et al. (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr. Biol. *19*, 1–8.
41. Finnilä, S., Lehtonen, M.S., and Majamaa, K. (2001). Phylogenetic network for European mtDNA. Am. J. Hum. Genet. *68*, 1475–1484.
42. Ingman, M., Kaessmann, H., Pääbo, S., and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. Nature *408*, 708–713.
43. Kumar, S., Padmanabham, P.B., Ravuri, R.R., Uttaravalli, K., Koneru, P., Mukherjee, P.A., Das, B., Kotal, M., Xaviour, D., Saheb, S.Y., et al. (2008). The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. BMC Evol. Biol. *8*, 230.
44. Hartmann, A., Thieme, M., Nanduri, L.K., Stempfl, T., Moehle, C., Kivisild, T., and Oefner, P.J. (2009). Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. Hum. Mutat. *30*, 115–122.

DOI 10.1016/j.ajhg.2009.10.024.

Response to Yao et al.

To the Editor: We are also concerned about errors in GenBank sequences, and that is why we took precautions to evaluate the effects of potential sequence errors.[1] But many of the potential errors reported by Yao et al. are highly subjective. They defined "phantom mutations" as (with exceptions) the exclusive presence of rare transversions in a specific data set. Although it is reasonable to be skeptical of such variations, surely such rare variations do actually occur without being errors. To deal with potential sequence errors, we took the step of doing the analysis twice; once for all reported variations and once for only variations present in more than 0.1% of the sequences. We made the latter choice to filter out sequencing errors, assuming that specific errors would not be repeated in many different sequences. This filtering process did remove 94% of their listed "phantom mutations." As Yao et al. acknowledge, the removal of these rare variations (some of which may be sequencing errors) had little effect on most of our results.

Yao et al. define "missing variants" as those variants expected to be seen in a particular haplogroup but not reported in a sequence assigned to it. The problem with this definition is that it presupposes that we already have a complete picture of mtDNA variation and that all deviations from it are errors. There are many examples of such "missing variants" being true variations. It was once thought that all L- sub-Saharan haplogroups had the substitution at position 16223, but later some lineages were characterized without it (L0d1a, L1c1a1, L2d, L3x2a). Also, the M1- defining substitution at position 16249 is absent in the branch M1a1a.

After the careful data mining of Yao et al., potential errors were found in < 200 of the 5140 sequences. So, ~96% of the sequences deposited in GenBank by the end of August 2008 did pass their extreme quality filter. Yao et al. list many cases in which errors in the original sequences have been acknowledged and corrected by authors but the GenBank sequence has not been updated. GenBank[2] allows the sequence depositor to update that sequence, but it depends on each depositor to carry out this procedure. Identifying these possible sequence errors is complex and is arguably highly subjective. To expect every author of a sequence data-mining project to carry out such a very subjective quality-control step is not reasonable, in our opinion.

Though we may disagree on specifics raised by Yao et al., we do share with them a concern about mtDNA sequence quality. Spirited discussions such as this one have been going on for the past decade. It is time to provide the mtDNA research community with analysis tools that allow them to efficiently check their sequences for potential problems, such as sequencing errors or unusual variations. We tried to go forward in this direction with our paper[1] by providing the mtDNA Gene-Syn software. Fortunately, others are also advancing along the same path.[3–5]

Luísa Pereira[1,2] and David C. Samuels[3]
[1]Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto 4200-465 Porto, Portugal; [2]Faculdade de Medicina da Universidade do Porto, 4200-465 Porto, Portugal; [3]Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN 37232, USA

References

1. Pereira, L., Freitas, F., Fernandes, V., Pereira, J.B., Costa, M.D., Costa, S., Maximo, V., Macaulay, V., Rocha, R., and Samuels, D.C. (2009). The Diversity Present in 5140 Human Mitochondrial Genomes. Am. J. Hum. Genet. *84*, 628–640.
2. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2009). GenBank. Nucleic Acids Res. *37*, D26–D31.
3. Brandon, M.C., Ruiz-Pesini, E., Mishmar, D., Procaccio, V., Lott, M.T., Nguyen, K.C., Spolim, S., Patil, U., Baldi, P., and Wallace, D.C. (2009). MITOMASTER: A Bioinformatics Tool for the Analysis of Mitochondrial DNA Sequences. Hum. Mutat. *30*, 1–6.
4. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–E394.
5. Lee, H.Y., Song, I., Ha, E., Cho, S.B., Yang, W.I., and Shin, K.J. (2008). mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. BMC Bioinformatics *9*, 483.

DOI 10.1016/j.ajhg.2009.10.022.