New Autosomal and Y-Chromosome STR Loci: Characterization and Potential Uses

John M. Butler,
Carolyn R. "Becky" Hill, Amy E. Decker,
Margaret C. Kline, Peter M. Vallone

National Institute of Standards and Technology

18th International Symposium on Human Identification
Hollywood, CA
October 4, 2007

Final version of this presentation available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

Questions to Be Addressed

- Why consider new STR loci?
- What has NIST accomplished with new autosomal STRs?
- Is there value in examining additional Y-STRs?
- Where can one learn more about these topics?

Why consider new STR loci?

Aren't the current core loci good enough?

Aren't the Current STR Loci Good Enough?

- Depends on the question being asked…
- For general forensic matching of evidence to suspect, the 13 CODIS STR loci are sufficient
- For other human identity/relationship testing questions, more autosomal or Y-STR loci can be beneficial or even necessary

How would additional STRs be useful?

How Would Additional STR Loci Be Useful?

- **Databases**: More loci to help resolve relatives in growing national DNA databases (UK went from 6 to 10 STRs in 1999; future Pan-European database will include >10 loci)
- **Casework**: Obtaining additional information with degraded DNA samples (miniSTRs); rapid screening of multiple crime scene samples
- **Identity/Relationship Testing**: Kinship analysis, parentage testing, complex criminal paternity, **missing persons/mass disasters**, immigration testing

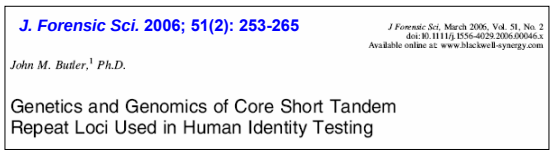
Call for More Loci
in Situations Involving Relatives

- **Missing Persons** and Disaster Victim Identification (kinship analysis)
- Immigration Testing (often limited references)
  - Recommendations for 25 STR loci
- Deficient Parentage Testing
  - often needed if only one parent and child are tested

Relationship testing labs are being pushed to answer more difficult genetic questions…and **we want to make sure the right tools are in place**

How are genetic loci introduced and adopted by the forensic/HID community?

History Plays a Role…



J. Forensic Sci. 2006; 51(2): 253-265

J Forensic Sci, March 2006, Vol. 51, No. 2
doi:10.1111/j.1556-4029.2006.00046.x
Available online at: www.blackwell-synergy.com

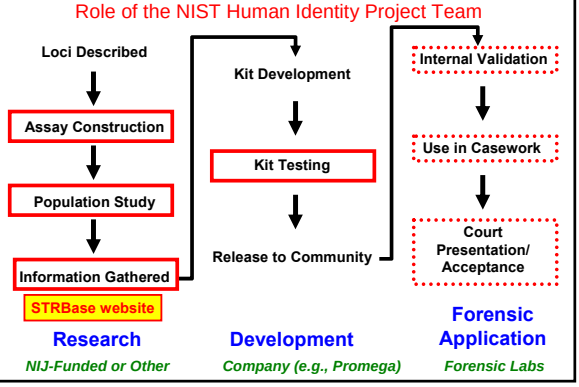
John M. Butler,[1] Ph.D.

Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing

- Only 17 STRs, which were available from Applied Biosystems and Promega in kit or prototype kit form (in 1996-97), were evaluated as part of the selection process for the 13 CODIS core loci
- Human Genome Project has increased knowledge…now thousands of STRs are known

Steps in Adopting Genetic Markers

Role of the NIST Human Identity Project Team



Justice for All Act of 2004

- **If additional loci are desired as core or supplementary loci on the national DNA database, the FBI must inform Congress six months prior to doing so…**
- "REPORT TO CONGRESS- If the Department of Justice plans to modify **or supplement the core genetic markers** needed for compatibility with the CODIS system, it shall notify the Judiciary Committee of the Senate and the Judiciary Committee of the House of Representatives in writing not later than 180 days before any change is made and explain the reasons for such change." (Section 203f)

What are important characteristics to consider in new loci?

Primary Characteristics in New STRs

- Genomic position
  - Adequate spacing from other (and current) loci to enable product rule use with autosomal markers
- Avoid known disease genes or linkage
  - To protect privacy concerns
- Polymorphic content (high heterozygosity)
  - More variable markers mean less can be used to reach desired rarity in full profile

Valuable Characteristics in New STRs

- Span/Range of observed alleles
  - Impacts electrophoretic real-estate
  - Tighter range makes differential amplification less likely
- Clean flanking region
  - To enable primer design near repeat (miniSTRs)
- Mutation rate known when trying to address multi-generational questions
- Provides benefit to haplotype resolution (Y-STRs)

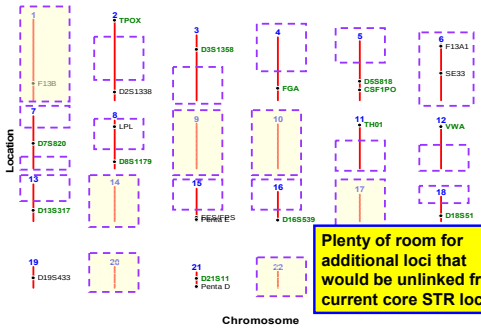
Steps We Use in Characterizing New Loci

- ✓ Select genetic loci
- ✓ Design primers – optimize multiplex assay
- ✓ Type population samples to examine variation
- ✓ Sequence alleles to establish nomenclature
- ✓ Develop bins and panels for genotyping
- ✓ Construct allelic ladders
- ✓ Evaluate RMP or ability to separate common types
- ✓ Perform mutation rate studies
- ✓ Perform concordance studies (when applicable)
- ✓ Calibrate genotypes with NIST SRM components
- ✓ Work with companies/collaborators
- ✓ Publish details on loci and assays

AUTOSOMAL STRs

Unused Chromosomal Locations

(relative to CODIS 13 STRs)



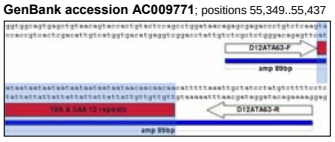
Butler, J.M. (2006) Genetics and genomics of core STR loci used in human identity testing. *J. Forensic Sci.* 51(2): 253-265.

Characterization of New Autosomal Loci

(miniSTR **D12ATA63**)

GenBank accession AC009771; positions 55,349..55,437

[FAM] – GAGCGAGACCCTGTCTCAAG
GAAAAGACATAGGATAGCAATTT



Chr 12 106.825 Mb
(12q23.3)

Trinucleotide
[TAA][CAA] repeat
76 -106 bp
Alleles 9 -19

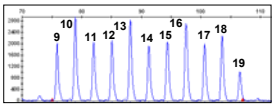
Allele	Caucasian (N = 260)	African Am (N = 259)	Hispanic (N = 140)
9	--	--	0.0036
10	0.0019	0.0154	0.0036
11	0.1385	0.1525	
12	0.2154	0.1	
13	0.0173		0.0286
14		0340	0.2214
15		0.0772	0.0714
16	**0.2981**	0.1004	**0.2643**
17	0.0981	0.0521	0.0679
18	0.0096	0.0058	0.0071
19	0.0019	0.0058	0.0036

~660 U.S. population samples

Heterozygosity Values

U.S. Caucasian	**0.842**
African American	**0.788**
U.S. Hispanic	**0.879**

D12ATA63 Allelic Ladder



To Appear in Jan 2008 Issue of *J. Forensic Sci.*

J Forensic Sci, January 2008, Vol. 53, No. 1
doi: 10.1111/j.1556-4029.2008.00595.x
Available online at: www.blackwell-synergy.com

Carolyn R. Hill, M.S.; Margaret C. Kline, M.S.; Michael D. Coble,† Ph.D.; and John M. Butler, Ph.D.

Characterization of 26 MiniSTR Loci for Improved Analysis of Degraded DNA Samples

- Primer sequences, GeneMapper bins and panels, genotypes on common samples, and allele frequency information **already available on STRBase**

How much DNA is required to obtain results with these new loci?

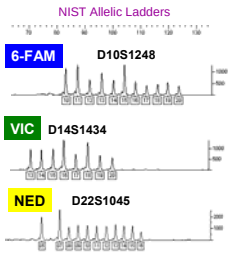
Assay Performance

- Our multiplex assays are designed to perform similarly to commercial kits
  - PCR Reaction (buffer, fluorescent dyes, volume)
  - PCR thermal cycling conditions
  - Work robustly on 0.5 to 1 ng of template DNA (or lower)
- Multiple miniplexes and a single megaplex developed to study **26 autosomal STRs**

Multiple Miniplexes

- **26 characterized loci** divided into nine 3plexes
- One locus per dye color
- Allelic ladders created
- **Amplicons <140 bp**
- miniSTRs
- Work with 100 pg DNA
- **For degraded samples**
  (bones in missing persons cases)

NC = Non-CODIS or non-core



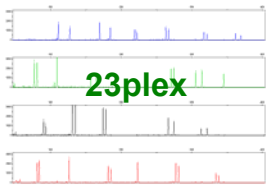
NC01 Loci

See Dixon et al. (2006) *Forensic Sci. Int.* 164: 33-44.

Single Megaplex

- So far **22 STRs and amelogenin** in single multiplex (Eventual goal to have all 26 loci)
- Multiple loci in four dye channels
- **Amplicons 70 to 400 bp**
  (No longer 'miniSTRs')
- Typically use 1 ng DNA
- **For reference samples**
  (a missing person's relatives)

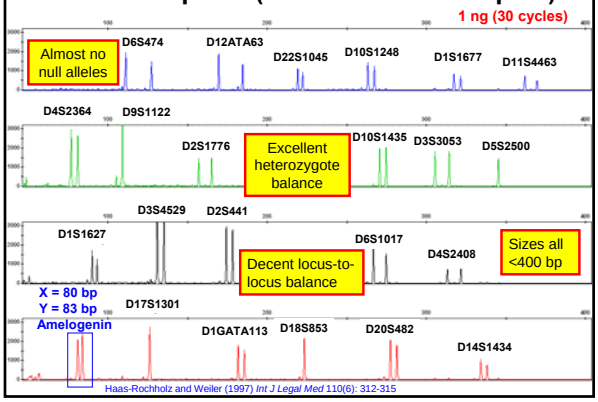
"Autoplex" or "miniMegaplex"



All loci unlinked from core (CODIS) STRs

How does this megaplex perform?

NIST "Autoplex" (Autosomal STR 23plex)



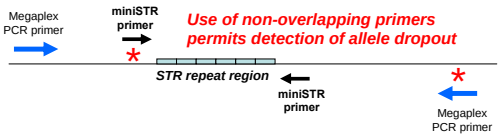
Evaluation of Autoplex (23plex)

- **660 U.S. population samples**
  - U.S. Caucasian, African American, Hispanic
  - **Concordance testing** compared to miniSTR results
- **790 father/son samples**
  - U.S. Caucasian, African American, Hispanic, Asian
  - **Mutation rate determination**
- 12 samples for **extended family testing**

>1450 samples examined so far
(multiple primer batches prepared)

Concordance Study to Check for Null Alleles

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



"Autoplex" vs miniSTRs

639 samples compared
Total types (639 x 22 loci): 14,058
28 types discordant (0.20%)*
99.80% concordance
*discordance not confirmed yet with sequencing

Identifiler vs MiniFiler

1308 samples compared
Total types (1308 x 8 loci): 10,464
27 types discordant (0.26%)
99.74% concordance
Hill et al. (2007) JFS 52(4): 870-873

Conclusions: (1) Our PCR primers have been well-designed and have almost no primer binding site mutations. (2) Roughly half of dropout is from megaplex primers – flanking regions near STR repeat do not appear to have a higher level of mutation

Mutation Rates Measured for New STRs

- **395 father/son pairs** (790 samples total)
- 22 STR loci examined
- 8690 allelic transfers
- Only **6 mutations** were observed in total

- **0.069%**
- (2-3 times less than typical 0.2% for common STRs)

Mutation rates generally track with heterozygosity (locus variability)

Locus	Mutation Rate
SE33	0.64%
FGA	0.28%
D18S51	0.22%
…	…
TPOX	0.01%

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

Conclusions: Mutation rates are lower than commonly used STRs likely due to selection of loci for miniSTR application with tighter allele ranges, more moderate heterozygosities, and more stable flanking regions.

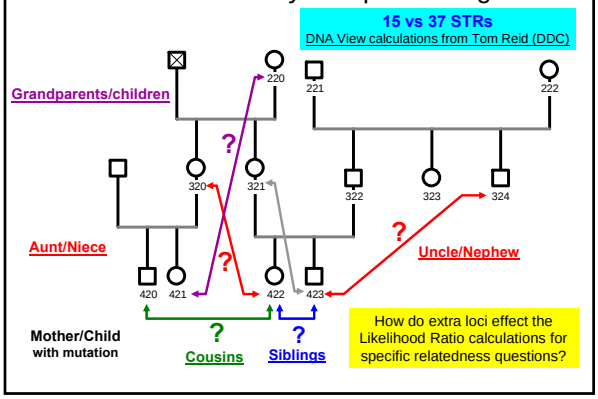
Population Data on New STRs

- **~660 samples** with three major U.S. populations on **all 26 autosomal STR loci**
  - Available on STRBase
    - **http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/Allele_Frequencies_for_26miniSTRs.pdf**
- **>3,000 samples tested world-wide** (Spain, Italy, Japan, Malaysia, Korea) on **first 6 loci** (NC01 & NC02)
  - **D2, D10, D22 now recommended European loci**

Gill *et al.* (2006) *Forensic Sci Int* **156(2):** 242-244

Can these new STRs help in missing persons cases or other forms of relationship testing?

Extended Family Sample Testing



Comparison of Likelihood Ratios

Relationship Examined	15 STRs (Identifiler, ID15)	ID15 + Autoplex 22 STRs = 37 loci (A37)
Mother/Child* (*with single mutation)	0.214	**5,200,000** Extra loci help…
Siblings	477	**113,000** Extra loci help…
Uncle/Nephew	824	**247,000** Extra loci help…
Cousins	0.45	**2.25**
Grandparents/ Grandchildren	0.53	**1.42**

Need lineage markers like Y-STRs

Y-CHROMOSOME STRs

Are Y-STRs more sensitive than autosomal STRs?

Are Y-STRs More Sensitive?

- Y-chromosome markers (kits) are **more selective as they offer male-specific amplification** but the loci (kits) themselves are NOT more sensitive.
- **Y-STRs have the same stochastic limitations with low-level DNA as autosomal markers**
- However, allele dropout of heterozygote sister alleles (false homozygosity) is not an issue with single-copy Y-STRs

What Y-STR loci and kits are commonly used today?

Review of Available Y-STR Loci and Data

Loci	Grouping (# Loci)	Available Data
DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385 a/b	Minimal Haplotype (9)	http://www.YHRD.org **50,867 haplotypes** (464 populations from around the world) NIJ-funded **US Database** at UCF: **14,015 haplotypes**
DYS438 DYS439	SWGDAM Core (11)	http://www.YHRD.org 23,981 haplotypes
DYS437	PowerPlex Y (12)	Promega website: 4004 haplotypes
DYS448 DYS456 DYS458 DYS635 GATA-H4	Yfiler (17)	Applied Biosystems website: **3561 haplotypes**

~400 additional Y-STRs currently known
Hanson & Ballantyne, *Legal Med* 2006;8(2):110-20

What is NIST doing with Y-STRs?

NIST Activities with Y-STRs

- SRM 2395 (Human Y Chromosome Standard)
  - http://www.cstl.nist.gov/biotech/strbase/SRM2395.htm
- Characterizing duplications and deletions
  - Butler et al. (2005) *J. Forensic Sci.* 50(4): 853-859
- Sequencing variant alleles
  - http://www.cstl.nist.gov/biotech/strbase/STRseq.htm
- Supplied ~20% of Yfiler 3561 database
  - http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm
- Concordance studies between Yfiler and NIST 20plex and 11plex assays

22 publications since 2002 on NIST Y-chromosome work

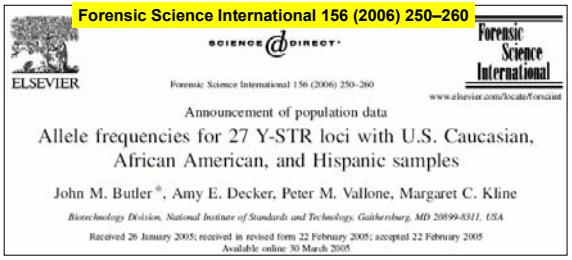
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

NIST Work with Additional Y-STR Loci

- Studies of Locus Variation
  - **37 Y-STRs** have been examined in all 665 NIST U.S. population samples and **92 Y-STRs** in a subset (32 C, 32 AA, 31 H) using previously published primers and 3-5plexes
- Analysis of Mutation Rates
  - **389 father/son pairs** with **17 Yfiler loci**
- Further characterization of SRM 2395 components
  - To enable calibration with additional Y-STRs
- Defining allele nomenclature on **144 Y-STRs**
  - To aid on-going genetic genealogy work

Publication with Additional Y-STR Loci

Forensic Science International 156 (2006) 250–260


ELSEVIER

SCIENCE DIRECT

Forensic Science International

Forensic Science International 156 (2006) 250–260

Announcement of population data

Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples

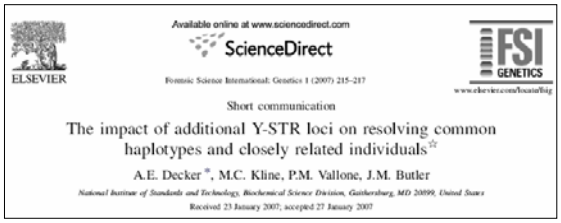
John M. Butler *, Amy E. Decker, Peter M. Vallone, Margaret C. Kline

Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8311, USA

Received 26 January 2005; received in revised form 22 February 2005; accepted 22 February 2005
Available online 30 March 2005

Full Haplotypes Available on http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm

Are there advantages to typing additional loci beyond the PowerPlex Y 12 or the Yfiler 17 Y-STRs?


ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Forensic Science International: Genetics 1 (2007) 215–217

FSI GENETICS

Short communication

The impact of additional Y-STR loci on resolving common haplotypes and closely related individuals☆

A.E. Decker *, M.C. Kline, P.M. Vallone, J.M. Butler

National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, MD 20899, United States

Received 23 January 2007; accepted 27 January 2007

Full 37 locus haplotypes available on STRBase:
http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/HispanicsHaplotype37.pdf
http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/CaucasiansHaplotype37.pdf
http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/AfricanAmericansHaplotype37.pdf

# times haplotype observed	12 loci PPY	17 loci Yfiler	37 loci ALL 37
1	505	626	652
2	34	12	2
3	14	2	.
4	3	.	.
5	2	.	.
6	.	.	.
7	1	.	.
8	.	.	.
9	.	.	.
10	.	.	.
11	.	.	.
12	1	.	.
HD	0.99906	0.99992	0.99999
DC	0.85366	0.97561	0.99695
# HT	560	640	654

NIST U.S. Pop (C, AA, H)
Total # samples: **656**

unique
unresolved haplotypes

One set of three unseparated Yfiler types will be examined next

Most common type is observed 12 times

With the 17 loci in Yfiler across the 656 samples, there are 626 unique haplotypes, 12 haplotypes that were observed twice and 2 haplotypes that were observed three times

95% (626/656) Yfiler haplotypes were unique

Subdividing Unresolved Yfiler Haplotypes (1)

Most Common Type

Sample Info	DYS 19	DYS 385a/b	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	H4
MT97185	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12
ZT79333	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12
TT51702	14	11,14	13	29	24	11	13	13	12	12	15	19	16	17	23	12

Locus	MT97185	ZT79333	TT51702
~~DYS444~~	~~12~~	~~12~~	~~12~~
~~DYS446~~	~~13~~	~~13~~	~~13~~
DYS449	30	30	31
DYS463	24	24	23
~~DYS485~~	~~15~~	~~15~~	~~15~~
~~DYS495~~	~~16~~	~~16~~	~~16~~
~~DYS505~~	~~12~~	~~12~~	~~12~~
~~DYS506~~	~~11~~	~~11~~	~~11~~
DYS520	21	22	21
DYS522	10	12	11

Locus	MT97185	ZT79333	TT51702
DYS532	14	14	13
DYS533	13	12	13
~~DYS534~~	~~15~~	~~15~~	~~15~~
~~DYS540~~	~~12~~	~~12~~	~~12~~
~~DYS556~~	~~11~~	~~11~~	~~11~~
DYS557	15	17	17
DYS570	16	17	17
DYS576	17	20	18
DYS594	9	10	10
DYS643	10	11	10

Either DYS522 or DYS576 will fully resolve all three of these samples

Lessons Learned from NIST Data Set

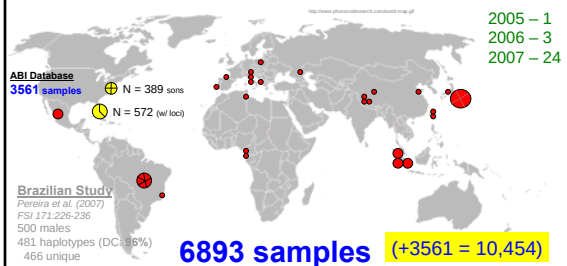
- Some Y-STRs that are more useful than others in sub-dividing common haplotypes (e.g., DYS576)
- You don't gain much by typing additional Y-STRs (most unresolved types only occur twice)
- 95% of 17 locus Yfiler haplotypes are unique

What additional population data exists with Yfiler?

And how does it compare to our NIST data?

Sources of Yfiler Worldwide Population Data

28 published population studies with Yfiler data



6893 samples (+3561 = 10,454)
6514 haplotypes (discrimination capacity 94.5%)
6257 unique haplotypes **(96.0% unique)**

Summaries of Recent Worldwide Yfiler Data

- **10,454 Yfiler profiles** now available
  - 3561 current Yfiler database + 6893 published data
- ~95% of the time a complete 17 locus Yfiler profile will be unique
- However, just like mtDNA, common types do exist so many of the remaining Yfiler haplotypes are shared (present in multiple individuals)

What do mutation rates look like for Y-STR markers?

Y-STR Mutation Rates Measured at NIST

- **389 father/son sample pairs**
  - U.S. Caucasians, African Americans, Hispanics and Asians
- **17 Y-STR loci** in the Yfiler kit
- **24 differences** between father and son
  - 13 mutations resulted in the gain of a repeat in the son
  - 11 resulted in a loss of a repeat
- All single step repeat mutations
  - except a two repeat loss at Y-GATA-H4
- **2 sample pairs were found to have two mutations**
  - African American pair: mutations at DYS458 and DYS635
  - Asian pair: mutations at DYS439 and Y-GATA-H4
- Also observed 4 duplications, 1 triplication, and 4 deletions that were seen in both father and son

Decker, A.E., Kline, M.C., Redman, J.W., Reid, T.M., Butler, J.M. (2007) Analysis of mutations in father-son pairs with 17 Y-STR loci. *FSI Genetics (in press)*

What are the "best" additional Y-STRs?

The Next Best Y-STRs…Beyond the Kits

- Decker *et al*. (2007) FSI Genetics 1:215-217
  - **DYS449**, DYS505, DYS522, DYS532, DYS534, **DYS570**, **DYS576**
- Hanson and Ballantyne (2007) *PLoS ONE* 8:e688
  - DYS444, DYS446, **DYS449**, DYS459a/b, **DYS481**, DYS508, DYS522, DYS527a/b, DYS549, DYS552, **DYS570**, **DYS576**, DYS607, DYS627
- Rodig *et al.* (2007) *Forensic Sci Int* (in press)
  - DYS447, **DYS449**, **DYS481**, **DYS570**, **DYS576**

These loci are useful for subdividing common types and lineage testing…

To Summarize…

Autosomal STRs

- **26 unlinked loci** have been characterized and we have developed multiple miniplexes and a megaplex (23plex)
- Additional loci show value with relationship testing
- **NIST SRM 2391b** will include information on additional autosomal STR loci

Y-Chromosome STRs

- Studies at NIST and worldwide show ~95% of observed 17 locus Yfiler profiles are unique
- Additional loci can help with common types
- **NIST SRM 2395** will include information on additional Y-STR loci

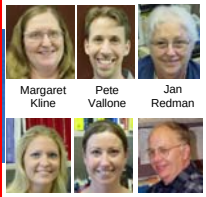
http://www.cstl.nist.gov/biotech/strbase/newSTRs.htm

Thank you for your attention…

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

Questions?

http://www.cstl.nist.gov/biotech/strbase
john.butler@nist.gov
301-975-4049



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

Collaborators

Mike Coble (now AFDIL) – early miniSTR work

Tom Reid (DDC) - father/son samples