

Reporting match error: casework, validation & language

Cybergenetics Webinar
User Group Meeting
December, 2018
Pittsburgh, PA

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



Cybergenetics

Cybergenetics © 2003-2018

How often would evidence
match the wrong person as
strongly as the defendant?

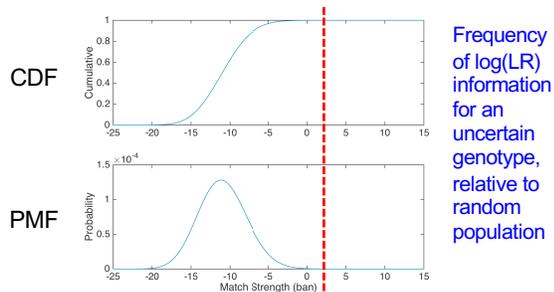
Evidence

information
likelihood ratio
unfamiliar concept

How often

probability
frequency
familiar concept

Non-contributor distribution



Tail region gives error: probability of misleading evidence (PME)

Genotype **prior** probability

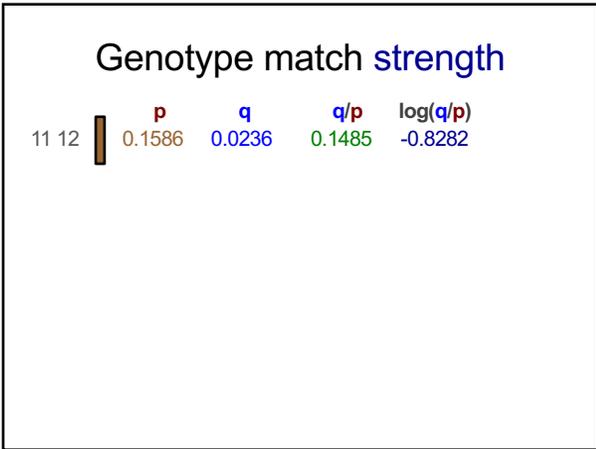
11 12  ^p
0.1586

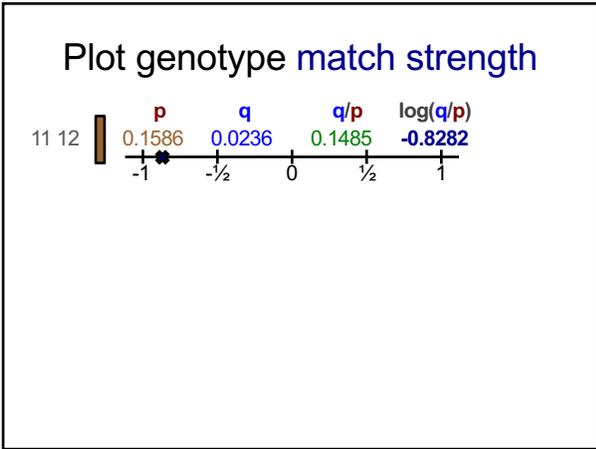
Genotype **posterior** probability

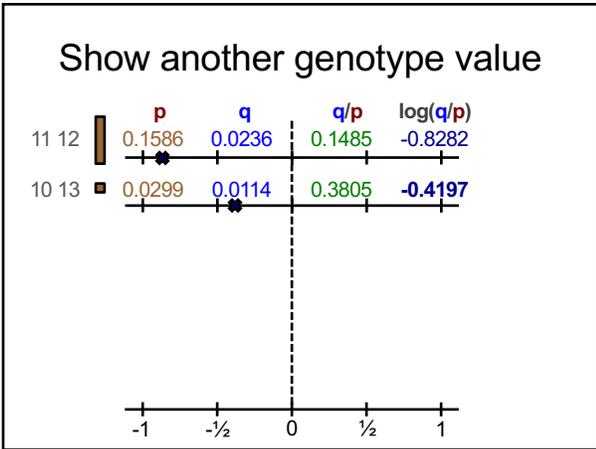
11 12  ^p 0.1586 ^q 0.0236

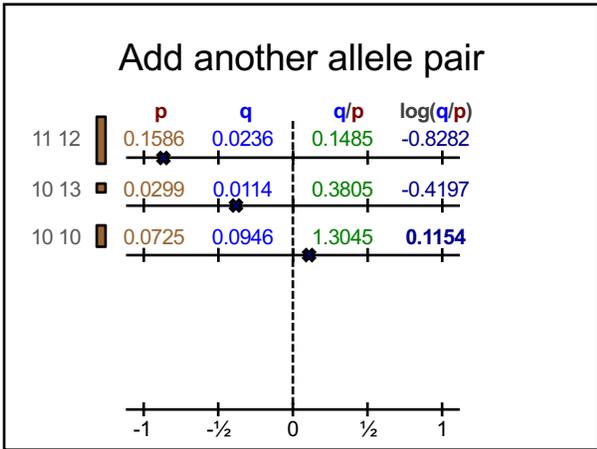
Genotype probability **ratio** (LR)

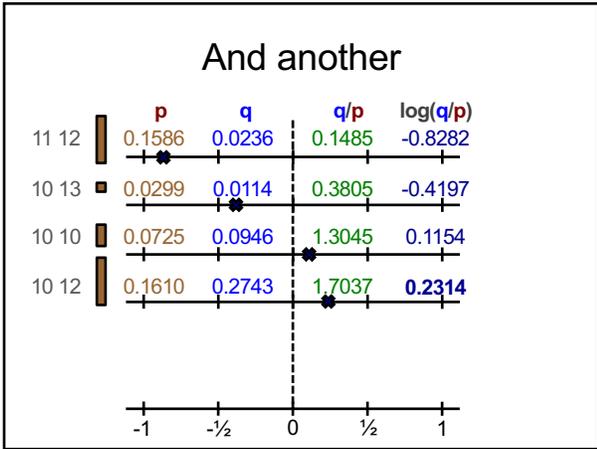
11 12  ^p 0.1586 ^q 0.0236 ^{q/p} 0.1485

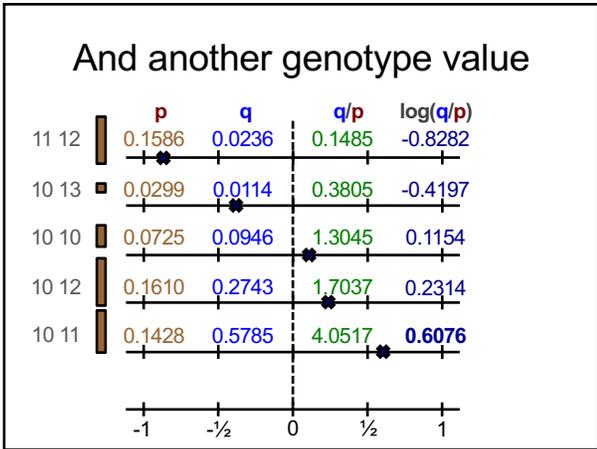


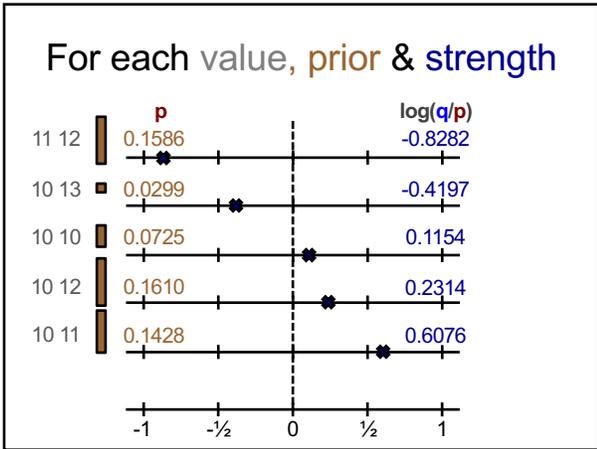


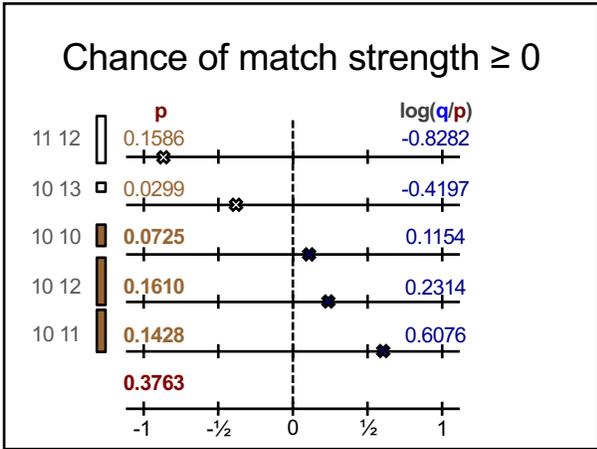


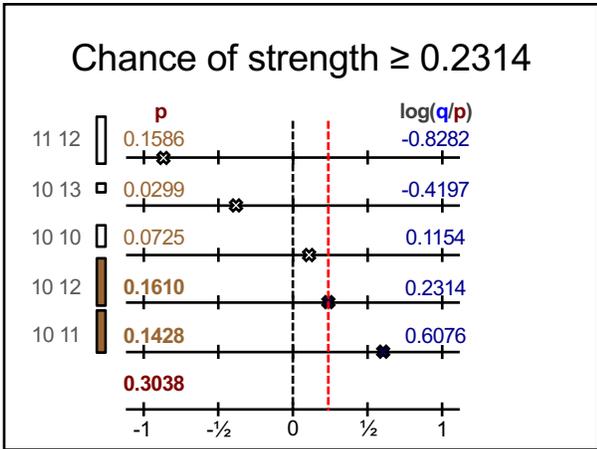


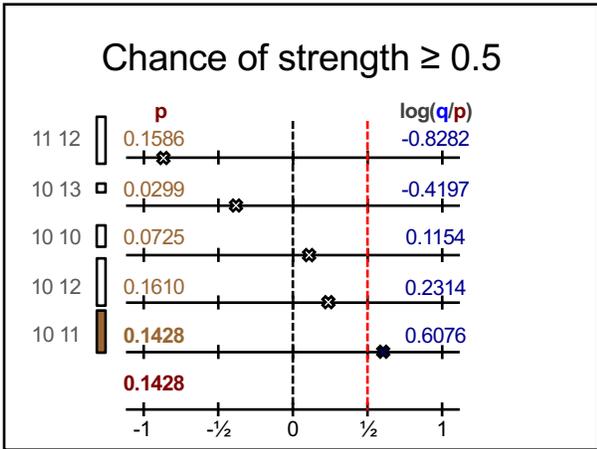


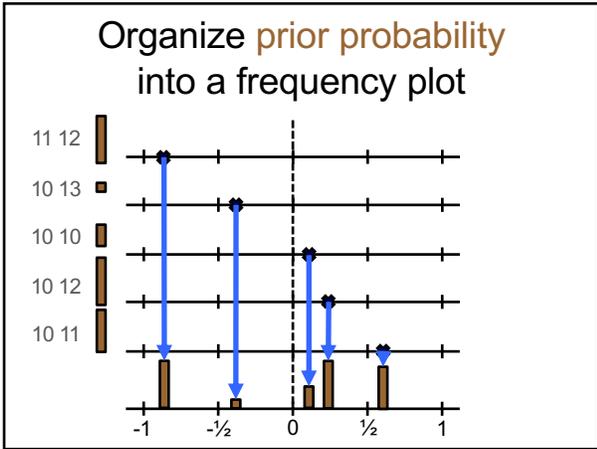


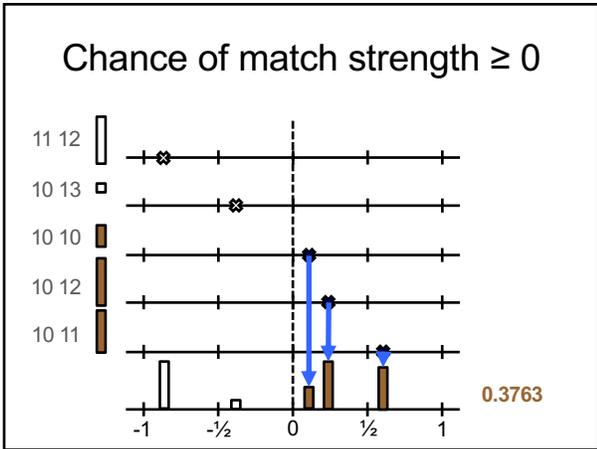




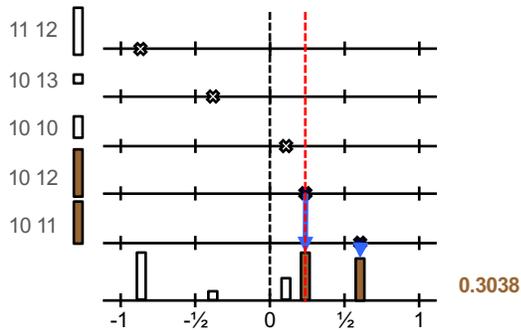






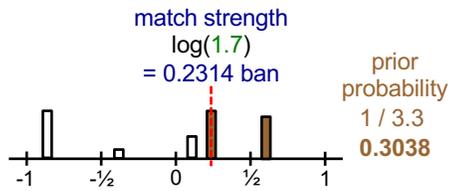


Chance of strength ≥ 0.2314

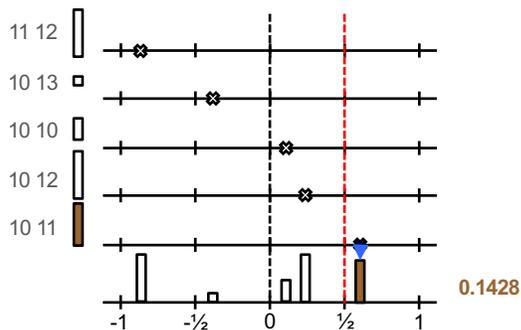


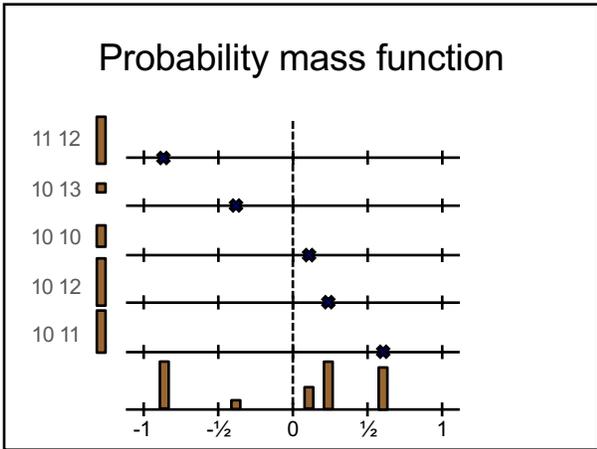
Error statement (one locus)

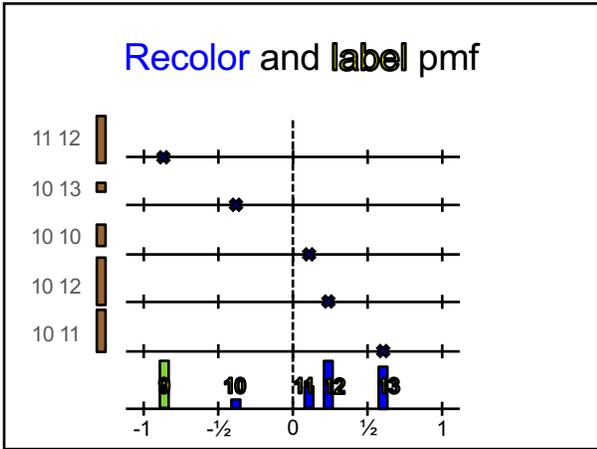
For a match strength of 1.7,
only 1 in 3.3 people would match as strongly

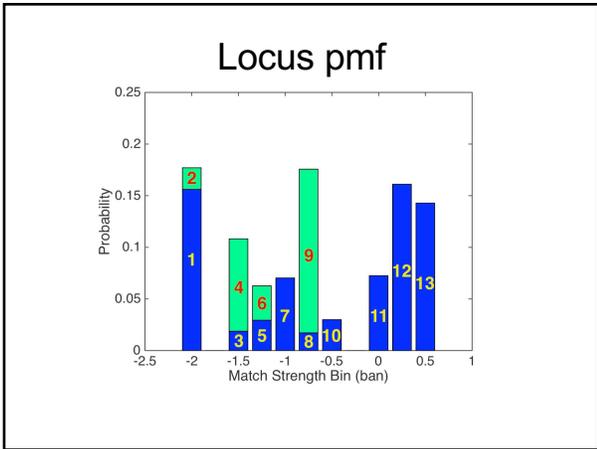


Chance of strength ≥ 0.5







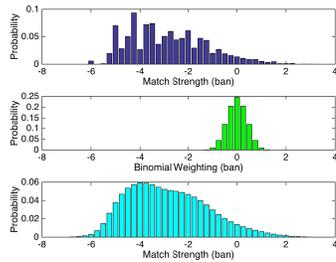


Multi-locus genotype information

1. STR loci are independent
2. Total $\log(\text{LR}) = \text{sum of locus } \log(\text{LR})\text{'s}$

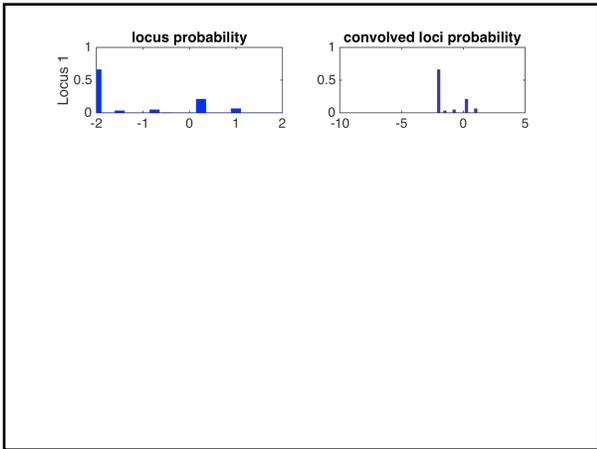
Therefore, pmf of total $\log(\text{LR})$ is the **convolution** of locus $\log(\text{LR})$ pmf's
(math fact, from 1718)

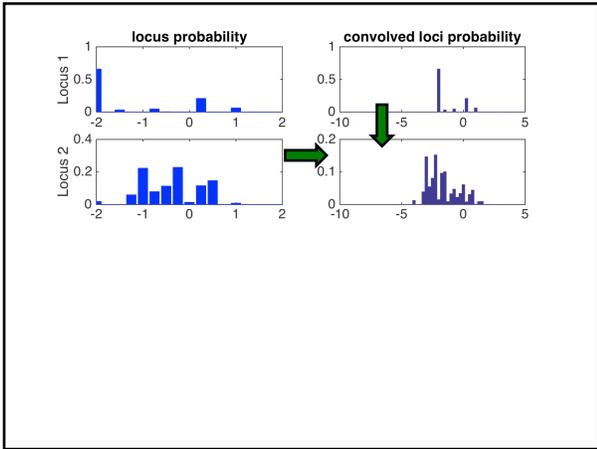
Convolution of functions

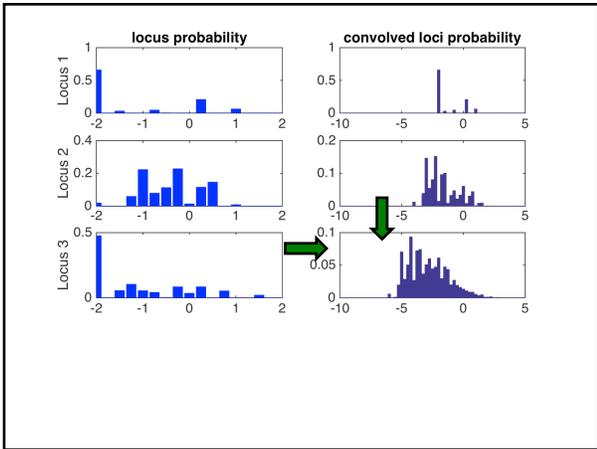


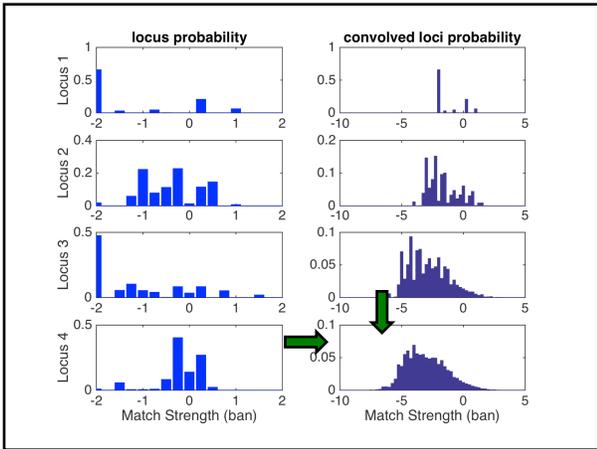
Sequential locus pmf convolution

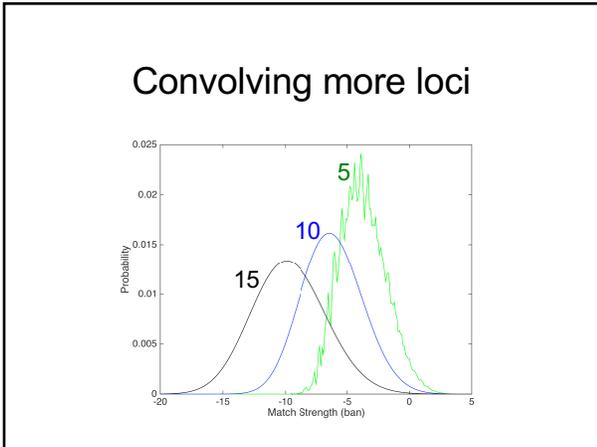
- pmf of one locus
 - convolve with second locus pmf
- pmf of two loci
 - convolve with third locus pmf
- pmf of three loci
 - convolve with fourth locus pmf
- pmf of four loci
 - convolve
- ...
- pmf across all loci

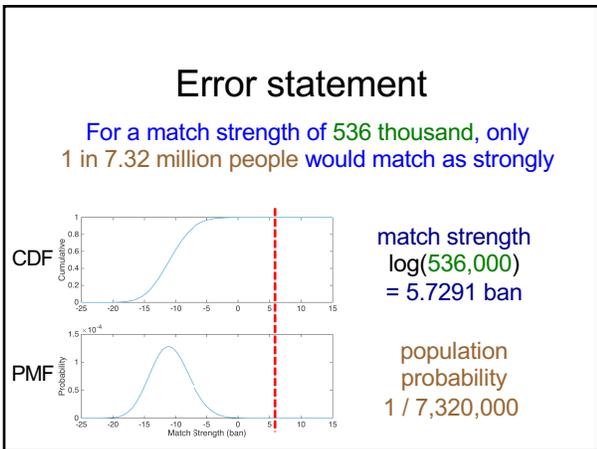












Exact vs. sampled

Exact

- all – 10^{24} genotypes
- accurate
- exact probability function
- convolution – fast

Sampled

- some – 10^4 genotypes
- approximate
- sample using random profiles
- Monte Carlo – slow



How can TrueAllele® help in reporting match error?

A homeless man took a woman into an alleyway and sexually assaulted her. He stole her phone so she couldn't call for help.

He threatened her, saying, "Don't tell anyone about this or I will kill you" and "You are never going to see your mother again."

Fearing for her life, she followed him across a bridge and into a downtown Pittsburgh park. He sexually assaulted her again, but she screamed and ran toward a hotel.

Hotel workers came to her aid, and chased after him. Police officers caught him a few blocks away.



Crime lab DNA analysis

The Allegheny County Medical Examiner's Office developed informative DNA data from the evidence.

Using limited DNA mixture interpretation methods, the lab said that **no conclusion can be made** due to **insufficient data** on some items, and the **complexity of the data** on others.

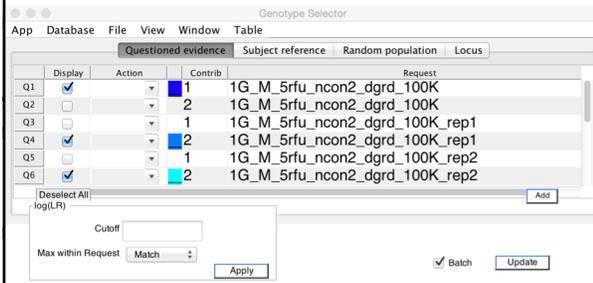
They did not report DNA match statistics.

TrueAllele® interpretation

Description	log(LR)	
	Victim	Suspect
non-sperm rectal swabs	16.06	25.81
sperm rectal swabs		3.69
right hand fingernails	30.72	21.31
left hand fingernails	29.97	16.30

LR values range from thousands to nonillions

Genotype selector



Match table

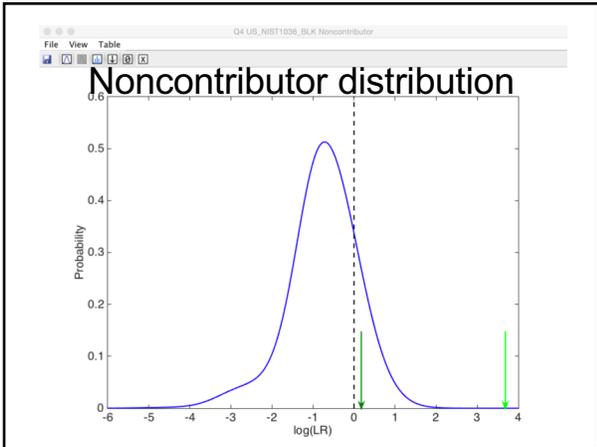
	Evidence	Contributor	N Contrib	Weight	Std Dev	KL	1N	2
Q1	1G_M_5rfu_ncon2_dgrd_100K	1	2	0.5300	0.1177	1.0086	0.1890	3.6451
Q4	1G_M_5rfu_ncon2_dgrd_100K_rep1	2	2	0.5073	0.1354	0.9191	0.1821	3.6896
Q6	1G_M_5rfu_ncon2_dgrd_100K_rep2	2	2	0.5282	0.1199	0.9655	0.1098	3.8186

Genotype info

Select the "Noncontributor" log(LR) distribution

System	system41.trueallele.net
Database	ACDA58
Software	3.25.5840.1
Person	erin
Client	ACDA
Reqname	1G_M_5rfu_ncon2_dgrd...
Process	twounknown
Type	evidence
Reqid	20
Contrib	2
Ncon	2
Nref	0
Wght	0.50726
Wstd	0.13544
Wtemname	1G_M
Wtemid	20
Power	0.91912
Casename	ACDA58
Casepart	evidence
Isref	0

Buttons: Spec, Data, Close, log(LR) Distribution, Noncontrib..., Contributor



Noncontributor statistics

stat	value
minim...	-40
mean	-0.7568
maxi...	44.2780
std dev	0.8910
positive	0.1834

Noncontributor tail statistics

Item	LR	log(LR)	left tail	L: one in	right tail	R: one in
1	1.52	0.1821	8.7151e-01	1.15	1.2849e-01	7.78
2	4.89 thousand	3.6896	1.0000e+00	1	3.4800e-08	28.7 million

$$PME = 1/28,700,000 \ll 1/4,890 = 1/LR$$

For a match strength of 4,890, only
1 in 28.7 million people would match as strongly

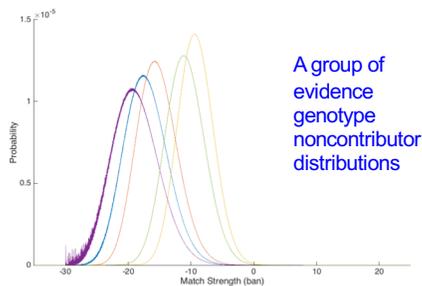
Why are verbal equivalents unnecessary?

LR for (H ₁) and 1/LR for (H ₂)	Qualitative Equivalent
1	Uninformative
2 to <100	Limited Support
100 to <10,000	Moderate Support
10,000 to <1,000,000	Strong Support
≥1,000,000	Very Strong Support

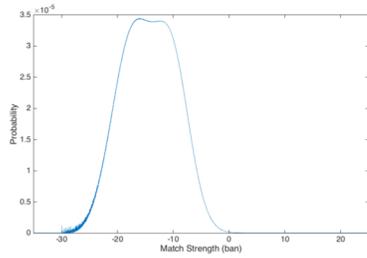
- hides the real match strength information
- not what a DNA expert actually believes
- misleads the jury about "million" (Koehler)

Just report PME error, along with the LR,
when the match strength is under a million

Why are error rates from DNA evidence and validation studies similar?



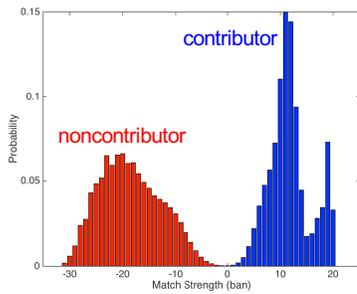
Validation pmf is the average



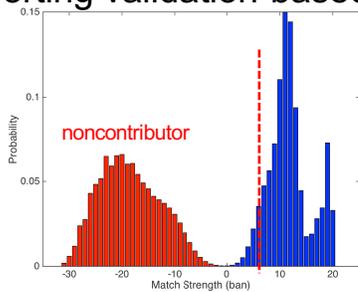
- exact: average the evidence distributions (a second)
- sample: compare evidence vs. random profiles (weeks)

Distribution curve (milliban)

Match strength histogram (ban)

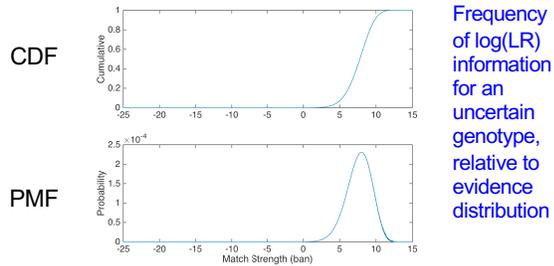


Reporting validation-based error



For a match strength of 536 thousand, only 1 in 9.65 million people would match as strongly

Reporting exclusionary error



Conclusions

- measuring error is built into genotype probability
- always report the LR; can also report PME
- verbal equivalents are not good science
- validation is easy – average the evidence curves
no “right” match answer is needed,
just the evidence genotype distributions

Perlin MW. Efficient construction of match strength distributions for uncertain multi-locus genotypes. *Heliyon*. 2018;4(10):e00824.
