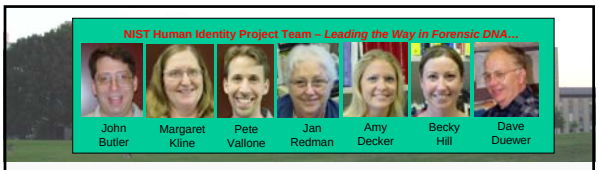


DNA Quality in the Context of Biometrics

John M. Butler, Ph.D.
Human Identity Project Team
National Institute of Standards and Technology

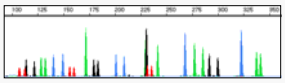
Biometric Quality Workshop II
Gaithersburg, MD
November 7, 2007

Questions to Address on DNA Quality and Potential Use in Biometrics

- How are DNA profiles generated and what information is stored?
- How long does it take to generate a DNA profile using current and near-term technologies?
- What are the primary issues impacting quality of DNA results?

Presentation Outline

- Intro to NIST Human Identity Project Team
- Overview of DNA testing process
- Efforts to ensure quality results with DNA testing



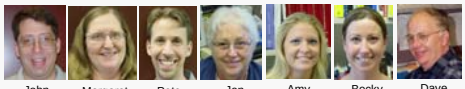
NIST and NIJ Disclaimer

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

NIST Human Identity Project Team



Publications and presentations available on STRBase: http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

- **19 publications** from Jan-Oct 2007
- **38 presentations** and **9 workshops** to the community from Jan-Oct 2007

NIST Human Identity Team Projects

Funded by the National Institute of Justice

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

Projects

33 different projects are described

[Human DNA Quantitation] [Mitochondrial DNA] [Y Chromosome] [Compromised DNA Evidence] [Miniaturization and Automation] [General Tools and Information] [Non-Human DNA] [Alternative Forensic DNA Markers]

Alphabetical Listing of Projects

ABI 3100 performance with various STR typing systems (April 2001-June 2003)

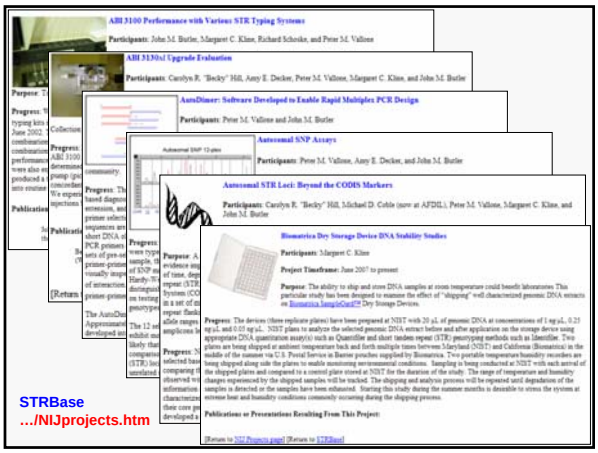
ABI 3130xl upgrade evaluation (Sept 2005-May 2006)

AutoDimer: software to enable rapid multiplex PCR design (2000-2005) [see also software.htm]

Autosomal SNP loci (July 2002-present)

Autosomal STR loci beyond the CODIS markers (Jan 2004-present) [see also newSTRs.htm]

Biomatrica dry storage device DNA stability studies (June 2007-present)



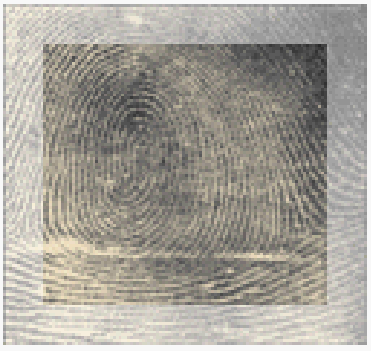
National Institute of Justice
The Research, Development, and Evaluation Agency of the U.S. Department of Justice

Current Areas of NIST Effort with Forensic DNA

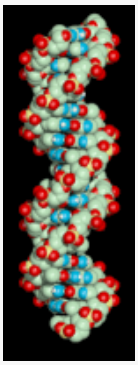
- **Standards** http://www.cstl.nist.gov/biotech/strbase/
  - Standard Reference Materials
  - Standard Information Resources (STRBase website)
  - Interlaboratory Studies
- **Technology**
  - Research programs in SNPs, miniSTRs, Y-STRs, mtDNA, qPCR
  - Assay and software development
- **Training Materials**
  - Review articles and workshops on STRs, CE, validation
  - PowerPoint and pdf files available for download

Overview of DNA Typing Process

Methods for Human Identification



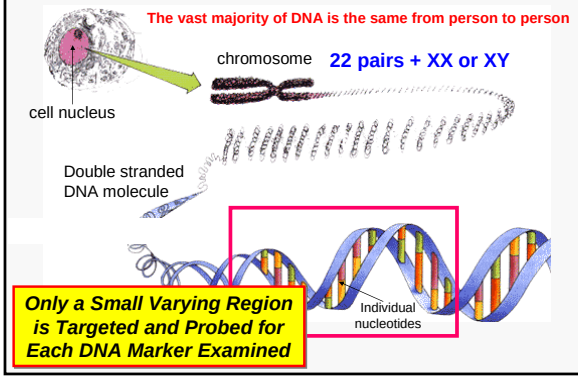
Fingerprints have been used since 1901



DNA since 1986

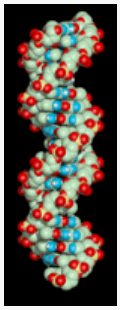
DNA in the Cell

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



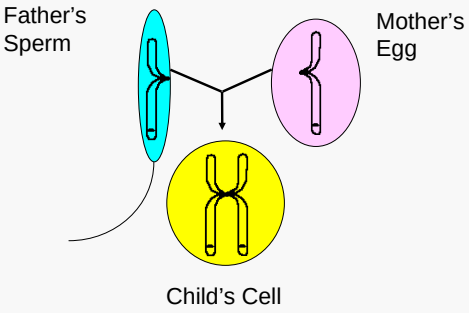
Only a Small Varying Region is Targeted and Probed for Each DNA Marker Examined

Characteristics of DNA

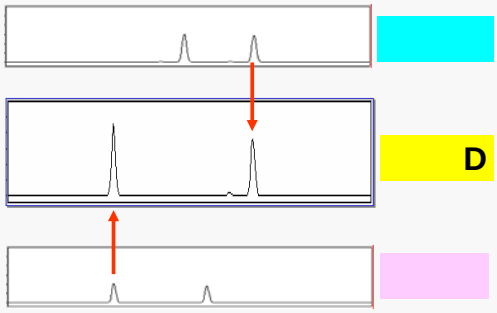


- Each person has a unique DNA profile (except identical twins).
- Each person's DNA is the same in every cell.
- An individual's DNA profile remains the same throughout life.
- Half of your DNA comes from your mother and half from your father.

Our DNA Comes from our Parents



Inheritance Pattern of DNA Profiles



Basis of DNA Profiling

The genome of **each individual is unique** (with the exception of identical twins) and **is inherited from parents**

Probe subsets of genetic variation in order to differentiate between individuals (statistical probabilities of a random match are used)

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

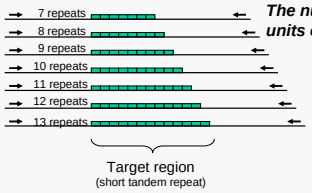
Current standard DNA tests **DO NOT look at genes** – little/no information about race, predisposal to disease, or phenotypical information (eye color, height, hair color) is obtained

Short Tandem Repeat (STR) Markers

An accordion-like DNA sequence that occurs between genes

TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGAAGACA
GGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGA**
TAGATATCATTGAAAGACAAAACAGAGATGGATGATAGATACAT
GCTTACAGATGCACAC

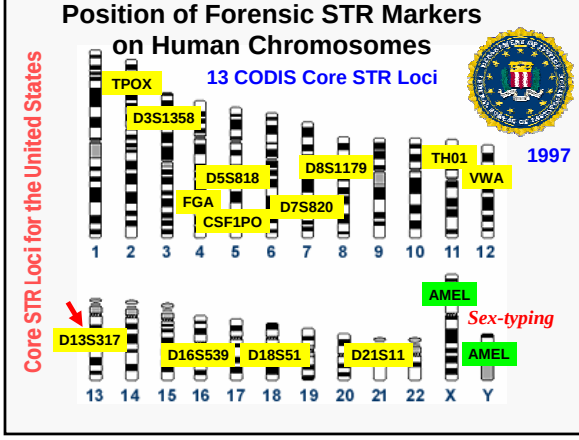
= 11 GATA repeats ("11" is all that is reported)



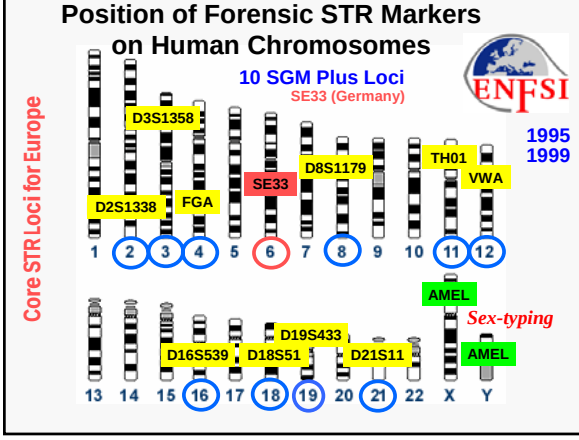
The number of consecutive repeat units can vary between people

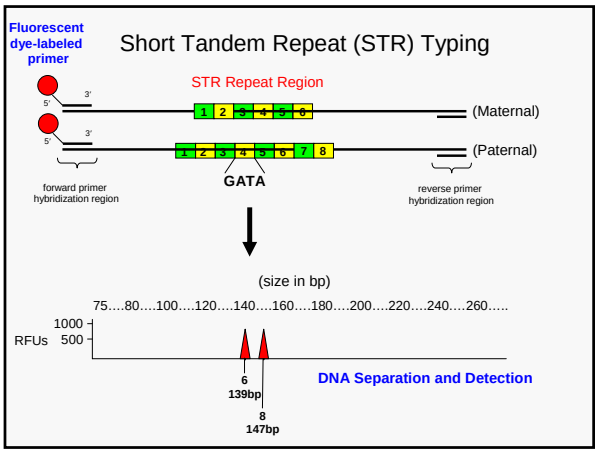
The FBI has selected **13 core STR loci** that must be run in all DNA tests in order to provide a common currency with DNA profiles

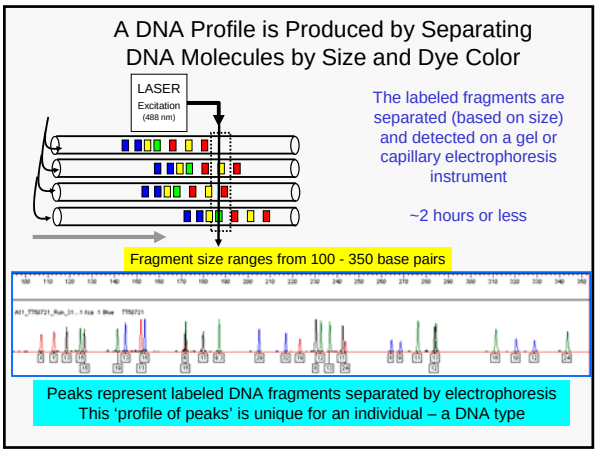
Position of Forensic STR Markers on Human Chromosomes

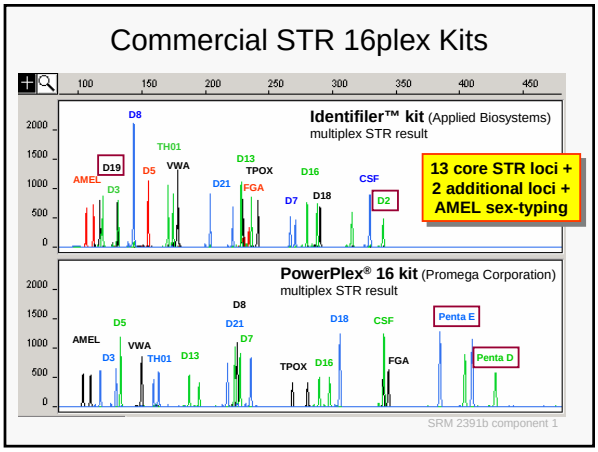


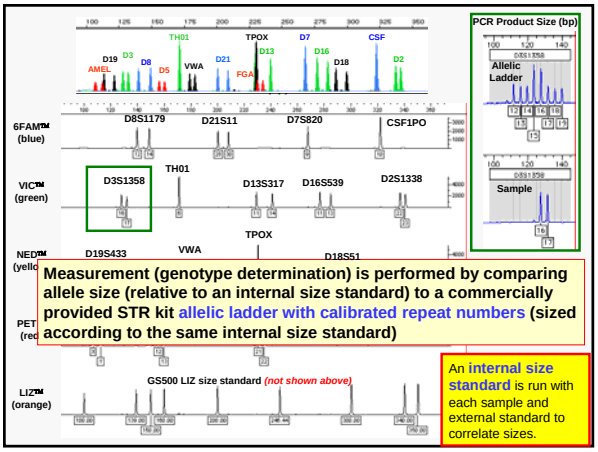
Position of Forensic STR Markers on Human Chromosomes







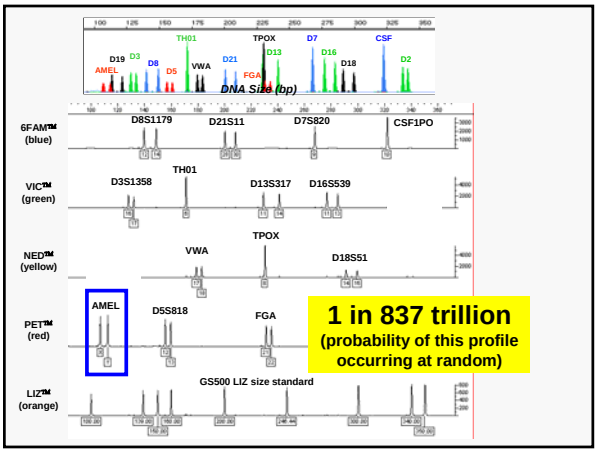




How Statistical Calculations are Made

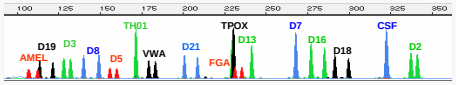
- Generate data with set(s) of samples from desired population group(s)
  - Generally only 100-150 samples are needed to obtain reliable allele frequency estimates
- Determine allele frequencies at each locus
  - Count number of each allele seen
- Allele frequency information is used to estimate the rarity of a particular DNA profile
  - Homozygotes (p^2), Heterozygotes (2pq)
  - Product rule used (multiply locus frequency estimates)

For more information, see Chapters 20 and 21 in *Forensic DNA Typing*, 2nd Edition



DNA Profile Frequency with all 13 CODIS STR loci

AmpFISTR® Identifiler™ (Applied Biosystems)



What would be entered into a DNA database for searching: 16,17-17,18-21,22-12,14-28,30-14,16-12,13-11,14-9,9-9,11-6,6-8,8-10,10

Locus	allele	value	allele	value	1 in	Combined
D3S1358	16	0.2533	17	0.2152	9.17	9.17
VWA	17	0.2815	18	0.2003	8.87	81
FGA	21	0.1854	22	0.2185	12.35	1005
D8S1179	12	0.1854	14	0.1656	16.29	16,364
D21S11	28	0.1589	30	0.2782	11.31	185,073
D18S51	14	0.1374	16	0.1391	26.18	4,845,217
D5S818	12	0.3841	13	0.1407	9.25	44,818,259
D13S317	11	0.3394	14	0.0480	30.69	1.38 x 10^9
D7S820	9	0.1772			31.85	4.38 x 10^{10}
D16S539	9	0.1126	11	0.3212	13.8	6.05 x 10^{11}
THO1	6	0.2318			18.62	1.13 x 10^{13}
TPOX	8	0.5348			3.50	3.94 x 10^{13}
CSF1PO	10	0.2169			21.28	8.37 x 10^{14}

PRODUCT RULE

The Random Match Probability for this profile in the U.S. Caucasian population is **1 in 837 trillion (10^{12})**

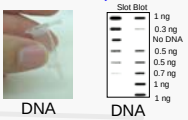
Steps in DNA Analysis

Usually 1-2 day process (a minimum of ~5 hours)

Steps Involved

- Collection
- Specimen Storage
- Extraction
- Quantitation
- Multiplex PCR
- STR Typing
- Interpretation of Results
- Database Storage & Searching
- Calculation of Match Probability

Blood Stain Buccal swab
Sample Collection & Storage



DNA Extraction DNA Quantitation

Biology

Multiplex PCR Amplification

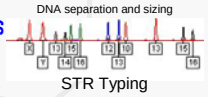
Genetics

If a match occurs, comparison of DNA profile to population allele frequencies to generate a case report with probability of a random match to an unrelated individual

DNA Database Search

Technology

DNA separation and sizing



STR Typing

Male: 13,14-15,16-12,13-10,13-15,16

Interpretation of Results

National DNA Index System (NDIS)

No names are associated with DNA profiles uploaded to NDIS
If my profile was entered for searching:
16,17-17,18-21,22-12,14-28,30-14,16-12,13-11,14-9,9-9,11-6,6-8,8-10,10

http://www.fbi.gov/hq/lab/codis/index1.htm

Combined DNA Index System (CODIS)

- Launched in October 1998 and now links all 50 states
- Used for linking serial crimes and unsolved cases with repeat offenders
- Convicted offender and forensic case samples along with a missing persons index

Requires 13 core STR markers

>50,000 investigations aided nationwide as of Nov 2007

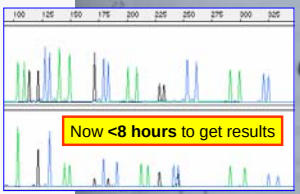
Contains more than 5 million DNA profiles

How Long Does It Take to Get DNA Results?

And At What Cost?

Progress Since 1995…

Almost **8 weeks** needed to get results



Now **<8 hours** to get results

O.J. Simpson DNA testing was performed with RFLP

Time Required for Testing

Now typically a minimum of 4-5 hours

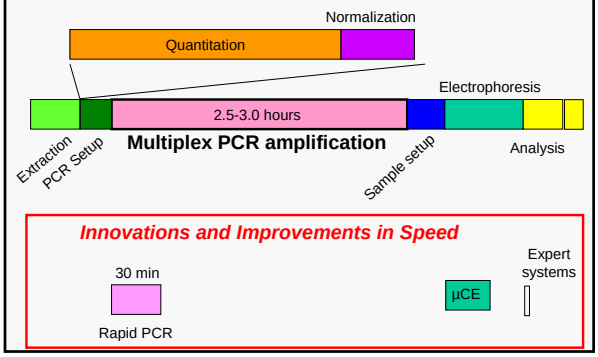
Collection	
Extraction	Could be <5 minutes
Quantitation	Not necessary if samples are uniform in amount
Amplification	Rapid thermal cycling to-date done with singleplexes; **typically 2-3 hours**
Genotyping	DNA separations (STR analysis) of <5 minutes have been demonstrated; **typically ~30 minutes**
Interpretation of Results	Currently performed manually in most labs; expert systems are under development to enable rapid interpretation
Database Storage & Searching	Search could be similar to fingerprint search in terms of speed

Biggest problem is length of time for PCR (with multiplex amplification)

Comparison a DNA profile to a reference or database

Male: 13,14-15,16-12,13-10,13-15,16-.....

Relative Time for Overall DNA Process



Cost of DNA Tests

- In high-throughput databanking laboratories today, a DNA profile can be generated for ~$20-30 per 13-locus STR profile (single source sample)
- Forensic casework or parentage testing work typically costs more.

DNA Quality Issues

Brief Historical Overview

Profiles in DNA (Sept 1999) 3(2): 10-11

CURRENT EVENTS

The Evolution of Quality Standards for Forensic DNA Analyses in the United States

By Special Agent Lawrence A. Presley, MS, MA
Federal Bureau of Investigation Laboratory, Washington, DC
lpresley@fbi.gov

Quality problems in late 1980s with DNA testing
TWGDAM established under FBI Lab sponsorship in 1988
NRC I (1992) and NRC II (1996) issued reports recommending formal QA programs
DNA Identification Act of 1994 lead to formation of DNA Advisory Board (DAB)
DAB Standards issued in Oct 1998 and Apr 1999
When DAB was dissolved in 2000, SWGDAM assumed leadership role

NIST had membership on the DNA Advisory Board and actively participates in SWGDAM

Scientific Working Group on DNA Analysis Methods (SWGDAM)

- Organized originally by FBI Laboratory as Technical Working Group on DNA Analysis Methods (TWGDAM) in 1988
- Meets semiannually – each January and July
- **Organized into eight subcommittees**:
  - Quality Assurance, CODIS, mtDNA, Mass Disasters/Missing Persons, Expert Systems, Serology, Y-STRs, and Mixture Interpretation
- Membership (usually ~50 attend) from public forensic DNA laboratories around the U.S.

Organizations Aiding Forensic DNA Standardization

The NIST Human Identity Project Team participates in EDNAP, ENFSI, and ISFG.

http://www.isfg.org/ednap/ednap.htm

- European DNA Profiling Group (EDNAP)
  - Working group of **International Society of Forensic Genetics (ISFG)**
  - Examine technologies and run interlab studies
  - 28 participants from 19 different countries

http://www.isfg.org/

- European Network of Forensic Science Institutes (ENFSI)
  - Defines policy within European Union
  - **ENFSI DNA Working Group** equivalent of SWGDAM
  - 85 participants from 32 different countries

http://www.enfsi.eu/

Have challenges with language differences due to many countries involved

DNA Identification Act (1994)

Public Law 103-322

42 § 14131. Quality assurance and proficiency testing standards

(a) Publication of quality assurance and proficiency testing standards

(1) (A) Not later than 180 days after September 13, 1994, the Director of the Federal Bureau of Investigation shall appoint an advisory board on DNA quality assurance methods from among nominations proposed by the head of the National Academy of Sciences and professional societies of crime laboratory officials.

(B) The advisory board shall include as members scientists from State, local, and private forensic laboratories, molecular geneticists and population geneticists not affiliated with a forensic laboratory, and a representative from the National Institute of Standards and Technology.

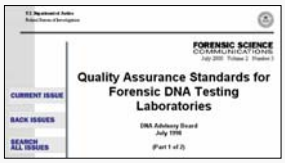
(C) **The advisory board shall develop, and if appropriate, periodically revise, recommended standards for quality assurance**, including standards for testing the proficiency of forensic laboratories, and forensic analysts, in conducting analyses of DNA.

DNA Advisory Board (DAB)

DAB Standards issued in 1998-1999

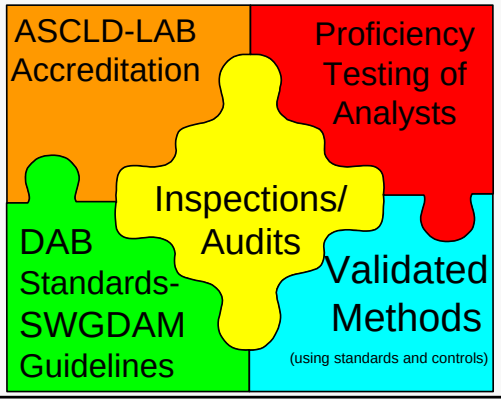
Quality Assurance Standards (QAS)

1. SCOPE
2. DEFINITIONS
3. QUALITY ASSURANCE PROGRAM
4. ORGANIZATION AND MANAGEMENT
5. PERSONNEL
6. FACILITIES
7. EVIDENCE *(SAMPLE)* CONTROL
8. VALIDATION
9. ANALYTICAL PROCEDURES
10. EQUIPMENT CALIBRATION AND MAINTENANCE
11. REPORTS
12. REVIEW
13. PROFICIENCY TESTING
14. CORRECTIVE ACTION
15. AUDITS
16. SAFETY
17. SUBCONTRACTOR OF ANALYTICAL TESTING FOR WHICH VALIDATED PROCEDURES EXIST



http://www.fbi.gov/hq/lab/fsc/backissu/july2000/codis2a.htm
http://www.fbi.gov/hq/lab/fsc/backissu/july2000/codis1a.htm

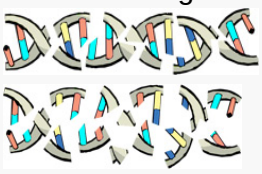
Ensuring Accurate Forensic DNA Results



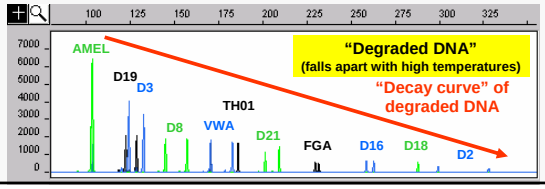
Checks and Controls on DNA Results

Community	FBI DNA Advisory Board's Quality Assurance Standards *(also interlaboratory studies)* ISO17025
Laboratory	ASCLD/LAB Accreditation and Audits
Analyst	Proficiency Tests & Continuing Education
Method/Instrument	**Validation of Performance** ***(along with traceable standard sample)***
Protocol	Standard Operating Procedure is followed
Data Sets	Allelic ladders, positive and negative amplification controls, and reagent blanks are used
Individual Sample	Internal size standard present in every sample
Interpretation of Result	Second review by qualified analyst/supervisor
Court Presentation of Evidence	Defense attorneys and experts with power of discovery requests

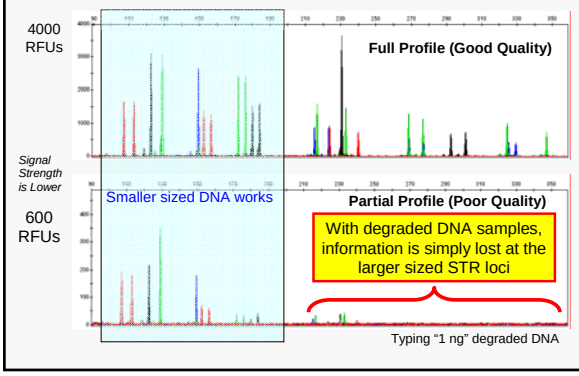
Degraded DNA



Larger segments of DNA cannot be recovered when DNA molecules have fragmented into small pieces (caused by heat, water, or bacteria)



Same DNA with Different Quality



Impact of Degraded DNA Samples

- Comparison to a phone number (string of 13 numbers)

001-301-975-4049

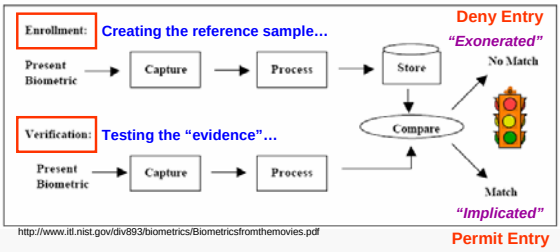
- If you only had "4049"…this information would be of limited value since it is not as specific (and could match other phone numbers from different area codes)
- DNA profiles are essentially a string of numbers – **if the DNA is damaged, then the string of numbers is shorter and less informative…**

----------4049 or ----301-9-------

DNA Data Quality

- The raw DNA data itself does not have quality scores directly attached to it.
- **Only the STR allele designations are stored without an indication of data quality.**
- Checks and balances exist in the entire system to try and ensure good quality data.
- Retesting of offender database sample is performed when a DNA database hit is observed.

DNA within the Biometric Model



http://www.itl.nist.gov/div893/biometrics/Biometricsfromthemovies.pdf

Match of 13 points (each with 2 variable alleles) within DNA

String of 26 numbers (order of listing DNA results would have to be standardized)
16,17-17,18-21,22-12,14-28,30-14,16-12,13-11,14-9,9-11,13-6,6-8,8-10,10

Summary

- Short tandem repeat (STR) markers are widely used for human identity testing applications.
- Core STR loci have been settled upon with some overlap between the U.S. and Europe.
- DNA analysis involving STR typing currently takes multiple hours to complete at a minimum cost of $20 with no near-term solution to speed up this process.
- Standards for quality assurance are in place but quality scores are not used on individual DNA data as only STR allele calls are stored.

Thank you for your attention…

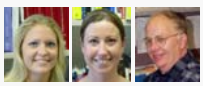
Our team publications and presentations are available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

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

Questions?

john.butler@nist.gov
301-975-4049

Margaret Kline | Pete Vallone | Jan Redman



Amy Decker | Becky Hill | Dave Duewer

Funding from the **National Institute of Justice (NIJ)** through NIST Office of Law Enforcement Standards