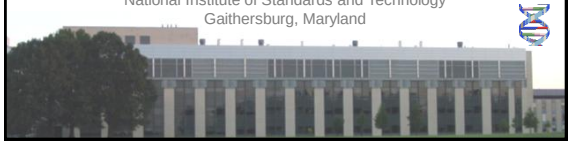
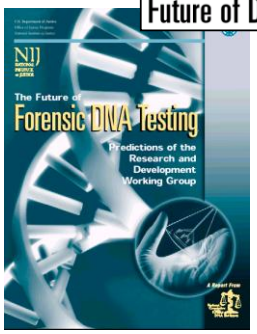
CIB Forensic Science Center
Training Seminar (Taipei, Taiwan)
June 6-7, 2012

The Future of Forensic DNA Typing

John M. Butler
NIST Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland



National Commission on the Future of DNA Evidence



http://www.ojp.usdoj.gov/nij/pubs-sum/183697.htm

- Report published in Nov 2000
- Asked to estimate where DNA testing would be 2, 5, and 10 years into the future

Conclusions

STR typing is here to stay for a few years because of DNA databases that have grown to contain millions of profiles

My Thoughts on the Future

- Near-term future
  - New autosomal STR loci for expanded core loci
  - Expanded use of databases (e.g., familial searching)
  - Rapid DNA testing
- More distant future
  - Next-generation DNA sequencing?
  - Loci besides STRs for identity testing?
  - Phenotyping capabilities?

STRs vs SNPs Article

Butler et al. (2007) STRs vs SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine and Pathology* 3:200-205


Forensic Sci Med Pathol (2007) 3:200–205
DOI 10.1007/s12024-007-0018-1

ORIGINAL PAPER

STRs vs. SNPs: thoughts on the future of forensic DNA testing

John M. Butler · Michael D. Coble · Peter M. Vallone

- **SNPs are unlikely to replace STRs** for routine forensic DNA testing due to challenges with high-level multiplexing and mixture detection/interpretation
- Most likely use of SNPs will be as ancestry-informative markers (AIMs) **for sample ethnicity estimation**

Harry T. Edwards
U.S. Court of Appeals (DC)
Co-Chair, Forensic Science Committee

National Academies Report on Forensic Science

- Released February 18, 2009
- Entitled "Strengthening Forensic Science in the United States: A Path Forward"
- 13 recommendations provided to Congress
- **Recommends establishing a National Institute of Forensic Science (NIFS)**
- **NIST will have a role in NIFS and our group has been asked to contribute expertise** regarding validation and testing of DNA systems as a model for other forensic disciplines

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Forensic Science Review Article

See June 15, 2009 issue of *Analytical Chemistry*


Anal. Chem. **2007**, *79*, 4365–4384

Forensic Science

T. A. Brettell*
Department of Chemical and Physical Sciences, Cedar Crest College, 100 College Drive, Allentown, Pennsylvania 18104-6196

J. M. Butler
Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8311

J. R. Almirall
Department of Chemistry and Biochemistry and International Forensic Research Institute, Florida International University, University Park, Miami, Florida 33199

2009 review article covers 160 DNA articles published in 2007-2008

Value of a Historical Review

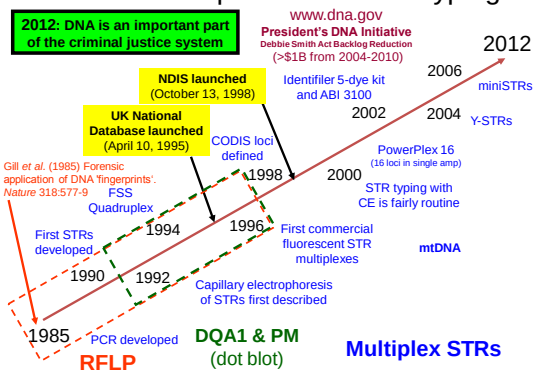
"If you want to understand today, you have to search yesterday."

– Attributed to Pearl Buck
(http://www.quotegarden.com/history.html)

Pearl Buck

The Nobel Prize in Literature 1938

Historical Perspective on DNA Typing



Stages of Forensic DNA Progression

Stages	Time Frame	Description
Exploration	1985-1995	Beginnings, different methods tried (RFLP and early PCR)
Stabilization	1995-2005	Standardization to STRs, selection of core loci, implementation of Quality Assurance Standards
Growth	2005-2012	Rapid growth of DNA databases, extended applications pursued
Sophistication	*The Future*	*Expanding tools available, confronting privacy concerns*

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)

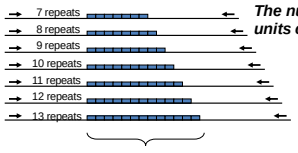
Current standard DNA tests **DO NOT look at genes** – little/no information about race, predisposal to disease, or phenotypical information (eye color, height, hair color) is obtained

Short Tandem Repeat (STR) Markers

An accordion-like DNA sequence that occurs between genes

TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGAAGACA
GGTG**GATAGATAGATAGATAGATAGATAGATAGATAGATAGATA**
GATATCATTGAAAGACAAAACAGAGATGGATGATAGATACATGCT
TACAGATGCACAC

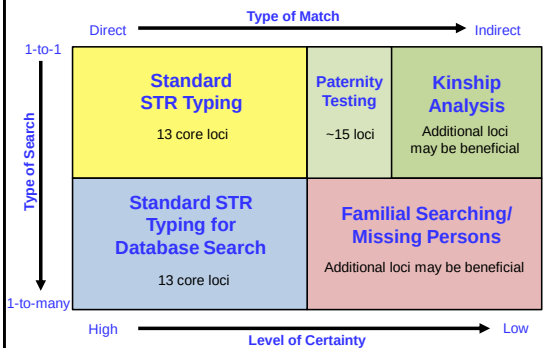
= 11 GATA repeats ("11" is all that is reported)



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

Expanding the Forensic Core Competency



Growth of DNA Databases

- Within the U.S., we have benefited from significant federal funding over the past seven years
- Expanded laws now enable more offenders to be included
- Have effectively locked technology with core STR markers used to generate DNA profiles that now number greater than 10 million profiles

Growth in Numbers of U.S. States **Requiring DNA Collection** for Various Offenses

Table 12.5 in forthcoming Butler (2009) *Fundamentals of Forensic DNA Typing*

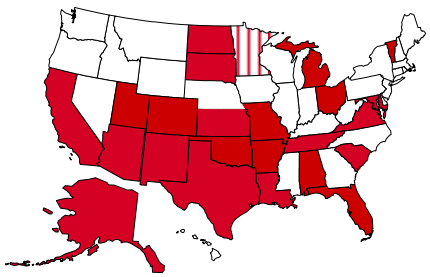
Offenses	Number of States		
	1999	2004	2008
Sex crimes	50	50	**50**
All violent crimes	36	48	**50**
Burglary	14	47	**50**
All felons	5	37	**47**
Juveniles	24	32	**32**
Arrestees/suspects	1	4	**14**

Sources: http://www.dnaresource.com and http://www.ncsl.org/programs/cj/dnadatabanks.htm

Starting initially with sex crimes, each category has grown in the past decade… burglary, all felons, arrestees…

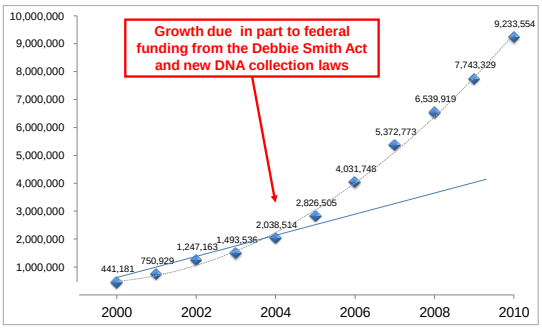
Half of the U.S. Requires Arrestee DNA Testing

+ Federal & DoD



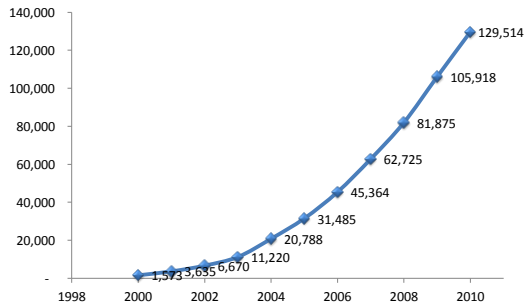
Data as of July 2010

Number of Offender DNA Profiles in the U.S. National DNA Database



Source: FBI Laboratory's CODIS Unit

Number of Investigations Aided in the U.S. National DNA Database



Source: FBI Laboratory's CODIS Unit

National DNA Index System (NDIS)

FBI LABORATORY

CODIS COMBINED DNA INDEX SYSTEM

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

>170,000 investigations aided nationwide as of April 2012

Contains more than 11 million DNA profiles

Growth in Numbers of DNA Profiles Present in Various NDIS Indices

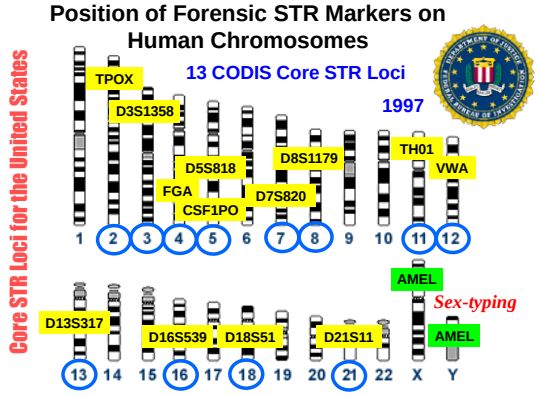
(cumulative totals by year)

Year ending Dec 31	Forensic	Convicted Offender	Arrestee	Total Offender*
2000	21,625	441,181	--	441,181
2001	27,897	750,929	--	750,929
2002	46,177	1,247,163	--	1,247,163
2003				
2004				
2005				
2006				
2007				
2008	248,943	6,398,874	140,719	6,539,919
2009	298,369	7,389,917	351,926	7,743,329
2010	**351,951**	8,559,841	668,849	**9,233,554**

Source: FBI Laboratory's CODIS Unit

In the last two years (2009, 2010):
103,008 forensic samples added
2,693,635 offender samples added

Position of Forensic STR Markers on Human Chromosomes



Expanding the CODIS Core Loci

D.R. Hares (2012) Expanding the CODIS Core Loci in the United States. *Forensic Sci. Int. Genet.* 6: e52-e54
Addendum to expanding the CODIS core loci in the United States, Forensic Sci. Int. Genet. (2012) doi:10.1016/j.fsigen.2012.01.003

Letter to the Editor

Expanding the CODIS core loci in the United States

CODIS Core Loci Working Group
Formed in May 2010 to make recommendations to FBI CODIS Unit

Douglas Hares (Chair) – FBI
John Butler – NIST
Cecelia Crouse – FL PBSO
Brad Jenkins – VA DFS
Ken Konzak – CA DOJ
Taylor Scott – IL SP

major reasons for expanding the CODIS core loci in the United States:
(1) To reduce the likelihood of adventitious matches [7] as the number of profiles stored at NDIS continues to increase each year (expected to total over 10 million profiles by the time of this publication). There are no signs that this trend will slow down as States expand the coverage of their DNA database programs and increase laboratory efficiency and capacity.
(2) To increase international compatibility to assist law enforcement data sharing efforts.
(3) To increase discrimination power to aid missing persons cases.

Three major reasons for expanding the CODIS core loci in the United States

D.R. Hares (2012) *Forensic Sci. Int. Genet.* 6(1):e52-e54

- **To reduce the likelihood of adventitious matches** as the number of profiles stored at NDIS continues to increase each year
- **To increase international compatibility** to assist law enforcement data sharing efforts
- **To increase discrimination power to aid missing persons cases**

International Comparability

Currently there are 24 autosomal STR markers present in commercial kits

U.S.	Europe
TPOX	
CSF1PO	
D5S818	
D7S820	
D13S317	
FGA	FGA
vWA	vWA
D3S1358	D3S1358
D8S1179	D8S1179
D18S51	D18S51
D21S11	D21S11
TH01	TH01
D16S539	D16S539
D2S1338	D2S1338
D19S433	D19S433
Penta D	
Penta E	
	D12S391
	D1S1656
	D2S441
	D10S1248
	D22S1045
	SE33

13 CODIS loci

ESS = European Standard Set

7 ESS loci

3 miniSTR loci developed at NIST

5 loci adopted in 2009 to expand to 12 ESS loci

Locus used in China ← D6S1043

SE33 → Core locus for Germany

Possible scenarios for extending sets of genetic markers to be used in national DNA databases

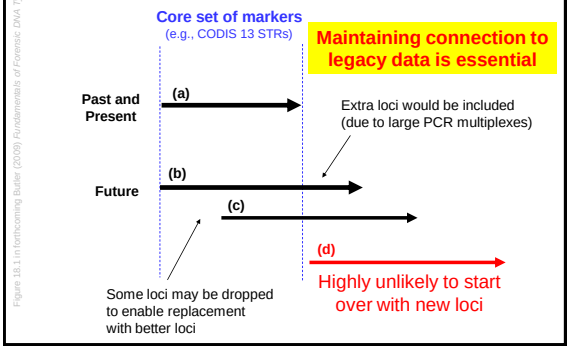


Figure 18.1 in forthcoming Butler (2009) Fundamentals of Forensic DNA Typing

Proposed Expanded CODIS Core Loci

D.R. Hares (2012) *Forensic Sci. Int. Genet.* 6(1):e52-e54

20 required loci
Amelogenin (for sex-typing)
18 autosomal STRs
1 Y-STR (DYS391)

	Locus
Section A (required)	Amelogenin
	D18S51
	FGA
	D21S11
	D8S1179
	vWA
	D13S317
	D16S539
	D7S820
	TH01
	D3S1358
	D5S818
	CSF1PO
	D2S1338
	D19S433
	D1S1656
	D12S391
	D2S441
	D10S1248
	~~Penta E~~
	DYS391

	Locus
Section B (in order of preference)	TPOX
	D22S1045
	SE33
	~~Penta D~~

Current CODIS 13 loci in red font

Penta D and Penta E removed from consideration due to possibly not being available to all manufacturers
D.R. Hares (2012). Addendum to expanding the CODIS core loci in the United States, *Forensic Sci. Int. Genet.* doi:10.1016/j.fsigen.2012.01.003

Criteria for Acceptance of Additional Loci

D.R. Hares (2012) *Forensic Sci. Int. Genet.* 6(1):e52-e54

Considered only short tandem repeat (STR) loci due to need for compatibility to existing database of >10 million STR profiles

- **STR Loci**
  - **No known association to medical conditions or defects**
  - **Low mutation rate**
  - **High level of independence**
  - **High level of discrimination**
  - **Use by international forensic DNA community**
  - **Number of loci vs. discrimination factor**
  - **Compliance with Quality Assurance Standards (QAS)**
- **Kit performance**
  - **Balance between loci**
  - **Reliable**
  - **Reproducible**
  - **Sensitive**
  - **Quality results**
  - **Adaptable for use by NDIS laboratories (# of amplifications, ability of kit manufacturers to produce)**
  - **QAS compliant (documentation and availability of validation requirements)**

Determination of Additional CODIS Core Loci

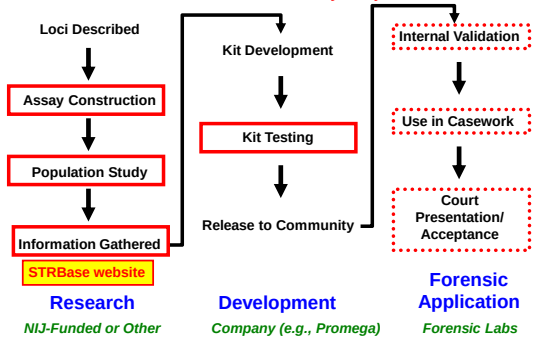
D.R. Hares (2012) Expanding the CODIS Core Loci in the United States. *Forensic Sci. Int. Genet.* 6: e52-e54
Addendum to expanding the CODIS core loci in the United States, Forensic Sci. Int. Genet. (2012) doi:10.1016/j.fsigen.2012.01.003

What	Why	Who/How	When
Form a Working Group (WG) to discuss initial selection	Establishes target goals	CODIS Core Loci Working Group with FBI Chair and 5 members; Web meetings	May 2010 - present
Announce proposed additional CODIS core loci	Sets desired target goals and informs manufacturers	WG Chair; Publish proposed listing of CODIS core loci	April 2011 online (published Jan 2012)
Ongoing Progress Reports	Provides updates for DNA community	WG Chair; Present updates on status of CODIS Core Loci project at meetings	2010-2012
Implementation Considerations & Strategy	Identify issues for implementation and timeline	WG	June 2011 - present
Manufacturers develop prototype kits	Creates tools to meet target goals	Manufacturers; Provide status reports to WG for timeline	2011-2012
Test and validate prototype kits	Examines if target goals can be met	Validation Laboratories; Follow QAS compliant validation plan	Beginning in 2012
Review and evaluate data from validation	Evaluates if desired performance is obtained	NIST, SWGDAM and FBI; Provide feedback, if any, to Manufacturers	In conjunction with and at the conclusion of validation
Selection of new CODIS core loci	Allows protocols to be established	FBI; seek input from DNA community and stakeholders; Notify Congress	After evaluation of validation data and kit production factors
Implementation of new CODIS core loci at the National DNA Index System	Enables target goals to be met	All NDIS-participating labs	~ 24 months after selection of new CODIS core loci

http://www.fbi.gov/about-us/lab/codis/planned-process-and-timeline-for-implementation-of-additional-codis-core-loci

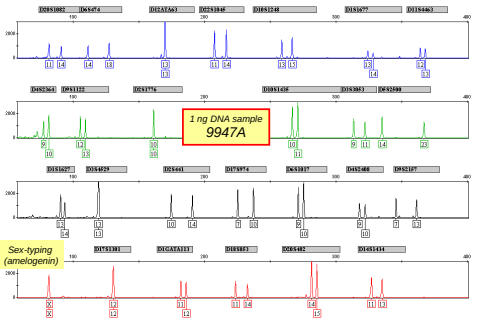
Steps in Adopting Genetic Markers

Role of the NIST Human Identity Project Team



NIST 26plex

Gender identification + 25 autosomal STR loci in a single amplification

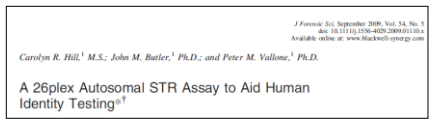


Hill, C.R., et al. (2009) A new STR 26plex assay for human identity testing. *J. Forensic Sci.* (in press)

Primer sequences and GM bins & panels available: **http://www.cstl.nist.gov/biotech/strbase/str26plex.htm**

NIST 26plex Demonstration

- Our group at NIST has demonstrated that 25 autosomal STRs and amelogenin (26plex with 52 PCR primers) can be co-amplified with sensitivities similar to commercial STR kits
  - Hill, C.R., Butler, J.M., Vallone, P.M. (2009) A 26plex autosomal STR assay to aid human identity testing. *J. Forensic Sci.* 54(5): 1008-1015.
  - See also http://www.cstl.nist.gov/biotech/strbase/str26plex.htm

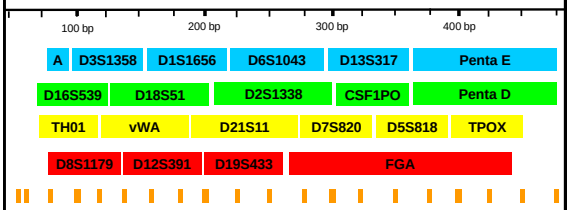


PowerPlex 21

- Promega STR kit was released in January 2012
  - NIST has been working with this kit since spring 2011 primarily for concordance testing and has permission from Promega to discuss results
- **Contains 20 autosomal STRs + amelogenin**
- **Enables examination of performance characteristics** similar to a future U.S. megaplex containing at least 20 loci

PowerPlex 21

(released by Promega in early 2012)

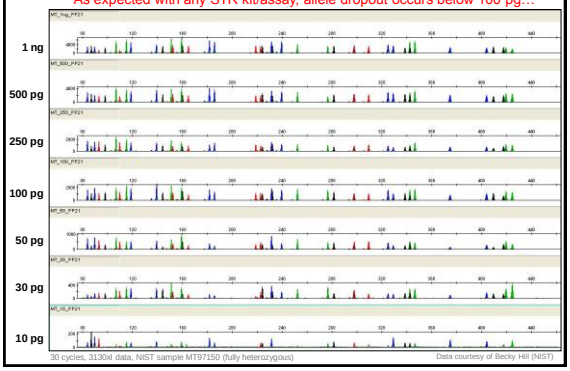


Promega 5-dye kit

13 CODIS STRs + amelogenin
Penta D & Penta E (PP16 loci)
D2S1338 & D19S433 (Identifiler loci)
D12S391 & D1S1656 (best new European loci)
D6S1043 (previously only used in China – ABI Sinofiler kit – highly polymorphic)

DNA Dilution Series with PowerPlex 21

As expected with any STR kit/assay, allele dropout occurs below 100 pg…



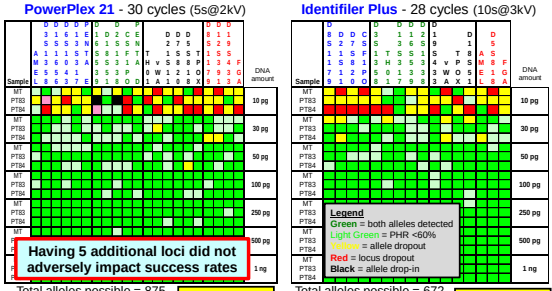
Data courtesy of Becky Hill (NIST)

Measurement of Allele Dropout and Extreme Peak Height Imbalance for 2 STR Kits

Three fully heterozygous (except PT83 at Penta D) pristine DNA samples were examined in a dilution series with PowerPlex 21 and Identifiler Plus. Results are ordered by amplicon size and dye color.

PowerPlex 21 - 30 cycles (5s@2kV)

Identifiler Plus - 28 cycles (10s@3kV)



Having 5 additional loci did not adversely impact success rates

Total alleles possible = 875
Total alleles present = 805
92% detected

Total alleles possible = 672
Total alleles present = 619
92% detected

Going Beyond the Core Competencies of Forensic DNA Testing

Core Competency

Standard STR Typing (DNA Profile)

Direct Matching (or Parentage)

Sufficient DNA quantity (ng)

Challenging kinship search questions → **Familial Searching Attempts (fishing for brothers or other relatives)**

Solution: Additional Markers (Y-chromosome, more STRs) and Multiple Reference Samples

Lower amounts of DNA being tested → **Touch DNA Attempts (poor quality, mixtures, low-level stochastic effects)**

Solution: Replicate Testing

Be very cautious when outside the box…

Figure 18.3 in forthcoming Butler (2009) Fundamentals of Forensic DNA Typing

CRIME & COURTS

July 7, 2010

Arrest Made in L.A. 'Grim Sleeper' Killings

Published July 07, 2010 | Associated Press

Print Email Share Comments (0) Recommend 42 Text Size

LOS ANGELES -- A one-time police mechanic was arrested and charged Wednesday in the serial killing of 10 people over 25 years after a DNA sample from his son was found to bear a close resemblance to DNA found on the victims.

Lonnie Franklin Jr., 57, was charged with 10 counts of murder, one count of attempted murder and special circumstance allegations of multiple murders that could make him eligible for the death penalty if convicted, District Attorney Steve Cooley said.

He is charged with 10 counts of murder and one count of attempted murder for a series of killings that date back to 1985.

Lonnie David Franklin Jr.

Victims of the Grim Sleeper

http://www.laweekly.com/2008-08-28/news/eleven-lives-stolen-and-one-lucky-survivor/

The Grim Sleeper's Victims

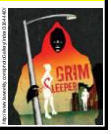
1) Debra Jackson (age 29) – August 10, 1985
2) Henrietta Wright (age 35) – August 12, 1986
3) Thomas Steele (age 36) – August 14, 1986
4) Barbara Ware (age 23) – January 10, 1987
5) Bernita Sparks (age 25) – April 15, 1987
6) Mary Lowe (age 26) – October 31, 1987
7) Lachrica Jefferson (age 22) - January 30, 1988
8) Monique Alexander (age 18) – September 11, 1988
9) Enietra Washington (raped but survived) – November 1988

10) Princess Berthomieux (age 14) – March 19, 2002
11) Valerie McCorvey (age 35) – July 11, 2003
12) Janecia Peters (age 25) – January 1, 2007

Ballistics on bullets recovered from the victim's bodies matched

DNA evidence recovered

Over a 13 year gap in detected crimes, hence the "Sleeper" nickname



http://blogs.laweekly.com/informer/crime/grim-sleeper-son-dna-trail-led/

Putative Relative Is Found

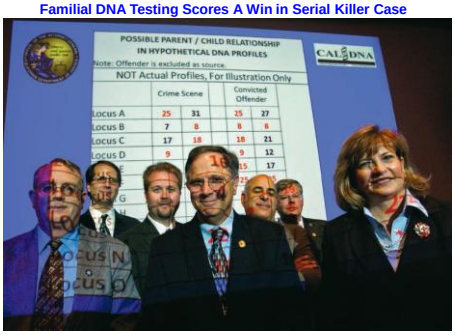
- June 30, 2010: Second familial search of the California database yielded one likely relative
- Database profile belonged to Christopher Franklin (31 years old)
  - Profile added to the database in 2009 after a felony weapons possession charge
- Grim Sleeper profile matched C. Franklin's profile with one allele at all 15 loci
- Both individuals shared the same Y-STR profile, indicating a possible paternal relationship

Identifying the Grim Sleeper

- Given that the murders spanned at least 25 years, the paternal relationship was likely father-son
- Undercover police shadowed C. Franklin's father, Lonnie David Franklin, Jr., who lived in the vicinity of the murders
- Police collected a DNA sample from Lonnie Franklin
  - **Direct match between L. Franklin and the Grim Sleeper**

California Familial DNA Search Team

Familial DNA Testing Scores A Win in Serial Killer Case



http://www.sciencemag.org/cgi/reprint/329/5989/262.pdf

Familial Searching in the U.S.

High-profile success in the Grim Sleeper case has led other states to consider familial searching

Experts say Texas might solve Twilight Serial Rapist cases with family DNA

http://www.examiner.com/law-enforcement-in-wichita-falls/experts-say-texas-might-solve-twilight-serial-rapist-cases-with-family-dna

DNA DATABASE

Milwaukee police on hunt for serial killer linked to 7 deaths

May 19, 2009 http://articles.cnn.com/2009-05-19/justice/wisconsin.serial.killer_1_dna-technology-dna-database-prostitutes?_s=PM:CRIME

Familial DNA hunt sought in East Coast rape case

http://www2.insidenova.com/news/2010/aug/04/familial_dna_hunt_sought_in_east_coast_rape_case-ar-428231/

March 21, 2011
Virginia announced familial searching capability

Wednesday December 1, 2010

Virginia could become 3rd state to use familial DNA searches

Some concerned practice could stigmatize those related to criminals

http://www.fairfaxtimes.com/cms/story.php?id=2600

Common PCR Thermal Cycling Times

Can we reduce PCR cycling times? What are the effects or limitations?

Thermal Cycling Times for Current STR Typing Kits

Year	Run on a 9700 thermal cycler	Hot start	Time per cycle	Cycles	Post soak	Total time
1997/98	Profiler Plus/Cofiler	11 min	3 min	28	60 min	2:52
1999	SGM Plus	11 min	3 min	28	45 min	2:53
2000	PowerPlex 16	12 min	1 min 45 s	32	30 min	3:00
2001	Identifiler	11 min	3 min	28	60 min	2:58
2003	PowerPlex Y	12 min	1 min 45 s	32	30 min	3:18
2004	Yfiler	11 min	3 min	30	80 min	2:45
2007	PowerPlex S5	2 min	4 min	30	45 min	3:21
2007	minifiler	11 min	3 min 20 s	30	45 min	3:16
2009	ESI 16, 17 ESX 16,17	2 min	4 min	30	45 min	3:22
2009	PowerPlex 16 HS	2 min	1 min 45 s	32	30 min	2:42
2009	NGM	11 min	3 min 20 s	29	10 min	2:33
2009	Identifiler Direct	11 min	3 min	26	25 min	2:34
2010	Idenfiler Plus	11 min	3 min 20 s	28	10 min	2:18
2011	PowerPlex 18D	2 min	1 min 10s	27	20 min	1:25

Thermal Cyclers

1. GeneAmp 9700 (Applied Biosystems)
2. Mastercycler Pro S (Eppendorf)
   - Peltier based
3. Rotor-Gene Q (Qiagen)
   - Air heated and cooled
4. SmartCycler (Cepheid)
   - Hot plates for heating, fans for cooling

Intended for real-time PCR

- Cycling for most STR kits is run in the
- '9600 emulation mode' (1°C/s)

PCR Thermal Cycling Profile

Identifiler STR kit

28 cycles of PCR

95°C	95°C	59°C	72°C	60°C
10 min	1 min	1 min	1 min	60 min

95°C	95°C	58°C	72°C	72°C
1 min	5 s	10 s	10 s	1 min

Sub 36 min run time

Maximum heating/cooling rate of ~2 to 6°C/s (cycler dependent)

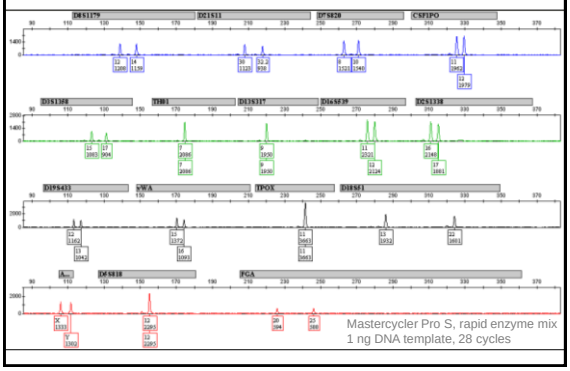
Rapid PCR Conditions

- 1 X Takara PCR mastermix, 1 U SpeedStar polymerase
  - *Premix Ex Taq*™ (Perfect Real Time)
- 10 μL total reaction in a thin walled tube (8-strip)
- 2 μL of Identifiler PCR primer mix
- ~1 ng of template DNA

- Utilize maximum ramp rate on thermal cyclers
  - GeneAmp 9700 = 1.6°C/s (36 min)
  - Rotor-Gene Q = 1.6°C/s (36 min)
  - SmartCycler = 5.8°C/s (20 min)
  - Mastercycler Pro S = 6.8°C/s (19 min)

Effective heating/cooling rates

Full Identifiler STR Profile with **19 min PCR**



Potential Applications with Rapid PCR Capabilities

- **Improve overall laboratory throughput**
  - Multiplex PCR amplification is already in many situations the longest part of the DNA analysis process (depending on DNA extraction and DNA quantitation methods)
  - With increased use of robotic sample preparation and expert system data analysis, bottleneck for sample processing will shift to time for PCR amplification…
- **Enable new potential DNA biometric applications** (because the overall DNA analysis process is faster)
  - Permit analysis of individuals at a point of interest such as an embassy, an airport, or a country border

A "Crystal Ball" to the Future?

- Better DNA extraction/recovery
- Continued use of miniSTRs
  - to improve success rates for recovery of information from compromised DNA evidence
- Replicate results for reproducibility
  - to improve reliability with low-template DNA testing

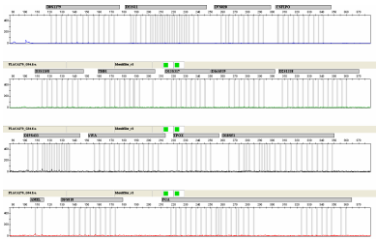
Highly degraded DNA

SNP genotyping in an extreme degradation case

Corpse half buried in a forest for ten years

- **Uncovered by a forest fire**
- **Calcinated remains**

Identifiler success 0%



Slide from Manuel Fondevila (NIST, USC)

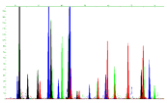
Highly degraded DNA

SNP genotyping in extreme degradation case

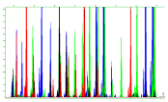
Corpse half buried in a forest for ten years

- **Uncovered by a forest fire**
- **Calcinated remains**

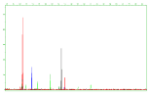
HID 52plex Auto 1: success 100%

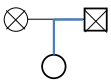


HID 52plex Auto 2 success 100%



MiniFiler success 30%





STRs +SNPs

P: - 99.993

Slide from Manuel Fondevila (NIST, USC)

Highly degraded DNA

Carlos Vullo's group from Argentina has published similar results with both SNPforID 52plex and IPATIMUP 38plex InDel reaction on common graves from Argentinian dictatorship period

Typing short amplicon binary polymorphisms: Supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains

C. Romanini[a], M.L. Catelli[a], A. Borosky[b], R. Pereira[d], M. Romero[a], M. Salado Puerto[c], C. Phillips[e], M. Fondevila[e], A. Freire[e], C. Santos[e], A. Carracedo[e], M.V. Lareu[e], L. Gusmao[d], C.M. Vullo[a,b,*]

Slide from Manuel Fondevila (NIST, USC)

Geographical Origin Prediction

- Lao O, van Duijn K, et al. (2006) Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. Am J Hum Genet 78: 680-90.
- Phillips, C., Salas, A., et al. (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. FSI: Genetics 1: 273-280.
- Halder, I., Shriver, M., et al. (2008) A Panel of Ancestry Informative Markers for Estimating Individual Biogeographical Ancestry and Admixture From Four Continents: Utility and Applications. Hum Mut 29: 648-658.
- Pereira R., Phillips C., et al. (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS One;7(1):e29684.

Slide from Manuel Fondevila (NIST, USC)

Phenotypic Trait Prediction

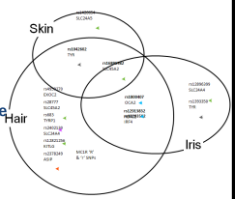
Traits of interest

- **Traits whose variation may be classified on discreet categories.**
- **Regulated by a relatively low number of genes.**
- **Fine example: Iris and hair pigmentation.**

Blue

Intermediate

Brown



Slide from Manuel Fondevila (NIST, USC)

Phenotypic trait prediction

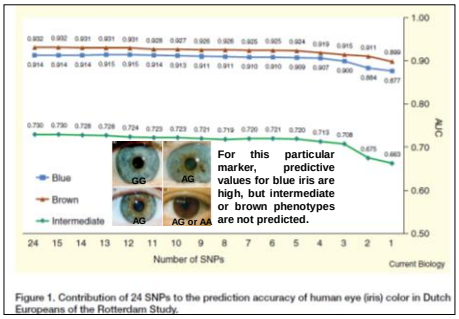


Figure 1. Contribution of 24 SNPs to the prediction accuracy of human eye (iris) color in Dutch Europeans of the Rotterdam Study.

Liu F., et al. (2009). Eye color and the prediction of complex phenotypes from genotypes, *Curr. Biol.* 19:R192–R193

Slide from Manuel Fondevila (NIST, USC)

Phenotypic trait prediction

- **Currently several research groups are working on the prediction of phenotypical traits by SNP typing.**
- **Best predictions have been achieved on iris pigmentation.**
- **However the achieved predictive values are still different for each variant. Research is not yet completed.**

- Branicki W, Kayser M et al. (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics;* DOI 10.1007/s00439-010-0939-8.
- Walsh S., et al. (2010) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci. Int. Genet. (Epub)
- Kayser M., Schneider P.M. (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci. Int. Genet. 3(3):154-61.
- Ruiz Y., C. Phillips et al.(2012) Further development of forensic eye color predictive tests. Forensic Sci. Int. Genet. (accepted for publication).

Slide from Manuel Fondevila (NIST, USC)

Next Generation Sequencing

- High throughput or ultra-high throughput sequencing
- Thousands or millions of sequencing reads in parallel
- DNA sequencing, RNA expression, clinical diagnostics, microbial forensics, …

1. Library preparation (genomic DNA or PCR amplicons)
2. Sequencing
3. Data analysis (assembly of reads)

Slide from Peter Vallone (NIST)

NGS Platforms

- Roche
  - 454 FLX
  - **454 GS Junio**
- PacificBio
  - Pacbio RS
- Illumina
  - GAIIx
  - HiSeq
  - HiScanSQ
  - **MiSeq**
- Life Tech
  - 5500 series
  - Ion torrent Proton
  - Ion torrent **PGM** (personal genome machine)




Slide from Peter Vallone (NIST)

Nature Biotechnology 2012

Performance comparison of benchtop high-throughput sequencing platforms

Nicholas J Loman[1], Raju V Misra[2], Timothy J Dallman[2], Chrystala Constantinidou[1], Saheer E Gharbia[2], John Wain[2,3] & Mark J Pallen[1]

1. 454 GS Junior (Roche)
2. MiSeq (Illumina)
3. Ion Torrent PGM (Life tech)

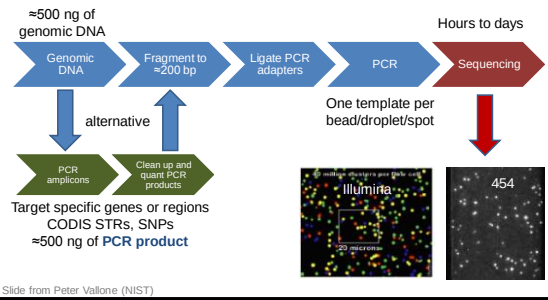
Table 1 Price comparison of benchtop instruments and sequencing runs

Platform	List price	Approximate cost per run	Minimum throughput (read length)	Run time	Cost/Mb	Mb/h
454 GS Junior	$108,000	$1,100	35 Mb (400 bases)	8 h	$31	4.4
Ion Torrent PGM						
(314 chip)	$80,490[a,b]	$225[c]	10 Mb (100 bases)	3 h	$22.5	3.3
(316 chip)		$425	100 Mb[2] (100 bases)	3 h	$4.25	33.3
(318 chip)		$625	1,000 Mb (100 bases)	3 h	$0.63	333.3
MiSeq	$125,000	$750	1,500 Mb (2 × 150 bases)	27 h	$0.5	55.5

Note pricing may vary between countries and/or sales territories. Instrument prices do not include service contracts. Sample prices do not include the cost of generating the initial fragmented genomic DNA library with adaptors (an additional cost of between $75–200 depending on method used). Cost per megabase assumes one sample and one sample sequencing kit per run. Unless stated, pricing information is from the online supplement of ref. 3. [a]Ion Torrent PGM pricing from Invitrogen US territory website (http://www.invitrogen.com/, accessed 21 February 2012). [b]Price includes Ion Torrent PGM, server, OneTouch and OneTouch ES sample automation systems. [c]Ion Torrent PGM prices include chip and sample preparation kit. [d]Configuration used in this study.

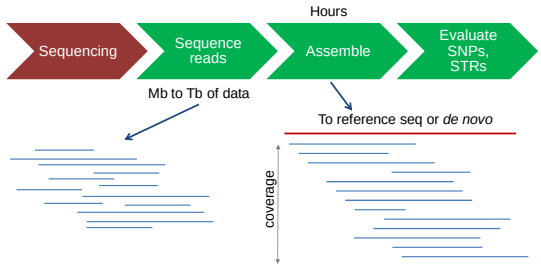
Slide from Peter Vallone (NIST)

Generalized NGS Workflow



Slide from Peter Vallone (NIST)

Generalized NGS Workflow



Slide from Peter Vallone (NIST)

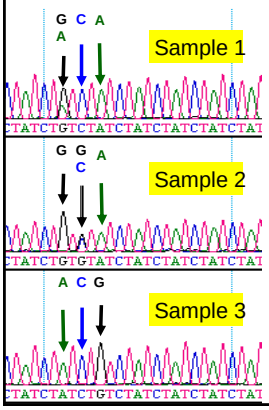
Next Generation Sequencing
Forensic Applications

- Going in depth into STR loci and beyond
  - STRs are useful for legacy (databases)
  - SNPs within STRs identify 'sub-alleles'
  - Millions of bases of sequence variants (SNPs)
- Opens up new human identity applications: biogeographical ancestry, externally visible traits, complex kinship, degraded samples, mixtures, other applications

Applications are currently being addressed by the forensic genetics community (*Kayser and deKnijff 2011)*

Slide from Peter Vallone (NIST)

SNPs within the D8S1179 repeat
All 3 samples '13,13' $[TCTA]_{13}$



Allele A - $[TCTA]_{13}$
Allele B - TCTA TCT**G** $[TCTA]_{11}$

Allele B - TCTA TCT**G** $[TCTA]_{11}$
Allele C - TCTA TCT**G** T**G**TA $[TCTA]_{10}$

There are **4** different '13' alleles in these 3 samples.

Allele D - $[TCTA]_2$ TCT**G** $[TCTA]_{10}$
Allele D - $[TCTA]_2$ TCT**G** $[TCTA]_{10}$

Data provided by Margaret Kline and Becky Hill

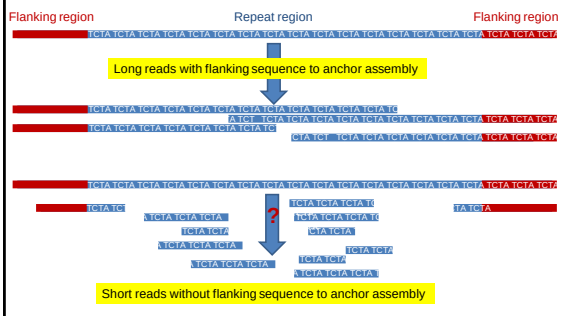
Specific issues with STRs

- Typically comprised of tetra nucleotide repeats
- Range 70 - 450+ bp regions
- Longer STRs can be difficult to assemble based on read length
- Illumina GAIIx (read length 150 bp)
  - Generated 1000-2500 bp amplicons (13 core loci)
  - Problems detecting D21S11 **32.2** and **34.2** alleles
  - Issues detecting D18S51
  - Custom informatics tools for assembling STRs

Bornman et al., 2012 Biotechniques Rapid Dispatch: 1-6

Slide from Peter Vallone (NIST)

Assembling STRs



Slide from Peter Vallone (NIST)

Next Generation Sequencing

- Challenges
  - Repeating sequences (STRs) and read lengths
  - Sample amount requirements (10 ng to 5 µg)
  - **Cost** and **time** per unit of information
  - Data analysis (storage, assembly, interpretation)
  - Policy, privacy, disease related markers
  - Validation
  - Standards/reference materials
    - Nomenclature
    - Accuracy of sequence information
    - Errors, platform and bioinformatics-based bias

Slide from Peter Vallone (NIST)

Next Generation Sequencing Workshop

- Interagency Workshop on the use of Next-Generation DNA Sequencing for Human Identification and Characterization (Jan 31 2012)
- Discussion of forensic applications of NGS (NIST, DoD, FBI, DHS) – materials can be found at:
  - http://www.nist.gov/mml/biochemical/genetics/ngs_hid_workshop.cfm
- We are in the process of looking at platforms to characterize forensic markers (mitochondrial, STRs, SNPs)
- Evaluate accuracy, reproducibility, identify initial requirements for a NGS forensic reference material

Slide from Peter Vallone (NIST)

Improvements in Forensic DNA Analysis

- Biology
  - Improved DNA extraction with automation
  - New capabilities for recovery of information from degraded DNA samples (e.g., miniSTRs)
- Technology
  - Parallel processing of DNA with capillary arrays
  - Expert systems for automated data interpretation
- Genetics
  - Ethnicity estimations (with STRs and/or SNPs)
  - Larger Y-STR and mtDNA population databases

Effective Training is Needed in All Areas

Some Thoughts on the Future…

- **PCR amplification**
  - Faster enzymes to enable rapid PCR
  - More robust enzymes and master mixes to overcome inhibition
- **Instrumentation**
  - More dye colors to aid higher levels of multiplexing
  - Rapid, integrated devices
  - Alternatives to capillary electrophoresis: PLEX-ID and NGS
- **Quantitative information**
  - qPCR and digital PCR
- **Marker systems**
  - Expanding sets of STR loci for growing DNA databases
  - Other marker systems: SNPs, InDels, X-STRs, RM Y-STRs
  - Body fluid identification with mRNA, miRNA, and DNA methylation
  - Phenotyping for external visible characteristics
  - Challenges with potential whole genome information
- **Data interpretation**
  - Probabilistic genotyping for low-level DNA and mixture interpretation
  - Probability of dropout

AAFS 2009 Topics Regarding Forensic DNA

From abstracts of presentations at AAFS meeting in Denver, CO (Feb 2009)

- Improved DNA extraction
- Predicting hair color and ancestry with SNPs
- X-chromosome STRs
- **Familial searching**
- Y-STRs and mixtures
- **Low level DNA samples**
- miniSTRs
- DNA screening assays
- Optimizing database labs
- Microfluidic biochip systems
- Use with property crimes
- Recovery from handguns
- DNA from IEDs
- Expert systems
- Automation with robotics
- DNA quantitation – qPCR
- PCR directly from blood
- mtDNA
- RNA
- Non-human DNA (dogs & cows)
- **Mixture interpretation**

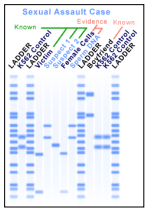
www.DNA.gov Website

Summary of NIJ-Funded Research

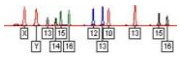
http://www.dna.gov/research/

The DNA Field Moves Forward…

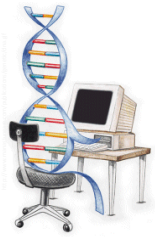
The Past | **The Present** | **The Future**



RFLP

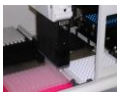


STRs



The Future

- More Robotics



- Expert Systems

- Animal & Plant DNA
- Physical Characteristics
- Ethnicity Estimation

The Future of Forensic DNA

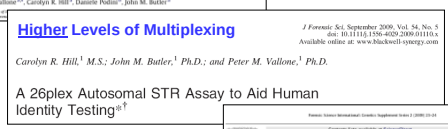
is Similar to the Olympic Motto of
"Swifter, Higher, Stronger"

Resources **Training** **Action**

Recent NIST Publications Demonstrating "Swifter, Higher, Stronger" DNA Analysis

Swifter PCR Amplification

Higher Levels of Multiplexing



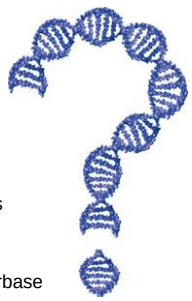
Stronger Powers of Discrimination

Acknowledgments

- Slides from Pete Vallone (NIST) and Manuel Fondevila (NIST, USC)
- Funding from National Institute of Justice and FBI Biometrics Center of Excellence for work performed within the NIST Applied Genetics Group

Thank you for your attention

Acknowledgments: A great team of scientists within our NIST Applied Genetics Group and funding from the National Institute of Justice and the FBI



Contact Information

John Butler
NIST Fellow
Group Leader of Applied Genetics
john.butler@nist.gov
301-975-4049

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

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