How TrueAllele® Computing Automates DNA Analysis and Databasing for Mass Disasters

Wecht Institute of Forensic Science & Law
Duquesne University
September, 2021
Pittsburgh, PA

Mark W Perlin, PhD, MD, PhD
Pittsburgh, PA

Cybergenetics

JUSTICE THROUGH SCIENCE

Meet Sarah

World Trade Center attack

18,000 victim remains collected
2,700 missing people

Identifying victim remains

Sarah and others' remains became charred pebbles

Task: associate the victim remains with the missing people

Closure
- Medical examiner confirms death
- Returns victim remains to family

DNA from biological evidence

Same DNA in victim remains as DNA from missing people

Approach
Compare victim remains DNA with missing people DNA to connect them through same DNA

Not the usual DNA problem

Victim remains
Small amounts of damaged charred DNA

Missing people
No person's DNA to compare with, just relatives & personal effects

Need a sophisticated DNA approach to associate remains with missing

Sarah's genotype

In nature, we inherit two genomes,
one from Mom and one from Dad

In science, we test a dozen locations,
to find out the pair of inherited alleles

At each genetic locus, that pair of alleles
is the person's **genotype**

Amplifying the DNA

We can't detect just dozens of cells,
but PCR can make us billions of copies

a+b 2a+2b 4a+4b 8a+8b 16a+16b etc

But PCR is an imperfect amplifier,
and so introduces **randomness**

a+b 2a+2b 3a+4b 5a+7b 9a+13b etc

Inferring genotypes

Inferred genotype is only known up to probability

Absolute genotyping
Human review fails when not one clear answer

Determine PCR randomness from DNA data
to produce accurate genotype probability
(Perlin filed patent in 2001)

Probabilistic genotyping
Assigns probability to every allele pair

Comparing genotypes

Inferred genotype vs. reference genotype provides match strength (LR statistic)

$LR = \frac{q}{p}$

Inferred genotype

Reference genotype

WTC victim remains

Which charred pebble came from Sarah?

18,000 victim remains
100,000 DNA tests
(multiple tests on one item)

Limited genotype inference

Older absolute methods can't interpret imperfect data

	Absolute
Input	Discard low data
Output	Alleles
Decisions	Make choices
DNA data	Change data
Approach	Black or white
Result	None, inaccurate
Tests	One at a time

Modern genotype inference

Newer probabilistic methods can interpret any data

	Absolute	Probabilistic
Input	Discard low data	Use all data
Output	Alleles	Genotypes
Decisions	Make choices	Consider everything
DNA data	Change data	Preserve data
Approach	Black or white	Shades of gray
Result	None, inaccurate	Accurate answers
Tests	One at a time	Joint analysis

Joint data analysis

Combine DNA data from different tests for more information

Multiple tests
Profiler Plus
Cofiler
PowerPlex 16
BodePlex 1
BodePlex 2
Big Mini

Produces more definite genotypes, stronger associations

More testing, more information

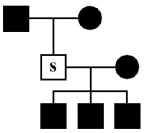
WTC missing people

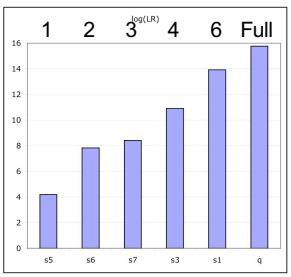
Relatives: share inherited DNA
why spouse is important
Personal effects: mixtures, low level

Reconstruct genotypes from the missing people
For comparison with victim remains genotypes

Genotypes from family

More relatives, more information

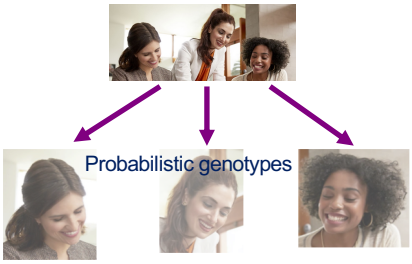




Genotypes from personal effects

Unmixing mixtures

Separating genotypes from mixed DNA items



Relatives + Effects = Information

Reconstructing the missing: Sarah

Comparing uncertain genotypes

Inferred VR genotype vs. inferred MP genotype provides match strength (LR statistic)

$LR = \sum \frac{q \cdot r}{p}$

Inferred genotype from victim remains

Inferred genotype from relatives & effects

Just probability: no reference needed

TrueAllele® System

Matching Database

User Interface

Genotype Inference

Entering DNA data

Upload data
Upload requests
(visual check)

User Interface

Coordinating computing

Matching Database

Store data & requests
Store genotypes
Compare genotypes

Inferring genotypes

Victim remains

Missing persons

Probabilistic genotypes
preserve identification information

Parallel computing

Genotype Inference

Matching genotypes

Automated
Incremental
Informative

Matching Database

LR = $\sum \frac{q \cdot r}{p}$

Returning Sarah to her family

Identified victim remains

Validating TrueAllele

Concordance on a set of 150 comparisons
between WTC victim remains and missing persons,
using both TrueAllele and NYC lab methods

Dickerson TM, Gajewski C, Ishii A, Desire M, Prinz MK.
"Renewed efforts to identify the victims of the
World Trade Center disaster via DNA testing" (A81).
American Academy of Forensic Sciences 64th Annual Meeting;
Atlanta, GA: AAFS; 2012. p. 73.

Cybergenetics then filed the full WTC comparison results of
18,000 victim remains versus 2,700 missing persons

Scalable database

Rack system, to desktop, to cloud computing
Scalable to any size problem
Can use any DNA marker test (NGS)
Automated computer operation
Genotype matching database
Designed for probabilistic genotypes

Use in criminal justice

- prosecution, defense, innocence
- reported in 45 states
- over 40 validation studies
  (8 peer reviewed)
- over 30 admissibility decisions
- used in 10,000's of cases

J. Oblock and N. Butt,
"The use of a database feature in the TrueAllele®
Casework system to cross-reference DNA cases",
American Academy of Forensic Sciences 71th Annual Meeting,
Baltimore, MD, 23-Feb-2019.

32

TrueAllele® for mass disasters

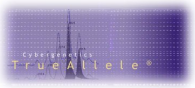
Matching Database

User Interface

Genotype Inference

More information

http://www.cybgen.com/information



- Courses
- Newsletters
- Newsroom
- Presentations
- Publications
- Webinars

http://www.youtube.com/user/TrueAllele
TrueAllele YouTube channel

Cybergenetics

YouTube

perlin@cybgen.com

JUSTICE THROUGH SCIENCE