Positive selection on mitochondrial haplogroup M9a in Tibet

Siqi Huang, Chuan-Chao Wang, Hui Li*

State Key Laboratory of Genetic Engineering and Ministry of Education Key, Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200438, China.

***Corresponding author. Tel: +86 21 51630427 (H.Li). E-mail addresses: LHCA@Fudan.edu.cn**

Abstract: Tibetan is considered as a good population example for analyzing the mechanism of possible adaptive evolution in the special highland environment (e.g., high attitude, hypoxia, and extreme cold). There might be many genetic variations contributed to this adaption. Mitochondrial DNA (mtDNA) variations might be important for the adaption given its role in coding core subunits of oxidative phosphorylation in mitochondria. Mitochondrial haplogroup M9a reaches the highest frequency in Tibetans and therefore is most probable to be selected for. In this study, we collected 152 complete mtDNA genomes belonging to haplogroup M9a. By whole sequence analyses we found the existence of adaptive selection for *ND5* (C12362T) and *CYB* (A15671G) genes in Tibetans. Meanwhile, we predicted the enhancement of stability upon mutations. These mutations may have certain effect on OXPHOS function and might be important in highlands adaptation.

Keyword: mitochondrial genome, adaptive selection, stability changes, hypoxia adaption

Introduction

Tibetan population is a good candidate for studying the genetic mechanism of highland adaptation, since they live on the highest plateau in the world. In this special environment, including thin oxygen and extreme coldness, Tibetan people are also famous for their living ability without suffering from chronic mountain diseases and frostbite. All the above traits are distinct characteristics of well adaptation to extreme environment of highland.

To answer how Tibetans adapt well to the extreme environment of highland, previous studies of humans on the Tibetan plateau have identified statistical signatures of positive selection in different sets of loci. Tibetans have also shown different genetic patterns compared with lowland populations on several genes, such as *HIF-1α*, a hypoxia-inducible transcription factor (Suzuki et al. 2003), *endothelial nitric oxide synthase* (*eNOS*) (Droma et

al. 2006), *ACE* (Droma et al. 2008), and *EPAS1*, which directly targets on *lysyl oxidase* (*LOX*) gene and be responsible for extracellular matrix protein cross-linking of amnion (Hanaoka et al. 2012). Functional mutations in these genes can all indicate the adaptations in highland. However, given the fact of genetic diversity, Tibetans' good ability of adapting to highland may be caused by mutations more than currently studied. Besides, since that trait is relevant to the generation of energy and the respiratory chain, genetic markers closely linked to these functions might be studied as well.

Thus, we chose mtDNA haplogroup as the candidate of our study. First, mtDNA is a better genetic marker, since it presents in thousands of copies per cell, and is capable of encompassing various percentages of mutant and normal molecules (or heteroplasmy) (Wallace et al. 2005). Such characteristics of mtDNA have provided variable resources for the operation of natural selection. In addition, since selection may have caused the haplogroup to flourish in the new environment, different mtDNA haplogroups may be functionally different in its specific region (Ruiz-Pesini et al. 2004; 2006). Thus, we suppose that mtDNA haplogroup is a good candidate for studying the relationship between Tibetan population and their high latitude living conditions.

In former researches, Sherpa people, one of the special population living on the Tibetan Plateau with extraordinary mountaineering, was chosen as a good candidate for hypoxia adaption (Suzuki et al. 2003; Droma et al. 2006; Droma et al. 2008; Hanaoka et al. 2012). Recently, a research has utilized mtDNA to unveil the secret why Sherpa population could survive in highland, and established the connection between mtDNA haplogroup and highland adaption (Kang et al. 2013). In that research, haplogroup M9a, A4e3a (18.4%) and C4a3b1 (18.4%) showed high frequency among all the samples. Especially, 54% samples belong to haplogroup M, and 34% of those belong to M9a, consisting of the majority parts of Sherpa population. After careful analysis, two nonsynonymous mutations (G3745A and T4216C) were found in two Sherpa-specific lineages A4e3a and C4a3b1. They might have certain effect on Complex I function and therefore are candidate variants for mtDNA adaptation of Sherpas. However, no candidate variants for mtDNA adaption of haplogroup M9 were found. Actually, haplogroup M9a'b traced its origin in North Asia/ northern China (Peng et al. 2011) and its sub-haplogroup M9a is well-known for its high frequency in Tibetan Plateau, which indicates the good fitness between this haplotype and highland environment. It is possible that during the course of human evolution, other events like genetic drift has interfered the signal of selection, so that variants of haplogroup M9a could not be detected in the above research. Thus, we suppose that study on people belonging to haplotype M9a could better illustrate the mechanism how Tibetans adapt to highland environment.

Moreover, to better illustrate the adaptation during peopling in Tibet Plateau, the analysis of stability changes

upon mtDNA variants is of great importance. The mtDNA encompasses the 13 core protein subunits that define the efficiency of the mitochondrial energy generating system oxidative phosphorylation (OXPHOS) (Wallace et al. 2005; Wallace et al. 2007). Hence, the stability changes on those proteins might influence the systemic function of energy production, metabolic rate modulation, oxygen utilization, and hypoxia adaptation.

To identify the potential adaptive variants of whole mtDNA, in this study, we sequenced the entire mitochondrial genomes specifically of 152 people belonging to haplogroup M9a. After manually picking mutations of different population, we managed to find unique and frequent nonsynonymous mutations in Tibet population according to the result of phylogenetic tree and median-joining network. The SVM algorithm was further used to predict the stability changes upon the mutations we found. Also, we constructed Bayesian Skyline Plots (BSP) and the time to the most recent common ancestor (TMRCA) of Tibet population to reveal the operation of positive selection.

According to our result, the nonsynonymous mutations (T12362C andG15671A) harbored on Tibetan population are included in the coding-region of mtDNA, and might have certain effect on Complex I and Complex V function of respiratory chain. Thus, we proposed that they were candidate variants for mtDNA adaptation of Tibetans and might reveal the mechanism how Tibetans adapt to plateau well.

Method

Populations and Samples

We collected 152 samples belonging to haplogroup M9a and its sub-haplogroups from published papers. All these samples were maternally unrelated and mainly from Tibet, Japan, and Southeast Asia (Figure 1).

Haplogroup Assignment and data analysis

We aligned complete sequences to RSRS by Mega5.05 (Tamura et al. 2011) and manually checked, then assigned them to the haplogroups according to PhyloTree Build 16 (van Oven and Kayser 2009; Behar et al. 2012). As in PhyloTree, positions 309.1C(C), 16182C, 16183C, 16193.1C(C), and 16519 were not used for haplogroup assignment as these were subject to highly recurrent mutations.

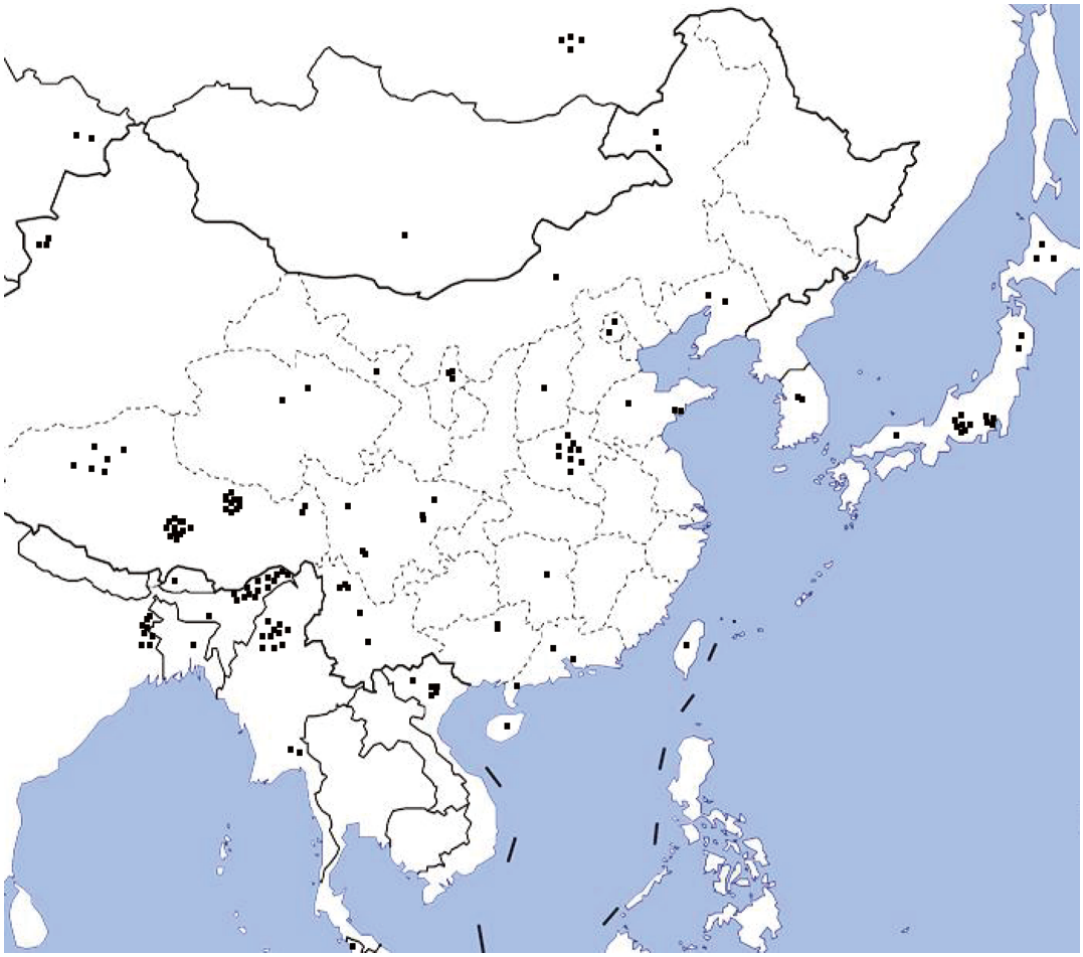

Figure 1 **Geographic locations of populations surveyed in this study.** For more details regarding the populations, refer to **Appendix A**.

We then excluded hyper variable region (**Appendix A**) and made the Neighbor-joining (NJ) unrooted tree of haplogroup M9a by Mega5.05 (Tamura et al. 2011) with 152 samples. The model we adopted was Kimura 2-parameter, including both transitions and transversions and completely deleting gaps or missing data, and we ran 100 replications of bootstrapping method. We then constructed the median-joining network by Network v4.6 (Bandelt et al. 1999) also using the 152 samples with their coding region sequences.

Neutralitytest

We first used DnaSP 5.10.01 (Libradoet al. 2009; Rozas 2009) to detect whether the samples have experienced selection in both Tibet and Japan population. Tajima's D test, Fu and Li's test were all used to confirm the results (Tajima 1989; Fu et al. 1993).

We then constructed Heat map to test if Tibet population bears much stronger effect of adaptive selection. We set the variable D as the difference value between neutrality result of Tibet and Japan (D=#Tibet-#Japan), and

used different colors to indicate the D values on eight genes (*CO1*, *CO2*, *ATP6*, *CO3*, *ND4*, *ND5*, *ND6*, and *CYB*) by five methods in all. The rest five genes were excluded in this research due to the lack of polymorphisms.

Selection Analysis by Population and Gene-specific Variation

According to the results of Heat map and mutation frequency, we supposed that *ND5* and *ATP8* variations might experience adaptive selection. We then conducted the nonsynonymous and synonymous analysis on *ND5* and *ATP8* genes to figure out the actual effect of selection on mtDNA variation (Sun et al. 2007; Mishmaret al. 2003; Ruiz-Pesini et al. 2004; Kivisild et al. 2006; Elson et al. 2004; Yang et al. 2000). In general, when positive selection dominates, the Ka/Ks (Nonsynonymous and synonymous substitutions) ratio is greater than 1; in this case, persity at the amino acid level is favored, likely due to the fitness advantage provided by the mutations. Conversely, when negative selection dominates, the Ka/Ks ratio is less than 1; in this case, most amino acid changes are deleterious, and therefore, are selected against. When the positive and negative selection forces balance each other, the Ka/Ks ratio is close to 1.The sites of nonsynonymous and synonymous substitutions were identified by DnaSP 5.10.01 (Librado P et al. 2009; Rozas J. 2009).

Structural Prediction and Analysis

We then picked out specific variations in Japan and Tibet and their coding proteins from the network and phylogenetic tree. We then used SSPro and SSPro8 to predict the secondary structures of these proteins (peptides) (Pollastri et al. 2002). Based on the predictions, we manually compared the differences caused by the variations. For high structures, the coordinates of alpha-C were first calculated by 3Dpro in PDB forms (Cheng et al. 2005), while information about other atoms was provided by MaxSprout (Holm et al. 1991). *ND2*, *CYB*, *ATP6*, and *ATP8* were simulated with their full length protein. Due to the test limitation of protein length, the structure changes of *ND4* were predicted based on 60-459aa from its N-terminal, and the first 400aa from N-terminal of *ND5* was taken into consideration. All the results were subsequently analyzed using Rasmol (Sayle et al. 1995). Stability change of each protein upon single-site mutations was calculated using SVM algorithm, which was exploited by Capriotti et al. (2005).

Coalescence age and population expansions estimates

The coalescence time of Japanese and Tibetan expansion was estimated using r statistic-based method and Bayesian MCMC method. For r statistic-based method, standard deviation was calculated following Saillard et al. (2000). Then TMRCA of each expansion was estimated using Soares rate for mtDNA coding regions (576-16023) (Soares et al. 2009), and a corrected rate of Mishmar's rate for coding regions respectively. For Bayesian MCMC analysis, the time of each distinct expansion was estimated using BEAST v1.8.0 (Drummond and Rambaut 2007). Each MCMC sample of each cluster with distinct expansion was based on a run of 10 million generations sampled every 1,000 steps with the first one million generations regarded as burn-in. We used the HKY+G model of nucleotide substitution without partitioning the coding region. A strict clock was used and prior substitution rate was assumed to be normally distributed, with a mean of 1.90×10^{-8} subs/site/year and an SD of 1.92×10^{-9} subs/site/year adopted from former published values (Zheng et al. 2012). Specifically, the BSPwas constructed according to a relatively high rate 1.90×10^{-8} subs/site/year, making our results more reliable because higher rate would result in lower time estimates. Each run was subsequently analyzed using Tracer v1.6.

Result

M9a phylogenetic tree based on mtDNA genome information

We constructed the phylogeny of haplogroup M9a (mainly M9a1) with152 sequenced mtDNA genomes from East Asia (especially Tibet and parts of East Europe(Figure 1). Discerned from the overall structure of the tree (Appendix B), most basal branches of M9a were distributed in Southeast Asia (4/5) with the exception of M9a1, this pattern suggested that M9a might originate from the south (just as Kang et al. 2013). The geographic distribution of M9a1 was rather complicated. Although this haplogroup did bear some genetic imprints of southern origin by harboring a basal lineage (i.e. 12817 and 12818) from southern China, this haplogroup had actually diffused to northern China (M9a1a1a and M9a1a1c1a), Japan (M9a1a1a),southwestern China, Tibet and northeast part of the South Asian Subcontinent (M9a1b1b and M9a1a1c1b), among which southern Himalaya region bears the most M9a1a1c1b haplogroup. Based on this pattern, it seems that haplogroup M9a1 had most likely been involved in some northward and westward dispersal(s) in East Asia, which is consistent with the finding of Peng et al. 2011.

Phylogeographic distribution

The phylogenetic tree of haplogroup M9a provided a basis for us to perform a well-defined phylogeographic analysis of this haplogroup. To better characterize the demographic history of M9a, we constructed the median-joining network (Bandelt et al. 1999) based on all 152 M9a mtDNAs (Figure 2). Similar to the observation from the phylogenetic tree (Appendix B), the median-network (based on the information of coding region) showed that within haplogroup M9a1b, the basal lineages were mainly restricted to Southeast Asia, central India, and northern China including Tibet, whereasM9a1b1 spread not only in the above regions, but also in northeast India and Bangladesh (Figure 2).

The basal lineages belonging to M9a1a*were mainly found in southern China (Figure 2). One of its derivatives, haplogroup M9a1a2, displayed a restricted distribution in southwestern China, Tibet, and northeast India (Figure 2; Appendix C), and presented a similar pattern to that of haplogroup M9a1a1c1b.

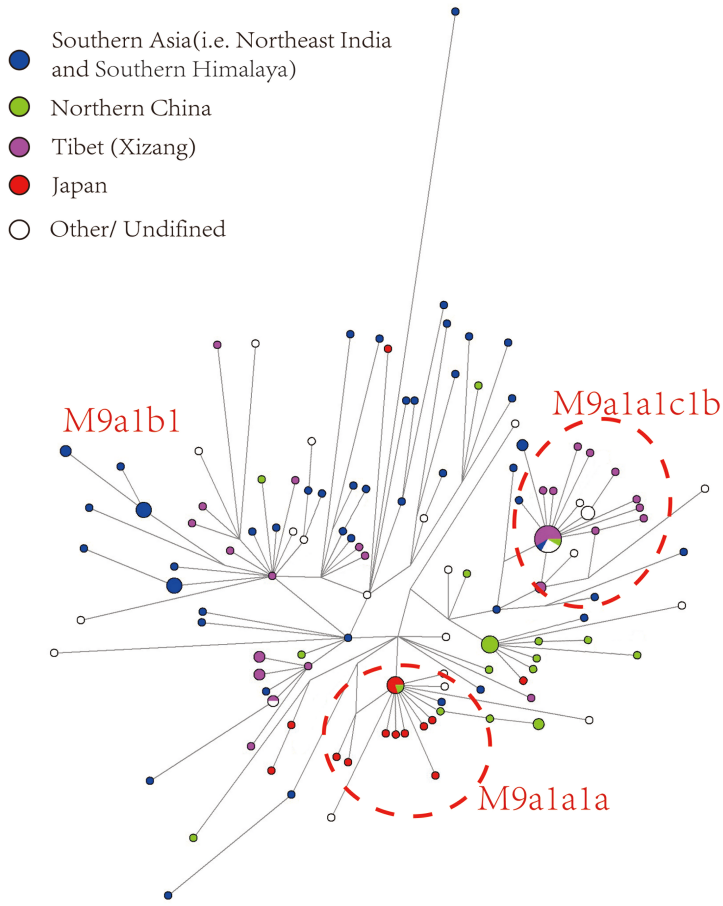

Figure 2 **Median-joining network of coding region haplotypes observed in 152 M9a mtDNAs.** (a) The median-joining network based on mtDNA coding region sites. The geographic origin of samples is shown by different colors corresponding to their respective locations by name, and the star-like expansion populations of Japan and Tibet are circled by red dotted line.

These results are also supported by recent research (such as that of Kang et al. 2014). Nevertheless, haplogroup M9a1a1a showed a distinct distribution pattern: most of M9a1a1basal lineages were distributed in northern China, southwestern China, southern Himalaya, northeast India, central India, and Japan, whereas its major sub-haplogroup M9a1a1c was prevalent in northern China, southwestern China, and Japan (Figure 2; Appendix C). Remarkably, the M9a1a1 lineages found in Tibet were mostly clustered into haplogroup M9a1a1c1b (54.55%), which might imply the possible operation of selection.

Neutrality tests on whole samples and those in major regions

In our research, the network represents a typical star-like pattern of lineage expansion. The two major nodes, which suggest the origin of expansion, are made up of Japanese and Tibetan samples respectively. Thus, we used Tajima's D, Fu and Li's D and F analyses (Tajima 1989; Fu et al. 1993) according to the model of neutral selection to test whether selections happened in the above two populations. All values turned out to be negative, which suggests that the values are more deviated from neutral model and have significant differences statistically. Then, we conducted the same tests on single gene and compared the result within Japan and Tibet. For eleven genes, except *ND5* and *CYB*, the results for Japanese and Tibetan samples are almost all negative and the same in absolute value, which is compatible with purifying selection theory. However, we note that for *ND5* and *CYB*, the values generated by all methods tend to be less negative (closer to zero) within Tibetans than those of Japanese (Appendix D).In addition, by calculating the difference value between Japan and Tibet results, *ATP6*,*ND5*, and *CYB* are the only three genes whose D-value is much higher than the average by all methods (Figure 3), that is, people in Tibet might experience the adaptive selection.

Meanwhile, we found that *ND5* C12362T and *CYB* A15671G mutations are most frequent in Tibet (80% and 76% respectively); while no specific mutations were found in Japanese. It is also interesting that the above variants are the decisive mutation for haplogroup M9a1b. Thus, we did further Ka/Ks analysis on *ND5* and *CYB* genes (Sun et al. 2007; Mishmar et al. 2003, Ruiz-Pesini et al. 2004; Kivisild et al. 2006; Elson et al. 2004; Yang et al. 2000) in the samples that belong to haplogroupM9a1b to double check whether these variations are adaptive. The common value for Ka/Ks of *ND5* is 0.5, while that of sample 12924 (Tibetan) is 0.682. Also, Ka/Ks result of *CYB* gene turns out to be 0.113, whereas, that value of 12932 (Tibetan) is 0.163. Thus, we can propose that C12362T and A15671G bear the evidence of adaptive evolution in Tibetan highlands, which is possibly driven by the local environment (high latitude, hypoxia, extreme cold, etc.) (Meiklejohn et al. 2007; Rozas et al. 2009; Librado et al. 2009).

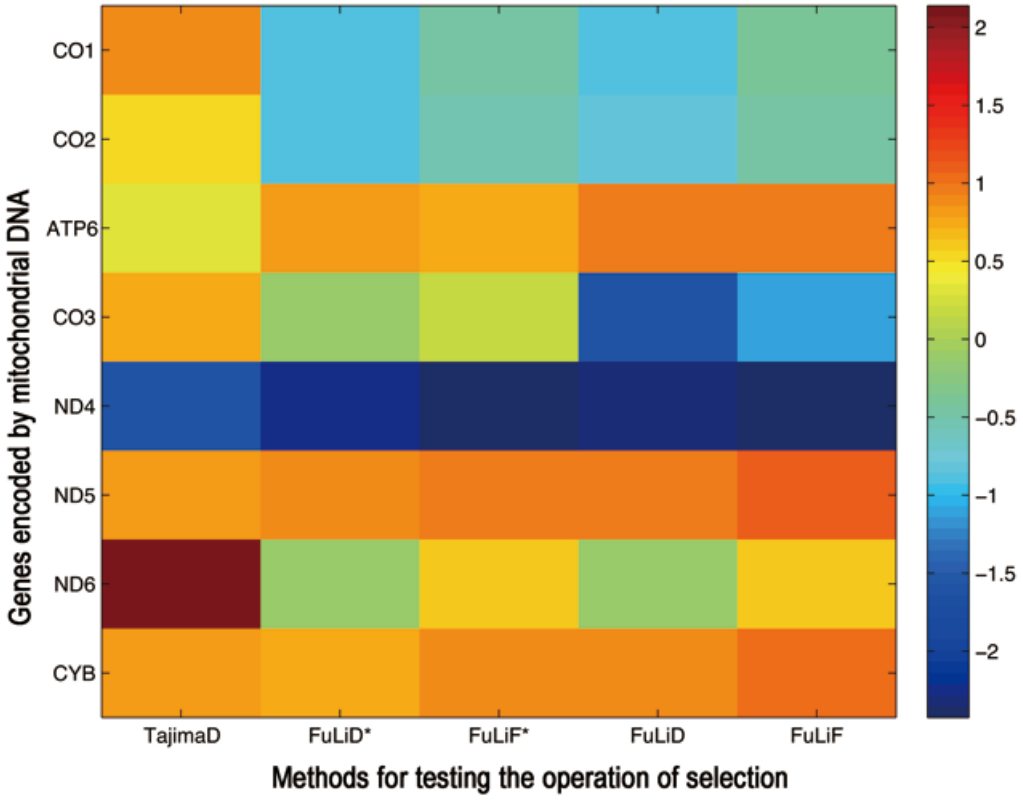

Figure 3 **Heat map of difference values between Tibet and Japan population.** The y-axis is the list of 8 genes encoded by mitochondrial DNA. The x-axis is the five methods for testing the operation of selection. The difference value between Tibet and Japan population (D=#Tibet-#Japan) is indicated by different colors, and the corresponding values are listed besides. Refer to Methods for detailed settings.

Protein analyses on variances detected in Tibet

To further illustrate the effects of gene mutations on their encoding proteins, we predicted and then analyzed the structural changes upon *ND5* C12362T and *CYB* A15671Gmutations. We found that some structural parameters have changed since the mutations took place.

ND5 C12362T is notable in Tibetan populations with its unique existence and high frequency. The ninth threonine (Thr) turns into isoleucine (Ile) because of the mutation. Although the two amino acids are both neutral, threonine is hydrophilic amino acid because of its hydroxyl group, while isoleucine is hydrophobic and aliphatic. Substitutions of turns and folds as well as alpha helix were observed around the mutation site according to secondary structures (Pollastri et al. 2002) (Appendix E). The prediction for high structures shows that there is remarkable difference between proteins with and without the mutation according to the former researches on protein structure prediction and structural genomic. (Cheng et al. 2005; Holm et al. 1991; Sayle et al. 1995). The groups become much tighter and the interaction is stronger owing to mutations (Figure 4a and 4b, Table 1).

Table 1 Structural Parameter of ND5, CYB Proteins between ancestral and derived alleles

	ND5	*ND5*-T9I	*CYB*	*CYB*-M309V
H-Bonds	297	286	292	291
Helices	20	19	18	18
Strands	0	2	0	0
Turns	27	25	31	29

We then used I-Mutant2.0 (Capriotti et al. 2005) to predict protein stability changes upon single point mutations. It is a support vector machine (SVM)-based tool for the automatic prediction starting from the mtDNA sequence whose result is clear to interpret, for example, the negative related G values imply the decrease of stability (see Material and Method). It is important that ΔΔG caused by *ND5* C12362Tis 0.69, which reflects the decrease in protein free energy and the increase in its stability. This positive value implies that the mutation should be adaptive and related to the highland environment.

Another mutation found in Tibetan is *CYB* A15671G, which makes the 309th methionine (Met) change into valine (Val). Being hydrophobic and aliphatic, the two amino acids are similar in properties and are both neutral. What's more, the mutation almost causes no difference in secondary structure with regard to alpha helix or beta-sheet (Appendix E). However, the mutated protein tends to be loose in its domains, compared with the higher structures (Figure 4c and 4d, Table 1). The stability of the protein is slightly decreased with ΔΔG=-0.30. Nevertheless, *ND5* C12362T and *CYB* A15671G mutations take place simultaneously in most cases, which

might be caused by the package of protein subunits. It is also interesting that this mutation is often correlated with *ND4* G11778A mutation, which causes low penetrance of Leber's Hereditary Optic Neuropathy (LHON) (Bandelt HJ et al. 2005, Qu J et al. 2009). Thus, A15671G is also likely to be a possible signal of adaptive evolution despite its effect of decreasing protein stability. Its mechanism needs further research.

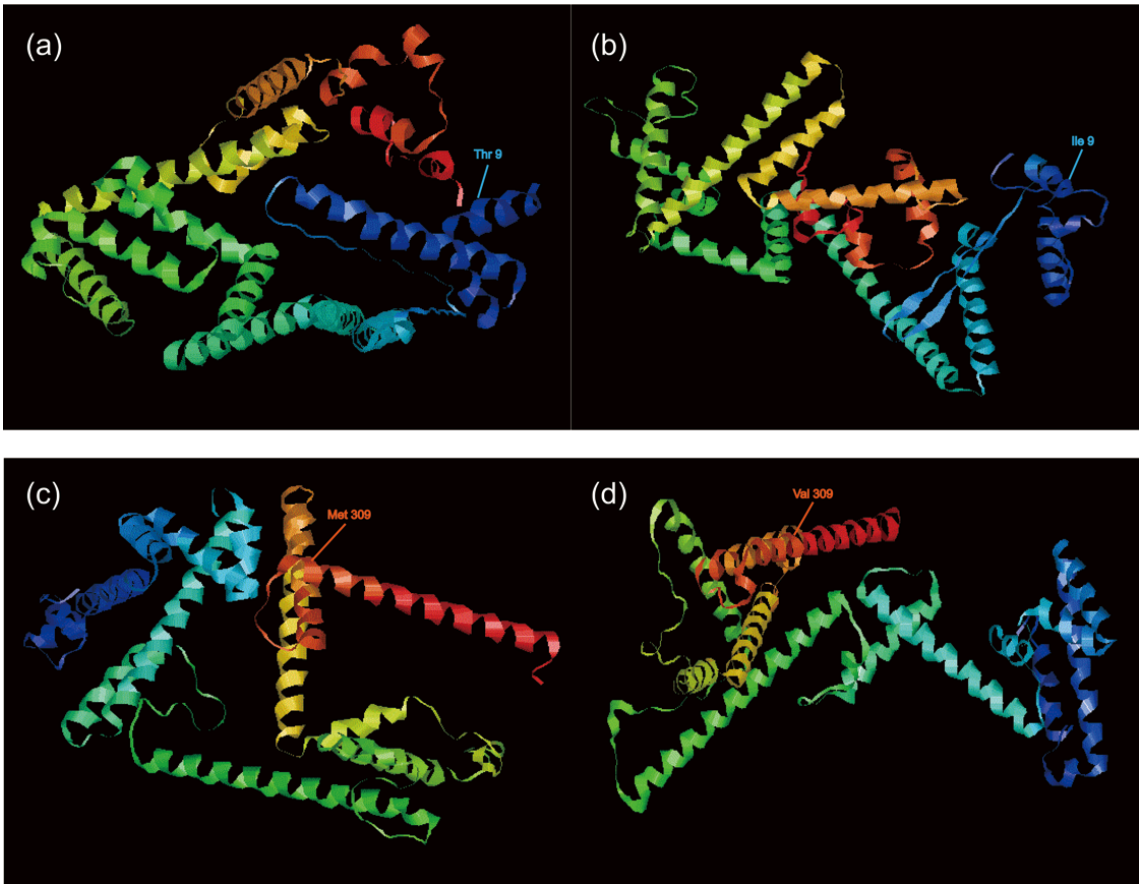

Figure 4 **Ribbon diagrams of original and mutated protein structures.** Amino acids grouped to different domains are shown in distinct colors. (a) and (b) Structures of *ND5* and *ND5*-T9I, respectively. The 9th residues from N-end (Thr in *ND5* and Ile in *ND5*-T9I) are indicated. (c) and (d) Structures of *CYB* and *CYB*-M309V, respectively. The 309th residues from N-end (Met in *CYB* and Val in *CYB*-T9I) are indicated.

Coalescence age and population expansions estimates

The large number of M9a'b samples with complete mtDNA genome information, as well as the network, allowed us to estimate the coalescence ages and expansions of haplogroup M9a1b, which is decided by variations C12362T and A15671G. We estimated the coalescence age of haplogroup M9a1b to beapproximately6.8to 17.5kya with geometric mean being 11.7kya (Figure 5a).

We then constructed BSP for this haplogroup to describe the historical maternal effective variation trends. BSPs estimate effective population size through time from random sequences of a population. However, haplogroups in general do not equate to population data, but the signals associated with a haplogroup might nevertheless reveal demographic processes in the populations carrying it, as previously suggested (Achilliet al. 2013; Soares et al. 2012).The main purpose of this analysis was to provide BSPs based on population of haplogroup M9a1b. It is clear that this haplogroup experienced a distinct trend of growth about 5kya and extended to 2.5 kya (Figure 5b). However, these ages must be received with caution (Cox. 2008; Endicott et al. 2009).

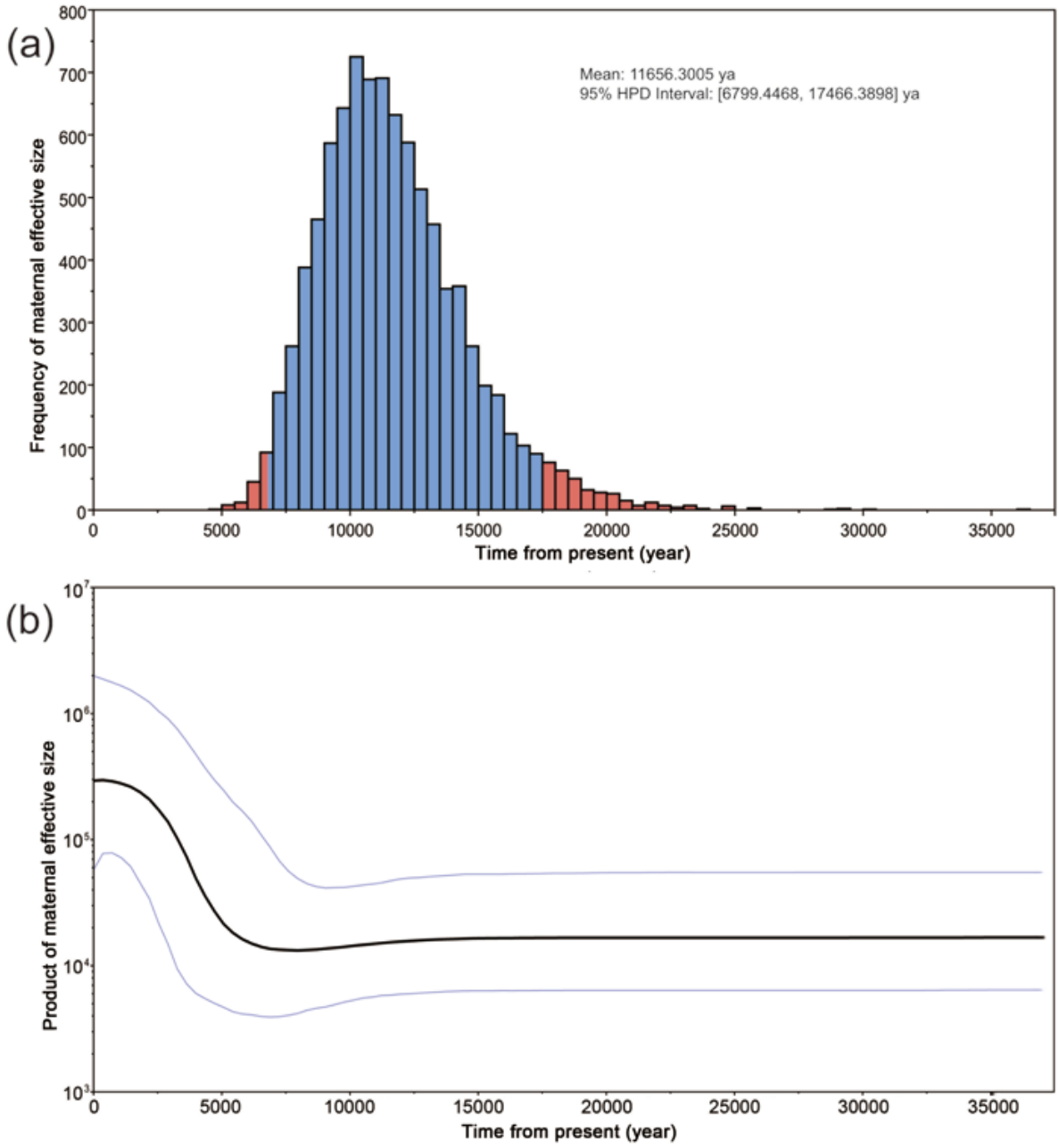

Figure5 Coalescence age estimates and mtDNA Bayesian skyline plot showing the size trend of Haplogroup M9a1b. (a) TMRCA for Haplogroup M9a1b. The blue bars show the 95% confidence interval. Detailed settings refer to Methods. (b) BSP for Haplogroup M9a1b. The thick solid line is the median estimate and the thin lines (blue) show the 95% highest posterior density limits. Detailed settings refer to Methods. Refer to Methods for detailed settings.

The time of the expansion lineage estimated was then compared to previous coalescence age estimate (Figure 5a and 5b). As a result, haplogroup M9a1b originated more than 10 kya, while it experienced a significant increase in population size only about 5 kya, which happens to be the time when Tibetans began to live in plateau. It then implied that the expansion might be driven by the mutations observed and there might be adaptive selection.

Discussion

Adjustment of Results on Neutrality Test

We have applied a number of tests of selection, including Tajima's D test, Fu and Li's test (Tajima 1989; Fu et al. 1993) as well as the Ka/Ks analysis on single-gene (Nei et al. 1986;Elson et al. 2004), to the coding regions of 51 human mtDNAs that encompass two major populations (Tibetan and Japanese). Evidence for selection was only found in some of the tests. Furthermore, it is often difficult to "isolate" the effects of selection from other processes, such as recent population expansions or bottleneck effects. Nevertheless, the preponderance of evidence (Peng et al. 2011; Kang et al. 2013) indicates that positive selection operated on the coding region of the Tibetan mitochondrial genome. Moreover, selection appears to have influenced the pattern of sequence divergence for most, and perhaps all, of the 13 protein-encoding genes, especially on gene *ND5* and *CYB* (Rand et al. 1996; Hasegawa et al. 1998). These conclusions are supported by other recent analyses, such as those of Moilanen and Majamaa (2003) and Moilanen et al. (2003).

Thus, further functional experiment should be fundamental to validate the adaptive biochemical changes induced by T12362C and A15671G, in case that these two variants were neutral mutations.

Potential Adaptation Mechanism for *ND5* and *CYB* Variants

Interpreted by Bioenergetics, both *ND5* C12362T (T9I) and *CYB* A15671G (M309V) found in the study, tend to be more stable in structure and property after mutations, indicating the operation of adaptive selection. And

potential adaptation mechanisms for *ND5* and *CYB* variants are supported by other recent analyses. (Drose et al. 2012; Murphy et al.2009; Brand et al. 2010; Chandel et al. 2000; Guzy et al.2005; Guzy et al. 2007; Klimova et al. 2008). However, the effect of variation on activity of holoenzyme assembling needs to be further studied. Moreover, according to phylogenetic analysis, *ND5* C12362T (T9I) and *CYB* A15671G (M309V) both define the haplogroup M9a1b, which might also be an evidence of adaptive selection.

When associated with energy production, the result is also in accord with the above conclusions. OXPHOS is a system that generates ATP to supply the demand of body and entitles mitochondria to be an energy-producing organelle (Baradaran et al. 2013). If the consisting units of OXPHOS (for example, *ND5* and *CYB*) tend to be stable in structures, the consumption of ATP would be decreased and the stability and functions of other units of Complex could be maintained with less energy. Especially under extreme environment (high altitude, cold, and hypoxia), the synthesis of ATP can reduces and its utility can be more efficient. In addition, the stability of OXPHOS units might promote heat generation for the resistance of extreme cold.

In our study, the stability of *ND5* C12362T (T9I) and *CYB* A15671G (M309V) increase after mutation, this might influence the energy metabolism, indicating the adaptive selection in Tibetan.

Evidence of Selection from Population Genetics

We also analyzed the association between two mutations (C12362T, A15671G) from population genetics. From our result, haplogroup M9a1b, which is defined by C12362T and A15671G, experienced a significant increase in population size only about 5 kya. It was long after its origination: over 10 kya (Figure 5a and 5b).

According to many archaeological evidences, Tibetans were believed to first live in plateau approximately 5 kya (Chen et al. 2015), which happens to be the time when population size of haplogroup M9a1b began to increase. In addition, Tibetans seemed to have experienced selective constraints of low oxygen concentration in extreme high altitude.

By comprehensively analyzing the two mutations from both gene level and population genetic level, we seem to find out the way how Tibetans successfully adapt to plateau environment. It is encouraging that such result can be applied to other biologic and medical fields, such as overcoming altitude sickness from gene level.

Conclusion

Our comprehensive phylogeographic analyses of mtDNA haplogroup M9a and the protein structures for

variances revealed that *ND5* C12362T and *CYB*A15671Gmutations had certain effect on Complex I and Complex III function and thus are selected for in Tibetan highlands. This adaptive selection is mainly driven by the specific geometric environment (attitude, climate changes, etc.), and consequently, leads to a significant increase in population size.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the National Excellent Youth Science Foundation of China (31222030), Shanghai Rising-Star Program (12QA1400300), Shanghai Shuguang Project(14SG05), MOE University Doctoral Research Supervisor's Funds (20120071110021), MOE Scientific Research Project (113022A),and FDUROP Wangdao Scholar Project (14068).

REFERENCE

Bandelt HJ, Forster P &Rohl A 1999. Median-joining networks for inferring intraspecific phylogenies. MolBiolEvol, 16, 37-48

Bandelt HJ, Achilli A, Kong QP, Salas A, Lutz-Bonengel S, Sun C, Zhang YP, Torroni A, Yao YG (2005) Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies. BiochemBiophys Res Commun, 333, 122-30

Baradaran R, Berrisford JM, Minhas GS,Sazanov LA (2013) Crystal structure of the entire respiratory complex I. Nature, 494, 443-8

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, Mccormack M, Montgomery HE, Pan H, Robbins PA, Shianna KV, Tam SC, Tsering N, Veeramah KR, Wang W, Wangdui P, Weale ME, Xu Y, Xu Z, Yang L, Zaman MJ, Zeng C, Zhang L, Zhang X, Zhaxi P, Zheng YT (2010) Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl AcadSci U S A, 107, 11459-64

Behar DM, Van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, Kivisild T, Torroni A,Villems R (2012) A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet,

90, 675-84

Brand MD (2010)The sites and topology of mitochondrial superoxide production. ExpGerontol, 45, 466-72

Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res, 33, W306-10

Chandel NS, Mcclintock DS, Feliciano CE, Wood TM, Melendez JA, Rodriguez AM, Schumacker PT (2000) Reactive oxygen species generated at mitochondrial complex III stabilize hypoxia-inducible factor-1alpha during hypoxia: a mechanism of O2 sensing. J BiolChem, 275, 25130-8

Chen F H, Dong GH, Zhang DJ, Liu XY, Jia X, An CB, Ma MM, Xie YW, Barton L, Ren XY, Zhao ZJ, Wu XH, Jones MK(2015) Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P.Science, 347, 248-250

Cheng J, Randall AZ, Sweredoski MJ,Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res, 33, W72-6

Cox MP (2008) Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. Hum Biol, 80, 335-57

Droma Y, Hanaoka M, Basnyat B, Arjyal A, Neupane P, Pandit A, Sharma D, Miwa N, Ito M, Katsuyama Y, Ota M, Kubo K 2006. Genetic contribution of the endothelial nitric oxide synthase gene to high altitude adaptation in Sherpas. High Alt Med Biol, 7, 209-20

Droma Y, Hanaoka M, Basnyat B, Arjyal A, Neupane P, Pandit A, Sharma D, Ito M, Miwa N, Katsuyama Y, Ota M, Kubo K (2008) Adaptation to high altitude in Sherpas: association with the insertion/deletion polymorphism in the Angiotensin-converting enzyme gene. Wilderness Environ Med, 19, 22-9

Drose S, Brandt U (2012) Molecular mechanisms of superoxide production by the mitochondrial respiratory chain. AdvExp Med Biol, 748, 145-69

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC EvolBiol, 7, 214

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7.MolBiolEvol, 29, 1969-73

Elson JL, Turnbull DM, Howell N (2004) Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet, 74, 229-38

Endicott P, Ho SY, Metspalu M, Stringer C (2009) Evaluating the mitochondrial timescale of human evolution. Trends EcolEvol, 24, 515-21

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics, 133, 693-709

Guzy RD, Hoyos B, Robin E, Chen H, Liu L, Mansfield KD, Simon MC, Hammerling U,Schumacker PT (2005) Mitochondrial complex III is required for hypoxia-induced ROS production and cellular oxygen sensing. Cell Metab, 1, 401-8

Guzy RD, Mack MM,Schumacker PT (2007) Mitochondrial complex III is required for hypoxia-induced ROS production and gene transcription in yeast. Antioxid Redox Signal, 9, 1317-28

Hanaoka M, Droma Y, Basnyat B, Ito M, Kobayashi N, Katsuyama Y, Kubo K, Ota M (2012) Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. PLoS One, 7, e50566

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

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS,Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet, 70, 1152-71

Holm L, Sander C (1991) Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors. J MolBiol, 218, 183-94

Ji F, Sharpley MS, Derbeneva O, Alves LS, Qian P, Wang Y, Chalkia D, Lvova M, Xu J, Yao W, Simon M, Platt J, Xu S, Angelin A, Davila A, Huang T, Wang PH, Chuang LM, Moore LG, Qian G, Wallace DC (2012) Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and high-altitude Tibetans. Proc Natl AcadSci U S A, 109, 7391-6

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, De Knijff P, Feldman M, Cavalli-Sforza LL,Oefner PJ (2006) The role of selection in the evolution of human mitochondrial genomes. Genetics, 172, 373-87

Kang L, Zheng HX, Chen F, Yan S, Liu K, Qin Z, Liu L, Zhao Z, Li L, Wang X, He Y, Jin L (2013) mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. MolBiolEvol, 30, 2579-87

Klimova T, Chandel NS (2008) Mitochondrial complex III regulates hypoxic activation of HIF. Cell Death Differ, 15, 660-6

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25, 1451-2

Lopes A, Schmidt Am Busch M, Simonson T (2010) Computational design of protein-ligand binding: modifying the specificity of asparaginyl-tRNAsynthetase. J ComputChem, 31, 1273-86

Meiklejohn CD, Montooth KL, Rand DM (2007) Positive and negative selection on the mitochondrial genome. Trends Genet, 23, 259-63

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. ProcNatl AcadSci U S A, 100, 171-6

Moilanen JS, Finnila S,Majamaa K (2003) Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J. MolBiolEvol, 20, 2132–2142

Moilanen JS, Majamaa K (2003) Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA. MolBiolEvol, 20, 1195–1210

Murphy MP (2009) How mitochondria produce reactive oxygen species.Biochem J, 417, 1-13

Nei M,Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. MolBiolEvol, 3, 418-26

Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu, Basang, Ciwangsangbu, Danzengduojie, Chen H, Shi H, Su B (2011) Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. MolBiolEvol, 28, 1075-81

Pollastri G, Przybylski D, Rost B,Baldi P (2002)Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins, 47, 228-35

Qu J, Zhou X, Zhang J, Zhao F, Sun YH, Tong Y, Wei QP, Cai W, Yang L, West CE, Guan MX (2009) Extremely low penetrance of Leber's hereditary optic neuropathy in 8 Han Chinese families carrying the ND4 G11778A mutation. Ophthalmology, 116, 558-564 e3

Rand DM,Kann LM (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. MolBiolEvol, 13, 735-48

Richards M, Macaulay V (2001)The mitochondrial gene tree comes of age. Am J Hum Genet, 68, 1315–1320.

Rozas J 2009. DNA sequence polymorphism analysis using DnaSP. Methods MolBiol, 537, 337-50

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, 223-6

Saillard J, Forster P, Lynnerup N, Bandelt HJ,Norby S (2000)mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet, 67, 718-26

Sayle RA, Milner-White EJ (1995) RASMOL: biomolecular graphics for all. Trends BiochemSci, 20, 374.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet, 84, 740-59

Sun C, Kong QP, Zhang YP (2007) The role of climate in human mitochondrial DNA evolution: a reappraisal. Genomics, 89, 338-42

Suzuki K, Kizaki T, Hitomi Y, Nukita M, Kimoto K, Miyazawa N, Kobayashi K, Ohnuki Y,Ohno H (2003) Genetic variation in hypoxia-inducible factor 1alpha and its possible association with high altitude adaptation in Sherpas. Med Hypotheses, 61, 385-9

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123, 585-95

Van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat, 30, E386-94

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

Wallace DC, Lott MT,Procaccio V (2007) Mitochondrial genes in degenerative diseases, cancer and aging (Chapter 13). In Rimoin DL, Connor JM, Pyeritz RE &Korf BR. Emery and Rimoin's Principles and Practice of Medical Genetics, 5th Edition. Churchill Livingstone, London, pp. 194–298

Wallace DC, Fan W,Procaccio V (2010)Mitochondrial energetics and therapeutics. Annu Rev Pathol, 5, 297-348

Xu S, Li S, Yang Y, Tan J, Lou H, Jin W, Yang L, Pan X, Wang J, Shen Y, Wu B, Wang H, Jin L (2011) A genome-wide search for signals of high-altitude adaptation in Tibetans. MolBiolEvol, 28, 1003-11

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 155, 431-49

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan, Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, Zhou G, Tang M, Qin J, Wang T, Feng S, Li G, Huasang, Luosang J, Wang W, Chen F, Wang Y, Zheng X, Li Z, Bianba Z, Yang G, Wang X, Tang S, Gao G, Chen Y, Luo Z, Gusang L, Cao Z, Zhang Q, Ouyang W, Ren X, Liang H, Zheng H, Huang Y, Li J, Bolund L, Kristiansen K, Li Y, Zhang Y, Zhang X, Li R, Li S, Yang H, Nielsen R, Wang J, Wang J (2010)

Sequencing of 50 human exomes reveals adaptation to high altitude. Science, 329, 75-8

Zheng HX, Yan S, Qin ZD, Wang Y, Tan JZ, Li H, Jin L (2011) Major population expansion of East Asians began before neolithic time: evidence of mtDNA genomes. PLoS One, 6, e25835

Zheng HX, Yan S, Qin ZD, Jin L (2012) MtDNA analysis of global populations support that major population expansions began before Neolithic Time. Sci Rep, 2, 745