Mitochondrial DNA Analysis of Forensic Evidence: Jack the Ripper or Ripoff?

Felice L. Bedford, University of Arizona

Abstract A widely publicized study claimed to finally identify the infamous serial killer Jack the Ripper based on a forensic analysis using mitochondrial DNA (mtDNA) extracted from a possession of a victim. However, only control regions of mtDNA were sequenced which greatly raises the odds of a match to a purported relative from chance alone. In addition, rather than simply stating the locations of the DNA polymorphisms or even the number of variants found as would be expected, they show only confusing crude graphical blocks that are further misleading about the odds of a chance match and do not allow independent verification of calculations. The haplogroup (for example, outlaw Jesse James is Haplogroup T2) is also withheld despite its usefulness towards evaluating a claim that the identified murderer was of Russian Jewish descent, as well as of public interest that they claim was a purpose of their report. They attribute all the secrecy to the Data Protection Act of 2018 but our search of the 354 page document does not preclude (or even mention) prohibition of publishing at the nucleotide level as claimed and if it did, hundreds of thousands of mtDNA sequences in publications would be in violation. Overall, no evidence is presented from mtDNA to implicate the identity of Jack the Ripper. Even mitochondrial DNA is innocent until proven guilty.

Keywords: Forensics, Haplogroup, Jack the Ripper, Jewish genetics, mtDNA

Sir,

The genetic analysis of cells from purported Jack the Ripper[1] is a disappointment to the mitochondrial DNA (mtDNA) enthusiast. While the authors state that a goal of their data presentation was to be of interest to the general public, they fail to provide the critical piece of information that would do so: The mtDNA haplogroup of Jack the Ripper. Haplogroups reflect the most basic division of maternal ancestry in population genetics. With more than half a million people choosing to have their own DNA tested at one commercial testing company alone, exchanging haplogroup information is becoming as commonplace as exchanging zodiac signs once was. And with nearly all people of Europe sharing one of only seven haplogroups, there is no call for secrecy. The mtDNA testing that the researchers conducted should have allowed sufficient prediction of haplogroup, akin to the low-resolution testing that was standard until relatively recently. This would at least allow Jack the Ripper to be added to the database of other notables such as outlaw Jessie James (Haplogroup T, more specifically T2), president Abraham Lincoln (X1c) and King Richard III (J1c2c).

Also of interest is that the suspect allegedly implicated by DNA, Aaron Kosminski, was Jewish of Russian ancestry. In the population genetics literature, some sequences have thus far only been found in Jewish groups; see for instance, K1a1b1a [2] and our investigations of T2e1b

[3] although these would require more detailed testing. In general, results pointing to a sequence with known high or low prevalence in the Russian Jewish population would be eyebrow raising.

Turning to the task of convincing identification from mtDNA, here too, the researchers' choices were puzzling, especially if they were not going to disclose the predicted haplogroup anyway. Based on the primer pairs they provide in a table, we can deduce that they tested a total of 400 positions from HVI (16000-16400) and 328 positions (48-376) from HVII control regions out of the more than 16,000 positions that comprise the human mitochondrial DNA. These are small regions that would be expected to have many thousands of matches based on chance alone. They come from the control regions of mtDNA which is subject to a rapid rate of mutation - HV stands for "hypervariable" - and therefore with a known inability to prove real connections between two matching sequences. The same technique used of amplifying small fragments of degraded DNA could just as easily have been applied to the slower mutating coding regions, which is especially of value if there were any private mutations (see below).

Concerning the data they did collect, the analysis and presentation remain unconvincing. They present only a graphical depiction of shaded matching "blocks" among mtDNA samples in order to argue for a match between the crime scene cells and the maternal relatives of a suspect. This chopping up of a continuous sequence of mtDNA into blocks is an artificial invention that not only turns a fine chisel instrument into a sledgehammer but one we suggest can be misleading. For instance, suppose on one block there is an "exact match" in which both DNA samples have the identical variant of, say, T16126C, that is, a change in nucleotide base from T to C at position 16126 compared to the Cambridge Reference Sequence (CRS), while at all the other positions of the block, both samples share the CRS. (The number of positions in each block was not provided but a guestimate might be 33 for HVI based on their figure showing 12 blocks and the 400 positions inferred from the table.) This exact match could be because both individuals share haplogroup known as T, which is found in nearly 10% of Europeans. The odds the two samples have an exact match on this one block based on chance alone would be 1 in 10 or 10%. Now let's add that they match identically again on a second block, this time both sharing the variant C16294T as well as matching CRS everywhere else. This matching would be expected because both individuals have haplogroup T, which harbors a genetic variant at position 16294 as well as 16126. So, the odds of a match on chance alone does not change with the addition of a second block match and remains 10%. Continuing to match additional blocks, if they both share a variant in a block from HV2, at A73G, once again this would be expected from variants found in the T haplogroup and the probability remains the same. Thus, the block analysis gives the illusion of each one being independent of the others and that each additional matching block would multiply the probabilities and decrease the odds of a match from chance alone. Instead, the odds of a match based on chance alone can stay identical even with exact matches on all of their 24 blocks. A chance match of 10% is a high number and far from the fraction of a percent desired to confirm a match.

On the decision to present this block data rather than the nucleotide sequences and variants, the authors offer: "Second, due to the restrictions set by the Data Protection Act, detailed nucleotide-level DNA information of living individuals should not be published". If this were true, then the several hundred thousand mtDNA detailed control region sequences published in the literature would be in violation, nearly all of which were collected and published while the donor was living. We searched a pdf version of the 354 page Data Protection Act 2018 (https://www.legislation.gov.uk/ukpga/2018 /12/contents) for "nucleotide" with zero hits and no mention of nucleotide-level DNA information. Instead, the Data Protection act refers only in very general terms to genetic information: "…'genetic data' means personal data relating to the inherited or acquired genetic characteristics of an individual which gives unique information about the physiology or the health of that individual." We note that health information is found only in the coding regions of mtDNA, not the control regions that the researchers tested. Moreover, there is no unique information provided by a mtDNA haplogroup because it is shared by millions of people, or even a more specific subhaplogroup, which is usually shared by many thousands.

It is also worth noting a basic fact about MtDNA transmission that was not discussed explicitly in the article. The living DNA donor could not have been a direct descendent of Jack the Ripper because males do not pass on their mtDNA to the next generation. Instead, the living donor is only at most a very general "cousin", descended from perhaps a sister or a maternal aunt, or great aunt, and so on, further obscuring a unique identification. Nor does the Data Act preclude publication and, in fact, there are even exemptions for "academic purposes" (Part V, 26b exemptions based on freedom of information) Overall, we therefore see no restrictions provided by the new Data Act of 2018 on publishing data more specific than block amalgams and more in keeping with standard mtDNA scientific publishing. (Note that this was their second reason for presenting only thick chunks of matching data; their first reason was to be of interest to the general public, but as already noted at the outset this would be accomplished with the revelation of Jack the Ripper's haplogroup, not shaded rectangles.)

Continued concern that the identity of alleged Jack the Ripper cousin might somehow be recreated would still not preclude providing the number (if not the positions) of private mutations that were found as well as the frequency of each mutation. That would allow calculation of the odds of a chance match between the mtDNA sequences in this particular case. Private mutations are variants that are not found in the majority of others with the same haplogroup, occurring because of time of origin, population isolation, or just plain luck. If Jack the Ripper's mtDNA happened to have infrequent private mutations, then the probability of a true relation to any matching sample is strengthened even though the mutations come from the low-resolution control regions as tested by the authors. But the converse is true as well and thus it is critical to present this data. They state that the frequency of the mtDNA "profile" for the suspect was only 1.9 x 10^{-2} (i.e. 0.019) which they said was obtained from EMPOP; however, EMPOP does not provide frequencies for arbitrary blocks of DNA sequences and the authors do not provide any specific information on how they arrived at this number. Are they hinting by "profile" that they found Jack the Ripper to have a relatively infrequent mtDNA haplotype with private mutations?

Finally, while detailed primers for mtDNA were included in methodology, basic information was withheld: How many other Jack the Ripper suspects were considered and checked for a mtDNA match against the scarf sample? The greater the number of non-matches that were ruled out for other suspects, the more compelling the match they did find. How many maternal relatives were tested for each of these suspects to be certain there were not any NPEs (non-parental events) so as to verify the maternal descent lineage? As a side note, it would also be useful for the reader to learn why the authors considered it an unknown "academic exercise" as to whether small amounts of 120-year-old cells could be tested for mtDNA when mtDNA has been extracted successfully from ancient remains. The oldest thus far dates to 5000 years before present (Otzi the Iceman, subhaplogroup K1f).

We suspect many of the issues discussed here stem from mtDNA being put to different uses by forensic and population genetic scientists. Having now seen the other perspective, we are asking the authors to provide from their fascinating research the predicted haplogroup they discovered for Jack the Ripper, the subhaplogroup branch information as per Phylotree v.17 [4], the number of private mutations, the frequencies of each of these mutations, and basic procedural information, so as to finally settle this historical and infamous Jack the Ripper case for scientists and public enthusiasts alike. Even mitochondrial DNA is innocent until proven guilty.

References

1. Louhelainen, J. & Miller, D. Forensic Investigation of a Shawl Linked to the "Jack the Ripper" Murders. *J. Forensic Sci.* **65**, 295–303 (2020). **Epub 2019 March 12**

2. Behar, D. M. *et al.* MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur. J. Hum. Genet. EJHG* **12**, 355–364 (2004).

3. Bedford, F. L., Yacobi, D., Felix, G. & Garza, F. Clarifying mitochondrial DNA subclades of T2e from Mideast to Mexico. *Journal of Phylogenetics and Evolutionary Biology* **2**, 1–8 (2013).

4. Oven, M. van. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int. Genet. Suppl. Ser.* **5**, e392–e394 (2015).

Submitted copy now published in J Forensic Sci, 2019, doi: 10.1111/1556-4029.14192 Available online at: onlinelibrary.wiley.com