



1

AI delivers accurate and objective forensic DNA evidence




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



Cybergenetics

Cybergenetics © 2003-2024



JUSTICE
THROUGH
SCIENCE

2

DNA evidence is powerful

Inculpatory evidence
A DNA *match* can find people where they shouldn't be.

Exculpatory evidence
A DNA *nonmatch* can show that someone wasn't there.

DNA is easy with simple evidence

3

Types of DNA evidence

Single source
Occurs when one person leaves their DNA on an item.
Simple evidence.

DNA mixture
Occurs when two or more people leave their DNA.
Most common form of DNA evidence.

DNA mixtures are hard for people to analyze

Solving complex mixtures

Human intelligence

Simplifies the DNA data to simplify the problem.
And usually gets the wrong answer.

Computer intelligence

Explains the DNA data to unmix the mixture.
And gets the right answer.

DNA is easy for complex evidence,
but only with smart AI computers.

Bayesian probability equations

$$\begin{aligned} \mu_i &= m_i \cdot \sum_{l=1}^K w_{l,i} \cdot \mathbf{R}_{l,i} \\ \mathbf{w}_i &\sim N(\mathbf{0}, \psi^{-2} \cdot \mathbf{I}) \\ \Sigma_i &= \sigma^2 \cdot \mathbf{V}_i + \tau^2 \\ \mathbf{d}_i &\sim N_i(\mu_i, \Sigma_i) \\ \mathbf{R}_{i,j} &\sim \begin{cases} f_i^j, & i = j \\ 2f_i f_j, & i \neq j \end{cases} \\ \mathbf{w} &\sim \text{Dir}(\mathbf{1}) \\ m_i &\sim N_i(5000, 5000^2) \\ \sigma^{-2} &\sim \text{Gam}(10, 20) \\ \tau^{-2} &\sim \text{Gam}(10, 500) \\ \psi^{-2} &\sim \text{Gam}(1/2, 1/200) \end{aligned}$$

Smart computers can solve math
that is impossible for people.

Focus of the talk

1. Human DNA interpretation fails
Losing truth harms justice
2. Computer artificial intelligence succeeds
Finding truth helps justice
3. Humans suppress computer intelligence
Hiding truth harms justice

New York v John Wakefield

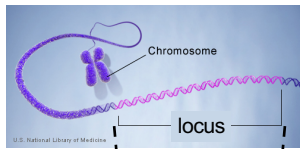
In April, 2010, Brett Wentworth (41) was found dead in his apartment, strangled with an electric guitar cord.

The police collected biological evidence was from amp cord sections, plus his shirt collar and forearm.

The New York State Police lab examined the DNA mixtures.

Comparing the evidence with suspect John Wakefield (44) found very little DNA match information.

DNA genotype



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

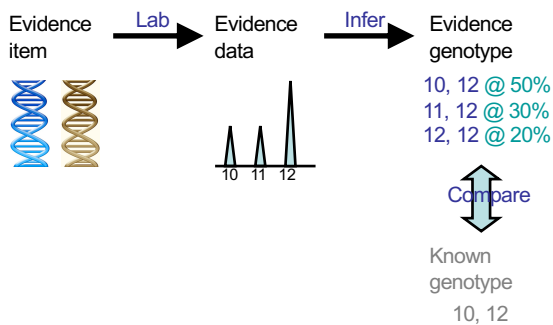
Many alleles allow for many many allele pairs. A person's genotype is relatively unique.

mother allele

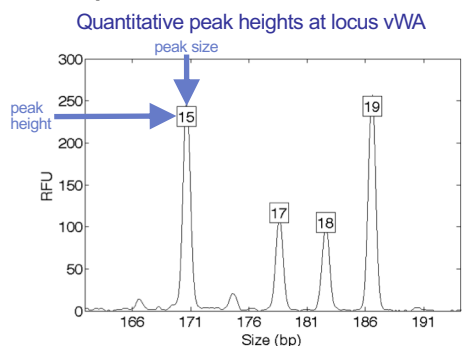
father allele

ACGT repeated word

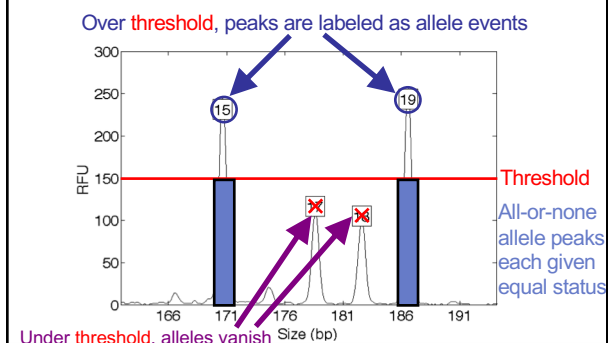
DNA evidence interpretation



Computers can use all the data



People may use less of the data



Human review CPI match statistic: subjective, biased & uninformative

J Pathol Inform

Editor-in-Chief: Justin Patterson, Pittsburgh, PA, USA
 OPEN ACCESS HTML format

Research Article
Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information

Mark William Perlin¹

¹Cybergenetics, Pittsburgh, USA
 E-mail: *Dr. Mark William Perlin - perlin@cygno.com
 *Corresponding author

Received 14 July 2015 Accepted 21 September 2015 Published 28 October 2015

Finding truth through science

- 16th – Francis Bacon; empiricism, testing
- 17th – Isaac Newton; math & physics
- 18th – Thomas Bayes; inverse probability

H hypothesis, D data
 How data updates our belief in a hypothesis

$$\Pr\{H | D\} \propto \Pr\{D | H\} \times \Pr\{H\}$$

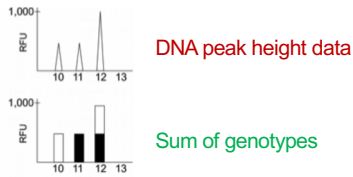
Posterior probability Likelihood Prior probability

Explaining DNA mixtures



Mark W. Perlin,¹ Ph.D., M.D., Ph.D. and Beata Szabady,¹ Ph.D. 2001

Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples

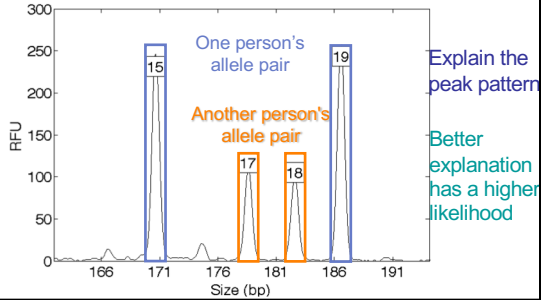


TrueAllele® computer solution

- Accurate. 43 validation studies, 8 published
- Objective. Workflow removes human bias
- Accepted. Reported in 47 states, used by 10 labs
- Transparent. Give math, software, 4GB DVD
- Neutral. Can statistically include or exclude

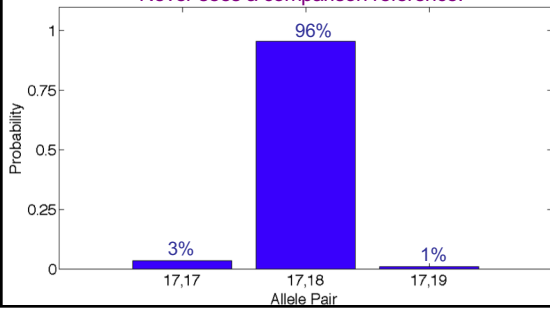
How the computer thinks

Consider every possible genotype solution



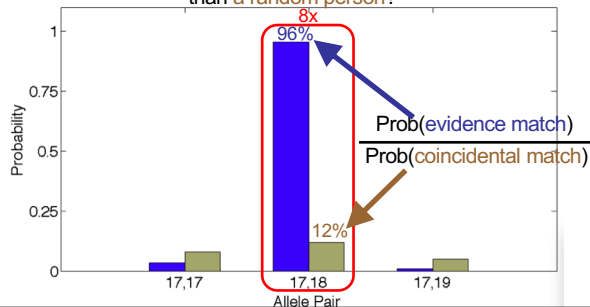
Evidence genotype - probability

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



DNA match information

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



Is the suspect in the evidence?

A match between the amp cord
and John Wakefield is:

5.88 billion times more probable than
a coincidental match to an unrelated Black person

300 million times more probable than
a coincidental match to an unrelated Caucasian person

2.25 billion times more probable than
a coincidental match to an unrelated Hispanic person

Match statistics

Item	Description	052B	188, 189
		Brett Wentworth	John Wakefield
004A-C	Swabs 0-3 ft of amp cord	18.81	-0.10
004D-F	Swabs 3-6 ft of amp cord	18.81	0.15
004G-I	Swabs 6-9 ft of amp cord	18.81	2.90
004J-L	Swabs 9-12 ft of amp cord	18.81	-16.69
004M-O	Swabs 12-15 ft of amp cord	17.68	8.48
004P-R	Swabs 15-18 ft of amp cord	18.70	-1.49
004S-T	Swabs 18-20 ft of amp cord	18.81	-1.09
045A	Shirt collar, outside rear	7.92	18.88
045C	Shirt collar, outside front	18.81	10.07
052F1-2	Victim forearm swabs	18.81	6.36

Heliyon

Received: 7 May 2024
Revised: 10 June 2024
Accepted: 20 September 2024

Efficient construction of match strength distributions for uncertain multi-locus genotypes

Mark W. Peltier
Chromatics, Pittsburgh, PA, USA
*Corresponding author.
Email address: peltier@chromatics.com (M.W. Peltier)

Error rate

Peer-reviewed validation studies²²

Perlin MW, Sinenikov A. An information gap in DNA evidence interpretation. *PLoS ONE*. 2009;4(12):e8327.

Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. *Science & Justice*. 2013;53(2):103-114.

Perlin MW, Hornyak J, Sugimoto G, Miller K. TrueAllele[®] genotype identification on DNA mixtures containing up to five unknown contributors. *Journal of Forensic Sciences*. 2015;60(4):857-868.

Greenspoon SA, Schiermeier-Wood L, Jenkins BC. Establishing the limits of TrueAllele[®] Casework: a validation study. *Journal of Forensic Sciences*. 2015;60(5):1263-1276.

Bauer DW, Butt N, Hornyak JM, Perlin MW. Validating TrueAllele[®] interpretation of DNA mixtures containing up to ten unknown contributors. *Journal of Forensic Sciences*. 2020; 65(2):380-398.

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele[®] DNA mixture interpretation. *Journal of Forensic Sciences*. 2011;56(6):1430-1447.

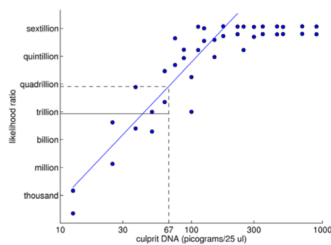
Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele[®] Casework validation study. *Journal of Forensic Sciences*. 2013;58(6):1458-1466.

Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. TrueAllele[®] Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE*. 2014;9(3):e92837.

TrueAllele predictability²³

An Information Gap in DNA Evidence Interpretation

Mark W. Perlin^{1*}, Alexander Sinenikov² 



TrueAllele reliability²⁴

TrueAllele Casework on Virginia DNA Mixture Evidence: Computer and Manual Interpretation in 72 Reported Criminal Cases

Mark W. Perlin^{1*}, Kiersten Dormer¹, Jennifer Hornyak¹, Lisa Schiermeier-Wood², Susan Greenspoon²

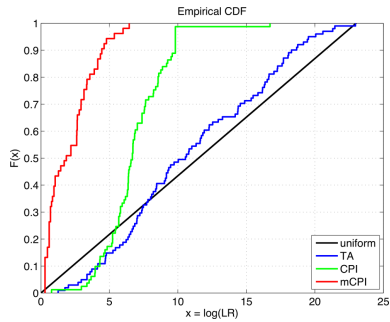
1 Cybergnetics, Pittsburgh, Pennsylvania, United States of America, 2 Department of Forensic Science, Richmond, Virginia, United States of America



Validation axes

- sensitive
- specific
- reproducible

TrueAllele accuracy



Higher human error rate

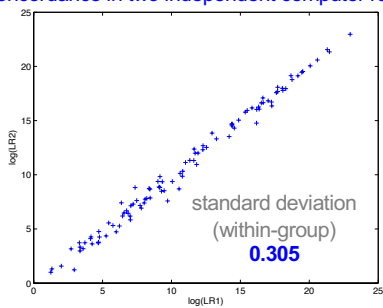
TrueAllele specificity (million samples)
 From noncontributor distribution, for LR > 100:
 Error rate = 1 in 1,000,000 (0.0001)%

CPI – analytical threshold
 5 false positives in 81 comparisons
 Error rate = 5 in 81 (6%)

mCPI – stochastic threshold
 17 inconclusive results
 1 false positive in 53 comparisons
 Error rate = 1 in 53 (2%)

TrueAllele reproducibility

Concordance in two independent computer runs



JOURNAL OF FORENSIC SCIENCES 28
J Forensic Sci., November 2013, Vol. 58, No. 6
 doi: 10.1111/1556-4029.12223
 Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Mark W. Perlin,¹ M.D., Ph.D.; Jamie L. Belrose,² M.S.; and Barry W. Duceman,³ Ph.D.

New York State TrueAllele® Casework Validation Study*

Computers find information
 Humans fail 70% of the time

JOURNAL OF FORENSIC SCIENCES 29
J Forensic Sci., 2019
 doi: 10.1111/1556-4029.14204
 Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

David W. Bauer,¹ Ph.D.; Nasir Butt,² Ph.D.; Jennifer M. Hornyak,¹ M.S.; and Mark W. Perlin,¹ Ph.D., M.D., Ph.D.

Validating TrueAllele® Interpretation of DNA Mixtures Containing up to Ten Unknown Contributors*

TABLE 4—Independent analysis.

	Operator Site		Mixture weight (%)	Peeling Round							
	Cybergenetics	CCRFSL		0	1	2	3	4	5	6	
Genotypes	78	78	13	7	K	K	K	K	K	K	K
Minimum	-5.16	-9.14	22	6	7	K	K	K	K	K	K
Mean	8.36	8.48	12	5	4	5	K	K	K	K	
Median	5.98	5.61	16	4	4	5	6	K	K	K	
Maximum	29.03	29.12	13	4	3	2	1	6	K	K	
SD	8.37	8.54	4	3	3	4	1	6	8	K	
σ_e		0.70	2	1	1	1	1	3	3	4	
			2	0	2	2	3	2	3	4	
			4	0	1	1	1	2	0	2	
			1	0	0	0	0	-1	-1	0	

TABLE 7—Peeling sensitivity.

30

TrueAllele acceptance

- Invented math & algorithms 30 years
- Developed computer systems 25 years
- Support users and workflow 10 laboratories; 100,000 items
- Routinely used in casework 525 agencies (FBI)
- Validate system reliability 43 studies
- Educate the community 175 talks
- Train or certify analysts 400 students
- Admissibility challenges 44 rulings, 15 states and federal
- Testify about LR results 145 trials
- Educate lawyers and public 1,000 people
- Make the ideas understandable 1,250 cases, 47 states

Frye & Daubert

- Testing
- Error rate
- Peer review
- General acceptance


Wakefield Frye ruling

Accordingly, the Court finds that Cybergene's TrueAllele Casework is not novel but instead is "generally accepted" under the Frye standard. The Court therefore DENIES the Defendant's Motion to Preclude, subject to sufficient foundational showings by the People as to their experts' qualifications and adherence to accepted procedures for collection, storage, or analysis of such evidence (cf. *People v. Kelly*, 288 AD2d 695 [3rd Dept 2001]).

THIS SHALL CONSTITUTE THE DECISION AND ORDER OF THE COURT.

Dated: February 9, 2015
at Cooperstown, New York

ENTER


Hon. Michael V. Coccoma
Supreme Court Justice

Verdict & sentence

27-May-2015



**Wakefield sentenced to life in prison
for murder in upstate New York**

Schenectady, NY

A Schenectady man was **sentenced** to life in prison without parole on May 27. John Wakefield had been convicted of strangling Brett Wentworth in his home with a guitar amplifier cord. DNA mixtures on the cord, as well as the victim's clothing and skin, tied Wakefield to the murder. The state crime lab could not resolve the mixtures, so the prosecutor asked Cybergene to solve the problem. TrueAllele separated the DNA mixture data into the genotypes of Wentworth and Wakefield. Following a successful Frye hearing in October, TrueAllele was **admitted** into evidence and Cybergene's Dr. Mark Perlin testified in March about the match results.

World Trade Center

18,000
victim remains

2,700
missing people



match



September 11, 2001 – New York City

Indiana v. Darryl Pinkins



New York v. Nick Hillary



Garrett Phillips (12)
Died from strangulation
October 24, 2011



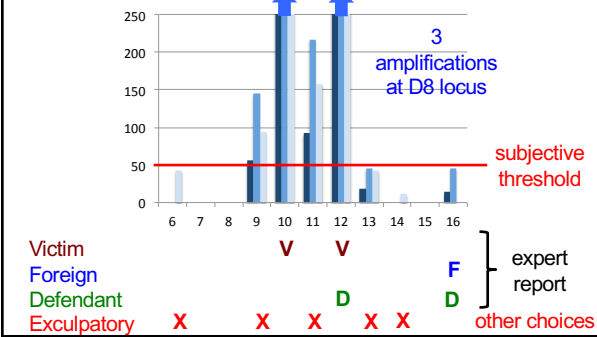
Oral "Nick" Hillary
Arrested for murder
May 15, 2013

TrueAllele findings in Hillary

2013. **26 Identifier tests** on left fingernails
 Mixture of 95% victim + 5% other
No statistical connection to Hillary
 We advised Minifiler for degraded DNA

2014. More lab data on left fingernails
9 tests using new kit & machine
 NYSP requested TrueAllele analysis
 Again, **no connection to Hillary**

Other PG software forces users to choose their data



Thresholds change PGS answers

RFU	Data choices	LR	All stutters
80		0	9
70		30	51
60		250	1,660
50	15,500,000		69,200 include
40		0	0
30		0	0 exclude

Different data choices, contradictory answers.
 PGS STRmix™ does not agree with itself.

Judge does not admit STRmix

The Expert was forced to **pick and choose data** from different “reliable sources” and **input parameters** into the [STRmix] program in such a way that he believed the system would tolerate.

ORDERED that the **defendant’s motion to preclude** the prosecution from calling an expert witness to testify regarding any conclusion reached by the use of the [STRmix] Software is **granted**.

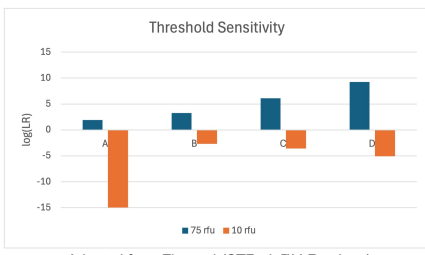
Nick Hillary acquitted

The New York Times

Oral Nicholas Hillary Acquitted in Potsdam Boy’s Killing



Technical Note
A tale of two PG systems: A comparison of the two most widely used continuous probabilistic genotyping systems in the United States
Susan A. Greenspoon PhD | Lisa Schliermeier-Wood MS | Bradford C. Jenkins MS



Adapted from Figure 4 (STRmix™ LR values)

How to hide truth from courts

- Make false statements
- Distract from truth
- Refuse to test method
- Accuse others of lying

Try to block AI computers

- Source code is needed to cross-examine software.
- Insist on "ground truth" to dismiss valid testing studies.
- Mislabeled transparent software as a "black box".
- PA v. Washington. Make up incorrect LR definitions.
- US v. Anderson. Demand impossible discovery items.
- US v. Mills. Focus on small LRs, ignore error rate.
- * FL v. Daniels. Demand irrelevant "internal" validation.
- US v. Johnson. Pretend low-level DNA is different.
- NIST. Ignore government agency conflicts and bias.
- * NE v. Simmer. Laud *ad hoc* PCAST, ignore standards.
- * US v. Gissantaner. Change Daubert prong meaning.
- US v. Sandoval. Ignore how thresholds discard data.
- NY v. Hillary. Claim different methods are the same.

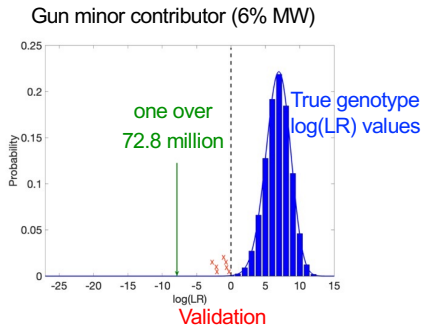
PA v. Washington

M.W. Perlin, "**Distorting DNA evidence: methods of math distraction**", *American Academy of Forensic Sciences 70th Annual Meeting*, Seattle, WA, 22-Feb-2018.

Focus on the LR *numerator*, ignore the *ratio*:

1. The defendant does not have the *highest probability genotype*.
2. *Other genotypes* have probabilities that add up to over half.
3. The *match probability* between the evidence and defendant is small.

US v. Mills



National Institute of Standards and Technology (NIST)

- Fear.** *Create false crisis*
CPI failure, MIX05 (2005)
Add new threshold (2010)
- Bias.** *Promote foreign product*
Product launch (2013)
Company on-site (2014)
- Deceit.** *Misquote science* (2016)
- Goal.** *Regulatory funding & power*

NE v. Simmer

The two most widely used methods (STRmix and TrueAllele) appear to be reliable within a certain range, based on the available evidence and the inherent difficulty of the problem. Specifically, these methods appear to be reliable for **three-person mixtures** in which the **minor contributor constitutes at least 20 percent of the intact DNA in the mixture** and in which the DNA amount exceeds the minimum level required for the method. – PCAST, relying on NIST’s Dr. John Butler

(Cites 2015 TrueAllele® validation paper in JFS on five-contributor mixtures that demonstrates reliability beyond these artificial limits.)

Recommends: The President should request and Congress should provide increased appropriations to NIST of (a) **\$4 million** to support the evaluation activities described above and (b) **\$10 million** to support increased research activities in forensic science, including on complex DNA mixtures, latent fingerprints, voice/speaker recognition, and face/iris biometrics.

[Visit to Washington, DC](#)

US v. Sandoval

William Thompson. *J Forensic Sci.* 2023.
Uncertainty in probabilistic genotyping of low template DNA:
A case study comparing STRmix™ and TrueAllele®.

20 conceptual errors
120 mistaken assertions

Data issue: change threshold, change STRmix answer

"To expect competing for-profit companies to refrain from overclaiming and to fully disclose all uncertainties surrounding their findings is apparently expecting too much. To expect courts to regulate these matters as part of their review of admissibility apparently is also expecting too much. If these matters are to be addressed at all, they will need to be addressed by the forensic science community through the standards development process." - Thompson

Man limits machines

In 1901, Connecticut passed a speed limit law for automobiles. The city speed limit was 12 miles per hour. The rural limit was 15 miles per hour.

Near a horse, cars had to slow down. If the horse was scared, the car had to stop.



Why man restrains the machine

- Pride
- Fear
- Envy
- Greed
- Power

But limiting truth in forensic DNA science harms justice.
Wrongful convictions, wrongful acquittals.

The *information* incentive

Human-based labs are funded by failure.
They are paid to process DNA samples.

Computer-based methods deliver success.
They produce accurate & objective information.

Finding DNA information delivers justice;
losing or hiding information does not.

Government should fund DNA labs based on
how much useable *information* they produce.

More information

<http://www.cybgen.com/information>



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

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



Cybergenetics



perlin@cybgen.com



JUSTICE
THROUGH
SCIENCE
