*Alma Mater Studiorum – Università di Bologna*

DOTTORATO DI RICERCA IN

**Oncologia e Patologia Sperimentale: progetto n.2**

**Ciclo XXIII**

**MED/04**

**Dipartimento di Patologia Sperimentale**

# The mitochondrial genetic in the Integrated European Project "GEHA – *GEnetics of Healthy Aging*"

**Dott.ssa ANNALAURA BARBIERI**

**Coordinatore:**
**Prof. SANDRO GRILLI**

**Relatore:**
**Prof. CLAUDIO FRANCESCHI**

*Esame finale anno 2011*

# INDEX

# PREFACE

Europe is the oldest continent in the world, being a quarter of the entire population more than 60 years old. This "demographic explosion" makes important to identify factors (biological and not biological) involved in aging, with the main goal of increasing the number of elderly persons in good health.

The present study is part of the Integrated European Project "**GEHA - *GEnetics of Healthy Aging***" (Franceschi et al. 2007), whose aim is to identify genes involved in healthy aging and longevity, which allow individuals to survive to advanced age in good cognitive and physical conditions and in absence of the major age-related diseases.

To achieve this aim, it was important to: (1) collect information on health status and DNA from 2192 long-lived sibpairs (90+) and an equal number of ethnically-matched younger controls from the 11 European countries; (2) perform a genome-wide linkage scanning in all the sibpairs and a LD mapping (linkage disequilibrium) of the candidate chromosomal region; (3) analyse three chromosomal regions (region D4S1564 in chromosome 4, region 11.p15.5 in chromosome 11, APO-E in chromosome 19), which were previously seen to be associated with longevity; (4) genotype all recruited subjects for APOE polymorphisms; (5) genotype all the recruited subjects for mtDNA haplogroups; (6) sequence a subset of the mtDNAs in order to evaluate the genetic and epigenetic variability.

In this context, the purpose of the Project is to examine the association of haplogroups with healthy aging and to sequence the entire mitochondrial genome in order to study new mutations and polymorphisms related to aging.

# Chapter 1

# INTRODUCTION

## 1. The aging process and mortality deceleration

Aging may be considered as the most common disease affecting inexorably 100% of the animal world and is the accumulation of changes in an organism over time. Aging can be defined as a multifactorial phenomenon characterized by a time-dependent decline in physiological function. The individual age can be defined with three parameters: **chronological ageing**, referring to how old a person is, is arguably the most straightforward definition of ageing and may be distinguished from "**social ageing**" (society's expectations of how people should act as they grow older) and "**biological ageing**" (an organism's physical state as it ages).

Differences are sometimes made between populations of elderly people. Divisions are sometimes made between the young old (65–74), the middle old (75–84) and the oldest old (85+). However, problematic in this is that chronological age does not correlate perfectly with functional age, i.e. two people may be of the same chronologic age, but differ in their mental and physical capacities, so they have extremely different biological age.

It is very difficult to define when ageing starts, because it is deeply influenced by culture, society, economic well-being and by the society self. Moreover, it is natural to associate aging with degenerative pathologies: some presents a mental decay, others only inferior articulation weakness or in particular one sensorial organ. From biomedical point of view, aging is a universal and progressive change which leads to a loss of individual capacity of adapting in the environment in which it lives. This event inexorably leads the organism to a major susceptibility condition against disease; consequently mortality arises in an age-dependent way.

Senescence is the state or process of aging. In particular *cellular senescence* is a phenomenon where isolated cells demonstrate a limitated ability to divide in culture (discovered by Hayflick in 1961) while *organismal senescence* is the aging of organisms. This senescence is characterized by a declining ability to respond to stress, increasing

homeostatic imbalance and increased risk of disease. Today, it is still not discovered one single aging biomarker. In Humans, this phenomenon is characterized by an enormous heterogeneity which appear not only among different organisms, but also in the single organism: tissue age in different way, producing a proliferative mosaic (Franceschi et al., 1999), where also single cells have different replicative potential. One example of cellular senescence has been attributed to the shortening, *in vitro,* of telomeres with each cell cycle in peripheral blood leucocyte and not in centenarians fibroblast; when telomeres become too short, the cells die. The length of telomeres is therefore the "molecular clock," predicted by Hayflick.

From a demographic view, in the latest one hundred years, and in particular from the second half of XX century and in industrialized countries, a demographic transition has happened, characterized by a drastic decrease of infant mortality and a rising of aging rate.

In the world, Europe is the area with the highest aging rate: a quarter of the entire European population has more than 60 years of age and it is estimated that it could reach a third in the next 30 years.

So, far from being fixed, the mortality of people over 80 years decrease dramatically since 1950, especially since 1970 in developed countries. The mortality curve was represented by Gompertz in 1825: it shows a semi-logarithmic scale in which death rate is in function with Age (years) and it explains why mortality rates increase exponentially with age (the Gompertz law) in many species, by taking into account the initial flaws (defects) in newly formed systems. He also suggested that it is a feature of all organisms but this could be true at his time, when people's lifespan was 80 years maximum. The Gompertz–Makeham law of mortality describes the age dynamics of human mortality rather accurately in the age window from about 30 to 80 years of age. At more advanced ages, death rates do not increase as fast as predicted by this mortality law – a phenomenon known as the late-life mortality deceleration. Now people can reach the venerable age of 100 and "real" data appreciably differ from Gompertz curve (in blue in fig.1). Vaupel et al (1998) tested the hypothesis that mortality accelerates with age as reproduction declines, estimating age trajectories of death rates for Humans and other organisms. As **Fig. 1.1** shows, death rates increase at a slowing rate after age 80. A logistic curve (in blue), that fits the data well from age 80 to 105, indicates that death rates may reach a plateau. A quadratic curve (in green) fit to the data at ages 105+ suggests a decline in mortality after age 110. The red line is the aggregation of 14

countries (Japan and 13 Western European countries) with reliable data, over the period from 1950 to 1990 for ages 80 to 109 and to 1997 for ages 110 and over.

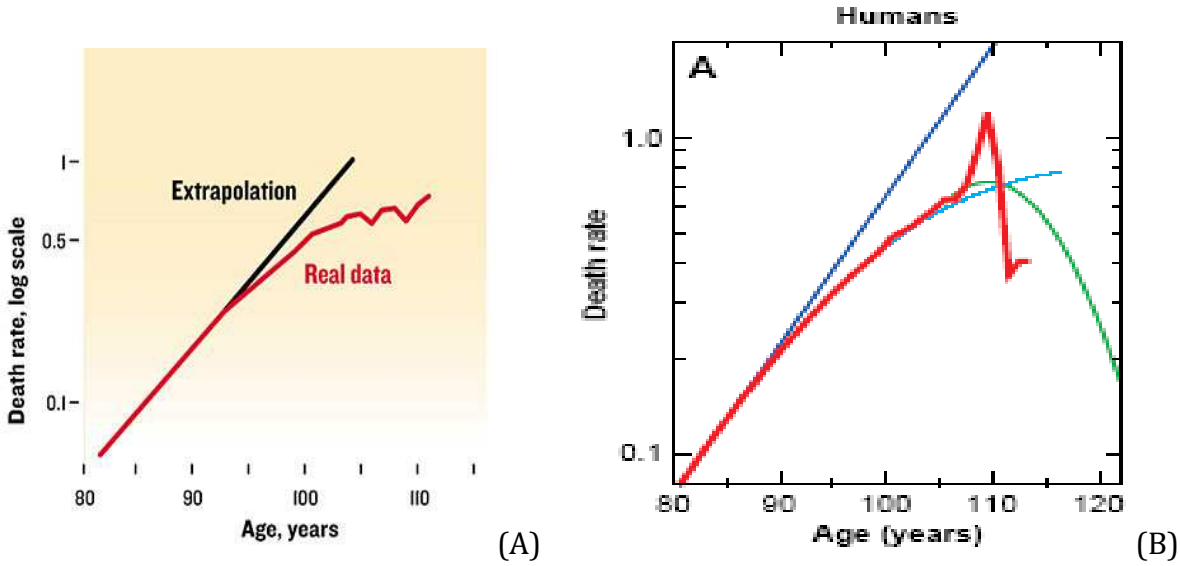



**Fig. 1.1**: (A) Figure inspired by Gavrilov and Gavrilova. It represents the Gompetz law and the real data trend. After 90 years, the mortality rate decreases. (B) Age trajectories of death rates from 80 to 122 years for females. The logistic curve (in blue) fits the data well from age 80 to 105 and it indicates that death rates may reach a plateau. A quadratic curve (in green) fits the data at ages 105+ suggesting a decline in mortality after age 110. The red line is the aggregation of 14 countries (Japan and 13 Western European countries) with reliable data, over the period from 1950 to 1990 for ages 80 to 109 and to 1997 for ages 110 and over (Source: Vaupel *et al*. 1998)

Mortality deceleration was observed not only in humans, but also in organisms, such as *Ceratitis capitata* (the Mediterranean fruit fly), *Anastrepha ludens*, *Anastrepha obliqua*, and *Anastrepha serpentina* (three other species of true fruit fly), *Diachasmimorpha longiacaudtis* (a parasitoid wasp), *Drosophila melanogaster*, *Caenorhabditis elegans* (a nematode worm), and *Saccharomyces cerevisiae* (baker's yeast) (Vaupel et al., 1998).

There is not still a clear explanation for this phenomenon, but this exponential rising of old persons could be attributable to changes at the individual level on the one hand and to changes in the composition of the surviving cohort on the other. In 2000, 69 million people world wide were aged 80 or over. By 2050, the 80+ year-old people are projected to be more than 370 million, representing 4,4% of the entire population. Similarly, the number of nonagenarians will reach 63 million by 2050, and centenarians will reach 5.3 million. These improvements in life expectancy came as the consequence of rising standard quality of life, environment hygiene, alimentation (both quality and

total caloric introduction of food), public health interventions, and medical developments that reduced death from infectious diseases.

The increase in life expectancy in developed countries leads to one hand extremely complicated demographic phenomenon, and on the other hand new problems regarding the allocation of resources for old age pensions and care for the elderly. In order to provide a suitable health care support, it is necessary a better comprehension of factors (biological and non-biological) involved in aging devoid of major diseases and disabilities, contributing to a much more well-being status.

The aging phenomenon is very complicate and it requires an integrated approach, or rather to combine biological studies with social sciences, demographic, historic and anthropological studies (Franceschi et al., 2000a). Accordingly the term "***successful aging***" identifies a multidimensional phenomenon, encompassing the avoidance of disease and disability, the maintenance of high physical and cognitive function, and sustained engagement in social and productive activities. Successful ageing may be viewed an interdisciplinary concept, spanning both psychology and sociology, where it is seen as the transaction between society and individuals across the life span with specific focus on the later years of life (Fentleman et al. 1990)

In this perspective, it is presumed that reaching an old age free from the most common pathologies is possible. Following studies have tried to separate the concept of "successful aging" from "pathological status" and "functional decline" (Rowe e Kahn, 2004). Thus, aging can not be determined only by genetic control (30%) but also by environmental and social-sanitary factors (social problems, stress, poor-feeding, pathologies in general). All these factors may modify every individual vulnerability with a consequent loss of adaptability (Motta *et al*., 2005; Perls *et al*., 2002).

Style of life seems to have an important role for aging as it is demonstrated by longitudinal studies; one of that was the European longitudinal study HALE (Knopps *et al*., 2004) conducted on subjects with a range age of 70-90 years, and a second study was conducted on subjects with a mean age of 80 years (Woo *et al*., 2002). The first has demonstrated that an optimal diet (Mediterranean diet) leads to a drastic decreasing of 50% of mortality; the second that an intense physical activity and a major consumption of fish, associated with less cigarette and alcohol, dramatically reduce mortality.

Day by day it has understood that "successful aging" studies could be the key to understand the difficult, and till now not completely clear, process of senescence. Today we speak about longevity.

**2. The extreme longevity**

Longevity is a multifactorial process in which genetic and environmental factors, as well epigenetic and stochastic element, seem to interact and each making variable contribution to the overall presentation of the phenotype (Candore *et al*., 2006).

$$\mathbf{L = En + S + G + Ep}$$

**(Longevity = Environment + Genetics + Epigenetics + Stochasticity)**

1) Environment: the environment represents about **20 to 30 percent** of longevity, meaning non genetic survival attributes that are fixed for individuals by the time they are 30 years old (Vaupel *et al*., 1998). Among these **non genetic fixed factors** are healthy conditions early in life, socio-economic conditions in childhood, the socio-economic position a person attained at about age 30, and physical environment (personality, intelligence, health behavior and everyday activities, mental and physical health), each contributing to attain longevity. Several studies in medicine and epidemiology, like Elo *et al*. (1992), and Evans *et al*. (2005), show that environmental quality is a very important factor affecting health and morbidity: air and water pollution, depletion of natural resources, are all capable of increasing human mortality. In today's western world, many illnesses today emerge from poor eating habits, lack of exercise, poor sleeping habits, and bad habits such as excessive drinking, drugs and nicotine. Declining cigarette smoking as well as changes in diet habits may be important factors in rising life expectancy, at least in some countries. It has been widely reported that environment plays an important role during pregnancy and during the first year, because in this period neuron's number and a lot of parameters are established. Barker's (1992) "**fetal-origins hypothesis**" suggests that nourishment and infections *in utero* and during infancy program the development of risk factors for several important diseases of middle and old age. Longevity may in part be determined by conditions in early childhood and perhaps before birth (Kanaka-Gantenbein, 2010).

Another important trait is the **psychological and dispositions towards life**. A higher longevity makes people more sympathetic to future generations and/or their future selves. Therefore, if someone expects to live longer, they should be willing to invest more in environmental quality. Lower cognitive functioning was consistently associated with an increased mortality risk across a number of studies (Small and

Backman, 1999). There is also some evidence that aspects of subjective positive and happiness are related to a longer life, while psychological distress and depression may increase the risk of death (Huppert and Whittington, 1995).

2) Stochasticity: another aspect to be taken into account is the **stochasticity**, meaning the wide variation of life span of genetically identical organisms even if reared in a constant environment. Kirkwood (2005) has shown that an isogenic population of the nematode *C.Elegans* has a striking intrinsic variability of life span (from 8 to 32 days, depending also on the strain). The whole process depend on chance and these stochastic events are absolutely random.

3) Genetic and epigenetic: studies on twins estimate that unexpectedly less than 10 % of individual variation in life expectancy after 30 years of age, depend on early event (intrauterine life and infancy), while the 65 % on late event, happened after 30 years of age. Thus, the same studies on twins indicate that 25-30% of longevity variability depends on genetic-epigenetic pool as quantitative multifactorial trait.

Epigenetic refers to phenotypic changes caused by mechanisms that are unrelated to changes in the underlying DNA sequence, most notably chromatin remodeling driven by histone modifications, and DNA methylation. An increasing body of evidence supports a role for epigenetic changes in the etiology of aging and its associated disease sequelae (Silvia Gravina *and* Jan Vijg, 2010). While epigenetic changes are essential for development and differentiation, they can also arise later in life either by non-random mechanisms, such as responses to environmental change, or through stochastic errors in maintaining fixed patterns of DNA or histone modification.

What is the theoric end of human life? A French woman has reached 122 years and surely it will not be an isolated case. The aging rate and the maximum life span are extremely variable, in fact genetic factors control long survival (90 years and more) explaining why nonagenarians and centenarians are numerous in the same family (cluster). Stocasticity seems to be important during the first years of life, while genetic gives its contribute after sixty years.

Interestingly, the aspect of sex differences in longevity must be analyzed. Oksuzyan et al. (2008) examined sex differences in health and survival, with a focus on Nordic countries, finding that men are physically stronger with fewer disabilities, but have substantially higher mortality rate at all ages compared with women. It is clearly a male-female health-survival paradox. It is probably due to multiple causes that include fundamental biological differences (immune factors and responses, hormones, and

disease patterns), behavioural differences (for example risk-taking, reluctance to seek and comply with medical treatment) and a methodological bias (such as selective non-participation and under-reporting of health problems, and delayed seeking of treatment by men) between sexes.

In recent decades the research on aging has expanded quickly, probably as a consequence of the lengthening of the average human life span and the increasing percentage of elderly population. There is a huge number of hypothesis and theories, divided according to the basic idea of aging being a programmed process or not, as the table shows (**Tab. 1.1**).

| Biological Level/Theory | Description |
|---|---|
| Evolutionary | |
| Mutation accumulation* | Mutations that affect health at older ages are not selected against. |
| Disposable soma* | Somatic cells are maintained only to ensure continued reproductive success; after reproduction, soma becomes disposable. |
| Antagonistic pleiotropy* | Genes beneficial at younger age become deleterious at older ages. |
| Molecular | |
| Gene regulation* | Aging is caused by changes in the expression of genes regulating both development and aging. |
| Codon restriction | Fidelity/accuracy of mRNA translation is impaired due to inability to decode codons in mRNA. |
| Error catastrophe | Decline in fidelity of gene expression with aging results in increased fraction of abnormal proteins. |
| Somatic mutation | Molecular damage accumulates, primarily to DNA/genetic material. |
| Dysdifferentiation | Gradual accumulation of random molecular damage impairs regulation of gene expression. |
| Cellular | |
| Cellular senescence-Telomere theory* | Phenotypes of aging are caused by an increase in frequency of senescent cells. Senescence may result from telomere loss (replicative senescence) or cell stress (cellular senescence). |
| Free radical* | Oxidative metabolism produces highly reactive free radicals that subsequently damage lipids, protein and DNA. |
| Wear-and-tear | Accumulation of normal injury. |
| Apoptosis | Programmed cell death from genetic events or genome crisis. |
| System | |
| Neuroendocrine* | Alterations in neuroendocrine control of homeostasis results in aging-related physiological changes. |
| Immunologic* | Decline of immune function with aging results in decreased incidence of infectious diseases but increased incidence of autoimmunity. |
| Rate-of-living | Assumes a fixed amount of metabolic potential for every living organism (live fast, die young). |

**Tab. 1.1** Classification of the most important theories of ageing (Source: Weinert and Timiras, 2003)

### 3. The evolutionary theories

Before the theory of evolution, the process of aging was conceived in the same way that all things deteriorate slowly over time. The first evolutionary theory by Charles Darwin was based on the concept that random and heritable variation of biological traits (caused by mutations) will lead to natural selection for preferential reproduction of those individuals who are particularly fit in a given environment. After the development of the theory of evolution, scientists began to wonder why evolution had produced such complex and well-adapted creatures that were so successful at surviving from conception through to adulthood, but which then fell into decay and died. August Weismann in 1891 developed the theory of programmed death which proposed that

aging evolved to the advantage of the species, not the individual, and that there just be an evolutionary advantage to having only a limited lifespan.

But contemporary theories propose two models for how aging can evolve: one is the theory of mutation accumulation and the second is the antagonistic pleiotropy hypothesis. The *mutation accumulation theory* of aging was proposed by Medawar in 1952 and considers aging as a bio product of natural selection. Aging in fact has no adaptive traits because natural selection does not occur in long lived animals and provides little additional contribution to offspring numbers. It means that old age is not under selective pressure per se. On the other hand the *antagonistic pleiotropy theory* proposed by Williams in 1957 says that some genes are beneficial at earlier ages but harmful at later ages (the genes with age related opposite effects are called pleiotropic genes).

There were attempts to define better the antagonistic pleiotropy theory; the *disposable soma theory* proposed by Kirkwood and Holliday (2005) predict that aging occurs due to the accumulation of damage during life and that failures of defensive or repair mechanisms contribute to aging. It postulated a special class of gene mutations with antagonistic pleiotropic effects in which hypothetical mutations save energy for reproduction (positive effect) by partially disabling molecular proofreading and other accuracy promoting devices in somatic cells (negative effect). In other words, given finite resources, the more an animal expends on bodily maintenance, the less it can expend on reproduction, and vice versa. The distinction between somatic and reproductive tissues is therefore important because the reproductive cell lineage, or germ line, must be maintained at a level that preserves viability across the generations, whereas the soma needs only to support the survival of a single generation. According to Kirkwood aging is a sort of strategy prefixed by natural selection in order to preserve energy (**Fig.1.2**). Organisms' life, who live in wild conditions (in a not protected environment), is drastically reduced by extrinsic phenomena (incidents, starvation, cold, predation, infection, and so on) and in consequence they die before aging. In a protected environment, organism can reproduce but it does not die, experiencing the aging process. In conclusion, senescence is not programmed because it hits only individuals who live long.

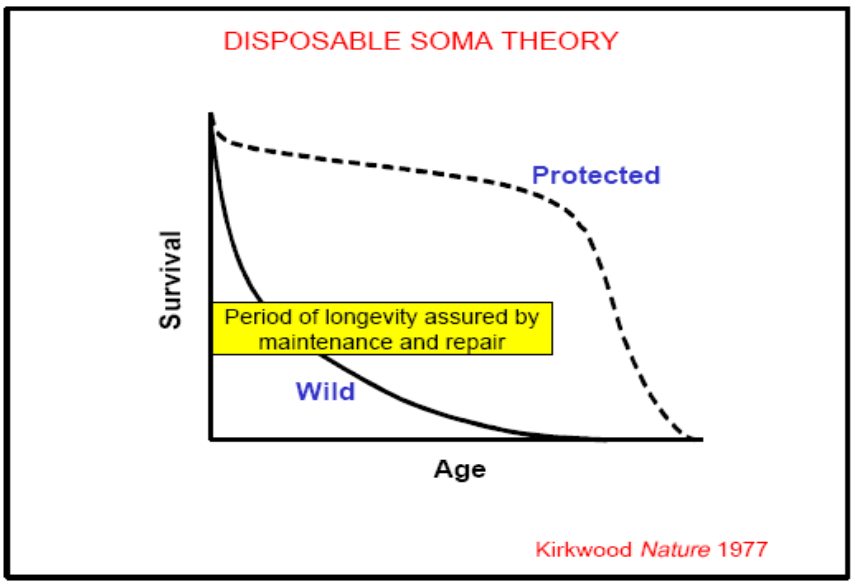



**Fig. 1.2** In a not protected environment, organisms' life is drastically reduced by extrinsic phenomena. In a protected environment organisms can experience aging.

### 4. The network theory

In 1989 Franceschi proposed a general theory of aging suggesting that this process is controlled by a *network* of cellular and molecular defence mechanisms (Franceschi, 1989). Among molecular and cellular defence network there are (**Fig. 1.3**):

- DNA repair mechanisms
- Antioxidant defence system, enzymatic or not
- Production of heat shock protein (HSPs)
- Activation of poly(ADP-ribose)polymerase (PARP)
- Apoptosis, an ancestral process which permit to eliminate damaged cells, mutated, or transformed (Franceschi *et al.*,1995a).

These mechanisms cope with such a variety of potentially harmful agents, such as internal or external stressors, dangerous for the maintenance of cell functional integrity, physical stressors (UV, gamma radiation, heat), chemical stressors (oxygen free radicals and reducing sugars), biological agents (Bacteria and viruses).

The network system of cellular and molecular basis of anti-aging (preserved by evolution)

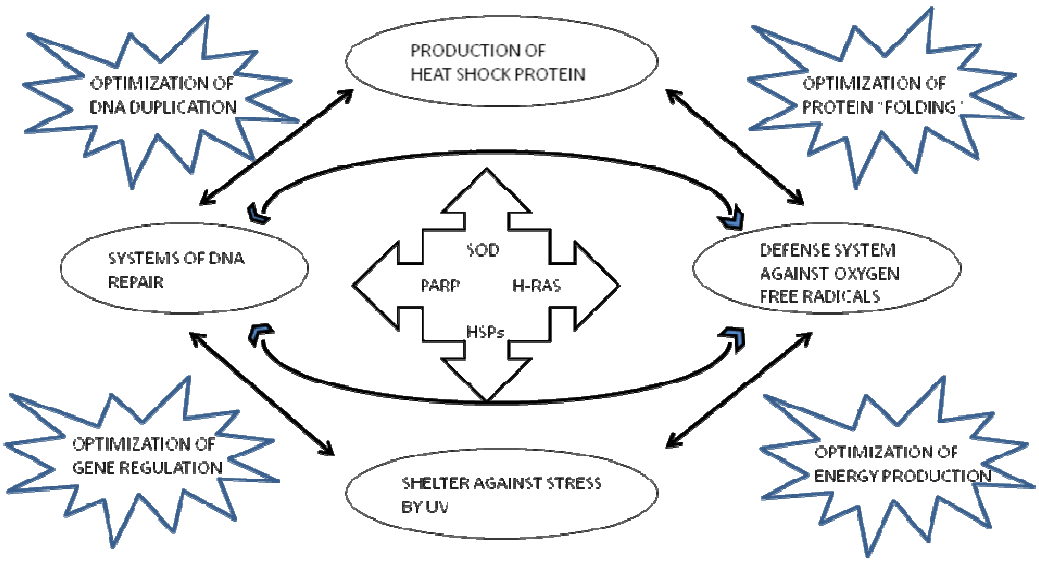



**Fig. 1.3:** The network of cellular and molecular system anti-aging (Original fugure modified by "Inflamm-aging. An evolutionary perspective on immunosenescence", Franceschi *et al.*, 2000)

The aim of the network theory is to combine evolutionary theories of aging proposed by Kirkwood (1977) with molecular biology data. An alteration of this defence system leads to a homeostasis loss and in consequence to senescence determined by physiological modification.

The way that every single mechanism contribute to network and how it is organized are still not clear, in particular in superior organisms. Some of the necessary functions involved in stressors defence were conserved during evolution, making an improvement and expansion of mechanisms, producing new and higher levels of organization (Franceschi *et al*., 2000b).

### 5. The remodelling theory

The necessity to conceptualize studies on human immunosenescence and the new model of healthy centenarians has lead to the elaboration of remodelling theory in aging (Franceschi et al. 1995, Franceschi and Cossarizza, 1995); the main question was to evaluate which was the contributor of the immune system to longevity.

In this approach only centenarians in good health, without principal pathologies and physic and cognitive disability were considered as a model of successful aging and successful physiological immunosenescence. Healthy centenarians are quite rare, so the assessment of their health status is methodologically difficult. Franceschi et al. (2000a)

studied 382 centenarians and found that 22 % of them were in good health status and they were compared with healthy people of different ages obtaining that some immune responses were unexpectedly well conserved in centenarians and that immune responses were differently affected by the aging process. Immunosenescence could be the result of continuous remodelling or *adaptation* of the entire system to the deteriorative changes occurring over time. In consequence, healthy centenarians are those who have the capacity of adapting their body to damaging agents and to immunological stressors as well. From an immunological point of view, centenarians are not "the best" but the "best adapted" to environment. From an evolutionary point of view, we have to consider that the human immunological system were programmed to survive till reproductive age and then to reach senescence.

The increasing percentage of 80-20 years people and the fact that environment conditions are in rapidly changing, in which the probability to encounter new antigens is relatively higher than last centuries, force the immune system to operate in new conditions, not planned by evolution and thus to modify itself in a still not clear way.

### 6. The replicative senescence

A telomere is a region of repetitive DNA at the end of a chromosome, which protects the end of the chromosome from deterioration, and of some non coding genes. In Humans, the repetitive sequence in telomeres is constituted by six nucleotides TTAGGG, repeated for a total length of 3-20 kb. There are also additional 100-300 kb between telomere and the rest of chromosome.

Telomere length is maintained in immortal cells by the telomerase enzyme. In laboratory, mortal cell lines can be immortalized by the activation of their telomerase gene, present in all cells but active in few cell types. What is the connection between telomere and aging process? Telomeres can be defined "molecular clock": as a consequence of cellular replication, their length reduce dramatically till when they can not perform their protective function for chromosomes. Cells can not reproduce correctly because the telomerase enzyme, an inverse transcriptase, is not able to regenerate the telomere till the end and its structure reduces every time cell divides itself. It means that each DNA replication leads to an important loss of genetic information, causing cell death.

To keep the same length of telomeres means confer eternity to cells. Joeng at al. (2004) conducted a study with two groups of worms belonging to *C.Elegans* and with

two different telomeric length. The research has demonstrated that the group with longer telomeres lived with a life expectancy more than 20%. But keeping the same telomere length means produce abnormal cells, leading to cancer.

### 7. MtDNA damage and the mitochondrial theory of aging

Among all aging theories, the ***free radical theory*** is one of the most gained, proposed for the first time by Harman in 1957. It explains how aging is strictly associated with high level of free radicals, which are chemical species with a free electron in their external orbital and, for this reason, they are extremely reactive and instable. Different conditions related to oxidative stress could interact with aging: an increasing speed of producing reactive oxygen species (ROS), a decline of antioxidant defense system and a low efficiency in repairing damaged molecules.

Although the available evidence strongly suggests that mitochondria play a role in this process, there appears to be a wide range of opinions as to the exact nature of the involvement of mitochondria in aging. The ***mitochondrial theory***, proposed by Miquel in 1980, is considered an extension of the free radical theory. As it is commonly known, nDNA is protected by histone proteins and various repair enzymes, which minimizes damage to nDNA from free radicals/oxidants. mtDNA has no histone protection or significant enzymes repair systems to offer free radical protection. Therefore, mtDNA is far more subject to free radical damage than nDNA. The commonest form of free radical damage to mtDNA molecules is the production of 80HdG, an oxidized guanine base. Even in young (3 month old) rats, the level of 80HdG is already 16 times higher in mtDNA than nDNA (Richter et al. 1995). As mtDNA damage accumulates over the lifetime of an individual, the functionality of the ETC enzyme complexes that produce ATP, and are encoded for (in part) by mtDNA, decreases dramatically and gradually produces a cellular energy crisis. The system is not capable to keep the equilibrium and it leads to cellular aging first, organism aging then. The dramatic mutation increasing causes defective structures formation in the respiratory chain and in consequence a defective functionality which creates a rising in ROS production. The mitochondrial theory of aging is based around the idea of a vicious cycle, in which somatic mutation of mtDNA engenders respiratory chain dysfunction, enhancing the production of DNA-damaging oxygen radicals.

The mtDNA has a very high mutation rate due to its chronic exposure to mitochondrial ROS. When a new mtDNA mutation arises in a cell, a mixed intracellular population of

mtDNAs is generated (wild-type and mutated mtDNAS), a state known as eteroplasmy. The mutant and normal molecules are randomly distributed into the daughter cells and as a consequence of this replicative segregation, the proportion of mutant and normal mtDNAs can drift toward an homoplasmic condition (all mutant or all wild type).

Furthermore, it has seen that in order to compensate the energy deficiency, mtDNAs have been found to be preferentially, clonally, amplified within cells (Coskun et al. 2003).

Therefore, cells with defective mitochondria are preferentially stimulated to replicate their mitochondria and mtDNAs, the mitochondrial energetic output declines, ROS production increases, and the propensity for apoptosis increases. Thus the accumulation of mutant mtDNA creates the aging clock (**Fig. 1.4**).

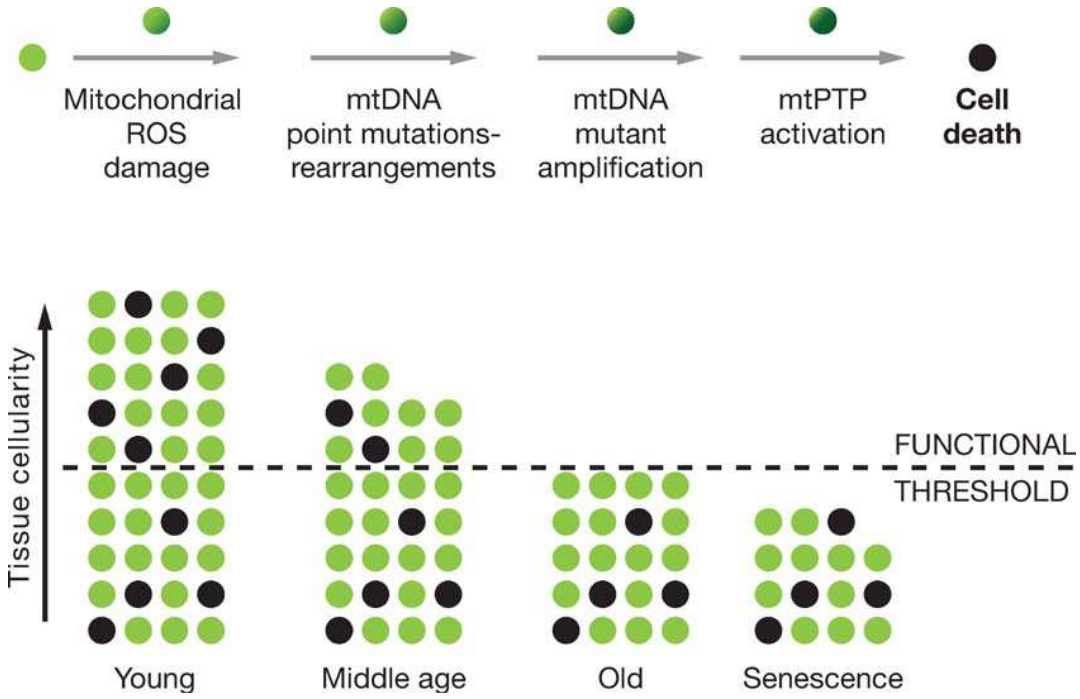



**Fig. 1.4:** Aging model. The dot line shows the minimum number of cells for the tissue to function normally. In black are died cells in a mitochondrial-mediate process. In green cells with optimal function.

### 8. The genetics of Human Longevity

A recent research conducted in Boston has revealed genetic profile of persons predispose to reach 100 years or more even if till now, a unique variant related to longevity has not been discovered but Family studies of exceptional longevity can potentially identify genetic and other factors contributing to long life and healthy aging.

The two main concepts arisen from recent studies on the genetics of human longevity are the following:

1) Human longevity clusters in families
2) Long-living siblings are likely enriched in longevity genes

Sebastiani (Science, 2010) has analyzed genetic profiles of centenarians and she has discovered 19 variants shared by people with similar characteristic, i.e. survival age, delay in developing Alzheimer disease, cardiovascular diseases and hypertension.

Therefore, natural selection does not select aging genes but aging is the result of pleiotropic effect of genes that operate in other areas. It is very interested that numerous studies have evidenced how life span is due to ipomorphic or nullomorphic mutations; the "*wild-type*" genes seem to codify processes with negative effect on longevity, thus defined "gerontogenes". The overwhelming majority of gerontogenes mutations increase the capacity of coping with stress and specifically with oxidative stress and caloric restriction (Christensen *et al*., 2006).

So, longevity is under genetic control (nuclear and mitochondrial), suggesting a substantial hereditability of healthy aging, in a complex interaction with a great variety of environmental factors (life style, nutrition and culture as well):

1. **Studies on evolutionistic biology** indicate that life span is species-specific and has a strong genetic basis. There are also mutations in specific genes which are able to increase life span, being involved in some principle metabolic ways (resistance to oxidative stress, insulin way, energetic metabolism). We talk about "*major genes*", genes conserved during the evolution and that control life span (Tatar *et al*., 2003).
1. **Studies on long-lived families and sibpairs** have evidenced that the survival rates of siblings of centenarians and siblings of a similar birth cohort, who died in their early seventies, had a four times greater chance of surviving to their early nineties (Perls *et al*., 1998). The same group in 2002 discovered that males sibling of centenarians, comparing with a cohort of 1900, had a seventeen times greater change to reach 100 years, while sisters had a probability eight times greater (Perls *et al*., 2002).
2. A great number of **association studies** have been conducted on candidate genes of centenarians. Some genes like ApoE, ApoB, ApoA1, H-Ras, IL-6, IL-10, IFNγ,

PON1, Tyrosine Hydroxylase, INS, IGF2, IGF-1R, Glutathione transferase (GSTT1) SIRT3 seem to play an important role in healthy aging and longevity (De Benedictis *et al*., 2001; Bonafè *et al*., 2003; Rose *et al*., 2003, Bellizzi *et al*. 2005, Glatt *et al*. 2007) (**Fig. 1.5**). In particular, studies on -174 C/G polymorphism of IL-6 promoter in male centenarians has revealed a significant increasing of person carrying mutation CG or CC (Franceschi *et al*., 2000; Bonafè *et al*., 2001). Probably the presence of this mutation could be surviving favorable in the last part of life, since significative differences among classes of inferior ages have not been observed.

Moreover, a relationship between plasmatic concentration of *IL-6* and susceptibility to principal age-related pathologies, such as Alzheimer disease (Franceschi *et al*., 2001), diabetes (Fernandez-Real *et al*., 2000), osteoporosis (Ferrari *et al*., 2001; McLean, 2009) and cardiovascular pathologies (Basso *et al*., 2002) has come out.

Another study was conducted on *PON1 gene* (Paraoxonase1), a major anti-atherosclerotic component of high-density lipoprotein (HDL), responsible for hydrolyising organophosphate pesticides and nerve gasses. Studies on two polymorphisms of PON1 gene (one in codon 55 and one in 192) evidenced that allele B frequency in codon 192 is higher in centenarians than in controls and that B+ subjects are carrying the M allele in codon 55 (Bonafè *et al*., 2003; Rea *et al*., 2004).

It has also been evidenced that the presence of allelic variant ε4 of ApoE gene (ApoE4) plays a negative role in reaching longevity: biomedical researchers have demonstrated how ε4 frequency in Italian and French centenarians is significantly lower than in younger controls. Moreover, it has been noticed that subjects affected by Alzheimer disease have a higher frequencies of ε4 allele, demonstrating that this allelic variant represents a high risk for the pathology (Carrieri *et al*., 2001; Blanchè *et al*., 2001).

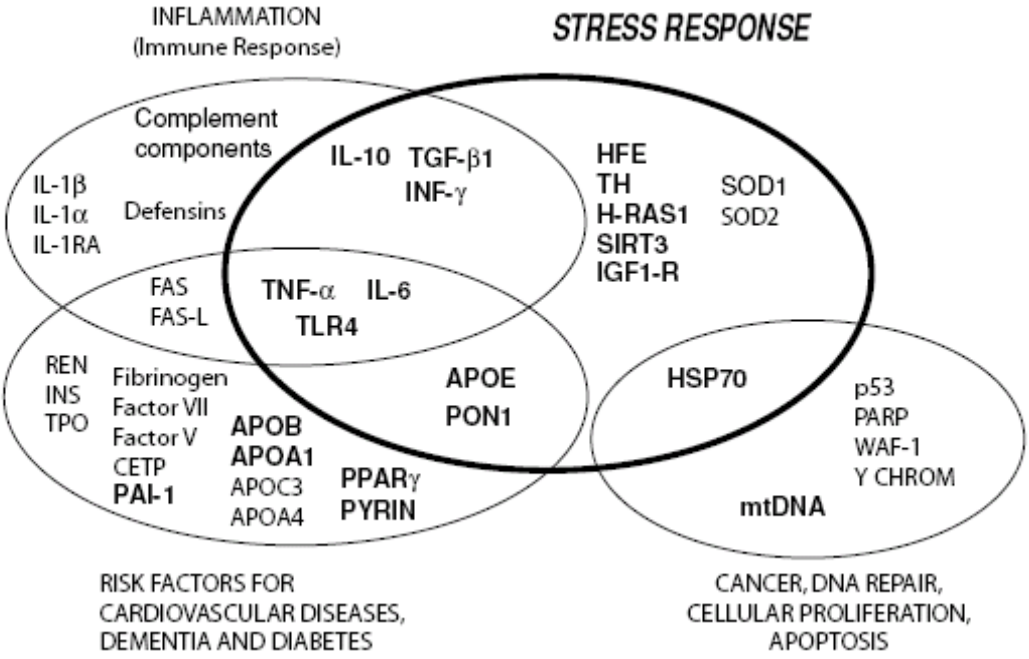



**Fig. 1.5:** Candidate genes in association studies related to longevity (from "The genetics of human longevity", Capri *et al*., 2006)

3. **Studies on chromosome 4**: a genome wide scan, with "Affected Sibpairs Analysis" (ASP), has been performed on 308 individuals belonging to 137 sibships demonstrating exceptional longevity and a significantly statistical association (P=0.044) between longevity and the genetic marker D4S1564 microsatellite on chromosome 4, that is underrepresented among long-living individuals when compared with younger controls (Puca *et al*., 2001). In order to identify gene variants affecting life span, the same group performed an haplotype-based fine-mapping study of the fragment. What they found was a haplotype marker within microsomal transfer protein (MTP) as a modifier of human life span. Nebel *et al*. (2005) performed a study on 1039 unrelated subjects of German ancestry with a mean age of 98 years, 373 of them were centenarians. They found that the MTP haplotype was over-represented only in controls, in comparison with studies reported in the literature.

   Beekman *et al*. (2006) investigated the linkage to 4q25 in 164 nonagenarian sibships of Leiden Longevity Study (LLS), comparing the MTP -493G/T and Q95H allele and haplotype frequencies in the Leiden Longevity Study (379 nonagenarians, 525 of their offspring, and 251 partners of their offspring) and in the Leiden 85+ Study (655 octogenarians and 244 young controls). This study

population was followed for at least 7 years, providing the opportunity to perform also prospective analyses using the longitudinal data. They found neither evidence for linkage at 4q25 nor association of the MTP locus with longevity in nonagenarian individuals. After a meta-analysis of all previous studies the authors concluded that the association in U.S. Caucasians may have its source in admixture of the U.S. control population rather than in the genetic effect of the locus on exceptional longevity.

4. **Studies on chromosome 11**: it is becoming more and more evident that chromosome 11 could be involved in human longevity because several studies have shown a large number of variants in such a region. The 11p15.5 chromosomal region, about 2.8 Mb, is of particular interest as it encloses five genes (HRAS1, SIRT3, TH, INS and IGF2) and its variability of which was found to be associated with life extension by association studies. These genes are homologous of genes that modulate lifespan in model organisms (De Benedictis *et al*., 1998; De Luca et a., 2002; Bonafè *et al*. 2002, Rose *et al*., 2003; De Rango *et al*., 2008). A recent study was perfomed by Lescai *et al*. (2009) who scanned the area in four European sample groups for a total of 1321 centenarians and 1140 younger subjects, matched for ethnicity and geographical origin, with a set of 239 SNPs. They didn't confirm the earlier findings of the literature because no significant results ($P<0.05$) have been found on the earlier associated loci (ie, TH, IGF2, INS and HRAS1). They performed a meta-analysis on the SIRT3 SNP data and the other 229 markers including 2461 samples. For SIRT3 no positive association was found except for one SNP having a significant effect (rs939915); for the other 229 markers, six SNPs have been found significant for the frequent genotype (rs4073591, DEAF1-rs4073590, KRTAP5-6-rs11040489, rs4930001, TSPAN32-rs800140 and rs16928120).
5. Studies on European centenarians suggest an association between hereditable variants of **mitochondrial DNA** (mtDNA) and longevity. In particular the group leaded by De Benedictis (De Benedictis *et al*., 1999) collected and analysed individuals selected for successful aging and longevity (212 subjects older than 100 years and in good clinical condition) and a sample of 275 younger individuals (median age 38 years) carefully matched as to sex and geographic origin (northern and southern Italy). They carried out that mtDNA haplogroup frequency distribution was different between centenarians and younger

individuals (P=0.017); and that the frequency of the J haplogroup was notably higher in centenarians than in younger individuals (P=0.0052). Further data showed that this distribution is population-specific, being also present in long-lived subjects from Ireland (Ross *et al*., 2001) and from Finland (Niemi *et al*., 2003), but not in subjects coming from southern Italy (Dato *et al*., 2004). Morever, a C150T mutation in mtDNA is much more frequent in centenarians than in younger subjects. This mutation is so important because it causes a remodelling of the replication origin at position 151, instead of 149 and can be inherited (as polymorphism) or acquired during life (mutation) (Rose et al, 2007).

# Chapter 2

# THE MITOCHONDRIAL GENOME

## 1. Mitochondrial genetics

Mitochondria are semi-autonomously functioning organelles, harbouring some life important cellular process, like apoptosis, regulation of cell cycle, lipidic metabolism, the citric acid cycle, the respiratory chain and the oxidative phosphorylation (OXPHOS), the last one generates approximately 90% of cellular adenosine triphosphate (ATP) (Wallace, 1997).

Mitochondria are involved in aging process and, as nucleous, contain a resident genome having unique genetic features, in fact it is independently replicated, translated and transcribed, is maternally inherited, because the cytoplasmic location of mitochondria, does not recombine and undergoes replicative segregation during both mitosis and meiosis (Anderson et al, 1981; Taanman, 1999). MtDNA comprise 0,1-1.0% of the total DNA in most mammalian cells, each organelle contains 2-10 copies of mtDNA molecules and each cell contain several mitochondria, so each human cell contains more than 1000 copies of mtDNA (Penta et al, 2001).

Another characteristic is that the human mtDNA is a supercoiled, double-stranded circular molecule of 16,569 base pairs (bp) composed of a control region or D-loop (displacement loop) and a coding region. Introns and intergenic sequences are absent so genes overlap. MtDNA contains 37 genes coding for 13 polypeptides of the mitochondrial electron respiratory chain, 22 tRNAs and 2 rRNAs.

The mitochondrial proteome consists of 1500 polypeptides, whose only 13 codified directly by mtDNA, and in particular by mitochondrial ribosome or *mitoribosome.*

The 13 mtDNA-encoded polypeptide genes are translated on mitochondrial ribosomes and all are structural subunits of OXPHOS enzyme complexes. These include 7 (ND1, 2, 3, 4L, 4, 5, 6) of the 46 polypeptides of complex I (NADH dehydrogenase), one (cytochrome b, cytb) of the 11 polypeptides of complex III (bc1 complex), 3 (COI, II, III)

of the 13 polypeptides of complex IV (cytochrome c oxidase), and 2 (ATP 6 and 8) of the 16 proteins of complex V (ATP
synthetase). The nDNA codes for all other mitochondrial proteins including all four subunits of complex II (succinate dehydrogenase), the mitochondrial DNA polymerase γ (POLG) subunits, the mitochondrial RNA polymerase components, the mitochondrial transcription factor (mtTFA), the mitochondrial ribosomal proteins and elongation factors, and the mitochondrial metabolic enzymes. These factors are recognized thanks to a leader sequence in the N-term and then transported in the mitochondrion.

MtDNA undergoes replication utilizing a different origin for each of the two DNA strands, the purine abundant heavy strand (H-strand), and the pyrimidine abundant light strand (L-strand). In addition to its mRNA, rRNA, and tRNA genes, the mtDNA comprises a 1,121-np control region (CR), also called displacement loop (D-loop, nt 16024-576) (Taanman, 1999). This CR includes the L- and H-strand promoters ($P_L$ and $P_H$); their mitochondrial transcription factor A (mtTFA) binding sites; the downstream conserved sequence blocks (CSB) I, II, and III; and the origins of H-strand replication ($O_{H1}$ and $O_{H2}$) (Shadel and Clayton, 1997) (**Fig. 2.1)**.

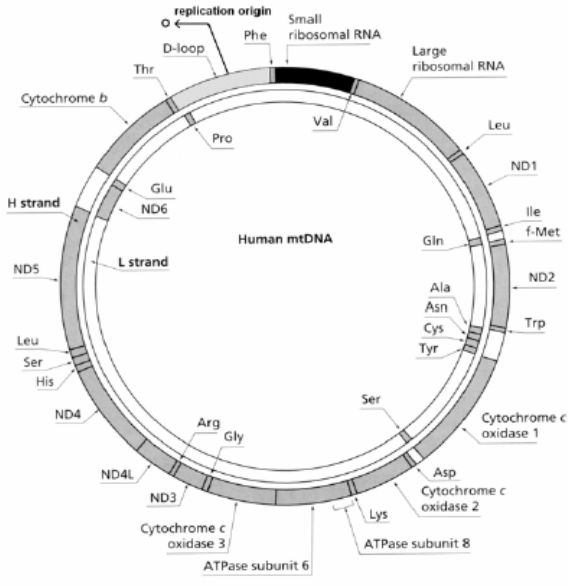



**Fig. 2.1:** Scheme of the mitochondrial DNA molecole organization. ND1, ND2, ND3, ND4, ND4L, ND5, ND6 are subunits of complex I; COI, COII, COIII are subunits of complex IV; cyt b is the complex III; genes codifying tRNAs are defyined with a single letter of the corrispondine amminoacid.

Unlike nuclear DNA, mtDNA may replicate more than once during each cell cycle, or not at all, and may undergo replication in non-dividing cells.

The human mtDNAs are strictly maternally inherited, thus the offspring mtDNA is identical to maternal mtDNA (except some mutations). When a mutation arises in an mtDNA, it creates a mixed population of normal and mutant type of mtDNAs, a state known as heteroplasmy. When a heteroplasmic cell divides, the two types of mtDNAs are randomly distributed into the daughter cells, which drifts toward either pure mutant or wild type. Over time, this replicative segregation results in segregation of the mutant mtDNAs into pure mutant or into normal population, this event is called homoplasmy (Wallace, 2007).

However, there are highly conserved regions, while other regions show a very high variability.

The overall mutation rate is ten to twenty times higher than the nuclear genome and most of the replacements that have been found in mitochondrial DNA are point mutations, with a strong preponderance of transitions (substitutions purine → purine or pirimidine → pirimidine) respect to transvertions (sostitutions purine → pirimidine or vice versa).

It has a very high mutation rate, due to the lack of histones and to an inefficient repair system. Moreover mtDNA is located in the mitochondrial matrix, near the mitochondrial respiratory chain, that is a potent source of DNA damaging free radicals. Particularly, the 1121 nucleotides of D-loop region, are the most sensitive to mutagenesis (Chinnery et al, 1999). In addition, mtDNA has a high mutation fixation rate, which explains the high level of mtDNA substitutions.

As a direct consequence of these mtDNA features, there are two levels of mtDNA variability: an hereditary or inter-individual variability, and a somatic or intra-individual variability.

Because of strict maternal inheritance, mtDNAs can only evolve by the sequential accumulation of mutations along radiating maternal lineages and the number of mutation, that differentiate the mtDNA of an individual from that of his ancestor, can be used as a molecular clock, providing a useful and thorny tool to the phylogenetic reconstruction.

This means that the human mtDNA is a molecular archive of the history and migration of women who have passed on to subsequent generations. If a mtDNA mutation arises that is beneficial in a particular environment, it and its descendants will increase in frequency in that environment. Thus, different subsets of the variation in

mtDNA tend to be confined to different regions and different human populations (Torroni, 2006).

## 2. The D-loop

The mtDNA comprises a 1,121-np control region (CR), also called displacement loop (D-loop, nt 16024-576) (Taanman, 1999). This CR includes the L- and H-strand promoters ($P_L$ and $P_H$); their mitochondrial transcription factor A (mtTFA) binding sites; the downstream conserved sequence blocks (CSB) I, II, and III; and the origins of H-strand replication ($O_{H1}$ and $O_{H2}$) (Shadel and Clayton, 1997) **(Fig. 2.2)**. In this particular trait three hypervariable sequences are present, called HVSI, HVSII and HVSIII, characterized by a high level of mutation, higher than mtDNA variability, showing a high number of *hotspot* (Malyarchuk and Rogozin, 2004).

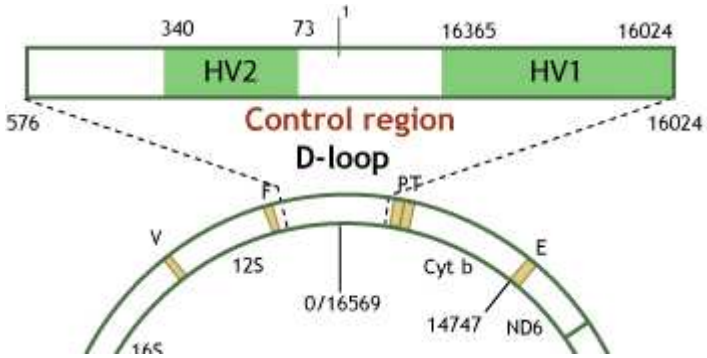



**Fig. 2.2:** Schematic representation of the D-loop region. HV1 and HV2 are hypervariable regions, $O_H$ is the origino f replication of the heavy chain and P are promoters of transcription.

Particularly, the 1121 nucleotides of D-loop region, are the most sensitive to mutagenesis (Chinnery et al, 1999). In addition, mtDNA has a high mutation fixation rate, and the high level of mtDNA substitutions and their relative stability make the d-loop the ideal candidate for the identification and classification of mitochondrial haplogroups (Kivisild *et al*, 2005; Torroni *et al*, 2006).

## 3. Using of mitochondrial DNA in a population's study

The uniparental inheritance is one of the greatest advantages of this marker because it allows us to trace the ancestral information in the DNA of a population based

solely on the maternal contribution and eliminating the effects of recombination, which characterize the nuclear DNA. The phylogenetic trees constructed on the casis of mitochondrial DNA data can therefore be interpreted as genealogies that trace the history of a species and/or a population by female lineage.

The first completely sequenced mitochondrial DNA was obtained in 1981 by Anderson and still we use it as a reference sequence (Cambridge Reference Sequence, CRS; Anderson *et al*., 1981; subsequently recise and corrected by Andrews *et al*., 1999).

The analysis of mitochondrial DNA variability in human populations is expressed at three levels:

- Analysis of D-loop sequenze;
- Analysis of restriction sites polymorphisms (RFLPs) in the coding region;
- Sequencing of the entire genome.

## 4. MtDNA inherited variability: the haplogroups

Testing mutations of the coding region, it is possible to attribute to individuals or people the belonging to specific haplogroups, i.e. monophyletic lines defined by the presence or absence of restriction sites in specific locations of the genome. The members of a haplogroup are descended from a single common female ancestor who first had this particular set of polymorphisms. Haplogroups can be thought as the arms of the family tree that correspond to various human migrations made by people during the expansion in the world.

The mitochondrial genome inherited variability has been studied for many years in relation to human history, on the assumption that the polymorphic inherited variants were neutral from the selection point of view and the populationistic arrangement of human mitochondrial genomes derived solely from the phenomena drift. And now it is established that the polymorphic variants of the mtDNA, that is mitochondrial haplogroups, are not neutral and therefore mitochondrial types have different efficiencies on different functional basic characters. In particular it was noted that there is a correlation between haplogroups and diseases, haplogroups and longevity.

To perform a detailed and thorough study, which can provide not only the variability of the oldest but also the most recent, it is necessary to dissect haplogroups in subhaplogroups, i.e. smaller phylogenetic entities.

All the information necessary to the reconstruction of phylogenetic trees, today in our possession, were found by sequencing a large number of samples; this analysis has

gone through a number of technological and methodological stages and is now in the era of complete sequence.

Even during the nineties emerged a defined phylogenetic picture of the mtDNA thanks to the mtDNA digestion with specific restriction enzymes and the study of the length of the fragments obtained after enzymatic cleavage (restriction fragment length polymorphism-enzyme, RFLP).

This method was applied for the first time by Wallace revealing that the mtDNAs can be classified into small monophyletic lines, i.e. haplogroups. More recent studies have been conducted by Torroni (Torroni et al. 1994) whose studies on 175 European American individuals mtDNAs revealed 117 haplotypes (different sequences). More related haplotypes were grouped into four haplogroups, referred to by capital letters (H, I, J, K).

Summarizing the data obtained from restriction, the main RFLP markers that define the major haplogroups observed in Europe are the following:

- **Haplogroup J**: presence of *NlaIII* site at position 4216, presence of *DdeI* site at position 10394 and absence of *BstI* and *HinfI* site, respectively at positions 13704 and 16065.
- **Haplogroup H**: absence of *AluI* site at position 7025, absence of *DdeI* site at position 10394 and absence of *MseI* site at position 14766.
- **Haplogroup I**: absence of *DdeI* site at position 1715, absence of *HaeII* site at position 4529, simultaneous presence of *AvaII* site at position 8249 and absence of *HaeIII* site at 8250, presence of *AluI* site at 10032, presence of *DdeI* site at 10394, simultaneous presence of *BamHI*/*MboI* sites at position 16389 and absence of *AvaII* site at position 16390.
- **Haplogroup K**: simultaneous presence of *HaeII* site at 9052 and *HhaI* at 9053, presence of *DdeI* site at position 10394, presence of *HinfI* site at position 12308.

The RFLP analysis of 49 Finns, 37 Swedes and 48 Tuscans mtDNAs has revealed the presence of additional European haplogroups that have been called T, U, V, W and X (Torroni et al. 1996):

- **Haplogroup T**: presence of *NlaIII* site at position 4216, presence of *BfaI* site at position 4917, absence of *DdeI* site at 10394, simultaneous presence of

*BamHI* site at 13366, MboI at 13367 and absence of *AvaII* site at 13367, *MspI* at 15925 and presence of *AluI* site at 15606.

- **Haplogroup U:** absence of *DdeI* site at position 10394, presence of *HinfI* at position 12308.
- **Haplogroup V**: absence of *NlaIII* site at position 4577, *DdeI* at 10394, *MseI* at 14766.
- **Haplogroup W**: simultaneous presence of *AvaII* site at position 8249 and absence of *HaeIII* at 8250 and 8994, *DdeI* at 10394.
- **Haplogroup X**: absence of *DdeI* site at 1715 e 10394, presence of *AccI* site at position 14465.

So the nine haplogroups listed above, along with a few representatives L African haplogroup and M Asian haplogroup, collect almost all of the variation of mtDNA in Europe.

Until recently three African haplogroups (L1, L2, L3), seven Asian (A, B, C, D, E, F, G) and nine in Europe (H, T, U, V, W, X, J, I, K) (Torroni et al, 1996) were identified, but now great strides are made in the phylogenetic analysis of mitochondrial haplogroups whose studies have provided a variation of the internal monophyletic haplogroups by dividing them into smaller subsets called **subhaplogroups** (Kivisild et al. 1999, Macaulay et al, 1999, Torroni et al. 2001).

The latest studies are examples of the ability of genetics to make inferences about the origin of humans and studies of evolutionary biology. These analysis show us that the data provided by mtDNA can be used not only to evaluate models based on direct inquiry of ancient material offered by other disciplines, but also to identify relationships previously never assumed among very different populations and different geographical areas. Thus, human genetics can now directly promote the development of new areas of research in paleontological, archaeological, linguistic and historical areas, not to mention the important discoveries in medicine, which led to relate the mitochondrion and its functions to multiple diseases and disorders.

Alongside the increased phylogenetic definition, even the nomenclature of the various haplogroups was amended and supplemented: haplogroup R0 replaced the pre-HV, pre-V replaced the HV0 and HV0a gathers everything that is not V, and all that is in HV0 branch (Torroni et al, 2006).

## 5. The phylogenetic trees

MtDNA variations studies in human populations have identified peculiar mutations that are neutral or near neutral, avoid elimination by selection and may thus become prevalent through genetic drift.

Hence, mutations which occurred tens thousands of years ago will nowadays be present in high frequency, population and continent specific mtDNA polymorphisms, creating groups of related mtDNA haplotypes, or haplogroups sharing a specific set of stable polymorphic restriction sites (Torroni and Wallace, 1994; Wallace 1995; Wallace, 1994). The classification of mtDNA haplogroups is based on information gained from RFLP analysis of the coding region and from the nucleotide sequence of the hypervariable segments I (HVSI) in the control region (Torroni et al., 1996).

In all the studies and construction of phylogenetic trees it is considered a mutation all the differences between the sequence to be analyzed and the sequence of Cambridge, and thanks to these mutations it is possible to identify the haplogroup and its subclass. Haplogroups are coded with capital letters and subclusters with a running number (Ballinger et al., 1992; Torroni et al., 1996).

According to the most recent classification, the phylogenetic tree of mitochondrial DNA originated in Africa about 150-200 thousand years ago (**Fig. 2.3**); African haplogroups are the oldest and fall into four main haplogroups: $L_0$ (the oldest), $L_1$, $L_2$ and $L_3$ (the younger). $L_0$, $L_1$ and $L_2$ represent about 76% of all mtDNA sub-Saharan Africa and are defined by the HpaII restriction site in position 3592. All non-Africans mtDNAs are descended from the branch $L_3$ and they are divided into the superfamilies M and N about 65,000 years ago.

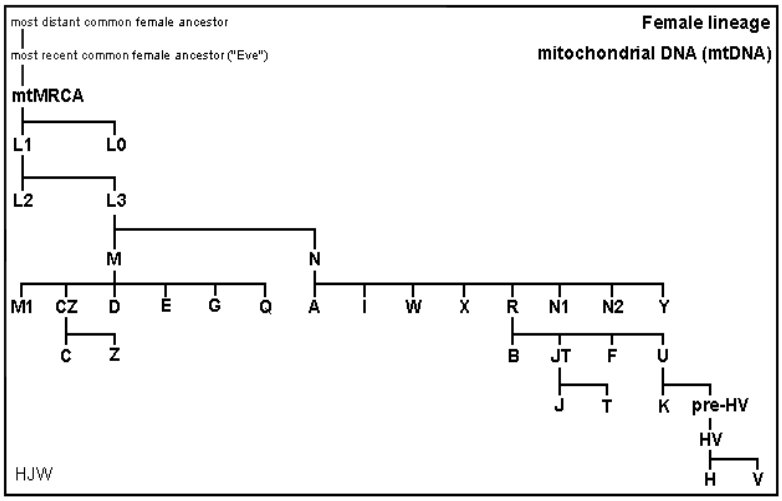



**Fig. 2.3**: Representation of the most important haplogroups. It is noticed the diramation of European haplogroups M and N from L3.

Haplogroups H, I, J, N1b, T, U, V, W are characteristic of populations of European descent, while macrohaplogroups A, B, C, D are present in the New World and Asia; G, Y, and Z are mainly present in Siberia (Mishmar et al., 2003) and the haplogroup X present with low frequency in northern Africa, western and central Asia, Europe and North America (Reidla et al, 2003).

The eight most common European haplogroups (H, U - which also includes K - J, T, V, X, I, W) all originate from the branch N. In Asia there is a multitude of different haplogroups, which includes more than thirty subdivisions, most of which belong to class M. N originated from haplogroups A, B and F while C, D and G separated from M haplogroup (**Fig. 2.4**). Of all these haplogroups, A, C and D were the first to overcome the Bering Strait about 20-30 thousand years ago, followed by haplogroup X 15,000 years ago and B 14,000 years ago.

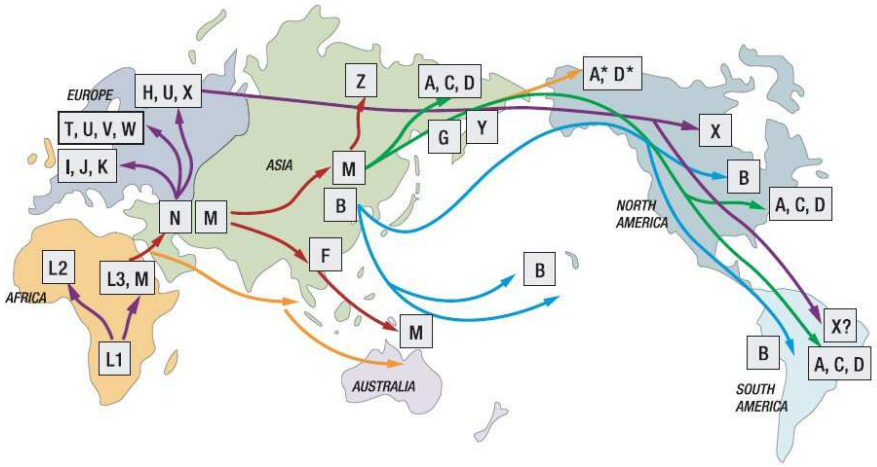



**Fig. 2.4**: Diagram outlining the migratory history of female mtDNA haplogroups. All mtDNAs arose in Africa with the first L0 branch, followed by lineages L1, L2 and L3. In the North L3 gave origine to M and N, which succeeded in leaving Africa and colonizing Eurasia. In Europe, N gave rise to the H, I, J, Uk, T, U, V,W, and X haplogroups. In Asia, M and N gave rise to a diverse range of mtDNA lineages including A, B, and F from N and C, D, and G from M.

Also today it was recognized that H and V, as well as J and T, form lines sisters (Umetsu et al., 2005) and K is a subgroup dell'aplogruppo U (UK).

## 6. Geographic selection theory

The geographic selection theory is the most recent theory that seeks to explain the origin of the various polymorphisms and their distribution; it hypothesizes that the polymorphisms, and therefore the protein variants, have been selected by climate

adaptation (Mishmar et al., 2003) and confer a genetic advantage for survival. In fact, in the case of efficient respiratory chain, most of the proton gradient developed by the various mitochondrial complexes is converted in ATP, whereas in the case that some mutation makes this process less effective, the proton gradient is dissipated, producing heat. Such mutations would be a selective advantage for populations living in cold climate regions (Wallace, 2005).

Other researchers (Kivisild et al., 2006) assume that if a polymorphism has an adaptive function, then it must be placed in a more ancient tree of mtDNA. On the contrary an excess of non synonymous mutations is characteristic of a younger line of the phylogenetic tree.

Studies about proteins encoded by mitochondrial DNA have shown, in favor of this thesis, that the amino acid sequence of the ATP6 gene is highly variable in the Arctic and it is highly conserved in the tropics and temperate zones; cytb is hypervariable and conserved in temperate in the tropics and the Arctic; COI is variable in tropical while it is kept in temperate zones and in the Arctic (Wallace, 2005; Raule et al, 2007).

These polymorphisms are therefore widely used to identify the race an individual belong and with the maternity test in forensic medicine, thanks to the unique features of mitochondrial DNA: maternal transmission, a high number of copies per cell, high-mutation and the absence of recombination (Yuasa and Umetsu, 2005).

## 7. The importance of "Mitochondrial Eve"

A comparison of mitochondrial DNA belonging to human races of different ethnicities and regions suggests that these DNA sequences have evolved molecularly from the sequence of a common ancestor. On the assumption that an individual inherits mitochondria only from his mother, this finding implies that all human beings have a female descent deriving from a female that researchers have named Mitochondrial Eve.

Based on the molecular clock technique, which correlates over time with genetic drift observed, it is believed that Eve has lived about 150,000 years ago. Although the name suggests the biblical Eve, the mitochondrial Eve was not the only female of her time in fact it is assumed that there were more than 20,000 individuals of the species of Eve. Only Eva gave rise to a single unbroken line of daughters that persists today. Thus considering how many and which mutations characterize an individual, it can be traced back to the genetic history of her female ancestors: the number of mutations separating the two individuals is indicative of the temporal distance between them and the

common ancestor, while the analysis of haplogroups allows us to reconstruct displacements of ancient men, or better women, across continents and regions of the world.

The phylogeny suggests that the first woman lived in Africa and that living beings, whose mitochondrial lineages branched first, are those of indigenous Africans. All other indigenous peoples of other continents have branched from the African.

All descended from Africans, some of them migrated out of Africa to populate the rest of the world. Because the mitochondrial Eve represents the root of the mitochondrial family tree, she must have lived in Africa before the exodus. Both archaeological discoveries, and anthropological examinations on the skull, as well the latest information generated by DNA research, confirms the theory of "**Out of Africa**". Obviously, the mutations accumulated in hundreds of thousands of years, represent the key to the reconstruction of our genetic history and the identification of mitochondrial lineages.

From the historical point of view the genetic data suggest that modern man's passage from East Africa to Asia probably occurred in two ways. One route took him to India and South-east Asia and from there further splits: to south, ie Australia, and to north, ie China and Japan.

The northern route took him through the Middle East, Persia, Central Asia and from there expanded in several directions: towards west (Europe) and to the east and north-east (Siberia and America, no later than 15,000 years ago, through Beringia, strip of land covering the present Bering Strait).

Population growth was a bit slow, also affected by the alternation of more or less favorable climatic junctures. The end of the last glacial period, about 14,000 years ago, was followed by a period of climatic stability. From 13 to 10,000 years ago in various parts of the world agriculture and pastoralism began to grow, providing significant new sources of food to groups of men devoted to hunting and gathering wild fruits. The emergence of agriculture involved a considerable development of new tools, the emergence of a new era in human history, the Neolithic Age.

## 8. Correlation between haplogroups and longevity, haplogroups and pathologies

The analysis of mtDNA haplogroups is currently providing new insights into the role of mtDNA-inherited variability in several complex traits like aging and neurodegenerative disease. It was observed that in Italian centenarians mtDNA haplogroup J was overexpressed (De Benedictis et al., 1999), suggesting a protective role for this mtDNA variant against aging and this observation was confirmed in a study on Finnish (Niemi et al., 2003). Other studies refute this association (Dato et al, 2004 and Ross et al, 2001). In Japanese centenarians a sublineage of haplogroup D was more frequent (Tanaka et., al 1998; Tanaka et al., 2000).

We have previously argued that mitochondrial DNA has a high rate of mutation which leads not only to the accumulation of a wide range of sequence polymorphisms that define haplogroups and subhaplogroups, but also to mutations that cause diseases. These diseases arise particularly in tissues and organs with high energy demand: the central nervous system, heart, skeletal muscle and heart, just to name a few. In addition to a wide spectrum of pathological changes, there are many other mtDNA polymorphisms that, alone or in combination with other polymorphisms appear to modulate the risk of complex diseases. Mutations characterizing haplogroups and subhaplogroups and considered "neutral" may instead play a role in some complex diseases.

It seems that certain polymorphisms are associated with one or more haplogroups and that they confer a protection against the onset of these neurodegenerative diseases, while others seem to facilitate their development (Santoro et al, 2006, Reeve et al, 2008). For example, some studies have shown that the J haplogroup could affect the expression of Leber's optic neuropathy (LHON) (Chinney et al, 2001) and in particular that some polymorphisms characteristic of J may increase the penetrance of pathogenic mutations (Torroni et al. 1997) such as, diabetes (Mohlke et al, 2005), and optic neuritis (Reynier et al, 1999).

It has been proposed that J haplogroup can have two different aspects: on one hand it can lead to longevity by minimizing the ROS production and it can favour a decoupling situation as **Fig. 2.5** shows. In fact some authors suggested that this OXPHOS uncoupling characteristic was an ancient mitochondrial adaptation to the cold, as it promoted heat production (Wallace et al, 1999). The same authors suggested that this condition would reduce ROS production, decreasing the gradual mitochondria

degradation thus favouring aging. This could explain the overrepresentation of J among populations of people over 100 years.

On the other hand, J has a general low ATP production that, in addition to environmental factors, causes ah higher proportion of diabetes patients. Furthermore it seems to amplificate some pathologies, as happens in LHON patients, in association with 11778 mutation.

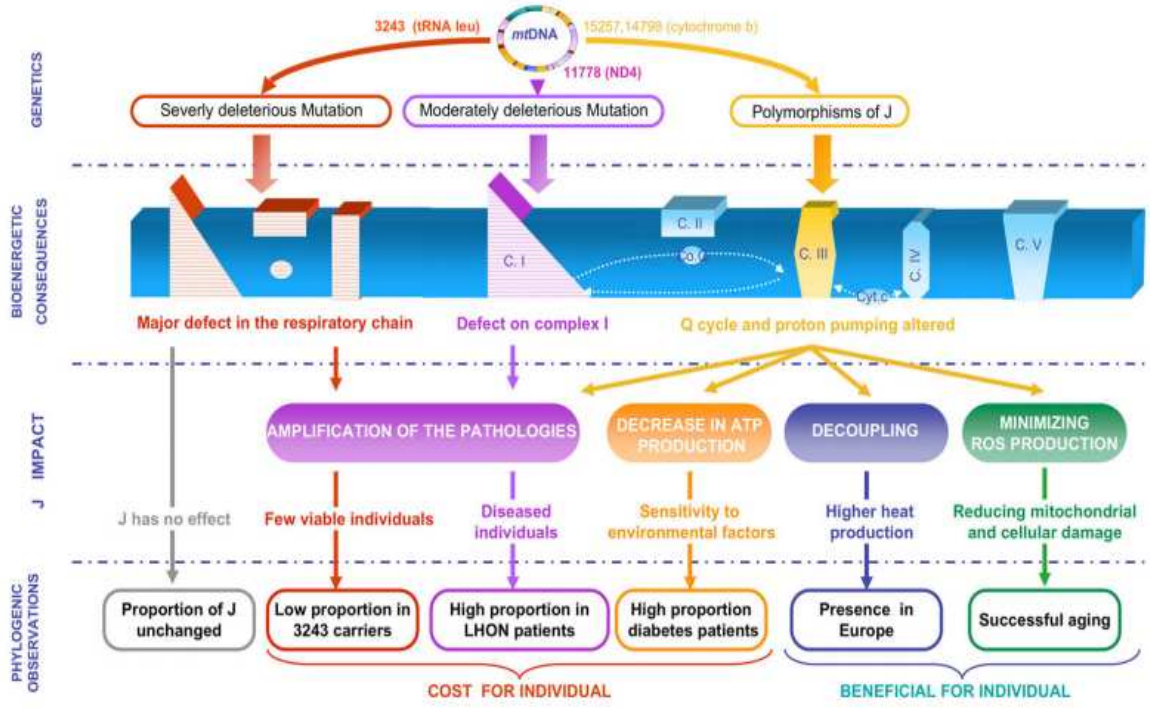



**Fig. 2.5**: Hypothetical mechanism of the "J paradox". The figure summarizes the impact of J on the OXPOS level, the implications on an individual level and the impact on the distribution of populations studied within the phylogeny.

Other associations were studied: haplogroups J and K are underrepresented in Parkinson's disease (van der Walt et al., 2003), and haplogroup T is underrepresentedin AD patients (Chagnon et al., 1999). Hence, it is likely to think that the different mtDNA lineages are qualitatively different from each other, bearing mutations able to improve the OXPHOS efficiency, and consequentely reducing the risk of some pathologies. Furthermore, it has been reported by Ruiz Pesini et al., that haplogroups H and T displayed a significant difference in the activity of complex I and IV of OXPHOS (Ruiz Pesini et al., 2000).

## 9. Point mutation of mtDNA

The mitochondrial theory of aging (Hamilton et al., 2001) proposes that accumulation of mutations in mtDNA and consequent mitochondrial dysfunction are the major contributors to aging and age-related neurodegenerative disease.

The polymorphisms and single base mutations in mtDNA are very probable because of the proximity of a source of ROS and the limited ability to correct from the enzyme deputy to the replication of mtDNA, polymerase γ (Kaguni 2004, Kujoth et al. 2005, Johnson et al. 2001).

Recently, somatic mutations in the Control Region (CR) of the mtDNA have been associated with aging (Coskun et al., 2003). The A189G and T408A CR mutations accumulate with age in skeletal muscle (Wang et al., 2001), and a T150C mutation accumulates in white blood cells (Zhang et al., 2003).

The C150T mutation (where a cytosine (C) is replaced with a thymine (T) in position 150 of the control region CR) is of particular interest in the study of aging.

It has been identified not only in skin fibroblasts but also in lymphocytes and granulocytes of peripheral blood in centenarians and twins. This led to hypothesize a link between mutation and longevity, and between mutation and resistance to stress-induced ROS production (Coskun et al. 2003).

The C150T mutation seems to promote longevity causing the displacement of the origin of replication from 151 to 149 in the control region (Zhang et al. 2003). In fact, the C150T is located near one of the main origins of replication of the H strand of mtDNA (**Fig. 2.6**).

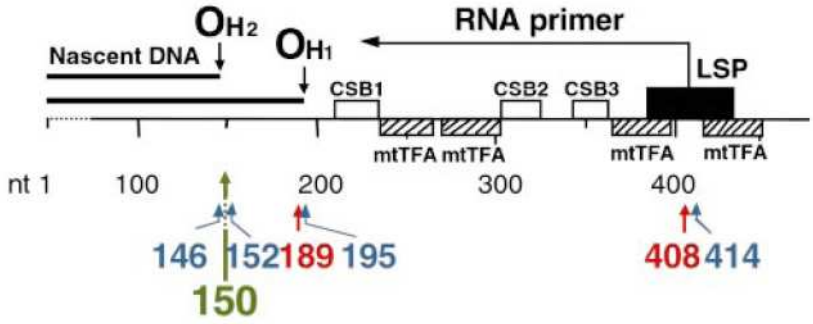



**Fig. 2.6**: Position of the C150T mutation in the D-loop and other specific tissue mutations age-dependent.

Zhang et al. (2003) argue that the somatic event in or near the 150 position could lead to a general remodeling of the entire system of replication, probably controlled by the nucleus. Therefore this remodeling could accelerate mtDNA replication, to compensate the oxidative damage of the same DNA and its functional deterioration occurring in old age. The age-dependent accumulation of point mutations with age, the previously identified in fibroblasts and skeletal muscle in critical sites for mtDNA replication, can be considered part of the remodeling.

The data also showed that C150T variant causes a remodelling of the replication origin at position 151 and can be either inherited (polymorphism) or somatically acquired (mutation).It also may cause an activity decreasing. Coskun *et al*. (2003) suggested that mtDNA haplogroups are likely not neutral and that C150T variant associated with haplogroup J (De Benedictis *et al*, 1999) might have changed oxidative phosphorylation efficiency (OXPHOS) by reducing the activity. Moreover J haplogroup is associated to the presence of at least four mutation near the replication origin, C150T included (Niemi *et al*., 2005).

How might C150T and/or haplogroup J contribute to longevity? The C150T variant or a linked polymorphism in haplogroup J might change OXPHOS efficiency and thus ROS production, reducing oxidation stress and increasing longevity. In fact human mtDNAs radiated out of Africa and colonized the world through successive accumulation of sequential mutations along radiating maternal linage, from mother to daughter. Transitions in mtDNA types between Africa and Eurasia and north and south seem to correlate with latitude indicating that mtDNA diversity has been subjected to climatic selection (Mishmar *et al*., 2003). This means that mitochondria burn calories to make ATP to do work, and generate heat to maintain the inside body temperature. The balance between these two mutually exclusive processes is determined by the OXPHOS efficiency. Highly efficient OXPHOS generates ATP with little waste heat. Less efficient OXPHOS generates more heat producing the same amount of ATP. The most efficient activity is preferable in the tropics, whereas the less would be critical for survival in the arctic. What is more in the arctic area the amino acid sequence of the human mtDNA ATP6 protein is hypervariable, implying that mutations in this mtDNA gene have been important for human adaptation to extreme cold. A reduction in OXPHOS efficiency would also burn more calories. As a result, fewer reducing electrons are introduced from the diet to make ROS. A life-long reduction in mitochondrial ROS stress would, in turn, decrease apoptosis and increase longevity.

Rather frequent is heteroplasmic somatic mutation T414G (**fig. 2.6**), a transversion at position 414 in the D-loop region of the mitochondrial genome, normally present in high proportions (more than 50%) in cultured fibroblasts of individuals with age over 65 years (Michikawa et al., 1999).

Wang et al. (2001), conducting studies on skeletal muscle, discovered that while the 414G mutation is absent, there are two other heteroplasmic mutations, A189G and T408A. Unlike somatic T414G, A189G and T408A polymorphisms are transmitted by germline and thus associated with specific mitochondrial haplogroups (Rose et al., 2002).

These muscle-specific mutations accumulate with age in the mitochondrial genome (Wang et al., 2001, Del Bo et al., 2002), particularly in the control region, which is important for replication. Another interesting observation is that the A189G is located in a position very close to one of the main origins of replication for the heavy chain H (position 191). It is curious to note that two of the main origins of replication have been identified as sites of accumulation of age-related mutations.

# Chapter 3

# The GEHA Project – Genetic of Healthy Aging

## 1. The GEHA Project

Europe is the oldest continent and in the last decade the number of people aged more than 90 years is rapidly rising. The actual proportion of people with more than 90 years is about 50% of the total population. There is a small proportion of elders that apparently undergoes an aging process and they surprisingly appear deprived of the most common age-related disease (cardiovascular disease, stroke, type II diabetes, cancer and dementia). In this scenario it is important to study causes and mechanisms of the aging, that in 2001 the **5-year European Union Integrated Project GEnetics of Healthy Aging (GEHA)** born. The most important aim of this study is to identify genes involved in healthy aging and longevity, which allows individuals to reach advanced old age in good cognitive and physical conditions, without the major age-related diseases.

The GEHA Project represents the strongest and the most competitive consortium ever realized in Europe to investigate genetic bases of human aging process, capable of reaching results that is impossible to obtain in a single European country.

The 5-year GEHA Project was supported through Priority 1 (Life Sciences, Genomics and Biotechnology for Health) of EU's FP6 (Project Number LSHM-CT-2004-503270) and approved by European Commission. The project started on May 1, 2004 and ended ion April 30, 2010 (with a 1-year delay).

The Project can be divided in several steps:

1) Standardization of all procedures: two Informed Consent Form (one for 90+ sibpairs and one for the younger controls), three Questionnaires (for 90+ sibpairs, for the younger control and for the family of the 90+ sibpair, in the National language and in English), phenotypic, genetic database plus a database for mtDNA, set up a collection of biological material (blood samples and cheek swab), a protocol for extraction and for APOE genotyping;

2) the **recruitment** of subjects, divided in 90+ sibpairs and an equal number of controls (TRIOS is composed of at least two 90+ sibs and 1 younger ethnically-matched control subject);
3) the **DNA extraction**, the quality control and shipment to the GEHA Partners in order to conduct genetic analysis (nuclear and mitochondrial genomes);
4) **Genetic analysis** on nuclear and mitochondrial DNA
5) **Statistical analysis** of the collected data with mathematical methods.

The GEHA Project is the largest international collaborative research group for the theoretical and sperimental capacity in the biogerontology and genetic are of aging.

## 2. The objectives of the GEHA Project

The most important objective of the Project is to identify gene which influence healthy aging and longevity, and that protect individuals from major age-related diseases. Accordingly the major goals of the Project are the following:

1. To overcome the **fragmentation of the research** on the genetics of aging in Europe;
2. to **set up a coherent, tightly integrated program** of research that unites demographers, geriatricians, geneticists, epidemiologists, molecular biologists, bioinformaticians and statisticians;
3. to **recruit an unprecedented number of long-living sibpairs (2192)** were both members are aged 90 years of age and more from 11 Euperan countries in 15 geographic areas;
4. to perform a **genome-wide scan on the DNA** of all the recruited subjects (ASP, Affected Sibpairs analysis, or Linkage Analysis);
5. to **recruit 2192 etnically-mached control subjects** (mean age 50-75 years) from the same geographic area in order to fine-map the chromosomal regions identified by Linkage analysis and the three candidate chromosomal regions (see n.8);
6. to **perform bioinformatics, functionale genomics, proteomics and molecular biology** studies on the putative longevity genes and gene variants resulting from ASP and LD mapping;
7. to test whether **ethnically different European population** (including Sardinia and Finland) share the same genes involved in aging and longevity;

8. to study the role played in human longevity by **nuclear regions** (**D4S1564** in chromosome 4, **11p15.5** in chromosome 11, and **ApoE** in chromosome 19);
9. to study the **role of mitochondrial DNA** through the analysis of germline variants (haplogroups) and mutations (C150T) in human longevity;
10. to identify **gender-specificity genes** associated with longevity;
11. to stratify the samples according to **APOE genotype;**
12. to develop **innovative analytical strategies** (based on statistical and mathematical models) capable of combining all the collected data;
13. to perform a **short longitudinal study** to evaluate the importance of genetic factors on the mortality of the recruited 90+ sibpairs;
14. to organize **activities** for young researchers and to disseminate the results of the GEHA activity, discussing ethical implications.

### 3. The GEHA consortium and its bodies

The GEHA consortium is composed of 25 partners of 12 European Countries (Italy, France, Germany, Denmark, Finland, Greece, England, North Ireland, Belgium, Nertherlands, Poland and Ukraina) and 1 Partner from China, who collaborates in an interdisciplinary study (**Fig. 3.1**).

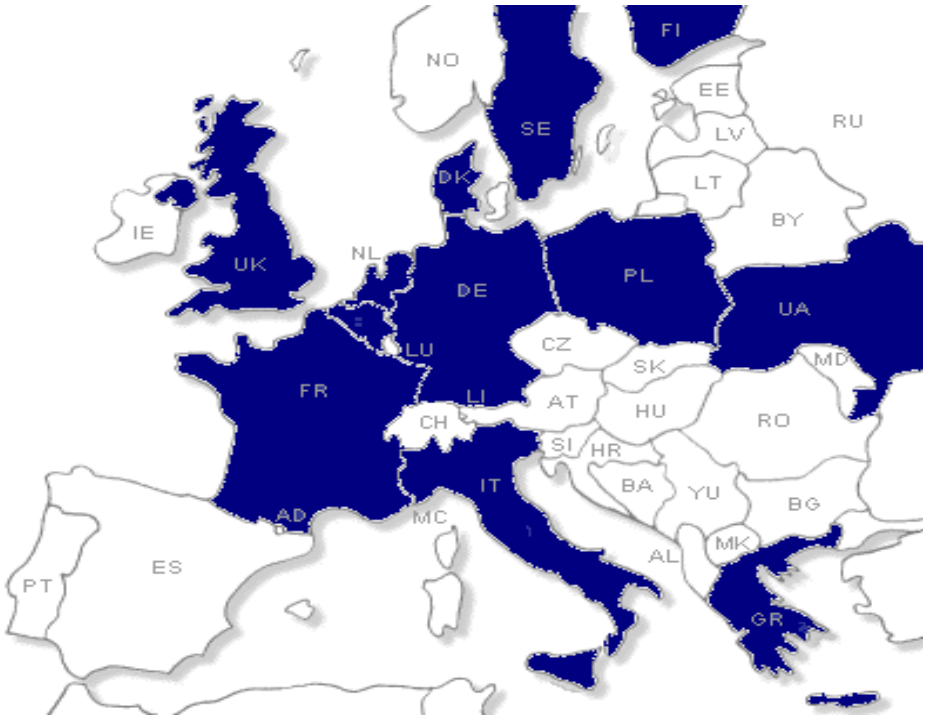



**Fig. 3.1:** representation of the 25 Partners who collaborates in the GEHA project

All these countries have different laws and traditions about privacy, ethical recommendation, access to demographic source, IPR (Intellectual Property Rules) rules, among others. The GEHA project regarding the genetics of human longevity requires the recruitment of very old sibpairs and the donation of blood or other biological material on which carry out the genetic analysis. Moreover, GEHA deals with sensitive issues, which requires attention and care as much as possible. For these reasons it was necessary to standardize all the tools and the fulfilment or ethical requirements both essential to start the first phase of the Project, i.e. the recruitment of 90+ sibpairs and younger controls.

In order to fulfil all the scientific, ethical, financial and IPR requirements, and following the guidelines of the EU, the GEHA project was endowed with a complex organization structure composed by the following bodies:

- **Coordinator** (Professor Claudio Franceschi),
- **Project Manager** (Dr. Alessandra Malavolta),
- **Scientific Manager** (Dr. Silvana Valensin),
- **General Assembly** composed of 25 members (i.e. all the Principal Investigator, one person from each Partner)
- **Steering Committee** composed of 9 members (i.e. the leaders of the 12 WPs),
- **Ethics Steering Group** composed of 3 internal members and 2 external member,
- **External Advisory and Gender Board** composed of eminent scientist from United States and Europe
- **Financial Management Board** composed of 5 members,
- **Legal and IPR Board** composed of 3 members.

The GEHA Consortium was composed as following:

- PARTNER 1: UNIBO **(GEHA Project Coordinator)** University of Bologna, **Italy**, PI: Prof. C. Franceschi
- PARTNER 2: CRLC, University of Montpellier, **France**, I: Prof. J.M. Robine
- PARTNER 3: CAU, Kiel Centre for Functional Genomics, **Germany**, PI: Prof. S. Schreiber

- PARTNER 4: CEPH, Foundation Jean Dausset, **France**, PI: Dr. H. Blanché
- PARTNER 5: ISS, Istituto Superiore di Sanità, **Italy**, PI: Dr. A. Stazi
- PARTNER 6: LUMC, Leiden University Medical Centre, LUMC, **The Netherlands** PI: Prof. E. Slagboom
- PARTNER 7: MPIDR, Max Planck Institute for Demographic Research, **Germany**, PI: Prof. J. Vaupel
- PARTNER 8: NHRF, National Hellenic Researcher Foundation, **Greece**, PI: Dr. E. Gonos
- PARTNER 9: KTL, Tampere School of Public Health, **Finland**, PI: Prof. A. Hervonen
- PARTNER 10: NENCKI, Nencki Istitute for Experimental Biology, **Poland**, PI: Prof. E. Sikora
- PARTNER 11: QUB, Queen's University of Belfast, **UK**, PI: Prof. I. M. Rea
- PARTNER 12: UNICAL, University of Calabria, **Italy**, PI: Prof. G.De Benedictis
- PARTNER 13: IFOM, Institute of Milan, **Italy**, PI: Prof. G. Pelicci
- PARTNER 14: UNISS, University of Sassari, **Italy**, PI: Prof. L. Deiana
- PARTNER 15: UCL, Catholic University of Louvain, **Belgium**, PI: Prof. M. Poulain
- PARTNER 16 : FUNDP, Facultes Universitaire Notre Dame de la Paix, **Belgium**, PI: Prof. O. Toussaint
- PARTNER 17: UNEW, University of Newcastle, **UK**, PI: Prof. T. Kirkwood
- PARTNER 18: SDU, University of Southern Denmark, **Denmark**, PI: Prof. B. Jeune
- PARTNER 19: TAMPERE, National Public Health Institute, **Finland**, PI: Prof. L. Peltonen
- PARTNER 20: R&I, Research Innovation s.r.l., **Italy**, PI: Dr. A. Leon
- PARTNER 21: INRCA, Italian National Research Centre on Aging, **Italy**, PI: Dr. L. Spazzafumo
- PARTNER 22: UAAR, University of Aarhus, **Denmark**, PI: Dr. P. Kristensen
- PARTNER 23: BGI, Bejing Genomics Institute, BGI, **China**, PI: Prof. L. Bolund
- PARTNER 24: EAT, Eppendorf Array Technologies, **Belgium**, PI: Prof. J. Remacle
- PARTNER 25: IG, Institute of Gerontology, Kiev, **Ukraine**,PI: Professor V.V. Bezrukov

### 4. GEHA databases

In order to ensure data privacy to the participant subjects and at the same time ensures full availability of samples, phenotypes and molecular data to all the Partners, GEHA envisages a peculiar centralization of the different types of data collected. The GEHA consortium set up three main databases:

- The **Phenotypic Database** containing clinical and demographic data on the basis of GEHA questionnaires (Odense, Denmark);
- The **Genotypic Database** containing genotyping data (Kiel, Germany);
- **mtDNA Database** containing data related to studies on mtDNA (Tampere, Finland). It allows to assign **haplogroups** automatically**,** to recognize **aminoacidic variations** automatically and to obtain all the **references related** to a **polimorphism** automatically (www.mitomap.org).

Even if these three databases are physically separated, they are strictly interconnected, allowing to all GEHA Partner to perform all types of analysis and protecting privacy of participants. It gives the possibility to introduce independent projects and treat data separately, connecting with the **other GEHA databases** in progress.

### 5. Genetic analysis (nuclear and mitochondrial genome)

Nuclear genome

In the last few years an enormous quantity of data are presented about human genome on millions of new single polymorphism (SNPs) variants. The main goal of GEHA is to perform Linkage analysis but surprising data were obtained with association studies (**Fig. 3.2**). After a preliminary investigation, the scan of the 11p15.5 region was performed, using DNA from centenarians and younger controls. A density of 1 SNP per 10.037 bp was obtained in the 2.4 Mb region. A variation in a gene already implicated in longevity was replicated in the German samples, and other three genes gave positive signals in both German and Central Italians, which render these genes important candidates gehe involved in longevity. Also the Linkage analysis was performed including 15 centers together and separate. Only at chromosome 19 they observed a borderline significant linkage result. Linkage analysis looks for co-inheritance of chromosomal regions with the trai in families. It is not only more powerful than

association studies for identifying rare-risk disease alleles but also it is not influenced by population admixture.

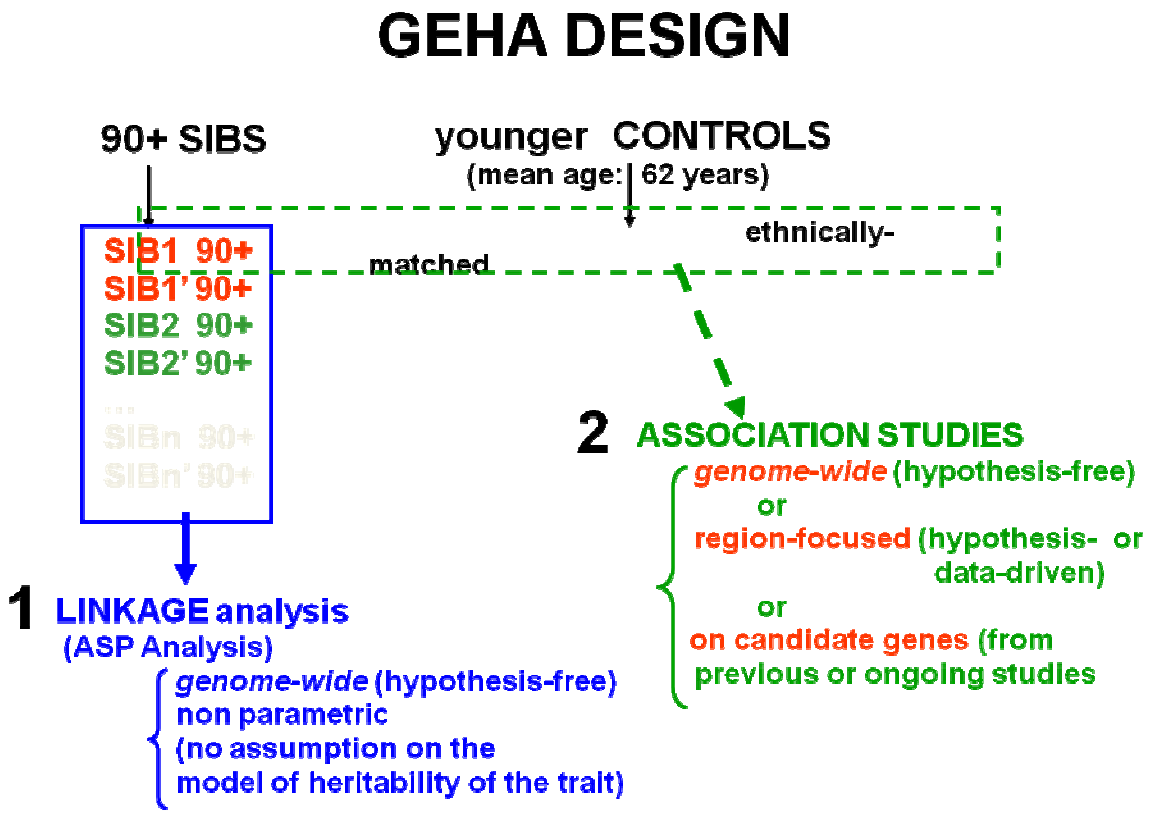



**Fig. 3.2:** graphic representation of the GEHA design. Linkage analysis (1) and association studies (2) were conducted on all recruited samples.

Mitochondrial DNA

About the study of mitochondrial DNA, the main activities were:

1**. mtDNA resequencing.** Different approaches were developed by the GEHA consortium in order to obtain complete sequences. The first methods uses 3 sets of primers producing fragments of about 6kb, sequenced with 39 primers, the second method uses 55 primers for each strand and a kit provided by APPLERA and standardized in the lab of Partner n.1 (UNIBO), and the productes are sequenced by a single universal primers. GEHA sample belong to the specific population of Southern Italy and Greece (sequenced by partner n.1), Finland and Denmark (sequenced by Partner n.23) for a total of 1000 mtDNAs. In particular Partner n. 1 should sequence 660 subjects, divided in 300 nonagenarians (90+) and 330 younger controls.

2. All other GEHA samples are genotyped for **mtDNA haplogroups and subhaplogroups**, using a protocol based on PCR amplification and sequencing of the mtDNA D-loop.

Results are showed in this thesis. The list of GEHA TRIOS (each constituted by two 90+ sibpairs and one younger ethnically matched control, Tab. 3.1 and Fig. 3.3) is the following:

| GEHA center | Country | 2 sibs and one control | 3 sibs and one control | 4 sibs and one control | 5 sibs and one control |
|---|---|---|---|---|---|
| UNIBO | Italy | 182 | 23 | 6 | 1 |
| CRLC | France | 241 | 29 | 3 | 1 |
| CAU | Germany | 93 | 5 | | |
| ISS | Italy | 73 | 2 | | |
| LUMC | Netherland | 148 | 14 | 2 | |
| SDU | DK | 392 | 46 | 5 | |
| NENKI | Poland | 133 | 4 | | |
| NHRF | Greece | 95 | 3 | | |
| UNICAL | Italy | 185 | 7 | 1 | |
| UNISS | Italy | 50 | 2 | | |
| UCL | Belgium | 79 | 1 | | |
| QUB | Ireland | 58 | 4 | 2 | |
| UNEW | UK | 100 | 1 | | |
| TAMPERE | Finland | 127 | 24 | 1 | 1 |
| INST.GERONT | Ukraine | 46 | 2 | | |
| **TOTAL** | | **2002** | **167** | **20** | **3** |

**Tab. 3.1:** list of all recruited samples coming from all Europe. The GEHA center, country and the number of trios are reported.

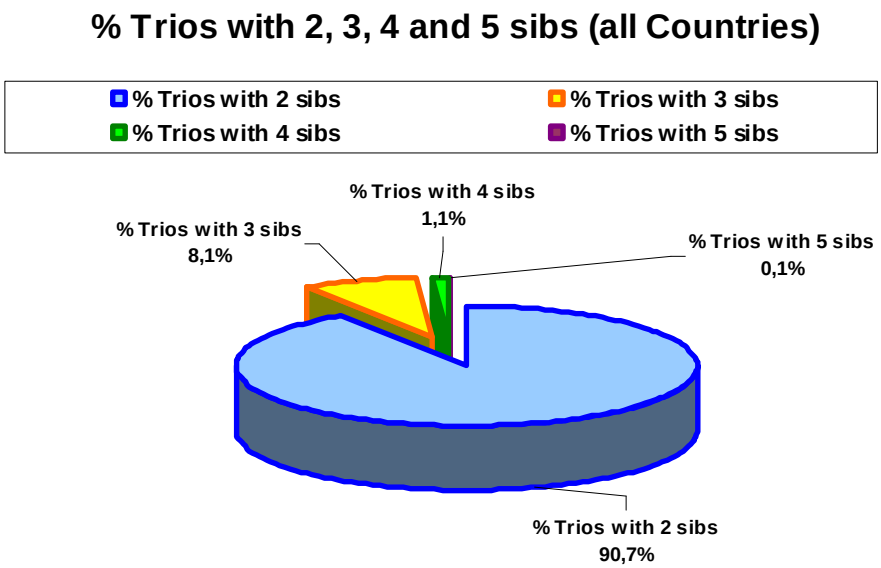



**Fig. 3.3:** a trios is formed by at least two 90+ sibpairs and one younger ethnically matched control. In our data sibships with two sibs are 90.7%, sibships with more than two sibs are 9.3%.

3. Analysis of **heteroplasmy.** A new protocol to highlight the heteroplasmy of the C150T point mutation by DHPLC technology has been developed and standardized. The results show that a correlation between heteroplasmy levels and longevity exists and is highly significant (Partner n.12).

# Chapter 4

## AIM OF THE STUDY

In the present study, we investigate the association of haplogroups, mitochondrial polymorphisms and mutation with longevity. It has been discovered that longevity is essentially a familiar characteristic: centenarians offspring lived more than others, centenarians offsprings are presumed to reach aging in physical and cognitive conditions better than their coetaneous. In this scenario, mitochondrial DNA and its variability plays an important role.

The aim of the Project GEHA is to identify genes involved in healthy aging and longevity, allowing individuals to survive to advanced old age in good cognitive and physical function and in the absence of major age-related diseases, such as type II diabetes, neurodegenerative diseases, cardiovascular diseases and osteoporosis.

The principal aim of this study is to analyse molecular markers that could play a role in healthy aging, proposing:

1. Amplification of control region and RFLP in order to study haplogroups and the relationship with longevity. The technique is applied on subjects coming from 11 European countries adherent to the Project;
2. Standardization of the entire methodology used to analyze mtDNAs, thanks to robotized platform Hamilton MicroLab Star;
3. Statistical analysis of obtained results and research of polymorphisms and mutations associated with longevity and healthy aging;
4. Analysis of complete sequences
5. Cluster analysis and Network analysis

The 2192 recruited candidates represent a great number of subjects, allowing to obtain sufficient data for statistical analysis. Moreover the extension of the recruitment area allow not to limit the analysis to a very restricted haplogroup numbers, exceeding all the geographic problems (because mtDNA is localized in limited geographic area).

# Chapter 5

# MATHERIALS AND METHODS

### 1. Samples

The available DNA samples belong to subjects that are divided in two groups: probands and controls. Proband is the oldest sibling of the sibship that was recruited, should have 90 years or more and should have at least one sibling of the same surname and an age above 90. The control subject is ethnically-matched with proband and should be aged 50-75 years. Subjects come from 11 countries, corresponding to 15 different geographic areas. Since the two members of the sibpair must have the same mtDNA due to maternal inheritance, the Project decided that only the oldest of the sibpairs must be genotyped for mtDNA haplogroup. Our sample is divided as follows:

| ANALYSIS | SIBS | CONTROLS | TOT |
|---|---|---|---|
| Complete sequencing | 637 | 655 | **1,292** |
| D-loop sequencing | 1449 | 1498 | **2,947** |
| **TOT** | **2,086** | **2,153** | **4,239** |

All subject's DNA has been extracted with an automate and standardized procedure (Gentra) in order to guarantee the uniformity of all concentrations by one Partner of the Project (KTL - Helsinki, Finland). For our analysis, 4 μg of genomic DNA have been provided by KTL.

### 2. General protocol

The general plan for the definition of mtDNA subhaplogroups is first the amplification and then the sequencing of D-loop fragment. Then we are able to assign a preliminary haplogroups which must be confirmed by amplification of a specific trait in the coding region, and a digestion with a specific enzyme.

The general scheme of the protocol is as follows:

- Amplification of the D-loop region (about 1500 bp)
- Agarose gel electrophoresis

- PCR clean-up reaction using ExoSAP-IT®
- Sequencing reaction
- Sequence analysis with SeqScape software in order to attribute a preliminary subhaplogroup
- Amplification of a coding region trait specific for each haplogroup
- Digestion using specific restriction enzymes
- To confirm or not the predicted subhaplogroup

PCR amplification

A pair of primers were used for Dloop amplification reaction: 15877F (CAAATGGGCCTGTCCTTGTA) and 770R (GCTGCGTGCTTGATG CTTGT) which are specific for the interested segment of mtDNA (the D-loop, about 1500 bp).

Amplification was performed in a 25 µl volume. Each amplification will contain at final concentration:

| **REAGENTS** | **[STOCK]** | **[FINAL]** | **VOLUME** |
|---|---|---|---|
| Buffer 10X (Invitrogen) | 200 mM Tris-HCl (pH8) 500 mM KCl | 20 mM Tris-HCl (pH8) 50 mM KCl | **2.5 µl** |
| $MgCl_2$ (Invitrogen) | 50 mM | 1 mM | **0.5 µl** |
| dNTP (Invitrogen) | 100 mM | 16 µM | **4 µl** |
| Taq (Invitrogen) | 5 U/µl | 0.04 U/µl | **0.3 µl** |
| Primers: 15877F, 770R | 10 µM | 0.4 µM | **1 µl** |
| DNA | 60 ng/µl | 0.72 ng/µl | **2 µl** |
| Water | / | / | **13.55 µl** |
| Final volume | | | **25 µl** |

The PCR cycling profile (using T1 Thermocycler Biometria) is as follows:

| **Temperature** | | **Cycle number** |
|---|---|---|
| 95°C | 5 min | 1 |
| 95°C<br>68°C<br>72°C | 40 sec<br>40 sec<br>2 min | 35 |

| | | |
|---|---|---|
| 72°C | 12 min | 1 |

Agarose gel electrophoresis

It is very important to verify the amplification on a agarose gel 1.5%. Materials we need are:

- 1.5 agarose gel made up with standard 0.5X TBE (see APPENDIX)
- Ethidium Bromide solution if this has not been pre-incorporated into the gel
- Tracking dye (Bromophenol Blue and/or xylene cyanol + glycerol)

In each well pipet 1.2 µl of the cleaned-up PCR reactions and 8.8 µl (6.2 µl $H_2O$ and 2.6 µl BF6X) to reach a volume of 10 µl. Load in a well a marker which contains a mix of known fragments and used to compare our bands. Run the gel at 150 V for half an hour.

PCR clean-up

The clean-up reaction is important because it purify PCR products from primers or nucleotides not incorporated. For PCR clean-up add 2 µl of Exosap-IT (directly to PCR product) each 5 µl of PCR product (**Fig. 5.1**). We experimented that only 1 µl of ExoSAP each 25 µl PCR product is sufficient and it allows to reduce costs. ExoSAP-IT® (USB, PN 78200) is a mix containing two enzymes: an exonuclease (ExoI) degrades residual single-stranded primers and any extraneous single-stranded DNA produced by PCR, and a shrimp alkaline phosphates (SAP) hydrolyzes remaining dNTPs from the PCR mixture which could interfere with the sequencing reaction.

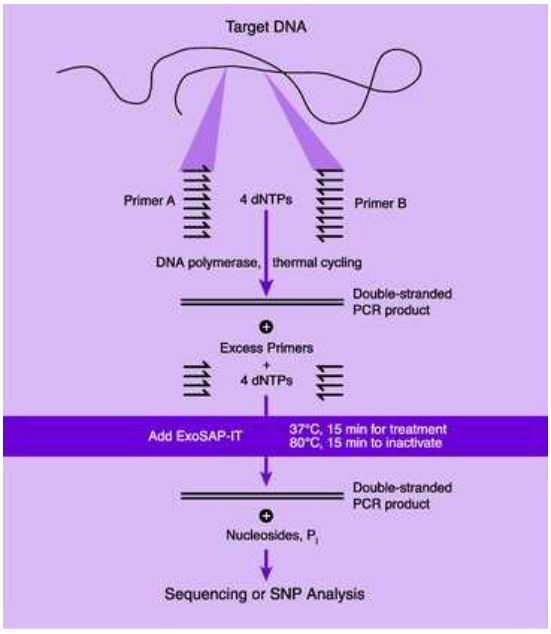



**Fig. 5.1**: PCR clean-up

The PCR cycling profile (using T1 Thermocycler Biometria) is as follows:

1. activation of enzyme at 37°C per 30 minutes
2. inactivation at 80°C per 15'

Sequencing reaction

The sequencing technique has been developed by Torroni's lab of University of Pavia. He has provided us a list of the most used primers to be tested on our samples. To execute the sequencing reaction, we have used three primers, each at concentration of 3.2 pmol/µl.

| Primers | | Sequence | Lenght (bp) | Tm |
|---|---|---|---|---|
| **15973** | For | AACTCCACCATTAGCACCCA | 20 | 60.38 |
| **16522** | For | TAAAGCCTAAATAGCCCACA | 20 | 55.27 |
| **13** | For | ATCACCCTATTAACCACTCACG | 22 | 58.01 |

The d-loop is sequenced in three different fragments, partially overlapping.

The reaction was performed in a 10 ul volume. Each reaction will contain at final concentration:

| REAGENTS | [STOCK] | [FINAL] | VOLUME |
|---|---|---|---|
| Buffer 5X (Applied Biosystems) | 5X | 0.8X | 1.6 µl |
| Big Dye Terminator v.3.1 (Applied Biosystems) | 8 µl in a total volume of 20 µl per reaction | 1 µl | 0.8 µl |
| Primers: 13F, 15973F,16522F | 10 µM | 0.32 µM | 1 µl |
| Cleaned-up PCR product | / | / | 2 µl |
| Water | / | / | 4,6 µl |
| Final volume | | | 10 µl |

The thermal profile is as follows:

| Temperature | | Cycle number |
|---|---|---|
| 96°C | 1 min | 1 |
| 96°C<br>50°C<br>60°C | 10 sec<br>5 sec<br>4 min | 25 |

The BigDye Terminator reagent contains a mix of nucleotide. Free bases that match the template sequence can attach to the new strand's growing (3') end, but among all the free bases swimming in the solution, there are a few having a chemical fluorescent part, i.e. dye. When the coloured bases attach to the growing strand, the extra chemical part keeps the new DNA strand from growing any further. A different coloured dye is attached to each of the four kinds of bases.

A completed sequencing reaction contains an array of coloured DNA fragments. The shortest are the length of the primer plus one coloured base. The longest fragments are usually between 500 and 800 bases long, which is when the sequencing reaction runs out of steam.

Sequencing clean-up reaction

To perform the sequencing clean-up reaction, it is important to add in each well the following reagents:

- 2.0 µl of sodium acetate (NaAc);

- 30 µl Ethanol 100% stored at –20°C, in order to precipitate sequenced fragments;
- Incubate at room temperature for 10 min;
- Spin the plate at 1100 xg for 42 min;
- Remove cover and invert onto paper towel and spin at 100xg for no more than 1 min
- 70 µl ethanol 70% stored at –20°C
- Spin at 1100 xg for 14 min;
- Remove cover and invert onto paper towel and spin at 100xg for no more than 1 min;
- Store the plate at 37°C for at least 1h in order to evaporate all the ethanol;
- Resuspend all wells with 10µl Hi Di Formamide (denaturing agent which allows to maintain a denaturing state for capillary elechtrophoresis);
- Denaturing at 95°C for 2 min.

**3. Analysis at ABI3730 sequencer**

The sample are then loaded on the automatic sequencer (ABI 3730, 48 capillaries). Inside the sequencer the division of the fragment depends on their molecular weight and are separated by a process called capillary electrophoresis. The sequencing machine sets up an electric field; all the DNA moves down through a porous gel toward the positive charge. Shorter fragments of DNA move more quickly through the matrix of the gel than larger fragments do (**Fig. 5.2**).

In the sequencing machine, a laser excites the fluorescent dyes, and a camera detects the lights that the excited dyes emit. One by one, the sequencing machine reads the DNA molecules passing down the gel, and sends the information to a computer. Each nucleotide have a different emission spectrum allowing to identify different bases.

A computer program helps integrate the information from individual sequencing reactions. It spots where fragments overlap, to puzzle the pieces back together.

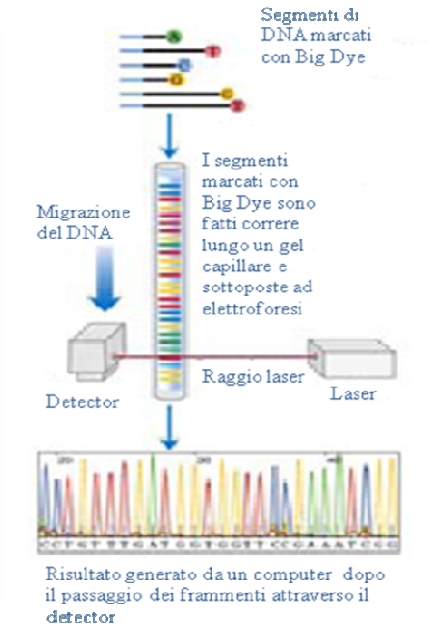



**Fig. 5.2:** principle of sequenze function. The sequencing fragments obtained (a) pass through a laser (b), while a system provides to record the Energy emitted by the excited fluorescent molecole

Many overlapping sequencing reads are needed to reveal the uninterrupted sequence of the original stretch of DNA. On average, every base pair of human DNA will be sequenced nine times. Some stretches of DNA are easier to read and need to be sequenced a little less often to get high-quality sequence. Some stretches need to be analyzed more exhaustively to get finished high-quality sequence.

The sequencer gives data in electropherogram form, through a file (.ab1) that shows four different peaks (green for adenine, red for timine, blue for cytosine and black for guanosine) and the height of each peak depends on the detected intensity.

Now, we are able to read file .ab1 with the software SeqScape v2.1.1 (Applied Biosystems) which allows the alignment, the assembling of the overlapping fragments and a comparison with the Reference Sequence of Cambridge (**Fig. 5.3).**

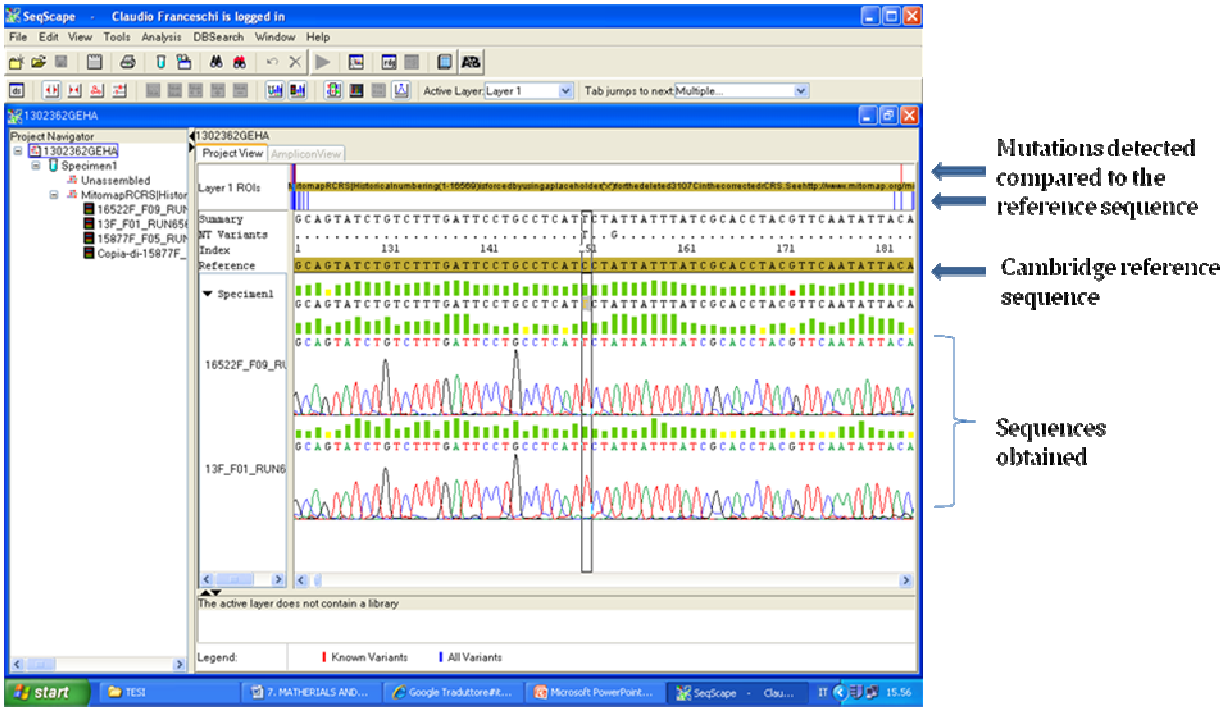



**Fig. 5.3**: Image of an analysis with Seqscape. The Cambridge reference sequence is in brown, coloured peak represent all the four nitrogenous bases and the green bars indicates the quality of the peak relative to the background noise.

### 4. Preliminary assignment of subhaplogroups

After having compared our sequences with the reference one, it is indispensable to define every single subhaplogroups to use specific phylogenetic trees.

The position 73 is the most ancient and whose transition A/G (adenine/guanosine) allows to distinguish between HV branch and U, JT and N branches.

If in position 73 there is an A (equal to Cambridge sequence), our interest is direct to HV branch, if there is a G we will take into account U, JT and N branches (for phylogenetic trees see ADDENDUM).

All mutations must be researched in its specific phylogenetic tree, till it is possible to obtain a more detailed description. Then a specific enzymatic digestion is needed.

The following table summarize the most important position of each subhaplogroup.

**Branch H-V**:

| Subhaplogroup | Mutational sites |
|---|---|
| **V** | 72, 16298, 4580 |
| **Pre*V2** | 72,16298, 15904 |

| | | |
|---|---|---|
| **Pre*V1** | | 72, 16298, 14766 |
| **HV1** | | 16067, 14766 |
| **PreHV1** | | 16126, 16362, 11719 |
| **HV*** | | 16168 |
| **H** | | 7025 |
| → | H1 | 3010 |
| → | H3 | 6776 |
| → | H5 | 456-16304 |
| → | H6 | 239-16362-16482 |

**Branch U**

| **Subhaplogroup** | | **Mutational sites** |
|---|---|---|
| **U1** | | 16249, 285, 13104 |
| → | U1a | 16183, 16189, 4991 |
| **U2** | | |
| → | U2e | 16051, 16129, 16189,16362 |
| **U3** | | 16343, 14139 |
| → | U3a | 16390 |
| **U4** | | 16356, 499, 4646 |
| **U5** | | 16270 |
| → | U5a | 16256, 14793 |
| → | U5b | 150, 14182 |
| **U6** | | 16172, 3348 |
| **U7** | | 16318, 5360 |
| **U8** | | |
| → | K | 16244, 16311, 9055 |
| → | U8b | 16311, 16234, 9055 |
| **U9** | | 6386, 12308 |

**Branch JT**

| **Subhaplogroup** | | **Mutational sites** |
|---|---|---|
| **JT** | | 16126 |
| **J** | | 16069 |
| → | J1 | 3010 |
| → | J2 | 7476 |
| **T** | | 16294 |
| → | T1 | 16163, 16189, 12633 |
| → | T2 | 16296, 14233 |

**Branch N**

| **Subhaplogroup** | | **Mutational sites** |
|---|---|---|
| **N** | | 16233 |
| → | X | 16189, 16278, 153, 14470 |
| → | I | 16129, 16392, 199, 204, 250, 10034 |

| | | |
|---|---|---|
| → | W | 16292, 189, 195, 204, 8994 |
| → | N1b | 16145, 16176G, 11362 |

**Amplification of coding region trait**

Once each subhaplogroup was hypothesed, it is indispensable to confirm this hypothesis trough the presence of particular polymorphisms in the coding region.

Amplification was performed in a 25 ul volume. Each amplification will contain at final concentration:

| REAGENTS | **[STOCK]** | **[FINAL]** | VOLUME |
|---|---|---|---|
| Buffer 10X (Invitrogen) | 200 mM Tris-HCl (pH8) 500 mM KCl | 20 mM Tris-HCl (pH8) 50 mM KCl | 2.5 µl |
| $MgCl_2$ (Invitrogen) | 50 mM | 1.5 mM | 0.75 µl |
| dNTP (Invitrogen) | 100 mM | 16 µM | 4 µl |
| Taq (Invitrogen) | 5 U/µl | 0.04 U/µl | 0.2 µl |
| Specific primers(*) | 10 µM | specific | specific |
| DNA | 60 ng/µl | 0.72 ng/µl | 2 µl |
| water | / | / | variable |
| Final volume | | | 25 µl |

The thermal profile is as follows:

| **Temperature** | | **Cycle number** |
|---|---|---|
| 93°C | 30 sec | 1 |
| 93°C<br>T°C<br>72°C | 15 sec<br>20 sec<br>1 min | 35 |
| 72°C | 12 min | 1 |

The following table (**Tab.5.1**) reports all the pair of primers necessary to produce the amplification, with the relative annealing temperature and the amplificated fragment length. Only for U4 and J2 haplogroups the cycle number of annealing is 44 instead of 35.

| Subhapl | Primers and their sequence | Tann | Length |
|---|---|---|---|

| | | | |
|---|---|---|---|
| **H** | **6892F:** GCAATATGAAATGATCTGCTGC<br>**7860R:** TCGTTGACCTCGTCTGTTATGT | 54,5°C | 242 bp |
| **H1** | **2988F:** CGATGTTGGATCAGGACATCTC<br>**3235R:** CTTAACAAACCCTGTTCTTGGG | 60°C | 248 bp |
| **H3** | **6637F:** TTCTTATCCTACCAGGCTTCG<br>**6807R:** GTGTGTCTACGTCTATTCCTACTGTAAACA | 62°C | 170 bp |
| **R0a** | **11491F:** ACGCCTCACACTCATTCTCA<br>**11750R:** TGCTAGGCAGAATAGTAATGAGGATGTAGG | 64°C | 260 bp |
| **V** | **4381F:** ACCTATCACACCCCATCCTAAA<br>**4683R:** TTAGAAGGATTATGGATGCGGT | 57°C | 302 bp |
| **HV1-HV0*** | **14676F:** TTCTCGCACGGACTACAACC<br>**14996R:** CGTGAAGGTAGCGGATGATT | 55°C | 321 bp |
| **HV0a** | **15796F:** GGACAAGTAGCATCCGTACT<br>**16439R:** GCACTCTTGTGCGGGATATT | 55°C | 643 bp |
| **U1** | **12744F:** CCTATTCCAACTGTTCATCG<br>**13192R:** GAGTGGTGATAGCGCCTAAG | 55°C | 448 bp |
| **U1a** | **4381F:** ACCTATCACACCCCATCCTAAA<br>**5210R:** GGTGGATGGAATTAAGGGTGTT | 59°C | 829 bp |
| **U2e** | **13134F:** AGCAGAAAATAGCCCACTAA<br>**13829R:** AGTCCTAGGAAAGTGACAGCGA | 56°C | 695 bp |
| **U3** | **14103F:** CTCTTTCTTCTTCCCACTCA<br>**14996R:** CGTGAAGGTAGCGGATGATT | 53°C | 893 bp |
| **U4** | **4381F:** ACCTATCACACCCCATCCTAAA<br>**5073R:** TGGTTATGTTAGGGTTGTACGG | 59°C | 656 bp |
| **U5a** | **14676F:** TTCTCGCACGGACTACAACC<br>**14996R:** CGTGAAGGTAGCGGATGATT | 55°C | 321 bp |
| **U5b** | **14158F:** CCGAGCAATCTCAATTACAATATG<br>**14996R:** CGTGAAGGTAGCGGATGATT | 57°C | 835 bp |
| **U6** | **3085F:** ATCCAGGTCGGTTTCTATCT<br>**3693R:** CAGGGCGTAGTTTGAGTTTGA | 54°C | 608 bp |
| **U7** | **5199F:** ATTCCATCCACCCTCCTCTC<br>**5607R:** AGTGGGGTTTTGCAGTCCTT | 56°C | 409 bp |
| **U8b-K** | **8829F:** CCTAGCCATGGCCATCC<br>**9184R:** GGCTTACTAGAAGTGTGAAAAC | 50°C | 356 bp |
| **U9** | **6120F:** ATCATAATCGGAGGCTTTGG<br>**6455R:** GAAGAGGGGCGTTTGGTATT | 54°C | 335 bp |
| **TJ** | **4142F:** GATTCCGCTACGACCAACTC<br>**4365R:** GGATTCTCAGGGATGGGTTC | 57°C | 224 bp |
| **T1** | **12014F:** CTCACCCACCACATTAACAACA<br>**13829R:** AGTCCTAGGAAAGTGACAGCGA | 60°C | 1816 bp |
| **T2** | **13957F:** GGCCTTCTTACGAGCCAAAA<br>**14257R:** TATTGGTGCGGGGGCTTTGTATAA | 60°C | 300 bp |
| **J1** | **2988F:** CGATGTTGGATCAGGACATCTC<br>**3235R:** CTTAACAAACCCTGTTCTTGGG | 60°C | 248 bp |
| **J2** | **7239F:** GCATACACCACATGAAACATCC<br>**7860R:** TCGTTGACCTCGTCTGTTATGT | 59°C | 622 bp |
| **I** | **9893F:** CAAACATCACTTTGGCTTCG<br>**10280R:** GGGTAAAAGGAGGGCAATTT | 53°C | 388 bp |
| **X** | **14379F:** CCATCGCTAACCCCACTAAA<br>**14996R:** CGTGAAGGTAGCGGATGATT | 54°C | 618 bp |
| **W** | **8910F:** CTTACCACAAGGCACACCTACA<br>**9230R:** ATAGGCATGTGATTGGTGGG | 58°C | 320 bp |
| **M1** | **10147F:** ACATAGAAAAATCCACCCCT<br>**10569R:** CTAGGCATAGTAGGGAGGAT | 53°C | 422 bp |

| | | | |
|---|---|---|---|
| **N1b** | **10457F:** TCATATTTACCAAATGCCCCTC<br>**11750R:** TGCTAGGCAGAATAGTAATGAGGATGTAGG | 64°C | 1294 bp |
| **L** | **10498F:** TAGCATTTACCATCTCACTTCT<br>**10930R:** GGAAAAGGTTGGGGAACAGC | 56°C | 433 bp |

**Tab. 5.1**: primers and related sequences, annealing temperature and length of each fragment.

### Enzymatic digestion reaction

The reaction has a final quantity of 25 µl composed of 7 µl of PCR product, 1 U of enzyme (New England) and related buffer 1X, and for a few enzyme also BSA 0.6X.

The list of all enzyme is reported in **Tab. 5.2**, with information about their temperature, the dimension of digested and not digested fragments.

| Sub-hapl | Enzyme | Temp | Digested fragments | Not digested fragments | Cut |
|---|---|---|---|---|---|
| **H** | AluI | 37°C | 137-75-30 | 167-75 | - |
| **H1, J1** | TaqI | 65°C | 228-20 | 248 | - |
| **H3** | NlaIII + BSA | 37°C | 139-31 | 170 | + |
| **R0a** | HaeIII | 37°C | 228-32 | 260 | + |
| **HV1-HV0*** | MseI + BSA | 37°C | 209-90-17-4 | 209-107-4 | - |
| **HV0a** | MseI + BSA | 37°C | 430-175-142-108 | 430-108-67-38 | + |
| **U1** | HinfI | 37°C | 287-89-72 | 287-161 | - |
| **U1a** | AluI | 37°C | 305-186-116 | 491-116 | - |
| **U2e** | HinfI | 37°C | 462-134-99-86 | 561-134 | + |
| **U3** | MboI | 37°C | 610-127-120-36 | 610-156-27 | + |
| **U4** | RsaI | 37°C | 411-179-83-19 | 590-83-19 | + |
| **U5a** | BsrbI | 37°C | 205-115 | 321 | + |
| **U5b** | RsaI+AluI | 37°C | 700-123-23 | 700-146 | - |
| **U6** | MboI | 37°C | 311-263-34 | 574-34 | + |
| **U7** | Tsp509I | 65°C | 158-82-76-57-35 | 240-76-57-35 | + |
| **U8b-K** | HaeII + BSA | 37°C | 224-132 | 356 | - |
| **U9** | HaeIII | 37°C | 140-123-72 | 195-140 | - |
| **TJ** | NlaIII + BSA | 37°C | 149-74 | 224 | + |
| **T1** | AvaII | 37°C | 1200-615 o 738-615-462 | 1816 | - |
| **T2** | Tsp509I | 65°C | 85-73-62-54-26 | 88-85-73-54 | - |
| **J2** | AluI | 37°C | 235-219-167 | 409-219 | - |
| **I** | AluI | 37°C | 200-139-48 | 339-48 | + |
| **V** | NlaIII + BSA | 37°C | 269-163 | 432 | - |
| **X** | AccI | 37°C | 531-86 | 618 | + |
| **W** | HaeIII | 37°C | 205-84-31 | 205-115 | - |
| **M1** | AluI | 37°C | 172-165-85 | 343-85 | + |
| **L** | MnlI + BSA | 37°C | 243-61-58 | 301-61 | - |

| | | | | | |
|---|---|---|---|---|---|
| **N1b** | AluI | 37°C | 441-313-215-142-106-70 | 441-313-227-142-106-70 | - |

**Tab. 5.2**: list of haplogroups and their specific enzymes. Incubation temperature and fragments digested or not are reported.

The digested product are visualized on agarose gel 2.5%, only T1 and T2 subhaplogroups need particolar conditions:

- Agarose gel 0,8% for T1 subhaplogroup (because of the great dimension of fragments)
- agarose gel 4% for T2 subhaplogroup (because of the little dimension of fragments).

If the first digestion does not confirm the hypotesis, it is necessari to revaluate the sequence, to suppose other subhaplogroup and digest the fragment with other enzyme.

## 5. Automated procedure

We also have automated the amplification and sequencing reaction with the MICROLAB STAR workstation, equipped with 16 channels and able to arrenge all the operation, reducing time and manual errors. The platform can prepare reactions for 96 different samples, can distribute the reagent mix (manually prepared) in each well and 96 different DNAs both stored in plates and in 2 ml tubes. The latest must be collocated in specific adaptor with the barcode turned to laser (**Fig. 5.4**).

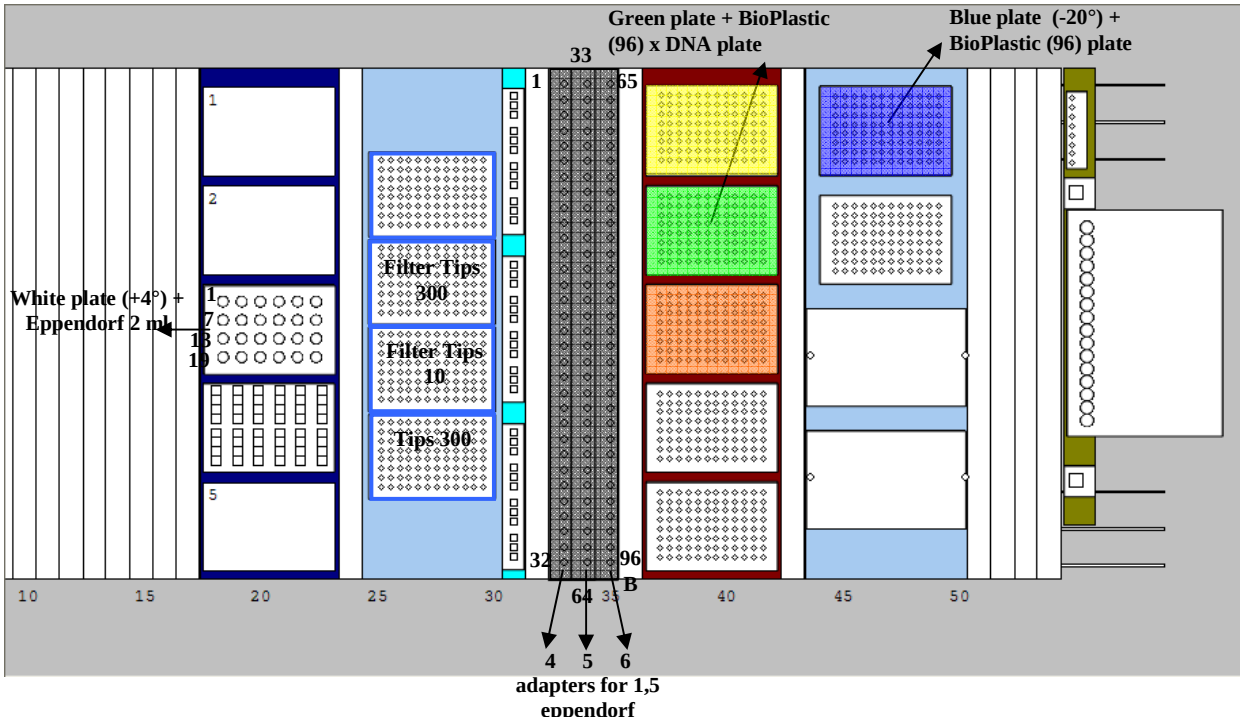



**Fig. 5.4:** robotic platform and the localization of all components and reagents.

The MICROLAB STAR workstation also allows to perform the sequencing reaction (**Fig. 5.5**). As this reaction is performed with three primers for each sample, all the 96 samples are divided in three plates, each carrying 32 samples as follows:

| | Primer 1 | | | | Primer 2 | | | | Primer 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A1 | A2 | A3 | A4 | A1 | A2 | A3 | A4 | A1 | A2 | A3 | A4 |
| | B1 | B2 | B3 | B4 | B1 | B2 | B3 | B4 | B1 | B2 | B3 | B4 |
| | C1 | C2 | C3 | C4 | C1 | C2 | C3 | C4 | C1 | C2 | C3 | C4 |
| | D1 | D2 | D3 | D4 | D1 | D2 | D3 | D4 | D1 | D2 | D3 | D4 |
| | E1 | E2 | E3 | E4 | E1 | E2 | E3 | E4 | E1 | E2 | E3 | E4 |
| | F1 | F2 | F3 | F4 | F1 | F2 | F3 | F4 | F1 | F2 | F3 | F4 |
| | G1 | G2 | G3 | G4 | G1 | G2 | G3 | G4 | G1 | G2 | G3 | G4 |
| | H1 | H2 | H3 | H4 | H1 | H2 | H3 | H4 | H1 | H2 | H3 | H4 |

**Samples 1-32**

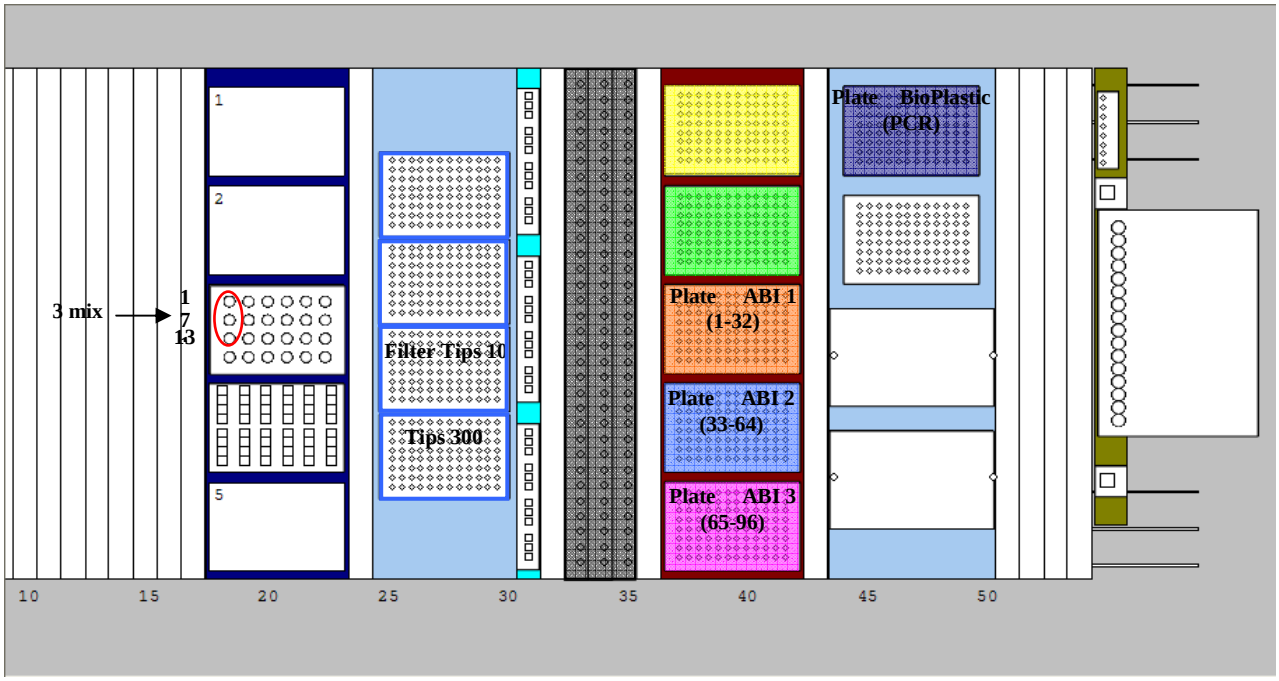



**Fig. 5.5**: robotic platform and the localization of all components and reagents for sequencing.

## 6. Example of RFLP for H haplogroup

In order to analyze the mtDNA genetic variability in our population and their relationship with longevity, different diagnostic sites has been identyfied through restriction analysis of the mtDNA coding region. This restriction allowed attributing at a single sample one specific haplogroup.

An example of H haplogroup analysis (attributed with a first preliminary analysis by SeqScape) is reported.

DNA amplification

As expected, the agarose gel shows an amplified band of 242 bp (**Fig. 5.6**). The PCR program is the same used for D-loop amplification, only annealing temperature has been modified, because it depends on primers.

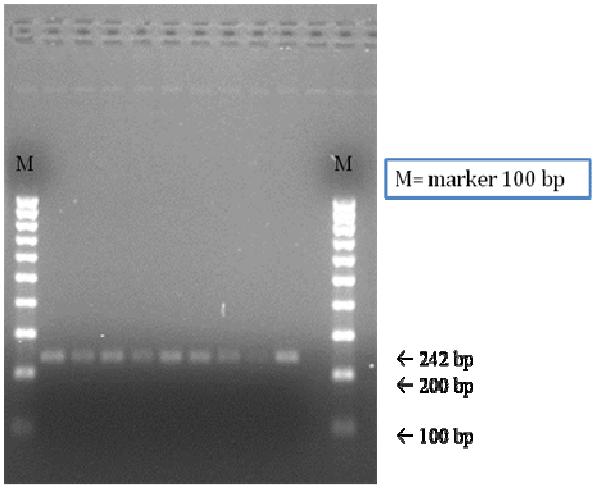



**Fig. 5.6**: Amplification of a coding region fragment specific for H haplogroup.

Digestion of amplified fragments

The digestion of amplified regions is performed using AluI enzyme (specific for the H haplogroup identification) (**Fig. 5.7**).

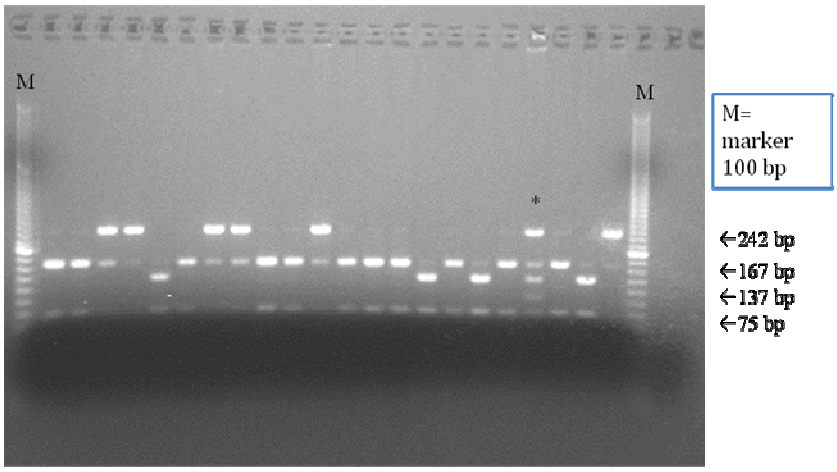



**Fig. 5.7**: AluI digestion of specific amplificated traits for H haplogroup.

It is known that AluI must not cut in 7025 position to confirm the H haplogroup hypothesis. The 242 bp fragment can be cut:

1) In only two fragments (one in 167 bp, and the other in 75 bp) if the enzyme has not digested in 7025 position;
2) In three fragments (on in 137 bp, one in 75 bp and in 30 bp) if the enzyme has digested in 7025 position;

The overwhelming majority is not digested confirming our hypothesis and so it is necessary to continue with other amplifications and digestion with specific enzyme for H1 and H3. Samples which show 137 bp and 75 bp bands do not belong to H haplogroup, but to V haplogroup for example, which presents in position 73 the A nucleotide.

Sample marked with (*) in Fig. 5.2 has probably endured a contamination; it is advisable to repeat the digestion.

**7. Statistical analysis**

The analysis of the sequences to determine haplogroups and somatic mtDNA variability was carried out by SeqScape software v 2.1.1 (Applied Biosystem).

Mitochondrial sub-haplogroups, genotype frequencies and gender were compared between 90+ subjects and their ethnically-matched controls using the $\chi^2$ with Pearson correction or Fisher-exact test.

The mitochondrial subhaplogroups variability envisages at least 80 subhaplogroups, which have been further grouped into 27 subgroups considering their frequencies and phylogenetic relationship. The comparison between each of the 27 mtDNA subhaplogroups in 90+ cases and younger ethnically-matched controls have been computed by applying Pearson's Chi-squared test with Yates' continuity correction.

All analyses were performed separately for women and men, as well as for the total group. Tests for statistical significance were two-sided with = 0.05; we performed also a logistic regression to generate odds ratio (OR) with their associated 95% confidence intervals (CI), to asses odds of carrying each mtDNA haplogroups in cases compared with controls. R statistical software was used for all statistical analyses.

The phylogenetic relation and the distribution of haplogroup J were represented by mathematical method of **Median-Joining Network** (Bandelt et al., 1999). This method combines two algorithms, one is Kruskal to calculate the *minimum spanning trees* and one is Farris euristic to calculate the *maximum parsimony*. The first was used to construct all the phylogenetic trees minimizing the sum of distances among

haplotypes, the latter to create *median vectors* which allow to group trees. We can consider median vectors as eventual haplotype not sampled or ancestral haplotype become extinct. Phylogenetic networks were constructed with **Network 4.1.1.1 program** (http://www.fluxus-engineering.com). Polymorphisms were divided into 4 classes according to their rate of evolution (Hasegawa et al. 1993; Malyarchuk and Derenko 2001; Allard et al. 2002). Fast positions (16093, 16129, 16189, 16311, 16362) were weighted by one, intermediate positions (16051, 16126, 16145, 16168, 16172, 16184, 16192, 16209, 16218, 16223, 16256, 16261, 16278, 16291, 16293, 16294, 16304, 16320, 16325) by 2, and slow positions (all other transitions between 16024 and 16383 as well as 16482) by 4.

**Multidimensional scaling** (**MDS**) is a set of related statistical techniques used in information visualization for exploring similarities or dissimilarities in data. In general, the goal of the analysis is to detect meaningful underlying dimensions that allow to explain observed similarities or dissimilarities (distances) between the investigated objects. The distances used in MDS need not be metric, as non-metric distances such as ranking can be used and the output coordinates are in the standard Euclidean space of the user-chosen dimension. The program used to obtain the distribution gives *"stress"* values, related to the good quality of distances data. The lower is the stress value, better is the adaptation of the reproduced distance matrix. Conventionally, a map is considered acceptable if the stress value is less than 0.1. In this thesis, the not metric MDS analysis was implemented in the sammon function in MASS library of R program.

# Chapter 6

# RESULTS

### 1. Distribution of samples

In this study it has been applied a high resolution analysis, through sequencing of d-loop and restriction analysis of specific markers in the coding region of mtDNA. We explore the mitochondriale haplogroups and the possible association between mtDNA-inherited sequence variation and longevity in a total of 4,239 samples. In 1,292 it has been conducted the entire sequencing, in 2,947 it has been performed the RFLP analysis.

The distribution between the sinlings and controls is shown in the table below.

The **Tab. 6.1** shows males/females proportion with its associated percentage. The haplogroup analysis is conducted on all our subjects and includes also samples whose mtDNA was completely sequenced. In particular our sample are divided as follows:

| ANALYSIS | SIBS | CONTROLS | TOT |
|---|---|---|---|
| Complete sequencing | 637 | 655 | **1,292** |
| Dloop sequencing | 1449 | 1498 | **2,947** |
| **TOT** | **2,086** | **2,153** | **4,239** |

In particular, the 1,292 samples are divided in 1,167 sequenced by the Chinese Partner (BGI) and 125 sequences analyzed in UNIBO.

In our total sample, there is a higher proportion of female (69%) subjects but, in both categories (male and female), proportion is equilibrate (about 50% cases and controls).

Moreover, subjects who can reach 100 years are higher in female than in male (respectively 4.8% e 2.7%). One possible reason is that males are under a much more stronger selection than females, so males can hardly reach the same female ages. This consideration reflects the general proportion in the population.

| | Males (N = 1,327) | | Females (N = 2,860) | |
|---|---|---|---|---|
| | **90+ (%)** | **Controls (%)** | **90+ (%)** | **Controls (%)** |
| **N** | 628 (47) | 699 (53) | 1458 (50) | 1454 (50) |
| **Mean Range** | 94.1 ± 2.5 | 61.7 ± 6.2 | 94.5 ± 2.6 | 61.7 ± 6.2 |
| **Age range** | 89-103 | 43-79 | 88-107 | 49-83 |
| | 17>100 years | | 70> 100 years | |

**Tab. 6.1**: sample distribution and general characteristic of the study participants.

All the analyzed samples come from European countries and from different geographic areas. For example Italy is represented by Bologna, Rome, Sassari and from Calabria region.

It is extremely important for statistical analysis to balance the number of 90+ subjects with their matched controls as reported in **Table 6.2** and **Fig. 6.1**.

| Country | Center name | No. siblings | No. younger controls |
|---|---|---|---|
| Belfast | QUB | 64 | 64 |
| Belgium | UCL | 80 | 88 |
| Bologna | UNIBO | 213 | 214 |
| Germany | CAU | 94 | 96 |
| Finland | TAMPERE | 153 | 145 |
| Montpellier | CRLC | 274 | 275 |
| Newcastle | UNEW | 99 | 100 |
| Denmark | SDU | 428 | 441 |
| Netherland | LUMC | 162 | 167 |
| Poland | NENKI | 129 | 132 |
| Roma | ISS | 75 | 75 |
| Sassari | UNISS | 47 | 52 |
| Ukraina | INST GERONT | 49 | 49 |
| Calabria | UNICAL | 125 | 152 |
| Greece | NHRF | 94 | 103 |
| **Total** | | **2,086** | **2,153** |

**Table 6.2:** distribution of samples and their different countries.

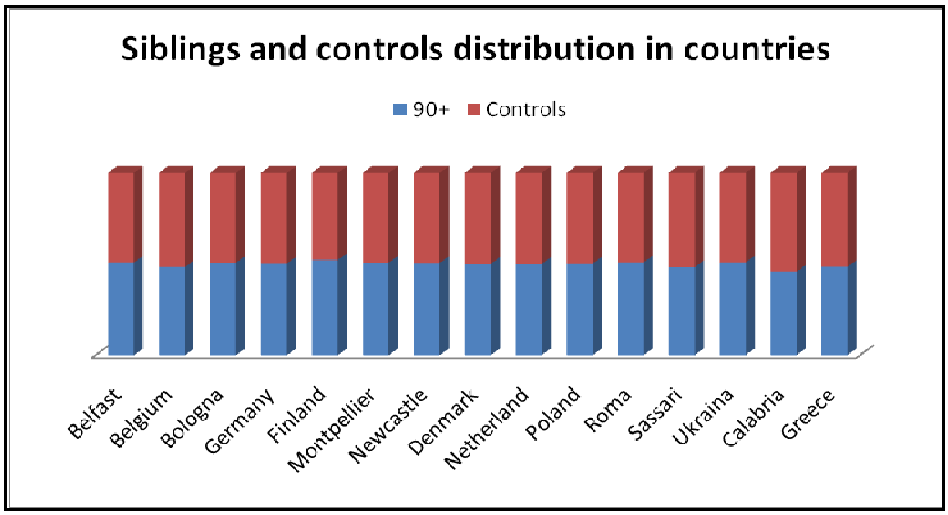



**Fig. 6.1**: Graphic distribution of 90+ subjects and controls as reported in Tab.1. They are correctedly bilanced.

### 1.1 Sex distribution of the participants

As we observed previously, females are more numerous than males, except for Greece where more males than females were recruited. The total number of males and females divided per country is reported in **Tab. 6.3** and in the following figures.

| | *Sibs* | | *Controls* | |
|---|---|---|---|---|
| | Male | Female | Male | Female |
| Belfast | 16 | 48 | 15 | 49 |
| Belgium | 25 | 55 | 26 | 62 |
| Bologna | 55 | 158 | 54 | 160 |
| Calabria | 41 | 84 | 66 | 86 |
| Denmark | 138 | 290 | 141 | 300 |
| Finland | 40 | 113 | 46 | 99 |
| Germany | 31 | 63 | 30 | 66 |
| Greece | 67 | 27 | 72 | 31 |
| Montpellier | 75 | 199 | 77 | 198 |
| Netherland | 52 | 110 | 74 | 93 |
| Newcastle | 21 | 78 | 25 | 75 |
| Poland | 24 | 105 | 26 | 106 |
| Roma | 25 | 50 | 27 | 48 |
| Sassari | 11 | 36 | 12 | 40 |
| Ukraina | 7 | 42 | 8 | 41 |

**Tab. 6.3:** Sex distribution (in %)

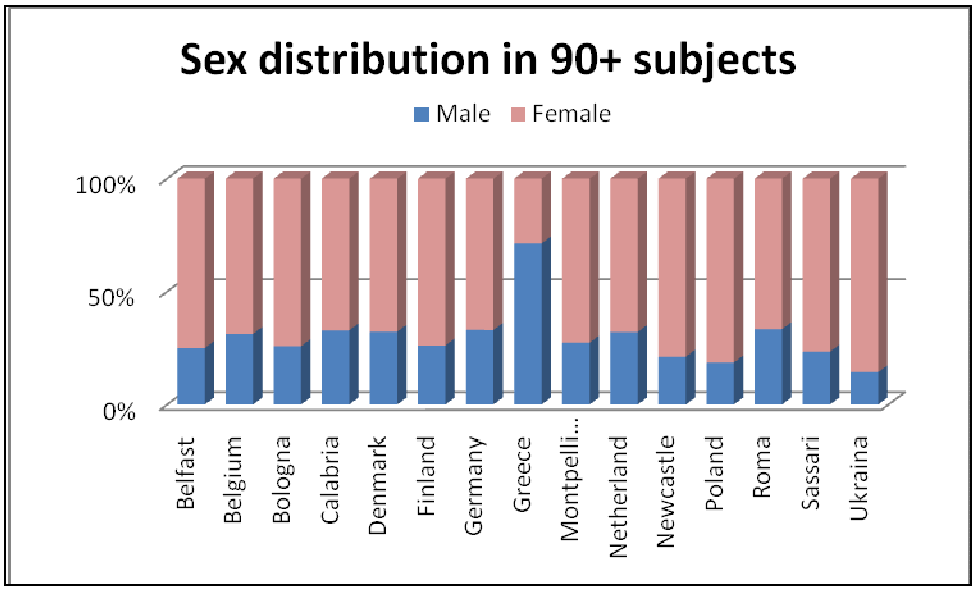



**Fig. 6.2** : Sex distribution in each countries in all 90+ subjects (in %)

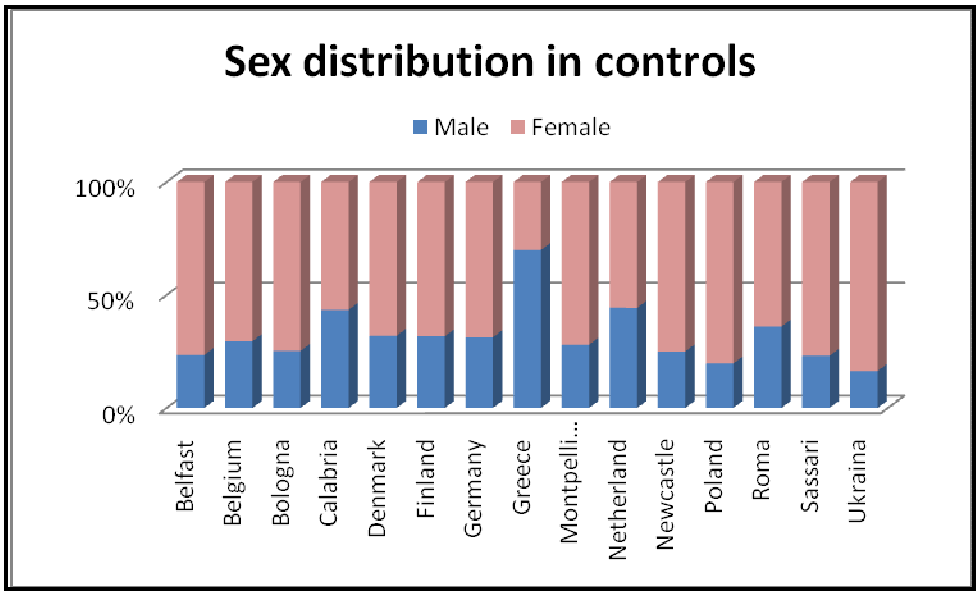



**Figure 6.3** : Sex distribution in each countries in all control subjects (in %)

### 1.2 Age Distribution

The following tables (**Tab. 6.4** and **Tab 6.5**) summaries the minimum, median, and maximum age of the older siblings and controls, separately for females and males.

| | *Female* | | | *Male* | | |
|---|---|---|---|---|---|---|
| | Min | Median | Max | Min | Median | Max |
| Belfast | 89 | 96.5 | 104 | 89 | 93.5 | 98 |
| Belgium | 90 | 95.5 | 101 | 91 | 94 | 97 |
| Bologna | 90 | 96 | 102 | 91 | 97 | 103 |
| Calabria | 90 | 96 | 102 | 91 | 97 | 103 |
| Denmark | 90 | 96.5 | 103 | 90 | 95 | 100 |
| Finland | 88 | 95 | 102 | 89 | 96 | 103 |
| Germany | 91 | 96.5 | 102 | 90 | 96 | 102 |
| Greece | 91 | 99 | 107 | 90 | 96 | 102 |
| Montpellier | 90 | 97.5 | 105 | 91 | 95.5 | 100 |
| Netherland | 90 | 96 | 102 | 91 | 97 | 103 |
| Newcastle | 90 | 96.5 | 103 | 91 | 95.5 | 100 |
| Poland | 91 | 96.5 | 102 | 91 | 95 | 99 |
| Roma | 91 | 95.5 | 100 | 91 | 95.5 | 100 |
| Sassari | 91 | 96 | 101 | 91 | 94.5 | 98 |
| Ukraina | 92 | 98 | 104 | 92 | 94.5 | 97 |

**Tab 6.4**: Summary of age-distribution for old siblings by sex and country (age in years)

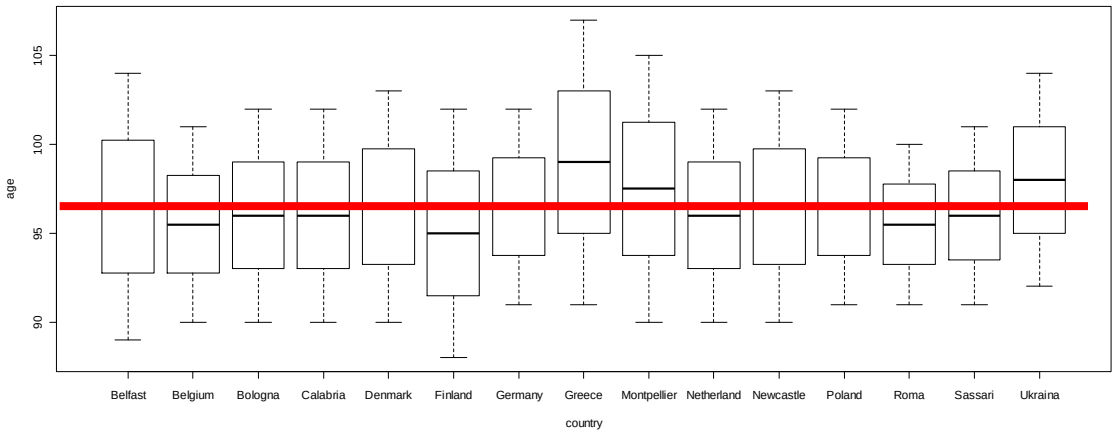



**Fig. 6.4**: Age distribution boxplot of female siblings by countries

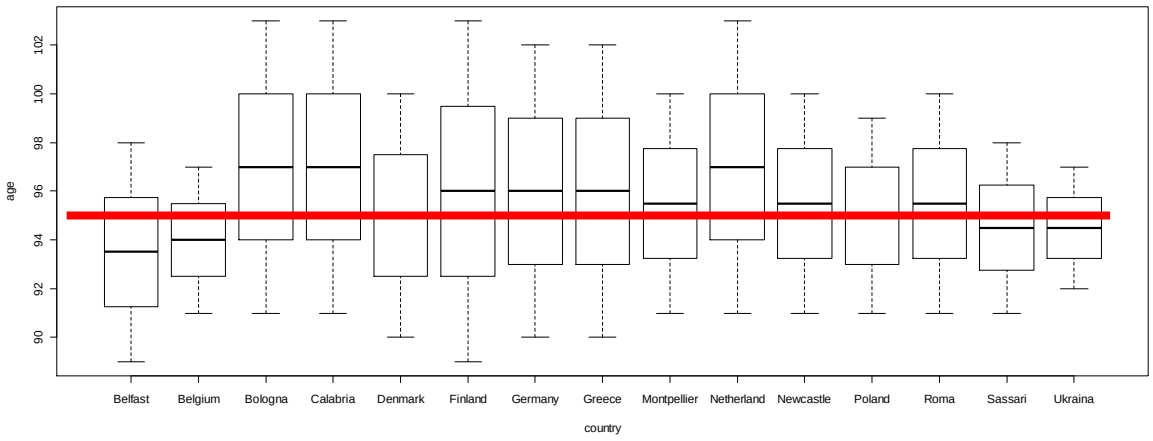



**Fig. 6.5**: Age distribution boxplot of male siblings by countries

| | *Female* | | | *Male* | | |
|---|---|---|---|---|---|---|
| | Min | Median | Max | Min | Median | Max |
| Belfast | 50 | 62.5 | 75 | 56 | 65 | 74 |
| Belgium | 50 | 61.5 | 73 | 50 | 62.5 | 75 |
| Bologna | 50 | 62.5 | 75 | 51 | 62 | 73 |
| Calabria | 50 | 62.5 | 75 | 50 | 62.5 | 75 |
| Denmark | 50 | 66.5 | 83 | 43 | 61 | 79 |
| Finland | 50 | 62.5 | 75 | 50 | 61.5 | 73 |
| Germany | 53 | 64 | 75 | 51 | 61.5 | 72 |
| Greece | 50 | 62.5 | 75 | 50 | 63 | 76 |
| Montpellier | 50 | 62 | 74 | 50 | 63 | 76 |
| Netherland | 50 | 65 | 80 | 51 | 63.5 | 76 |
| Newcastle | 50 | 62.5 | 75 | 51 | 63.5 | 76 |
| Poland | 49 | 61 | 73 | 49 | 61.5 | 74 |
| Roma | 50 | 61.5 | 73 | 52 | 60.5 | 69 |
| Sassari | 54 | 65 | 76 | 61 | 64.5 | 68 |
| Ukraina | 56 | 66 | 76 | 61 | 64.5 | 68 |

**Tab. 6.5:** Summary of age-distribution for the controls by sex and country (age in years)

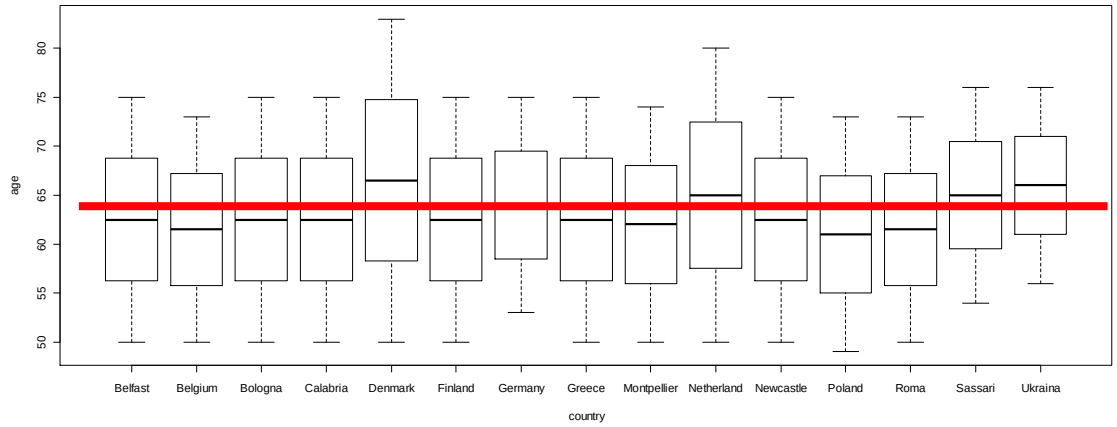



**Fig. 6.6**: Age distribution boxplot of female controls by countries.

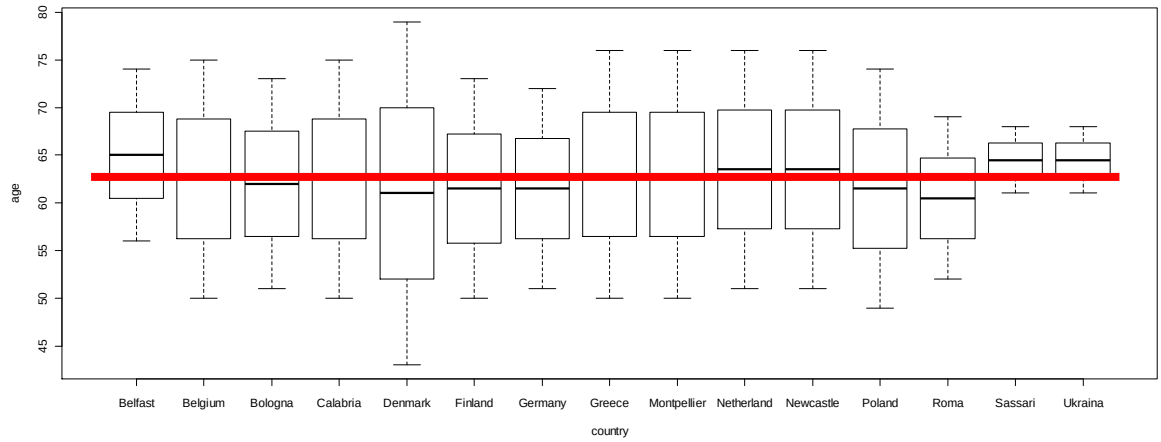



**Fig. 6.7**: Age distribution boxplot of male controls by countries.

The red bars in the figures above indicate the general average age and we can observe that, specially among 90+ males, dividing in geographical areas, Bologna, Calabria, Rome and Greece (which belong to Mediterranean area) and Germany, Finland and Netherland (to the northern areas) have an higher average age than other countries and other subjects.

### 2. Distribution of haplogroups

Among our total samples (4,239 samples), the hierarchical survey of diagnostic markers, present in the coding region, allowed the classification of mtDNA from 90+ and controls into more than 40 haplogroups and sub-haplogroups. Most of these are typical of modern European population, but a few percentage of East Asian (M1a, D5a), sub-Saharian African (L1b1) and North American (A4, C1d) mtDNAs has been also detected. This latter finding is not so unexpected because in Europe, East Asian and in particular

African haplogroups are quite common. We decided to group all the sub-haplogroups with frequencies lower than 1.5%, thus reducing the overall number of categories from 85 (see Tab.1 in the *Appendix*) to 22 (**Tab. 6.6** and **Fig. 6.8**).

| 90+ sibpair (n=2,086) | | | | controls (n=2,153) | | | |
|---|---|---|---|---|---|---|---|
| **sub-haplogroups** | **N** | **%** | **SE** | **sub-haplogroups** | **N** | **%** | **SE** |
| H* (1) | 402 | 19.27 | 0.0086 | H* (1) | 381 | 17.70 | 0.0082 |
| H1 | 292 | 14.00 | 0.0076 | H1 | 325 | 15.10 | 0.0077 |
| H2 | 42 | 2.01 | 0.0031 | H2 | 23 | 1.07 | 0.0022 |
| H3 | 71 | 3.40 | 0.0040 | H3 | 85 | 3.95 | 0.0042 |
| H5 | 68 | 3.26 | 0.0039 | H5 | 63 | 2.93 | 0.0036 |
| H6 | 52 | 2.49 | 0.0034 | H6 | 47 | 2.18 | 0.0031 |
| HV0* (2) | 83 | 3.98 | 0.0043 | HV0* (2) | 68 | 3.16 | 0.0038 |
| HV* | 42 | 2.01 | 0.0031 | HV* | 50 | 2.32 | 0.0032 |
| I | 46 | 2.21 | 0.0032 | I | 44 | 2.04 | 0.0030 |
| J1 | 144 | 6.90 | 0.0056 | J1 | 153 | 7.11 | 0.0055 |
| J2 | 39 | 1.87 | 0.0030 | J2 | 57 | 2.65 | 0.0035 |
| K1 | 117 | 5.61 | 0.0050 | K1 | 140 | 6.50 | 0.0053 |
| T1 | 43 | 2.06 | 0.0000 | T1 | 53 | 2.46 | 0.0033 |
| T2 | 174 | 8.34 | 0.0061 | T2 | 152 | 7.06 | 0.0055 |
| U (3) | 59 | 2.83 | 0.0036 | U (3) | 61 | 2.83 | 0.0036 |
| U2 | 38 | 1.82 | 0.0029 | U2 | 39 | 1.81 | 0.0029 |
| U4 | 34 | 1.63 | 0.0028 | U4 | 48 | 2.23 | 0.0032 |
| U5a | 90 | 4.31 | 0.0044 | U5a | 109 | 5.06 | 0.0047 |
| U5b | 51 | 2.44 | 0.0034 | U5b | 56 | 2.60 | 0.0034 |
| W | 51 | 2.44 | 0.0034 | W | 39 | 1.81 | 0.0029 |
| X | 42 | 2.01 | 0.0031 | X | 41 | 1.90 | 0.0029 |
| OTHER (4) | 106 | 5.08 | 0.0048 | OTHER (4) | 119 | 5.53 | 0.0049 |

**Tab. 6.6:** list of sub-haplogroups with the related frequencies and Standard Error (SE). (1) H* includes all mtDNAs belonging to H, except those classified as H1, H3, H5 and H6. (2) HV0* includes HV0a, (3) U includes U1, U3, U6, U7, U8, (4) OTHER includes K, K2, T, R0, R0a, R1, R2, HV1, HV2, N1a, N1b, N1c,N9a, A4, D5, C1d, M1, L1b1, with frequencies lower than 1,5%.

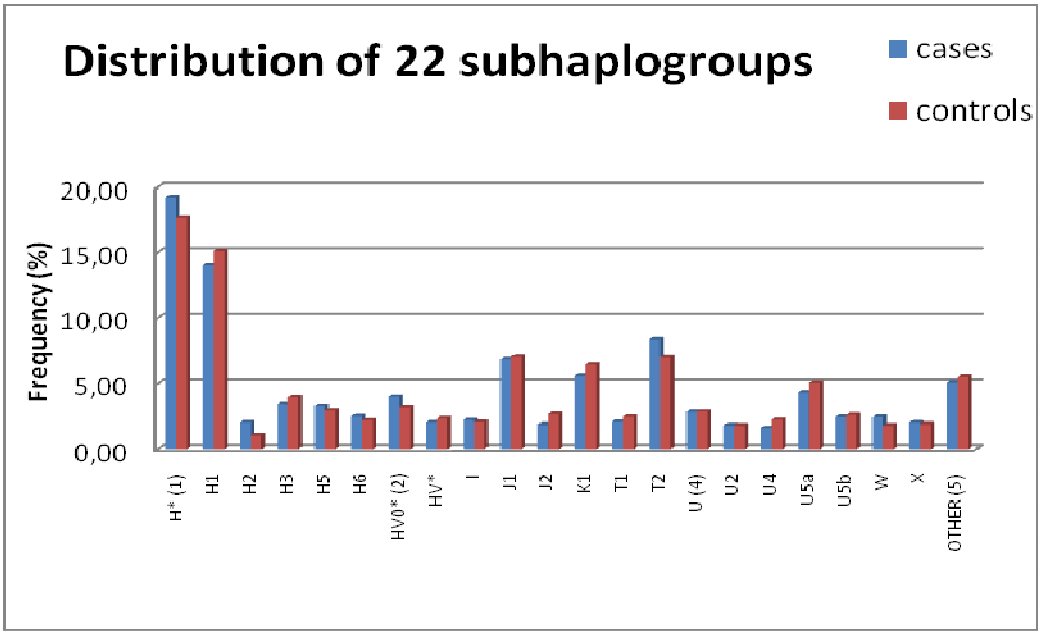



**Fig. 6.8:** Distribution of the most important haplogroups.

No significant difference was found between cases and controls distribution in mtDNAs classification.

Comparing our frequencies with those previously reported in literature **Tab. 6.7** (Richards et al. 2000), we verified that our results correspond with a good approximation to the reference values referred to populations.

| **Haplogroup** | **90+ subjects** | **Controls** | **Reference (*)** |
|---|---|---|---|
| HV | 51.7% | 48.3% | 50.4-54.1% |
| H | 44.4% | 42.9% | 44.5-48.2% |
| V | 1.3% | 1.5% | 3.9-5.4% |
| J | 8.7% | 9.7% | 8.3-10.4% |
| T | 10.6% | 9.8% | 7.2-9.2% |
| U | 13.0% | 14.5% | 20.1-23.2% |
| K | 6.5% | 7.4% | 4.9-6.6% |
| I | 2.2% | 2.0% | 1.6-2.7% |
| W | 2.4% | 1.8% | 1.5-2.5% |

| | | | |
|---|---|---|---|
| X | 2.0% | 1.9% | 1.2-2.0% |

(*) Richards et al. 2000

**Tab. 6.7**: The table shows the haplogrops' frequencies compared with those reported by Richards et al. 2000.

In general, frequencies are similar to those reported in literature except for frequencies of V, T and U haplogroups that are more distant from Richards data. In particular haplogroup HV resulted to be the most frequent, present in 51.7% nonagenarians and 48.3% controls in agreement to previous studies (Torroni et al. 1996). Haplogroup T has a higher frequency both in 90+ subjects and controls than in reference population, while haplogroups V and U have an opposite trend. Probably, it could be due to the fact that we have studied new population previously never explored.

It is also extremely important to perform analysis data divided for gender. The sub-haplogroups, the frequencies, the Standard Error and the comparison between male and female are shown in **Tab. 6.8** and **Tab. 6.9** (and respectively in **Fig. 6.9** and **Fig. 6.10**).

| Males (N = 1327) | | | | | | |
|---|---|---|---|---|---|---|
| | 90+ subjects (N=628) | | | Controls (N= 699) | | |
| | N | % | SE | N | % | SE |
| H* (1) | 134 | 21.3 | 0.0163 | 135 | 19.3 | 0.0149 |
| H1 | 75 | 11.9 | 0.0129 | 108 | 15.5 | 0.0137 |
| H2 | 10 | 1.6 | 0.0050 | 11 | 1.6 | 0.0047 |
| H3 | 18 | 2.9 | 0.0067 | 16 | 2.3 | 0.0057 |
| H5 | 24 | 3.8 | 0.0077 | 20 | 2.9 | 0.0063 |
| H6 | 19 | 3.0 | 0.0068 | 16 | 2.3 | 0.0057 |
| HV0* (2) | 24 | 3.8 | 0.0077 | 24 | 3.4 | 0.0069 |
| HV* | 14 | 2.2 | 0.0059 | 17 | 2.4 | 0.0058 |
| I | 7 | 1.1 | 0.0042 | 13 | 1.9 | 0.0051 |
| J1 | 40 | 6.4 | 0.0097 | 44 | 6.3 | 0.0092 |
| J2 | 6 | 1.0 | 0.0039 | 22 | 3.1 | 0.0066 |
| K1 | 47 | 7.5 | 0.0105 | 46 | 6.6 | 0.0094 |
| T1 | 11 | 1.8 | 0.0052 | 11 | 1.6 | 0.0047 |
| T2 | 46 | 7.3 | 0.0104 | 54 | 7.7 | 0.0101 |
| U (3) | 21 | 3.3 | 0.0072 | 20 | 2.9 | 0.0063 |
| U2 | 12 | 1.9 | 0.0055 | 13 | 1.9 | 0.0051 |
| U4 | 13 | 2.1 | 0.0057 | 13 | 1.9 | 0.0051 |
| U5a | 24 | 3.8 | 0.0077 | 30 | 4.3 | 0.0077 |
| U5b | 18 | 2.9 | 0.0067 | 17 | 2.4 | 0.0058 |

| | | | | | | |
|---|---|---|---|---|---|---|
| W | 13 | 2.1 | 0.0057 | 16 | 2.3 | 0.0057 |
| X | 18 | 2.9 | 0.0067 | 19 | 2.7 | 0.0062 |
| OTHER (4) | 34 | 5.4 | 0.0090 | 34 | 4.9 | 0.0081 |

**Tab. 6.8:** frequencies of male mtDNA sub-haplogroups of 628 90+subjects and 699 controls from all European countries. SE=Standard Error; Sub-haplogroups with frequencies lower than 1.5% were grouped. (1) H* includes all mtDNAs belonging to H, except those classified as H1, H3, H5 and H6. (2) HV0* includes HV0a, (3) U includes U1, U3, U6, U7, U8, (4) OTHER includes K, K2, T, R0, R0a, R1, R2, HV1, HV2, N1a, N1b, N1c,N9a, A4, D5, C1d, M1, L1b1, with frequencies lower than 1.5%.

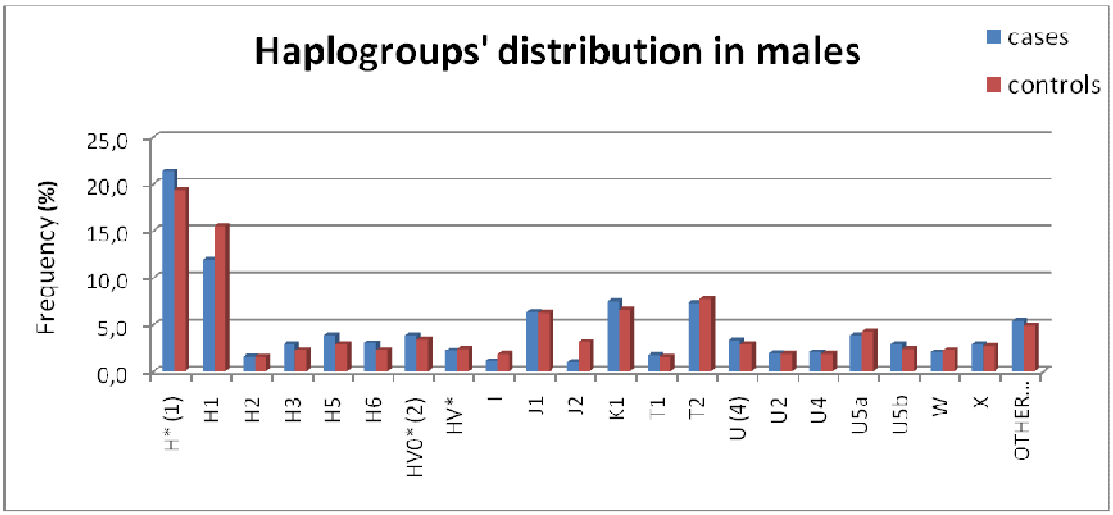



**Fig. 6.9:** Haplogroup distribution of frequencies in males.

The Pearson chi-squared test has demonstrated that haplogroup distribution in males is not significant (X-squared = 22.6211, p-value = 0.6543).

| Females (N = 2912) | | | | | | |
|---|---|---|---|---|---|---|
| | 90+ subjects (N= 1458) | | | Controls (N= 1454) | | |
| | N | % | SE | N | % | SE |
| H* | 268 | 18.4 | 0.0101 | 246 | 16.9 | 0.0098 |
| H1 | 217 | 14.9 | 0.0093 | 217 | 14.9 | 0.0093 |
| H2 | 32 | 2.2 | 0.0038 | 12 | 0.8 | 0.0024 |
| H3 | 53 | 3.6 | 0.0049 | 69 | 4.7 | 0.0056 |
| H5 | 44 | 3.0 | 0.0045 | 43 | 3.0 | 0.0044 |
| H6 | 33 | 2.3 | 0.0039 | 31 | 2.1 | 0.0038 |
| HV0* | 59 | 4.0 | 0.0052 | 44 | 3.0 | 0.0045 |
| HV* | 28 | 1.9 | 0.0036 | 33 | 2.3 | 0.0039 |
| I | 39 | 2.7 | 0.0042 | 31 | 2.1 | 0.0038 |
| J1 | 104 | 7.1 | 0.0067 | 109 | 7.5 | 0.0069 |
| J2 | 33 | 2.3 | 0.0039 | 35 | 2.4 | 0.0040 |
| K1 | 70 | 4.8 | 0.0056 | 94 | 6.5 | 0.0064 |

| | | | | | | |
|---|---|---|---|---|---|---|
| T1 | 32 | 2.2 | 0.0038 | 42 | 2.9 | 0.0044 |
| T2 | 128 | 8.8 | 0.0074 | 98 | 6.7 | 0.0066 |
| U | 38 | 2.6 | 0.0042 | 41 | 2.8 | 0.0043 |
| U2 | 26 | 1.8 | 0.0035 | 26 | 1.8 | 0.0035 |
| U4 | 21 | 1.4 | 0.0031 | 35 | 2.4 | 0.0040 |
| U5a | 66 | 4.5 | 0.0054 | 79 | 5.4 | 0.0059 |
| U5b | 33 | 2.3 | 0.0039 | 39 | 2.7 | 0.0042 |
| W | 38 | 2.6 | 0.0042 | 23 | 1.6 | 0.0033 |
| X | 24 | 1.6 | 0.0033 | 22 | 1.5 | 0.0032 |
| OTHER | 72 | 4.9 | 0.0057 | 85 | 5.8 | 0.0062 |

**Tab. 6.9:** frequencies of female mtDNA sub-haplogroups in 1458 90+ subjects and 1,454 controls from 11 European countries. SE=Standard Error; Sub-haplogroups with frequencies lower than 1.5% were grouped. H* includes all mtDNAs belonging to haplogroup H, except those further classified (H1, H3, H5 and H6). The same rationale has been used for HV0*, U and Other.

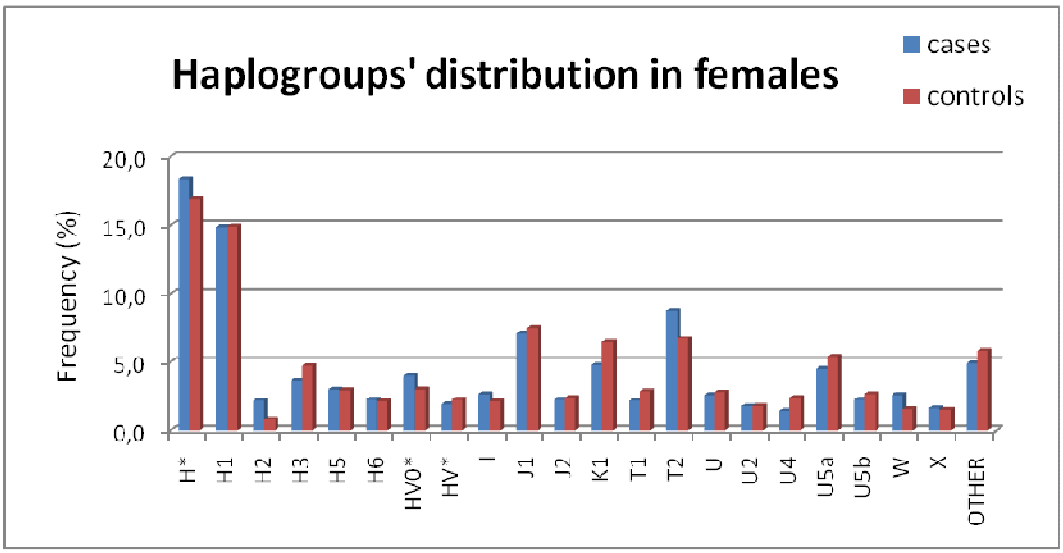



**Fig. 6.10:** HaplogroupS' distribution of frequencies in females.

The Pearson chi-squared test has demonstrated that haplogroup distribution in females is slightly statistical significant (X-squared = 39.5362, p-value = 0.04329). Also a XY conditioning plot analysis was conducted: thit is a plot of two variables conditional on the value of a third variable (called the conditioning variable). The conditioning variable may be either a variable that takes on only a few discrete values or a continuous variable that is divided into a limited number of subsets. In the graph below (**Fig. 6.11**) it is possible to notice that for both cases and controls the majority of haplogroup frequencies are around the 5%, only a few are near 10% and two are more than 15%.

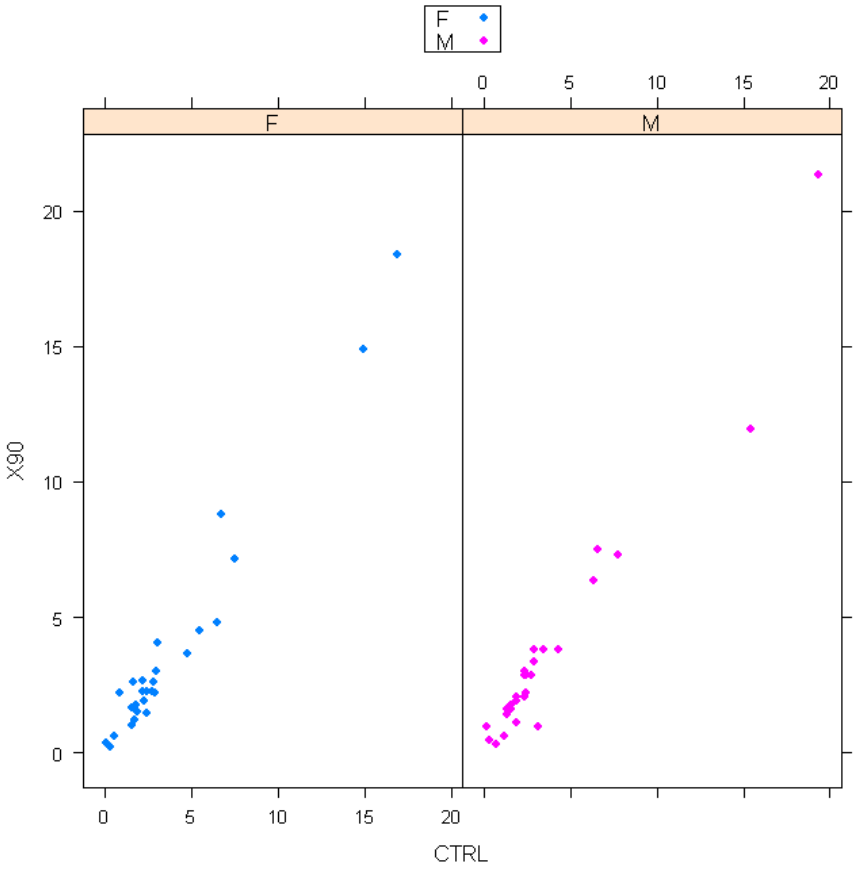



**Fig. 6.11:** a XY conditioning plot divided in Females and Males. On the ordinate there are the haplogroup frequencies of 90+ subjects, on the x-axis the controls frequencies (there are 26 dots, each corresponding to an haplogroup). On the x-axis the frequency of controls is reported, on the y-axis the frequency of siblings.

### 3. Associations between haplogroups and longevity

As no difference has been revealed from the distribution between 90+ cases and controls even when frequencies were compared separately for gender, we further proceeded with statistical analysis performing an association test (Fisher test) between cases/controls and gender (**Tab. 6.10**) for each subhaplogroup (high resolution analysis), to identify the potential sub-haplogroup associated with healthy aging and longevity.

| Subhaplogroups | OR | 95% CI | Fisher exact test p value |
|---|---|---|---|
| H | 0.874 | 0.71 - 1.06 | 0.1938 |
| **H1** | **0.695** | **0.48 - 0.99** | **0.0427** |
| H3 | 1.461 | 0.63 - 3.38 | 0.3379 |
| H5 | 1.171 | 0.53 - 2.59 | 0.7137 |
| H6 | 1.114 | 0.45 - 2.77 | 0.8357 |

| | | | |
|---|---|---|---|
| HV | 0.971 | 0.37 - 2.52 | 1 |
| HV0* | 0.884 | 0.47 - 1.64 | 0.7665 |
| I | 0.484 | 0.08 - 2.45 | 0.4748 |
| I1 | 0.392 | 0.006 - 6.09 | 0.6030 |
| I3 | 0.367 | 0.02 - 3.41 | 0.3707 |
| J1 | 0.953 | 0.55 - 1.62 | 0.8978 |
| **J2** | **0.293** | **0.08 - 0.86** | **0.0214** |
| K | 0.403 | 0.060 - 2.31 | 0.2761 |
| K1 | 1.370 | 0.79 - 2.36 | 0.2425 |
| K2 | 0.102 | 0.001 - 1.73 | 0.1026 |
| N | 3 | 0.078 - 234.45 | 1 |
| N1 | 0.728 | 0.13 - 3.64 | 0.7311 |
| R0 | 4.156 | 0.34 - 232.12 | 0.3333 |
| T | 1.886 | 0.12 - 37.90 | 1 |
| T1 | 1.309 | 0.45 - 3.80 | 0.6303 |
| T2 | 0.653 | 0.39 - 1.07 | 0.0918 |
| U1 | 1.137 | 0.17 - 7.49 | 1 |
| U2 | 0.924 | 0.31 - 2.66 | 1 |
| U3 | 2.023 | 0.28 - 24.13 | 0.6800 |
| U4 | 1.656 | 0.58 - 4.72 | 0.3392 |
| U5 | 1.060 | 0.62 - 1.78 | 0.9000 |
| U6 | 3.237 | 0.29 - 51.96 | 0.3348 |
| U8 | 2.307 | 0.33 - 17.92 | 0.4136 |
| W | 0.724 | 0.20 - 2.53 | 0.5843 |
| W5 | 0.141 | 0.0014 - 4.93 | 0.2262 |
| X | 1.570 | 0.25 - 10.38 | 0.6951 |
| X2 | 0.577 | 0.17 - 1.87 | 0.4210 |

**Tab. 6.10**: Fisher test (p-value), Odds ratio (OR) and 95% Confidence Intervals (95%CI) for subhaplogroups in the entire sample.

In **Tab. 6.10** only the principal subhaplogroups are reported, the other are too few to give a significative result (for example A4, C, D, L, M, N, R1, R2, U7, U9, W1, W4, W6 and X1).

Of all haplogroups, only H1 and J2 resulted significant.

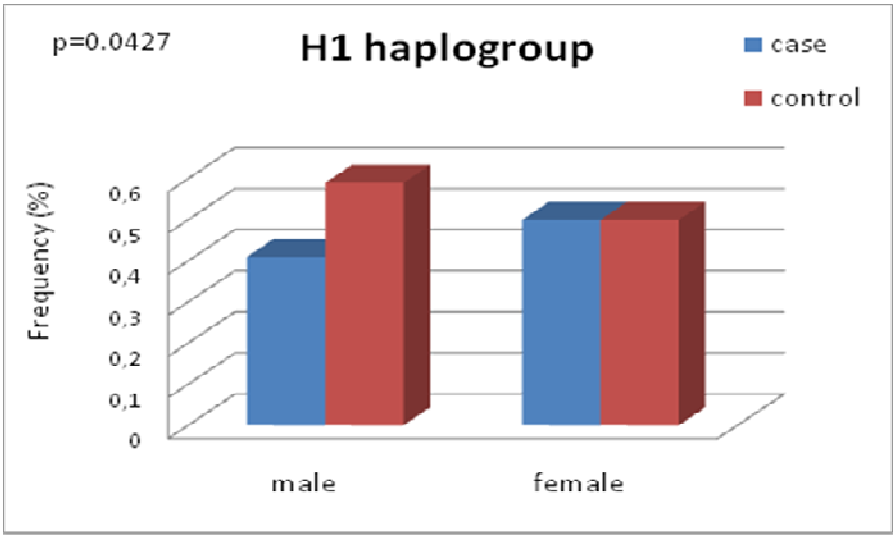



**Fig. 6.12:** distribution of males and females, cases and controls belonging to H1 haplogroup.

Among all males, H1 is more represented in controls than cases, while among females the frequency is similar. The p value is slightly significant (p=0.0427) (**Fig. 6.12**).

As regard haplogroup J2, differences between cases and controls resulted significant (p=0.0214). If we take into account gender considering the distribution among males shows that more than 80% are cases, as **Fig. 6.13** reports, while among females the frequency is equally distributed.

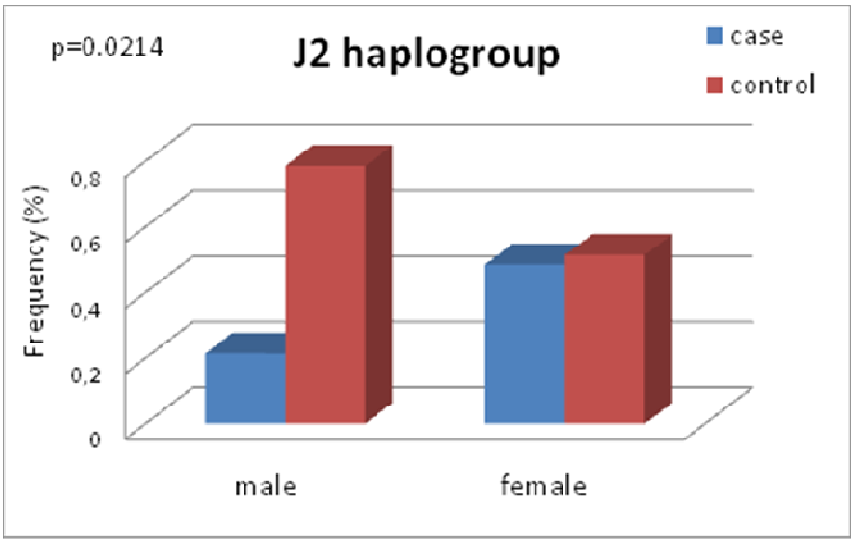



**Fig. 6.13:** distribution of male and female belonging to J2 haplogroup. It is represented with a high frequency among male controls.

These percentages mean that haplogroup J2 has an higher frequency among male controls than cases and that we cannot observe any difference between the cases and controls among females.

### 4. H1 and J2 distribution in countries

We also analyzed the distribution of H1 and J2 haplogroup in the 11 European countries in order to investigate the possible association with one or more countries. By applying a Pearson's Chi-squared test to each country in the H1 distribution (**Tab. 6.11** and **Fig. 6.14**), without taking into account gender, we verified that H1 is significantly associated with control belonging to Greece (p=0.0455), Poland (p=0.01631) and highly associated with Newcastle (p=0.00604). But H1 inverts its trend in Belfast, Belgium, Montpellier, Finland and Sassari even if they are not significant (being more represented in 90+ subjects of different countries than controls). If we stratify for gender, we could find that H1 in Greece, Poland and Newcastle, is significantly associated with male controls confirming this association (data not reported).

| | **90+** | | | **controls** | | |
|---|---|---|---|---|---|---|
| | **N** | **%** | **SE** | **N** | **%** | **SE** |
| Belfast | 12 | 4.1 | 0.0150 | 8 | 2.5 | 0.0126 |
| Belgium | 17 | 5.8 | 0.0178 | 15 | 4.6 | 0.0170 |
| Bologna | 20 | 6.8 | 0.0191 | 29 | 8.9 | 0.0231 |
| Calabria | 9 | 3.1 | 0.0131 | 14 | 4.3 | 0.0165 |
| Denmark | 75 | 25.7 | 0.0331 | 81 | 24.9 | 0.0351 |
| Finland | 25 | 8.6 | 0.0212 | 22 | 6.8 | 0.0204 |
| Germany | 13 | 4.5 | 0.0156 | 14 | 4.3 | 0.0165 |
| **Greece** | **4** | **1.4** | **0.0088** | **12** | **3.7** | **0.0153** |
| Montpellier | 48 | 16.4 | 0.0281 | 44 | 13.5 | 0.0278 |
| Netherland | 19 | 6.5 | 0.0187 | 21 | 6.5 | 0.0199 |
| **Newcastle** | **6** | **2.1** | **0.0108** | **20** | **6.2** | **0.0195** |
| **Poland** | **12** | **4.1** | **0.0150** | **27** | **8.3** | **0.0224** |
| Roma | 9 | 3.1 | 0.0131 | 4 | 1.2 | 0.0089 |
| Sassari | 16 | 5.5 | 0.0173 | 7 | 2.2 | 0.0118 |
| Ukraina | 7 | 2.4 | 0.0116 | 7 | 2.2 | 0.0118 |

**Tab. 6.11**: distribution of haplogroup H1 in all the 11 countries. The frequencies and the Standard Error (SE) are reported.

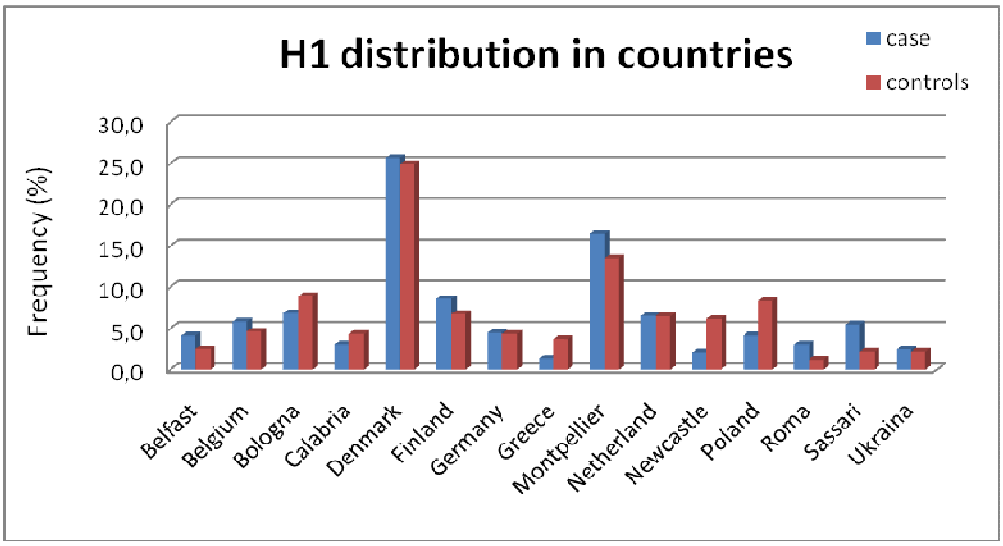



**Fig. 6.14**: H1 distribution for cases and controls in all countries.

The same rationale has been used for J2 haplogroup distribution (**Tab. 6.12** and **Fig. 6.15**) and the same Pearson's Chi-squared test was conducted, showing that J2 is associated with male controls belonging to Calabria (p=0.008151) and Greece (p=0.02535), which surprisingly represent the South Europe. As the distribution of this haplogroup evidences, J2 is not represented among 90+ subjects in Calabria and Greece but only among controls subjects. The absence of association of mtDNA haplogroup J with longevity in southern Italian population was previously observed (De Benedictis et al, 1999 and Dato et al, 2004).

| | **90+** | | | **controls** | | |
|---|---|---|---|---|---|---|
| | **N** | **%** | **SE** | **N** | **%** | **SE** |
| Belfast | 1 | 2.6 | 0.0120 | 3 | 5.3 | 0.0181 |
| Belgium | 2 | 5.1 | 0.0167 | 2 | 3.5 | 0.0149 |
| Bologna | 3 | 7.7 | 0.0202 | 5 | 8.8 | 0.0229 |
| **Calabria** | **0** | **0.0** | **0.0000** | **7** | **12.3** | **0.0266** |
| Denmark | 9 | 23.1 | 0.0319 | 12 | 21.1 | 0.0331 |
| Finland | 3 | 7.7 | 0.0202 | 5 | 8.8 | 0.0229 |
| Germany | 4 | 10.3 | 0.0230 | 1 | 1.8 | 0.0106 |
| **Greece** | **0** | **0.0** | **0.0000** | **5** | **8.8** | **0.0229** |
| Montpellier | 4 | 10.3 | 0.0230 | 6 | 10.5 | 0.0249 |
| Netherland | 2 | 5.1 | 0.0167 | 4 | 7.0 | 0.0207 |
| Newcastle | 2 | 5.1 | 0.0167 | 1 | 1.8 | 0.0106 |
| Poland | 3 | 7.7 | 0.0202 | 2 | 3.5 | 0.0149 |
| Roma | 1 | 2.6 | 0.0120 | 0 | 0.0 | 0.0000 |
| Sassari | 4 | 10.3 | 0.0230 | 3 | 5.3 | 0.0181 |
| Ukraina | 1 | 2.6 | 0.0120 | 1 | 1.8 | 0.0106 |

**Tab. 6.12**: distribution of haplogroup J2 in all the 11 countries. We reported the frequencies and the Standard Error (SE).

As we can observe from the distribution of J2 in each country, the distribution of this subhaplogroup, not considering Calabria and Greece, is nearly balanced (for a total of 39 J2 cases and 45 J2 controls), but when considering also these two areas, J2 seems to be more prevalent in controls. We conclude that Calabria and Greece determined this difference. The frequency in Calabria and Greece is high (respectively 4.6% and 4.8%) considering that J2 represents only 2% of the entire distribution (2.6% in controls and 1.8% in siblings).

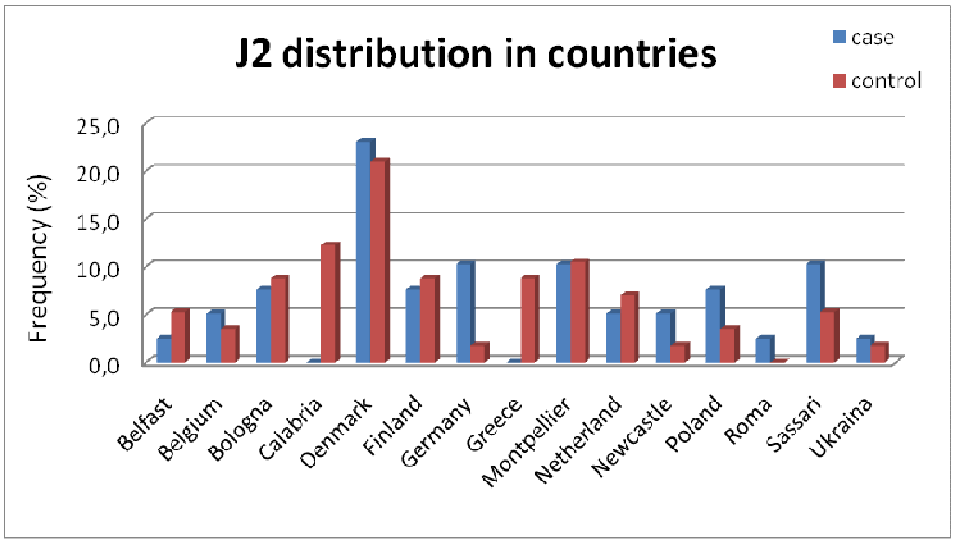



**Fig. 6.15:** Distribution of J2 in all the 11 countries adherent to the Project.

Also, if we stratify for gender, we could better appreciate the association between J2 and male controls but the number of samples belonging to this haplogroup is not enough to conduct exhaustive statistical analysis.

**5. How subhaplogroups can influence associations?**

Furthermore, we have studied the distribution of all subhaplogroups of H1 (H1a, H1a1, H1a2, H1a3, H1b, H1c, H1c1, H1e, H1f, H1n) in order to verify which subhaplogroup, if present, could influence the association (**Fig. 6.16**). The analysis was conducted only taking into account the complete sequences. In fact, simply analyzing the Dlopp sequences we are not able to assign the subhaplogroups of H1. Only by sequencing the entire sequence we can go more deeply in the analysis.

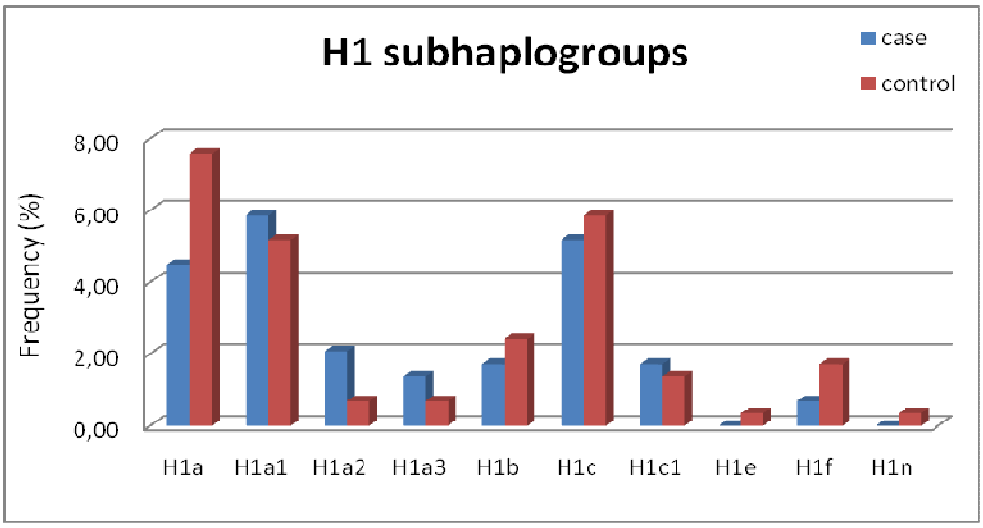



**Fig. 6.16:** distribution of all H1 subhaplogroups. None of the subhaplogroups resulted significant.

The calculation of p value for all the subhaplogroups, taking into account male/female and case/control, reported no significant data.

Furthermore, we have studied the distribution of the subhaplogroups J2a and J2b, and we observed that only J2a is really significant (**X-squared = 5.6958, p-value = 0.01701**) rather than J2b subhaplogroup (X-squared = 1.1722, p-value = 0.2789). If we consider the distribution in male and females (**Fig. 6.17** and **Fig. 6.18**), among males about 80% are controls, confirming the J2 haplogroup trend. This analysis was conducted both on complete sequences and on d-loop sequences.

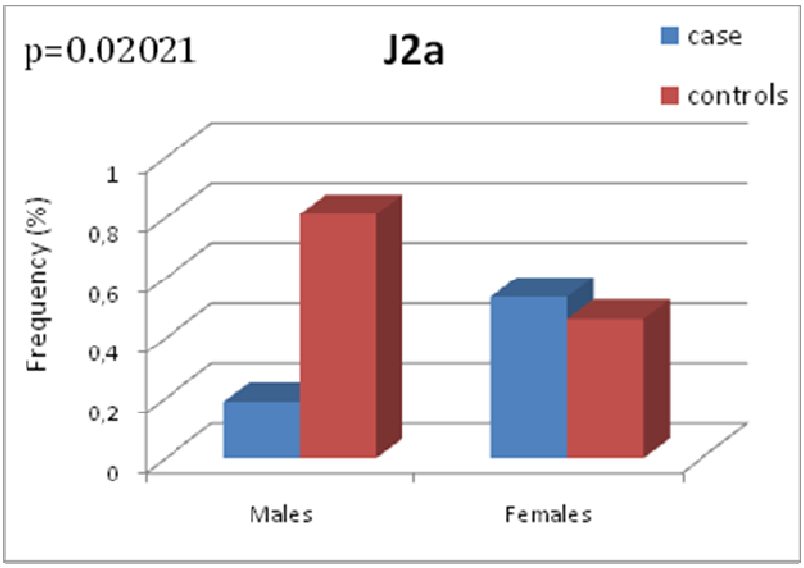



**Fig. 6.17:** distribution of J2a among gender.

This fact suggests that the higher frequency of haplogroup J2 in male controls is attributable to an increase in subcluster J2a, rather than to J2b.

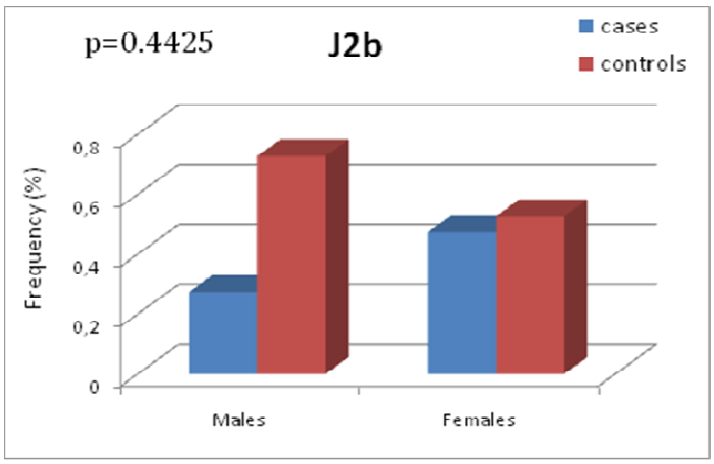



**Fig. 6.18:** distribution of J2b among gender.

These data are in contrast with data reported by literature. The fact that J2 has a higher frequency in male controls is not supported by previous studies conducted in North Italy and in Finland (De Benedictis, 1999, Niemi et al. 2003). Other studies are in contrast with the findings of De Benedictis et al. (1999). The study conducted by Ross et al. (2001) on Irish failed to show any J haplogroup association with either age and gender. We found the association between H1 and male controls, and with a highly significance, J2 with male controls.

It is very interesting to note that in general J2 haplogroup appears to be slightly more frequent in 90+ females than in control females, even thought this data is not significant. But analyzing J2a and J2b subhaplogroups, we can see an opposite trend: J2a seems to be more frequent in cases while J2b in controls.

### 6. Analysis of quartiles

A quartile is one of the three values that divide a range of data into four equal parts. The first quartile (also defined “lower quartile”) is the number below which lies the 25 percent of the bottom data. The second quartile (the 'median') divides the range in the middle and has 50 percent of the data below it. The third quartile (also called 'upper quartile') has 75 percent of the data below it and the top 25 percent of the data above it.

We tried to analyze our 90+ subjects samples using quartiles in order to discover association never seen. For men, we identified different values: the I quartile is represented by samples with age below 92 years, the II quartile by subjects with 92 and 93 years, the III quartile by samples with 94 and 95 years and the IV quartile by subjects with age higher than 96 years (to 103 years). For women, the I quartile groups samples with age below 93 years, the II quartile is represented by subjects with 93 years, the III quartile by samples with 94 and 95 years (Median value is 94) and the IV quartile by subjects with age higher than 96 years (to 107 years). The values of median, lower quartile and upper quartile are as follows:

| | *Male* | *Female* |
|---|---|---|
| Median | 94 | 94 |
| Lower quartile | 92 | 93 |
| **Upper quartile** | **96** | **96** |

We wanted to study the association between J2 and subjects with more than 96 years (the upper quartile value) and eventually apply the same association study to all other haplogroups.

Firstly we analyzed J haplogroup in a case/control study, taking into account the gender. By applying a Fisher exact test, J1 subhaplogroup seems to be no significantly associated with this group of subjects (OR=0.78, 95%CI 0.39-1.51, p=0.5337), J2 is again significant and associated with male controls (OR=0.24, 95%CI 0.041-0.96, p=0.03302), in fact among males, 88% is represented by controls, as **Fig. 6.19** shows.

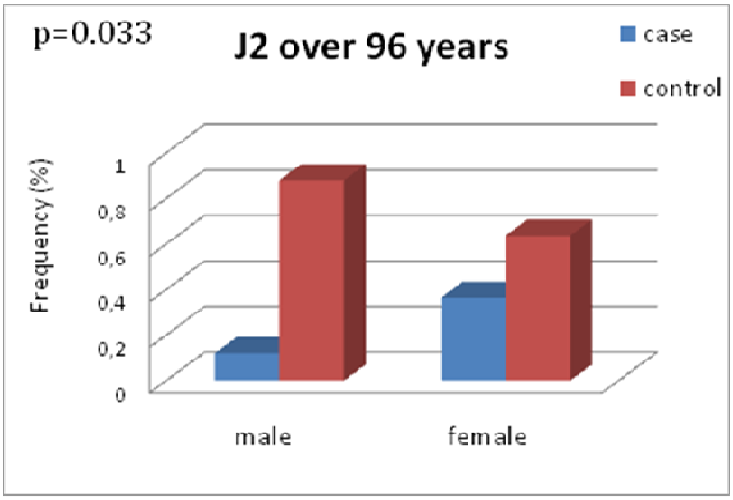



**Fig. 6.19:** distribution of males and females to J2 haplogroup in cases and controls. 88% of males are controls.

By extending this analysis on this restricted group to all haplogroups in this restricted group (over 96 years), also T2 haplogroup resulted significant (OR=0.566, 95%CI 0.30-1.02, p=0.05) as **Fig. 6.20** shows. Among cases, 76% are females while males are only 23%. T2 is tendentially more represented in females both cases and controls. In T2 subjects younger than 96 years old no significant association is observed.

Instead, H1 is now not significant (p=0.09).

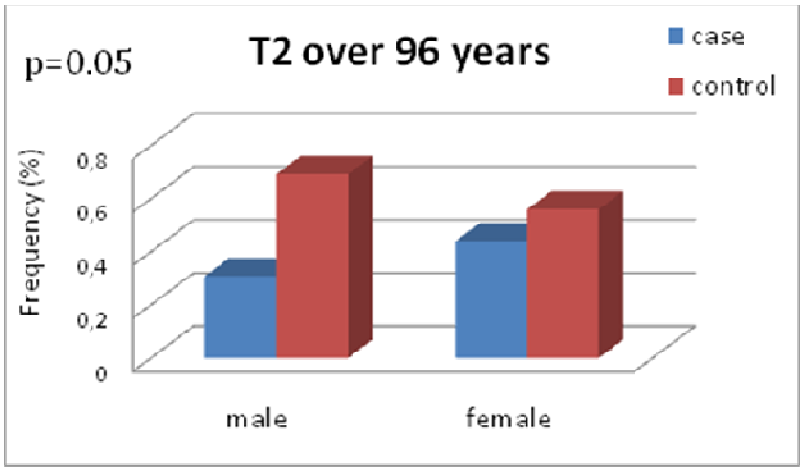



**Fig. 6.20:** distribution of males and females belonging to T2 haplogroup in cases and controls. 69% of males are represented by controls.

## 7. Analysis of mutations in complete sequences

### 7.1 J2 haplogroup and complete sequencing

In order to identify whether the association of J2 with longevity could be attributed to specific mutations, the complete sequences of Finland and Denmark samples were analyzed. The Operative Unit of Bologna has sequenced all mtDNAs from Calabria and Greece but J2 is not represented among sibs, while among controls there were only two sequences for Greece and one for Calabria, not sufficient for statistical analysis or more simply for a comparison. The Chinese Partner has sequenced mtDNAs from Finland and Denmark and we have used these data in order to verify whether could exist a mutation or mutations more frequent in 90+ subjects than controls or *viceversa*. In order to assess if the J2 background of the patients harboured mutations responsible for the longevity phenotype, we investigated the complete sequences of J2 haplogroup subjects (compared to controls). The result of the sequence analysis is summarized in **Table**

**6.13**. A total of 110 mutated positions relative to the reference sequence of Cambridge were detected and analyzed. Overall 23 of the total mutations (underlined and in italics) were not previously reported in either MITOMAP (www.mitomap.org) or mtDB (www.genpat.uu.se/mtDB) and each mutation was observed only in a single J2 mtDNA.

| **Nucleotide position** | **Locus** | **Nucleotide Change** | **Aminoacid Change** | **90+ subjects N=12 (%)** | **Controls N=17 (%)** |
|---|---|---|---|---|---|
| 73 | D-loop (HVSII) | A>G | non coding | 12 (100) | 17 (100) |
| 146 | D-loop (HVSII) | T>C | non coding | 1 (8,3) | 0 (0) |
| 150 | D-loop (HVSII) | C>T | non coding | 12 (100) | 17 (100) |
| 152 | D-loop (HVSII) | T>C | non coding | 12 (100) | 17 (100) |
| 189 | D-loop (HVSII) | A>G | non coding | 1 (8,3) | 3 (17,6) |
| 195 | D-loop (HVSII) | T>C | non coding | 9 (75) | 13 (76,5) |
| 203 | D-loop (HVSII) | G>A | non coding | 1 (8,3) | 0 (0) |
| 215 | D-loop (HVSII) | A>G | non coding | 9 (75) | 12 (70,6) |
| 295 | D-loop (HVSII) | C>T | non coding | 12 (100) | 17 (100) |
| 319 | D-loop (HVSII) | T>C | non coding | 9 (75) | 13 (76,5) |
| 430 | D-loop | T>C | non coding | 0 (0) | 2 (11,8) |
| 489 | D-loop (HVSIII) | T>C | non coding | 12 (100) | 17 (100) |
| 513 | D-loop (HVSIII) | G>A | non coding | 9 (75) | 13 (76,5) |
| 569 | D-loop (HVSIII) | C>T | non coding | 1 (8,3) | 0 (0) |
| 709 | rRNA 12S | G>A | - | 1 (8,3) | 1 (5,9) |
| 750 | rRNA 12S | A>G | - | 12 (100) | 17 (100) |
| 1438 | rRNA 12S | A>G | - | 12 (100) | 17 (100) |
| 1850 | rRNA 16S | T>C | - | 9 (75) | 13 (76,5) |
| 2706 | rRNA 16S | A>G | - | 12 (100) | 17 (100) |
| 2824 | rRNA 16S | C>G | - | 1 (8,3) | 0 (0) |
| 3447 | NADH dehydrogenase subunit 1 | A>G | synonymous | 4 (33,3) | 5 (29,4) |
| 3915 | NADH dehydrogenase subunit 1 | G>A | synonymous | 1 (8,3) | 0 (0) |
| *3930* | NADH dehydrogenase subunit 1 | C>T | synonymous | 0 (0) | 1 (5,9) |
| **4216** | NADH dehydrogenase subunit 1 | T>C | Tyr Y > His H | 12 (100) | 17 (100) |
| **4232** | NADH dehydrogenase subunit 1 | T>C | Ile I > Thr T | 0 (0) | 1 (5,9) |
| 4769 | tRNa isoleucine | A>G | - | 12 (100) | 17 (100) |
| *5290* | NADH dehydrogenase subunit 2 | A>G | Asn N > Ser S | 0 (0) | 1 (5,9) |
| *5307* | NADH dehydrogenase subunit 2 | A>G | Thr T > Ala A | 0 (0) | 1 (5,9) |
| 5585 | - | G>A | - | 2 (16,7) | 0 (0) |
| 5633 | tRNa alanine | C>T | - | 3 (25) | 4 (23,5) |
| *5936* | Cytochrome c oxidase subunit I | C>A | Asn N > Lys K | 0 (0) | 1 (5,9) |
| *6019* | Cytochrome c oxidase subunit I | C>G | Ala A > Gly G | 0 (0) | 1 (5,9) |
| *6020* | Cytochrome c oxidase subunit I | C>G | synonymous | 0 (0) | 1 (5,9) |
| *6024* | Cytochrome c oxidase subunit I | C>A | Leu L > Met M | 0 (0) | 1 (5,9) |
| *6025* | Cytochrome c oxidase subunit I | T>A | Leu L > Gln Q | 0 (0) | 1 (5,9) |
| 6026 | Cytochrome c oxidase subunit I | G>A | synonymous | 0 (0) | 1 (5,9) |
| *6027* | Cytochrome c oxidase subunit I | G>A | Gly G > Ser S | 0 (0) | 1 (5,9) |
| 6029 | Cytochrome c oxidase subunit I | C>A | synonymous | 0 (0) | 1 (5,9) |

| | | | | | |
|---|---|---|---|---|---|
| 6378 | Cytochrome c oxidase subunit I | T>C | synonymous | 1 (8,3) | 0 (0) |
| *6447* | Cytochrome c oxidase subunit I | C>T | Pro P > Ser S | 0 (0) | 1 (5,9) |
| *6448* | Cytochrome c oxidase subunit I | C>A | Pro P > His H | 0 (0) | 1 (5,9) |
| *6731* | Cytochrome c oxidase subunit I | T>C | synonymous | 1 (8,3) | 0 (0) |
| 7028 | Cytochrome c oxidase subunit I | C>T | synonymous | 12 (100) | 16 (94,1) |
| *7302* | Cytochrome c oxidase subunit I | T>C | synonymous | 1 (8,3) | 0 (0) |
| 7476 | tRNA serine | C>T | - | 12 (100) | 17 (100) |
| 7501 | tRNA serine | T>C | - | 2 (16,7) | 2 (11,8) |
| *7690* | Cytochrome c oxidase subunit II | C>T | synonymous | 0 (0) | 1 (5,9) |
| 7789 | Cytochrome c oxidase subunit II | G>A | synonymous | 9 (75) | 12 (70,6) |
| 7960 | Cytochrome c oxidase subunit II | A>G | synonymous | 1 (8,3) | 0 (0) |
| *8245* | Cytochrome c oxidase subunit II | A>G | synonymous | 1 (8,3) | 0 (0) |
| **8860** | ATP synthase F0 subunit 6 | A>G | Thr T > Ala A | 12 (100) | 17 (100) |
| *8904* | ATP synthase F0 subunit 6 | C>T | synonymous | 0 (0) | 1 (5,9) |
| **9145** | ATP synthase F0 subunit 6 | G>A | Ala A> Thr T | 1 (8,3) | 0 (0) |
| *9344* | Cytochrome c oxidase subunit III | C>T | synonymous | 0 (0) | 1 (5,9) |
| **9477** | Cytochrome c oxidase subunit III | G>A | Val V> Ile I | 1 (8,3) | 0 (0) |
| *9788* | Cytochrome c oxidase subunit III | C>G | synonymous | 1 (8,3) | 0 (0) |
| *9791* | Cytochrome c oxidase subunit III | A>T | synonymous | 2 (16,7) | 0 (0) |
| **9856** | Cytochrome c oxidase subunit III | T>G | Ile I > Ser S | 0 (0) | 1 (5,9) |
| *9857* | Cytochrome c oxidase subunit III | C>G | Ile I > Met M | 0 (0) | 1 (5,9) |
| **9861** | Cytochrome c oxidase subunit III | T>C | Phe F > Leu L | 0 (0) | 1 (5,9) |
| 10172 | NADH dehydrogenase subunit 3 | G>A | synonymous | 3 (25) | 4 (23,5) |
| **10237** | NADH dehydrogenase subunit 4 | T>C | Ile I > Thr T | 1 (8,3) | 3 (17,6) |
| **10398** | NADH dehydrogenase subunit 5 | A>G | Thr T > Ala A | 12 (100) | 17 (100) |
| 10448 | tRNA arginine | T>C | - | 1 (8,3) | 0 (0) |
| 10499 | NADH dehydrogenase subunit 4L | A>G | synonymous | 9 (75) | 13 (76,5) |
| 10801 | NADH dehydrogenase subunit 4 | G>A | synonymous | 0 (0) | 1 (5,9) |
| 10961 | NADH dehydrogenase subunit 4 | C>T | synonymous | 1 (8,3) | 0 (0) |
| 10966 | NADH dehydrogenase subunit 4 | T>C | synonymous | 1 (8,3) | 0 (0) |
| 11251 | NADH dehydrogenase subunit 4 | A>G | synonymous | 12 (100) | 17 (100) |
| 11377 | NADH dehydrogenase subunit 4 | G>A | synonymous | 9 (75) | 13 (76,5) |
| 11719 | NADH dehydrogenase subunit 4 | G>A | synonymous | 12 (100) | 17 (100) |
| *11900* | NADH dehydrogenase subunit 4 | G>A | Val V > Met M | 1 (8,3) | 0 (0) |
| 12528 | NADH dehydrogenase subunit 5 | G>A | synonymous | 12 (100) | 17 (100) |
| 12612 | NADH dehydrogenase subunit 5 | A>G | synonymous | 12 (100) | 17 (100) |
| *13026* | NADH dehydrogenase subunit 5 | C>T | synonymous | 1 (8,3) | 0 (0) |
| **13708** | NADH dehydrogenase subunit 5 | G>A | Ala A> Thr T | 12 (100) | 17 (100) |
| 13722 | NADH dehydrogenase subunit 5 | A>G | synonymous | 9 (75) | 13 (76,5) |
| 14133 | NADH dehydrogenase subunit 5 | A>G | synonymous | 9 (75) | 13 (76,5) |
| **14180** | NADH dehydrogenase subunit 6 | T>C | synonymous | 0 (0) | 1 (5,9) |
| 14194 | NADH dehydrogenase subunit 6 | C>T | synonymous | 1 (8,3) | 0 (0) |
| **14759** | cytochrome b | C>A | Arg R> Gly G | 0 (0) | 1 (5,9) |
| **14766** | cytochrome b | C>T | Ile I > Thr T | 12 (100) | 17 (100) |
| *15014* | cytochrome b | T>C | Phe F > Leu L | 1 (8,3) | 0 (0) |
| **15191** | cytochrome b | T>C | Leu L > Met M | 1 (8,3) | 3 (17,6) |
| **15213** | cytochrome b | T>C | Ile I > Thr T | 0 (0) | 1 (5,9) |

| | | | | | |
|---|---|---|---|---|---|
| 15217 | cytochrome b | G>A | synonymous | 2 (16,7) | 0 (0) |
| **15257** | cytochrome b | G>A | Asp D > Asn N | 12 (100) | 17 (100) |
| **15326** | cytochrome b | A>G | Thr T > Ala A | 12 (100) | 17 (100) |
| **15452** | cytochrome b | C>A | Leu L > Ile I | 12 (100) | 17 (100) |
| 15613 | cytochrome b | A>G | synonymous | 0 (0) | 1 (5,9) |
| **15812** | cytochrome b | G>A | Val V > Met M | 3 (25) | 4 (23,5) |
| 15930 | tRNA threonine | G>A | - | 1 (8,3) | 0 (0) |
| 15983 | tRNA proline | T>C | - | 1 (8,3) | 0 (0) |
| 16037 | D-loop | A>G | non coding | 0 (0) | 1 (5,9) |
| 16069 | D-loop (HVS I) | C>T | non coding | 12 (100) | 17 (100) |
| 16086 | D-loop (HVS I) | T>C | non coding | 1 (8,3) | 0 (0) |
| 16126 | D-loop (HVS I) | T>C | non coding | 12 (100) | 17 (100) |
| 16145 | D-loop (HVS I) | G>A | non coding | 9 (75) | 13 (76,5) |
| 16168 | D-loop (HVS I) | C>T | non coding | 0 (0) | 1 (5,9) |
| 16172 | D-loop (HVS I) | T>C | non coding | 1 (8,3) | 3 (17,6) |
| 16193 | D-loop (HVS I) | C>T | non coding | 3 (25) | 4 (23,5) |
| 16220 | D-loop (HVS I) | A>G | non coding | 0 (0) | 1 (5,9) |
| 16231 | D-loop (HVS I) | T>C | non coding | 9 (75) | 13 (76,5) |
| 16261 | D-loop (HVS I) | C>T | non coding | 9 (75) | 13 (76,5) |
| 16278 | D-loop (HVS I) | C>T | non coding | 3 (25) | 3 (17,6) |
| 16299 | D-loop (HVS I) | A>G | non coding | 0 (0) | 1 (5,9) |
| 16301 | D-loop (HVS I) | C>T | non coding | 1 (8,3) | 1 (5,9) |
| 16311 | D-loop (HVS I) | T>C | non coding | 0 (0) | 1 (5,9) |
| 16355 | D-loop (HVS I) | C>T | non coding | 0 (0) | 2 (11,8) |
| 16519 | D-loop | T>C | non coding | 0 (0) | 3 (17,6) |

**Tab. 6.13**: Novel mutation are underlined and in italics, mutations resulting in an amino acid change are in bold.

In the table above, we can identify mutations which were previously observed to be connected to longevity. In particular T489C, G13708A and A14133G were associated to longevity by a study conducted by Bilal et al. (2008) and the mutation A10398G by Tanaka et al. (2002), even though they are all characteristic polymorphisms of J2 haplogroup.

We also analyzed the novel mutations and we found that there are eight new mutation in old subjects (two of them modify the aminoacid) and fifteen are in controls (ten of them imply an aminoacid modification) but we haven't found nothing of significative.

We have tried to compare the number of mutations along mtDNA molecule by mtDNA regions (**Tab. 6.14**) and we found that the accumulation of mutations in the Dloop is significantly higher in controls than in cases (p=0.0001).

| J2 | | |
|---|---|---|
| **mtDNA region** | **N. mutations in 90+** | **N. mutations in controls** |
| D-loop | 160 | 235 |
| ND1 | 17 | 24 |
| ND2 | 0 | 2 |
| ND3 | 3 | 4 |
| ND4L | 9 | 13 |
| ND4 | 39 | 51 |
| ND5 | 55 | 78 |
| ND6 | 1 | 1 |
| COI | 15 | 26 |
| COII | 11 | 13 |
| COIII | 4 | 4 |
| ATPase6 | 13 | 18 |
| ATPase8 | 0 | 0 |
| cyt b | 55 | 78 |
| rRNA total | 47 | 65 |
| tRNA total | 32 | 40 |

**Tab. 6.14:** Number of mutations found along the mtDNA molecule regions in 90+ subjects and younger controls.

In general, there are more mutation in controls than in 90+ subjects and in particular in COI, which is a part of IV complex, even when we consider the frequency of mutation in each mitochondrial gene.

As regard sporadic mutations (showing up in one sample only), they are observed in 23 mtDNAs positions in 90+ subjects while they are observed in 30 positions in controls subjects. Even though this difference between cases and controls is not significant (X-squared = 0.9245, p-value = 0.3363), it is very interesting to notice how mutations accumulate in each group. In general, In controls there are much more mutations than in nonagenarians (**Fig. 6.21**). These could mean that there are more deleterious mutations wich lead to a further decreasing of the performance. In fact controls have 5 non coding mutations, 10 synonymous mutations and 15 aminoacid changes. These mutations which change the aminoacid sequence hit NADH dehydrogenase subunits (I, II and VI), cytochrome c (subunit I and III) and cytochrome b. At the same time we found also among cases some 90+ subjects belonging to J2 haplogroup. First of all they have different mutations in different positions (ATP synthase F0 subunit VI, Cytochrome c oxidase subunit III, NADH dehydrogenase subunit

IV and cytochrome b), 4 non coding mutations and 11 synonymous mutations. Intringuingly, we notice that 90+ subjects' mtDNAs have 4 sporadic mutations hitting tRNA and rRNA genes, absolutely absent in the control subjects.

**SPORADIC MUTATIONS**

| | 90+ | controls |
|---|---|---|
| tRNA | **3** | 0 |
| rRNA | **1** | 0 |

We verified the significance of this finding by comparing the number of the tRNA + rRNA mutations with the number of mutations falling in the remaining of coding region (23 in siblings and 30 in controls). The $\chi^2$ test showed a strong significance (p=0.017).

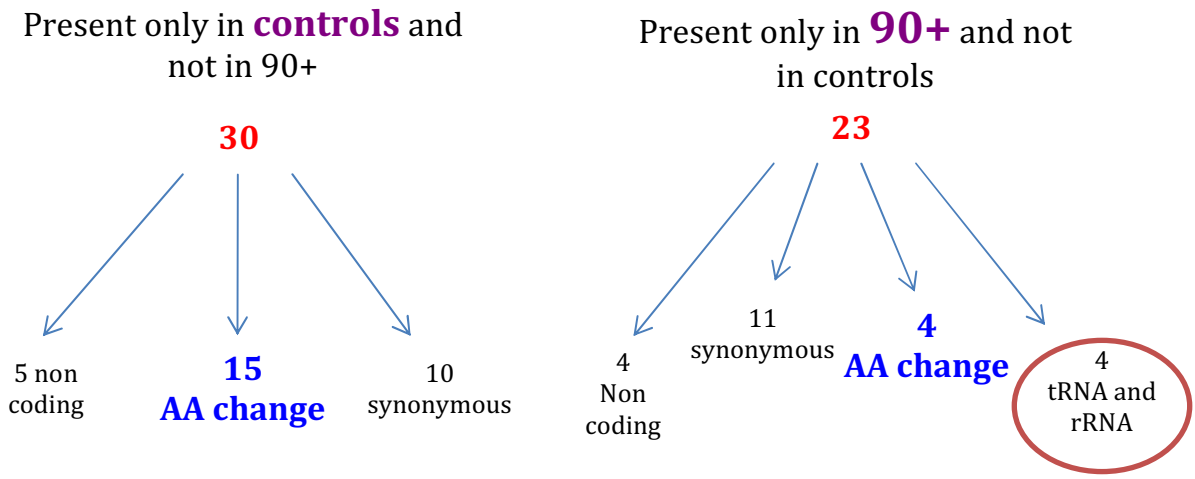



**Fig.6.21**: List of sporadic mutations found in controls and 90+ subjects. For details see the text above.

We then analyzed the non-synonymous mutations that affect subjects belonging to haplogroup J2, and we observed that the frequency of these mutations (the number of mutations in a given complex divided by the total number of only non-synonymous mutations) is larger and is statistically significant in the respiratory chain complex IV of the control subjects. In the other complexes it appears that the non synonymous mutations accumulate more among the nineties, even though there are not significant differences. But at what subunits of the respiratory chain the significance is due to? The significance is attributable to cytochrome c subunit I, which is significant with a p value = 0.0081 (**Fig. 6.22**).

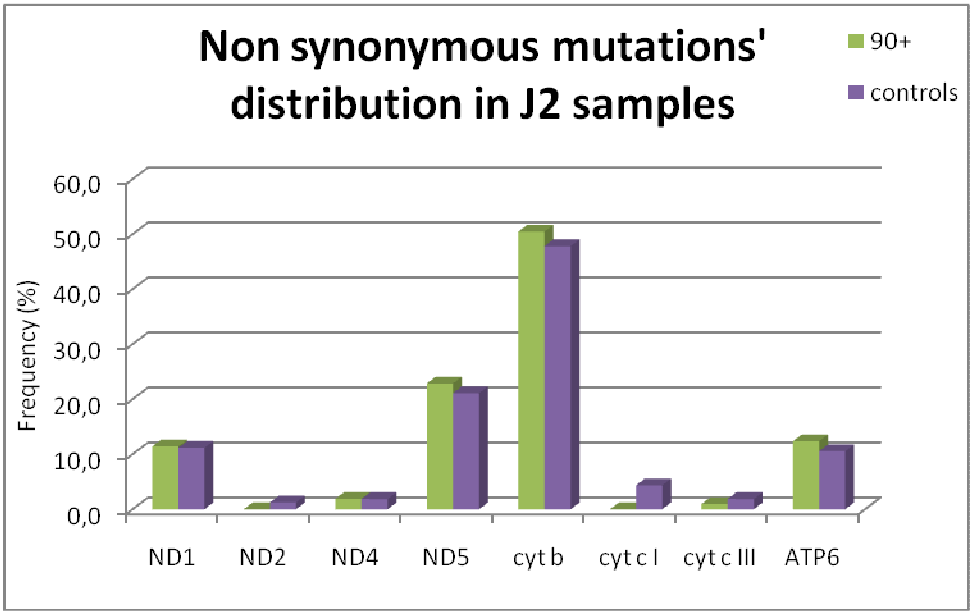



**Fig. 6.22**: distribution of non synonymous mutations in each gene. ND1, ND2, ND4, ND5 belong to complex I, cyt b to complex III, cyt c I and cyt c III to complex IV and ATP6 to complex V.

### 7.2 T2 haplogroup and complete sequencing

In order to assess if the T2 background of the patients harboured mutations responsible for the longevity phenotype, we investigated the complete sequences of T2 haplogroup subjects (compared to controls). The same analysis of J2 was performed for T2 haplogroup. The result of the sequence analysis are summarized in **Tab. 6.15**.

A total of 251 mutated positions relative to the reference sequence of Cambridge were detected and analyzed. Overall 46 of the total mutations (underlined and in italics) were not previously reported in either MITOMAP (www.mitomap.org) or mtDB (www.genpat.uu.se/mtDB). The majority of mutations were observed in a single mtDNA (or in 90+ subjects or in controls). There are other mutations that were detected both in 90+ and in controls (position 5322, 8492, 9788, 9790, 9791, 10496, 12363, 16236, 16276), other mutations are present only in 90+ subjects and other in controls (see beyond in sporadic mutations). This could explain how the increasing of mutation numbers in 90+ subjects is a favourable aspect and it is genetically positive.

Among all mutations observed in the 86 T2 mtDNA sequences, the mutation in position 189 was observed in 17.3% in 90+ subjects and in 2.9% in controls subjects and this difference is slightly significant (p=0.046). Also mutation in position 195 was observed only in 23.6% of controls subjects (p=0.011) and mutation C150T already associated to longevity (Rose et al. 2007) is present only in mtDNAs of 90+ subjects than

controls even if it is not resulted significant. Also mutation in position 189 is present in eight mtDNAs of sibling and only in one of controls (p=0.01141). All these mutations hit the dloop region. There is another mutation in position 3350 (ND1 subunit) which is present only in four mtDNAs of 90+ subjects, it does not change an amino acid and hence is not considered a candidate mutation even if it could have a functional role.

| **Nucleotide position** | **Locus** | **Nucleotide Change** | **Aminoacid Change** | **90+ subjects N=52 (%)** | **Controls N=34 (%)** |
|---|---|---|---|---|---|
| 41 | D-loop | C>T | non coding | 0 (0) | 1 (2,9) |
| 57 | D-loop (HVSII) | T>C | non coding | 1 (1,9) | 0 (0) |
| 61 | D-loop (HVSII) | C>T | non coding | 2 (3,8) | 1 (2,9) |
| 64 | D-loop (HVSII) | C>T | non coding | 1 (1,9) | 1 (2,9) |
| 73 | D-loop (HVSII) | A>G | non coding | 52 (100) | 33 (97,0) |
| 93 | D-loop (HVSII) | A>G | non coding | 1 (1,9) | 1 (2,9) |
| 95 | D-loop (HVSII) | A>C | non coding | 1 (1,9) | 1 (2,9) |
| 146 | D-loop (HVSII) | T>C | non coding | 8 (15,3) | 4 (11,8) |
| 150 | D-loop (HVSII) | C>T | non coding | 3 (5,8) | 0 (0) |
| 151 | D-loop (HVSII) | C>T | non coding | 1 (1,9) | 0 (0) |
| 152 | D-loop (HVSII) | T>C | non coding | 5 (9,6) | 8 (23,6) |
| 153 | D-loop (HVSII) | A>G | non coding | 2 (3,8) | 0 (0) |
| 189 | D-loop (HVSII) | A>G | non coding | 9 (17,3) | 1 (2,9) |
| 195 | D-loop (HVSII) | T>C | non coding | 0 (0) | 8 (23,6) |
| 198 | D-loop (HVSII) | C>T | non coding | 3 (5,8) | 1 (2,9) |
| 204 | D-loop (HVSII) | T>C | non coding | 1 (1,9) | 0 (0) |
| 215 | D-loop (HVSII) | A>G | non coding | 3 (5,8) | 1 (2,9) |
| 225 | D-loop (HVSII) | G>A | non coding | 1 (1,9) | 0 (0) |
| 227 | D-loop (HVSII) | A>G | non coding | 1 (1,9) | 0 (0) |
| 279 | D-loop (HVSII) | T>C | non coding | 4 (7,7) | 3 (8,8) |
| 297 | D-loop (HVSII) | A>C | non coding | 0 (0) | 1 (2,9) |
| 310 | D-loop (HVSII) | T>C | non coding | 0 (0) | 1 (2,9) |
| 316 | D-loop (HVSII) | G>A | non coding | 1 (1,9) | 1 (2,9) |
| 316 | D-loop (HVSII) | G>C | non coding | 1 (1,9) | 1 (2,9) |
| *321* | D-loop (HVSII) | T>C | non coding | 1 (1,9) | 0 (0) |
| 324 | D-loop (HVSII) | C>G | non coding | 0 (0) | 1 (2,9) |
| 330 | D-loop (HVSII) | C>A | non coding | 0 (0) | 1 (2,9) |
| *350* | D-loop (HVSII) | A>C | non coding | 0 (0) | 1 (2,9) |
| 385 | D-loop | A>G | non coding | 1 (1,9) | 0 (0) |
| 389 | D-loop | G>A | non coding | 2 (3,8) | 0 (0) |
| 513 | D-loop (HVSIII) | G>A | non coding | 1 (1,9) | 0 (0) |
| 709 | rRNA 12S | G>A | - | 52 (100) | 34 (100) |
| 750 | rRNA 12S | A>G | - | 52 (100) | 34 (100) |
| 930 | rRNA 12S | G>A | - | 28 (53,8) | 14 (41,2) |
| 1420 | rRNA 12S | T>C | - | 1 (1,9) | 0 (0) |
| 1438 | rRNA 12S | A>G | - | 52 (100) | 34 (100) |

| | | | | | |
|---|---|---|---|---|---|
| *1530* | rRNA 12S | A>G | - | 1 (1,9) | 0 (0) |
| 1625 | tRNA valine | A>G | - | 1 (1,9) | 0 (0) |
| 1888 | rRNA 16S | G>A | - | 52 (100) | 34 (100) |
| 2141 | rRNA 16S | T>C | - | 6 (11,5) | 2 (5,9) |
| *2412* | rRNA 16S | A>G | - | 1 (1,9) | 0 (0) |
| 2706 | rRNA 16S | A>G | - | 52 (100) | 34 (100) |
| *2707* | rRNA 16S | A>G | - | 1 (1,9) | 0 (0) |
| *2780* | rRNA 16S | C>G | - | 1 (1,9) | 0 (0) |
| 2850 | rRNA 16S | T>C | - | 1 (1,9) | 3 (8,8) |
| 3010 | rRNA 16S | G>A | - | 1 (1,9) | 0 (0) |
| 3105 | rRNA 16S | A>G | - | 1 (1,9) | 0 (0) |
| 3335 | NADH dehydrogenase subunit 1 | T>C | synonymous | 1 (1,9) | 0 (0) |
| 3338 | NADH dehydrogenase subunit 1 | T>C | synonymous | 1 (1,9) | 1 (2,9) |
| 3350 | NADH dehydrogenase subunit 1 | T>C | synonymous | 4 (7,7) | 0 (0) |
| **3394** | NADH dehydrogenase subunit 1 | T>C | Tyr Y> His H | 0 (0) | 1 (2,9) |
| **3398** | NADH dehydrogenase subunit 1 | T>C | Met M > Thr T | 1 (1,9) | 0 (0) |
| **3511** | NADH dehydrogenase subunit 1 | A>G | Thr T > Ala A | 1 (1,9) | 0 (0) |
| 3549 | NADH dehydrogenase subunit 1 | C>T | synonymous | 0 (0) | 1 (2,9) |
| 3552 | NADH dehydrogenase subunit 1 | T>C | synonymous | 1 (1,9) | 0 (0) |
| 3633 | NADH dehydrogenase subunit 1 | T>C | synonymous | 0 (0) | 1 (2,9) |
| 3826 | NADH dehydrogenase subunit 1 | T>C | synonymous | 0 (0) | 3 (8,8) |
| *3867* | NADH dehydrogenase subunit 1 | C>T | synonymous | 0 (0) | 1 (2,9) |
| **4216** | NADH dehydrogenase subunit 1 | T>C | Tyr Y > His H | 52 (100) | 34 (100) |
| *4246* | NADH dehydrogenase subunit 1 | A>C | Ile I > Leu L | 1 (1,9) | 0 (0) |
| **4491** | NADH dehydrogenase subunit 2 | G>A | Val V > Ile I | 1 (1,9) | 1 (2,9) |
| 4688 | NADH dehydrogenase subunit 2 | T>C | synonymous | 1 (1,9) | 1 (2,9) |
| 4769 | NADH dehydrogenase subunit 2 | A>G | synonymous | 52 (100) | 34 (100) |
| 4859 | NADH dehydrogenase subunit 2 | T>C | synonymous | 1 (1,9) | 1 (2,9) |
| 4913 | NADH dehydrogenase subunit 2 | A>C | synonymous | 0 (0) | 1 (2,9) |
| **4917** | NADH dehydrogenase subunit 2 | A>G | Asn N-Asp D | 52 (100) | 34 (100) |
| **4924** | NADH dehydrogenase subunit 2 | G>C | Ser S > Thr T | 1 (1,9) | 0 (0) |
| *4961* | NADH dehydrogenase subunit 2 | A>G | synonymous | 1 (1,9) | 0 (0) |
| 5147 | NADH dehydrogenase subunit 2 | G>A | synonymous | 24 (46,1) | 13 (38,2) |
| 5187 | NADH dehydrogenase subunit 2 | C>T | synonymous | 4 (7,7) | 3 (8,8) |
| **5277** | NADH dehydrogenase subunit 2 | T>C | Phe F > Leu L | 2 (3,8) | 5 (14,7) |
| **5319** | NADH dehydrogenase subunit 2 | A>G | Thr T > Ala A | 1 (1,9) | 1 (2,9) |
| *5322* | NADH dehydrogenase subunit 2 | A>C | synonymous | 2 (3,8) | 1 (2,9) |
| 5426 | NADH dehydrogenase subunit 2 | T>C | synonymous | 2 (3,8) | 5 (14,7) |
| **5480** | NADH dehydrogenase subunit 2 | A>G | Ala A > Thr T | 1 (1,9) | 0 (0) |
| 5527 | tRNA tryptophan | A>G | - | 1 (1,9) | 0 (0) |
| 5567 | tRNA tryptophan | T>C | - | 0 (0) | 1 (2,9) |
| 5580 | - | T>C | - | 2 (3,8) | 0 (0) |
| **6249** | Cytochrome c oxidase subunit I | G>A | Ala A > Thr T | 1 (1,9) | 0 (0) |
| **6261** | Cytochrome c oxidase subunit I | G>A | Ala A > Thr T | 4 (7,7) | 3 (8,8) |
| 6293 | Cytochrome c oxidase subunit I | T>C | synonymous | 0 (0) | 1 (2,9) |
| **6489** | Cytochrome c oxidase subunit I | C>A | Leu L>Ile I | 2 (3,8) | 5 (14,7) |
| 6524 | Cytochrome c oxidase subunit I | T>C | synonymous | 1 (1,9) | 0 (0) |

| | | | | | |
|---|---|---|---|---|---|
| 6899 | Cytochrome c oxidase subunit I | G>A | synonymous | 1 (1,9) | 0 (0) |
| 7022 | Cytochrome c oxidase subunit I | T>C | synonymous | 1 (1,9) | 3 (8,8) |
| 7028 | Cytochrome c oxidase subunit I | C>T | synonymous | 52 (100) | 34 (100) |
| **7041** | Cytochrome c oxidase subunit I | G>A | Val V > Ile I | 1 (1,9) | 0 (0) |
| 7076 | Cytochrome c oxidase subunit I | A>G | synonymous | 0 (0) | 1 (2,9) |
| *7100* | Cytochrome c oxidase subunit I | A>G | synonymous | 1 (1,9) | 0 (0) |
| 7268 | Cytochrome c oxidase subunit I | T>C | synonymous | 0 (0) | 1 (2,9) |
| **7684** | Cytochrome c oxidase subunit II | T>C | synonymous | 1 (1,9) | 0 (0) |
| *7685* | Cytochrome c oxidase subunit II | A>G | Ile I > Val V | 1 (1,9) | 0 (0) |
| **7853** | Cytochrome c oxidase subunit II | G>A | Val V > Ile I | 1 (1,9) | 0 (0) |
| 7873 | Cytochrome c oxidase subunit II | C>T | synonymous | 4 (7,7) | 3 (8,8) |
| 7891 | Cytochrome c oxidase subunit II | C>T | synonymous | 1 (1,9) | 1 (2,9) |
| *7979* | Cytochrome c oxidase subunit II | G>A | Asp D > Asp N | 0 (0) | 1 (2,9) |
| *8041* | Cytochrome c oxidase subunit II | A>G | Met M > Val V | 1 (1,9) | 1 (2,9) |
| *8256* | Cytochrome c oxidase subunit II | T>C | Val V > Ala A | 2 (3,8) | 0 (0) |
| 8270 | - | C>T | - | 1 (1,9) | 0 (0) |
| *8416* | ATP synthase F0 subunit 8 | C>T | synonymous | 0 (0) | 1 (2,9) |
| *8492* | ATP synthase F0 subunit 8 | A>G | Lys K > Asp D | 1 (1,9) | 1 (2,9) |
| 8572 | ATP synthase F0 subunit 8 | G>C | synonymous | 0 (0) | 2 (5,9) |
| 8697 | ATP synthase F0 subunit 6 | G>A | synonymous | 52 (100) | 34 (100) |
| **8860** | ATP synthase F0 subunit 6 | A>G | Thr T > Ala A | 52 (100) | 34 (100) |
| **8944** | ATP synthase F0 subunit 6 | A>G | Met M > Val V | 1 (1,9) | 1 (2,9) |
| **9053** | ATP synthase F0 subunit 6 | G>A | Ser S > Asn N | 1 (1,9) | 0 (0) |
| 9117 | ATP synthase F0 subunit 6 | T>C | synonymous | 6 (11,5) | 2 (5,9) |
| 9254 | Cytochrome c oxidase subunit III | A>G | synonymous | 6 (11,5) | 4 (11,8) |
| *9719* | Cytochrome c oxidase subunit III | C>A | synonymous | 1 (1,9) | 0 (0) |
| *9788* | Cytochrome c oxidase subunit III | C>G | synonymous | 1 (1,9) | 2 (5,9) |
| *9790* | Cytochrome c oxidase subunit III | C>T | Ser S > Stop | 1 (1,9) | 2 (5,9) |
| *9791* | Cytochrome c oxidase subunit III | A>T | synonymous | 1 (1,9) | 2 (5,9) |
| **9843** | Cytochrome c oxidase subunit III | A>G | Thr T > Ala A | 0 (0) | 1 (2,9) |
| 10000 | tRNA glycine | G>T | - | 1 (1,9) | 0 (0) |
| 10005 | tRNA glycine | A>G | - | 1 (1,9) | 0 (0) |
| *10111* | NADH dehydrogenase subunit 3 | T>A | Met M > Lys K | 2 (3,8) | 0 (0) |
| *10116* | NADH dehydrogenase subunit 3 | A>G | Ile I > Val V | 1 (1,9) | 0 (0) |
| *10243* | NADH dehydrogenase subunit 3 | T>C | Phe F > Ser S | 0 (0) | 1 (2,9) |
| 10403 | NADH dehydrogenase subunit 3 | A>G | synonymous | 1 (1,9) | 0 (0) |
| 10463 | tRNA arginine | T>C | - | 52 (100) | 34 (100) |
| *10496* | NADH dehydrogenase subunit 4L | A>G | synonymous | 1 (1,9) | 1 (2,9) |
| *10559* | NADH dehydrogenase subunit 4L | A>G | synonymous | 2 (3,8) | 0 (0) |
| 10589 | NADH dehydrogenase subunit 4L | G>A | synonymous | 1 (1,9) | 0 (0) |
| *10746* | NADH dehydrogenase subunit 4L | C>T | synonymous | 0 (0) | 1 (2,9) |
| **10750** | NADH dehydrogenase subunit 4L | A>G | Asn N > Ser S | 2 (3,8) | 3 (8,8) |
| 10822 | NADH dehydrogenase subunit 4 | C>T | synonymous | 4 (7,7) | 3 (8,8) |
| 10876 | NADH dehydrogenase subunit 4 | A>G | synonymous | 0 (0) | 1 (2,9) |
| *10879* | NADH dehydrogenase subunit 4 | A>C | synonymous | 0 (0) | 1 (2,9) |
| 10993 | NADH dehydrogenase subunit 4 | G>A | synonymous | 0 (0) | 1 (2,9) |
| 10997 | NADH dehydrogenase subunit 4 | A>C | Ser S > Arg R | 1 (1,9) | 0 (0) |

| | | | | | |
|---|---|---|---|---|---|
| **11016** | NADH dehydrogenase subunit 4 | G>A | Ser S > Asn N | 0 (0) | 1 (2,9) |
| 11020 | NADH dehydrogenase subunit 4 | A>G | synonymous | 1 (1,9) | 0 (0) |
| 11176 | NADH dehydrogenase subunit 4 | G>A | synonymous | 2 (3,8) | 0 (0) |
| 11251 | NADH dehydrogenase subunit 4 | A>G | synonymous | 52 (100) | 34 (100) |
| 11260 | NADH dehydrogenase subunit 4 | T>C | synonymous | 1 (1,9) | 0 (0) |
| 11290 | NADH dehydrogenase subunit 4 | A>G | synonymous | 1 (1,9) | 0 (0) |
| *11344* | NADH dehydrogenase subunit 4 | A>G | synonymous | 0 (0) | 1 (2,9) |
| *11395* | NADH dehydrogenase subunit 4 | C>T | synonymous | 0 (0) | 1 (2,9) |
| 11719 | NADH dehydrogenase subunit 4 | G>A | synonymous | 52 (100) | 34 (100) |
| 11812 | NADH dehydrogenase subunit 4 | A>G | synonymous | 52 (100) | 34 (100) |
| 11914 | NADH dehydrogenase subunit 4 | G>A | synonymous | 5 (9,6) | 3 (8,8) |
| 11944 | NADH dehydrogenase subunit 4 | T>C | synonymous | 0 (0) | 3 (8,8) |
| 12172 | tRNA histidine | A>G | - | 0 (0) | 2 (5,9) |
| **12341** | NADH dehydrogenase subunit 5 | C>T | Thr T > Ile I | 2 (3,8) | 0 (0) |
| **12358** | NADH dehydrogenase subunit 5 | A>G | Thr T > Ala A | 1 (1,9) | 0 (0) |
| *12363* | NADH dehydrogenase subunit 5 | C>T | synonymous | 3 (5,8) | 2 (5,9) |
| **12397** | NADH dehydrogenase subunit 5 | A>G | Thr T > Ala A | 1 (1,9) | 1 (2,9) |
| 12408 | NADH dehydrogenase subunit 5 | T>C | synonymous | 1 (1,9) | 0 (0) |
| 12481 | NADH dehydrogenase subunit 5 | T>A | Phe F > Ile I | 1 (1,9) | 0 (0) |
| 12501 | NADH dehydrogenase subunit 5 | G>A | synonymous | 0 (0) | 1 (2,9) |
| 12741 | NADH dehydrogenase subunit 5 | C>T | synonymous | 5 (9,6) | 1 (2,9) |
| 12771 | NADH dehydrogenase subunit 5 | G>A | synonymous | 1 (1,9) | 1 (2,9) |
| 13020 | NADH dehydrogenase subunit 5 | T>C | synonymous | 3 (5,8) | 1 (2,9) |
| 13050 | NADH dehydrogenase subunit 5 | A>G | synonymous | 1 (1,9) | 0 (0) |
| **13105** | NADH dehydrogenase subunit 5 | A>C | Ile I > Leu L | 0 (0) | 1 (2,9) |
| **13359** | NADH dehydrogenase subunit 5 | G>A | synonymous | 0 (0) | 1 (2,9) |
| 13368 | NADH dehydrogenase subunit 5 | G>A | synonymous | 52 (100) | 34 (100) |
| 13692 | NADH dehydrogenase subunit 5 | C>T | synonymous | 1 (1,9) | 1 (2,9) |
| 13722 | NADH dehydrogenase subunit 5 | A>G | synonymous | 2 (3,8) | 0 (0) |
| **13934** | NADH dehydrogenase subunit 5 | C>T | Thr T > Met M | 1 (1,9) | 0 (0) |
| 13965 | NADH dehydrogenase subunit 5 | T>C | synonymous | 11 (21,1) | 6 (17,6) |
| **13966** | NADH dehydrogenase subunit 5 | A>G | Thr T > Ala A | 6 (11,5) | 2 (5,9) |
| 13980 | NADH dehydrogenase subunit 5 | G>A | synonymous | 1 (1,9) | 0 (0) |
| 14097 | NADH dehydrogenase subunit 5 | C>T | synonymous | 1 (1,9) | 0 (0) |
| 14118 | NADH dehydrogenase subunit 5 | A>G | synonymous | 2 (3,8) | 1 (2,9) |
| 14233 | NADH dehydrogenase subunit 6 | A>G | synonymous | 52 (100) | 34 (100) |
| 14587 | NADH dehydrogenase subunit 6 | A>G | synonymous | 1 (1,9) | 0 (0) |
| 14687 | NADH dehydrogenase subunit 6 | A>G | synonymous | 8 (15,3) | 5 (14,7) |
| 14720 | tRNA glutamic acid | C>G | - | 1 (1,9) | 0 (0) |
| 14722 | tRNA glutamic acid | T>G | - | 1 (1,9) | 0 (0) |
| 14727 | tRNA glutamic acid | T>C | - | 2 (3,8) | 0 (0) |
| **14759** | cytochrome b | C>A | Arg R > Ser S | 1 (1,9) | 0 (0) |
| *14762* | cytochrome b | A>G | Lys K > Glu E | 2 (3,8) | 0 (0) |
| 14766 | cytochrome b | C>T | synonymous | 51 (98,1) | 34 (100) |
| *14819* | cytochrome b | T>C | Ser S > Pro P | 1 (1,9) | 0 (0) |
| 14905 | cytochrome b | G>A | synonymous | 52 (100) | 34 (100) |
| 14954 | cytochrome b | A>G | Thr T > Ala A | 0 (0) | 1 (2,9) |

| | | | | | |
|---|---|---|---|---|---|
| 15028 | cytochrome b | C>A | synonymous | 2 (3,8) | 5 (14,7) |
| 15043 | cytochrome b | G>A | synonymous | 2 (3,8) | 5 (14,7) |
| **15110** | cytochrome b | G>A | Ala A > Thr T | 1 (1,9) | 1 (2,9) |
| **15326** | cytochrome b | A>G | Thr T > Ala A | 52 (100) | 34 (100) |
| **15381** | cytochrome b | C>T | Thr T > Ile I | 1 (1,9) | 0 (0) |
| **15452** | cytochrome b | C>A | Leu L- Ile I | 52 (100) | 34 (100) |
| **15479** | cytochrome b | T>C | Phe F > Leu L | 2 (3,8) | 0 (0) |
| 15607 | cytochrome b | A>G | synonymous | 52 (100) | 34 (100) |
| *15608* | cytochrome b | C>G | Leu L > Val V | 0 (0) | 1 (2,9) |
| *15609* | cytochrome b | T>C | Leu L > Pro P | 0 (0) | 1 (2,9) |
| *15610* | cytochrome b | A>G | synonymous | 0 (0) | 1 (2,9) |
| **15758** | cytochrome b | A>G | Ile I > Val V | 2 (3,8) | 0 (0) |
| **15884** | cytochrome b | G>A | Ala A > Thr T | 1 (1,9) | 0 (0) |
| 15928 | - | G>A | - | 52 (100) | 34 (100) |
| 16037 | D-loop (HVSI) | A>G | non coding | 0 (0) | 1 (2,9) |
| 16126 | D-loop (HVSI) | T>C | non coding | 52 (100) | 34 (100) |
| 16129 | D-loop (HVSI) | G>A | non coding | 1 (1,9) | 0 (0) |
| 16140 | D-loop (HVSI) | T>C | non coding | 1 (1,9) | 0 (0) |
| 16153 | D-loop (HVSI) | G>A | non coding | 1 (1,9) | 0 (0) |
| 16172 | D-loop (HVSI) | T>C | non coding | 5 (9,6) | 3 (8,8) |
| 16182 | D-loop (HVSI) | A>C | non coding | 2 (3,8) | 5 (14,7) |
| 16183 | D-loop (HVSI) | A>C | non coding | 3 (5,8) | 6 (17,6) |
| 16184 | D-loop (HVSI) | C>T | non coding | 0 (0) | 1 (2,9) |
| 16189 | D-loop (HVSI) | T>C | non coding | 7 (13,4) | 7 (20,6) |
| 16194 | D-loop (HVSI) | A>C | non coding | 1 (1,9) | 0 (0) |
| 16195 | D-loop (HVSI) | T>G | non coding | 0 (0) | 1 (2,9) |
| *16197* | D-loop (HVSI) | C>G | non coding | 1 (1,9) | 0 (0) |
| 16201 | D-loop (HVSI) | C>A | non coding | 1 (1,9) | 0 (0) |
| *16204* | D-loop (HVSI) | G>A | non coding | 1 (1,9) | 0 (0) |
| *16205* | D-loop (HVSI) | C>A | non coding | 1 (1,9) | 0 (0) |
| *16208* | D-loop (HVSI) | G>A | non coding | 1 (1,9) | 0 (0) |
| 16209 | D-loop (HVSI) | T>A | non coding | 1 (1,9) | 0 (0) |
| *16211* | D-loop (HVSI) | C>A | non coding | 1 (1,9) | 0 (0) |
| 16213 | D-loop (HVSI) | G>A | non coding | 1 (1,9) | 0 (0) |
| 16214 | D-loop (HVSI) | C>A | non coding | 1 (1,9) | 1 (2,9) |
| 16218 | D-loop (HVSI) | C>A | non coding | 0 (0) | 1 (2,9) |
| 16224 | D-loop (HVSI) | T>C | non coding | 1 (1,9) | 0 (0) |
| *16228* | D-loop (HVSI) | C>A | non coding | 0 (0) | 1 (2,9) |
| *16228* | D-loop (HVSI) | C>T | non coding | 1 (1,9) | 0 (0) |
| 16232 | D-loop (HVSI) | C>A | non coding | 1 (1,9) | 0 (0) |
| 16234 | D-loop (HVSI) | C>A | non coding | 0 (0) | 1 (2,9) |
| *16236* | D-loop (HVSI) | C>A | non coding | 1 (1,9) | 1 (2,9) |
| 16245 | D-loop (HVSI) | C>G | non coding | 2 (3,8) | 1 (2,9) |
| 16245 | D-loop (HVSI) | C>T | non coding | 2 (3,8) | 1 (2,9) |
| 16247 | D-loop (HVSI) | A>G | non coding | 1 (1,9) | 0 (0) |
| 16255 | D-loop (HVSI) | G>A | non coding | 1 (1,9) | 0 (0) |
| 16258 | D-loop (HVSI) | A>C | non coding | 1 (1,9) | 1 (2,9) |

| | | | | | |
|---|---|---|---|---|---|
| 16261 | D-loop (HVSI) | C>T | non coding | 1 (1,9) | 0 (0) |
| 16263 | D-loop (HVSI) | T>C | non coding | 2 (3,8) | 0 (0) |
| 16265 | D-loop (HVSI) | A>C | non coding | 1 (1,9) | 1 (2,9) |
| 16266 | D-loop (HVSI) | C>T | non coding | 1 (1,9) | 0 (0) |
| 16269 | D-loop (HVSI) | A>C | non coding | 1 (1,9) | 1 (2,9) |
| *16276* | D-loop (HVSI) | T>A | non coding | 1 (1,9) | 1 (2,9) |
| 16282 | D-loop (HVSI) | C>A | non coding | 1 (1,9) | 1 (2,9) |
| 16291 | D-loop (HVSI) | C>T | non coding | 0 (0) | 1 (2,9) |
| 16292 | D-loop (HVSI) | C>T | non coding | 5 (9,6) | 4 (11,8) |
| 16294 | D-loop (HVSI) | C>T | non coding | 52 (100) | 34 (100) |
| 16295 | D-loop (HVSI) | C>T | non coding | 0 (0) | 1 (2,9) |
| 16296 | D-loop (HVSI) | C>T | non coding | 33 (63,5) | 20 (58,8) |
| 16297 | D-loop (HVSI) | T>C | non coding | 0 (0) | 1 (2,9) |
| 16298 | D-loop (HVSI) | T>C | non coding | 2 (3,8) | 5 (14,7) |
| 16299 | D-loop (HVSI) | A>G | non coding | 1 (1,9) | 1 (2,9) |
| 16304 | D-loop (HVSI) | T>C | non coding | 27 (51,9) | 14 (41,2) |
| *16308* | D-loop (HVSI) | T>A | non coding | 0 (0) | 1 (2,9) |
| 16310 | D-loop (HVSI) | G>T | non coding | 0 (0) | 1 (2,9) |
| 16311 | D-loop (HVSI) | T>C | non coding | 1 (1,9) | 0 (0) |
| 16313 | D-loop (HVSI) | C>A | non coding | 0 (0) | 1 (2,9) |
| *16315* | D-loop (HVSI) | T>A | non coding | 0 (0) | 1 (2,9) |
| 16320 | D-loop (HVSI) | C>T | non coding | 2 (3,8) | 0 (0) |
| 16322 | D-loop (HVSI) | A>T | non coding | 0 (0) | 1 (2,9) |
| 16324 | D-loop (HVSI) | T>C | non coding | 7 (13,4) | 3 (8,8) |
| 16368 | D-loop | T>C | non coding | 1 (1,9) | 0 (0) |
| 16519 | D-loop | T>C | non coding | 45 (86,5) | 30 (88,2) |

**Tab. 6.15**: Novel mutation are underlined and in italics, mutations resulting in an amino acid change are in bold.

We noticed that among all mutations listed in the above Tab.6.15, some of them were previously associated with longevity. In particular the mutation G3010A was connected to longevity by Tanaka et al (2002) an Bilal et al. (2008), such as for G15043A, G16129A and T16297C. They are all sporadic mutation present only in 90+ subjects except for the mutation T16297C, present in one control sequence.

We also tried to compare the number of mutations along mtDNA molecule by mtDNA regions (**Tab. 6.16**) and we did not find any statistical significant result. The complete re-sequencing of the 86 mtDNAs belonging to T2 revealed that 90+ subjects showed a trend towards a higher number of mutations in all genes when compared with controls. But if we consider the total number of mutations, the frequency of mutations in each region is higher in controls than in our cases.

| T2 | | |
|---|---|---|
| **mtDNA region** | **N. mutations in 90+** | **N. mutations in controls** |
| D-loop | 380 | 258 |
| ND1 | 62 | 42 |
| ND2 | 121 | 87 |
| ND3 | 4 | 1 |
| ND4L | 6 | 5 |
| ND4 | 171 | 117 |
| ND5 | 96 | 53 |
| ND6 | 61 | 39 |
| COI | 64 | 48 |
| COII | 11 | 6 |
| COIII | 10 | 11 |
| ATPase6 | 112 | 71 |
| ATPase8 | 1 | 4 |
| cyt b | 274 | 151 |
| rRNA total | 302 | 189 |
| tRNA total | 60 | 37 |

**Tab. 6.16:** Number of mutations found along the mtDNA molecule regions in 90+ subjects and younger controls.

Subsequently, we searched for groups of singleton mutations falling in specific mtDNA regions that may be associated with longevity. They are observed in 79 mtDNAs positions in 90+ subjects while they are observed in 47 positions in controls subjects. This difference between cases and controls is strongly significant (X-squared = 8.127, df = 1, p-value = 0.004361) and it is very interesting to notice how mutations accumulate in each group. In general, taking into account simply absolut numbers it appears that mutations seem to accumulate much more in cases than in controls, but if frequency is considered, the trend is opposite. In other words, mutations accumulate with higher frequency in controls than in cases, reflecting the same pathway of J2 haplogroup mutations. As **figure 6.23** shows, 29 non coding mutations, 18 synonymous mutations and 19 mutations which cause the aminoacid changes accumulate in 90+ subjects, whereas 20 non coding mutations, 17 synonymous mutations and 9 mutations changing the sequence of aminoacids accumulate in controls. As for the analysis of entire sequences resulting J2, also here we assist to a high accumulations of mutations in tRNA and rRNA genes in 90+ subjects.

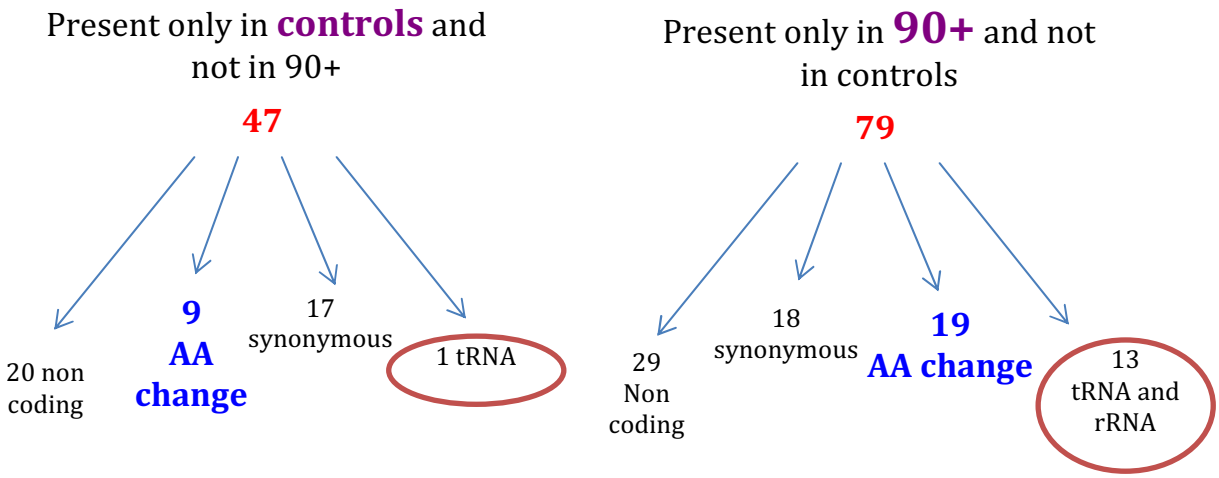



**Fig.6.23:** List of sporadic mutations found in controls and 90+ subjects. For details see the text above.

We found that 13 sporadic mutations (single occurrences) were present in tRNA and rRNA genes from 90+ subjects mtDNAs and only one sporadic mutation in controls.

**SPORADIC MUTATIONS**

| | 90+ | controls |
|---|---|---|
| tRNA | **6** | 1 |
| rRNA | **7** | 0 |

We verified the significance of this finding by comparing the number of the tRNA+rRNA mutations with the number of mutations falling in the remaining of coding region (79 in siblings and 47 in controls). The $\chi^2$ test showed a strong significance (p=0.014).

Finally, we have investigated the non-synonymous mutations that affect persons belonging haplogroup T2, and we discovered that the frequency of these mutations, as well as already performed for J2, is larger, but not significant, in the respiratory chain complex IV of the control subjects. Also in this haplogroup frequency of non-synonymous mutations in other genes seem to accumulate more in the nineties. In particular, once again there are more mutations in genes cyt c I and cyt c III but nothing is statistically significant (**Fig. 6.24**).

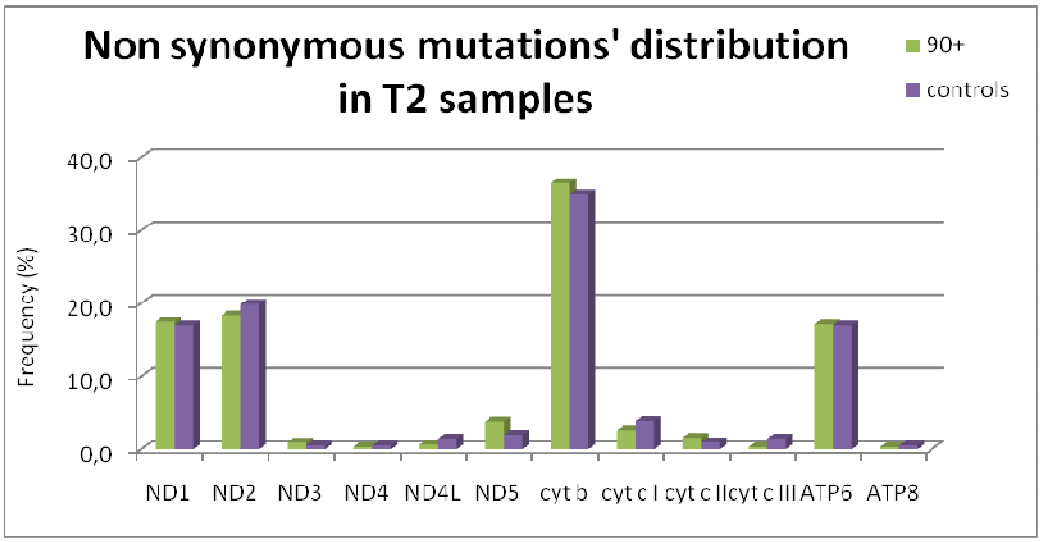



**Fig. 6.24**: distribution of non synonymous mutations in each gene. ND1, ND2, ND3, ND4, ND4L and ND5 belong to the complex I, cyt b to complex III, cyt c I, II and III to complex IV and ATP6 and ATP8 to complex V.

### 8. Cluster analysis

Cluster analysis or clustering (introduced by Robert Tryon in 1939) or group analysis, is the assignment of a set of observations into subsets (called *clusters*) so that observations in the same cluster are similar in some sense. All the clustering techniques calculate the distance between two elements based on Euclidean distance, or more simply the geometric distance in the multidimensional space. The good quality of analysis depends on how the distance is calculated. Clustering algorithms group together elements on the strength of their mutual distance, thus depends on how much the element is distant from the set.

We have applied this cluster analysis to our samples taking into account all subhaplogroups frequency distribution in different geographic areas (**Fig. 6.25**). To justify the countries aggregations, it has been conducted a similarity study among countries and parameters were haplogroups frequencies in cases and controls.

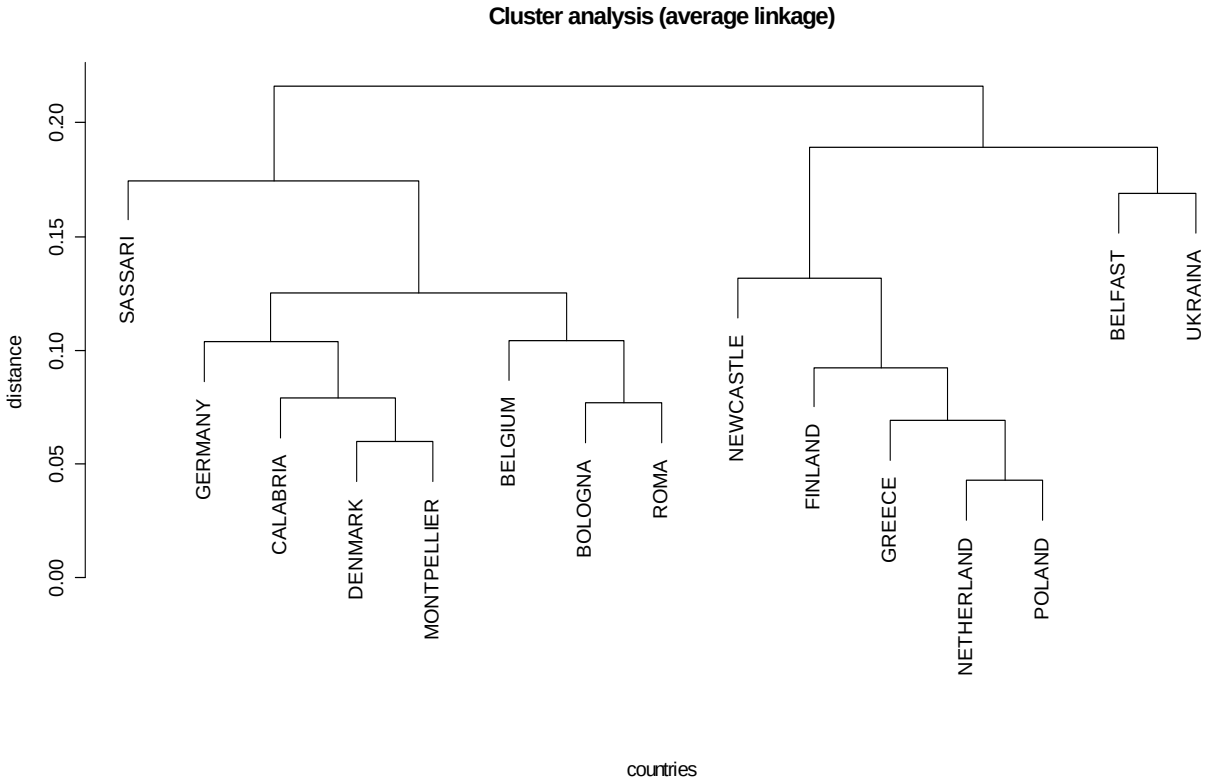



**Fig. 6.25:** Cluster analysis of all 15 geographic areas. On the left the first group includes Italian regions and Central Europe, on the right the second group includes North European countries, and the unexpected Greece.

More precisely, this aggregation has been built recording for each nation differences among frequencies related to cases and controls for all the haplogroups, except for N1, OTHER and R0A because too less numerous.

The analysis resulted in the two groups of samples, using a Chi-squared test to try out the significativity of cases/controls vs. haplogroups. In detail, we have obtained the following results. The first group includes Sassari, all Italian regions (Calabria, Bologna and Roma) and Central Europe countries (Denmark, Germany, Montpellier and Belgium). Sassari is the only geographic area that distance itself from the others, Bologna and Roma on one hand and Denmark and Montpellier on the other hand result similar. The distribution of this first group results significative (X-squared = 17.6093, p-value = 0.01386).

The only anomaly in the distribution of the second group is Greece which is thought to be more associated to Mediterranean countries (in particular to Calabria, representative of the South Europe) than to North European countries. Also in this case the distribution is statistically significative (X-squared = 19.513, p-value = 0.006723). In

order to conduct the same cluster analysis, we have forced the analysis by dividing the 11 countries into three groups, each representative of a single geographic area, as follows:

**1**. Finland

**2**. Europe: Denmark, Belfast, Newcastle, Netherland, Belgium, Germany, France, Poland, Ukraine

**3**. Mediterraneo: Bologna, Roma, Calabria, Sassari, Greece.

We decided to set Finland alone on the basis of a study conducted by Perola in 2008 on GEHA nuclear data. According to this study, Finland may distance itself from the other countries, as **Fig. 6.26**:

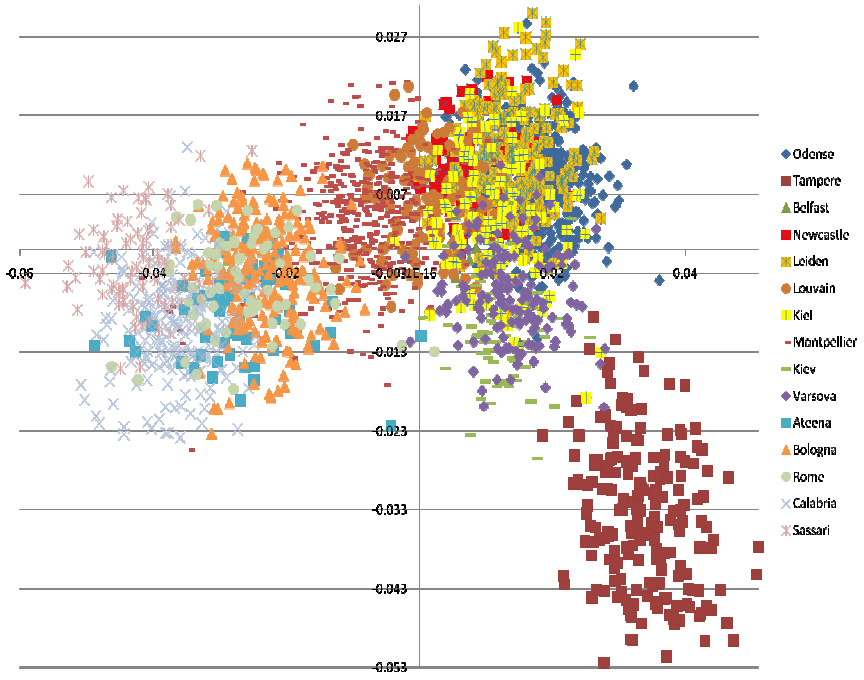



**Fig. 6.26**: PCA analysis on the 15 geographic areas (11 countries adherent to the Project) on the nuclear DNA. As we immediately notice Finland distance itself from other countries (red squared), while Greece is similar to Calabria, Sassari, Rome and Bologna (in the left side of the figure). We do not know the parameters used.

Then, we have obtained a matrix of distances in which value change from 0 (maximum similarity) to 1 (no similarity).

| | **Finland** | **Europe** | **Mediterranean area** |
|---|---|---|---|
| **Finland** | 0.000 | 0.083 | 0.115 |
| **Europe** | 0.083 | 0.000 | 0.030 |
| **Mediterranean area** | 0.115 | 0.030 | 0.000 |

As we expected, Finland is the country with the maximum divergence rate (red value) and whose haplogroup distribution differ from the rest of the analyzed countries.

The Mediterranean area (included Greece) is deeply divergent from Finland (value in orange), while the rest of Europe has an intermediate trend between Finland and Mediterranean area.

### 9. MDS – Multidimensional scaling

Multidimensional scaling (*MDS*) allows to explain observed similarities or dissimilarities (distances) between the investigated objects, by analyzing any kind of similarity or dissimilarity matrix, in addition to correlation matrices. In order to compare the distribution of haplotype variability in our samples and to confirm the genetic similarity between Finland and Greece, we calculated the genetic distances. Distance matrix was represented with not metric MDS using MASS library of R.

**Fig. 6.27** reports the distribution of all countries in controls subjects. We report only controls distribution because the high number of samples is more statistically significant than 90+ subjects.

All the dissimilarities were calculated using Hellinger distance, which is used to quantify the similarity between two probability distributions and it is defined in term of the "Hellinger integral". It considers the relative frequencies of control haplogroups related to the two compared countries.

We can notice that all countries are well distributed in a homogeneous cluster. All the Northern countries are in the middle of the graph, near among them. Belfast samples seems to be more different than other countries because it shows a high grade of isolation, such as Finland. In this MDS, again Greece is the nearest country to Finland, confirming our similarity between these two countries.

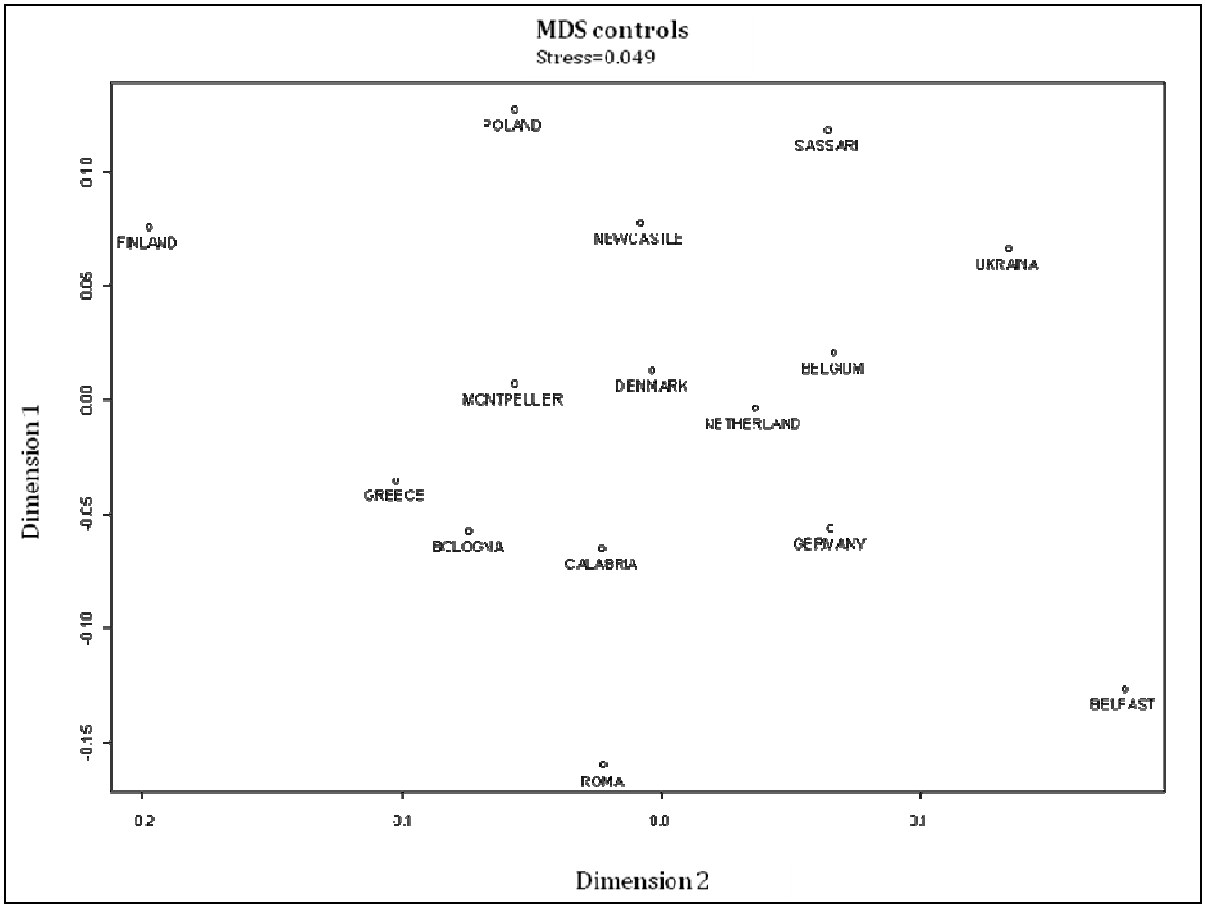



**Fig. 6.27:** MDS constructed on the basis of distance matrix of all our controls subjects belonging to all European areas.

### 10. Network of J haplogroup

The overall phylogenetic network, based on 281 J haplogroup females (137 sibs and 144 controls), 112 J haplogroup males (46 sibs and 66 controls) and their corresponding haplotypes is shown in **Fig. 6.28** (females) and **Fig. 6.31** (males). In order to analyze the different haplotype distribution among nations, a median-joining network was also constructed for the HVS-I and HVS-II sequence data (16080-300 segment region) of all samples resulted J.

Each node and solid edge could be labeled, but even with a small font the labels may detract from the overall presentation. Above on the left the legend has been reported.

A phylogenetic network for the variation scored in the 393 haplogroup J mtDNAs is shown.

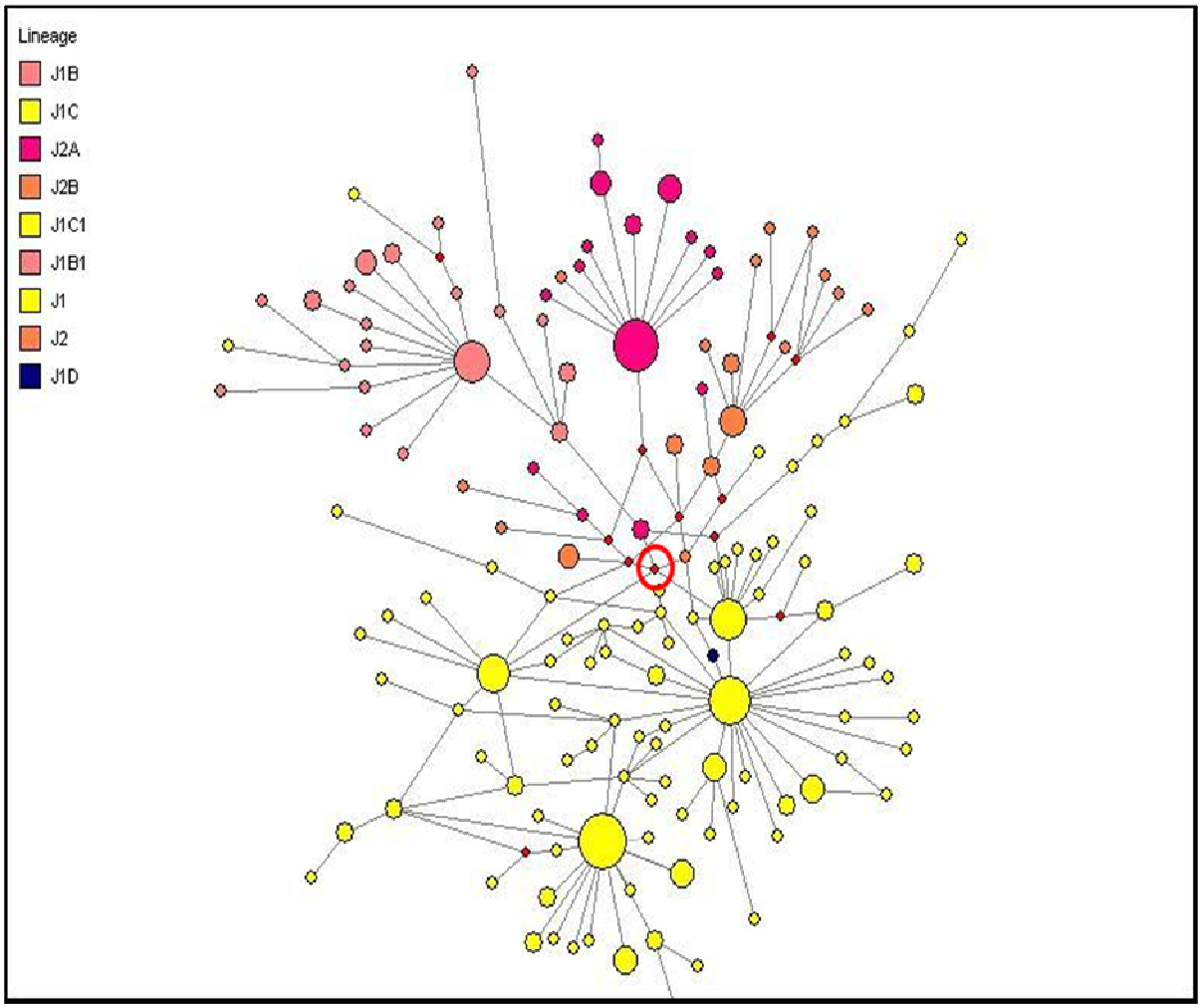



**Figure 6.28:** The J phylogenetic network (unlabelled) of all sibs and controls females. Circles are proportional to lineage frequencies. The red circle is the median vector 8, which connects the two phylogenetic branch of J2a and J2b. The area of each circle is proportional to the number of mtDNAs in the total sample harboring the corresponding haplotype**.** Lines represent one mutational step and red dots are hypothetical missing intermediates (median vectors).

The node corresponding to mv3 (median vector3 in red, **Fig. 6.28**) has been evidenced to emphasize that it is the point at which the network connects the two phylogenetic branch of J2a and J2b. In particular J2 is characterize by 150 and 152 mutations, J2a is defined by polymorphisms 195, 215, 16145, 16231 and 16261, J2b by 16193. With the red line under mutations in **Fig. 6.29**, there are typical polimorphisms of J. With the blue circle we indicate the median vector 9, which really divides J2a from J2b (also visible and clear in **Fig. 6.30**).

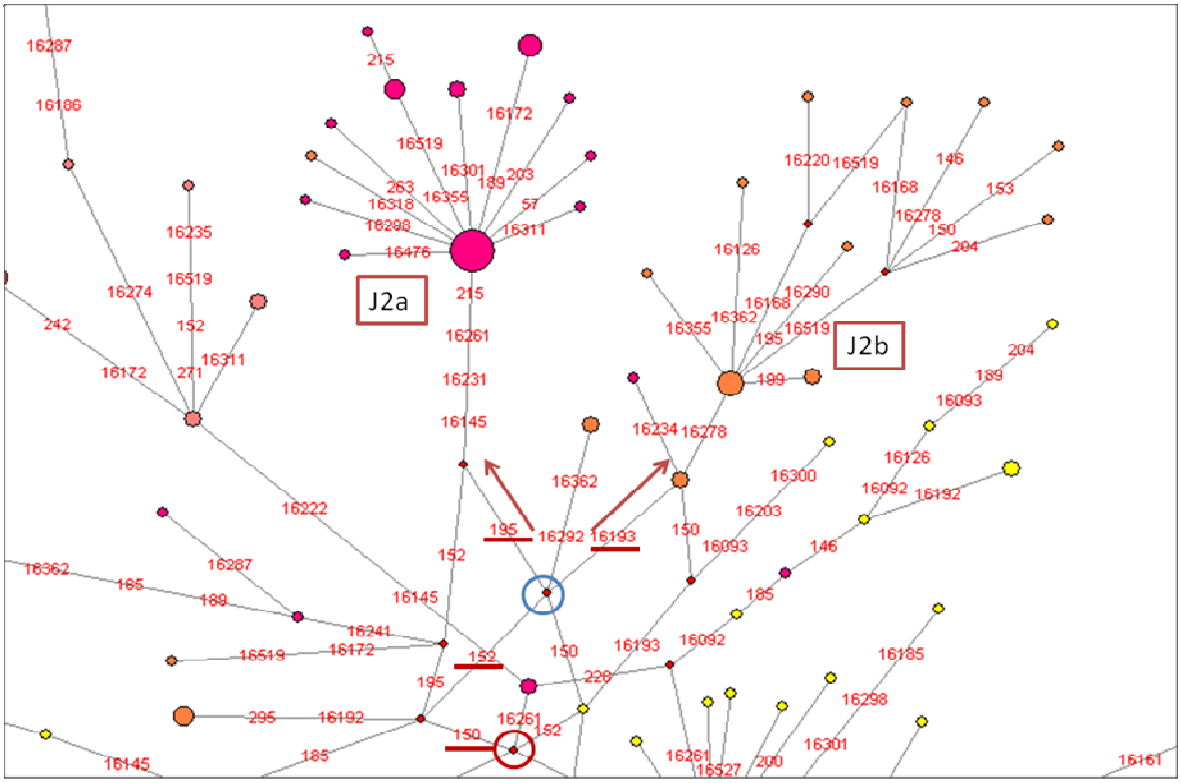



**Fig. 6.29:** The J phylogenetic network (unlabelled) of all sibs and controls females. Circles are proportional to lineage frequencies. The red circle is the median vector 3, which represent the start point of J haplogroup. The blue circle indicates the median vector 8, which divides J2a from J2b branch. The red lines under mutations are ancient polymorphisms typical of haplogroup J. The area of each circle is proportional to the number of mtDNAs in the total sample harboring the corresponding haplotype**.** Lines represent one mutational step and red dots are hypothetical missing intermediates.

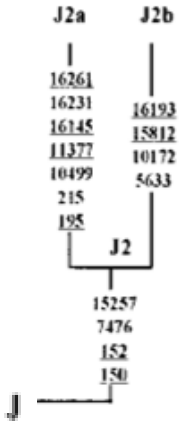



**Fig. 6.30**: J2 haplogroup branch and typical polymorphisms.

There are several obvious features of the phylogenetic network for the haplogroup. One of the most apparent is the subgroups J1 and J2 dominating the haplogroup. In female network (**Fig. 6.28**) subhaplogroup J2a, J2b and J1b take the form of star-like

clusters, whereas J1c do not exhibit the pattern of a dominant central node with more than one node, radiating out from it. The star-like shape of network indicates population expansion.

The network analysis has allowed us not only to revise and correct wrong haplogroups but also to collocate samples, whose haplogroup were general, in specific cluster (in our case, samples defined as J1 were collocated by Network in J1b subcluster). Another striking aspect of the network is that it is not a tree, but instead it contains many cycles or *reticulations.* A closer inspection of the edges in the network reveals that several of these cycles contain edges that correspond to mutations at nucleotide 228, typical of J1c.

As regard male network (**Fig. 6.31**), we observed that only J2a and J1b have a star-like cluster, J1c has different central nodes, as in females network, while J2b do not present a specific cluster. It is interesting to notice that in J2b cluster different cycles are present and in particular the most recurring mutation is 16261, representative of this haplogroup.

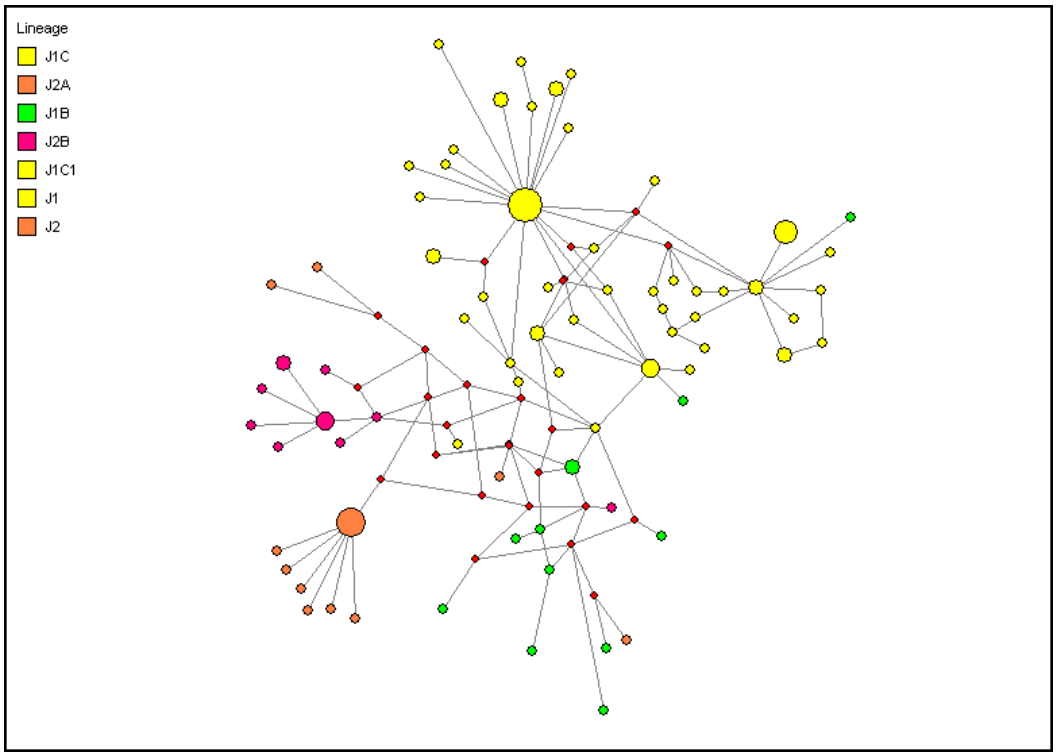



**Fig. 6.31:** The J phylogenetic network (unlabelled) of all sibs and controls males. Circles are proportional to lineage frequencies. The area of each circle is proportional to the number of mtDNAs in the total sample harboring the corresponding haplotype**.** Lines represent one mutational step and red dots are hypothetical missing intermediates.

The node corresponding to mv10 (median vector10, **Fig. 6.32**) has been evidenced to emphasize that it is the point at which the network connects the two phylogenetic branch of J2a and J2b. As for female network, J2 is characterized by 150 and 152 mutations, J2a is defined in this case by polymorphisms 195, 16145 and 16261, J2b by 16193. With the red line under mutations, there are typical polimorphisms of J.

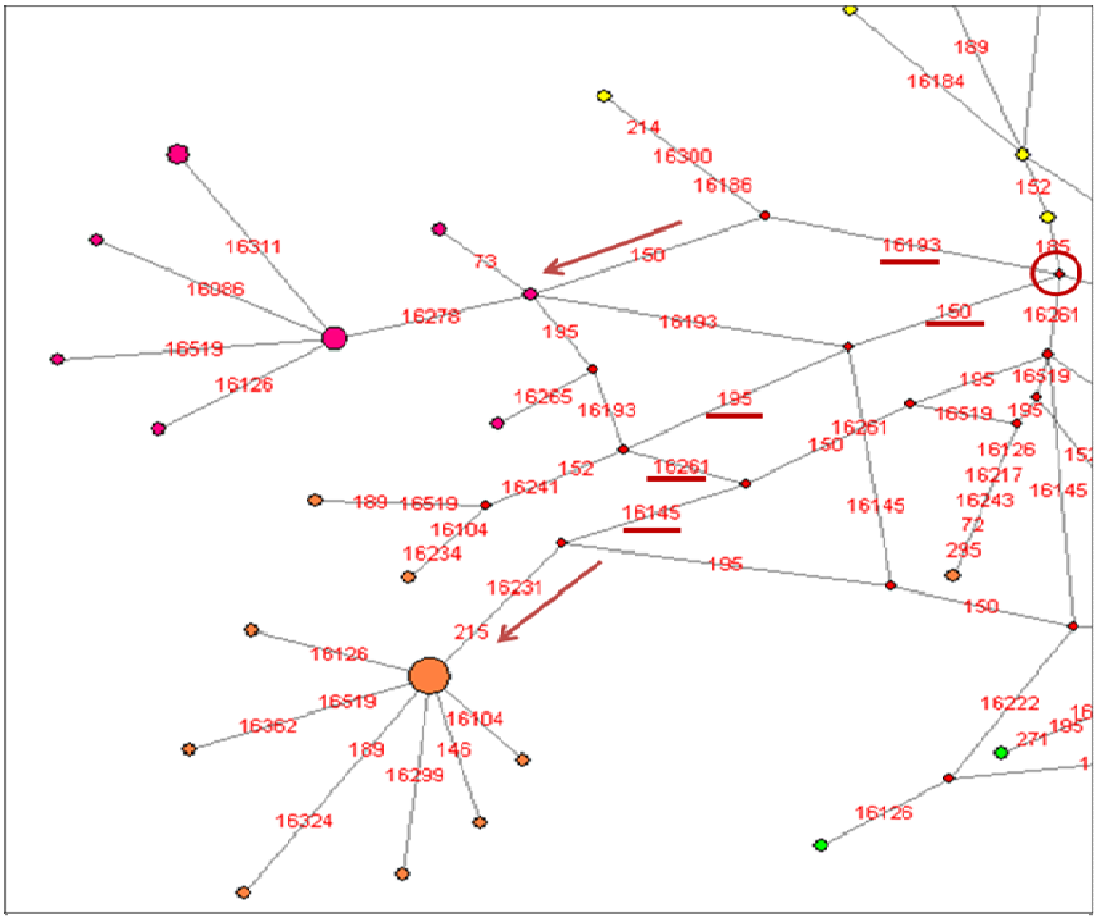



**Fig. 6.32:** The J phylogenetic network (unlabelled) of males. Circles are proportional to lineage frequencies. The red circle is the median vector 10, which connects the two phylogenetic branch of J2a and J2b. The red lines under mutations are ancient polymorphisms typical of haplogroup J.The area of each circle is proportional to the number of mtDNAs in the total sample harboring the corresponding haplotype**.** Lines represent one mutational step and red dots are hypothetical missing intermediates.

We have also tried to focus on geographical distribution in J network, in order to evidence a particular trend of Finland in females and males (**Fig. 6.33** and **Fig. 6.34**).

The distribution of countries haplotypes in the European-wide minimum spanning-network suggests multiple independent colonization events. Multiple unrelated founding events is one possible explanation for this pattern. In particular Finland seems to localize in the external branches, suggesting the fact that Finland is an

isolated country both in females and male networks. Instead Denmark is represented with higher frequency more than other countries in J2a and J1c in both gender. The star-like shape of the European-wide haplotype network strongly suggest sudden expansion.

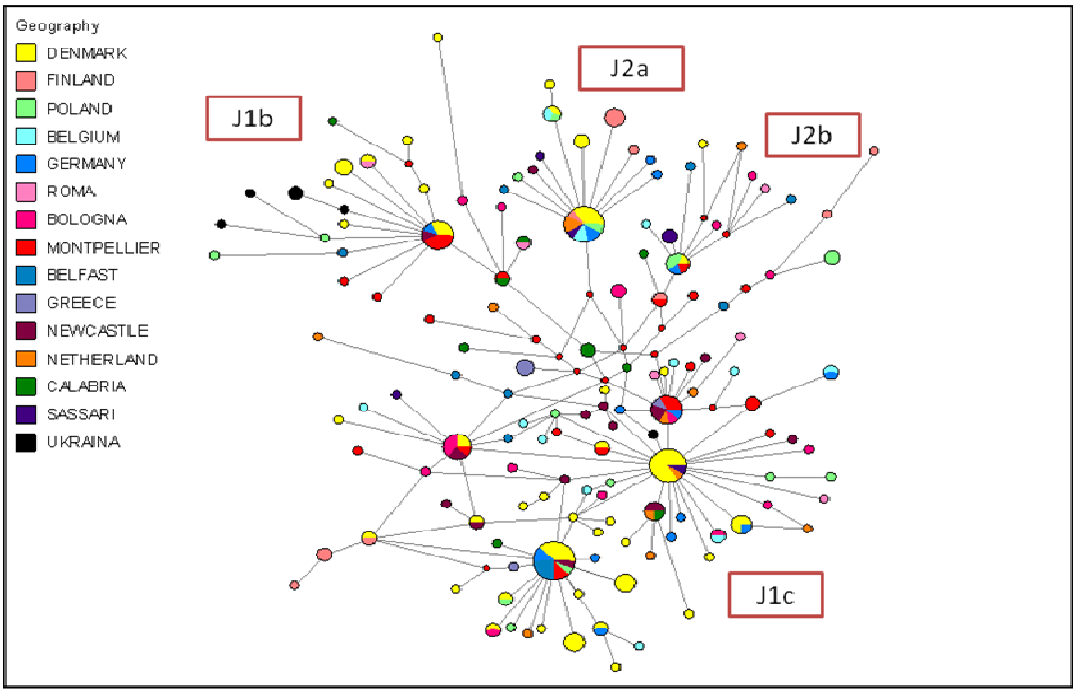



**Fig. 6.33:** The geographic representation of J females. Circles are proportional to lineage frequencies.

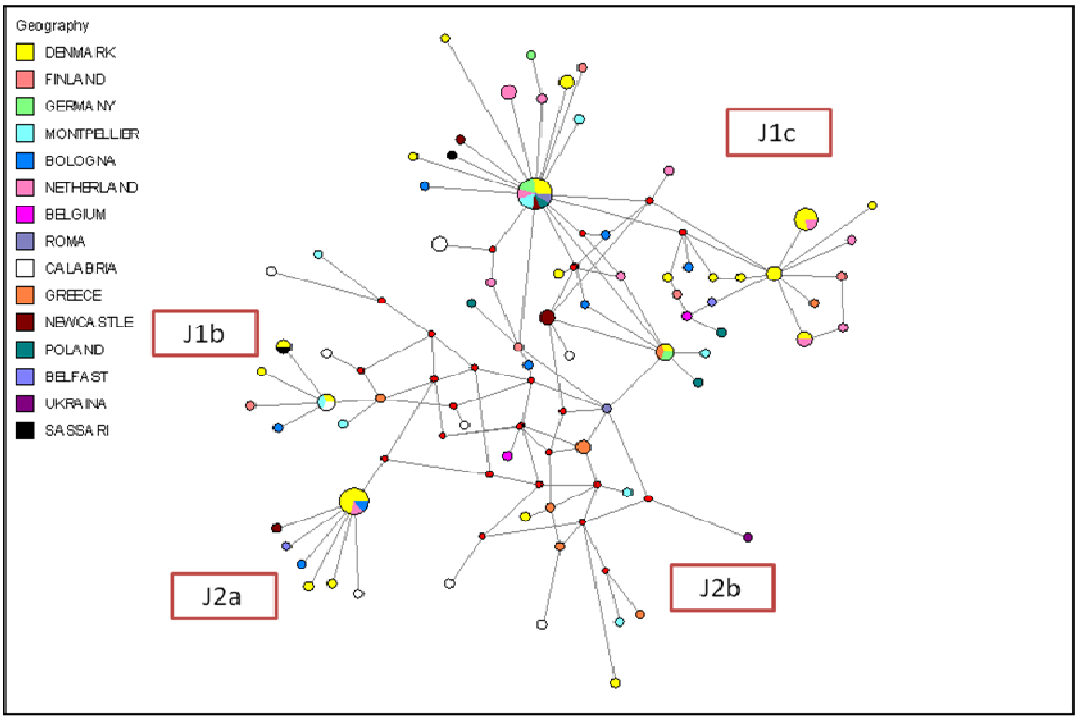



**Fig. 6.34:** The geographic representation of J males. Circles are proportional to lineage frequencies.

# Chapter 6

# DISCUSSION

Much evidence has accumulated on the association between mitochondrial DNA and the aging process. It is known that somatic mutations accumulate with aging suggesting a possible pathophysiology role for mtDNA in aging and senescence. On the other hand, several data demonstrate that the inherited mtDNA variability plays a role in longevity.

The present study is part of the European Project GEHA – GEnetic of Healthy Aging – whose the most important aim is to identify genes involved in healthy aging and longevity, which allows individuals to reach advanced old age in good cognitive and physical conditions, without the major age-related diseases. The GEHA Project represents the strongest and the most competitive consortium ever realized in Europe to investigate genetic bases of human aging process, capable of reaching results that is impossible to obtain in a single European country. The aim of the Project GEHA is to identify genes involved in healthy aging and longevity, allowing individuals to survive to advanced old age in good cognitive and physical function and in the absence of major age-related diseases, such as type II diabetes, neurodegenerative diseases, cardiovascular diseases and osteoporosis.

In the present study, we investigate the association of haplogroups, mitochondrial polymorphisms and mutation with longevity.

## 1. Haplogroups and association with male controls

In this study it has been applied a high resolution analysis, through the complete sequencing, the D-loop region sequencing and the restriction analysis of specific markers in the coding region of mtDNA. We determined and analyzed the haplogroup of a large cohort of 90+ subjects (N=2,086) and controls (N=2158) comparable for ethnicity and sex from 11 European countries, adherent to the GEHA Project and we wanted to test whether the analysis of mtDNA haplogroups is able to reveal any association between mtDNA inherited variability and longevity. The approach was a

comparative analysis between mtDNA of healthy 90+ subjects and younger controls matched for sex and geographic area, avoiding possible bias related to a founder effect or population heterogeneity.

From a descriptive analysis, we can observe that recruited females are in general more numerous than males in all countries except for Greece where the number of males is incredibly higher than females, both in cases and controls. We can say that the distribution is balanced among recruitment centre. The approach was a comparative analysis between our cases (90+ subjects) and the younger controls, matched for sex and geographic area.

Then we wanted to test whether the analysis of mtDNA haplogroups is able to reveal any association between mtDNA inherited variability and longevity.

The mitochondrial theory of aging proposes that the accumulation of mutations in mtDNA, caused by ROS (Reactive Species of Oxygen), is the mayor contributor to the cellular deterioration, leading to the aging process (Kowald and Kirkwood, 2000). Consistent with this theory is the enormous number of data in literature identifying mutations occurring with age. The fact that there is a possibility to inherit mtDNA polymorphisms which may predispose certain individuals to become nonagenarians or centenarians is supported by some studies. In fact the question of whether unusual longevity is linked to certain genetic markers has sparked much interest and resulted in a sizeable literature. From an evolutionary view, if there is a clear genetic component to longevity, and if longevity benefits fitness, long-lived individuals should have an evolutionary advantage, and their genes should be expected to become more frequent in a population across several generations.

Previous reviews (Madrigal et al, 2008; Capri et al, 2006) on candidate genes, which may result in unusual longevity, conclude that there is a consistent association between longevity and some apolipoprotein genes, genes involved in stress-response, and mtDNA. Of these, mitochondrial DNA has been the most frequently researched system. The hypothesis that longevity is associated with advantageous mtDNA markers has been tested mainly in individuals with unusual longevity, taken to mean individuals who live over 90 years of age (although a few studies focus on centenarians).

Specifically, three mutations (mt5178A, mt8414T, mt3010A) were found in significantly higher frequencies in Japanese centenarians (Tanaka et al, 1998; Alexe et al, 2007), while another variant (mt9055A) was found to be significantly more frequent in French centenarians (Ivanova et al, 1998).

Other studies have shown that centenarians from Northern Italy have a significantly different frequency of the J haplogroup than do younger controls (20 vs. about 2%) (De Benedictis et al, 1999). A higher frequency of the J haplogroup and a significantly high frequency of three mtDNA polymorphisms (150T, 489C, 10398G) has also been reported in Finnish long-lived subjects (Niemi et al, 2003).

Lastly, other studies report a significantly higher frequency of the 150T mutation in aged individuals in comparison with younger subjects in Finnish (Niemi et al, 2003), Japanese (Zhang et al, 2003) and Italian subjects (De Benedictis et al, 1999), although this association was not replicated by a study with Ashkenazi Jews (Shlush et al, 2008 ), who is a homogeneous population, due to a strong founder effect, followed by a rapid population expansions and characterized by high levels of consanguinity and endogamy.

The fact that both Finnish as well as the northern Italian study found an association of longevity, with the same mtDNA haplogroup J, certainly motivates further investigation. Several of these associations were not replicated in other studies (Iwata et al, 2007, Castri et al, 2009) suggesting that the association between mitochondrial DNA variants and longevity could be population-dependent (Dato et al, 2003). In this regard, Dominguez-Garrido et al (2009) found that J2 was overrepresented in elderly people in Pyrenees but not in people coming from the Ebro's Valley in Spain and they discovered that the former population have a lower mtDNA damage. It means that environmental condition can have a phenotypic survival advantage or disadvantage on population in study, demonstrating that the geographical altitude (Pyrenees Mountains), causing a lower oxygen pressure, determines lesser ROS production and reduces levels of mtDNA damage than in Valley one.

Our data demonstrates that there is not an increased frequency of haplogroup J within the aged population. In fact, in contrast to the findings of De Benedictis et al. who found that the J haplogroup was significantly associated with male centenarians of northern Italy, our study failed to show this association, even though we found that H1 and J2 significantly increased in the control males, thus representing a risk factor.

Among all males, H1 is more represented in controls than cases ($p=0.0427$), while among females the frequency is the same. As regard J2 haplogroup, it is resulted significant ($p=0.0214$). If we take into account gender considering the distribution among males, more than 80% are cases, as Fig.16 reports, while among females the frequency is equally distributed. These percentages mean that haplogroup J2 has an

higher frequency among male controls than cases and that J2 is not associated with female gender.

Furthermore, we have studied the distribution of H1 and J2 subhaplogroups in order to verify which subhaplogroup, if present, could influence the association. We studied H1 subhaplogroups (H1a, H1a1, H1a2, H1a3, H1b, H1c, H1c1, H1e, H1f, H1n) but we found no significant data. What about J2, we have studied the distribution of the subhaplogroups J2a and J2b, and we observed that only J2a is highly significant (p-value = 0.01701) rather than J2b subhaplogroup (p-value = 0.2789) and in particular among males about 80% are controls, confirming the J2 haplogroup trend. This fact suggests that the higher frequency of haplogroup J2 in male controls is attributable to an increase in subcluster J2a, rather than to J2b.

As these results demonstrates, it is very important to conduct a high resolution analysis by stratifying in subhaplogroups and for gender. For example, haplogroup H, the most common in Europe with a frequency of 30%-50%, is divided in numerous subhaplogroups, whose frequency is geographic-specific. Such a diversity could explain how differences in European countries are effectively noticed in association studies. Also, the strategy to collect subhaplogroups phylogenetically related, could be statistical informative, but not sufficiently biological exhaustive.

Therefore, we investigated the relationship between haplogroups and aging through the analysis of age quartile, which has directed us to study subjects with an age superior to 96 years. For men, we identified 94 years as median, 92 years is the lower quartile value and 96 years is the upper quartile value. For women, we identified 94 years as median, 93 years is the lower quartile value and 96 years is the upper quartile value. We decided to analyze the association between haplogroups and all 96+ years old subjects.

Firstly we verified the association of this group of individuals with J2 haplogroup. By applying a Fisher exact test, J1 seems to be no significantly associated with this group of subjects (p=0.5337), J2 is again significant and associated with male controls (p=0.03302), in fact among males, 88% is represented by controls.

Secondly, we calculated the significance for all subhaplogroups and also T2 haplogroup resulted slightly significant (p=0.05). T2 is more represented in male controls than females even though it is tendentially over-represented in females either cases and controls. In T2 subjects younger than 96 years old no significant association was observed. Instead, H1 is now not significant (p=0.09).

This finding is almost expected because in the overall phylogenetic tree, haplogroup T is closest to haplogroup J, which is characterised by the HVR1 motif 16069–16126 (Torroni et al. 1994; Richards et al. 1996) as well as coding region mutations at 4216, 10398, 11251, 12612, 13708, and 15452 (Torroni et al. 1994; Macaulay et al. 1999; Finnilä and Majamaa 2001). When considering HVR1 mutations, it is therefore the additional mutation at 16294 that defines haplogroup T, whereas haplogroup J is distinguished by the mutation at 16069.

As we can see from these data, age is a strong limit and this interaction suggests that age 96, corresponding to upper quartile of our samples, could be considered as a threshold. These data allowed us to identify the subgroup of 90+ subjects (older than 96 years of age) where J2 and T2 haplogroup has a stronger effect.

**2. Analysis of H1 and J2 in all European countries**

Therefore, we analyzed the distribution of H1 and J2 haplogroup in the 11 European countries in order to discover, where it is possible, an association with one or more countries. By applying a Pearson's Chi-squared test to each country in the H1 distribution, without taking into account gender, we verified that H1 is significantly associated with control belonging to Greece ($p=0.0455$), Poland ($p=0.01631$) and highly associated with Newcastle ($p=0.00604$). But H1 inverts its trend in Belfast, Belgium, Montpellier, Finland and Sassari (being more represented in 90+ subjects of different countries than controls) even if they are not significant. If we stratify for gender, we could find that H1 is significantly associated with male controls confirming this association (data not reported).

The same analysis has been used for J2 haplogroup distribution showing that J2 is associated with male controls belonging to Calabria ($p=0.008151$) and Greece ($p=0.02535$), which surprisingly represent the South Europe. As the distribution of this haplogroup evidences, J2 is not represented among 90+ subjects in Calabria and Greece but only among controls subjects. Even discarding Calabria and Greece, the distribution of J2 haplogroup is nearly balanced (for a total of 39 J2 cases and 45 J2 controls), but when considering these two areas, J2 seems to be more prevalent in controls. We conclude that Calabria and Greece determined this difference. The frequency in Calabria and Greece is high (respectively 4.6% and 4.8%) considering that J2 represents only 2% of the entire distribution (2.6% in controls and 1.8% in siblings).

The absence of association of mtDNA haplogroup J with longevity in southern Italian population (in Calabria) was in keeping with a study on southern Italian population (Dato et al. 2004) which revealed the absence of association of mtDNA haplogroup J with longevity in a southern Italian population. After that this association has been reported in three independent studies in northern Europeans, suggesting that the influence of mtDNA variability on longevity is population specific. Sequencing of HVS-I in Italian centenarians and controls has not revealed any clustering into a specific haplotype within haplogroup J (Rose et al. 2001). These data suggest that the association between mtDNA variants and longevity could be highly geographically or population dependent, as could be seen from other genetic studies on longevity (Franceschi et al. 2005).

Similarly, in the Japanese population, the C5178A transversion (characteristic of haplogroup D), was reported to be associated with longevity, being more frequent in centenarians than in a control group of younger subjects. But the same study confirmed an absence of association of haplogroup D with southern Chinese. This is likely to be due to the absence of old persons in the Chinese sample (the maximum age was 75 years old), but it is true that an association between haplogroup D and longevity exists and is specific of the Japanese population and thus absent in the Chinese group.

### 3. The analysis of complete sequences

Most of the pathological mtDNA mutations identified so far were probably the easiest to evidence at first, but according to many,they are only the most extreme fraction of a much larger group mutations that although "natural" are not necessarily "neutral". In recent years, for many other diseases and phenotypes, which lack a clear pattern of transmission, has been postulated a role for sequence variation in the "natural" sequence of the mtDNA and it was assumed that the "natural "forms of mtDNA, which can be very different from each other because of the high evolutionary rate of mtDNA, may modulate the expression not only of pathological mtDNA mutations, but also of nuclear genotypes. It is generally accepted that mtDNAs should be entirely sequenced. It is believed, in fact, that the entire sequence can hide non-functional polymorphisms that may be associated with a particular character and whose D-loop sequencing data does not give enough clarification on SNPs that characterize the genomic sequence.

The entire sequence also allows to identify all the polymorphisms, searching those really functional, and to identify any new mutations that may play a role in longevity and aging.

Finally, the entire sequence, although much more expensive and laborious, allow a better resolution of haplogroups' phylogeographic trees, which were widely documented (Torroni et al, 2006).

In order to assess whether J2 and T2 haplogroups harboured mutations involved in longevity, we investigated the complete mtDNA sequences of our J2 and T2 subjects among complete sequences. Since J2 and T2 are not present among 90+ subjects of Greece and Calabria, we cannot analyze them. We focused on the complete sequences of Denmark and Finland, sequenced by Chinese Partner, and more numerous than our samples. We did not find any particular mutation in J2, while in T2 we found some mutated position hitting the control region. In particular the mutation A189G is observed in 17.3% of siblings and in 2.9% controls and this difference is slightly statistically significant (p=0.046). It has been reported that CR mutation A189G accumulates with age in skeletal muscle (Wang et al. 2001, Zhang et al. 2003) and is germline transmitted polymorphism associated with specific mtDNA haplogroups (in this case with haplogroup T). Thus, it appears that each tissue may accumulate its own unique somatic mtDNA CR mutations with age, but some of these same variants might also be inherited.

Two additional mtDNA CR mutations has been detected, T195C and C150T. The first was observed only in mtDNAs controls in a percentage of 23.6% (p=0.011), the second is present only in mtDNAs of 90+ subjects and not in controls even if this distribution is not significant. The story of C150T is a bit complicated; it accumulates with age in skin fibroblasts, but is also present in the blood cell lymphocytes of centenarians and twins. Rose et al. (2007) reported the association of the C150T mutation with centenarians (and in particular the high level of heteroplasmy) suggesting to the authors that the C150T imparts resistance to stress and thus promotes longevity. Yet, this mutation has also been reported to be an inherited polymorphism in some instances.

We have compared the number of mutations along mtDNA molecule by mtDNA regions for J2 and T2 haplogroups. We found that the percentage of mutations in the D-loop of subjects resulted in J2 is significantly higher in controls than in cases (p=0.0001). In almost all regions, mutations seem to be more numerous in controls than cases. In

particular, ND1 and ND2 show a slightly higher percentage of mutation in controls than in cases, as well as the COI region shows a similar trend (4.29% vs controls. 3.50% in cases). While for T2 the frequency of mutation accumulation in each region is quite similar both in cases and controls. As for J2, we do not notice any significant difference between cases and controls in the regions of ND1 and ND2, while the percentage of mutations in the COI is higher (3.99% vs. 3.0%), once again, in controls than in cases. These data are not statistically significant but a trend that favors mutations in controls compared to cases is clear.

ND1 is thought that, if hit by mutations, it is able to bypass the "obstacle" or preventing the onset of OXPHOS itself, or using other mechanisms. The situation is quite different for the COI, which belongs to complex IV. If the COI is affected by many mutations, it is unable to cope with this situation and to bypass the site of damage. The OXPHOS crashes and this leads to an event even more dangerous with further accumulation of ROS.

In addition we found that sporadic mutations are more numerous in controls than in cases both in J2 and T2. As regard J2 complete sequences, 23 sporadic mutations are observed in in 90+ subjects while 30 in controls. This fact could mean that there are more deleterious mutations which could lead to a further decreasing of the performance in a very negative way. Controls have 5 non coding mutations, 9 synonymou mutations and 16 amminoacid changes. But we have found that there are 90+ subjects belonging to J2 haplogroup. Why? We investigated tha mutations which lead to amminoacid change and we discovered that they hit different genes. Intringuingly, we notice that 90+ subjects' mtDNAs have 4 sporadic mutations hitting tRNA and rRNA genes, totally absent in the control subjects. We already know that J2 haplogroup is characterized by a low OXPHOS performance and by a low ROS production as a consequence, as previously anticipated, and we hypotize that these mutations in tRNA and rRNAs could lead to a further decreasing of ROS production. This advantageous situation could compense the disadvantageous effect of J2. We investigated the accumulation of non-synonymous mutations both in nonagenarians and control subjects and we discovered that there is an higher number of these mutations in controls than in nonagenarians in complex IV and in particular in cyt c subunitI. The same was observed for T2 even though this data is not significant.

As regard T2 sporadic mutations' accumulation, we found the same trend as seen in J2. A number of 13 sporadic mutations tRNA plus rRNA genes from 90+ subjects

mtDNAs and only one sporadic mutation in controls. The same mechanisms can be hypotized for T2 haplogroup.

So the complete re-sequencing revealed that 90+ subjects showed a trend towards a higher number of sporadic mutations in tRNA and rRNA genes when compared with controls.

### 4. Final consideration on the association study

We have found an association between H1, J2 and T2 haplogroups and male controls.

Each of the mtDNA haplogroups is determined by a few ancient polymorphism, even though they harbour a great number of other nucleotide variants. Polymorphisms in mtDNA may be mildly deleterious, causing a subtle decrease in OXPHOS activity and an increase in the frequency of somatic mtDNA mutations. The differences in mtDNA haplogroup frequencies between the present 90+ group subjects and the controls suggest either a contribution from mildly deleterious polymorphisms that shorten the life span in the younger age groups or from advantageous polymorphisms that lengthen the life span in the elderly. Our data appear to favour the presence of disadvantageous polymorphisms and support a role for mitochondria and mtDNA in the degenerative processes involved in ageing.

It has been proposed that J haplogroup represents a *paradox* (Rose et al, 2001) because it is associated with longevity on one hand, but on the other hand it shows similar characteristic to that found in association with several complex diseases, for example the Leber Hereditary Optic Neuropathy (LHON). In fact, it has been found that haplogroup J seems to boost the effect of mutations causing this disease and thus contributing to optic neuritis in multiple sclerosis patients. Indeed, a relationship between OXPHOS performance and haplogroups exists.

Haplogroup J seems to have a *border line status* because a low OXPHOS performance could lead to a reducing in the production of ROS (through an increase in detoxifying enzymes, due to nuclear genes). In this way, a low OXPHOS performance may not necessarily be detrimental for the cell. On the other hand, this situation could put the cell in a vulnerable situation where a further single mutation (for example 11778 in LHON) would be even more damaging (see paragraph 8 of Chapter 2).

Intriguingly, haplogroup T shows a significantly less efficient OXPHOS respect to H (Ruiz-Pesini et al, 2000). Also the mutations defining the haplogroup J hit particularly the complex I protein subunits and partly share with haplogroup T.

It seems that there is a strong similarity between these two haplogroups, in fact we found that both are significantly associated with male controls.

But why were these phenomena not observed in females? Different findings support the difference in gene/aging association studies in males and females (Ivanova et al 1998). Longevity is a multifactorial trait in which a phenotypic effect of a gene depends on the physiological background where the gene is expressed. The effect of mtDNA variability on successful aging could vary between sexes, since males and females have a different physiological aspect. It is known that life expectancy is significantly higher in females than in males for a gender effect and probably for this reason we have found the association of J2, as risk factor, with male controls.

Another question is: why have not we found J2 represented in Southern Europe? The mitochondrial genome is highly variable and a continent-specific haplogroup may include mtDNA mutations that only occur in a specific ethnic group. Also genetic and environmental background can influence the effect of mtDNA mutations on a complex trait, such as longevity. It is extremely important to remember that the mtDNA haplogroup J, which is believed to have entered Europe about 10 000 years ago from the Near East, is characterized by a low efficiency of oxidative phosphorylation. This may favors the onset of either complex diseases or longevity, according to the genetic background of the carriers. It has been proposed that a low efficiency of oxidative phosphorylation leads to a waste of heat which represents an advantage in the cold climate of northern Europe; on turn, the cold climate of northern Europe seems to have favored the accumulation of further mutations emphasizing this feature of the J molecules. Therefore, the population-specific association of mtDNA haplogroup J with longevity may be due to population-specific genetic backgrounds, to particular interactions between haplogroup J and different environments, and/or to diversity of the J molecules between northern and southern European populations. In this regard, it is important to continue to have rapid and cheap genetic markers, such as haplogroups, to test the role of mtDNA on longevity in various populations from various geographical areas. For example, the Mt5178A mutation found in Japanese centenarians (Tanaka et al, 2000) and postulated to decelerate the accumulation of mtDNA mutations in somatic cells with advancing age, is included in the M haplogroup, but virtually absent in Europe.

Similarly, it is important to point out that any specific nucleotide, or haplogroup defining, mtDNA polymorphism associated with aging, may not directly involved but simply acting as a marker for other tightly linked polymorphisms occurring elsewhere in the mtDNA genome that directly could affect longevity.

### 5. The cluster analysis and MDS analysis

In this thesis, we have also conducted a cluster analysis on the differences of all haplogroups frequencies in all countries. In general the calculation of cluster analysis is based on the distance between two elements based on Euclidean distance, or more simply the geometric distance in the multidimensional space. The good quality of analysis depends on how the distance is calculated.

More precisely, this aggregation has been built recording for each nation differences among frequencies related to cases and controls for all the haplogroups, except for N1, OTHER and R0A because too less numerous. The cluster analysis evidenced two groups, both statistical significative: one including all Calabria, Bologna, Roma and Sassari belonging to Italy, Germany, Montpellier and Belgium belonging to Central Europe and Denmark (p=0.01386), the second including all the northern European countries plus Greece, which might have been collocated in the first group, near to Calabria and Mediterranean area (p-value=0.006723).

Perola et al. (2008, data not yet published) conducted the same analysis on the nuclear genome of GEHA samples and found that Finland may distance itself from the other countries. On the basis of this result, we decided to force our analysis by grouping countries into three clusters: Finland, Europe (including Denmark, Belfast, Newcastle, Netherland, Belgium, Germany, France, Poland, Ukraine) and Mediterranean area (Bologna, Roma, Calabria, Sassari, Greece). As expected, Finland is the country with the maximum divergence rate, it means that its overall haplogroup distributions deeply differ from that of the rest of countries. The same result was obtained by conducting a MDS (Multidimensional scaling) analysis. Again Finland is isolated from other countries, while all the Northern countries are in the middle of the graph, near among them. Also Mediterranean area countries are close by each other. Greece is one more time the nearest country to Finland.

Why Finland is so isolated? It is known that geographical and cultural isolation has greatly shaped Finnish gene pool towards homogeneity, as can be seen for example in certain recessive diseases which are infrequent elsewhere (Norio *et al.* 1973). The

oldest settlement in Finland dates back approximately 9,000 years and a second wave of settlers arrived in the southern parts of Finland around 5,500 years ago. Permanent settlement extended across southern Finland and along the coast and riversides of Ostrobothnia in the 16th century (Norio *et al.* 1973). About 40% of the Finns belong to haplogroup H, which is the most common haplogroup in Europe but rare among Asians (Torroni *et al.* 1996, Richards et al. 1998). Also haplogroup U (16-28%), J (4.5-14%), W (4.1-9.2%) and T (2.5-6.1%) are frequent among Finns. The remaining European haplogroups, I, K, V and X, are less common, as in other parts of Europe, each with frequency below 5.5% in both studies.

This means that Finland has different haplogroup frequencies from other European countries, for examples, haplogroup J in Finland has a frequency of 8%, in Italy about 2%.

From the cluster analysis and MDS analysis we have noticed that the nearest country to Finland was Greece. Why? There are not scientific evidences or literature supporting this similarity, but an emerging theory could explain it. About the historical studious Felice Vinci, the Achaeans would have lived in the early II millennium B.C. on the Baltic coast and in the middle of the millennium, following a tightening of the climate, as identified in this age by paleoclimatology, they would have moved southward along the Dnepr river reaching the Black Sea and the Aegean. The newcomers have founded the Mycenaean City and they would have given to the new places the same names of northern cities, but they are not fully responsive to their original geographical location, due to differences in conformation of the two regions. The main argument is represented by inconsistencies detected by Vinci between the geography described by Homer in his Iliad and Odyssey and the conformation on the Mediterranean lands, already noticed by Strabo. Also the climate description in Homeric poems would better adapt to Baltic region rather than Mediterranean.

This is only a theory since we have no scientific support to our observation.

## 6. The network analysis

The haplogroup J network based on sequence variation in the control region could be divided into two subclusters that confirmed the subdivision proposed previously (Torroni et al. 1997).

The overall phylogenetic network was based on 281 J haplogroup females (137 sibs and 144 controls) and 112 J haplogroup males (46 sibs and 66 controls). In order to

analyze the different haplotype distribution among nations, a median-joining network was also constructed for the HVS-I and HVS-II sequence data (16080-300 segment region) of all J samples. We have individuate the median vectors in female and male network representing a branch point between J2a and J2b. In particular in female network we have identified the median vector 3 from which we found the ancient polymorphisms 150 and 152, then the median vector 9 which divides J2a from J2b. J2a is defined by polymorphisms 195, 215, 16145, 16231 and 16261, J2b by 16193. In male network we see that median vector 10 is responsible of the division between J2a and J2b. As for female network, J2 is characterize by 150 and 152 mutations, J2a is defined in this case by polymorphisms 195, 16145 and 16261, J2b by 16193.

There are several obvious features of the phylogenetic network for the haplogroup J. In female network subhaplogroup J2a, J2b and J1b take the form of star-like clusters, whereas J1c do not exhibit the pattern of a dominant central node with more than one node, radiating out from it. The star-like shape of network indicates population expansion.

As regard male network (figure 28), we observed that only J2a and J1b have a star-like cluster, J1c has different central nodes, as in females network, while J2b do not present a specific cluster. It is interesting to notice that in J2b cluster different cycles are present and in particular the most recurring mutation is 16261, representative of this haplogroup.

Another striking aspect of these networks is that there are areas with a form different from tree branch, rather they contain reticulations. This kind of conformation is more frequent in males than in females, in fact a closer inspection of the edges in the male network reveals that several of these cycles contain edges that correspond to mutations at nucleotide 228, typical of J1c.

At the end, we tried to focus on geographical distribution of J haplogroup in all the European countries, members of the European Project GEHA. We cannot observe any particular distribution; we only can this nation, as previously reported (Finnila et al. 2000).

# Chapter 7

# CONCLUSIONS

It can be concluded that the studies of associations between haplogroups and longevity or healthy aging is still an open field of research.

The development of high-throughput genotyping technologies has greatly increased the feasibility of comprehensive associatin studies of the mitochondrial genome.

Stratifing by sex, significant differences were found between controls and 90+ subjects and we found the association of J2 with males controls. Even when we consider the age and particularly subjects with an age over 96 years, in addition to J2, whose significance was confirmed, T2 was significant.

In general, we can confirm that the association of J2 is population dependent and that the population-specific association of mtDNA haplogroup J with longevity may be due to population-specific genetic backgrounds, to particular interactions between haplogroup J and different environments, and/or to diversity of the J molecules between northern and southern European populations.

These observation were possible because the number of recruited samples were too high that it allowed to obtain sufficient data for statistical analysis; this is defined as the *power of GEHA.*

We also analyzed the entire sequences in oder to understand if there were any particular mutations in the coding region which can justify such associations. In fact, from this analysis we observed a higher percentage of mutations in controls compared with sibs, which could have a negative role and could justify the association of J2 with the controls as a risk factor. The fact that a higher percentage of mutations in tRNA and rRNA accumulate only in sibs, might have a beneficial effect on them.

Finally, the GEHA project has planned the collection of phenotypic data of each patient (BMI, handgrip, cognitive and functional Activities, hypercholesterolemia, diseases, etc. ..). It would be interesting to stratify for these data because there could be a hidden association between haplogroups and a particular phenotype.

# APPENDIX

TBE 5X (TRIS BORATO EDTA) 1L

54 gr TRIS BASE

27,5 gr boric acid

20 ml  EDTA 0,5X pH=8

TBE 0,5X 1L

100 ml TBE 5X

900 ml distilled $H_2O$

AGAROSE GEL 1,5% IN TBE 0.5X 250 ML

250 ml TBE 0,5X

3,75 gr agarose 1.5%

12,5 ml Ethidium Bromide (5 µl in 100 ml)

EDTA

P.M.=372.24 g/mol

A final volume of 50 ml at concentration 125 mM:

Moles=0.125M·0.050 l= 0.00625 moles

Grams: 0.00625·372.24=2.33g

NaAc

P.M.=82.03 g/moli

A final concetration of 3M pH=4.6.

Moles= 3M·0.2 l=0.6 moles

Grams= 0.6 moli·82.03 g/moles=49.218g

| 90+ sibpair (n=2,086) | | | | controls (n=2,153) | | | |
|---|---|---|---|---|---|---|---|
| **sub-haplogroups** | **N** | **%** | **SE** | **sub-haplogroups** | **N** | **%** | **SE** |
| H* | 366 | 17,55 | 0,0083 | H* | 357 | 16,58 | 0,0080 |
| H1 | 292 | 14,00 | 0,0076 | H1 | 325 | 15,10 | 0,0077 |
| H2 | 42 | 2,01 | 0,0031 | H2 | 23 | 1,07 | 0,0022 |
| H3 | 71 | 3,40 | 0,0040 | H3 | 85 | 3,95 | 0,0042 |
| H4 | 5 | 0,24 | 0,0011 | H4 | 9 | 0,42 | 0,0014 |
| H5 | 68 | 3,26 | 0,0039 | H5 | 63 | 2,93 | 0,0036 |
| H6 | 52 | 2,49 | 0,0034 | H6 | 47 | 2,18 | 0,0031 |
| H7 | 18 | 0,86 | 0,0020 | H7 | 10 | 0,46 | 0,0015 |
| H8 | 4 | 0,19 | 0,0010 | H8 | 0 | 0,00 | 0,0000 |
| H9 | 9 | 0,43 | 0,0014 | H9 | 5 | 0,23 | 0,0010 |
| HV0a | 13 | 0,62 | 0,0017 | HV0a | 6 | 0,28 | 0,0011 |
| HV0* | 70 | 3,36 | 0,0039 | HV0* | 62 | 2,88 | 0,0036 |
| HV1 | 4 | 0,19 | 0,0010 | HV1 | 5 | 0,23 | 0,0010 |
| HV2 | 1 | 0,05 | 0,0005 | HV2 | 0 | 0,00 | 0,0000 |
| HV* | 42 | 2,01 | 0,0031 | HV* | 50 | 2,32 | 0,0032 |
| V | 27 | 1,29 | 0,0025 | V | 33 | 1,53 | 0,0026 |
| I | 25 | 1,20 | 0,0024 | I | 21 | 0,98 | 0,0021 |
| I1 | 9 | 0,43 | 0,0014 | I1 | 12 | 0,56 | 0,0016 |
| I3 | 12 | 0,58 | 0,0017 | I3 | 11 | 0,51 | 0,0015 |
| J1 | 9 | 0,43 | 0,0014 | J1 | 8 | 0,37 | 0,0013 |
| J1b | 25 | 1,20 | 0,0024 | J1b | 27 | 1,25 | 0,0024 |
| J1c | 109 | 5,23 | 0,0049 | J1c | 117 | 5,43 | 0,0049 |
| J1d | 1 | 0,05 | 0,0005 | J1d | 1 | 0,05 | 0,0005 |
| J2 | 2 | 0,10 | 0,0007 | J2 | 7 | 0,33 | 0,0012 |
| J2a | 25 | 1,20 | 0,0024 | J2a | 32 | 1,49 | 0,0026 |
| J2b | 12 | 0,58 | 0,0017 | J2b | 18 | 0,84 | 0,0020 |
| K | 13 | 0,62 | 0,0017 | K | 15 | 0,70 | 0,0018 |
| K1 | 13 | 0,62 | 0,0017 | K1 | 12 | 0,56 | 0,0016 |
| K1a | 79 | 3,79 | 0,0042 | K1a | 91 | 4,23 | 0,0043 |
| K1b | 0 | 0,00 | 0,0000 | K1b | 1 | 0,05 | 0,0005 |
| K1c | 25 | 1,20 | 0,0024 | K1c | 36 | 1,67 | 0,0028 |
| K2 | 7 | 0,34 | 0,0013 | K2 | 6 | 0,28 | 0,0011 |
| N1a | 8 | 0,38 | 0,0014 | N1a | 5 | 0,23 | 0,0010 |
| N1b | 12 | 0,58 | 0,0017 | N1b | 15 | 0,70 | 0,0018 |
| N1c | 2 | 0,10 | 0,0007 | N1c | 3 | 0,14 | 0,0008 |
| N9a | 2 | 0,10 | 0,0007 | N9a | 1 | 0,05 | 0,0005 |
| R0 | 3 | 0,14 | 0,0008 | R0 | 3 | 0,14 | 0,0008 |
| R0a | 12 | 0,58 | 0,0017 | R0a | 10 | 0,46 | 0,0015 |
| R1a | 1 | 0,05 | 0,0005 | R1a | 1 | 0,05 | 0,0005 |
| R2 | 0 | 0,00 | 0,0000 | R2 | 4 | 0,19 | 0,0009 |
| T | 6 | 0,29 | 0,0012 | T | 6 | 0,28 | 0,0011 |
| T1 | 4 | 0,19 | 0,0010 | T1 | 5 | 0,23 | 0,0010 |
| T1a | 38 | 1,82 | 0,0029 | T1a | 44 | 2,04 | 0,0030 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| T1b | 1 | 0,05 | 0,0005 | T1b | 4 | 0,19 | 0,0009 |
| T2 | 50 | 2,40 | 0,0033 | T2 | 44 | 2,04 | 0,0030 |
| T2a | 7 | 0,34 | 0,0013 | T2a | 2 | 0,09 | 0,0007 |
| T2b | 104 | 4,99 | 0,0048 | T2b | 94 | 4,37 | 0,0044 |
| T2c | 7 | 0,34 | 0,0013 | T2c | 5 | 0,23 | 0,0010 |
| T2e | 6 | 0,29 | 0,0012 | T2e | 6 | 0,28 | 0,0011 |
| T2f | 0 | 0,00 | 0,0000 | T2f | 1 | 0,05 | 0,0005 |
| U | 2 | 0,10 | 0,0007 | U | 4 | 0,19 | 0,0009 |
| U1 | 4 | 0,19 | 0,0010 | U1 | 3 | 0,14 | 0,0008 |
| U1a | 4 | 0,19 | 0,0010 | U1a | 11 | 0,51 | 0,0015 |
| U1b | 2 | 0,10 | 0,0007 | U1b | 0 | 0,00 | 0,0000 |
| U1c | 0 | 0,00 | 0,0000 | U1c | 1 | 0,05 | 0,0005 |
| U2 | 13 | 0,62 | 0,0017 | U2 | 13 | 0,60 | 0,0017 |
| U2d | 1 | 0,05 | 0,0005 | U2d | 0 | 0,00 | 0,0000 |
| U2e | 24 | 1,15 | 0,0023 | U2e | 26 | 1,21 | 0,0024 |
| U3 | 11 | 0,53 | 0,0016 | U3 | 5 | 0,23 | 0,0010 |
| U3a | 11 | 0,53 | 0,0016 | U3a | 7 | 0,33 | 0,0012 |
| U3b | 0 | 0,00 | 0,0000 | U3b | 1 | 0,05 | 0,0005 |
| U4 | 14 | 0,67 | 0,0018 | U4 | 21 | 0,98 | 0,0021 |
| U4a | 15 | 0,72 | 0,0018 | U4a | 17 | 0,79 | 0,0019 |
| U4b | 5 | 0,24 | 0,0011 | U4b | 10 | 0,46 | 0,0015 |
| U5 | 0 | 0,00 | 0,0000 | U5 | 3 | 0,14 | 0,0008 |
| U5a | 90 | 4,31 | 0,0044 | U5a | 109 | 5,06 | 0,0047 |
| U5b | 51 | 2,44 | 0,0034 | U5b | 56 | 2,60 | 0,0034 |
| U6a | 6 | 0,29 | 0,0012 | U6a | 11 | 0,51 | 0,0015 |
| U7 | 7 | 0,34 | 0,0013 | U7 | 3 | 0,14 | 0,0008 |
| U8 | 3 | 0,14 | 0,0008 | U8 | 3 | 0,14 | 0,0008 |
| U8a | 4 | 0,19 | 0,0010 | U8a | 1 | 0,05 | 0,0005 |
| U8b | 4 | 0,19 | 0,0010 | U8b | 8 | 0,37 | 0,0013 |
| U9 | 1 | 0,05 | 0,0005 | U9 | 0 | 0,00 | 0,0000 |
| W | 29 | 1,39 | 0,0026 | W | 26 | 1,21 | 0,0024 |
| W1 | 1 | 0,05 | 0,0005 | W1 | 1 | 0,05 | 0,0005 |
| W4 | 13 | 0,62 | 0,0017 | W4 | 1 | 0,05 | 0,0005 |
| W5 | 6 | 0,29 | 0,0012 | W5 | 3 | 0,14 | 0,0008 |
| W6 | 2 | 0,10 | 0,0007 | W6 | 8 | 0,37 | 0,0013 |
| X | 12 | 0,58 | 0,0017 | X | 13 | 0,60 | 0,0017 |
| X1 | 0 | 0,00 | 0,0000 | X1 | 2 | 0,09 | 0,0007 |
| X2 | 11 | 0,53 | 0,0016 | X2 | 6 | 0,28 | 0,0011 |
| X2a | 2 | 0,10 | 0,0007 | X2a | 0 | 0,00 | 0,0000 |
| X2b | 15 | 0,72 | 0,0018 | X2b | 12 | 0,56 | 0,0016 |
| X2c | 2 | 0,10 | 0,0007 | X2c | 8 | 0,37 | 0,0013 |
| OTHER | 8 | 0,38 | 0,0014 | OTHER | 12 | 0,56 | 0,0016 |

**Tab. 1:** Frequencies of all mtDNA sub-haplogroups in 2,086 90+ sibpairs and 2,153 controls from all over Europe.

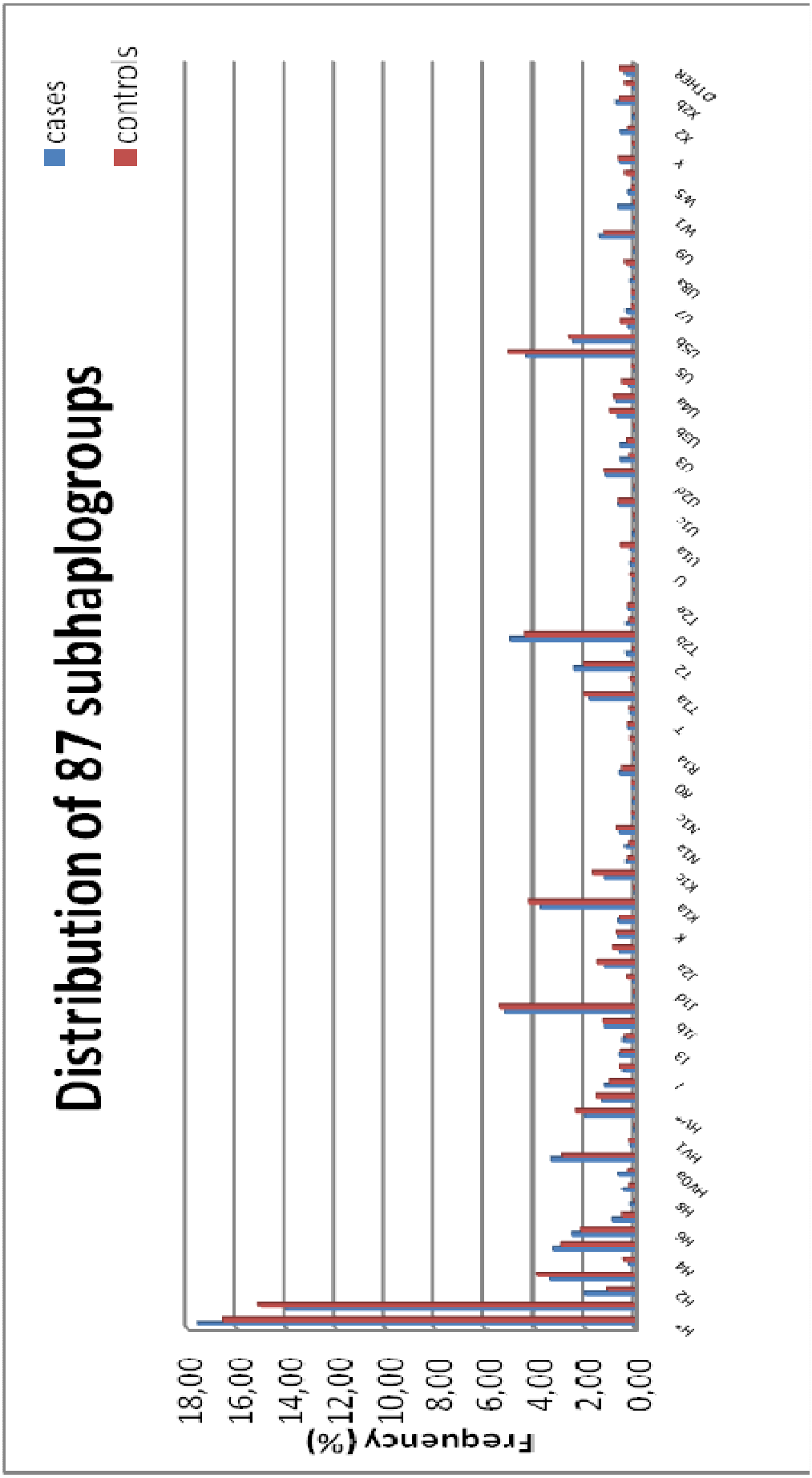
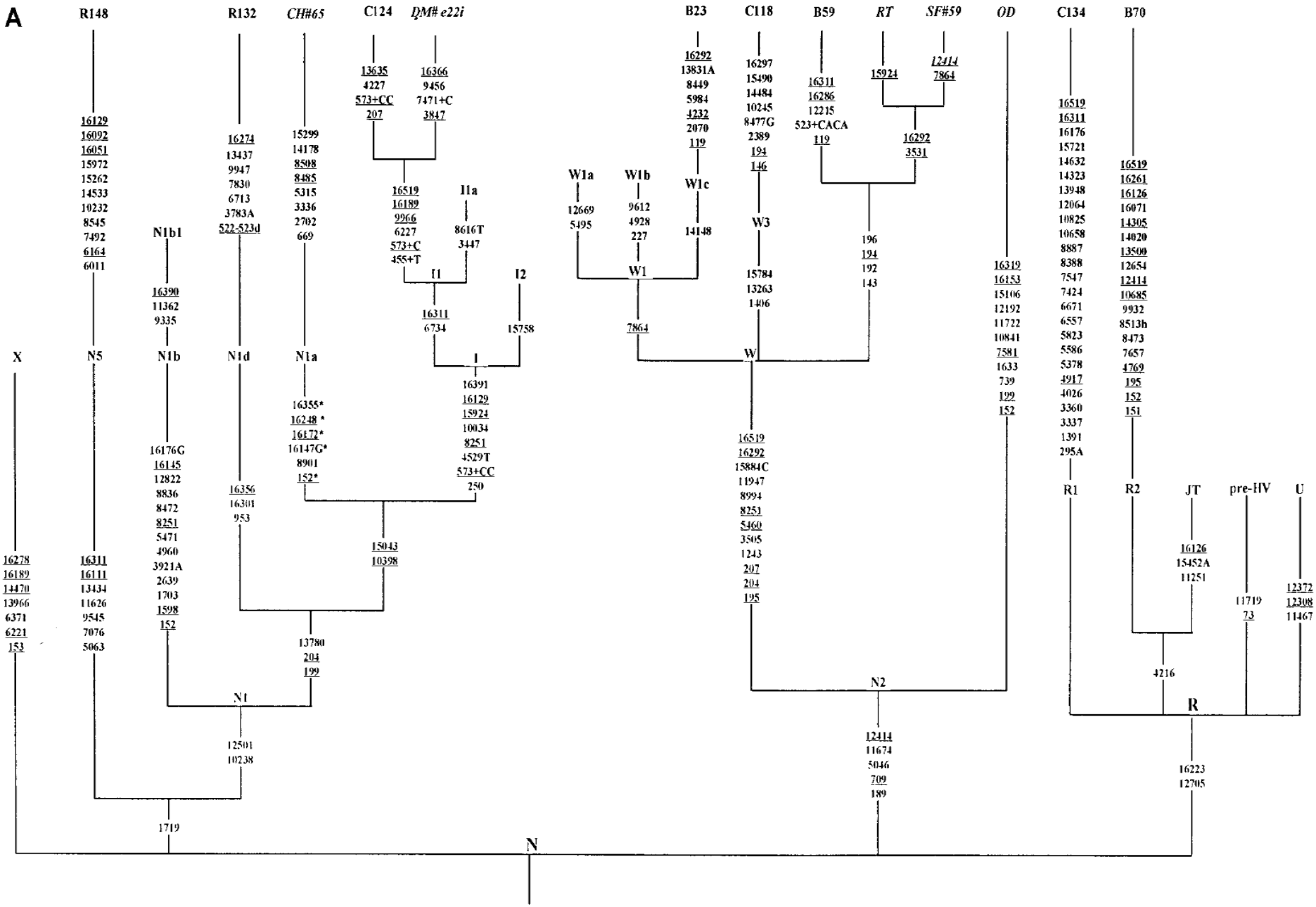
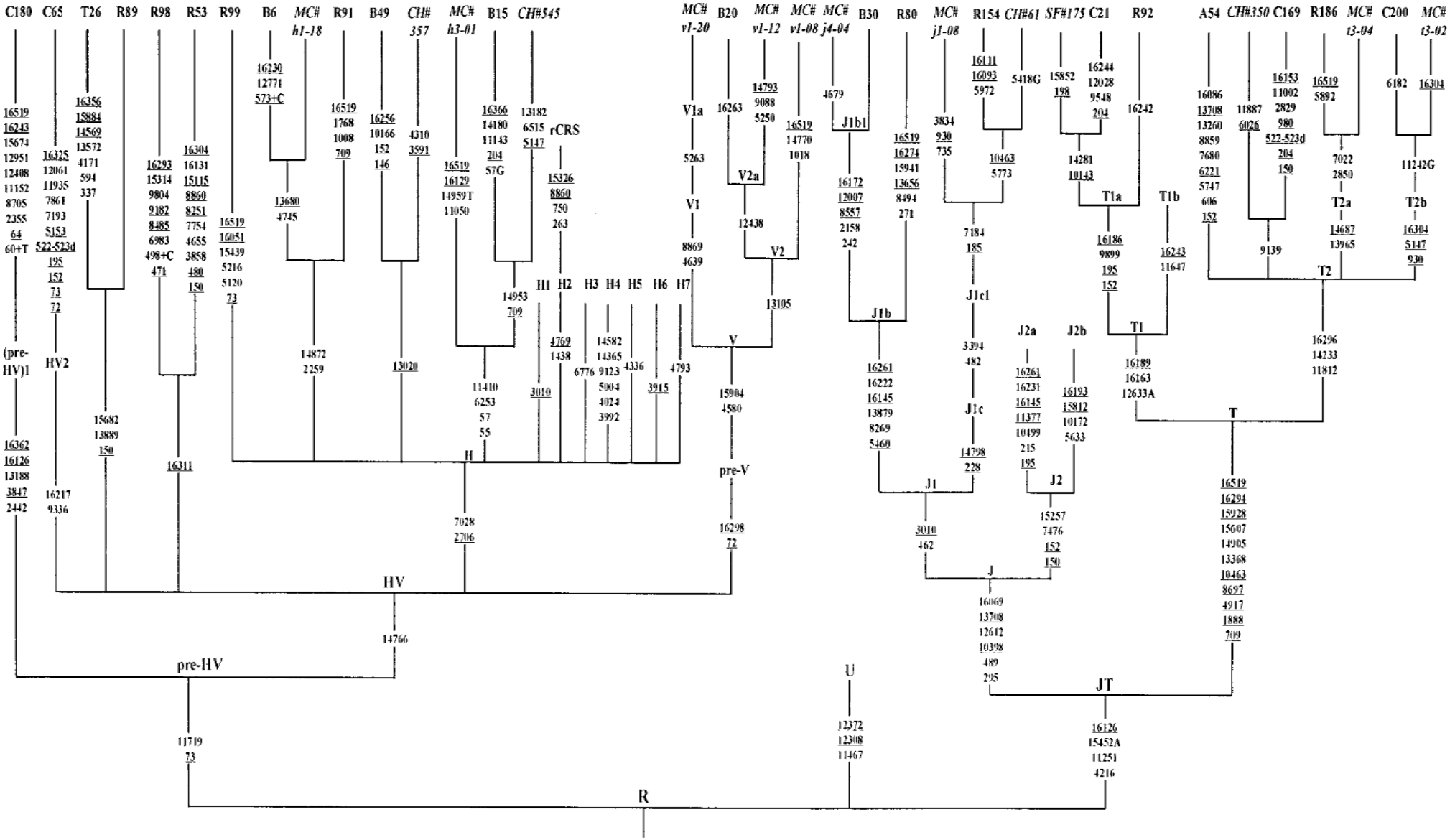
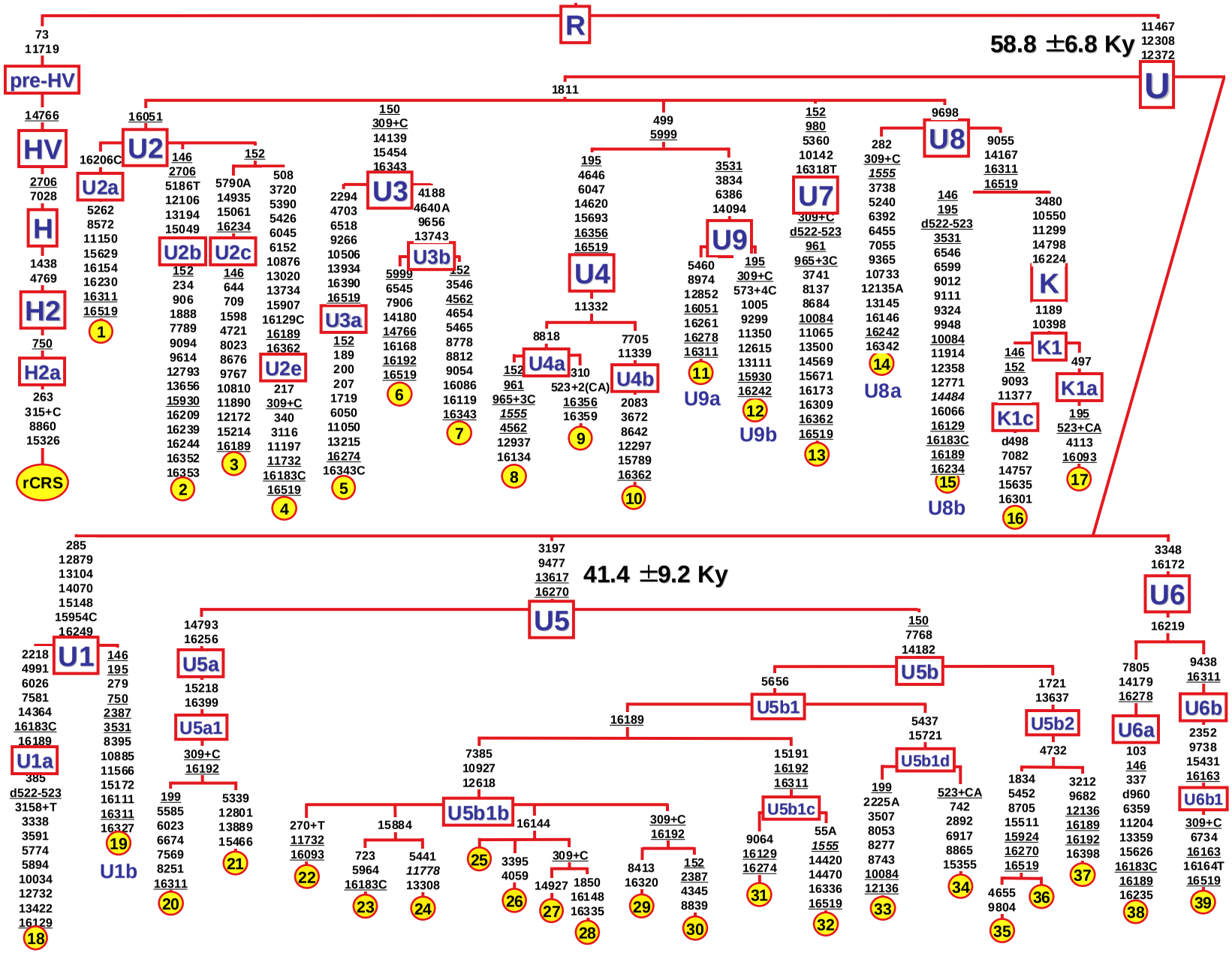


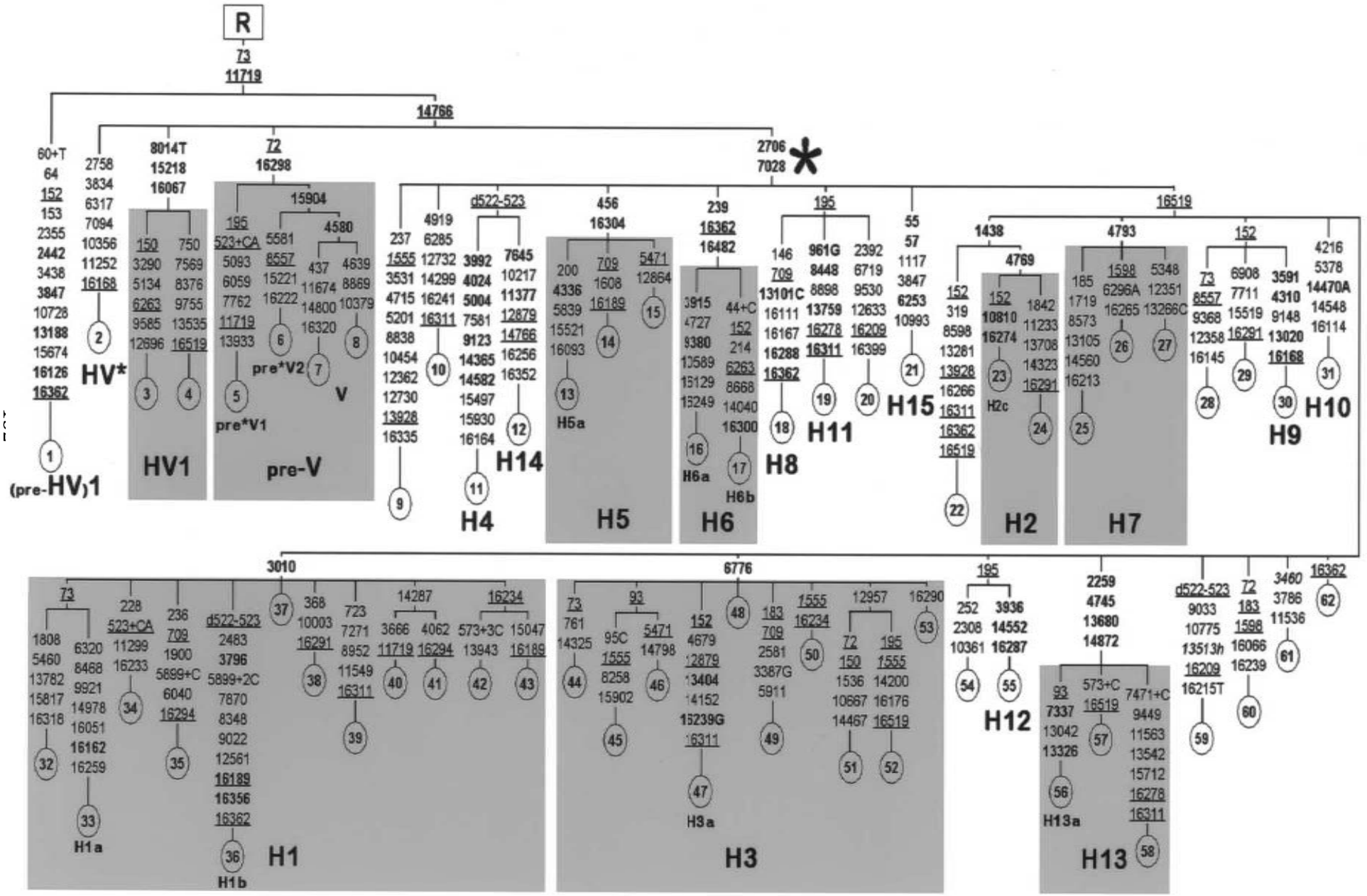

**Fig. 1**: Graphical distribution of all 90+ and controls subhaplogroups.


A
R148
R132
CH#65
C124
DM# e22i
B23
C118
B59
RT
SF#59
OD
C134
B70
16129 16092 16051 15972 15262 14533 10232 8545 7492 6164 6011
16274 13437 9947 7830 6713 3783A 522-523d
15299 14178 8508 8485 5315 3336 2702 669
13635 4227 573+CC 207
16366 9456 7471+C 3847
16292 13831A 8449 5984 4232 2070 119
16297 15490 14484 10245 8477G 2389 194 146
16311 16286 12215 523+CACA 119
15924
12414 7864
16292 3531
16519 16311 16176 15721 14632 14323 13948 12064 10825 10658 8887 8388 7547 7424 6671 6557 5823 5586 5378 4917 4026 3360 3337 1391 295A
16519 16261 16126 16071 14305 14020 13500 12654 12414 10685 9932 8513h 8473 7657 4769 195 152 151
N1b1
16390 11362 9335
16519 16189 9966 6227 573+C 455+T
11a
8616T 3447
I1
I2
16311 6734
15758
I
16391 16129 15924 10034 8251 4529T 573+CC 250
W1a
12669 5495
W1b
9612 4928 227
W1c
14148
W1
7864
W3
15784 13263 1406
W
196 194 192 143
16319 16153 15106 12192 11722 10841 7581 1633 739 129 152
X
N5
N1b
N1d
N1a
16278 16189 14470 13966 6371 6221 153
16311 16111 13434 11626 9545 7076 5063
16176G 16145 12822 8836 8472 8251 5471 4960 3921A 2639 1703 1598 152
16356 16301 953
16355* 16248 * 16172* 16147G* 8901 152*
15043 10398
13780 204 199
16519 16292 15884C 11947 8994 8251 5460 3505 1243 207 204 195
R1
R2
JT
pre-HV
U
16126 15452A 11251
11719 73
12372 12308 11467
4216
N1
N2
R
12501 10238
12414 11674 5046 709 189
16223 12705
1719
N

**B**


C180
C65
T26
R89
R98
R53
R99
B6
MC# h1-18
R91
B49
CH# 357
MC# h3-01
B15
CH#545
rCRS
MC# v1-20
B20
MC# v1-12
MC# v1-08
MC# j4-04
B30
R80
MC# j1-08
R154
CH#61
SF#175
C21
R92
A54
CH#350
C169
R186
MC# t3-04
C200
MC# t3-02
16519 16243 15674 12951 12408 11152 8705 2355 64 60+T
(pre-HV)1
16362 16126 13188 3847 2442
16325 12061 11935 7861 7193 5153 522-523d 195 152 73 72
HV2
16217 9336
16356 15884 14569 13572 4171 594 337
15682 13889 150
16293 15314 9804 9182 8485 6983 498+C 471
16304 16131 15115 8860 8251 7754 4655 3858 480 150
16311
16519 16051 15439 5216 5120 73
16230 12771 573+C
13680 4745
14872 2259
16519 1768 1008 709
16256 10166 152 146
4310 3591
13020
16519 16129 14959T 11050
14953 709
11410 6253 57 55
H
7028 2706
16366 14180 11143 204 57G
13182 6515 5147
15326 8860 750 263
H1 H2 H3 H4 H5 H6 H7
4769 1438
3010
14582 14365 9123 5004 4024 3992
6776
4336
3915
4793
V1a
5263
V1
8869 4639
V
15904 4580
pre-V
16298 72
14793 9088 5250
V2a
12438
V2
13105
16519 14770 1018
4679
J1b1
16172 12007 8557 2158 242
J1b
16261 16222 16145 13879 8269 5460
J1
3010 462
J
16069 13708 12612 10398 489 295
3834 930 735
7184 185
J1c1
3394 482
J1c
14798 228
16111 16093 5972
10463 5773
5418G
16519 16274 15941 13656 8494 271
15852 198
14281 10143
16244 12028 9548 204
T1a
16186 9899 195 152
T1
16189 16163 12633A
16242
T1b
16243 11647
J2a
16261 16231 16145 11377 10499 215 195
J2
15257 7476 152 150
J2b
16193 15812 10172 5633
16086 13708 13260 8859 7680 6221 5747 606 152
11887 6026
9139
16153 11002 2829 980 522-523d 204 150
16519 5892
7022 2850
T2a
14687 13965
T2
16296 14233 11812
6182
11242G
T2b
16304 5147 930
T
16519 16294 15928 15607 14905 13368 10463 8697 4917 1888 709
JT
16126 15452A 11251 4216
U
12372 12308 11467
pre-HV
11719 73
14766
HV
R


R
58.8 ±6.8 Ky
U
pre-HV
HV
H
H2
H2a
rCRS
U2
U2a
U2b
U2c
U2e
U3
U3a
U3b
U4
U4a
U4b
U5
41.4 ±9.2 Ky
U1
U1a
U1b
U5a
U5a1
U5b
U5b1
U5b1b
U5b1c
U5b1d
U5b2
U6
U6a
U6b
U6b1
U7
U8
U8a
U8b
U9
U9a
U9b
K
K1
K1a
K1c
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39


R
73
11719
14766
60+T
64
152
153
2355
2442
3438
3847
10728
13188
15674
16126
16362
1
(pre-HV)1
2758
3834
6317
7094
10356
11252
16168
2
HV*
8014T
15218
16067
150
3290
5134
6263
9585
12696
3
750
7569
8376
9755
13535
16519
4
HV1
72
16298
15904
195
523+CA
5093
6059
7762
11719
13933
5
pre*V1
5581
8557
15221
16222
6
pre*V2
4580
437
11674
14800
16320
7
V
4639
8869
10379
8
pre-V
2706
7028
*
237
1555
3531
4715
5201
8838
10454
12362
12730
13928
16335
9
4919
6285
12732
14299
16241
16311
10
d522-523
3992
4024
5004
7581
9123
14365
14582
15497
15930
16164
11
H4
7645
10217
11377
12879
14766
16256
16352
12
H14
456
16304
200
4336
5839
15521
16093
13
H5a
709
1608
16189
14
5471
12864
15
H5
239
16362
16482
3915
4727
9380
10589
15129
15249
16
H6a
44+C
152
214
6263
8668
14040
16300
17
H6b
H6
195
146
709
13101C
16111
16167
16288
16362
18
H8
961G
8448
8898
13759
16278
16311
19
2392
6719
9530
12633
16209
16399
20
H11
55
57
1117
3847
6253
10993
21
H15
16519
1438
152
319
8598
13261
13928
16266
16311
16362
16519
22
4769
152
10810
16274
23
H2c
1842
11233
13708
14323
16291
24
H2
4793
185
1719
8573
13105
14560
16213
25
1598
6296A
16265
26
5348
12351
13266C
27
H7
152
73
8557
9368
12358
16145
28
6908
7711
15519
16291
29
3591
4310
9148
13020
16168
30
H9
4216
5378
14470A
14548
16114
31
H10
3010
73
1808
5460
13782
15817
16318
32
6320
8468
9921
14978
16051
16162
16259
33
H1a
228
523+CA
11299
16233
34
236
709
1900
5899+C
6040
16294
35
d522-523
2483
3796
5899+2C
7870
8348
9022
12561
16189
16356
16362
36
H1b
37
368
10003
16291
38
723
7271
8952
11549
16311
39
H1
14287
3666
11719
40
4062
16294
41
16234
573+3C
13943
42
15047
16189
43
6776
73
761
14325
44
93
95C
1555
8258
15902
45
5471
14798
46
152
4679
12879
13404
14152
15239G
16311
47
H3a
48
183
709
2581
3387G
5911
49
1555
16234
50
12957
72
150
1536
10667
14467
51
195
1555
14200
16176
16519
52
16290
53
H3
195
252
2308
10361
54
3936
14552
16287
55
H12
2259
4745
13680
14872
93
7337
13042
13326
56
H13a
573+C
16519
57
7471+C
9449
11563
13542
15712
16278
16311
58
H13
d522-523
9033
10775
13513h
16209
16215T
59
72
183
1598
16066
16239
60
3460
3786
11536
61
16362
62
H2
H7

# Bibliography

**Alexe** G, Fuku N, Bilal E, Ueno H, Nishigaki Y, Fujita Y, Ito M, Arai Y, N NH, Bhanot G, Tanaka M. *Enrichment of longevity phenotype in mtdna haplogroups d4b2b, d4a, and d5 in the Japanese population*. Hum Genet 2007; 121: 347–356.

**Allard** MW, Miller K, Wilson M, Monson K, Budowle B. 2002 *Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human contro region sequences.* Scientific working group on DNA analysis methods. J Forensic Sci. 47:1215–1223.

**Anderson** S, Bankier AT, Barrel BG, De Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) *Sequence and organization of the human mitochondrial genome*, Nature 290: 457-65

**Andrews** RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) *Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA* Nature America Inc 23: 147

**Autere** J, Moilanen JS, Finnila S, Soininen H, Mannermaa A, Hartikainen P, Hallikainen M, Majamaa K (2004) *Mitochondrial DNA polymorphisms as risk factors for Parkinson's disease and Parkinson's disease dementia*, Hum. Genet., 115: 29-35

**Ballinger** SW, Shoffner JM, Gebhart S, Koontz DA, Wallace DC (1994) *Mitochondrial diabetes revisited*, Nat. Genet., 7:458-59

**Ballinger** SW, Shoffner JM, Hedaya EV, Trounce I, Polak MA, et al.: *Maternally transmitted diabetes and deafness associated with a 10.4 kb mitochondrial DNA deletion*, Nat. Genet., 1992; 1: 11-15

**Barja** G (2004) *Free radicals and aging* TRENDS in Neurosciences 27: 595-600

**Barker** DJ. *The fetal origins of adult hypertension.* J Hypertens Suppl. 1992 Dec;10(7):S39-44.

**Basso** F, Lowe GD, Rumley A, McMahon AD, Humphries SE, *Interleukin-6 -174 G > C polymorphism and risk of coronary heart disease in West of Scotland coronary prevention study (WOSCOPS),* (2002) Arterioscler. Thromb. Vasc. Biol., 22: 599-604,

**Beekman** M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE. *Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies*. J Gerontol A Biol Sci Med Sci. 2006 Apr;61(4):355-62.

**Bellizzi** D, Rose G, Cavalcante P, Covello G, Dato S, De Rango F, *et al. A novel VNTR enhancer within the SIRT3 gene, a human homologue of SIR2, is associated with survival at oldest ages.* (2005) Genomics 85:258 –263.

**Bilal** E, Rabadan R, Alexe G, Fuku N, Ueno H, Nishigaki Y, Fujita Y, Ito M, Arai Y, Hirose N, Ruckenstein A, Bhanot G, Tanaka M. *Mitochondrial DNA haplogroup D4a is a marker for extreme longevity in Japan.* (2008) PLoS One. Jun 11;3(6):e2421.

**Blanche** H, Cabanne L, Sabhatou M, Thomas G*. A study of French centenarians : are ACE and APOE associated with longevity ?* (2001) CR Acad Sci III ;324 :129-35

**Bonafè** M., Olivieri F., Cavallone L., Giovagnetti S., Cardelli M. et al., *A gender-dependent genetic predisposition to produce high levels of IL-6 is detrimental for longevity* (2001) Eur. J. Immunol., 31 (8): 2357-2361.

**Bonafè** M, Marchegiani F, Cardelli M, Olivieri F, Cavallone L, Giovagnetti S, Pieri C, Marra M, Antonicelli R, Troiano L, Gueresi P, Passeri G, Berardelli M, Paolisso G, Barbieri M, Tesei S, Lisa R, De Benedictis G, Franceschi C*. Genetic analysis of Paraoxonase (PON1) locus reveals an increased frequency of Arg192 allele in centenarians.* Eur J Hum Genet. 2002 May;10(5):292-6.

**Bonafè** M. et al. *Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: cues for*

*an evolutionarily conserved mechanism of lifespan control.* (2003) J. Clin. Endocrinol. Metab. 88, 3299–3304

**Brown**, M. D., Hosseini, S., Steiner, I., Wallace, D. C., Korn-Lubetzki, I. (2004) *Complete mitochondrial DNA sequence analysis in a family with early-onset dystonia and optic atrophy* Movement Disorders . 19 (2): 235-237

**Candore** G, Balistreri CR, Listì F, Grimaldi MP, Vasto S, Colonna-Romano G, Franceschi C, Lio D, Caselli G, Caruso C*. Immunogenetics, gender, and longevity*. 2006 Ann N Y Acad Sci.;1089:516-37.

**Cann** RL, Stoneking M, Wilson AC: *Mitochondrial DNA and human evolution,* Nature, 1987; 325(6099):31-6

**Capri** M, Salvioli S, Federica S, Valensin S, Celani L, Monti D, Pawelec G, Benedictis GD, Efstathios SG, Franceschi C. *The genetics of human longevity* Ann N Y Acad Sci 2006; 1067: 252–263.

**Carelli** V, Giordano C, D'Amati G (2003) *Pathogenic expression of homoplastic mtDNA mutations needs a complex nuclear-mitochondrial interaction* TRENDS in Genetics 19: 257-262

**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. *Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease.* (2001) Hum Genet; 108:194–8

**Castri** L, Melendez-Obando M, Villegas-Palma R, Barrantes R, Raventos H, Pereira R, Luiselli D, Pettener D, Madrigal L. (2009) *Mitochondrial Polymorphisms Are Associated Both with Increased and Decreased Longevity,* Hum Hered 67:147–153

**Castro** MG, Huerta C, Reguero JR, Soto MI, Domènech E, Alvarez V, Gomez-Zaera M, Nunes V, Gonzàlez P, Corao A, Coto E. (2006) *Mitochondrial DNA haplogroups in Spanish patients with hypertrophic cardiomyopathy International* Journal of Cardiology 112: 202 – 206

**Chagnon** P, Gee M, Filion M, Robitaille Y, Belouchi M, Gauvreau D. (1999) *Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population.* Am J Med Genet. 2;85(1):20-30.

**Chen** XJ, Butow RA (2005) *The organization and inheritance of the mitochondrial genome* Nature Genetics 6, 815-825

**Chinney** PF, Brown DT, Andrews RM, Singh-Kler R, Riordan-Eva P, Lindley J et al. (2001) *The mitochondrial ND6 gene in a hot spot for mutations that cause Leber's hereditary optic neuropathy* Brain 124: 209-18

**Chomyn** A, Attardi G (2003) *MtDNA mutations in aging and apoptosis* Biochemical and Biophysical Research Communication 304, 519-529

**Christensen** K., Johnson TE., Vaupel JW. *The quest for genetic determinants of human longevity: challenges and insights (*2006) Nat Rev Genet 7(6):436-48

**Clayton** DA (1982) *Replication of animal mitochondrial DNA* Cell 28: 693-705
Cortopassi CA, Arnheim N (1990) *Detection of a specific mitochondrial DNA deletion in tissues of older individuals* Nucleic Acid Res, 18: 6927-33

**Coskun** PE, Pesini ER, Wallace DC (2003) *Control region mtDNA variants: longevity, climatic adaptation, and a forensic conundrum* PNAS 100, 2174-2176

**Dato** S, Passarino G, Rose G, Altomare K, Bellizzi D, Mari V, Feraco E, Franceschi C, De Benedictis G: *Association of nitochondrial DNA haplogroup J with longevity is population specific*, Eur. J. Hum. Genet., 2004; 12: 1080-1082

**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, Bonafè M, Franceschi C: *Gene/longevity association studies at four autosomal loci (REN, THO, PARP, SOD2),* Eur. J. Hum. Genet., 1998; 6: 534-541

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

**De Benedictis** G, Tan Q, Jeune B, Christensen K, Ukraintseva SV, Bonafè M, Franceschi C, Vaupel JW, Yashin AI. *Recent advances in human gene-longevity association studies.* (2001) Mech Ageing Dev 122(9):909-20

**De Luca** M, Rose G, Bonafè M, Garasto S, Greco V, Weir BS, Franceschi C, De Benedictis G. *Sex-specific longevity associations defined by Tyrosine Hydroxylase-Insulin-Insulin Growth Factor 2 haplotypes on the 11p15.5 chromosomal region.* Exp Gerontol. 2001 Nov;36(10):1663-71.

**Del Bo** R, Bordoni A, Boneschi FM, Crimi M, Sciacco M, Bresolin N, et al. *Evidence and age-related distribution of mtDNA D-loop point mutations in skeletal muscle from healthy subjects and mitochondria patients* J Neurol Sci 202: 85–91.

**De Rango** F, Dato S, Bellizzi D, Rose G, Marzi E, Cavallone L, Franceschi C, Skytthe A, Jeune B, Cournil A, Robine JM, Gampe J, Vaupel JW, Mari V, Feraco E, Passarino G, Novelletto A, De Benedictis G. *A novel sampling design to explore gene-longevity associations: the ECHA study.* Eur J Hum Genet. 2008 Feb;16(2):236-42. Epub 2007 Nov 7.

**Di Mauro** S, Mancuso M, Filosto M. (2004) *Le malattie mitocondriali* Neural Sci 25: 51-52

**Elo** IT, Preston SH. *Effects of early-life conditions on adult mortality: a review*. Popul Index. 1992 Summer;58(2):186-212.

**Endo** T, Yamamoto H, Esaki M (2003) *Functional cooperation and separation of translocators in protein import into mitochondria, the double-membrane bounded organelles Journal of* Cell Science 116: 3259-67

**Evans** L, Kennedy GA, Wertheim EH. *An examination of the association between eating problems, negative mood, weight and sleeping quality in young women and men.* Eat Weight Disord. 2005 Dec;10(4):245-50.

**Fentleman**, DL; Smith, J & Peterson, J (1990). *Successful ageing in a postretirement society*; in Baltes, Margret M.; Baltes, Paul B. (1990). *Successful ageing: perspectives from the behavioral sciences*. Cambridge, UK: Cambridge University Press.

**Fernández-Real** JM., Broch M., Vendrell J., Gutiérrez C., Casamitjana R., Pugeat M., Richart C., Ricart W. *Interleukin-6 Gene Polymorphism and Insulin Sensitivity* (2000) Diabetes, Vol. 49

**Ferrari** S.L., Garnero P., Emond S., Montgomery H., Humphries S.E. et al., *A functional polymorphic variant in the interleukin-6 gene promoter associated with low bone resorption in postmenopausal women* (2001) Arthritis Rheum, 44: 196-201

**Fesahat** F, Houshmand M, Panahi MS, Gharagozli K, Mirzajani F *Do haplogroups H and U act to increase the penetrance of Alzheimer's disease?.* (2007) Cell Mol Neurobiol. 27(3):329-34.

**Finnilä** S, Majamaa K. *Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population.* J Hum Genet. 2001;46(2):64-9.

**Fish** J, Raule N, and Attardi G: *Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis*, Science, 2004: 306: 2098-101.

**Forster** P. (2003) *To err is human* Annals of Human Genetics 67: 2-4

**Franceschi** C: *Cell proliferation and cell death in the aging process*, Aging Clin. Exp. Res., 1989; 1: 3-13

**Franceschi** C, Monti D, Sansoni P, Cossarizza A: *The immunology of exceptional individuals: the lesson of centenarians*, Immunol. Today, 1995; 16: 12-16

**Franceschi** C., Monti D., Barbieri D., Grassilli E., Troiano L. et al., *Immunosenescence in humans: deterioration or remodelling***?** (1995) Intern. Rev. Immunol., 12: 57-74

**Franceschi** C and Cossarizza A: *The reshaping of the immune system with age*, Int. Rev. Immunol, 1995; 12: 1-4

**Franceschi** C, Valesin S, Bonafè M, Paolisso G, Yashin AI, Monti D, De Benedictis G: *The network and the remodeling theories of aging: historical background and new perspectives*, Experimental Gerontology, 2000; 35:879-896

**Franceschi** C, Bonafè M, Valesin S, Olivieri F, De Luca M, Ottaviani E, De Benedictis G: *Inflamm-aging. An evolutionary perspective on immunosenescence,* Ann. N. Y. Acad. Sci., 2000; 908: 244-254

**Franceschi** C., Motta L., Valensin S., Rapisarda R., Franzone A. et al., *Do men and women follow different trajectories to reach extreme longevity?* (2000) Aging Clin. Exp. Res, 12: 77-84.

**Franceschi** C., Valensin S., Lescai F., Olivieri F., Licastro F. et al., *Neuroinflammation and the genetics of Alzheimer's disease: The search for a pro-inflammatory phenotype*, (2001) Aging Clin. Exp. Res, 13: 163-170

**Franceschi** C and Bonafè M: *Centenarians as a model of healthy aging*, Biochemical Society, 2003; 31: 457-461

**Franceschi** C, Olivieri F, Marchegiani F, Cardelli M, Cavallone L, et al. (2005) Genes involved in immune response/inflammation, IGF1/insulin pathway and response to oxidative stress play a major role in the genetics of human longevity: the lesson of centenarians. Mech Ageing Dev 126: 351–361.

**Fuku** N, Park KS, Yamada Y, Nishigaki Y, Cho YM et al. (2007) *Mitocohndrial haplogroup N9a confers resistance against type 2 diabetes in Asians The* Amer Journ of Human Genetics 80: 407-415

**Gadaleta** MN et al: *Mitochondrial-DNA copy number and mitochondrial-DNA deletion in adult and senescent rats*, Mutat. Res., 1992; 275, 181-193

**Gaweda-Walerych** K., Maruszak A, Safranow K. Bialecka M., Klodowska-Duda G. et al. *Mitochondrial DNA haplogroups and subhaplogroups are associated with Parkinson's disease risk in a Polish PD cohort* (2008) J Neural Transm 115:1521–1526

**Ghezzi** D, Marelli C, Achilli A, Goldwurm S, Pezzoli G, Barone P, Pellecchia MT, Stanzione P, Brusa L, Bentivoglio AR, et al. *Mitochondrial DNA haplogroup K is associated with a lower risk of parkinson's disease in Italians*, Eur. J. Hum. Genet., 2005; 13: 748-752

**Glatt** SJ, Chayavichitsilp P, Depp C, Schork NJ, Jeste DV. *Successful aging: from phenotype to genotype. (*2007) Biol Psychiatry 62(4):282-93

**Gravina** S & Vijg J *Epigenetic factors in aging and longevity* Pflugers Arch - Eur J Physiol (2010) 459:247–258

**Goto** Y, Nonaka I, Horai S (1990) *A mutation in the tRNA$^{(LEU)(UUR)}$ gene associated with MELAS subgroup of mitochondrial encephalomyopathies* Nature 348:651-653

**Harman** D (1972) *The biologic clock: the mitochondria***?** J Am Geriatr *Soc* 20: 145-7

**Harman** D (2006) *Free radical theory of aging: an update: increasing the functional life* span Ann N Y Acad Sci 1067:10-21

**Hasegawa** M, Di Rienzo A, Kocher TD, Wilson A. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol. 37:347–354.

**Herrnstadt** C, Preston G, Howell N (2002) *Reduced-median-network analysis of complete mitochondrial DNA coding region sequences for the major African, Asian, and European haplogroups***,** Am. J. Hum. Genet; 70: 1152-1171

**Herrnstadt** C, Howell N (2004) *An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders* Mitochondrion 4: 791-798

**Hirst** J, Carroll J, Fearnley IM, Shannon RJ, Walker JE (2003) *The nuclear encoded subunits of complex I from bovine heart mitochondria* Biochimica et Biophysica Acta 1604: 135– 150

**Huerta** C, Castro MG, Coto E, Blazquez M, Ribacoba R, Guisasola LM, Salvador C, Martinez C, Lahoz CH, Alvarez V (2005) *Mitochondrial DNA polymoorphisms and risk of Parkinson's disease in Spanish population*, J. Neurol. Sci.; 236: 49-54

**Holt** IJ, Harding AE, Petty RK, Morgan-Hughes JA (1990) *A new mitochondrial disease associated with mitochondrial DNA heteroplasmy*, Am. J. Hum. Genet.; 46: 428-33

**Holt** IJ, Lorimer HE, Jacobs HT (2000) *Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA* Cell 100: 515-24

**Hoth** M, Fanger CM, Lewis RS (1997) *Mitochondrial regulation of store-operated calcium signaling in T lymphocytes*, J. Cell. Biol.; 137: 633-648

**Huppert** FA, Whittington JC, (1995) *"Symptons of psychological distress predict 7-year mortality*, Psycological Medicine, 25: 1073-1086

**Ide** T, Tsutsui H, Hayashidani S, Kang D, Suematsu N, et al. (2001) *Mitochondrial DNA damage and dysfunction associated with oxidative stress in failing hearts after myocardial infarction Circulation Research* 88:529-535

**Ivanova** R, Lepage V, Charron D, Schächter F. *Mitochondrial genotype associated with French Caucasian centenarians.* Gerontology. 1998;44(6):349.

**Iwata** N, Zhang J, Atzmon G, Leanza S, Cho J, Chomyn A, Burk R, Barzilai N, Attardi G. *Aging-related occurrence in Ashkenazi jews of leukocyte heteroplasmic mtDNA mutation adjacent to replication origin frequently remodelled in Italian centenarians.* Mitochondrion 2007; 7: 267–272.

**Jacq** C, Miller JR, et al. (1977) *A pseudogene structure in 5S DNA of Xenopus laevis,* Cell; 12: 109-20

**Johnson**, A.A. and K.A. Johnson (2001) *Exonuclease proofreading by human mitochondrial DNA polymerase.* J Biol Chem, **276**(41): p. 38097-107.

**Johnson** MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL (1983) *Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns,* J. Mol. Evol; 19: 255-71

**Joeng** KS, Song EJ, Lee KJ, Lee J. *Long lifespan in worms with long telomeric DNA.* Nat Genet. 2004, 36(6): 607-11.

**Kaguni,** L.S., (2004) *DNA polymerase gamma, the mitochondrial replicase.* Annu Rev Biochem, 2004. **73**: p. 293-320.

**Kanaka-Gantenbein** C*. Fetal origins of adult diabetes*. Ann N Y Acad Sci. 2010 Sep;1205(1):99-105

**Kato** T (2001) *The other, forgotten genome: mitochondrial DNA and mental disorders* Molecular Psychiatry 6:625-633

**Katoh** K, Misawa K, Kuma K, Miyata T. (2002) *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform* Nucleic Acids Res*.* 30: 3059-66

**Kazuno** A, Munakata K, Nagai T, Shimozono S, Tanaka M, Yoneda M, Kato N, Miyawaki A, Kato T (2006) *Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics* Plos Genetics 2: 1167-1177

**Kirkwood** TB, Feder M, Finch CE, Franceschi C, Globerson A, Klingenberg CP, LaMarco K, Omholt S, Westendorp RG. *What accounts for the wide variation in life span of genetically identical organisms reared in a constant environment?* Mech Ageing Dev. 2005 Mar;126(3):439-43.

**Kivisild** T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scherfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL,

Oefner PJ (2006) *The role of selection in the evolution of human mitochondrial genomes* Genetics 172: 373-387

**Knoops** KT, De Groot LC, Kromhout D, Perrin AE, Moreiras- Varela O, Menotti A, et al. *Mediterranean diet, lifestyle factors, and 10-years mortality in elderly European men and women: the HALE project*. (2004) JAMA 292: 1433-9.

**Kowald** A, Kirkwood TB. *Accumulation of defective mitochondria through delayed degradation of damaged organelles and its possible role in the ageing of post-mitotic and dividing cells.* J Theor Biol. 2000 Jan 21;202(2):145-60.

**Kujoth** GC, Hiona A, Pugh TD, Someya S, Panzer K,Wohlgemuth SE, Hofer T, Seo AY, Sullivan R, Jobling WA, Morrow JD, Van Remmen H, Sedivy JM, Yamasoba T, Tanokura M, Weindruch R, Leeuwenburgh C, Prolla TA. (2005) *Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging* Science 309, 481–484.

**Kujoth** GC, Leeuwenburgh W, Prolla TA (2006) *Mitochondrial DNA Mutations and Apoptosis in Mammalian Aging Cancer* Res 66, 7386-7389

**Lee** CM et al: *Age-associated alterations of the mitochondrial genome*, Free Radic. Biol. Med., 1997; 22: 1259-1269

**Ling** F and Shibata T: *Recombination-dependent mtDNA partitioning: in vivo role of Mhr1p to promote pairing of homologous DNA*, EMBO J., 2002; 21: 4730-4740

**Ling** F and Shibata T: *Mhr1p-dependent concatemeric mitochondrial DNA formation for generating yeast mitochondrial homoplasmic cells,* Mol. Cell. Biol., 2004; 15: 310-322

**Macaulay** V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A. *The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs.* Am J Hum Genet. 1999 Jan;64(1):232-49.

**Madrigal** L, Meléndez-Obando M: *Grandmothers' longevity negatively affects daughters' fertility*. Am J Phys Anthropol 2008

**Malyarchuk** BA, Derenko MV. 2001. Variation of human mitochondrial DNA: distribution of hot spots in hypervariable segment I of the major noncoding region. Genetika. 37: 991–1001.

**Man** PY, Turnbull D.M., Chinnery P.F. (2002) *Leber hereditary optic neuropathy* J Med Genet 39: 162-169

**McKenzie** D., Bua E., McKiernan S., Cao Z., Wanagat J. and Aiken M.J. (2002) *Mitochondrial DNA deletion mutations* Eur J Biochem 269

**McLean** RR. *Proinflammatory cytokines and osteoporosis* (2009) Curr Osteoporos Rep. 7(4):134-9

**Mehta** P., Mellick GD, Rowe DB, Halliday GM, Jones MM, Manwaring N, Vandebona H, Silburn PA, Wang JJ, Mitchell P, Sue CM. *Mitochondrial DNA haplogroups J and K are not protective for Parkinson's disease in the Australian community*. (2009) Mov Disord 24(2):290-2

**Merriwether** DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC: *The structure of human mitochondrial DNA variation,* J. Mol. Evol., 1991; 33: 543-555

**Michikawa** Y, Mazzucchelli F, Bresolin N, Scarlato G, Attardi G: *Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication*, Science, 1999; 286: 774-779

**Miquel** J et al. (1980) *Mitochondrial role in cell aging* Exp Gerontol 15: 575-91.

**Mishmar** D, Ruiz-Pesini E, Golik P, Macaulay V, Clark A, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) *Natural selection shaped regional mtDNA variation in humans* PNAS 100: 171-176

**Mohlke** KL, Jackson AU, Scott LJ, Peck EC, Suh YD et al. (2005) *Mitochondrial polymorphisms and susceptibility to type 2 diabetes-related traits in Finns* Hum Genet 118: 245-254

**Motta** M., Bennati E., Ferlito L., Malaguarnera M., Motta L. (2005) *Successful aging in centenarians: myths and reality* Archives of Gerontology and Geriatrics *40: 241-251*.

**Niemi** AK, Hervonen A, Hurme M, Karhunen PJ, Jylha M, Majamaa K: *Mitochondrial DNA polymorphisms associated with longevity in a Finnish population*, Hum. Genet., 2003; 112: 29-33

**Niemi** AK, Moilanen JS, Tanaka M, Hervonen A, Hurme M, Lehtimaki T, Arai Y, Hirose N, Majamaa K: *A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects*, Eur. J. Hum. Genet., 2005; 13: 166-170

**Nijtmans** LG, Henderson NS, Attardi G, Holt IJ (2001*) Impaired ATP synthase assembly associated with a mutation in the human ATP synthase subunit 6 gene* J. Biol. Chem. 276: 6755-62

**Nishigaki** Y, Tadesse S, Bonilla E, Shungu D, Hersh S, Keats JB, Berlin CI, Goldberg MF, Vockley J, Di Mauro S, Hirano M (2003) *A novel mitochondrial tRNA$^{Leu(UUR)}$ mutation in a patient with features of MERRF and Kearns–Sayre syndrome* Neuromuscular Disorders 13: 334–340

**Nishigaki** Y, Yamada Y, Fuku N, Matsuo H, Segawa T, Watanabe S, Kato K, Yokoi K, Yamaguchi S, Nozawa Y, Tanaka M (2006) *Mitochondrial haplogroup N9b is protective against myocardial infarction in Japanese male* Hum Genet 120: 827-36

**Norio** R, Nevanlinna HR, Perheentupa J. *Hereditary diseases in Finland; rare flora in rare soul.* Ann Clin Res. 1973 Jun;5(3):109-41..

**Olivieri** A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt HJ,

Torroni A: *The mtDNA legacy of the levantine early upper palaeolithic in Africa*, Science, 2006; 314: 1766-1770

**Olsen** RB, Olsen J, Gunner-Svensson, Waldstrom B**.** *Social networks and longevity. A 14 year follow-up study among elderly in Denmark.* (1991) Soc. Sci. Med.; 33:1189-1195

**Panahi** MSS, Houshmand M, Tabassi AR (2006) *Mitochondrial D-loop variation in Leber Hereditary Neuropathy patients harboring primary G11778A, G3460A, T14484C mutations: J and W haplogroups as high-risk factors* Archives of Medical Research 37: 1028-1033

**Perls** TT, Bubrick E, Wager CG, Vijg J, Kruglyak L. *Siblings of centenarians live longer.* (1998) Lancet 351(9115):1560

**Perls** T, Kunkel LM, Puca A. *The genetics of exceptional human longevity.* (2002) Rev. J. Am. Geriatr. Soc. 50:359-368

**Pesole** G, Saccone C. (2001) *A novel method for estimating substitution rate variation among sites in a large dataset of homologous DNA sequences* Genetics 157: 859-65

**Pitkanen** S, Robinson BH (1997) *Mitochondrial complex I deficiency leads to increased production of superoxide radicals and induction of superoxide dismutase* J. Clin. Invest. 98: 345–351

**Poulton** J, Bednarz AL, Scott-Brown M, Thompson C, Macaulay VA, Simmons D. (2002) *The presence of a common mitochondrial DNA variant is associated with fasting insulin levels in Europeans in Auckland* Diabet Med. 19: 969-71

**Puca**, AA et al. *A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4***.** (2001) Proc. Natl Acad. Sci. USA 98, 10505–10508

**Pyle** A, Foltynie T, Tiangyou W, Lambert C, Keers SM, Allcock LM, Davison J, Lewis SJ, Perry RH, Barker R, Burn DJ, Chinnery PF*. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD.* (2005) Ann Neurol 57(4):564-7

**Raule** N, Sevini F, Santoro A, Altilia S, Franceschi C: *Association studies on human mitochondrial DNA: methodological aspects and result in the most common age-related diseases*, Mitochondrion, 2007; 7: 29-38

**Rea** IM, Mc Dowell I, McMaster D, Smye M, Stout R, Evans A; MONICA group (Belfast). Monitoring of Cardiovascular trends study group. *Apolipoprotein E alleles in nonagenarian subjects in the Belfast Elderly Longitudinal Free-living Ageing Study (BELFAST).* (2001) Mech Ageing Dev. 122:1367-72

**Reeve** AK, Krishnan K.J, Turnbull D (2008) *Age related mitochondrial degenerative disorders in humans* Biotechnol. J,

**Reich** DE, Lander ES (2001) *On the allelic spectrum of human disease* TRENDS in Genetics 17: 502-510

**Reynier** P, Penisson-Besnier I, Moreau C, Savagner F, Vielle B, Emile J, Bubas F, Malthiery Y: *mtDNA haplogroup J: a contributing factor of optic neuritis*, Eur. J. Hum. Genet., 1999; 7: 404-406

**Richards** MB, Macaulay VA, Bandelt HJ, Sykes BC. *Phylogeography of mitochondrial DNA in western Europe.* Ann Hum Genet. 1998 May;62(Pt 3):241-60.

**Richards** M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, et al. *Tracing European founder lineages in the Near Eastern mtDNA pool.* Am J Hum Genet. 2000 Nov;67(5):1251-76. Epub 2000 Oct 16.

**Richter**, C. (1995) "Oxidative damage to mitochondrial DNA and its relationship to ageing" Int J Biochem Cell Biol 27: 647-53.

**Rose** G, Passarino G, Franceschi C, De Benedictis G. (2002) *The variability of the mitochondrial genome in human aging: a key for life or death?* The international journal of Biochemistry and Cell Biology 34, 1449-1460

**Rose** G, Dato S, Altomare K, Bellizzi D, Garasto S, Greco V, Passarino G, Feraco E, Mari V, Barbi C, BonaFe M, Franceschi C, Tan Q, Boiko S, Yashin AI, De Benedictis G. *Variability of the SIRT3 gene, human silent information regulator Sir2 homologue, and survivorship in the elderly.* Exp Gerontol. 2003 Oct;38(10):1065-70.

**Rose** G, Passarino G, Scornaienchi V, Romeo G, Dato S, Bellizzi D, Mari V, Feraco E, Maletta R, Bruni A, Franceschi C, De Benedictis G. *The mitochondrial DNA control region shows genetically correlated levels of heteroplasmy in leukocytes of centenarians and their offspring* BMC Genomics. 2007 Aug 29;8:293.

**Roses** AD, Einstein G, Gilbert J, Goedert M, Han SH, et al: *Morphological, biochemical, and genetic support for an apolipoprotein E effect on microtubular metabolism*, Ann. NY Acad., 1996; 777: 146-57

**Ross** OA, McCormack R, Curran MD, Duguid RA, Barnett YA, Rea IM; Middleton D: *Mitochondrial DNA polymorphism: its role in longevity of the Irish population*, Exp. Gerontol., 2001; 36: 1161-1178

**Ross** OA, McCormack R, Maxwell LD, Duguid RA, Quinn DJ, Barnett YA, Rea IM, El-Agnaf OM, Gibson JM, Wallace A, Middleton D, Curran MD: *mt4216C variant in linkage with the mtDNA TJ cluster may confer a susceptibility to mitochondrial dysfunction resulting in an icreased risk of Parkinson's disease in the Irish*, Exp. Gerontol, 2003; 38: 397-405

**Ruiz-Pesini** E, Lapena AC, Diez-Sanchez C et al *Human mtDNA haplogroups associated with high or reduced spermatozoa motility***.** Am J Hum Genet 2000, 67: 682-696

**Samuels** DC: *Mitochondrial DNA repeats constrain the life span of mammals*, Trends Genet., 2005; 20: 226-228

**Santoro** A, Salvioli S, Raule N, Capri M, Sevini F, Valesin S, Monti D, Bellizzi D, Passarino G, Rose G, De Benedictis G, Franceschi C: *Mitochondrial DNA involvement in human longevity*, Biochim. Biophys. Acta, 2006; 1757: 1388-99

**Saraste** M. (1999) *Oxidative Phosphorylation at the fin de siècle* Science, 283: 1488-93

**Scaglia** F and Wong LJ. *Human mitochondrial transfer RNAs: role of pathogenic mutation in disease Muscle Nerve* 2008, 37 (2): 150-71

**Sebastiani** P, Hadley EC, Province M, Christensen K, Rossi W, Perls TT, Ash AS. *A family longevity selection score: ranking sibships by their longevity, size, and availability for study.* (2009) Am J Epidemiol. 170(12):1555-62.

**Sebastiani** P, Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB, Myers RH, Steinberg MH, Montano M, Baldwin CT, Perls TT. *Genetic Signatures of Exceptional Longevity in Humans.* (2010) Science

**Shoffner** JM, Lott MT, Lezza AM, Seibel P, Ballinger SW, Wallace DC: *Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA$^{Lys}$ mutation*, Cell, 1990; 61: 931-37

**Shoffner** JM, Brown MD, Torroni A, Lott MT, Cabell MF, Mirra SS, Beal MF, Yang CC, Gearing M, Salvo R. et al. (1993) *Mitochondrial DNA variants observed in Alzheimer disease and Parkinson disease patients* Genomics 17: 171-84

**Small** BJ ans Backman L (1999) "*Time to death and cognitive performance*" Current Directions in Psycological Science, 8: 168-172

**Smeitink** JA, Zeviani M, Turnbull DM, Jacobs HT (2006) *Mitochondrial medicine: a metabolic perspective on the pathology of oxidative phosphorylation disorders* Cell metabolism 3, 9-13

**Smith** MA, Casadesus G, Joseph JA, Perry G: *Amyloid-beta and tau serve antioxidant functions in the aging and Alzheimer brain*, Free Radic. Biol. Med., 2002; 33: 1194-99

**Sturtz** LA, Diekert K, Jensen LT, Lill R, Culotta VC: *A fraction of yeast Cu, Zn-superoxide dismutase and its metallochaperone, CSS, localize to the intermembrane space of mitochondria. A physiological role for SOD1 in guarding against mitochondrial oxidative damage*, J. Biol. Chem., 2001; 276: 38084-89

**Taanman** JW (1999) *The mitochondrial genome: structure, transcription, translation and replication* Biochimica et Biophysica Acta 1410: 103-123

**Tanaka** M, Gong JS, Zhang J, Yamada Y,Borgeld HJ, Yagi K. (2000) *Mitochondrial genotype associated with longevity and its inhibitory effect on mutagenesis* Mechanism of ageing and development 116: 65-76

**Tanaka** M, Jia-Shun G, Jie Z, Masato Y, Kunimasa Y. *Mitochondrial genotype associated with longevity*. The Lancet 1998; 351: 185–186.

**Tanaka** M (2002) Mitochondrial genotypes and cytochrome b variants associated with longevity or Parkinson's disease, J Neurol 249: II/11-II/18

**Tatar**, M., Bartke, A., Antebi, A. *The endocrine regulation of aging by insulin-like signals.* (2003) Science, 1346–1351

**Tapper** DP and Clayton DA: *Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands*, J. Biol. Chem., 1981; 256: 5109-15

**Tatuch** Y, Christodoulou J, Feigenbum A, Clarke JTR, Wherret J et al: *Heteroplasmic mtDNA mutation (T-G) at 8993 can cause Liegh disease when the percentage of abnormal mtDNA is high*, Am. J. Hum. Genet., 1992; 50: 852-58

**Theves** C, Keyser-Tracqui C, Crubezy E, Salles JP, Ludes B, Telmon N. (2006) *Detection and Quantification of the Age-Related Point Mutation A189G in the Human Mitochondrial DNA* J Forensic Sci 51: 865-873

**Torroni** A and Wallace DC: *Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance*, J. Bioenerg. Biomembranes, 1994; 26: 261-271

**Torroni** A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC. *Classification of European mtDNAs from an analysis of three European populations.* Genetics. 1996 Dec;144(4):1835-50.

**Torroni** A, Petrozzi M, D'Urbano L, Sellito D, Zeviani M, Carrara F, Carducci C, Leuzzi V, Carelli V, Barboni P, De Negri A, Scozzari R: *Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484*, Am. J. Hum. Genet., 1997; 60: 1107-1121

**Torroni** A, Campos Y, Rengo C, Sellitto D, Achilli A, Magri C, Semino O, Garcìa A, Jara P, Arenas J, Scozzari R. (2003) *Mitochondrial DNA haplogroups do not play a role in the variable phenotypic presentation of the A3243G mutation* Am. J. Hum. Genet. 72: 1005-1012

**Torroni** A, Achilli A, Macaulay V, Richards M, Bandelt HJ: *Harvesting the fruit of the human mtDNA tree*, Trends in Genetics, 2006; 22: 339-345

**Trifunovic** A, Wredenberg A, Farkenberg M et al (2004) *Premature aging in mice expressing defective mitochondrial DNA polymerase* Nature 429: 417-423

**Trifunovic** A, Hansson A, Wredenberg A, Rovio AT, Dufour E, Khvorostov I, Spelbrink JN, Wibon R, Jacobs HT, Larsson NG. (2005) *Somatic mtDNA mutations cause aging phenotypes without affecting reactive oxygen species production* PNAS 102: 17993-17998

**Umetsu** K, Yuasa I. (2005) *Recent progress in mithocondrial DNA analysis* Legal Medicine 7, 259–262

**Van der Walt** JM, Nicodemus KK, Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Haines JL, Koller WC, Lyons K, Pahwa R, Stern MB, Colcher A, Hiner BC, Jankovic J, Ondo WG, Allen Jr. FH, Goetz CG, Small GW, Mastaglia F, Stajich JM, McLaurin AC, Middleton LT, Scott BL, Schmechel DE, Pericak-Vance MA, Vance JM: *Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease*, Am. J. Hum. Genet., 2003; 72: 804-811

**van der Walt** JM, Scott WK, Slifer S, Gaskell PC, Martin ER, Welsh-Bohmer K, et al. *Maternal lineages and Alzheimer disease risk in the Old Order Amish.* (2005) Hum Genet. 118(1):115-22.

**Vanin** EF: *Processed pseudogenes: characteristics and evolution*, Annu Rev Genet, 1985; 19: 253-72

**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. *Biodemographic trajectories of longevity*. (1998) Science. 280(5365):855-60.

**Vellai** T, Takacs K, Vida G. (1998) *A new aspect to the origin and evolution of eukaryotes* J Mol Evol 46:499–507

**Wallace** DC, Singh G, Lott MT, Hodge JA, Schurr TG: *Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy*, Science, 1988; 242: 1427-30

**Wallace** DC, Zheng X, Lott MT, Shoffner JM, Hodge JA et al: *Familial mitochondrial encephalomyopathy (MERRF): genetic, pathophysiological, and biochemical characterization of a mitochondrial DNA disease*, Cell, 1988; 55: 601-10

**Wallace** DC: *Diseases of the mitochondrial DNA,* Annu. Rev. Biochem., 1992; 61: 1175-212

**Wallace** DC: *Mitochondrial DNA sequence variation in human evolution and disease*, Proc. Natl. Acad. Sci. U.S.A., 1994; 91: 8739-8746

**Wallace** DC: *Mitochondrial DNA variation in human evolution, degenerative disease, and aging*, Am. J. Hum. Genet., 1995; 57: 201-223

**Wallace** DC: *Mitochondrial diseases in man and mouse,* Science, 1999; 283: 1482-88

**Wallace** DC, Brown MD, Lott MT: *Mitochondrial DNA variation in human evolution and disease*, Gene, 1999; 238: 211-30

**Wallace** DC, Lott MT, Brown MD, Kerstann K: *Mitochondria and neuro-ophthalmogical diseases*. In *The Metabolic and Molecular Basis of Inherited Disease*, ed. CR Scriver, AL Beaudet, WS Sly, D Valle, 2001; pp. 2425-512. New York: McGraw-Hill

**Wallace** DC, Lott MT: *Mitochondrial genes in degenerative diseases, cancer and aging*. In *Emery and Rimoin's Principles and Practice of Medical Genetics*, ed. DL Rimoin, JM Connor, RE Pyeritz, BR Korf, 2002; pp. 299-409. London: Churchill Livingstone

**Wallace** DC, *A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine*, Annu. Rev. Genet., 2005; 39: 359-407

**Wang** H, Kazemi-Esfarjani P, Benzer S. *Multiple-stress analysis for isolation of Drosophila longevity genes* PNAS, 2004; 101: 12610-12615

**Wang** Y, Michikawa Y, Mallidis C, Bai Y, Woodhouse L, Yarasheski KE, Miller CA, Askanas V, Engel WK, Bhasin S & Attardi G. (2001) *Muscle-specific mutations accumulate with aging in critical human mtDNA control sites for replication.* Proc Natl Acad Sci U S A. 98(7): 4022-7.

**Wei** YH, Lee HC. (2002) *Oxidative Stress, Mitochondrial DNA Mutation, and Impairment of Antioxidant Enzymes in Aging* Exp Biol Med 227: 671–682, 2002

**Weinert** BT, Timiras PS. *Invited review: Theories of aging*. J Appl Physiol. 2003 Oct;95(4):1706-16.

**Woo** J, Ho SC, Yu AL. *Lifestyle factors and health otcomes in elderly Hong Kong chinese aged 70 years and over.* (2002) Gerontology; 48: 234-40.

**Zeviani** M, Di Donato S. (2004) *Mitochondrial disorders* Brain 127:2153-2172

**Zhang** J, Asin-Cayuela J, Fish J, Michikawa Y, Bonafè M, Olivieri F, Passarino G, De Benedictis G, Franceschi C, Attardi G, *Strikingly higher frequency in centenarians and twins of mtDNA mutation causing remodeling of replication origin in leukocytes*, Proc. Natl. Acad. Sci., 2003; 100: 1116-1121

**Zsurka** G, Kàlmàn J, Csaszar A, Rasko I, Janka Z, Venetianer P. (1999) *No mitochondrial haplotype was found to increase risk for Alzheimer's disease* Society of Biological Psychiatry 44: 371-37