SHORT COMMUNICATION

Haplotype diversity in mitochondrial genome in a Chinese Han population

Ke Ma[1], Hui Li[1], Yu Cao[2,3], Xuejun Zhao[1], Wenbin Liu[1] and Xueying Zhao[1]

Investigations into the use of mitochondrial genome (mtGenome) typing by massively parallel sequencing technologies are well underway in many areas, including forensic genetics. Previous studies have demonstrated that mtGenome sequencing data generated from Ion torrent personal genome machine (PGM) system were highly viable and reliable in forensic research. In this study, 145 whole mtGenomes from unrelated Chinese Han population were sequenced using the Ion PGM system. Results showed that 145 distinct haplotypes were obtained at a relatively high coverage with limited strand bias. The distribution of variants across the entire mtGenomes was illustrated and 70.74% of the variants were observed outside of the control region. An overall increase in the number of unique haplotypes as well as haplotype diversity were observed by detection of mtGenome compared with hypervariable region I/II (HV I/II) and control region (CR). This study demonstrates the substantially higher degree of haplotype resolution with whole-mtGenome sequences in comparison to HV I/II or CR that historically targeted for forensic testing, which shows the potential value of mtGenome typing in forensic testing in the future.
Journal of Human Genetics (2016) **61,** 903–906; doi:10.1038/jhg.2016.74; published online 16 June 2016

INTRODUCTION

Mitochondrial DNA (mtDNA) typing is extensively used in many areas including medical genetics, evolutionary anthropology, genetic genealogy and forensic genetics.[1–3] mtDNA is present in higher copy number in the human cell than nuclear DNA and thus particularly useful for forensic testing and ancient DNA analyses,[4] where DNA may be highly fragmented and damaged or specimens may contain little or no nuclear DNA such as a fingernail and a hair shaft without root.[5] Additional features of mtDNA relate to its maternal mode of inheritance and lack of recombination, which make it particularly distinctive and informative in kinship analyses.[6,7] The traditional method for forensic testing of mtDNA is Sanger-type sequencing (STS) of the hypervariable regions I and II (HVI and HVII).[4,8] According to the latest recommendations for mtDNA typing, the entire control region (CR) of the mtDNA should be considered for forensic genetic casework and databasing purposes.[9] To date, forensic databases contain limited coding region data of mtDNA, whereas some population genetic databases and medical data sets contain whole-mitochondrial genome (mtGenome) sequences.[8,10] Recent studies have demonstrated that >70% of the mtDNA variants can be located outside HVI/II for some haplogroups,[11,12] thus whole-mtGenome sequences provide far greater discrimination power for forensic testing than traditional mtDNA typing data.[12]

Massively parallel sequencing (MPS), also termed next generation sequencing, hold great potential to expand forensic mtDNA testing beyond current capacity. Compared to traditional STS, MPS technologies make possible higher throughput sequencing at a substantially reduced cost, thus facilitating the establishment of larger mtGenome databases in relatively short terms.[10,13] Recent studies have demonstrated that mtGenome sequences obtained by MPS were highly concordant with those obtained by STS;[4,13–16] moreover, MPS can effectively recover usable mtDNA profiles even from highly degraded and damaged specimens.[11] To date, Chinese population mtGenome data generated from MPS that analysis in forensic application have not been described.[8,10] In this study, we report 145 distinct mtGenome haplotypes from Chinese population, and evaluate the mtGenome typing by comparison of statistics including variant distribution, haplotype assignment and discrimination power to traditionally HVI/II and CR typing.

MATERIALS AND METHODS

Sample collection and DNA extraction

Peripheral blood samples were collected from 145 unrelated Chinese Han individuals from Shanghai City, Eastern China. Informed consent was obtained from each participant. Genomic DNA was extracted using the BioRobot EZ1 Advanced XL and the EZ1 DNA Investigator Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol. The quantity of recovered DNA was determined using the Qubit dsDNA BR Assay Kit on a Qubit 2.0 Fluorometer (Thermo Fisher, Foster City, CA, USA). This study was approved by the Ethical Committee of Fudan University.

[1]Shanghai Key Laboratory of Crime Scene Evidence, Shanghai Research Institute of Criminal Science and Technology, Shanghai, China; [2]Key Laboratory of Forensic Evidence and Science Technology, Ministry of Public Security, Institute of Forensic Science, Shanghai Public Security Bureau, Shanghai, China and [3]State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai, China
Correspondence: Dr W Liu or Dr X Zhao, Shanghai Key Laboratory of Crime Scene Evidence, Shanghai Research Institute of Criminal Science and Technology, North Zhongshan No.1 Road 803, Shanghai 200083, China.
E-mail: wbliu1981@163.com or xiaoxuexx588@126.com

Long-range PCR

Amplification of the mtGenome was accomplished by long-range PCR in four separate reactions using the KOD FX Neo PCR Kit (Toyobo, Japan). The primers sequences and amplicon sizes are listed in Supplementary Table S1. Four independent amplification reactions were performed, each in a total volume of 25 μl containing 12.5 μl of 2 × KOD FX Neo Buffer, 2 μl of dNTP Mix, 0.5 μl of KOD FX Neo polymerase, 0.2 μM of each of the forward and reverse primer, and 1–3 ng of DNA templates. Control DNA of 9947A (Thermo Fisher) and 007 (Thermo Fisher) human cell line samples were used as positive controls. Negative controls had no added template DNA. PCR was performed in a GeneAmp PCR System 9700 thermal cycler (Thermo Fisher) using the following conditions: 94 °C for 5 min; 30 cycles of 94 °C for 10 s, 60 °C for 30 s, 68 °C for 5 min; a final extension at 72 °C for 5 min and hold at 4 °C. The PCR products were purified using the Agencourt AMPure XP PCR Purification System (Beckman Coulte, Fullerton, CA, USA) as recommended by the manufacturer. The concentrations of the products were quantified by the Qubit 2.0 Fluorometer with the Qubit dsDNA BR Assay Kit (Thermo Fisher). For each sample, four fragments were normalized and then pooled in equimolar amounts.

Library construction

DNA Libraries were prepared using the Ion Xpress Plus Fragment Library Kit (Thermo Fisher Scientific, Waltham, MA, USA) following the library preparation protocol provided by Ion Torrent (Thermo Fisher). The PCR products were enzymatically fragmented by preparing a mastermix of 35 μl of DNA solution, 5 μl of Ion Shear Plus 10 × reaction buffer and 10 μl of Ion Shear Plus Enzyme Mix II. The reactions were incubated at 37 °C for 10 min to yield fragments ~ 260 bp. Adapters with barcodes were ligated to the sheared fragments using the Ion Xpress Barcode Adapters (Thermo Fisher). A mastermix of 74 μl of DNA solution, 10 μl of 10 × Ligase Buffer, 2 μl of Ion P1 Adapter, 2 μl of Ion Xpress Barcode X (X = number of the used barcode for each sample), 2 μl of dNTP Mix, 2 μl of DNA Ligase and 8 μl of Nick Repair Polymerase were prepared for each sample and incubated at 25 °C for 15 min and 72 °C for 5 min. The ligated fragments were size-selected at ~ 330 bp using the E-Gel SizeSelect Agarose Gel (Thermo Fisher). Amplification was then performed with a mastermix of 100 μl of Platinum PCR SuperMix High Fidelity, 5 μl of Library Amplification Primer Mix and 25 μl of DNA solution. The amplification conditions were 5 min at 95 °C, 5 cycles of 15 s at 95 °C, 15 s at 58 °C and 60 s at 70 °C. Purification of library was performed using the Agencourt AMPure XP PCR Purification System (Beckman Coulter). Libraries were assessed using the Agilent 2100 Bioanalyzer with the Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA) and were quantified using the Qubit 2.0 Fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher).

Template preparation and sequencing

Libraries were normalized and pooled to an equimolar concentration as recommended by the manufacturer. The pooled libraries was used to generate template-positive Ion Sphere Particles containing clonally amplified DNA. Emulsion PCR was performed on the Ion OneTouch 2 instrument with the Ion PGM Template OT2 200 Kit according the template preparation protocol provided by Ion Torrent (Thermo Fisher). Template-positive Ion Sphere Particles were enriched using the Ion OneTouch Enrichment System and subsequently loaded onto Ion 318 v2 Chips (Thermo Fisher). Sequencing was performed using the Ion PGM System with the Ion PGM Sequencing 200 Kit v2 following the recommended protocol (Thermo Fisher).

Data analysis and statistics

The mtGenome sequencing data were analyzed with the Ion Torrent Software Suite (v 4.2.1) using the plug-in variant caller (v 4.2). The output of the variant caller was presented as a list of base differences relative to the revised Cambridge Reference Sequence (NC_012920.1).[17] Concordance testing was performed for HVI and HVII in a subset of samples (n = 24) using STS, as described by King et al.[4] Mitochondrial haplogroups were assigned to haplotypes for each individual using Mito Tool,[18] a web server based on PhyloTree Build 17. The haplogroup assignments were re-evaluated by manual inference with PhyloTree Build 17 to improve the HaploGrep predictions. Variants observed but not known to be associated with a haplogroup, not previously observed in the database of published mtGenomes (http://www.phylotree.org/mtDNA_seqs.htm), or variants expected, but not observed, for each haplotype were validated by visualizing BAM files using Integrative Genomics Viewer.[4,19]

For the 145 Chinese samples, the number of different haplotypes and unique haplotypes were counted for mtGenome and CR/HVI-II, respectively. The random match probabilities, haplotype diversities (HD) and power of discrimination were then calculated as described by Stoneking et al.[10] and Tajima.[20] The RMP was calculated as $RMP = \Sigma pi^2$, where pi is the frequency of the ith haplotype. The HD was calculated as $HD = (1 - RMP) \times (n/(n-1))$, where n is the sample size.[10] The power of discrimination was calculated as $PD = 1 - \Sigma pi^2$ (Power Stats v12, Promega, Madison WI, USA).

Sanger-type sequencing

For all the 145 samples, STS was employed to confirm sequence data from poly-C stretches covering positions 303–316 and 16 184–16 193. Primers for amplification and sequencing were listed in Supplementary Table S1. Sequencing was performed on the 3130*xl* Genetic Analyzer (Thermo Fisher) with the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher). Raw data was analyzed using the Vector NTI Advance11 Software v11 (Thermo Fisher).

RESULTS AND DISCUSSION

The Ion PGM system was demonstrated to be feasible and reliable in forensic research of mtGenome sequencing.[13,15] Concordance testing of mtGenome sequences obtained with PGM and STS was performed by Parson et al., which showed a generally high level of consistency between the two derived data, with the total number of differences below 0.02%.[13] In this study, 145 whole mtGenomes from unrelated Chinese Han population were sequenced using the Ion PGM System. Concordance testing obtained complete consistency between the PGM and STS data, excluding the number of Cs in the homopolymeric stretches around positions 303–316 and 16 184–16 193, similar to the results reported by Parson et al.,[13] Mikkelsen et al.,[14] Seo et al. [15] and McElhoe et al.[16] Thus, sequence data from positions 303–316 and 16 184–16 193 for all the 145 samples were verified using STS in this study.

Coverage analysis

In this study, samples were involved for further analysis with a coverage threshold of 40 × .[4] Sufficient data were obtained to reliably determine the mtGenome sequence, and the coverage was similar among all 145 samples. In average, each sample was represented with 74 827 (±26 819) sequence reads, and the coverage across samples was 668 × (±261). The coverage across the mtGenomes varied, and both forward and reverse strands illustrated a similar trend. The average coverage and s.d. for each base position of the mtGenomes was normalized and shown in Figure 1.[21]

Strand bias

The average coverage of both forward and reverse strands at each nucleotide position was displayed in Figure 1. The strand balance percentage (lower coverage/higher coverage) of all samples at all positions were calculated and displayed in Figure 2. In all 145 samples, 95% of all positions had a strand balance percentage above 40%. Positions of high strand bias (< 40%) coincide with positions of relatively low coverage, which might be attributed to homopolymeric stretches since these regions may be difficult to sequence or data analyze because of technique-related limitations.[4] For example, in all 145 mtGenomes, positions around the poly-C stretches 303–316 and 16 184–16 193, and low areas ~ 3552–3575 and 8605–8625, showed high strand bias as well as low coverage. Potential methods to overcome the problem of bias and coverage variation would improve both quantity and quality of sequencing data in mtGenome research.
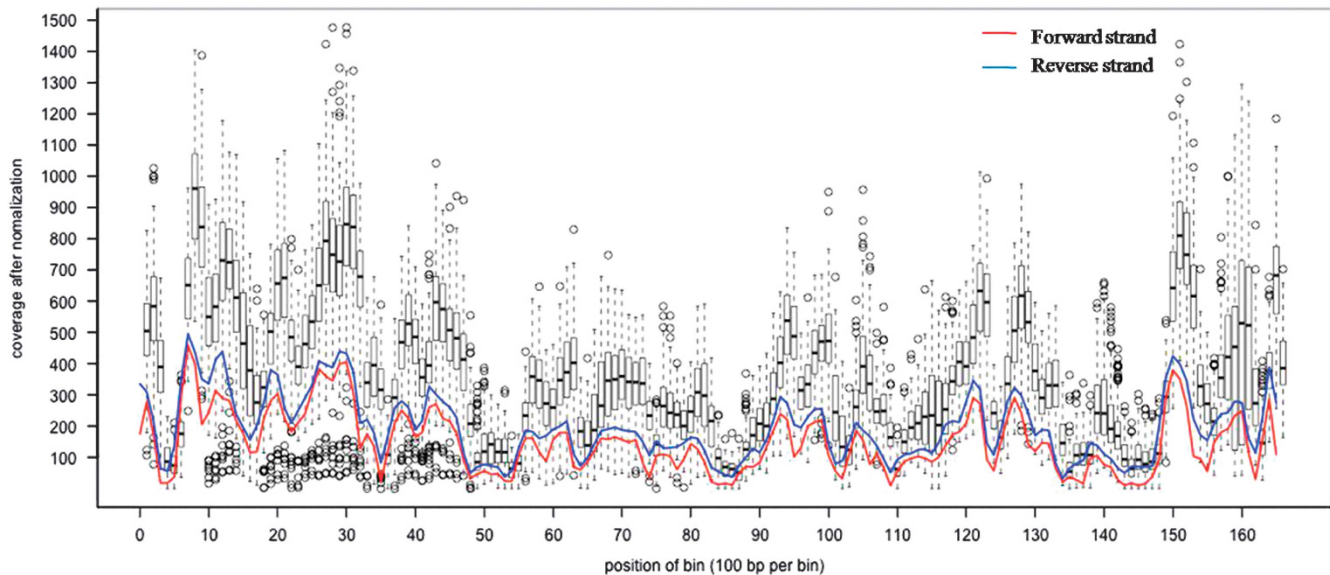
Figure 1 The average coverage of both forward and reverse strands at each nucleotide position after normalization.
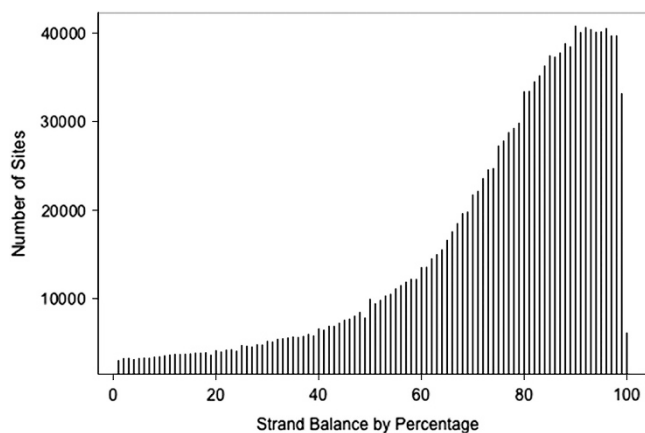
Figure 2 The overall strand bias display for all 145 samples. X axis is the ratio of coverage between the forward and reverse strands at each nucleotide position (lower coverage/higher coverage). Y axis is the number of positions with specific percentages of strands bias.

Table 1 Summary statistics

	HVI/II	CR	mtGenome
Different haplotypes	136	138	145
Unique haplotypes	127	131	145
Power of discrimination	0.9912	0.9916	0.9931
Random match probability (%)	0.88	0.84	0.69
HD	0.9981	0.9985	1

Abbreviations: CR, control region; HD, haplotype diversity; HVI, hypervariable regions I; mtGenome, mitochondrial genome. Summary statistics were calculated for Chinese Han population for several regions historically targeted for forensic typing: HVI/II, the entire CR and the full mtGenome.

Variants calling

In this study, the sequence variants from each sample were presented in Supplementary Table S2, and are available in Sequence Read Archive database in National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/sra/; accession number: SRP072199). From the 145 mtGenomes sequenced, a total of 5625 variants were observed relative to the revised Cambridge Reference Sequence, which were distributed across 782 base positions throughout the mtGenome. As would be expected, the variants were not distributed equally across the entire mtGenome, but were clustered heavily in the hypervariable region. Out of all 5625 variants, 1321 of the polymorphisms (23.48%) were observed in the HVI/II and 1646 of the variants (29.26%) resided in the CR. Thus, 3979 of the variants (70.74%) were located outside of the CR of mtGenome, which demonstrated the potential value of coding region to improve the discrimination power of mtDNA typing in forensic testing.

Haplotype assessment and statistics

For all 145 individual mtGenomes, the HD and summary statistics were compared among HVI/II, CR and mtGenome (Table 1). In total, 145 unique haplotypes were generated from the 145 Chinese mtGenomes sequenced, whereas 127 (87.6%) and 131 (90.3%) unique haplotypes were detected by HVI/II and CR, respectively. These results demonstrated the increase in the number of unique haplotypes detected by CR compared with HVI/II was 3.1%, and in comparison to CR typing, complete mtGenome sequencing increased the number of unique haplotypes by 10.7% for the 145 Chinese population samples. These improvements in haplotype resolution are consistent with two recent studies of mtGenome haplotypes from three U.S. populations.[4,10]

In this study, the random match probabilities, HD and power of discrimination were calculated for the 145 Chinese samples (Table 1). The RMP for mtGenome data were 0.69%, compared with 0.88% for HVI/II and 0.84% for CR data. A similar pattern was observed with HD, with 0.9981, 0.9985 and 1 for HVI/II, CR and mtGenome data, respectively. These results demonstrated the substantially higher discrimination power with entire mtGenome sequences in comparison to the smaller portions of the molecule historically targeted for forensic testing.[4,10]

Haplogroup assignment was performed for all 145 haplotypes using the online software Mito Tool.[18] A total of 108 distinct (nested-) haplogroups were identified. The haplogroup frequencies from the Chinese Han population reported here and from five other East/Southeast Asian population previously reported were presented and compared in Supplementary Table S3.[22–26] For Eastern Chinese Han

population data set, the majority of haplotypes (80.0%) were assigned to haplogroups B, D, F, M* (11.7%, 31.7%, 13.8 and 22.8%, respectively). The majority haplotypes of Japanese and Koreans population were assigned to Haplogroups B, D and M* similarly, whereas haplotypes of Filipino clustered to the Haplogroups B, E, M* and N*, Southwest Chinese Han population were assigned to Haplogroups M*, R, F and B. When compared the forensic statistics, complete mtGenome sequencing data (Chinese Han and Filipino population) showed higher HD (1 and 0.998, respectively) in comparison to sequence data based on CR (Koreans and Myanmar population, with HD of 0.998 and 0.997, respectively). Phylogenetic resolution of the haplogroups showed direct comparison of mtGenome haplogroups to HVI/II or CR-based haplogroup data, with more highly resolved haplogroup categorizations included (Supplementary Table S4).[4,10] In this study, 44 out of 145 samples showed higher resolved haplogroups of mtGenome than CR-based haplogroups, and 11 samples changed clades between CR and mtGenome data. Thirty-six samples changed clades between HVI/II and CR/mtGenome, 35 of which changed macrohaplogroups. These observations can be explained by the fact that haplogroup assignment based on HVI/II cannot distinguish all haplogroups as detailed as a larger segment can. The results demonstrated that haplogroup assignment based on PhyloTree polymorphisms can be incorrect or coarse for those haplogroups defined by (partial) CR variants, for example, HVI/II, which are still mainly used in forensic genetics, similar to the previously reported results.[4,27,28]

CONCLUSION

Previous studies have demonstrated that mtGenome sequencing data generated from Ion PGM System was highly reliable and feasible in forensic research.[13,15] In this study, a total of 145 mtGenomes from Chinese population were sequenced in a high-throughput fashion using the Ion PGM System. Some strand bias and coverage variation was observed but generally did not diminish the ability to assign variant calls. Compared with the limited variants observed in the hypervariable region, 73.73% of the variants resided outside of the CR of mtGenome. Moreover, a great improvement in haplotype resolution was observed between HVI/II and mtGenome, which dramatically increased the discrimination power of mtDNA in forensic research. This MPS approach, with advantages of short sequencing time (< 8 h for sequencing reaction), low running costs and high scalability (three different chip sizes available), may facilitate generation of mtGenome population data to support mtDNA research of forensic, evolutionary anthropology and medical genetics.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. Cell 148, 1145–1159 (2012).

2 Zheng, H. X., Yan, S., Qin, Z. D., Wang, Y., Tan, J. Z., Li, H. et al. Major population expansion of East Asians began before neolithic time: evidence of mtDNA genomes. PLoS ONE 6, e25835 (2011).

3 Ye, F., Samuels, D. C., Clark, T. & Guo, Y. High-throughput sequencing in mitochondrial DNA research. Mitochondrion 17, 157–163 (2014).

4 King, J. L., LaRue, B. L., Novroski, N. M., Stoljarova, M., Seo, S. B., Zeng, X. et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Sci. Int. Genet. 12, 128–135 (2014).

5 Just, R. S., Irwin, J. A. & Parson, W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. Forensic Sci. Int. Genet. 18, 131–139 (2015).

6 Asari, M., Azumi, J., Shimizu, K. & Shiono, H. Differences in tissue distribution of HV2 length heteroplasmy in mitochondrial DNA between mothers and children. Forensic Sci. Int. 175, 155–159 (2008).

7 Bini, C. & Pappalardo, G. mtDNA HVI length heteroplasmic profile in different tissues of maternally related members. Forensic Sci. Int. 152, 35–38 (2005).

8 Irwin, J. A., Parson, W., Coble, M. D. & Just, R. S. mtGenome reference population databases and the future of forensic mtDNA analysis. Forensic Sci. Int. Genet. 5, 222–225 (2011).

9 Parson, W., Gusmão, L., Hares, D. R., Irwin, J. A., Mayr, W. R., Morling, N. et al. DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing. Forensic Sci. Int. Genet. 13, 134–142 (2014).

10 Just, R. S., Scheible, M. K., Fast, S. A., Sturk-Andreaggi, K., Röck, A. W., Bush, J. M. et al. Full mtGenome reference data: development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. Forensic Sci. Int. Genet. 14, 141–155 (2015).

11 Templeton, J. E., Brotherton, P. M., Llamas, B., Soubrier, J., Haak, W., Cooper, A. et al. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. Investig. Genet. 4, 26 (2013).

12 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. 4, 1764 (2013)

13 Parson, W., Strobl, C., Huber, G., Zimmermann, B., Gomes, S. M., Souto, L. et al. Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). Forensic Sci. Int. Genet. 7, 543–549 (2013).

14 Mikkelsen, M., Frank-Hansen, R., Hansen, A. J. & Morling, N. Massively parallel pyrosequencing of the mitochondrial genome with the 454 methodology in forensic genetics. Forensic Sci. Int. Genet. 12, 30–37 (2014).

15 Seo, S. B., Zeng, X., King, J. L., Larue, B. L., Assidi, M., Al-Qahtani, M. H. et al. Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform ion torrent™ PGM™. BMC Genomics 16 (Suppl 1), S4 (2015).

16 McElhoe, J. A., Holland, M. M., Makova, K. D., Su, M. S., Paul, I. M., Baker, C. H. et al. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Sci. Int. Genet. 13, 20–29 (2014).

17 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).

18 Fan, L. & Yao, Y. G. MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. Mitochondrion 11, 351–356 (2011).

19 Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14, 178–192 (2013).

20 Chaitanya, L., VanOven, M., Brauer, S., Zimmermann, B., Huber, G., Xavier, C. et al. High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database. Forensic Sci. Int. Genet. 21, 158–167 (2016).

21 Janevski, A., Varadan, V., Kamalakaran, S., Banerjee, N. & Dimitrova, N. Effective normalization for copy number variation detection from whole genome sequencing. BMC Genomics 13, S16 (2012).

22 Gunnarsdóttir, E. D., Li, M., Bauchet, M., Finstermeier, K. & Stoneking, M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res. 21, 1–11 (2011).

23 Tanaka, M., Cabrera, V. M., González, 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).

24 Lee, H. Y., Yoo, J. E., Park, M. J., Chung, U. & Shin, K. J. Mitochondrial DNA control region sequences in Koreans: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. Int. J. Legal Med. 120, 5–14 (2006).

25 Summerer, M., Horst, J., Erhart, G., Weißensteiner, H., Schönherr, S., Pacher, D. et al. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. BMC Evol. Biol. 28, 14–17 (2014).

26 Wang, D., Su, L. Y., Zhang, A. M., Li, Y. Y., Li,.X. A., Chen, L. L. et al. Mitochondrial DNA copy number, but not haplogroup, confers a genetic susceptibility to leprosy in Han Chinese from Southwest China. PLoS ONE 7, e38848 (2012).

27 Bandelt, H. J. & Salas, A. Current next generation sequencing technology may not meet forensic standards. Forensic Sci. Int. Genet. 6, 143–145 (2012).

28 Röck, A. W., Dür, A., VanOven, M. & Parson, W. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). Forensic Sci. Int. Genet. 7, 601–609 (2013).

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