



Assessing relationships in an ancient skeletal collection by the number of alleles shared identical by state among pairs of individuals

S. Presciuttini ^{a,*}, B. Bramanti ^b, S. Hummel ^b, B. Herrmann ^b

^a*Department of Biomedicine, University of Pisa, c/o Centro Retrovirusus, S.S. Abetone e Brennero 2, 56127, Pisa, Italy*

^b*Institute of Zoology and Anthropology, University of Göttingen, Göttingen, Germany*

Abstract

A sample of 26 skeletal remains from an 18th century graveyard were typed with the Profiler Plus™ kit. To assess the possible relationships existing among these individuals, we applied a method based on the probabilities of sharing 0, 1 or 2 alleles identical by state (IBS) per locus among all possible pairs. We identified with high confidence ($P > 99\%$) a trio of two brothers with a child and a second pair of two siblings.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Relationship; Skeletal; IBS

1. Introduction

In a previous study [1], we described a collection of skeletons from an 18th century burial ground at Goslar, Germany, and showed genotype results of 26 individuals typed for up to 9 STRs with the Profiler Plus™ kit, plus amelogenin. We used stringent criteria in assigning genotypes (each allele had to be identified at least twice in at least four amplifications, generally from two different DNA extracts). No significant differences between allele and genotype frequency distribution were found in our sample compared with present-day German speaking populations; however, the level of allele sharing among some of the skeletons suggested that the sample could include some close relatives. In the present work, we reworked at the genotype database and addressed the problem of

* Corresponding author. Tel.: +39-050-555234; fax: +39-050-560276.

E-mail address: sprex@biomed.unipi.it (S. Presciuttini).

inferring the relationship between pairs of individuals using a novel approach. This is based on the probabilities of sharing 0, 1 or 2 alleles identical by state (IBS) at a number of loci, and is conceptually and computationally easier and more robust than the conventional method based on the population frequencies of the observed alleles.

2. Method of analysis

The first step of our method was to obtain the probabilities z_0 , z_1 and z_2 of a pair of individuals sharing 0, 1 or 2 alleles IBS for each of the typed loci, and for each of the following five relationships: (1) nonrelatives (NR); (2) first cousins (FC); (3) second degree: grandparent/grandson, uncle–aunt/niece–nephew, half-sibs, here collectively called HS; (4) full sibs (FS); (5) parent/child (PC). We simulated 100,000 pairs for each relationship, based on the weighted mean allele frequencies in three present-day German-speaking populations, and obtained the z_i values as the proportions of pairs sharing 0, 1 or 2 alleles.

The second step was to pair each individual of the sample with all others, to determine the number of shared alleles at each locus, and to substitute in each pair the value of the observed number of shared alleles at each locus with the corresponding estimated probability. Probabilities were then multiplied across loci, separately for the five relationships. Thus, the final value was the probability of observing the particular configuration of shared alleles between two individuals, given each specified relationship.

3. Results

Because of amplification failures, some of the samples had incomplete genotypes; 7 samples (30%) had all 9 genotypes, and the other 11 (50%) had 7 or 8 genotypes. We paired all individuals from our database with all others, and selected the pairs with at least five common typed markers (including 22 out of 26 examined individuals, for a total of 227 pairs). For each pair, we computed four likelihood ratios (LR), using the obtained PC, FS, HS and FC probabilities in the numerator, respectively, and the NR probability in the

Table 1

Topmost LRs that the specified pairs (first column) are first-degree relatives rather than unrelated, sorted in decreasing order of the largest among the two FS and PC LR values (boldface)

Pair	Likelihood ratios				Nominal <i>P</i>
	FC/NR	HS/NR	FS/NR	PC/NR	
21–24	16.7	135.6	10,468.3	2822.7	0.9999
3–23	5.8	20.4	490.1	117.1	0.9980
19–24	7.8	38.2	72.6	458.1	0.9978
12–17	2.8	3.8	28.5	0.0	0.9661
18–25	2.8	6.3	9.9	21.5	0.9555
10–25	4.1	10.0	20.7	0.0	0.9540
16–20	2.6	5.8	6.5	19.4	0.9511

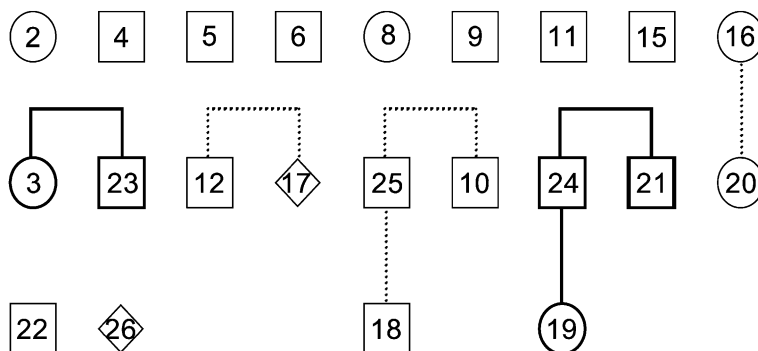


Fig. 1. First-degree relationships identified among 22 skeletal remains. Square: male; circle: female; diamond: sex uncertain; unbroken lines: relationships established at the 99% significance level; broken lines: more uncertain relationships.

denominator. We arranged the pairs in decreasing order of the largest between the two PC and FS LR values, thus selecting the pairs with the highest evidence of a first-degree relationship. Table 1 shows the results for the first seven pairs, chosen for having a nominal P value $>95\%$.

The nominal P values were corrected for the multiple comparisons by generating 100 exact replicates of our database, in which each individual genotype was randomly chosen from the general population, and applying to each replicate our method of calculation. In this way, we obtained the empirical distribution of the first-degree vs. nonrelative LR among truly unrelated individuals. The three top ranking pairs exceeded the 99% corrected significance level, and we accepted them as being first-degree relatives. The best evidence indicated that pair 21–24 were full sibs, pair 3–23 were full sibs, and pair 19–24 were parent–child. Since individual 24 was a member of two of these pairs, we looked at the LRs of pair 19–21. This pair also ranked high for an LR, suggestive of biological relationship. Its maximum LR was 7.9 for the HS relationship (second-degree). Thus, the pedigree of maximum likelihood for this trio of skeletons implied that 21 and 24 were brothers and 19 was a child of 24. Ten other pairs with LRs included between 10 and 30 (4.3% of our data) were included in the top 2.5% tail of the empirical distribution. This suggests that some of them could also be first-degree pairs, though the evidence must be considered tentative at the present stage. Nevertheless, it is interesting to note that they included a second trio of two full sibs with a child (10–25 and 25–18, respectively, see Table 1); even in this case, the LR of 10–18 (not shown) supported this relationship. Fig. 1 summarizes our results.

4. Discussion

The present work shows that with a relatively low number of markers (the 9 STRs included in the Profiler Plus™ kit) and a simple method based on IBS allele sharing, it is possible to approach the problem of assessing the relationships existing among samples of

unknown relatedness. Other evidence, either genetic in nature or nongenetic, may then be used to corroborate the obtained results. For example, although the maximum likelihood pedigree structure of our accepted trio was in favor of two brothers with a child, a possible alternative was that they were a grandfather with a child and a grandchild. The likelihood of this second possibility was only about three times lower than the first. Use of mitochondrial DNA could discriminate between these two possibilities.

In the present paper, we evaluated the probabilities z_0 , z_1 and z_2 for each relationship and for each locus by computer simulations. However, it is possible (though tedious) to obtain exact values via full enumeration of the relevant offspring of all possible genotype pairings in a population. On the other hand, it is easy to show that in the PC case, these probabilities are linear functions of H , the locus heterozygosity; specifically, $z_0=0$, $z_1=H$, and $z_2=1-H$, respectively. This suggests that the z_i values generally depend on the heterozygosity of a locus rather than on the number of alleles and their frequency distribution. We are working at this issue.

The functional dependence of the z_i values on H represents one of the major advantages of our method with respect to the conventional approach. This is based on the population frequency of the particular alleles observed in a given pair of individuals, and it is thus highly dependent on the correct estimation of allele frequencies, which are subject to much larger standard errors than H . In addition, the conventional method requires application of different equations, depending on the particular configuration of the alleles shared by a given pair. Furthermore, preliminary evidence suggests that the IBS method is more powerful when H is higher than 0.75. In conclusion, when highly variable STR are used for genotyping, the IBS method appears to be computationally easier, more robust and more powerful than the conventional approach in assessing relationships among pairs of individuals of unknown relatedness.

Reference

- [1] B. Bramanti, S. Hummel, T. Schultes, B. Herrmann, STR allelic frequencies in a German skeleton collection, *Anthropol. Anz.* 58 (1) (2000) 51–59.