Journal of Human Genetics (2013) 58, 780–787

www.nature.com/jhg

ORIGINAL ARTICLE

Comprehensive analysis of common and rare mitochondrial DNA variants in elite Japanese athletes: a case–control study

Eri Mikami[1,2,3], Noriyuki Fuku[3], Qing-Peng Kong[4], Hideyuki Takahashi[5], Nao Ohiwa[5], Haruka Murakami[6], Motohiko Miyachi[6], Mitsuru Higuchi[7], Masashi Tanaka[3], Yannis P Pitsiladis[8] and Takashi Kawahara[9]

The purpose of the present study was to identify mitochondrial DNA (mtDNA) polymorphisms and rare variants that associate with elite Japanese athletic status. Subjects comprised 185 elite Japanese athletes who had represented Japan at international competitions (that is, 100 endurance/middle-power athletes: EMA; 85 sprint/power athletes: SPA) and 672 Japanese controls (CON). The entire mtDNA sequences (16 569 bp) were analyzed by direct sequencing. Nucleotide variants were detected at 1488 sites in the 857 entire mtDNA sequences. A total of 311 variants were polymorphisms (minor allele frequency $\geqslant$ 1% in CON), and the frequencies of these polymorphisms were compared among the three groups. The EMA displayed excess of seven polymorphisms, including subhaplogroup D4e2- and D4g-specific polymorphisms, compared with CON ($P<0.05$), whereas SPA displayed excess of three polymorphisms and dearth of nine polymorphisms, including haplogroup G- and subhaplogroup G2a-specific polymorphisms, compared with CON ($P<0.05$). The frequencies of 10 polymorphisms, including haplogroup G- and subhaplogroup G2a-specific polymorphisms, were different between EMA and SPA ($P<0.05$): although none of these polymorphisms differed significantly between groups after correcting for multiple comparison (false discovery rate q-value $\geqslant$ 0.05). The number of rare variants in the 12S ribosomal RNA and NADH dehydrogenase subunit I genes were also higher in SPA than in CON ($P<0.05$). Analysis of the entire mtDNA of elite Japanese athletes revealed several haplogroup- and subhaplogroup-specific polymorphisms to be potentially associated with elite Japanese athletic status.
Journal of Human Genetics (2013) **58,** 780–787; doi:10.1038/jhg.2013.102; published online 10 October 2013

Keywords: elite athlete status; haplogroup; mitochondrial genome; polymorphism; physical performance

INTRODUCTION

Environmental factors such as physical training, nutrition, injury prevention and recovery have an important role in sporting success. Sporting success is also dependent on genetic factors with some 66% of the variance in athlete status being attributed to genetic factors.[1] To date, numerous studies have attempted to identify genetic polymorphisms that relate to physical training and performance. Over 200 genes in both nuclear DNA and mitochondrial DNA (mtDNA) have been reported to associate with physical performance and health-related fitness,[2,3] and the number of genes associated with physical performance-related phenotypes are expected to increase dramatically with the application of genome-wide methods to elite athlete cohorts; albeit such studies are only in their infancy.[4]

Almost all eukaryotic cells contain mitochondria with the primary purpose to produce ATP through respiration and to regulate energy metabolism. The human mitochondrial genome is 16 569-bp in length and comprises a total of 37 genes essential for mitochondrial function (13 proteins of mitochondrial oxidative phosphorylation (OXPHOS), 2 rRNAs and 22 tRNAs).[5] MtDNA is inherited almost exclusively through the maternal lineage. As several familial studies have found aerobic capacity to be more strongly influenced by maternal rather than paternal inheritance,[6–8] maternally inherited mtDNA is likely to contain candidate genes influencing aerobic capacity and therefore potentially elite athlete status. In support of this idea, we recently reported that mitochondrial haplogroups, which are a set of tightly linked mtDNA polymorphisms, were associated with elite endurance athlete status in elite Kenyan[9] and Japanese[10] athletes. We also found intriguing associations between mitochondrial haplogroups with sprint/power athlete (SPA) status in elite Japanese[10] and African–American[11] athletes and with muscle power in Japanese

[1]Graduate School of Sport Sciences, Waseda University, Saitama, Japan; [2]Japan Society for the Promotion of Science, Tokyo, Japan; [3]Department of Genomics for Longevity and Health, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan; [4]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China; [5]Department of Sports Sciences, Japan Institute of Sports Sciences, Tokyo, Japan; [6]Department of Health Promotion and Exercise, National Institute of Health and Nutrition, Tokyo, Japan; [7]Faculty of Sport Sciences, Waseda University, Saitama, Japan; [8]College of Medicine, Veterinary and Life Sciences, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK and [9]Department of Sports Medicine, Japan Institute of Sports Sciences, Tokyo, Japan
Correspondence: Dr N Fuku, Department of Genomics for Longevity and Health, Tokyo Metropolitan Institute of Gerontology, 35-2 Sakae-cho, Itabashi-ku, Tokyo 173-0015, Japan.
E-mail: nfuku@tmig.or.jp
Received 10 June 2013; revised 2 September 2013; accepted 5 September 2013; published online 10 October 2013

non-athlete individuals.[12] Associations of mitochondrial haplogroups with elite Caucasian and Asian athlete status have also been reported by others.[13–16]

The matrilineal inheritance of mtDNA and linear accumulation of polymorphisms have allowed the construction of detailed mtDNA phylogenies.[17] These phylogenies allow haplogroup identification through the analysis of a small number of haplogroup-specific polymorphisms in the hypervariable region-I and in certain gene-coding regions. For practical/economic reasons, this approach has been adopted in all previous studies that examined associations between mitochondrial haplogroups and elite athlete status. From this analysis, it is possible to infer the existence of haplogroup-linked polymorphisms in gene-coding regions; however, this inference is imperfect because of incomplete linkage between the polymorphisms in the non-coding regions and those in the gene-coding regions.[18] Consequently, the functional polymorphisms, which are responsible for the previously reported associations between haplogroups and elite athlete status have not been identified to date.

Each mitochondrial haplogroup comprises several subhaplogroups, which are younger branches than haplogroups in the mtDNA phylogenetic tree. Influential variants have been eliminated from the older branches of the tree by 'purifying selection' in evolutionary terms. Consequently, young branches (subhaplogroups) in the mtDNA phylogenetic tree contain a higher proportion of non-synonymous substitutions in the protein-coding genes and substitutions in the rRNA and tRNA genes than old branches (haplogroups).[19,20] Therefore, subhaplogroup-specific substitutions are more likely to be associated with various health- and performance-related phenotypes. Indeed, we previously reported that subhaplogroup D4a belonging to the haplogroup D was selected for longevity from the analysis of complete mtDNA sequences in Japanese centenarians and semi-supercentenarians,[21,22] although we originally reported that the m.5178C>A polymorphism, which characterizes mitochondrial haplogroup D, was associated with longevity.[23] Therefore, the polymorphisms defining each haplogroup but also those characterizing the subhaplogroups could have functional influences on elite athletic performance.

On the other hand, homoplasic (haplogroup-nonspecific) polymorphisms, which are inferred from the presence of a substitution in >1 lineages in the mtDNA phylogenetic tree, may also represent good candidates of elite athletic performance. Although we previously found that several haplogroup-nonspecific polymorphisms in the control region were associated with elite Japanese athletic status,[24] it remains to be determined whether haplogroup-nonspecific polymorphisms in the gene-coding region are associated with elite athletic performance. Previous studies that attempted to identify performance-associated variants have focused on common variants, that is, polymorphisms. However, rare variants may account for the 'missing heritability' of common diseases.[25] Therefore, there is an urgent need to investigate both common but also rare variants that may associate with elite athletic performance.

The aim of the present study was to revisit the previously published cohort[24] and identify the precise mtDNA polymorphisms and possibly rare variants that associate with elite Japanese athletic status.

MATERIALS AND METHODS

Subjects

The subjects in the present study comprised 185 elite Japanese athletes (146 males and 39 females) from all over Japan. All athletes were international athletes (participants in Olympic Games and/or World and Asian Championships held in 1964–2009) and the group included several medalists at these international competitions. A total of 185 athletes were classified as endurance/middle-power athletes (EMA) ($n=100$) and SPA ($n=85$) based on the criteria of Yang *et al*.[26] The EMA group included 21 endurance runners competing in events of $\geqslant$800 m, 10 sailing athletes, 9 swimmers competing in events of $\geqslant$200 m, 10 rowers, 5 long-distance cyclists, 7 canoeists, 13 volleyball players, 6 basketball players, 6 hockey players, 5 soccer players, 5 water polo players, 2 boxers and 1 modern pentathlete. The SPA comprised 38 track and field athletes (23 sprinters competing in events of $\leqslant$400 m, 9 jumpers, 5 throwers and 1 decathlete), 9 swimmers competing in events of $\leqslant$100 m, 9 gymnasts, 8 competitive fencers, 7 divers, 5 wrestlers, 5 weightlifters, 2 short-distance track cyclists and 2 judo athletes. The control group (CON) consisted of 672 Japanese individuals (387 males and 285 females), whose entire mtDNA sequences[18] were registered in our Human Mitochondrial Genome Single Nucleotide Polymorphism Database (http://mtsnp.tmig.or.jp/mtsnp/index_e.shtml). In order to ascertain whether the control group used in the present study represented an appropriate non-athletic control, we compared mitochondrial haplogroup distribution between this control group and another confirmed non-athletic healthy Japanese control ($n=480$) with normal physical activity levels, and confirm that the frequencies of mitochondrial haplogroups did not differ between the two groups ($P=0.352$) (Supplementary Table S1).

Written informed consent was obtained from each subject, and the study was approved by the Ethics Committees of the Japan Institute of Sports Sciences and Tokyo Metropolitan Institute of Gerontology.

Data collection and analysis

Total DNA was isolated from the venous blood or saliva using QIAamp DNA Blood Maxi Kit (QIAGEN, Hilden, Germany) or Oragene•DNA (DNA Genotek, Ottawa, Ontario, Canada), respectively. Complete mtDNA sequencing was carried out as previously described[27] and was referred by our website (http://mtsnp.tmig.or.jp/mtsnp/annex_e.html). Briefly, the entire mitochondrial genome was amplified as six fragments ($\sim$3 kb) by the first polymerase chain reaction (PCR) and 60 overlapping segments ($\sim$1 kb) by the second PCR. The primer pairs and PCR conditions are shown in Supplementary Tables S2 and S3, respectively. The sequences of primers are shown in Supplementary Tables S4, S5. These second PCR products were purified with MultiScreen-PCR Plates (Millipore, Bedford, MA, USA). Sequence reactions were carried out with a BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA). Sequences were analyzed with an Applied Biosystems 3130xl Genetic Analyzer (Applied Biosystems). In some cases, the light-strand sequences could not be determined on the 3′ side of long stretches of C due to T>C transitions at positions m.310, m.961 and m.16189 or due to the poly-C sequences at m.568-m.573 and m.5895-m.5899. For such cases, the heavy-strand (reverse) sequences were determined for five segments. The mtDNA sequences of 185 elite Japanese athletes were deposited in GenBank as the following accession numbers: AP013108–AP013292, and the sequences are attached in FASTA format in Supplementary Data 1. Each of the mtDNA sequences was compared with the original[5] and the revised[28] Cambridge reference sequences, and nucleotide variants were confirmed visually using DNA Sequencing Software Sequencher version 4.2.2 (Gene Codes Corporation, Ann Arbor, MI, USA).

On the basis of the obtained mitochondrial genomes of 185 elite Japanese athletes, a phylogenetic tree was reconstructed manually according to the previously reported mtDNA scheme of East Asians.[18,29]

Statistical analysis

MtDNA polymorphism (minor allele frequency (MAF)$\geqslant$1% in CON) and rare variant (MAF<1% in CON) frequencies between the three groups were compared using the χ^2 test and Fisher's exact test, respectively. The *P*-value, odds ratio (OR) and 95% confidence intervals (CI) were calculated. The differences in mean number of rare variants between the three groups were examined by performing the Wilcoxson rank sum test. False discovery rate *q*-values were calculated (QVALUE software, version 1.0[30]), to adjust for multiple comparisons (*q*-value<0.05 was considered significant). All other tests were conducted using JMP Genomics version 4 or JMP version 8 (SAS Institute Japan, Tokyo, Japan).

RESULTS

Definition of mtDNA polymorphisms and rare variants

All mtDNA variants detected in 185 elite Japanese athletes and haplogroup classification are shown in the mtDNA phylogenetic tree (Supplementary Figure 1). Sequence analysis of the entire mtDNA of 185 elite Japanese athletes and 672 control subjects detected a total of 1488 variants (1458 single-nucleotide variants, 7 length variants, 15 insertions and 8 deletions). Among these variants, those with a MAF of equal to or higher than 1% in CON were defined as 'polymorphisms', whereas those with a MAF of lower than 1% in CON were defined as 'rare variants'. Consequently, we detected 311 polymorphisms and 1177 rare variants in all participants.

Association analysis of mtDNA variants with elite athlete status

Table 1 displays the results of polymorphisms that reached $P<0.05$ when the frequencies of polymorphisms were compared between the three groups by χ^2 tests. For the mtDNA control region, we previously examined associations between high-frequency polymorphisms (MAF$\geqslant$5% at least in one group: CON, EMA or SPA) and elite Japanese athletic status, and reported that the frequencies of five polymorphisms (m.152T>C, m.204T>C, m.514(CA)n repeat (n$\geqslant$5), poly-C stretch at m.568–573 (C$\geqslant$7) and m.16278C>T) were significantly different between elite Japanese athletes and CON at the level of $P<0.05$.[24] However, the data of low-frequency polymorphisms (MAF = not less than 1% and less than 5%) in the control region and the data of comparisons between EMA and SPA are presented here for the first time. When the frequencies of polymorphisms were compared between the three groups, EMA displayed excess of seven polymorphisms (m.4343A>G, m.11215C>T, m.15518C>T, m.15874A>G, and our previously reported three polymorphisms: m.152T>C, m.514(CA)$_{\geqslant 5}$, poly-C stretch at m.568–573) compared with CON ($P<0.05$). Among these polymorphisms, the frequencies of three polymorphisms (poly-C stretch at m.568-573, m.4343A>G and m.15518C>T) in EMA and SPA deviated in opposite directions ($P<0.05$). On the other hand, SPA displayed excess of three polymorphisms (m.151C>T, m.15314G>A and our previously reported m.204T>C) and dearth of 9 polymorphisms (m.4833A>G, m.5108T>C, m.5601C>T, m.7600G>A, m.9377A>G, m.13563A>G, m.14200T>C, m.14569G>A and our previously reported m.16278C>T) compared with CON ($P<0.05$). Among them, the frequencies of five polymorphisms (m.16278C>T, m.204T>C, m.4833A>G, m.5108T>C and m.14569G>A) were different between EMA and SPA ($P<0.05$) (Table 1). In addition, when the frequencies of polymorphisms were compared between EMA and SPA, other two polymorphisms (m.16140T>C and m.13104A>G) were different ($P<0.05$). However, none of these polymorphisms differed significantly between groups after correcting for multiple comparisons (false discovery rate q-value$\geqslant$0.05).

To examine whether these polymorphisms are subhaplogroup-specific polymorphisms or haplogroup-nonspecific polymorphisms, we determined the locations of these polymorphisms in the mtDNA phylogenetic tree (Supplementary Figure 1 and Table 1). In all figures, the numbers in blue indicate the variants that showed higher frequencies in EMA than in SPA/CON, and those in red indicate the variants that showed higher frequencies in SPA than in EMA/CON. The polymorphisms with high frequencies in EMA are clustered in the branches of haplogroup G (Figure 1), subhaplogroup D4e/D4e2 (Figure 2a) and subhaplogroup D4g (Figure 2b). On the other hand, in gene-coding regions, we did not find any haplogroup-nonspecific polymorphisms that are associated with the elite Japanese athletic status (Table 1).

The frequencies of rare variants were also compared between the three groups by the Fisher's exact test. The list of rare variants that reached statistical significance at the level of $P<0.05$ is shown in Supplementary Table S6. Frequencies of 8 and 5 rare variants were significantly different at the level of $P<0.05$ between EMA and CON and between SPA and CON, respectively. When the frequencies of rare variants were compared between EMA and SPA, there were no variants that reached $P<0.05$.

Association of the number of rare variants per person with the elite athlete status

To determine the effect of clustering of rare variants in certain genes, we examined the association of the number of rare variants per person with elite athletic status. When we analyzed each gene region separately, there were two statistically significant differences between SPA and CON. In the region of the 12S ribosomal RNA gene (MT-RNR1) and the NADH dehydrogenase subunit I (MT-ND1) gene, the numbers of rare variants were significantly larger in SPA than in CON ($P=0.002$ and 0.049, respectively, Table 2). For protein-coding region, rare variants that cause amino-acid replacements were only included in this analysis.

DISCUSSION

From the analysis of entire mtDNA sequences of 185 elite Japanese athletes and 672 Japanese controls, we found that 7 polymorphisms, including subhaplogroup D4e2- and D4g-specific polymorphisms, were associated with the elite EMA status, whereas 12 polymorphisms, including haplogroup G- and subhaplogroup G2a-specific polymorphisms, were associated with the elite SPA status. In addition, the frequencies of 10 polymorphisms, including haplogroup G- and subhaplogroup G2a-specific polymorphisms, were different between EMA and SPA. Although none of these differences remained significant after correcting for multiple comparisons (false discovery rate q-value$\geqslant$0.05), these polymorphisms are likely candidates influencing elite Japanese athletic performance. Furthermore, the numbers of rare variants in the MT-RNR1 and MT-ND1 genes were significantly higher in SPA than in CON. To the best of our knowledge, this is the first case–control study investigating, at the level of subhaplogroups and/or rare variants, associations between mtDNA variants and elite athletic status.

The frequencies of seven polymorphisms, that is, m.4343A>G, m.11215C>T, m.15518C>T, m.15874A>G and three polymorphisms (m.152T>C, m.514(CA)$_{\geqslant 5}$, and poly-C stretch at m.568-573), which we previously reported,[24] were higher in EMA than in CON at the level of $P<0.05$. Two of these polymorphisms, namely, m.4343A>G and m.15518C>T polymorphisms, were specific for the subhaplogroup D4g, and three polymorphisms, namely, poly-C stretch at m.568-573, m.13104A>G and m.16278C>T, were mainly linked with subhaplogroup D4g (Figure 2b and Supplementary Figure 1). These EMA-associated polymorphisms in subhaplogroup D4g may have functional influences on endurance performance. One of these polymorphisms, namely m.4343A>G polymorphism, is located at the TψC loop region of the tRNA for glutamine. tRNAs have cloverleaf secondary structure due to four base-paired stems. This cloverleaf structure comprises three non-base-paired loops: D, anticodon and TψC loop. As pathogenic mutations are often located at stem structures and tended to disrupt Watson–Crick nucleotide paring in stem,[31] the m.4343A>G located at the TψC loop region is probably not deleterious. Indeed, this polymorphism has been

Table 1 Polymorphisms in the entire mtDNA with differences between groups

Polymorphism (amino-acid change)	*Minor allele*	*Gene region*	*Haplogroup/ subhaplogroup specificity*	*CON (n = 672) % (n)*	*EMA (n = 100) % (n)*	*SPA (n = 85) % (n)*	*EMA vs CON* P-*value*	*EMA vs CON* *OR (95% CI)*	*EMA vs CON* q-*value*	*SPA vs CON* P-*value*	*SPA vs CON* *OR (95% CI)*	*SPA vs CON* q-*value*	*EMA vs SPA* P-*value*	*EMA vs SPA* *OR (95% CI)*	*EMA vs SPA* q-*value*
Control region															
m.16140T>C	C	Control region	B5, B4c1b, M7a2	5.7 (38)	1.0 (1)	9.4 (8)	0.047	0.17 (0.02–1.24)	0.805	0.172	1.73 (0.78–3.85)	0.932	**0.008**	**0.10 (0.10–0.79)**	0.524
m.16278C>T	T	Control region	B4d, D4g1, G2	8.3 (56)	10.0 (10)	0.0 (0)	0.578	1.22 (0.60–2.48)	0.829	**0.006**	**0.00**	0.918	**0.003**	**INF**	0.524
m.151C>T	T	Control region	D4a2a, D5c, B4d3, F1b1a1a1, M7b2b, Z1	1.2 (8)	1.0 (1)	4.7 (4)	0.869	0.84 (0.10–6.78)	0.872	**0.014**	**4.10 (1.21–13.91)**	0.918	0.121	0.20 (0.02–1.87)	0.673
m.152T>C	C	Control region	A, A1a, A1b, A2a, A3, A5c, B4b1b, B5b3, D4a, D4b1a1,D4f, D4l1a, D5b1a, D5c, F1, F1b1, F1c, F4a, G2a1, G3a, G4a, N1b, N9b, M7a1a6, M7a1b, M8a2, M12, Y1b, Z	18.5 (124)	28.0 (28)	24.7 (21)	**0.025**	**1.72 (1.07–2.77)**	0.805	0.167	1.45 (0.85–2.46)	0.932	0.613	1.19 (0.61–2.29)	0.950
m.204T>C	C	Control region	B4d3, B5, D4d1a, M7a1a, M7a1b, N9a2, Z4	4.9 (33)	2.0 (2)	10.7 (9)	0.192	0.40 (0.09–1.67)	0.805	**0.029**	**2.32 (1.07–5.02)**	0.918	**0.013**	**0.17 (0.04–0.81)**	0.524
m.514(CA)n	(CA)⩾5	Control region	A, B4a, B4c1b1, B4e, B5, C1, D4b, D4c, D5a2, F1, M7a1a, M7c, M10a, N1b, Z5, Z3	40.6 (273)	27.0 (27)	40.0 (34)	**0.009**	**1.85 (1.16–2.95)**	0.805	0.912	1.03 (0.65–1.63)	0.982	0.061	0.55 (0.30–1.03)	0.641
Poly-C stretch at m.568-573	C⩾7	Control region	C5, D4g1, F4b, G4a, M10	4.0 (27)	11.0 (11)	1.2 (1)	**0.003**	**2.95 (1.42–6.16)**	0.417	0.191	0.28 (0.04–2.12)	0.932	**0.007**	**10.38 (1.31–82.17)**	0.524
RNA-coding region															
m.4343A>G	G	tRNA Gln	D4g	2.4 (16)	6.0 (6)	0.0 (0)	**0.042**	**2.62 (1.00–6.85)**	0.805	0.150	0.00	0.918	**0.022**	**INF**	0.524
m.5601C>T	T	tRNA Ala	G2a	4.8 (32)	3.0 (3)	0.0 (0)	0.429	0.62 (0.19–2.06)	0.805	**0.040**	**0.00**	0.918	0.107	INF	0.673
Protein-coding region															
m.4833A>G (Thr122Ala)	G	ND2	G	8.6 (58)	11.0 (11)	2.4 (2)	0.438	1.31 (0.66–2.59)	0.805	**0.044**	**0.26 (0.06–1.06)**	0.918	**0.022**	**5.13 (1.10–23.83)**	0.524
m.5108T>C	C	ND2	G, B4c2, M7a1b	9.4 (63)	12.0 (12)	2.4 (2)	0.408	1.32 (0.68–2.54)	0.805	**0.029**	**0.23 (0.06–0.97)**	0.918	**0.013**	**5.66 (1.23–26.05)**	0.524
m.7600G>A	A	COII	G2a	4.8 (32)	3.0 (3)	0.0 (0)	0.429	0.62 (0.19–2.06)	0.805	**0.040**	**0.00**	0.918	0.107	INF	0.673
m.9377A>G	G	COIII	G2a, D5b2	5.0 (34)	3.0 (3)	0.0 (0)	0.368	0.58 (0.17–1.93)	0.805	**0.034**	**0.00**	0.918	0.107	INF	0.673
m.11215C>T	T	ND4	D4e	4.8 (32)	10.0 (10)	3.5 (3)	**0.031**	**2.22 (1.06–4.67)**	0.805	0.610	0.73 (0.22–2.44)	0.982	0.086	3.04 (0.81–11.42)	0.673
m.13104A>G	G	ND5	D4g, D4k3	3.0 (20)	6.0 (6)	0.0 (0)	0.118	2.08 (0.81–5.31)	0.805	0.107	0.00	0.918	**0.022**	**INF**	0.524
m.13563A>G	G	ND5	G2	4.8 (32)	3.0 (3)	0.0 (0)	0.429	0.62 (0.19–2.06)	0.805	**0.040**	**0.00**	0.918	0.107	INF	0.673
m.14200T>C	C	ND6	G2a	4.5 (30)	3.0 (3)	0.0 (0)	0.499	0.966 (0.20–2.21)	0.805	**0.047**	**0.00**	0.918	0.107	INF	0.673
m.14569G>A	A	ND6	G, B4b1b, N9a2c	9.5 (64)	13.0 (13)	2.4 (2)	0.279	1.42 (0.75–2.68)	0.805	**0.027**	**0.23 (0.06–0.95)**	0.918	**0.008**	**6.20 (1.36–28.32)**	0.524
m.15314G>A (Ala190Thr)	A	Cyt*b*	D4a1a1	1.0 (7)	1.0 (1)	4.7 (4)	0.969	0.96 (0.12–7.88)	0.878	**0.008**	**4.69 (1.34–16.37)**	0.918	0.121	0.20 (0.02–1.87)	0.673
m.15518C>T	T	Cyt*b*	D4g	2.4 (16)	6.0 (6)	0.0 (0)	**0.042**	**2.62 (1.00–6.85)**	0.805	0.150	0.00	0.918	**0.022**	**INF**	0.524
m.15874A>G	G	Cyt*b*	D4e2, A5	3.4 (23)	10.0 (10)	3.5 (3)	**0.002**	**3.14 (1.45–6.80)**	0.417	0.959	1.03 (0.30–3.51)	0.991	0.086	3.04 (0.81–11.42)	0.673

Abbreviations: Ala, alanine; CO, cytochrome c oxidase; CON, controls; Cyt*b*, cytochrome *b*; Gln, glutamine; EMA, endurance/middle-power athletes; mtDNA, mitochondrial DNA; ND, NADH dehydrogenase; SPA, sprint/power athletes; INF, infinity; Thr, threonine.
Values in bold indicate significant differences between groups.

predicted to be benign by an evolutionarily-based computational analysis.[32] Although there are no functional studies, Zhu *et al.*[33] reported that m.4343A>G polymorphism was associated with essential hypertension and was highly conserved from bacteria to humans. This substitution may result in a functional difference of tRNA for glutamine. tRNAs encoded by mtDNA are essential for the synthesis of the 13 subunits of mitochondrial OXPHOS. Therefore, it is possible that this polymorphism could influence elite EMA status through improved efficiency in protein synthesis within the mitochondria, that is, mitochondrial biogenesis, although detailed mechanism of the association between m.4343A>G polymorphism and elite EMA status remains unclear. On the other hand, m.11215C>T and m.15874A>G were mainly linked with subhaplogroups D4e and D4e2, respectively (Figure 2a and Supplementary Figure 1). As both polymorphisms are synonymous substitutions, it is unclear why these polymorphisms are enriched in the elite EMA. One possibility is the 'golden mean hypothesis'. We previously reported the absence of certain mtDNA variants in centenarians and their presence in patients with Parkinson's disease.[34] These findings imply that these variants are not beneficial for long-term survival but appear to predispose individuals to certain adult-onset diseases, and that centenarians are genetically hitting the golden mean. In the present study, the association between subhaplogroup D4e and elite EMA status could be explained by improved mitochondrial function through the absence of non-synonymous variants in subhaplogroup D4e.

The frequencies of nine polymorphisms (m.4833A>G, m.5108T>C, m.5601C>T, m.7600G>A, m.9377A>G, m.13563A>G, m.14200T>C, m.14569G>A and our previously reported m.16278C>T) were lower in SPA than in CON at the level of $P<0.05$. These polymorphisms were all observed at the branches of haplogroup G or subhaplogroup G2a (Figure 1 and Supplementary Figure 1) and may therefore suggest that haplogroup G and/or subhaplogroup G2a possess variants not beneficial/essential for sprint/power performance. One of these polymorphisms, namely, m.4833A>G, is a non-synonymous substitution. This polymorphism causes the Thr122Ala replacement in NADH dehydrogenase subunit 2, which is a subunit of the complex I. It is possible that this replacement is radical, because the physicochemical difference between the original and altered amino-acid residues is relatively high, with a Grantham value of 58;[35] the average of Grantham value is 50. Furthermore, this replacement is predicted to be deleterious with a PolyPhen-2 score of 0.665. Notably, impairment in complex I activity

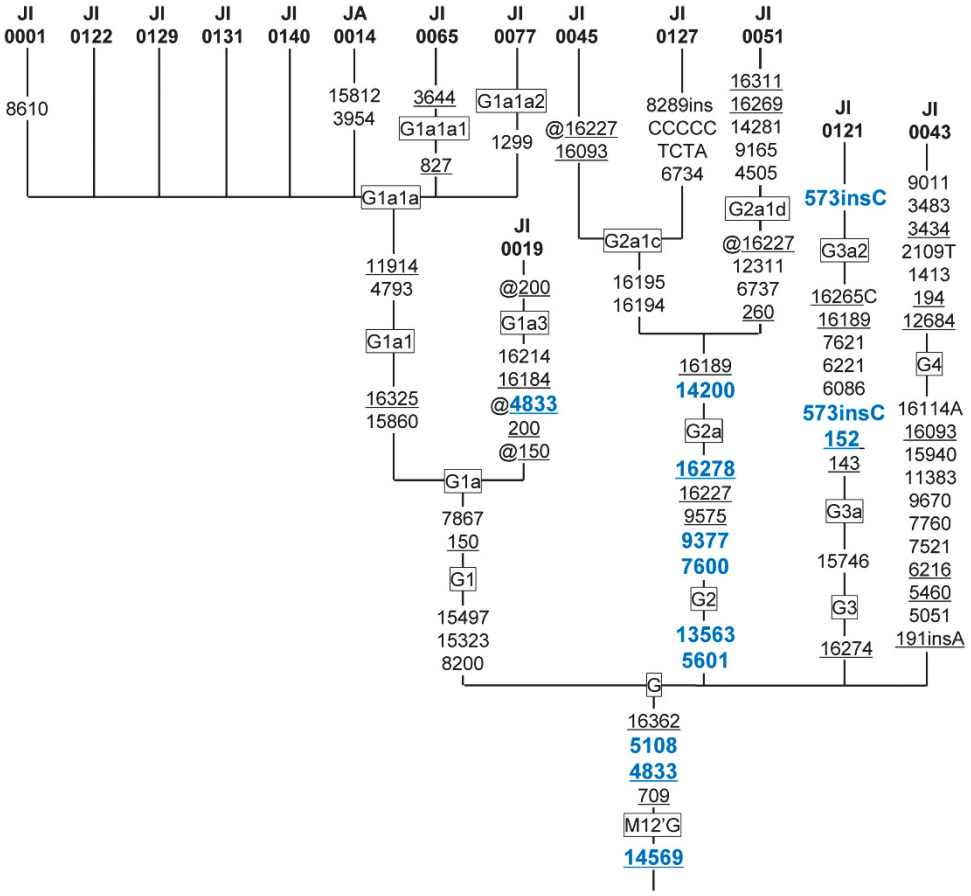

Figure 1 The phylogenetic tree of haplogroup G. Nine polymorphisms (m.4833A>G, m.5108T>C, m.5601C>T, m.7600G>A, m.9377A>G, m.13563A>G, m.14200T>C, m.14569G>A and m.16278C>T), which were significantly lower in SPA than in CON, were all observed at the branches of haplogroup G or subhaplogroup G2a. The numbers along the links refer to nucleotide positions. Combination of alphabetic and numeric characters in square boxes indicate haplogroup and subhaplogroup names. Alphanumeric characters with black and bold indicate subject ID. Suffixes A, C, G and T refer to transversions; 'Y' means heteroplasmic variations (that is, C/T); 'ins' indicates an insertion event (the exact number of the inserted nucleotide(s) was disregarded). Recurrent (homoplasic) variations are underlined. At marks (@) indicate revertants. '573insC' indicates poly-C stretch at m.568-573. The numbers in blue indicate the variants that showed higher frequencies in EMA than in SPA/CON.

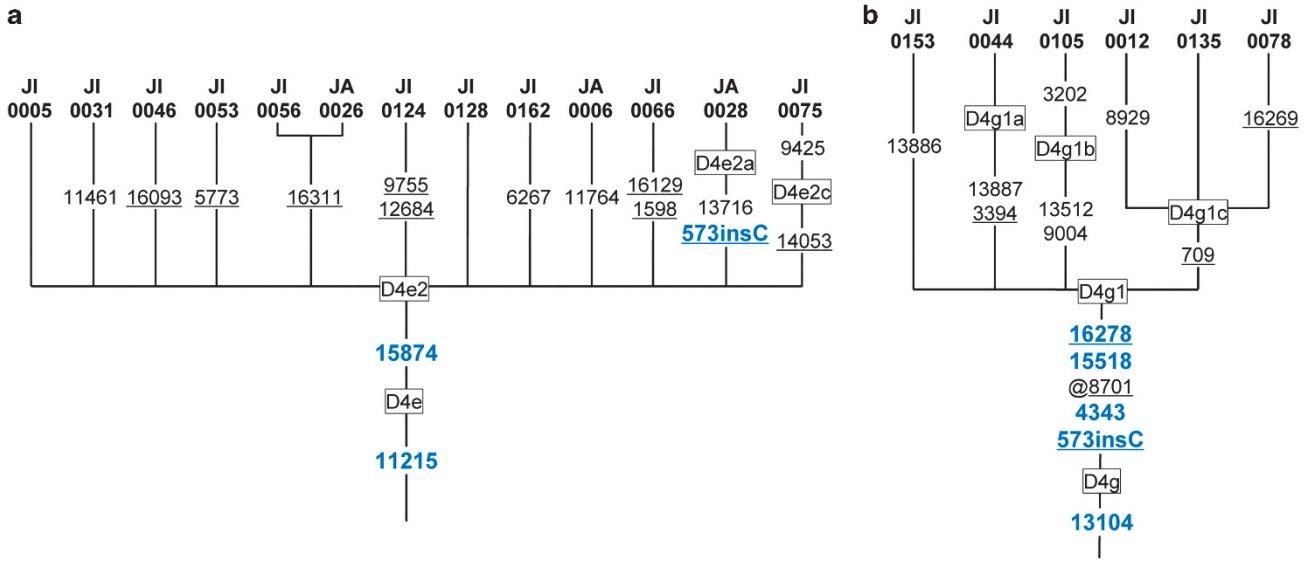

Figure 2 The phylogenetic trees of subhaplogroup D4e (**a**) and D4g (**b**). EMA associated with two polymorphisms (m.11215C>T and m.15874A>G) and five polymorphisms (m.4343A>G, m.15518C>T, poly-C stretch at m.568-573, m.13104A>G and m.16278C>T) were linked with subhaplogroups D4e/D4e2 and D4g, respectively. The numbers along the links refer to nucleotide positions. Combination of alphabetic and numeric characters in square boxes indicates haplogroup and subhaplogroup names. Alphanumeric characters with black and bold indicate subject ID. Recurrent (homoplasic) variations are underlined. At marks (@) indicate revertants. '573insC' indicates poly-C stretch at m.568-573. The numbers in blue indicate the variants which showed higher frequencies in EMA than in SPA/CON.

Table 2 Distribution of the number of rare variations between groups

Region	*Gene*	*CON* (n = *672*)				*EMA* (n = *100*)				*SPA* (n = *85*)				*EMA vs CON*	*SPA vs CON*	*EMA vs SPA*
		Mean	*Median*	*s.d.*	*Range*	*Mean*	*Median*	*s.d.*	*Range*	*Mean*	*Median*	*s.d*	*Range*	P-*value*	P-*value*	P-*value*
All region		1.38	1	1.67	0–10	1.52	1	1.98	0–8	1.49	1	1.61	0–8	0.973	0.343	0.500
Control region		0.47	0	0.81	0–7	0.4	0	0.75	0–4	0.47	0	0.73	0–3	0.367	0.752	0.359
rRNA	All rRNA	0.18	0	0.45	0–3	0.24	0	0.57	0–3	0.22	0	0.50	0–2	0.455	0.408	0.933
	12S rRNA	**0.06**	**0**	**0.25**	**0–2**	0.08	0	0.31	0–2	**0.14**	**0**	**0.35**	**0–1**	0.456	**0.002**	0.120
	16S rRNA	0.12	0	0.36	0–3	0.16	0	0.42	0–2	0.08	0	0.28	0–1	0.392	0.402	0.207
tRNA	All tRNA	0.17	0	0.42	0–2	0.26	0	0.63	0–3	0.14	0	0.38	0–2	0.502	0.555	0.356
Protein-coding region	All genes	0.56	0	0.89	0–6	0.62	0	0.99	0–4	0.66	0	0.85	0–4	0.823	0.115	0.290
	ND1	**0.04**	**0**	**0.21**	**0–1**	0.07	0	0.26	0–1	**0.09**	**0**	**0.29**	**0–1**	0.268	**0.049**	0.550
	ND2	0.05	0	0.24	0–2	0.04	0	0.24	0–2	0.04	0	0.19	0–1	0.439	0.606	0.850
	ND3	0.00	0	0.04	0–1	0.01	0	0.10	0–1	0.00	0	0.00	0	0.119	0.722	0.357
	ND4	0.04	0	0.21	0–2	0.03	0	0.17	0–1	0.08	0	0.28	0–1	0.578	0.094	0.118
	ND4L	0.00	0	0.05	0–1	0.01	0	0.10	0–1	0.00	0	0.00	0	0.293	0.615	0.357
	ND5	0.06	0	0.25	0–2	0.09	0	0.32	0–2	0.08	0	0.28	0	0.315	0.316	0.970
	ND6	0.05	0	0.24	0–2	0.02	0	0.14	0–1	0.01	0	0.11	0–1	0.191	0.117	0.659
	Cyt*b*	0.09	0	0.29	0–2	0.08	0	0.27	0–1	0.07	0	0.26	0–1	0.722	0.535	0.810
	CO1	0.04	0	0.22	0–2	0.07	0	0.26	0–1	0.11	0	0.38	0–2	0.208	0.087	0.721
	CO2	0.04	0	0.19	0–1	0.05	0	0.22	0–1	0.02	0	0.15	0–1	0.591	0.486	0.348
	CO3	0.04	0	0.20	0–2	0.06	0	0.24	0–1	0.05	0	0.21	0–1	0.281	0.658	0.699
	ATPase6	0.08	0	0.29	0–2	0.09	0	0.29	0–1	0.09	0	0.29	0–1	0.753	0.673	0.923
	ATPase8	0.01	0	0.09	0–1	0	0	0.00	0	0.01	0	0.11	0–1	0.343	0.797	0.278

Abbreviations: Cyt*b*, cytochrome *b*; CO, cytochrome c oxidase; CON, controls; EMA, endurance/middle-power athletes; ND, NADH dehydrogenase; SPA, sprint/power athletes.
For protein-coding region, the distribution and *P*-value of the number of non-synonymous variations per person is shown. Values in bold indicate significant differences between groups.

has been reported in cybrid cells with the m.4833A>G polymorphism;[36] albeit mtDNA with m.4833A>G polymorphism was obtained from only one patient with type 2 diabetes mellitus and therefore other polymorphisms or rare variants may have caused the impairment in complex I. Further extensive studies are required to clarify the functional significance of m.4833A>G polymorphism.

Similarly, m.5601C>T polymorphism in the TψC loop region of tRNA for alanine and m.16278C>T polymorphism in the control region also appear to be non beneficial for sprint/power performance.

The frequencies of three polymorphisms (m.151C>T, m.15314G>A and our previously reported m.204T>C) were higher in SPA than in CON at the level of $P<0.05$. Among these polymorphisms, m.151C>T and m.204T>C are located in the control region and the frequencies of these polymorphisms tended to deviate in opposite directions in SPA and EMA. These results could be explained, in part at least, by our previous contention that different genotypes at a single locus in the mtDNA control region may have opposite effects on elite SPA and EMA status through alterations in muscle fiber-type composition and the capacity for OXPHOS and glycolytic flux.[24] The other polymorphism, namely, m.15314G>A, is a non-synonymous substitution causing the Ala190Thr replacement in cytochrome *b*, which is a subunit of the complex III. Although the Grantham value of this substitution is relatively high, having a value of 58, this replacement was predicted to be benign with a PolyPhen-2 score of 0.00. Thus, the impact of this Ala190Thr replacement is largely unknown.

The frequencies of 10 polymorphisms were different and in opposite directions between EMA and SPA ($P<0.05$). Similarly, we previously reported that mitochondrial haplogroup F and N9a were associated with type 2 diabetes in opposite directions, namely, haplogroup F was associated with an increased risk of type 2 diabetes mellitus, whereas haplogroup N9a was associated with resistance against type 2 diabetes mellitus.[37,38] Interestingly, Hwang *et al.*[39] reported that cybrid cells harboring these haplogroups, that is, F and N9a, exhibited significant differences in their nuclear gene expression pattern; mitochondrial haplogroup F showed a decreased gene expression of the mitochondrial OXPHOS pathway and an increased gene expression of the cytosolic glycolysis pathway compared with mitochondrial haplogroup N9a. This observation can be regarded as a compensatory response for decreased ATP production caused by a defective mitochondrial haplogroup, resulting in an increased expression of nuclear genes involved in glycolysis. This phenomenon might explain, at least partly, the opposite effects of mtDNA polymorphisms on elite SPA and EMA status.

Our results showed that the mtDNA sequences of elite SPA contained a higher number of rare variants in the MT-RNR1 and MT-ND1 genes than in CON. It has been argued that a higher number of rare variants in certain genes could influence susceptibility to Alzheimer's disease[40,41] and hypertriglyceridemia.[25] These rare variants were predicted to have comprised function.[25] In addition, it was suggested that rare variants comprise the largest part of disease-associated variants by use of a comprehensive weighted scoring system (MtSNPscore).[42] In the present study, three rare variants with amino-acid replacements in the MT-ND1 observed in SPA (m.3410A>T (Lys35Met), m.3571C>T (Lue89Phe) and m.4136A>G (Tyr277Cys) were predicted to be deleterious with PolyPhen-2 scores of 0.984–1. Therefore, it is possible that these rare variants reduce mitochondrial function and enhance the potential for glycolytic flux as compensation, and hence the association with elite SPA status. In the present study, we also compared the frequencies of rare variants between groups and found eight variants (m.2857T>C, m.3434A>G, m.8296A>G, m.8962A>G, m.9711C>T, m.12630G>A, m.12684G>A and m.13140A>G) were different between EMA and CON and five variants (m.5261G>A, m.7694C>T, m.7852G>A, m.10103A>G and m.11150G>A) between SPA and CON; whereas no differences between EMA and SPA were found. Among these variants, four variants (m.2857T>C, m.8962A>G, m.9711C>T and m.13140A>G) with high frequencies in EMA are clustered in the branches of haplogroup A3 (Supplementary Figure 1 and Supplementary Table S6). We previously reported that haplogroup A was associated with elite endurance status.[43] Therefore, these rare variants may influence elite athletic performance solely and/or through clustering.

We previously reported that mitochondrial haplogroups F and G1 were associated with elite Japanese SPA and EMA status, respectively.[10] The distribution of major mitochondrial haplogroups in EMA, SPA and CON in the present study is shown in Supplementary Table S7. In the present study, the frequency of haplogroup F-specific polymorphism (m.3970C>T) was not significantly different between SPA and CON, although this polymorphism tended to be different (SPA: 9/85 vs CON: 40/672, $P=0.1017$, OR: 1.87 (95% CI 0.87–4.01), data not shown). However, when we combined SPA with EMA into one group as an all-athlete group, this group showed higher frequency of haplogroup F-specific polymorphism (m.3970C>T) compared with CON (all athlete: 20/185 vs CON: 40/672, $P=0.0218$, OR: 1.92 (95% CI 1.09–3.36), data not shown). These results imply that haplogroup F is associated with elite performance regardless of event specialization. On the other hand, the frequency of haplogroup G in EMA was significantly higher than in SPA. Further replication studies are necessary to confirm these associations.

Although the results of the present study are intriguing, there are several limitations. The first limitation is that our athlete cohort is made up of athletes from a variety of different sporting disciplines. However, the truly elite nature of our athlete cohort combine with the technical approaches used provides a fairly unique opportunity to study genetic associations with sporting success. A second key limitation is the problem of multiple comparisons and the finding that no polymorphisms differed significantly between groups after correcting for multiple comparisons; a reflection most likely of a lack of sufficient statistical power, hence all associations significant at $P<0.05$ have been reported in the present study in order to avert type 2 errors. Therefore, replication studies are required to confirm the presently reported associations between mtDNA variants and elite athletic performance and functional studies to investigate the underlying mechanisms of these associations. In the present study, we focused exclusively on the mitochondrial genome. However, it is acknowledged that a comprehensive analysis of the nuclear genome will be essential in order to significantly enhance our understanding of the genetic basis of elite performance. Elucidating these genetic contributions to athletic performance will substantially enhance the possibility of using genetics to enhance the detection of sporting talent.

In conclusion, analysis of the entire mtDNA of elite Japanese athletes revealed several haplogroup- and subhaplogroup-specific polymorphisms to be potentially associated with elite Japanese athletic status. Furthermore, the presence of rare variants in the MT-RNR1 and MT-ND1 genes may also contribute to sprint/power performance.

ACKNOWLEDGEMENTS

This work was supported in part by a Grant-in-Aid for Young Scientists (A-21680050 to NF); by a Grant-in-Aid for Scientific Research (A-22240072 to MT); and a Grant-in-Aid for the Global COE (Sport Sciences for the Promotion of Active Life to Waseda University) from the Ministry of Education, Culture, Sports, Science and Technology; and by a research project in Japan Institute of Sports Sciences (to TK). EM was the recipient of Grant-in-Aid for JSPS Fellows from the Japan Society for the Promotion of Science.

1 De Moor, M. H., Spector, T. D., Cherkas, L. F., Falchi, M., Hottenga, J. J., Boomsma, D. I. *et al*. Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Res. Hum. Genet.* **10,** 812–820 (2007).
2 Bray, M. S., Hagberg, J. M., Perusse, L., Rankinen, T., Roth, S. M., Wolfarth, B. *et al.* The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. *Med. Sci. Sports Exerc.* **41,** 35–73 (2009).
3 Roth, S. M., Rankinen, T., Hagberg, J. M., Loos, R. J., Perusse, L., Sarzynski, M. A. *et al.* Advances in exercise, fitness, and performance genomics in 2011. *Med. Sci. Sports Exerc.* **44,** 809–817 (2012).
4 Pitsiladis, Y. & Wang, G. Necessary advances in exercise genomics and likely pitfalls. *J. Appl. Physiol.* **110,** 1150–1151 (2011).
5 Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290,** 457–465 (1981).
6 Bouchard, C., Daw, E. W., Rice, T., Perusse, L., Gagnon, J., Province, M. A. *et al.* Familial resemblance for VO2max in the sedentary state: the HERITAGE family study. *Med. Sci. Sports Exerc.* **30,** 252–258 (1998).
7 Lesage, R., Simoneau, J. A., Jobin, J., Leblanc, J. & Bouchard, C. Familial resemblance in maximal heart rate, blood lactate and aerobic power. *Hum. Hered.* **35,** 182–189 (1985).
8 Perusse, L., Gagnon, J., Province, M. A., Rao, D. C., Wilmore, J. H., Leon, A. S. *et al.* Familial aggregation of submaximal aerobic performance in the HERITAGE Family study. *Med. Sci. Sports Exerc.* **33,** 597–604 (2001).
9 Scott, R. A., Fuku, N., Onywera, V. O., Boit, M., Wilson, R. H., Tanaka, M. *et al.* Mitochondrial haplogroups associated with elite Kenyan athlete status. *Med. Sci. Sports Exerc.* **41,** 123–128 (2009).
10 Mikami, E., Fuku, N., Takahashi, H., Ohiwa, N., Scott, R., Pitsiladis, Y. *et al.* Mitochondrial haplogroups associated with elite Japanese athlete status. *Br. J. Sports Med.* 1179–1183 (2011).
11 Deason, M., Scott, R., Irwin, L., Macaulay, V., Fuku, N., Tanaka, M. *et al.* Importance of mitochondrial haplotypes and maternal lineage in sprint performance among individuals of West African ancestry. *Scand. J. Med. Sci. Sports* **22,** 217–223 (2012).
12 Fuku, N., Murakami, H., Iemitsu, M., Sanada, K., Tanaka, M. & Miyachi, M. Mitochondrial macrohaplogroup associated with muscle power in healthy adults. *Int. J. Sports Med.* **33,** 410–414 (2012).
13 Castro, M. G., Terrados, N., Reguero, J. R., Alvarez, V. & Coto, E. Mitochondrial haplogroup T is negatively associated with the status of elite endurance athlete. *Mitochondrion* **7,** 354–357 (2007).
14 Kim, K. C., Cho, H. I. & Kim, W. MtDNA Haplogroups and Elite Korean Athlete Status. *Int. J. Sports Med.* **33,** 76–80 (2012).
15 Niemi, A. K. & Majamaa, K. Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *Eur. J. Hum. Genet.* **13,** 965–969 (2005).
16 Nogales-Gadea, G., Pinos, T., Ruiz, J. R., Marzo, P. F., Fiuza-Luces, C., Lopez-Gallardo, E. *et al.* Are mitochondrial haplogroups associated with elite athletic status? A study on a Spanish cohort. *Mitochondrion* **11,** 905–908 (2011).
17 Maca-Meyer, N., Gonzalez, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2,** 13 (2001).
18 Tanaka, M., Cabrera, V. M., Gonzalez, A. M., Larruga, J. M., Takeyasu, T., Fuku, N. *et al.* Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.* **14,** 1832–1850 (2004).
19 Elson, J. L., Turnbull, D. M. & Howell, N. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am. J. Hum. Genet.* **74,** 229–238 (2004).
20 Ruiz-Pesini, E. & Wallace, D. C. Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA. *Hum. Mutat.* **27,** 1072–1081 (2006).
21 Alexe, G., Fuku, N., Bilal, E., Ueno, H., Nishigaki, Y., Fujita, Y. *et al.* Enrichment of longevity phenotype in mtDNA haplogroups D4b2b, D4a, and D5 in the Japanese population. *Hum. Genet.* **121,** 347–356 (2007).
22 Bilal, E., Rabadan, R., Alexe, G., Fuku, N., Ueno, H., Nishigaki, Y. *et al.* Mitochondrial DNA haplogroup D4a is a marker for extreme longevity in Japan. *PLoS One* **3,** e2421 (2008).
23 Tanaka, M., Gong, J. S., Zhang, J., Yoneda, M. & Yagi, K. Mitochondrial genotype associated with longevity. *Lancet* **351,** 185–186 (1998).
24 Mikami, E., Fuku, N., Takahashi, H., Ohiwa, N., Scott, R., Pitsiladis, Y. *et al.* Polymorphisms in the control region of mitochondrial DNA associated with elite Japanese athlete status. *Scand. J. Med. Sci. Sports* (e-pub ahead of print 31 January 2012; doi:10.1111/j.1600-0838.2011.01424.X).
25 Johansen, C. T., Wang, J., Lanktree, M. B., Cao, H., McIntyre, A. D., Ban, M. R. *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat. Genet.* **42,** 684–687 (2010).
26 Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S. *et al.* ACTN3 genotype is associated with human elite athletic performance. *Am. J. Hum. Genet.* **73,** 627–631 (2003).
27 Tanaka, M., Hayakawa, M. & Ozawa, T. Automated sequencing of mitochondrial DNA. *Methods Enzymol.* **264,** 407–421 (1996).
28 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23,** 147 (1999).
29 Kong, Q. P., Bandelt, H. J., Sun, C., Yao, Y. G., Salas, A., Achilli, A. *et al.* Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum. Mol. Genet.* **15,** 2076–2086 (2006).
30 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100,** 9440–9445 (2003).
31 McFarland, R., Elson, J. L., Taylor, R. W., Howell, N. & Turnbull, D. M. Assigning pathogenicity to mitochondrial tRNA mutations: when "definitely maybe" is not good enough. *Trends Genet.* **20,** 591–596 (2004).
32 Kondrashov, F. A. Prediction of pathogenic mutations in mitochondrially encoded human tRNAs. *Hum. Mol. Genet.* **14,** 2415–2419 (2005).
33 Zhu, H. Y., Wang, S. W., Liu, L., Chen, R., Wang, L., Gong, X. L. *et al.* Genetic variants in mitochondrial tRNA genes are associated with essential hypertension in a Chinese Han population. *Clin. Chim. Acta.* **410,** 64–69 (2009).
34 Tanaka, M., Fuku, N., Takeyasu, T., Guo, L. J., Hirose, R., Kurata, M. *et al.* Golden mean to longevity: rareness of mitochondrial cytochrome b variants in centenarians but not in patients with Parkinson's disease. *J. Neurosci. Res.* **70,** 347–355 (2002).
35 Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185,** 862–864 (1974).
36 Ohkubo, E., Aida, K., Chen, J., Hayashi, J. I., Isobe, K., Tawata, M. *et al.* A patient with type 2 diabetes mellitus associated with mutations in calcium sensing receptor gene and mitochondrial DNA. *Biochem. Biophys. Res. Commun.* **278,** 808–813 (2000).
37 Fuku, N., Nishigaki, Y. & Tanaka, M. Mitochondrial haplogroup N9a confers resistance against metabolic syndrome and type 2 diabetes mellitus in Asian individuals. *Asia Pac. J. Endocrinol.* **2009,** 65–73 (2009).
38 Fuku, N., Park, K. S., Yamada, Y., Nishigaki, Y., Cho, Y. M., Matsuo, H. *et al.* Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians. *Am. J. Hum. Genet.* **80,** 407–415 (2007).
39 Hwang, S., Kwak, S. H., Bhak, J., Kang, H. S., Lee, Y. R., Koo, B. K. *et al.* Gene expression pattern in transmitochondrial cytoplasmic hybrid cells harboring type 2 diabetes-associated mitochondrial DNA haplogroups. *PLoS One* **6,** e22116 (2011).
40 Elson, J. L., Herrnstadt, C., Preston, G., Thal, L., Morris, C. M., Edwardson, J. A. *et al.* Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? *Hum. Genet.* **119,** 241–254 (2006).
41 Tanaka, N., Goto, Y., Akanuma, J., Kato, M., Kinoshita, T., Yamashita, F. *et al.* Mitochondrial DNA variants in a Japanese population of patients with Alzheimer's disease. *Mitochondrion* **10,** 32–37 (2010).
42 Bhardwaj, A., Mukerji, M., Sharma, S., Paul, J., Gokhale, C. S., Srivastava, A. K. *et al.* MtSNPscore: a combined evidence approach for assessing cumulative impact of mitochondrial variations in disease. *BMC Bioinformatics* **10** (Suppl 8), S7 (2009).
43 Tanaka, M., Takeyasu, T., Fuku, N., Li-Jun, G. & Kurata, M. Mitochondrial genome single nucleotide polymorphisms and their phenotypes in the Japanese. *Ann. N. Y. Acad. Sci.* **1011,** 7–20 (2004).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)