A novel strategy for sibship determination in trio sibling model

**Aim** To use a virtually simulated population, generated from published allele frequencies based on 15 short tandem repeats (STR), to evaluate the efficacy of trio sibship testing and sibling assignment for forensic purposes.

**Methods** Virtual populations were generated using 15 STR loci to create a large number of related and unrelated genotypes (10,000 trio combinations). Using these virtual populations, the probability of related and unrelated profiles can be compared to determine the chance of inclusion of being siblings if they are true siblings and the chance of inclusion if they are unrelated. Two specific relationships were tested – two reference siblings were compared to a third true sibling (3S trio, sibling trio) and two reference siblings were compared to an unrelated individual (2S1U trio, non-sibling trio).

**Results** When the likelihood ratio was greater than 1, 99.87% of siblings in the 3S trio population were considered as siblings (sensitivity); 99.88% of non-siblings in the 2S1U trio population were considered as non-siblings (specificity); 99.9% of both populations were identified correctly as siblings and non-siblings; and the accuracy of the test was 99.88%.

**Conclusions** The high sensitivity and specificity figures when using two known siblings compared to a putative sibling are significantly greater than when using only one known relative. The data also support the use of increasing number of loci allowing for greater confidence in genetic identification. The system established in this study could be used as the model for evaluating and simulating the cases with multiple relatives.
DNA profiling when applied to the identification of human remains requires comparison to reference data generated from either ante mortem samples or from known living relatives (1,2). If there are direct genetic relatives, such as an offspring and both genetic parents, then this process is well established; however if there are only putative siblings then the opportunity for misidentification increases. Such instances arose in the identification of deceased from a mass disaster (3). The comparison of the DNA profiles of two putative siblings can result in a likelihood ratio (LR) or combined sibling index (SI), being the odds favoring the tested samples originating from siblings compared to the samples originating from two genetically unrelated individuals. Previous studies have shown that the combined SI values will increase if more short tandem repeat (STR) loci are used (4,5) and reported a method for determining sibling (6). Recently, the universal algorithms for commonly used kinship indexes between two individuals have been established (7), and it has been suggested from the comparison study that the power of the identity by state (IBS) method was similar to that of ITO method (8) in full sibling determination, with advantages of convenience in calculation and independence on the allele frequency of STR loci (9). This study extends the data to cases where an unknown sample is tested and compared to two reference siblings.

Basic laws of inheritance are used to assess the probability that common alleles in the tested sample and a reference sample are inherited identical by decent (IBD). There is a 25% chance that two siblings will inherit both alleles IBD from common parents; a 50% chance that two siblings will inherit one parental allele IBD, and a 25% chance that two siblings will inherit IBD no parental alleles and therefore will share no alleles at a diploid locus. This last possibility, where there is a 25% chance that at any one locus the two true siblings may share neither allele, is problematic when identifying deceased in a mass disaster when there are only living siblings for genetic linkage. While the probability that one sibling will inherit 0 alleles IBD from their common parents is constant at 25% for each locus, the probability that two full siblings will not share an allele/locus partly depends on locus polymorphism information content (PIC) or heterozygosity. Greater PIC/locus will increase the chance that 2 siblings have neither allele in common (10). Additionally, greater confidence in sibling assignment of the tested sample might be obtained if there are 2 reference siblings for comparison. In such a situation, a sibling trio, the DNA type of the tested sample can be compared to those of the 2 reference siblings to determine the likelihood of either being a sibling or being genetically unrelated. This may provide confidence in an exclusion, based on sufficient inconsistencies, if there is no expected allele observed in the unknown (or tested) sibling. For inclusion, the LR will increase dramatically (10).

Previous studies (6-13) undertook DNA typing of known genetic relatives from which they made conclusions of putative sibling identification. These studies are necessarily limited by the availability of data. Computer-simulated populations have been also used (2,5,14) to generate virtual populations, with the limitation that these are generated rather than real data, but with the great benefit of an increase in the size of the data available. In this study, we report on the use of a virtual simulated population generated from published allele frequencies based on 15 STR to evaluate the efficacy of trio frequencies based on 15 STR to evaluate the efficacy of trio sibship testing and sibling assignment (15).

MATERIALS AND METHODS

Populations

Two simulated populations with genotypes of 15 STR loci were created from members of a previously described population (15) to produce trio sibling combinations. These data were generated by Microsoft® Office Excel 2007. The two populations were designed to create two distinct situations: 2 reference siblings with a true sibling (3S trio, sibling trio) and 2 reference siblings with an unrelated individual (2S1U trio, non-sibling trio). Each population contained 10,000 trio data sets based on the 15 STR loci generated by the AmpFISTR® Identifiler™ kit (Applied Biosystems, Foster City, CA, USA). The population was created by starting with the assignment of alleles for both father and mother randomly chosen from the population data (15) in a spreadsheet of Microsoft® Office Excel 2007. After genotypes of parents are assigned, genotypes of their offsprings can be assigned by randomly selecting one allele from each locus from each parent. Three offsprings from each “family” were selected to generate the 3S trio population. Two offsprings from each “family” were selected and combined with an unrelated individual to generate the 2S1U trio population.

Calculations

The trio genotype combinations were entered into a spreadsheet and all calculations were performed using Microsoft® Office Excel 2007. The likelihood ratio (LR) of each unique combination was calculated using relat-
ed (R) and unrelated (U) hypotheses. R is the probability of observing the genotype of S3 given that reference siblings S1 and S2 are truly full siblings of S3. U is the probability of observing the genotype of S3 given that the reference genotypes of full siblings S1 and S2 are unrelated to S3. R is the sum of the genotype probabilities deduced from three siblings S1, S2, and S3, as shown in the example (Table 1). U is the sum of the possible probabilities of genotypes deduced from two siblings, S1 and S2, and the probability of an unrelated S3. All LR formulas of possible trio combinations were generated and built in the spreadsheet. The allele frequencies used are for the Taiwan population (15). As this was a virtual population, any substructure and possible mutational events were ignored. The distributions of likelihood ratio (LR) values were plotted using the SigmaPlot 9 (Systat software, Inc., San Jose, CA, USA), with the curve fitting analysis.

### Results

Initial results generated 64 genotype combinations resulting from 2 reference siblings and one tested sample. Within these combinations, there were 37 combinations with common parents where S3 could not be excluded as being a sibling of the other 2 reference siblings (S1 and S2) in trio paternity test (Table 2). The LR formulas for these combinations (Table 3) were generated according to standard methods of genetic inheritance (Table 1). For the other 27 combinations, sample S3 within the trio did not possess the expected alleles and fell into the category of an exclusion (or genetic inconsistency) as being a sibling of the other two (Table 4).

The LR values and distribution for populations 3S and 2S1U were calculated (Figure 1). For this calculation, the 27 trio combinations (Table 4), where there was a genetic inconsistency, were given a value of 0.001, rather than zero. The log LR values of population 3S ranged from -2.24 to 15.96 and for population 2S1U they ranged from -32.57 to 2.15.
The simulated method was validated by using 75 real trio siblings, and the log LR ranged from 2.33 to 13.20. This distribution for the real population showed concordance with the simulated population 3S (Figure 1), indicating the accuracy of the simulation model. In Figure 1, the 2U and 2S duo populations (one known sibling compared to one tested sample) were also included and used as the simulated controls. The distributions of log LR were closer for 2U and 2S than for 2S1U and 3S. These data indicated that greater confidence in sibling assignment of the tested sample could be obtained if there were two reference siblings rather than one for comparison.

The data for the 2S1U trio population showed that there was only 1.09% (109 in 10,000) of the genotype combina-

### Table 3. The formulas to derive the likelihood ration (LR) for 37 non-excluded combinations in a trio sibling test with 2 reference siblings (S1 and S2) and one putative sibling (S3)

<table>
<thead>
<tr>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>LR formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>AA</td>
<td>AA</td>
<td>(\frac{1 + 3a + 9a^2}{4 + 8a^2 + 8a^3 + 4a^4})</td>
</tr>
<tr>
<td>AA</td>
<td>AA</td>
<td>AB</td>
<td>(\frac{1 + 3a}{4a^2 + 4a^3})</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>AA</td>
<td>(\frac{1}{4 + 8a + 4a^2})</td>
</tr>
<tr>
<td>AA</td>
<td>AA</td>
<td>BC</td>
<td>(\frac{1}{4 + 8a + 4a^2})</td>
</tr>
<tr>
<td>AA</td>
<td>AB</td>
<td>AA</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AA</td>
<td>AB</td>
<td>BB</td>
<td>(\frac{1}{4ab + 8a^2 + 8a^2b})</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>AA</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>AA</td>
<td>(\frac{1}{4a + 8a^2})</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>AB</td>
<td>(\frac{1}{4a + 8a^2})</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>BB</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>AB</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>AB</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>AC</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AB</td>
<td>AA</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AB</td>
<td>BB</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>BB</td>
<td>AB</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AC</td>
<td>AA</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AC</td>
<td>AB</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AC</td>
<td>BC</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AC</td>
<td>CC</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AC</td>
<td>CD</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>AC</td>
<td>BD</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>CD</td>
<td>AB</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>CD</td>
<td>AC</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>CD</td>
<td>AD</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
<tr>
<td>AB</td>
<td>CD</td>
<td>BC</td>
<td>(\frac{1}{4a + 8a^2 + 8a^3})</td>
</tr>
</tbody>
</table>

*a, b, c, and d represent the frequencies of allele A, B, C, and D respectively.
tions where the unrelated individual in the trio was not excluded as being a sibling. These data indicated that a test for sibship based on these genotype combinations would result in 98.91% of non-sibling trios being excluded correctly. The distribution of LR (log) values for these combinations that were not excluded in the 2S1U population ranged from -6.49 to 1.83 (Figure 2). Analysis of the number of loci indicating sibship exclusion showed that out of the 10,000 combinations there were 520 instances of the 2S1U scenario, with only one locus exhibiting a genetic inconsistency. These data highlight a possible risk of making a positive sibship identification based on one genetic inconsistency, such as an assumption of this being the result of a mutation. There were 1432 and 2111 instances for two and three loci exhibiting a genetic inconsistency, respectively. The STR loci of D18S51, D2S1338, and FGA were found to be most informative in excluding sibship. The power of exclusion (PE) (16) for nonparent of these loci was 0.727, 0.726, and 0.722 respectively (15). The PE of the least informative locus TPOX was only 0.338.

The sensitivity and specificity of the test was determined by comparing the populations 3S and 2S1U generated using these 15 STR loci. Table 5 shows the results for the trio siblings, where an LR of at least 100 was obtained; we used 100 as this was a figure suggested by the AABB for non-exclusions (17). The sensitivity, specificity, and accuracy were 98.50%, 99.99% and 99.25%, respectively; with a PPV of 99.99% and NPV of 98.52%. If an LR threshold value was greater than 1, 99.87% of siblings in the 3S trio population were considered as siblings (sensitivity); 99.88% of non-siblings in 2S1U trio population were considered as non-siblings (specificity); 99.9% of both populations were identified correctly as siblings and non-siblings; and the accuracy of the test was 99.88%. If the LR value was greater than 1000, then there were no expected false inclusions as the specificity was 100%, with an accuracy of 97.83%.

**Table 4.** Genotype combinations in which the alleged sibling (S3) can be excluded as being a sibling of the other 2 reference siblings (S1 and S2) in trio sibship test* 

<table>
<thead>
<tr>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>AB</td>
<td>XX</td>
<td>AA</td>
<td>BC</td>
<td>CC</td>
<td>AB</td>
<td>CD</td>
<td>BB</td>
</tr>
<tr>
<td>AA</td>
<td>AB</td>
<td>XY</td>
<td>AA</td>
<td>BC</td>
<td>CX</td>
<td>AB</td>
<td>CD</td>
<td>BC</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>AX</td>
<td>AA</td>
<td>BC</td>
<td>XX</td>
<td>AB</td>
<td>CD</td>
<td>BX</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>BX</td>
<td>AA</td>
<td>BC</td>
<td>XY</td>
<td>AB</td>
<td>CD</td>
<td>CC</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>XX</td>
<td>AB</td>
<td>AC</td>
<td>AX</td>
<td>AB</td>
<td>CD</td>
<td>CX</td>
</tr>
<tr>
<td>AA</td>
<td>BB</td>
<td>XY</td>
<td>AB</td>
<td>AC</td>
<td>AX</td>
<td>AB</td>
<td>CD</td>
<td>DD</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>AX</td>
<td>AB</td>
<td>AC</td>
<td>XY</td>
<td>AB</td>
<td>CD</td>
<td>DX</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>BX</td>
<td>AB</td>
<td>AC</td>
<td>AX</td>
<td>AB</td>
<td>CD</td>
<td>CX</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>BB</td>
<td>AB</td>
<td>AC</td>
<td>AX</td>
<td>AB</td>
<td>CD</td>
<td>YY</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>AX</td>
<td>AB</td>
<td>AC</td>
<td>XY</td>
<td>AB</td>
<td>CD</td>
<td>DX</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>BB</td>
<td>AB</td>
<td>AC</td>
<td>AX</td>
<td>AB</td>
<td>CD</td>
<td>YY</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>XX</td>
<td>AB</td>
<td>AC</td>
<td>AX</td>
<td>AB</td>
<td>CD</td>
<td>YY</td>
</tr>
<tr>
<td>AA</td>
<td>BC</td>
<td>XY</td>
<td>AB</td>
<td>AC</td>
<td>AX</td>
<td>AB</td>
<td>CD</td>
<td>YY</td>
</tr>
</tbody>
</table>

*Alleles X and Y in S3 are those not observed in siblings S1 and S2.

**Table 5.** The value of test result in trio siblings with the likelihood ratio (LR) of at least 100

<table>
<thead>
<tr>
<th>Test result</th>
<th>True status</th>
<th>Sibling</th>
<th>Unrelated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sibling</td>
<td>9850</td>
<td>1</td>
<td>9851</td>
<td></td>
</tr>
<tr>
<td>Unrelated</td>
<td>150</td>
<td>9999</td>
<td>10,149</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10,000</td>
<td>10,000</td>
<td>20,000</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.** Case study of the combined sibling index (SI) in duo and trio sibling situations from a family of 2 true siblings (S1 and S2) and a putative third sibling (S3)

<table>
<thead>
<tr>
<th>STR loci</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>SI between S1 and S3</th>
<th>SI between S2 and S3</th>
<th>SI between S1, S2, and S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>12,17</td>
<td>13,15</td>
<td>12,15</td>
<td>1.22</td>
<td>1.002</td>
<td>2.917</td>
</tr>
<tr>
<td>D21S11</td>
<td>29,31</td>
<td>29,30.2</td>
<td>30,2,31</td>
<td>1.457</td>
<td>12.876</td>
<td>8.381</td>
</tr>
<tr>
<td>D7S820</td>
<td>8,11</td>
<td>8,11</td>
<td>8,11</td>
<td>4.353</td>
<td>4.353</td>
<td>5.265</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>12,13</td>
<td>9,11</td>
<td>9,11</td>
<td>0.25</td>
<td>15.228</td>
<td>11.631</td>
</tr>
<tr>
<td>D31358</td>
<td>15,17</td>
<td>15,15</td>
<td>17,17</td>
<td>1.315</td>
<td>0.25</td>
<td>0.791</td>
</tr>
<tr>
<td>TH01</td>
<td>9.9</td>
<td>9.9</td>
<td>9.9</td>
<td>2.412</td>
<td>2.412</td>
<td>2.997</td>
</tr>
<tr>
<td>D13337</td>
<td>10,11</td>
<td>10,12</td>
<td>10,11</td>
<td>5.233</td>
<td>1.113</td>
<td>4.406</td>
</tr>
<tr>
<td>D165339</td>
<td>12,12</td>
<td>11,12</td>
<td>9,12</td>
<td>1.423</td>
<td>0.836</td>
<td>0.483</td>
</tr>
<tr>
<td>D251338</td>
<td>20,20</td>
<td>22,25</td>
<td>20,25</td>
<td>2.492</td>
<td>2.194</td>
<td>17.435</td>
</tr>
<tr>
<td>D195433</td>
<td>13,15</td>
<td>13,13</td>
<td>13,15</td>
<td>8.483</td>
<td>1.075</td>
<td>9.11</td>
</tr>
<tr>
<td>VWA</td>
<td>14,17</td>
<td>17,18</td>
<td>14,17</td>
<td>3.403</td>
<td>0.757</td>
<td>2.824</td>
</tr>
<tr>
<td>TPOX</td>
<td>11,11</td>
<td>8,11</td>
<td>8,11</td>
<td>1.121</td>
<td>1.697</td>
<td>1.474</td>
</tr>
<tr>
<td>D18S51</td>
<td>13,15</td>
<td>15,19</td>
<td>13,15</td>
<td>5.426</td>
<td>0.929</td>
<td>4.82</td>
</tr>
<tr>
<td>D5S818</td>
<td>7,12</td>
<td>11,12</td>
<td>11,12</td>
<td>0.833</td>
<td>3.071</td>
<td>2.396</td>
</tr>
<tr>
<td>FGA</td>
<td>21,21</td>
<td>21,22</td>
<td>19,22</td>
<td>0.25</td>
<td>0.959</td>
<td>0.634</td>
</tr>
<tr>
<td>AMEL</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Combined SI| 4163.289 | 3977.735 | 36561.850.243     |}

www.cmj.hr
Application to a real-world case

An illustration of the application of a combined SI is provided (Table 6) in the following real case where a putative third sibling (S3) was tested against two confirmed siblings (S1 and S2). The figure obtained for the combined SI using the trio sibship test (36 561 850.243) was far higher than for the duo test (4163.289 or 3977.735).

DISCUSSION

We studied sibship assignment based on 15 STR loci; where the tested sample was a putative sibling of 2 reference siblings. The use of a virtual population starting with known alleles allows for up to 10 000 trio combinations to be used in such a study. In addition to the high exclusion rate for non-siblings, this study showed a high degree of sensitivity, specificity, and accuracy in sibling identification. In our previous study (14), the variations of the distribution for paternity index (PI) and random man not excluded (RMNE) values in paternity test were evaluated based on 1244 virtual families. There were minor variations between the PI and 1/RMNE values in trio parentage testing compared with duo parentage testing. Also, the distribution of PI/(1/RMNE) for duo families exhibited greater variation than that for trio families. This highlighted the effect that different mathematical methods can have on the results using either of these tests; this effect was found to be greater in the duo cases. A consequence is that with more individuals being tested greater confidence in the results will be obtained.

Our data for trio situations are much higher than those reported previously in duo sibship tests (11,12,18). In the report using 33 duo pairs and 15 STR loci, sensitivity of 93.94% and specificity of 90.91% was reported (19). This study provides evidence that analysis of trio sibship testing with a fixed number of STR loci is more powerful than analysis of duo using the same loci. The trio sibling model described is a cost-effective way to screen disaster samples for sibling assignment and identification. The system established in this study could be used as the model for evaluating and simulating the cases with multiple relatives.

References