Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II

Pavel L. Ivanov[1], Mark J. Wadhams[2], Rhonda K. Roby[2], Mitchell M. Holland[2], Victor W. Weedn[2] & Thomas J. Parsons[2]

In 1991, nine sets of skeletal remains were excavated from a mass grave near Yekaterinburg, Russia which were believed to include the Russian Tsar Nicholas II, the Tsarina Alexandra, and three of their daughters[1]. Nuclear DNA testing of the remains verified such a family group, and mitochondrial DNA (mtDNA) sequences of the presumed Tsarina matched a known maternal relative, Prince Philip[2]. mtDNA sequences from bone of the presumed Tsar matched two living maternal relatives except at a single position, where the bone sample had a mixture of matching (T) and mismatching (C) bases. Cloning experiments indicated that this mixture was due to heteroplasmy within the Tsar; nevertheless, the 'mismatch' fueled a lingering controversy concerning the authenticity of these remains. As a result, the official final report on the fate of the last Russian Royals has been postponed by Russian authorities pending additional, convincing DNA evidence. At the request of the Russian Federation government, we analysed the skeletal remains of the Tsar's brother Georgij Romanov, in order to gain further insight into the occurrence and segregation of heteroplasmic mtDNA variants in the Tsar's maternal lineage. The mtDNA sequence of Georgij Romanov matched that of the putative Tsar, and was heteroplasmic at the same position. This confirms heteroplasmy in the Tsar's lineage, and is powerful evidence supporting the identification of Tsar Nicholas II. The rapid intergenerational shift from heteroplasmy to homoplasmy, and the different heteroplasmic ratios in the brothers, is consistent with a 'bottleneck' mechanism of mtDNA segregation.

The remains of the Grand Duke of Russia Georgij Romanov, who died in 1899, were recently exhumed from St. Peter and Paul Cathedral in St. Petersburg, with documentation by representatives of the Russian Federation government and the Russian Orthodox Church. Analyses yielding mtDNA displacement loop (D-loop) sequences were performed on femur and tibia samples (Fig. 1). Subsequently, sequences were determined from femur and tibia samples of the putative Tsar Nicholas, and finally from a blood sample from Countess Xenia Cheremeteff-Sfiri, a living maternal relative of the Tsar. The sequence analysis of the putative Tsar replicated previous work[2], to confirm the results in an additional independent laboratory — it has been alleged that the apparent heteroplasmy in the presumed Tsar was due to contamination[3]. The same bones (left femur and tibia) from both sets of remains were analysed to eliminate the possibility that the bones could have been from the same individual.

DNA was extracted from bone specimens using organic and silica-based methods, and the hypervariable regions (HV1, HV2) of the mtDNA D-loop were amplified with overlapping polymerase chain reaction primer sets[4–6]. Approximately 340 base pairs from HV1 and 270 base pairs from HV2 were determined from each specimen by automated direct sequencing. The sequence of Xenia Cheremeteff-Sfiri was the same as previously determined, and had substitutions (six transitions and one insertion) at positions 73, 263, 315.1 (insertion), 16126, 16169, 16294, and 16296 relative to the standard reference sequence[7]. The sequence from two bones of Georgij Romanov matched that of Countess Xenia except at position 16169, where a mixture of both C and T was observed (Fig 2). The two bones of the presumed Tsar gave sequences that matched those of Georgij Romanov, including the occurrence of C/T heteroplasmy at position 16169. This confirmed, now in a third independent laboratory, the previously reported sequence. Results from multiple sequencing reactions for both sets of skeletal remains were free of any significant underlying peaks at other base positions. The ratios of C to T at 16169 were consistent for each individual, but differed between the individuals: on the light strand, Georgij Romanov had an average of 62% T and 38% C, while the putative Tsar had 28% T and 72% C.
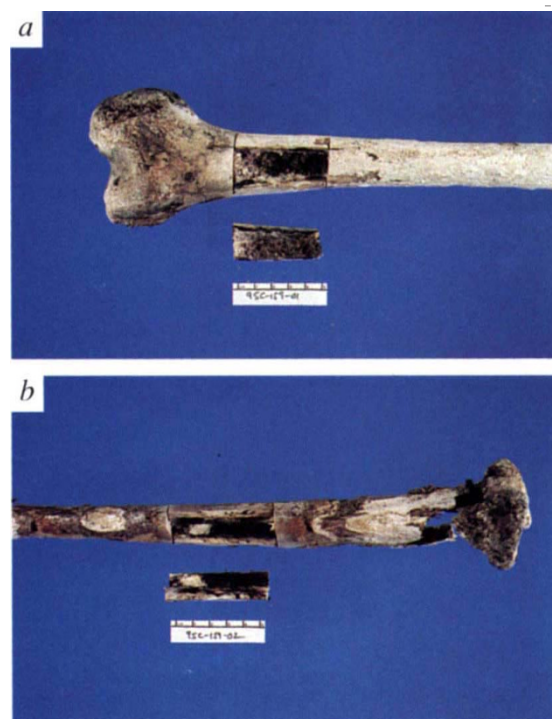
Fig. 1 Sample of bone removed from exhumed remains of Georgij Romanov. *a*, Femur; *b*, tibia.

[1] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov St., 117984 Moscow, Russia
[2] Armed Forces DNA Identification Laboratory, Office of the Armed Forces Medical Examiner, Armed Forces Institute of Pathology, 1413 Research Blvd., Rockville, Maryland 20850, USA

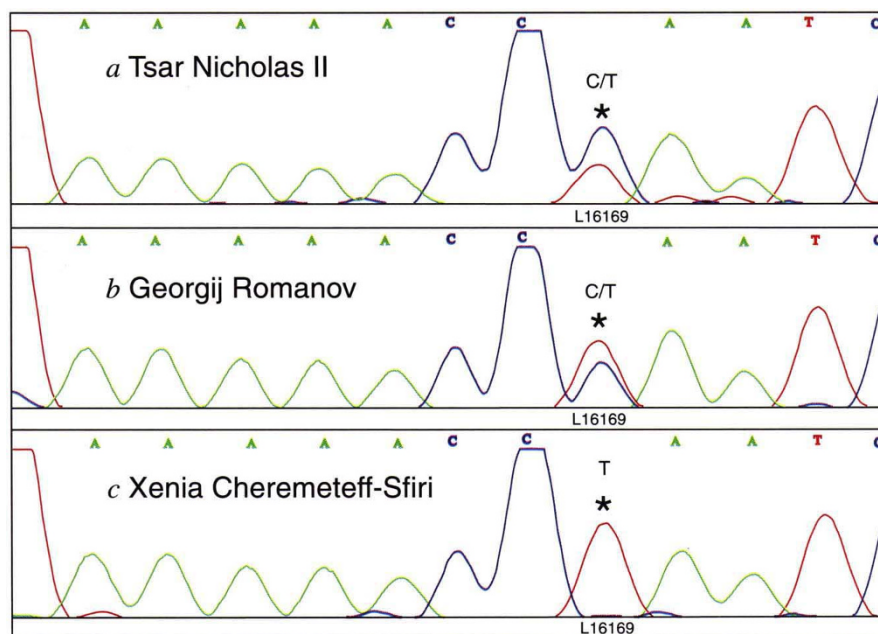Correspondence should be addressed to T.J.P. or P.L.I.

Fig. 2 Automated sequence chromatographs comparing mtDNA sequences at position 16169. *a*, Sequence from bones of putative Tsar Nicholas II, showing heteroplasmy with cytosine predominating thymine; *b*, sequence from bones of Grand Duke Georgij Romanov, showing heteroplasmy with thymine predominating cytosine; *c*, sequence from Countess Xenia Cheremeteff-Sfiri, homoplasmic for thymine.

The mtDNA sequence from Georgij Romanov confirms the presence of heteroplasmy in a known, close maternal relative of the Tsar. However, this heteroplasmy is not apparent in two living members of this matrilineage, Countess Cheremeteff-Sfiri and the Duke of Fife. Hence, the previous discrepancy between the putative Tsar and these relatives is due to heteroplasmy that was apparently passed from the Tsar's mother Maria Feodorovna to her sons Georgij and Nicholas, but which segregated to homoplasmy in the course of more extended intergenerational transmission (Fig. 3).

Current information on the mechanism of mtDNA transmission in humans is limited and partly contra-dictory[8]. A number of lines of evidence suggest that a very small proportion of a female's mtDNA pool is passed between generations. Such a bottleneck in mtDNA transmission would account for the rarity with which heteroplasmy is observed[9], despite a high mtDNA mutation rate. A bottleneck hypothesis also fits the patterns of rapid, differential segregation of heteroplasmic variants seen in bovines[10,11], and in some human pedigrees involving pathogenic mtDNA mutations[12,13]. However, other studies indicate that heteroplasmic variants do not always segregate rapidly between generations[14,15]. Processes of mtDNA segregation may thus be complex, producing different patterns in different species or circumstances (for example, inter- versus intra-organellar heteroplasmy)[8,16,17]. In this study, the different heteroplasmic ratios in the Romanov siblings, and the progression to apparent homoplasmy within four generational events, indicates a rapid segregation of heteroplasmic variants. This is consistent with a simple bottleneck mechanism. It would be of great interest to track the segregation of position 16169 variants in additional members of this mtDNA lineage: under a bottleneck mechanism, other closely related members would be expected to show variable mixtures of C and T, or may be homoplasmic for either base.

The significance of the mtDNA matches in verifying the identity of the Romanov remains can be evaluated following a Bayesian approach[2]. Considering only a sequence match between the putative Tsar and Georgij Romanov gives a likelihood ratio of 150: the observed data are 150 times more likely if the remains are those of Tsar Nicholas II, than if the remains were those of another, unrelated individual. This does not take into account the remarkable circumstance that the ques-
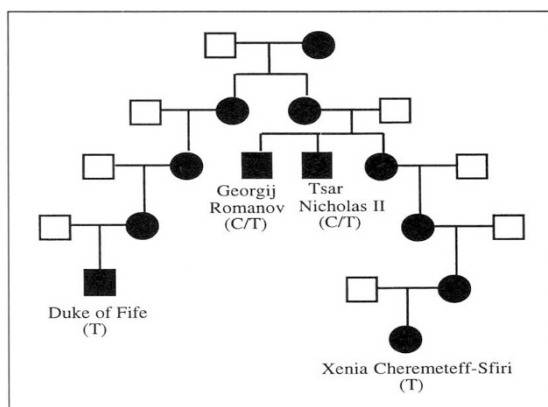

Fig. 3 Lineage of Tsar Nicholas II, indicating individuals whose mtDNA sequences were determined (sequence at position 16169 is listed beneath name). Black symbols indicate the Hessian maternal lineage. D-loop sequences from the Duke of Fife were determined in a previous study, and match those of Countess Xenia Cheremeteff-Sfiri[2].

tioned remains and Georgij Romanov are heteroplasmic at the same position. To consider this, we have estimated the chance of observing heteroplasmy at the same position in random, unrelated individuals to be 2.4×10^{-4}, resulting in a likelihood ratio of 3.8×10^5. When combined with the mtDNA match of the putative Tsarina to Prince Philip, the total mtDNA data are 1.3×10^8 times more likely if the remains are those of the Romanovs, than if they were an unrelated family. Given that anthropological and circumstantial evidence was also considered conclusive for the identification of the Romanovs[18], there now appears to be no reasonable scientific objection to accepting the authenticity of the remains.

Methods

DNA extraction from skeletal material. Bone powder was incubated with proteinase K, phenol-chloroform extracted, and concentrated as described[4]. Silica-based DNA extraction was performed with modifications of described methods[5,6]: all solutions were aliquotted in single-use amounts to 1.7 ml eppendorf tubes and UV-irradiated with 3 j/cm² at 254 nm. Sanded bone was soaked for two min in a 0.5% sodium hypochlorite solution to ensure surface decontamination, followed by extensive water and ethanol rinses. Approximately 0.5 g of bone was ground and incubated in 1 ml guanidinium thiocyanate (GuSCN) extraction buffer (4.7 M GuSCN, 20 mM EDTA, 46 mM Tris pH 8.0, 1.2% Triton X-100) for ~16 h at 56 °C. Insoluble material was pelleted and discarded, and the supernatant mixed with 50 μl silica suspension (Sigma S-5631, prepared as described[5]). After incubation at room temperature for 20 min, the silica matrix was pelleted and washed twice with washing buffer (GuSCN extraction buffer lacking Triton X-100 and EDTA) and 70% ethanol. The final pellet was air dried, and the DNA eluted in 110 μl TE with incubation at 56 °C for 20 min. The silica matrix was pelleted again, and the supernatant, containing DNA, transferred to a fresh tube. DNA from the blood reference sample was extracted using the Chelex® method[19].

Amplification. The two D-loop hypervariable regions were amplified with two overlapping primer sets each, as described[4]. Some reactions employed hot start PCR using HotStart® 50 wax microtubes (Molecular Bio-Products). Primer set I: F15990 (5'–TTAACTCCACCATTAGCACC–3') and R16239 (5'–TGG-CTTTGGAGTTGCAGTTG–3'), Primer set II: F16163 (5'–TGACCACCTGTAGTACATAA–3') and R16391 (5'–GAG-GATGGTGGTCAAGGGAC–3'), Primer set III: F34 (5'–CACC-CTATTAACCACTCACG–3') and R251 (5'–TGGAAAGTG-GCTGTGCAGAC–3'), Primer set IV: F174 (5'–TATTTATCG-CACCTACGTTC–3') and R465 (5'–TGAGATTAGTAG-TATGGGAG–3').

Sequence analysis. Sequence was determined from PCR products by flourescently-labelled cycle sequencing with an ABI 373A automated DNA sequencer (Applied Biosystems Division of Perkin-Elmer), as described[4].

Laboratory facilities and contamination avoidance. Work was performed in laboratories designed exclusively for mtDNA analysis from aged bone. Safeguards against cross-over contamination include the following: rigorous physical separation of pre- and post-PCR laboratories, including separate ventilation systems with controlled air pressure, and high-air-exchange laboratory anterooms with adhesive floor mats; set-up of amplifications in sterile biological hoods with UV lamps; complete decontamination with 0.5% sodium hypochlorite of work surfaces (daily) and floors (weekly); sterile manipulations performed wearing clean gowns and single-use disposable face masks, sleeves, and gloves. Processing of different samples was

separated in time to avoid the possibility of cross-contamination, with analysis of Georgij Romanov completed prior to work on the putative Tsar, and analysis of the blood reference performed after all bone testing, in a separate laboratory. For all extractions and amplifications a 'reagent blank' negative control, lacking any added sample, was processed in parallel with the samples, as well as amplification negative controls run without addition of extract. For the first organic extraction of Georgij Romanov, faint amplification products were visible in several reagent blank controls; when sequenced, these clearly differed from the sequence derived from the bone extracts. However, extractions were repeated for Georgij Romanov using a silica extraction method, and controls were completely negative throughout. Sequence from the silica bone extracts matched the sequence from the organic bone extracts.

Data interpretation. Bayesian inference was used to assess the significance of the mtDNA match between the questioned male remains (the putative Tsar) and Georgij Romanov. We consider two events in this analysis: E1, sequence match between Georgij Romanov and the remains of the putative Tsar; and E2, the co-occurrence of heteroplasmy at position 16169 in Georgij Romanov and the putative Tsar. The likelihood ratio is the probability of E1 and E2 if the remains are those of Tsar Nicholas (hypothesis R), divided by the probability of E1 and E2 if the remains are unrelated (hypothesis R'). If we assume that E1 and E2 are independent, the likelihood ratio can be written:

$$LR = \frac{p(E1|R)}{p(E1|R')} \times \frac{p(E2|R)}{p(E2|R')} \qquad (Eq.1)$$

The assumption of independence between E1 and E2 means that the chances of heteroplasmy at a given position are not dependent on D-loop polymorphisms at other sites; although there is no specific information validating this assumption, it seems quite reasonable to a first approximation.

The first term in Eq.1 can be evaluated in a manner directly analogous to the 'upper bound' value previously calculated[2]: $p(E1|R) = e^{-g\mu} = 0.96$, where g=2 generational events separating the tested individuals, and μ is the estimated mutation rate of the D-loop =1/50. (Varying estimates of the D-loop mutation rate have been inferred from evolutionary analyses, with a recently reported value of 1 per 300 generations[20]. However, studies at the U.S. Armed Forces DNA Identification Laboratory (AFDIL) involving many sequence comparisons between mothers and offspring indicate an empirically-determined mutation rate closer to 1 per 50 generations (manuscript in preparation) — this value is adopted for our current considerations.)

$p(E1|R')$ is the probability that the mtDNA sequences of two unrelated individuals match each other, and is simply the population frequency of the haplotype in question. Ignoring position 16169, the sequence from the remains matches that of one individual in the combined database of random mtDNA sequences, from the British Home Office ($n = 100$, British whites) and the AFDIL ($n = 207$, European Americans). Hence the estimated haplotype frequency is $2/308 = 0.0065 = p(E1|R')$. Plugging these values into the first term of Eq.1 gives:

$$\frac{p(E1|R)}{p(E1|R')} = 150$$

Hence, based on sequence matching alone, ignoring heteroplasmy, the data are 150 times more likely if the remains are those of Tsar Nicholas than if they are not.

The second term of Eq.1 is the probability of sharing heteroplasmy at position 16169 if the male remains are from a sibling of Georgij Romanov, divided by that probability if they are unrelated. To evaluate this, we must estimate the rate with which heteroplasmy arises, and its intergenerational persistence. As above, the mutation rate of the human D-loop is estimated at one complete substitution every 50 generations. Each substitution is presumed to occur via a heteroplasmic transformation spanning one or a small number of generations. In the case of the Tsar, apparent heteroplasmy in the Tsar's mother segregated to homoplasmy within four generations. Adopting

four generations as a rough estimate of the window within which heteroplasmic intermediates may be observed, the chance that a randomly sampled individual is heteroplasmic anywhere in the D-loop may be estimated as $4/50 = 0.08$. This is certainly much higher than the incidence of documented D-loop heteroplasmies (resulting in a figure that is, for the purposes of our evaluation, conservative), but current direct sequencing methods may overlook unbalanced heteroplasmic mixtures with some frequency, failing to distinguish them from sequence 'background'[21].

When we consider that the heteroplasmy occurs at the same position, there are 610 positions in the hypervariable regions, of which 332 are polymorphic in our database and therefore presumably free to mutate to a heteroplasmic condition. (We note that only five of 715 individuals in the total AFDIL human database have a polymorphism (T) relative to the reference sequence (C) at position 16169, indicating that it is not a hot spot for mutation.) Thus, given that a reference individual has a heteroplasmic mutation at a particular position, we can roughly estimate the chance of randomly sampling another individual with heteroplasmy at that position as $p(E2|R') = (4/50) \times (1/332) = 2.4 \times 10^{-4}$.

The probability that the male remains share heteroplasmy with Georgij Romanov if the two are siblings (separated by two generational events) is $e^{-2\beta}$, where ß is the rate of fixation of homoplasmy. Above, we estimated $\beta = 1/4$ (time to homoplasmic fixation = four generations), so $p(E2|R) = 0.61$. Plugging the above values into the second term of Eq.1 gives

$$\frac{p(E2|R)}{p(E2|R')} = 2.5 \times 10^3$$

Using the values derived above, Eq.1 gives a likelihood ratio of 3.8×10^5. This result can be multiplied by the likelihood ratio derived from the match between the putative Tsarina and Prince Philip (E3), separated by four generational events. Since this sequence is unique in our database, we lack a suitable estimate of the haplotype frequency: to estimate random match frequency we instead use the observed number of matches in all pair-wise comparisons of sequences in the Caucasian database,

$$p(E1|R') = \frac{133 \text{ matches}}{46,971 \text{ comparisons}} = 0.0028$$

As before,

$$\frac{p(E3|R)}{p(E3|R')} = \frac{(e^{-4\mu})}{(0.0028)} = 330$$

Hence, the total DNA evidence is $(330) \times (3.8 \times 10^5) = 1.3 \times 10^8$ times more likely if the remains are indeed those of the Romanovs.

1. Massie, R.K. *The Romanovs: the Final Chapter.* (Random House, New York, 1995).
2. Gill, P. *et al.* Identification of the remains of the Romanov family by DNA analysis. *Nature Genet.* **6**, 130–135 (1994).
3. Maples, W.R. Affidavit, Circuit Court for the city of Charlottesville, Case No. 8021, November 12, 1993.
4. Holland, M.M. *et al.* Mitochondrial DNA sequence analysis of human remains. *Crime Lab. Digest* **22**, 3–8 (1995).
5. Boom R. *et al.* Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* **28**, 495–503 (1990).
6. Hoss, M. & Pääbo, S. DNA extraction from Pleistocene bones by a silica-based purification method. *Nucl. Acids Res.* **16**, 3913–3914 (1993).
7. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1991).
8. Poulton, J. Transmission of mtDNA: cracks in the bottleneck. *Am. J. Hum. Genet.* **57**, 224–226 (1995).
9. Monnat, R.J. & Loeb, L.A. Nucleotide sequence preservation of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **82**, 2895–2899 (1985).
10. Laipis, P., Hauswirth, W., O'Brian, T., & Michaels, G. Unequal partitioning of bovine mitochondrial genotypes among siblings. *Proc. Natl. Acad. Sci. USA* **85**, 8107–8110 (1988).
11. Koehler, C.M. *et al.* Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics* **129**, 247–255 (1991).
12. Bolhuis, P.A. *et al.* Rapid shift in genotype of human mitochondrial DNA in a family with Leber's hereditary optic neuropathy. *Biochem. Biophys.*
Res. Comm. **170**, 994–997 (1990).
13. Wallace, D.C. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am. J. Hum. Genet.* **57**, 201–223 (1995).
14. Howell, N. *et al.* Mitochondrial gene segregation in mammals: is the bottleneck always narrow? *Hum. Genet.* **90**, 117–120 (1992).
15. Bendall, K.E. & Sykes, B.C. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am. J. Hum. Genet.* **57**, 248–256 (1995).
16. Vilkki, J., Savonataus, M. & Singh, G. Segregation of mitochondrial genomes in a heteroplasmic lineage with Leber hereditary optic neuroretinopathy. *Am. J. Hum. Genet.* **47**, 95–100 (1990).
17. Solignac, M., Monnerot, M. & Mounolou, J.C. Mitochondrial DNA heteroplasmy in *Drosophila mauritiana. Proc. Natl. Acad. Sci. USA* **80**, 6942–6946 (1983).
18. Case Report of Office of Procurator General of Russian Federation, No. 18/123666–93, 1993.
19. Walsh, P.S., Metzger, D. A. & Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10**, 506–513 (1991).
20. Stoneking, M., Sherry, S.T., Redd, A.J. & Vigilant, L. New approaches to dating suggest a recent age for the human mtDNA ancestor. *Phil. Trans. R. Soc. Lond.* B **337**, 167–175 (1992).
21. Comas, D., Pääbo, S. & Bertranpetit, J. Heteroplasmy in the control region of human mitochondrial DNA. *Genome Res.* **5**, 89–90 (1995).