistic analysis. Let us consider the curve representing the survival function of males: since $S(x)$ is decreasing, the

slope is always negative, but in the 65-70 year age interval we can observe the largest negative change of slope, whereas in the 85-90 age interval we have the largest positive change of slope. A similar observation can be seen in females. A quantitative measure of the changes of the slope is given by the second derivative of the survival function: $S''(x) = \frac{d^2 S}{dx^2}$. The graphs of the second derivative of the survival functions of both males and females are shown in Fig. 2. The minimum value of $S''(x)$ occurs at x=66 years of age in males and x=73 years in females, corresponding to the largest change of slope, in the negative sense, of the curves in Fig. 1. The maximum value of $S''(x)$ is obtained at x=88 and x=91 years of age in males and females respectively, corresponding to the largest positive change in slope. On the basis of such results we defined three sex-specific age groups: Group 1 (G1) included males younger than 66 years and females less than 73 years of age; Group 2 (G2) included males within the age range 66-88 years and females between the ages of 73-91 years; Group 3 (G3) included males older than 88 years and females over 91 years of age.

Figure 1. Synthetic survival functions built for males and females in the Italian population from 1890 onward. Demographic data were obtained from www.mortality.org. Since the minimum age in our samples is 18 years the functions are normalised so that the value at age 18 was 1.

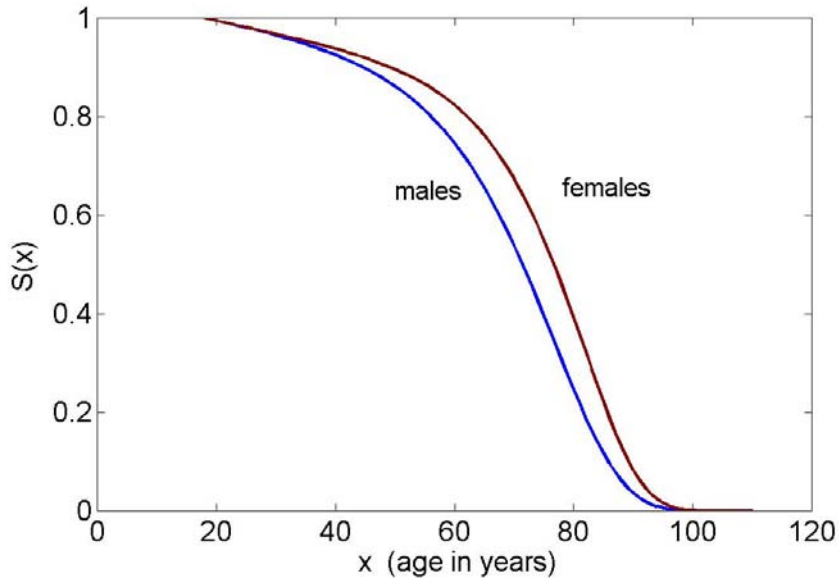
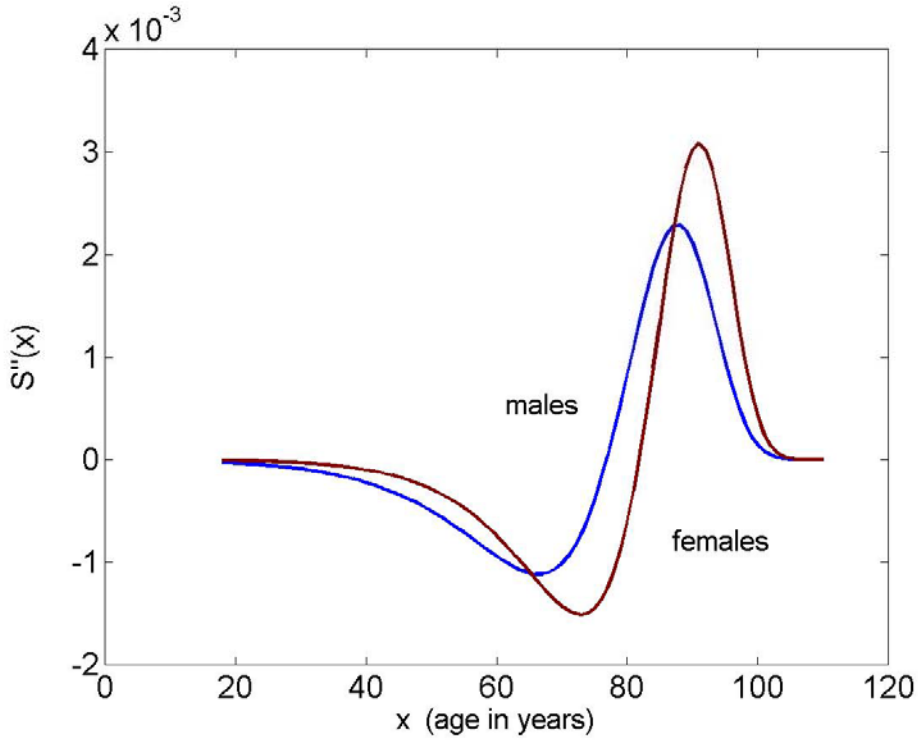


Figure 2. Second derivative (*S''*) of the survival function built from Fig.1.



4.3.1 Logistic analysis

Logistic multivariate analysis was performed by the forward stepwise procedure. The youngest group (G1) was considered as the reference group. The variables added to the logistic model at each step are reported in Table 3, where it can be noted that the "importance of variables" (Hosmer and Lemeshow, 2000) clearly differed between females and males.

Table 3: Loci (variables) included in the final regression model.
The asterisk indicates interaction between loci.

Females

Step	Variable	χ^2	df	p-value
1	APOE	8,997	2	0,011
2	HSP70-1	8.157	2	0,017
3	HSP90β	7.757	2	0,021
4	mtDNA	7.065	2	0,029
5	HSP70-1*mtDNA	5.333	2	0.069

Males

Step	Variable	χ^2	df	p-value
1	APOA4	40.475	2	0,000
2	APOA1	14.449	2	0.001
3	SIRT3	8.363	2	0,015
4	TH	6.804	2	0,033
5	APOE	1.809	2	0.405
6	APOA4*APOE	6.191	2	0.045
7	APOA1*TH	5.431	2	0.066

Note. The APOE main effect is included in the model as it is part of a significant interaction.

Finally, by using the variables (loci) reported in Table 3, we compared G2 with G1, G3 with G2, and G3 with G1 (model 1, 2 and 3, respectively) in order to estimate if, and to what extent, the probability to be assigned to the different age classes could be related to these variables. The final models with the relevant Odd Ratios (O.R.) are shown in Table 4 where we highlighted, for each comparison, genotypes which are unfavorable (yellow) or favorable (red) to the probability to be assigned to the older age group as compared to the probability to be part of the younger one, as estimated on the basis of O.R. values. Finally

we verified the Goodness-of-fit of the models by the Hosmer and Lemeshow test (Hosmer and Lemeshow, 2000). The goodness-of-fit statistic was significant in every comparison (in females: p=0.986, p=0.999, p=0.992 in comparisons 1, 2 and 3, respectively; in males: p=0.801, p=0.652, p=0.602 in comparisons 1, 2 and 3, respectively).

Table 4. Final regression models with estimated Odds Ratio (O.R.) and p-value for the 3 comparisons included in each mutinomial model. For comparison 1 and 3 (both using G1 as reference group) O.R. were obtained directly from the equations included in the models, for comparison 2 O.R. were obtained by difference of equations included in the model (Agresti, 1996). O.R. lower (yellow) and higher (red) than 1 indicates negative and positive effect respectively on the probability to be assigned to the older "survival phenotype in healthy status".

Females (G1: 18-72 years; G2: 73-91 years; G3: 92-106 years).

	Comparison 1 (G2 on G1)			Comparison 2 (G3 on G2)			Comparison 3 (G3 on G1)		
Variable	**O.R.**	**95% CI**	**p-value**	**O.R.**	**95% CI**	**p-value**	**O.R.**	**95% CI**	**p-value**
APOE	0.854	0.464-1.570	0.611	0.333	0.113-0.979	0.046	0.285	0.110-0.739	0.010
HSP70-1	1.851	1.148-2.987	0.012	0.853	0.446-1.629	0.629	1.565	0.941-2.603	0.085
HSP90β	0.923	0.557-1.529	0.755	0.465	0.217-0.994	0.048	0.440	0.231-0.839	0.013
mtDNA	0.710	0.439-1.146	0.161	2.379	1.290-4.387	0.006	1.700	1.046-2.763	0.032
HSP70-1*mtDNA	0.651	0.237-1.786	0.404	4.937	1.294-18.843	0.019	2.826	0.977-8.171	0.055

Males (G1: 18-65 years; G2: 66-88 years; G3: 89-102)

	Comparison 1 (G2 on G1)			Comparison 2 (G3 on G2)			Comparison 3 (G3 on G1)		
Variable	**O.R.**	**95% C.I.**	**p-value**	**O.R.**	**95% C.I.**	**p-value**	**O.R.**	**95% C.I.**	**p-value**
SIRT3	0.929	0.546-1.581	0.786	0.464	0.206-1.043	0.053	0.376	0.182-0.776	0.008
APOA1	0.575	0.365-0.908	0.017	0.696	0.383-1.267	0.236	0.430	0.259-0.714	0.001
APOA4	0.927	0.558-1.541	0.770	4.015	2.252-7.159	< 0.001	3.801	2.309-6.259	< 0.001
APOE	0.786	0.445-1.387	0.405	0.888	0.395-1.995	0.773	0.541	0.270-1.084	0.083
TH	2.004	0.983-4.085	0.056	0.301	0.099-0.918	0.035	0.571	0.194-1.682	0.309
APOA1*TH	0.156	0.029-0.834	0.030	2.929	0.204-42.014	0.429	0.292	0.023-3.664	0.340
APOA4*APOE	0.108	0.012-0.948	0.045	8.449	0.752-94.948	0.084	0.718	0.178-2.892	0.641

In order to draw attention to whether the impact of the analyzed genetic variability was different in the two sexes we calculated the Nagerlkerke index, which estimates the overall impact of the multivariate models on the total variance, for both models obtained in males and females. The index was 19.6% in males and 8.2% in females. Next, to evaluate the impact of the genetic variability at different ages we estimated Nagerlkerke indexes for comparisons 1 and 2 reported in Table 4. In females we found that the variance explained by the genetic data is 3.2% in comparison 1, and 14.6% in comparison 2; in males we obtained 9.1% in comparison 1, and 23.4% in comparison 2.

4.4 Discussion

Taking advantage of a large genotype dataset assembled in 1996-2004, the aim of the present work was to estimate by multilocus analysis the influence of the genetic variability of ten candidate genes on survival at advanced ages in good health. To our knowledge, multilocus analysis has never been applied to this trait.

Before discussing the results we obtained, we wish to point out two features of the present study. First, as we were dealing with cross-sectional data, we paid particular attention to the quality of the sampling in order to avoid false positive results due to population stratification. In fact the study was carried out in a genetically homogeneous population (Calabrians) in which the origin of every subject was meticulously verified. Second, we used a rational criterion for categorizing the entire sample into age classes on the basis of a synthetic survival curve assembled by demographic data (Yashin et al., 1999). Although the definition of three survival groups based on a synthetic survival curve may appear arbitrary, the age cutoffs we identified by this approach are consistent with literature data (Yashin et al., 1999; Ukrainsteva and Yashin, 2001). Indeed, the negative change of slope of the survival curve may be related to the re-setting of biological parameters following the reproductive period; the further re-setting of biological parameters that occurs at old age may contribute to the positive change of slope of the synthetic survival curve (Franceschi et al., 2000). Finally, as the opportunity to use a stepwise procedure is debated, chiefly when it is used for blind data mining, we want to point out the following points: i) the covariates were not gathered by chance; by contrast we selected ten loci for which a possible role in longevity had been previously hypothesized on the basis of biological data; ii) the genes we selected are independent from each other and the analysis was carried out in a homogeneous population. This rules out the possibility of colinearity between covariates, which is the main problem when dealing with stepwise regression; iii) we

checked that regression models including all the covariates showed the same results of the two models selected by stepwise regression (data not shown).

On the whole, since the subjects included in the analysis were selected for healthy status (all of them were free of overt pathologies), the procedure we used for selecting age classes may represent "survival phenotypes in healthy status". The entire set of data reported in Tables 3 and 4 provides two indicators: i) sex-specific gene effects; and ii) age-specific gene effects.

Sex-specific gene effects

Both Tables 3 and 4 show that the regression models include loci (regressors) which are different in males and females. This finding implies that a susceptibility locus affects the probability of survival in healthy status by means of loci and genotypes which are sex-specific. The only exception is the *APOE* gene which acts in both females and males. Therefore, also through this approach, the crucial role played by *APOE* variability on healthy status and survival is confirmed (Gerdes et al., 2000). In addition, the finding that both *APOA1* and *APOA4* significantly contribute to the regression model in males (Table 4) is in line with the important role of lipoprotein genotypes in survival (Atzmon et al., 2006). Sex-specific gene effects are confirmed by the observation that the gene impact on the "survival phenotype in healthy status" is higher in males (Nagerlkerke index=19.6%) compared to females (Nagerlkerke index=8.2%). This implies that the variability of the genetic factors affects the trait under study more in males than in females. This finding is also in line with literature data (Ivanova et al., 1998; Tan et al., 2001; Bellizzi et al., 2005).

Although in our models genetic effects are overestimated (environmental and stochastic factors are not considered), the above results indicate that the genetic component is more important in male than in female longevity. This finding parallels demographic data. In fact, it has been reported that the continued progress of medical and social assistance in developed countries is favoring female longevity more than male longevity. For instance, in most northern European countries, the increase in number of nonagenarians and centenarians is higher for women than men, and the female/male ratio for this segment of the population is continuously growing (Robine and Paccaud, 2005).

On the whole, the general agreement between our results and current literature data confirms the validity of the multilocus approach in research concerning the genetics of human aging.

Age-specific gene effects

Table 4 clearly shows that although more than one gene affects the probability of young subjects of attaining longevity (comparison 3) these genes have a different impact on survival in healthy status depending on the different ages. For example, the *APOE* gene (females) mainly acts on the probability of subjects of the second "survival phenotype" to be part of the third "survival phenotype" (comparison 2), while the *APOA1* gene (males) acts on the first "survival phenotype" (comparison 1).

Furthermore, the finding that the Nagerlkerke index is higher for comparison 2 (G3 *versus* G2) than for comparison 1 (*G2 versus G1*) in both sexes (for females 3.2% in comparison 1, and 14.6% in comparison 2; in males 9.1% in comparison 1, and 23.4% in comparison 2) shows that the genes under study modulate the probability of attaining longevity chiefly by acting on the second "survival phenotype". These observations are paralleled by data gathered by Lipsi and Caselli (2002) regarding causes of death. In fact, these data show that prevalent causes of death are age-specific, and that the incidence of the different causes of death leads to mortality trajectories typical of each population. On the other hand, different demographic studies on the emergence of centenarians in developed countries from 1950 onward have shown that the prevalence of centenarians in a population is mainly caused by the improvement of the survival rate of the octogenarians in the population (Vaupel and Jeune, 1995). Thus, although demographic selection is active since the birth, the prevalence of long-lived subjects in a population is related to the survival rate at about 80 years, which may be related not only to environmental factors (e.g. social and medical assistance), but also to the presence of genetic factors which predispose to longevity and act in a sex and age specific way. Our findings are also in line with literature data suggesting that a given allele may change its adaptive significance according to cell microenvironments that characterize the various ages in the life (De Benedictis and Franceschi, 2006). An important consequence of this is that the medical approach to the elderly should be age and sex specific. This may have important consequences with reference to social and medical strategies aimed at improving the quality of life in a segment of population which is expected to increase dramatically in next years.

In conclusion, the analyses presented here show the validity of multilocus approaches in research concerning the genetics of human aging and reveal a sophisticated picture which underlies the complexity of the gene network modulating human survival at advanced age in healthy status.

5. CONCLUSIONS

Aging represents the major socio-economical challenge for western societies. One of the main results of the research in this field is that aging is a very complex trait, which is influenced by a number of factors. The analysis of some of these factors was the aim of the work I carried out during my PhD appointment. The main results of such work can be summarized as follows:

- **THE SPATIAL ANALYSIS OF THE DISTRIBUTION OF MALE LONGEVITY PHENOTYPE IN CALABRIA.**

  The Spatial Analysis of the distribution of long-lived individuals in Calabria revealed not uniform pattern. In particular, it was possible to identify two areas of male longevity located in two provinces of the region. As index of longevity the Nonagenarian Rate (mNR) was taken. To explore the possible effects of population inbreeding on the distribution of long-lived individuals in such provinces, we carried out a surname analysis which allows to identify areas characterized by high inbreeding levels. The relationship between inbreeding and male longevity was tested by computing the correlation coefficients between mNR and surname abundance in Reggio Calabria and Cosenza provinces. We found a significant negative correlation in one of the two areas. This result suggests a possible effect played by the genetic structure of the population on the spatial distribution of male longevity in that area.

- **A CLUSTER ANALYSIS TO DEFINE HUMAN AGING PHENOTYPES.**

  The definition of a precise and consistent aging phenotype that allows to measure the physical and cognitive decline, as well as the increase of mortality hazard late in life, is a major problem for studies aimed at finding the genetic factors modulating rate and quality of human aging. In this frame, it seems promising the concept of *frailty* which tends to figure out the subjects who are more vulnerable and more prone to negative outcomes, such as death or hospitalization. Cognitive, functional and psychological measures turned out to be the most effective measures to define *frailty*, as they condense most of the *frailty cycle* that occurs in the elderly and is probably responsible of the aging related physical decline. We used MMSE, Hand Grip strength, and GDS as variable parameters in a hierarchical Cluster Analysis (CA) in order to recognise aging phenotypes. By using a sample of 65-85 years old subjects we identified three

frailty phenotypes that were consistent from both geriatric and genetic perspectives. Therefore, the method we propose may provide unbiased phenotypes suitable for the identification of genetic variants affecting the quality of aging in this age range. The CA method was less effective in ultranonagenarians, probably due to the high prevalence of *frail* subjects in this age group that makes difficult to distinguish discrete phenotypes.

- **SEX-AND-AGE SPECIFICITY OF SUSCEPTIBILITY GENES MODULATING SURVIVAL AT OLD AGE.**

  We aimed to investigate the influence of the genetic variability of candidate genes on survival at old age in good health. First, on the basis of a synthetic survival curve built by using historic mortality data taken from the Italian population from 1890 onward, we defined three age classes ranging from 18-106 years. Second, we assembled a multinomial logistic regression model to evaluate the effect of dichotomous variables (genotypes) on the probability to be assigned to a specific category (age class). Third, we applied the regression model to a cross-sectional dataset (10 genes; 972 subjects selected for healthy status) categorized according to age and sex. We found that genetic factors influence survival at advanced ages in good health in a sex and age specific way. Furthermore, we found that genetic variability plays a stronger role in males than in females and that, in both genders, its impact is especially important at very old ages.The analyses presented here underline the age-specific effect of the gene network in modulating survival at advanced age in good health.

The above findings reprensent different aspects of the aging process. Due to the multidimensional aspect of such phenomenon it is crucial look at them not separately, but in a global context.

5. REFERENCES

Agresti A. (2000) Categorical Data Analysis, Wiley, New York.

Akey J, Jin L, Xiong M. (2001) Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 9: 91-300.

Altomare K, Greco V, Bellizzi D, Berardelli M, Dato S, DeRango F, Garasto S, Rose G, Feraco E, Mari V, Passarino G, Franceschi C, De Benedictis G. (2003) The allele (A)(-110) in the promoter region of the HSP70-1 gene is unfavorable to longevity in women. Biogerontology 4: 215-220.

Arantes-Oliveira N, Berman JR, Kenyon C. (2003) Healthy animals with extreme longevity. Science 302: 611.

Armstrong, CM, Kaeberlein M, Imai SI, Guarente L. (2002) Mutations in Saccharomyces cerevisiae gene SIR2 can have differential effects on in vivo silencing phenotypes and in vitro histone deacetylation activity. Mol Biol Cell 13: 1427-1438.

Atzmon G, Rincon M, Schechter CB, Shuldiner AR, Lipton RB, Bergman A, Barzilai N (2006) Lipoprotein genotype and conserved pathway for exceptional longevity in humans. PLoS Biol 4:e113.

Bartali B, Frongillo EA, Bandinelli S, Lauretani F, Semba RD, Fried LP, Ferrucci L. (2006) Low nutrient intake is an essential component of frailty in older persons. J Gerontol A Biol Sci Med Sci 61: 589-593.

Bartke A, Chandrashekar V, Dominici F, Turyn D, Kinney B, Steger R, Kopchick JJ. (2003) Insulin-like growth factor 1 (IGF-1) and aging: controversies and new insights. Biogerontology 4: 1-8.

Bathum L, Christiansen L, Tan Q, Vaupel J, Jeune B, Christensen K. (2005) No evidence for an association between extreme longevity and microsomal transfer protein polymorphisms in a longitudinal study of 1651 nonagenarians. Eur J Hum Genet 13:1154-1158.

Bellizzi D, Rose G, Cavalcante P, Covello G, Dato S, De Rango F, Greco V, Maggiolini M, Feraco E, Mari V, Franceschi C, Passarino G, De Benedictis G. (2005) A novel VNTR enhancer within the SIRT3 gene, a human homologue of SIR2, is associated with survival at oldest ages. Genomics 85: 258-263.

Bohringer S, Hardt C, Miterski B, Steland A, Epplen JT. (2003) Multilocus statistics to uncover epistasis and heterogeneity in complex diseases: revisiting a set of multiple sclerosis data Eur J Hum Genet 11: 573-584.

Bonafe M, Cardelli M, Marchegiani F, Cavallone L, Giovagnetti S, Olivieri F, Lisa R, Pieri C, Franceschi C. (2001) Increase of homozygosity in centenarians revealed by a new inter-Alu PCR technique. Exp Gerontol 36: 1063-1073.

Bortz WM. (2002) A conceptual framework of frailty: a review. J Gerontol A Biol Sci Med Sci 57: M283-M288.

Brooks A, Lithgow GJ, Johnson TE. (1994) Mortality rates in a genetically heterogeneous population of Caenorhabditis elegans. Science 263: 668–671.

Brown-Borg HM, Borg KE, Meliska CJ, Bartke A. (1996) Dwarf mice and the aging process. Nature 384: 33–36.

Browner WS, Kahn AJ, Ziv E, Reiner AP, Oshima J, Cawthon RM, Hsueh WC, Cummings SR. (2004) The genetics of human longevity. Am J Med 117: 851-860

Brunet, G., P. Darlu, G. Zei, eds. 2001. Le patronyme. Histoire, anthropologie, société. Paris: CNRS Editions.

Carey JR, Liedo P, Orozco D, Vaupel JW. (1992) Slowing of mortality rates at older ages in large medfly cohorts. Science 258: 457–461.

Cardelli M, Marchegiani F, Cavallone L, Olivieri F, Giovagnetti S, Mugianesi E, Moresi R, Lisa R, Franceschi C. (2006) A polymorphism of the YTHDF2 gene (1p35) located in an Alu-rich genomic domain is associated with human longevity. J Gerontol A Biol Sci Med Sci 61: 547-556.

Carrieri G, Bonafe M, De Luca M, Rose G, Varcasia O, Bruni A, Maletta R, Nacmias B, Sorbi S, Corsonello F, Feraco E, Andreev KF, Yashin AI, Franceschi C, De Benedictis G. (2001) Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. Hum Genet 108:194-198.

Cavalli-Sforza LL. (2001) Pourqoi les patronymes? In: Le Patronyme – Historie, anthropologie, société. Brunett G, Darlu P, and Zei G, eds. pp 407-418. Paris: CNRS Editions.

Cavalli Sforza LL, Moroni A, Zei, G. (2004) Consanguinity, Inbreeding, and Genetic Drift in Italy. Princeton University Press, Princeton NJ (MPB:39).

Christensen K, Johnson TE, Vaupel JW. (2006) The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet 7: 436-448.

Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafen E, Leevers SJ, Partridge L. (2001) Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. Science 292: 104-106.

Clayton D. (1999) A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission Am J Hum Genet 65: 1170-1177

Cohen HJ, Pieper CF, Harris T, RaoKM and Currie MS. (1997) The association of plasma IL-6 levels with functional disability in community-dwelling elderly. J Gerontol A Biol Sci Med Sci 52: M201-M208.

Cordell HJ. (2002) Epistasis: what it means what it doesn't mean and statistical methods to detect it in humans Hum Mol Genet 11: 2463-2468

Cohen HJ, Harris T, and Pieper CF. (2003) Coagulation and activation of inflammatory pathways in the development of functional decline and mortality in the elderly. Am J Med 114: 180-187.

Corti MC, Guralnik JM, Salive ME and Sorkin, JD. (1994) Serum albumin level and physical disability as predictors of mortality in older persons. JAMA 272: 1036-1042.

Christensen K, Johnson TE, Vaupel JW. (2006) The quest for genetic determinants of human longevity: challenges and insights. Nat Rev Genet 7: 436-448.

Dato S, Carotenuto L, De Benedictis G. (2006) Genes and longevity: a genetic-demographic approach reveals sex and age specific gene effects not shown by the case-control approach (APOE and HSP70.1 loci). Biogerontology, in press.

De Benedictis G, Falcone E, Rose G, Ruffolo R, Spadafora P, Baggio G, Bertolini S, Mari D, Mattace R, Monti D, Morellini M, Sansoni P, Franceschi C. (1997) DNA multiallelic systems reveal gene/longevity associations not detected by diallelic systems: The APOB locus. Hum Genet 99: 312-318.

De Benedictis G, Carotenuto L, Carrieri G, De Luca M, Falcone E, Rose G, Cavalcanti S, Corsonello F, Feraco E, Baggio G, Bertolini S, Mari D, Mattace R, Yashin AI, Bonafe M, Franceschi C. (1998) Gene/longevity association studies at four autosomal loci (REN, THO, PARP, SOD2). Eur J Hum Genet 6: 534-541.

De Benedictis G, Rose G, Carrieri G, De Luca M, Falcone E, Passarino G, Bonafe M, Monti D, Baggio G, Bertolini S, Mari D, Mattace R, Franceschi C. (1999) Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. FASEB J 13: 1532-1536.

De Benedictis G, Tan Q, Jeune B, Christensen K, Ukraintseva SV, Bonafè M, Franceschi C, Vaupel JW, Yashin AI. (2001) Recent advances in human gene-longevity association studies. Mech Aging Dev 1: 909-920.

De Benedictis G and Franceschi C. (2006) The unusual genetics of human longevity. Sci Aging Knowledge Environ. 10:pe20.

Du RF, Yuan YD, Hwang J, Mountain J, Cavalli-Sforza LL. (1992) Chinese Surnames and the Genetic Difference between North and South China. Journal of Chinese Linguistics. Monograph Series 5.

Epstein MP and Satten GA. (2003) Inference on haplotype effects in case-control studies using unphased genotype data Am J Hum Genet 73: 1316-1329

Excoffier L and Slatkin M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population Mol Biol Evol 12: 921-927.

Finkel D, Whitfield K, McGue M. (1995) Genetic and environmental influences on functional age: a twin study. J Gerontol B Psychol Sci Soc Sci 50: P104–P113.

Fisher RA, Corbet AS and Williams CB. (1943) The relation between the number of species and the number of individuals in a random sample of animal population. J Anim Ecol 12: 42–58.

Folstein MF, Folstein SE, McHugh PR. (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12: 189-198.

Franceschi C, Valensin S, Bonafe M, Paolisso G, Yashin AI, Monti D, De Benedictis G. (2000) The network and the remodeling theories of aging: historical background and new perspectives. Exp Gerontol 35: 879-896.

Frederiksen H, McGue M, Jeune B, Gaist D, Nybo H, Skytthe AE, Vaupel JW, Christensen K. (2002) Do children of long-lived parents age more successfully? Epidemiology 13: 334-339.

Fried LP and Walston J.(1998) Frailty and failure to thrive. In: Hazzard WR, Blass JP, Ettinger WH Jr, Halter JB, Ouslander J, eds. Principles of Geriatric Medicine and Gerontology. 4th ed. New York: McGraw Hill: 1387–1402.

Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA. (2001) Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 56: M146-M156.

Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. (2004) Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. J Gerontol A Biol Sci Med Sci 59:255-263.

Garasto S, Rose G, DeRango F, Berardelli M, Corsonello A, Feraco E, Mari V, Maletta R, Bruni A, Franceschi C, Carotenuto L, De Benedictis G. (2003) The study of APOA1, APOC3 and APOA4 variability in healthy aging people reveals another paradox in the oldest old subjects. Ann Hum Genet 67:54-62.

Garasto S, Berardelli M, DeRango F, Mari V, Feraco E, De Benedictis G. (2004) A study of the average effect of the 3'APOB-VNTR polymorphism on lipidemic parameters could explain why the short alleles (<35 repeats) are rare in centenarians BMC Med Genet 5:3.

Gerdes LU, Jeune B, Ranberg KA, Nybo H, Vaupel JW. (2000) Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene is a "frailty gene," not a "longevity gene". Genet Epidemiol 19: 202-210.

Gompertz B. (1825) On the nature of the function expressive of the law of human motality, and on a new mode of determining the value of life contingencies. Phil Trans Roy Soc 115: 513-585.

Gottlieb K, ed. 1983. Surnames as markers of inbreeding and migration. Hum Biol 55 (2).

Guarente L and Kenyon C. (2000) Genetic pathways that regulate aging in model organisms. Nature 408: 255-262.

Herskind AM, McGue M, Holm NV, Sørensen TI, Harvald B and Vaupel JW. (1996) The heritability of human longevity, a population-based study of 2872 Danish twin pairs born 1870-1900. Hum Genet 97: 319-323.

Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, Kaprio J, Pedersen NL, Christensen K. (2006) Genetic influence on human lifespan and longevity. Hum Genet 119:312-321.

Hoffman C, Rice D, Sung HY. (1996) Persons with chronic conditions: their prevalence and costs. JAMA 276: 1473–1479.

Hoh J and Ott J. (2003) Mathematical multi-locus approaches to localizing complex human trait genes. Nat Rev Genet 4: 701-709.

Horiuchi S and Coale AJ. (1990) Age patterns of mortality for older women: an analysis using the age-specific rate of mortality change with age. Math Popul Stud 2: 245-267.

Hosmer DW Jr and Lemeshow S. (2000) Applied Logistic Regression. 2d ed. Wiley, New York.

Hsu AL, Murphy CT, Kenyon C. (2003) Regulation of aging and age-related disease by DAF-16 and heat-shock factor. Science 300: 1142-1145.

Istituto Centrale di Statistica. (1986) Sommario di Statistiche Storiche, 1926-1985 [Summary of Historical Statistics, 1926-1985] . Rome: ISTAT.

Istituto Centrale di Statistica. (1974) XI Censimento generale della popolazione 24 ottobre 1971, Vol. V, Sesso, eta` , stato civile [11th National Census of Population, 24 October 1971, 5th Vol., Sex, Age, Marital Status]. Rome: ISTAT.

Ivanova R, Henon N, Lepage V, Charron D, Vicaut E, Schachter F. (1998) HLA-DR alleles display sex-dependent effects on survival and discriminate between individual and familial longevity. Hum Mol Genet 7: 187-194.

Jeune B and Vaupel JW. (1995) Exceptional Longevity: From the Preistory to the Present. Odense Univ. Press, Odense, Denmark.

Jeune B. (2002) Living longer - but better?Aging Clin Exp Res 14: 72-93.

Jeune B, Skytthe A, Cournil A, Greco V, Gampe J, Berardelli M, Andersen-Ranberg K, Passarino G, De Benedictis G, Robine JM. (2006) Hand-grip strength among nonagenarians and centenarians in three European regions. J Gerontol Med Sci, in press.

Jian-Gang Z, Yong-Xing M, Chuan-Fu W, Pei-Fang L, Song-Bai Z, Nui-Fan G, Guo-Yin F, Lin H. (1998) Apolipoprotein E and longevity among Han Chinese population. Mech Aging Dev 104: 159-167

Johnson TE. (1990) Increased life-span of age-1 mutant in Caenorhabditis elegans and lower Gompertz rate of aging. Science 249: 908-912.

Jones DM, Song X, Rockwood K. (2004) Operationalizing a frailty index from a standardized comprehensive geriatric assessment. J Am Geriatr Soc 52: 1929-1933.

Kang JH, Logroscino G, De Vivo I, Hunter D, Grodstein F. (2005) Apolipoprotein E, cardiovascular disease and cognitive function in aging women. Neurobiol Aging 26: 475-484.

Kannisto V. (1994) *Development of Oldest-Old Mortality, 1950-1990,* Odense University Press, Odense, Denmark.

Kannisto V. (1996) Development of Oldest-Old Mortality 1950-1990. Odense University Press, Odense Denmark.

Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. (1963) Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. JAMA 21: 914-919.

Kervinen K, Savolainen MJ, Salokannel J, Hynninen A, Heikkinen J, Ehnholm C, Koistinen MJ, Kesaniemi YA. (1994) Apolipoprotein E and B polymorphisms--longevity factors assessed in nonagenarians. Atherosclerosis 105: 89-95.

Kimura K, Tissenbaum H, Liu Y, Ruvkun G. (1997) The daf-2 insulin receptor family member regulates longevity and diapause in C. elegans. Science 277: 942-946.

Kirkwood TB. (2005) Time of our lives. What controls the length of life? EMBO Rep 6: S4-S8.

Lachin JL. (2000) Statistical considerations in the intent-to-treat principle. Control Clin Trials 21: 167-189.

Lao JI, Montoriol C, Morer I, Beyer K. (2005) Genetic contribution to aging: deleterious and helpful genes define life expectancy. Ann N Y Acad Sci 1057: 50-63.

Lasker GW (1985) Surnames and Genetic Structure. Cambridge, England: Cambridge University Press.

Leng S, Chaves P, Koenig K and Walston J. (2002) Serum interleukin-6 and hemoglobin as physiological correlates in the geriatric syndrome of frailty: a pilot study. J Am Geriatr Soc 50: 1268-1271.

Leroi AM, Bartke A, De Benedictis G, Franceschi C, Gartner A, Gonos ES, Fedei ME, Kivisild T, Lee S, Kartaf-Ozer N, Schumacher M, Sikora E, Slagboom E, Tatar M, Yashin AI, Vijg J, Zwaan B. (2005) What evidence is there for the existence of individual genes with antagonistic pleiotropic effects? Mech Aging Dev 126: 421-429.

Lin K, Dorman JB, Rodan A, Kenyon C. (1997) daf-16: An HNF-3/forkhead family member that can function to double the life-span of Caenorhabditis elegans. Science 278: 1319-1322.

Lin YJ, Seroude L, Benzer S. (1998) Extended life-span and stress resistance in the Drosophila mutant methuselah. Science 282: 943-946.

Lipsi M and Caselli G. (2002) Evoluzione della geografia della mortalità in Italia : tavole provinciali e probabilità di morte per causa. Anni: 1971-1973, 1981-1983, 1991-1993. Università degli Studi "La Sapienza", Dipartimento di Scienze Demografiche, Roma.

Marini M, Lapalombella R, Canaider S, Farina A, Monti D, De Vescovi V,Morellini M, Bellizzi D, Dato S, De Benedictis G, Passarino G, Moresi R, Tesei S, Franceschi C. (2004) Heat shock response by EBV-immortalized B-lymphocytes from centenarians and control subjects: a model to study the relevance of stress response in longevity. Exp Gerontol. 39: 83-90.

McClearn GE. (1997). Prospects for quantitative trait locus methodology in Gerontology. Exp Gerontol 32: 49-54

McClearn GE, Johansson B, Berg S, Pedersen NL, Ahern F, Petrill SA, Plomin R. (1997) Substantial genetic influence on cognitive abilities in twins 80 or more years old. Science 276: 1560-1563.

McGue M, Vaupel JW, Holm N, Harvald B. (1993) Longevity is moderately heritable in a sample of Danish twins born 1870-1880. J Gerontol 48: B237–B244.

Meloni R, Albanese V, Ravassard P, Treilhou F, Mallet J. (1998) A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. Hum Mol Genet 7: 423-428.

Meloni R, Biguet NF, Mallet J. (2002) Post-genomic era and gene discovery for psychiatric diseases: there is a new art of the trade? The example of the HUMTH01 microsatellite in the Tyrosine Hydroxylase gene. Mol Neurobiol 26: 389-403.

Melzer D, Dik MG, van Kamp GJ, Jonker C, Deeg DJ. (2005) The apolipoprotein E e4 polymorphism is strongly associated with poor mobility performance test results but not self-reported limitation in older people. J Gerontol A Biol Sci Med Sci 60: 1319-1323.

Mitnitski AB, Graham JE, Mogilner AJ, Rockwood K. (2002) Frailty, fitness and late-life mortality in relation to chronological and biological age. BMC Geriatr 27: 1.

Miyata M and Smith JD. (1996) Apolipoprotein E allele-specific antioxidant activity and effects oncytotoxicity by oxidative insults and beta-amyloid peptides. Nat Genet 14: 55-61

Murakami S, Otsuka K, Hotta N, Yamanaka G, Kubo Y, Matsuoka O, Yamanaka T, Shinagawa M, Nunoda S, Nishimura Y, Shibata K, Takasugi E, Nishinaga M, Ishine M, Wada T, Okumiya K, Matsubayashi K, Yano S, Ichihara K, Cornelissen G, Halberg F. (2005) Common carotid intima-media thickness is predictive of all-cause and cardiovascular mortality in elderly community-dwelling people: Longitudinal Investigation for the Longevity and Aging in Hokkaido County (LILAC) study. Biomed Pharmacother 59 Suppl 1:S49-S53.

Murphy CT, Mc Carroll SA, Bargmann CI., Fraser A, Kamath RS, Ahringer J, Li H, Kenyon C. (2003) Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. Nature 17: 277-283.

Nagerlkerke NJD. (1991) A note on a general definition of the coefficient of the determination. Biometrika 78: 691-692.

Passarino G, Calignano C, Vallone A, Franceschi C, Jeune B, Robine JM, Yashin AI, Cavalli Sforza LL, De Benedictis G. (2002) Male/female ratio in centenarians: a possible role played by population genetic structure. Exp Gerontol 37: 1283-1289.

Passarino G, Cavalleri GL, Stecconi R, Franceschi C, Altomare K, Dato S, Greco V, Cavalli Sforza LL, Underhill PA, De Benedictis G. (2003) Molecular variation of human HSP90alpha and HSP90beta genes in Caucasians. Hum Mutat 21: 554-555.

Peel N, Bartlett H and McClure R. (2004) Healthy aging: how is it defined and measured? Aust J Aging 23: 115-119.

Pepe G, Di Perna V, Resta F, Lovecchio M, Chimienti G, Colacicco AM, Capurso A. (1998) In search of a biological pattern for human longevity: impact of apo A-IV genetic polymorphisms on lipoproteins and the hyper-Lp(a) in centenarians. Atherosclerosis 137: 407-417.

Perls T, Kunkel L, Puca A. (2002) The genetics of aging. Curr Opin Genet Dev 12: 362-369.

Pope A and Tarlov A. (1991) Disability in America: Toward a National Agenda for Prevention. Institute of Medicine (U.S.) Committee on a National Agenda for the Prevention of Disabilities. Washington, DC: National Academy Press.

Poulain M, Pes GM, Grasland C, Carru C, Ferrucci L, Baggio G, Franceschi C, Deiana L. (2004) Identification of a geographic area characterized by extreme longevity in the Sardinia island: the AKEA study. Exp Gerontol 39: 1423-1429.

Ranieri P, Rozzini R, Franzoni S, Barbisoni P and Trabucchi M. (1998) Serum cholesterol levels as a measure of frailty in elderly patients. Exp Aging Res 24: 169-179.

Relton CL, Wilding CS, Pearce MS, Laffling AJ, Jonas PA, Lynch SA, Tawn EJ, Burn J. (2004) Gene-gene interaction in folate-related genes and risk of neural tube defects in a UK population. J Med Genet 41: 256-260.

Robine JM, Paccaud F. (2005) Nonagenarians and centenarians in Switzerland, 1860-2001: a demographic analysis. J Epidemiol Community Health 59: 31-37.

Robine JM, Caselli G, Rasulo D, Cournil A. (2006) Differentials in the femininity ratio among centenarians: variations between northern and southern Italy from 1870. Popul Stud (Camb) 60: 99-113.

Rockwood K. (2005) What would make a definition of frailty successful? Age Aging 34: 432-434

Rose MR. (1984) Laboratory selection of postponed senescence in Drosophila melanogaster. Evolution 38: 1004–1010.

Rowe JW and Kahn RL. (1987) Human aging: usual and successful aging. Science 237: 143-149.

Rowe JW and Kahn RL. (1998) Successful aging. New York: Pantheon Books.

Royston P and Altman D. (1994) Regression using fractional polynomials of continuous: Parsimonious parametric modeling. Applied Statistics 43: 429-467.

Ruiz-Pesini E, Lapena AC, Diez-Sanchez C, Perez-Martos A, Montoya J, Alvarez E, Diaz M, Urries A, Montoro L, Lopez-Perez MJ, Enriquez JA. (2000) Human mtDNA haplogroups associated with high or reduced spermatozoa motility. Am J Hum Genet 67: 682-696

Schachter F, Faure-Delaneff L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L, Cohen D. (1994) Genetic associations with human longevity at the APOE and ACE loci. Nat Genet 6: 29-32.

Schaid DJ Rowland CM Tines DE Jacobson RM Poland GA. (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous Am J Hum Genet 70: 425-434

Schork NJ, Fallin D, Lanchbury JS. (2000) Single nucleotide polymorphisms and the future of genetic epidemiology Clin Genet 58: 250-264

Sgrò C and Partridge L. (1999) A delayed wave of death from reproduction in Drosophila. Science 286: 2521-2524.

Sheikh JI and Yesavage JA. (1986) Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontology : A Guide to Assessment and Intervention 165-173, NY: The Haworth Press.

Tan Q, De Benedictis G, Yashi AI, Bonafe M, De Luca M, Valensin S, Vaupel JW, Franceschi C. (2001) Measuring the genetic influence in modulating the human life span: gene-environment interaction and the sex-specific genetic effect. Biogerontology 2: 141-153.

Tan Q, Bathum L, Christiansen L, De Benedictis G, Dahlgaard J, Frizner N, Vach W, Vaupel JW, Yashin AI, Christensen K, Kruse TA. (2003) Logistic regression models for polymorphic and antagonistic pleiotropic gene action on human aging and longevity. Ann Hum Genet 67: 598-607.

Tan Q, Christiansen L, Bathum L, Zhao JH, Yashin AI, Vaupel JW, Christensen K, Kruse TA. (2005) Estimating haplotype relative risks on human survival in population-based association studies. Hum Hered 59: 88-97.

Tan Q, Christiansen L, Bathum L, Li S, Kruse TA, Christensen K. (2006) Genetic association analysis of human survival in cohort studies of elderly subjects: an example of PON1 gene in the Danish 1905 birth cohort. Genetics 172: 1821-1828.

Terry DF, Wilcox M, McCormick MA, Lawler E, Perls TT. (2003) Cardiovascular advantages among the offspring of centenarians. J Gerontol A Biol Sci Med Sci 58: M425-M431.

Terry DF, Wilcox MA, McCormick MA, Pennington JY, Schoenhofen EA, Andersen SL, Perls TT. (2004) Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. J Am Geriatr Soc 52: 2074-2076.

Terwilliger JD and Ott J. (1992) A haplotype-based “haplotype relative risk” approach to detecting allelic associations Hum Hered 42: 337-346

Thatcher AR, Kannisto V, Vaupel JW. (1996) The force of mortality from age 80 to 120. University Press. Odense, Denmark.

Ukraintseva SV and Yashin AI. (2001) How individual age-associated changes may influence human morbidity and mortality patterns. Mech Aging Dev 122: 1447-1460.

Vaupel JW, Manton KG, Stallard E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16: 439–454.

Vaupel JW and Yashin AI. (1985) Heterogeneity’s ruses: some surprising effects of selection on population dynamics. Am Stat 39: 176–195.

Vaupel JW and Jeune B. (1995) The emergence and proliferation of centenarians. In Jeune B and Vaupel JW (Eds.) *Exceptional longevity: from prehistory to the present*. Odense University Press, Odense.

Vaupel JW. (1997) The remarkable improvements in survival at older ages. Philos Trans R Soc London Ser B 352: 1799-1804.

Vaupel, JW, Carey JR, Christensen K, Johnson TE, Yashin AI, Holm NV, Iachine IA, Kannisto V, Khazaeli AA, Liedo P, Longo VD, Zeng Y, Manton KG, Curtsinger JW. (1998) Biodemographic trajectories of longevity. Science 280: 855-60.

Wagner EH, Austin BT, Von Korff M. (1996) Organizing care for patients with chronic illness. Milbank Q 74:511-544.

Walston J, McBurnie MA, Newman A, Tracy RP, Kop W J, Hirsch CH, Gottdiener J, and Fried LP. (2002) Frailty and activation of the inflammation and coagulation systems with and without clinical comorbidities: results from the Cardiovascular Health Study. Arch Intern Med 162: 2333-2341.

Walston J, Hadley EC, Ferrucci L, Guralnik JM, Newman AB, Studenski SA, Ershler WB, Harris T, Fried LP. (2006) Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. J Am Geriatr Soc 54: 991-1001.

Ward JH. (1963) Hierarchical Grouping to Optimize an Objective Function. J Am Stat Assoc 58: 236-244.

Williams GC. (1957) Pleiotropy, natural selection, and the evolution of senescence. Evolution 11: 398-411.

Yashin AI, De Benedictis G, Vaupel JW, Tan Q, Andreev KF, Iachine IA, Bonafe M, DeLuca M, Valensin S, Carotenuto L, Franceschi C. (1999) Genes, demography, and life span: the contribution of demographic data in genetic studies on aging and longevity. Am J Hum Genet 65: 1178-1193.

Zei G, Matessi RG, Siri E, Moroni A, Cavalli-Sforza LL. (1983) Surnames in Sardinia. I. Fit of frequency distributions for neutral alleles and genetic population structure. Ann Hum Genet 47: 329-352.

Zei G, Piazza A, Moroni A, Cavalli-Sforza LL. (1986) Surnames in Sardinia. III. The spatial distribution of surnames for testing neutrality of genes. Ann Hum Genet 50: 169-180.

7. SUPPLEMENTARY MATERIAL

Table 1a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of APOA1-MspI-RelLP 628 A>G polymorphism in the three age groups in females.

	Age Group							
Genotype*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
PP	198	0.600 ± 0.027	69	0.663 ± 0.046	53	0.609 ± 0.052	320	0.614 ± 0.021
AP	114	0.345 ± 0.026	31	0.298 ± 0.045	31	0.356 ± 0.051	176	0.338 ± 0.021
AA	18	0.055 ± 0.013	4	0.038 ± 0.019	3	0.034 ± 0.020	25	0.048 ± 0.009

*A/P denotes Absence/Presence of the restriction site.

Table 1b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of APOA1-MspI-RelLP 628 A>G polymorphism in the three age groups in females

	Age Group							
Allele*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
P	510	0.773 ± 0.016	169	0.813 ± 0.027	137	0.787 ± 0.031	816	0.783 ± 0.013
A	150	0.227 ± 0.016	39	0.188 ± 0.027	37	0.213 ± 0.031	226	0.217 ± 0.013

*A/P denotes Absence/Presence of the restriction site.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0. 090, df = 1, p = 0. 7642
- Group 2: Chi-square = 0. 049, df = 1, p = 0. 8248
- Group 3: Chi-square = 0. 358, df = 1, p = 0. 5496

Table 2a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of APOA1-MspI-RelLP 628 A>G polymorphism in the three age groups in males.

	Age Group							
Genotype*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
PP	105	0.475 ± 0.034	74	0.617 ± 0.040	75	0.682 ± 0.044	254	0.563 ± 0.023
AP	97	0.439 ± 0.033	44	0.367 ± 0.040	30	0.273 ± 0.042	171	0.379 ± 0.023
AA	19	0.086 ± 0.019	2	0.017 ± 0.012	5	0.045 ± 0.020	26	0.058 ± 0.011

*A/P denotes Absence/Presence of the restriction site.

Table 2b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of APOA1-MspI-RelLP 628 A>G polymorphism in the three age groups in females

	Age Group							
Allele*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
P	307	0.695 ± 0.022	192	0.800 ± 0.026	180	0.818 ± 0.026	679	0.753 ± 0.014
A	135	0.305 ± 0.022	48	0.200 ± 0.026	40	0.182 ± 0.026	223	0.247 ± 0.014

*A/P denotes Absence/Presence of the restriction site.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0.263, df = 1, p = 06081
- Group 2: Chi-square = 2.552, df = 1, p = 0.1102
- Group 3: Chi-square = 0.764, df = 1, p = 0. 3821

Table 3a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of APOA4-HincII-RelLP 1033 A>G polymorphism for three age groups in females.

	Age Group							
Genotype*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
PP	241	0.730 ± 0.024	84	0.808 ± 0.039	63	0.724 ± 0.048	388	0.745 ± 0.019
AP	84	0.255 ± 0.024	18	0.173 ± 0.037	23	0.264 ± 0.047	125	0.240 ± 0.019
AA	5	0.015 ± 0.007	2	0.019 ± 0.013	1	0.011 ± 0.011	8	0.015 ± 0.005

*A/P denotes Absence/Presence of the restriction site.

Table 3b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of APOA4-HincII-RelLP 1033 A>G polymorphism for three age groups in females.

	Age Group							
Allele*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
P	566	0.858 ± 0.014	186	0.894 ± 0.021	149	0.856 ± 0.027	901	0.865 ± 0.011
A	94	0.142 ± 0.014	22	0.106 ± 0.021	25	0.144 ± 0.027	141	0.135 ± 0.011

*A/P denotes Absence/Presence of the restriction site.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0.583, df = 1, p = 0.4451
- Group 2: Chi-square = 0.752, df = 1, p = 0.3858
- Group 3: Chi-square = 0.481, df = 1, p = 0.4880

Table 4a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of APOA4-HincII-RelLP 1033 A>G polymorphism for three age groups in males.

	Age Group							
Genotype*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
PP	158	0.715 ± 0.030	88	0.733 ± 0.040	42	0.382 ± 0.046	288	0.639 ± 0.023
AP	61	0.276 ± 0.030	30	0.250 ± 0.040	53	0.482 ± 0.048	144	0.319 ± 0.022
AA	2	0.009 ± 0.006	2	0.017 ± 0.012	15	0.136 ± 0.033	19	0.042 ± 0.009

*A/P denotes Absence/Presence of the restriction site.

Table 4b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of APOA4-HincII-RelLP 1033 A>G polymorphism for three age groups in males.

	Age Group							
Allele*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
P	377	0.853 ± 0.017	206	0.858 ± 0.023	137	0.623 ± 0.033	720	0.798 ± 0.013
A	65	0.147 ± 0.017	34	0.142 ± 0.023	83	0.377 ± 0.033	182	0.202 ± 0.013

*A/P denotes Absence/Presence of the restriction site.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square =2.222, df = 1, p = 0.1361
- Group 2: Chi-square = 0. 094, df = 1, p = 0.7592
- Group 3: Chi-square = 0. 071, df = 1, p = 0.7899

Table 5a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of 3'APOB-VNTR polymorphism for three age groups in females.

	Age Group							
Genotype*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
SS	15	0.045 ± 0.011	4	0.038 ± 0.019	1	0.011 ± 0.011	20	0.038 ± 0.008
SM	83	0.252 ± 0.024	25	0.240 ± 0.042	18	0.207 ± 0.043	126	0.242 ± 0.019
SL	18	0.055 ± 0.013	6	0.058 ± 0.023	6	0.069 ± 0.027	30	0.058 ± 0.010
MM	135	0.409 ± 0.027	50	0.481 ± 0.049	39	0.448 ± 0.053	224	0.430 ± 0.022
ML	67	0.203 ± 0.022	16	0.154 ± 0.035	21	0.241 ± 0.046	104	0.200 ± 0.018
LL	12	0.036 ± 0.010	3	0.029 ± 0.016	2	0.023 ± 0.016	17	0.033 ± 0.008

*Alleles: S < 35 repeats; M 35–39 repeats; L > 39 repeats.

Table 5b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of 3'APOB-VNTR polymorphism for three age groups in females.

	Age Group							
Allele*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
S	131	0.198 ± 0.016	39	0.188 ± 0.027	26	0.149 ± 0.027	196	0.188 ± 0.012
M	420	0.636 ± 0.019	141	0.678 ± 0.032	117	0.672 ± 0.036	678	0.651 ± 0.015
L	109	0.165 ± 0.014	28	0.135 ± 0.024	31	0.178 ± 0.029	168	0.161 ± 0.011

*Alleles: S < 35 repeats; M 35–39 repeats; L > 39 repeats.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 2.014, df = 3, p = 0.5695
- Group 2: Chi-square =1.448, df = 3, p = 0.6943
- Group 3: Chi-square =1.091, df = 3, p = 0.7792

Table 6a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of 3'APOB-VNTR polymorphism for three age groups in males.

	Age Group							
Genotype*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
SS	11	0.050 ± 0.015	3	0.025 ± 0.014	7	0.064 ± 0.023	21	0.047 ± 0.010
SM	58	0.262 ± 0.030	27	0.225 ± 0.038	33	0.300 ± 0.044	118	0.262 ± 0.021
SL	8	0.036 ± 0.013	7	0.058 ± 0.021	6	0.055 ± 0.022	21	0.047 ± 0.010
MM	84	0.380 ± 0.033	47	0.392 ± 0.045	33	0.300 ± 0.044	164	0.364 ± 0.023
ML	52	0.235 ± 0.029	28	0.233 ± 0.039	28	0.255 ± 0.042	108	0.239 ± 0.020
LL	8	0.036 ± 0.013	8	0.067 ± 0.023	3	0.027 ± 0.016	19	0.042 ± 0.009

*Alleles: S < 35 repeats; M 35–39 repeats; L > 39 repeats.

Table 6b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of 3'APOB-VNTR polymorphism for three age groups in males.

	Age Group							
Allele*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
S	88	0.199 ± 0.019	40	0.167 ± 0.024	53	0.241 ± 0.029	181	0.201 ± 0.013
M	278	0.629 ± 0.023	149	0.621 ± 0.031	127	0.577 ± 0.033	554	0.614 ± 0.016
L	76	0.172 ± 0.018	51	0.213 ± 0.026	40	0.182 ± 0.026	167	0.185 ± 0.013

*Alleles: S < 35 repeats; M 35–39 repeats; L > 39 repeats.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 4.893, df = 3, p = 0.1798
- Group 2: Chi-square = 2.152, df = 3, p = 0.5415
- Group 3: Chi-square = 3.140, df = 3, p = 0.3705

Table 7a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of APOE ε2, ε3, ε4 alleles for three age groups in females.

	Age Group							
Genotype	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
ε2 ε2	2	0.006 ± 0.004	0	0.000	1	0.011 ± 0.011	3	0.006 ± 0.003
ε2 ε3	36	0.109 ± 0.017	10	0.096 ± 0.029	14	0.161 ± 0.039	60	0.115 ± 0.014
ε2 ε4	6	0.018 ± 0.007	1	0.010 ± 0.010	1	0.011 ± 0.011	8	0.015 ± 0.005
ε3 ε3	234	0.709 ± 0.025	78	0.750 ± 0.042	67	0.770 ± 0.045	379	0.727 ± 0.020
ε3 ε4	47	0.142 ± 0.019	15	0.144 ± 0.034	4	0.046 ± 0.022	66	0.127 ± 0.015
ε4 ε4	5	0.015 ± 0.007	0	0.000	0	0.000	5	0.010 ± 0.004

Table 7b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of APOE ε2, ε3, ε4 alleles for three age groups in females.

	Age Group							
Allele	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
ε2	46	0.070 ± 0.010	11	0.053 ± 0.016	17	0.098 ± 0.023	74	0.071 ± 0.008
ε3	551	0.835 ± 0.014	181	0.870 ± 0.023	152	0.874 ± 0.025	884	0.848 ± 0.011
ε4	63	0.095 ± 0.011	16	0.077 ± 0.018	5	0.029 ± 0.013	84	0.081 ± 0.008

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square =2.824, df = 3, p = 0.4196
- Group 2: Chi-square =1.044, df = 3, p = 0.7906
- Group 3: Chi-square =0.727, df = 3, p = 0.8668

Table 8a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of APOE ε2, ε3, ε4 alleles for three age groups in males.

	Age Group							
Genotype	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
ε2 ε2	0	0.000	0	0.000	0	0.000	0	0.000
ε2 ε3	26	0.118 ± 0.022	21	0.175 ± 0.035	19	0.173 ± 0.036	66	0.146 ± 0.017
ε2 ε4	9	0.041 ± 0.013	5	0.042 ± 0.018	1	0.009 ± 0.009	15	0.033 ± 0.008
ε3 ε3	145	0.656 ± 0.032	77	0.642 ± 0.044	78	0.709 ± 0.043	300	0.665 ± 0.022
ε3 ε4	40	0.181 ± 0.026	17	0.142 ± 0.032	12	0.109 ± 0.030	69	0.153 ± 0.017
ε4 ε4	1	0.005 ± 0.005	0	0.000	0	0.000	1	0.002 ± 0.002

Table 8b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of APOE ε2, ε3, ε4 alleles for three age groups in males.

	Age Group							
Allele	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
ε2	35	0.079 ± 0.013	26	0.108 ± 0.020	20	0.091 ± 0.019	81	0.090 ± 0.010
ε3	356	0.805 ± 0.019	192	0.800 ± 0.026	187	0.850 ± 0.024	735	0.815 ± 0.013
ε4	51	0.115 ± 0.015	22	0.092 ± 0.019	13	0.059 ± 0.016	86	0.095 ± 0.010

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 8.981, df = 3, p = 0.0295
- Group 2: Chi-square =5.312, df = 3, p = 0.1503
- Group 3: Chi-square = 1.665, df = 3, p = 0.6447

Table 9a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of HSP70-1 - 110A/C polymorphism for three age groups in females.

	Age Group							
Genotype	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
AA	141	0.427 ± 0.027	30	0.288 ± 0.044	28	0.322 ± 0.050	199	0.382 ± 0.021
AC	152	0.461 ± 0.027	60	0.577 ± 0.048	45	0.517 ± 0.054	257	0.493 ± 0.022
CC	37	0.112 ± 0.017	14	0.135 ± 0.033	14	0.161 ± 0.039	65	0.125 ± 0.014

Table 9b: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of HSP70-1 - 110A/C polymorphism for three age groups in females.

	Age Group							
Allele	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
A	434	0.658 ± 0.018	120	0.577 ± 0.034	101	0.580 ± 0.037	655	0.629 ± 0.015
C	226	0.342 ± 0.018	88	0.423 ± 0.034	73	0.420 ± 0.037	387	0.371 ± 0.015

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0.171, df = 1, p = 0.6792
- Group 2: Chi-square = 3.438, df = 1, p = 0.0637
- Group 3: Chi-square = 0.334, df = 1, p = 0.5633

Table 10a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of HSP70-1 -110A/C polymorphism for three age groups in males.

	Age Group							
Genotype	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
AA	84	0.380 ± 0.033	45	0.375 ± 0.044	40	0.364 ± 0.046	169	0.375 ± 0.023
AC	99	0.448 ± 0.033	62	0.517 ± 0.046	54	0.491 ± 0.048	215	0.477 ± 0.024
CC	38	0.172 ± 0.025	13	0.108 ± 0.028	16	0.145 ± 0.034	67	0.149 ± 0.017

Table 10b: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of HSP70-1 -110A/C polymorphism for three age groups in males.

	Age Group							
Allele	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
A	267	0.604 ± 0.023	152	0.633 ± 0.031	134	0.609 ± 0.033	553	0.613 ± 0.016
C	175	0.396 ± 0.023	88	0.367 ± 0.031	86	0.391 ± 0.033	349	0.387 ± 0.016

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0.891, df = 1, p = 0.3452
- Group 2: Chi-square =1.517, df = 1, p = 0.2181
- Group 3: Chi-square = 0.105, df = 1, p = 0.7459

Table 11a: Absolute (Abs) and relative(Rel) genotypic frequencies ± standard errors (SE) of HSP90-alfa-HinfI-RelLP 888 C>T polymorphism for three age groups in females.

	Age Group							
Genotype*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
PP	253	0.767 ± 0.023	75	0.721 ± 0.044	71	0.816 ± 0.042	399	0.766 ± 0.019
AP	73	0.221 ± 0.023	28	0.269 ± 0.043	15	0.172 ± 0.040	116	0.223 ± 0.018
AA	4	0.012 ± 0.006	1	0.010 ± 0.010	1	0.011 ± 0.011	6	0.012 ± 0.005

*A/P denotes Absence/Presence of the restriction site.

Table 11b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of HSP90-alfa-HinfI-RelLP 888 C>T polymorphism for three age groups in females.

	Age Group							
Allele*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
P	579	0.877 ± 0.013	178	0.856 ± 0.024	157	0.902 ± 0.023	914	0.877 ± 0.010
A	81	0.123 ± 0.013	30	0.144 ± 0.024	17	0.098 ± 0.023	128	0.123 ± 0.010

*A/P denotes Absence/Presence of the restriction site.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0.246, df = 1, p = 0.6199
- Group 2: Chi-square = 0.854, df = 1, p = 0.3544
- Group 3: Chi-square = 0.043, df = 1, p = 0.8357

Table 12a: Absolute (Abs) and relative(Rel) genotypic frequencies ± standard errors (SE) of HSP90-alfa-HinfI-RelLP 888 C>T polymorphism for three age groups in males.

	Age Group							
Genotype*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
PP	167	0.756 ± 0.029	96	0.800 ± 0.037	91	0.827 ± 0.036	354	0.785 ± 0.019
AP	53	0.240 ± 0.029	24	0.200 ± 0.037	19	0.173 ± 0.036	96	0.213 ± 0.019
AA	1	0.005 ± 0.005	0	0.000	0	0.000	1	0.002 ± 0.002

*A/P denotes Absence/Presence of the restriction site.

Table 12b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of HSP90-alfa-HinfI-RelLP 888 C>T polymorphism for three age groups in males.

	Age Group							
Allele*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
P	387	0.876 ± 0.016	216	0.900 ± 0.019	201	0.914 ± 0.019	804	0.891 ± 0.010
A	55	0.124 ± 0.016	24	0.100 ± 0.019	19	0.086 ± 0.019	98	0.109 ± 0.010

*A/P denotes Absence/Presence of the restriction site.

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square =2.236, df = 1, p = 0.1348
- Group 2: Chi-square =1.481, df = 1, p = 0.2236
- Group 3: Chi-square =0.983, df = 1, p = 0.3215

Table 13a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of HSP90-Beta 5976 A>C polymorphism for three age groups in females.

	Age Group							
Genotype	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
AA	238	0.721 ± 0.025	77	0.740 ± 0.043	74	0.851 ± 0.038	389	0.747 ± 0.019
AC	86	0.261 ± 0.024	24	0.231 ± 0.041	11	0.126 ± 0.036	121	0.232 ± 0.018
CC	6	0.018 ± 0.007	3	0.029 ± 0.016	2	0.023 ± 0.016	11	0.021 ± 0.006

Table 13b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of HSP90-Beta 5976 A>C polymorphism for three age groups in females.

	Age Group							
Allele	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
A	562	0.852 ± 0.014	178	0.856 ± 0.024	159	0.914 ± 0.021	899	0.863 ± 0.011
C	98	0.148 ± 0.014	30	0.144 ± 0.024	15	0.086 ± 0.021	143	0.137 ± 0.011

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0.309, df = 1, p = 0.5783
- Group 2: Chi-square = 0.442, df = 1, p = 0.5062
- Group 3: Chi-square =3.393, df = 1, p = 0.0655

Table 14a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of HSP90-Beta 5976 A>C polymorphism for three age groups in males.

	Age Group							
Genotype	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
AA	167	0.756 ± 0.029	92	0.767 ± 0.039	77	0.700 ± 0.044	336	0.745 ± 0.021
AC	52	0.235 ± 0.029	23	0.192 ± 0.036	33	0.300 ± 0.044	108	0.239 ± 0.020
CC	2	0.009 ± 0.006	5	0.042 ± 0.018	0	0.000	7	0.016 ± 0.006

Table 14b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of HSP90-Beta 5976 A>C polymorphism for three age groups in males.

	Age Group							
Allele	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
A	386	0.873 ± 0.016	207	0.863 ± 0.022	187	0.850 ± 0.024	780	0.865 ± 0.011
C	56	0.127 ± 0.016	33	0.138 ± 0.022	33	0.150 ± 0.024	122	0.135 ± 0.011

Test for Hardy-Weinberg equilibrium:

- Group 1: Chi-square = 0.885, df = 1, p = 0.3468
- Group 2: Chi-square =4.420, df = 1, p = 0.0355
- Group 3: Chi-square = 3.426, df = 1, p = 0.0642

Table 15a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of SIRT3 VNTR intron 5 polymorphism for three age groups in females.

	Age Group							
Genotype*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
1 1	65	0.197 ± 0.022	21	0.202 ± 0.039	19	0.218 ± 0.044	105	0.202 ± 0.018
1 2	24	0.073 ± 0.014	13	0.125 ± 0.032	8	0.092 ± 0.031	45	0.086 ± 0.012
1 3	62	0.188 ± 0.022	15	0.144 ± 0.034	14	0.161 ± 0.039	91	0.175 ± 0.017
1 4	74	0.224 ± 0.023	20	0.192 ± 0.039	20	0.230 ± 0.045	114	0.219 ± 0.018
1 5	1	0.003 ± 0.003	1	0.010 ± 0.010	0	0.000	2	0.004 ± 0.003
1 6	3	0.009 ± 0.005	1	0.010 ± 0.010	0	0.000	4	0.008 ± 0.004
1 7	0	0.000	1	0.010 ± 0.010	0	0.000	1	0.002 ± 0.002
2 2	2	0.006 ± 0.004	1	0.010 ± 0.010	3	0.034 ± 0.020	6	0.012 ± 0.005
2 3	11	0.033 ± 0.010	5	0.048 ± 0.021	2	0.023 ± 0.016	18	0.035 ± 0.008
2 4	17	0.052 ± 0.012	4	0.038 ± 0.019	1	0.011 ± 0.011	22	0.042 ± 0.009
2 5	0	0.000	1	0.010 ± 0.010	0	0.000	1	0.002 ± 0.002
2 6	0	0.000	1	0.010 ± 0.010	0	0.000	1	0.002 ± 0.002
3 3	16	0.048 ± 0.012	6	0.058 ± 0.023	5	0.057 ± 0.025	27	0.052 ± 0.010
3 4	28	0.085 ± 0.015	7	0.067 ± 0.025	7	0.080 ± 0.029	42	0.081 ± 0.012
3 5	1	0.003 ± 0.003	0	0.000	0	0.000	1	0.002 ± 0.002
3 6	0	0.000	1	0.010 ± 0.010	0	0.000	1	0.002 ± 0.002
4 4	26	0.079 ± 0.015	5	0.048 ± 0.021	6	0.069 ± 0.027	37	0.071 ± 0.001
4 5	0	0.000	0	0.000	2	0.023 ± 0.016	2	0.004 ± 0.003
4 6	0	0.000	1	0.010 ± 0.010	0	0.000	1	0.002 ± 0.002

*Allele nomenclature refers to the repeat number.

Table 15b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of SIRT3 VNTR intron 5 polymorphism for three age groups in females.

	Age Group							
Allele*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
1	294	0.445 ± 0.019	93	0.447 ± 0.034	80	0.460 ± 0.038	467	0.448 ± 0.015
2	56	0.085 ± 0.011	26	0.125 ± 0.023	17	0.098 ± 0.023	99	0.095 ± 0.009
3	134	0.203 ± 0.016	40	0.192 ± 0.027	33	0.190 ± 0.030	207	0.199 ± 0.012
4	171	0.259 ± 0.017	42	0.202 ± 0.028	42	0.241 ± 0.032	255	0.245 ± 0.013
5	2	0.003 ± 0.002	2	0.010 ± 0.007	2	0.011 ± 0.008	6	0.006 ± 0.002
6	3	0.005 ± 0.003	4	0.019 ± 0.010	0	0.000	7	0.007 ± 0.003
7	0	0.000	1	0.005 ± 0.005	0	0.000	1	0.001 ± 0.001

*Allele nomenclature refers to the repeat number.

Test for Hardy-Weinberg equilibrium (5000 permutations):

- Group 1: p = 0.6672
- Group 2: p = 0.9616
- Group 3: p = 0.1170

Table 16a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of SIRT3 VNTR intron 5 polymorphism for three age groups in males.

	Age Group							
Genotype*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
1 1	33	0.149 ± 0.024	16	0.133 ± 0.031	18	0.164 ± 0.035	67	0.149 ± 0.017
1 2	27	0.122 ± 0.022	12	0.100 ± 0.027	7	0.064 ± 0.023	46	0.102 ± 0.014
1 3	36	0.163 ± 0.025	12	0.100 ± 0.027	19	0.173 ± 0.036	67	0.149 ± 0.017
1 4	48	0.217 ± 0.028	28	0.233 ± 0.039	27	0.245 ± 0.041	103	0.228 ± 0.020
1 5	0	0.000	1	0.008 ± 0.008	1	0.009 ± 0.009	2	0.004 ± 0.003
2 2	5	0.023 ± 0.010	3	0.025 ± 0.014	0	0.000	8	0.018 ± 0.006
2 3	8	0.036 ± 0.013	5	0.042 ± 0.018	0	0.000	13	0.029 ± 0.008
2 4	14	0.063 ± 0.016	8	0.067 ± 0.023	4	0.036 ± 0.018	26	0.058 ± 0.011
2 5	1	0.005 ± 0.005	0	0.000	0	0.000	1	0.002 ± 0.002
3 3	15	0.068 ± 0.017	7	0.058 ± 0.021	5	0.045 ± 0.020	27	0.060 ± 0.011
3 4	19	0.086 ± 0.019	19	0.158 ± 0.033	14	0.127 ± 0.032	52	0.115 ± 0.015
3 5	1	0.005 ± 0.005	0	0.000	2	0.018 ± 0.013	3	0.007 ± 0.004
4 4	13	0.059 ± 0.016	9	0.075 ± 0.024	11	0.100 ± 0.029	33	0.073 ± 0.012
4 5	1	0.005 ± 0.005	0	0.000	1	0.009 ± 0.009	2	0.004 ± 0.003
4 6	0	0.000	0	0.000	1	0.009 ± 0.009	1	0.002 ± 0.002

*Allele nomenclature refers to the repeat number.

Table 16b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of SIRT3 VNTR intron 5 polymorphism for three age groups in males.

	Age Group							
Allele*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
1	177	0.400 ± 0.023	85	0.354 ± 0.031	90	0.409 ± 0.033	352	0.390 ± 0.016
2	60	0.136 ± 0.016	31	0.129 ± 0.022	11	0.050 ± 0.015	102	0.113 ± 0.011
3	94	0.213 ± 0.019	50	0.208 ± 0.026	45	0.205 ± 0.027	189	0.210 ± 0.014
4	108	0.244 ± 0.020	73	0.304 ± 0.030	69	0.314 ± 0.031	250	0.277 ± 0.015
5	3	0.007 ± 0.004	1	0.004 ± 0.004	4	0.018 ± 0.009	8	0.009 ± 0.003
6	0	0.000	0	0.000	1	0.005 ± 0.005	1	0.001 ± 0.001

*Allele nomenclature refers to the repeat number.

Test for Hardy-Weinberg equilibrium (5000 permutations):

- Group 1: p = 0.4848
- Group 2: p = 0.5512
- Group 3: p = 0.4850

Table 17a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of human tyrosine (HUMTHO.1) STR polymorphism for three age groups in females.

	Age Group							
Genotype	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
6 6	24	0.073 ± 0.014	6	0.058 ± 0.023	8	0.092 ± 0.031	38	0.073 ± 0.011
6 7	29	0.088 ± 0.016	11	0.106 ± 0.030	6	0.069 ± 0.027	46	0.088 ± 0.012
6 8	35	0.106 ± 0.017	3	0.029 ± 0.016	9	0.103 ± 0.033	47	0.090 ± 0.013
6 9	43	0.130 ± 0.019	16	0.154 ± 0.035	10	0.115 ± 0.034	69	0.132 ± 0.015
6 10	38	0.115 ± 0.018	13	0.125 ± 0.032	12	0.138 ± 0.037	63	0.121 ± 0.014
6 10*	3	0.009 ± 0.005	0	0.000	0	0.000	3	0.006 ± 0.003
7 7	7	0.021 ± 0.008	2	0.019 ± 0.013	3	0.034 ± 0.020	12	0.023 ± 0.007
7 8	19	0.058 ± 0.013	4	0.038 ± 0.019	3	0.034 ± 0.020	26	0.050 ± 0.010
7 9	14	0.042 ± 0.011	10	0.096 ± 0.029	7	0.080 ± 0.029	31	0.060 ± 0.010
7 10	26	0.079 ± 0.015	5	0.048 ± 0.021	3	0.034 ± 0.020	34	0.065 ± 0.011
7 10*	0	0.000	0	0.000	1	0.011 ± 0.011	1	0.002 ± 0.002
8 8	5	0.015 ± 0.007	1	0.010 ± 0.010	0	0.000	6	0.012 ± 0.005
8 9	16	0.048 ± 0.012	4	0.038 ± 0.019	4	0.046 ± 0.022	24	0.046 ± 0.009
8 10	18	0.055 ± 0.013	8	0.077 ± 0.026	3	0.034 ± 0.020	29	0.056 ± 0.010
8 10*	2	0.006 ± 0.004	0	0.000	2	0.023 ± 0.016	4	0.008 ± 0.004
9 9	9	0.027 ± 0.009	2	0.019 ± 0.013	1	0.011 ± 0.011	12	0.023 ± 0.007
9 10	29	0.088 ± 0.016	15	0.144 ± 0.034	10	0.115 ± 0.034	54	0.104 ± 0.013
9 10*	1	0.003 ± 0.003	0	0.000	1	0.011 ± 0.011	2	0.004 ± 0.003
10 10	11	0.033 ± 0.010	4	0.038 ± 0.019	3	0.034 ± 0.020	18	0.035 ± 0.008
10 10*	1	0.003 ± 0.003	0	0.000	0	0.000	1	0.002 ± 0.002
10* 10*	0	0.000	0	0.000	1	0.011 ± 0.011	1	0.002 ± 0.002

Table 17b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of human tyrosine (HUMTHO.1) STR polymorphism for three age groups in females.

	Age Group							
Allele**	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
6	196	0.297 ± 0.018	55	0.264 ± 0.031	53	0.305 ± 0.035	304	0.292 ± 0.014
7	102	0.155 ± 0.014	34	0.163 ± 0.026	26	0.149 ± 0.027	162	0.155 ± 0.011
8	100	0.152 ± 0.014	21	0.101 ± 0.021	21	0.121 ± 0.025	142	0.136 ± 0.011
9	121	0.183 ± 0.015	49	0.236 ± 0.029	34	0.195 ± 0.030	204	0.196± 0.012
10	134	0.203 ± 0.016	49	0.236 ± 0.029	34	0.195 ± 0.030	217	0.208 ± 0.013
10*	7	0.011 ± 0.004	0	0.000	6	0.034 ± 0.014	13	0.012 ± 0.003

** TH-STR alleles are indicated according to the number of the repeats. 10* is the imperfect HUMTHO.1 allele 10 (Puers et al., 1993)

Test for Hardy-Weinberg equilibrium (5000 permutations):

- Group 1: p = 0.5712
- Group 2: p = 0.4128
- Group 3: p = 0.1856

Table 18a: Absolute (Abs) and relative (Rel) genotypic frequencies ± standard errors (SE) of human tyrosine (HUMTHO.1) STR polymorphism for three age groups in males.

	Age Group							
Genotype	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
6 6	17	0.077 ± 0.018	18	0.150 ± 0.029	5	0.045 ± 0.020	40	0.089 ± 0.013
6 7	12	0.054 ± 0.015	6	0.050 ± 0.020	14	0.127 ± 0.032	32	0.071 ± 0.012
6 8	15	0.068 ± 0.017	15	0.125 ± 0.030	7	0.064 ± 0.023	37	0.082 ± 0.013
6 9	35	0.158 ± 0.025	15	0.125 ± 0.030	20	0.182 ± 0.037	70	0.155 ± 0.017
6 10	33	0.149 ± 0.024	16	0.133 ± 0.031	7	0.064 ± 0.023	56	0.124 ± 0.016
6 10*	1	0.005 ± 0.005	0	0.000	0	0.000	1	0.002 ± 0.002
7 7	2	0.009 ± 0.006	2	0.017 ± 0.012	4	0.036 ± 0.018	8	0.018 ± 0.006
7 8	13	0.059 ± 0.016	6	0.050 ± 0.020	4	0.036 ± 0.018	23	0.051 ± 0.010
7 9	15	0.068 ± 0.017	7	0.058 ± 0.021	9	0.082 ± 0.026	31	0.069 ± 0.012
7 10	17	0.077 ± 0.018	5	0.042 ± 0.018	7	0.064 ± 0.023	29	0.064 ± 0.012
7 10*	1	0.005 ± 0.005	0	0.000	1	0.009 ± 0.009	2	0.004 ± 0.003
8 8	1	0.005 ± 0.005	0	0.000	1	0.009 ± 0.009	2	0.004 ± 0.003
8 9	11	0.050 ± 0.015	4	0.033 ± 0.016	5	0.045 ± 0.020	20	0.044 ± 0.010
8 10	10	0.045 ± 0.014	9	0.075 ± 0.024	5	0.045 ± 0.020	24	0.053 ± 0.011
9 9	9	0.041 ± 0.013	3	0.025 ± 0.014	5	0.045 ± 0.020	17	0.038 ± 0.009
9 10	20	0.090 ± 0.019	7	0.058 ± 0.021	7	0.064 ± 0.023	34	0.075 ± 0.012
9 10*	1	0.005 ± 0.005	1	0.008 ± 0.008	3	0.027 ± 0.016	5	0.011 ± 0.005
10 10	7	0.032 ± 0.012	5	0.042 ± 0.018	6	0.055 ± 0.022	18	0.040 ± 0.009
10 10*	1	0.005 ±0.005	1	0.008 ± 0.008	0	0.000	2	0.004 ± 0.003

Table 18b: Absolute (Abs) and relative (Rel) allelic frequencies ± standard errors (SE) of human tyrosine (HUMTHO.1) STR polymorphism for three age groups in males.

	Age Group							
Allele	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
6	130	0.294 ± 0.022	88	0.367 ± 0.031	58	0.264 ± 0.030	276	0.306 ± 0.015
7	62	0.140 ± 0.017	28	0.117 ± 0.021	43	0.195 ± 0.027	133	0.147 ± 0.012
8	51	0.115 ± 0.015	34	0.142 ± 0.023	23	0.105 ± 0.021	108	0.120 ± 0.011
9	100	0.226 ± 0.020	40	0.167 ± 0.024	54	0.245 ± 0.029	194	0.215 ± 0.014
10	95	0.215 ± 0.020	48	0.200 ± 0.026	38	0.173 ± 0.025	181	0.201 ± 0.013
10*	4	0.009 ± 0.005	2	0.008 ± 0.006	4	0.018 ± 0.009	10	0.011 ± 0.003

** TH-STR alleles are indicated according to the repeat number. 10* is the imperfect HUMTHO.1 allele 10 (Puers et al., 1993)

Test for Hardy-Weinberg equilibrium (5000 permutations) :

- Group 1: p = 0.3606
- Group 2: p = 0.3594
- Group 3: p = 0.3732

Table 19a: Absolute (Abs) and relative (Rel) haplogroup frequencies ± standard errors (SE) of Mitochondrial DNA for three age groups in females.

	Age Group							
Haplogroups*	Group 1 (< 73 years)		Group 2 (73-90 years)		Group 3 (>90 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
H,V	123	0.373 ± 0.027	31	0.094 ± 0.016	43	0.130 ± 0.019	197	0.597 ± 0.027
I,V,X	27	0.0818 ± 0.015	5	0.015 ± 0.007	9	0.027 ± 0.009	41	0.124 ± 0.018
U,K	75	0.2273 ± 0.023	34	0.103 ± 0.017	18	0.055 ± 0.013	127	0.385 ± 0.027
J,T	65	0.1970 ± 0.022	14	0.042 ± 0.011	11	0.033 ± 0.010	90	0.273 ± 0.025
Others	40	0.1212 ± 0.018	20	0.061 ± 0.013	6	0.018 ± 0.007	66	0.200 ± 0.022

Table 19b: Absolute and relative haplogroup frequencies ± standard errors of Mitochondrial DNA for three age groups in males.

	Age Group							
Haplogroups*	Group 1 (<66 years)		Group 2 (66-87 years)		Group 3 (> 87 years)		Total sample	
	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE	Abs	Rel ± SE
H,V	89	0.403 ± 0.033	41	0.186 ± 0.026	38	0.172 ± 0.025	168	0.760 ± 0.029
I,V,X	16	0.072 ± 0.017	13	0.059 ± 0.016	12	0.054 ± 0.015	41	0.186 ± 0.026
U,K	47	0.213 ± 0.028	23	0.104 ± 0.021	30	0.136 ± 0.023	100	0.452 ± 0.033
J,T	43	0.195 ± 0.027	26	0.118 ± 0.022	19	0.086 ± 0.019	88	0.398 ± 0.033
Others	26	0.118 ± 0.022	17	0.077 ± 0.018	11	0.050 ± 0.015	54	0.244 ± 0.029

*Mitochondrial DNA haplogroups were pooled according to the phylogenetic proximity

Statistical Analyses

Multinomial logistic regression

A logistic regression for a polytomous response variable has shown to be extremely useful to explain the influence of some genetic polymorphisms on the genetic traits. Multinomial logistic regression model allows to evaluate the effect of one (or a few) dichotomous variable(s) (genotypes) on the probability to belong to different age groups.

Suppose we have a sample of *n* independent observations (y_i, $\mathbf{x_i}$), i=1,…,n, where y_i denotes the category of the outcome variable and $\mathbf{x_i}$ is a *p*-covariate vector $\mathbf{x_i} = (x_{i1},...,x_{ip})$ of independent variables, for the *i-th* subject. Furthermore, assume that the categories of the outcome variable Y, are coded at three levels 0, 1 or 2 indicating the age group which the subjects belong to, and each covariate equals 1 if the genetic characteristics is present and zero otherwise.

The probability $P(Y = j/\mathbf{x})$ to belong to any of the three age groups j=0,1,2 can be expressed as function of the *p* genotype information **x**.

Setting the group coded Y=0 as reference outcome value (baseline), the three category logit model pairs each other response category Y=1,2 with the baseline category Y=0. So, two logit functions are not redundant at a fixed setting **x** for explanatory variables:

$$g_1(\mathbf{x}) = \ln\frac{P(Y=1/\mathbf{x})}{P(Y=0/\mathbf{x})} \text{ and } g_2(\mathbf{x}) = \ln\frac{P(Y=2/\mathbf{x})}{P(Y=0/\mathbf{x})}.$$

The model (Agresti, 2000)

$$g_j(\mathbf{x}) = \beta_{0j} + \beta_{j1}x_1 + \beta_{j2}x_2 + ... + \beta_{jp}x_p = \beta_{0j} + \boldsymbol{\beta}'_j\,\mathbf{x} \qquad j = 1, 2 \qquad (A.1)$$

simultaneously describes the effect of **x** on these two logits, $(\beta_{0j}, \boldsymbol{\beta}'_j)$' is the vector of the unknown parameters in the linear predictor for the j-th logit. The effects vary according to the response paired with baseline. The logit for comparing Y=2 to Y=1 may be obtained as difference between the logit of Y=2 versus Y=0 and the logit of Y=1 versus Y=0

$$\ln\frac{P(Y=2/\mathbf{x})}{P(Y=1/\mathbf{x})} = \ln\frac{P(Y=2/\mathbf{x})}{P(Y=0/\mathbf{x})} - \ln\frac{P(Y=1/\mathbf{x})}{P(Y=0/\mathbf{x})}.$$

Multinomial logit model can be expressed in terms of response probabilities:

$$P(Y=0/\mathbf{x}) = \frac{1}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}};\ P(Y=1/\mathbf{x}) = \frac{e^{g_1(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}} \text{ and } P(Y=2/\mathbf{x}) = \frac{e^{g_2(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}$$

A general expression for the conditional probability in the three category model is:

$$P(Y = j/\mathbf{x}) = \frac{e^{g_j(\mathbf{x})}}{\sum_{k=0}^{2} e^{g_k(\mathbf{x})}} = \frac{\exp(\beta_{0j} + \boldsymbol{\beta}'_j \mathbf{x})}{1 + \sum_{k=1}^{2} \exp(\beta_{0k} + \boldsymbol{\beta}'_k \mathbf{x})}, \qquad \text{for } j = 0, 1, 2$$

where the baseline category vector $\beta_0 = (\beta_{01},...,\beta_{0p})' = \mathbf{0}$ and $g_0(\mathbf{x}) = \mathbf{0}$ for identifiably reasons.

Interpretation of the coefficients

The odds of the dependent variable being at level j, with j=0 as the baseline category, among subjects with binary covariate x_k=1 is defined as $\frac{P(Y = j/x_k = 1)}{P(Y = 0/x_k = 1)}$. Similarly, the odds of the dependent variable at category j among subjects with x_k=0 is defined as $\frac{P(Y = j/x_k = 0)}{P(Y = 0/x_k = 0)}$.

In a Multinomial logistic regression model, each estimated coefficient equals the estimated log odds ratio associated with a unit increase in the value of a covariate assuming that the values of all other covariates are held fixed. For considering the effect of a single independent variable, say the k-th, the effect of all the independent variables in the model with values $(x_1,...,x_{j-1},x_{j+1},...,x_p)$ are held constant, expressed as $K = \sum_{h=1}^{j-1} x_h\beta_h + \sum_{h=j+1}^{p} x_h\beta_h$.

Consider now that the k-th independent variable is dichotomous with categories 1 or 0, indicating if a particular feature is present or not, then the ratio of the j-th odds for an individual who has this feature to that of an individual without it, is

$$\frac{P(Y = j/x_k = 1; K, \boldsymbol{\beta}) / P(Y = 0/x_k = 1; K, \boldsymbol{\beta})}{P(Y = j/x_k = 0; K, \boldsymbol{\beta}) / P(Y = 0/x_k = 0; K, \boldsymbol{\beta})} = \exp(\beta_{jk}),$$

consequently, the log odds ratio is β_{jk}. It approximates how much more likely (or unlikely) it is for the dependent variable to be at level j instead of at the baseline category when the subject has the k-th characteristics x_k=1 than subjects do not have it x_k=0.

Maximum Likelihood Estimation

For i=1,…,n, let $\mathbf{y_i} = (y_{i0}y_{i1}y_{i2})'$ represent the multinomial trial for the subject i, where y_{ij}=1 when the response for the i-th subject is in category j and zero otherwise, with $\pi_j(x_i) = P[Y_{ij} = 1|\mathbf{x_i}]$, for j=0,1,2 each of which is a function of the vector of the 2(p+1) parameters $\boldsymbol{\beta}$, (Hosmer and Lemeshow, 2000).

The likelihood function for a sample of n independent observations is

$$L(\boldsymbol{\beta}) \propto \prod_{i=1}^{n} \left(\pi_0(\mathbf{x_i})^{y_{i0}} \pi_1(\mathbf{x_i})^{y_{i1}} \pi_2(\mathbf{x_i})^{y_{i2}} \right).$$

Using the fact that $\sum y_{ij} = 1$ for each i, the log-likelihood function is

$$l(\boldsymbol{\beta}) \propto \sum_{i=1}^{n} \left(y_{i1} g_1(\mathbf{x_i}) + y_{i2} g_2(\mathbf{x_i}) - \ln(1 + e^{g_1(\mathbf{x_i})} + e^{g_2(\mathbf{x_i})}) \right).$$

Fitting the three category logistic regression model maximizes the log-likelihood subject to multinomial probabilities $\pi_j(\mathbf{x})$, j=0,1,2, simultaneously satisfying the equations in (A.1) that specify the model.

Differentiating the $l(\boldsymbol{\beta})$ with respect to each of $2(p+1)$ unknown parameters gives the log-likelihood equations (to simplify the notation, we let $\pi_{ij} = \pi_j(\mathbf{x_i})$)

$$\frac{\partial l(\boldsymbol{\beta})}{\partial(\beta_{jk})} = \sum_{i=1}^{n} x_{ik}(y_{ij} - \pi_{ij}) = 0$$

for j = 1, 2 and k = 0 , 1, 2, …, p, with $x_{i0} = 1$ for each subject.

The solution requires iterative methods for solving non linear equations (as Newton-Raphson method) in order to determine the (ML) maximum likelihood parameter estimates $\hat{\boldsymbol{\beta}}$.

Measures of Goodness-of-fit

Let *M* be the maximum likelihood estimated model. We use the deviance statistic *D* for testing the null hypothesis that the model (*M*) holds against the alternative that the most general model holds (Fabbris, 1997)

$$D = -2 \ln \left[\frac{\text{likelihood of the estimated model}}{\text{likelihood of the saturated model}} \right].$$

In testing whether the estimated model (*M*) fits, we test whether all parameters in the saturated model (*S*) but not in (*M*) equal zero. The saturated model contains as many parameters as there are data points and the fitted model is nested within (*S*).

Under the null distribution, *D* is approximately chi-squared and the degrees of freedom are the difference in the number of parameters in the two models (*M*) and (*S*). Small values of *D* suggest that the model (*M*) has a good fitting.

By checking the fit of model, we can also compare the fitted model (*M*) with the same model in which only the intercept appears. In that way under null hypothesis all parameters

of the fitted model are supposed null except the intercept. The likelihood ratio statistic, denoted with G^2, is defined as difference

$$G^2 = D(\text{model with only intercept}) - D(\text{estimated model})$$

It allows to test whether the difference in goodness of fit for the two models is significantly high, so that it shows evidence that adding regressors in the model the fitting has improved.

The chi-squared asymptotic distribution of G^2 has p degrees of freedom as many as coefficients in the estimated model. High values for the statistics G^2 suggest evidence against the null hypothesis.

In order to construct the multinomial logistic regression model it is important to answer questions about the contributions of each independent variable to the prediction of the response. The null hypothesis for this test is $H_0 : \beta_{jk} = 0$ for the k-th covariate in the j-th logit, j=1,2 and k=0,1,…,p. To test such a null hypothesis, one can perform a likelihood ratio chi-squared test with 1 degree of freedom comparing the measures *D* of the models defined with and without the independent variable at issue. The change in the goodness-of-fit due to the inclusion of the independent variable is obtained as follows:

$$G^2 = D(\text{model without the variable}) - D(\text{model with the variable}).$$

It can also be expressed by:

$$G^2 = -2\ln\left[\frac{\text{likelihood of the model without the variable}}{\text{likelihood of the model with the variable}}\right].$$

This statistic is large when model without variable fits poorly compared to model which contains the variable.

An easier alternative test for the hypothesis $H_0 : \beta_{jk} = 0$, with j = 1,2 and k = 0,1,2,…,p, for each coefficient in the model is using the Wald statistic test computed as

$$W_{jk}^2 = \frac{\hat{\beta}_{jk}^2}{se(\hat{\beta}_{jk})}$$

where $se(\hat{\beta}_{jk})$ is the standard error of $\hat{\beta}_{jk}$. In performing this test, we consider that under null hypothesis, the Wald statistic follows a chi-squared distribution with 1 degree of freedom. The p-value is then the right-tailed chi-squared probability above the observed value and we reject H_0 when p-value is lower than the fixed level of significance α.

The Nagelkerke R^2 is a statistic that attempts to quantify the proportion of explained "variation" in the logistic regression model, likewise to the R^2 in a linear regression model.

The Nagelkerke index is $R^2 = \frac{\widetilde{R}^2}{\max(\widetilde{R}^2)}$ where $\widetilde{R}^2 = 1 - \left(\frac{l(0)}{l(\boldsymbol{\beta})}\right)^{\frac{2}{n}}$, $\max \widetilde{R}^2 = 1 - [l(0)]^{\frac{2}{n}}$, $l(0)$ is the log-likelihood for the model with only a constant, $l(\boldsymbol{\beta})$ is the log-likelihood for the model under consideration. The closer to one R^2 is, the better the model fits.