

REVIEW

Open Access

Maternal ancestry and population history from whole mitochondrial genomes

Toomas Kivisild

Abstract

MtDNA has been a widely used tool in human evolutionary and population genetic studies over the past three decades. Its maternal inheritance and lack of recombination have offered the opportunity to explore genealogical relationships among individuals and to study the frequency differences of matrilineal clades among human populations at continental and regional scales. The whole mtDNA genome sequencing delivers molecular resolution that is sufficient to distinguish patterns that have arisen over thousands of years. However, mutation rate is highly variable among the functional and non-coding domains of mtDNA which makes it challenging to obtain accurate split dates of the mitochondrial clades. Due to the shallow coalescent time of mitochondrial TMRCA at approximately 100 to 200 thousand years (ky), mtDNA data have only limited power to inform us about the more distant past and the early stages of human evolutionary history. The variation shared by mitochondrial genomes of individuals drawn from different continents outside Africa has been used to illuminate the details of the colonization process of the Old World, whereas regional patterns of variation have been at the focus of studies addressing questions of a more recent time scale. In the era of whole nuclear genome sequencing, mitochondrial genomes are continuing to be informative as a unique tool for the assessment of female-specific aspects of the demographic history of human populations.

Keywords: Maternal ancestry, MtDNA, Population history

Review

Introduction

Maternal inheritance [1], fast mutation rate [2], high copy number per cell [3,4], and the lack of recombination [5,6] were the features that brought mtDNA at the focus of evolutionary genetic studies in the 1980's and 1990's when the human genome sequencing had not been completed yet and the idea of whole nuclear genome level population genetics was only a daydream for population geneticists. The presence of mitochondria as energy producing small bacteria-like 'power cells' within our cells is one of the defining features of eukaryotes. The adoption of this organelle was a critical step in the earliest stages of our evolutionary history that allowed the cells of our ancestors to diversify in size and shape and to develop their characteristic feeding mode of a phagotrophic predator [7]. The special relationship between the hosting cell and the mitochondria also determines the

specific aspects of the replication, transmission and population genetics of the DNA molecules in mitochondria, the variation of the mtDNA copy number by cell types and developmental stages and the small size and high gene density of mitochondrial genome (for review see [8]).

Humans along with western chimpanzees and eastern gorillas have remarkably low genetic diversity compared to other great apes [9]. Low genetic diversity means that for any nuclear gene one needs to sequence thousands or tens of thousands of base pairs to have a chance of finding SNPs that are informative for population genetic purposes. In the era of PCR and Sanger sequencing the high mutation rate made it more cost effective to uncover DNA sequence variation at the population scale from mtDNA than from any nuclear locus. Furthermore, the lack of recombination allowed the data from coding and non-coding regions of mtDNA to be combined into the shape of a phylogenetic tree. The branches of this ever-growing tree, as more data became available, could be labelled by distinctive restriction fragment length polymorphisms (RFLPs). As a result, the most common

Correspondence: tk331@cam.ac.uk
Division of Biological Anthropology, University of Cambridge, CB2 1QH
Cambridge, UK

branches were assigned alphabetic labels that became to be known as mtDNA haplogroups [10].

The nomenclature of mtDNA haplogroups was introduced in the mid-1990s with A-G labels assigned to variation observed in Asian and American lineages [10,11], H-K to Europe [12] whereas only a single letter, L, was assigned to describe the highest level of variation observed in Africa in a study using an Asian outgroup [13]. The mtDNA nomenclature that is currently used (<http://www.phylotree.org/>) has a robust branch structure that has been determined through the rigorous and detailed analyses of the whole mtDNA genomes [14]. These topological details of the mtDNA phylogeny have been revealed step by step over the last two decades thanks to the contributions of many groups in covering with data ever increasing numbers of populations across the world and thanks to the advances in technology that eventually have led to the use of whole mtDNA sequencing as a routine approach in the field.

Robust inference of the phylogenetic tree and its high resolution has been important for various reasons. The initial RFLP based studies, for example, with limited number of polymorphic sites that were known in the early 1980's had concluded that the root of human mtDNA was in Asia [15]. However, more comprehensive analyses of 195 polymorphic RFLP sites across the whole mtDNA sequence determined in 145 human placentas and two cell lines drawn from five geographically distinct populations [16] suggested that all variants observed in present day populations can be inferred to derive from a single female ancestor who was postulated to have lived approximately 200,000 years ago in Africa. However, these early phylogenies were not sufficiently robust, so that critics were able to produce alternative root topologies and African origins were repeatedly challenged and reclaimed in the following decade [17-20]. Although the RFLP studies and HVS-I sequencing based work often ended up showing high level of phylogenetic uncertainty they were the approaches taken at the time that provided the first insights into the mtDNA variation at continental scales. These efforts led to the formulation of research hypotheses that became actively debated and subject to further scrutiny, including, for example, the earliest attempts to define the genetic source and number of founding lineages of Native Americans [21] and of Polynesians [22,23], and relative contributions of Palaeolithic, Mesolithic and Neolithic gene flow in the peopling of Europe [24].

Mutation rates and TMRCA of mtDNA variation

All evolutionary genetic studies that associate the patterns of mtDNA variation observed in human populations with time explicit models make assumptions about the molecular clock. The mutation rate of mtDNA in

animals is known to be higher by at least an order of magnitude than the mutation rate in nuclear genes [2]. In vertebrates the mitochondrial mutation rate, in fact, is $\times 25$ higher than nuclear DNA mutation rate whereas the opposite is true for most of the plants whose mitochondria evolve approximately $\times 20$ slower than their nuclear genes [25]. However, the rates at which mutations occur or get fixed in mitochondria are not uniformly high along the molecule and its functional domains. The rate variation among sites and time dependence of substitution rates at the intra- and interspecies scales [26-29], along with issues related to germ line and somatic heteroplasmy [30] have been major challenges for getting accurate estimates of human mtDNA mutation rate. Heteroplasmy refers to the existence of different types of mtDNA in the same individual. Because of high copy number in most human tissues the levels of mtDNA heteroplasmy may vary from very low, $<5\%$, that can be detected and studied now with the next generation sequencing methods (reviewed in [31]), to those up to 1:1 ratio. Most heteroplasmies are resolved within a few generations by the severe germ-line bottlenecks leading to the loss of many de novo mutations, an effect that needs to be considered when calibrating mutation rates from pedigree data [30]. Somatic heteroplasmies do not contribute to mutation rate and only a small fraction of germ-line mutations get fixed in genealogies. Further complicating factors include the directionality of the mutations [32] – most hypervariable positions are unstable only in the G- \rightarrow A, T- \rightarrow C direction (according to the L-strand convention of the reference sequence) and the 60 fold or higher effective transition/transversion rate biases [33].

Mechanisms emphasizing the damage exposure of one of the strands of the mtDNA molecule during the replication and/or transcription processes have been put forward to explain the high mutation rate of mtDNA, being both transition biased and strand specific [32,34,35]. Damage patterns that are caused by the deamination of the heavy strand lead to the excess of A to G and C to T transitions. Notably, transition hotspot patterns observed in aDNA are similar to those observed to be hypervariable in living populations suggesting that the underlying mechanism as how mutations accumulate in germ line is similar to the build-up of post-mortem damage [36].

The first estimates of mutation rate of the whole mtDNA that were used for the estimation of the TMRCA age were based on the divergence estimates of humans from the chimpanzee outgroup [37,38]. The apparent problem with this phylogenetic approach that used a distant outgroup for calibration of the mtDNA mutation rate was that it produced estimates which were at odds with the mutation rates estimated from pedigree data. In case of the hypervariable regions of the D-loop, several

pedigree studies [39–42] had inferred mutation rates that were up to an order of magnitude higher than the phylogenetic rate [43] (Table 1). More recent studies using high coverage mtDNA sequence data suggest that these differences are mainly due to the detection of heteroplasmic states of somatic mutations which never get fixed in the germ lines [30]. Although it is encouraging to see recent aDNA based studies yielding concordant mutation rates for the whole mtDNA genome, substantial differences are still noted among functional domains of the molecule (Table 1).

Overall, the mutation rate of human mtDNA is over an order of magnitude higher than nuclear rate mainly because of the deamination based high transition rates which are >60 times higher than the transition rate in nuclear genome while the transversion rates are more similar, with only approximately $\times 5$ higher rate than in nuclear genes. To put these rate estimates further into perspective, it is interesting to note that the per-generation mutation rate of mtDNA in humans, approximately 6×10^{-7} , is

approximately $\times 10$ faster than that of *Drosophila* [52] while the per year mutation rate is $\times 100$ slower because the generation time in *Drosophila* is just 10 days.

One of the questions addressed in mtDNA studies on global scale has been the age of the diversity in the locus. Different studies have yielded mtDNA TMRCA age estimates that are young relative to autosomal data and vary (depending on the dating technique and mutation rate being use) in the range of 100 to 200 thousand years ago (kya) [26,37,38,53–55]. These estimates are generally similar [47,56] to those based on Y chromosome or slightly younger [57] when considering the rare Y chromosome haplogroup A00 lineages that were recently found restricted to West Africans. The upper end of these time estimates falls to a period in the African fossil record that is associated with the first appearance of anatomically modern humans [58]. Considering that the time back to TMRCA of a genetic locus is determined primarily by the long term effective population size of the species, the age of TMRCA does not necessarily

Table 1 Pedigree, phylogeny and aDNA-based estimates of mtDNA mutation rates (per bp per year $\times 10^{-8}$)

Domain	mtDNA	coding	RNA	PC1 + 2	PC3	syn	HVS-I	HVS-II	D-loop
No sites	16569	15447	4021	6034	3017	4170	360	316	1122
Pedigree data									
Howell <i>et al.</i> 2003 [40]									47
Sigurdardottir <i>et al.</i> 2000 [42]							21		
Heyer <i>et al.</i> 2001 [39]							39		
Santos <i>et al.</i> 2005 [41]							7.4	43.3	17
Rebolledo <i>et al.</i> 2014 [30]	1.3								7.7
Phylogenetic									
Vigilant <i>et al.</i> 1991 [43]									5.7
Horai <i>et al.</i> 1995 [37]						3.9			
Ingman <i>et al.</i> 2000 [38]		1.7							
Tang <i>et al.</i> 2002 [44]		1							
Mishmar <i>et al.</i> 2003 [45]		1.26							
Kivisild <i>et al.</i> 2006 [26]						3.41			
Loogvali <i>et al.</i> 2009 [27]						3.00			
Soares <i>et al.</i> 2009 [28]	1.67	1.71	0.77	0.89	1.93		16.4	22.9	12.3
Phylogeographic									
Forster <i>et al.</i> 1996 [21]							18		
Ho and Endicott 2008 [46]		2.04							
Poznik <i>et al.</i> 2013 [47]	2.3								
Ancient DNA									
Fu <i>et al.</i> 2013 [48]	2.67	1.57		0.82	3.27				
Brotherton <i>et al.</i> 2013 [49]	2.4								
Fu <i>et al.</i> 2014 [50]	2.53								
Rieus <i>et al.</i> 2014 [51]	2.14		1.01	0.76	3.32		31.4		

^aPC1 + 2; second and third codon positions of protein coding genes; PC3, third codon positions; syn, total number of PC1 and PC3 positions at which transitions will not lead to amino acid replacement which is equal to the sum over all fourfold (2026) and twofold (2144) degenerate sites.

inform us about a biologically significant event, such as the origin of the species, unless the species went through a speciation bottleneck and was founded from a very small number of individuals. Genetic and fossil evidence for such major founder event after the split of human and Neanderthal/Denisovan ancestors or a sudden change in morphology at this critical period of time has been lacking [59,60].

The need for whole mtDNA sequences

Two major limitations of the RFLP approach and D-loop sequencing were the small number of bases and therefore limited molecular resolution for distinguishing variation at sub-regional level, and, secondly, low robustness of the phylogenetic inferences caused by the high mutation rate of the hypervariable regions. Hypervariable positions are known to undergo multiple parallel mutations in many lineages and this parallelism becomes a significant confounding factor even within a short time scale of few tens of thousands of years of evolutionary history. These recurrent mutations generate phylogenetic uncertainty, also known as homoplasy, which even in case of the presence of only a few tens of such sites and sample size of few tens of individuals can lead to the problem of millions of trees having equal length or likelihood to be consistent with the data. Network approaches [61] were developed to visualize the complexity of parallel relationships among the mitochondrial lineages but for solving them more data from the conservative regions of mtDNA were required. Further improvements of the classical Sanger sequencing technology in the end of the last century enabled the sequencing of the whole mtDNA for the purpose of human evolutionary studies. Progress in the technology use was significantly motivated by our need to understand the genetics of disease.

When deleterious mutations occur over time natural selection prohibits them reaching high frequency and removes them from circulation. One of the key drivers of the study of full mtDNA sequences has been medical genetics and, in particular, the need to understand the genetic basis of mitochondrial disorders and deleterious mutations. Compared to our nuclear genes, those residing in mitochondria do not have introns and much non-coding sequence around them - the whole mitochondrial genome is densely (93%) packed with protein coding, ribosomal and transport RNA genes (Figure 1). A large proportion of positions in these genes are known to be highly conserved across different species, implying strong purifying selection, and invariable in large human cohorts likely because of being fatally deleterious or associated with disease (see MITOMAP [62]). All mitochondrial genes are viably important and diseases associated with impaired function of mitochondrial protein coding genes affect primarily muscular and neural function (for review, see [63]).

Therefore, unsurprisingly, the first studies to employ the whole mtDNA sequencing approach were those attempting to uncover the causative mutations of neurodegenerative diseases [64-66].

Besides the motivation for disease studies the sequencing of whole mtDNA provided also the means for getting statistically better supported phylogenetic trees to study the history of human populations. The first worldwide survey of mtDNA whole genome sequences [38] showed with a robust bootstrap support of the internal branches that the root of the human mtDNA variation lies in Africa with TMRCA date of $171,500 \pm 50,000$ years and that the age of the youngest clade with African and non-African sequences was $52,000 \pm 27,500$ years. Other whole mtDNA studies, for example [26,45,56,67-69], based on global sampling have generally agreed with these structural findings and revealed more details of the regional patterns of diversity, time scale of the accumulation of diversity, and the female effective population size changes over time. It should be noted, though, before exploring the geographic distribution of its variation that mtDNA molecule, however well resolved its phylogeny and no matter how large the sample size, remains to be just one single genetic locus which is subject to large stochastic variation and that population level inferences of demographic history require the synthesis of evidence from many loci.

Distribution of variation in mtDNA genomes among human populations

Compared to the estimates based on autosomal data the observed differences in mitochondrial sequences among human populations on a global scale are significantly higher and second only to the differences based on Y chromosomes, with Africa showing the highest within region diversity and Native Americans having the lowest [56]. As it has been repeatedly shown with ever increasing sample sizes that are reaching tens of thousands of individuals now [68], the root of the mtDNA phylogeny and the most diverse branches are restricted to African populations (Figure 2). Using the maximum molecular resolution enabled by the analysis of whole mtDNA genomes, the first seven bifurcations in this tree, in fact, define the distinction of strictly sub-Saharan African branches (L0-L6) from those that are shared by Africans and non-African populations. Analyses of whole mtDNA sequences of sub-Saharan Africans have revealed early, ca 90 to 150 thousand years (ky) old divergence of the L0d and L0k lineages that are specific to the Khoisan populations from South Africa and it has been estimated that during this time period at least six additional lineages existed in Africa with living descendants [53,54]. In contrast to the overall high basal clade diversity and geographic structure some terminal branches from

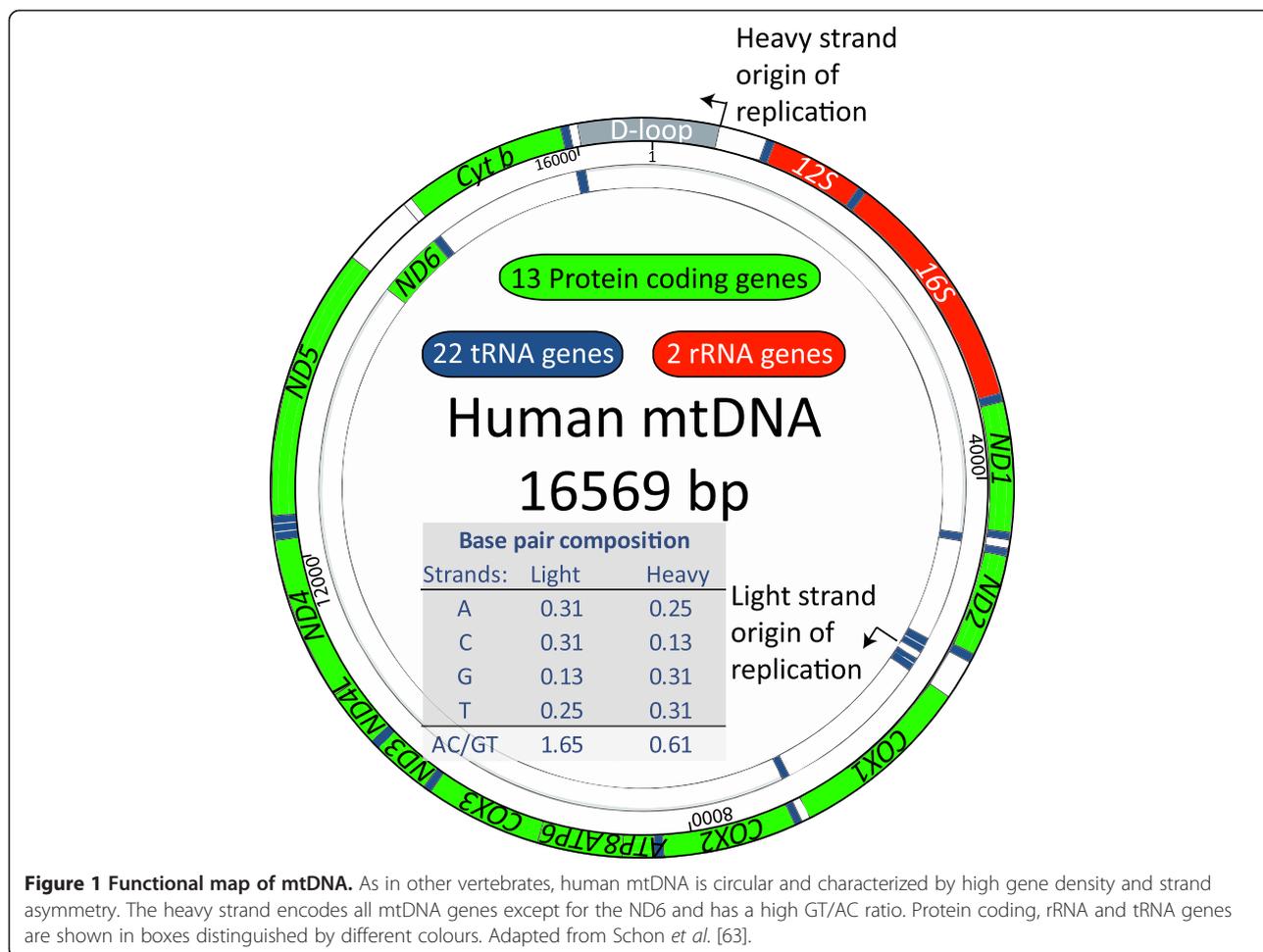
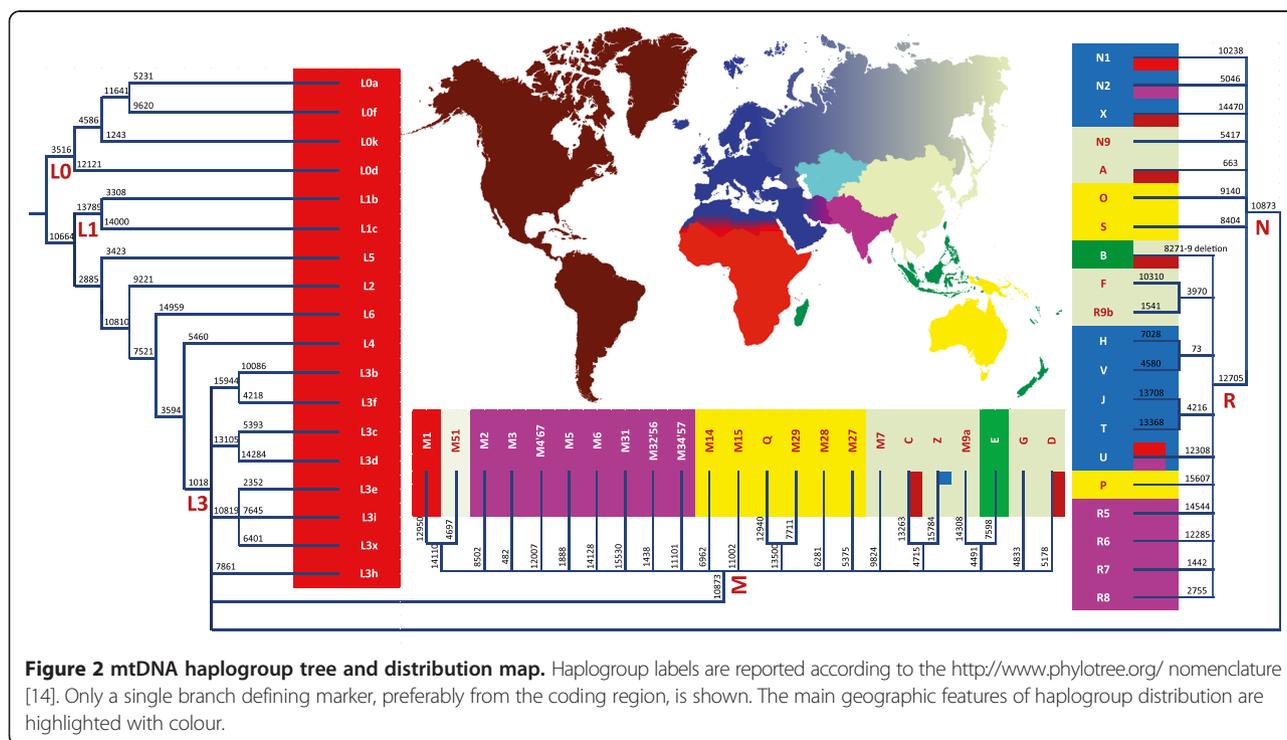


Figure 1 Functional map of mtDNA. As in other vertebrates, human mtDNA is circular and characterized by high gene density and strand asymmetry. The heavy strand encodes all mtDNA genes except for the ND6 and has a high GT/AC ratio. Protein coding, rRNA and tRNA genes are shown in boxes distinguished by different colours. Adapted from Schon *et al.* [63].

haplogroups L0a, L1c, L2a, and L3e show recent coalescent times and wide geographical distribution in Africa, likely due to the recent Bantu expansion [70-72]. Given the complexity of admixture of the Bantu-speaking populations the use of whole mtDNA sequences in these studies have been instrumental in revealing the distinct autochthonous sources and ancient substructure at the background of the overall high genetic homogeneity of the Bantu speakers [70]. Outside Africa, haplogroup L0-L6 lineages are extremely rare and restricted to geographic areas that have received historic gene flow from Africa, such as Mediterranean Europe, West Asia, and Americas. On the basis of analyses of high resolution whole mtDNA sequences it has been estimated that approximately two thirds of the rare African L lineages that are found at combined frequency of <1% in Europe were brought in from Africa during the Roman times, Arab conquests and Atlantic slave trade while just one third are more likely to have been introduced earlier during pre-historic times [73].

The fact that virtually every non-African mtDNA lineage derives from just one of the two sub-clades of the African haplogroup L3 (Figure 2) has been interpreted as an evidence of a major bottleneck of mtDNA diversity at the onset of the out of Africa dispersal [74]. The magnitude of this bottleneck has been estimated from the whole mtDNA sequence data yielding the estimates of the effective population size which range between several hundred [75] and only few tens of females [56]. The separation of these two sub-clades, M and N, from their African sister-clades in L3 can be dated back to 62 to 95 kya [48] whereas the internal coalescent time estimates of the M and N founders have been estimated in the range of 40 to 70 ky [26,28,75] and suggest that their dispersal occurred probably after rather than before the eruption of Mount Toba 74 kya in Indonesia, one of the Earth's largest known volcanic events in human history. Archaeological evidence from Jurreru River valley, India, has shown the presence of artefacts right above and below the layers of ash associated with the Toba eruption [76]. It is not clear whether the makers of these



artefacts were archaic or anatomically modern humans. As in case of the global TMRCA estimate considered above the wide error ranges around the age estimates of haplogroups M and N reflect primarily the uncertainties of the mutation rate - in relative terms, the age estimates of M and N, as determined from whole mtDNA sequences form approximately one third of the total depth of the global mtDNA tree. Claims for relatively recent, post-Toba, time depth of the non-African founder-haplogroups have been recently supported by the aDNA evidence of the 45 kya Ust-Ishim skeleton whose whole mtDNA sequence falls at the root of haplogroup R [50]. While haplogroups M and N are widely spread in Asia, Australia, Oceania and Americas, the geographic distribution of each of their sub-clades has more specific regional configuration (Figure 2).

In Eurasia haplogroups U, HV, JT, N1, N2 and X are today common in Europe, Southwest Asia and North Africa [77]; haplogroups R5-R8, M2-M6 and M4'67 are restricted to South Asia [78], while haplogroups A-G, Z and M7-M9 are widespread in East Asia [79] (Figure 2). Despite the clear and distinct geographic spread patterns in extant populations it is not simple and straightforward to make inferences about the origin of these patterns and to associate the haplogroup labels with specific prehistoric events or time periods. Phylogeographic inferences made from extant variation both at low and high molecular resolution have suggested that majority of the haplogroups that are common today throughout Europe derive from the Late Glacial

re-colonization event [77]. ADNA evidence, however, shows [80] that only a subset of haplogroup U variation is likely to have ancestry in pre-Neolithic Europe while other haplogroups are likely to be related with more recent episodes of gene flow and demographic events which, apparently, have quite dramatically changed the genetic landscape of the region in the past 10,000 years. ADNA analyses of the nuclear genomes of Mesolithic and Neolithic samples from Europe have suggested that the discontinuity observed in central European mtDNA types may be echoed by the appearance approximately 4,500 years ago in Europe of an ancient Near Eastern component in the autosomal genes [81].

MtDNA variation in Native Americans variation primarily falls to haplogroups A to D; X and that with the exclusion of X form a subset of the East Asian diversity [10]. Since the initial attempts to define the number of Native American founder lineages within these five basic haplogroups at low resolution attainable with RFLP and hypervariable region sequencing approaches [10,21], at least 16 sub-clades have been assigned now the founder status on the basis of whole mtDNA genome sequence analysis [82-87]. The spread of these sub-clades in North and South America has been associated with at least three distinct demographic events: (1) the main wave of the spread of the ancestors of both North and South American native populations 15–18 kya involving nine Pan-American founders A2*, B2*, C1b, C1c, C1d*, C1d1, D1, D4h3a, and D4e1c, followed potentially approximately at the same time by an inland route dispersal of

C4c, X2a and X2g carriers to the east coast of the USA; (2) the spread of Paleo-Eskimo D2a [88] lineages ca 5 kya along the Arctic through northern Canada and Greenland, which were replaced, in the same region, by (3) the spread of Neo-Eskimos carrying A2a, A2b, and D3 lineages. Phylogeographic inferences from modern whole mtDNA sequence data associating the spread of haplogroup A2a lineages with Paleo-Eskimos [83] have not been supported by aDNA evidence which instead points to all available skeletal evidence that is associated with the Paleo-Eskimo cultures Saqqaq and Dorset having unusually low mtDNA diversity restricted only to haplogroup D2a [89].

The whole mtDNA sequencing of Oceanians has revealed a number of distinct mtDNA lineages that were undistinguishable at lower resolution from those spread in Mainland Asia. The peopling of Oceania has been modelled to involve at least two major demographic events: firstly, the initial settlement of Sahul (Papua New Guinea and Australia) by anatomically modern humans explains the presence of mtDNA haplogroups M14-M15, M27-M29, Q, P, O, and S only in Australia and Melanesia; secondly, this was followed by a more recent Holocene dispersal of the populations speaking Austronesian languages who would have extended widely the geographic distribution of haplogroup B4a1a1 lineages [90]. Although the high frequency of an intergenic 9-bp deletion together with a specific D-loop motif, that is characteristic to the haplogroup B4a1a1 mtDNA molecules of all Austronesian speaking populations, was noticed already in the low resolution studies of 1990's, the employment of whole mtDNA sequencing, in combination with aDNA evidence, has made it possible now to narrow substantially down the geographic regions in Island Southeast Asia that carried the sequences directly ancestral to those of the majority of Austronesians [91-94].

The future of whole mtDNA analyses in the era of next generation sequencing of whole nuclear genomes

Now that tens of thousands of whole mitochondrial genome sequences are already publicly available and cover virtually all extant population of the world, is there still a need for more mtDNA data and room for novel findings? Whole mitochondrial sequencing certainly continues to have an important role in forensics, in medical genetics and in ancestry and genealogy related applications because of the specific needs for mtDNA evidence in these fields. Although questions about demographic history of populations, natural selection, the extent of admixture and many other relevant aspects of genetic research of human populations can now be addressed at the level of whole genome sequences, mtDNA has continued to play an important role in the evolutionary genetic studies. MtDNA sequence variation is used in

aDNA studies for the estimation of contamination levels (for example [60]) and, in turn, the accumulating aDNA evidence allows us to get increasingly more accurate insights into the complexities of mitochondrial mutation rate (Table 1). ADNA evidence combined with data from extant populations enables us, as described above, to better understand the temporal dynamics of the change of genetic diversity in regions such as Europe [80,81].

Whole mtDNA sequencing will continue to inform us about the sex specific patterns of human migrations and admixture. Consistent with the evidence from nuclear genetic loci and historical records whole mtDNA sequences of the Siddis from India have been shown to include a substantial proportion of lineages that have the closest affinity with those of the Bantu speaking populations of East Africa [95]. Because this admixture dates back only a couple of centuries it is not surprising that both sex specific loci and autosomes show consistent patterns. In contrast, other South Asian populations, such as Santhals and Mundas who speak Austroasiatic languages, have maintained the evidence of their admixed origins and Southeast Asian descent only in their Y chromosome while their mtDNA lineages cluster most closely with neighboring Indian populations [96].

The inferences of long-term effective population size from whole mtDNA and Y chromosome sequence data are continuing to provide new insights into the social behaviour of the past populations. The comparisons of female (N_f) and male (N_m) effective population size estimates suggest that N_f/N_m ratio has been higher than 1 over the course of our evolutionary history and showing an increase in more recent times [56]. Several factors can explain N_f/N_m deviations from 1, including selection, mobility and residence patterns. Analyses of populations from the Indonesian archipelago have shown that during the historic times the contacts with foreigners, such as Chinese, Indians, Arabs and Europeans, have left a noticeable imprint in the Y-chromosome variation of these indigenous populations whereas these patterns are not reflected in their mtDNA data. Whole mtDNA sequence data, on the other hand, have retained more clearly the evidence of a major geographic expansion of specific founder types, suggesting that in pre-historic times the women were more mobile than men in spreading their mitochondria from island to island [97]. This together with the findings of sex specific patterns of the Asian versus Papuan ancestry components suggests that the predominant residence pattern of the proto-Oceanic speaking populations who spread the Austronesian languages in the Pacific may have been matrilocal [90,92,98-100]. Matrilocal residence in today's world is rare and restricted to a small number of populations, some of which have been studied to explore the effect of residence patterns on our genetic diversity [101].

Due to prevailing patrilocality the genetic differences among population are typically higher for Y chromosome than for mtDNA, although this effect has been mostly noticed at local rather than global scale [102]. It has been shown that it is crucial to use the full power of whole mtDNA sequences to reveal such differences [103].

Conclusions

In sum, mtDNA evidence will probably continue to be important for various facets of population genetic research in the coming decades. Because of its high copy number, it will be used routinely in aDNA studies for the preliminary assessment of the quality of DNA preservation and for the evaluation of contamination. And, because of its maternal inheritance, it will continue to be informative tool for the study of sex-specific patterns in and among human populations.

Abbreviations

aDNA: ancient DNA; HVS: hypervariable segment; mtDNA: mitochondrial DNA; N_f : female effective population size; N_m : male effective population size; RFLP: restriction fragment length polymorphisms; TMRCA: the most recent common ancestor.

Competing interests

The author declares that he has no competing interests.

Received: 26 November 2014 Accepted: 4 February 2015

Published online: 10 March 2015

References

- Hutchison 3rd CA, Newbold JE, Potter SS, Edgell MH. Maternal inheritance of mammalian mitochondrial DNA. *Nature*. 1974;251:536–8.
- Brown WM, George Jr M, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A*. 1979;76:1967–71.
- Michaels GS, Hauswirth WW, Laipis PJ. Mitochondrial DNA copy number in bovine oocytes and somatic cells. *Dev Biol*. 1982;94:246–51.
- Piko L, Matsumoto L. Number of mitochondria and some properties of mitochondrial DNA in the mouse egg. *Dev Biol*. 1976;49:1–10.
- Hagstrom E, Freyer C, Battersby BJ, Stewart JB, Larsson NG. No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucleic Acids Res*. 2014;42:1111–6.
- Merrifether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, et al. The structure of human mitochondrial DNA variation. *J Mol Evol*. 1991;33:543–55.
- Kurland CG, Collins LJ, Penny D. Genomics and the irreducible nature of eukaryote cells. *Science*. 2006;312:1011–4.
- Stewart JB, Larsson NG. Keeping mtDNA in shape between generations. *PLoS Genet*. 2014;10:e1004670.
- Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. *Nature*. 2013;499:471–5.
- Torrioni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, et al. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet*. 1993;53:563–90.
- Torrioni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, et al. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol*. 1994;93:189–99.
- Torrioni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet*. 1994;55:760–76.
- Chen YS, Torrioni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet*. 1995;57:133–49.
- van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 2009;30:E386–394.
- Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL, et al. Ethnic variation in Hpa I endonuclease cleavage patterns of human mitochondrial DNA. *Proc Natl Acad Sci U S A*. 1981;78:5768–72.
- Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature*. 1987;325:31–6.
- Excoffier L, Langaney A. Origin and differentiation of human mitochondrial DNA. *Am J Hum Genet*. 1989;44:73–85.
- Saitou N, Omoto K. Time and place of human origins from mt DNA data. *Nature*. 1987;327:288.
- Wolpoff MH. Multiregional evolution—the fossil alternative to Eden. In: Mellars P, Stringer C, editors. *Human Revolution*. Princeton: Princeton University Press; 1989. p. 62–108.
- Maddison DR. African origin of human mitochondrial-DNA reexamined. *Syst Zool*. 1991;40:355–63.
- Forster P, Harding R, Torroni A, Bandelt HJ. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet*. 1996;59:935–45.
- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, et al. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics*. 1992;130:139–52.
- Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M. Evolutionary history of the COII/trnAlys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol*. 1995;12:604–15.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet*. 2000;67:1251–76.
- Lynch M, Koskella B, Schaack S. Mutation pressure and the evolution of organelle genomic architecture. *Science*. 2006;311:1727–30.
- Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, et al. The role of selection in the evolution of human mitochondrial genomes. *Genetics*. 2006;172:373–87.
- Loogvali EL, Kivisild T, Margus T, Villems R. Explaining the imperfection of the molecular clock of hominid mitochondria. *PLoS One*. 2009;4:e8260.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohlf A, et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*. 2009;84:740–59.
- Soares P, Abrantes D, Rito T, Thomson N, Radivojac P, Li B, et al. Evaluating purifying selection in the mitochondrial DNA of various mammalian species. *PLoS One*. 2013;8:e58993.
- Rebolledo-Jaramillo B, Su MS, Stoler N, McElhoo JA, Dickens B, Blankenberg D, et al. Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A*. 2014;111:15474–9.
- Ye F, Samuels DC, Clark T, Guo Y. High-throughput sequencing in mitochondrial DNA research. *Mitochondrion*. 2014;17:157–63.
- Frank AC, Lobry JR. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene*. 1999;238:65–77.
- Bandelt H-J, Kong Q-P, Richards MB, Macaulay V. Estimation of mutation rates and coalescence times: some caveats. In: Bandelt H-J, Richards MB, Macaulay V, editors. *Human Mitochondrial DNA and the Evolution of Homo Sapiens*. Volume 18. Berlin, Heidelberg: Springer-Verlag; 2006. p. 47–90.
- Xia X. DNA replication and strand asymmetry in prokaryotic and mitochondrial genomes. *Current Genomics*. 2012;13:16–27.
- Lin Q, Cui P, Ding F, Hu S, Yu J. Replication-associated mutational pressure (RMP) governs strand-biased compositional asymmetry (SCA) and gene organization in animal mitochondrial genomes. *Current Genomics*. 2012;13:28–36.
- Gilbert MT, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, et al. Distribution patterns of postmortem damage in human mitochondrial DNA. *Am J Hum Genet*. 2003;72:32–47.
- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A*. 1995;92:532–6.
- Ingman M, Kaessmann H, Pääbo S, Gyllenstein U. Mitochondrial genome variation and the origin of modern humans. *Nature*. 2000;408:708–13.
- Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet*. 2001;69:1113–26.

40. Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet.* 2003;72:659–70.
41. Santos C, Montiel R, Sierra B, Bettencourt C, Fernandez E, Alvarez L, et al. Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). *Mol Biol Evol.* 2005;22:1490–505.
42. Sigurgardottir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P. The mutation rate in the human mtDNA control region. *Am J Hum Genet.* 2000;66:1599–609.
43. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. African populations and the evolution of human mitochondrial DNA. *Science.* 1991;253:1503–7.
44. Tang H, Siegmund DO, Shen P, Oefner PJ, Feldman MW. Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. *Genetics.* 2002;161:447–59.
45. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, et al. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A.* 2003;100:171–6.
46. Ho SY, Endicott P. The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am J Hum Genet.* 2008;83:142–6. author reply 146–147.
47. Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science.* 2013;341:562–5.
48. Fu Q, Mitnik A, Johnson PL, Bos K, Lari M, Bollongino R, et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol.* 2013;23:553–9.
49. Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, et al. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun.* 2013;4:1764.
50. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature.* 2014;514:445–9.
51. Rieux A, Eriksson A, Li M, Sobkowiak B, Weinert LA, Warmuth V, et al. Improved calibration of the human mitochondrial clock using ancient genomes. *Mol Biol Evol.* 2014;31:2780–92.
52. Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol.* 2008;6:e204.
53. Barbieri C, Vicente M, Rocha J, Mpoloka SW, Stoneking M, Pakendorf B. Ancient substructure in early mtDNA lineages of southern Africa. *Am J Hum Genet.* 2013;92:285–92.
54. Behar DM, Villemers R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, et al. The dawn of human matrilineal diversity. *Am J Hum Genet.* 2008;82:1130–40.
55. Rito T, Richards MB, Fernandes V, Alshamali F, Cerny V, Pereira L, et al. The first modern human dispersals across Africa. *PLoS One.* 2013;8:e80031.
56. Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Invest Genet.* 2014;5:13.
57. Mendez FL, Krahn T, Schrack B, Krahn AM, Veeramah KR, Woerner AE, et al. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am J Hum Genet.* 2013;92:454–9.
58. McDougall I, Brown FH, Fleagle JG. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature.* 2005;433:733–6.
59. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature.* 2011;475:493–6.
60. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science.* 2012;338:222–6.
61. Bandelt HJ, Forster P, Sykes BC, Richards MB. Mitochondrial portraits of human populations using median networks. *Genetics.* 1995;141:743–53.
62. Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.* 2007;35:D823–828.
63. Schon EA, DiMauro S, Hirano M. Human mitochondrial DNA: roles of inherited and somatic mutations. *Nature Rev Genet.* 2012;13:878–90.
64. Ozawa T, Tanaka M, Ino H, Ohno K, Sano T, Wada Y, et al. Distinct clustering of point mutations in mitochondrial DNA among patients with mitochondrial encephalomyopathies and with Parkinson's disease. *Biochem Biophys Res Commun.* 1991;176:938–46.
65. Ikebe S, Tanaka M, Ozawa T. Point mutations of mitochondrial genome in Parkinson's disease. *Brain Res Mol Brain Res.* 1995;28:281–95.
66. Yoneda M, Tanno Y, Horai S, Ozawa T, Miyatake T, Tsuji S. A common mitochondrial DNA mutation in the t-RNA(Lys) of patients with myoclonus epilepsy associated with ragged-red fibers. *Biochem Int.* 1990;21:789–96.
67. Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* 2001;2:13.
68. Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet.* 2012;90:675–84.
69. Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, et al. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet.* 2002;70:1152–71.
70. Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, et al. Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. *PLoS One.* 2014;9:e99117.
71. Marks SJ, Montinaro F, Levy H, Brisighelli F, Ferri G, Bertoni S, et al. Static and moving frontiers: the genetic landscape of Southern African Bantu-speaking populations. *Mol Biol Evol.* 2015;32:29–43.
72. de Filippo C, Bostoen K, Stoneking M, Pakendorf B. Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc Biol Sci/R Soc.* 2012;279:3256–63.
73. Cerezo M, Achilli A, Olivieri A, Perego UA, Gomez-Carballa A, Brisighelli F, et al. Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res.* 2012;22:821–6.
74. Underhill PA, Kivisild T. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet.* 2007;41:539–64.
75. Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, et al. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science.* 2005;308:1034–6.
76. Petraglia M, Korisettar R, Boivin N, Clarkson C, Ditchfield P, Jones S, et al. Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. *Science.* 2007;317:114–6.
77. Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, et al. The archaeogenetics of Europe. *Curr Biol.* 2010;20:R174–183.
78. Chaubey G, Metspalu M, Kivisild T, Villemers R. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays.* 2007;29:91–100.
79. Stoneking M, Delfin F. The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol.* 2010;20:R188–193.
80. Brandt G, Haak W, Adler CJ, Roth C, Szecsenyi-Nagy A, Karimnia S, et al. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science.* 2013;342:257–61.
81. Lazaridis I, Patterson N, Mitnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature.* 2014;513:409–13.
82. Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, et al. The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One.* 2008;3:e1764.
83. Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshier Kashani B, et al. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci U S A.* 2013;110:14308–13.
84. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, et al. Beringian standstill and spread of Native American founders. *PLoS One.* 2007;2:e829.
85. O'Rourke DH, Raff JA. The human genetic history of the Americas: the final frontier. *Curr Biol.* 2010;20:R202–207.
86. Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, et al. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol.* 2009;19:1–8.
87. Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Hooshier Kashani B, et al. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res.* 2010;20:1174–9.
88. Gilbert MT, Kivisild T, Gronnow B, Andersen PK, Metspalu E, Reidla M, et al. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science.* 2008;320:1787–9.

89. Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, et al. The genetic prehistory of the New World Arctic. *Science*. 2014;345:1255832.
90. Kayser M. The human genetic history of Oceania: near and remote views of dispersal. *Curr Biol*. 2010;20:R194–201.
91. Soares P, Rito T, Trejaut J, Mormina M, Hill C, Tinkler-Hundal E, et al. Ancient voyaging and Polynesian origins. *Am J Hum Genet*. 2011;88:239–47.
92. Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, et al. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol*. 2005;3:e247.
93. Ko AM, Chen CY, Fu Q, Delfin F, Li M, Chiu HL, et al. Early Austronesians: into and out of Taiwan. *Am J Hum Genet*. 2014;94:426–36.
94. Duggan AT, Stoneking M. Recent developments in the genetic history of East Asia and Oceania. *Curr Opin Genet Dev*. 2014;29C:9–14.
95. Shah AM, Tamang R, Moorjani P, Rani DS, Govindaraj P, Kulkarni G, et al. Indian Siddis: African descendants with Indian admixture. *Am J Hum Genet*. 2011;89:154–61.
96. Chaubey G, Metspalu M, Choi Y, Magi R, Romero IG, Soares P, et al. Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol Evol*. 2011;28:1013–24.
97. Tumonggor MK, Karafet TM, Hallmark B, Lansing JS, Sudoyo H, Hammer MF, et al. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. *J Hum Genet*. 2013;58:165–73.
98. Hage P, Marck J. Matrilineality and the melanesian origin of Polynesian Y chromosomes. *Curr Anthropol*. 2003;44:S121–7.
99. Jordan FM, Gray RD, Greenhill SJ, Mace R. Matrilocal residence is ancestral in Austronesian societies. *Proc Biol Sci/R Soc*. 2009;276:1957–64.
100. Tumonggor MK, Karafet TM, Downey S, Lansing JS, Norquest P, Sudoyo H, et al. Isolation, contact and social behavior shaped genetic diversity in West Timor. *J Hum Genet*. 2014;59:494–503.
101. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocality. *Nat Genet*. 2001;29:20–1.
102. Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat Genet*. 2004;36:1122–5.
103. Gunnarsdottir ED, Nandineni MR, Li M, Myles S, Gil D, Pakendorf B, et al. Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocality groups from Sumatra. *Nat Commun*. 2011;2:228.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

