

Beyond traditional paternity and identification cases: Selecting the most probable pedigree

T. Egeland, *Section of Medical Statistics, University of Oslo, and
National Hospital, Centre for Epidemiology and Hospital Statistics, 0027 Oslo, Norway.*

Phone +47 22867919, fax +47 22867662, email: egeland@basalmed.uio.no

P. F. Mostad, *Norwegian Computing Center, Oslo.*

B. Mevåg, *Institute of Forensic Medicine, Oslo.*

M. Stenersen, *Institute of Forensic Medicine, Oslo.*

Abstract

The paper extends on the traditional methodology used to quantify DNA evidence in paternity or identification cases. By extending we imply that there are more than two alternatives to choose between. In a standard paternity case the two competing explanations H_1 : "John Doe is the father of the child" and H_2 : "A random man is the father of the child", are typically considered. A paternity index of 100,000 implies that the data is 100,000 more likely assuming hypothesis H_1 rather than H_2 . If H_2 is replaced by "A brother of John Doe is the father", the LR may change dramatically. The main topic of this paper is to determine the most probable pedigree given a certain set of data including DNA profiles. In the previous example this corresponds to determining the most likely relation between John Doe and the child. Based on DNA obtained from victims of a fire, bodies found in an ancient grave or from individuals seeking to confirm their anticipated family relations, we would like to determine the most probable pedigree. The approach we present provides the possibility to combine non-DNA evidence, say age of individuals, and DNA profiles. The program **familias**, obtainable as shareware from <http://www.nr.no/familias>, delivers the probabilities for the various family constellations. More precisely, the information (if any) prior to DNA is combined with the DNA-profiles in a Bayesian manner to deliver the posterior probabilities. We exemplify using the well published Romanov data where the accepted solution emerges among 4536 possibilities considered. Various other applications based on forensic case work are discussed. In addition we have simulated data to resemble an incest case. Since the true family relation is known in this case, we may evaluate the method.

Keywords: Forensic genetics; Likelihood ratios; Identification

1. Introduction

There is a long tradition in forensics to report the weight of DNA evidence by means of paternity indices or likelihood-ratios (LR) ([4], [5]). For instance, in a standard paternity case the two competing explanations H_1 : “John Doe is the father of the child” and H_2 : “A random man is the father of the child”, are typically considered. A paternity index or a LR of 100,000 is then interpreted to mean that the data is 100,000 more likely assuming explanation H_1 rather than H_2 . As pointed out by several authors, the phrasing of the hypotheses is of critical importance. If H_2 is replaced by “A brother of John Doe is the father”, the LR may change considerably. The topic of this paper may be formulated in several ways, one is the following: We would like to determine the most probable pedigree given a certain set of data. In the former example this corresponds to determining the most probable relation between John Doe and the child; there is obviously a large number of pedigrees to choose from ranging from a parent-child relation to a pair of unrelated individuals.

There are several applications of our approach including identification following disasters, resolving family relations when incest is suspected and determining the most probable relation between a person applying for immigration and claimed relatives of the individual. To be more specific, we outline a case which will be detailed later in the paper. Following a fire, four dead bodies were recovered. They were believed to be a mother, her two children, as well as a more distant relative. Based on the autopsy, it was apparent that one of the persons could be old enough to be a parent, whereas the other three persons were most likely not old enough to have children. DNA data from the mentioned bodies were obtained as well as data from five putative close relatives. The pedigree is shown in Figure 1. Based on this information, the **familias** program presented in this paper verifies the assumed relations between the bodies found. In this example, it may well be fairly obvious who the four individuals are. There may however be some doubt as to the precise identification of the four beyond the fact that the four missing persons have been identified. Obviously, similar calculations may be required in immigration cases. Indeed, in the seminal paper by Jeffreys et al. [8] one such case is discussed. The need to have some numerical quantification may be stronger in immigration cases: There are no known missing persons and the authorities may find it easier to relate to numbers than imprecise statements summarizing the views of the forensic experts.

The outline of the paper is as follows: In the next section we describe the method followed by examples, all but one based on real data. In Example 2 we discuss the well published Romanov case and fill in a detail left open in [6]. Here we evaluate 4536 different pedigrees showing that the accepted solution is the clearly most probable. Example 3 is based on simulated data and so we are able to compare the results of our approach with the truth. The last two examples both deal with the problem of determining the most probable relation between two persons without having access to data from additional family members.

The most recent version of the software is available from <http://www.nr.no/familias> and extends our previous program described in [3]. Obviously, there are several computer programs (e.g. [2]), but as far as we know, none solving the problems we address. Throughout we focus on identification cases. Related problems for stains, addressed for instance in [10], may be handled similarly.

2. Methods: A Bayesian approach

Our method may be divided into the following stages: First, we describe the set of “possible” pedigrees involving the relevant persons. Clearly, this set is extremely large, but for practical reasons we need to limit the number. The examples we have run so far have involved less than 10,000 pedigrees. Secondly, we assign a prior probability distribution to this set of pedigrees, based on non-DNA evidence. Finally, we introduce DNA measurements and mutation parameters, obtaining a posterior probability distribution on the pedigree set.

2.1. Selecting pedigrees

Many different methods could be used to specify, and then generate, a set of pedigrees containing a given group of individuals. We use the following approach: First, we distinguish between children, i.e., persons that can not have children of their own, and persons that may have children. This distinction will typically be made based on age determination. If no such information is available, then a safe procedure would be to classify all persons at this stage as adults. Next, adults are characterized according to gender.

Based on the information above, one may generate all possible pedigrees containing only these individuals. However, one will frequently be interested in pedigrees involving persons not included in the original group. For example, to describe that a woman has three children with the same man, it is necessary to include this man in the pedigree, even though his DNA is unavailable. If, for instance, two persons found following a fire are believed to be sisters, two additional persons are needed to describe the pedigree where they are full sisters; if they may also be cousins, clearly more persons will be required. The implemented approach introduces a number of “extra” men and “extra” women and generates all possible, different pedigrees.

2.2. Prior model

The set of pedigrees generated should contain the pedigrees we consider probable given background information, but it will also contain a large number of pedigrees that are unlikely for different reasons. For example, many very incestuous pedigrees will be generated; in most cases, they should not be considered a priori as

likely as non-incestuous pedigrees. Similarly, most generated pedigrees will indicate a more promiscuous behavior than is usual in most cultures.

In our method, we try to generate a probability distribution on the set of pedigrees reflecting such considerations. Starting with a uniform probability distribution on the pedigree set, we may choose to modify the prior probabilities of different pedigrees using the three options *inbreeding*, *promiscuity* and *generations*. The first of these may be used to increase or decrease likelihoods of pedigrees involving inbreeding. A similar comment applies to promiscuity, while generations alludes to the modification of likelihoods of pedigrees extending over several generations. For instance, five persons may constitute five generations. However, such a pedigree may be inconsistent with age information. In more precise terms, the prior distribution is proportional to

$$M_I^{b_I} M_P^{b_P} M_G^{b_G}, \quad (2.1)$$

where M_I , M_P , and M_G are non-negative parameters provided by the user of the program; the subscripts refer to the three mentioned options. The corresponding integer exponents b_I , b_P , and b_G explained next are calculated by **familias**. b_I is the number of children where both parents are present in the pedigree and where the parents have a common ancestor in the pedigree. For *promiscuity*, the number of pairs having precisely one parent in common is calculated and denoted b_P . Finally, turning to *generations*, the longest chain of generations starting with a named person and ending in an adult, is calculated and assigned the value b_G . In addition, it is possible to discard automatically all pedigrees where the number of generations exceeds a prescribed level.

The user may assign 0 prior probability to all incestuous pedigrees by letting $M_I = 0$. A value of the parameter between 0 and 1 decreases the likelihood of incestuous alternatives while a value exceeding 1 increases incestuous constellations. A similar comment applies to the other options. A small, artificial example, illustrates some of the concepts above. Assume three men, M1, M2 and M3 are found dead and two alternatives are considered: H_1 *M1 is the father of M2 who is the father of M3* and H_2 *M1 is the father of M2. M3 is unrelated to M1 and M2*. The ratio of the priors corresponding to alternatives H_1 and H_2 follows from (2.1) as

$$\frac{M_I^0 M_P^0 M_G^3}{M_I^0 M_P^0 M_G^2} = M_G.$$

We emphasize that this prior is but one pragmatic suggestions among many others possible, in many cases they are not needed.

Example 1a. There are 4536 possible family relations among one male and four females, according to our implemented algorithm. A large number of these families may be ruled out prior to acquiring DNA

data in the presence of some additional information on the individuals. Discarding or down weighting of pedigrees may be done informally or, as exemplified towards the end of this example, by use of the prior distribution (2.1). Assuming for instance that three of the females are children, i.e., they have no children of their own, the number of possible family relations is reduced to 192. In this case the number may be deduced by a combinatorial argument: There are 3 relations between the adults (one may be parent of the other or they are unrelated) and each child may have 0, 1 or 2 parents among the two adults resulting in 4 combinations. Consequently we realize that there are $3*4*4*4=192$ possible constellations. If these 192 pedigrees are considered equally likely prior to the introduction of DNA-data, we assign a *prior probability* of $1/192=0.0052$ to each. A different prior model is reasonable if the adults are of similar age so that one may not be parent of the other. The number of possible constellations is then reduced to 64 ($=1*4*4*4$) and the prior probability is in this case $1/64$. Finally, a paternity case may be viewed as the particular case where there are only two possibilities depending on whether or not the adult male is the father of the children or not. We may alternatively use the prior model (2.1) to modify the flat prior distributions assigned to the 4536 possible pedigrees. The inbreeding option with parameter 0 reduces the number of admissible cases to 2020. In other words, more than half of the initial cases involve cases where parents are related within the pedigree. The promiscuity option with 0 parameter reduces the number further to 817. Limiting the number of generations to 2 and 3 reduces this figure to 133 and 241 respectively.

2.3. Posterior model

The DNA for each locus for all persons announced to have such data in the prior model as well as a mutation rate for each system is used to compute the likelihood of all pedigrees considered in the prior model. These likelihoods are multiplied with the priors to obtain the posterior probability. All details related to the computation of the likelihoods are provided in [3].

2.4. Terminology

The odds form of Bayes's theorem may be written ([5], p. 17)

$$\frac{\Pr(H_p | E, I)}{\Pr(H_d | E, I)} = \frac{\Pr(E | H_p, I)}{\Pr(E | H_d, I)} \times \frac{\Pr(H_p | I)}{\Pr(H_d | I)},$$

where H_p and H_d are the prosecution and defence hypotheses, I some conditioning information, like age, and E additional information, typically DNA evidence. The conventional usage of the term *odds* implies that $\Pr(H_p | \cdot) = 1 - \Pr(H_d | \cdot)$, where the dot indicates some general conditioning. However, the theorem

is true also if this last relation does not apply. In such cases we will use the term *posterior probability ratio* (PPR) rather than posterior odds. The analogue of the usual informal version of the theorem then reads

$$\text{Posterior probability ratio} = \text{Likelihood ratio} \times \text{prior probability ratio}$$

and so the PPR coincides with the LR whenever a flat prior is used.

3. Results

Example 1b (Example 1a continued). Consider a group of five persons, three small girls and two adults, one male and one female. If we assume that all persons are heterozygous with alleles A and B in a system and that the two adults are parents of the three children the likelihood of the data becomes $\frac{1}{2}\frac{1}{2}\frac{1}{2}2p_Ap_B2p_Ap_B$ which equals 0.000183 for allele frequencies of $p_A = 0.087$ and $p_B = 0.22$ for A and B respectively. The numbers are chosen to agree with the system HUMVWA in the Romanov case discussed below. In this simple case analytical computation was possible; in more complex cases or if mutations are involved, we resort to numerical answers from **familias**. The LR comparing the alternatives H_1 : “The two adults are parents of the children” and H_2 : “The female and a random man are parents of the children” obtained by dividing the likelihood assuming H_1 (given above) by the likelihood assuming H_2 (details omitted) equals

$$LR = \frac{4}{(p_A + p_B)(1 + 3(p_A + p_B))} = 6.8. \quad (3.1)$$

familias computes the likelihoods 0.000183 and 0.000027 and the posterior probability corresponding to H_1 becomes

$$\frac{0.5 * 0.000183}{0.5 * 0.000183 + 0.5 * 0.000027} = 0.872$$

while the posterior probability assuming H_2 is calculated similarly resulting in 0.128. The posterior odds becomes $0.872/0.128 = 6.8$ coinciding with the analytical expression (3.1) as it should.

Example 2 (Example 1b continued). By now the identification of the Romanovs, i.e., Tsar Nicholay II, Tsarina and three of their five children is well established. DNA analysis played an important role as documented in [6] and subsequent papers including [7]. The identification used STR (Short Tandem Repeat) analysis to determine the relations between the 9 bodies (the mentioned Romanovs, three servants and a doctor) found in the grave in Ekaterinburg 1991 and mtDNA to demonstrate that the royal family had been found by comparing mtDNA from known relatives like Prince Philip, Duke of Edinburgh. Regarding the STR analysis, Gill et al. [6] write (p. 133) “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”. Such a paper has to our knowledge not appeared. The data of the Romanov case in Table 1 is a reduced version of a Table 1 in [6].

We pretend that the case is not resolved (and hence that the parenthesized information of Table 1 giving the family relations are unknown to us) and ask for the most probable family relation between the five persons assuming only that persons 4 and 7 are male and female respectively. Recall that there is a total of 4536 family constellations. Introducing the DNA data, we use a Norwegian database which may or may not be appropriate for nobility. However, we emphasize that our calculations primarily serve as an exercise ¹. It turns out that there are 1439 constellations compatible with the data and that **familias** singles out the family indicated in Table 1 as the one with the highest posterior probability (0.186). The two closest competitors have posterior probabilities of 0.093. These alternatives arise if there is a parent child relation between persons 3 and 5 of Table 1. If we are willing to assume that 3, 5 and 6 are children, and consequently a priori rule out the aforementioned alternatives, the same conclusion is reached and the posterior probability ratio (PPR) comparing the two most probable alternatives, is increased to 61.6. The accepted conclusion is reached in this case without modifying priors using the options *inbreeding*, *promiscuity* and *generations* discussed in Section 2.2. Even so, we tried the three with all parameters equal to 0.1 arriving at the same result, this time with a PPR of 0.991/0.00161=615.9.

Example 3. The definition of incest have changed in history and differs from country to country. We will consider an example which involves illegal relations by most definitions. A woman (M) gives birth to a child (C). The father of M, AF, is suspected to be the father of C, but there may be other explanations, incestuous or not. DNA data is available only for the three mentioned persons. Consider the hypotheses

H_1 : AF, the father of M, is the father of C.

H_2 : TF, a brother of M and son of AF, is the father of C.

Assuming that M, C and AF have genotypes A_1A_3 , A_1A_2 and A_1A_2 respectively, the maternal and paternal alleles of C can be deduced without ambiguity and the likelihood ratio is ([5], p. 176),

$$LR = \frac{2}{1 + p_2}, \quad (3.2)$$

where p_2 is the frequency of the paternal allele. The *disadvantage* of using **familias**, is that we obtain no analytical formulae. The *advantage* is that more complex situations involving other genotypes, mutations

¹The system HUMACTBP2 is not included since we lack a reference database appropriately standardized. For allele 10 of Table 1 we have used the allele frequency for 9.3 since 10 is extremely rare and some or all of the 10s of Table 1 would today be typed as 9.3. Note that both of the above decisions are conservative in the sense that the evidence in favor of the now accepted conclusion is underestimated.

and many alternatives may be handled. The following 3 alternatives have been selected in addition to the two above, see Figure 2:

3. A random man is the father of the child,
4. a half-brother of M is the father,
5. a half-brother of C is the father.

A priori the alternatives are assumed equally likely and we do not question the relations between AF and M and M and C. We consider a system with 20 equally likely alleles, denoted 1,2, ..., 20. To be specific assume first the data is as in the middle column of Table 2.

Using **familias** we got the following posterior probabilities for the five alternatives: 0.4494, 0.2360, 0.2472, 0.0449 and 0.0225. The PPR comparing alternatives 1 and 2 is $0.4494/0.2360=1.905$, coinciding as it should with the result obtained by using Equation (3.2): $2/(1+0.05)=1.905$. There are other genotypes consistent with H_1 . We simulated 100 such allele combinations. For 76 cases, the correct answer emerged as the most probable, while for the remaining simulations Alternative 5 proved to be the most probable. For the alleles corresponding to the rightmost column of Table 2, three alternatives appear to be approximately equiprobable. The posterior probabilities for the five alternatives come out as: 0.2532, 0.2658, 0.1519, 0.0606 and 0.2785. This result underscores the importance of considering many loci. Based on the 100 simulations, the estimated mean posterior probabilities for the 5 alternatives are 0.3790, 0.0426, 0.2108, 0.2383 and 0.1293. The corresponding numbers for the correct alternative based on these numbers increase to 0.87 and 0.99 using 5 and 10 independent loci with similar allele distributions.

Example 4. Four dead bodies were found following a fire in Norway. The persons were believed to be a mother (called 943 in Figure 1), two daughters (941, 944) and a relative (942). DNA data from the 9 numbered persons in Figure 1 was obtained. We use data from 5 systems (HUMACTBP2, HUMTHO11, HUMFES, HUMVWA and HUMF13A1). There is an enormous number of families that can be constructed from the 11 persons. We have considered 16: the one in Figure 1, the four alternatives obtained by replacing one of the numbered persons by a random, unrelated person, the six constellations appearing if a pair, say 941 and 942 are replaced similarly by random persons, the four constellations achieved by removing the four triplets and finally the case were bodies 941, 942, 943 and 944 are all unrelated random persons. Other possibilities, like interchanging 941 and 942, lead to 0 probability (mutations are disregarded in this example) and are not considered. We extend on the discussion of the various possibilities to consider in the next section. Assigning a uniform prior of $1/16$ to all alternatives the accepted solution represented by

Figure 1 is chosen since it is given the highest posterior probability. This is consistent with all other evidence of the case. The PPR of this alternative to the second most probable is $0.9987/0.00118=850$ while the PPR compared to the least likely alternative, the dead persons being unrelated to all other, is enormously large, $1.5 * 10^{16}$.

In some cases, there may be insufficient information to resolve family relations with acceptable certainty. This may be the case when there is DNA-data only from two persons. Below we outline two cases; one where the analysis points to a clearly most probable pedigree *among the ones considered* and one where no conclusion is reached. Obviously, potential problems could be solved provided we had access to more DNA-data.

Example 5. We evaluate the likelihood of possible family relations between the female F1 and the male M1. F1 is believed to be 20 years older than M1. We base our analysis on 13 loci and assume for simplicity allele frequencies of 0.05. In this example we also include the possibility for mutations following the model described in [3] and assume mutation rates of 0.005. The DNA-data may be summarized as in Table 3.

The individuals share both alleles in 4 loci, one in 6 and none in three. The likelihood of a mother-child relation must be small, but it will not be 0 since mutations are accounted for in the calculations. To accommodate a reasonable amount of close family relations, we included three additional persons. **familias** was used to evaluate 5 a priori equiprobable alternatives. The pedigrees are followed by the posterior probability below:

- M1 and F1 are siblings: 0.9974
- M1 and F1 are half siblings: 0.00133
- F1 is the aunt of M1: 0.00125
- M1 is the mother of F1: 3.1416e-009
- M1 and F1 are unrelated: 1.0246e-008

For instance, the PPR comparing the first two alternatives amounts to $0.9974/0.00132=750$ while the two last alternatives are extremely unlikely. If we alternatively let **familias** generate all possible relations between 5 persons and remove incestuous cases, the program singles out the sibling-case as the clearly most probable with the same PPR's; the posterior probabilities will obviously differ.

Example 6. Assume we are asked to determine the most probable family relation between the males NN and NN1 of ages 1 and 28 respectively. The individuals share both alleles in one system; one in five loci and none in the last three. Table 4 provides complete information.

Consider the following alternatives:

1a NN1 is the half-brother of the father or mother of NN.

1b NN1 and NN are cousins.

2a NN1 is the uncle of NN.

2b NN1 is the grandfather of NN.

3a NN1 and NN are unrelated.

3b NN1 and NN are siblings

3c NN1 and NN are half-siblings.

Alternatives 1a and b are equally probable and slightly more probable than the equiprobable alternatives 2a and b which in turn are more probable than the equally plausible alternatives 3a, 3b and 3c. We have also considered all families that can be made based on 2 females, 3 males and a child. Among these there are none with higher posterior probabilities than 1a and 1b. There are some equally probable pedigrees, but these are ruled out because they are improbable for other reasons. For instance, NN1 may not be the great-grandfather of NN considering that they are separated in age by 27 years. Moreover, the age information indicates that 1a is more likely than 1b and that 2a is more plausible than 2b. Summing up, the analysis is of limited conclusive value since the PPR comparing alternatives 1 and 3 is only 1.5. (However, the original question of paternity could be resolved based on the above data.)

4. Discussion

We discuss some remaining issues addressing first the complexity of the problems that may be taken care of by **familias**. In some cases, e.g., the Romanov example, a brute force approach is possible. All family relations may be evaluated and the approach concludes with the most probable pedigree without using prior information or modification of the flat priors. In other cases, as in Example 4, generation of all possible family relations between 11 persons exhausts what is currently possible. In the paper we restricted attention

to 16 alternatives, obviously we could have considered, say, some thousand alternatives as in the Romanov case. However, there remains work to be done to handle this and larger examples in a more optimal way. One larger example is described in [9] which involved the identification of 141 individuals. Relatives of persons believed to have died in the disaster provided blood samples and DNA-profiles were obtained. A program that takes the DNA-data from the deceased and their relatives as input and outputs the most probable family relations or confirms what is believed, is well beyond what **familias** presently can handle in cases of this magnitude. This and similar problems would have to be broken down to many smaller identification cases, possible using statistical clustering methods, to be within the reach of **familias**. An extension of **familias** to handle problems of this size would be of obvious interest. A partial solution would be to extend the possibilities to constrain the generation of possible pedigrees, i.e., certain family relations can be assumed fixed. For instance, one could avoid questioning accepted family relations between living relatives of persons believed to have died in a disaster.

The prior (2.1) requires parameters. We have not addressed the problem of assessing these parameter beyond providing some simple ad-hoc rules. For instance the values 0, between 0 and 1 and above 1 respectively removes incestuous pedigrees, decrease and increase their likelihood. Further studies could be carried out to indicate reasonable ways of assigning values to a parameter or to find better parametrizations.

Our examples do not include mitochondrial and Y-chromosome DNA. We believe that the approach we have described may be extended to be useful for such data as well. Kinship is another issue that has not been discussed in this paper. We are working on extensions in this direction following the recommendations of [1].

familias gives the user the possibility to determine the most likely family relations between a number of persons. This should be done with care, particularly if nothing is known but the DNA-data. The suggested prior may be used to disregard some unreasonable alternatives, but not all. When hypotheses are not specified on beforehand, there will always be a problem of interpreting results that appear to favor certain explanations, particularly when a large number of explanations have been investigated. This cautionary remark is relevant in our setting as it is generally when multiple comparisons are involved. We emphasize that a program like ours normally only constitute one part in a complex process of identification.

In conclusion, the examples show that useful statements may be derived from the method we have described. Moreover, the strength of the conclusion is quantified by posterior probabilities and posterior probability ratios and depends on the DNA data and the pedigree as well possible prior (non-DNA) data that is used. This last point was clearly demonstrated by the Romanov case: the correct or accepted answer is obtained based solely on DNA data, but more convincing evidence is obtained if unreasonable possibilities are down weighted or removed.

References

- [1] *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press, 1996.
- [2] C. Brenner. Symbolic kinship program. *Genetics*, 145:535–542, 1997.
- [3] T. Egeland, P. Mostad, and B. Olaisen. Computerized probability assessments of family relations. *Science and Justice*, pages 269–275, 1997.
- [4] E. Essen-Möller. Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis. Theoretische Grundlagen. *Mitt. Anthropol. Ges (Wien)*, 68:9–53, 1938.
- [5] I. Evett and B. Weir. *Interpreting DNA evidence*. Sinauer, MA, USA, 1998.
- [6] P. Gill, P. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, I. Evett, E. Hagelberg, and K. Sullivan. Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics*, 6:130–135, 1994.
- [7] P. Ivanov, M. Wadhams, R. Roby, M. Holland, V. Weedn, and T. Parsons. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genetics*, 12:417–420, 1996.
- [8] A. Jeffreys, J. Brookfield, and R. Semeonoff. Positive identification of an immigration test-case using DNA fingerprints. *Nature*, 317(31):818–819, 1985.
- [9] B. Olaisen, M. Stenersen, and B. Mevåg. Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster. *Nat Genet*, 15, 1997.
- [10] M. Sjerps and A. Kloosterman. On the consequences of DNA profile mismatches for close relatives of an excluded suspect. *Int J Legal Med*, 112:176–180, 1999.

Skeleton	HUMVWA/31	HUMTH01	HUMF13A1	HUMFES
3 (child)	15,16	8,10	5,7	12,13
4 (Tsar)	15,16	7,10	7,7	12,13
5 (child)	15,16	7,8	5,7	12,13
6 (child)	15,16	8,10	3,7	12,13
7 (Tsarina)	15,16	8,8	3,5	12,13

Table 1. STR genotypes for the nine skeletons [6].

Person	Alleles, version 1	Alleles, version 2
AF	1,2	1,2
M	1,3	2,3
C	1,2	2,3

Table 2 Samples of 2 simulations, denoted version 1 and 2, assuming H_1 : AF, the father of M, is the father of C.

Locus	1	2	3	4	5	6	7	8	9	10	11	12	13
F1	ab	ac	bc	ab	ab	ac	ab	ab	ab	ac	ab	ac	aa
M1	ab	bb	ac	ab	bc	bc	ab	bc	bc	bb	ab	bb	ab

Table 3 The alleles of the female F1 and male M1 are shown for 13 loci.

Locus	NN	NN1
D3S1358	17 18	16 17
HUMVWA	15 16	15 16
FGA	24 25	21 23
D8S1179	10 12	13 15
D21S11	30 30	30 31
D18S51	16 17	14 16
D13S317	10 12	12 12
D7S820	8 12	10 10

Table 4. Data for person NN and NN1 is shown.

5. Legends

Figure 1: Four dead bodies were found following a fire in Norway. The persons were believed to be a mother (called 943 in the figure), two daughters (941, 944) and a relative (942). DNA data from the nine numbered persons in Figure 1 was obtained.

Figure 2: A woman (M) gives birth to a child (C). The father of M, AF, is suspected to be the father of C (Alternative 1 in the figure), but there may be other explanations (Alternatives 2-5) , incestuous or not. DNA data is available only for the three mentioned persons.