Forensic Science International: Genetics 33 (2018) 10–16

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Discrimination of relationships with the same degree of kinship using chromosomal sharing patterns estimated from high-density SNPs

Chie Morimoto[a,b], Sho Manabe[a], Shuntaro Fujimoto[a], Yuya Hamano[a,c], Keiji Tamaki[a,*]

[a] Department of Forensic Medicine, Kyoto University Graduate School of Medicine, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
[b] Research Fellow of the Japan Society for the Promotion of Science, Japan
[c] Forensic Science Laboratory, Kyoto Prefectural Police Headquarters, 85-3, 85-4, Yabunouchi-cho, Kamigyo-ku, Kyoto 602-8550, Japan

ARTICLE INFO

Keywords:
Kinship analysis
Single nucleotide polymorphisms
Same degree of kinship
Chromosomal sharing
DNA microarray

ABSTRACT

Distinguishing relationships with the same degree of kinship (e.g., uncle–nephew and grandfather–grandson) is generally difficult in forensic genetics by using the commonly employed short tandem repeat loci. In this study, we developed a new method for discerning such relationships between two individuals by examining the number of chromosomal shared segments estimated from high-density single nucleotide polymorphisms (SNPs).

We computationally generated second-degree kinships (i.e., uncle–nephew and grandfather–grandson) and third-degree kinships (i.e., first cousins and great-grandfather–great-grandson) for 174,254 autosomal SNPs considering the effect of linkage disequilibrium and recombination for each SNP. We investigated shared chromosomal segments between two individuals that were estimated based on identity by state regions. We then counted the number of segments in each pair.

Based on our results, the number of shared chromosomal segments in collateral relationships was larger than that in lineal relationships with both the second-degree and third-degree kinships. This was probably caused by differences involving chromosomal transitions and recombination between relationships. As we probabilistically evaluated the relationships between simulated pairs based on the number of shared segments using logistic regression, we could determine accurate relationships in > 90% of second-degree relatives and > 70% of third-degree relatives, using a probability criterion for the relationship ≥0.9. Furthermore, we could judge the true relationships of actual sample pairs from volunteers, as well as simulated data. Therefore, this method can be useful for discerning relationships between two individuals with the same degree of kinship.

1. Introduction

In forensic genetics, a pairwise kinship analysis is typically performed to investigate a relationship between two individuals by calculating the likelihood ratio [1]. The likelihood ratio is generally the ratio between the probability that certain DNA typing results would be obtained if the pair is related versus that obtained if the pair is unrelated. Meanwhile, when no predicted relationship exists between the pair (e.g., identification of bone remains, without personal belongings), all possible relationships should be compared simultaneously. We have previously proposed a method for pairwise kinship analysis using the "index of chromosome sharing" (*ICS*) calculated from high-density autosomal single nucleotide polymorphisms (SNPs) [2]. In the *ICS* calculation, we investigate the genetic length (centi-Morgan (cM)) of all identity by state (IBS) segments between the two individuals, and then sum the genetic length of the IBS segments longer than a given threshold (i.e., 4 cM), which is applied for removing coincidental matches. Because the *ICS* values reflect the total genetic length of shared chromosomal regions between two individuals, the values relate to the expected proportion of the shared genome (i.e., degree of kinship). Thus, the *ICS* values in sibling (first-degree relative), uncle–nephew (second-degree relative), first cousin (third-degree relative), first cousin once removed (fourth-degree relative), and second cousin (fifth-degree relative) were distributed as log-normal distributions, centering around the median values of 2805, 1919, 1018, 544, and 311, respectively. Using this approach, we could determine the degrees of kinship up to the third-degree, regardless of sex, even when the kinship of the pair was totally unpredictable.

However, we probably cannot distinguish relatives with the same degree of kinship (e.g., uncle–nephew and grandfather–grandson) using this approach, because these relatives have the same expected total genetic length of shared chromosomal regions. If possible, adding key individuals is effective for increasing information [3]. Otherwise, investigating markers on the sex chromosomes [4–6] might be useful to

* Corresponding author.
E-mail address: ktamaki@fp.med.kyoto-u.ac.jp (K. Tamaki).

https://doi.org/10.1016/j.fsigen.2017.11.010
Received 11 September 2017; Received in revised form 7 November 2017; Accepted 14 November 2017
Available online 22 November 2017
1872-4973/

discriminate such relationships. For example, Y chromosomal markers are useful for distinguishing paternal relatives from maternal relatives. However, it is impossible to discriminate within paternal relationships such as uncle–nephew and grandfather–grandson. Then, we can theoretically distinguish between second-degree relatives by using identity by descent (IBD) probabilities of two linked loci [7–9]. However, these studies were not verified by practical data, and the accuracy is unknown.

Chromosomal sharing patterns, including the number and the genetic length of the actual shared segments (i.e., IBD regions) on the chromosomes, are expected to differ between relatives, even for the same degree of kinship, due to differences in the frequency of meioses and chromosomal transitions from their common ancestors [10–14]. If two individuals have a collateral relationship, the shared chromosomal segments are expected to be larger in number and shorter in genetic length than those in lineal relationship with the same degree of kinship. This is because the number of recombination events from the common ancestors is larger in a collateral relationship than that in a lineal relationship with the same degree of kinship [11,13].

In this study, we further developed our previous approach to distinguish between relationships with the same degree of kinship by investigating the differences between chromosomal sharing patterns. We targeted the second and third-degree of kinship, including collateral and lineal relationships: uncle–nephew vs. grandfather–grandson as second-degree kinships, and first cousins vs. great-grandfather–great-grandson as third-degree kinships. We focused on the number of shared chromosomal segments between two individuals. The shared segments were estimated on the basis of IBS regions, the effect of coincidental matches being minimized. For a probabilistic prediction of the relationships between two individuals, we proposed a logistic regression model, using computationally generated genotypes. We also confirmed the validity of the proposed method employing actual samples that were genotyped using DNA microarray.

2. Methods

2.1. Generating computer-based familial genotypes of high-density SNPs

We utilized 249 computationally synthesized familial genotypes (Fig. 1) of 174,254 autosomal SNPs on the HumanCore BeadChip (Illumina, San Diego, CA, USA) used in our previous study [2].

The SNP haplotypes of the founders (i.e., Nos. 1, 2, 3, 6, 7, and 10 in Fig. 1) were estimated using the Shape-IT algorithm [15] and 1498 Japanese individuals in the Nagahama Prospective Genome Cohort [16], to incorporate linkage disequilibrium into the computational genotypes. We assumed that the founders were unrelated (i.e., non-inbred pedigrees). We then generated the haplotypes of the descendants by taking into account recombination events between each SNP based on sex-averaged recombination rates [17]. The family included two types of second-degree relative pairs: uncle–nephew (Nos. 5–8 in Fig. 1) as a collateral relationship, and grandfather–grandson (Nos. 1–8 in Fig. 1) as a lineal relationship. It also included two types of third-degree relative pairs: first cousins (Nos. 8–9 in Fig. 1) as a collateral relationship, and great-grandfather–great-grandson (Nos. 1–11 in Fig. 1) as a lineal relationship.

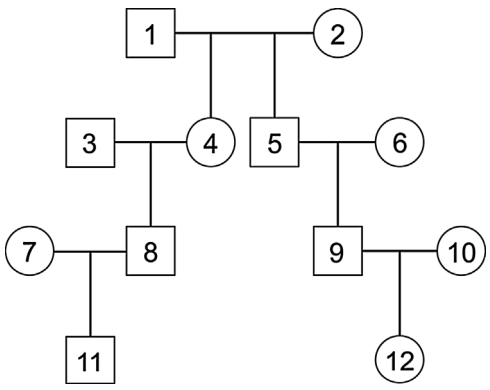

Fig. 1. Family tree of a family comprising 12 individuals.

In the SNP typing system based on DNA microarrays, uncalled events (i.e., genotypes that cannot be obtained) or genotyping errors (e.g., a true genotype, AB, is miscalled as AA) occur infrequently even in pristine DNA samples. In this study, we also included uncalled events and genotyping errors into the simulated genotypes for each probability. The values of the two probabilities were determined from the actual genotyping results of 67 pristine DNA samples typed on the HumanCore BeadChip (Illumina) in our previous study [2]. Detailed methods are presented in the Supplementary information and Supplementary Table 1. Uncalled events were randomly generated at all SNPs, with a probability of 5.97×10^{-4}. We assumed that genotyping errors occurred only in heterozygous genotypes through allele imbalance (i.e., one allele in the heterozygous genotype is undetected). Therefore, genotyping errors were introduced by converting heterozygous genotypes into randomly chosen homozygous genotypes, with a probability of 2.28×10^{-4}. All programs used for simulations were run using the statistical software R, version 3.3.2 [18].

2.2. Analyzing chromosomal sharing patterns between individual pairs

We investigated the differences in chromosomal sharing patterns between collateral relatives and lineal relatives with the same degree of kinship. We ordered all 174,254 SNPs according to their chromosomal positions and then counted the IBS state at each locus between two individuals. The IBS state has three possible outcomes: 0, 1, or 2. SNPs with uncalled events in either individual are ignored in the subsequent analysis. We then chose IBS regions (i.e., regions where 1 or 2 are continuously aligned), which could reflect chromosomal sharing. However, some IBS regions include coincidental matches, because even unrelated pairs can share SNPs by chance. Therefore, we have to exclude comparatively shorter IBS regions to minimize the effect of coincidental matches. We set a threshold for the genetic length of IBS regions between two individuals using receiver operating characteristic curve analysis. We identified IBS regions longer than the threshold (i.e., 4 cM) as apparent "shared segments."

To analyze the total amounts of chromosomal sharing, *ICS* values in each pair were calculated by summing the genetic length of the shared segments in the same way as our previous study [2]. In this study, we also counted the number of the shared segments in each relationship. We then compared the *ICS* values and the number of shared segments between collateral relatives and lineal relatives with the same degree of kinship: uncle–nephew vs. grandfather–grandson and first cousins vs. great-grandfather–great-grandson.

2.3. Logistic regression analysis

We used a binary logistic regression analysis to determine whether the relationship of the pair is collateral or lineal because we assumed two outcomes (collateral or lineal). We modeled the probability that the pair has a collateral relationship as a function of the number of shared segments, using the following equation:

$$\Pr(C|x) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x)} \quad (1)$$

where $\Pr(C|x)$ is the probability that the relationship of the pair is collateral, and x is the number of shared segments. In this model, β_0 and β_1 are the logistic parameters. We estimated β_0 and β_1 for each degree of kinship. $\Pr(L|x)$, the probability that the relationship of the pair is lineal, can be calculated by $\Pr(L|x) = 1 - \Pr(C|x)$.

We randomly selected 2/3 of the computationally generated