Identification of the remains of the Romanov family by DNA analysis

Peter Gill[1], Pavel L. Ivanov[2], Colin Kimpton[1], Romelle Piercy[1], Nicola Benson[1], Gillian Tully[1], Ian Evett[1], Erika Hagelberg[3] & Kevin Sullivan[1]

Nine skeletons found in a shallow grave in Ekaterinburg, Russia, in July 1991, were tentatively identified by Russian forensic authorities as the remains of the last Tsar, Tsarina, three of their five children, the Royal Physician and three servants. We have performed DNA based sex testing and short tandem repeat (STR) analysis and confirm that a family group was present in the grave. Analysis of mitochondrial (mt) DNA reveals an exact sequence match between the putative Tsarina and the three children with a living maternal relative. Amplified mtDNA extracted from the remains of the putative Tsar has been cloned to demonstrate heteroplasmy at a single base within the mtDNA control region. One of these sequences matches two living maternal relatives of the Tsar. We conclude that the DNA evidence supports the hypothesis that the remains are those of the Romanov family.

[1]*Central Research and Support Establishment, Forensic Science Service, Aldermaston, Reading, Berkshire, RG7 4PN, UK*
[2]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 117984, Moscow, Russia*
[3]*University of Cambridge, Department of Biological Anthropology, Downing Street, Cambridge CB2 3DZ, UK*

P.G., C.K., R.P., N.B., G.T., I.E. & K.S. present address: Forensic Science Service, Priory House, Gooch Street North, Birmingham B5 6QQ, UK

The official records of the fate of the Russian royal family are scant, but it is known that prior to their deaths they were held prisoner in Ipatiev House at Ekaterinburg in the Urals of Central Russia. It is believed that shortly after the night of July 16th, 1918, Tsar Nicholas II, his wife, Tsarina Alexandra, their four daughters, Olga, Tatyana, Maria and Anastasia and their only son, Alexei, were herded into the cellar together with three of their servants and the family doctor, Eugeny Botkin. They were shot by the Bolshevik firing squad, although a number of the victims were allegedly stabbed to death when gunfire failed to kill them. The bodies were stripped and placed onto a truck with the intention of disposing them down a mine shaft. However, the truck developed a mechanical fault during the journey; the Bolsheviks placed the victims into a hastily dug pit in a road and to hinder identification, sulphuric acid was thrown into the open grave. After the bodies were covered, a truck was driven backwards and forwards over the site to flatten the area.

This account is supported by forensic evidence collected in 1918–19 by Nikolai Sokolov (a White Russian monarchist investigator) whose seven-volume dossier has become the basis for the historical version of the fate of the Romanovs. However, the version of events described above has never been positively verified.

After referring to archival materials and photographs, which gave an indication of a burial site, two Russian amateur historian investigators, Gely Ryabov and Alexander Avdonin, announced that they had discovered a communal grave approximately 20 miles from Ekaterinburg. Consequently, the Russian government authorised an official investigation coordinated by the Chief Forensic Medical Examiner of the Russian Federation. The grave consisted of a shallow pit (less than 1 m deep) and contained human skeletal remains. Many of the bones were badly damaged, but it was possible to identify nine corpses. All of the skeletons showed evidence of violence and mistreatment before death. Some of the skulls had bullet wounds; bayonet marks were also found. Facial areas of the skulls were destroyed, rendering classical facial identification techniques difficult.

Extensive work was carried out by Russian forensic experts, involving computer aided facial reconstruction, odontology, age estimation and sexing of the remains. Significantly, some of the bodies had gold, platinum and porcelain dental work, which indicated that at least some of the remains were aristocrats. Tests by Russian scientists tentatively indicated that the pit contained the remains of the Tsar, Tsarina and three of the five children. It was concluded that two bodies were missing, namely the Tsarevich, Alexei, and one of the daughters. In 1992, we were approached by the Chief Forensic Medical Examiner of the Russian federation with a request to initiate an official Anglo-Russian investigation to verify the authenticity of the remains using DNA based techniques.

Identification of human remains by DNA analysis has proven to be a powerful tool in forensic investigations. The amplification of chromosomal short tandem repeat (STR) loci has been shown to have considerable potential for individual identification[1,2] and has been successfully employed in the typing of bone material up to 15 years old[3,4]. Analysis of mtDNA is even more powerful because it is usually present at a high copy number in cells[5] and is more likely to survive for prolonged periods compared to chromosomal DNA. Accordingly, mtDNA analysis has been successfully applied to human remains such as 7,000

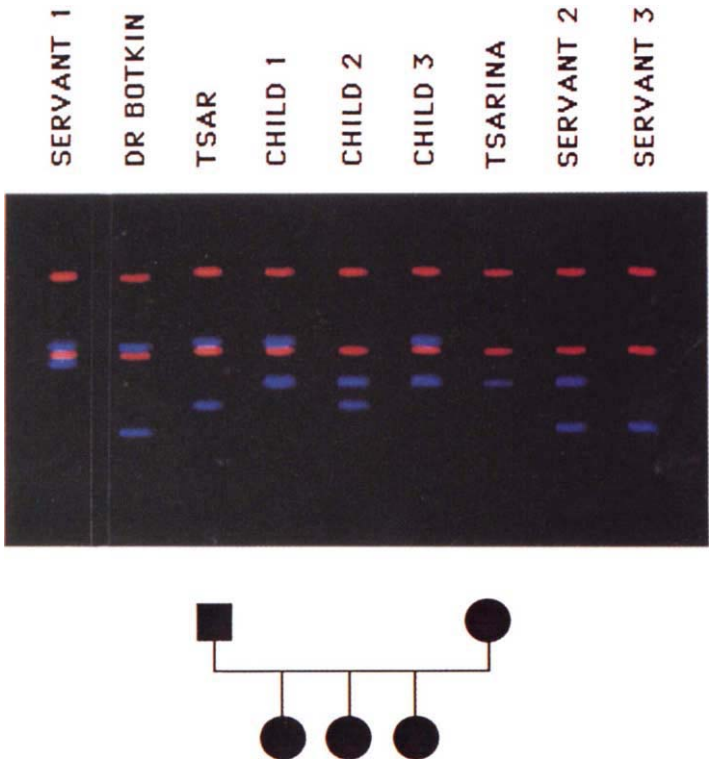

Fig. 1 Analysis of skeletal remains at STR locus HUMTHO1, with the postulated relationship of five samples shown below. Blue bands are amplified HUMTHO1 alleles and red bands are internal electrophoresis size markers.

year-old human brain tissue[6] and 5,500 year-old bone[7]. It can also be used to trace maternal inheritance and is the only option to test relatedness if there are several generations between ancestor and living descendant. To establish paternity from 75-year-old bones, the best approach is to analyse short tandem repeats (STRs).

We followed a dual strategy using STR analysis to establish whether a family group was present and mtDNA analysis to determine relatedness with known maternally related descendants of the Romanov family.

Chromosomal DNA analysis

Sex testing. The sex of the bones was determined by amplification of a portion of the X-Y homologous gene, amelogenin[8,9], which provides a robust method for typing samples of a very degraded nature: X and Y-specific products of 106 and 112 basepairs (bp), respectively were generated from a single primer pair. DNA extracts from all nine skeletons were readily typed and the results confirmed conclusions drawn from physical examination of the bones regarding their sex. In total, four male and five female bodies were identified using this method.

Short tandem repeats. Five tetrameric STR loci HUMTH01[10], HUMVWA31[11], HUMF13A1[12] HUMFES/FPS[13] and HUMACTPB2[14], were amplified from each of the nine skeletons. The amount of template DNA added to each amplification reaction was between 20–40 pg which corresponded to 3–7 diploid genome copies. At such low copy number stochastic variation can result in unequal amplification of individual alleles at a given locus. Indeed, we have observed that at genomic DNA levels of 50 pg or less, individual alleles from known heterozygotes may completely fail to be amplified. Therefore, in order to avoid mistyping, all samples were analysed a minimum of four times and any which were apparent homozygotes, a minimum of six times.

All skeletal samples yielded amplification products corresponding to expected allele sizes. Typical results are shown in Fig. 1. Amplification of HUMF13A1 and HUMFES/FPS loci occasionally resulted in the production of additional artifactual bands. Such bands tended to be of constant size and occurred outside the expected allele size range for each of these loci, hence the identification of the true allele bands was not affected. STR genotypes of the nine skeletal samples are given in Table 1. The observed genotypes of skeletons 3–7 exhibit patterns which would be expected in a family group where 4 and 7 were the parents of children 3, 5 and 6. All other adults were excluded as possible parents.

If the remains are those of the Romanovs then the STR and sex test data indicate that one of the princesses and Tsarevitch Alexei were missing from the grave. This would support some historical accounts which indicate that two bodies were either burned or buried separately. Alternatively, two individuals may have survived the massacre.

MtDNA analysis

Both DNA strands of both hypervariable mtDNA regions were sequenced for all samples, with the exception of skeleton 9 (a probable servant) for which sequence from nt 16216 to 16360 was determined by manual sequencing from one strand only. Duplicate extractions and sequence determinations also were performed on the 'family group'.

No sequence differences were observed between duplicate samples from the same individual. In general, 380 nucleotides in the first hypervariable region (bases 16020 to 16400) and 360 nucleotides in the second hypervariable region (bases 48 to 408) were determined from the amplified bone DNA extracts. The quality of the sequence was generally comparable to that produced from the fresh blood samples (Fig. 2). Pairwise comparisons from the nine bone samples indicated that six different sequences were present in the group which varied on average by six nucleotides and identical

Table 1 STR genotypes[a] for the nine skeletons

Skeleton	HUMVWA/31	HUMTH01	HUMF13A1	HUMFES/FPS	HUMACTBP2
1 (servant)	14,20	9,10	6,16	10,11	ND
2 (doctor)	17,17	6,10	5,7	10,11	11,30
3 (child)	15,16	8,10	5,7	12,13	11,32
4 (Tsar)	15,16	7,10	7,7	12,12	11,32
5 (child)	15,16	7,8	5,7	12,13	11,36
6 (child)	15,16	8,10	3,7	12,13	32,36
7 (Tsarina)	15,16	8,8	3,5	12,13	32,36
8 (servant)	15,17	6,9	5,7	8,10	ND
9 (servant)	16,17	6,6	6,7	11,12	ND

[a]Allele designation for all loci except HUMACTBP2 is based on the number of repeat units (determined by sequencing of specific alleles — data not shown). The allele designation for HUMACTBP2 is based on an arbitrary scale identical to that of Kimpton *et al.*[2].

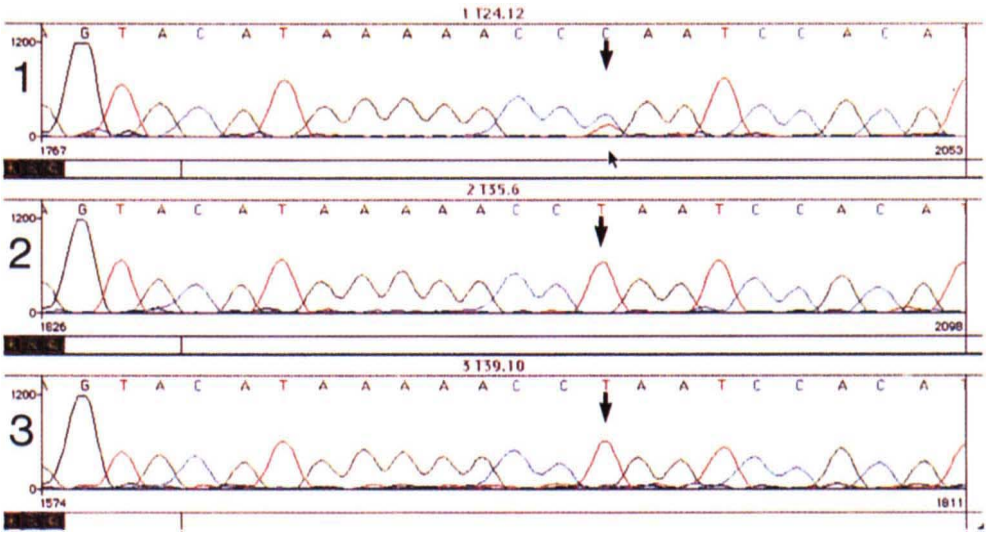

Fig. 2 Comparison of mtDNA sequence (nt 16155–16178) of the putative Tsar with two maternal relatives. Sequence from bone extract of putative Tsar (1) displays heteroplasmy at position 16169 (arrowed) in contrast with sequence from blood samples of relatives (2 and 3).

sequences were generated from the putative Tsarina and three children (Table 2). HRH Prince Philip, the Duke of Edinburgh, is a grand-nephew of unbroken maternal descent from Tsarina Alexandra (Fig. 3). He provided a sample of blood for comparison purposes which enabled us to confirm the sibling status of the children and the identification of the mother; all of the mtDNA sequences were the same (Table 2).

The mtDNA sequence of the putative Tsar was compared with two relatives of unbroken maternal descent from the Tsar's maternal grandmother (Fig. 3). The two relatives had the same sequence as the putative Tsar, with the exception of a single nucleotide (nt) at position 16169. Closer examination of the latter sequence (Fig. 2) revealed heteroplasmy at this position comprising nts C and T present at a ratio of approximately 3.4:1, estimated from the peak areas (corrected for the relative signal intensities of C and T residues proximal to nt 16169). Heteroplasmy was observed in duplicate extracts from skeleton 4, and this was confirmed by sequencing individual recombinant clones which were constructed by ligating polymerase chain reaction (PCR) products into plasmid pUC18, followed by transformation into *E. coli*. Upon sequencing, a total of 28 clones were found to have C at nt 16169, while 7 had T (giving a C:T ratio of 4:1).

At present, no data are available on the rate at which mutations within the non-coding region of human mtDNA become fixed within a given maternal lineage, or on the frequency of heteroplasmy in this region. A difficulty with measuring the latter is the relatively high background signal generated by many sequencing techniques which may mask the presence of all but the most pronounced heterogeneities.

MtDNA sequences of the mother (Dagmar, daughter of Louise of Hesse Cassel and Christian IX, King of Denmark) and grandmother (Louise of Hesse Cassel) of Tsar Nicholas II are not available, so it is not known whether either woman exhibited this sequence heteroplasmy. Furthermore, evidence is conflicting regarding the rate at which heteroplasmy converts to homoplasmy in mtDNA lineages: segregation of a

Table 2 Summary of mtDNA differences compared to the Anderson[21] reference sequence

Origin of sample	DNA source	Length sequenced (bp)	Positions within hypervariable regions (HVR) of mitochondrial DNA																		
			HVR 1												HVR 2						
			16111	16126	16169	16261	16264	16278	16293	16294	16296	16304	16311	16357	73	146	195	263	309.1	309.2	315.1
			C	T	C	C	C	C	A	C	C	T	T	T	A	T	T	A	*	*	*
Servant 1 (?)	Femur skeleton 1	760	.	.	.	.	.	.	.	.	.	C	.	.	–	.	.	G	C	C	C
Servant 2 (?)	Femur skeleton 8	742	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	C
Servant 3 (?)	Femur skeleton 9	650	.	.	.	T	.	T	G	.	.	.	C	.	–	.	C	G	.	.	C
Royal Physician Dr Botkin (?)	Femur skeleton 2	736	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	G	C	C	C
Daughter 1 of Tsar/Tsarina (?)	Femur skeleton 3	755	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Daughter 2 of Tsar/Tsarina (?)	Femur skeleton 5	634	T	.	.	.	.	.	.	.	.	.	.	C	.	.	–	G	.	.	C
Daughter 3 of Tsar/Tsarina (?)	Femur skeleton 6	760	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Tsarina Alexandra (?)	Femur skeleton 7	744	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Duke of Edinburgh (Grand nephew of Tsarina)	Blood sample	760	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Tsar Nicholas II (?)	Femur skeleton 4	782	.	C	Y	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C
Gt. Gt. grandson of Louise of Hesse-Cassel	Blood sample	781	.	C	T	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C
Gt. Gt. Gt. granddaughter of Louise of Hesse-Cassel	Blood sample	782	.	C	T	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C

., Sequence unchanged from reference sequence; –, No nucleotide assignment; *, Nucleotide absent from reference sequence; Y, C/T heteroplasmy.

particular heteroplasmic silent polymorphism in a human mtDNA coding region persisted for at least three generations[15], while a *de novo* mutation in the mtDNA non-coding region appeared to become fixed in a single generation of a Holstein cattle maternal lineage[16]. We are undertaking an extensive survey of human maternal lineages to investigate this phenomenon.

Statistical evaluation of the evidence

The STR analysis supports the hypothesis that bodies 3–7 were related, although the probabilistic analysis is extremely complex and will be the subject of a separate paper. The mtDNA data also provided evidence to support the hypothesis that a mother and three of her children were present. However, STR data cannot provide information which specifically identifies whether this is the Romanov family. This problem can only be addressed by comparing mtDNA sequences with known maternal relatives.

Probabilistic analysis of the mtDNA proceeds as follows: R, The group is the Romanov family; R', The group is an unknown family unrelated to the Romanovs.

A third possibliity — that the group is an unknown family related to the Romanovs — is discounted because of the historical evidence to show that there were no related candidate families of the correct ages and sexes. A fourth suggestion — an elaborate hoax — was not considered in the absence of a realistic scenario. Finally, the possibility that contamination occurred is considered highly unlikely for the following reasons: i) Identical results were obtained from two different bones, each extracted in duplicate and at different times. In addition, for each set of extractions a negative control (no bone powder added) was taken through the entire procedure in parallel with the main tests. There is a need to interpret negative controls with some caution because it has been observed that reagents containing very low levels of contaminating DNA may preferentially amplify in the test sample. However, other bones from the grave, processed in parallel did not show evidence of a similar 'contaminant'. ii) Bone samples were extracted, amplified and sequenced separately, and blindly in a different laboratory (by E. Hagelberg). iii) STR results were not mixtures. iv) On average we observe 8.5 base differences between random individuals from a white Caucasian population from the UK[17]; it is improbable that two randomly chosen sequences would differ by just one basepair ($p=0.014$ using pairwise comparisons as described[17]).

We consider therefore, that the mtDNA extracted from the Tsar was genetically heteroplasmic. This complicates the interpretation because the strength of the evidence depends upon whether we accept *a priori* that a mutation has occurred in the Tsar. The probability of a single mutation was calculated by Stoneking[23] to be approximately 1/300 per generation, but this estimate does not take account of the incidence of heteroplasmy (much of which may be undetected). We use classical Bayesian inference for interpretation purposes. A standard short-hand notation is used: p(E|R) is the probability of the evidence if the group is the Romanov family, whereas p(E|R') is the probability of the evidence if the group is an unknown, unrelated family. The analysis we used followed two different approaches (see Methodology). The first calculation (lower bound) determined the likelihood ratio ($LR=p(E|R)/p(E|R')$) where the numerator p(E|R) is the probability of a single mutation occurring in the Tsar and no mutations in the generations between the Tsar, Tsarina and their living relatives ($p=3.2\times10^{-3}$); the denominator p(E|R') is the probability of two unrelated individuals differing by just one base in a white Caucasian database of 100 individuals ($p=4.8\times10^{-5}$)[17]. Under these assumptions, the evidence is 70 times more likely if R is true than if R' is true.

Alternatively, the calculation can proceed using the assumption that no mutations were observed in the Tsar, Tsarina and living relatives (the cloning experiments demonstrated that the mtDNA sequence of the Tsar's relatives is present as a minor component in the Tsar). This (upper bound) calculation evaluates the numerator p(E|R), the probability of observing no mutations in the Tsar, Tsarina and living relatives as $p=0.95$; the denominator p(E|R') was calculated from the observation

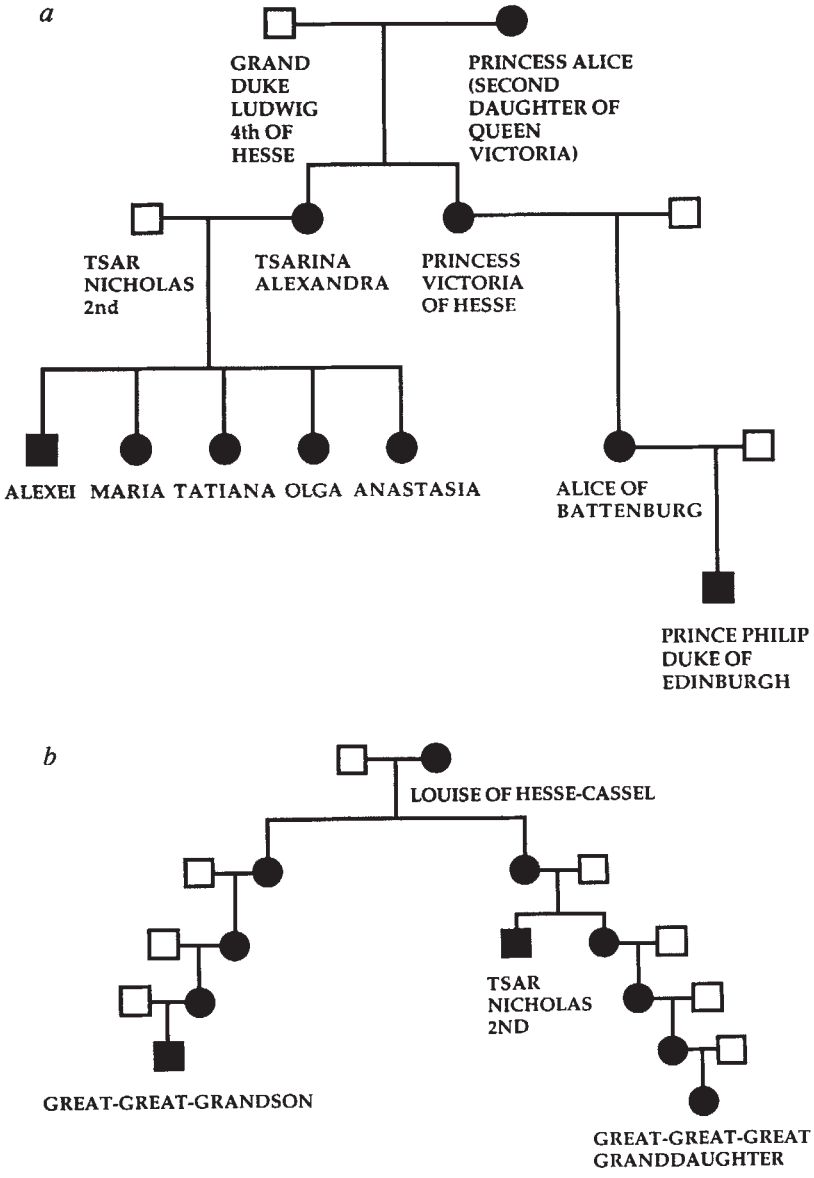

Fig. 3 a, Lineage of Tsarina Alexandra, showing relationship to HRH Prince Philip (Duke of Edinburgh). *b*, Lineage of Tsar Nicholas II, showing relationship to two maternal relatives tested.

of chance matches in a database ($p=1.2 \times 10^{-5}$)[17]. Under these circumstances, the evidence is 8×10^4 more likely if R is true than if R' is true.

None of the sequences obtained from the bone samples have been observed in our white Caucasian database[17] or of those held by the US Armed Forces Institute of Pathology (M. Holland; and M. Stoneking, personal communication); a total of 215 samples. Therefore, a likelihood ratio p(E|R)/p(E|R') based on the actual observed frequency of each mitochondrial genotype can be calculated as $0.95/(1/215)^2 = 4.4 \times 10^4$ (p(E|R) is unchanged, p(E|R') is simply the square of the observed frequency).

In our database[17], 78 out of 100 sequences were unique, as there are a large number of rare sequences in the population. If the total number of different sequences in the population is greater than the size of a small database (n), the observed frequency (1/n) of a typical (rare) sequence is directly dependent upon the actual size of the database and will therefore tend to overestimate the actual frequency. We argue that best estimates are given by pairwise comparisons because it has been previously demonstrated that overall, chance matches ($p=0.003$)[17] are a rare occurrence. However, both estimates derived from pairwise comparisons and are similar to those from observed frequencies, although we do not have a Russian nobility database to compare.

Discussion

The identification of victims of mass-disasters or murders where bodies are undiscovered for many years is one of the most challenging fields of forensic medicine. The powerful techniques offered by DNA analysis, used in conjunction with conventional methods greatly increases the probability of otherwise difficult identifications, provided that suitable relatives can be found for purposes of comparison. MtDNA analysis can be used effectively where victims and living descendants are separated by many generations, whereas STR analysis can be used to demonstrate closer relationships of family groups consisting of siblings and their parents.

This is the first major historical investigation in which both STR and mtDNA have been used as investigative tools. STR evidence demonstrated the presence of a family among the remains found in a mass grave at Ekaterinburg. Sex testing and mtDNA sequencing have established that the remains were almost certainly those of the Russian royal family. Although two of the Romanov children were missing from the grave, we cannot speculate on the identity of the missing princess. The use of DNA analysis in this extraordinary case provides a dramatic insight into the power of these techniques in solving historical questions.

As with any forensic case, interpretation of the evidence cannot be considered in isolation of other scientific and non-scientific evidence. Anthropological evidence gives prior information of the probability of the remains being those of the Tsar's family. *A priori*, there was dental evidence that the family was aristocratic (fillings were made of gold and platinum); all the bodies showed evidence of wounds; the bodies were estimated to be the correct age and sex; and were found in the correct location (Plaksin, V.O., personal communication).

We have examined several different scenarios, and evaluated the likelihood of the evidence for each of them. We have been widely reported by the news media for quoting a 98.5% probability that the remains are those of the Romanovs (equivalent to posterior odds of 70 to 1 on — see Methodology). We believe this to be a cautious statement, which will be revised upwards when all of the available evidence is brought together for critical examination and as our understanding of the population genetics and mutation biology of mtDNA improves. Bayesian inference places the prior odds (or the evidence prior to the DNA analysis) into the context of the two different likelihood rations derived from the DNA evidence. If we assume modest prior odds of 10 to 1, based upon the anthropological evidence (R.P. Helmer, President of the Craniofacial Identification Society of the International Association of Forensic Sciences; personal communication) this translates into posterior odds, of the remains being those of the Romanovs, of 700 to 1 on (lower bound) and 8.4×10^5 on (upper bound), respectively.

Methodology

Laboratory organisation. To minimise the possibility of contamination, all extractions were set up in a laminar flow cabinet in a dedicated laboratory (with dedicated equipment). Amplification was carried out in a different laboratory, so that amplified products never entered the extraction laboratory. Negative controls (where no bone powder was added to the extraction mixture) were used in all experiments. On the rare occasion that the negative controls gave a signal, the experiment was rejected. All of the results given for the putative family members have been duplicated. Sequence results have been confirmed by sequencing in both directions.

DNA extraction and quantitation. DNA was extracted from bone fragments using a modification of the method described by Hagelberg & Clegg[18]. The outer surfaces of bone fragments from the skeletons were removed by sanding with a flap-wheel attached to a high-speed electric drill and the remaining bone, approximately 1 g, was frozen in liquid nitrogen, then ground to a fine powder in a 6700 freezer mill (Glen Creston). The powder was mixed thoroughly with 2 ml 0.5 M EDTA (pH 8.0) containing 1 mg Proteinase K, plus 0.5% Tween 20, and incubated overnight at 37 °C. The mixture was then extracted twice with phenol, twice with phenol/chloroform and once with chloroform before centrifuging in a Centricon 30 microconcentrator (Amicon) for 1 h. The concentrates were washed 2–3 times by adding 3 ml distilled water and centrifuging for 1 h. DNA extracts from bones were quantitated by hybridization with a human-specific DNA probe kit (Gibco BRL, cat. no 4220 SA). Total human genomic DNA yield from the bones averaged 50–100 pg g^{-1} of bone. Blood samples from maternal relatives of the Tsar and Tsarina were supplied as a liquid or as stains on cotton cloth. 1µl (1mm²) samples were extracted by boiling for 10 min in a 20% Chelex (Biorad) solution and were amplified directly.

Sex determination. Sex was determined by amplification of a segment of the X-Y homologous gene, amelogenin, as described[9] except that samples (20 pg total human DNA) were amplified through 39 cycles comprising 30 s at 94 °C, 50 s at 60 °C and 50 s at 72 °C in a Perkin Elmer 9600 thermal cycler. 20 µl aliquots of PCR products were then loaded in a 4% agarose gel and separated by electophoresis for 90 min at 100 V and visualised by ethidium bromide staining under UV light.

Amplification of STRs. Each locus was amplified individually in a 50 µl reaction volume consisting of 1× PARR buffer (Cambio Laboratories), 1.25 U *Taq* polymerase, 200 µM dNTPs and 0.25 µM of each primer (one of which was labelled). Primer sequences and fluorescent dye labels were as previously described[2]. PCR cycling conditions were as follows: denaturing temperature 94 °C for 45 s, annealing temperature 54 °C for 60 s, extension temperature 72 °C for 60 s (HUMVWA31 — 38 cycles; HUMTH01 — 35 cycles; HUMF13A1 — 43 cycles; HUMFES/FPS — 40 cycles; HUMACTBP2 — 38 cycles). Following PCR, fragment sizes were determined as described[2].

Amplification of mtDNA. Two rounds of nested PCR from 50 pg of total extracted human DNA were performed as described[19] except

that the first round amplification was carried out over 30 cycles using primers L15926 and H00580 (ref. 20). 4 pairs of internal primers were used in separate second round amplification reactions with each pair comprising one 5' biotinylated oligonucleotide plus one chimaeric oligonucleotide incorporating the M13 universal sequencing primer sequence at the 5' end and sequencing was carried out essentially as described[17].

Quantitation of sequence microheterogeneity. Second round amplification of mtDNA from the putative Tsar was performed as described above using primers L15997 and H16401. The amplification was carried out for 30 cycles, from 1 μl first round PCR product. Second round PCR product was purified from a 3% Nu-Seive GTG agarose gel (FMC) using the GENECLEAN IIR Kit (BIO 101) according to the manufacturer's instructions. Approximately 100 ng of purified DNA was then ligated into the *Sma*I site of pUC18 using the SureCloneTM Ligation Kit (Pharmacia) as directed, and 1/4 of this reaction was used to transform Epicurian Coli SURE™ competent cells according to the given protocol. Colonies which contained recombinant plasmids were identified by ampicillin resistance and α-complementation using standard protocols[22]. Small-scale plasmid purifications were carried out using the Magic™ Minipreps DNA Purification System (Promega) according to the manufacturer's instructions. *Taq* cycle sequencing of 12% of the plasmid DNA was carried out using a Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's protocol. Electrophoresis and analysis were carried out as given above.

Interpretation. *Analysis incorporating mutation rates:* The mutation rate in the 800 bp control region has been estimated as once in every 6,000 years[23]. This figure translates into approximately 200–300 generations. On this we base the assumption that mutation events within the control region occur with a Poisson distribution with parameter (gμ) where g is the number of generations and μ is conservatively taken to be 1/300. The probability of a single mutation occurring in one generation is therefore $\mu e^{-\mu}$.

Let M and F denote the mtDNA sequences of the male and female adult members of the family group. Let X denote the common sequences of two maternal descendants of the Tsar's maternal grandmother (Fig. 3) and let D denote the sequence of the Duke of Edinburgh. The analysis needs to take account of the following events:

E_1: the sequence (M) of the male corresponds to the common sequence (X) of 2 confirmed maternal relatives of the Tsar with the exception of a difference at one site, E_1; M≈X.

E_2: the sequence (F) of the female corresponds to the sequence (D) of the Duke of Edinburgh; E_2: F=D.

Bayes theorem gives: $\frac{p(R|E_1E_2)}{p(R'|E_1E_2)} = \frac{p(E_1E_2|R)}{p(E_1E_2|R')} \cdot \frac{p(R)}{p(R')}$, where p(R)/p(R'), the prior odds, are based on all of the other non-DNA evidence. The likelihood ratio (LR) is evaluated as follows:

$$LR = \frac{p(E_1E_2|R)}{p(E_1E_2|R')} = \frac{p(E_1|E_2R)}{p(E_1|E_2R')} \cdot \frac{p(E_2|R)}{p(E_2|R')}$$

If it is accepted that, given R or R', the male and female are from independent female lines, then

$p(E_1|E_2R) = p(E_1|R)$ and $p(E_1|E_2R') = p(E_1|R')$ and:

$$LR = \frac{p(E_1|R)}{p(E_1|R')} \cdot \frac{p(E_2|R)}{p(E_2|R')}$$

To evaluate $p(E_1|R)/p(E_1|R')$: E_1 is the event of 1 mutation occurring between the Tsar and his mother and no mutations occurring between the Tsar's mother and each of her descendants. There are 5 generations and 4 generations, respectively between the Tsar's mother and living relatives (Fig. 3), hence $p(E_1|R) = \mu e^{-\mu}.e^{-5\mu}.e^{-4\mu} = \mu e^{-10\mu}$.

If R' is the case then there is a probability[19] of 69/4950 that the mtDNA sequences of 2 unrelated individuals would differ by only one base.

To evaluate $p(E_2|R)/p(E_2|R')$: If R is the case, then the probability that no mutations will occur in the 4 generations between the Tsarina and the Duke of Edinburgh is $e^{-4\mu}$ (Fig. 3). If R' is the case, then there is a probability[19] of 17/4950 that the mtDNA sequences of two unrelated individuals would correspond exactly. Then:

$$LR = \frac{\mu e^{-14\mu}}{(69/4950)(17/4950)} = 70 \text{ (to 1 significant figure)}$$

Analysis without incorporating mutation rate: Given heteroplasmy, which itself suggests that the mutation is recent, we argue that this is an understatement of the evidence and could be regarded as a lower limit; we can postulate an upper bound for the LR by carrying out a calculation as though no mutation was observed. Under this circumstance the probability that no mutations have occurred in the 14 generations between the Tsar, Tsarina and their living relatives is $e^{-14\mu}$, hence the numerator of the likelihood ratio (LR) = 0.95 and the denominator is simply the square of the frequency of observation of chance matches in our database of white Caucasians $(17/4950)^2$, therefore the upper bound LR = 8×10^4.

Acknowledgements

We thank HRH Prince Philip, the Foreign and Commonwealth Office, Maestro Mstislav L. Rostropovich, the Englehardt Institute of Molecular Biology, Russian Academy of Sciences, Professor V.O. Plaksin (Chief Forensic Medical Examiner of Russian Federation), D. Lindley, N. McCrery, J. Nott, Scientific Council on the Russian Human Genome Project, Wink Inc., Consulate General of Japan in Britain and Biwako cultural museum in Shiga Prefecture of Japan and Russian International Airlines (AEROFLOT) for their help and contribution to this work. P.L.I was supported by research grants from Applied Biosystems (ABI), Perkin Elmer Corporation and the Foreign and Commonwealth Office.

Received 10 August; accepted 8 November 1993.

1. Fregeau, C.J. & Fourney R.M. DNA typing with a fluorescently-tagged short tandem repeats: A sensitive and accurate approach to human identification. *Biotechniques* **15**, 100–119 (1993).
2. Kimpton, C.P. *et al.* Automated DNA profiling employing 'multiplex' amplification of short tandem repeat loci. *PCR Meth. Applic.* **13**, 13–22 (1993).
3. Hagelberg, E., Gray I.C. & Jeffreys A.J. Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* **352**, 427–429 (1991).
4. Jeffreys, A.J., Allen, M.J., Hagelberg, E. & Sonnberg, A. Identification of the skeletal remains of Josef Mengele by DNA analysis. *For. Sci. Int.* **56**, 65–76 (1992).
5. Bogenhagen, D. & Clayton, D.A. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. *J. biol. Chem.* **249**, 7991–7995 (1974).
6. Paabo, S., Gifford, J.A. & Wilson, A.C. Mitochondrial DNA sequences from 7000 year old brain. *Nucl. Acids Res.* **16**, 9775–9785 (1988).
7. Hagelberg, E., Sykes, B. & Hedges, R. Ancient bone DNA amplified. *Nature* **342**, 485 (1989).
8. Sullivan, K.M. Mannucci, A. Kimpton, C. & Gill, P. A Rapid and Quantitative DNA Sex Test: Fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques* (in the press).
9. Mannucci, A., Sullivan, K.M., Ivanov, P.L. & Gill, P. Forensic application of a rapid and quantitative DNA sex text by amplification of the X-Y homologous gene amelogenin. *Int. J. leg. Med.* **15**, 636–641 (1993).
10. Edwards, A., Civitello A., Hammond, H.A. & Caskey, C.T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J. hum. Genet.* **49**, 746–756. (1991).
11. Kimpton, C.P., Walton, A. & Gill, P. A further tetranucleotide repeat polymorphism in the vWF gene. *Hum. molec. Genet.* **1**, 287 (1992).
12. Polymeropoulos, M.H., Rath, D.S., Xiao,H. & Merril, C.R. Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). *Nucl. Acids Res.* **19**, 4036 (1991).
13. Polymeropoulos, M.H., Rath, D.S., Xiao, H. & Merril, C.R. Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). *Nucl. Acids Res.* **19**, 4791 (1991).
14. Warne, D., Warkins, P., Bodfish, P., Nyberg, K. & Spurr, N.K. Tetranucleotide repeat polymorphism at the human beta actin related pseudogene 2 (ACTBP2) detected using the polymerase chain reaction. *Nucl. Acids Res.* **19**, 6980 (1991).
15. Howell, N. *et al.* Mitochondrial gene segregation in mammals: is the bottleneck always narrow? *Hum. Genet.* **90**, 117–120 (1992).
16. Koehler, C.M. *et al.* Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics* **129**, 247–255 (1991).
17. Piercy, R., Sullivan, K.M., Benson, N. & Gill, P. The application of mitochondrial DNA typing to the study of white caucasian identification. *Int. J. leg. Med.* **106**, 85–90 (1993).
18. Hagelberg, E. & Clegg J. Isolation and characterisation of DNA from archeaological bone. *Proc. R. Soc. Lond. B.* **244**, 45–50 (1991).
19. Sullivan, K.M., Hopgood R., Lang, B. & Gill, P. Automated amplification and sequencing of human mitochondrial DNA. *Electrophoresis* **12**, 17–21 (1991).
20. Orrego, C. & King, M.-C. Determination of familial relationships. In PCR protocols (eds M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White) 416–426 (Academic Press, London, 1990).
21. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981).
22. Maniatis, T., Fritsch, E.F. & Sambrook, J. Molecular cloning: a laboratory manual. (Cold Spring Harbor Laboratory, New York, 1982).
23. 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).