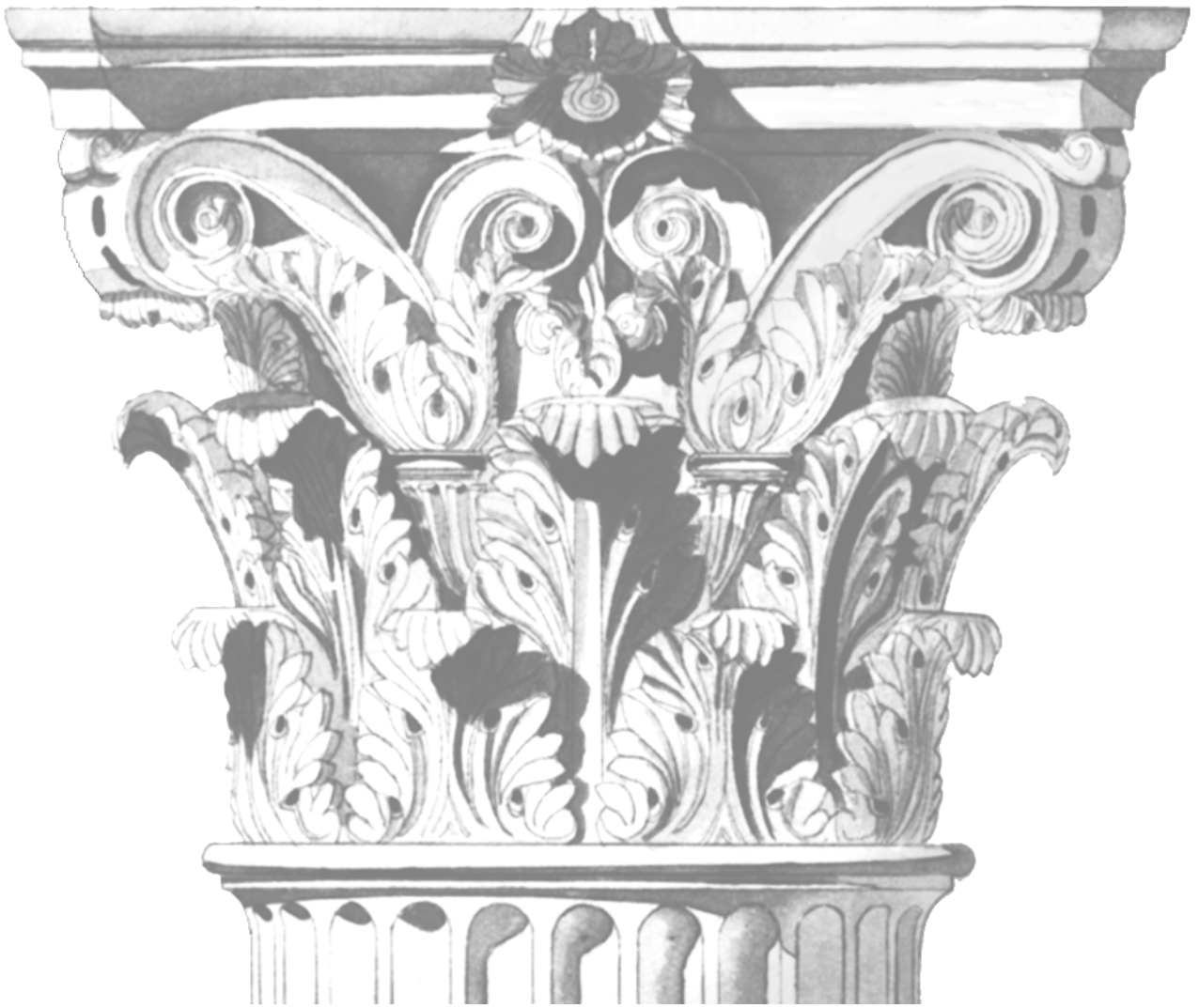
NON-NEUTRAL SEQUENCE VARIATION IN HUMAN MITOCHONDRIAL DNA: SELECTION AGAINST DELETERIOUS MUTATIONS AND HAPLOGROUP-RELATED POLYMORPHISMS

JUKKA MOILANEN

Department of Neurology,
Biocenter Oulu,
University of Oulu

OULU 2003



JUKKA MOILANEN

NON-NEUTRAL SEQUENCE VARIATION IN HUMAN MITOCHONDRIAL DNA: SELECTION AGAINST DELETERIOUS MUTATIONS AND HAPLOGROUP-RELATED POLYMORPHISMS

Academic Dissertation to be presented with the assent of the Faculty of Medicine, University of Oulu, for public discussion in the Auditorium 8 of Oulu University Hospital, on October 31st, 2003, at 12 noon.

OULUN YLIOPISTO, OULU 2003

Supervised by
Professor Kari Majamaa

Reviewed by
Professor Jaakko Kaprio
Professor Mauno Vihinen

ISBN 951-42-7166-1 (URL: http://herkules.oulu.fi/isbn9514271661/)

ALSO AVAILABLE IN PRINTED FORMAT
Acta Univ. Oul. D 757, 2003
ISBN 951-42-7165-3
ISSN 0355-3221 (URL: http://herkules.oulu.fi/issn03553221/)

OULU UNIVERSITY PRESS
OULU 2003

Moilanen, Jukka, Non-neutral sequence variation in human mitochondrial DNA: selection against deleterious mutations and haplogroup-related polymorphisms
Department of Neurology; Biocenter Oulu, University of Oulu, P.O.Box 5000, FIN-90014 University of Oulu, Finland
Oulu, Finland
2003

Abstract

Mitochondrial DNA (mtDNA) is a maternally inherited 16.6 kbp circular genome that codes for 13 subunits of the mitochondrial respiratory chain, 2 rRNAs and 22 tRNAs. The mutation rate in mtDNA is high and therefore, mutations have accumulated sequentially to lineages that have diverged tens of thousands of years ago. The neutral theory predicts that a proportion of these variations may be slightly deleterious, associated with diseases and selected against, but the issue is still controversial.

This study reports an analysis of selection against mutations in mtDNA. First, the population prevalence of one of the most pathogenic mtDNA mutations, the common MELAS mutation (3243A>G), was determined in a population-based screening setting in Northern Ostrobothnia, and the reproductive capacity, or genetic fitness, of women with the mutation was estimated in order to measure for the first time the degree of host-level selection against this highly pathogenic mutation. The frequency of 3243A>G was high, as the minimum estimate for the prevalence was 10.2/100,000, and this together with the geographical distribution of maternal ancestors of the mutation carriers suggested that nuclear genes may be involved in the population history of the mutation. Surprisingly, the genetic fitness of mutation carriers was not reduced, suggesting that the average host-level selection against carriers is not strong. Second, all available complete human mtDNA sequences worldwide (*N*=847) were collected into a database and analysed for evidence to support the hypothesis concerning slightly deleterious mutations and selective constraints imposed by lineage-specific interactions. 465 distinct missense and 6 nonsense mutations were identified. 48% of the amino acid replacements changed the polarity, 44% hydropathy, 32% aliphaticity, 26% size, 13% aromaticity, and 8% charge. Nonconservative amino acid replacements were found to be more common among the evolutionarily recent mutations than among the older ones, and mutations that have arisen more than once during human evolution showed different properties from the remaining ones. The major continent-specific mtDNA lineages were analysed in terms of nucleotide diversity indices, neutrality tests and nonsynonymous/synonymous rate ratios, and patterns suggesting selective constraints possibly due to lineage-specific interactions were identified. Moreover, a general correlation between nucleotide position and nucleotide polymorphism was identified in the mtDNA.

The results are compatible with the assumption that selection has a marked role in human mtDNA evolution and that selective constraints may vary between populations, so that the pathogenic potential of a given mutation may depend markedly on the presence of other, interacting mutations.

Keywords: genetic epidemiology, mitochondrial DNA, mitochondrial encephalomyopathies, phylogenetic network, population genetics, selection

Acknowledgements

This research was carried out at Biocenter Oulu and the Department of Neurology at the University of Oulu during the years 1996–2003.

I wish to express my deepest gratitude to my supervisor Professor Kari Majamaa, M.D., Ph.D., who introduced me to scientific thinking and provided good possibilities to carry out this work. His unique enthusiasm, optimism, sense of humour, support and constant willingness to share his vast knowledge on all aspects of mitochondrial genetics, neurology and statistics have been greatly appreciated and invaluable for the completion of this work.

I wish to thank Professor Vilho Myllylä, M.D., Ph.D., head of the Department of Neurology, for giving me the opportunity to carry out the research in his department. I also wish to thank Professor Jaakko Kaprio, M.D., Ph.D., and Professor Mauno Vihinen, Ph.D., who revised the manuscript, for their constructive criticism and kind advice. I am also grateful to Malcolm Hicks, M.A., for revision of the English language of the original articles and of this thesis.

I thank all my co-authors for their important collaboration, and I also wish to thank all the members of the Northern Mitochondria team for creating an inspiring and pleasant atmosphere during these years. Most importantly, my family members deserve my warmest thanks for their support and patience.

I sincerely thank all the patients and their family members who kindly participated in this research. It is my understanding that we share the hope that this research will be a small step on a path which will eventually lead to a complete understanding of mitochondrial genetics and efficient treatments for mitochondrial disorders.

This work was supported financially by the Sigrid Juselius Foundation, the Maud Kuistila Memorial Foundation, Oulun Lääketieteellinen Tutkimussäätiö, and Neurologiasäätiö.

Abbreviations

ATP	adenosine triphosphate
bp	base pair
D-loop	displacement loop
DNA	deoxyribonucleic acid
HVS	hypervariable segment
LHON	Leber's hereditary optic neuropathy
MELAS	mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes
MERRF	myoclonus epilepsy with ragged-red fibres
mRNA	messenger RNA
mtDNA	mitochondrial DNA
mTERF	transcription termination factor
NADH	reduced nicotinamide adenine dinucleotide
NARP	neurogenic weakness, ataxia and retinitis pigmentosa
NRR	net reproduction rate
OXPHOS	oxidative phosphorylation
OR	odds ratio
P	P value
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
ROS	radical oxygen species
rRNA	ribosomal ribonucleic acid
tRNA	transfer ribonucleic acid
YBP	years before present

List of original articles

I Majamaa K, Moilanen JS, Uimonen S, Remes AM, Salmela PI, Kärppä M, Majamaa-Voltti KA, Rusanen H, Sorri M, Peuhkurinen KJ & Hassinen IE (1998) Epidemiology of A3243G, the mutation for mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes: prevalence of the mutation in an adult population. Am J Hum Genet 63:447-454.

II Moilanen JS & Majamaa K (2001) Relative fitness of carriers of the mitochondrial DNA mutation 3243A>G. Eur J Hum Genet 9:59-62.

III Moilanen JS & Majamaa K (2003) Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA. Mol Biol Evol 20:1195-1210.

IV Moilanen JS, Finnilä S & Majamaa K. Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J. Mol Biol Evol, in press.

Contents

1 Introduction

Mitochondria are intracellular organelles that are present in eukaryotic cells, and a typical human cell has several hundred of these. They have a central role in cellular energy metabolism and apoptosis, which is a particular form of programmed cell death. Mitochondria are evolutionarily ancient, and it is assumed that their life as intracellular endosymbionts started at the very beginning of the evolution of eukaryotes. They still harbour their own DNA, however, known as mitochondrial DNA (mtDNA), and although many of its genes have been transferred to the nucleus during the millions of years of co-evolution, the mtDNA still codes for certain essential components of the mitochondrial respiratory chain and protein synthesis apparatus.

The human mtDNA is a maternally inherited, circular 16.6 kbp molecule which is present in several copies in each mitochondrion. The maternal inheritance and high mutation rate have resulted in maternal lineages which have diverged thousands or tens of thousands of years ago and accumulated mutations sequentially thereafter. While there are many examples of pathogenic mtDNA mutations, there is also evidence that some of the mtDNA variation that is observed in populations might be slightly deleterious, that is, it might include susceptibility alleles for certain complex diseases. Mutations in nuclear genes of the mitochondrial machinery can also be pathogenic, and both the nuclear-encoded and mtDNA-encoded components of the mitochondria are very probably connected by complex interactions with each other and with the environment, so that such interactions impose structural and functional constraints on the genes. Consequently, structural and functional constraints may vary between populations and mtDNA genotypes, possibly explaining why the risks of certain traits appear to be higher in some mtDNA lineages than in others.

The aims of this work were to study selective constraints against 3243A>G, a definitely deleterious mtDNA mutation, in the population of Northern Ostrobothnia, and to find evidence for slightly deleterious mutations and selective constraints imposed by lineage-specific interactions in a data set consisting of all available complete mtDNA sequences worldwide.

2 Review of the literature

2.1 Mitochondria

2.1.1 Structure

The mitochondrion is a cytoplasmic organelle that is present in eukaryotic cells. A typical human cell has several hundred of these. Mitochondria are 0.5–1 μm in size and are bound by two membranes, the inner one of which forms tubular or lamellar structures called cristae (Reichert & Neupert 2002, Riva *et al.* 2003). The space inside the inner membrane, called the matrix, contains various enzymes, proteins, ribosomes, tRNAs and mitochondrial DNA (mtDNA) molecules. Mitochondrial biogenesis is a complex process producing about 20% of cellular protein, and the number of nuclear genes involved is approximately 1000 (Scarpulla 2002, Goffart & Wiesner 2003). Nuclear-encoded intra-mitochondrial proteins are imported into mitochondria through membrane-bound channel proteins (Muro *et al.* 2003, Truscott *et al.* 2003), the energy-generating apparatus being located in the inner membrane and being partially encoded by the mtDNA. Mitochondria are morphologically heterogeneous within cells (Collins & Bootman 2003). They can fuse and form complex reticulations, and their morphology is under nuclear genetic control (Capaldi *et al.* 2002, Santel *et al.* 2003). The location of mitochondria in the cytosol is non-random. They can move, and they are transported in response to physiological changes (Chada & Hollenbeck 2003)

2.1.2 Function

2.1.2.1 Respiratory chain

Mitochondria have a central role in cellular energy metabolism. The oxidative phosphorylation system (OXPHOS system), which is the final common pathway in the production of ATP (adenosine triphosphate) from glucose, is located in the inner mitochondrial membrane and consists of five protein complexes. Complex

I (NADH:quinone oxidoreductase, NADH dehydrogenase), the largest and least understood, consists of at least 42 subunits (Saraste 1999), seven of which are encoded by mtDNA (MTND1, MTND2, MTND3, MTND4, MTND4L, MTND5, and MTND6). It is L-shaped and has two major domains that are separated by a thin collar (Guenebaut *et al.* 1998), and it may exist in two distinct forms (Grivennikova *et al.* 2003). All four subunits of Complex II (succinate-ubiquinone reductase) are encoded by nuclear genes. Complex III (ubiquinol:ferrocytochrome c oxidoreductase, cytochrome bc1) has one mtDNA-encoded subunit, cytochrome b (MTCYB), and Complex IV (cytochrome c oxidase, COX) has three mtDNA-encoded subunits (MTCO1, MTCO2, and MTCO3), whereas ten are encoded by nuclear DNA. Complex V (ATP synthase), the structure and function of which are quite well known (Capaldi & Aggeler 2002), has two mtDNA-encoded subunits, MTATP6 and MTATP8. The subunits of the respiratory chain are transported into the inner membrane by a membrane insertion machinery (Stuart 2002), and the precursor polypeptides are processed by peptidases (Gakh *et al.* 2002). Some subunits of the respiratory complexes of mitochondria are of an ancient evolutionary origin, as they have homologues in bacteria, for example (Baker *et al.* 1999, Mathiesen & Hägerhäll 2002).

2.1.2.2 Other cellular functions

In addition to their function in respiration, the mitochondria have an important role in apoptosis, which is a form of programmed cell death that is important in the development and tissue homeostasis of multicellular organisms. Malfunctioning at any level of the cellular apoptotic-signalling pathway is eventually translated into release of apoptogenic factors from the mitochondrial intermembrane space, which results in organized cell death. The role of mitochondria in the apoptotic pathway is complex and not fully understood (Robertson & Orrenius 2002, van Gurp *et al.* 2003). Mitochondrial energy production may also be linked to apoptosis. The respiratory chain produces radical oxygen species (ROS), and its impairment increases this production (Pitkänen & Robinson 1996, Barrientos & Moraes 1999, Adam-Vizi 2003). On the other hand, the level of ROS production may influence the rate of apoptosis (Barrientos & Moraes 1999, Chomyn & Attardi 2003). Mitochondrial factors that influence the regulation of apoptosis may be important in determining susceptibility to certain diseases where apoptosis is involved in the pathogenesis.

Mitochondria have a large capacity to accumulate and retain calcium (Chalmers & Nicholls 2003), and this buffering capability is important in Ca^{2+} signalling, the regulation of metabolism, ATP production and other processes (Park *et al.* 2001), including neurosecretion (Montero *et al.* 2002). An overload of the Ca^{2+} buffering capacity, ischaemia/reperfusion and oxidative stress can lead to the opening of the mitochondrial permeability transition pore, and consequently to the mitochondrial permeability transition (MPT), which is a common pathway to necrosis and apoptosis (Kim *et al.* 2003).

2.1.2.3 Role in ageing

Mitochondrial dysfunction and the accumulation of protein damage are thought to contribute to ageing (Mandavilli *et al.* 2002, Rose *et al.* 2002). The biology of radical oxygen species may be the main determinant of the life span, at least in some organisms (Hekimi & Guarente 2003), but it is probably not the only mechanism relevant to ageing (Lee *et al.* 2003). Somatic mutations accumulate in mtDNA during ageing (Bo *et al.* 2002, Chinnery *et al.* 2002, Fayet *et al.* 2002, Lin *et al.* 2002), and the level of mtDNA decreases (Welle *et al.* 2003). This may be due to diminished mtDNA repair activity (Chen *et al.* 2002). The average level of the accumulation of somatic mutations is not high, but it has been suggested that the relatively few cells that have lost their oxidative phosphorylation capacity may be toxic to the rest of the body (de Grey 2002). The accumulation of somatic mutations with age may also be involved in the pathogenesis of certain diseases, such as Parkinson's disease (Gu *et al.* 2002, Tanaka *et al.* 2002). The role of mitochondria in ageing is further supported by the observation that particular mtDNA haplogroups or polymorphisms may be associated with successful ageing (Khaidakov *et al.* 2003, Niemi *et al.* 2003). It has been suggested that the accumulation of somatic mutations and other age-related processes takes place because there is no evolutionary selection against such processes in multicellular organisms after the reproductive period (Hughes *et al.* 2002, Martin 2002).

2.1.3 Evolutionary origin

It is generally assumed that the mitochondrion is a remnant of a prokaryotic bacterium which became endosymbiotic with the eukaryotic cell soon after the advent of the first eukaryotes. Alternatively, the fusion of an anaerobic archaebacterium with a respiration-capable proteobacterium may have been the event that initiated the evolution of eukaryotes (Roger *et al.* 1998, Vellai *et al.* 1998, Vellai & Vida 1999, Gray *et al.* 2001, Simpson & Roger 2002, Amiri *et al.* 2003, Emelyanov 2003). Various forms of mitochondria-like organelles with different functional properties exist in diverse eukaryotic organisms, e.g. anaerobic mitochondria and hydrogenosomes, and these share common ancestry with the mitochondria (Dyall *et al.* 2000, Tielens *et al.* 2002). Mitochondria have co-evolved with their host organisms, and during this process many genes that have presumably been present in the original endosymbiont have been transferred to the nuclear DNA of the eukaryotic cell. Such copies of mtDNA genes (*numts*) have also been identified in humans. They are usually nonfunctional pseudogenes, but it has been suggested that some *numts* might even be functional (Tourmen *et al.* 2002, Woischnik & Moraes 2002). The transfer of mtDNA to nuclear DNA seems to be an ongoing process, as *de novo* transfer of mtDNA to nuclear DNA has been shown to cause a genetic disease by introducing a premature stop codon within a functional nuclear gene (Turner *et al.* 2003). If the transfer from mtDNA to nuclear DNA is such a common process, there are probably important functional reasons why the 13 mtDNA-encoded proteins and other mtDNA genes are still encoded by mtDNA. Such reasons may be related to restrictions on protein import into mitochondria, for instance.

2.1.4 Mitochondrial DNA

2.1.4.1 Genome organization

The long co-evolution of mitochondria and their host organisms has resulted in different organizations of the mtDNA in the major eukaryotic kingdoms. In animals the mtDNAs are typically small (approximately 16 kbp), circular molecules that encode 37 or fewer tightly packed genes, but even the closest unicellular relatives of animals have highly different genome organizations (Burger *et al.* 2003). The mitochondrial DNA in mammals is organized into protein-rich structures within mitochondria (nucleoids), which may also be the units of mtDNA inheritance (Garrido *et al.* 2003), and mitochondrial transcription factor A (TFAM) seems to be at least one of the proteins that are involved in the packaging of mtDNA (Alam *et al.* 2003). The human mtDNA is a circular, double-stranded DNA molecule (Anderson *et al.* 1981) that is present in 2–10 copies in each mitochondrion. It is typically 16,568 bp in length, but the total number of bases varies slightly depending on the presence of insertions/deletions and the lengths of certain highly variable tandem repeats. The mtDNA molecule has two strands, a guanine-rich heavy strand (H) and cytosine-rich light strand (L), and it codes for 13 proteins, 22 tRNAs and 2 rRNAs. Most genes are transcribed from the H-strand, but one subunit of Complex I (MTND6) and eight tRNAs are transcribed from the L-strand. MtDNA has no introns, and with the exception of a few noncoding nucleotides between some genes, the only noncoding region in mtDNA is the displacement loop (D-loop), a region of 1121 bp which contains the promoters for L and H-strand transcription and the origin of replication of the H-strand (Figure 1). The transcription of both strands is directed by the promoters in the D-loop, and almost the entire length of both strands is used to produce polycistronic RNA transcripts, which are subsequently cleaved to produce tRNAs, rRNAs and mRNAs. The transcription and translation of mtDNA are controlled by nuclear genes (Falkenberg *et al.* 2002, Shoubridge 2002, Casas *et al.* 2003, Rodeheffer 2003). It appears, however, that the nuclear transcription of the respiratory chain subunits may not be coordinated with the transcription of mtDNA-encoded subunits (Duborjal *et al.* 2002).

The genetic code of mtDNA has evolved to become different from that of nuclear DNA, and furthermore, there are several variations of the mtDNA genetic code (Barrell *et al.* 1979, Knight *et al.* 2001). In mammalian mtDNA, UGA codes for tryptophan instead of termination, AUA codes for methionine instead of isoleucine, AGA and AGG are terminators instead of coding for arginine, and AUA or AUU is sometimes used as an initiation codon instead of AUG. These differences make the protein-coding genes of mtDNA unintelligible to the nucleocytosolic system, except for the MTND4L gene which does not contain any UGA, AGA or AGG codons.

2.1.4.2 Relaxed replication

The widely accepted hypothesis of mtDNA replication states that the mtDNA is replicated from two origins (Figure 1) and that the replication is bidirectional and asynchronous (Clayton 1982). Replication is initiated at the heavy strand replication origin using a short

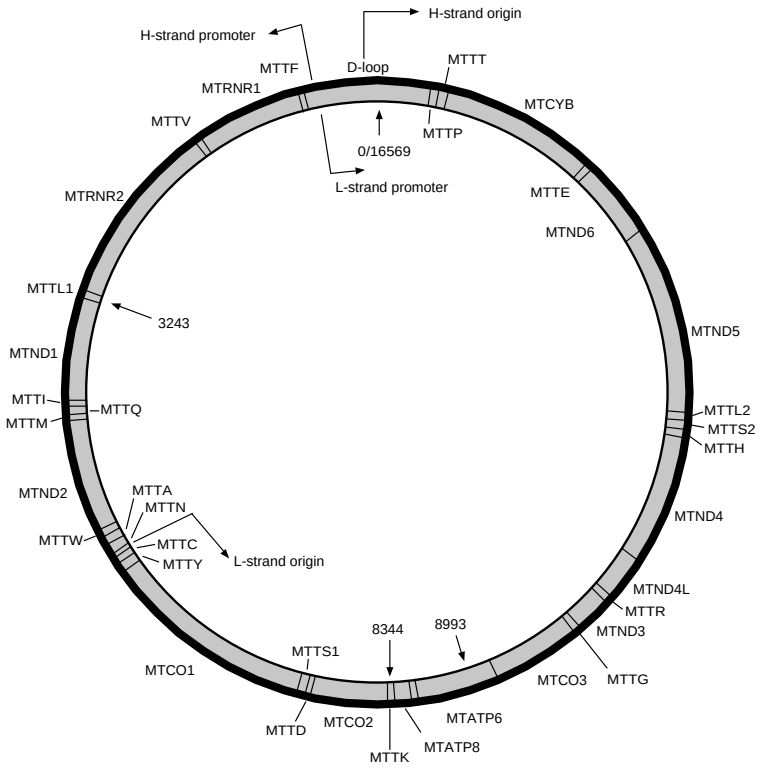


Fig. 1. **The human mitochondrial DNA genome. Genes that are transcribed from the H-strand and L-strand are shown outside and inside the circle, respectively. The locations of the common MELAS (3243A>G), MERRF (8344A>G) and NARP (8993T>C, 8993T>G) mutations are also shown.**

RNA primer which is transcribed from the L-strand. H-strand synthesis proceeds until the L-strand origin is exposed, which results in the initiation of L-strand synthesis. At the final stages of replication the primers are removed and the completed DNA circles are ligated. It has been suggested recently, however, that mammalian mitochondrial DNA replication proceeds mainly, or exclusively, by a strand-coupled mechanism and that ribonucleotides are incorporated on the L-strand during the process (Yang *et al.* 2002, Cerritelli *et al.* 2003), implying that the replication process may in fact be more complex than previously assumed.

The replication is relaxed in the sense that it can take place in any phase of cell division, but its regulation is under nuclear genetic control (Capps *et al.* 2003). The probability of replication increases in the vicinity of the nucleus, and TFAM is involved in its regulation, because it is required for the initiation of replication and also seems to control the maintenance level of mtDNA (Larsson *et al.* 1998). The relaxed replication

of mtDNA leads to clonally expanding mtDNA lineages within body organs and tissues, lineages that may differ because of somatic mutations (Khrapko *et al.* 2003). The mtDNA molecules within a multicellular organism form a large population which has evolved from a small number of founder molecules and which is also subject to other population-genetic phenomena such as mutation, drift and selection (Jenuth *et al.* 1997). It is only the population of mtDNAs in the germ line cells that contributes to the future generations of the organism, however.

2.1.4.3 Maternal inheritance

The mtDNA in mammals is maternally inherited (Giles *et al.* 1980). The number of mtDNA molecules in oocytes exceeds 100,000, whereas sperm contains only 100–1500 mtDNAs (Chen *et al.* 1995a, Manfredi *et al.* 1997, Diez-Sanchez *et al.* 2003). Paternal mitochondria enter the oocyte, but they are promptly removed, possibly by ubiquitin-dependent proteolysis (Manfredi *et al.* 1997, Sutovsky *et al.* 2000). The presence of paternal mtDNA has been observed in abnormal embryos in association with certain *in vitro* fertilization techniques (John *et al.* 2000), but no paternal mtDNA has been found in children born after intracytoplasmic sperm injection (ICSI), suggesting that paternal mtDNA is normally removed successfully even when introduced into the oocyte by this method (Marchington *et al.* 2002). It is possible, however, that the mechanism responsible for the elimination of paternal mtDNA may fail occasionally, potentially leading to maternal/paternal mtDNA mosaicism in an individual (Schwartz & Vissing 2002).

The number of mtDNA molecules within a cell experiences radical changes during the development of the oocytes, resulting in the sampling of only a small proportion of mtDNAs and subsequent replication of the sampled molecules. This phenomenon is called the mitochondrial “bottleneck” and its most likely purpose is to lower the probability of new mutations in mtDNA being allowed to pass on along the germ line (Jenuth *et al.* 1996, Jansen & de Boer 1998, Marchington *et al.* 1998). Furthermore, it is possible that the partitioning of mitochondria into daughter cells by cell division is non-random and that developing oocytes may contain mechanisms that favour the segregation of highly functional organelles (Cox & Spradling 2003). Consistent with this hypothesis, there is evidence that the segregation of mtDNA is under nuclear genetic control (Battersby *et al.* 2003). Errors in such mechanisms could result in a “leakage” of pathogenic mutations into the germ line from individuals who harbour mutations in the associated nuclear genes.

2.1.4.4 Lack of recombination

Recombination due to crossing-over between the loci (breakage and reunion during meiosis) is common in nuclear DNA and leads to the formation of new combinations of genes that are different from that of either of the parents. This mechanism accelerates the elimination of deleterious mutations and the incorporation of beneficial alleles in

organisms that reproduce sexually (Butlin 2002, Bachtrog 2003), but mtDNA seems to lack this benefit because of its clonal inheritance. Paternal mitochondria nevertheless enter the oocyte during fertilization, and it is theoretically possible that paternal mtDNA may somehow end up in the vicinity of a maternal mtDNA molecule and that the molecules could recombine by crossing-over. This would result in partial mixing of sequences between mtDNA lineages. Considerable efforts to find evidence for mtDNA recombination in humans have not been successful (Merriweather & Kaestle 1999, Morris & Lightowlers 2000, Elson *et al.* 2001, Innan & Nordborg 2002), and it is likely that recombination in human mtDNA is very rare, if it has occurred at all.

2.1.4.5 High mutation rate

The mutation rate in mtDNA is at least 10-fold relative to that in nuclear DNA, and consequently mtDNA evolves rapidly (Brown *et al.* 1979, Wallace *et al.* 1987). In addition to environmental factors common to nuclear DNA, such as radiation (Forster *et al.* 2002), mtDNA is susceptible to oxidative damage by reactive oxygen species (ROS), because it is physically near the oxidative phosphorylation system which generates these highly active molecules. It has been suggested that exposure to oxygen radicals may lead to gene-specific substitution rates (Nedbal & Flynn 1998, Williams & Hurst 2002). Furthermore, it has been assumed that mtDNA lacks effective DNA repair mechanisms, although recent studies have indicated that at least some repair activity exists but it may be less well developed than in nuclear DNA (Croteau *et al.* 1999, Kang & Hamasaki 2002, Mason *et al.* 2003). The role of possible changes in repair activity with age is not well understood (Chen *et al.* 2002, Stevnsner *et al.* 2002). Moreover, at least some mutations may be explained by context-specific transient failures of the mtDNA replication system (Malyarchuk *et al.* 2002), and it is plausible that mutations in any nuclear genes that code for components of replication or repair systems may expose some individuals to the development of mtDNA errors.

Interestingly, the mutation rate estimated from pedigree analyses is considerably higher than that estimated from phylogenetic studies (Parsons *et al.* 1997, Heyer *et al.* 2001, Howell *et al.* 2003c), suggesting that a large proportion of mutations that are observed in pedigrees are removed by selection before they can be observed as fixed polymorphisms in phylogenetic studies.

2.2 Mitochondrial DNA variation in human populations

The evolutionary relationships between species can be inferred from comparisons and phylogenetic reconstructions of DNA sequences, and mtDNA has been used extensively for this purpose because of its high mutation rate, maternal inheritance and lack of recombination. The evolutionary orders of mammals (Arnason *et al.* 2002) and the divergence times of humans and other primates (Glazko & Nei 2003) have been studied using mtDNA, for example, and similar studies have also been carried out to assess

the evolution of different human populations (Richards *et al.* 2000, Cavalli-Sforza & Feldman 2003). Molecular genetic analyses of the haploid Y chromosome have provided information on the evolution of paternal lineages and on the differences between maternal and paternal lineages (Lell & Wallace 2000, Ke *et al.* 2001, Helgason *et al.* 2003). Evolutionary mtDNA, Y-chromosomal, and autosomal analyses have generally confirmed the "out of Africa" hypothesis, and human mtDNA lineages form a single, monophyletic phylogenetic tree which is rooted in Africa (Johnson *et al.* 1983, Ingman *et al.* 2000, Jorde *et al.* 2000, Caramelli *et al.* 2003). The high mutation rate has resulted in the sequential accumulation of a large number of nucleotide substitutions in mtDNA lineages that have diverged on the same time scale as human populations have colonized the different geographical regions of the world. These mutations are often assumed to be neutral and to have evolved mostly by genetic drift, but evidence is now accumulating that selection may have had a significant role during human mtDNA evolution (Mishmar *et al.* 2003).

2.2.1 Mitochondrial DNA haplogroups

The initial analyses of human mtDNA lineages were based on sequence variation in the D-loop, in particular in the two hypervariable segments HVS-I and HVS-II (Richards *et al.* 1996, Wilkinson-Herbots *et al.* 1996), and on restriction fragment length polymorphism (RFLP) analyses of the entire genome, allowing screening of 15-20% of the mtDNA sequence for variations (Chen *et al.* 1995b). For these historical reasons, the major human mtDNA lineages have been classified into haplogroups according to the information provided by RFLPs and the sequence of the hypervariable regions in the D-loop. The principal clusters have been denoted with capital letters, with additional letters or numbers used to denote lineages within the principal clusters (Torroni *et al.* 1996). More recently, phylogenetic reconstruction of the human mtDNA using complete mtDNA sequences has allowed haplogroups to be defined by individual nucleotide substitutions, which provides the highest possible level of accuracy (Ingman *et al.* 2000, Finnilä *et al.* 2001, Maca-Meyer *et al.* 2001, Herrnstadt *et al.* 2002a, Mishmar *et al.* 2003).

2.2.1.1 Continent-specific mtDNA lineages

Analyses of the distribution of the major mtDNA haplogroups in different continents and phylogenetic analyses have facilitated reconstruction of the human mtDNA dispersal out of Africa during evolution (Figure 2). Haplogroup L, which is prominent in Africa, comprising about 3/4 of all African mtDNAs, is the oldest of the human mtDNA lineages (Wallace *et al.* 1999, Ingman *et al.* 2000), and closest to the most recent common ancestor of all human populations ("the mitochondrial Eve"), who presumably lived in central Africa approximately 130,000–200,000 years before present (YBP). Haplogroup L is further divided into several sublineages (L0, L1, L2, and L3). Two macrolineages (M and N) diverged from L, presumably in northeastern Africa or in the Middle East,

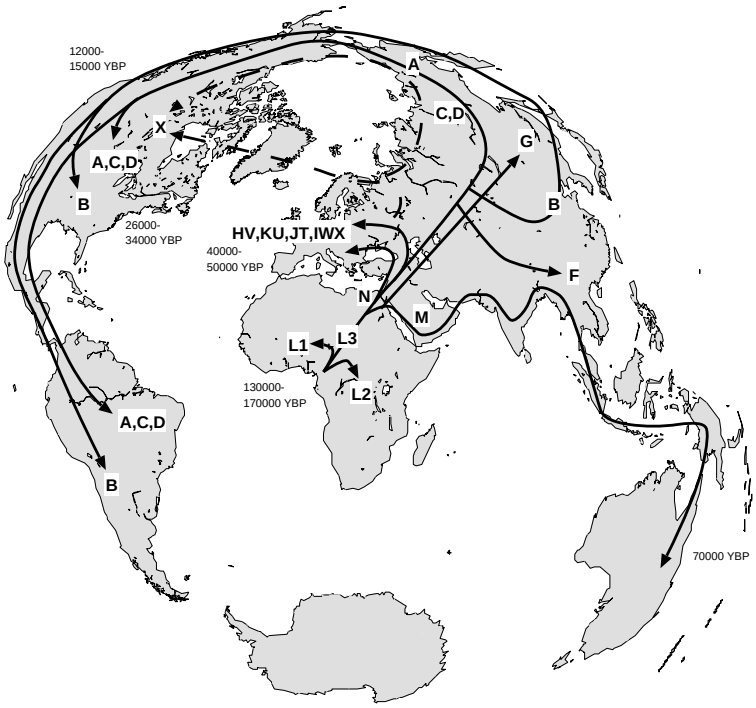


Fig. 2. Human mtDNA migrations. YBP, years before present. Capital letters indicate the major mtDNA haplogroups.

approximately 65,000 years ago, and the European haplogroups H, I, J, K, T, U, V, W, and X (Torroni *et al.* 1996) were subsequently derived primarily from macrolineage N, whereas M and N contributed equally to the radiation of mtDNA into the Asian-specific haplogroups A, C, D, G, Z, and Y. The American continent was populated from northeastern Asia through the Bering land bridge by humans with haplogroups A, C, and D (Derbeneva *et al.* 2002, Silva *et al.* 2002), whereas haplogroup B may have arrived later and via a coastal route (Starikovskaya *et al.* 1998, Mishmar *et al.* 2003). Thus, all the major continents were inhabited and the haplogroups had achieved their present geographical distribution some 35,000–15,000 YBP, although demographic shifts have also occurred since then (Bandelt *et al.* 2001, Salas *et al.* 2002).

2.2.1.2 Finnish mtDNA lineages

The oldest archaeological evidence of Finnish settlement dates back approximately 10,000 years, suggesting that Finland was settled as soon as the last of the seven

major ice ages ended, but the origin of these people is unknown. The Combed Ware culture arrived 5000–6000 years ago, followed by the Corded Ware (Battle-Axe) culture about 1000 years later, but it is not known for certain to what extent these cultural changes represented population immigrations. Nevertheless, later immigrations have been genetically fairly insignificant (Norio 2003b). Southern Finland and the coast and river valleys of Ostrobothnia were permanently inhabited by the 16th century, and the remaining parts of the country were settled soon after. The Finnish population has been subject to cultural and geographical isolation, and it has evolved from a relatively small group of founders, which has resulted in "the Finnish Disease Heritage", an enrichment of certain genetic diseases which are rare elsewhere, while other genetic diseases are rare in Finland (Norio 2003a).

Studies on Finnish mtDNA variation have revealed a western European pattern of polymorphisms (Vilkki *et al.* 1988, Sajantila *et al.* 1996, Torroni *et al.* 1996), suggesting a close relationship between the Finns and other Europeans. The frequencies of the European haplogroup clusters are similar to those in other European populations, the most common haplogroups being H, U, J, and T (Torroni *et al.* 1996, Finnilä *et al.* 2001). There are geographical differences in the frequencies of individual haplogroups, however (Torroni *et al.* 1996, Meinilä *et al.* 2001), and some ethnic groups such as the Saami appear to have been originally sharply differentiated but to have subsequently mixed with the remaining population (Sajantila *et al.* 1995, Lahermo *et al.* 1996, Meinilä *et al.* 2001). Analyses of Y chromosomal variation have provided similar results regarding the Saami (Lahermo *et al.* 1999, Raitio *et al.* 2001). The mitochondrial haplogroup Z represents an Asian contribution to the Finnish mtDNA pool and is present at a low frequency among the Finns, but the frequency of the Asian-specific Y chromosomal haplotypes is higher. One explanation for this discrepancy is that the Y chromosome and the mtDNA may have been subject to differential drifting (Meinilä *et al.* 2001).

2.3 Mitochondrial disorders

2.3.1 Point mutations in mtDNA

Nucleotide substitutions in mtDNA are maternally inherited, and many pathogenic point mutations are heteroplasmic, i.e. wild-type and mutant mtDNA molecules are both present in mitochondria and in cells in varying proportions. A clinical disease ensues when the proportion of mutant mtDNA exceeds a certain threshold at which the deleterious consequences of the mutation are no longer compensated for by the wild-type mtDNA (Rossignol *et al.* 2003). Point mutations in mtDNA occur in all genes, and their consequences may depend on the changes they impose on the products of the respective genes, the tissue distribution of the mutation, the relative reliance of tissues on ATP production, other mtDNA polymorphisms (Lamminen *et al.* 1997, Howell *et al.* 2003a), tissue-specific (Bykhovskaya *et al.* 2000) and developmentally regulated (Sasarman *et al.* 2002) nuclear-encoded factors that interact with the mtDNA gene products, and on environmental factors (Estivill *et al.* 1998). The intramitochondrial, intracellular and intercellular environments as such constitute a playground of complex interactions, and

it is therefore no surprise that the pathomechanisms of mtDNA point mutations are still not very well understood. The pathogenicity of a particular allele can be inferred from certain properties, however. First, heteroplasmy may indicate that the mutation is faced with intracellular selective pressure, although neutral mutations may sometimes show heteroplasmy as well. Second, pathogenic mutations face strong negative selection on an evolutionary time-scale, and they are therefore evolutionarily recent and not observed as fixed polymorphisms within haplogroups (Abe *et al.* 1998, Hutchin *et al.* 2001, Howell *et al.* 2003b, Torroni *et al.* 2003). Third, sequence positions that are functionally important are conserved between species, again suggesting the role of selection against deleterious alleles. Fourth, it may be possible to infer the pathogenicity of a mutation from its consequences for the properties or structure of the respective gene product, e.g. from the physicochemical properties of amino acids involved in a missense mutation. And finally, the genotype-phenotype correlations are sometimes so strong that they do not leave much doubt as to the pathogenicity of the mutation. At least 100 such deleterious mtDNA point mutations have been identified so far (DiMauro & Schon 2001), the most frequent ones being the 3243A>G "MELAS" mutation, the LHON primary mutations, the 8344A>G "MERRF" mutation, and the 8993T>G "NARP" mutation. Furthermore, there appears to be a class of slightly deleterious mutations that modify the risks of contracting certain complex diseases or traits.

2.3.1.1 Mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS)

The classic clinical features of the MELAS syndrome were described in 1984 (Pavlakis *et al.* 1984), but the molecular aetiology was not revealed until 1990, when the 3243A>G transition in the MTTL1 transfer RNA gene was discovered by two Japanese groups (Goto *et al.* 1990, Kobayashi *et al.* 1990). Several other mutations causing this syndrome have also been identified, but 3243A>G is by far the most common, occurring in approximately 80% of MELAS patients. The classic MELAS syndrome includes myopathy, encephalopathy, stroke-like episodes and lactic acidosis, but not all patients harbouring 3243A>G develop these symptoms (Hammans *et al.* 1995). Other typical symptoms include sensorineural hearing impairment (Uimonen *et al.* 2001), diabetes mellitus (Guillausseau *et al.* 2001, Choo-Kang *et al.* 2002, Maassen 2002), hypertrophic cardiomyopathy, ataxia, basal ganglia calcifications, ophthalmoplegia and neuropathy (Kärppä *et al.* 2003), which may occur alone, without any other symptoms, and some patients may be asymptomatic. Such phenotypic variability is at least partly attributed to variation in the degree of mutation heteroplasmy between individuals and between different tissues, accompanied by biochemical threshold effects. The pathogenetic mechanism of 3243A>G remains unclear. The mutation alters a conserved dihydrouridine stem in MTTL1, and this may alter the functional stability of the tRNA molecule. The clinical features, rate of protein synthesis and activity of the respiratory chain do not correlate well, however, suggesting that other mechanisms are probably also involved. The 3243A>G mutation lies within the binding site for mTERF, which promotes termination of translation at the MTRNR2/MTTL1 gene boundary, and

precursor RNA corresponding to a contiguous transcript of the MTRNR2, MTTL1, and MTND1 genes has been observed in cytoplasmic hybrid cell lines (cybrids) harbouring the 3243A>G mutation (Koga *et al.* 1995). Other mutations in MTTL1 (such as 3271T>C) which are not within the mTERF binding site can cause clinically indistinguishable phenotypes, however. This discrepancy may be explained by an associated defect in post-transcriptional modifications of the tRNA molecule (Yasukawa *et al.* 2000). Despite the similar phenotype, the 3243A>G seems to cause an aminoacylation defect which could interfere with protein synthesis (Chomyn *et al.* 2000), while 3271T>C does not affect aminoacylation efficiency (Park *et al.* 2003). The 3243A>G mutation is clinically one of the most common pathogenic mtDNA mutations, but population-based studies regarding its epidemiology have been lacking, with the exception of the North East of England where the minimum estimate for its prevalence is 0.95/100,000 (Chinnery *et al.* 2000). Furthermore, it is not associated with any particular mtDNA haplogroup, and haplotype analyses suggest that in most instances the mutation has arisen recently, implying that it is subject to strong negative selection (Torroni *et al.* 2003). An apparent *de novo* occurrence of the mutation has also been observed (Maassen *et al.* 2002).

2.3.1.2 Leber's hereditary optic neuropathy (LHON)

LHON is an enigmatic disorder that manifests itself in mid-life in the form of acute or subacute central vision loss, leading to central scotoma and blindness (Man *et al.* 2002). It is relatively common with a minimum point prevalence of 3.22/100,000 in the North East of England (Man *et al.* 2003). The mean age at onset varies from 27 to 34 years, and the eyes can be affected simultaneously or sequentially, with an average interval of two months. The severity and final outcome of the visual impairment varies considerably. Ancillary symptoms are usually mild, but a severe neurological disease is also possible. The risk of blindness is considerably higher in males than in females. At least 18 allelic variants of mtDNA have been associated with LHON, most of these being in the genes coding for Complex I subunits, and these mutations are assumed to act autonomously or in association with each other to cause the disease. Four mutations, 3460G>A, 11778G>A, 14459G>A, and 14484T>C appear to have the highest pathogenic potential, and are therefore called "primary" mutations, whereas the remaining "secondary" mutations are generally assumed to interact with the primary ones to increase the probability of clinical disease. The biochemical defect in LHON is probably related to a deficiency in Complex I caused by structural changes in the subunits (Qi *et al.* 2003), and the subsequent pathomechanism may involve increased ROS production (Wong *et al.* 2002), apoptosis (Danielson *et al.* 2002, Ghelli *et al.* 2003), or both. The highly tissue-specific phenotype might be explained by hypothetical neuron-specific alterations in Complex I structure (Wong *et al.* 2002). No attempt has apparently yet been made to examine whether the male bias in penetrance can be explained by the finding that the expression of antioxidant genes is higher in females than in males and consequently the amount of oxidative damage is lower (Borras *et al.* 2003).

2.3.1.3 MERRF

MERRF (myoclonic epilepsy associated with ragged-red fibres) can be produced by mutations in at least two tRNA genes, MTTK (Shoffner *et al.* 1990) and MTTL1 (Moraes *et al.* 1993). The maternally inherited heteroplasmic 8344A>G transition, which accounts for 80–90% of cases (Shoffner & Wallace 1992), produces multiple deficiencies in the enzyme complexes of the respiratory chain consistent with a general defect in translation (Bindoff *et al.* 1991). Its population prevalence has been estimated to be approximately one-fourth of the prevalence of the 3243A>G mutation or the various LHON mutations (Chinnery *et al.* 2000, Remes *et al.* 2003) The biochemical defect is often segmental, suggesting a non-random distribution of mutant and wild-type mtDNAs within muscle cells (Matsuoka *et al.* 1991).

2.3.1.4 NARP

NARP (neuropathy, ataxia and retinitis pigmentosa) and some cases of the Leigh syndrome (subacute necrotizing encephalomyopathy) result from missense 8993T>C or 8993T>G mutations in the MTATP6 gene (Holt *et al.* 1990), which introduce an incorrect amino acid into a highly conserved position in the hydrophobic sequence of the ATPase 6 subunit. This apparently induces a structural defect in ATP synthase that causes impairment of ATP synthesis and increased ROS production (Carelli *et al.* 2002), which leads to the clinical symptoms.

2.3.1.5 Interactions and slightly deleterious mutations

Distinct mtDNA mutations may act synergistically to modulate disease expression (Lamminen *et al.* 1997, Rahman *et al.* 1999, Brown *et al.* 2002, Howell *et al.* 2002, Sudoyo *et al.* 2002). Furthermore, there are also interactions between nuclear genes and environmental factors, which modulate the severity of hearing loss that is associated with the mitochondrial 1555A>G mutation, for example (Abe *et al.* 2001, Guan *et al.* 2001). MtDNA variants are thought to be associated with susceptibility to several disorders, including hypertension (Watson *et al.* 2001), Parkinson's disease (Tanaka 2002, Walt *et al.* 2003), Alzheimer's disease (Chagnon *et al.* 1999, Carrieri *et al.* 2001, Cottrell *et al.* 2001), fasting insulin levels (Poulton *et al.* 2002) or glucose (Kim *et al.* 2002), ageing (Tanaka 2002), sudden infant death syndrome (Opdal *et al.* 2002), bipolar disorder (Kato *et al.* 2000) and unsuccessful outcome of pregnancy at a high altitude (Myres *et al.* 2000). While the nature of such associations is not yet clear, it is possible that mtDNA variants may modify susceptibility to disorders in which the mitochondria are somehow involved in the pathogenesis, e.g. by apoptosis or calcium regulation.

2.3.2 Rearrangements of mtDNA

Numerous deletions, duplications and insertions have been observed in mtDNA, and these cause various syndromes which generally affect post-mitotic tissues such as the central nervous system, muscles, and heart. The most frequent deletion is the "Common deletion", which spans 4977 bps between nucleotide positions 8469–13466. The breakpoints are mostly associated with repeat sequences, suggesting that the rearrangements probably result from transient failures of the mtDNA replication system (Rocher *et al.* 2002). Large deletions generally remove at least one tRNA gene, and are therefore likely to cause a translational defect and dysfunction of the respiratory chain (Gellerich *et al.* 2002). The most important clinical syndromes caused by large-scale mtDNA deletions are Kearns-Sayre syndrome (KSS), Pearson syndrome and chronic progressive external ophthalmoplegia (CPEO). KSS is characterized by pigmentary retinopathy, progressive external ophthalmoplegia and onset before the age of 20 years. The Pearson bone marrow-pancreas syndrome occurs in early infancy and is characterized by anaemia, pancytopaenia and exocrine insufficiency of the pancreas. Surviving patients may later develop features of KSS. The clinical features of CPEO include ptosis, ophthalmoplegia, exercise intolerance and proximal muscle wasting and weakness. Single mtDNA deletions are almost invariably sporadic, but it is possible that they may also be transmitted through the germ line in rare cases (Shanske *et al.* 2002). Multiple deletions have been observed in several syndromes, and these appear to result from mutations in nuclear genes (Vissing *et al.* 2002, Agostino *et al.* 2003).

2.3.3 Mutations in nuclear genes

The great majority of components of the mitochondrial respiratory chain complexes, proteins involved in the assembly and the maintenance of these complexes, the mitochondrial protein import machinery and the mtDNA replication and error-correction systems are encoded by nuclear DNA, and errors in these genes can cause diseases. In fact the majority of cases with a respiratory chain deficiency may result from errors in nuclear-encoded subunits of the OXPHOS system (Triepels *et al.* 2001). Isolated Complex I deficiency with recessive inheritance can result from a mutation in several genes, including NDUFV1 (Schuelke *et al.* 1999), NDUFV2 (Benit *et al.* 2003), NDUFS1 (Benit *et al.* 2001), NDUFS2 (Loeffen *et al.* 2001), NDUFS4 (van den Heuvel *et al.* 1998), NDUFS7 (Smeitink & van den Heuvel 1999) and NDUFS8 (Loeffen *et al.* 1998). The Complex II subunits SDHA (Bourgeron *et al.* 1995), SDHB (Astuti *et al.* 2001), SDHC (Niemann & Muller 2000), and SDHD (Baysal *et al.* 2000) have also been found to carry pathogenic mutations, and mutations in BCS1L, which is involved in the assembly of Complex III, have caused Complex III deficiency (de Lonlay *et al.* 2001). Complex IV deficiency can be caused by mutations in the SURF1 (Zhu *et al.* 1998), SCO2 (Papadopoulou *et al.* 1999), COX10 (Valnot *et al.* 2000b) and SCO1 (Valnot *et al.* 2000a) genes involved in the assembly of this enzyme. Gene defects in ANT1, Twinkle and POLG (Goethem *et al.* 2001, Spelbrink *et al.* 2001, Agostino *et al.* 2003) have been identified in patients with autosomal recessive or sporadic progressive

ophthalmoplegia in association with multiple mtDNA deletions. Multiple deletions may also be associated with MNGIE (myoneurogastrointestinal encephalopathy), which can be caused by mutations in ECGF1 (Nishino *et al.* 1999) or in other genes (Vissing *et al.* 2002). The mitochondrial DNA depletion syndrome can result from mutations in DGUOK or TK2, for example (Elpeleg *et al.* 2002). It is likely that only a small minority of nuclear genes associated with mitochondrial diseases have been identified, and the products of such genes may also be subject to deterioration by other processes within the mitochondrial matrix (Graziewicz *et al.* 2002), implying that they may be involved in pathogenetic interactions with mitochondrial and other nuclear genes. On the other hand, the nuclear DNA may also contain genes that can suppress defects in mtDNA genes (Rinaldi *et al.* 2003).

2.4 Selection against mutations in mtDNA

The central role of mitochondria in cellular energy metabolism and the role of mtDNA mutations in human disease clearly imply that natural selection has a significant role in mtDNA evolution. The stringency of the varying structural and functional constraints is most likely the main determinant of the intensity of selection against particular mtDNA mutations (Tourasse & Li 2000). For example, the amino acid substitutions in the cytosolic portion of primate MTCO2 appear to have been more deleterious than those in the transmembrane portion (Templeton 1996), although transmembrane regions may generally be more conserved than the nontransmembraneal regions (Naylor *et al.* 1995, Tourasse & Li 2000). Furthermore, there are differences between genes, since nonsynonymous substitution rates have been shown to vary among the protein-coding genes of mtDNA, whereas synonymous rates are more uniform (Pesole *et al.* 1999). Functional and structural constraints are certainly also important in the rRNA and tRNA genes of mtDNA. A proportion of the selective constraints are also likely to arise from interactions between the mtDNA-encoded gene products and the nuclear-encoded components of the mitochondria, leading to natural selection which will favour evolutionary co-adaptation of interacting molecules coded by these two distinct genomes (Blier *et al.* 2001). Environmental factors may also have imposed selective constraints on mtDNA during human evolution (Mishmar *et al.* 2003), and therefore evolutionary constraints may also have implications for susceptibility to disease.

2.4.1 The neutral theory

The neutral mutation-random drift hypothesis aroused controversy in 1968 when it was first introduced (Kimura 1968), but it has since become a cornerstone of molecular evolution. The neutral theory states that the majority of polymorphisms observed within and between species are the result of random drift or neutral mutations rather than of natural selection. Deleterious mutations are assumed to occur, but they are promptly removed by purifying selection. A variation of the neutral theory, the *nearly*

neutral theory, differs slightly in certain assumptions regarding the quantity and fate of slightly deleterious alleles (Ohta 2002). Both theories assume that the observed patterns of molecular variation result from drift, mutation and/or purifying selection, whereas positive selection or adaptive evolution is nonneutral and not consistent with these neutral theories. Positive selection occurs when a selectively advantageous mutation is segregating in the population, and this may also result in a selective sweep, for example, which refers to the elimination of variation at linked neutral sites as the frequency of a positively selected allele increases in the population (Bamshad & Wooding 2003).

An important prediction of the neutral theory is that evolutionarily recent mutations, i.e. mutations in the peripheral parts of the phylogeny, may include slightly deleterious alleles which have not yet been removed by selection. This could also imply that evolutionarily recent mutations in human mtDNA may have more severe structural or functional consequences for the products of the respective genes than older mutations, e.g. in terms of the physicochemical properties of the amino acids involved in nonsynonymous mutations, but there are no reports in which this question is systematically addressed.

The neutral theory has during the last decade become a null model against which specific occurrences of selection can be detected. The possibility of identifying genomic regions that have been subject to selection is important, because it may allow the identification of disease-causing mutations, for example. Furthermore, assessment of neutrality provides a tool for developing hypotheses regarding function from genomic data (Nielsen 2001). Statistical methods for detecting the signature of selection from population-based data can be assigned to two main categories: comparisons of observed differences in performance (genetic fitness) between genotypes, and comparisons of observed patterns of nucleotide sequence variation with those expected under the neutral model.

2.4.2 Relative fitness

The intensity of selection against a particular genotype is measured in terms of fitness, which indicates the degree of adaptation of the genotype to its environment, and consequently the probability that the genotype will pass on to future generations (Bamshad & Wooding 2003). Deleterious mutations tend to reduce the fitness and reproductive capacity of their carriers. Fitness comprises three components: fertility, survival and generation time (Morton 1982), and the relative fitness of a genotype is usually measured as its fitness relative to that of another genotype in the same gene system. Comparison of fertility, survival and generation time between mutation carriers and the remaining population therefore provides a direct measurement of the intensity of selection against mutation carriers (Cavalli-Sforza & Bodmer 1971). A convenient summary measure for the three components of fitness is the net reproduction rate (*NRR*), which is a measure of the extent to which one female generation is reproducing itself. It can be interpreted as the average number of daughters a woman would have during her reproductive years, assuming fertility and mortality at given rates (Newell 1994), and the relative reproductive capacity, or fitness, of mutation carriers is given directly by the ratio of the *NRR* of mutation carriers to that of the general population, assuming that the

generation time does not differ between these (Cavalli-Sforza & Bodmer 1971). This method is particularly suitable for mtDNA, because its evolution proceeds exclusively by mother-daughter transmissions, and because *NRR*s are usually published in official population statistics. It is surprising, therefore, that there are no previous reports in which the relative fitness of carriers of pathogenic mtDNA mutations is estimated from pedigree data. This would allow inferences to be made on the degree and nature of natural selection that is currently present against this class of highly deleterious mtDNA mutations.

2.4.3 Nucleotide polymorphism indices

Natural selection produces patterns of sequence polymorphisms that are different from those observed when selection is not present, and therefore nucleotide polymorphism data can be used to identify loci that are subject to selection. Several descriptive statistics are commonly used to summarize nucleotide polymorphism data and to compare data sets. The number of polymorphic or segregating sites, denoted by S, is calculated simply by identifying all sites that show variation in the sample. Similarly, it is possible to infer the total number of mutations in the sample (η), which is an estimate of the minimum number of mutations required to explain the observed variation, as it does not include any correction for multiple substitutions at a site (Fu & Li 1993). The total number of singleton mutations (η_s) is calculated by identifying all those alleles that are present in only one sequence in the sample (Fu & Li 1993). The diversity between sequences can also be measured in terms of pairwise sequence comparisons. The average number of pairwise nucleotide differences is calculated as

$$k = \frac{1}{[n(n-1)/2]} \sum_{i<j} k_{ij}, \tag{2.1}$$

where k_{ij} is the number of nucleotide differences between the i-th and j-th sequences and n is the number of sequences (Tajima 1983). These statistics can be used to derive estimates for the population mutation parameter θ, which is $4N\mu$ for nuclear DNA, $3N\mu$ for X-chromosomal DNA and $2N\mu$ for mitochondrial DNA, where N is the effective population size and μ is the neutral mutation rate per generation (Fu 1997). The population mutation parameter for the number of segregating sites (S) and its variance are given by

$$\theta_S = \frac{S}{a_n}, \tag{2.2}$$

$$Var(\theta_S) = \frac{{a_n}^2 S + {b_n}^2 S^2}{{a_n}^2({a_n}^2 + b_n)}, \tag{2.3}$$

where

$$a_n = \sum_{i=1}^{n-1} \frac{1}{i}, \tag{2.4}$$

$$b_n = \sum_{i=1}^{n-1} \frac{1}{i^2}, \tag{2.5}$$

whereas the population mutation parameter for pairwise differences and its variance are estimated as

$$\theta_\pi = k, \tag{2.6}$$

$$Var(\theta_\pi) = \frac{(n+1)\theta_\pi}{3(n-1)} + \frac{2(n^2+n+3){\theta_\pi}^2}{9n(n-1)}. \tag{2.7}$$

Correspondingly, the formula for the population mutation parameter for the total number of mutations (η) is

$$\theta_\eta = \frac{\eta}{a_n} \tag{2.8}$$

and that for singleton mutations (η_s) is

$$\theta_{\eta s} = \frac{n-1}{n}\eta_s \tag{2.9}$$

(Tajima 1983, Fu & Li 1993, Simonsen *et al.* 1995). Each θ can additionally be standardized by the length of the sequence analysed (Li 1997). Nucleotide diversity per site, π (θ_π/site), is the frequency with which any two sequences in the sample differ at a site, and is most often used for direct comparisons between sequence sets and classes of nucleotide sites, because it does not depend on the sample size or the length of the sequence analysed (Nei 1987).

Sequence diversity in mtDNA is higher in African haplogroups than in other continent-specific haplogroups, and Africa encompasses the greatest diversity of all continents, consistent with the assumption that the African haplogroups and populations are the oldest of all and have therefore accumulated such a high degree of sequence variation (Wallace *et al.* 1999). Besides such comparisons between populations, the nucleotide polymorphism indices also allow inferences to be made regarding selection against different classes of sites. For example, the neutral theory states that in the subset of third codon positions containing fourfold degenerate sites the fate of almost all mutations will be governed by neutral mutation-random drift, because they do not manifest themselves as amino acid changes and therefore have no effect on fitness. On the other hand, purifying selection presumably weeds out a significant proportion of mutations in the nondegenerate second codon positions, in which all mutations result in amino acid replacements, and therefore the overall fraction of neutral mutations and the average sequence diversity would be smaller at nondegenerate sites than at fourfold degenerate ones (Gerber *et al.* 2001). Analyses of nucleotide diversity therefore facilitate inferences on the relative degree of selection between genes, genome regions or classes of sites within sequences that share a common history. The mtDNA genes are most likely subject to different selective constraints, but the relative rank order of genes or classes of sites should be similar in all phylogenetic lineages, unless there are lineage-specific factors that impose different constraints on different lineages (Rand 2001).

2.4.4 Neutrality tests

Under the assumptions of the standard neutral equilibrium model, the different θ estimators should yield similar values when the mutations are neutral (Simonsen *et al.*

1995). However, θ_S and $\theta_{\eta s}$ are affected by the presence of low-frequency alleles in the sample, whereas these have little impact on θ_π, leading to differences between the estimates when selection is present (Fu & Li 1993). Such differences form the basis of various statistical tests of neutrality, including Tajima's *D* (Tajima 1989), which is based on the difference between θ_π and θ_S and its variance,

$$D = \frac{\theta_\pi - \theta_S}{\sqrt{Var(\theta_\pi - \theta_S)}}, \tag{2.10}$$

Fu and Li's $F*$ (Fu & Li 1993), which is based on the difference between θ_π and $\theta_{\eta s}$ and its variance,

$$F* = \frac{\theta_\pi - \theta_{\eta s}}{\sqrt{Var(\theta_\pi - \theta_{\eta s})}}, \tag{2.11}$$

and Fu's Fs (Fu 1997), which is based on the probability of the observed number of haplotypes or more being observed under conditions of neutrality,

$$S' = \sum_{m \geq k_0} \frac{|S_n^m| {\theta_\pi}^m}{\theta_\pi(\theta_\pi - 1)\cdots(\theta_\pi - n + 1)}, \tag{2.12}$$

$$Fs = \ln \frac{S'}{1 - S'}, \tag{2.13}$$

where k_0 is the number of haplotypes and S_n^m is $s(n,m)$, the Stirling number of the first kind indicating the number of permutations of *n* elements that contain exactly *m* cycles.

The existence of many deleterious mutations in a population will produce an excess of rare alleles as compared with that expected under conditions of neutrality, and this will result in negative values for these tests, whereas positive values are associated with a relative lack of low-frequency alleles. Analyses of primate or human mtDNAs using neutrality tests have mostly been based on single genes and the results have varied, as significant departures from neutrality have been observed in some studies but not in others (Gerber *et al.* 2001). Neutrality tests are sensitive to several assumptions, including demographic factors such as population history, and violation of these assumptions can produce patterns that may be misinterpreted as evidence of selection (Wayne & Simonsen 1998). However, as neutrality tests are measures of the allele frequency spectrum (Przeworski *et al.* 2000), they could also be used as summary statistics to identify genome regions that show patterns of nucleotide substitution that are not consistent with the average pattern for the genome (Nielsen 2001, Rand 2001). Mitochondrial DNA should be a particularly successful candidate for such analyses because of its lack of recombination and uniparental clonal inheritance.

2.4.5 Silent and replacement mutations

2.4.5.1 Nonsynonymous/synonymous rate ratio

Synonymous mutations in protein-encoding genes are mostly free of selection, whereas a large proportion of nonsynonymous mutations are likely to be deleterious and selected

against (Majewski & Ott 2003). The nonsynonymous substitution rate is therefore generally lower than the synonymous substitution rate. These rates may be compared by calculating $\omega = d_N/d_S$, where d_N is the number of nonsynonymous substitutions per nonsynonymous sites and d_S is the number of synonymous substitutions per synonymous sites. ω indicates the relative level of selection against nonsynonymous mutations, as compared with the synonymous variation. A value of $\omega > 1$ implies that nonsynonymous mutations are fixed with higher probability than neutral mutations, because of positive selection.

2.4.5.2 McDonald-Kreitman test

The McDonald-Kreitman (MK) test (McDonald & Kreitman 1991) analyses whether ω is similar within and between species, as is expected when a gene evolves neutrally. This test has been used extensively to analyse mtDNA sequence variation, and an excess of amino acid polymorphisms within species as compared with that between species has often been found (Nachman 1998, Rand & Kann 1998, Weinreich & Rand 2000, Gerber *et al.* 2001). This finding has been interpreted as evidence for the nearly neutral theory, in which weakly selected amino acid variants are observed within populations but removed by selection before they have the opportunity to become fixed. The MK test has been criticized, however (Graur & Li 1991, Maynard Smith 1994, Wayne & Simonsen 1998, Eyre-Walker 2002), and a significant result of a MK test may have several interpretations (Gerber *et al.* 2001), but the finding of higher ω within species than between species also seems to persist in at least some analyses that attempt to account for the specific problems of the MK test (Hasegawa *et al.* 1998).

2.4.5.3 d_N/d_S distributions between lineages

Distributions of pairwise comparisons of ω between sequences can be compared between lineages to make inferences regarding the relative selective constraints within each lineage. This method has recently been used to test the hypothesis that the evolution of major continent-specific human mtDNA haplogroups has been shaped by environmental factors, and the authors found evidence for different selective constraints for the protein-coding genes of mtDNA in different haplogroups that were selected according to their geographical origin, suggesting that climate may have had a role in the evolution of the major human mtDNA lineages (Mishmar *et al.* 2003). The authors did not explicitly consider, however, the possibility that similar differences might exist also between lineages that have not been selected according to their origin, and furthermore, the analysis included only a subset of human mtDNA sequences that were available in sequence databases.

2.4.6 Other methods

The Hudson-Kreitman-Aguade (HKA) test (Hudson *et al.* 1987) compares the variability within and between species in two or more loci simultaneously. The requirement of free recombination limits its use to comparisons between mtDNA and nuclear loci (Kreitman 2000). Selection may also be detected from differences in the estimated ages of nonsynonymous and synonymous mutations (Nielsen & Weinreich 1999), for instance, or from the allelic distribution of nonsynonymous and synonymous polymorphisms (Akashi 1999). Furthermore, it is possible to detect positive selection or recombination using a method which tests whether there are regions within the analysed sequence that do not fit with a single phylogenetic topology and nucleotide substitution process estimated for the entire sequence (Grassly & Holmes 1997).

3 Aims of the present research

Mutations in mtDNA are important causes of disease, but data concerning their epidemiology have been limited. Pathogenic mutations in mtDNA are observed around the world, but they are rare in all populations, suggesting strong negative selection against them. No attempt has been made, however, to estimate the selection against any pathogenic mutation in mtDNA empirically from pedigree data. A large number of polymorphisms have accumulated into continent-specific mtDNA lineages during human evolution, but it is not known whether the observed nonsynonymous mutations are completely neutral or whether some of them are slightly deleterious. The genes in mtDNA are subject to different selective constraints and are therefore expected to evolve at different rates, but the rank order of these rates should be the same in all lineages of the phylogeny. Evidence is accumulating, however, that some mtDNA regions may have evolved at different rates in different lineages. Moreover, the risks of contracting certain complex diseases may be haplogroup-specific. These findings suggest that selective constraints against mtDNA mutations may be lineage-specific, possibly due to mitochondrial-mitochondrial or mitochondrial-nuclear interactions in the corresponding lineages.

The specific aims of this research were:

1. to estimate the prevalence of 3243A>G, one of the most deleterious mtDNA point mutations, in the population of Northern Ostrobothnia
2. to estimate the host-level selection against individuals harbouring the 3243A>G mutation
3. to identify and characterize all sequence variations in published mtDNA sequences
4. to develop a test for the hypothesis that phylogenetically recent nonsynonymous mutations are more deleterious than ancient mutations, and to apply the test to the available mtDNA sequences, and
5. to compare polymorphism indices and patterns of polymorphism between haplogroups and mtDNA regions to find possible evidence for lineage-specific selective influences

4 Subjects and methods

4.1 Genetic epidemiology of the 3243A>G mutation (I, II)

4.1.1 Ascertainment of patients (I)

The phenotypic variability of the disease caused by the 3243A>G mutation is high. In addition to the classic MELAS syndrome (Pavlakis *et al.* 1984), clinical evaluation of patients carrying this mutation has revealed a wide array of milder phenotypes, such as hypertrophic cardiomyopathy, ataxia, basal ganglia calcifications, ophthalmoplegia, sensorineural hearing impairment and diabetes mellitus (Kadowaki *et al.* 1994, Mariotti *et al.* 1995, Morgan-Hughes *et al.* 1995), many of which are also common in the general population. A minimum estimate of the prevalence of the 3243A>G mutation was therefore obtained by identifying as many patients as possible from the adult population of Northern Ostrobothnia (total population 353,895 on December 31, 1994).

Regional and municipal health-care registers were used to identify patients with phenotypes commonly associated with mtDNA disorders (Johns 1995, Shoffner 1996). Patients >20 years of age and with diabetes mellitus, hearing loss, epilepsy, brain infarct, ophthalmoplegia, basal ganglia calcification, white-matter disease, hypertrophic or dilated cardiomyopathy, or ataxia, were first identified, and subgroups were then ascertained by reference to clinical criteria and family history data and analysed for the presence of the 3243A>G mutation.

The number of patients in the first three diagnostic groups was high, and therefore a family history of any of the three disorders in first or second-degree maternal relatives was used as an additional inclusion criterion. The family history was collected by means of a questionnaire. The author was responsible for the preparation and mailing of these questionnaires. In the two largest patient groups, epilepsy and diabetes mellitus, the author was also responsible for patient identification and for collection, management and statistical analyses of the data, and contributed to similar tasks in the third group, patients with hearing loss.

4.1.2 Analysis of families with the 3243A>G mutation (I)

A blood sample was requested from each patient fulfilling the selection criteria. DNA was purified from these samples and screened for the 3243A>G mutation by restriction-fragment analysis using the restriction enzyme *Apa*I (Kobayashi *et al.* 1990). After the mutation had been found, a muscle biopsy sample or buccal-epithelial cell sample was also taken, and adult relatives were requested to visit the outpatient clinic, where similar samples were obtained from those who volunteered.

The pedigrees of the probands with the mutation were compiled from official population records, noting all first and second-degree maternal relatives and a maternal ancestor born during the 19th century. In addition to this genealogical analysis, mitochondrial DNA haplogroups were determined by a RFLP analysis using 14 restriction enzymes (Torroni *et al.* 1996), and the 22 tRNA genes and the D-loop were sequenced to determine whether the mtDNA genotypes were consistent with the assumption that the 3243A>G mutation was identical by descent in some or all of the families.

The prevalence of the 3243A>G mutation was estimated as $N/245,201$, where N was the number of verified mutation carriers and the denominator was the total adult (>20 years) population of Northern Ostrobothnia on December 31, 1994. Furthermore, a similar estimate was obtained by assuming that the unexamined first-degree maternal relatives of the verified mutation carriers also harboured the mutation.

4.1.3 Relative fitness of 3243A>G carriers (II)

The relative reproductive capacity, or genetic fitness, of mutation carriers is given by NRR_m/NRR_p, where NRR_m is the net reproduction rate of the mutation carriers and NRR_p is the net reproduction rate of women in the general population (Cavalli-Sforza & Bodmer 1971). NRR_m and NRR_p were estimated retrospectively for the population of the province of Oulu during the period 1 January 1973 to 31 December 1996. Annual net reproduction rates for the general population (NRR_p) were obtained from the official population statistics, and that for women carrying the 3243A>G mutation was estimated on the basis of 16 families, 5 of which had been found after the population-based screening for the mutation (I). Sisters of the probands, female probands, mothers of the probands and maternal aunts of the probands were considered eligible for the analysis. This resulted in 92 individuals, 32 of which had been examined with respect to the 3243A>G mutation. The number of reproductive years, defined as the age between 15 and 49 and coinciding with the years 1973–1996, was recorded for each woman harbouring the mutation. These data, the number and date of live births during that period, and the person-years method (Kahn & Sempos 1989) were used to estimate age-specific birth rates at five-year intervals. The probability of survival to the midpoint of each 5-year interval was estimated by Kaplan-Meier survival analysis for the female mutation carriers in the 16 families. NRR_m was then calculated by multiplying the age-specific birth rate for each 5-year interval by the probability of survival to the midpoint of the interval and summing the products over all the intervals (Cavalli-Sforza & Bodmer 1971). The mean of the age-specific fertility distribution ($\overline{m}$), a commonly used measure of the timing of fertility (Newell 1994), was

calculated for the mutation carriers and for the general population and compared using a one-way analysis of variance (ANOVA) to exclude the possibility that NRR_m and NRR_p differed simply because the generation times for these two groups were different. The LIFETEST and GLM procedures of SAS 6.12 (SAS Institute 1997) were used for the Kaplan-Meier survival analysis and ANOVA, respectively.

4.2 Mitochondrial DNA sequence sets (III and IV)

The analysis of variation in currently available mtDNA sequences requires the ability to align and annotate a large number of sequences that have been obtained from different sources and therefore have varying degrees of pre-existing annotations. The absolute sequence positions of features vary due to the presence of insertions, deletions and highly variable repeat length polymorphisms. Moreover, the identification of codon positions, for example, is problematic because the mtDNA contains overlapping gene regions with different reading frames for different genes. No existing software packages were found which would have facilitated the identification and extraction of particular classes of mutations or subsequences of particular genes or functional classes of sites, as required by the analyses presented here. Therefore, a custom database system for such mtDNA sequence analyses was developed and implemented using MySQL (`http://www.mysql.com`), which is an open-source relational database system based on the standardized SQL query language. A relational database stores data in separate tables that are linked by defined relations, making it possible to combine data from several tables on request. The database was constructed to contain tables for the reference sequence, feature locations, raw sequence data, sequence identifiers and haplogroups, mutations identified, and physicochemical properties of the nonsynonymous mutations, and these were managed using the SQL language, Perl/DBI constructs and a web-based interface (Figure 3). All the variations in published mtDNA sequences were identified and stored in the database, including the sequences themselves.

Human mitochondrial DNA sequences that were available in public databases were downloaded from the Internet, resulting in a data set of 847 sequences with varying degrees of annotation (III, table 1). Three sequences (GenBank NC_001807.4, J01415.1, V00662.1) were excluded as being historical or current reference sequences, and three additional sequences (GenBank E27669.1–E27671.1) were excluded because they were patent-associated and demonstrated variation only on certain positions, being otherwise similar to V00662.1, including its errors, which suggests that they were not genuine human sequences. One of the sequences was from the CCL2 HeLa lineage, and it was excluded because of an unusually high rate of divergence (Herrnstadt *et al.* 2002b). This resulted in 840 complete coding region sequences, which were compared with the reference sequence using the diffseq utility of the EMBOSS software package (Rice *et al.* 2000). Diffseq takes two overlapping, nearly identical sequences and reports the differences between them. The reference was the Mitomap reference sequence (available at `http://www.mitomap.org`), a slightly modified version of the 2001 Revised Cambridge Reference Sequence (Andrews *et al.* 1999).

The differences reported by diffseq were converted to lists of standardized mutation

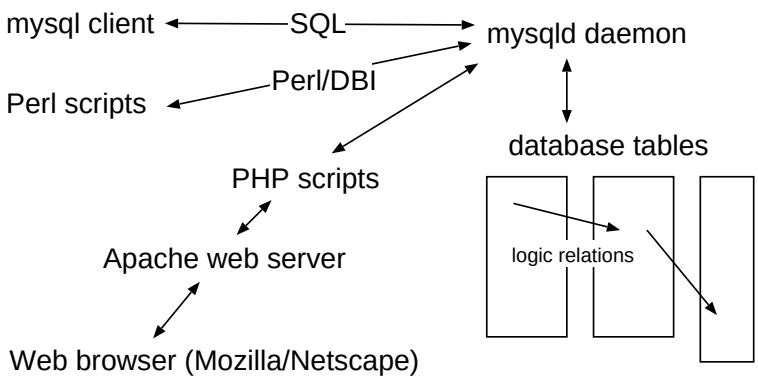


Fig. 3. **Diagram of the SQL database system.**

names (den Dunnen & Antonarakis 2001), which were then used, with the reference sequence, to populate the SQL database with the nucleotide data of the sequences. This novel approach allowed the position of each nucleotide in each sequence to be determined with respect to the reference sequence, so that the sequence data could be aligned, handled and extracted using standard SQL and Perl DBI programming language constructs. The stored sequences were compared with the original ones to ensure that the algorithms for storing the data had run without errors. The sequences were annotated with features according to the Mitomap function locations (`http://www.mitomap.org`).

4.2.1 Characterization of nonsynonymous mutations (III)

Amino acid translations of protein-coding genes were obtained from the database using the Bio::PrimarySeqI interface of Bioperl (Stajich *et al.* 2002, `http://www.bioperl.org`) and nonsynonymous changes were subsequently identified. Observed nonsense mutations were verified manually, and the amino acid translations of 325 mutations were compared with those found in Mitomap, all the observed differences being found to be errors in Mitomap. The amino acids involved in the nonsynonymous mutations were characterized in terms of six physicochemical properties, namely polarity (Grantham 1974), size (Grantham 1974), isoelectric point (Alff-Steinberger 1969), aliphatic and aromatic nature, and hydropathy (Kyte & Doolittle 1982), and the amino acid replacements were assigned to categories according to changes in these physicochemical properties. Furthermore, each replacement was defined as conservative or nonconservative according to the BLOSUM62 matrix used for sequence comparisons (Henikoff & Henikoff 1992), nonconservative replacements having a negative value in the matrix (Cargill *et al.* 1999).

The distribution of mutations within genes was assessed by identifying hydrophobic

and hydrophilic regions of genes, defined by comparison of the average hydropathy of each 19-amino acid segment with the mean for all segments in the respective gene. The Kyte-Doolittle hydropathy index for 19 neighbouring amino acids was calculated for each amino acid position using the pepinfo utility of the EMBOSS package. This segment size was used because it has been proposed as achieving good sensitivity in the identification of transmembrane regions (Kyte & Doolittle 1982).

4.2.2 Reduced-median networks of nonsynonymous mutations (III)

Six reduced-median networks (Bandelt *et al.* 1995) were constructed from the nonsynonymous mutations in the 840 sequences. All the coding region variations (Finnilä *et al.* 2001, Herrnstadt *et al.* 2002a) were used to assign each sequence to one of the six networks. The sequences were converted to a binary data matrix (Bandelt *et al.* 1995), and reduced-median networks were constructed from the binary data using Network 2.1 (available at `http://fluxus-engineering.com`). All binary characters were weighted equally, and the default reduction threshold $r = 2$ (Bandelt *et al.* 2000) was used in the analysis.

The physicochemical properties of the amino acid replacements were compared in terms of their position in the network in order to assess the hypothesis of selection against mildly deleterious replacements, which maintains that the mutations in the periphery of the phylogenetic network should include more deleterious mutations with dissimilar amino acid properties than those in the central parts of the network. Private mutations occupying the peripheral tips of the network were inferred from alleles that were present in only one sequence, whereas homoplasic mutations were inferred from the presence of a mutation in >1 lineage in the networks. The frequencies of the mutation categories among private amino acid replacements, homoplasic replacements and replacements in hydrophobic regions were compared with those among the remaining ones using the two-tailed Fisher's exact test as implemented in R 1.4.1 (Ihaka & Gentleman 1996).

4.2.3 Rate of detection of new nonsynonymous mutations in European sequences (III)

The total number of amino acid polymorphisms in the European populations is not known. Estimates for the cumulative rate of discovery of new nonsynonymous mutations and for the total number of mutations were derived by taking 500 permutations of the 647 European sequences and examining the sequences contained in each permutation consecutively, calculating for each sequence the cumulative sum of mutations that had not occurred in the previous sequences. Sampling was performed without replacement. Statistical models having an asymptotic maximum were fitted to the mean of the 500 cumulative sum curves by the nonlinear least squares method as implemented in R 1.4.1 to predict the total number of mutations and the number of sequences required for identifying most of them.

4.2.4 Polymorphism indices and phylogenetic tests for neutrality (IV)

The European haplogroup clusters encompassed 646 sequences (HV 300, KU 142, JT 113, IWX 91), whereas 108 belonged to the Asian haplogroups and 82 to the African haplogroup cluster L. One European sequence could not be assigned into any of the four European clusters, and it was therefore excluded from comparisons between these. Genes, codon positions and nondegenerate, twofold degenerate and fourfold degenerate sites were identified according to the Mitomap reference sequence and function locations. Sequence data sets were defined according to their haplogroup assignment and functional classes of sites and analysed in terms of nucleotide polymorphism indices and neutrality tests. Nucleotide diversity π, Watterson's estimate θ_S, and $\theta_{\eta s}$ and the neutrality tests D, $F*$ and Fs were calculated for each sequence data set. A novel program was written in C to calculate the descriptive statistics and tests required in the analyses at a high level of numerical precision, and the source code was released publicly (dnastats, available at `http://cc.oulu.fi/~jukkamoi/mtres`). The program was also designed to permit arbitrary sliding-window analyses. π and the other statistics were compared between the haplogroup-defined sequence sets to identify exceptions from the general rank order of classes of sites.

4.2.5 Nonsynonymous/synonymous rate ratios (IV)

Numbers of nonsynonymous mutations per nonsynonymous sites (d_N) and numbers of synonymous mutations per synonymous sites (d_S) in the 13 protein-coding genes between sequences in each haplogroup-defined sequence set were calculated by the method of Nei & Gojobori (1986) as implemented in DnaSP 3.53 (Rozas & Rozas 1999). The distributions of $d_N/(d_N + d_S)$ values were compared between the haplogroups and the significances of differences between lineages were assessed using the Kruskal-Wallis rank sum test as implemented in R 1.4.1.

4.2.6 Sliding-window analyses (IV)

Two sliding-window methods were used to identify genome regions that showed patterns of nucleotide substitution that were not consistent with the average pattern for the genome.

4.2.6.1 Allele frequency distribution

Neutrality tests are summaries of the allele frequency spectrum (Przeworski *et al.* 2000). Nucleotide diversity and the neutrality tests D, $F*$ and Fs were calculated for each haplogroup-defined sequence set in terms of 1000-bp segments shifted along the mtDNA genome in 40-bp steps. These sliding-window analyses were carried out by using DnaSP

and dnastats with consistent results. The hypothesis of a correlation between nucleotide position and the sliding-window estimates was assessed by linear regression. One of the most extreme lineage-specific local deviations in the allele frequency distribution was examined in more detail to assess its possible causes. A median network (Bandelt *et al.* 1995) was constructed for the respective gene and segregating sites within and outside the observed region were identified, after which Fisher's exact test was used to assess the hypothesis that the frequencies of segregating sites and nonsynonymous and synonymous mutations were distributed evenly between lineages and the respective gene.

4.2.6.2 Partial likelihood optimization

The coding region sequence data sets were analysed by a method which identifies regions which do not fit with a single phylogenetic topology and nucleotide substitution process along the entire sequence. The method uses a maximum-likelihood phylogeny to calculate the likelihoods for each site in the sequence data, and calculates a measure of the average likelihood of a window with respect to the rest of the sequence for windows of varying sizes and positions. Maximum-likelihood trees were reconstructed for each sequence data set using TREE-PUZZLE 5.0 (Schmidt *et al.* 2002) and enforcing the HKY substitution model (Hasegawa *et al.* 1985) with a uniform substitution rate. The transition/transversion parameter was estimated from the data. The sequence data, maximum-likelihood tree and ti/tv parameter were given as input to PLATO 2.11 (Grassly & Holmes 1997) to identify genome regions with a significantly low likelihood, and mutations within such regions were identified. Homologies in the other lineages were screened for by examining each pair of mutations within the region, in order to assess the hypothesis that the significantly low likelihood had resulted from an ancient recombination event.

5 Results

5.1 Prevalence of 3243A>G in Northern Ostrobothnia (I)

Five pedigrees with the 3243A>G mutation in the population of Northern Ostrobothnia were known before the present study. A total of 615 patients fulfilled the selection criteria, and 480 samples were analysed for the 3243A>G mutation. The screening revealed a total of 14 patients, including members of five newly identified families, and one additional family was identified outside the screening setting. The total number of verified mutation carriers >20 years of age in these 11 families who were living in the province of Northern Ostrobothnia was 25. The minimum estimate for the prevalence of the 3243A>G mutation in the adult population of the area was therefore 10.2/100,000 (95% confidence interval 6.2–14.2/100,000). Unexamined first-degree maternal relatives of verified mutation carriers were then defined as obligatory carriers. The total number of verified and obligatory carriers aged >20 years and living in the province of Northern Ostrobothnia was 40, resulting in an estimate of 16.3/100,000 (95% confidence interval 11.3–21.4/100,000) for the prevalence of the 3243A>G mutation.

RFLP and sequencing analysis revealed nine mtDNA genotypes in the 11 families. Six probands belonged to mtDNA haplogroup U, two to haplogroup H, two to haplogroup T and one to haplogroup I. Three probands belonging to haplogroup U had identical sequences. Genealogical analysis showed that two of these three families had a common ancestor born in 1864, but no ancestors common to this family and the third family sharing the mtDNA genotype were found even though the genealogical analyses were extended back to 1808 and 1837.

5.2 Relative fitness of females carrying 3243A>G is not markedly reduced (II)

The reproductive period of 31 female carriers of the 3243A>G mutation coincided with the years 1973-1996. Six had the MELAS syndrome, 13 had diabetes mellitus or hearing impairment but no MELAS syndrome, and the remaining women had milder symptoms or were unaffected. The net reproduction rate calculated for the mutation carriers (NRR_m)

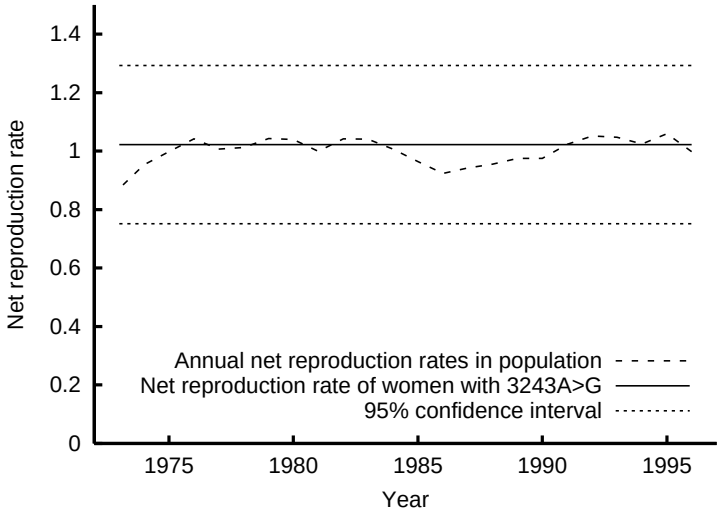


Fig. 4. **Comparison of net reproduction rates between the general population and women with 3243A>G.**

was 1.02 (0.75–1.29, 95% confidence interval), while the annual net reproduction rate in the general population varied between 0.87 and 1.06 during the years 1973–1996, the median (NRR_p) being 1.01 (Figure 4). The mean of the age-specific fertility distribution ($\overline{m}$) did not differ between mutation carriers and the general population (28.0 vs. 28.4 years; $P = 0.83$, one-way analysis of variance).

5.3 Recent nonsynonymous mutations differ from the ancient ones (III)

A total of 988 synonymous, 465 nonsynonymous missense, and 6 nonsense mutations were identified in the protein-coding genes of 840 mtDNA sequences, when the differences were defined relative to the reference sequence. One-third of all the mutations were nonsynonymous. The six nonsense mutations included two in the initiator codon of the MTND1 gene (3308T>C and 3308T>A), a mutation in the initiator codon of MTND5 (12338T>C), a mutation in the stop codon of MTCO1 (7444G>A; MTCO1:S513_X514insKQK), and a frameshift deletion in the middle of MTCO1 (6577delG; MTCO1:G225fsX28). Half of the missense mutations (48%) changed polarity, 26% size, 8% charge, 32% aliphaticity, 13% aromaticity and 44% hydropathy, indicating that such changes are common in human mtDNA and that evaluation of the pathogenicity of a mutation should not rely solely on these structural considerations.

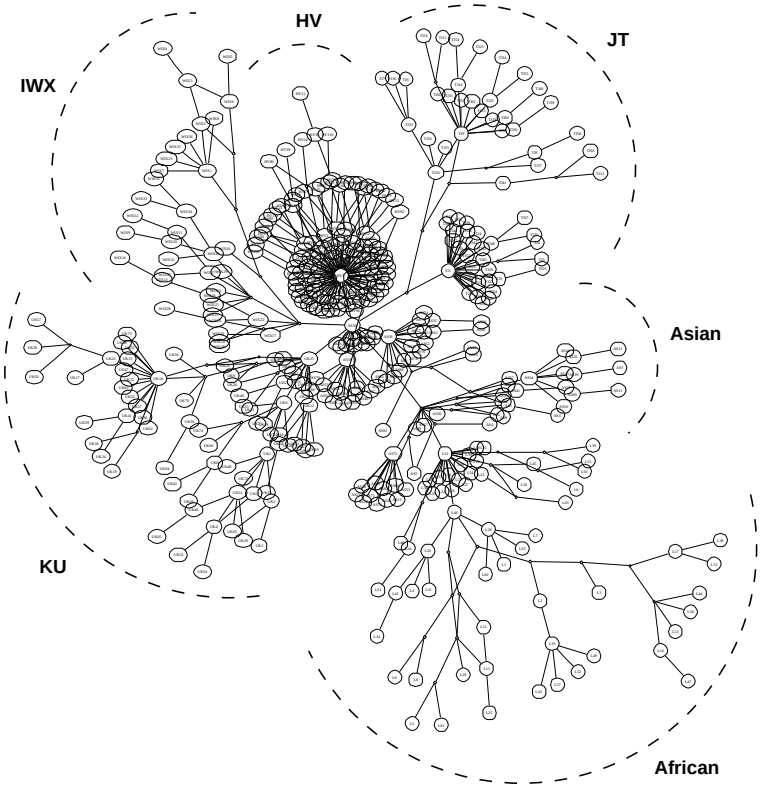


Fig. 5. Unrooted reduced-median network of amino acid haplotypes of 840 human mtDNA sequences. An overview.

The reduced-median networks of the amino acid changes showed few differences between the major continent-specific haplogroups, but there were high degrees of variation and highly starlike phylogenies within the haplogroups. More than half of the nonsynonymous mutations (56%) were private, that is, they were present in only one sequence. One-fourth of the mutations (25%) were homoplasic, indicating that they had arisen more than once during human evolution. Comparison of the physicochemical properties of amino acid changes revealed that nonconservative changes were almost twice as common among private mutations than among nonprivate ones (OR 1.83; 95% confidence interval 1.18–2.85; $P = 0.005$). Homoplasic mutations were less often nonconservative (OR 0.44; 95% C.I. 0.25–0.76; $P = 0.002$) and less often involved changes in the amino acid size (OR 0.58; 95% C.I. 0.33–1.00; $P = 0.04$), aliphaticity (OR 0.49; 95% C.I. 0.29–0.82; $P = 0.004$) and aromaticity (OR 0.44; 95% C.I. 0.17–0.97; $P = 0.04$) than the nonhomoplasic ones.

Analysis of the distribution of the nonsynonymous mutations along the 13 protein-coding genes revealed that they were not distributed evenly across or within the genes.

The density of mutations was generally high in MTATP6 and MTATP8, whereas most of the other genes showed at least some regions which were almost free of amino acid replacements. Such regions were particularly evident in the central parts of MTCO1, MTND4 and MTND5, suggesting that these are conserved because of their functional importance. Nonsynonymous mutations were found to occupy both hydrophobic and hydrophilic regions of genes when the regions were defined according to the average 19-aa hydropathy for the respective gene. The mutations in hydrophobic regions involved less changes in charge and more changes in aliphaticity than expected and were less often nonconservative than those in hydrophilic regions, but such differences apparently resulted from differences in amino acid content between the hydrophobic and hydrophilic regions, because $103/381$ (27%) of the charged amino acids and $697/1,065$ (65%) of the aliphatic amino acids in the reference sequence were found to be located in the hydrophobic regions of the genes.

5.4 Rate of detection of new nonsynonymous mutations in European sequences (III)

The Weibull growth curve $y = \alpha - \beta \exp[-\exp(\delta)x^{\varepsilon}]$ provided the best fit for the mean of the cumulative sum curves, with parameters $\alpha = 1080.84$, $\beta = 1080.31$, $\delta = -5.435$, and $\varepsilon = 0.6664$. The fit was almost perfect (residual sum of squares 16.86). α represents the asymptotic maximum of the Weibull growth curve. According to the estimated model, the 301 mutations detected in 647 European sequences encompass approximately 28% of all the nonsynonymous mutations that may be present in European populations, and 12,200 sequences will be required to identify 90% of the 1,081 mutations, and 18,100 sequences to identify 95%.

5.5 Polymorphism indices, neutrality tests and sliding-window analyses suggest lineage-specific differences in selection (IV)

5.5.1 Comparisons of diversity indices and d_N/d_S ratios

Analysis of the set of all sequences revealed a consistently high haplotype diversity, and all three neutrality tests (*D*, $F*$, and *Fs*) were generally highly negative, suggesting selection against deleterious alleles, population expansion, or both. Comparison of π between functional classes of sites showed a rank order which was in accordance with previous reports, as π increased in the order nondegenerate sites/codon position 1-2 < tRNA < rRNA < coding sequence (CDS) < twofold degenerate sites < codon position 3 < fourfold degenerate sites < D-loop. These differences are assumed to result from differential selection against these classes of sites. Comparison of our estimates with those reported for nuclear sequence variation (Przeworski *et al.* 2000) indicated that π for the coding-region genes of mtDNA is generally not much higher than that found for

many genes in nuclear DNA. The Asian tRNA sequence set showed a low haplotype and nucleotide diversity and a highly negative Tajima's *D*, possibly indicating that mutations in tRNA genes may have been more deleterious in Asian populations than elsewhere.

The nonsynonymous/synonymous rate ratios differed between lineages in several genes when the rate ratio distributions were calculated from pairwise comparisons within each lineage, also suggesting lineage-specific differences in selection against nonsynonymous mutations in the protein-coding genes of mtDNA.

5.5.2 Sliding-window analyses

The sliding-window analysis of allele frequency distribution showed a correlation between nucleotide position and the three population mutation parameters and neutrality tests, this correlation being closer for the number of singleton mutations and for the number of segregating sites than for nucleotide diversity, and consequently closer for $F*$ and Fs than for Tajima's *D*. This may indicate systematically different selection or mutation rates for genes in the 5' and 3' ends of human mtDNA. Furthermore, the analysis allowed identification of regions in which the allele frequency distribution differed from the surrounding regions of the genome and the other haplogroup clusters, the most conspicuous of the regions being between nucleotide positions 12478-13611 within MTND5 in haplogroup cluster JT. Detailed analysis of this region indicated that the finding was mostly explained by an unusual distribution of mutations in haplogroup J. This region had fewer segregating sites and nonsynonymous mutations in haplogroup J than would be expected from their distribution outside the region and in other clusters, suggesting that it may have been under stronger selective pressure in haplogroup J than in others, possibly because of mutual interactions between mutations in the subunits of the OXPHOS complex I.

The maximum-likelihood sliding-window analysis indicated that several sequence data sets included one or more regions which did not fit with the phylogenetic topology and nucleotide substitution process estimated for the entire set. Some of the regions identified differed in position between sequence sets, whereas some were common to one or more haplogroup clusters. There were not many regions which had a significantly low likelihood in the analysis of protein-coding sequences, and these did not generally span entire genes, suggesting that lineage-specific selective forces may have influenced only certain regions of genes and not entire genes. The regions with low likelihood included no homologous pairs of alleles consistent with ancient recombination events between lineages.

6 Discussion

6.1 Selection against the 3243A>G mutation

The common MELAS mutation 3243A>G in mtDNA was found in 11 pedigrees from the province of Northern Ostrobothnia, and its frequency in the adult population was calculated to be at least 10.2/100,000. This is the highest published prevalence figure worldwide for this mutation and indicates that it is surprisingly common in Northern Finland, and common by comparison with other pathogenic mutations in mtDNA (Chinnery *et al.* 2000, Chinnery & Turnbull 2001, Darin *et al.* 2001, Man *et al.* 2003). The true prevalence may be even higher, because only selected patient groups were analysed, and cases were lost due to nonresponse to questionnaires, nonrecovery of patient charts and nonparticipation in blood-sample collection.

Interestingly, 6 out of the 11 maternal ancestors had been born in the province of Kainuu. Finland is known for >30 heritable disorders that are unusually prevalent here due to founder effects and population isolation, "the Finnish Disease Heritage" (Norio 2003a), but the enrichment of rare alleles generally leads to an increased risk of recessively inherited disease and not mitochondrial disease. It is possible that the apparent clustering of the 3243A>G mutation could be explained by the recent finding that errors in the nuclear genes coding for mtDNA repair and maintenance mechanisms may provoke mtDNA mutation or transmission to the germ line, and such nuclear errors are often recessive. The frequency of the MERRF mutation 8344A>G in Northern Ostrobothnia is similar to that in other populations, however, and much lower than that of 3243A>G, even though both have a similar phenotype in adulthood and both are mutations in a tRNA gene. The upper limit of the 95% confidence interval for the prevalence of 8344A>G is 1.5/100,000 (Remes *et al.* 2003), whereas the lower limit for 3243A>G in this study was 6.2/100,000, indicating that 3243A>G is at least four times more common than 8344A>G in Northern Ostrobothnia. Therefore, if a nuclear factor is responsible for the high prevalence of 3243A>G, it should be fairly specific to that mutation. Alternatively, population-specific differences in selection against these mutations could explain the observed differences in prevalence. The ascertained families with the 3243A>G mutation did not include any childhood-onset MELAS patients, and the frequency of the mutation among patients with childhood encephalopathies and myopathies in the same geographical region is low (Uusimaa *et al.* 2000), suggesting

that the onset of symptoms is generally delayed until adulthood.

This was the first attempt to assess the reproductive capacity, or fitness, of carriers of a highly pathogenic mtDNA mutation. Surprisingly, it was found that the female carriers of 3243A>G were on average fairly mildly affected, and they were able to have as many children as women in the general population. The estimated confidence interval for their relative fitness was wide, but a severe reduction was nevertheless excluded. This is a seemingly contradictory finding, in that the 3243A>G is generally considered to be one of the most deleterious mtDNA mutations and has all the typical characteristics of a deleterious mutation in phylogenetic analyses (Torroni *et al.* 2003). This discrepancy has at least two possible explanations. First, the majority of mutation carriers in the population may generally have such a low degree of mutant heteroplasmy that they are affected only mildly, if at all. A severe reduction in fitness may occur only when someone inherits a high degree of mutant heteroplasmy, and this event, albeit rare, may still be common enough that, on an evolutionary time scale, it always removes the mutation from each lineage where it has occurred. Second, selection may also operate on levels other than reproductive fitness. If mtDNAs, mitochondria, or germ line cells such as oocytes harbouring the 3243A>G mutation were somehow recognized and selected against in mtDNA replication, replicative segregation, cell division, or fertilization, this would result in a powerful negative selection against the mutation in evolutionary analyses, even though the consequences at the level of the entire organism would be negligible. Interestingly, there is some evidence that the segregation of mtDNA genotypes may be nonrandom (Battersby *et al.* 2003, Cox & Spradling 2003).

6.2 A novel phylogenetic test for the neutrality of amino acid replacements

The number of complete mtDNA sequences has increased exponentially during the last couple of years, marking the start of mitochondrial population genomics (Hedges 2000). Thus the mtDNA shares the “genomic revolution” with nuclear DNA, and the large number of mtDNA sequences that are emerging are opening up new possibilities for statistical analyses of genome-wide evolutionary patterns (Wolfe & Li 2003) and attempts to assign functional relevance to the numerous variations observed in these sequences. This study is the first to use bioinformatic tools to collect all the available complete human mtDNA coding region sequences into a single database and to identify and characterize all nonsynonymous variation in these sequences. Moreover, a novel phylogenetic test for the neutrality of amino acid replacements was constructed by formulating the hypothesis that the evolutionarily recent nonsynonymous mutations, which, according to the neutral theory, might include some slightly deleterious alleles, could also show different physicochemical properties from the older mutations.

The finding that private nonsynonymous mutations were more often nonconservative, while homoplasic mutations were more often conservative and included fewer changes involving small, aliphatic and aromatic amino acids than the remaining mutations, suggests that the physicochemical properties of amino acid replacements are relevant to natural selection and protein evolution, and that a proportion of the nonsynonymous

mutations that are observed in human populations are deleterious enough to be eventually removed by selection, implying that they may also be deleterious enough to be associated with diseases. The finding that not all carriers of the 3243A>G mutation were severely affected and that their reproductive fitness was not markedly reduced even though the mutation is obviously subject to purifying selection (Torroni *et al.* 2003) suggests that mutations which appear to be slightly deleterious in phylogenetic analyses may well modify the risks of contracting certain disorders and manifest themselves in clinical illness only in some individuals. Such slightly deleterious mutations might explain the observation that some traits seem to be associated with particular mtDNA lineages, and it is also possible that certain combinations of otherwise neutral polymorphisms are slightly deleterious because of their cumulative effect (Lertrit *et al.* 1994, Lehtonen *et al.* 2003) or mutual interactions (Rahman *et al.* 1999, Brown *et al.* 2002, Howell *et al.* 2002, Sudoyo *et al.* 2002).

6.3 Characterization of mtDNA sequence variation

The reduced-median networks and tabulated physicochemical properties of the nonsynonymous mutations should have practical applications, because they facilitate the identification of haplogroup-specific fixed polymorphisms and comparisons of the physicochemical properties of the mutations. The frequency of changes in the individual physicochemical properties of amino acid replacements was high, suggesting that such changes are quite common in human mtDNA and that a polar-nonpolar amino acid replacement may generally be no more pathogenic than a polar-polar replacement, for example.

The sequences analysed here contained six nonsense mutations which had not been mentioned in the original articles, and in one case an association between a haplogroup-related polymorphism and cardiomyopathy had been reported whereas the published sequence contained an uncommented frameshift deletion in the middle of the MTCO1 gene (Shin *et al.* 2000). It is difficult to exclude the possibility of sequencing errors in such instances, but they nevertheless highlight the importance of a population-genetic approach for recognizing neutral variants that may be present in the population, and the importance of examining the entire mtDNA when a particular variant is suspected to be pathogenic. It is also possible that the sequence databases may contain errors which could potentially influence the results of statistical analyses of mtDNA, but the effects of possible random sequencing errors on the analyses presented here should be small, due to the very large number of sequences included.

6.4 Fingerprints of mutation and selection in mtDNA

The distribution of nonsynonymous mutations was assessed by a simple transmembrane prediction method based on the average Kyte-Doolittle hydropathy scale of 19 adjacent amino acids. This analysis revealed differences in the density of amino acid replacements

between and within genes, as MTATP6 and MTATP8 appeared to be more variable overall than the remaining genes, and some genes such as MTCO1, MTND4 and MTND5 showed regions which were almost invariable. Such differences are likely to represent different functional constraints in the respective subunits, and the high variation in MTATP6 is consistent with the finding that the functional constraints for this subunit seem to be so low that even the nuclear-encoded and evolutionarily distant subunit of the green alga *Chlamydomonas reinhardtii* has been reported to function in human cells (Ojaimi *et al.* 2002).

The analysis of nucleotide diversity between mtDNA lineages and classes of nucleotide sites indicated that the general rank order of sites and genome regions was consistent with the assumption that a certain proportion of nonsynonymous variation is deleterious and is removed by selection, since the lowest diversity was observed for the nondegenerate sites and the highest for the fourfold degenerate sites and the D-loop. Somewhat surprisingly, the nucleotide diversity per site for the coding region of mtDNA was not much higher than that found for many genes in nuclear DNA (Przeworski *et al.* 2000). The Asian tRNA sequences had a lower nucleotide diversity than the Asian nondegenerate sites, and Tajima's *D* was also highly significantly negative. This could imply that selection against mutations in tRNA genes has been stronger in Asian populations than elsewhere.

This first sliding-window analysis of nucleotide polymorphism indices and neutrality tests to be conducted for a large set of human mtDNA sequences produced surprising results. It is commonly assumed that the African haplogroup L has the highest nucleotide diversity of all continent-specific haplogroups, but it turned out that this is not equally true for all regions of mtDNA, and the regions of highest diversity occupied different positions in the African and Asian sequences. Even more surprisingly, there was a positive correlation between nucleotide position and the three population mutation parameters and a negative correlation for the neutrality tests $F*$ and Fs. This might imply systematically different selection or mutation rates between genes in the 3' and 5' ends of the human mtDNA coding-region sequence. The reason for the observed correlation is not known, and it does not correspond well to the time for which mtDNA is single-stranded during replication (Nedbal & Flynn 1998).

The sliding-window analysis also revealed genome regions in which the polymorphism indices and neutrality tests differed from those for the surrounding regions and for other haplogroup clusters. This was most conspicuous for the region spanning amino acids 48-425 of MTND5 in haplogroup cluster JT, where it was found to result from the relative lack of segregating sites and nonsynonymous mutations within this region, especially in haplogroup J. Interestingly, this same region of MTND5 was found to be relatively well conserved in the analysis of the distribution of nonsynonymous mutations within and between genes, and, furthermore, a segment of the region has been assigned particular functional importance in the bacterial Complex I NuoL subunit, which is analogous to the mammalian MTND5 subunit (Mathiesen & Hägerhäll 2002). It is also interesting that the pattern was observed in haplogroup J, because this haplogroup has been associated with susceptibility to certain complex diseases (Lamminen *et al.* 1997, Torroni *et al.* 1997, Wallace *et al.* 1999, Brown *et al.* 2002) and with longevity (Rose *et al.* 2001, Niemi *et al.* 2003), and it is defined by three amino acid replacements, two of which are within the subunits of Complex I and one within MTND5 and close to the identified aberrant segment. It is therefore possible that the pattern represents an example of mitochondrial-

mitochondrial or nuclear-mitochondrial interaction that is specific to this lineage and which has exposed the region to unusually powerful selective pressure.

The analysis of nonsynonymous/synonymous rate ratios in the 13 protein-coding genes of mtDNA indicated differences between the African, Asian and European clusters, suggesting that such differences may result from something other than the climate, unless it is assumed that the European haplogroup clusters have also evolved at different temperatures (Mishmar *et al.* 2003). Finally, the maximum-likelihood sliding-window analysis suggested that the regions with the highest diversity differ between haplogroup clusters. Overall, several lines of evidence suggest that selection against mtDNA may differ between lineages, possibly due to mitochondrial-mitochondrial or mitochondrial-nuclear interactions, environmental interactions, or combinations of these. Understanding of such interactions will be crucial to our attempts to resolve the pathogenesis of various mitochondrial disorders.

7 Conclusions

Mutations in mitochondrial DNA are important determinants of disease. The MELAS mutation 3243A>G is one of the most common pathogenic mutations, and its genetic epidemiology in Northern Ostrobothnia suggests that nuclear genes may be involved in its high prevalence in this area. The 3243A>G mutation did not cause any marked reproductive disadvantage to its female carriers, however, suggesting that most carriers in the population are relatively mildly affected. This does not necessarily contradict the notion of strong evolutionary selection against the mutation as inferred from phylogenetic studies.

General-purpose bioinformatic tools were well suited for the analysis of >14 Mb of mtDNA sequence data, although it was necessary to build a custom database system and write supporting software to complete the analyses. It became obvious during the analyses that the importance of minimizing human intervention, and consequently the number of human-introduced errors, is crucial in large-scale sequence analyses. The possibilities for assigning functional relevance to sequence data, which is one of the central goals of bioinformatics, are rapidly increasing with the growing quantities of human mtDNA sequence data.

This was the first attempt to assign functional relevance to the mtDNA variation that is present in the >800 human mtDNA sequences that have been published during the last couple of years. A novel hypothesis regarding the physicochemical properties of amino acid replacements in a phylogenetic context was formulated and tested, and the results supported the neutral model of evolution and the assumption that slightly deleterious amino acid polymorphisms, possibly associated with the risks of contracting certain complex diseases, are present in populations. Moreover, all the nonsynonymous variation that was identified was characterized, and the distribution of mutations within and between genes was assessed to identify protein regions that appear to be constrained by selection. The distribution was still relatively sparse, however, suggesting that many more sequences still need to be analysed for definite mapping of the constrained and nonconstrained regions in proteins encoded by human mtDNA.

The exploratory analysis, which included nucleotide polymorphism indices, neutrality tests and comparisons of nonsynonymous and synonymous mutations between mtDNA lineages and genes, provided several lines of evidence suggesting that the selective constraints against particular regions of the mtDNA genome may differ between lineages,

and that lineage-specific interactions are a plausible explanation for such differences. This is an important finding, because it implies that the pathogenic potential of a mutation in mtDNA or in mtDNA-related nuclear genes may well depend markedly on the presence of other, interacting mutations. The understanding of such interactions and constraints will be a major challenge in the field of mitochondrial bioinformatics in the future.

References

Abe, S., Usami, S., Shinkawa, H., Weston, M., Overbeck, L., Hoover, D., Kenyon, J., Horai, S. & Kimberling, W. (1998) Phylogenetic analysis of mitochondrial DNA in Japanese pedigrees of sensorineural hearing loss associated with the A1555G mutation. Eur J Hum Genet 6:563–569.

Abe, S., Kelley, P., Kimberling, W. & Usami, S. (2001) Connexin 26 gene (GJB2) mutation modulates the severity of hearing loss associated with the 1555A–>G mitochondrial mutation. Am J Med Genet 103:334–338.

Adam-Vizi, V. (2003) Mitochondria: inducers and targets of oxidative stress. J Neurochem 85 Suppl 2:3.

Agostino, A., Chinnery, P., Carrara, F., Schaefer, A. & Tiranti, V. (2003) Mutations of ANT1, Twinkle, and POLG1 in sporadic progressive external ophthalmoplegia (PEO). Neurology 60:1354–1356.

Akashi, H. (1999) Within- and between-species DNA sequence variation and the 'footprint' of natural selection. Gene 238:39–51.

Alam, T., Kanki, T., Muta, T., Ukaji, K., Abe, Y., Nakayama, H., Takio, K., Hamasaki, N. & Kang, D. (2003) Human mitochondrial DNA is packaged with TFAM. Nucleic Acids Res 31:1640–1645.

Alff-Steinberger, C. (1969) The genetic code and error transmission. Proc Natl Acad Sci U S A 64:584–591.

Amiri, H., Karlberg, O. & Andersson, S. (2003) Deep origin of plastid/parasite ATP/ADP translocases. J Mol Evol 56:137–150.

Anderson, S., Bankier, A., Barrell, B., de Bruijn, M., Coulson, A., Drouin, J., Eperon, I., Nierlich, D., Roe, B., Sanger, F., Schreier, P., Smith, A., Staden, R. & Young, I. (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Andrews, R., Kubacka, I., Chinnery, P., Lightowlers, R., Turnbull, D. & Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147.

Arnason, U., Adegoke, J., Bodin, K., Born, E., Esa, Y., Gullberg, A., Nilsson, M., Short, R., Xu, X. & Janke, A. (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. Proc Natl Acad Sci U S A 99:8151–8156.

Astuti, D., Latif, F., Dallol, A., Dahia, P., Douglas, F., George, E., Skoldberg, F., Husebye, E., Eng, C. & Maher, E. (2001) Gene mutations in the succinate dehydrogenase subunit SDHB cause susceptibility to familial pheochromocytoma and to familial paraganglioma. Am J Hum Genet 69:49–54.

Bachtrog, D. (2003) Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. Nat Genet 34:215–219.

Baker, M., Grundy, W. & Elkan, C. (1999) A common ancestor for a subunit in the mitochondrial proton-translocating NADH:ubiquinone oxidoreductase (complex I) and short-chain dehydrogenases/reductases. Cell Mol Life Sci 55:450–455.

Bamshad, M. & Wooding, S. (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111.

Bandelt, H., Forster, P., Sykes, B. & Richards, M. (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753.

Bandelt, H., Macaulay, V. & Richards, M. (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol Phylogenet Evol 16:8–28.

Bandelt, H., Alves-Silva, J., Guimaraes, P., Santos, M., Brehm, A., Pereira, L., Coppa, A., Larruga, J., Rengo, C., Scozzari, R., Torroni, A., Prata, M., Amorim, A., Prado, V. & Pena, S. (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet 65:549–563.

Barrell, B., Bankier, A. & Drouin, J. (1979) A different genetic code in human mitochondria. Nature 282:189–194.

Barrientos, A. & Moraes, C. (1999) Titrating the effects of mitochondrial complex I impairment in the cell physiology. J Biol Chem 274:16188–16197.

Battersby, B., Loredo-Osti, J. & Shoubridge, E. (2003) Nuclear genetic control of mitochondrial DNA segregation. Nat Genet 33:183–186.

Baysal, B., Ferrell, R., Willett-Brozick, J., Lawrence, E., Myssiorek, D., Bosch, A., van der May, A., Taschner, P., Rubinstein, W., Myers, E., Richards, C., Cornelisee, C., Devilee, P. & Devlin, B. (2000) Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. Science 287:848–851.

Benit, P., Chretien, D., Kadhom, N., de Lonlay-Debeney, P., Cormier-Daire, V., Cabral, A., Peudenier, S., Rustin, P., Munnich, A. & Rotig, A. (2001) Large-scale deletion and point mutations of the nuclear NDUFV1 and NDUFS1 genes in mitochondrial complex I deficiency. Am J Hum Genet 68:1344–1352.

Benit, P., Beugnot, R., Chretien, D., Giurgea, I., Lonlay-Debeney, P. D., Issartel, J., Corral-Debrinski, M., Kerscher, S., Rustin, P., Rotig, A. & Munnich, A. (2003) Mutant NDUFV2 subunit of mitochondrial complex I causes early onset hypertrophic cardiomyopathy and encephalopathy. Hum Mutat 21:582–586.

Bindoff, L., Desnuelle, C., Birch-Machin, M., Pellissier, J., Serratrice, G., Dravet, C., Bureau, M., Howell, N. & Turnbull, D. (1991) Multiple defects of the mitochondrial respiratory chain in a mitochondrial encephalopathy (MERRF): a clinical, biochemical and molecular study. J Neurol Sci 102:17–24.

Blier, P., Dufresne, F. & Burton, R. (2001) Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. Trends Genet 17:400–406.

Bo, R. D., Bordoni, A., Boneschi, F. M., Crimi, M., Sciacco, M., Bresolin, N., Scarlato, G. & Comi, G. (2002) Evidence and age-related distribution of mtDNA D-loop point mutations in skeletal muscle from healthy subjects and mitochondrial patients. J Neurol Sci 202:85–91.

Borras, C., Sastre, J., Garcia-Sala, D., Lloret, A., Pallardo, F. & Vina, J. (2003) Mitochondria from females exhibit higher antioxidant gene expression and lower oxidative damage than males. Free Radic Biol Med 34:546–552.

Bourgeron, T., Rustin, P., Chretien, D., Birch-Machin, M., Bourgeois, M., Viegas-Pequignot, E., Munnich, A. & Rotig, A. (1995) Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. Nat Genet 11:144–149.

Brown, M., Starikovskaya, E., Derbeneva, O., Hosseini, S., Allen, J., Mikhailovskaya, I., Sukernik, R. & Wallace, D. (2002) The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J. Hum Genet 110:130–138.

Brown, W., M, J. G. & Wilson, A. (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76:1967–1971.

Burger, G., Forget, L., Zhu, Y., Gray, M. & Lang, B. (2003) Unique mitochondrial genome architecture in unicellular relatives of animals. Proc Natl Acad Sci U S A 100:892–897.

Butlin, R. (2002) The costs and benefits of sex: new insights from old asexual lineages. Nat Rev Genet 3:311–317.

Bykhovskaya, Y., Estivill, X., Taylor, K., Hang, T., Hamon, M., Casano, R., Yang, H., Rotter, J., Shohat, M. & Fischel-Ghodsian, N. (2000) Candidate locus for a nuclear modifier gene for maternally inherited deafness. Am J Hum Genet 66:1905–1910.

Capaldi, R. & Aggeler, R. (2002) Mechanism of the F1F0-type ATP synthase, a biological rotary motor. Trends Biochem Sci 27:154–160.

Capaldi, R., Aggeler, R., Gilkerson, R., Hanson, G., Knowles, M., Marcus, A., Margineantu, D., Marusich, M., Murray, J., Oglesbee, D., Remington, S. & Rossignol, R. (2002) A replicating module as the unit of mitochondrial structure and functioning. Biochim Biophys Acta 1555:192–195.

Capps, G., Samuels, D. & Chinnery, P. (2003) A model of the nuclear control of mitochondrial DNA replication. J Theor Biol 221:565–583.

Caramelli, D., Laluela-Fox, C., Vernesi, C., Lari, M., Casoli, A., Mallegni, F., Chiarelli, B., Dupanloup, I., Bertranpetit, J., Barbujani, G. & Bertorelle, G. (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. Proc Natl Acad Sci U S A 100:6593–6597.

Carelli, V., Baracca, A., Barogi, S., Pallotti, F., Valentino, M., Montagna, P., Zeviani, M., Pini, A., Lenaz, G., Baruzzi, A. & Solaini, G. (2002) Biochemical-clinical correlation in patients with different loads of the mitochondrial DNA T8993G mutation. Arch Neurol 59:264–270.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C., Lim, E., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. & Lander, E. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238.

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

Casas, F., Daury, L., Grandemange, S., Busson, M., Seyer, P., Hatier, R., Carazo, A., Cabello, G. & Wrutniak-Cabello, C. (2003) Endocrine regulation of mitochondrial activity: involvement of truncated RXRα and c-Erb Aα1 proteins. FASEB J 17:426–436.

Cavalli-Sforza, L. & Bodmer, W. (1971) The genetics of human populations. W.H. Freeman and Company, San Francisco.

Cavalli-Sforza, L. & Feldman, M. (2003) The application of molecular genetic approaches to the study of human evolution. Nat Genet 33 Suppl:266–275.

Cerritelli, S., Frolova, E., Feng, C., Grinberg, A., Love, P. & Crouch, R. (2003) Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. Mol Cell 11:807–815.

Chada, S. & Hollenbeck, P. (2003) Mitochondrial movement and positioning in axons: the role of growth factor signaling. J Exp Biol 206:1985–1992.

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 85:20–30.

Chalmers, S. & Nicholls, D. (2003) The relationship between free and total calcium concentrations in the matrix of liver and brain mitochondria. J Biol Chem 278:19062–19070.

Chen, D., Cao, G., Hastings, T., Feng, Y., Pei, W., O'Horo, C. & Chen, J. (2002) Age-dependent decline of DNA repair activity for oxidative lesions in rat brain mitochondria. J Neurochem 81:1273–1284.

Chen, X., Prosser, R., Simonetti, S., Sadlock, J., Jagiello, G. & Schon, E. (1995a) Rearranged mitochondrial genomes are present in human oocytes. Am J Hum Genet 57:239–247.

Chen, Y., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. & Wallace, D. (1995b) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57:133–149.

Chinnery, P. & Turnbull, D. (2001) Epidemiology and treatment of mitochondrial disorders. Am J Med Genet 106:94–101.

Chinnery, P., Johnson, M., Wardell, T., Singh-Kler, R., Hayes, C., Brown, D., Taylor, R., Bindoff, L. & Turnbull, D. (2000) The epidemiology of pathogenic mitochondrial DNA mutations. Ann Neurol 48:188–193.

Chinnery, P., Samuels, D., Elson, J. & Turnbull, D. (2002) Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? Lancet 360:1323–1325.

Chomyn, A. & Attardi, G. (2003) MtDNA mutations in aging and apoptosis. Biochem Biophys Res Commun 304:519–529.

Chomyn, A., Enriquez, J., Micol, V., Fernandez-Silva, P. & Attardi, G. (2000) The mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episode syndrome-associated human mitochondrial tRNALeu(UUR) mutation causes aminoacylation deficiency and concomitant reduced association of mRNA with ribosomes. J Biol Chem 275:19198–19209.

Choo-Kang, A., Lynn, S., Taylor, G., Daly, M., Sihota, S., Wardell, T., Chinnery, P., Turnbull, D. & Walker, M. (2002) Defining the importance of mitochondrial gene defects in maternally inherited diabetes by sequencing the entire mitochondrial genome. Diabetes 51:2317–2320.

Clayton, D. (1982) Replication of animal mitochondrial DNA. Cell 28:693–705.

Collins, T. & Bootman, M. (2003) Mitochondria are morphologically heterogeneous within cells. J Exp Biol 206:1993–2000.

Cottrell, D., Blakely, E., Johnson, M., Ince, P. & Turnbull, D. (2001) Mitochondrial enzyme-deficient hippocampal neurons and choroidal cells in AD. Neurology 57:260–264.

Cox, R. & Spradling, A. (2003) A Balbiani body and the fusome mediate mitochondrial inheritance during Drosophila oogenesis. Development 130:1579–1590.

Croteau, D., Stierum, R. & Bohr, V. (1999) Mitochondrial DNA repair pathways. Mutat Res 434:137–148.

Danielson, S., Wong, A., Carelli, V., Martinuzzi, A., Schapira, A. & Cortopassi, G. (2002) Cells bearing mutations causing Leber's hereditary optic neuropathy are sensitized to Fas-induced apoptosis. J Biol Chem 277:5810–5815.

Darin, N., Oldfors, A., Moslemi, A., Holme, E. & Tulinius, M. (2001) The incidence of mitochondrial encephalomyopathies in childhood: clinical features and morphological, biochemical, and DNA abnormalities. Ann Neurol 49:377–383.

de Grey, A. (2002) The reductive hotspot hypothesis of mammalian aging: membrane metabolism magnifies mutant mitochondrial mischief. Eur J Biochem 269:2003–2009.

de Lonlay, P., Valnot, I., Barrientos, A., Gorbatyuk, M., Tzagoloff, A., Taanman, J., Benayoun, E., Chretien, D., Kadhom, N., Lombes, A., de Baulny, H., Niaudet, P., Munnich, A., Rustin, P. & Rotig, A. (2001) A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. Nat Genet 29:57–60.

den Dunnen, J. & Antonarakis, S. (2001) Nomenclature for the description of human sequence variations. Hum Genet 109:121–124.

Derbeneva, O., Sukernik, R., Volodko, N., Hosseini, S., Lott, M. & Wallace, D. (2002) Analysis of mitochondrial DNA diversity in the Aleuts of the Commander Islands and its implications for the genetic history of Beringia. Am J Hum Genet 71:415–421.

Diez-Sanchez, C., Ruiz-Pesini, E., Lapena, A., Montoya, J., Perez-Martos, A., Enriquez, J. & Lopez-Perez, M. (2003) Mitochondrial DNA content of human spermatozoa. Biol Reprod 68:180–185.

DiMauro, S. & Schon, E. (2001) Mitochondrial DNA mutations in human disease. Am J Med Genet 106:18–26.

Duborjal, H., Beugnot, R., Camaret, B. D. & Issartel, J. (2002) Large functional range of steady-state levels of nuclear and mitochondrial transcripts coding for the subunits of the human mitochondrial OXPHOS system. Genome Res 12:1901–1909.

Dyall, S., Koehler, C., Delgadillo-Correa, M., Bradley, P., Plumper, E., Leuenberger, D., Turck, C. & Johnson, P. (2000) Presence of a member of the mitochondrial carrier family in hydrogenosomes: conservation of membrane-targeting pathways between hydrogenosomes and mitochondria. Mol Cell Biol 20:2488–2497.

Elpeleg, O., Mandel, H. & Saada, A. (2002) Depletion of the other genome-mitochondrial DNA depletion syndromes in humans. J Mol Med 80:389–396.

Elson, J., Andrews, R., Chinnery, P., Lightowlers, R., Turnbull, D. & Howell, N. (2001) Analysis of European mtDNAs for recombination. Am J Hum Genet 68:145–153.

Emelyanov, V. (2003) Mitochondrial connection to the origin of the eukaryotic cell. Eur J Biochem 270:1599–1618.

Estivill, X., Govea, N., Barcelo, E., Badenas, C., Romero, E., Moral, L., Scozzri, R., D'Urbano, L., Zeviani, M. & Torroni, A. (1998) Familial progressive sensorineural deafness is mainly due to the mtDNA A1555G mutation and is enhanced by treatment of aminoglycosides. Am J Hum Genet 62:27–35.

Eyre-Walker, A. (2002) Changing effective population size and the McDonald-Kreitman test. Genetics 162:2017–2024.

Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N. & Gustafsson, C. (2002) Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. Nat Genet 31:289–294.

Fayet, G., Jansson, M., Sternberg, D., Moslemi, A., Blondy, P., Lombes, A., Fardeau, M. & Oldfors, A. (2002) Ageing muscle: clonal expansions of mitochondrial DNA point mutations and deletions cause focal impairment of mitochondrial function. Neuromuscul Disord 12:484–493.

Finnilä, S., Lehtonen, M. & Majamaa, K. (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475–1484.

Forster, L., Forster, P., Lutz-Bonengel, S., Willkomm, H. & Brinkmann, B. (2002) Natural radioactivity and human mitochondrial DNA mutations. Proc Natl Acad Sci U S A 99:13950–13954.

Fu, Y. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925.

Fu, Y. & Li, W. (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709.

Gakh, O., Cavadini, P. & Isaya, G. (2002) Mitochondrial processing peptidases. Biochim Biophys Acta 1592:63–77.

Garrido, N., Griparic, L., Jokitalo, E., Wartiovaara, J., Bliek, A. V. D. & Spelbrink, J. (2003) Composition and dynamics of human mitochondrial nucleoids. Mol Biol Cell 14:1583–1596.

Gellerich, F., Deschauer, M., Chen, Y., Muller, T., Neudecker, S. & Zierz, S. (2002) Mitochondrial respiratory rates and activities of respiratory chain complexes correlate linearly with heteroplasmy of deleted mtDNA without threshold and independently of deletion size. Biochim Biophys Acta 1556:41–52.

Gerber, A., Loggins, R., Kumar, S. & Dowling, T. (2001) Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Annu Rev Genet 35:539–566.

Ghelli, A., Zanna, C., Porcelli, A., Schapira, A., Martinuzzi, A., Carelli, V. & Rugolo, M. (2003) Leber's hereditary optic neuropathy (LHON) pathogenic mutations induce mitochondrial-dependent apoptotic death in transmitochondrial cells incubated with galactose medium. J Biol Chem 278:4145–4150.

Giles, R., Blanc, H., Cann, H. & Wallace, D. (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A 77:6715–6719.

Glazko, G. & Nei, M. (2003) Estimation of divergence times for major lineages of primate species. Mol Biol Evol 20:424–434.

Goethem, G. V., Dermaut, B., Lofgren, A., Martin, J. & Broeckhoven, C. V. (2001) Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. Nat Genet 28:211–212.

Goffart, S. & Wiesner, R. (2003) Regulation and co-ordination of nuclear gene expression during mitochondrial biogenesis. Exp Physiol 88:33–40.

Goto, Y., Nonaka, I. & Horai, S. (1990) A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. Nature 348:651–653.

Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. Science 185:862–864.

Grassly, N. & Holmes, E. (1997) A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol Biol Evol 14:239–247.

Graur, D. & Li, W. (1991) Neutral mutation hypothesis test. Nature 354:115–116.

Gray, M., Burger, G. & Lang, B. (2001) The origin and early evolution of mitochondria. Genome Biol 2:reviews1018.1–1018.5.

Graziewicz, M., Day, B. & Copeland, W. (2002) The mitochondrial DNA polymerase as a target of oxidative damage. Nucleic Acids Res 30:2817–2824.

Grivennikova, V., Serebryanaya, D., Isakova, E., Belozerskaya, T. & Vinogradov, A. (2003) The transition between active and de-activated forms of NADH:ubiquinone oxidoreductase (Complex I) in the mitochondrial membrane of Neurospora crassa. Biochem J 369:619–626.

Gu, G., Reyes, P., Golden, G., Woltjer, R., Hulette, C., Montine, T. & Zhang, J. (2002) Mitochondrial DNA deletions/rearrangements in Parkinson disease and related neurodegenerative disorders. J Neuropathol Exp Neurol 61:634–639.

Guan, M., Fischel-Ghodsian, N. & Attardi, G. (2001) Nuclear background determines biochemical phenotype in the deafness-associated mitochondrial 12S rRNA mutation. Hum Mol Genet 10:573–580.

Guenebaut, V., Schlitt, A., Weiss, H., Leonard, K. & Friedrich, T. (1998) Consistent structure between bacterial and mitochondrial NADH:ubiquinone oxidoreductase (complex I). J Mol Biol 276:105–112.

Guillausseau, P., Massin, P., Dubois-LaForgue, D., Timsit, J., Virally, M., Gin, H., Bertin, E., Blickle, J., Bouhanick, B., Cahen, J., Caillat-Zucman, S., Charpentier, G., Chedin, P., Derrien, C., Ducluzeau, P., Grimaldi, A., Guerci, B., Kaloustian, E., Murat, A., Olivier, F., Paques, M., Paquis-Flucklinger, V., Porokhov, B., Samuel-Lajeunesse, J. & Vialettes, B. (2001) Maternally inherited diabetes and deafness: a multicenter study. Ann Intern Med 134:721–728.

Hammans, S., Sweeney, M., Hanna, M., Brockington, M., Morgan-Hughes, J. & Harding, A. (1995) The mitochondrial DNA transfer RNALeu(UUR) A–>G(3243) mutation. A clinical and genetic study. Brain 118:721–734.

Hasegawa, M., Kishino, H. & Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174.

Hasegawa, M., Cao, Y. & Yang, Z. (1998) Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol Biol Evol 15:1499–1505.

Hedges, S. (2000) Human evolution. A start for population genomics. Nature 408:652–653.

Hekimi, S. & Guarente, L. (2003) Genetics and the specificity of the aging process. Science 299:1351–1354.

Helgason, A., Hrafnkelsson, B., Gulcher, J., Ward, R. & Stefansson, K. (2003) A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. Am J Hum Genet 72:1370–1388.

Henikoff, S. & Henikoff, J. (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89:10915–10919.

Herrnstadt, C., Elson, J., Fahy, E., Preston, G., Turnbull, D., Anderson, C., Ghosh, S., Olefsky, J., Beal, M., Davis, R. & Howell, N. (2002a) 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., Preston, G., Andrews, R., Chinnery, P., Lightowlers, R., Turnbull, D., Kubacka, I. & Howell, N. (2002b) A high frequency of mtDNA polymorphisms in HeLa cell sublines. Mutat Res 501:19–28.

Heyer, E., Zietkiewicz, E., Rochowski, A., Yotova, V., Puymirat, J. & Labuda, D. (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. Am J Hum Genet 69:1113–1126.

Holt, I., Harding, A., Petty, R. & Morgan-Hughes, J. (1990) A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. Am J Hum Genet 46:428–433.

Howell, N., Miller, N., Mackey, D., Arnold, A., Herrnstadt, C., Williams, I. & Kubacka, I. (2002) Lightning strikes twice: Leber hereditary optic neuropathy families with two pathogenic mtDNA mutations. J Neuroophthalmol 22:262–269.

Howell, N., Herrnstadt, C., Shults, C. & Mackey, D. (2003a) Low penetrance of the 14484 LHON mutation when it arises in a non-haplogroup J mtDNA background. Am J Med Genet 119A:147–151.

Howell, N., Oostra, R., Bolhuis, P., Spruijt, L., Clarke, L., Mackey, D., Preston, G. & Herrnstadt, C. (2003b) Sequence analysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy. Am J Hum Genet 72:1460–1469.

Howell, N., Smejkal, C., Mackey, D., Chinnery, P., Turnbull, D. & Herrnstadt, C. (2003c) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet 72:659–670.

Hudson, R., Kreitman, M. & Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159.

Hughes, K., Alipaz, J., Drnevich, J. & Reynolds, R. (2002) A test of evolutionary theories of aging. Proc Natl Acad Sci U S A 99:14286–14291.

Hutchin, T., Navarro-Coy, N., Camp, G. V., Tiranti, V., Zeviani, M., Schuelke, M., Jaksch, M., Newton, V. & Mueller, R. (2001) Multiple origins of the mtDNA 7472insC mutation associated with hearing loss and neurological dysfunction. Eur J Hum Genet 9:385–387.

Ihaka, R. & Gentleman, R. (1996) R: a language for data analysis and graphics. J Comp Graph Stat 5:299–314.

Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713.

Innan, H. & Nordborg, M. (2002) Recombination or mutational hot spots in human mtDNA? Mol Biol Evol 19:1122–1127.

Jansen, R. & de Boer, K. (1998) The bottleneck: mitochondrial imperatives in oogenesis and ovarian follicular fate. Mol Cell Endocrinol 145:81–88.

Jenuth, J., Peterson, A., Fu, K. & Shoubridge, E. (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet 14:146–151.

Jenuth, J., Peterson, A. & Shoubridge, E. (1997) Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. Nat Genet 16:93–95.

John, J. S., Sakkas, D., Dimitriadi, K., Barnes, A., Maclin, V., Ramey, J., Barratt, C. & Jonge, C. D. (2000) Failure of elimination of paternal mitochondrial DNA in abnormal embryos. Lancet 355:200.

Johns, D. (1995) Mitochondrial DNA and disease. New Engl J Med 333:638–644.

Johnson, M., Wallace, D., Ferris, S., Rattazzi, M. & Cavalli-Sforza, L. (1983) Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol 19:255–271.

Jorde, L., Watkins, W., Bamshad, M., Dixon, M., Ricker, C., Seielstad, M. & Batzer, M. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66:979–988.

Kadowaki, T., Kadowaki, H., Mori, Y., Tobe, K., Sakuta, R., Suzuki, Y., Tanabe, Y., Sakura, H., Awata, T. & Goto, Y. (1994) A subtype of diabetes mellitus associated with a mutation of mitochondrial DNA. N Engl J Med 330:962–968.

Kahn, H. & Sempos, C. (1989) Statistical methods in epidemiology. Oxford University Press, New York.

Kang, D. & Hamasaki, N. (2002) Maintenance of mitochondrial DNA integrity: repair and degradation. Curr Genet 41:311–322.

Kato, T., Kunugi, H., Nanko, S. & Kato, N. (2000) Association of bipolar disorder with the 5178 polymorphism in mitochondrial DNA. Am J Med Genet 96:182–186.

Ke, Y., Su, B., Song, X., Lu, D., Chen, L., Li, H., Qi, C., Marzuki, S., Deka, R., Underhill, P., Xiao, C., Shriver, M., Lell, J., Wallace, D., Wells, R., Seielstad, M., Oefner, P., Zhu, D., Jin, J., Huang, W., Chakraborty, R., Chen, Z. & Jin, L. (2001) African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. Science 292:1151–1153.

Khaidakov, M., Heflich, R., Manjanatha, M., Myers, M. & Aidoo, A. (2003) Accumulation of point mutations in mitochondrial DNA of aging mice. Mutat Res 526:1–7.

Khrapko, K., Nekhaeva, E., Kraytsberg, Y. & Kunz, W. (2003) Clonal expansions of mitochondrial genomes: implications for in vivo mutational spectra. Mutat Res 522:13–19.

Kim, J., Park, K., Cho, Y., Kang, B., Kim, S., Jeon, H., Kim, S. & Lee, H. (2002) The prevalence of the mitochondrial DNA 16189 variant in non-diabetic Korean adults and its association with higher fasting glucose and body mass index. Diabet Med 19:681–684.

Kim, J., He, L. & Lemasters, J. (2003) Mitochondrial permeability transition: a common pathway to necrosis and apoptosis. Biochem Biophys Res Commun 304:463–470.

Kimura, M. (1968) Evolutionary rate at the molecular level. Nature 217:624–626.

Knight, R., Landweber, L. & Yarus, M. (2001) How mitochondria redefine the code. J Mol Evol 53:299–313.

Kobayashi, Y., Momoi, M., Tominaga, K., Momoi, T., Nihei, K., Yanagisawa, M., Kagawa, Y. & Ohta, S. (1990) A point mutation in the mitochondrial tRNA(Leu)(UUR) gene in MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes). Biochem Biophys Res Commun 173:816–822.

Koga, Y., Davidson, M., Schon, E. & King, M. (1995) Analysis of cybrids harboring MELAS mutations in the mitochondrial tRNA(Leu(UUR)) gene. Muscle Nerve 3:119–123.

Kreitman, M. (2000) Methods to detect selection in populations with applications to the human. Annu Rev Genomics Hum Genet 1:539–559.

Kärppä, M., Syrjälä, P., Tolonen, U. & Majamaa, K. (2003) Peripheral neuropathy in patients with the 3243A>G mutation in mitochondrial DNA. J Neurol 250:216–221.

Kyte, J. & Doolittle, R. (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132.

Lahermo, P., Sajantila, A., Sistonen, P., Lukka, M., Aula, P., Peltonen, L. & Savontaus, M. (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. Am J Hum Genet 58:1309–1322.

Lahermo, P., Savontaus, M., Sistonen, P., Beres, J., de Knijff, P., Aula, P. & Sajantila, A. (1999) Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. Eur J Hum Genet 7:447–458.

Lamminen, T., Huoponen, K., Sistonen, P., Juvonen, V., Lahermo, P., Aula, P., Nikoskelainen, E. & Savontaus, M. (1997) mtDNA haplotype analysis in Finnish families with Leber hereditary optic neuroretinopathy. Eur J Hum Genet 5:271–279.

Larsson, N., Wang, J., Wilhelmsson, H., Oldfors, A., Rustin, P., Lewandoski, M., Barsh, G. & Clayton, D. (1998) Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. Nat Genet 18:231–236.

Lee, S., Lee, R., Fraser, A., Kamath, R., Ahringer, J. & Ruvkun, G. (2003) A systematic RNAi screen identifies a critical role for mitochondria in C. elegans longevity. Nat Genet 33:40–48.

Lehtonen, M., Moilanen, J. & Majamaa, K. (2003) Increased variation in mtDNA in patients with familial sensorineural hearing hearing impairment. Hum Genet 113:220–227.

Lell, J. & Wallace, D. (2000) The peopling of Europe from the maternal and paternal perspectives. Am J Hum Genet 67:1376–1381.

Lertrit, P., Kapsa, R., Jean-Francois, M., Thyagarajan, D., Noer, A., Marzuki, S. & Byrne, E. (1994) Mitochondrial DNA polymorphism in disease: a possible contributor to respiratory dysfunction. Hum Mol Genet 3:1973–1981.

Li, W.-H. (1997) Molecular evolution. Sinauer Associates, Sunderland, Massachusetts.

Lin, M., Simon, D., Ahn, C., Kim, L. & Beal, M. (2002) High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. Hum Mol Genet 11:133–145.

Loeffen, J., Smeitink, J., Triepels, R., Smeets, R., Schuelke, M., Sengers, R., Trijbels, F., Hamel, B., Mullaart, R. & van den Heuvel, L. (1998) The first nuclear-encoded complex I mutation in a patient with Leigh syndrome. Am J Hum Genet 63:1598–1608.

Loeffen, J., Elpeleg, O., Smeitink, J., Smeets, R., Stockler-Ipsiroglu, S., Mandel, H., Sengers, R., Trijbels, F. & van den Heuvel, L. (2001) Mutations in the complex I NDUFS2 gene of patients with cardiomyopathy and encephalomyopathy. Ann Neurol 49:195–201.

Maassen, J. (2002) Mitochondrial diabetes: pathophysiology, clinical presentation, and genetic analysis. Am J Med Genet 115:66–70.

Maassen, J., Biberoglu, S., 't Hart, L., Bakker, E. & de Knijff, P. (2002) A case of a de novo A3243G mutation in mitochondrial DNA in a patient with diabetes and deafness. Arch Physiol Biochem 110:186–188.

Maca-Meyer, N., Larruga, J. & Cabrera, V. (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13.

Majewski, J. & Ott, J. (2003) Amino acid substitutions in the human genome: evolutionary implications of single nucleotide polymorphisms. Gene 305:167–173.

Malyarchuk, B., Rogozin, I., Berikov, V. & Derenko, M. (2002) Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. Hum Genet 111:46–53.

Man, P., Turnbull, D. & Chinnery, P. (2002) Leber hereditary optic neuropathy. J Med Genet 39:162–169.

Man, P., Griffiths, P., Brown, D., Howell, N., Turnbull, D. & Chinnery, P. (2003) The epidemiology of Leber hereditary optic neuropathy in the North East of England. Am J Hum Genet 72:333–339.

Mandavilli, B., Santos, J. & Houten, B. V. (2002) Mitochondrial DNA repair and aging. Mutat Res 509:127–151.

Manfredi, G., Thyagarajan, D., Papadopoulou, L., Pallotti, F. & Schon, E. (1997) The fate of human sperm-derived mtDNA in somatic cells. Am J Hum Genet 61:953–960.

Marchington, D., Macaulay, V., Hartshorne, G., Barlow, D. & Poulton, J. (1998) Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. Am J Hum Genet 63:769–775.

Marchington, D., Brown, M. S., Lamb, V., van Golde, R., Kremer, J., Tuerlings, J., Mariman, E., Balen, A. & Poulton, J. (2002) No evidence for paternal mtDNA transmission to offspring or extra-embryonic tissues after ICSI. Mol Hum Reprod 8:1046–1049.

Mariotti, C., Savarese, N., Suomalainen, A., Rimoldi, M., Comi, G., Prelle, A., Antozzi, C., Servidei, S., Jarre, L. & DiDonato, S. (1995) Genotype to phenotype correlations in mitochondrial encephalomyopathies associated with the A3243G mutation of mitochondrial DNA. J Neurol 242:304–312.

Martin, G. (2002) The evolutionary substrate of aging. Arch Neurol 59:1702–1705.

Mason, P., Matheson, E., Hall, A. & Lightowlers, R. (2003) Mismatch repair activity in mammalian mitochondria. Nucleic Acids Res 31:1052–1058.

Mathiesen, C. & Hägerhäll, C. (2002) Transmembrane topology of the NuoL, M and N subunits of NADH:quinone oxidoreductase and their homologues among membrane-bound hydrogenases and bona fide antiporters. Biochim Biophys Acta 1556:121–132.

Matsuoka, T., Goto, Y., Yoneda, M. & Nonaka, I. (1991) Muscle histopathology in myoclonus epilepsy with ragged-red fibers (MERRF). J Neurol Sci 106:193–198.

Maynard Smith, J. (1994) Estimating selection by comparing synonymous and substitutional changes. J Mol Evol 39:123–128.

McDonald, J. & Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.

Meinilä, M., Finnilä, S. & Majamaa, K. (2001) Evidence for mtDNA admixture between the Finns and the Saami. Hum Hered 52:160–170.

Merriweather, D. & Kaestle, F. (1999) Mitochondrial recombination? (continued). Science 285:837.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M., Sukernik, R., Olckers, A. & Wallace, D. (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A 100:171–176.

Montero, M., Alonso, M., Albillos, A., Cuchillo-Ibáñez, I., Olivares, R., Villalobos, C. & Alvarez, J. (2002) Effect of inositol 1,4,5-triphosphate receptor stimulation on mitochondrial [Ca^{2+}] and secretion in cromaffin cells. Biochem J 365:451–459.

Moraes, C., Ciacci, F., Bonilla, E., Jansen, C., Hirano, M., Rao, N., Lovelace, R., Rowland, L., Schon, E. & DiMauro, S. (1993) Two novel pathogenic mitochondrial DNA mutations affecting organelle number and protein synthesis. Is the tRNA(Leu(UUR)) gene an etiologic hot spot? J Clin Invest 92:2906–2915.

Morgan-Hughes, J., Sweeney, M., Cooper, J., Hammans, S., Brockington, M., Schapira, A., Harding, A. & Clark, J. (1995) Mitochondrial DNA (mtDNA) diseases: correlation of genotype to phenotype. Biochim Biophys Acta 1271:135–140.

Morris, A. & Lightowlers, R. (2000) Mitochondrial DNA recombination. Lancet 356:941.

Morton, N. (1982) Outline of genetic epidemiology. Karger, Basel.

Muro, C., Grigoriev, S., Pietkiewicz, D., Kinnally, K. & Campo, M. (2003) Comparison of the TIM and TOM channel activities of the mitochondrial protein import complexes. Biophys J 84:2981–2989.

Myres, J., Malan, M., Shumway, J., Rowe, M., Amon, E. & Woodward, S. (2000) Haplogroup-associated differences in neonatal death and incidence of low birth weight at elevation: a preliminary assessment. Am J Obstet Gynecol 182:1599–1605.

Nachman, M. (1998) Deleterious mutations in animal mitochondrial DNA. Genetica 102-103:61–69.

Naylor, G., Collins, T. & Brown, W. (1995) Hydrophobicity and phylogeny. Nature 373:565–566.

Nedbal, M. & Flynn, J. (1998) Do the combined effects of the asymmetric process of replication and DNA damage from oxygen radicals produce a mutation-rate signature in the mitochondrial genome? Mol Biol Evol 15:219–223.

Nei, M. (1987) Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M. & Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426.

Newell, C. (1994) Methods and models in demography. John Wiley & Sons, Chichester.

Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. Heredity 86:641–647.

Nielsen, R. & Weinreich, D. (1999) The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly deleterious theory. Genetics 153:497–506.

Niemann, S. & Muller, U. (2000) Mutations in SDHC cause autosomal dominant paraganglioma, type 3. Nat Genet 26:268–270.

Niemi, A., Hervonen, A., Hurme, M., Karhunen, P., Jylhä, M. & Majamaa, K. (2003) Mitochondrial DNA polymorphisms associated with longevity in a Finnish population. Hum Genet 112:29–33.

Nishino, I., Spinazzola, A. & Hirano, M. (1999) Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. Science 283:689–692.

Norio, R. (2003a) Finnish Disease Heritage I: characteristics, causes, background. Hum Genet 112:441–456.

Norio, R. (2003b) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. Hum Genet 112:457–469.

Ohta, T. (2002) Near-neutrality in evolution of genes and gene regulation. Proc Natl Acad Sci U S A 99:16134–16137.

Ojaimi, J., Pan, J., Santra, S., Snell, W. & Schon, E. (2002) An algal nucleus-encoded subunit of mitochondrial ATP synthase rescues a defect in the analogous human mitochondrial-encoded subunit. Mol Biol Cell 13:3836–3844.

Opdal, S., Vege, A., Egeland, T., Musse, M. & Rognum, T. (2002) Possible role of mtDNA mutations in sudden infant death. Pediatr Neurol 27:23–29.

Papadopoulou, L., Sue, C., Davidson, M., Tanji, K., Nishino, I., Sadlock, J., Krishna, S., Walker, W., Selby, J., Glerum, D., Coster, R., Lyon, G., Scalais, E., Lebel, R., Kaplan, P., Shanske, S., Vivo, D. D., Bonilla, E., Hirano, M., DiMauro, S. & Schon, E. (1999) Fatal infantile cardioencephalomyopathy with COX deficiency and mutations in SCO2, a COX assembly gene. Nat Genet 23:333–337.

Park, H., Davidson, E. & King, M. (2003) The pathogenic A3243G mutation in human mitochondrial tRNALeu(UUR) decreases the efficiency of aminoacylation. Biochemistry 42:958–964.

Park, M., Ashby, M., Erdemli, G., Petersen, O. & Tepikin, A. (2001) Perinuclear, perigranular and sub-plasmalemmal mitochondria have distinct functions in the regulation of cellular calcium transport. EMBO J 20:1863–1874.

Parsons, T., Muniec, D., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M., Berry, D., Holland, K., Weedn, V., Gill, P. & Holland, M. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368.

Pavlakis, S., Phillips, P., DiMauro, S., Vivo, D. D. & Rowland, L. (1984) Mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes: a distinctive clinical syndrome. Ann Neurol 16:481–488.

Pesole, G., Gissi, C., Chirico, A. D. & Saccone, C. (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. J Mol Evol 48:427–434.

Pitkänen, S. & Robinson, B. (1996) 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, A., Scott-Brown, M., Thompson, C., Macaulay, V. & Simmons, D. (2002) The presence of a common mitochondrial DNA variant is associated with fasting insulin levels in Europeans in Auckland. Diabet Med 19:969–971.

Przeworski, M., Hudson, R. & Rienzo, A. D. (2000) Adjusting the focus on human variation. Trends Genet 16:296–302.

Qi, X., Lewin, A., Hauswirth, W. & Guy, J. (2003) Suppression of complex I gene expression induces optic neuropathy. Ann Neurol 53:198–205.

Rahman, S., Taanman, J., Cooper, J., Nelson, I., Hargreaves, I., Meunier, B., Hanna, M., Garcia, J., Capaldi, R., Lake, B., Leonard, J. & Schapira, A. (1999) A missense mutation of cytochrome oxidase subunit II causes defective assembly and myopathy. Am J Hum Genet 65:1030–1039.

Raitio, M., Lindroos, K., Laukkanen, M., Pastinen, T., Sistonen, P., Sajantila, A. & Syvänen, A. (2001) Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. Genome Res 11:471–482.

Rand, D. (2001) Mitochondrial genomics flies high. Trends Ecol Evol 16:2–4.

Rand, D. & Kann, L. (1998) Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica 102-103:393–407.

Reichert, A. & Neupert, W. (2002) Contact sites between the outer and inner membrane of mitochondria-role in protein transport. Biochim Biophys Acta 1592:41–49.

Remes, A., Kärppä, M., Karvonen, S., Moilanen, J., Uimonen, S., Salmela, P., Rusanen, H., Sorri, M., Hassinen, I. & Majamaa, K. (2003) Epidemiology of the mitochondrial DNA 8344A>G mutation for myoclonus epilepsy and ragged red fibers (MERRF) syndrome. J Neurol Neurosurg Psychiatry 74:1158–1159.

Rice, P., Longden, I. & Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16:276–277.

Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H. & Sykes, B. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C. & Stefanescu, G. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276.

Rinaldi, T., Gambadoro, A., Francisci, S. & Frontali, L. (2003) Nucleo-mitochondrial interactions in Saccharomyces cerevisiae: characterization of a nuclear gene suppressing a defect in mitochondrial tRNA(Asp) processing. Gene 303:63–68.

Riva, A., Loffredo, F., Uccheddu, A., Riva, F. & Tandler, B. (2003) Mitochondria of human adrenal cortex have tubular cristae with bulbous tips. J Clin Endocrinol Metab 88:1903–1906.

Robertson, J. & Orrenius, S. (2002) Role of mitochondria in toxic cell death. Toxicology 181-182:491–496.

Rocher, C., Letellier, T., Copeland, W. & Lestienne, P. (2002) Base composition at mtDNA boundaries suggests a DNA triple helix model for human mitochondrial DNA large-scale rearrangements. Mol Genet Metab 76:123–132.

Rodeheffer, M. (2003) Multiple interactions involving the amino-terminal domain of yeast mtRNA polymerase determine the efficiency of mitochondrial protein synthesis. J Biol Chem 278:18695–18701.

Roger, A., Svard, S., Tovar, J., Clark, C., Smith, M., Gillin, F. & Sogin, M. (1998) A mitochondrial-like chaperonin 60 gene in Giardia lamblia: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci U S A 95:229–234.

Rose, G., Passarino, G., Carrieri, G., Altomare, K., Greco, V., Bertolini, S., Bonafe, M., Franceschi, C. & Benedictis, G. D. (2001) Paradoxes in longevity: sequence analysis of mtDNA haplogroup J in centenarians. Eur J Hum Genet 9:701–707.

Rose, G., Passarino, G., Franceschi, C. & Benedictis, G. D. (2002) The variability of the mitochondrial genome in human aging: a key for life and death? Int J Biochem Cell Biol 34:1449–1460.

Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J. & Letellier, T. (2003) Mitochondrial threshold effects. Biochem J 370:751–762.

Rozas, J. & Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175.

Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., Issel-Tarver, L., DiRienzo, A. & Pääbo, S. (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Res 5:42–52.

Sajantila, A., Salem, A., Savolainen, P., Bauer, K., Gierig, C. & Pääbo, S. (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci U S A 93:12035–12039.

Salas, A., Richards, M., la Fe, T. D., Lareu, M., Sobrino, B., Sanchez-Diz, P., Macaulay, V. & Carracedo, A. (2002) The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111.

Santel, A., Gaume, B. & Youle, R. (2003) Mitofusin-1 protein is a generally expressed mediator of mitochondrial fusion in mammalian cells. J Cell Sci 116:2763–2774.

Saraste, M. (1999) Oxidative phosphorylation at the fin de siecle. Science 283:1488–1493.

SAS Institute (1997) SAS release 6.12. Cary, NC.

Sasarman, F., Karpati, G. & Shoubridge, E. (2002) Nuclear genetic control of mitochondrial translation in skeletal muscle revealed in patients with mitochondrial myopathy. Hum Mol Genet 11:1669–1681.

Scarpulla, R. (2002) Nuclear activators and coactivators in mammalian mitochondrial biogenesis. Biochim Biophys Acta 1576:1–14.

Schmidt, H., Strimmer, K., Vingron, M. & von Haeseler, A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504.

Schuelke, M., Smeitink, J., Mariman, E., Loeffen, J., Plecko, B., Trijbels, F., Stockler-Ipsiroglu, S. & van den Heuvel, L. (1999) Mutant NDUFV1 subunit of mitochondrial complex I causes leukodystrophy and myoclonic epilepsy. Nat Genet 21:260–261.

Schwartz, M. & Vissing, J. (2002) Paternal inheritance of mitochondrial DNA. N Engl J Med 347:576–580.

Shanske, S., Tang, Y., Hirano, M., Nishigaki, Y., Tanji, K., Bonilla, E., Sue, C., Krishna, S., Carlo, J., Willner, J., Schon, E. & DiMauro, S. (2002) Identical mitochondrial DNA deletion in a woman with ocular myopathy and in her son with Pearson syndrome. Am J Hum Genet 71:679–683.

Shin, W., Tanaka, M., Suzuki, J., Hemmi, C. & Toyo-oka, T. (2000) A novel homoplasmic mutation in mtDNA with a single evolutionary origin as a risk factor for cardiomyopathy. Am J Hum Genet 67:1617–1620.

Shoffner, J. (1996) Maternal inheritance and the evaluation of oxidative phosphorylation diseases. Lancet 348:1283–1288.

Shoffner, J. & Wallace, D. (1992) Mitochondrial genetics: principles and practice. Am J Hum Genet 51:1179–1186.

Shoffner, J., Lott, M., Lezza, A., Seibel, P., Ballinger, S. & Wallace, D. (1990) Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA(Lys) mutation. Cell 61:931–937.

Shoubridge, E. (2002) The ABCs of mitochondrial transcription. Nat Genet 31:227–228.

Silva, W., Bonatto, S., Holanda, A., Ribeiro-Dos-Santos, A., Paixao, B., Goldman, G., Abe-Sandes, K., Rodriguez-Delfin, L., Barbosa, M., Paco-Larson, M., Petzl-Erler, M., Valente, V., Santos, S. & Zago, M. (2002) Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. Am J Hum Genet 71:187–192.

Simonsen, K., Churchill, G. & Aquadro, C. (1995) Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141:413–429.

Simpson, A. & Roger, A. (2002) Eukaryotic evolution: getting to the root of the problem. Curr Biol 12:691–693.

Smeitink, J. & van den Heuvel, L. (1999) Human mitochondrial complex I in health and disease. Am J Hum Genet 64:1505–1510.

Spelbrink, J., Li, F., Tiranti, V., Nikali, K., Yuan, Q., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H., Zeviani, M. & Larsson, C. (2001) Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. Nat Genet 28:223–231.

Stajich, J., Block, D., Boulez, K., Brenner, S., Chervitz, S., Dagdigian, C., Fuellen, G., Gilbert, J., Korf, I., Lapp, H., Lehväslaiho, H., Matsalla, C., Mungall, C., Osborne, B., Pocock, M., Schattner, P., Senger, M., Stein, L., Stupka, E., Wilkinson, M. & Birney, E. (2002) The Bioperl toolkit: Perl modules for the life sciences. Genome Res 12:1611–1618.

Starikovskaya, Y., Sukernik, R., Schurr, T., Kogelnik, A. & Wallace, D. (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. Am J Hum Genet 63:1473–1491.

Stevnsner, T., Thorslund, T., de Souza-Pinto, N. & Bohr, V. (2002) Mitochondrial repair of 8-oxoguanine and changes with aging. Exp Gerontol 37:1189–1196.

Stuart, R. (2002) Insertion of proteins into the inner membrane of mitochondria: the role of the Oxa1 complex. Biochim Biophys Acta 1592:79–87.

Sudoyo, H., Suryadi, H., Lertrit, P., Pramoonjago, P., Lyrawati, D. & Marzuki, S. (2002) Asian-specific mtDNA backgrounds associated with the primary G11778A mutation of Leber's hereditary optic neuropathy. J Hum Genet 47:594–604.

Sutovsky, P., Moreno, R., Ramalho-Santos, J., Dominko, T., Simerly, C. & Schatten, G. (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. Biol Reprod 63:582–590.

Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tanaka, M. (2002) Mitochondrial genotypes and cytochrome b variants associated with longevity or Parkinson's disease. J Neurol 249 Suppl 2:11–18.

Tanaka, M., Fuku, N., Takeyasu, T., Guo, L., Hirose, R., Kurata, M., Borgeld, H., Yamada, Y., Maruyama, W., Arai, Y., Hirose, N., Oshida, Y., Sato, Y., Hattori, N., Mizuno, Y., Iwata, S. & Yagi, K. (2002) Golden mean to longevity: rareness of mitochondrial cytochrome b variants in centenarians but not in patients with Parkinson's disease. J Neurosci Res 70:347–355.

Templeton, A. (1996) Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. Genetics 144:1263–1270.

Tielens, A., Rotte, C., van Hellemond, J. & Martin, W. (2002) Mitochondria as we don't know them. Trends Biochem Sci 27:564–572.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. & Wallace, D. (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850.

Torroni, A., Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P., Negri, A. D. & Scozzari, R. (1997) 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 60:1107–1121.

Torroni, A., Campos, Y., Rengo, C., Sellitto, D., Achilli, A., Magri, C., Semino, O., Garcia, 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.

Tourasse, N. & Li, W. (2000) Selective constraints, amino acid composition, and the rate of protein evolution. Mol Biol Evol 17:656–664.

Tourmen, Y., Baris, O., Dessen, P., Jacques, C., Malthiery, Y. & Reynier, P. (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics 80:71–77.

Triepels, R., Heuvel, L. V. D., Trijbels, J. & Smeitink, J. (2001) Respiratory chain complex I deficiency. Am J Med Genet 106:37–45.

Truscott, K., Brandner, K. & Pfanner, N. (2003) Mechanisms of protein import into mitochondria. Curr Biol 13:326–337.

Turner, C., Killoran, C., Thomas, N., Rosenberg, M., Chuzhanova, N., Johnston, J., Kemel, Y., Cooper, D. & Biesecker, L. (2003) Human genetic disease caused by de novo mitochondrial-nuclear DNA transfer. Hum Genet 112:303–309.

Uimonen, S., Moilanen, J., Sorri, M., Hassinen, I. & Majamaa, K. (2001) Hearing impairment in patients with 3243A>G mtDNA mutation: phenotype and rate of progression. Hum Genet 108:284–289.

Uusimaa, J., Remes, A., Rantala, H., Vainionpää, L., Herva, R., Vuopala, K., Nuutinen, M., Majamaa, K. & Hassinen, I. (2000) Childhood encephalopathies and myopathies: a prospective study in a defined population to assess the frequency of mitochondrial disorders. Pediatrics 105:598–603.

Valnot, I., Osmond, S., Gigarel, N., Mehaye, B., Amiel, J., Cormier-Daire, V., Munnich, A., Bonnefont, J., Rustin, P. & Rotig, A. (2000a) Mutations of the SCO1 gene in mitochondrial cytochrome c oxidase deficiency with neonatal-onset hepatic failure and encephalopathy. Am J Hum Genet 67:1104–1109.

Valnot, I., von Kleist-Retzow, J., Barrientos, A., Gorbatyuk, M., Taanman, J., Mehaye, B., Rustin, P., Tzagoloff, A., Munnich, A. & Rotig, A. (2000b) A mutation in the human heme A:farnesyltransferase gene (COX10) causes cytochrome c oxidase deficiency. Hum Mol Genet 9:1245–1249.

van den Heuvel, L., Ruitenbeek, W., Smeets, R., Gelman-Kohan, Z., Elpeleg, O., Loeffen, J., Trijbels, F., Mariman, E., de Bruijn, D. & Smeitink, J. (1998) Demonstration of a new pathogenic mutation in human complex I deficiency: a 5-bp duplication in the nuclear gene encoding the 18-kD (AQDQ) subunit. Am J Hum Genet 62:262–268.

van Gurp, M., Festjens, N., van Loo, G., Saelens, X. & Vandenabeele, P. (2003) Mitochondrial intermembrane proteins in cell death. Biochem Biophys Res Commun 304:487–497.

Vellai, T. & Vida, G. (1999) The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. Proc R Soc Lond B Biol Sci 266:1571–1577.

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

Vilkki, J., Savontaus, M. & Nikoskelainen, E. (1988) Human mitochondrial DNA types in Finland. Hum Genet 80:317–321.

Vissing, J., Ravn, K., Danielsen, E., Duno, M., Wibrand, F., Wevers, R. & Schwartz, M. (2002) Multiple mtDNA deletions with features of MNGIE. Neurology 59:926–929.

Wallace, D., Ye, J., Neckelmann, S., Singh, G., Webster, K. & Greenberg, B. (1987) Sequence analysis of cDNAs for the human and bovine ATP synthase β subunit: mitochondrial DNA genes sustain seventeen times more mutations. Curr Genet 12:81–90.

Wallace, D., Brown, M. & Lott, M. (1999) Mitochondrial DNA variation in human evolution and disease. Gene 238:211–230.

Walt, J. V. D., Nicodemus, K., Martin, E., Scott, W., Nance, M., Watts, R., Hubble, J., Haines, J., Koller, W., Lyons, K., Pahwa, R., Stern, M., Colcher, A., Hiner, B., Jankovic, J., Ondo, W., FH, J. A., Goetz, C., Small, G., Mastaglia, F., Stajich, J., McLaurin, A., Middleton, L., Scott, B., Schmechel, D., Pericak-Vance, M. & Vance, J. (2003) Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. Am J Hum Genet 72:804–811.

Watson, B., Khan, M., Desmond, R. & Bergman, S. (2001) Mitochondrial DNA mutations in black Americans with hypertension-associated end-stage renal disease. Am J Kidney Dis 38:529–536.

Wayne, M. & Simonsen, K. (1998) Statistical tests of neutrality in the age of weak selection. Trends Ecol Evol 13:236–240.

Weinreich, D. & Rand, D. (2000) Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics 156:385–399.

Welle, S., Bhatt, K., Shah, B., Needler, N., Delehanty, J. & Thornton, C. (2003) Reduced amount of mitochondrial DNA in aged human muscle. J Appl Physiol 94:1479–1484.

Wilkinson-Herbots, H., Richards, M., Forster, P. & Sykes, B. (1996) Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. Ann Hum Genet 60:499–508.

Williams, E. & Hurst, L. (2002) Is the synonymous substitution rate in mammals gene-specific? Mol Biol Evol 19:1395–1398.

Woischnik, M. & Moraes, C. (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res 12:885–893.

Wolfe, K. & Li, W. (2003) Molecular evolution meets the genomics revolution. Nat Genet 33 Suppl:255–265.

Wong, A., Cavelier, L., Collins-Schramm, H., Seldin, M., McGrogan, M., Savontaus, M. & Cortopassi, G. (2002) Differentiation-specific effects of LHON mutations introduced into neuronal NT2 cells. Hum Mol Genet 11:431–438.

Yang, M., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H. & Holt, I. (2002) Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. Cell 111:495–505.

Yasukawa, T., Suzuki, T., Ueda, T., Ohta, S. & Watanabe, K. (2000) Modification defect at anticodon wobble nucleotide of mitochondrial tRNAs(Leu)(UUR) with pathogenic mutations of mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes. J Biol Chem 275:4251–4257.

Zhu, Z., Yao, J., Johns, T., Fu, K., Bie, I. D., Macmillan, C., Cuthbert, A., Newbold, R., Wang, J., Chevrette, M., Brown, G., Brown, R. & Shoubridge, E. (1998) SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. Nat Genet 20:337–343.