The Romanov mystery:

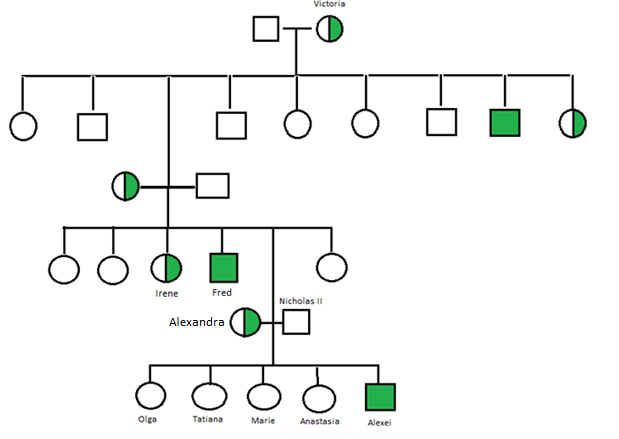


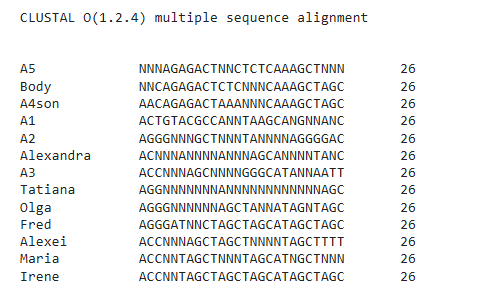
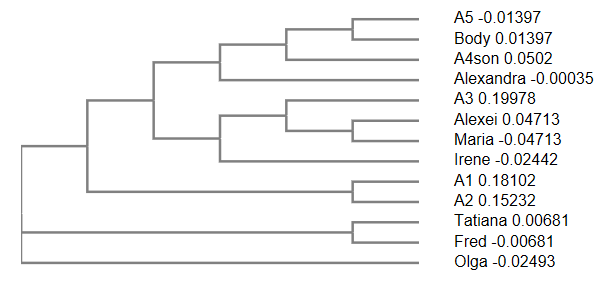
Figure - The Romanov family tree including details of known haemophiliac carriers/patients, represented with green

The Romanov family (Figure 1) had ruled Russia since 1613, however in 1918 in response to the Russian revolution, the family that included Tsar Nicolas, Tsarina Alexandra, their five children, Olga, Tatiana, Marie, Anastasia and Alexei were executed by a Bolshevik mob. It was reported that the bodies of the family were burned and buried in the nearby woods. For many years the location of the graves remained unknown, however, ten years following the murder, an amateur archaeologist discovered the bodies of 6 of the seven family members. The body of Anastasia, however, appeared to be missing.

In response to this discovery, there was much speculation about what happened to the family, with reports that Anastasia may have survived. In this report, we analyse the genomic data of 5 women who have since come forward claiming to be the missing Princess using genomic DNA analysis to determine the candour of their testimonies as well as DNA collected from an unidentified body found near the other graves.

Mitochondrial DNA (mtDNA) from the four women claiming to be Anastasia (A1, A2, A3, A5), as well as the son of a fifth woman, A4son (Figure 2) and sequences of the Romanov family were compared to look for levels of relatedness. MtDNA is beneficial over genomic DNA due to the lack of recombination between parental DNA making tracing phylogenies through maternal lineages more feasible (Taberlet 1996). As mtDNA is maternally inherited (Giles 1980), DNA samples were also compared to the sequences of Princess Irene and Prince Fred who are related to the Romanov’s through Alexandra. They will, therefore, be highly similar to Alexandra and her five children, but not to Nicolas.

Regrettably, the grave sites were compromised, with the bones of the unidentified body contaminated upon the farmer falling in, and the six bodies identified as the rest of the Romanov family heavily degraded by fire. As a result, the findings should be treated with caution as it meant there was less coverage for comparison. Further, due to a low budget, only low coverage sequencing could be afforded that included a 10% genotyping error. Many of the nucleotides collected from the family’s bodies could not be accurately determined and are represented below with an N.

Figure 2-multiple sequence alignment and phylogenetic comparisons between the maternal lineage of the Romanov family and the Anastasia candidates

The Clustal Omega programme was used to create a multiple sequence alignment between the Romanov’s and the Anastasia contenders (Figure 2) (ref?). The results demonstrate how degraded parts of the Romanov sequences were but suggest the son of A4 is the most closely related to the Tsarina (0.00035) while A3 showed a high level of similarity to Irene, Alexei and Marie.

Relatives of Queen Victoria, including members of the Romanov family (Figure 1) suffered from a rare genetic condition called haemophilia. Haemophilia is a recessive disease carried on the X chromosome that affects about 1 in 5000 males. While the disease affects males, it is carried by females (*Br Med J*). Alexandra was known to be a carrier of this allele, and Prince Alexei suffered from the disease, for which he was treated by the questionably trustworthy Grigori Rasputin. There is ~50% chance a carrier will have a son that suffers from the disease and ~50% chance her daughter will also be a carrier. Genetic analysis revealed that the unidentified body and A2 were both haemophilia carriers while A2 also had a haemophilic son. The finding of this rare heritable genetic condition fits with the Romanov family history. None of the other candidates were carriers, although the son of A4 was a haemophilic sufferer. As A4 herself was not a carrier, it may suggest that the boy is not her own son, therefore reducing the credibility of her testimony. However, in ~30% cases, the disease can arise as the result of new mutations leading to ‘sporadic haemophilia’(WFH 2012), which would explain her son’s condition without her being a carrier.

Despite the questionable testimony given by A4, she bears a great physical similarity to what is known about the Princess, being of the right age based on Rasputin’s claim the princess was born in 1901, as well as having blue eyes and strawberry blonde hair. A2 was also within an acceptable age bracket, with blue eyes and blonde hair. The physical features of these two women matched best to the photos of Anastasia. A1 and A5 were a little older than would have been expected, while A3 was younger than would have been expected.

Pairwise global alignment of the DNA sequences was also made between each candidate and the Tsarina (Figure 3). This revealed that none of the samples showed a high level of similarity to the Tsarina, entering the possibility that none of these women or the body are that of the missing Princess. However, it is also noteworthy that there were high levels of gaps in the sequences which could be attributed to the high level of degradation of the Tsarina’s DNA sequences. As only a short section of the DNA was available for comparison, it does not give conclusive results. Therefore, further pairwise global alignment comparisons were carried out using the sequences of Irene whose DNA was far more intact (Figure 4). Gaps in sequence alignment are the result of insertions and deletions.

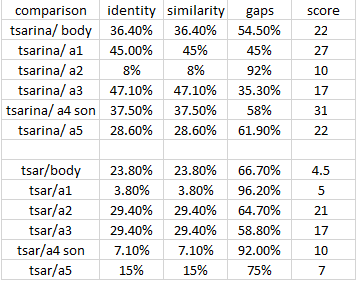


Figure 3-Pairwise global alignment using mtDNA for each Anastasia candidate in comparison to the Tsarina Alexandra showing the son of A4 to be the most closely related

The second pairwise global alignment resulted in generally higher scores as a result of more intact sequences that could be compared, but again showed that there was not a considerable match with any of the samples. The comparison between Irene and the son of A4 showed a 56.2% similarity which along with the highest score of 40.5, added to the highest score in comparison with the Tsarina suggests A4 is the most likely to be Anastasia given the level of similarity to maternal relatives of the Romanov family.

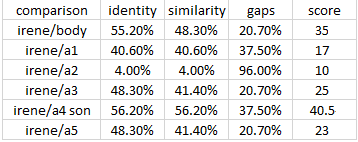
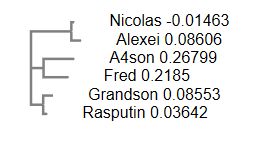


Figure 4-Pairwise global alignment using mtDNA for each Anastasia candidate in comparison to Princess Irene showing A4 to be the most closely related

A sequence alignment using Y chromosome sequences was performed using sequences collected from the males in question in this report, including Prince Fred, Nicolas, Alexei, the farmer's grandson and Rasputin (Figure 5). A phylogenetic tree confirmed Nicolas and Alexei to be closely linked. However interestingly the farmer's grandson and Rasputin appeared to be genetically related. This, along with knowledge of the farmer's daughter’s reaction to Rasputin’s name and the son’s great similarity to Rasputin suggest that the grandson is Rasputin’s child.

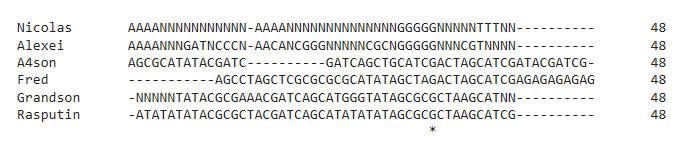


Figure 5- Multiple comparisons and sequence alignment of the Y chromosome sequences for 6 males associated with Anastasia

In a study by Schwark 2008, it was found that the reliability of genetic analysis reduces with the severity of fire damage a body suffers. Despite advances in DNA sequencing techniques that have aided in forensic investigations such as this one, it should be noted that using ancient DNA, especially of fire damaged samples, can be problematic due to contamination from recent DNA (Hagelberg 2015). As the DNA collected from the bodies of the Romanov family, including the body of the unidentified body was burnt with high levels of degradation, this, therefore, is a major limitation in the search for Anastasia as in many cases there were large sequence gaps.

In summary, each case had its credits and reservations. The unidentified body showed a high level of similarity to Irene in the global alignment and further was a carrier of haemophilia, however, possibly as a result of high levels of degradation to the sample, the multiple comparisons showed low levels of relatedness to the Romanov family. A1 was too old to match the description, with low scores in the global alignment comparison to Irene and in the multiple sequence comparisons. A2 fitted the physical description and was a haemophilia carrier. However the results of the multiple comparison and global alignment showed low relatedness scores. A3 was too young and showed low global alignment scores, but the results of the multiple comparisons showed a possible relationship to Irene, Alexei and Marie. A4 matched the physical description the best, and further had the highest score in the global alignment to both Irene and the Tsarina as well as having the closest link to the Tsarina in the multiple comparisons. A5 was too old and showed little relatedness in the multiple comparison, compiled with low/mid scores in the global alignment comparisons.

Based on this, although there remains some doubt over the identity of Anastasia, it appears the testimony of A4 is the most plausible from the evidence available. It should, however, be noted that a conclusive decision could not be made based on limited genetic evidence. In future, longer sequences should be obtained and compared to more Romanov family members.

**References:**

1. Br Med J 1956;1:675
2. Clustal Omega, <https://www.ebi.ac.uk/Tools/msa/clustalo/>, date of access 10/04/19
3. Giles, R.E., Blanc, H., Cann, H.M. and Wallace, D.C., 1980. Maternal inheritance of human mitochondrial DNA. *Proceedings of the National academy of Sciences*, *77*(11), pp.6715-6719.
4. Hagelberg, E., Hofreiter, M. and Keyser, C., 2015. Ancient DNA: the first three decades.
5. Schwark, T., Heinrich, A., Preuße-Prange, A. and von Wurmb-Schwark, N., 2011. Reliable genetic identification of burnt human remains. Forensic Science International: Genetics, 5(5), pp.393-399.
6. Taberlet, P., 1996. The use of mitochondrial DNA control region sequencing in conservation genetics. Molecular Genetic Approaches in Conservation. Oxford University Press, New York, pp.125-142.
7. World Federation of Haemophilia (WFH), ‘Carriers and Women with haemophilia’ (2012)