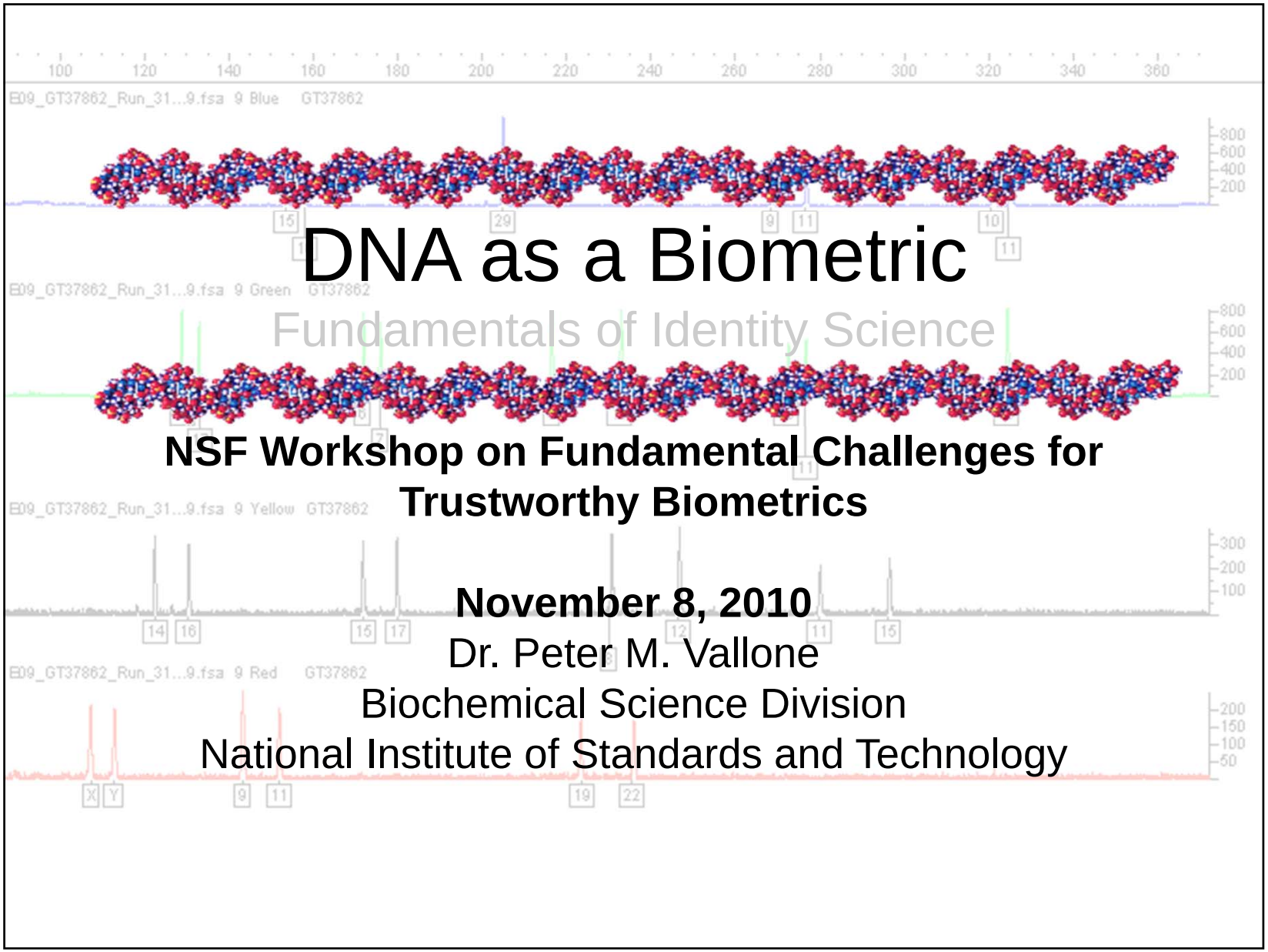


DNA as a Biometric

Fundamentals of Identity Science

NSF Workshop on Fundamental Challenges for Trustworthy Biometrics

November 8, 2010

Dr. Peter M. Vallone
Biochemical Science Division
National Institute of Standards and Technology

Outline

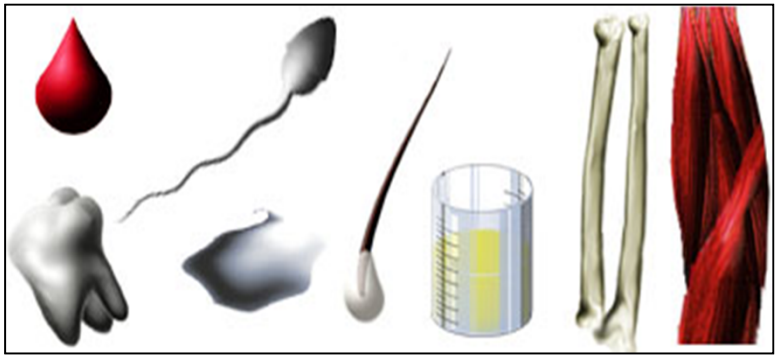
- Basics of DNA Typing
- DNA as a Biometric

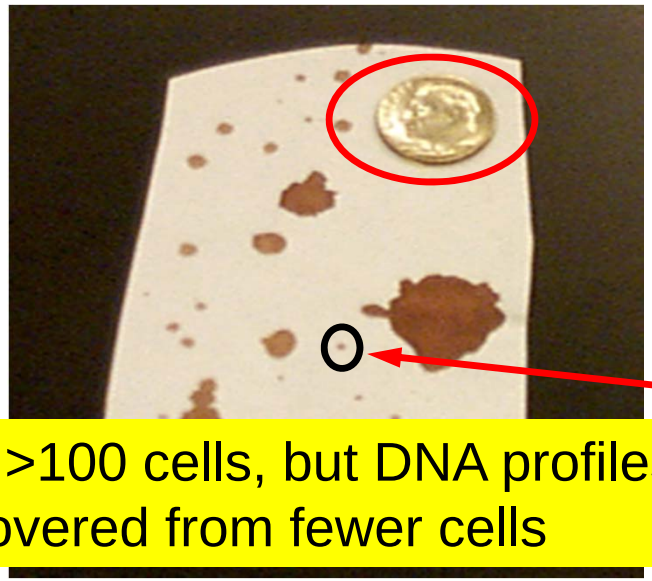
General Characteristics of Genomic DNA

- Each individual has a unique DNA profile
  - with exception of monozygotic siblings
- Each person's DNA is the same in every cell
  - DNA from skin cells will match DNA from blood cells
- An individual's DNA profile remains the same throughout life
- Half of your DNA comes from your mother and half from your father
  - implications for determining kinship

Sources of Biological Evidence

- **Blood**
- **Semen**
- **Saliva**
- **Urine**
- **Hair**
- **Teeth**
- **Bone**
- **Tissue**





Blood Sample

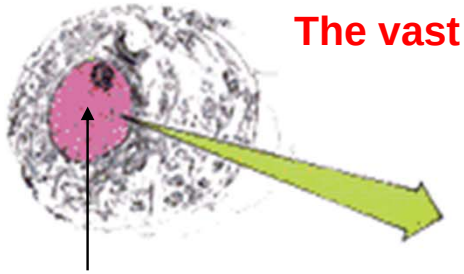
Only a very small amount of blood is needed to

ONA

best results with >100 cells, but DNA profiles can be recovered from fewer cells

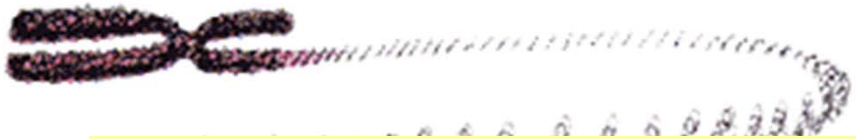
DNA in the Cell

The vast majority of DNA is the same from person to person



cell nucleus

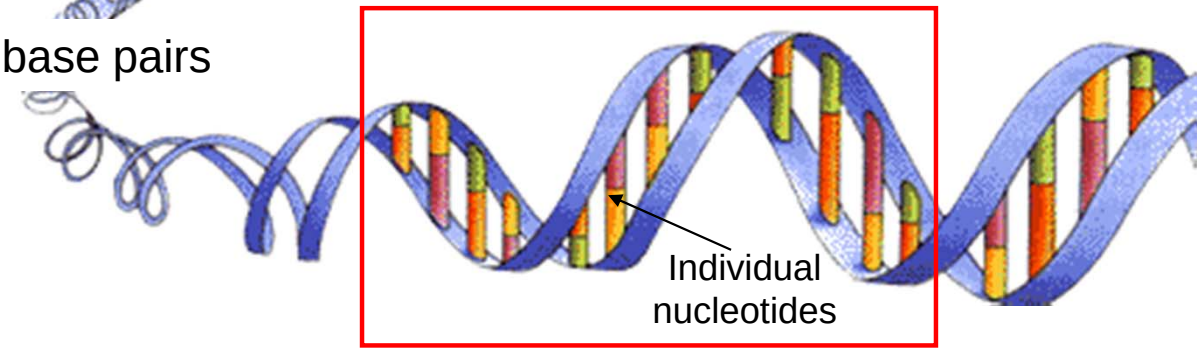
chromosome **22 pairs + XX or XY**



Only a small varying region is targeted and probed for each DNA marker examined

Double stranded DNA molecule

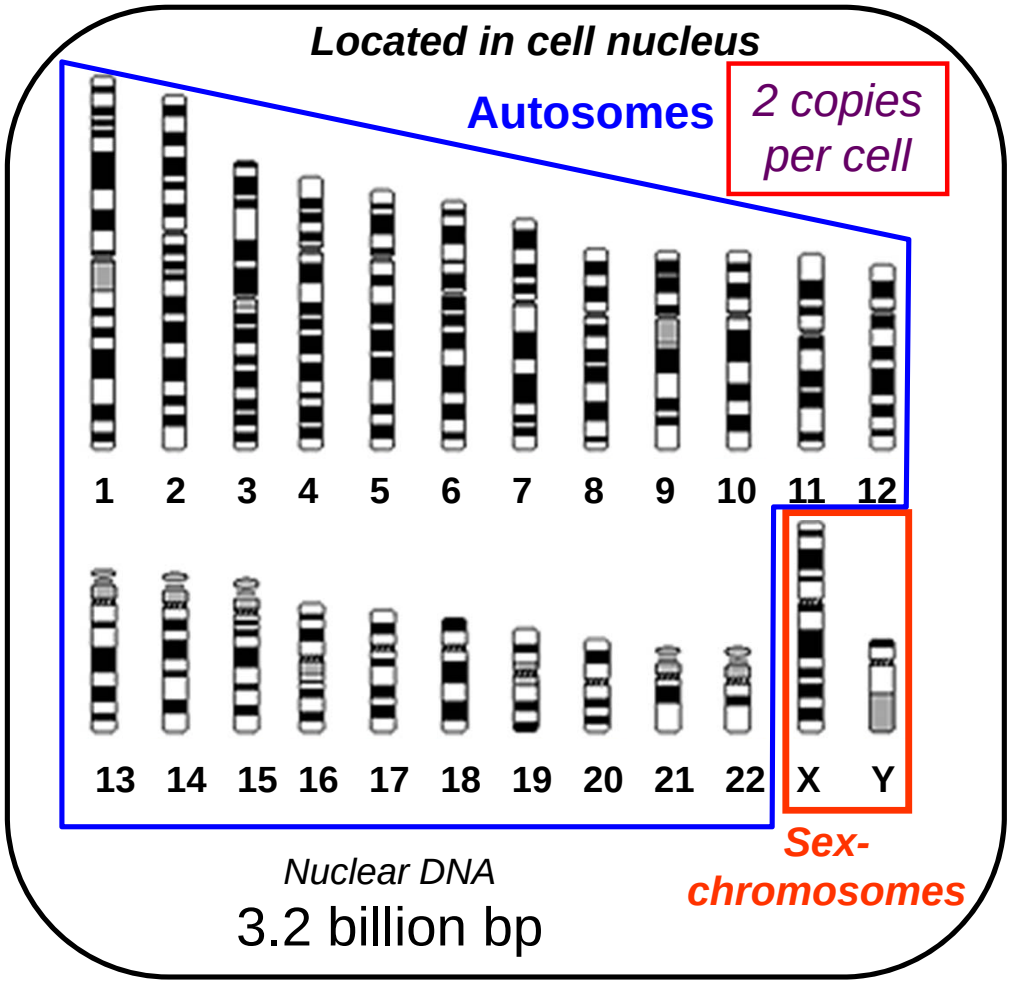
~3 billion total base pairs



Individual nucleotides

Human Genome

23 Pairs of Chromosomes + mtDNA



Located in mitochondria (multiple copies in cell cytoplasm)

100s of copies per cell

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 2.3, ©Elsevier Science/Academic Press

What Type of Genetic Variation?

- Sequence Variation
  - single nucleotide polymorphisms (SNPs)
  - insertions/deletions

GCTAGTCGATGCTC**(G/A)**GCGTATGCTGTAGC

- Length Variation
  - **short tandem repeats** (STRs)

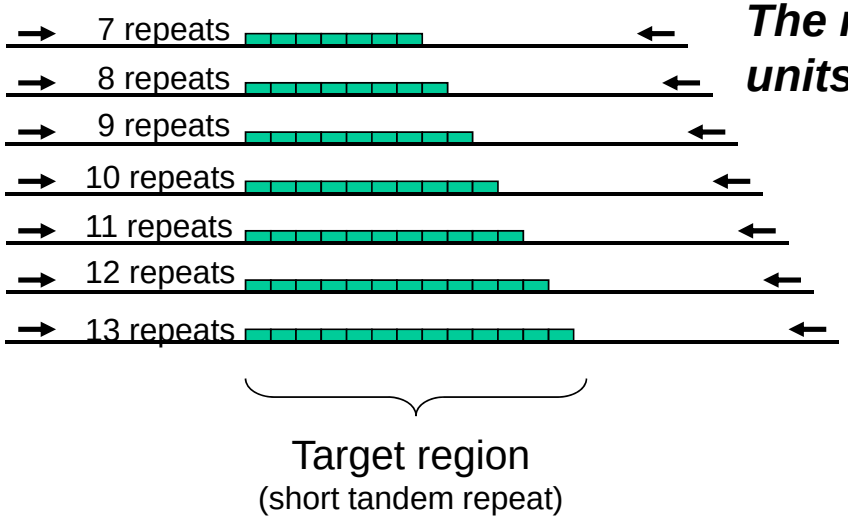
CTAGTCGT**(GATA)(GATA)(GATA)**GCGATCGT

Short Tandem Repeat (STR) Markers

An accordion-like DNA sequence that occurs between genes

TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGAAGACA
GGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**
GATAGATATCATTGAAAGACAAAACAGAGATGGATGATAGATACA
TGCTTACAGATGCACAC

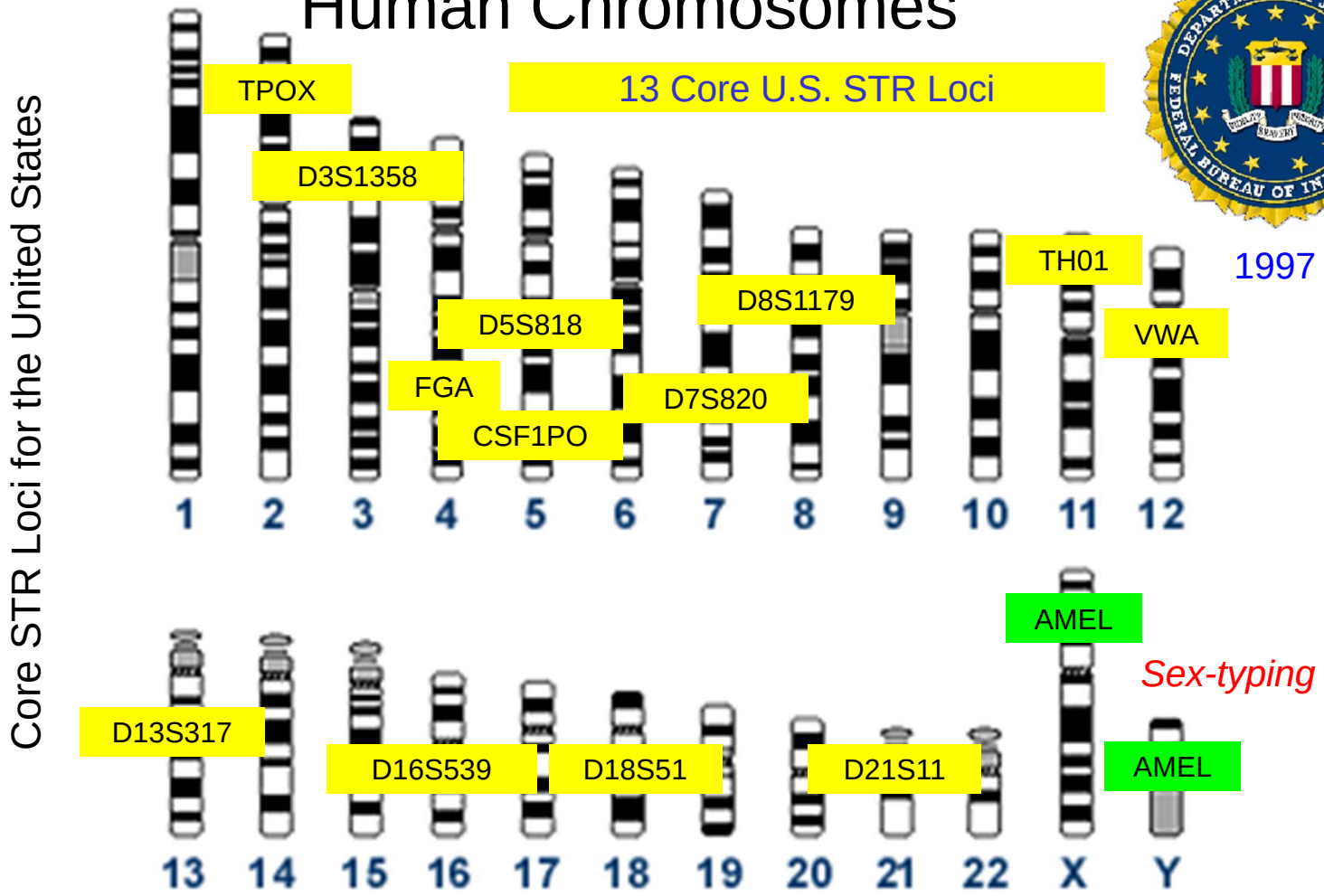
= 12 GATA repeats ("12" is all that is reported)



The number of consecutive repeat units can vary between people

The FBI has selected **13 core STR loci** that must be run in all DNA tests in order to provide a common currency with DNA profiles

Position of Forensic STR Markers on Human Chromosomes



Forensic DNA Testing

Probe subsets of genetic variation in order to differentiate between individuals (14 to 16 regions)

DNA typing must be done efficiently and reproducibly (information must hold up in court)

Over 8 million profiles in the national FBI database

Typically, we are *not* looking at genes – little/no information about race, predisposition to disease, or phenotypic information (eye color, height, hair color) is obtained

Applications

- Forensic cases: matching suspect with evidence
- Paternity testing: identifying father
- Missing persons investigations
- Military DNA “dog tag”
- Convicted offender DNA databases
- Mass fatalities
- Historical investigations
- Genetic genealogy
- DNA as a biometric tool

DNA Testing Requires a Reference Sample

A DNA profile by itself is fairly useless because it has no context…

DNA analysis for identity only works by comparison – **you need a reference sample**

Crime Scene Evidence compared to **Suspect(s)** (Forensic Case)
Child compared to **Alleged Father** (Paternity Case)
Victim's Remains compared to **Biological Relative** (Mass Disaster ID)
Soldier's Remains compared to **Direct Reference Sample** (Armed Forces ID)

Steps in Forensic DNA Analysis

Usually 1-2 day process (a minimum of ~8 hours)

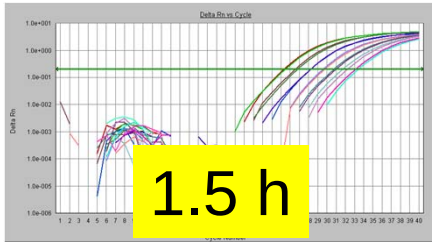

Blood Stain


Buccal swab

Sample Collection & Storage

Biology

DNA Extraction



DNA Quantitation

Multiplex PCR Amplification

DNA separation and sizing

Technology

STR Typing

Interpretation of Results

Genetics

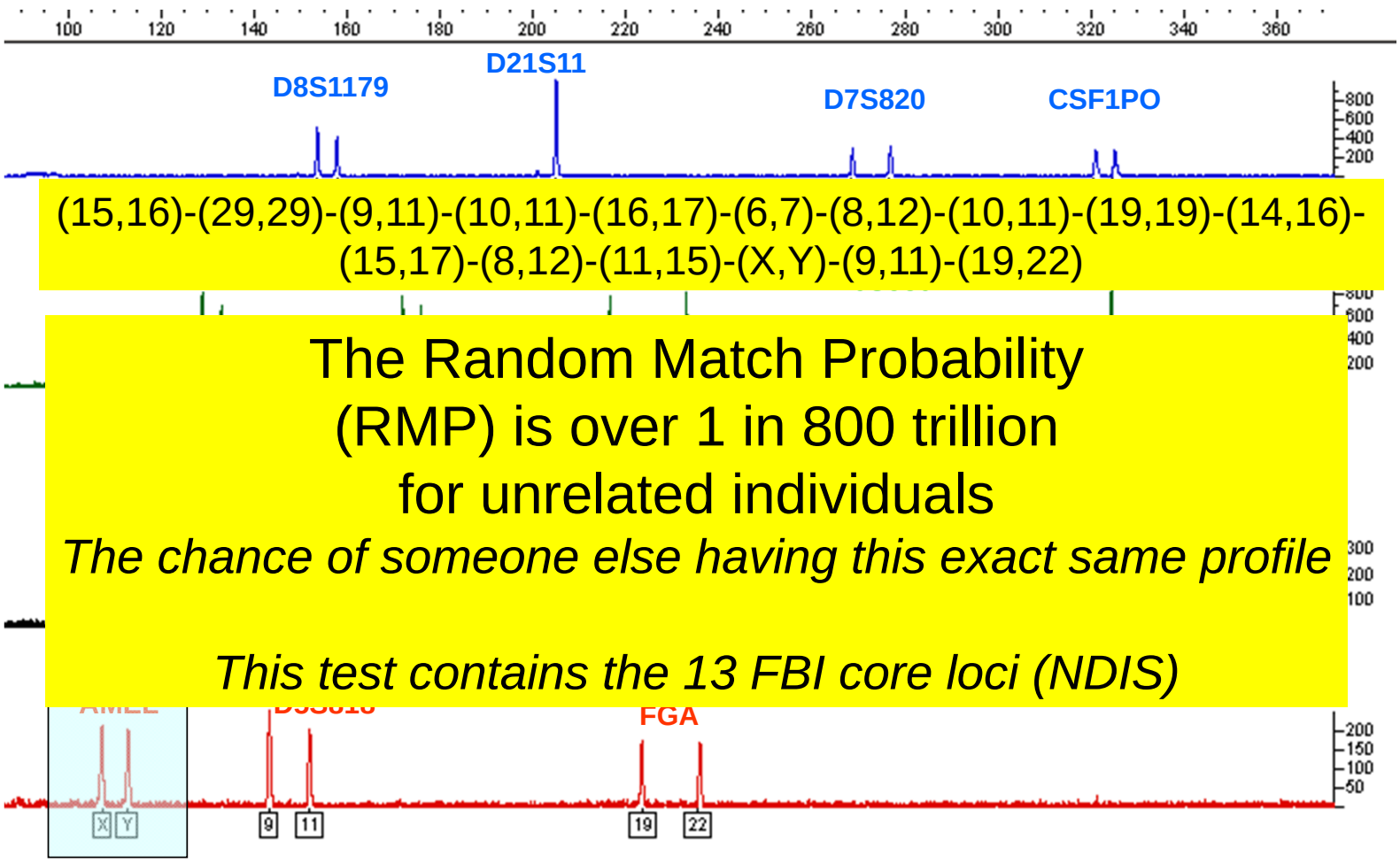
Statistics Calculated

DNA Database search
Paternity test
Reference sample

Applied Use of Information

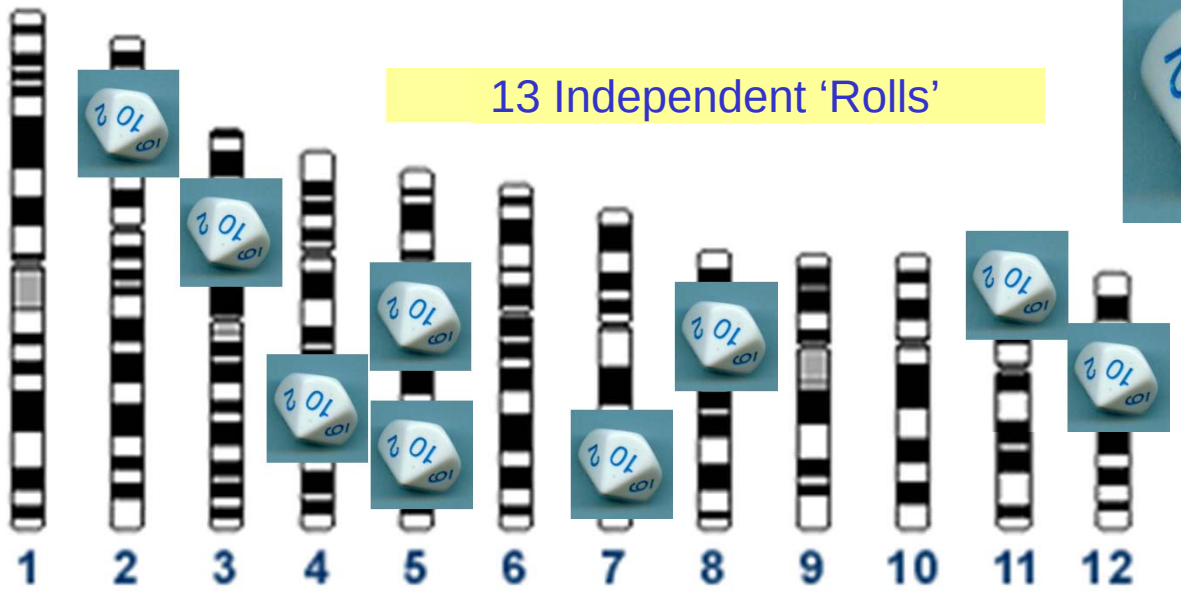
Identifiler (Applied Biosystems) 15 STR Loci Kit

Information is tied together with multiplex PCR and data analysis

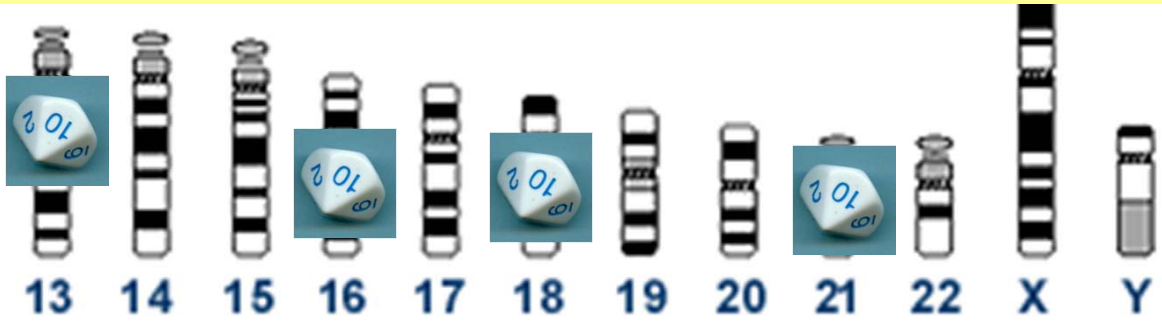


Product Rule

13 Independent ‘Rolls’



Assuming a 10 sided die (P=0.1) = 0.1 x 10^13 or **1 in 10 trillion**



Large Random Match Probabilities?

- Let's assume each marker could be one of 10 states (or alleles)
- All loci are independent – so we can multiply the
- 2 x 5 x 8 x 2 x 6 x 7 x 7 x 10 x 1 x 3
- 0.1 x 0.1 x 0.1 x 0.1 x 0.1 x 0.1 x 0.1 x 0.1 x 0.1 x 0.1

Product Rule

For heterozygous loci

$$P = 2pq$$

P = probability; p and q are frequencies of allele in a given population

Example: For the locus D3S1358 an individual is 15,18 with frequencies of 0.2825 and 0.1450 respectively

P = 2(0.2825)(0.1450) = 0.0819 or 1 in 12

For 5 loci the

Profile Probability = $(P_1)(P_2)\ldots(P_n)$

= (0.0819)(0.0875)(0.0687)(0.0245)(0.0984)

0.000001187 or 1 in 842,539

Kinship Testing

- DNA profiles can also be used to evaluate the probability of a specific familial relationship
- As a familial relationship becomes more distant, the ability of DNA to confirm the likelihood of that relationship decreases
  1. Parent-offspring
  2. Siblings
  3. Half siblings = uncle/nephew = grandchild
  4. Cousins

Autosomal Paternity Example

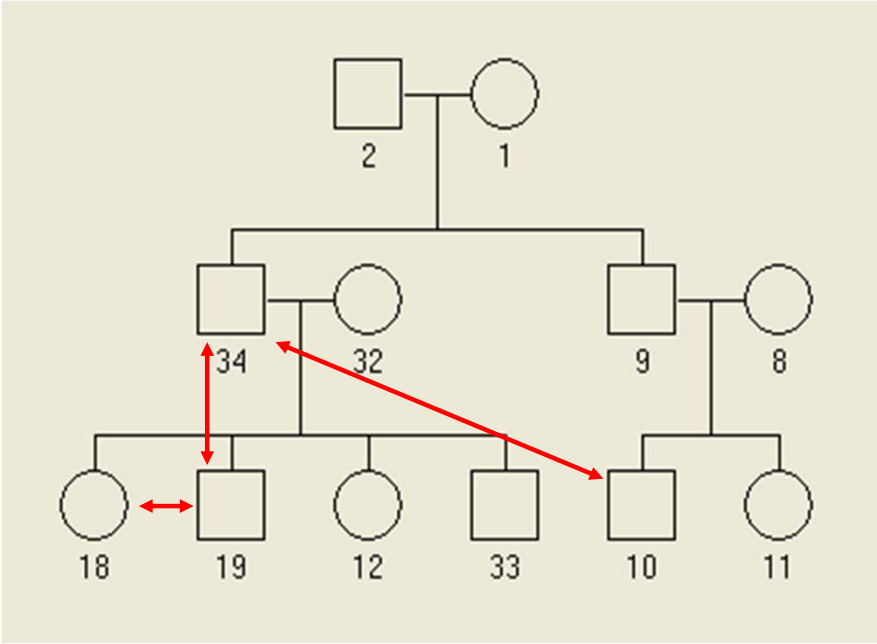
Dad

Child

Brother

Sister

Complex Kinship Testing



The statistical power for complex kinship testing significantly decreases compared to one-to-one matching

Requirements:

- Genotypes of individuals being tested
- Allele frequencies for the loci involved in the testing
- **A Hypothesis!**

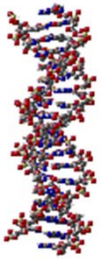
- Basic statistical equations are known
- Difficult to identify distant relationships
- Discriminatory power comes from multiple family members and the use of informative markers

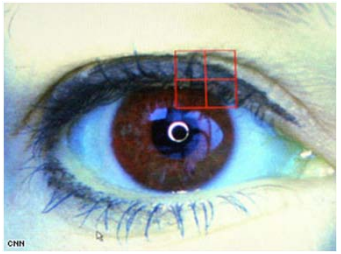
DNA as a Biometric

Current Biometrics

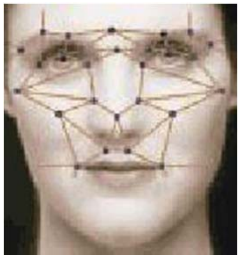
Some commonly measured features

- Physical
  - Fingerprints (Palm/hand geometry)
  - Iris, retinal
  - Face
  - Odor/scent
  - DNA?





- Behavioral
  - Gait
  - Voice
  - Vein (IR thermogram)
  - Hand geometry
  - Handwriting



Characteristics of a Biometric

- Universality
  - each person should have the characteristic
- Uniqueness
  - is how well the biometric separates individuals from another
- Permanence
  - measures how well a biometric resists aging and variance over time
- Collectability
  - ease of acquisition for measurement

Jain, A. K.; Ross, Arun; Prabhakar, Salil (January 2004), "An introduction to biometric recognition", *IEEE Transactions on Circuits and Systems for Video Technology* **14th** (1): 4–20

Characteristics of a Biometric

(practical considerations)

- Performance
  - accuracy, **speed**, and robustness of technology used
- Acceptability
  - degree of approval of a technology
- Circumvention
  - ease of use of a substitute

Jain, A. K.; Ross, Arun; Prabhakar, Salil (January 2004), "An introduction to biometric recognition", *IEEE Transactions on Circuits and Systems for Video Technology* **14th** (1): 4–20

DNA Typing as a Biometric

Advantages

- High level of accuracy
  (Gold Standard)
- Solid foundation of Forensic DNA Testing
  (pop stats, molecular biology, court acceptance, protocols, training, education)
- Kinship determination (unique to DNA)
- Potential use for:
  - Phenotype (traits; eye/hair color)
  - Biogeographical Ancestry

Challenges

- Expensive
- **Time consuming**
- Sample collection (invasive, stability issues)
- Technical expertise required for analysis
- Low level template, mixtures, PCR inhibition
- Policy/Privacy/Ethical issues

Interest in Rapid DNA Typing

- DoD (field testing, rapid intelligence, mass fatalities)
- DHS (kinship determination, border security, immigration)
- DoJ (law enforcement, initial information)
- Industry (security, authentication)

- Each customer will have specific requirements
  - sample input
  - information output
  - degrees of 'accuracy'

The time required for generating a STR profile will have to be significantly reduced

Goals for Rapid DNA Typing Systems

- Develop an integrated system capable of performing DNA testing in less than 1 hour

- Little user interaction (or experience)
- Rugged
- Robust
- Simple data interpretation
- 4-16 samples per run
- Disposable chips (with reagents on board)

Swab in…answer out

Rapid DNA Typing Systems Under Development

- Systems are currently under development and are not yet commercially available

- Network Biosystems (Woburn, MA) http://www.netbio.com
- ZyGEM and Lockheed Martin (Charlottesville,VA) http://www.zygem.com
- IntegenX (Pleasanton, CA) http://www.integenx.com
- Forensic Science Service (UK) http://www.forensic.gov.uk/

Use of DNA as a Biometric Tool American Academy of Forensic Science, Feb 22, 2010, Seattle, WA
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Biometrics Consortium Conference September, 2010 Tampa, FL
http://www.biometrics.org/bc2010/program.pdf

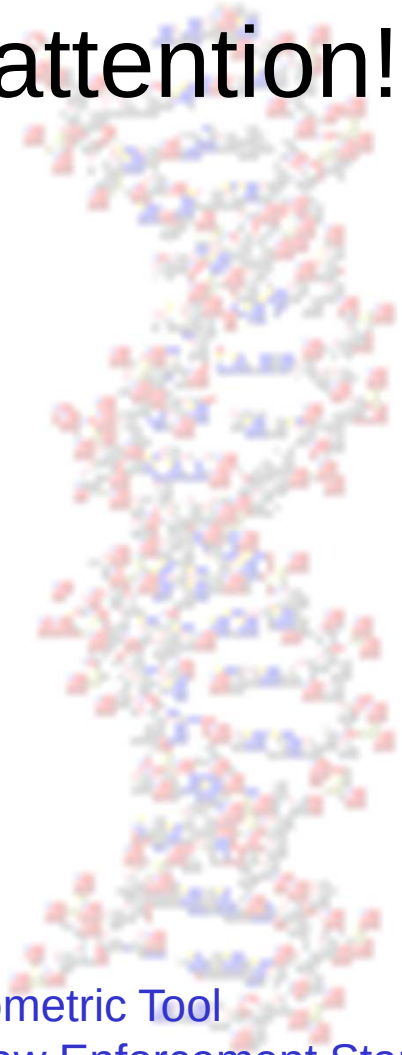
DNA Analysis Approach (integrated)

Steps Involved	Target Times	Challenges
Collection		Buccal swab, blood, other?
Extraction	~15 min	Rapid extraction (solid or liquid phase?) Reagents stable and compatible with device
Quantitation		Can be skipped for a reference sample **BUT** Does the extraction method allow for a target amount of DNA to be released? ~1 ng
Amplification	~20-30 min	Rapid PCR amplification of a commercial STR kit Locus balance, stutter, adenylation, heterozygote balance, reproducibility
Separation/ Detection	~20 min	Resolution, reproducibility, sensitivity, post-run signal processing
Data Interpretation		Expert system software? How much user intervention is needed?
Total Time	~1 hour	Can rapid typing be done reproducibly and accurately? Cost efficient? (instrumentation, reagents, consumables) ***General challenge of going from macro scale to micro scale!***

NIST Efforts with DNA Biometrics

- Developing rapid PCR protocols
- Evaluating kinship analysis software
- Support for external rapid DNA efforts
- Designing standards materials for device testing
- Testing prototype rapid DNA devices

Thank you for your attention!



Questions?

Peter.Vallone@nist.gov

301-975-4872

Acknowledgements

Erica Butts

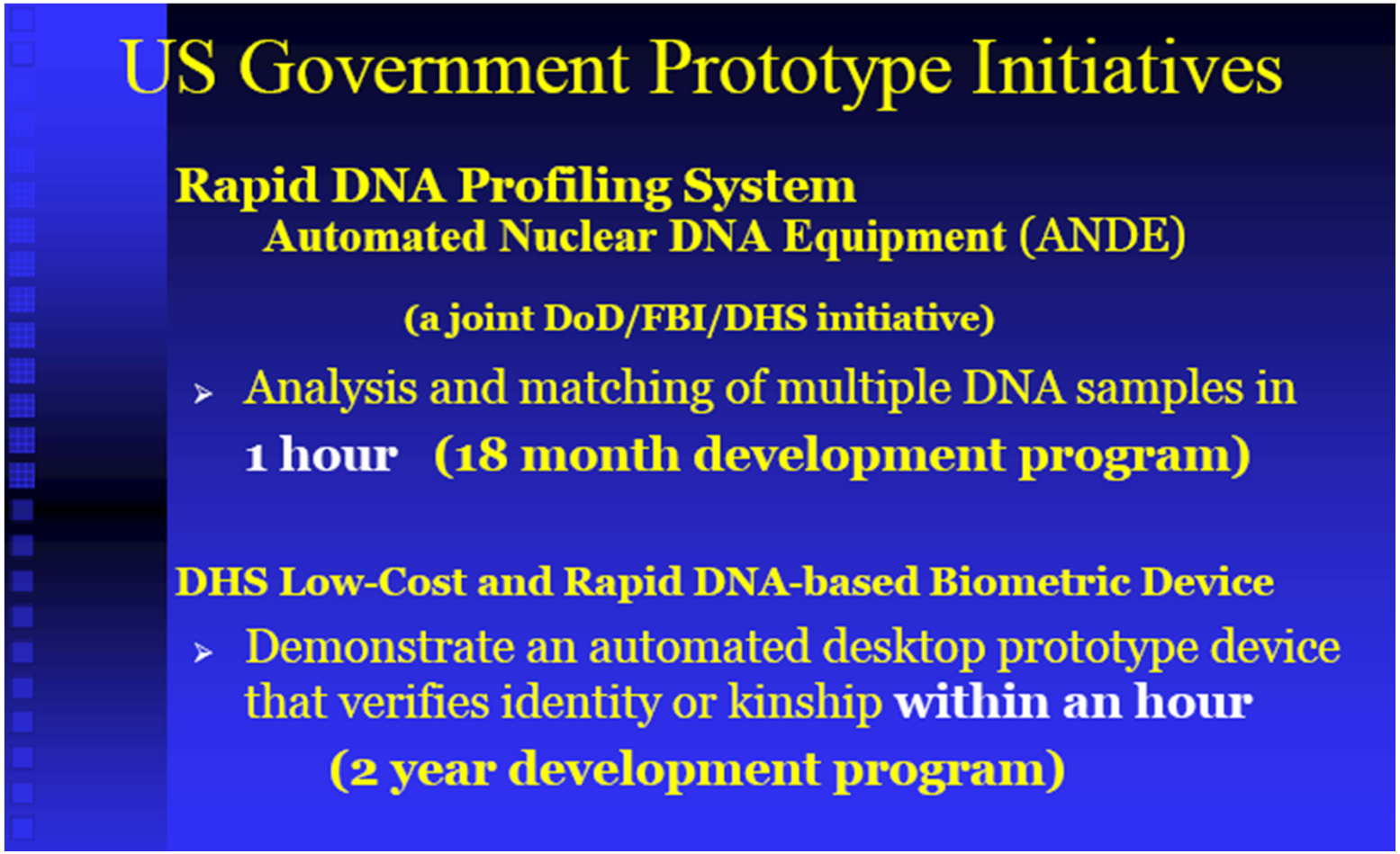
Kristen O'Connor

Outside funding agencies:

FBI - Evaluation of Forensic DNA Typing as a Biometric Tool

NIJ – Interagency Agreement with the Office of Law Enforcement Standards

ANDE (Automated Nuclear DNA Equipment)



http://biometrics.org/bc2009/presentations/wednesday/McCurdy%20BrA%20Wed%201040-1055.pdf

Recent Work with Rapid PCR

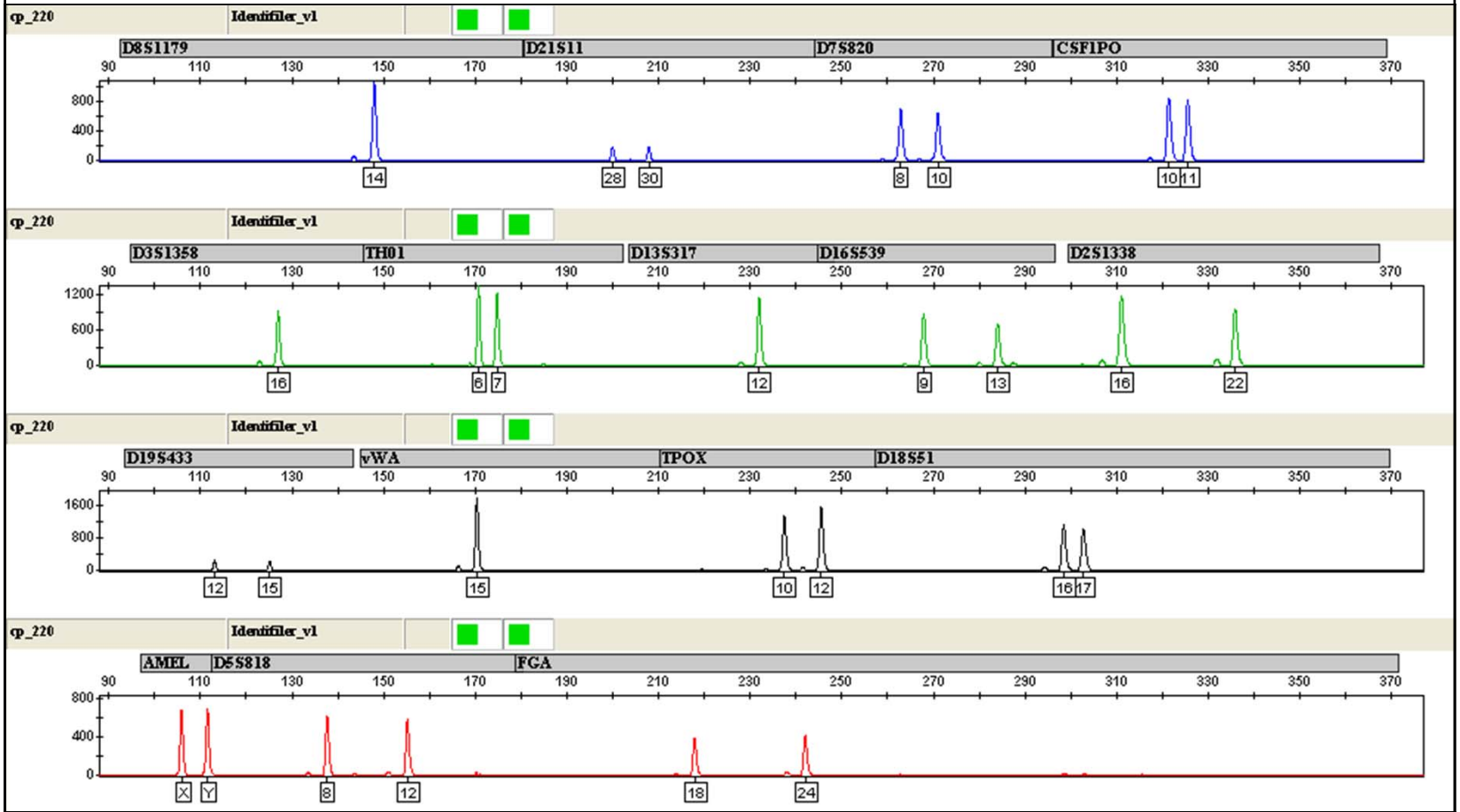
At NIST we are working on new PCR methods to reduce the time for PCR down to 20 minutes

Polymerase Chain Reaction (PCR) is a means to create billions of exact copies of the human genome – necessary/essential for DNA typing

~3.5 h → ~20 min?

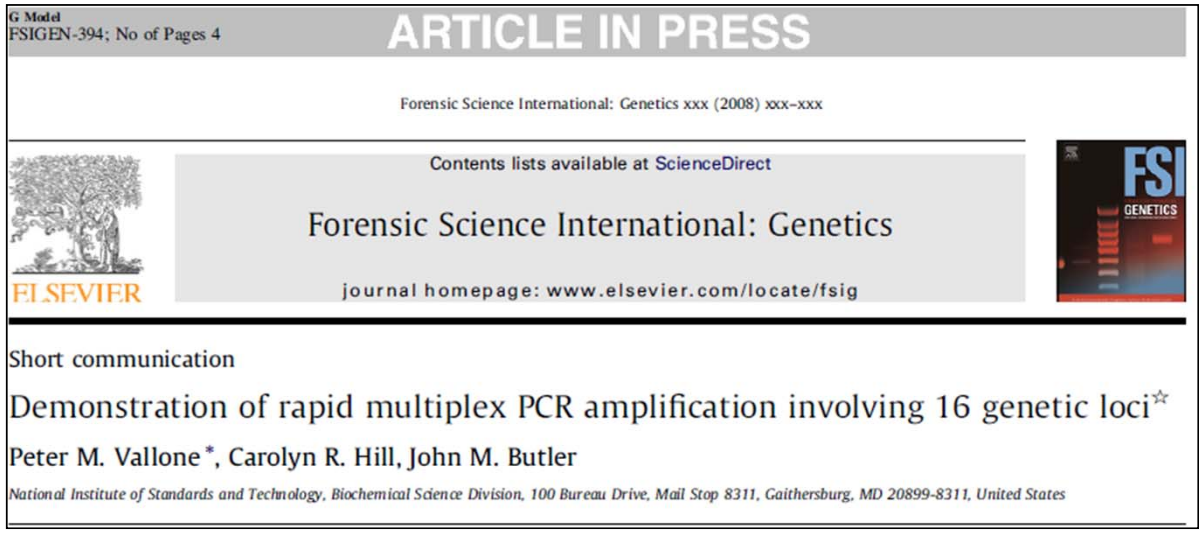
Multiplex PCR Amplification

20 Minute PCR Amplification on Cepheid Cycler



28 cycles, Identifiler STR kit, 1 ng of DNA

Rapid PCR Article



G Model
FSIGEN-394; No of Pages 4

ARTICLE IN PRESS

Forensic Science International: Genetics xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

ELSEVIER

Short communication

Demonstration of rapid multiplex PCR amplification involving 16 genetic loci☆

Peter M. Vallone*, Carolyn R. Hill, John M. Butler

National Institute of Standards and Technology, Biochemical Science Division, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899-8311, United States

Vallone, P.M., Hill, C.R., Butler, J.M. (2008) Demonstration of rapid multiplex PCR amplification involving 16 genetic loci. *FSI Genetics* 3(1): 42-45.

Rapid PCR Amplification of STR Typing Kits 20th Annual International Symposium on Human Identification (Promega Meeting) October 14, 2009, Las Vegas, NV

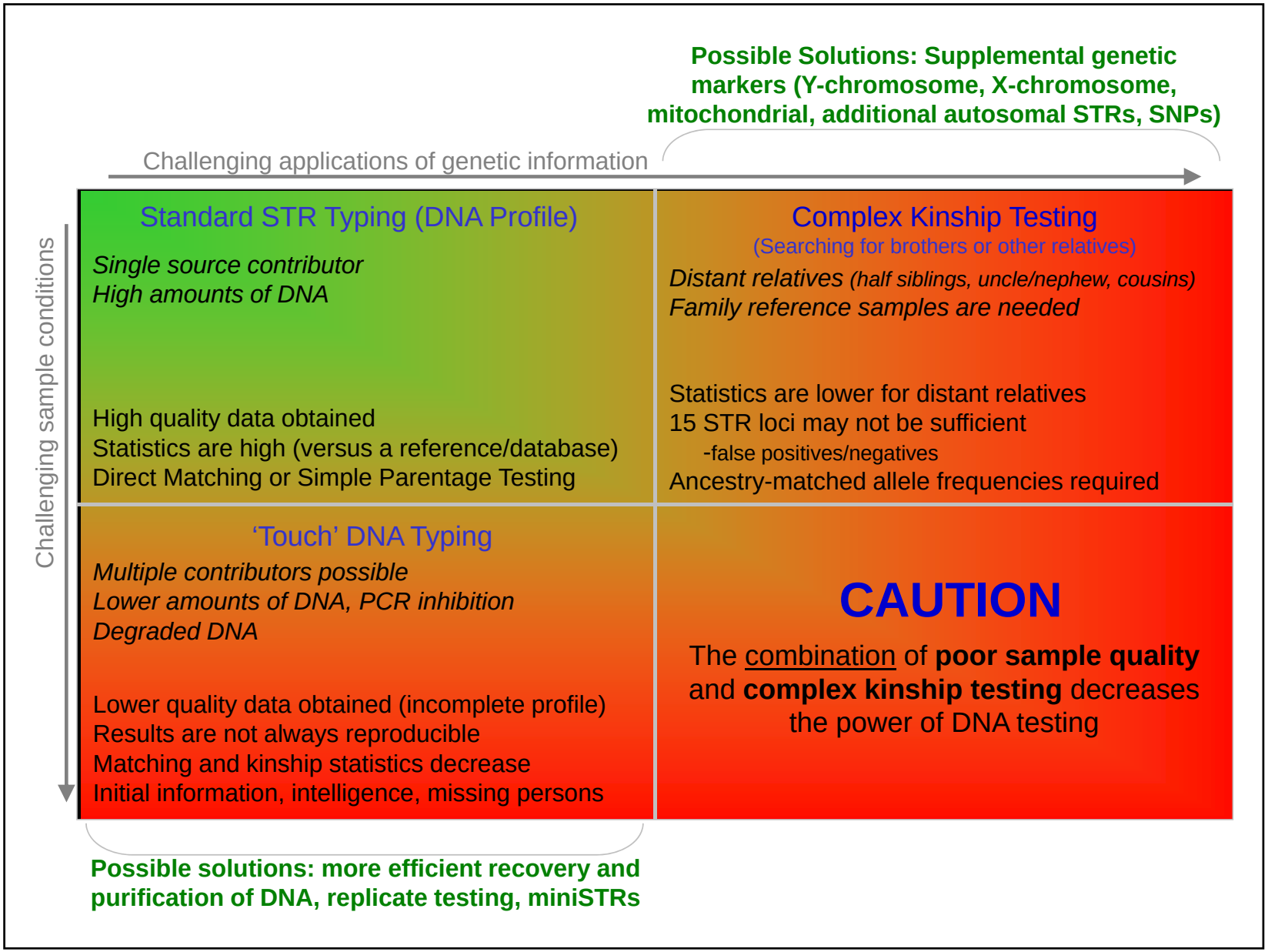
Rapid Amplification of Commercial STR Typing Kits, International Society of Forensic Genetics (ISFG), September 16, 2009, Buenos Aires, Argentina

Use of DNA as a Biometric Tool American Academy of Forensic Science, Feb 22, 2010, Seattle, WA

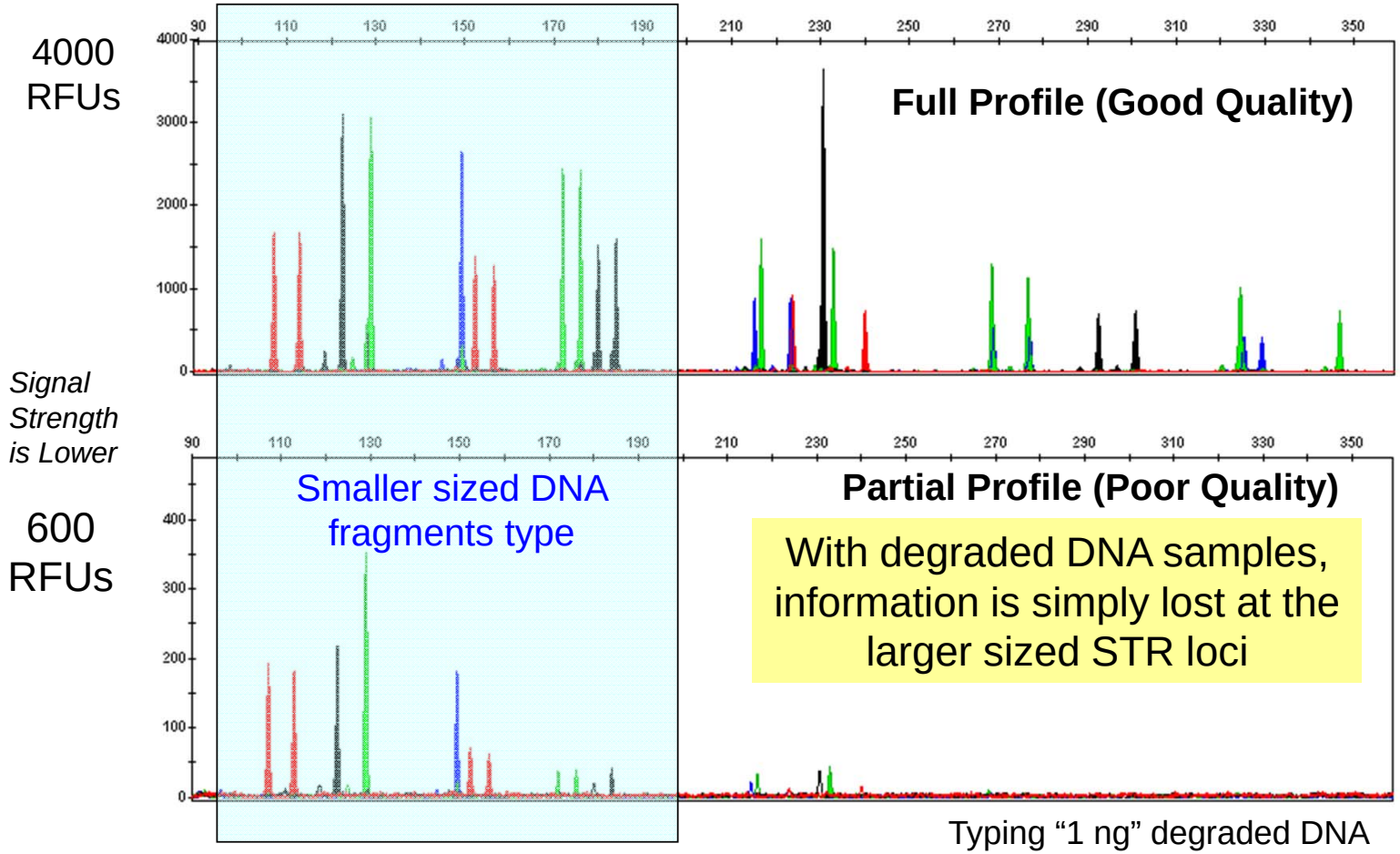
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

Future Directions

- Functional prototypes should be available for testing in the next 12-18 months
- 3-4 year horizon until concordance testing and validation
- Further education on the **strengths and limitations** of DNA

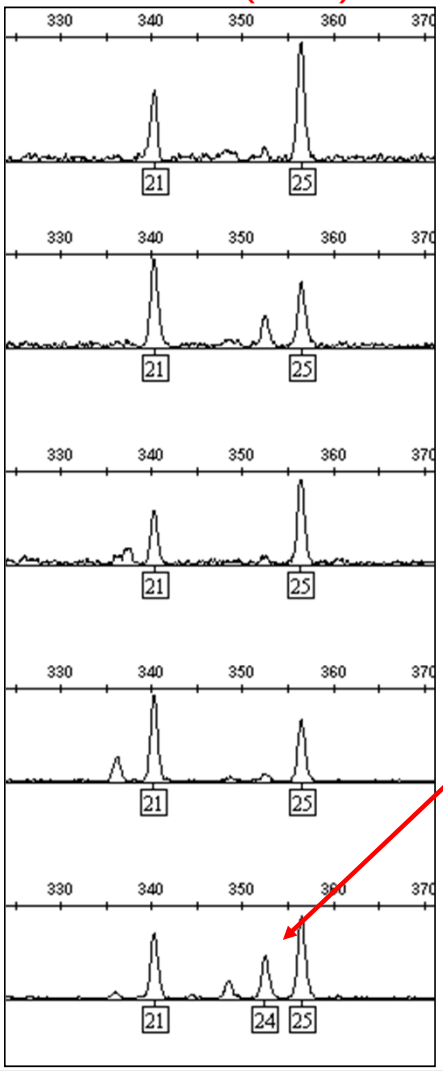

Possible Solutions: Supplemental genetic markers (Y-chromosome, X-chromosome, mitochondrial, additional autosomal STRs, SNPs)
Challenging applications of genetic information
Challenging sample conditions
Standard STR Typing (DNA Profile)
Single source contributor
High amounts of DNA
High quality data obtained
Statistics are high (versus a reference/database)
Direct Matching or Simple Parentage Testing
Complex Kinship Testing
(Searching for brothers or other relatives)
Distant relatives (half siblings, uncle/nephew, cousins)
Family reference samples are needed
Statistics are lower for distant relatives
15 STR loci may not be sufficient
-false positives/negatives
Ancestry-matched allele frequencies required
'Touch' DNA Typing
Multiple contributors possible
Lower amounts of DNA, PCR inhibition
Degraded DNA
Lower quality data obtained (incomplete profile)
Results are not always reproducible
Matching and kinship statistics decrease
Initial information, intelligence, missing persons
CAUTION
The combination of poor sample quality and complex kinship testing decreases the power of DNA testing
Possible solutions: more efficient recovery and purification of DNA, replicate testing, miniSTRs

Same DNA with Different Quality



Low Levels of DNA Template

Locus 1 (FGA)



At very low amounts of DNA template results are **not reproducible**

5 replicates of typing 10 pg of DNA are shown

Artifacts

Drop out (false negative)

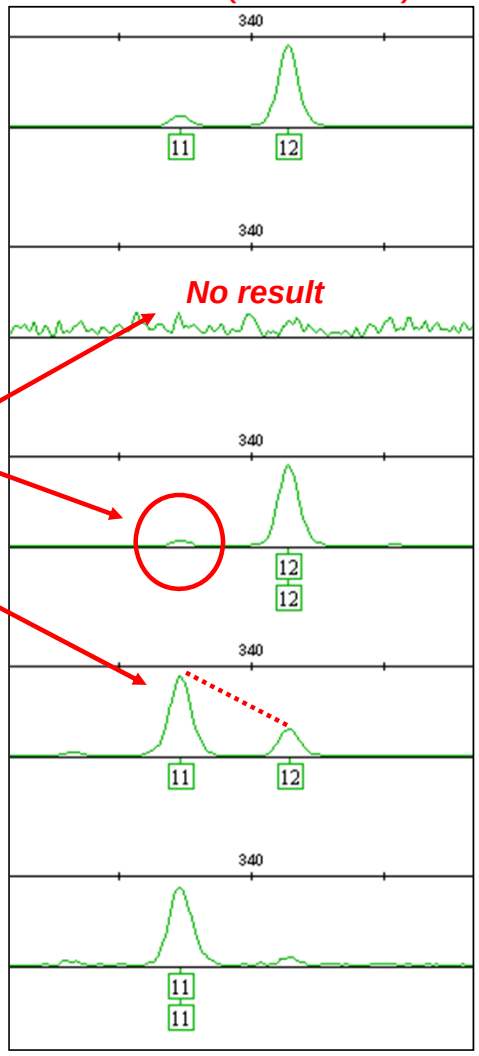
Drop in (false positive)

Peak imbalance

High stutter

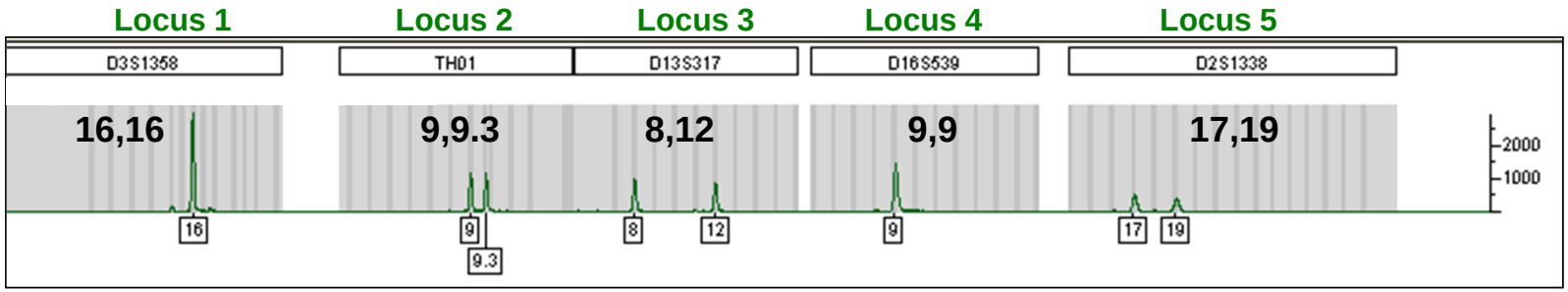
PowerPlex 16 HS (10 pg @ 34 cycles)

Locus 2 (CSF1PO)



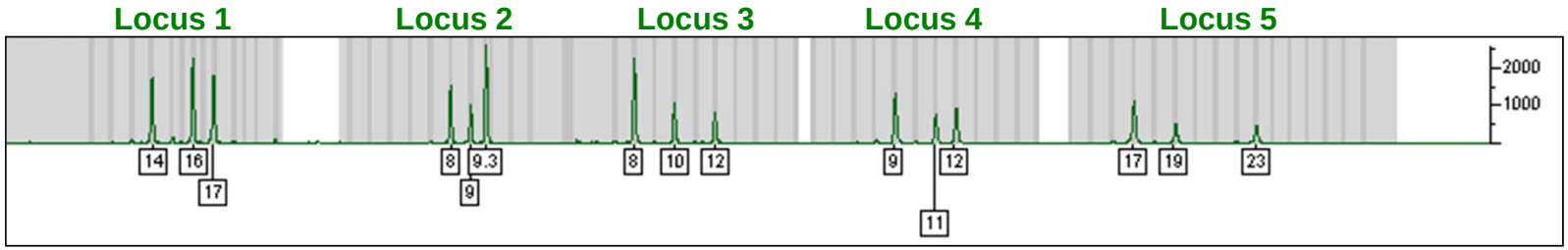
Mixture Interpretation – A Major Challenge…

Single Source Sample



One or two peaks observed at each locus (tested DNA region)

Mixture Sample



More than two peaks observed at more than two loci (tested DNA regions)

Different possible combinations could have given rise to the particular mixture observed