Automated Familial Search using a Probabilistic Genotype Database

Matthew Legler[1]; Kevin Miller[2], PhD;
Mark Perlin[1], PhD, MD, PhD; Garett Sugimoto[3], MS

[1]Cybergenetics Corp, Pennsylvania
[2]Hamilton Robotics, Nevada
[3]Kern Regional Crime Laboratory, California

Cybergenetics

Familial searching

Person left DNA

Who is he?

DNA evidence
from a
crime scene

Familial searching

Person left DNA

No

DNA evidence
from a
crime scene

In DNA database?
He's not.

Familial searching

Person left DNA

DNA evidence
from a
crime scene

Yes   No

In DNA database?

But his brother is!

Familial searching

Person left DNA

LR
Database match

DNA evidence
from a
crime scene

Yes   No

In DNA database?

Find his brother to find him.
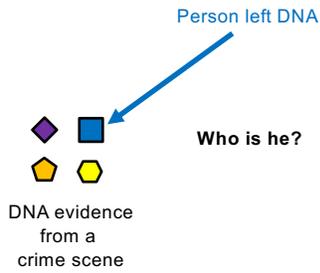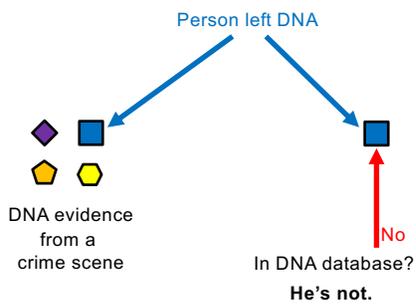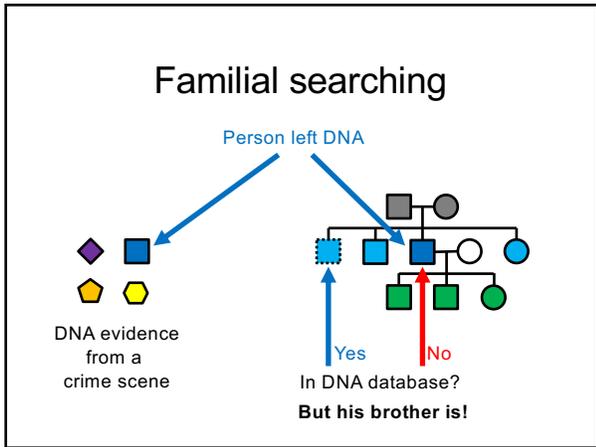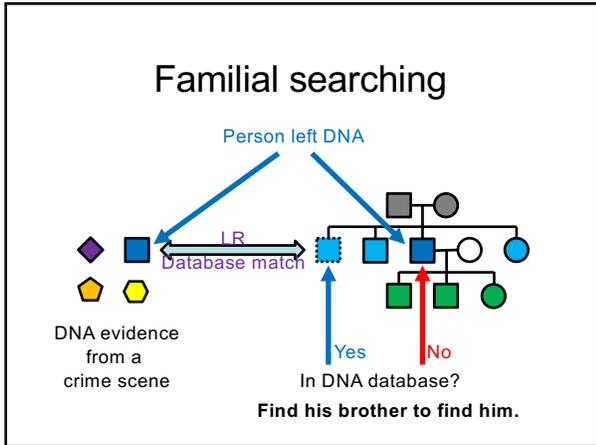
Probabilistic genotype database

|	Old	New
DNA unit	Allele	Genotype
Peak height	Not quantitative	Use all data
PCR variation	No data model	Model the experiment
Uncertainty	Discard data	Consider all genotypes
Partial profile	Allele list	Genotype probability
Partial match	Reference profile	Inferred relative genotype
Comparison	Allele inclusion	Likelihood ratio
Information	Lose	Preserve
Database	CODIS	TrueAllele

TrueAllele® automation

Start with a DNA database of reference genotypes

Reference genotype.

Upload reference genotype to TrueAllele database.

TrueAllele® automation

Add family prob genotypes to the DNA database
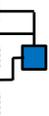
Parent	PCH	
Sibling	SIB	
Child	PCH	

Add prob genotype family.

Infer and upload family prob genotypes to TrueAllele database.

TrueAllele® automation

A person left their DNA at a crime scene

Parent	PCH	
Sibling	SIB	
Child	PCH	

DNA evidence (EVI) from a crime scene
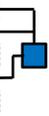
Have prob genotype family.

Inferred family in TrueAllele database.

TrueAllele® automation

Find the person through their relative's DNA

DNA evidence (EVI) from a crime scene

Parent — PCH
Sibling — SIB
Child — PCH

LR
Database match

Match prob genotype family

to find crime scene relative in TrueAllele database.
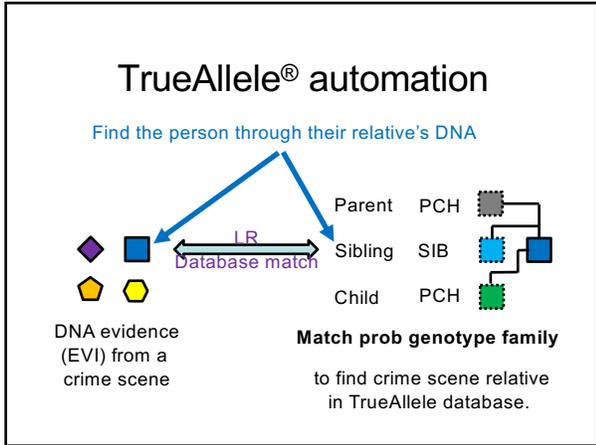
Kern lab's validation studies

The goal of this work was to validate the use of the TrueAllele® VUIer™ Software (version 3.3.6228.1) Kinship Application in forensic casework.

Studies:
1. Accuracy
2. Sensitivity and Specificity
3. Known and Nonprobative Evidence

---

1. Accuracy study

Purpose – to demonstrate how the kinship application may be used to match various inferred kinship genotypes to single source evidence genotypes.

Method – DNA samples from four families were tested:
1. Family 1 – 12 known individuals spanning 3 generations
2. Family 2 – 4 known individuals, all immediate family members
3. Family 3 – 6 known individuals with half siblings included
4. Family 4 - 5 known individuals, all immediate family members

Accuracy results

Using the 26 samples, there were 676 pairwise comparisons made for each assumed genotype group:
- PCH – Inferred genotypes of parent/child of known
- SIB – Inferred genotypes of full sibling(s) of known
- HAS – Inferred genotypes of half sibling(s), grandparent(s) / grandchildren, uncle-aunt / niece-nephew of known

for a total of more than 2,000 comparisons.

Result – EVI-PCH and EVI-SIB searches could possibly be used to identify matches between single source evidence items and inferred PCH or SIB genotypes of the parents, children and full siblings of the true contributor.

---

2. Sensitivity & specificity study

TrueAllele inferred genotypes (PCH and SIB) of true contributors were compared to various mixed DNA samples to evaluate the rate of true matches and the corresponding match scores (likelihood ratios).

31 single source profiles and 36 mixed samples (ranging from 2 to 5 contributors) were compared to the inferred genotypes of 44 single source known reference profiles; including many relatives of the true contributors.

Evidence vs. parent/child

Mixture weight of contributor	Count	PCH Search-Self Percent	PCH Search-Self Average	PCH Search-True PCH Percent	PCH Search-True PCH Average	PCH Search-True SIB Percent	PCH Search-True SIB Average	PCH Search-False + count	PCH Search-False + Average
0 -- 10	48	72.92%	5.61	14.63%	3.53	23.81%	3.04	1	2.12
>10 -- 20	11	100.00%	7.19	66.67%	3.87	100.00%	4.53	0.00	N/A
>20 -- 30	13	100.00%	7.76	75.00%	4.1	57.14%	4.72	0.00	N/A
>30 -- 40	24	100.00%	8.61	76.92%	4.41	80.00%	4.99	0.00	N/A
>40 -- 50	13	100.00%	9.47	100.00%	4.11	70.00%	4.62	0.00	N/A
>50 -- 60	2	100.00%	8.47	100.00%	4.13	100.00%	4.7	0.00	N/A
>60 -- 70	2	100.00%	11.98	100.00%	6.56	100.00%	6.47	0.00	N/A
>70 -- 80	2	100.00%	11.24	100.00%	6.47	N/A	N/A	0.00	N/A
>80 -- 90	5	100.00%	11.68	100.00%	5.205	0.00%	N/A	0.00	N/A
>90 -- 100	28	100.00%	13.52	100.00%	7.96	100.00%	6.71	0.00	N/A

EVI-PCH searches appeared to be more specific for inferred genotypes of closely related relatives of known contributors

Evidence vs. sibling

Mixture weight of contributor	Count	SIB Search-Self Percent	SIB Search-Self Average	SIB Search-True SIB Percent	SIB Search-True SIB Average	SIB Search-True PCH Percent	SIB Search-True PCH Average	Sib Search-true HAS Percent	Sib Search-true HAS Average	Sib Search-False + count	Sib Search-False + Average
0 – 10	48	70.83%	6.53	28.57%	3.021	17.07%	3.1	33.33%	2.12	0	N/A
>10 – 20	11	100.00%	8.24	100.00%	4.82	66.67%	3.32	75.00%	2.65	0	N/A
>20 – 30	13	100.00%	9.24	85.71%	4.88	75.00%	3.59	0.00%	N/A	1	2.65
>30 – 40	24	100.00%	10.32	80.00%	5.75	76.92%	3.87	50.00%	2.96	0	N/A
>40 – 50	13	100.00%	11.18	100.00%	4.43	92.31%	4.54	100.00%	3.56	0	N/A
>50 – 60	2	100.00%	10.29	100.00%	3.13	100.00%	4.14	0.00%	N/A	1	2.2
>60 – 70	2	100.00%	14.57	100.00%	7.96	100.00%	5.33	0.00%	N/A	1	2.48
>70 – 80	2	100.00%	14.25	N/A	N/A	100.00%	6.31	100.00%	3.58	1	2.55
>80 – 90	5	100.00%	13.74	100.00%	4.41	100.00%	4.75	33.33%	2.11	1	2.2
>90 – 100	28	100.00%	18.27	100.00%	10.22	100.00%	9.27	83.33%	4.07	11	2.88

EVI-SIB searches were more sensitive for inferred genotypes of more distant relatives of known contributors

---

3. Nonprobative evidence and known study

2012 sexual assault cold case

20 evidence items, including:
Bra cups – left & right
Unworn shirt – nsp & sp

2 non-suspect known references:
Consenting partner (CP)
Consenting partner's father (CPF)

Casework DNA matches

Evidence   **Reference**

Bra left ⟺ Consenting partner
Bra right

Shirt nsp ⟺ Consenting partner's father
Shirt sp

+ 16 case items   No matches to CP or CPF

Cybergenetics © 2003-2022

Automated familial search

Database match with LR over a million

Evidence   **Reference**

Bra left
Bra right   ⟷   Consenting partner's
                    4   inferred parent-child
Shirt nsp   *probabilistic genotype*
Shirt sp

+ 5,000 DB items   No other matches
                   Nor to CP or CPF

Fully automated familial database

Infer mixture & kinship genotypes; compare them.
Fully automated process; no human involvement.
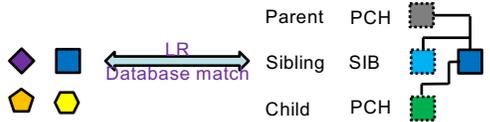
Parent   PCH

LR
Database match   Sibling   SIB

Child   PCH

DNA evidence
(EVI) from a   **Match prob genotype family**
crime scene
                   to find crime scene relative
                   in TrueAllele database.

Conclusions

• Familial search finds investigative leads
• Automated TrueAllele familial searching
• Kern laboratory protocol development
• Three familial search validation studies
• Parent/child better than siblings
• Reuses existing DNA evidence data
• Makes familial database search routine

Computer-automated familial search workflow