ORIGINAL ARTICLE

The limitations of kinship determinations using STR data in ill-defined populations

Vincent Zvénigorosky[1,2] · Audrey Sabbagh[3] · Angéla Gonzalez[2] · Jean-Luc Fausser[2] · Friso Palstra[3] ·
Georgii Romanov[4,5] · Aisen Solovyev[4,6] · Nikolay Barashkov[4,5] · Sardana Fedorova[4,5] · Éric Crubézy[7] ·
Bertrand Ludes[1] · Christine Keyser[1,2]

Abstract

The likelihood ratio (LR) method is commonly used to determine kinship in civil, criminal, or forensic cases. For the past 15 years, our research group has also applied LR to ancient STR data and obtained kinship results for collections of graves or necropolises. Although we were able to reconstruct large genealogies, some pairs of individuals showed ambiguous results. Second-degree relationships, half-sibling pairs for example, were often inconsistent with detected first-degree relationships, such as parent/child or brother/sister pairs. We therefore set about providing empirical estimations of the error rates for the LR method in living populations with STR allelic diversities comparable to that of the ancient populations we had previously studied. We collected biological samples in the field in North-Eastern Siberia and West Africa and studied more than 800 pairs of STR profiles from individuals with known relationships. Because commercial STR panels were constructed for specific regions (namely Europe and North America), their allelic makeup showed a significant deficit in diversity when compared to European populations, replicating a situation often faced in ancient DNA studies. We assessed the capacity of the LR method to confirm known relationships (effectiveness) and its capacity to detect those relationships (reliability). Concerns over the effectiveness of LR determinations are mostly an issue in forensic studies, while the reliability of the detection of kinship is an issue for the study of

✉ Vincent Zvénigorosky
z.vincent@live.fr

Audrey Sabbagh
audrey.sabbagh@ird.fr

Angéla Gonzalez
agonzalezmartin@unistra.fr

Jean-Luc Fausser
fausser@unistra.fr

Friso Palstra
f_palstra@yahoo.com

Georgii Romanov
gpromanov@gmail.com

Aisen Solovyev
nelloann@mail.ru

Nikolay Barashkov
barashkov2004@mail.ru

Sardana Fedorova
sardaanafedorova@mail.ru

Éric Crubézy
crubezy.eric@free.fr

Bertrand Ludes
bertrand.ludes@orange.fr

Christine Keyser
ckeyser@unistra.fr

[1] CNRS FRE 2029-BABEL,, Paris Descartes University, Paris, France

[2] Strasbourg Institute of Legal Medicine, Strasbourg, France

[3] UMR 261 MERIT, IRD, Paris Descartes University, Paris, France

[4] Laboratory of Molecular Biology, North-Eastern Federal University, Yakutsk, Sakha Republic, Russia

[5] Laboratory of Molecular Genetics, Yakut Science-Centre of Complex Medical Problems, Yakutsk, Sakha Republic, Russia

[6] Institute for Humanitarian Studies and Problems of Indigenous Peoples of the North, Yakutsk, Sakha Republic, Russia

[7] CNRS UMR 5288 AMIS, Toulouse Paul Sabatier University, Toulouse, France

necropolises or other large gatherings of unidentified individuals, such as disaster victims or mass graves. We show that the application of LR to both test populations highlights specific issues (both false positives and false negatives) that prevent the confirmation of second-degree kinship or even full siblingship in small populations. Up to 29% of detected full sibling relationships were either overestimated half-sibling relationships or underestimated parent-offspring relationships. The error rate for detected half-sibling relationships was even higher, reaching 41%. Only parent-offspring pairs were reliably detected or confirmed. This implies that, in populations that are small, ill-defined, or for which the STR loci analyzed are inappropriate, an examiner might not be able to distinguish a pair of full siblings from a pair of half-siblings. Furthermore, half-sibling pairs might be overlooked altogether, an issue that is exacerbated by the common confusion, in many languages and cultures, between half-siblings and full siblings. Consequently, in the study of ancient populations, human remains of unknown origins, or poorly surveyed modern populations, we recommend a conservative approach to kinship determined by LR. Next-generation sequencing data should be used when possible, but the costs and technology involved might be prohibitive. Therefore, in potentially contentious situations or cases lacking sufficient external information, uniparental markers should be analyzed: ideally, complete mitochondrial genomes and Y-chromosome haplotypes (STR, SNP, and/or sequencing).

Keywords Short tandem repeats · Kinship · Likelihood ratios · Ancient DNA · Forensic genetics

Introduction

Since the earliest studies of ancient human populations, it has been evident that the identification of genealogical links between the individuals discovered in funerary ensembles is an essential clue to their funerary, social, and cultural practices. In preliterate societies, or those that do not mark graves with the names of their occupants, or even those that use perishable materials for such markings, only the methods and techniques of anthropobiology can detect kinship between individuals.

At first, skeletal characteristics were used to propose the existence of biological links between individuals excavated in archeological digs, but the rise of molecular genetics was immediately recognized as a solution to the detection of kinship. Diverse markers were identified and analyzed in the light of transmission probabilities that had been described in the nineteenth century and the early twentieth century. This eventually led to the constitution of STR (short tandem repeats) panels that permitted the construction of large national and international databases, to be used in criminal or contentious cases, mainly to identify an individual or confirm paternity. To fulfill these roles, a Bayesian method was devised [1], which permitted the determination of kinship probabilities from STR profiles and their comparison with other available evidence through the calculation of likelihood ratios (LRs).

It is a necessity of their design that panels and databases are representative of the populations from and for which they were built. The LR method is very dependent on the allelic frequencies attached to each marker. Consequently, using it assumes the reliability of these frequencies or, more accurately, the representativity of the reference population used to compute these frequencies. This is an evident limitation, as even some modern populations are not well-represented in reference data. Ancient populations are often too far removed from modern populations for the LR method to provide

meaningful results when applying allelic frequencies from one to the other. It must be implemented using parameters computed on the ancient data itself (small numbers of individuals) or populations that are closely related (the strength of that relation being ultimately unknown).

Our team, as well as a number of others, has used likelihood ratios applied to STR data and produced satisfactory results within studies of collections of graves or large necropolises [2, 3]. Before the advent of next-generation sequencing (NGS), this approach was the most cost-effective and it yielded information that was highly valuable to archeologists. It also allowed research groups to procure data for the large majority of the individuals analyzed, contrary to what was then presented for whole genomes, with many samples having to be rejected before the analysis because of the poor quality or quantity of the DNA extracted.

Although we were able to construct genealogies comprising several individuals, it quickly appeared that some pairs showed ambiguous results [2, 4]. As could be expected given the variations in proportions of shared alleles within different levels of kinship [5]: second-degree relationships, half-sibling pairs for example, were often inconsistent with detected first-degree relationships, such as parent/child or brother/sister pairs. It was known that the theoretical precision of LR would decrease with the genetic distance between individuals, but it could not be quantified in the context of ancient DNA studies, especially since most modern work in criminal or paternity cases did not appear to suffer from the same limitations. There are however notable exceptions, such as the identification of shipwrecked migrants in Lampedusa [6], an actual case of genetic kinship determination in an ill-defined modern population.

The estimation of the reliability of LR in detecting kinship necessitated the study of a large number of related pairs with known biological relationships of different

degrees, which had to be collected among isolated populations whose genetic proximity with the modern references that are Europe and North America would be comparable to that of ancient populations. We therefore organized an expedition to North-Eastern Siberia (the Sakha Republic), a region inhabited by the supposed descendants of an ancient population (the Yakuts) we had already studied [2]. We spent several weeks in two localities and endeavored to construct the genealogies of all inhabitants [7], before collecting buccal swabs, with written informed consent (in Russian and Yakut), for all volunteers and their children. To these, we added data from individuals that had been previously studied and analyzed in two other villages from the same region [4]. In order to evaluate the method in two genetically distant and differently structured populations, we also gathered data (FTA cards) from rural villages in the Republic of Benin, in West Africa. Individual written informed consent was obtained from adults for themselves or their children, using a consent form translated in their native language (Ditamari, Bariba or Fulani). We analyzed the same 21 autosomal STR loci for all individuals.

We applied the LR method systematically to all pairs of individuals and quantified two metrics: effectiveness and reliability. Effectiveness is defined here as the capacity of the method to avoid false negatives (undetected known relationships) and reliability is defined as the capacity of the method to avoid false positives (detected relationships known to be false). We propose guidelines regarding the use and interpretation of LR on STR data in small, remote, or ancient populations, especially regarding the determination of genetic kinship and the constitution of genealogies.

Materials and methods

Summary of samples

The Yakuts constitute the majority ethnic group in the Sakha Republic (Russian Federation), in North-Eastern Siberia, which is also called Yakutia. The modern genetic pool of this population is the result of an influx of Altaic semi-nomadic groups that arrived in the region no sooner than the thirteenth century, and their subsequent admixture with local Siberian populations. Population density is extremely low (less than $2/km^2$ outside the capital and as low as $0.04/km^2$ in one of the regions surveyed) and the overwhelming majority of the participants self-identified as members of the Yakut ethnic group, although extended genealogies revealed occasional mixed marriages with Russians or Tunguses (Siberians). The total population of the Sakha Republic is about 1 million, while the Yakut population is about 500,000. Many archeological populations share the structure of the modern

Yakut population: very low density, complex networks of intermarriage and ethnic homogeneity.

We gathered familial information and collected biological samples for 311 modern Yakuts and implemented STR analyses on 219 of them, presenting 304 known biological relationships and 82 civil relationships with no supposed genetic link (Table 1). To this number, we added 64 unrelated modern Yakuts from previous studies, for a total number of 283 individuals. These individuals resided in one village in Central Yakutia (Churapchinsky district) and three villages in Northern Yakutia (Momsky and Oymyakonsky districts). All villages are situated at least 100 km away from the capital (Yakutsk) and are predominantly inhabited by Yakuts, rather than Russians or members of Siberian ethnic minorities. Complete genealogies were built through repeated interviews and the confrontation of conflicting information.

The Republic of Benin gained independence from France in 1960, with borders defined in 1894 by the French colonial empire, crossing at least three well-established ancient kingdoms. Ethnic diversity is high, and the samples presented originate from three broad groups: the Bariba, the Fulani (or Fula), and the Gando (themselves partly considered to be a subgroup of the Fulani). Population density in Benin is comparable to some European or other African countries, at around $100/km^2$, but the regions surveyed are more sparsely populated, with less than $30/km^2$. There again, many necropolises present the same structure as the modern Beninese population: diverse ethnicities, unknown rates of interethnic marriage, and complex matrimonial practices, such as polygamy.

We produced and analyzed STR data for 305 Beninese individuals, corresponding to 310 biological relationships and 144 civil relationships. Samples were collected in four rural villages, 40 km from the Natitingou city center, in the Birni district of the Atacora department, in Northwest Benin. Familial relationships were identified

Table 1 Declared relationships in each population

	Relationship category declared				
	PO	FS	HS/	CO/	U
Yakut population ($n = 283$)	116	59	105	54	82
Beninese population ($n = 305$)	169	56	65	20	144

PO parent-offspring pairs, *FS* full-siblings, *HS/...* half-siblings and same-level relationships (namely grandparent/grandchild and avuncular, or aunt-uncle/niece-nephew, relationships), *CO/...* cousins and same-level relationships, *CO2/...* second-cousins and same-level relationships, *U* unrelated

through interviews with members of the household and further confirmed by village authorities.

STR typing

In samples from the Yakut population, DNA was extracted from buccal swabs collected in the winter of 2016. All 219 swabs produced usable material. In samples from the Beninese population, DNA was extracted from FTA cards collected in November 2015 and also produced usable material for all 305 individuals.

All Yakut individuals were analyzed using the Qiagen Investigator 24plex QS1 kit at 21 autosomal STR loci (D8S1179, D7S820, D3S1358, D13S317, D16S539, D2S1338, D19S433, D5S818, D21S11, CSF1PO, vWA, THO1, TPOX, D18S51, FGA, D1S1656, D12S391, D2S44, D10S1248, D22S1045, and SE33) along with three gender identification markers (Amelogenin, DYS391 and a Y-indel). All Beninese individuals were analyzed using the GlobalFiler kit (Life Technologies™), which amplifies the same 21 STR loci and three gender identification markers. All autosomal STR profiles and known relationships are presented in Supplementary Tables S1a, S1b, S1c, and S1d.

The experimental conditions were those recommended by the manufacturer. All STR products were run on the 3100 or 3500 genetic analyzer (Life Technologies™) and analyzed using GeneMapper v. 4.1 (Life Technologies™).

Implementation of the LR method

Likelihood ratios are a probabilistic method predicated upon the computation of the probabilities of pairs of genotypes given different kinship hypotheses. With large numbers of individuals or markers, it must be implemented using programming languages and dedicated software. Results in this study were produced with the Familias Package for R [8] and the categories of kinship tested were the following: "PO" for parent-offspring pairs; "FS" for full-siblings, "HS/…" for half-siblings and same-level relationships (namely grandparent/grandchild and avuncular, or aunt-uncle/niece-nephew, relationships); "CO/…" for cousins and same-level relationships; "CO2/…" for second-cousins and same-level relationships; "U" for unrelated. All possible pairs were tested independently; we did not reconstruct trios (for example mother/father/child). Although trios and larger genealogies typically yield much better results, they cannot be used in archeological contexts and are often unavailable in some forensic contexts.

LR results are noted in the following manner: "LR-PO" is the ratio of the likelihood of PO and the likelihood of U. "Pr(PO)" is the posterior probability of a PO relationship for a given pair of individuals, if all possible kinship levels but PO

and U are excluded, prior probabilities being equal, $P(PO) = P(U) = 0.5$, in this Bayesian model.

Simulations

We performed the same LR tests on 10,000 pairs of simulated PO, FS, HS, CO, and U individuals, using frequencies for the 21 STR loci included in the empirical study, computed on the complete Beninese population, the complete Yakut population, and a collected European population from published data [9]. All simulations, kinship tests, and estimations of effectiveness and reliability were performed using the R programming language and the Familias Package for R [8].

Effectiveness and reliability

The result for a single pair of individuals counted positively toward global effectiveness if the previously known relationship (simulated or declared) was confirmed by LR. A pair of declared full siblings that resulted in a LR-FS superior to all other LR (LR-PO, LR-HS and LR-CO) therefore constituted a success, while any other result was considered a failure. Since LR-U cannot be computed (given the present method of testing all kinship levels against U), any pair of unrelated individuals was considered effectively confirmed if all LRs were inferior to 1, including especially LR-CO.

In the same way, results counted positively toward global reliability if the relationship detected in a pair of individuals was in fact the known relationship (simulated or declared). A pair where LR-FS was superior to all other LR was considered a success if the known relationship was in fact FS. Any other result (the known relationship being PO, HS, CO or U) was considered a failure.

Results

Allelic frequencies computed using unrelated individuals

The use of likelihood ratios is predicated upon the availability of reliable allelic frequencies, computed using a large population of unrelated individuals. While such allelic frequencies might be available when studying well-described geographic regions (notably Europe and North America), in some forensic cases and most archeological cases, frequencies must be computed using a limited dataset. In the case of the Yakut population presented here, the number of individuals is largely superior to what is presented in any previous study of this population. The Beninese dataset is also substantial. However, both sample sets include (by design) large numbers of related individuals.

We therefore implemented a method for removing the minimum number of related individuals from allelic frequencies computations. We did not use declared relationships as an indicator because some of them were likely mistakenly declared. We instead repeatedly computed all LRs using complete population allelic frequencies and eliminated the individuals implicated in first-degree (PO and FS) relationships, one-by-one, starting with those implicated in multiple relationships, until no first-degree relationship could be detected. This process removed 140 individuals from the Yakut population (143 were left) and 117 from the Beninese population (188 left). Even through this minimal elimination process, several alleles disappeared from the studied population. It was therefore necessary to reintroduce these alleles in the set of allelic frequencies. We set those frequency values at $1/2N$, where N is the size of the population.

Finally, we used a frequency set that is a compilation of all available allelic frequencies, from all populations (Autosomal Database (allstr.de), Qualitype GmbH, Germany), to test whether we would observe a significant drop in effectiveness or reliability. All LRs were computed for all pairs using the three allelic frequency sets: complete population ("Natural"), relationships eliminated ("Unrelated"), and compiled ("World").

Removal of irregular LR results

Some results were considered more likely to be the consequence of faulty declarations, rather than errors of the method. In order to not underestimate the success of the LR method for any level of relationship, we eliminated any result that was inferior to LR = 1. For example, a declared PO pair with a LR-PO of 0 was not considered a failure to detect or confirm PO and rather a probable error in the declaration, which would occur in the case of an (official or effective) adoption and the confusion between civil and biological kinship. FS and HS pairs were treated the same way, even though it is possible that some LR-HS would be inferior to 1 in a real HS pair. The results presented below are therefore the conservative estimates of the error rates, with actual error rates (for HS in particular) likely higher.

It should therefore be noted that 36 declared PO relationships had to be eliminated from the Beninese set because all LR values were inferior to 1. Since any LR-PO equal to zero is either a sign of an allelic exclusion caused by a mutation or the result of a faulty declaration (which is more likely), we did not take these into account to compute reliability or effectiveness. This was also the case for one PO pair in the Yakut population. We also removed pairs for the FS and HS levels (respectively 18 and 26 in the Beninese population and 4 and 10 in the Yakut population). We do not present results for CO pairs because most of them (19 out of 20) had to be removed from

the Beninese population since LR-CO values were all inferior to 1.

Given the removal of all ambiguous U calls, we cannot provide an estimation of the reliability of LR for unrelated pairs. Instead, the number of U calls will be considered, since its variations using different sets of allelic frequencies are informative regarding the reliability of kinship determinations. Although simulated relationships are known with certainty by definition, U calls for PO, FS, HS, and CO relationships were also removed from the final results, to allow for direct comparison between simulated and empirical results.

Effectiveness decreases with increased genetic distance

Declared PO relationships were verified in more than 99% of cases (Fig. 1, Supplementary Table S2), using any set of allelic frequencies. This corresponded to only one error in both the Yakut sample and the Beninese sample across frequency sets, except in the Beninese sample using world frequencies, where there were two errors. The issue occurs with the same pair in all cases, except in the Beninese sample using world frequencies. The situation (in all cases) is the following: the difference between LR-PO and LR-FS is close to zero, although LR-FS is slightly higher. Pr(PO) and Pr(FS) are both superior to 99.9999% and cannot be distinguished before the fifth decimal place. No declared PO relationship was underestimated further than the FS level.

Depending on the set of allelic frequencies used, declared FS relationships were confirmed in 81 to 86% of cases. It should be noted that neither the use of unrelated frequencies or world frequencies compared to the natural ones had a clearly negative or positive impact on the results from either population. In the Yakut sample, errors at the FS level were shared between overestimations into PO (between 5/8 and 5/9) and underestimations into HS (between 3/8 and 4/9). In the Beninese population, all 6 or 7 failures to confirm declared FS relationships (depending on the frequency set) were underestimations into HS. No declared FS relationship was underestimated further than the HS level.

In the same way, declared HS were confirmed in 63 to 80% of cases. There again, using different sets of allelic frequencies had no clear positive or negative effects. The use of world frequencies, for example, decreased effectiveness in the Beninese population while it increased it in the Yakut population. In the Yakut sample, errors at the HS level were also shared between overestimations (between 6/35 and 19/25, exclusively into FS) and underestimations (between 29/35 and 6/25, exclusively into CO). The same was true in the Beninese population (between 0/8 and 5/18 overestimations, between 8/8 and 13/18 underestimations). Because we eliminated all ambiguous cases, no declared HS was underestimated further than the CO level.

Fig. 1 Effectiveness of LR confirmations of declared relationships, measured as classification accuracy. YAK: Yakut, BEN: Beninese. Results are presented for three allelic frequency sets: complete population ("Natural"), relationships eliminated ("Unrelated"), and compiled ("World")
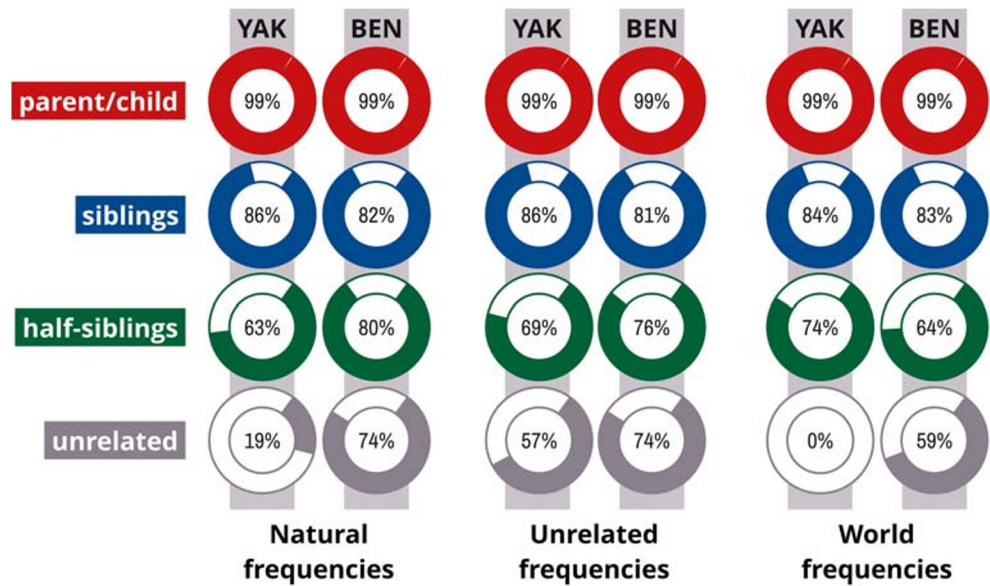
Most declared unrelated pairs in the Yakut population could not be confirmed by LR: between 43 and 100% were overestimated into CO (most errors) or HS (10% to 30% of errors). In the Beninese population, the error rate was lower (26% to 41%), with 7 to 15% being overestimations into HS and the rest (the majority), into CO.

Reliability decreases with increased genetic distance

In the Beninese population, all detected PO relationships ($n = $ 132) had been declared as such and were the same whatever the set of allelic frequencies used (Fig. 2, Supplementary Table S3). On the contrary, out of the 111 PO relationships detected in the Yakut population, 5 were overestimations of

FS relationships. Using different sets of allelic frequencies only had limited influence on LR values, and the 5 overestimated pairs were the same ones in every situation.

In the Yakut population, 54 FS relationships were detected, one being an underestimated PO pair and 6 overestimated HS pairs. The use of unrelated frequencies did not change the number or the nature of the errors, or which pairs were concerned. The use of world frequencies caused many new errors and left the previous errors unchanged. In this last case, out of 68 detected FS pairs, 19 were overestimations of HS pairs. In the Beninese population, the detection of FS pairs was overall more reliable, with only 1 pair out of 32 being an underestimation of a PO pair when using natural frequencies, and no HS overestimation. Unrelated frequencies, however, led to

Fig. 2 Reliability of LR detected kinships, measured as classification accuracy. YAK: Yakut, BEN: Beninese. Results are presented for three allelic frequency sets: complete population (Natural), relationships eliminated (Unrelated), and compiled (World)
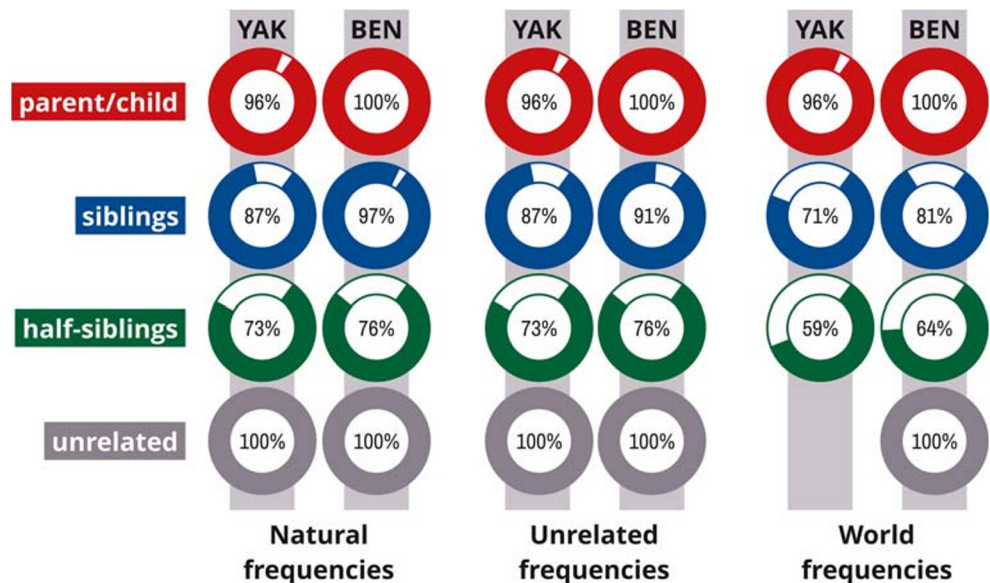
more errors, with two HS overestimations (3 errors out of 33 detected). Finally, world frequencies led to 5 new erroneous FS being detected (one declared PO and four declared HS pairs out of 37 detected FS pairs).

In the Yakut population, using the three different sets of frequencies, the numbers of detected HS pairs were 82, 85, and 123 for the natural, unrelated, and world frequencies, respectively. Errors included underestimations of FS pairs and overestimations of CO (cousins) or U (unrelated) pairs. There again, the use of natural or unrelated frequencies had little to no influence and the use of world frequencies led to the least reliable results. In the Beninese population, reliability was overall better, but it followed a very similar trend, with the numbers of HS pairs detected of 41, 41, and 50, respectively, and errors ranging from FS underestimation to U overestimation.

The results regarding detected unrelated pairs can appear very satisfactory (100% reliability in every case), but this result was caused by the elimination of ambiguous U calls, designed to avoid underestimating reliability for other kinship categories. Moreover, the number of detected pairs varied greatly. In the Yakut population, it was 4, 40, or 0, depending on the frequency set, while in the Beninese population, it remained 84, regardless of the frequency set used.

Simulated results replicate empirical results

The majority of empirical results showed no difference in proportions with simulated data (all chi-squared p values inferior to 0.05), for any set of frequencies for LR computations (natural population, unrelated individuals, world frequencies) or any basis for simulation (Beninese natural population, Yakut natural population, European population).

Notable exceptions in effectiveness results were (Supplementary Table S2): 0% effectiveness for CO relationships using world frequencies in the Yakut population and all empirical effectiveness results for unrelated (U) pairs lower than expected given any set of frequencies.

In reliability results (Supplementary Table S3): 100% reliability in PO relationships for the Beninese population is higher than expected; the reliability of FS calls in the Yakut population using world frequencies is much lower than expected with appropriate allelic frequencies; on the contrary, the reliability of HS calls is higher than expected. The reliability of CO calls is in all cases lower than expected, with minimums of 0% in both populations.

Lower allelic diversity predicts lower effectiveness and reliability

The comparison of simulated results between the European, Beninese, and Yakut populations shows that regarding both effectiveness and reliability and concerning all levels of kinship, the European simulation is expected to provide better results than the Beninese simulation, which is in turn expected to provide better results than the Yakut simulation (all chi-squared p values inferior to 0.05).

This implies that removing all ambiguous relationships and faulty declarations from the dataset exposes the limitations in LR determination of kinship that originate not in approximations but in reduced allelic diversity. In other words, a genetically homogenous population like the Yakuts, with little allelic diversity, will always lead to less effective and reliable determinations of kinship using LR on STR data. To a lesser extent, this is also true of the Beninese population presented here.

Discussion

LRs are reliable in detecting or confirming parent/child pairs

It appears that most parent/child pairs were reliably detected and successfully confirmed. There was no occurrence of a PO pair being underestimated further than FS. It should however be noted that the use of LR in PO situations is barely more efficient than a simple count of allelic exclusions. Out of 251 pairs without exclusions, 95.8% shared a PO relationship (as declared) and LR called PO for 97.9%. As previously mentioned, the difference is solely caused by the absence of exclusions in some FS pairs.

This implies that the use of LR in PO pairs halved the (already low) false-negative rate. However, in the common forensic situation where PO is compared to U but not to FS (a case with only two hypotheses), the success rate of a test consisting only in the counting of allelic exclusions would have been 100%.

We excluded ambiguous allelic exclusions from this study, since its aim was the evaluation of the LR method in small populations. These are cases where a mutation might have occurred, leading to the absence of a shared allele between parent and child, most often at only one locus. New methods do however propose the potential resolution of those cases through the analysis of STR sequences and show promising results [10, 11].

Previous studies, especially simulations, have not found such large numbers of ambiguities between PO and FS, but they often focus on a two-hypothesis approach [12]. Nevertheless, the overestimation of FS into PO due to the absence of allelic exclusions is the direct result of the lack of allelic diversity in the population under study. It can therefore often be overlooked by studies that have access to reliable allelic frequency data, collected or simulated.

In exploratory situations (where there are multiple hypotheses), any NGS-based method for the detection of PO

relationships would be both more effective and less subject to the problem of mutations, which are less frequent with SNPs. If hundreds or thousands of well-described independent SNPs are available, the detection of PO is entirely reliable [13, 14].

We should note here that many PO relationships declared were not biological relationships. This phenomenon was more prominent in the Beninese population, where it is not uncommon for some individuals to place their children in the care of biological (grandparents, aunts…) or allied (mothers-in-law…) family members. Although this also occurs in the Yakut population, it is much less frequent and most cases were identified in the field, through direct questioning and the explicit distinction between biological and non-biological offspring.

LR cannot reliably distinguish sibling and half-sibling pairs

The main issue highlighted in our study is the unreliability of FS and HS determinations. More than 13% (37/278, across populations and frequency sets) of FS pairs detected had been declared HS and more than 7% (33/422) of HS pairs detected had been declared FS.

Previous studies have shown that ambiguities in the second degree of kinship (typically, HS) are always present [15, 16]. This leads to situations where FS or HS can reliably be established against U but not against each other. This means that when external factors can guarantee that only two hypotheses are valid (i.e., "related" or "unrelated"), the LR method remains very reliable. It is again a common situation, with most forensic cases implying only one alternative hypothesis.

The effectiveness of second-degree LR can be improved using more STR markers, the sequences of those STR, additional SNP markers [14, 17], or combinations of all three [18]. While these approaches can be very successful (both effectiveness and reliability over 99.9%), they all require access to allelic frequencies for whatever additional markers are used, ideally along with information regarding heterozygosity and linkage disequilibrium. They do not therefore constitute a simple solution to the study of small or remote populations.

NGS-based methods, using very large numbers of SNPs, can be a solution less dependent on reliable allelic frequencies. They are however significantly more costly and seldom part of routine practice in forensic genetics laboratories, especially in countries still developing forensic infrastructure.

There is no error-proof way of selecting allelic frequencies

We show that the use of STR allelic frequencies computed on a population of unrelated individuals has not permitted significant improvement compared to the use of the natural population (all available individuals). There was no effect on the

detection or confirmation of PO relationships in either test population, and the effect on FS, HS, or U relationships was not consistent. Unrelated frequencies improved the effectiveness of HS and U calls in the Yakut population but reduced it or had no effect in the Beninese population. They also decreased reliability of FS calls in the Beninese population and had no effect in the Yakut population.

While the use of allelic frequencies computed on a population of unrelated individuals is a logical decision, it leads to arbitrarily underestimated frequencies for rare alleles. A common situation would be a trio of a mother and her two children, the siblings sharing one allele (at one locus) otherwise not observed in the population. The natural frequency (without eliminating any individual) in a population of 300 would be 2/600. If both children are eliminated from the computation of unrelated allelic frequencies, the allele is also eliminated. It must therefore be reintroduced with an arbitrary frequency, for example $1/2N$, where N is population size. If 100 individuals are eliminated in the process, the frequency of an already rare allele would be 1/400 (a 25% decrease), making it artificially rarer. This leads to an increase in the value of LR-HS between the siblings which is twice as important as the increase in the value of LR-FS, sometimes leading to new ambiguities.

While it is possible to include estimates of FST values in LR computations, they also suppose the representativity of the sample population and its uniformity. Using SNPs, some methods have been proposed to estimate frequencies where they are not readily available (for SNPs, see [19]) but they require access to NGS data and important computational capacities.

We can therefore not conclude on the use of different sets of allelic frequencies in small or ill-defined populations. There appears to be no satisfactory way of compensating for insufficient sampling, although it is common in the study of isolated populations and unavoidable in the study of ancient populations. In the case of a population displaying low diversity at the STR loci analyzed, the solution may not be the elimination of related individuals (at least as a first approach) but the constitution of appropriate STR panels. This, however, is also a costly endeavor for most research structures.

Conclusion

Our results show that the approximations of the LR method applied to STR in small or ill-defined populations cannot be reliably corrected. While reduced allelic diversity accounts for some errors and can be expected, the impact of population structure (i.e., unknown relationships between putative ancestors in genealogies) on LR determinations of kinship cannot be anticipated.

In the situations we have outlined (populations that are small, insufficiently surveyed, or for which the STR loci

analyzed are inappropriate) and the cases we present (ambiguous sibling/half-sibling pairs, high or low levels of inbreeding), an examiner might not be able to draw a conclusion one way or the other. Reliable determination for all degrees of kinship will be achieved using NGS and the analysis of very large numbers of SNPs or the combined analysis of autosomal, Y-chromosome and X-chromosome STRs, and of noncoding autosomal, Y-chromosome, and mitochondrial SNPs.

The LR method applied to STR will however not readily be abandoned by forensic geneticists around the world. Most national DNA databases are made up of STR profiles, and their conception has often required long and arduous scientific and legal work [20–22], and this effort has provided very reliable means of identification (especially direct individual identification). Furthermore, the current cost of NGS technologies keeps them inaccessible to countries that are only now beginning to constitute forensic databases, even for punctual cases.

Some countries have the resources to constitute these databases but only recently began the process of documenting STR allelic frequencies, a process that can be particularly demanding in nations like China which are home to very diverse ethnic groups. There, STR allelic frequencies are still being recorded even in the Han ethnic group [23, 24]. Even in countries where the use of a national forensic DNA database is routine, the results we present should inform the search for familial links rather than direct STR profile matches, especially since this is a recent and widespread development [25].

When studying ill-defined populations, in forensic or archeological cases, we therefore recommend that STR only be used to analyze first-degree kinship, especially parent/child pairs and that LR for all levels of kinship be computed and compared for each pair. Overlooking kinship levels through a two-hypothesis approach might lead to unseen errors if this approach is not warranted (that is, supported by unequivocal external data). This is especially true in differentiating siblings and half-siblings, two levels of kinship often designated by ambiguous terms or even not distinct whatsoever.

Sibling pairs should only be analyzed when no other level of kinship can be envisaged, since we found that determinations were unreliable in up to 20% of cases and that distinction between siblings and half-siblings is not possible using LR applied to STR in the absence of parental genotypes. Conclusions regarding any more distant relationship should not be attempted with that method. Investigators should also provide information concerning the population of origin of the individuals analyzed when it is available, so that some issues can be foreseen, and so reference allelic frequencies can be accurately selected when available. If allelic frequencies have to be computed on a population of related individuals, we recommend the use of both natural allelic frequencies and pruned (unrelated) frequencies. All ambiguities highlighted in that way should be discussed.

Finally, in contentious cases (or even routine if possible), uniparental markers should be analyzed, that is to say (at a minimum), the complete mitochondrial genome or a portion of it and Y-STR markers, as is done, is archeological genetics to confirm or reject kinship. These will permit the resolution of many cases and, in the case of shared complete mitochondrial haplotypes, can provide further arguments in favor of a recent common ancestor.

Compliance with ethical standards

This study was approved by the local Committee on Biomedical Ethics of the Yakut Scientific Centre of Complex Medical Problems. Informed consent was obtained from volunteers, or the parents of participating children where applicable, before sample collection. For the Beninese sample, the study was reviewed and approved by the ethics committee of the Research Institute of Applied Biomedical Sciences (CER-ISBA/ Institut des Sciences Biomédicales Appliquées) in Benin (No 61/CER/ ISBA/15). Individual written informed consent was obtained using a consent form translated in the subject's native language.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Buckleton JS, Bright J-A, Taylor D (2016) Forensic DNA evidence interpretation, Second edn. CRC Press, Taylor & Francis Group, Boca Raton
2. Keyser C, Hollard C, Gonzalez A, Fausser JL, Rivals E, Alexeev AN, Riberon A, Crubezy E, Ludes B (2015) The ancient Yakuts: a population genetic enigma. Philos Trans R Soc Lond Ser B Biol Sci 370(1660):20130385. https://doi.org/10.1098/rstb.2013.0385
3. Mary L, Zvenigorosky V, Kovalev A, Gonzalez A, Fausser JL, Jagorel F, Kilunovskaya M, Semenov V, Crubezy E, Ludes B, Keyser C (2019) Genetic kinship and admixture in Iron age Scytho-Siberians. Hum Genet 138(4):411–423. https://doi.org/10.1007/s00439-019-02002-y
4. Zvenigorosky V, Crubezy E, Gibert M, Theves C, Hollard C, Gonzalez A, Fedorova SA, Alexeev AN, Bravina RI, Ludes B, Keyser C (2016) The genetics of kinship in remote human groups. Forensic Sci Int Genet 25:52–62. https://doi.org/10.1016/j.fsigen.2016.07.018

5. Thompson EA (2013) Identity by descent: variation in meiosis, across genomes, and in populations. Genetics 194(2):301–326. https://doi.org/10.1534/genetics.112.148825

6. Olivieri L, Mazzarelli D, Bertoglio B, De Angelis D, Previdere C, Grignani P, Cappella A, Presciuttini S, Bertuglia C, Di Simone P, Polizzi N, Iadicicco A, Piscitelli V, Cattaneo C (2018) Challenges in the identification of dead migrants in the Mediterranean: the case study of the Lampedusa shipwreck of October 3rd 2013. Forensic Sci Int 285:121–128. https://doi.org/10.1016/j.forsciint.2018.01.029

7. Zvénigorosky V, Nikolaeva D, Romanov G, Solovev A, Barashkov N, Crubézy É, Fedorova S, Keyser C (2019) Persistence and disappearance of traditional patrilocality: matrimonial strategies and postnuptial residence patterns in two eastern Siberian communities of the twenty-first century. Sibirica 18(1):53–70

8. Egeland T (2015) Relationship inference with Familias and R : statistical methods in forensic genetics. First edition

9. Ruitberg CM, Reeder DJ, Butler JM (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Res 29(1):320–322. https://doi.org/10.1093/nar/29.1.320

10. Ma Y, Kuang JZ, Nie TG, Zhu W, Yang Z (2016) Next generation sequencing: improved resolution for paternal/maternal duos analysis. Forensic Sci Int Genet 24:83–85. https://doi.org/10.1016/j.fsigen.2016.05.015

11. Wang L, Chen M, Wu B, Liu YC, Zhang GF, Jiang L, Xu XL, Zhao XC, Ji AQ, Ye J (2018) Massively parallel sequencing of forensic STRs using the ion chef and the ion S5 XL systems. J Forensic Sci 63(6):1692–1703. https://doi.org/10.1111/1556-4029.13767

12. Tamura T, Osawa M, Ochiai E, Suzuki T, Nakamura T (2015) Evaluation of advanced multiplex short tandem repeat systems in pairwise kinship analysis. Leg Med (Tokyo) 17(5):320–325. https://doi.org/10.1016/j.legalmed.2015.03.005

13. Cho S, Shin ES, Yu HJ, Lee JH, Seo HJ, Kim MY, Lee SD (2017) Set up of cutoff thresholds for kinship determination using SNP loci. Forensic Sci Int Genet 29:1–8. https://doi.org/10.1016/j.fsigen.2017.03.009

14. Mo SK, Liu YC, Wang SQ, Bo XC, Li Z, Chen Y, Ni M (2016) Exploring the efficacy of paternity and kinship testing based on single nucleotide polymorphisms. Forensic Sci Int Genet 22:161–168. https://doi.org/10.1016/j.fsigen.2016.02.012

15. von Wurmb-Schwark N, Podruks E, Schwark T, Gopel W, Fimmers R, Poetsch M (2015) About the power of biostatistics in sibling analysis-comparison of empirical and simulated data. Int J Legal Med 129(6):1201–1209. https://doi.org/10.1007/s00414-015-1252-9

16. Turrina S, Ferrian M, Caratti S, Cosentino E, De Leo D (2016) Kinship analysis: assessment of related vs unrelated based on defined pedigrees. Int J Legal Med 130(1):113–119. https://doi.org/10.1007/s00414-015-1290-3

17. Mo SK, Ren ZL, Yang YR, Liu YC, Zhang JJ, Wu HJ, Li Z, Bo XC, Wang SQ, Yan JW, Ni M (2018) A 472-SNP panel for pairwise kinship testing of second-degree relatives. Forensic Sci Int Genet 34:178–185. https://doi.org/10.1016/j.fsigen.2018.02.019

18. Li R, Li H, Peng D, Hao B, Wang Z, Huang E, Wu R, Sun H (2019) Improved pairwise kinship analysis using massively parallel sequencing. Forensic Sci Int Genet 38:77–85. https://doi.org/10.1016/j.fsigen.2018.10.006

19. Samsi S, Yu B, Ricke DO, Fremont-Smith P, Kepner J, Reuther A (2018) Large-scale Bayesian kinship analysis. In: 2018 IEEE High Performance extreme Computing Conference (HPEC), 25–27 Sept. 2018, pp 1–4. https://doi.org/10.1109/HPEC.2018.8547549

20. Manus J-M (2001) Empreintes génétiques au Code pénal et au Code de la Santé publique. Rev Fr Lab 2001(333):9. https://doi.org/10.1016/S0338-9898(01)80138-4

21. Saint Dizier D (2002) Fichier national automatisé des empreintes génétiques FNAEG. Médecine & Droit 2002(53):1–5. https://doi.org/10.1016/S1246-7391(02)83011-1

22. Manaouil C, Werbrouck A, Traulle É, Cordier A, Gignon M, Jarde O (2008) Le Fichier national automatisé des empreintes génétiques (FNAEG) : fonctionnement, intérêts et limites. Médecine & Droit 2008(88):17–23. https://doi.org/10.1016/j.meddro.2007.12.006

23. Zhao ZL, Xia L, Zhao C, Ameen F, Alyahya S, Yao J (2019) Population genetics of 15 autosomal STR loci in the Han population of Ili Kazakh autonomous prefecture, northwestern China. Ann Hum Genet 83(5):318–324. https://doi.org/10.1111/ahg.12314

24. Pan M, Ye Q, Ju XB, Cui H, Zhou HY (2019) Population genetic polymorphism and mutation analysis of 19 autosomal STR loci in Jiangsu Han individuals in eastern China. Ann Hum Biol 46:1–7. https://doi.org/10.1080/03014460.2019.1632929

25. Debus-Sherrill S, Field MB (2019) Familial DNA searching- an emerging forensic investigative tool. Sci Justice 59(1):20–28. https://doi.org/10.1016/j.scijus.2018.07.006