Science & Sorcery in
Forensic DNA Evidence

Science in our World: Certainty & Controversy
Penn State Beaver
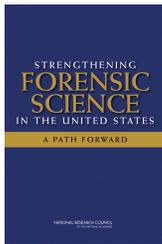**April, 2017**
Monaca, PA

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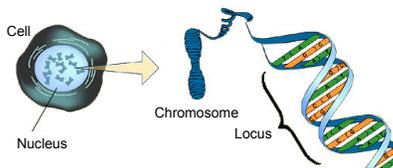
Cybergenetics

Cybergenetics © 2003-2017

National Academy of Sciences

STRENGTHENING
FORENSIC
SCIENCE
IN THE UNITED STATES
A PATH FORWARD

NATIONAL RESEARCH COUNCIL

Trouble in River City
2009

Where's the
"science" in
Forensic Science?

DNA biology

Cell

Nucleus

Chromosome
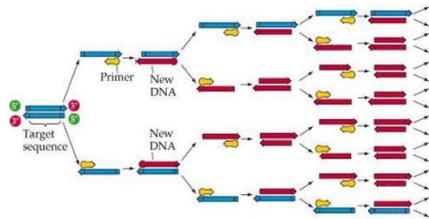
Locus

Short tandem repeat

DNA locus paragraph

Take me out to the ball game
take me out with the crowd
buy me some peanuts and Cracker Jack
I don't care if I never get back
let me
root root root root root root root root root root
for the home team,
if they don't win, it's a shame
for it's one, two, three strikes, you're out
at the old ball game

23 volumes in cell's
DNA encyclopedia

"root" repeated 10 times, so
allele length is 10 repeats

1985

Polymerase chain reaction

Primer
New DNA
Target sequence
New DNA

DNA genotype

Chromosome

locus

U.S. National Library of Medicine

mother allele
1 2 3 4 5 6 7 8 9 10

ACGT repeated word
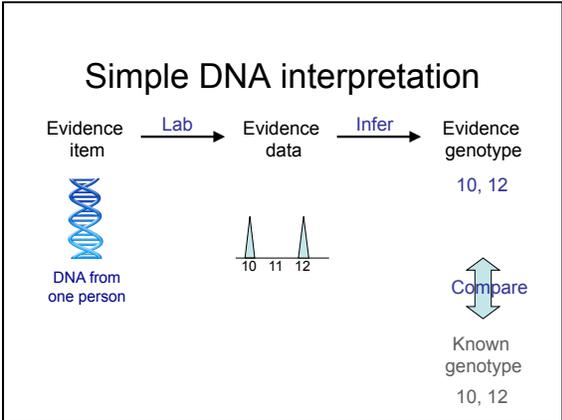
father allele
1 2 3 4 5 6 7 8 9 10 11 12

A genetic locus has
two DNA sentences,
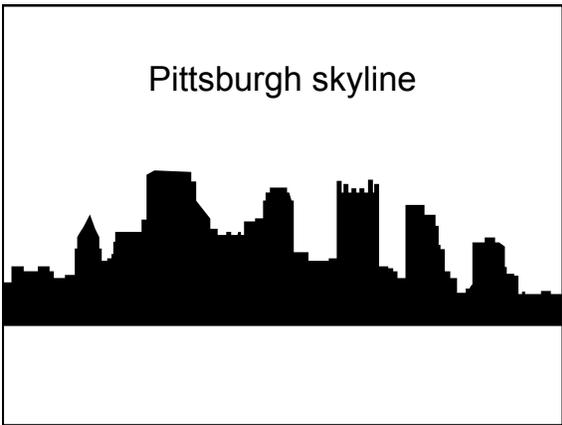one from each parent.

An allele is the number
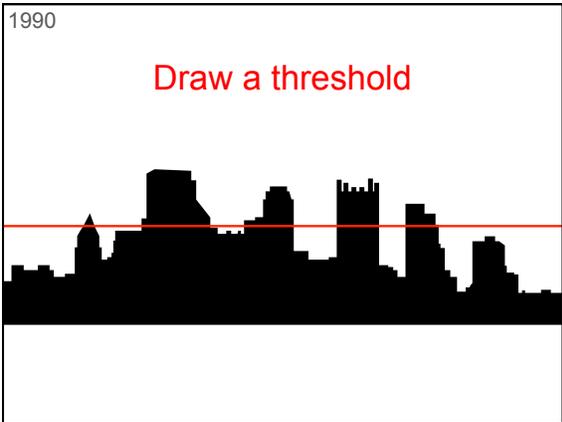of repeated words.

A genotype at a locus
is a pair of alleles.

10, 12

Many alleles allow for
many many allele pairs.
A person's genotype
is relatively unique.

Simple DNA interpretation

Evidence item →(Lab)→ Evidence data →(Infer)→ Evidence genotype

10, 12

DNA from one person

10 11 12

Compare

Known genotype

10, 12

Pittsburgh skyline

1990

Draw a threshold

Same heights & vacant lots

1995

DNA mixture

eye of newt   toe of frog

Double, double toil and trouble

2000

Simplify data to interpret mixture

f_1 f_2 f_3 f_4 f_5

1. Apply threshold to find "alleles"
2. Add allele frequencies ($f_1 + f_2 + \ldots$)
3. Square sum, take reciprocal
4. Locus "probability of inclusion" (PI)
5. Multiply locus PI values
6. Combined (CPI) **match statistic**

Natural variation

Unreliable DNA mixture statistics

NIST (Commerce Department) study in 2005
Two contributor mixture data, known victim
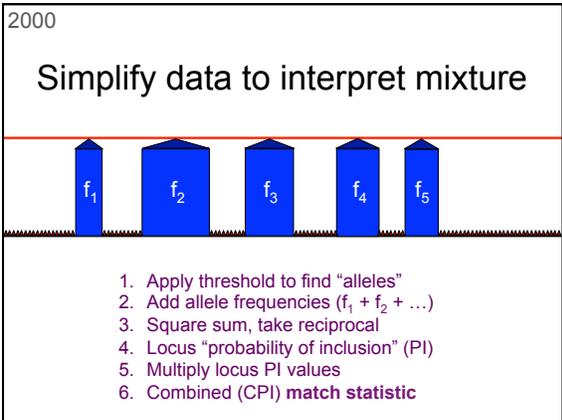
Some Differences in Reporting Statistics				
			Case1	
LabID	Kits Used	Caucasians	African Americans	Hispanics
90	ProPlus/Cofiler	1.18E+15	2.13E+14	3.09E+15
34	ProPlus/Cofiler	2.40E+11	7.90E+08	9.80E+10
33	ProPlus/Cofiler	2.94E+08	1.12E+08	1.74E+09
6	ProPlus/Cofiler	40,000,000	3,500,000	280,000,000
9	ProPlus/Cofiler	1.14E+07	1.97E+07	1.54E+08
79	ProPlus/Cofiler	930,000	47,900	1,350,000
16	ProPlus/Cofiler	434,600	31,710	399,100
Remember that these labs are interpreting the same MIX05 electropherograms				

When not
"inconclusive":

213 trillion (14)

31 thousand (4)

Forensic DNA labs put on notice ten years ago

Biased DNA workflow

(1)
Choose data

(3)
Person decides

(2)
Calculate statistic

1N
0UT

• Put people in the process
• To overcome software limits
• And introduce human bias

Falsely identify innocent people

MIX13 Case 5 Outcomes with Suspect C
(whose genotypes were not present in the mixture)

# Labs	Report Conclusions	Reasons given
6	Exclude Suspect C	detailed genotype checks (ID+); TrueAllele negative LR (ID+); assumed major/minor and suspects did not fit (ID+); 3 labs noted Penta E missing allele 15 (PP16HS)
3	Inconclusive with C only (A & B included)	All these labs used PP16HS
21	Inconclusive for A, B, and C	
70	**Include** & provide CPI statistics	*All over the road…*

Range of CPI stats for Caucasian population:
FBI allele frequencies: **1 in 9** to **1 in 344,000**

Mixture statistics shut down labs

"National accreditation board suspends all
DNA testing at D.C. crime lab"
𝕿𝖍𝖊 𝖂𝖆𝖘𝖍𝖎𝖓𝖌𝖙𝖔𝖓 𝕻𝖔𝖘𝖙 *April 27, 2015*
Did not comply with FBI standards

"New protocol leads to reviews of
'mixed DNA' evidence"
𝕿𝖍𝖊 𝕿𝖊𝖝𝖆𝖘 𝕿𝖗𝖎𝖇𝖚𝖓𝖊 *September 12, 2015*
24,468 lab tests affected

Statistics lack scientific basis

Misled courts for 15 years on countless DNA mixtures

J Pathol Inform

Research Article

Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information
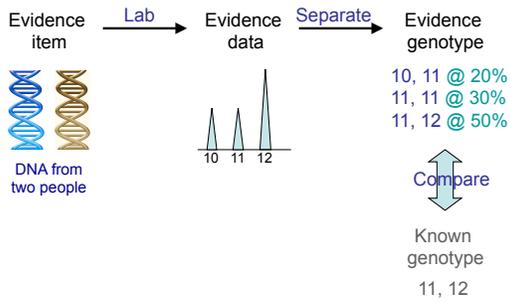
Mark William Perlin[1]

[1]Cybergenetics, Pittsburgh, USA

E-mail: *Dr. Mark William Perlin - perlin@cybgen.com
*Corresponding author

TrueAllele® computer technology

• Accurate. 34 validation studies, 7 published

• Objective. Workflow removes human bias

• Accepted. Reported in 37 states, used by labs

• Transparent. Give math, software (4GB DVD)

• Neutral. Can statistically include or exclude

DNA mixture interpretation

Evidence item → Lab → Evidence data → Separate → Evidence genotype

DNA from two people

10 11 12

10, 11 @ 20%
11, 11 @ 30%
11, 12 @ 50%

Compare

Known genotype

11, 12

Crime labs don't use all DNA data

all-or-none
1+1 = 1
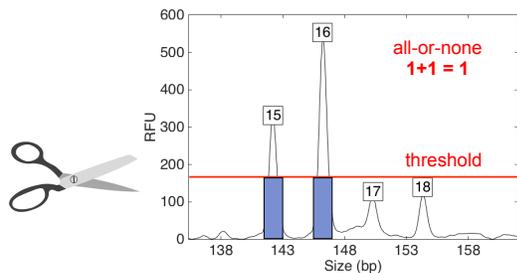
threshold

Threshold

Mark Perlin © 2011

I've been dealt a losing hand
Drawing lines across the sand
Whose meaning I can't understand
Cause mathematics has been banned

Used to think that one and one
Added up to two in sum
Now they say it's all or none
Arithmetic has been undone

Sing along

Na na-na-na na na
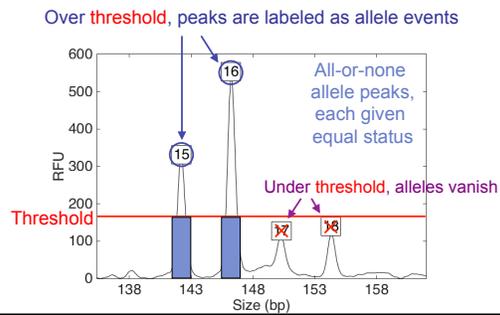Na-na na na na-na na
Na na-na-na na na
Na-na na na na-na na

soundcloud.com/markperlin/threshold

Wolfe sisters homicide

On February 6, 2014, Susan Wolfe (44)
and her younger sister Sarah (38, left)
were killed in their East Liberty home in Pittsburgh.

Crime lab didn't use all DNA data

Over threshold, peaks are labeled as allele events

All-or-none allele peaks, each given equal status

Under threshold, alleles vanish

Threshold

RFU

Size (bp)

Threshold 2

Science should let numbers talk
Not let perpetrators walk
Who can count the victims lost
To criminals who've not been caught

Draw a threshold throw away
The evidence from DNA
Release a killer who should stay
In prison till his dying day

Pennsylvania v. Allen Wade

Thresholds failed to interpret most DNA mixtures

Hat	No conclusions
Cup	Insufficient data
Fingernails	Contamination, insufficient data
Gear shift	Insufficient data
Seat lever	Cannot be excluded
Knit hat	Insufficient data
Sock	Too complex, no conclusions

Computers can use all the data

Quantitative peaks at locus vWA
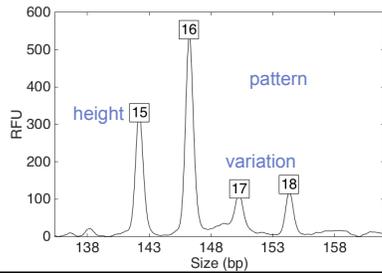
Threshold 3

I am a public employee
The people pay my salary
Yet I pretend I cannot see
The data right in front of me

The rules are written from on high
I follow them but know not why
Trained to make the numbers lie
As innocents are raped or die

Pennsylvania v. Allen Wade

The crime lab reported 5 DNA mixture matches
TrueAllele found 17 matches on the same data

Hat	65.3 thousand	Allen Wade
Cup	20.5 thousand	Susan Wolfe
Fingernails	6.06 trillion	Allen Wade
Gear shift	9.37 million	Sarah Wolfe
Seat lever	385 billion	Sarah Wolfe
Knit hat	25.7 thousand	Allen Wade
Sock	300	Sarah Wolfe

Allen Wade Found Guilty On All Counts In East Liberty Sisters' Slaying

CBS News, May 23, 2016

PITTSBURGH (KDKA/AP)
- A man accused of killing two sisters who lived next door to him in East Liberty has been found guilty on all counts.
- Allen Wade was accused of shooting Sarah and Susan Wolfe after they returned from work on Feb. 6, 2014, apparently to steal a bank card.
- On Monday morning, a jury found Wade guilty of first-degree murder, robbery, burglary and theft by unlawful taking.

Pennsylvania v. Allen Wade

Thresholds failed to interpret DNA mixture
TrueAllele succeeded on the same data

A hat left from a burglary of the Wolfe sister's home
six weeks before the murder matched
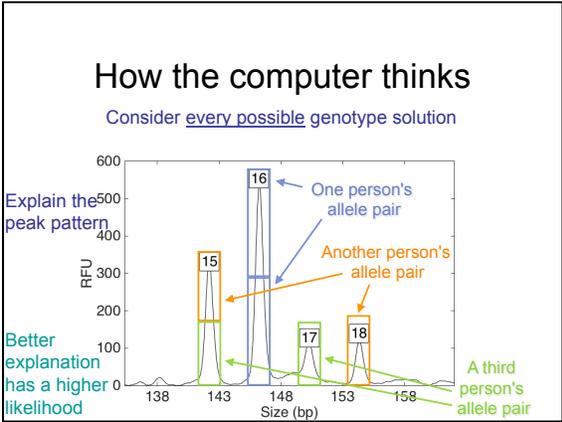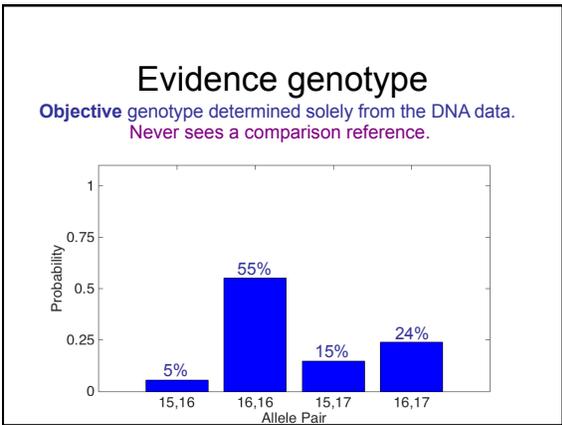Allen Wade with a 65.3 thousand statistic

Preventable Crime

Threshold 4

I rather would use TrueAllele
Interpret DNA for real
Let the evidence reveal
Thugs who slash and shoot and steal

But instead I make believe
Ignoring what I plainly see
Science slips to sorcery
As thresholds make a mockery
Of DNA as it should be

How the computer thinks

Consider <u>every possible</u> genotype solution

Explain the peak pattern

Better explanation has a higher likelihood

- One person's allele pair
- Another person's allele pair
- A third person's allele pair

Evidence genotype

Objective genotype determined solely from the DNA data. Never sees a comparison reference.

- 15,16 — 5%
- 16,16 — 55%
- 15,17 — 15%
- 16,17 — 24%

DNA match information

How much more does the suspect match the evidence than a random person?

- 11x
- 55% — Prob(evidence match)
- 5% — Prob(coincidental match)

Match information at 15 loci

DNA Locus	Likelihood Ratio
CSF1PO	
D13S317	
D16S539	
D18S51	
D21S11	
D3S1358	
D5S818	
D7S820	
D8S1179	
FGA	
Penta_D	
Penta_E	
TH01	
TPOX	
vWA	

(Horizontal axis: Likelihood Ratio, 0 to 180)

Is the suspect in the evidence?

A match between the right fingernails
and Allen Wade is:

6.06 trillion times more probable than
a coincidental match to an unrelated Black person

32.5 trillion times more probable than
a coincidental match to an unrelated Caucasian person

8 trillion times more probable than
a coincidental match to an unrelated Hispanic person

Allegheny County: 59 cases, 12 trials, 3 exonerations

Crime	Evidence	Defendant	Outcome	Sentence
rape	clothing	Ralph Skundrich	guilty	75 years
murder	gun, hat	Leland Davis	guilty	23 years
rape	clothing	Akaninyene Akan	guilty	32 years
murder	shotgun shells	James Yeckel, Jr.	guilty plea	25 years
murder	fingernail	Anthony Morgan	guilty	life
weapons	gun	Thomas Doswell	guilty plea	1 year
bank robbery	clothing	Jesse Lumberger	guilty	10 years
drugs	gun	Derek McKissick	guilty plea	2 1/2 years
drugs	gun	Steve Morgan	guilty plea	2 1/2 years
murder	door, clothing	Calvin Kane	guilty plea	20 years
murder	fingernail, clothing	Allen Wade	guilty	life
murder	gun	Jaykwaan Pinckney	guilty plea	10 years
child rape	clothing	Dhaque Jones	guilty plea	6 years
shooting	gun	Anthony Jefferson	guilty plea	4 years
weapons	gun	Zachary Blair	guilty plea	15 years
weapons	gun	Delmingo Williams	guilty plea	3 years
incest rape	clothing	Terry L.	guilty	40 years
bank robbery	hat	Robert Schatzman	guilty	1 1/2 years
weapons	gun	Rashawn Walker	guilty	1.5 years
robbery	hat	Lauren Peak	guilty plea	1 year
rape	body cavity	Freddie Cole	guilty plea	2 years
murder	gun	Chaz White	guilty plea	4 years
sex crime	clothing	Christopher Stavish	guilty plea	2 1/2 years
murder	gun	Jake Knight	guilty	life

Darryl Pinkins imprisoned

1989 – 5 men raped an Indiana woman
Darryl Pinkins and 2 others misidentified
1991 – wrongfully convicted, 65 year sentence

2001 – DNA mixture evidence
2 contributors found, not the accused
but 5 were needed, post-conviction relief denied

TrueAllele Pinkins findings

1. compared *evidence with evidence*
2. calculated *exclusionary match statistics*
3. revealed 5% *minor mixture contributor*
4. *jointly analyzed* DNA mixture data
5. showed three perpetrators were *brothers*

found 5 unidentified genotypes,
defendants not linked to the crime

Pinkins exonerated

April 25, 2016

Certainty & controversy

Why do crime labs fail to interpret DNA?

- Expertise in forensic biology
  - Great at generating DNA data
- Not in computational statistics
  - Poor at interpreting the data
- Adversarial nature of court
  - Cross-examination, termination
- Strong incentives on lab process
  - Focus on data, not information
- Fear of uncertainty, comfort in certainty
  - Saying nothing avoids controversy

Recommendations

1. Open DNA to public scrutiny
2. Revisit all past DNA mixture cases
3. Educate trial attorneys and judges
4. Fully automate mixture interpretation
5. Extensively validate DNA interpretation
6. Keep methods within their limits
7. Go beyond laboratory limits

Nonprofit organization

Justice Through Science

1. Reanalyze forgotten DNA evidence

2. Educate lawyers, scientists & society

More information

http://www.cybgen.com/information
- Courses
- Newsletters
- Newsroom
- Presentations
- Publications
- Webinars

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

Cybergenetics

You Tube

perlin@cybgen.com