Review

Author for correspondence:
Douglas C. Wallace
e-mail: wallaced1@email.chop.edu

Bioenergetics in human evolution and disease: implications for the origins of biological complexity and the missing genetic variation of common diseases

Douglas C. Wallace

Center of Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Colket Translational Research Building, Room 6060, 3501 Civic Center Boulevard, Philadelphia, PA 19104-4302, USA

Two major inconsistencies exist in the current neo-Darwinian evolutionary theory that random chromosomal mutations acted on by natural selection generate new species. First, natural selection does not require the evolution of ever increasing complexity, yet this is the hallmark of biology. Second, human chromosomal DNA sequence variation is predominantly either neutral or deleterious and is insufficient to provide the variation required for speciation or for predilection to common diseases. Complexity is explained by the continuous flow of energy through the biosphere that drives the accumulation of nucleic acids and information. Information then encodes complex forms. In animals, energy flow is primarily mediated by mitochondria whose maternally inherited mitochondrial DNA (mtDNA) codes for key genes for energy metabolism. In mammals, the mtDNA has a very high mutation rate, but the deleterious mutations are removed by an ovarian selection system. Hence, new mutations that subtly alter energy metabolism are continuously introduced into the species, permitting adaptation to regional differences in energy environments. Therefore, the most phenotypically significant gene variants arise in the mtDNA, are regional, and permit animals to occupy peripheral energy environments where rarer nuclear DNA (nDNA) variants can accumulate, leading to speciation. The neutralist–selectionist debate is then a consequence of mammals having two different evolutionary strategies: a fast mtDNA strategy for intra-specific radiation and a slow nDNA strategy for speciation. Furthermore, the missing genetic variation for common human diseases is primarily mtDNA variation plus regional nDNA variants, both of which have been missed by large, inter-population association studies.

1. Evolution and energetics

In the mid-nineteenth century, Charles Darwin and Alfred Russel Wallace proposed that species arose through natural selection, thus accounting for how one species can split into two similar species occupying adjacent niches. Darwin & Wallace [1,2] elaborated on this theory to explain many important biological questions. However, Wallace was concerned that natural selection was insufficient to explain the origins of consciousness and the human brain. At the time and even today, the primary debate is whether biology is a science that can be understood by natural principles. Hence, Wallace's concerns have been minimized by the natural scientists.

With the maturation of physical science theories—in particular, thermodynamics—an inconsistency arose. Thermodynamics taught that, in a closed system, entropy increases and complexity decays. Therefore, the stable existence of complex living organisms seemed to defy physics. This dilemma was partially

resolved by Schrödinger, who pointed out that living organisms were not closed systems, but rather acquire energy from their environment as what he called 'negative entropy' [3]. The concept that the flow of energy through a system can generate and sustain complex structures is the domain of non-equilibrium thermodynamics that is now understood to be central to life, the reason why we eat and breathe [4–6].

Still, energy flow only explains the maintenance of complexity. It does not necessitate the development of complex forms. Resolution of this dilemma comes from information theory. The discovery of the structure of DNA revealed that biology is primarily about the storage and retrieval of information. Within a narrow thermal range, such as exists on Earth, the flow of energy through organic systems generates ordered structures. One of these structures is nucleic acids, which can accumulate information. Each year, the flow of energy through the biosphere from sunlight on the Earth's surface or from geothermal vents on the ocean floor provides the energy to generate more nucleic acids, and the more nucleic acids the more information, the more information the greater the complexity [7]. Originally, energy flow created nucleic acids directly, though inefficiently [6,8]. With the advent of intra-cellular DNA replication, however, nucleic acids accumulated rapidly through cell proliferation. Hence, the complexity of modern organisms is a consequence of four billion years of energy flow and the resulting information accumulation, making biological information stored energy. This implies that one of the more important actions of natural selection is enrichment for the more energy-efficient individuals among organisms attempting to exploit the same energy resource [7].

2. Evolution and Mendelian genetics

In the later part of the nineteenth century, Gregor Mendel outlined the rules of inheritance for sexually reproducing organisms. Each parent provides one copy of each gene to an offspring through fusion of the male and female gametes, generating a 'diploid' individual. At sexual maturity, the individual separates the two gene copies into his/her sex cells in preparation for conception of the next generation.

Subsequently, it was discovered that the behaviour of the genes corresponded to the behaviour of the chromosomes, and later that chromosomes packaged DNA. Since nucleic acids can replicate and in the process mutate, this led to the concept that mutations in the nDNA generate the variation that is acted on by natural selection to create organismal diversity. This neo-Darwinian synthesis implied that a significant proportion of the nDNA variation must be functional and, in the right environment, beneficial.

This concept stood until the 1960s, when molecular genetic studies on human nDNA revealed that much of the genetic variation that differentiated human populations is due to differences in the frequency of alleles common to both populations. Furthermore, the differences in allelic frequencies between populations could be explained primarily by statistical fluctuations [9–11]. This led Kimura to propose that virtually all extant chromosomal genetic variation was neutral, because the vast majority of functional mutations would be deleterious and removed by purifying selection [12]. This 'neutralist' hypothesis precipitated the neutralist–selectionist debate. If all intra-specific genetic variation was neutral, where was the functional variation that could permit individuals to adapt to environmental changes and ultimately give rise to new species?

Since before the time of Darwin and Wallace, species have been defined primarily by anatomical differences and all anatomical traits are coded by nDNA genes. Consequently, analysis of nDNA variation has been highly informative in understanding the progressive changes in anatomical traits that occur during speciation. By contrast, so much of human nDNA variation could be explained by stochastic processes that it soon became dogma that all intra-specific genetic differences were the result of stochastic processes such as genetic drift and founder effects.

However, humans are a single species and anatomy does not vary markedly within a species. Therefore, intra-specific variation must affect functions other than anatomical traits. Since energy is also fundamental to the life process, bioenergetic changes could be the source of intra-specific variation.

A new opportunity arose to find adaptive variation within the human nDNA as a consequence of the Human Genome Project. The screening of multiple human nDNAs permitted the identification of millions of single nucleotide polymorphisms (SNPs). These SNPs were then used to screen populations to identify chromosomal regions that are linked (in linkage disequilibrium) with loci that alter the risk of developing metabolic diseases. Major human metabolic phenotypes include diabetes and obesity and these clinical manifestations are directly related to individual responses to environmental differences in energy resources. In fact, many important environmental differences are related to the type and availability of calories and demands for use of those calories for tissue maintenance, physical work, combating infections, reproduction and to cope with environmental limitations such as oxygen deprivation and toxins. The aggregate of all of these factors will be referred to here as the energetic environment. Since the energetic environment within a species' niche can change, chromosomal locus variants that are beneficial at one time in one energetic environment might become deleterious in another. Therefore, chromosomal loci related to diabetes and obesity should provide insight into the genetics of energy metabolism.

Genome Wide Association Studies (GWAS) have identified 63 chromosomal loci associated with type 2 diabetes. However, all of these loci have very weak phenotypic effects, so in aggregate they account for only 5.7 per cent of the variance in disease susceptibility. Simulation studies have suggested that an additional 488 loci may contribute to type 2 diabetes risk, but again the aggregated effect of all such loci would still explain only 10.7 per cent of risk. Additional projections suggest that if it were technically possible to identify them, perhaps approximately 49 per cent of the risk variance might be explained by common variants with low phenotype effect [13]. Taking into account body mass index, additional insulin resistance loci have been identified [14], but the chromosomal variants found by GWAS still fall far short of accounting for the approximately 3.5-fold increased risk faced by first degree relatives of diabetes patients [15].

Since the GWAS study design relies on linked DNA polymorphisms to identify disease risk loci, it requires that one or more SNPs be stably linked to the functional locus. Because the observed loci have weak phenotypic effects, large numbers of samples have been required to achieve statistical

significance. To obtain sufficient numbers, successful studies have combined the data from multiple geographically dispersed populations. This research design requires that the DNA polymorphism must have become associated with the functional locus early in human radiation so that the resulting linkage unit could become dispersed throughout global populations. Given that severely deleterious mutants would be eliminated by purifying selection, this research design necessitates that only the variants with the most modest phenotypic effects would survive long enough to remain associated within a single linkage group and become sufficiently dispersed throughout human populations to be detected by GWAS studies.

Since loci with large phenotypic effects would be rapidly eliminated by purifying selection, any extant high effect locus must be of recent origin. Having arisen recently, such large phenotypic effect loci must be confined to a single regional population and linked to a set of SNPs that are not associated with this functional locus anywhere else in the world. As a result, analysis of the relevant SNPs in a large multiple population study would dilute out any association between chromosomal SNPs and important regional phenotypic variants.

Such large phenotypic effect loci for diabetes and obesity loci have often been reported in regional association studies. One group of notable gene variants are those associated in the uncoupling protein (UCP) 1–3 genes [16–19]. However, these population-specific associations have often been dismissed as being 'non-replicateable' in other populations. But population-specific associations are precisely what would be expected for the most phenotypically significant gene variants.

One important type 2 diabetes locus is the peroxisome-proliferating-activated receptor γ (PPARγ) gene that has been associated with diabetes in GWAS studies [13,20,21]. However, a specific PPARγ variant, P121A, has also been associated with diabetes in specific populations and validated through family studies [22]. Hence, this locus encompasses both ancient, small-effect variants as well as recent, regional, large-effect variants. A G482S amino acid substitution in the functionally related PPARγ coactivator gene-1α (PGC-1α) gene has also been associated with metabolic alterations in regional populations, diabetes in the Danish [23] and altered lipid metabolism in the Pima Indians [24], but PGC-1α is not routinely detected by GWAS [13].

These observations suggest that there is a continuum of phenotypic effects among nDNA bioenergetic gene variants. The milder mutant phenotypes, which are less affected by purifying selection, may be retained for prolonged periods in the human population. Those that arose early in human radiation have remained with their original linkage group while being dispersed throughout the global population, and provide a signal in inter-population GWAS studies. More phenotypically significant variants have arisen throughout human history but have been acted on by purifying selection, eliminating them from the global population. Hence, these loci are not associated with widely dispersed chromosomal linkage groups. Some mutants, such as the PGC-1α G482S mutation, may be sufficiently adaptive to have arisen multiple independent times in different populations on different haplotypes, each new variant associated with a different linkage group rendering them undetectable by GWAS. To find these regional high impact variants it will

be necessary to study regional populations for functional genetic mutations, presumably by whole-genome sequencing.

3. Evolution and energetics

PPARγ and PGC-1α are nuclear transcription factors that play a major role in regulating bioenergetics and particularly mitochondrial biogenesis. The UCPs regulate the coupling efficiency of the mitochondrial energy production system, oxidative phosphorylation (OXPHOS). Hence, the importance of these and multiple other loci identified by GWAS implicate mitochondrial bioenergetics in diabetes and obesity and, by extension, mitochondrial functional variation in human regional environmental adaptation. Consistent with this supposition, the mitochondria are estimated to generate about 90 per cent of the cellular energy in differentiated tissue cells and the mitochondrial genome encompasses in the order of one to two thousand nDNA genes and thousands of copies of the mtDNA. Hence, a large number of mitochondrial gene targets can be mutated and have significant effects on cellular bioenergetics.

The mitochondria are the product of a symbiosis between two micro-organisms that occurred about two billion years ago. The nature of the original partner organisms is actively debated, but the progenitor of the mitochondrion is thought to have been an α-protobacterium that harboured a complete OXPHOS system. Both of these organisms alone were limited in their complexity since a single bacterial cell can generate only enough energy to sustain about 10 000 genes [5,6]. However, when the host cell acquired multiple oxidative bacteria, the bacterial energy could be pooled to provide the required energy for adding more genes to the host cell's DNA to create more complex anatomical structures. There was a problem, however. The oxidative bacteria needed most of their energy to sustain themselves. This dilemma was resolved since bacteria readily exchange genes. By transferring structural genes from the oxidative bacteria to the host cell's DNA, the number of bacterial gene copies was reduced from thousands to two, the pair of homologues in the nDNA. This reduced the amount of DNA to be replicated and reduced the complexity of mtDNA transcriptional regulation, with significant savings in energy. The accumulation of nDNA also permitted the nDNA genes to radiate and address new functional genetic space [5,25]. Hence, there was a significant selective pressure to transfer most of the oxidative bacterial genes to the host cell's DNA, which ultimately became the eukaryotic cell nDNA. As the number of its genes declined, the oxidative bacterium became progressively more integrated into the host cell.

Surprisingly, the progressive transfer of genes from the mtDNA to the nDNA did not go to completion and all oxidative eukaryotic cells still retain a mtDNA. In humans, the mtDNA codes for 13 polypeptide genes plus the rRNA and tRNA genes for the mitochondrial, bacteria-like, protein synthesis system [26,27].

To sustain a complete bacterial biogenesis apparatus is very energetically expensive. Therefore, there must be a strong evolutionary advantage for retaining the mtDNA. While several hypotheses have been put forward for why the mtDNA has been retained [5,28], the fact that all of the polypeptide genes retained by the mtDNA are central to

the mitochondrial energy-generating system OXPHOS provides one explanation [29].

All fungal and animal mtDNAs retain essentially the same set of OXPHOS polypeptide genes. In mammals, these include seven (ND1-3, 4L, 4-6) of the approximately 45 polypeptides of the electron transport chain (ETC) enzyme complex I, one (cytochrome b, cytb) of the 11 polypeptides of the ETC complex III, three (COI-III) of the 13 polypeptides of ETC complex IV (cytochrome c oxidase, COX) and two (ATP6 and 8) of the approximately 15 polypeptides of complex V, the ATP synthase.

Functionally, the mtDNA polypeptides are central to the electron and proton wiring system for mitochondrial energy production. In OXPHOS, reducing equivalents (electrons from reduced sources) derived from food flow from reduced to oxidized down the ETC that is embedded in the mitochondrial inner membrane. Starting with NADH, which is oxidized by complex I, and succinate by complex II, the electrons are transferred to coenzyme Q (CoQ), then to complex III, then cytocrome c, then to complex IV, and finally to oxygen to generate water. As the electrons traverse complexes I, III and IV, the energy released is used to pump protons from the mitochondrial matrix across the inner membrane to the inter-membrane space. This creates an electrochemical gradient that is acid and positive on the outside and alkaline and negative on the inside [30]. The resulting capacitance of about 0.2 V is the potential energy from which virtually all human biological processes are driven. Given that a human has in the order of 10^{17} mitochondrial capacitors, this is a great deal of potential energy, the vital force that animates our life. When breathing stops, the membrane potential collapses, energy transduction ceases and death ensues [27,31,32].

The potential energy stored in the mitochondrial capacitors can be used for many purposes: to take up Ca^{++} from the cytosol, modulate cellular REDOX status and reactive oxygen species (ROS) production, and transport proteins and substrates in and out of the mitochondrion. Particularly important, however, is that the mitochondrial inner membrane potential provides the motive force for driving complex V, the ATP synthase, to condense ADP and phosphate (Pi) to generate ATP [30]. ATP is the chemical energy carrier that is exported from the mitochondrion to the cytosol to energize cellular reactions and drive work.

Given the critical nature of the inner membrane potential, it is surprising that the efficiency by which the reducing equivalents from food calories are converted into ATP differs between different individuals and regional human populations. Some individuals are highly efficient at converting caloric energy into a proton gradient via the electron transport and by transforming the energy of the proton gradient into ATP. Consequently, these individuals need to burn the least number of calories for the required ATP. Since calories are a unit of heat, such 'tightly coupled' individuals generate the minimum heat for the ATP used. By contrast, some individuals are less efficient at converting reducing equivalents to membrane potential, and membrane potential to ATP. These 'loosely coupled' individuals burn more calories for the same amount of ATP and thus produce more core body heat per ATP used. This altered coupling efficiency can be modulated epigenetically, for example, by induction of uncoupling protein 1 (UCP1) in brown fat [33], or more stably by altering the sequence of the mtDNA and thus changing proton pumping efficiency [34,35].

In tropical and temperate environments, it is generally more advantageous to be more tightly coupled so as to produce the maximum ATP with the minimum heat. However, in the arctic the constraining factor is cold. Therefore, it can be beneficial to be less coupled so that more heat is generated to maintain core body temperature, provided sufficient dietary calories are available. That this type of adaptive variation is due to mtDNA changes has been supported by the demonstration that climate differences correlate with mtDNA rather than nDNA variation [36] and that the basal metabolic rate of Siberian populations is higher than that of more southern populations [37–39].

The importance of mtDNA variation in regional adaptation makes sense when it is realized that the mtDNA codes for the proteins that are central to the coupling of electron flow to proton pumping and thus ATP production. All four mitochondrial inner membrane complexes that incorporate mtDNA-coded polypeptides (I, III, IV and V) either generate or use the proton gradient. By contrast, complex II (succinate dehydrogenase), which transports electrons but does not pump protons, is composed of four nDNA polypeptides. Since the electrochemical gradient is a capacitor, the proton permeability of complexes I, III, IV and V must be balanced with each other. If any one of the complexes becomes leaky for protons, then the capacitor can short, which would be deleterious. Hence, the 13 polypeptides of the mtDNA represent an integrated electrical circuit in which each polypeptide must be functionally compatible with the other 12 mtDNA polypeptides of comparable coupling efficiency [27].

4. Evolution and mtDNA genetics

Given that the mtDNA polypeptides are highly variable in their energetic efficiency [40], then the random recombination of the 13 mtDNA polypeptides between different mtDNA lineages, for example, between mtDNAs that are more or less coupled, could erode energy efficiency. This dilemma is avoided by having these variable OXPHOS electron and proton transport genes linked together in a single non-recombining piece of DNA: the maternally inherited mtDNA. This forces all of the membrane potential elements to coevolve consistent with environmental factors such as climate and diet. Uniparental inheritance of the mtDNA is achieved by the concerted elimination of the paternal mtDNA at fertilization, thus blocking inter-individual mtDNA pairing and recombination [27]. Proof that absolute uniparental inheritance of the mtDNA is critical for animals has been obtained by artificially mixing two normal but different mtDNAs within the mouse female germline. The resulting 'heteroplasmic' mice manifested marked behavioural abnormalities and severe learning defects [41].

Because of strict maternal inheritance, the only way that the mtDNA sequence can change is by the sequential accumulation of mutations along radiating maternal lineages. Therefore, to a first approximation, the number of nucleotide differences between any two individuals is proportional to the time that they shared a common maternal ancestor. This unique feature of the mtDNA has permitted reconstruction

of the ancient origins and migrations of women. By overlaying the mtDNA mutational tree, which shows the genetic affinities between indigenous peoples with the geographic location of the populations that harbour those mtDNAs, the progressive movement of humans was mapped (figure 1). The mtDNA proved to be a particularly powerful tool for studying human radiation, since the mtDNA sequence evolution rate was found to be much greater than that of nDNA-coded mitochondrial genes [44–46]. Indeed, the sequence evolution rate of the mtDNA has resulted in critical mtDNA changes corresponding remarkably well with the times of prominent transitional events in human geographical radiation.

The striking correlation between mtDNA variation and geographic locale of indigenous populations stands in stark contrast to nDNA variation, where the common allelic variants are dispersed across most populations, differing primarily in allele frequencies [9–11]. For the mtDNA, each indigenous population has a limited number of clusters of related mtDNA haplotypes. Each of these haplotype clusters is generally descended from a founder mtDNA that harboured one or more functional variants. Therefore, the founder mtDNAs must have arisen in the population and, as the population grew, the descendants of the founding mtDNA acquired additional variants creating population-specific groups of related mtDNA haplotypes, designated haplogroups.

Haplogroups are also related to each other in higher-order clusters [43]. All of the haplogroups in Africa, designated 'L' haplogroups, are components of the wider African mtDNA lineage designated macrohaplogroup L that arose between 130 000 to 200 000 years before the present (YBP) (figure 1). Of the L haplogroups, haplogroup L3 spawned two new mtDNAs, designated M and N, in the Ethiopian area. About 65 000 YBP, only mtDNAs derived from M and N successfully departed Africa and colonized the rest of the world, generating macrohaplogroups M and N. Macrohaplogroup M moved along tropical Southeast Asia, ultimately reaching Australia, and later moved north out of Southeast Asia to form a plethora of central and eastern Asian mtDNA haplogroups, including C, D, G and M1–M20. By contrast, macrohaplogroup N went in two directions. In one, N moved along Southeast Asia to Australia and from southern Asia north into central Asia to generate haplogroups A and Z. In the second, N moved due north out of Africa to form the European haplogroups I, X and W and also into western Eurasia to form sub-macrohaplogroup R. R then gave rise to the European-specific haplogroups H, J, Uk, T, U and V, and R also moved east to produce the Asian mtDNA haplogroups B and F (figure 1).

Of all of the Asian mtDNA variants, only A, C and D became enriched in northeastern Siberia and were in a position to cross the Bering Land Bridge about 20 000 YBP to give rise to the Paleo-Indians. Later, A, C and D were joined by additional migrations bringing B and X to round out the Paleo-Indian mtDNA lineages. Subsequently, additional migrations brought haplogroups A and A + D to produce the Na-Dene and the Eskimos and Aleuts, respectively. Lastly, individuals carrying mtDNA haplogroup B mtDNA moved out from the Asian coast to colonize all of the Pacific Islands (figure 1) [47,48].

This regionality of the mtDNA haplogroups is extraordinary in several ways. First, of all of the African diversity, only two mtDNA lineages (M and N) colonized the rest of the world. This striking fact was missed in the first paper to report an African origin of the mtDNA [49], since these authors included African-American mtDNAs in their 'African' sample and about 30 per cent of African-American mtDNAs are of European, Native American and Asian origin introduced after the arrival of Africans in the Americas [50]. Second, of all of the Asian mtDNAs, only three mtDNA lineages (A, C and D) moved to extreme northeast Siberia to found the Paleo-Indians. Third, and most surprising, the mtDNA sequence evolution rate is such that it produced important mtDNA evolutionary changes that coincide with the major human geographical transitions. It is hard to imagine that such associations occurred by chance. It is much more probable that mtDNA variation was intimately involved in facilitating the human adaptation to different regional environments.

Support for the hypothesis that the major human population transitions were facilitated by mtDNA variation comes from the observations that mtDNA lineages that appear together with human transitions often harbour important mtDNA functional changes [34,35,51]. For example, the founding mtDNA for macrohaplogroup N, which moved directly from sub-tropical Africa north into the European temperate zone, harboured two polypeptide variants: ND3 nt 10398 G>A (A114T) and ATP6 nt 8701 G>A (A59T)[47,48], which changed the mitochondrial membrane potential and Ca^{2+} metabolism [52]. By contrast, macrohaplogroup M, which remained in semi-tropical Southeast Asia, was not founded by distinctive polypeptide variants. Significant mitochondrial biochemical differences have also been reported between other haplogroups, for example, European haplogroups H and Uk [53,54].

That mtDNA variants might have been adaptive has been substantiated by studying mtDNA variation in Tibetans, who are adapted to high altitude and thus low oxygen tension. In Tibetans, the ratio of macrohaplogroup M to N mtDNAs is greater than in low-land Chinese. Furthermore, an mtDNA ND1 nt 3394C (Y30H) variant has been found to have arisen three independent times on different macrohaplogroup M mtDNAs in Tibetans, but never to have become established on any Tibetan macrohaplogroup N mtDNAs. The 3394C variant progressively increases in frequency in villages with increasing altitude. For example, haplogroup M9 mtDNA, which encompasses the 3394C variant, is present at less than 2 per cent of the mtDNAs at sea level but increases to almost 35 per cent of the mtDNAs in the highest Tibetan villages.

Biochemical analysis has shown that when the 3394C variant arises on macrohaplogroup N mtDNAs, specifically haplogroups B or F, it results in a 15–28% reduction in complex I-specific activity. However, when the 3394C allele is present on haplogroup M9 mtDNA, it is associated with a complex I activity that is equal to or greater than any of the macrohaplogroup N mtDNAs with the wild-type 3394T variant [40]. Hence, the 3394C variant is adaptive at high altitude when arising on macrohaplogroup M mtDNAs.

mtDNA haplogroups have also been correlated with predisposition to a wide range of metabolic and degenerative diseases, cancers and longevity [48,53,55,56]. For example, the ND1 nt 3394C (Y30H) variant, when arising on macrohaplogroup N mtDNAs in European populations, is associated with increased penetrance of the milder primary mtDNA mutations that cause Leber Hereditary Optic Neuropathy

mutation rate = 2.2–2.9%/MYR
time estimates are YBP
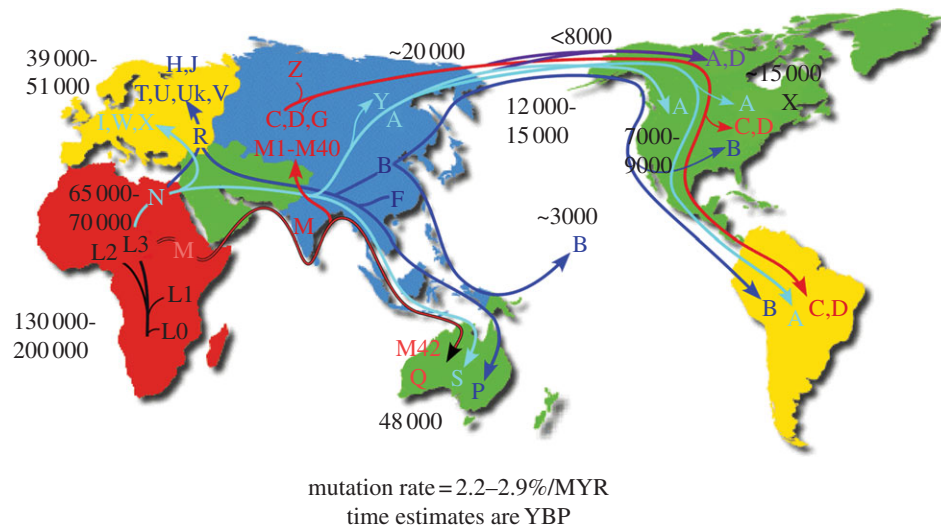
Figure 1. Diagram of the migratory history of the human mtDNA haplogroups. *Homo sapiens* mtDNAs arose in Africa about 130 000 – 200 000 years before the present (YBP), with the first African-specific haplogroup branch being L0, followed by the appearance in Africa of lineages L1, L2 and L3. In northeastern Africa, L3 gave rise to two new lineages M and N. Only M and N mtDNAs successfully left Africa about 65 000 YBP and colonized all of Eurasia and the Americas. The diverse array of mtDNA lineages that M and N spawned are clustered together as macrohaplogroups M and N. The founders of macrohaplogroup M moved out of Africa through India and along the Southeast Asian coast down along the Malaysian peninsula and into Australia, generating haplogroups Q and M42 around 48 000 YBP. Subsequently, M moved north out of Southeast Asia to produce a diverse array of mtDNA lineages including haplogroups C, D, G and many other M haplogroup lineages. In northeast Asia, haplogroup C gave rise to haplogroup Z. The founders of macrohaplogroup N also move though Southeast Asia and into Australia, generating haplogroup S. In Asia, macrohaplogroup N mtDNAs also moved north to generate central Asian haplogroup A and Siberian haplogroup Y. In western Eurasia, macrohaplogroup N founders also moved north to spawn European haplogroups I, W, and X and in western Eurasia gave rise to sub-macrohaplogroup R. R moved west to produce the European haplogroups H, J, Uk, T, U, and V and also moved east to generate Australian haplogroup P and eastern Asian haplogroups F and B. By 20 000 YBP, mtDNA haplogroups C and D from M and A from N were enriched in northeastern Siberia and thus were positioned to migrate across the Bering land bridge (Beringia) to give rise to the first Native American populations, the Paleo-Indians. Haplogroups A, C, and D migrated throughout North America and on through Central American to radiate into South America. Haplogroup X, which is most prevalent in Europe but is also found in Mongolia, though not in Siberia, arrived in North America about 15 000 YBP, but remained in northern North America. Haplogroup B, which is not found in Siberia but is prevalent along the coast of Asia, arrived in North America about 12 000 to 15 000 YBP and moved through North and Central America and into South America, combining with A, C, D and X to generate the five dominant Paleo-Indian haplogroups (A − D + X). A subsequent migration of haplogroup A out of the Chukotka peninsula about 7000 to 9000 YBP gave rise to the Na-Déné (Athabaskins, Navajo, Apache, etc.). Subsequent movement across the Bering Strait, primarily carrying haplogroups A and D after 6000 YBP, produced the Eskimo and Aleut populations. Most recently, eastern Asian haplogroup B migrated south along the Asian coast through Micronesia and out into the Pacific to colonize all of the Pacific islands. Ages of migrations are approximated using mtDNA sequence evolution rates determined by comparing regional archeological or physical anthropological data with corresponding mtDNA sequence diversity. Since selection may have limited the accumulation of diversity in certain contexts, ages for regional migrations can best be estimated from the diversity encompassed within an individual regional or continental lineage, since selection would have had its greatest effect in enriching for the founding mtDNA haplotype, after which mtDNA mutations would accumulate randomly and thus become clock-like [42,43] (reproduced from http://www.mitomap.org, with permission).

(LHON) [57,58]. Haplogroup F mtDNAs are associated with a predilection to diabetes [59] and analysis of the complex I-specific activities between mtDNA haplogroups B and F has revealed a 30 per cent lower complex I activity in haplogroup F [40]. Consequently, adaptive and clinically relevant mtDNA variants are also associated with functional changes in OXPHOS.

These data prove that at least some mtDNA variation is not neutral, but rather is functional and can be adaptive when present on the appropriate mtDNA background and environment. Yet the same variant in a different mtDNA background and/or environment can be deleterious and pre-dispose to metabolic and/or degenerative diseases [34,35,51].

According to Kimura [12], the high mtDNA mutation rate should create many deleterious mutations resulting in a genetic load that could destroy the species. This concern is particularly apt for the mtDNA since every one of the genes of the mtDNA is critical for life. However, this concern is unfounded since the mammalian ovary has a pre-fertilization selection system that removes those proto-oocytes that harbour severe mtDNA mutations before they can be ovulated and fertilized

[60,61]. Hence, these mutations never contribute to the genetic load of the species. The human female generates millions of proto-oocytes, but only successfully ovulates about 400. All the rest are eliminated by atresia. This apparent inefficiency might exist specifically to permit selection against severe mtDNA mutations, thus permitting a high mtDNA sequence evolution rate.

The high mtDNA mutation rate paired with the pre-fertilization selection means that de novo mtDNA mutations that subtly alter mitochondrial energy metabolism are constantly being introduced into the mammalian germline. These mtDNA variants then produce bioenergetic diversity that lets subpopulations within the species physiologically adapt to and occupy slightly different region energetic environments. Hence, mtDNA variation is an important component of the missing functional variation that allows subpopulations of humans and other animals to adapt to local environments.

Why does this system not work for nDNA anatomical gene mutations? The difference is that mitochondrial bioenergetics is manifest at the individual cellular level, so natural selection

can act directly on the metabolism of the unfertilized oocyte and/or its associated nurse cells. By contrast, to express anatomical genes, the oocyte must be fertilized and progress through development to generate a fully formed individual before the mutation can be manifest and acted on by natural selection. Therefore, even the most deleterious nDNA anatomical mutations are passed into the population where they can contribute to the genetic load. The only way of minimizing the lethal effects of deleterious anatomical mutations is to keep the nDNA mutation rate very low. This means that nDNA mutations are generally too rare to be the initiating factors in intra-specific functional radiation.

What about mutations in nDNA-coded bioenergetic genes such as those detected by GWAS? As Kimura [12] predicted, most nuclear mutations result in lethal phenotypes, for example, those that cause Maturity Onset Diabetes of the Young (MODY) [31]. Less severe but still functionally significant metabolic gene mutations might be beneficial in peripheral energetic environments and become locally enriched by adaptive selection. However, if the energetic environment changes or the individuals migrate to a new environment these functional variants become maladaptive and are removed by purifying selection causing metabolic disease. Only the mildest phenotypic variants in bioenergetic loci escape selection and can become dispersed throughout populations and detected by GWAS [13,14].

5. Evolution and the interaction of Mendelian and mitochondrial genetics

Animals thus have two primary evolutionary strategies: an energetic strategy that changes rapidly and is optimal for intra-specific radiation into various regional energetic environments and an anatomical strategy that changes slowly and is optimal for inter-specific structural radiation for exploitation of new niches. The short-term energetic adaptive strategy is primarily driven by the high mtDNA mutation rate and the resulting rapid intra-specific energetic radiation. The long-term anatomical adaptive strategy is rooted in the low nDNA mutation rate, leading to subtle inter-specific anatomical alterations [7,32].

The continuously arising mtDNA functional variation allows subpopulations of animal species to occupy a range of energetic environments within the species' niche. Those subpopulations that occupy peripheral energetic environments for prolonged periods will amass multiple mtDNA variants permitting the adjustment to local energy resources and demands. As the mtDNA sequence evolves, the slower-evolving nDNA bioenergetic genes will follow to consolidate survival in the peripheral energetic environment and to maintain optimal nDNA and mtDNA interactions [62].

The nDNA variants can have a range of phenotypic consequences from large to small. Variants that are of physiological significance and thus phenotypic importance in moderately divergent environments include the specific genetic variants in PPARγ [22], PGC-1α [23,24] and UCP1-3 [16–19]. Examples of nDNA variants that permitted human adaptation to extreme environments are found among Tibetans, who have adapted to low oxygen tensions for tens of thousands of years. Adaptive Tibetan nDNA loci include an intronic SNP within endothelial Per-Arnt-Sim (PAS) domain protein 1 (EPAS1), also known as hypoxia-inducible

factor 2α (HIF-2α). HIF-2α is a transcription factor that responds to hypoxia by inducing glycolysis and modifying mitochondrial energetics. The Tibetan SNP shows a 78 per cent frequency difference between Tibetan and Han samples [63]. Additional hypoxia pathway genetic variants found in Tibetans include the HIF-1α modifying prolyl hydroxylase domain-containing protein 2 (PHD2)(EGLN1) and the HIF-1α target gene PPARA (PPARα) [64].

However, when an environment changes those loci that were adaptive in the previous environment can become incompatible with the new environment. This can result in pathological conditions such as diabetes and obesity and these alleles are then removed by purifying selection. nDNA variants with more modest phenotypic effects can be retained and dispersed by stochastic processes.

Associations between mtDNA haplogroups and diabetes have been found when studying well-defined human populations [59]. However, studies attempting to correlate mtDNA SNPs to diabetes using an aggregate of multiple Europe and American populations have not been successful [65].

These considerations then provide a logic for linking the disparate observations about human genetic variation and the origin of intra-specific adaptive variation. The high mutation rate of the mtDNA together with the pre-fertilization ovarian selection provide a rapid adaptation system for each species to adjust to regional environmental changes. These variants have a very high functional effect and hence are relatively specific to particular regional environments and thus subpopulations. As a result, even relatively closely related populations have very different adaptive mtDNA variants and thus different mtDNA haplotypes, haplogroups and associated mtDNA SNPs. The high mtDNA mutation rate also means that the same adaptive variant can arise multiple independent times in disparate populations. However, in different genetic and environmental contexts the phenotypic effect of a particular mtDNA variant can be very different.

Once mtDNA variation has permitted a species' subpopulation to become established in a marginal energetic environment for prolonged periods, mutations that augment energetic adaptation also accumulate in the nDNA-coded bioenergetic genes. If the subpopulation remains indefinitely in the marginal environment, then sufficient nDNA changes will accumulate to render the mtDNA and nDNA variants of the isolated population incompatible with those of the parental population, causing speciation [62,66–73]. Consistent with this supposition, a number of nDNA OXPHOS gene variants have been found to correspond to primate speciation events [34,62,74–77].

As a subpopulation drifts away from the parent population energetically, new environmental energy reservoirs become accessible, frequently in significantly different physical forms. This selects for anatomical changes that optimize exploitation of the new energy reservoir.

Once the new species has arisen to exploit a new energy reservoir, the species will begin to expand its range to overlap that of the new energy resource. This will require the population to expand back into more median global energetic environments. The resulting reduction in the extreme environmental constraints will select for the return of the mitochondrial bioenergetic metabolism back to its more energy-efficient state. This revision back to the more energy-efficient mtDNA genotypes can occur relatively rapidly due to the high sequence evolution rate of the mtDNA and the strong selection for the

most efficient OXPHOS system when calorie limitation is the primary selective factor.

Only a portion of the mtDNA polypeptide amino acids are amenable to adaptive mutations, since these mutations must alter the efficiency of the OXPHOS enzyme complex without altering the enzyme's structure or assembly. Since these amino acid substitutions do not affect the capacity of individuals within the species to successfully mate with each other, these mtDNA variants can be maintained within a species and are not the final arbiters of speciation. Thus, these adaptive mtDNA variants can be polymorphic within a species but be essentially invariant between species. The reversion of these adaptive polypeptide variants to a more universal mtDNA genotype following speciation may explain why the mtDNAs from one species can introgress into another very closely related species through a hybridization zone [78].

In addition to the adaptive mtDNA polypeptide variants, the high mtDNA mutation rate also generates mtDNA variants that alter the interaction between the mtDNA and nDNA subunits of the same OXPHOS complex. This class of mtDNA variants can also arise in marginal energy environments and may be beneficial in certain contexts. However, since they reduce the stability of the OXPHOS complexes, they produce a strong selective pressure for the appearance of nDNA mutations in the OXPHOS enzyme polypeptide that interacts with the mtDNA mutant polypeptide to permit optimal enzyme assembly and function. Once such pairs of complementary mtDNA and nDNA OXPHOS polypeptide variants become established in a peripheral environment, they will begin to limit the success of matings of individuals from the peripheral environment with those of the parental environment. Hence, these mtDNA variants can drive speciation via nuclear–cytoplasmic incompatibility [79].

In summary, there are two major classes of evolutionarily relevant mtDNA variants, one that permits expansion of the species range into energetically peripheral environments and another that leads to the genetic isolation of the peripheral population through the accumulation of complementary mtDNA and nDNA OXPHOS enzyme variants. As more of the nuclear–cytoplasmic subunit interaction variants accumulate, the genetic barrier between successful matings of the parental and peripheral population increases until speciation becomes established. As the new species expands out from its peripheral energy environment, the adaptive mtDNA variants revert to permit the new species to optimally exploit its new energetic niche.

6. Energetic genetics and human health

The neutralist–selectionist debate arose because studies of nDNA genetic variation between human populations could not identify the required adaptive variation in the nDNA. The discovery and characterization of the mtDNA evolutionary system resolves this dilemma since it is mtDNA variation that provides the initial high impact variation for intraspecific adaption to different energy environments. The high mtDNA mutation rate permits physiological adaptation in the time span of thousands of years, which is the time frame for intra-specific radiation. Hence, it was mtDNA variation that permitted our ancestors to occupy new energetic environments, which is why mtDNA variation is regional and corresponds to major human geographical transitions.

Occupation of alternative energetic environments then enriched for rare nDNA variants, which consolidated the bioenergetic adaptation of the isolated animal populations to different energetic environments within the species' niche. Subsequent environmental changes and/or migration created gene–environment mismatches and bioenergetic stress that in humans manifests as metabolic disease. The most phenotypically significant adaptive loci generate the greatest functional disconnect and most severe disease when the environment changes. Hence, these are the risk factors for common diseases, not the much milder and ubiquitous variants detected across all populations by GWAS.

The clarification of the bioenergetic genetics of common metabolic diseases indicates that to identify high impact nDNA genetic loci for common diseases it will be necessary to sequence the entire genome of patients and controls from regional populations and look for novel regional genetic variants that correlate with the disease. With these new approaches, high impact loci should be found that will permit presymptomatic diagnosis in regional populations. Such advances will suggest new therapeutic approaches for the treatment and prevention of human disease [48].

References

1. Darwin C. 1859 *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, UK: John Murray of Albemarle Street.
2. Darwin CR, Wallace AR. 1858 On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *J. Proc. Linn. Soc. Lond. Zool.* **3**, 46–50.
3. Schrodinger E. 1944 *What is life? The physical aspect of the living cell*. Cambridge, UK: Cambridge University Press.
4. Morowitz HJ. 1968 *Energy flow in biology: biological organization as a problem in thermal physics*. New York, NY: Academic Press.
5. Lane N. 2011 Energetics and genetics across the prokaryote–eukaryote divide. *Biol. Direct* **6**, 35. (doi:10.1186/1745-6150-6-35)
6. Lane N, Martin WF. 2012 The origin of membrane bioenergetics. *Cell* **151**, 1406–1416. (doi:10.1016/j.cell.2012.11.050)
7. Wallace DC. 2010 Colloquium paper: bioenergetics, the origins of complexity, and the ascent of man. *Proc. Natl Acad. Sci. USA* **107**, 8947–8953. (doi:10.1073/pnas.0914635107)
8. Ricardo A, Szostak JW. 2009 Origin of life on earth. *Scient. Am.* **301**, 54–61. (doi:10.1038/scientificamerican0909-54)
9. Cavalli-Sforza LL. 1998 The DNA revolution in population genetics. *Trends Genet.* **14**, 60–65. (doi:10.1016/S0168-9525(97)01327-9)
10. Cavalli-Sforza LL, Bodmer WF. 1971 *The genetics of human populations*. San Francisco, CA: WH. Freeman.
11. Li JZ *et al*. 2008 Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104. (doi:10.1126/science.1153717)

12. Kimura M. 1983 *The neutral theory of molecular evolution*. Cambridge, UK: Cambridge University Press.

13. Morris AP *et al*. 2012 Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990. (doi:10.1038/ng.2383)

14. Manning AK *et al*. 2012 A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669. (doi:10.1038/ng.2274)

15. Scott LJ *et al*. 2007 A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345. (doi:10.1126/science.1142382)

16. Dalgaard LT. 2011 Genetic variance in uncoupling protein 2 in relation to obesity, type 2 diabetes, and related metabolic traits: focus on the functional −866G>a promoter variant (rs659366). *J. Obesity* **2011**, 340241. (doi:10.1155/2011/340241)

17. Jia JJ, Tian YB, Cao ZH, Tao LL, Zhang X, Gao SZ, Ge CR, Lin QY, Jois M. 2010 The polymorphisms of UCP1 genes associated with fat metabolism, obesity and diabetes. *Mol. Biol. Rep.* **37**, 1513–1522. (doi:10.1007/s11033-009-9550-2)

18. Jun HS, Kim IK, Lee HJ, Lee HJ, Kang JH, Kim JR, Shin HD, Song J. 2009 Effects of UCP2 and UCP3 variants on the manifestation of overweight in Korean children. *Obesity* (*Silver Spring*) **17**, 355–362. (doi:10.1038/oby.2008.531)

19. Srivastava N, Prakash J, Lakhan R, Agarwal CG, Pant DC, Mittal B. 2010 A common polymorphism in the promoter of UCP2 is associated with obesity and hyperinsulenemia in northern Indians. *Mol. Cell. Biochem.* **337**, 293–298. (doi:10.1007/s11010-009-0311-2)

20. Saxena R *et al*. 2007 Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336. (doi:10.1126/science.1142358)

21. Zeggini E *et al*. 2007 Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341. (doi:10.1126/science.1142364)

22. Altshuler D *et al*. 2000 The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80. (doi:10.1038/79216)

23. Ek J, Andersen G, Urhammer SA, Gaede PH, Drivsholm T, Borch-Johnsen K, Hansen T, Pedersen O. 2001 Mutation analysis of peroxisome proliferator-activated receptor-gamma coactivator-1 (PGC-1) and relationships of identified amino acid polymorphisms to type II diabetes mellitus. *Diabetologia* **44**, 2220–2226. (doi:10.1007/s001250100032)

24. Muller YL, Bogardus C, Pedersen O, Baier L. 2003 A Gly482Ser missense mutation in the peroxisome proliferator-activated receptor gamma coactivator-1 is associated with altered lipid oxidation and early

25. Lane N, Martin W. 2010 The energetics of genome complexity. *Nature* **467**, 929–934. (doi:10.1038/nature09486)

26. Anderson S *et al*. 1981 Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465. (doi:10.1038/290457a0)

27. Wallace DC. 2007 Why do we have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annu. Rev. Biochem.* **76**, 781–821. (doi:10.1146/annurev.biochem.76.081205.150955)

28. Allen JF. 2003 The function of genomes in bioenergetic organelles. *Phil. Trans. R. Soc. Lond. B* **358**, 19–37. (doi:10.1098/rstb.2002.1191)

29. Wallace DC. 1982 Structure and evolution of organelle genomes. *Microbiol. Rev.* **46**, 208–240.

30. Mitchell P. 1961 Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148. (doi:10.1038/191144a0)

31. Wallace DC. 2005 A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu. Rev. Genet.* **39**, 359–407. (doi:10.1146/annurev.genet.39.110304.095751)

32. Wallace DC. 2011 Bioenergetic origins of complexity and disease. *Cold Spring Harb. Symp. Quant. Biol.* **76**, 1–16. (doi:10.1101/sqb.2011.76.010462)

33. Puigserver P, Wu Z, Park CW, Graves R, Wright M, Spiegelman BM. 1998 A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839. (doi:10.1016/S0092-8674(00)81410-5)

34. Mishmar D, Ruiz-Pesini E, Mondragon-Palomino M, Procaccio V, Gaut B, Wallace DC. 2006 Adaptive selection of mitochondrial complex I subunits during primate radiation. *Gene* **378**, 11–18. (doi:10.1016/j.gene.2006.03.015)

35. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004 Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**, 223–226. (doi:10.1126/science.1088434)

36. Balloux F, Handley LJ, Jombart T, Liu H, Manica A. 2009 Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proc. R. Soc. B* **276**, 3447–3455. (doi:10.1098/rspb.2009.0752)

37. Leonard WR, Sorensen MV, Galloway VA, Spencer GJ, Mosher MJ, Osipova L, Spitsyn VA. 2002 Climatic influences on basal metabolic rates among circumpolar populations. *Am. J. Hum. Biol.* **14**, 609–620. (doi:10.1002/ajhb.10072)

38. Snodgrass JJ, Leonard WR, Sorensen MV, Tarskaia LA, Mosher MJ. 2008 The influence of basal metabolic rate on blood pressure among indigenous Siberians. *Am. J. Phys. Anthropol.* **137**, 145–155. (doi:10.1002/ajpa.20851)

39. Snodgrass JJ, Leonard WR, Tarskaia LA, Alekseev VP, Krivoshapkin VG. 2005 Basal metabolic rate in the

Yakut (Sakha) of Siberia. *Am. J. Hum. Biol.* **17**, 155–172. (doi:10.1002/ajhb.20106)

40. Ji F *et al*. 2012 Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and high-altitude Tibetans. *Proc. Natl Acad. Sci. USA* **109**, 7391–7396. (doi:10.1073/pnas.1202484109)

41. Sharpley MS *et al*. 2012 Heteroplasmy of mouse mtDNA Is genetically unstable and results in altered behavior and cognition. *Cell* **151**, 333–343. (doi:10.1016/j.cell.2012.09.004)

42. MITOMAP. 2012 A human mitochondrial genome database. See http://www.mitomap.org.

43. Ruiz-Pesini E *et al*. 2007 An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.* **35**, D823–D828. (doi:10.1093/nar/gkl927)

44. Brown WM, Prager EM, Wan A, Wilson AC. 1982 Mitochondrial DNA sequences in primates: tempo and mode of evolution. *J. Mol. Evol.* **18**, 225–239. (doi:10.1007/BF01734101)

45. Neckelmann N, Li K, Wade RP, Shuster R, Wallace DC. 1987 cDNA sequence of a human skeletal muscle ADP/ATP translocator: lack of a leader peptide, divergence from a fibroblast translocator cDNA, and coevolution with mitochondrial DNA genes. *Proc. Natl Acad. Sci. USA* **84**, 7580–7584. (doi:10.1073/pnas.84.21.7580)

46. Wallace DC, Ye JH, Neckelmann SN, Singh G, Webster KA, Greenberg BD. 1987 Sequence analysis of cDNAs for the human and bovine ATP synthase b-subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Curr. Genet.* **12**, 81–90. (doi:10.1007/BF00434661)

47. Wallace DC, Brown MD, Lott MT. 1999 Mitochondrial DNA variation in human evolution and disease. *Gene* **238**, 211–230. (doi:10.1016/S0378-1119(99)00295-4)

48. Wallace DC, Lott MT, Procaccio V. 2013 Mitochondrial medicine: the mitochondrial biology and genetics of metabolic and degenerative diseases, cancer, and aging. In *Emery and Rimoin's principles and practice of medical genetics*, vol. 1 (eds DL Rimoin, RE Pyeritz, BR Korf). Philadelphia, PA: Churchill Livingstone Elsevier.

49. Cann RL, Stoneking M, Wilson AC. 1987 Mitochondrial DNA and human evolution. *Nature* **325**, 31–36. (doi:10.1038/325031a0)

50. Lind JM *et al*. 2007 Elevated male European and female African contributions to the genomes of African American individuals. *Hum. Genet.* **120**, 713–722. (doi:10.1007/s00439-006-0261-7)

51. Ruiz-Pesini E, Wallace DC. 2006 Evidence for adaptive selection acting on the tRNA and rRNA genes of the human mitochondrial DNA. *Hum. Mutat.* **27**, 1072–1081. (doi:10.1002/humu.20378)

52. Kazuno AA, Munakata K, Nagai T, Shimozono S, Tanaka M, Yoneda M, Kato N, Miyawaki A, Kato T. 2006 Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics. *PLoS Genet.* **2**, e128. (doi:10.1371/journal.pgen.0020128)

53. Gomez-Duran A, Pacheu-Grau D, Lopez-Gallardo E, Diez-Sanchez C, Montoya J, Lopez-Perez MJ, Ruiz-Pesini E. 2010 Unmasking the causes of multifactorial disorders: OXPHOS differences

between mitochondrial haplogroups. *Hum. Mol. Genet.* **19**, 3343–3353. (doi:10.1093/hmg/ddq246)

54. Rollins B *et al*. 2009 Mitochondrial variants in schizophrenia, bipolar disorder, and major depressive disorder. *PLoS ONE* **4**, e4913. (doi:10.1371/journal.pone.0004913)

55. Khusnutdinova E, Gilyazova I, Ruiz-Pesini E, Derbeneva O, Khusainova R, Khidiyatova I, Magzhanov R, Wallace DC. 2008 A mitochondrial etiology of neurodegenerative diseases: evidence from Parkinson's disease. *Ann. NY Acad. Sci.* **1147**, 1–20. (doi:10.1196/annals.1427.001)

56. Wallace DC. 2008 Mitochondria as chi. *Genetics* **179**, 727–735. (doi:10.1534/genetics.104.91769)

57. Brown MD, Torroni A, Reckord CL, Wallace DC. 1995 Phylogenetic analysis of Leber's hereditary optic neuropathy mitochondrial DNA's indicates multiple independent occurrences of the common mutations. *Hum. Mutat.* **6**, 311–325. (doi:10.1002/humu.1380060405)

58. Liang M *et al*. 2009 Leber's hereditary optic neuropathy is associated with mitochondrial ND1 T3394C mutation. *Biochem. Biophys. Res. Commun.* **383**, 286–292. (doi:10.1016/j.bbrc.2009.03.097)

59. Fuku N *et al*. 2007 Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians. *Am. J. Hum. Genet.* **80**, 407–415. (doi:10.1086/512202)

60. Fan W *et al*. 2008 A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations. *Science* **319**, 958–962. (doi:10.1126/science.1147786)

61. Stewart JB, Freyer C, Elson JL, Larsson NG. 2008 Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease. *Nat. Rev. Genet.* **9**, 657–662. (doi:10.1038/nrg2396)

62. Burton RS, Barreto FS. 2012 A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Mol. Ecol.* **21**, 4942–4957. (doi:10.1111/mec.12006)

63. Yi X *et al*. 2010 Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78. (doi:10.1126/science.1190371)

64. Simonson TS *et al*. 2010 Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**, 72–75. (doi:10.1126/science.1189406)

65. Saxena R *et al*. 2006 Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am. J. Hum. Genet.* **79**, 54–61. (doi:10.1086/504926)

66. Barrientos A, Kenyon L, Moraes CT. 1998 Human xenomitochondrial cybrids. Cellular models of mitochondrial complex I deficiency. *J. Biol. Chem.* **273**, 14210–14217. (doi:10.1074/jbc.273.23.14210)

67. Bayona-Bafaluy MP, Blits B, Battersby BJ, Shoubridge EA, Moraes CT. 2005 Rapid directional shift of mitochondrial DNA heteroplasmy in animal tissues by a mitochondrially targeted restriction endonuclease. *Proc. Natl Acad. Sci. USA* **102**, 14392–14397. (doi:10.1073/pnas.0502896102)

68. Coyne JA, Orr HA. 1998 The evolutionary genetics of speciation. *Phil. Trans. R. Soc. Lond. B* **353**, 287–305. (doi:10.1098/rstb.1998.0210)

69. Kenyon L, Moraes CT. 1997 Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proc. Natl Acad. Sci. USA* **94**, 9131–9135. (doi:10.1073/pnas.94.17.9131)

70. McKenzie M, Chiotis M, Pinkert CA, Trounce IA. 2003 Functional respiratory chain analyses in murid xenomitochondrial cybrids expose coevolutionary constraints of cytochrome *b* and nuclear subunits of complex III. *Mol. Biol. Evol.* **20**, 1117–1124. (doi:10.1093/molbev/msg132)

71. McKenzie M, Trounce I. 2000 Expression of *Rattus norvegicus* mtDNA in *Mus musculus* cells results in multiple respiratory chain defects. *J. Biol. Chem.* **275**, 31514–31519. (doi:10.1074/jbc.M004070200)

72. McKenzie M, Trounce IA, Cassar CA, Pinkert CA. 2004 Production of homoplasmic xenomitochondrial mice. *Proc. Natl Acad. Sci. USA* **101**, 1685–1690. (doi:10.1073/pnas.0303184101)

73. Schmidt TR, Wu W, Goodman M, Grossman LI. 2001 Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome *c* oxidase. *Mol. Biol. Evol.* **18**, 563–569. (doi:10.1093/oxfordjournals.molbev.a003836)

74. Doan JW, Schmidt TR, Wildman DE, Uddin M, Goldberg A, Huttemann M, Goodman M, Weiss ML, Grossman LI. 2004 Coadaptive evolution in cytochrome *c* oxidase: 9 of 13 subunits show accelerated rates of nonsynonymous substitution in anthropoid primates. *Mol. Phylogenet. Evol.* **33**, 944–950. (doi:10.1016/j.ympev.2004.07.016)

75. Grossman LI, Wildman DE, Schmidt TR, Goodman M. 2004 Accelerated evolution of the electron transport chain in anthropoid primates. *Trends Genet.* **20**, 578–585. (doi:10.1016/j.tig.2004.09.002)

76. Schmidt TR, Wildman DE, Uddin M, Opazo JC, Goodman M, Grossman LI. 2005 Rapid electrostatic evolution at the binding site for cytochrome *c* on cytochrome *c* oxidase in anthropoid primates. *Proc. Natl Acad. Sci. USA* **102**, 6379–6384. (doi:10.1073/pnas.0409714102)

77. Uddin M, Opazo JC, Wildman DE, Sherwood CC, Hof PR, Goodman M, Grossman LI. 2008 Molecular evolution of the cytochrome *c* oxidase subunit 5A gene in primates. *BMC Evol. Biol.* **8**, 8. (doi:10.1186/1471-2148-8-8)

78. Bastos-Silveira C, Santos SM, Monarca R, Mathias Mda L, Heckel G. 2012 Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Mol. Ecol.* **21**, 5309–5323. (doi:10.1111/mec.12018)

79. Montooth KL, Meiklejohn CD, Abt DN, Rand DM. 2010 Mitochondrial–nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of *Drosophila*. *Evolution* **64**, 3364–3379. (doi:10.1111/j.1558-5646.2010.01077.x)