Assessing the Implications for Close Relatives in the Event of Similar but Non-Matching DNA Profiles

Dan E. Krane

Wright State University - Main Campus, dan.krane@wright.edu

Follow this and additional works at: https://corescholar.libraries.wright.edu/biology

Part of the Biology Commons, Medical Sciences Commons, and the Systems Biology Commons

Repository Citation

This Presentation is brought to you for free and open access by the Biological Sciences at CORE Scholar. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.
Assessing the implications for close relatives in the event of similar but non-matching DNA profiles

Dan Krane
Wright State University, Dayton, OH 45435

Forensic Bioinformatics
(www.bioforensics.com)
Familial search

- Database search yields a close but imperfect DNA match
- Can suggest a relative is the true perpetrator
- Great Britain performs them routinely
- Reluctance to perform them in US since 1992
- NRC report
- Current CODIS software cannot perform effective searches
Three approaches to familial searches

- Search for rare alleles (inefficient)
- Count matching alleles (arbitrary)
- Likelihood ratios with kinship analyses
Accounting for relatives

Graph showing the distribution of pairwise shared alleles for Randomized Individuals, Simulated Cousins, and Simulated Siblings. The x-axis represents the number of pairwise shared alleles, ranging from 0 to 24. The y-axis shows the percent of total, ranging from 0% to 20%. The graph includes peaks for each category, indicating the percentage of shared alleles for relatives.
Three approaches to familial searches

- Search for rare alleles (inefficient)
- Count matching alleles (arbitrary)
- Likelihood ratios with kinship analyses
Example

- 2003 North Carolina performed post-conviction DNA testing on evidence from a 1984 rape and murder
- Exonerated Darryl Hunt, who had served 18 years of a life sentence
- Database search yielded best match to Anthony Brown with 16/26 alleles
- Brother Willard Brown tested and found to be a perfect match
Thresholds for similarity

- Virginia: “be very, very close”
- California: “appear useful”
- Florida: match at least 21 out of 26 alleles
Is 16/26 close enough?

- How many pairs of individuals match at 16+ alleles with unrelated databases of size...

- 1,000: 562 pairs of individuals
- 5,000: 13,872 pairs of individuals
- 10,000: 52,982 pairs of individuals
Is the true DNA match a sibling or a random individual?

• Given a closely matching profile, who is more likely to match, a sibling or a randomly chosen, unrelated individual?

• Use a likelihood ratio

\[ LR = \frac{P(E \mid \text{relative})}{P(E \mid \text{random})} \]
Probabilities of siblings matching at 0, 1 or 2 alleles

- Weir and NRC I only present probabilities that siblings match perfectly.

\[
P(E \mid \text{sib}) = \begin{cases} 
\frac{P_a \cdot P_b \cdot HF}{4}, & \text{if shared} = 0 \\
\frac{P_b + P_a \cdot P_b \cdot HF}{4}, & \text{if shared} = 1 \\
\frac{1 + P_a + P_b + P_a \cdot P_b \cdot HF}{4}, & \text{if shared} = 2
\end{cases}
\]

HF = 1 for homozygous loci and 2 for heterozygous loci.
Probabilities of parent/child matching at 0, 1 or 2 alleles

- Weir and NRC I only present probabilities that parent/child match perfectly.

\[
P(E \mid \text{parent/child}) = \begin{cases} 
0, & \text{if shared} = 0 \\
\frac{P_b}{2}, & \text{if shared} = 1 \\
\frac{P_a + P_b}{2}, & \text{if shared} = 2
\end{cases}
\]
Other familial relationships

Cousins:

\[ P(E \mid \text{cousins}) = \begin{cases} \frac{6 \cdot P_a \cdot P_b \cdot HF}{8}, & \text{if shared} = 0 \\ \frac{P_b + 6 \cdot P_a \cdot P_b \cdot HF}{8}, & \text{if shared} = 1 \\ \frac{P_a + P_b + 6 \cdot P_a \cdot P_b \cdot HF}{8}, & \text{if shared} = 2 \end{cases} \]

Grandparent-grandchild; aunt/uncle-nephew-niece; half-siblings:

\[ P(E \mid \text{GG/AUNN/HS}) = \begin{cases} \frac{2 \cdot P_a \cdot P_b \cdot HF}{4}, & \text{if shared} = 0 \\ \frac{P_b + 2 \cdot P_a \cdot P_b \cdot HF}{4}, & \text{if shared} = 1 \\ \frac{P_a + P_b + 2 \cdot P_a \cdot P_b \cdot HF}{4}, & \text{if shared} = 2 \end{cases} \]

HF = 1 for homozygous loci and 2 for heterozygous loci
Familial search experiment

- Randomly pick sibling pair or unrelated pair from a synthetic database
- Choose one profile to be evidence and one profile to be initial suspect
- Test hypothesis:
  - $H_0$: A sibling is the source of the evidence
  - $H_A$: An unrelated person is the source of the evidence
Hypothesis testing using an LR threshold of 1

<table>
<thead>
<tr>
<th>Decision</th>
<th>True state</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Evidence from unrelated individual</td>
</tr>
<tr>
<td>Evidence from sibling</td>
<td>~2% [Type I error; false positive]</td>
</tr>
<tr>
<td>Evidence from sibling</td>
<td>~4% [Type II error; false negative]</td>
</tr>
<tr>
<td>Evidence from sibling</td>
<td>~98% [Correct decision]</td>
</tr>
<tr>
<td>Evidence from sibling</td>
<td>~96% [Correct decision]</td>
</tr>
<tr>
<td>Evidence from sibling</td>
<td>~96% [Correct decision]</td>
</tr>
</tbody>
</table>
Considering rarity of alleles

- As few as 5/26 rare alleles
- 13/26 average alleles
- 15/26 common alleles
Thresholds for similarity

- Virginia: “be very, very close”
- California: “appear useful”
- Florida: match at least 21 out of 26 alleles
- North Carolina: 16 out of 26 is enough
How well does an LR approach perform relative to alternatives?

- Low-stringency CODIS search identifies all 10,000 parent-child pairs (but only 1,183 sibling pairs and less than 3% of all other relationships and a high false positive rate)
- Moderate and high-stringency CODIS searches failed to identify any pairs for any relationship
- An allele count-threshold (set at 20 out of 30 alleles) identifies 4,233 siblings and 1,882 parent-child pairs (but fewer than 70 of any other relationship and with no false positives)
How well does an LR approach perform relative to alternatives?

- LR set at 1 identifies > 99% of both sibling and parent-child pairs (with false positive rates of 0.01% and 0.1%, respectively)
- LR set at 10,000 identifies 64% of siblings and 56% of parent-child pairs (with no false positives)
- Use of non-cognate allele frequencies results in an increase in false positives and a decrease in true positives (that are largely offset by either a ceiling or consensus approach)
How well does an LR approach perform relative to alternatives?

- LR set at 1 identifies > 78% of half-sibling, aunt-niece, and grandparent-grandchild pairs (with false positive rates at or below 9%)
- LR set at 1 identifies 58% of cousin pairs (with a 19% false positive rate)
- LR set at 10,000 identifies virtually no half-sibling, aunt-niece, grandparent-grandchild or cousin pairs (with no false positives)
How well does an LR approach perform with mixed samples?

- LR set at 1 identifies >99% of both sibling and parent-child pairs even in 2- and 3-person mixtures (with false positive rates of 10% and 15%, and of 0.01% and 0.07%, respectively).

- LR set at 1 identifies >86% of half-sibling, aunt-niece, and grandparent-grandchild pairs in 2- and 3-person mixtures (with false positive rates lower than 22% and 30%, respectively).

- LR set at 1 identifies >74% of cousin pairs in 2- and 3-person mixtures (with false positive rates of 41% and 49%, respectively).
“We’ve been doing familial searches for years. The difference between investigating identical twins and other siblings is just a matter of degree.”
Resources

• Internet

• Scientists
  – Jason Gilder (Forensic Bioinformatics)
  – Fred Bieber (Harvard University)
  – Sandy Zabel (Northwestern University)
  – Larry Mueller (UC, Irvine)
  – Keith Inman (Forensic Analytical, Haywood, CA)

• Publications