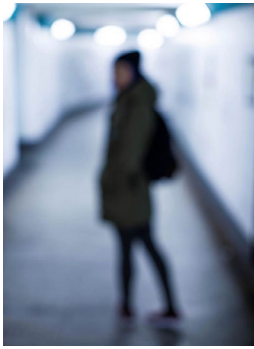


Interpretation of Y chromosome STRs for missing persons cases

Jianye Ge, Benjamin Crysup, Dixie Peters,
Meng Huang, Bruce Budowle

Center for Human Identification
University of North Texas Health Science Center (UNTHSC)

Y-STR for forensic applications



Most violent crimes are committed by men

- 99% of the forcible rapes
- 88% of the robberies
- 85% of the burglaries
- 88.8% of the homicide offenders

Y chromosome inheritance

- Only inherit from father to sons
- Can be used to trace male lineage

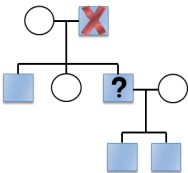
Y chromosome tests provide information for

- Kinship analysis
- Missing persons identification
- Familial searching
- Mixture in sexual assault cases
- Ancestry inference
- Link multiple cases
- Y database searching

Snyder HN. Arrest in the United States, 1990-2010. Washington, DC: Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice; 2012.
Homicide Trends in the United States, 1990-2008. Washington, DC: Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice; 2011.

Y-STR for missing persons cases

- Current autosomal STRs (20~25 loci) are insufficient to precisely determine distant relationships (e.g., first cousin)
- Y-STR can significantly increase or decrease the likelihood ratio (LR)
 - For 1st degree relatives, 6 most informative Y-STRs = 5 or 6 of the most informative autosomal STRs = 9 or 10 of the least informative autosomal STRs
 - Even better with distant relationships
 - LR = 1,000s with the major commercial Y-STR kits
- Probability of Exclusion (PE)
 - PE of 6 best Y-STRs = 99.7%
 - PE of 6 best autosomal STRs = 90% for unrelated, 99% for 1st degree relatives



1st degree = parent-child or full-sib

Ge J, Eisenberg A, Budowle B. Developing criteria and data to determine best options for expanding the core CODIS loci. Investigative Genetics. 2012 Dec;3(1):1-4.

Missing Persons Unit at UNTCHI hsc

- 650-1,000 remains samples processed/year
- 1,200-2,400 family reference samples/year
- Greater contribution of DNA profiles in CODIS/NDIS than all other MP lab

Remains at NDIS

Category	Count
UNTCHI	7526
Other US labs	6771

References at NDIS

Category	Count
UNTCHI	18358
Other US labs	10254

Current as of August 2021

SWGDAM guidelines hsc

- **SWGDAM Guidelines for Missing Persons Casework**

"For missing persons, relatives of missing persons and unidentified human remains samples, additional DNA methods other than autosomal STR typing (such as mtDNA or Y-STR typing) should always be considered, if relevant."

"A Y-STR LR and an mtDNA LR are computed for each of those systems."
- **SWGDAM Interpretation Guidelines for Y-Chromosome STR Typing by Forensic DNA Laboratories**
 - Focused on haplotype frequency estimation
 - No recommendation was given on how to calculate likelihood ratio for scenarios with multiple male references and/or mutations

A complex case - example hsc

The pedigree shows a family structure where individual 'a' is the father of 'b' and 'c', and 'c' is the father of 'd' and 'e'. Individual 'b' has a Y-STR haplotype of (11,20). Individual 'e' has a Y-STR haplotype of (11,21). Individual 'd' is missing, indicated by a yellow circle with a question mark. Individual 'a' is untyped, and 'c' is also untyped.

- A two-locus Y-STR pedigree including 2 typed references (b & e), 2 untyped individuals (a & c), and 1 missing person (d)
- The missing person is related to two references with different Y haplotypes

hsc⁺⁺

• Pedigree likelihood ratio (PLR) for lineage markers, similar to PLR for autosomal markers

Int J Legal Med (2011) 125:119–129
DOI 10.1007/s00414-010-0514-9

ORIGINAL ARTICLE

Pedigree likelihood ratio for lineage markers

Jianye Ge · Arthur Eisenberg · Jiangwei Yan · Ranajit Chakraborty · Bruce Budowle

• A convenient and accurate method to determine if two Y-STR profiles are from the same lineage

Electrophoresis 2010, 31, 1888–1890

Ha Liu¹,
 Jianyong Li¹,
 Junli He¹,
 Andrea Carbonari²,
 Marc Stone³,
 Jianye Ge¹

Research Article
A convenient guideline to determine if two Y-STR profiles are from the same lineage

The routine of forensic forensic and technology team. Y-STR is 200-300 profile analysis for criminal investigations. A large population of the Chinese population. Y-STR profiles to the database were generated using either the Applied Biosystems Y10 or Y10a1 Plus PCR Amplification kits. The additional for the Y10a1 Plus kit as compared to the Y10a1 kit results in a consistent candidate mutation rate increase across the loci.

hsc⁺⁺

Pedigree likelihood ratio (PLR)
for Y-STRs

hsc⁺⁺

Likelihood ratio with two profiles

• If two Y-STR profiles (A and B) match

$$LR = \frac{\Pr(A\&B|same\ lineage)}{\Pr(A\&B|different\ lineages)} = \frac{1}{p} \quad p \text{ is haplotype frequency}$$

• If two Y-STR profiles (A and B) do not match

$$LR_1 = \frac{\Pr(A\&B|same\ lineage)}{\Pr(A\&B|different\ lineages)} = \frac{\Pr(A) \times \Pr(A \rightarrow B)}{\Pr(A) \times \Pr(B)} = \frac{\Pr(A \rightarrow B)}{\Pr(B)}, \quad \text{if A is the ancestor}$$

$$LR_2 = \frac{\Pr(A\&B|same\ lineage)}{\Pr(A\&B|different\ lineages)} = \frac{\Pr(B) \times \Pr(B \rightarrow A)}{\Pr(A) \times \Pr(B)} = \frac{\Pr(B \rightarrow A)}{\Pr(A)}, \quad \text{if B is the ancestor}$$

Pr(A) = haplotype frequency of A
 Pr(B) = haplotype frequency of B
 Pr(A → B) = transmission probability from A to B with mutation considered
 Pr(B → A) = transmission probability from B to A with mutation considered

Pick the more conservative one, if we do not know which one is the ancestor

Example



Markers	DYS456	DYS389I	DYS390	DYS389II	Frequency
Profile 1	17	14	25	31	0.001
Profile 2	17	14	25	30	0.0005

• $Pr(31 \rightarrow 30) = Pr(30 \rightarrow 31) = 0.002$

$LR_1 = Pr(30 \rightarrow 31) / Pr(\text{Profile 1}) = 0.002 / 0.001 = 2$

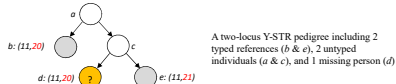
$LR_2 = Pr(31 \rightarrow 30) / Pr(\text{Profile 2}) = 0.002 / 0.0005 = 4$

• Use the more conservative LR (e.g., 2 in this example)

Pedigree likelihood ratio (PLR) with Y-STRs



• What if there are multiple Y-STR references with different haplotypes?



• Two competing hypotheses



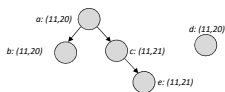
Pedigree likelihood (PL)



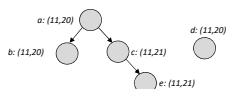
• PL = the cumulative haplotype frequency(ies) of the founder(s) × the cumulative transmission probabilities of all father-son pairs

• Predict possible haplotypes of the untyped individuals

- $a = (11,20), (11,21), \text{etc.}; c = (11,20), (11,21), \text{etc.}$
- Calculate likelihood for each haplotype combination (e.g., $a = (11,20)$ & $c = (11,21)$)



Pedigree likelihood (PL) – cont.

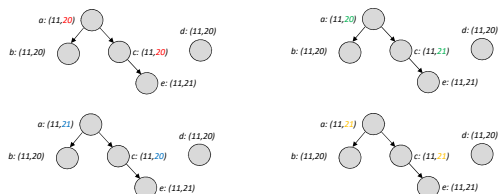


- Founders' haplotype frequencies = $\Pr(a) \times \Pr(d)$
 - Search Y-STR database (e.g., YHRD) to obtain haplotype frequencies
- Transmission probabilities = $\Pr(a \rightarrow b) \times \Pr(a \rightarrow c) \times \Pr(c \rightarrow e)$
 - $\Pr(c \rightarrow e) = \Pr(11 \rightarrow 11) \times \Pr(20 \rightarrow 21)$, assuming independent mutations between markers
 - Mutation rates and mutation model (e.g. Two Phase Model)

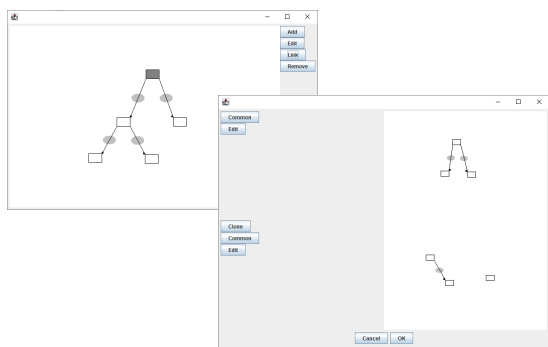
Pedigree likelihood (PL) – cont.



- Sum the likelihoods for all possible haplotype combinations for the untyped individuals
- $PL = L(G|a = (11,20) \ \& \ c = (11,20)) + L(G|a = (11,20) \ \& \ c = (11,21)) + L(G|a = (11,21) \ \& \ c = (11,20)) + L(G|a = (11,21) \ \& \ c = (11,21))$



Software under development



hsc⁺⁺

A convenient method to assess if two Y-STR profiles are from the same lineage

hsc⁺⁺

A convenient method to assess if two Y-STR profiles are from the same lineage

- PLR approach requires pedigree structure is defined, which may not be available in some cases
- Analysts just need to know if two Y-STR profiles are related or not
- A simple solution – use number of mismatched loci/steps
 - More distant relationship → More mismatched loci/steps

Example:
10 and 12 → 1 mismatched locus or 2 steps

- Y-STR profiles from the **same lineage** usually have **no or a small number of mismatched loci/steps**
- Y-STR profiles from **different lineages** usually have a **relatively large number of mismatched loci/steps**

hsc⁺⁺

Merge distributions and select a threshold

≤ 2 : More likely they are father-son
 ≥ 3 : More likely they are unrelated

How reliable is this method? hsc

- Study by Liu et al. (2016)
 - 7,405 Chinese unrelated Yfiler Plus profiles
 - Define a lineage as up to 4 meioses relationships (3rd degree or first-cousin)
 - ≤5: more likely from the same lineage
- The chance of being incorrect is extremely low
 - Two unrelated Yfiler Plus profiles, in most cases (i.e., >99.975%), will have more than 5 mismatched steps
 - For almost all close relatives, the mismatched steps are ≤5

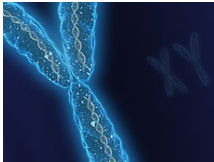
Relationship	Chance of mismatched steps ≤5
Father-son	99.999997%
Grandfather-Grandson; brother	99.999984%
Uncle-nephew	99.999950%
First-Cousin	99.999878%

- Genotyping error (i.e., ~0.1%) is greater than the error associated with this method
- Accuracy will be lower if the lineage definition includes more distant relatives




Liu H, Li X, Muller J, Carbonaro A, Short M, Ge J. A convenient guideline to determine if two Y-STR profiles are from the same lineage. *Electrophoresis*. 2016 Jul 23;37(12):2659-68. doi: 10.1002/elps.201600199.

Summary hsc

- Interpretation methods for complex cases
 - Pedigree likelihood ratio
 - A convenient method based on mismatches
- Interpretation software under development
 - Free to access
 - Ready by Q1 2022



ACKNOWLEDGMENTS

- UNTCHI
 - Bruce Budowle
 - Benjamin Crystup
 - Dixie Peters
 - Melody Josserand
 - Meng Huang

This study was supported in part by award 2020-DQ-BX-0018 (Interpretation of Y chromosome STRs for missing persons cases), awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the U.S. Department of Justice.
