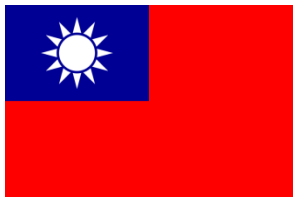


CIB Forensic Science Center
Training Seminar (Taipei, Taiwan)
June 6-7, 2012

NIST
National Institute of Standards and Technology

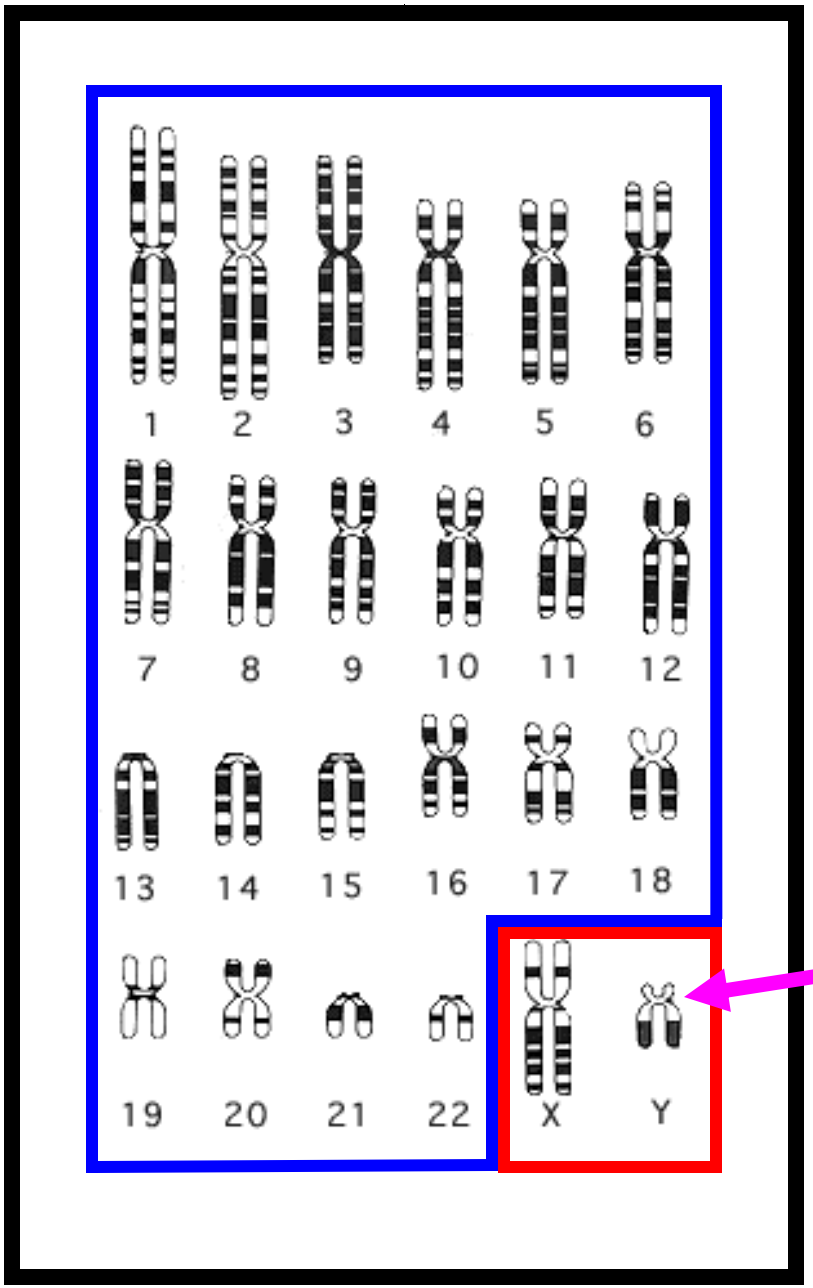
Lineage Markers:
Y-STRs, mtDNA, and X-STRs

John M. Butler

NIST
National Institute of Standards and Technology

NIST Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland

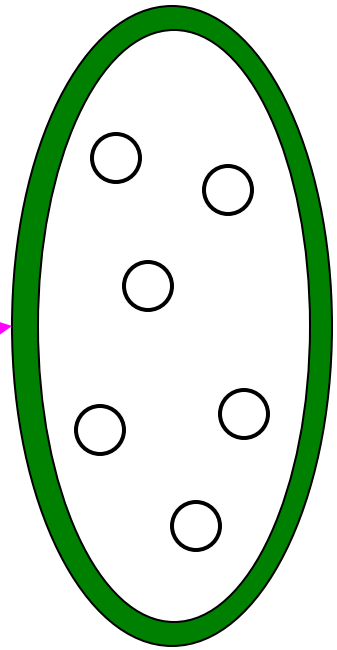
Cell Nucleus – 3.2 billion bp



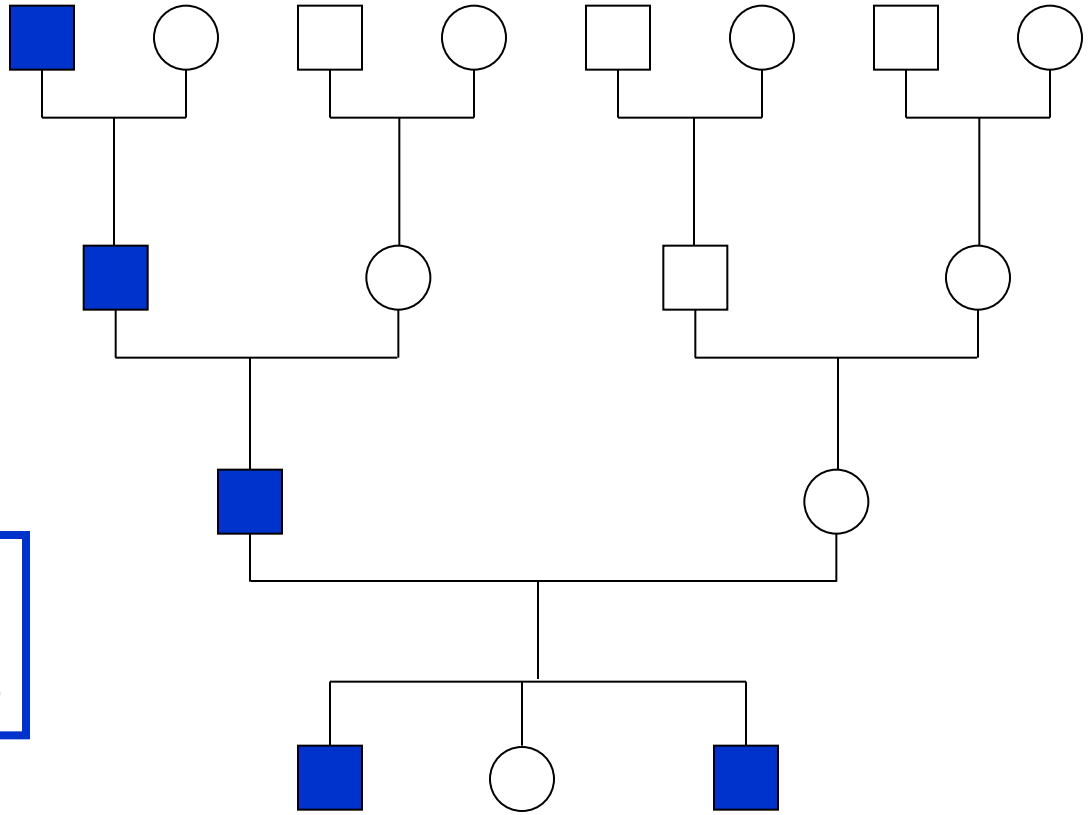
Autosomes – 22 pairs – 2 copies per cell

Sex Chromosomes (XX or XY)

mitochondria – in cell cytoplasm
100s of mtDNA copies per cell



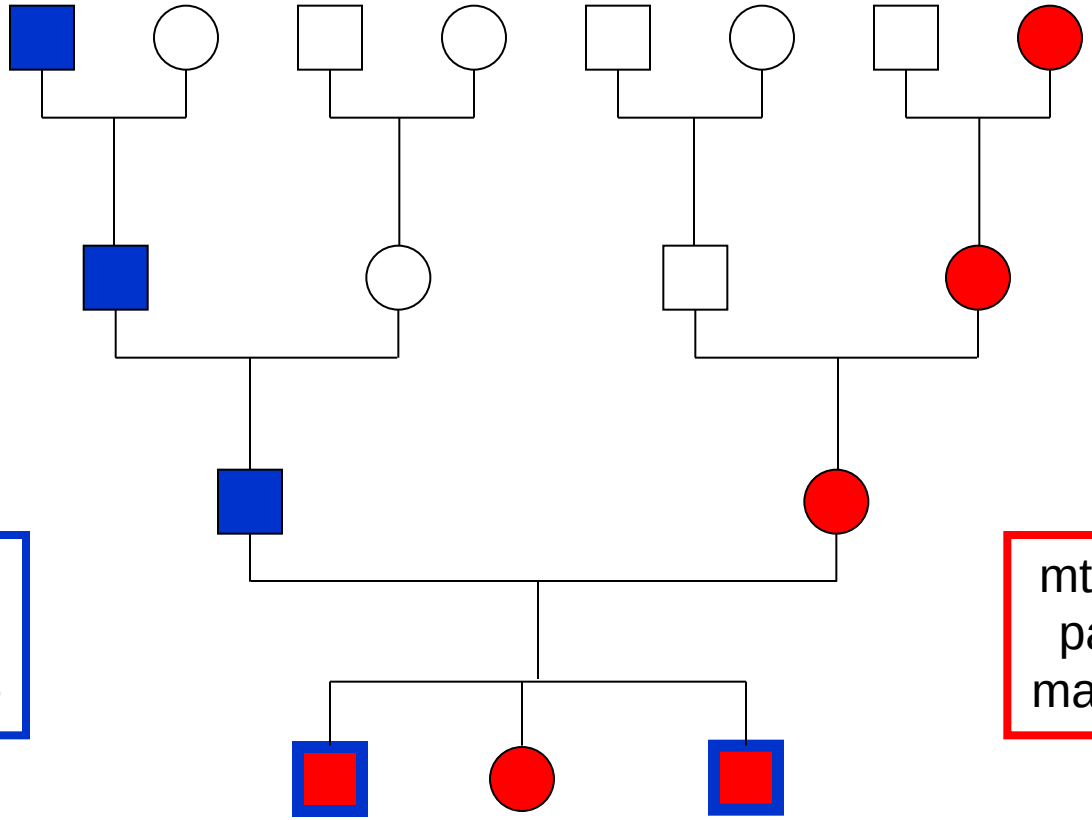
Lineage Markers: Y-Chromosome



Y chromosome passed along paternal lineage

Autosomal DNA
1/8 from Great-grandparents

Lineage Markers: mtDNA



Y chromosome passed along paternal lineage

mtDNA genome passed along maternal lineage

Autosomal DNA
1/8 from Great-grandparents

Different Inheritance Patterns

TABLE 15.1 Specific Relationships and the Probability of Transmitting Genetic Information (Barring Mutation). Some of the ChrY Information is Not Applicable (N/A) as Women Do Not Have a Y-Chromosome.

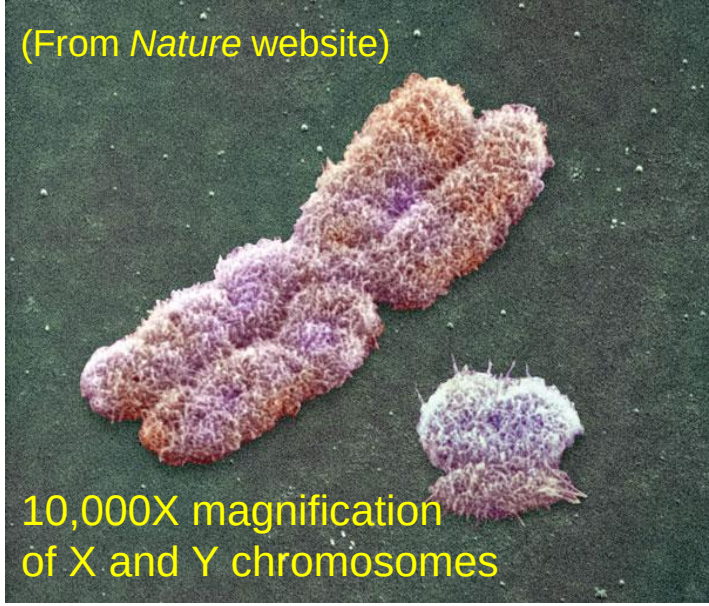
Inheritance	Autosomal Markers	ChrY Markers	mtDNA	ChrX Markers
Mother → Son	50%	N/A	100%	100%
Mother → Daughter	50%	N/A	100%	50%
Father → Son	50%	100%	0%	0%
Father → Daughter	50%	0%	0%	100%
Paternal Grandmother → Granddaughter	25%	N/A	0%	100%
Maternal Grandmother → Granddaughter	25%	N/A	100%	25%
Paternal Grandfather → Grandson	25%	100%	0%	0%

Butler, J.M. (2012) *Advanced Topics in Forensic DNA Typing: Methodology*, p. 457

THE HUMAN Y CHROMOSOME: AN EVOLUTIONARY MARKER COMES OF AGE

Mark A. Jobling & Chris Tyler-Smith

Nature Reviews Genetics (2003) 4, 598-612



Abstract

- **Until recently, the Y chromosome seemed to fulfill the role of juvenile delinquent among human chromosomes — rich in junk, poor in useful attributes, reluctant to socialize with its neighbors and with an inescapable tendency to degenerate. The availability of the near-complete chromosome sequence, plus many new polymorphisms, a highly resolved phylogeny and insights into its mutation processes, now provide new avenues for investigating human evolution. Y-chromosome research is growing up.**

What has happened in the past decade…

- **Selection of core Y-STR loci** (SWGDAM Jan 2003)
- "Full" Y-chromosome sequence became available in June 2003; over 400 Y-STR loci identified (only ~20 in 2000)
- **Commercial Y-STR kits released**
  - ~~Y-PLEX 6,5,12 (2001-03)~~, **PowerPlex Y** (9/03), **Yfiler** (12/04), **PPY23** (6/12)
- Many population studies performed and online databases generated with thousands of Y-STR haplotypes
- Forensic casework demonstrations showing value of Y-STR testing along with court acceptance
- Some renewed interest in Y-STRs to aid familial searching

Y-STR Information

- Why the Y?
- Y-STR Loci & Kits
- Y-STR Databases
- Y-STR Stats & Interpretation Issues
- Genetic Genealogy & Familial Searching

Value of Y-Chromosome Markers

J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition; Table 9.1

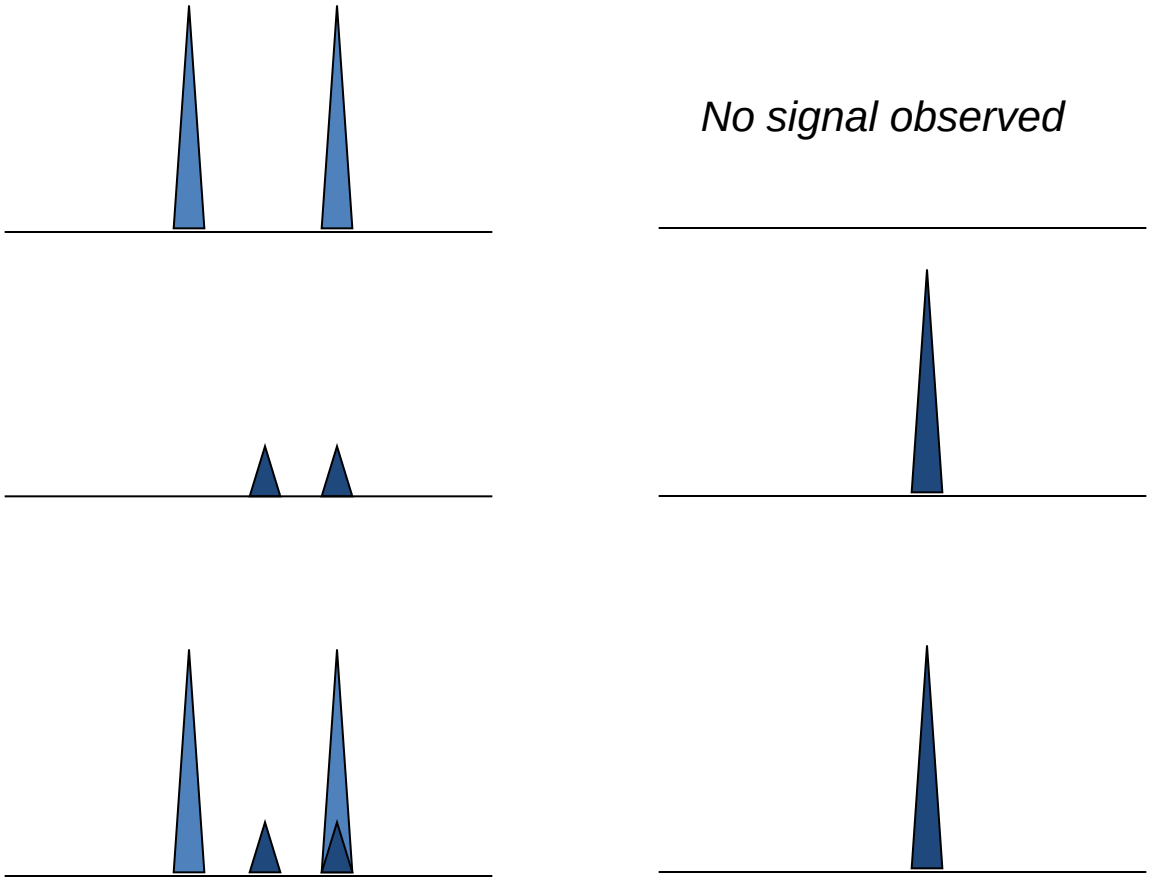
Application	Advantage
Forensic casework on sexual assault evidence	**Male-specific amplification** (can avoid differential extraction to separate sperm and epithelial cells)
Paternity testing	Male children can be tied to fathers in motherless paternity cases
Missing persons investigations	**Patrilineal male relatives may be used for reference samples**
Human migration and evolutionary studies	Lack of recombination enables comparison of male individuals separated by large periods of time
Historical and genealogical research	Surnames usually retained by males; can make links where paper trail is limited

Y-STRs can permit simplification of male DNA identification in sexual assault cases

Female Victim DNA Profile

Male Perpetrator DNA Profile

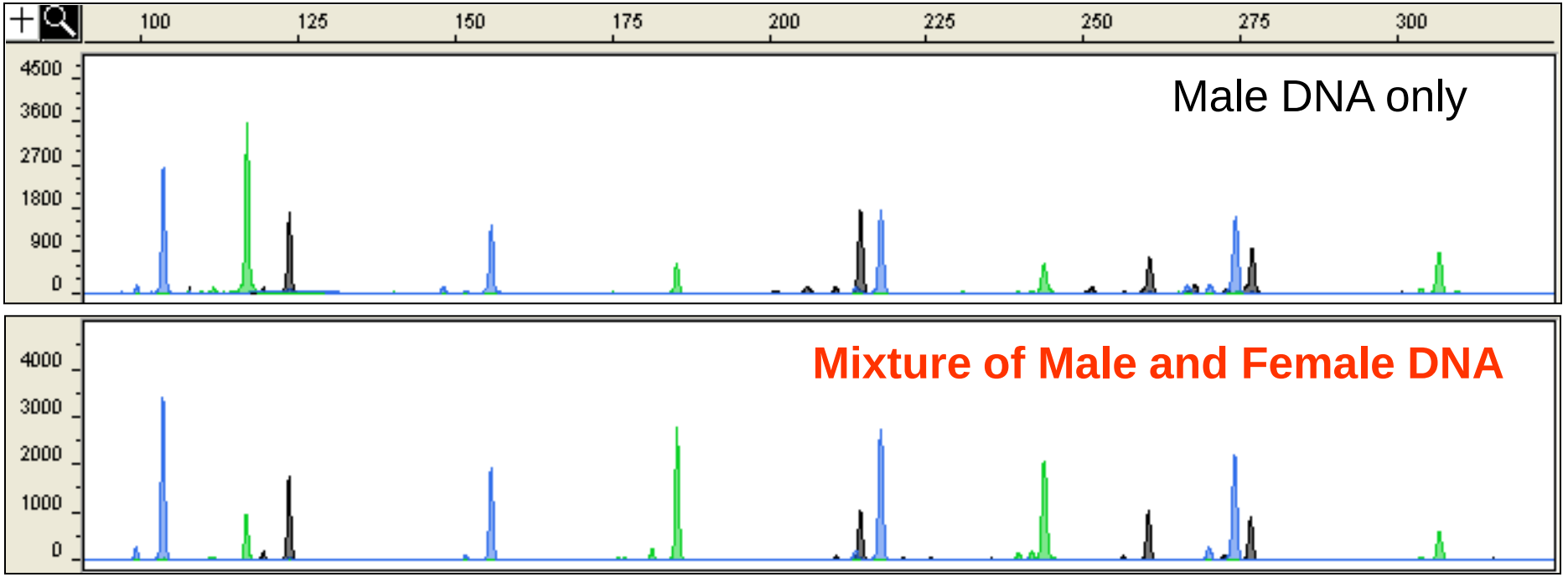
DNA Profile from Crime Scene



Autosomal STR Profile

Y-Chromosome STR Profile

Y-STRs Identify the Male Component even with Excess Female DNA



800X female DNA

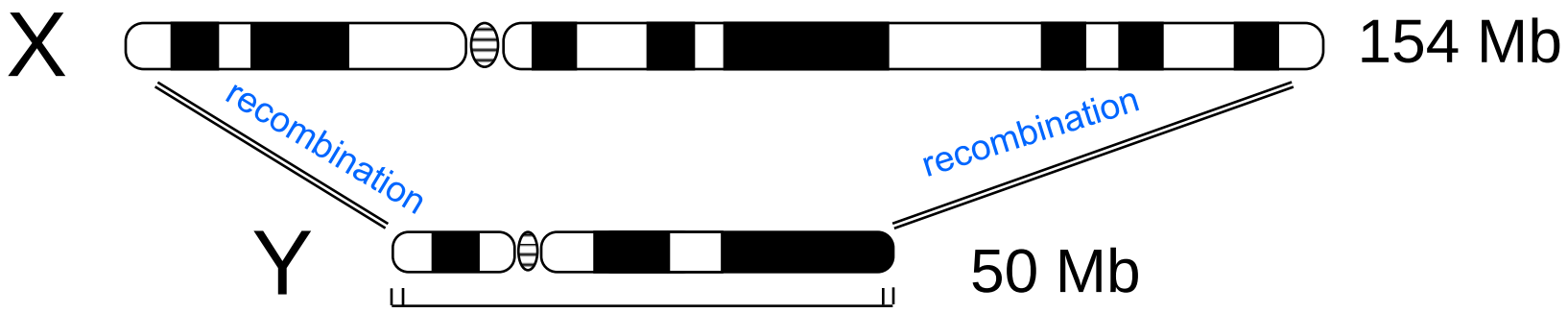
Disadvantages of the Y-Chromosome

- Loci are not independent of one another and therefore rare random match probabilities cannot be generated with the product rule; must use haplotypes (combination of alleles observed at all tested loci)

- **Paternal lineages possess the same Y-STR haplotype** (barring mutation) and thus fathers, sons, brothers, uncles, and paternal cousins cannot be distinguished from one another

- Not as informative as autosomal STR results
  - **More like addition (10 + 10 + 10 = 30) than multiplication (10 x 10 x 10 = 1,000)**

X- and Y-Chromosomes Recombine at Their Tips



Male-specific region of the Y (MSY)

Focus in Forensics is on the NRY

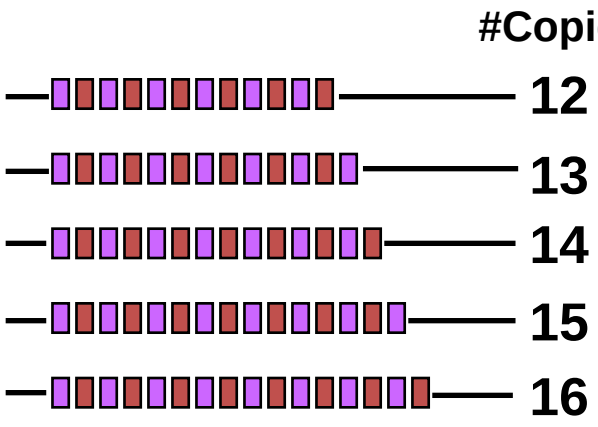
PCR primers for Y-STRs need to be carefully designed to avoid X-chromosome homology

Various Types of Genetic Markers on the Human Y-Chromosome

Y-STRs

Short Tandem Repeats

– GATAGATAGATAGATA –



Multi-state characters

Quickly evolving (**2 x 10^{-3}**/gen)

High resolution haplotypes

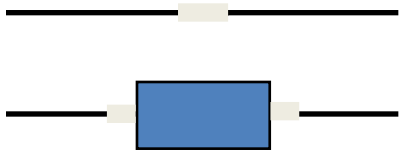
Y-SNPs

Single Nucleotide Polymorphisms

–CG**A**TG–

–CG**G**TG–

Insertion/deletions (indels)



Binary characters

Slowly evolving (**~10^{-8}**/gen)

Low resolution haplogroups

Slide from Alan Redd (University of Arizona) presentation at Promega Oct 2002

Y-STR Loci & Kits

History of Y-STR Marker Discovery

"Extended Haplotype"

1992 - DYS19 (Roewer et al.)

1994 - YCAI a/b, YCAII a/b, YCAIII a/b, DXYS156 (Mathias et al.)

1996 - DYS389I/II, DYS390, DYS391, DYS392, DYS393 (Roewer et al.)

1996 - DYF371, DYS425, DYS426 (Jobling et al.)

1997 - DYS288, DYS388 (Kayser et al.)

"Minimal Haplotype"

1998 - DYS385 a/b (Schneider et al.)

1999 - A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4 (White et al.)

2000 - DYS434, DYS435, DYS436, DYS437, DYS438, DYS439 (Ayub et al.)

2000 - G09411 (DYS462), G10123 (de Knijff unpublished)

U.S. Haplotype

2001 - DYS441, DYS442 (Iida et al.)

2002 - DYS443, DYS444, DYS445 (Iida et al.); DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d (Redd et al.)

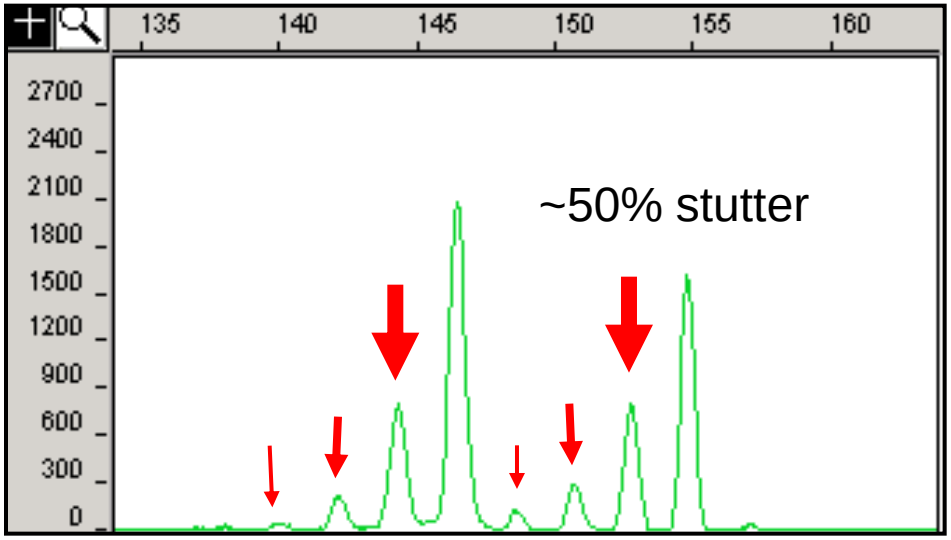
2002 – DYS468-DYS596 (**129 new Y STRs**; Manfred Kayser GDB entries)

2003 – DYS597-DYS645 (**50 new Y STRs**; Manfred Kayser GDB entries)

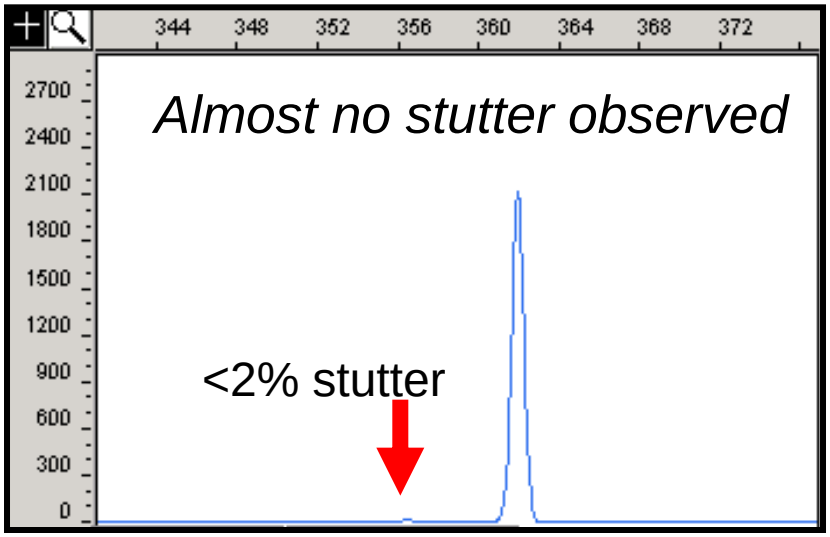
From J.M. Butler (2003) Recent developments in Y-STR and Y-SNP analysis. *Forensic Sci. Rev*. 15:91-111

STR Markers with Low Stutter Products Benefit Forensic Analysis where Mixtures might be Present

YCAII a/b (dinucleotide repeat)



DYS448 (hexanucleotide repeat)



[CA]

11-24 repeats

YCC gene diversity 0.908

(Redd et al. 2002)

[AGAGAT]

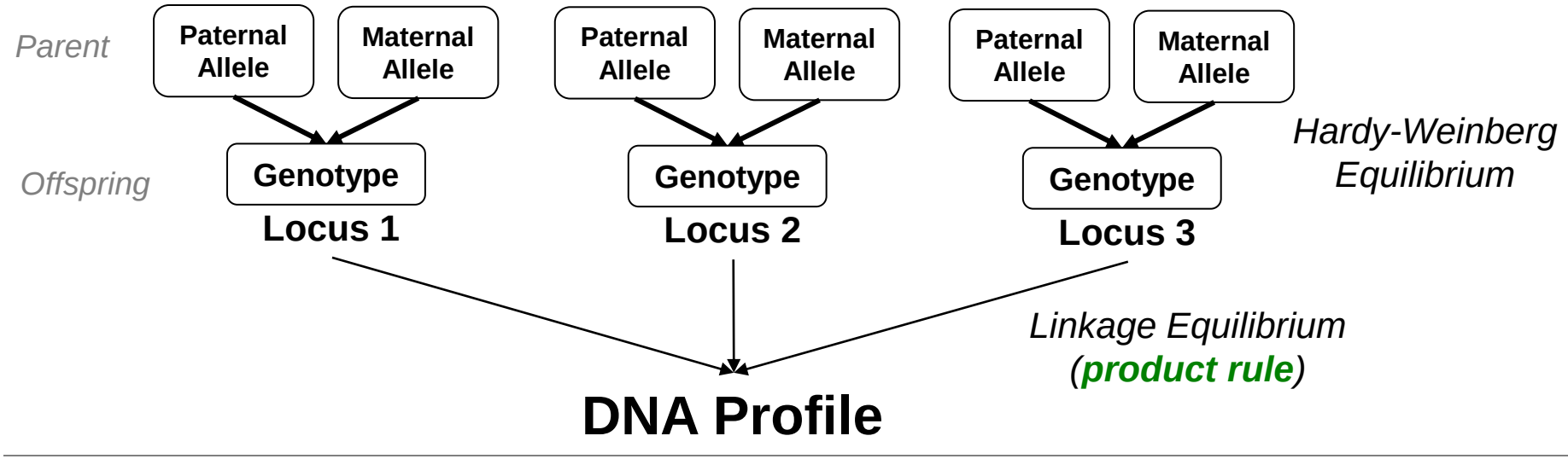
20-26 repeats

YCC gene diversity 0.782

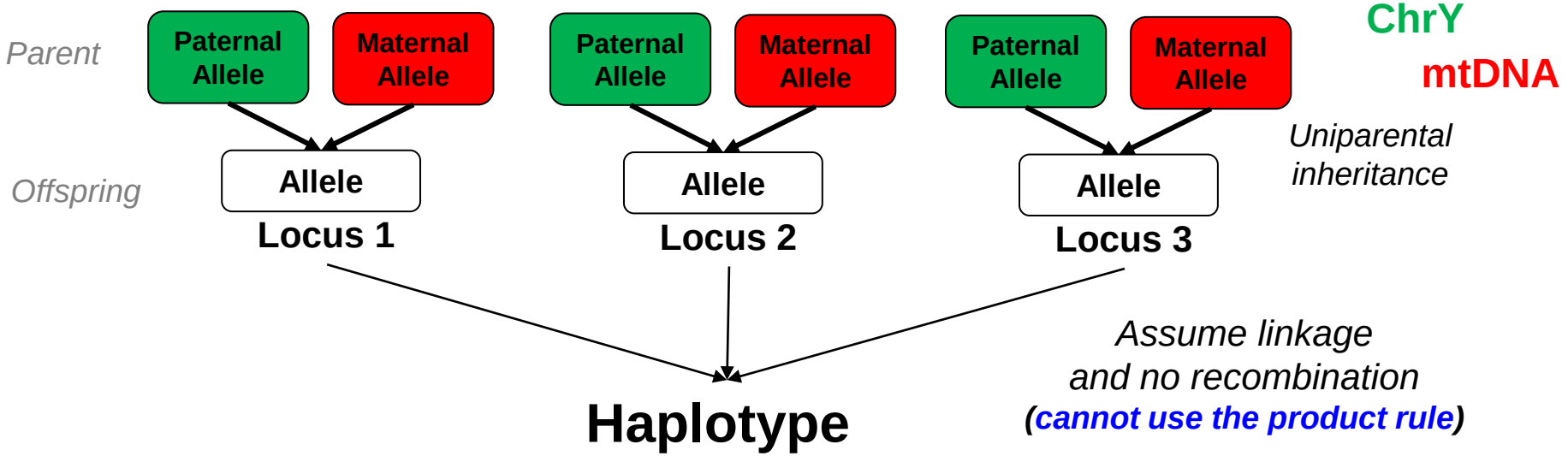
(Redd et al. 2002)

Differences between Autosomal and Lineage Markers

Autosomal Markers



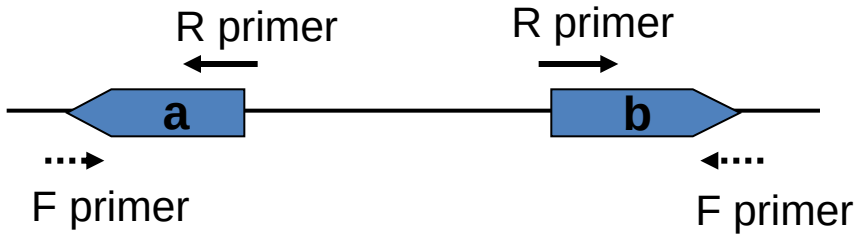
Lineage Markers



Single Primer Sets Produce Multiple PCR Products

(a) DYS385 a/b

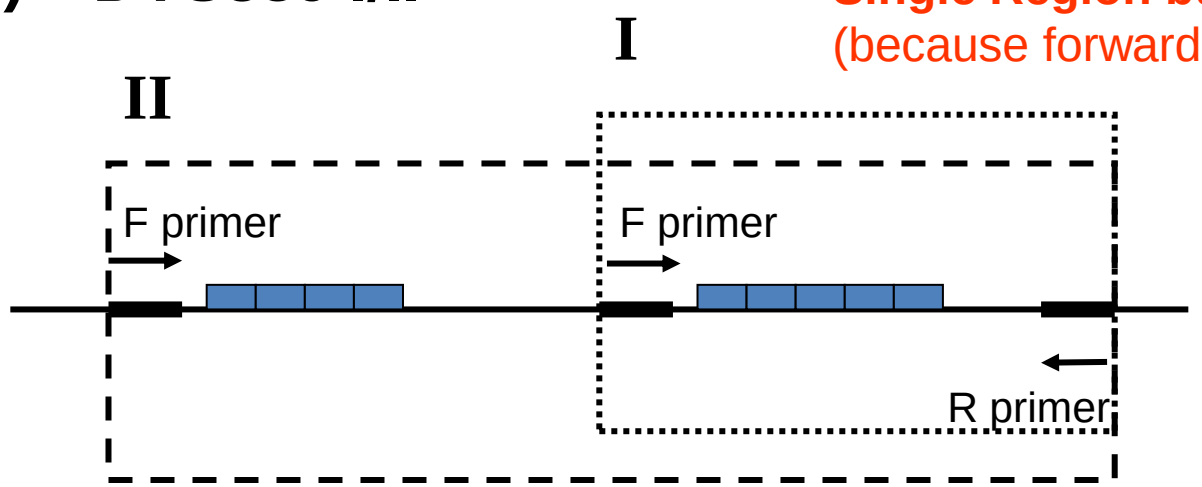
Multi-Copy (Duplicated) Marker

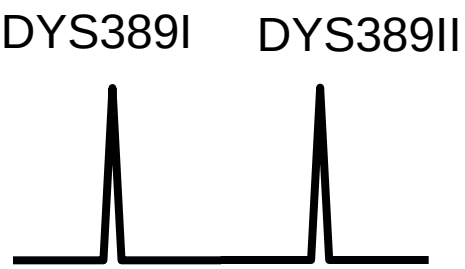


Duplicated regions are 40,775 bp apart and facing away from each other

(b) DYS389 I/II

Single Region but Two PCR Products
(because forward primers bind twice)





17 PCR products
15 primer sets

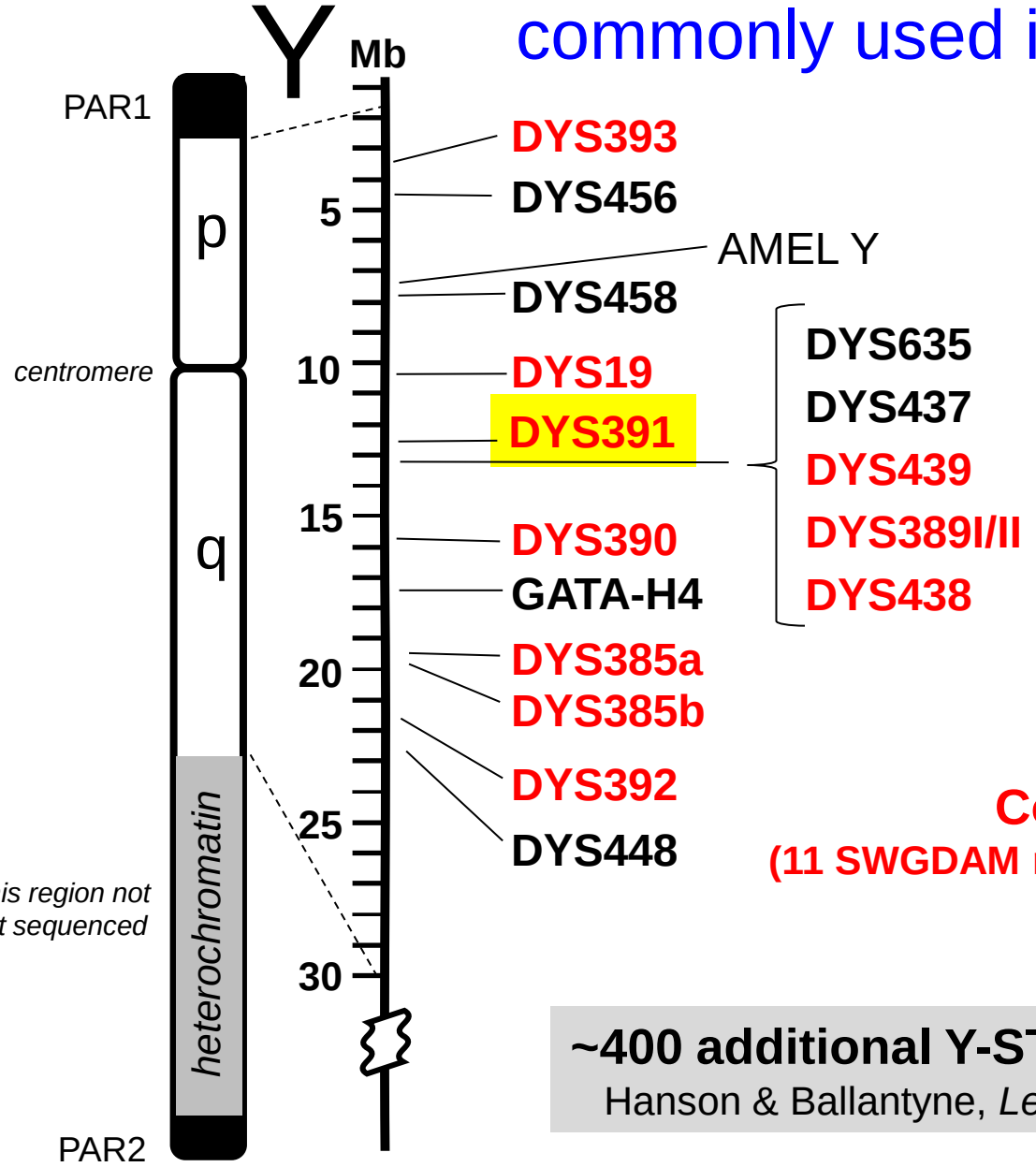
Characteristics of the 17 Commonly Used Y-STR Loci

STR Marker	Position (Mb)	Repeat Motif	Allele Range	Mutation Rate*
DYS393	3.19	AGAT	8-17	0.10 %
DYS456	4.33	AGAT	13-18	0.42 %
DYS458	7.93	GAAA	14-20	0.64 %
DYS19	10.13	TAGA	10-19	0.23 %
DYS391	12.61	TCTA	6-14	0.26 %
DYS635	12.89	TSTA	17-27	0.35 %
DYS437	12.98	TCTR	13-17	0.12 %
DYS439	13.03	AGAT	8-15	0.52 %
DYS389 I/II	13.12	TCTR	9-17 / 24-34	0.25 % / 0.36 %
DYS438	13.38	TTTTC	6-14	0.03 %
DYS390	15.78	TCTR	17-28	0.21 %
GATA-H4	17.25	TAGA	8-13	0.24 %
DYS385 a/b	19.26	GAAA	7-28	0.21 %
DYS392	21.04	TAT	6-20	0.04 %
DYS448	22.78	AGAGAT	17-24	0.16 %

*Mutation rates are from as many as 15000 meioses described in a YHRD summary of 23 publications in Jan 2011 (see (http://www.yhrd.org/Research/Loci/)

J.M. Butler (2012) *Advanced Topics in Forensic DNA Typing: Methodology*, Table 13.2

Relative positions of 17 Y-STR loci
commonly used in ChrY testing



DYS391 has been proposed for inclusion as a future CODIS core locus
(D.R. Hares, *FSI Genetics*, 2012, 6, e52-e54)

Core U.S. loci
(11 SWGDAM recommended in Jan 2003)

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

PAR = pseudo-autosomal region (recombines with X-chromosome)

J.M. Butler (2012) *Advanced Topics in Forensic DNA Typing: Methodology*, Figure 13.6

Recent Developments with Y-STR Typing

- **Promega Corporation** plans to release **PowerPlex Y23** (23 loci) in June 2012 which will enable further resolution of Y-STR haplotypes
  - Population databases will need to be developed with the new extended haplotypes

- Manfred Kayser's group has developed a set of **rapidly mutating (RM) Y-STR loci** that have the capability to resolve fathers and sons in many instances
  - An international collaboration is currently on-going to study these RM Y-STRs in more detail (14 RM Y-STRs in 3 multiplexes)

ELSEVIER

Forensic Science International 129 (2002) 10–24

Forensic Science International

www.elsevier.com/locate/forsciint

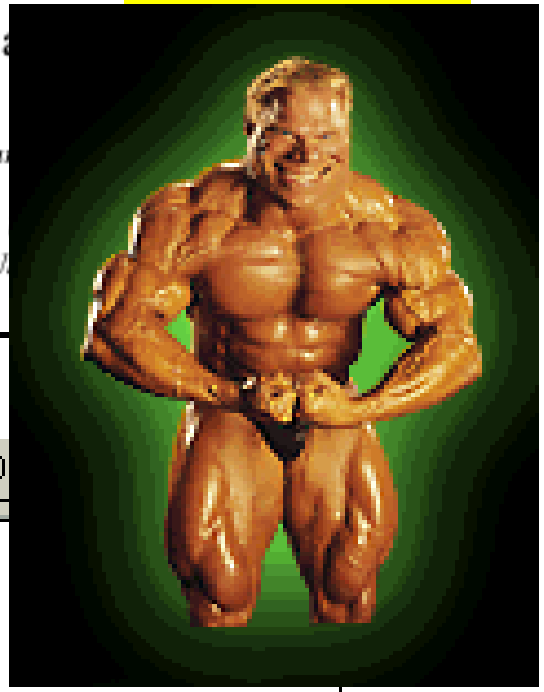
A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers

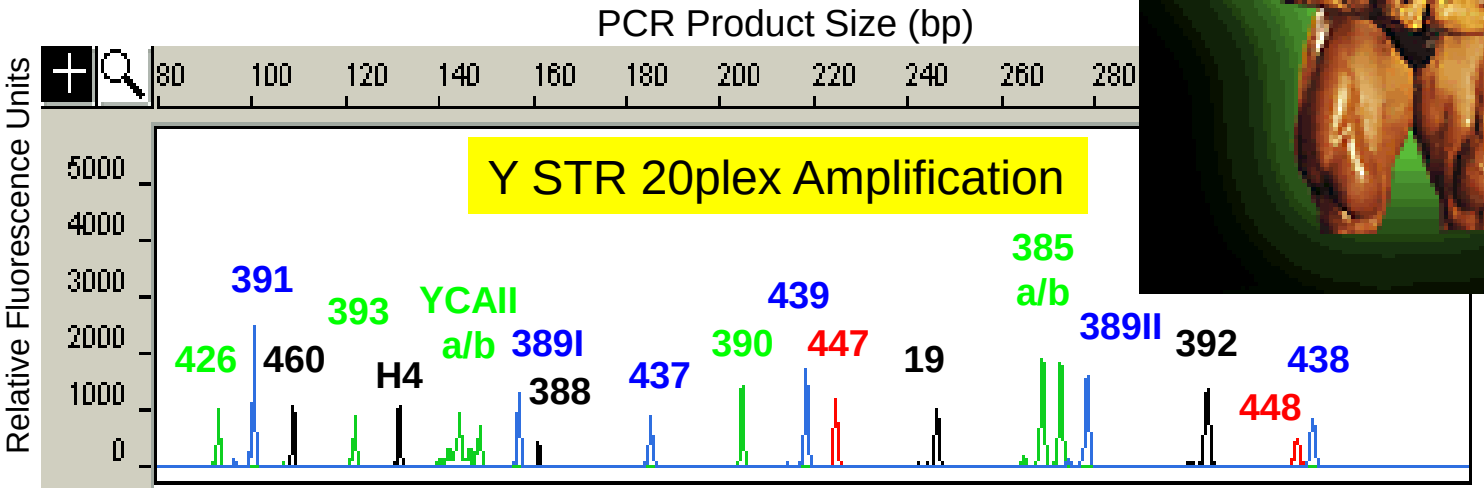
John M. Butler[a,*], Richard Schoske[a,b], Peter M. Vallone[a], Ma
Alan J. Redd[c], Michael F. Hammer[c]

[a]*Biotechnology Division, National Institute of Standards and Technology, 100 Bu*
Mail Stop 8311, Gaithersburg, MD 20899, USA
[b]*Department of Chemistry, American University, Washington, DC 20016,*
[c]*Division of Biotechnology, University of Arizona, Tucson, AZ 85721, U*

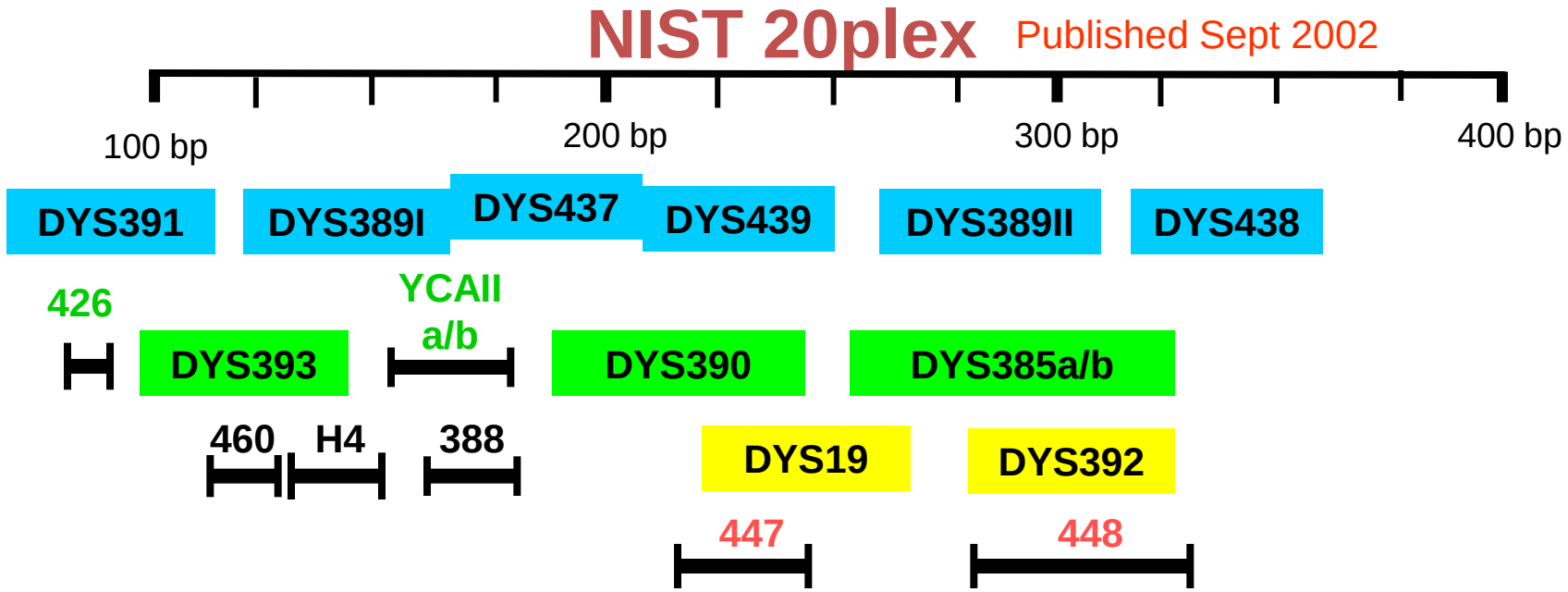
Received 22 February 2002; accepted 8 May 2002

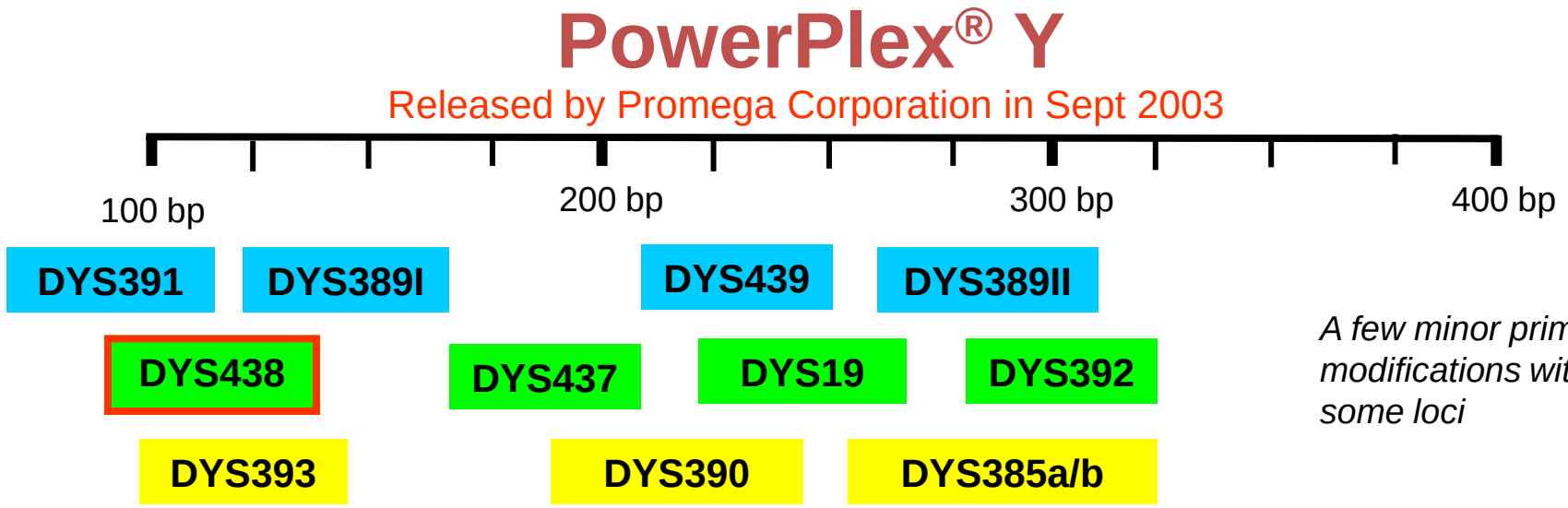
The Manly-plex





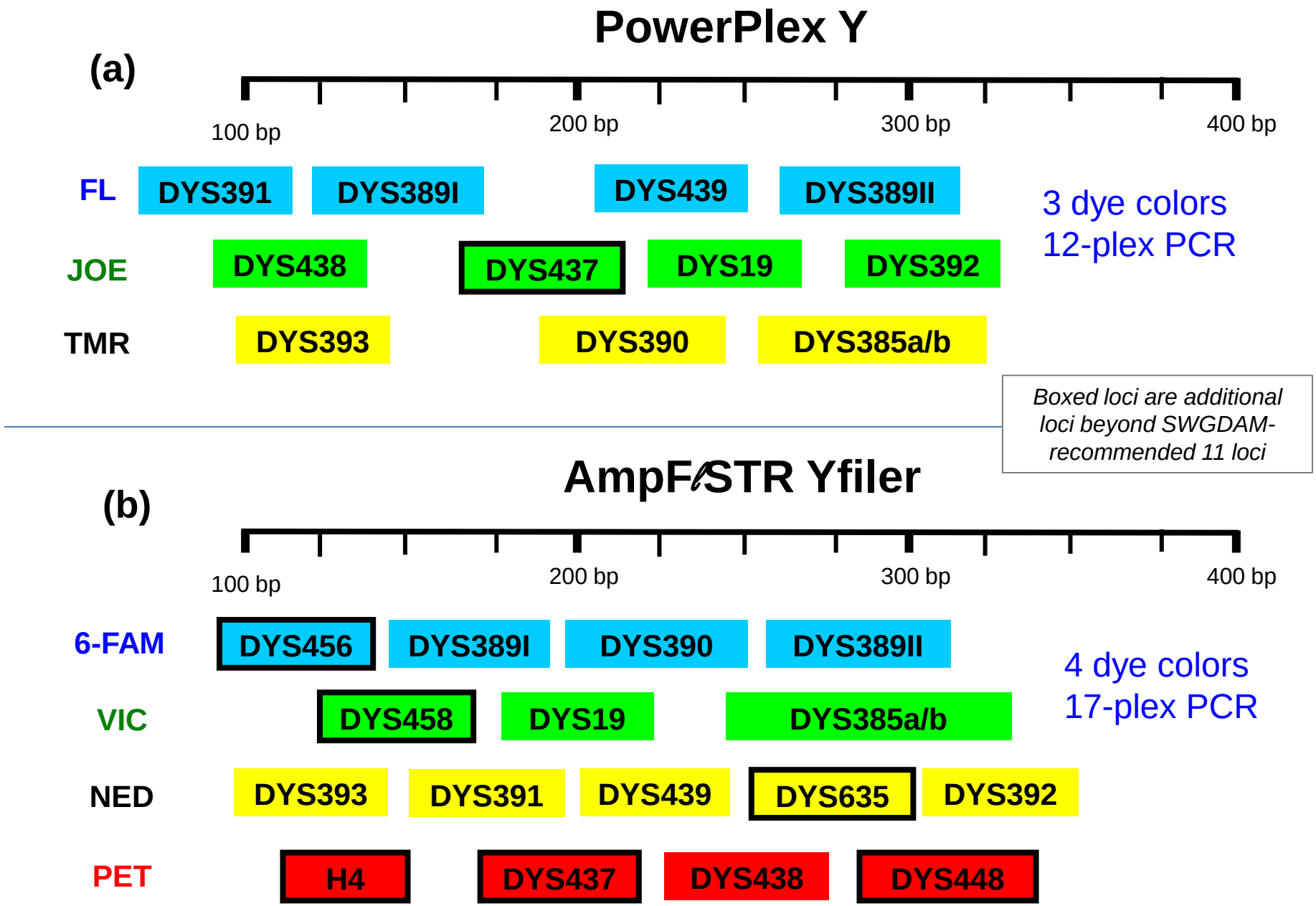
Allele size range and locus dye colors





A few minor primer modifications with some loci

Current Commercial Y-STR Kits (Loci, Dye Colors, Size Ranges)



# times haplotype observed	9 MHL
1	429
2	34
3	13
4	4
5	3
6	1
7	1
8	1
9	2
10	.
11	1
12	.
13	1
15	.
26	**1**
HD	0.996644
%DC	0.748476
# HT	491

429 of the 656 had a unique haplotype with the MHL loci, 34 sample haplotypes were observed twice in the sample set, 13 sample haplotypes were observed three times, etc.

With the 9 loci of the minimal haplotype (MHL) run on 656 samples, 26 samples had the most common type

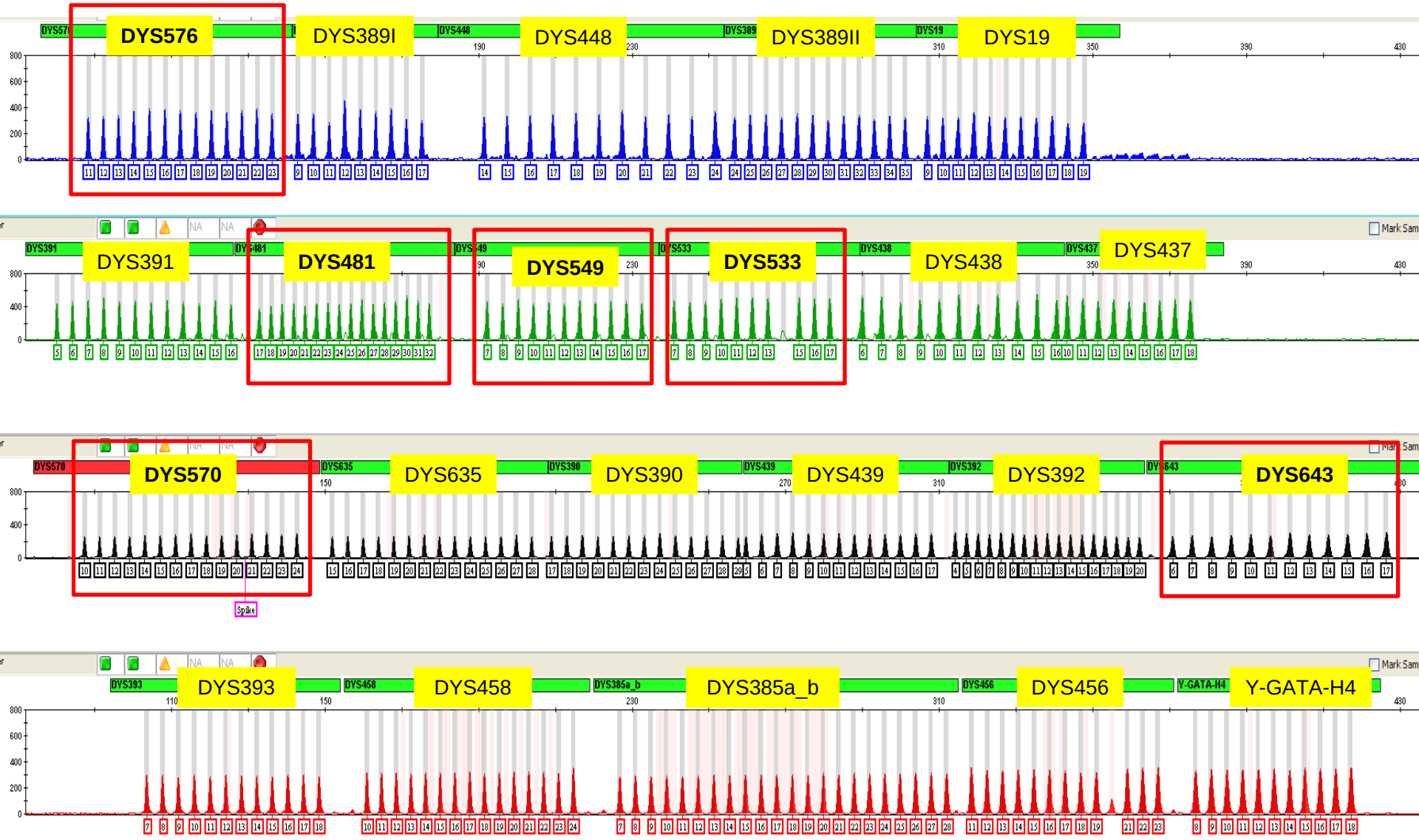
Total = 656 samples

# times haplotype observed	9 MHL	11 SWGDAM	12 PPY	17 Yfiler
1	**429**	**486**	**505**	**626**
2	34	33	34	**12**
3	13	10	14	**2**
4	4	6	3	.
5	3	1	2	.
6	1	1	.	.
7	1	2	1	.
8	1	.	.	.
9	2	.	.	.
10	.	1	.	.
11	1	.	.	.
12	.	.	**1**	.
13	1	.	.	.
15	.	**1**	.	.
26	**1**	.	.	.
HD	0.996644	0.998529	0.999064	0.999916
%DC	0.748476	0.824695	0.853659	0.97561
# HT	491	541	560	640

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

Total = 656 samples

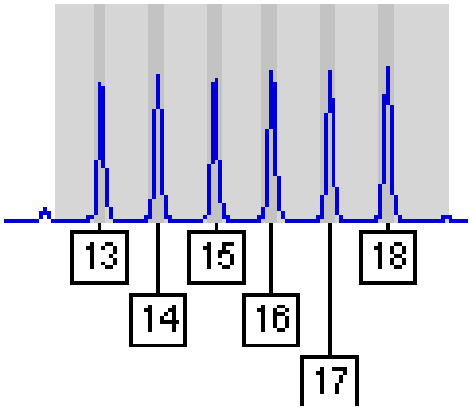
PowerPlex Y23 has 6 “new” loci



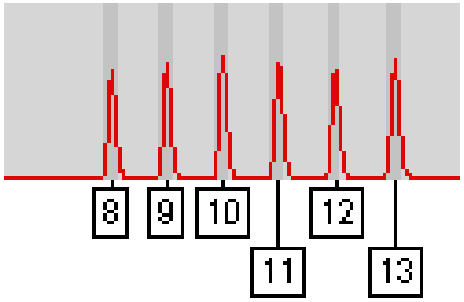
Improved Allelic Ladder Coverage

Yfiler

DYS456

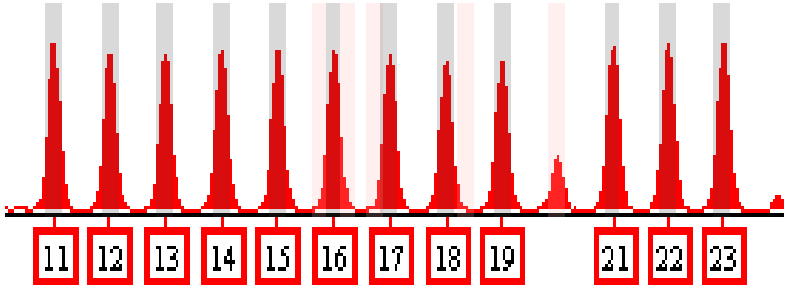


Y-GATA-H4

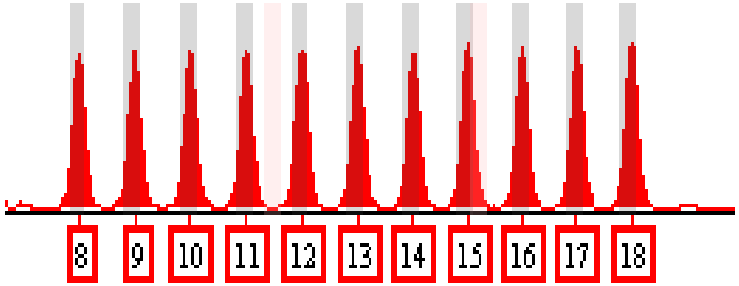


PowerPlex Y23

DYS456

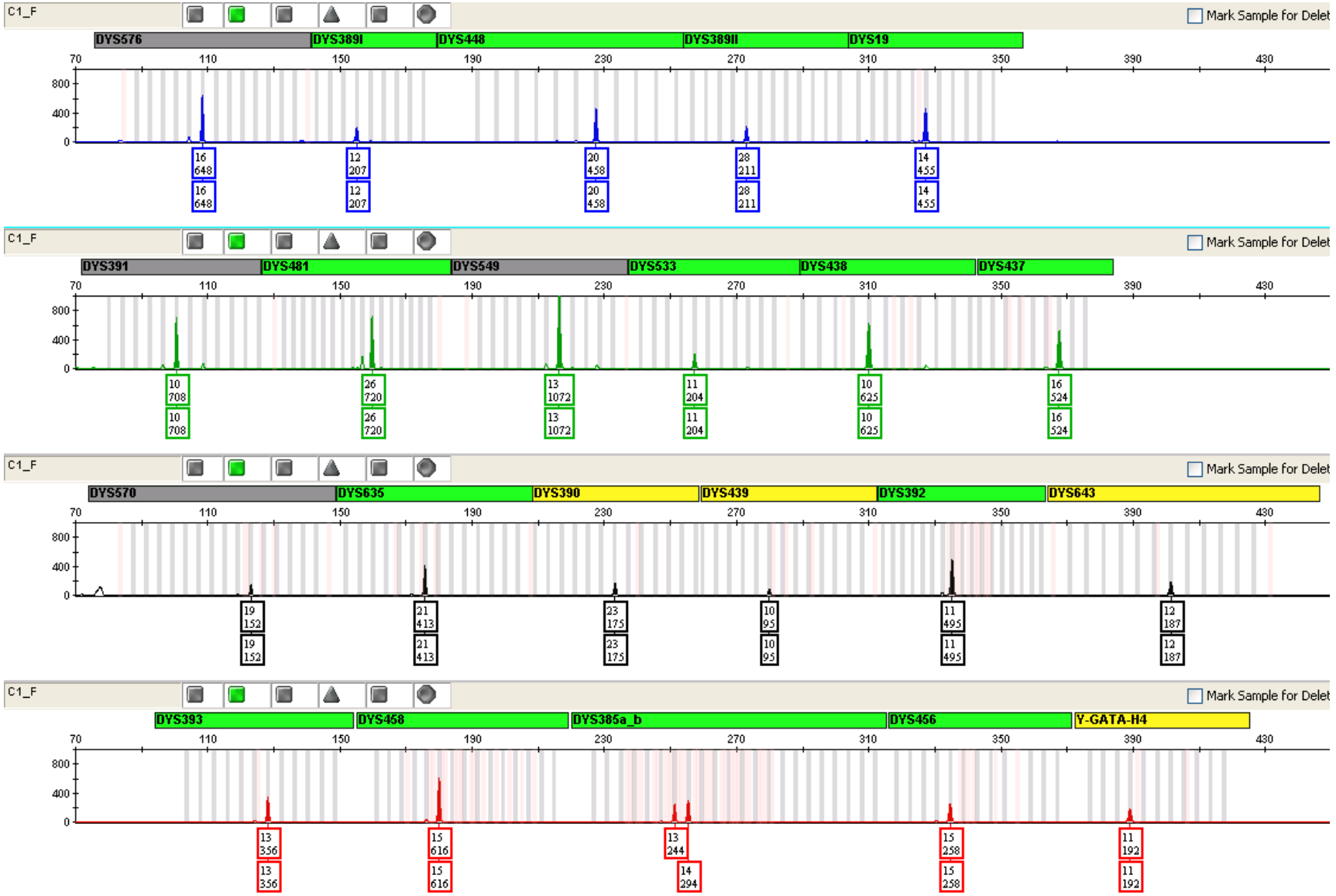


Y-GATA-H4



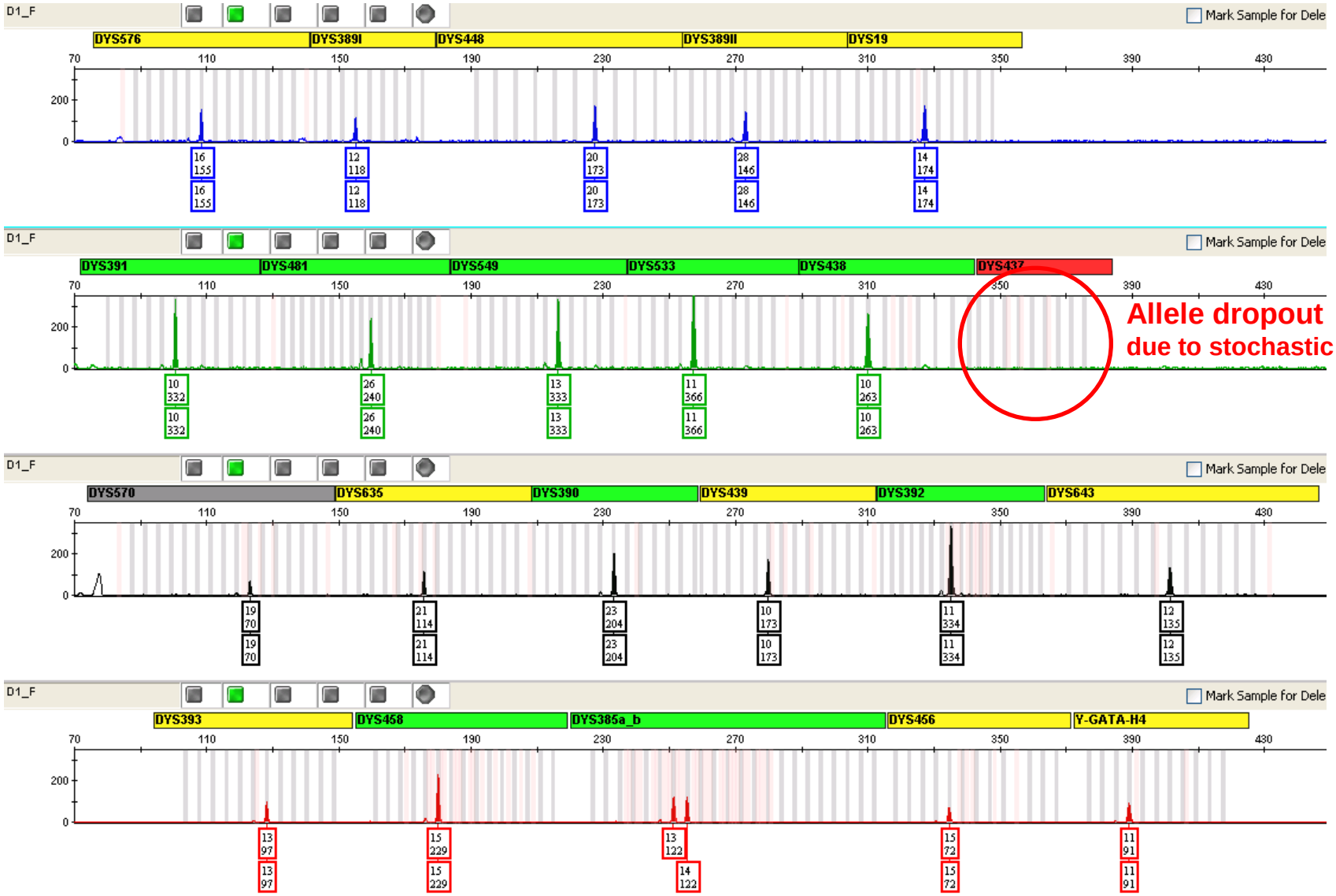
Excellent Sensitivity

Mixture of **125pg Male** + 400ng Female



Excellent Sensitivity

Mixture of **62.5pg Male** + 400ng Female



Y-STR Databases

YHRD has >100,000 haplotypes

On-line Y-STR Population Databases

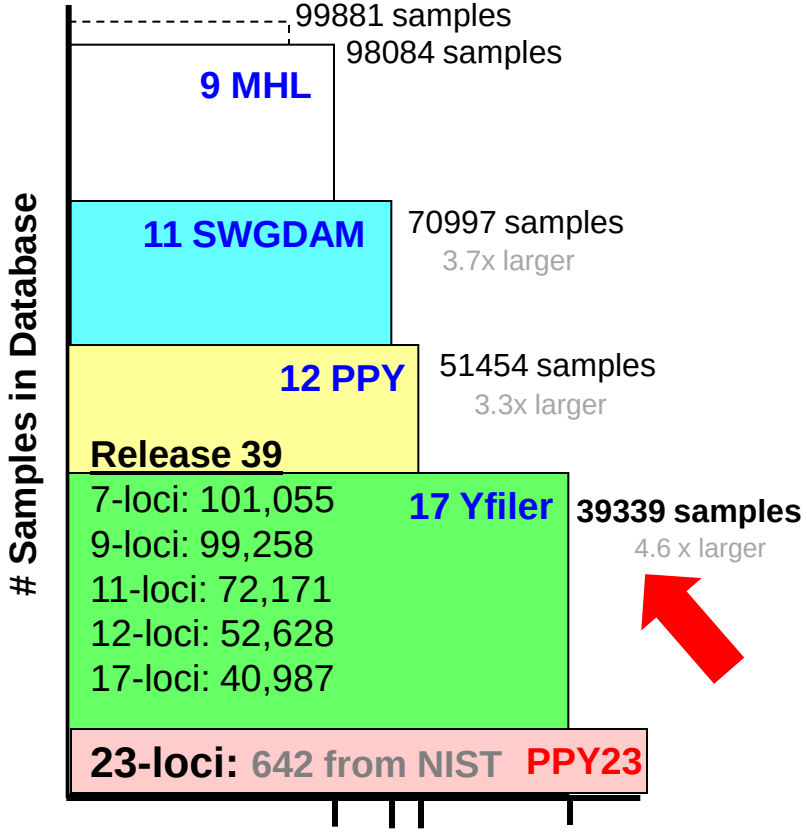
YHRD

Launched Feb 2000

http://www.yhrd.org

Release 38 Dec 30, 2011

750 Populations (109 countries)



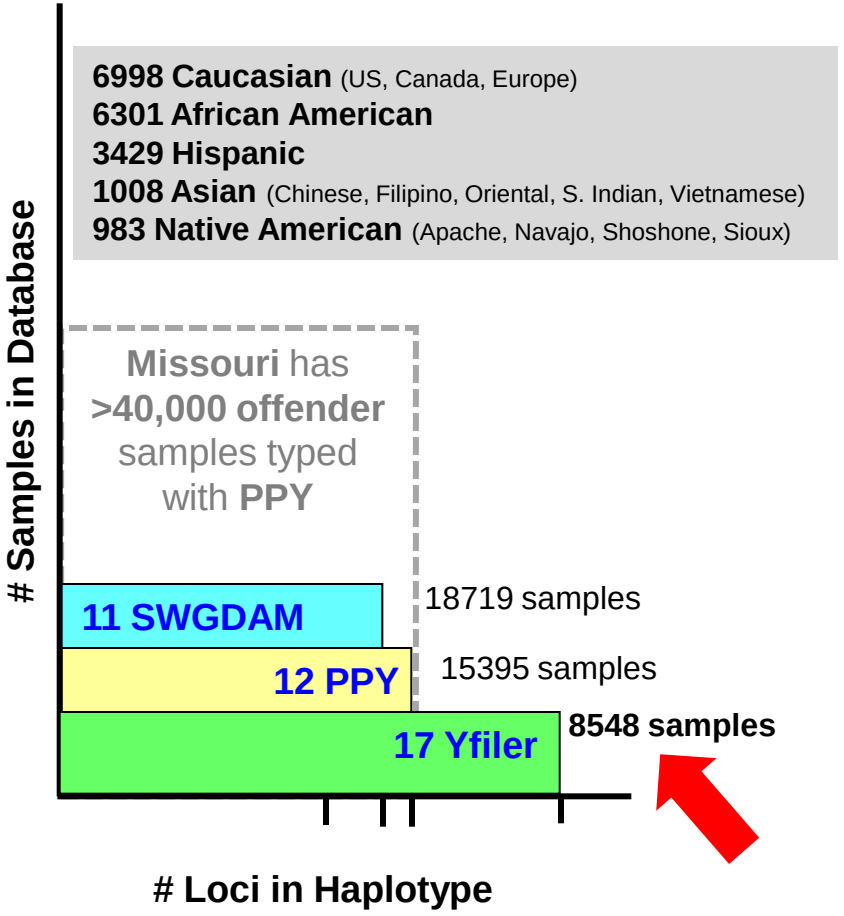
US YSTR

Launched Dec 2007

http://www.usystrdatabase.org

Release 2.6 Jan 3, 2012

Focus is on U.S. samples



Population Data Publications Describing Handling of Y-STR and mtDNA Haplotype Information

Forensic Science International: Genetics 4 (2010) 145–147

Contents lists available at ScienceDirect

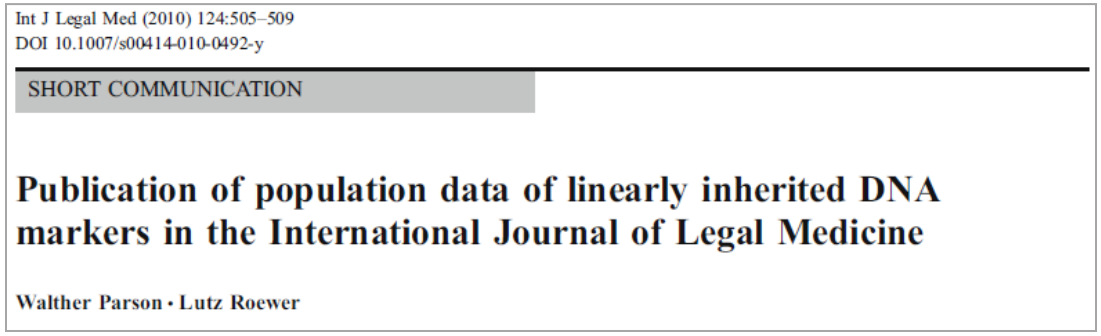
Forensic Science International: Genetics

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

Editorial

Publication of population data for forensic purposes

Carracedo, A., Butler, J.M., Gusmao, L., Parson, W., Roewer, L., Schneider, P.M. (2010) Editorial: Publication of population data for forensic purposes. ***Forensic Sci. Int. Genet***. 4: 145-147



Int J Legal Med (2010) 124:505–509
DOI 10.1007/s00414-010-0492-y

SHORT COMMUNICATION

Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine

Walther Parson · Lutz Roewer

Parson, W., Roewer, L. (2010) Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine. ***Int. J. Legal Med***. 124: 505-509

- The leading forensic journals *require* Y-STR and mtDNA population data to be reviewed by and submitted to YHRD and EMPOP

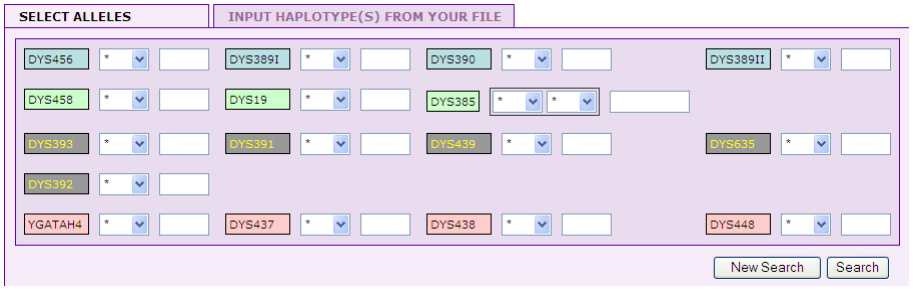
US YSTR Contributions

Contributor to US YSTR	# Samples	% of Database
Applied Biosystems (includes UNTHSC, NIST samples, …)	6,159	33%
Promega	3,800	20%
ReliaGene	3,037	16%
University of Arizona	2,462	13%
NCFS (University of Central Florida)	2,440	13%
Illinois State Police	398	2.1%
Santa Clara Co. CA Crime Lab	143	0.6%
Marshall University	113	0.6%
Washington State Patrol Crime Lab	40	0.2%
San Diego Sheriff's Regional Crime Lab	39	0.2%
CA DOJ	32	0.2%
Orange County CA Coroner	30	0.2%
Richland County Sheriff's Dept.	7	0.04%
Release 2.6 (Jan 3, 2012)	**18,719**	**8548** 17-locus profiles

http://www6.appliedbiosystems.com/yfilerdatabase/

Population	# Haplotypes
African American	1932
Asian	330
Asian Indian	564
Caucasian	4114
Chinese	577
Filipino	105
Hispanic	1601
Japanese	1078
Malay	579
Native American	105
Sub-saharan African	59
Thai	246
Vietnamese	103
All	**11393**

Applied Biosystems still maintains its Yfiler database



Y-STR Stats
& Interpretation Issues

Forensic Science International: Genetics 4 (2010) 281–291

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Fundamental problem of forensic mathematics—The evidential value of a rare haplotype

Charles H. Brenner [a,b,*]

[a] School of Public Health, Forensic Science Group, U.C. Berkeley, Berkeley, CA United States
[b] DNA·VIEW, 6801 Thornhill Drive, Oakland, CA 94611-1336, United States

"The fundamental question to decide the evidentiary significance of a trait linking suspect to crime is not one of frequency but of probability: What is the probability for such a match to happen by coincidence when the suspect is innocent?"

New Lineage Marker Interpretation Information

Forensic Science International: Genetics 5 (2011) 78–83

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

The interpretation of lineage markers in forensic DNA testing

J.S. Buckleton [a], M. Krawczak [b], B.S. Weir [c,*]

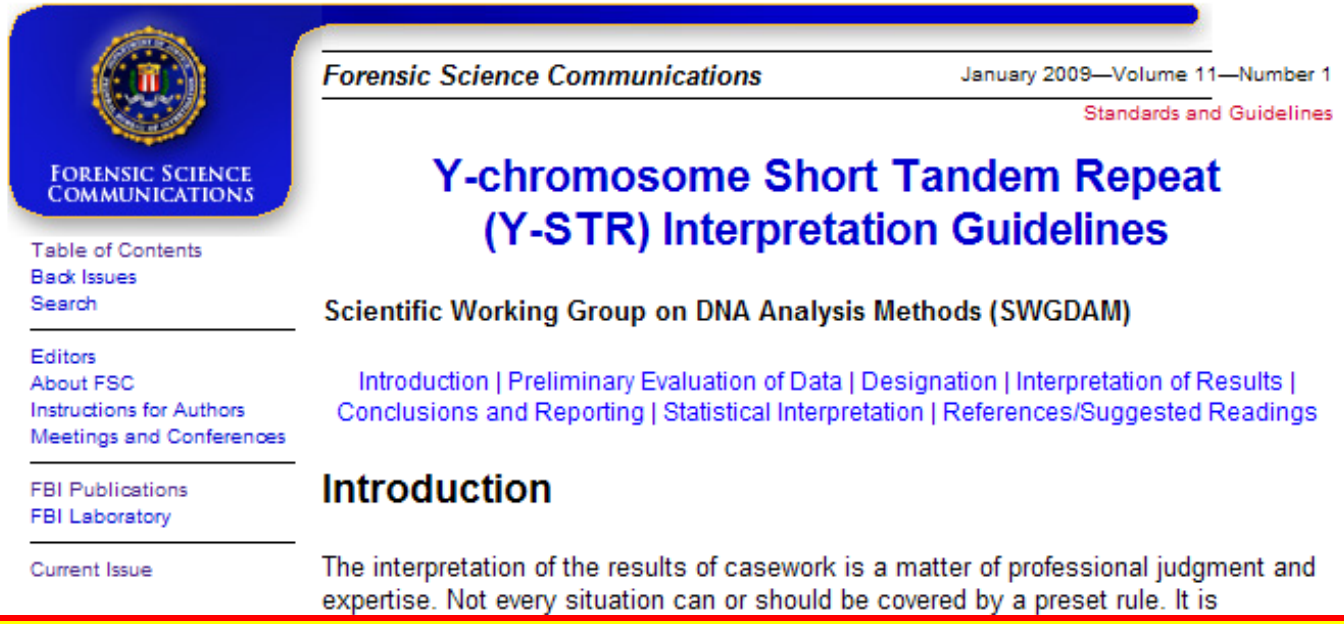
[a] ESR Ltd, Private Bag 92021, Auckland, New Zealand
[b] Institute of Medical Informatics and Statistics, Christian-Albrechts University, 24105 Kiel, Germany
[c] Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA

This article reviews and discusses a number of highly relevant topics:

- **Normal vs. binomial (Clopper-Pearson) sampling distributions**
- **Theta corrections**
- **Handling rare haplotypes (Charles Brenner approach)**
- **Combination of lineage and autosomal markers**

Current (2009) SWGDAM Y-STR Interpretation Guidelines

- **Approved July 15, 2008 by SWGDAM**
- Published in *Forensic Sci. Comm.* Jan 2009 issue



FORENSIC SCIENCE COMMUNICATIONS

Table of Contents
Back Issues
Search

Editors
About FSC
Instructions for Authors
Meetings and Conferences

FBI Publications
FBI Laboratory

Current Issue

Forensic Science Communications January 2009—Volume 11—Number 1

Standards and Guidelines

Y-chromosome Short Tandem Repeat (Y-STR) Interpretation Guidelines

Scientific Working Group on DNA Analysis Methods (SWGDAM)

Introduction | Preliminary Evaluation of Data | Designation | Interpretation of Results | Conclusions and Reporting | Statistical Interpretation | References/Suggested Readings

Introduction

The interpretation of the results of casework is a matter of professional judgment and expertise. Not every situation can or should be covered by a preset rule. It is

Will likely be updated soon to reflect change to Clopper-Pearson…

http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/jan2009/standards/2009_01_standards01.htm/

Results of Y-STR Profile Search

The following profile was searched on 15 January 2011 against several databases: DYS19 (14), DYS389I (13), DYS398II (29), DYS390 (24), DYS391 (11), DYS392 (13), DYS393 (13), DYS385 a/b (11,15), DYS438 (12), DYS439 (13), DYS437 (15), DYS448 (19), DYS456 (17), DYS458 (18), DYS635 (23), and GATA-H4 (12).

Database	Minimal haplotype (9 loci)	SWGDAM (11 loci)	PowerPlex Y (12 loci)	Yfiler (17 loci)	3/N for zero observations
YHRD	403/89804 = 0.45 %	29/62548 = 0.046 %	14/42277 = 0.033 %	0/30300 = <0.0033 %	3/30300 = 0.0099 %
US Y-STR	6/18547 = 0.032 %	1/18547 = 0.0054 %	1/15223 = 0.0066 %	0/8376 = <0.012 %	3/8376 = 0.036 %
Yfiler database	64/11393 = 0.56 %	4/11393 = 0.035 %	4/11393 =0.035 %	0/11393 = <0.0088 %	3/11393 = 0.026 %

Normal vs. Clopper-Pearson

In March 2010 the US Y-STR database changed its 95 % confidence interval calculations to the Clopper-Pearson method.

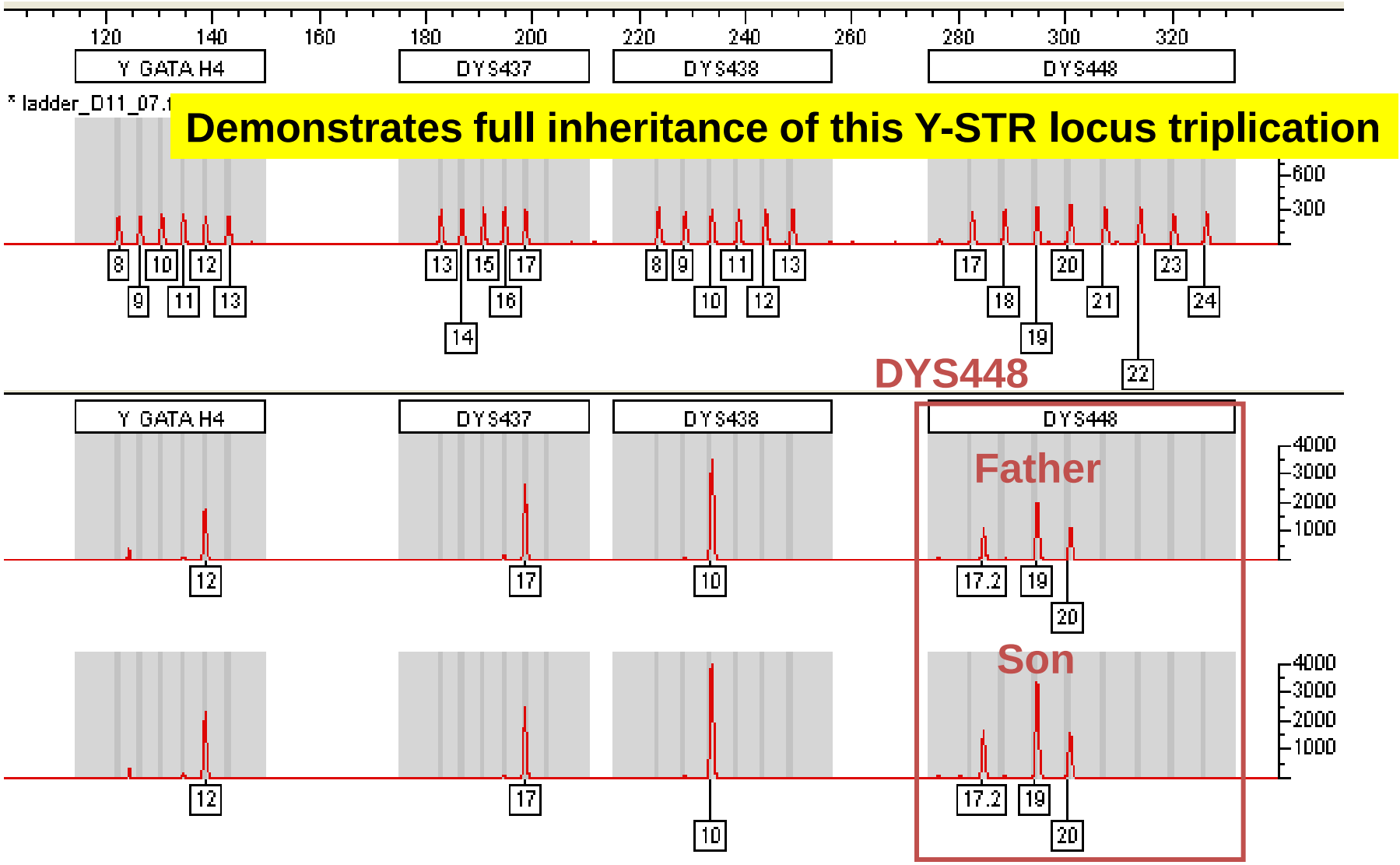
Count values	**Frequency** p = x/N	**Normal 95 % confidence interval**	**Clopper-Pearson** 95 % confidence interval*
YHRD 9 loci: **403**/89804	0.449 %	0.485 %	**0.487 %**
YHRD 12 loci: **14**/42277	0.0331 %	0.0477 %	**0.0518 %**
US Y-STR 12 loci: **1**/15223	0.0657 %	0.0174 %	**0.0317 %**

* Calculation performed with HaploCALc_1.0 Excel spreadsheet kindly provided by Steven P. Myers, CA DOJ

Note that with a large number of observations, such as 403 out of a database of 89804, there is almost no difference between the normal and Clopper-Pearson approaches. However, the normal method is less conservative (i.e., provides a more rare frequency) when the haplotype frequency is low, such as 1 out of 15223 or even 14 out of 42277. Although there are differences in these calculations, re-evaluation by the Clopper-Pearson method will not suddenly change a reported result by orders of magnitude or likely change the outcome of a report significantly.

DYS448 Triplication

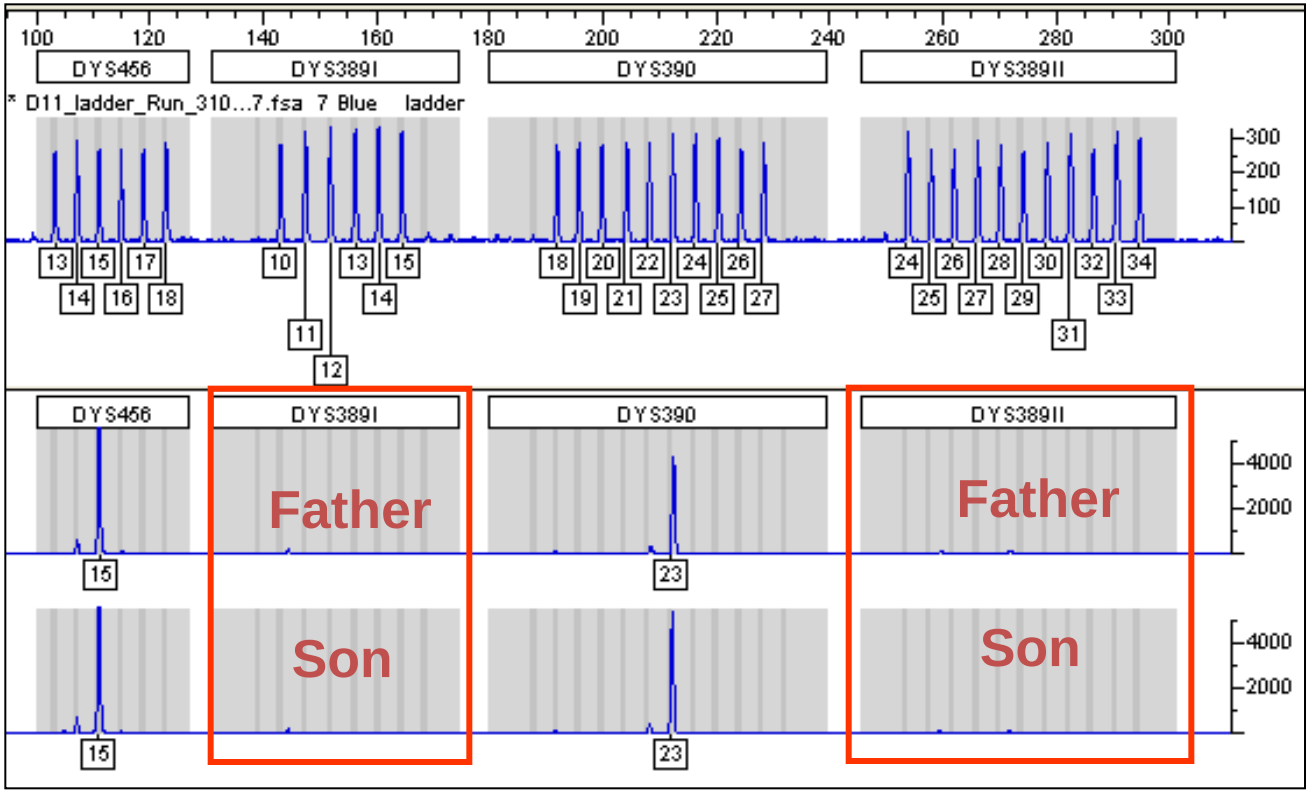
Seen in Both Father and Son

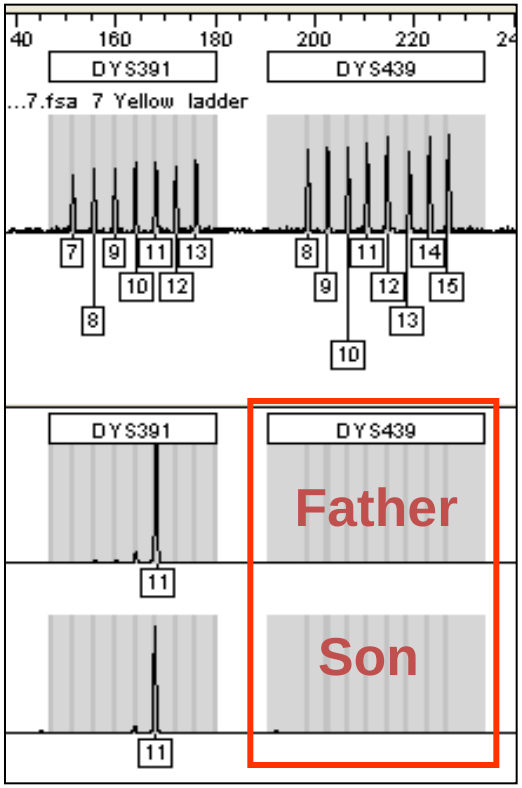


DYS389I, DYS389II, DYS439 Deletions

Seen in Both Father and Son

Yfiler data



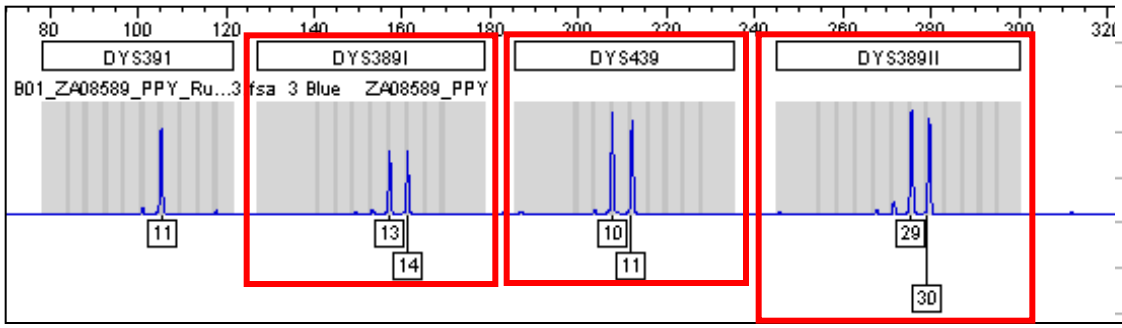


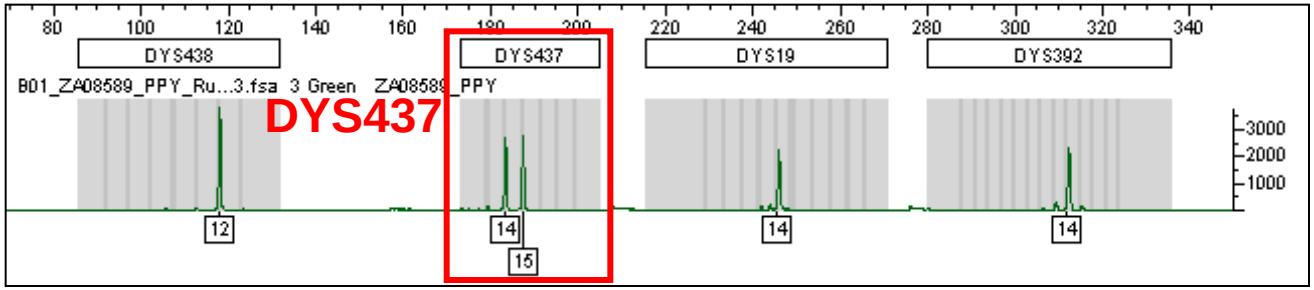
DYS389I **DYS389II** **DYS439**

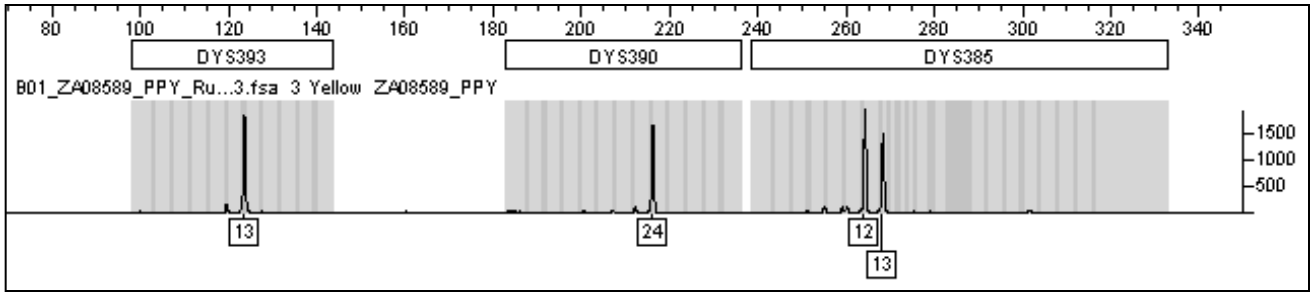
Full inheritance of these Y-STR locus deletions

Duplication at Multiple Loci with Single-Source Sample

PowerPlex Y data







Y-chromosome mapping

q-arm

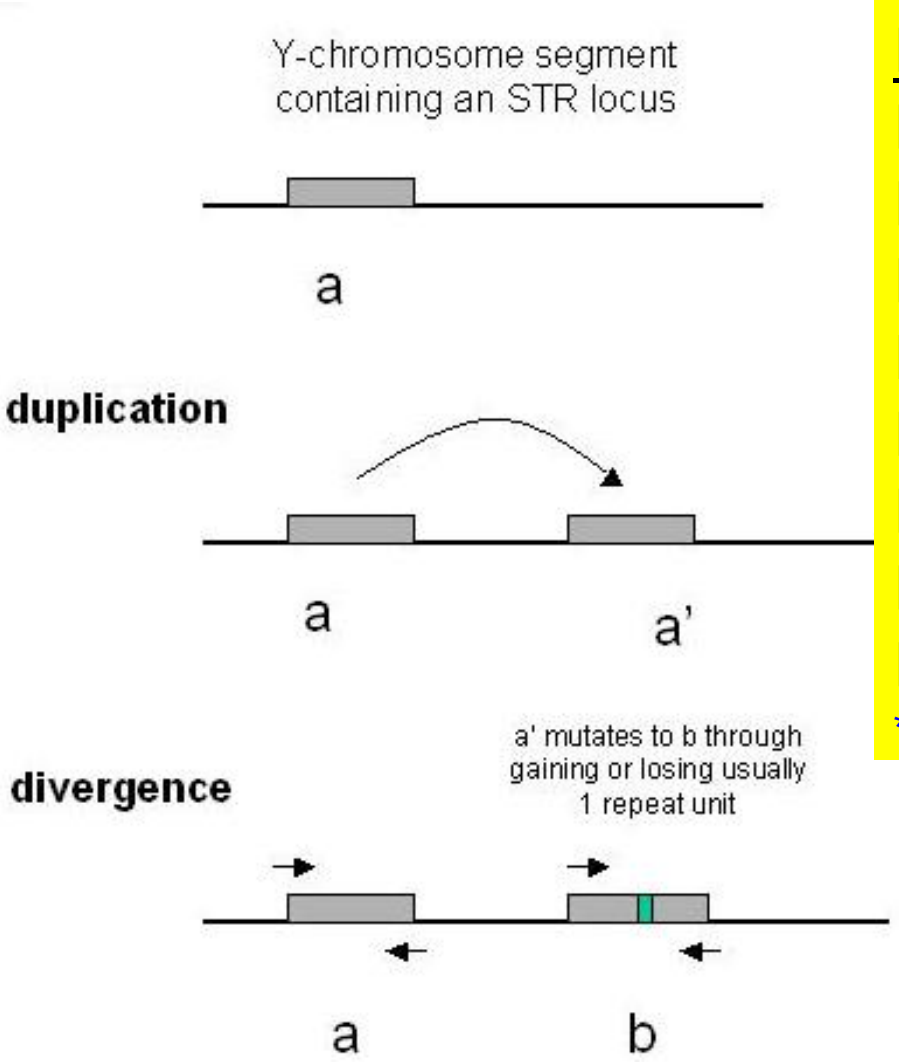
Y STR Marker	Position (Mb)
DYS391	13.413
DYS635 (C4)	13.690
DYS434	13.777
DYS437	13.778
DYS435	13.807
DYS439	13.826
DYS389 I/II	13.923
DYS388	14.057
DYS442	14.071
DYS438	14.248

Entire region of Y-chromosome has likely been duplicated and then diverged

Most duplications have a single repeat spread in allele patterns

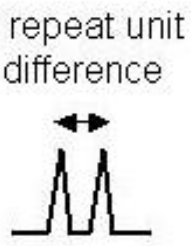
Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Duplication and Divergence Model



Locus	# dup*	>1 repeat
DYS19	23	2
DYS389I	5	0
DYS389II	9	2
DYS390	1	0
DYS391	3	1
DYS392	0	0
DYS393	3	0
DYS385a/b	17	0

*from www.yhrd.org, literature, and our work



92% have single repeat difference

Since single-step mutations are most common, then single repeat spacing in duplicated alleles is expected

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Deciphering between a Mixture of Multiple Males and Locus Duplication

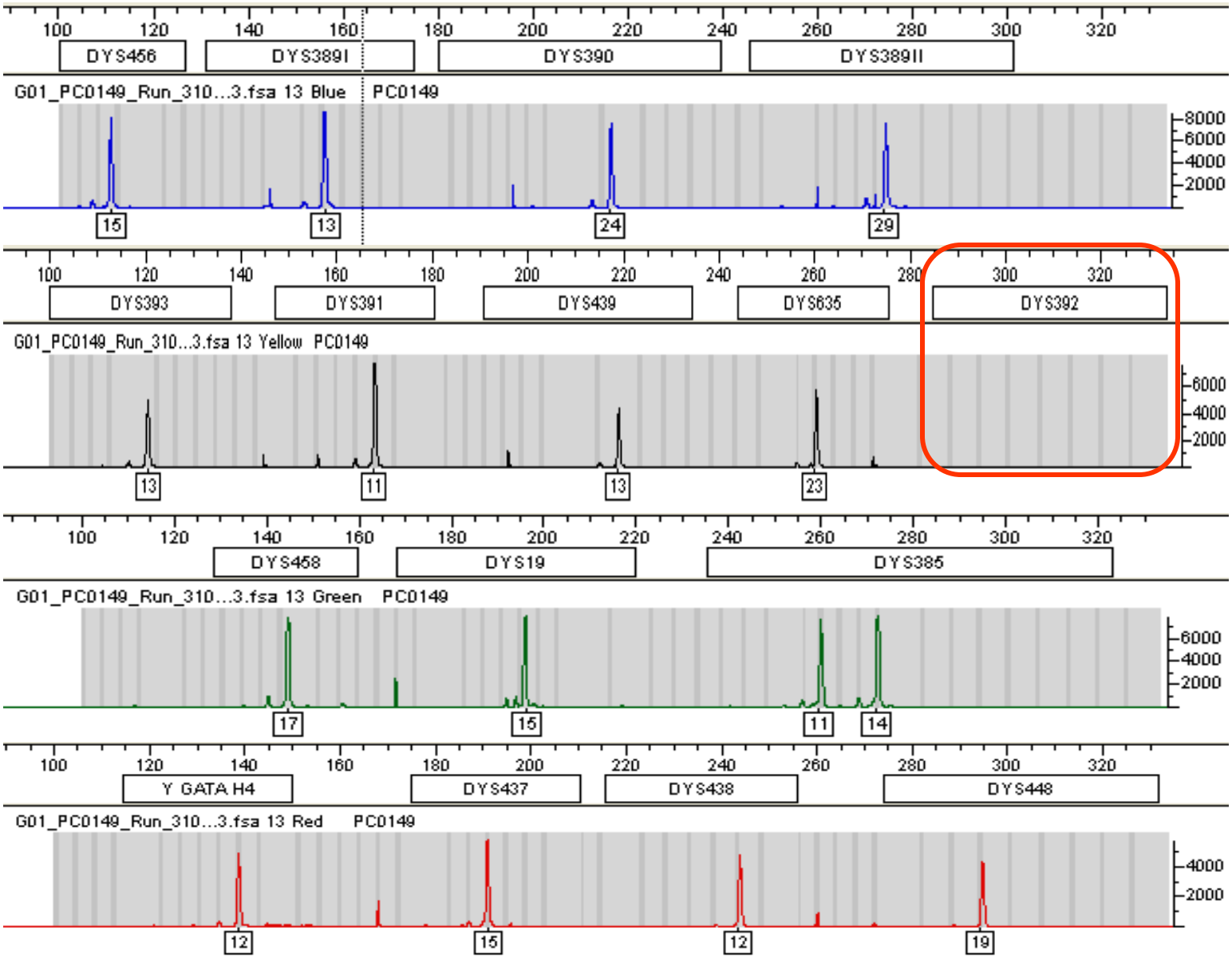
- ***Note the number of loci containing >1 allele (other than multi-copy DYS385)***
- ***Consider relative position on the Y-chromosome if multiple loci have two alleles***
- ***See if repeat spread is >1 repeat unit***
- ***Examine DYS385 for presence of >2 alleles***

Locus duplication along the Y-chromosome is in many ways analogous to heteroplasmy in mitochondrial DNA, which depending on the circumstances can provide greater strength to a match between two DNA samples.

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

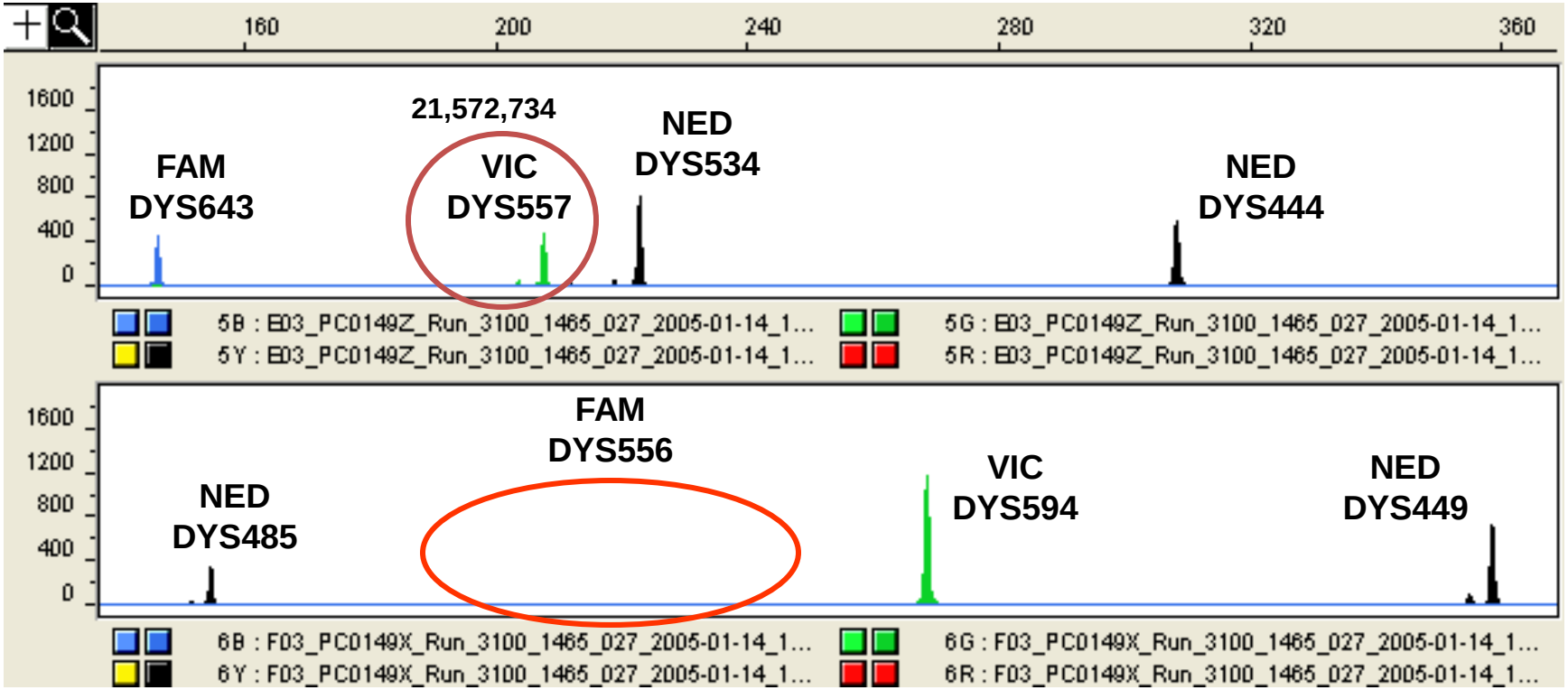
Sample PC0149 with Yfiler

DYS392 is deleted



PC0149 with Additional Y-STRs

One of the closest available loci fails



DYS556 ~32,000 bp away from DYS392 is **missing**

DYS557 ~600,000 bp away from DYS392 is present

Practical Information on Y Deletions

- If DYS458 is deleted in Yfiler, then your sample is likely to lack an Amelogenin Y amplicon as DYS458 and AMEL Y are 1.13 Mb apart on the short arm of the human Y-chromosome
  - Chang *et al.* (2007) *Forensic Sci. Int*. 166: 115-120

- Many Y-chromosomes are more complicated than originally thought!

Y-STR Summary

- Mutation rates are similar to autosomal STRs (~0.2%) – based on father-son studies
- Variant alleles are observed as in autosomal STRs due to flanking region mutations, etc.
- Regions of the Y-chromosome can be duplicated or deleted causing Y-STRs to be duplicated or deleted
- Careful primer design is important to avoid X-chromosome homology or Y-chromosome duplications

Standardization is Critical for Success and Data Sharing

Needs	How/When Accomplished
Core Y-STR loci	SWGDAM Y-STR Committee selected 11-loci in January 2003
Consistent allele nomenclature	NIST SRM 2395 (2003); kit allelic ladders; ISFG (2006) and NIST (2008) publications
Commercially available Y-STR kits	Early ReliaGene kits (2001-2003); **PowerPlex Y** (2003) and **Yfiler** (2004) **PowerPlex Y23** (2012)
Accessible, searchable population databases for haplotype frequency estimations	**YHRD** (72,171 11-locus haplotypes from 750 worldwide populations) **US YSTR** (18,719 11-locus haplotypes from primarily U.S. population groups)
Interpretation guidelines	SWGDAM Y-STR Interpretation Guidelines published in January 2009 (*will likely be revised soon*)

Predictions for **the Future** of Y-STR Analysis

- Continued use with casework (with excess female DNA)

- Improved frequency estimates with growing Y-STR databases
  - YHRD now at **70,997 11-locus profiles** (39,339 Yfiler)
  - USYSTR has **18,719 11-locus profiles** (8,548 Yfiler)

- Use with familial searching to eliminate false positives
  - Myers, S.P. et al. (2011) *FSI Genetics* 5(5): 493-500 – describes CA DOJ familial searching

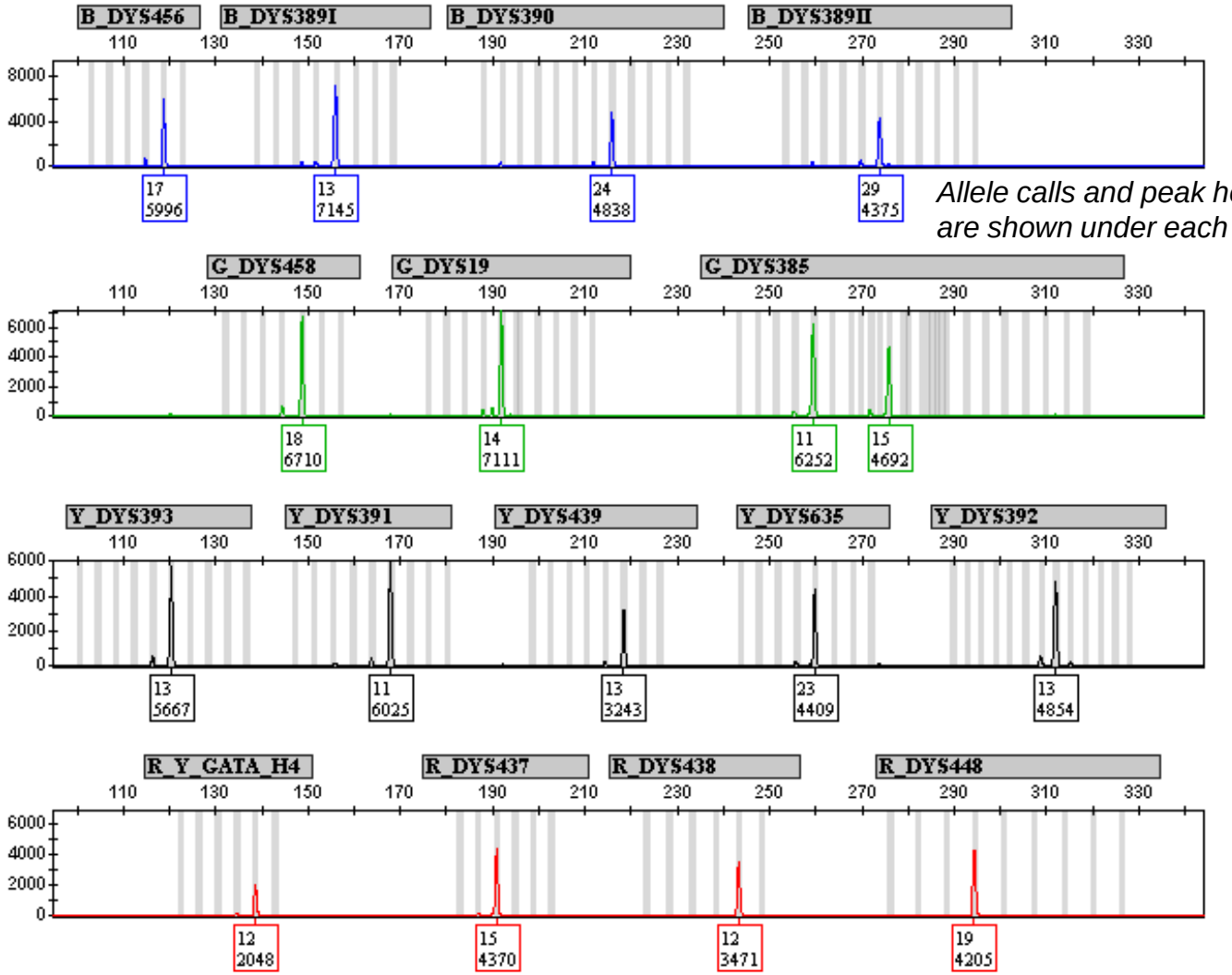
- **New Y-STR kits with additional loci**
  - At the ISHI meeting, Promega announced a Y-STR 23plex was being developed
  - Will take time though to grow large population databases that cover all of the new loci

- Use of fast mutating loci to help resolve paternal lineages (e.g., to separate brothers or father/son haplotypes)
  - Ballantyne, K.N. et al. (2010) *Am J Hum Genet* 87(3): 341-353
  - Ballantyne, K.N. et al. (2012) *FSI Genetics* (*in press*)

- **In some cases, being able to put a lineage name to an unknown Y-STR profiles using on-line genetic genealogy information**

Yfiler Result (17 Y-STRs)

from a Single-Source Male of European Ancestry



Allele calls and peak heights are shown under each peak

Haplotype

DYS19	389I	389II	390	391	392	393	385 a/b	438	439	437	448	456	458	635	H4
14	13	29	24	11	13	13	11-15	12	13	15	19	17	18	23	12

J.M. Butler (2010) *Fundamentals of Forensic DNA Typing*, Figure 16.4

Results of a Genetic Genealogy Search
with an "unknown" profile using (14 of 17) Yfiler loci

Compare	User ID	Pedigree	Last Name	Origin	Haplogroup	Tested With	Markers Compared	Genetic Distance
	KB56Q	Show	Smith	Slievenisky, Down, Northern Ireland	Unknown	Family Tree DNA	14	0
	XU3XE	Show	Butler	Ireland	Unknown	Family Tree DNA	14	0
	VAP7E		Butler	Ireland	R1b (tested)	Family Tree DNA	14	0
	74VV9	Show	Butler	Ireland	Unknown	Family Tree DNA	14	0
	T65UT		Butler	Ireland	Unknown	Family Tree DNA	14	0
	5BJX4		Butler	Ireland	Unknown	Family Tree DNA	14	0
	CYFNX		Butler	Unknown	Unknown	Family Tree DNA	14	0
	2B587		Butler	Unknown	Unknown	Family Tree DNA	14	0
	JSRJW		Butler	Unknown	Unknown	Family Tree DNA	14	0
	QWQG7		Butler	Unknown	Unknown	Family Tree DNA	14	0
	F9W7H		Butler	Unknown	Unknown	Family Tree DNA	14	0
	UXBFW		Butler	Unknown	Unknown	Family Tree DNA	14	0
	SFUSJ		Butler	Unknown	Unknown	Family Tree DNA	14	0
	4ZF4Z		Butler	Unknown	Unknown	Family Tree DNA	14	0
	P66AH		Harris	Unknown	R1b1a2*	Family Tree DNA	14	0
	W27DJ		Butler	Mississippi, USA	Unknown	Family Tree DNA	14	0
	VBVX9		Butler	South Carolina, USA	Unknown	Family Tree DNA	14	0
	FKNWZ		Butler	Mississippi, USA	Unknown	Family Tree DNA	14	0
	2NZ68		Butler	Quitman, Texas, USA	R1b*	Other - Anchestry by DNA	14	0
	ZFN67		Willhite (Adopted)	Tennessee, USA	Unknown	Family Tree DNA	14	0

17 of 20 full matches are "Butlers"

Other 3 are Butlers but didn't know it…
(adoption or other happenings in the gene pool of the past!)

www.Ysearch.org
Search conducted Jan 5, 2012

104,015 Records
80,143 Different Haplotypes
74,907 Surnames

Currently larger than YHRD – but serves a different purpose

YHRD Search Results (with 17 loci)

DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385
14	13	29	24	11	13	13	11,15

DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
12	13	15	19	17	18	23	12

National database | Metapopulations | SNP

Whole database

Search Reset

Matches grouped by Metapopulations | Matches grouped by Continents | Matches grouped by Haplogroups | Frequency surveying estimates

- **All Metapopulation:** Found 0 of 39339 matching haplotypes [f=0 (95% CI: 0 – 9.377×10^{-5})] in 0 of 263 populations.
  - **Eurasian Metapopulation:** Found 0 of 15455 matching haplotypes [f=0 (95% CI: 0 – 2.387×10^{-4})] in 0 of 113 populations.
  - **East Asian Metapopulation:** Found 0 of 12522 matching haplotypes [f=0 (95% CI: 0 – 2.945×10^{-4})] in 0 of 63 populations.
  - **Australian Aboriginal Metapopulation:** Found 0 of 766 matching haplotypes [f=0 (95% CI: 0 – 4.804×10^{-3})] in 0 of 1 populations.
  - **African Metapopulation:** Found 0 of 1533 matching haplotypes [f=0 (95% CI: 0 – 2.403×10^{-3})] in 0 of 10 populations.
  - **Native American Metapopulation:** Found 0 of 384 matching haplotypes [f=0 (95% CI: 0 – 9.56×10^{-3})] in 0 of 9 populations.
  - **Eskimo Aleut Metapopulation:** Found 0 of 301 matching haplotypes [f=0 (95% CI: 0 – 1.218×10^{-2})] in 0 of 2 populations.
  - **Afro-Asiatic Metapopulation:** Found 0 of 1636 matching haplotypes [f=0 (95% CI: 0 – 2.252×10^{-3})] in 0 of 20 populations.
  - **Admixed Metapopulation:** Found 0 of 6742 matching haplotypes [f=0 (95% CI: 0 – 5.47×10^{-4})] in 0 of 45 populations.

0 matches found in 39,339 Yfiler profiles searched from 263 populations worldwide

Geographical projection

With 95% confidence interval

≈3/n = 3/39,339 = **1 in 13,113 ≈ 1 in 13,000**

Primary Steps Involved:
1 - Generate profile (Y or mtDNA)
2 - Query population database
3 - Report frequency estimate (with adjustment?)

Summary of Issues

Want good quality data going into database

population studies

Real-World Population Variation

Forensic Sample

1

HV1/HV2 control region mtGenome?

PP Y
Yfiler

mtDNA sequence (or difference from reference)

Y-STR Profile

2

Online **Population Database**

EMPOP
YHRD
US YSTR

Can θ correction help with this sampling issue?

Assume pop data representative of real-world

3 Counting method

Sampling confidence interval (normal or C-P; 1-tail or 2-tail)

?

What the court wants to know…

Haplotype Frequency Estimate

(of lineage <u>not</u> the individual)

Genetic Genealogy
&
Familial Searching

Trying to connect

close male relatives

FamilyTree DNA tests 111 Y-STR Markers for Genetic Genealogy

Panel	#	STR Marker
Panel 1 (Y-DNA1-12)	1	**DYS393**
Panel 1 (Y-DNA1-12)	2	**DYS390**
Panel 1 (Y-DNA1-12)	3	**DYS19**
Panel 1 (Y-DNA1-12)	4	**DYS391**
Panel 1 (Y-DNA1-12)	5-6	**DYS385a-b**
Panel 1 (Y-DNA1-12)	7	DYS426
Panel 1 (Y-DNA1-12)	8	DYS388
Panel 1 (Y-DNA1-12)	9	**DYS439**
Panel 1 (Y-DNA1-12)	10	**DYS389I**
Panel 1 (Y-DNA1-12)	11	**DYS392**
Panel 1 (Y-DNA1-12)	12	**DYS389II**
Panel 2 (Y-DNA13-25)	13	**DYS458**
Panel 2 (Y-DNA13-25)	14-15	DYS459a-b
Panel 2 (Y-DNA13-25)	16	DYS455
Panel 2 (Y-DNA13-25)	17	DYS454
Panel 2 (Y-DNA13-25)	18	DYS447
Panel 2 (Y-DNA13-25)	19	**DYS437**
Panel 2 (Y-DNA13-25)	20	**DYS448**
Panel 2 (Y-DNA13-25)	21	DYS449
Panel 2 (Y-DNA13-25)	22-25	DYS464a-b-c-d
Panel 3 (Y-DNA26-37)	26	DYS460
Panel 3 (Y-DNA26-37)	27	**Y-GATA-H4**
Panel 3 (Y-DNA26-37)	28-29	YCA II a-b
Panel 3 (Y-DNA26-37)	30	**DYS456**
Panel 3 (Y-DNA26-37)	31	DYS607
Panel 3 (Y-DNA26-37)	32	**DYS576**
Panel 3 (Y-DNA26-37)	33	**DYS570**
Panel 3 (Y-DNA26-37)	34-35	CDY a-b
Panel 3 (Y-DNA26-37)	36	DYS442
Panel 3 (Y-DNA26-37)	37	**DYS438**

Panel	#	STR Marker
Panel 4 (Y-DNA38-67)	38	DYS531
Panel 4 (Y-DNA38-67)	39	DYS578
Panel 4 (Y-DNA38-67)	40-41	DYF395S1a-b
Panel 4 (Y-DNA38-67)	42	DYS590
Panel 4 (Y-DNA38-67)	43	DYS537
Panel 4 (Y-DNA38-67)	44	DYS641
Panel 4 (Y-DNA38-67)	45	DYS472
Panel 4 (Y-DNA38-67)	46	DYF406S1
Panel 4 (Y-DNA38-67)	47	DYS511
Panel 4 (Y-DNA38-67)	48	DYS425
Panel 4 (Y-DNA38-67)	49-50	DYS413a-b
Panel 4 (Y-DNA38-67)	51	DYS557
Panel 4 (Y-DNA38-67)	52	DYS594
Panel 4 (Y-DNA38-67)	53	DYS436
Panel 4 (Y-DNA38-67)	54	DYS490
Panel 4 (Y-DNA38-67)	55	DYS534
Panel 4 (Y-DNA38-67)	56	DYS450
Panel 4 (Y-DNA38-67)	57	DYS444
Panel 4 (Y-DNA38-67)	58	**DYS481**
Panel 4 (Y-DNA38-67)	59	DYS520
Panel 4 (Y-DNA38-67)	60	DYS446
Panel 4 (Y-DNA38-67)	61	DYS617
Panel 4 (Y-DNA38-67)	62	DYS568
Panel 4 (Y-DNA38-67)	63	DYS487
Panel 4 (Y-DNA38-67)	64	DYS572
Panel 4 (Y-DNA38-67)	65	DYS640
Panel 4 (Y-DNA38-67)	66	DYS492
Panel 4 (Y-DNA38-67)	67	DYS565

Panel	#	STR Marker
Panel 5 (Y-DNA68-111)	68	DYS710
Panel 5 (Y-DNA68-111)	69	DYS485
Panel 5 (Y-DNA68-111)	70	DYS632
Panel 5 (Y-DNA68-111)	71	DYS495
Panel 5 (Y-DNA68-111)	72	DYS540
Panel 5 (Y-DNA68-111)	73	DYS714
Panel 5 (Y-DNA68-111)	74	DYS716
Panel 5 (Y-DNA68-111)	75	DYS717
Panel 5 (Y-DNA68-111)	76	DYS505
Panel 5 (Y-DNA68-111)	77	DYS556
Panel 5 (Y-DNA68-111)	78	**DYS549**
Panel 5 (Y-DNA68-111)	79	DYS589
Panel 5 (Y-DNA68-111)	80	DYS522
Panel 5 (Y-DNA68-111)	81	DYS494
Panel 5 (Y-DNA68-111)	82	**DYS533**
Panel 5 (Y-DNA68-111)	83	DYS636
Panel 5 (Y-DNA68-111)	84	DYS575
Panel 5 (Y-DNA68-111)	85	DYS638
Panel 5 (Y-DNA68-111)	86	DYS462
Panel 5 (Y-DNA68-111)	87	DYS452
Panel 5 (Y-DNA68-111)	88	DYS445
Panel 5 (Y-DNA68-111)	89	Y-GATA-A10
Panel 5 (Y-DNA68-111)	90	DYS463
Panel 5 (Y-DNA68-111)	91	DYS441
Panel 5 (Y-DNA68-111)	92	Y-GGAAT-1B07
Panel 5 (Y-DNA68-111)	93	DYS525
Panel 5 (Y-DNA68-111)	94	DYS712
Panel 5 (Y-DNA68-111)	95	DYS593
Panel 5 (Y-DNA68-111)	96	DYS650
Panel 5 (Y-DNA68-111)	97	DYS532

Panel	#	STR Marker
Panel 5 (Y-DNA68-111)	98	DYS715
Panel 5 (Y-DNA68-111)	99	DYS504
Panel 5 (Y-DNA68-111)	100	DYS513
Panel 5 (Y-DNA68-111)	101	DYS561
Panel 5 (Y-DNA68-111)	102	DYS552
Panel 5 (Y-DNA68-111)	103	DYS726
Panel 5 (Y-DNA68-111)	104	**DYS635**
Panel 5 (Y-DNA68-111)	105	DYS587
Panel 5 (Y-DNA68-111)	106	**DYS643**
Panel 5 (Y-DNA68-111)	107	DYS497
Panel 5 (Y-DNA68-111)	108	DYS510
Panel 5 (Y-DNA68-111)	109	DYS434
Panel 5 (Y-DNA68-111)	110	DYS461
Panel 5 (Y-DNA68-111)	111	DYS435

Provides coverage of **17 Yfiler loci** and **6 additional loci** in **PowerPlex Y23 loci**

Expected Number of Y-STR Differences with Various Levels of Relatedness Between Tested Males

	12 loci	25 loci	37 loci	67 loci	111 loci	Interpretation
Very Tightly Related	N/A	N/A	0	0	0	Your exact match means your relatedness is extremely close. Few people achieve this close level of a match. All confidence levels are well within the time frame that surnames were adopted in Western Europe.
Tightly Related	N/A	N/A	1	1-2	1-2	Few people achieve this close level of a match. All confidence levels are well within the time frame that surnames were adopted in Western Europe.
Related	0	0-1	2-3	3-4	3-5	Your degree of matching is within the range of most well-established surname lineages in Western Europe. If you have tested with the Y-DNA12 or Y-DNA25 test, you should consider upgrading to additional STR markers. Doing so will improve your time to common ancestor calculations.
Probably Related	1	2	4	5-6	6-7	Without additional evidence, it is unlikely that you share a common ancestor in recent genealogical times (1 to 6 generations). You may have a connection in more distant genealogical times (less than 15 generations). If you have traditional genealogy records that indicate a relationship, then by testing additional individuals you will either prove or disprove the connection.
Only Possibly Related	2	3	5	7	8-10	It is unlikely that you share a common ancestor in genealogical times (1 to 15 generations). Should you have traditional genealogy records that indicate a relationship, then by testing additional individuals you will either prove or disprove the connection. A careful review of your genealogical records is also recommended.
Not Related	3	4	6	>7	>10	You are not related on your Y-chromosome lineage within recent or distant genealogical times (1 to 15 generations).

From http://www.familytreedna.com

If two men share a surname, how should the genetic distance at 25 Y-chromosome STR markers be interpreted?

Genetic Distance	Relationship	Interpretation
0	Related	A perfect 25/25 match between two men who share a surname (or variant) means they likely share a common male ancestor within the genealogical time frame. The probability of a close relationship is very high.
1	Related	A 24/25 match between two men who share a surname (or variant) means they likely share a common male ancestor within the genealogical time frame. For most closely related and same surnamed individuals, the mismatch markers are often DYS439, DYS385, DYS389i, DYS389ii, DYS458, DYS459, DYS449, and DYS464 which have shown themselves to move most rapidly.
2	Probably Related	A 23/25 match between two men who share a surname (or variant) means they may share a common male ancestor within the genealogical time frame. The probability of a relationship is good. However, your results show mutations and therefore more time between you and the other same surnamed person. For most closely related and same surnamed individuals, the mismatch markers are often DYS439, DYS385, DYS389i, DYS389ii, DYS458, DYS459, DYS449, and DYS464 which have shown themselves to move most rapidly.

From http://www.familytreedna.com

If two men share a surname, how should the genetic distance at 111 Y-chromosome STR markers be interpreted?

Genetic Distance	Relationship	Interpretation	Related in This Number of Generations or LESS			
			Confidence			
			50%	90%	95%	99%
0	Very Tightly Related	A 111/111 match indicates a very close or immediate relationship. Most exact matches are 3rd cousins or closer, and over half are related within two generations (1st cousins).	2	4	5	6
1	Tightly Related	A 110/111 match indicates a close relationship. Most one-off matches are 5th or more recent cousins, and over half are 2nd cousins or closer.	3	6	7	9
2	Tightly Related	A 109/111 match indicates a close relationship. Most matches are 7th cousins or closer, and over half are 4th or more recent cousins.	5	8	9	11
3	Related	A 108/111 match indicates a genealogical relationship. Most matches at this level are related as 9th cousins or closer, and over half will be 5th or more recent cousins. This is well within the range of traditional genealogy.	6	10	11	14

From http://www.familytreedna.com

Rapidly Mutating (RM) Y-STRs

Trying to separate

close male relatives

Mutations Seen in 100 African American Father-Son Pairs

Ethnicity	Sample	locus	Allele (father)	Allele (child)	Comments
African American	65B	Y GATA H4	11	9	**loss of 2 repeats**
African American	46B	DYS389I and DYS389II	14,30	13,29	loss of 1 repeat
African American	58B	DYS389I and DYS389II	14,32	15,33	gain of 1 repeat
African American	18B	DYS390	24	23	loss of 1 repeat
African American	90B	DYS456	15	16	gain of 1 repeat
African American	16B	**DYS458**	18	19	gain of 1 repeat
African American	39B	**DYS458**	18	19	gain of 1 repeat
African American	16B	**DYS635**	23	22	loss of 1 repeat
African American	47B	**DYS635**	22	23	gain of 1 repeat
African American	72B	**DYS635**	22	23	gain of 1 repeat
African American	22B	DYS448	19,20	19,20	Duplication
African American	72B	DYS448	19,20	19,20	Duplication
African American	97B	DYS448	17.2,19,20	17.2,19,20	Triplication *
African American	33B	DYS389I and DYS389II			Deletion *
African American	33B	DYS439			Deletion *

Mutations in both DYS458 and DYS635 were observed in father and son 16B

Rapidly Mutating Y-STRs for Separating Male Relatives

Locus (average mutation rate)

DYS449 (1.2%)
DYS518 (1.8%)
DYS547 (2.4%)
DYS570 (1.2%)
DYS576 (1.4%)
DYS612 (1.4%)
DYS626 (1.2%)
DYS627 (1.2%)

G Model
FSIGEN-744; No. of Pages 11
ARTICLE IN PRESS

Forensic Science International: Genetics xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages

Kaye N. Ballantyne [a,1,2], Victoria Keerl [a,1,3], Andreas Wollstein [a,b], Ying Choi [a], Sofia B. Zuniga [c], Arwin Ralf [a], Mark Vermeulen [a], Peter de Knijff [c], Manfred Kayser [a,*]

[a] Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, 3000 CA Rotterdam, The Netherlands
[b] Cologne Center for Genomics, University of Cologne, D-50674 Cologne, Germany
[c] Department of Human Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands

DYF387S1 (1.6%)
DYF399S1 (7.7%)
DYF403S1 a/b (3.1/1.2%)
DYF404S1 (1.3%)
DYS526 a/b (1.3%)

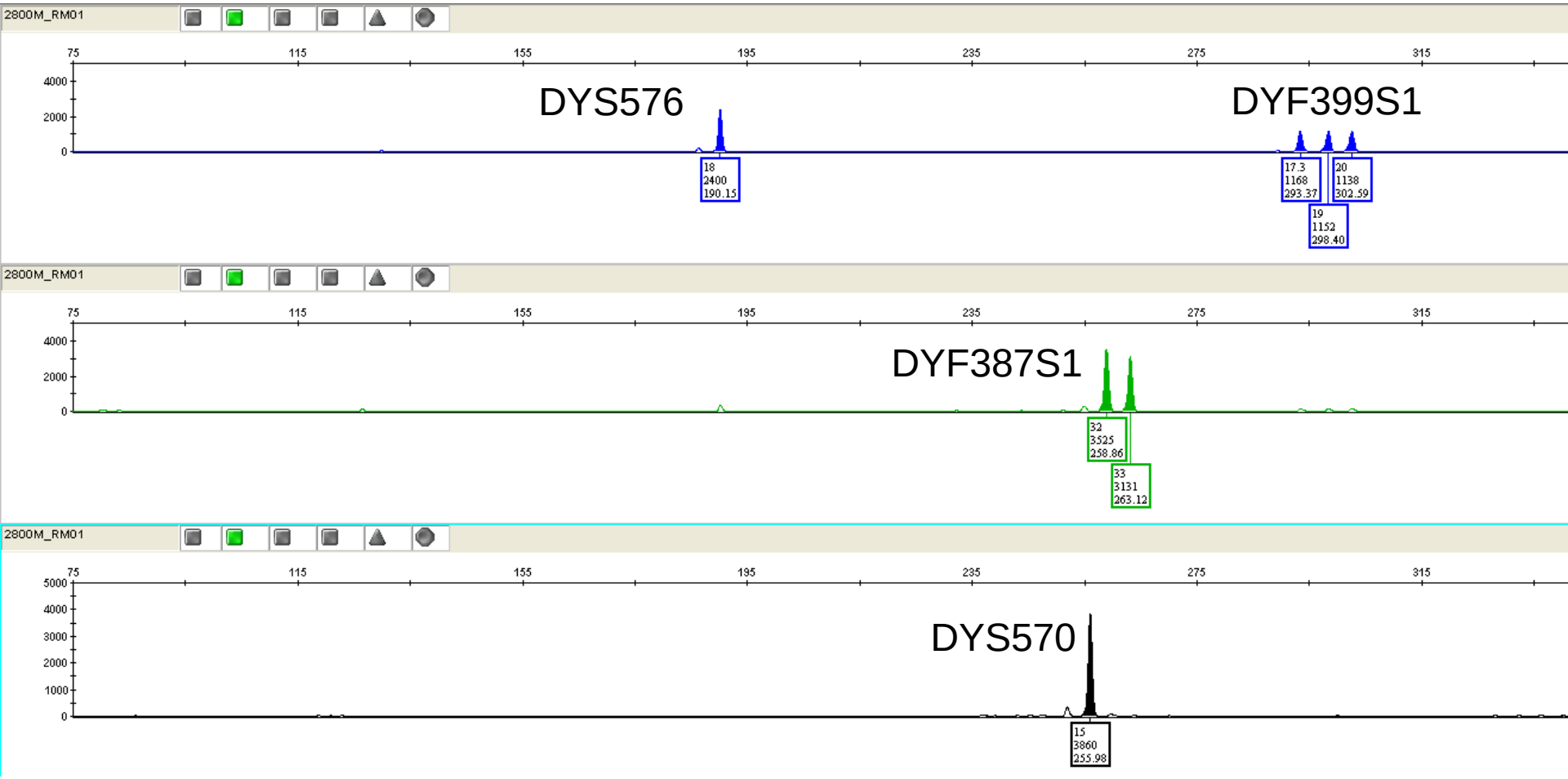
DYS458 (0.64%) is highest in Yfiler loci where average is ~0.2%

The American Journal of Human Genetics 87, 341–353, September 10, 2010 **ARTICLE**

Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications

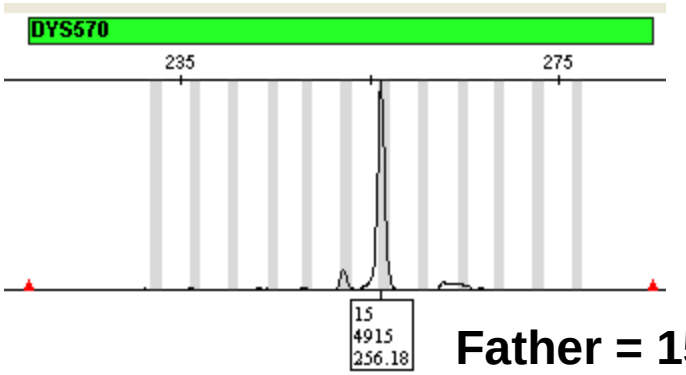
Kaye N. Ballantyne,[1] Miriam Goedbloed,[1] Rixun Fang,[2] Onno Schaap,[1] Oscar Lao,[1] Andreas Wollstein,[1,3] Ying Choi,[1] Kate van Duijn,[1] Mark Vermeulen,[1] Silke Brauer,[1,4] Ronny Decorte,[5] Micaela Poetsch,[6] Nicole von Wurmb-Schwark,[7] Peter de Knijff,[8] Damian Labuda,[9] Hélène Vézina,[10] Hans Knoblauch,[11] Rüdiger Lessig,[12] Lutz Roewer,[13] Rafal Ploski,[14] Tadeusz Dobosz,[15] Lotte Henke,[16] Jürgen Henke,[16] Manohar R. Furtado,[2] and Manfred Kayser[1,*]

RM_Multiplex01

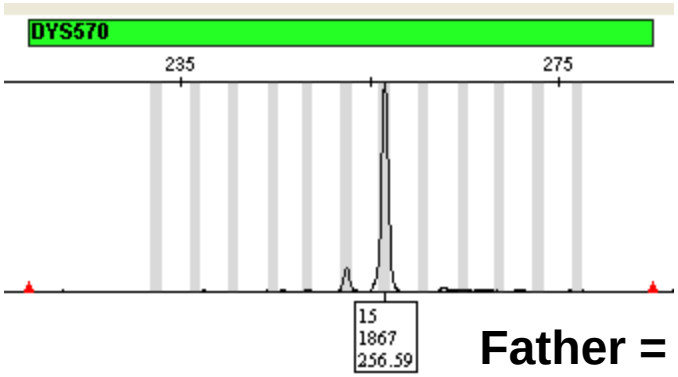


RM Y-STR Father-Son Mutations

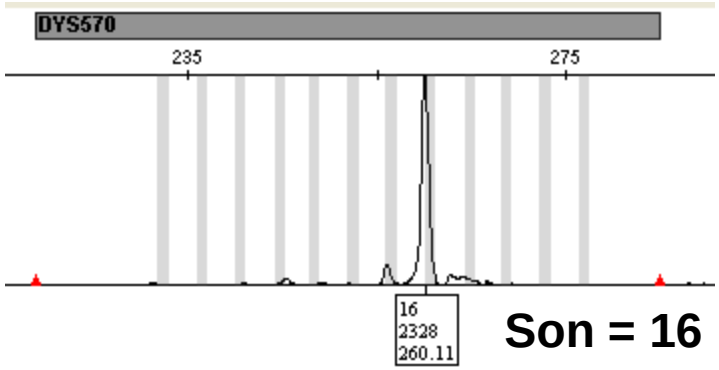
DYS570 is present in PowerPlex Y23



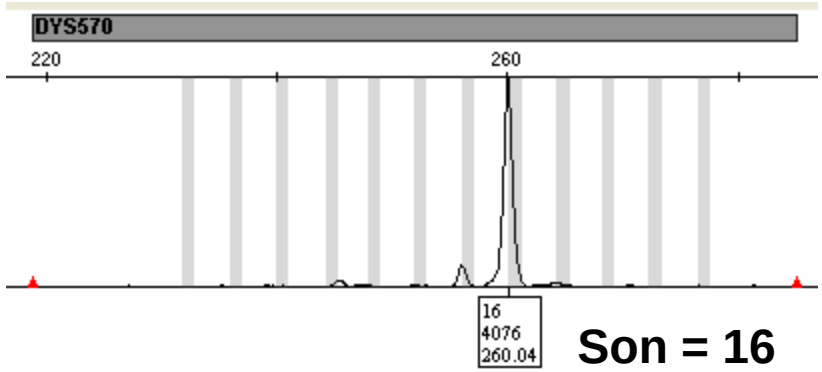
Father = 15



Father = 15



Son = 16



Son = 16

NIST Father/Son
AA sample 79

NIST Father/Son
AA sample 88

Statistical Calculations

Statistical Calculations on Y-STR Data

- **Locus (gene) Diversity** = $\mathbf{(n/n-1)(1 - \Sigma p_i^2)}$ where n is the number of samples in the dataset and p_i is the frequency of the i^{th} allele

- **Haplotype Diversity** $\mathbf{(HD) = (n/n-1)(1 - \Sigma p_i^2)}$ where n is the number of samples in the dataset and p_i is the frequency of the i^{th} haplotype

- **Random Match Probability** (RMP) = 1 - HD

- **Discrimination Capacity** (DC) – total number of observed haplotypes divided by the total number of individuals in the dataset

- **Unique Haplotypes** (UH) – number of haplotypes that occur only once in the dataset

Calculating Gene (STR) Diversity

Locus	Allele	Size Range (bp)	Count	Combined Freq (N = 661)
DYS463	17	222.45	1	0.0015
	18	227.34-227.44	27	0.0408
	19	232.30-232.39	7	0.0106
	20	237.24-237.44	151	0.2284
	21	242.21-242.41	67	0.1014
	22	247.12-247.40	74	0.1120
	23	252.13-252.33	35	0.0530
	24	257.05-257.49	256	**0.3873**
	25	262.01-262.26	37	0.0560
	26	267.05-267.21	5	0.0076
	28	277.22	1	0.0015
		failure	2	
				STR diversity
		TOTAL	**661**	**0.7684**

$$D = (n/n-1)(1 - \sum x^2)$$

Haplotype Diversity

- is a measure of the uniqueness of a particular haplotype in a given population

$$H = \frac{N}{N-1}\left(1 - \sum_i x_i^2\right)$$

Relative frequency

Population size

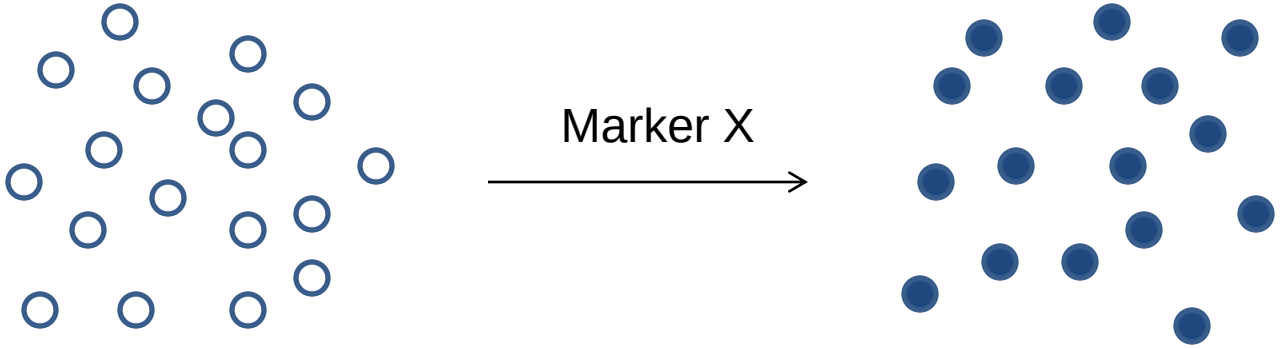
Discrimination Capacity

- is a measure of the number of unique haplotypes in a given population

$$DC = \frac{\#H}{N}$$

# of Haplotypes

Population size



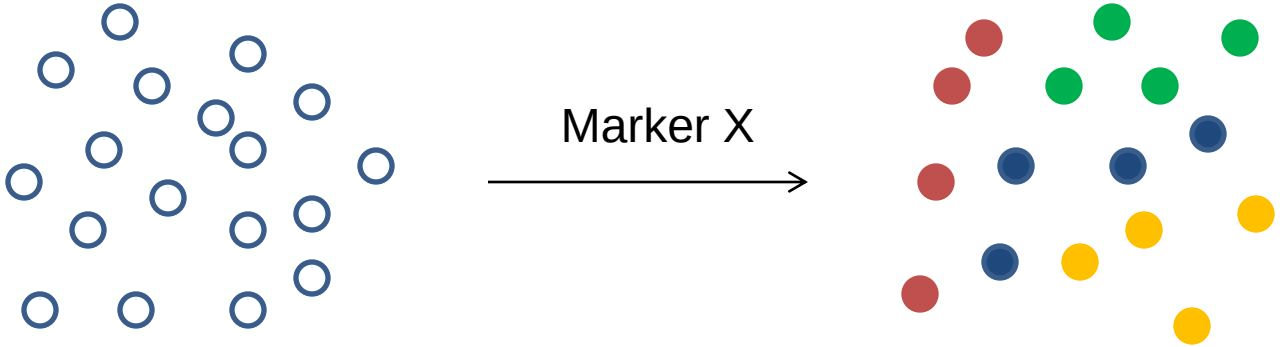
1 type = 100%

N = 100

$$H = \frac{N}{N-1}\underbrace{(1-\sum_i x_i^2)}$$

0

DC = 1/100 = 0.01



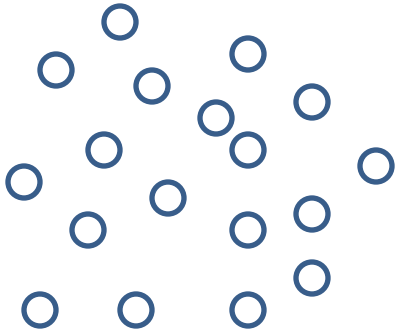
4 types = 25%

N = 100

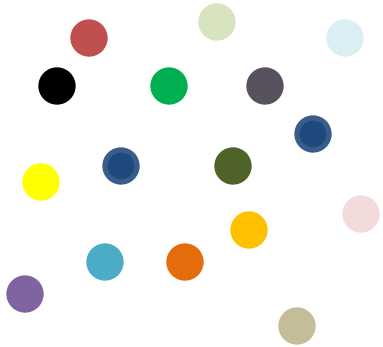
$$H = \frac{N}{N-1}\underbrace{(1-\sum_i x_i^2)}$$

0.758

DC = 4/100 = 0.04



Marker X



100 types = 0%

N = 100

$$H = \frac{N}{N-1}\underbrace{(1-\sum_i x_i^2)}$$

0.989

DC = 100/100 = 1.0

# times haplotype observed	MHL	SWGDAM	PPY	Yfiler	ALL 37
1	**429**	**486**	**505**	**626**	**652**
2	34	33	34	**12**	**2**
3	13	10	14	**2**	.
4	4	6	3	.	.
5	3	1	2	.	.
6	1	1	.	.	.
7	1	2	1	.	.
8	1	.	.	.	.
9	2	.	.	.	.
10	.	1	.	.	.
11	1	.	.	.	.
12	.	.	**1**	.	.
13	1	.	.	.	.
14	.	.	.	.	.
15	.	**1**	.	.	.
16	.	.	.	.	.
17	.	.	.	.	.
18	.	.	.	.	.
19	.	.	.	.	.
20	.	.	.	.	.
21	.	.	.	.	.
22	.	.	.	.	.
23	.	.	.	.	.
24	.	.	.	.	.
25	.	.	.	.	.
26	**1**	.	.	.	.
HD	0.996644	0.998529	0.999064	0.999916	0.999991
DC	0.748476	0.824695	0.853659	0.97561	0.996951
# HT	491	541	560	640	654

Haplotype Diversity (HD) vs. Discrimination Capacity (DC)

$$HD = (N/N-1)(1 - \sum x^2)$$

x = frequency of each haplotype

$$DC = (\#HT)/N$$

N = 656

Acknowledgments

Funding from interagency agreements 2008-IJ-R-029 (and previously 1999 & 2003) between the National Institute of Justice and the NIST Office of Law Enforcement Standards

NIST Past and Present Team Members:

Amy Decker, Richard Schoske, Christian Ruitberg, Jill Appleby,
Mike Coble, **Becky Hill**, Margaret Kline, Peter Vallone, Dave Duewer

Past Collaborators:

Mike Hammer, Alan Redd, Tom Reid,
ISFG DNA Commission, SWGDAM Y-STR Committee

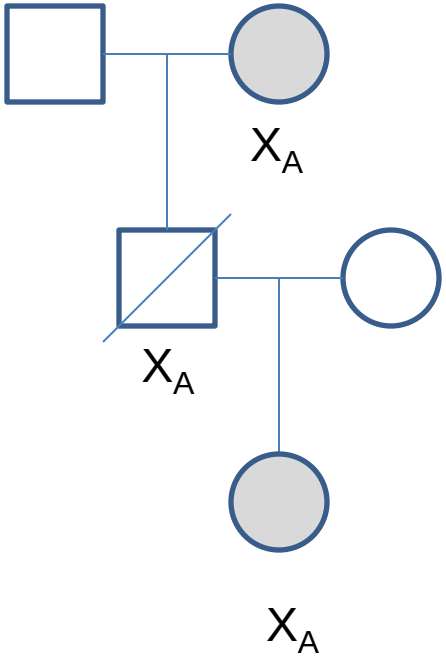
http://www.cstl.nist.gov/biotech/strbase/y_strs.htm

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

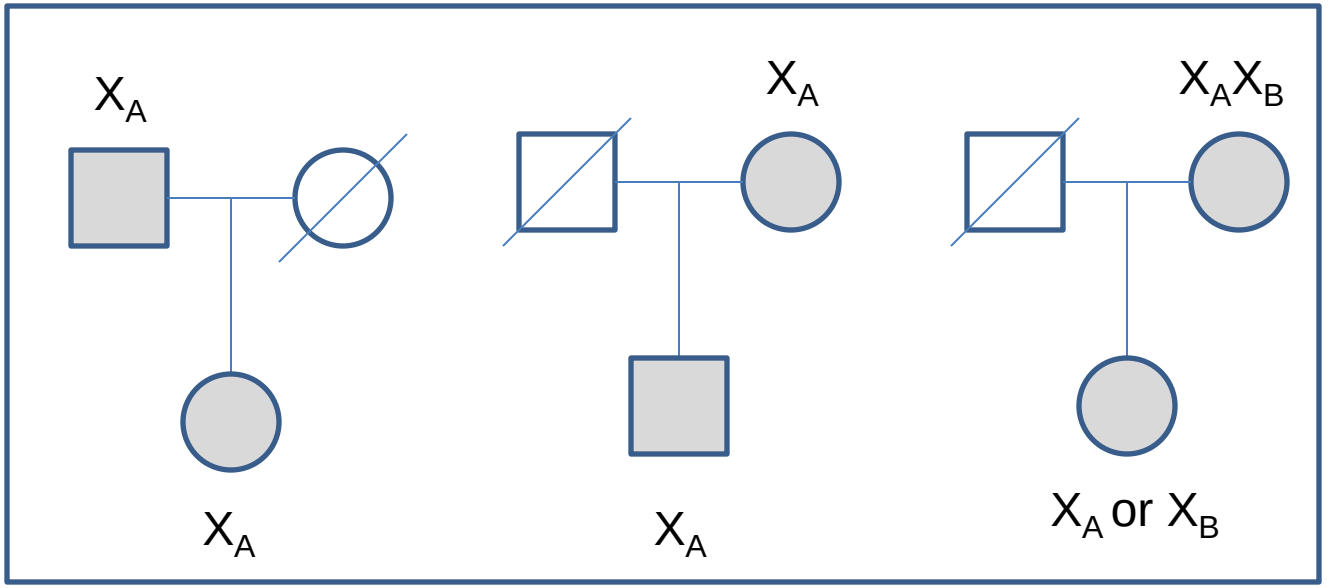
X-Chromosome Markers

Applications of X-Chromosome Analysis

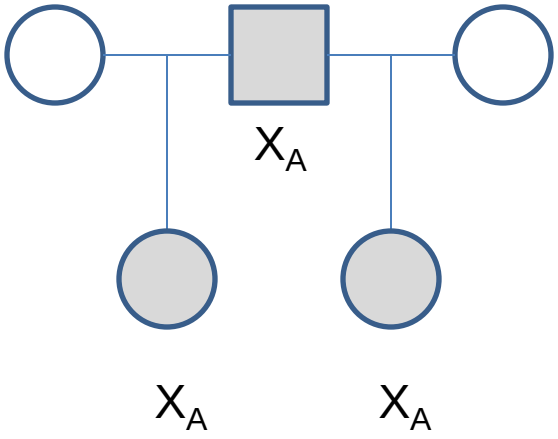
- Complex kinship cases involving at least one female
- Disputed paternity to a daughter (especially in motherless cases)
- Half-sister testing where the father is the common relative
- Grandparent—grandchild comparisons
- Paternity testing in incest cases



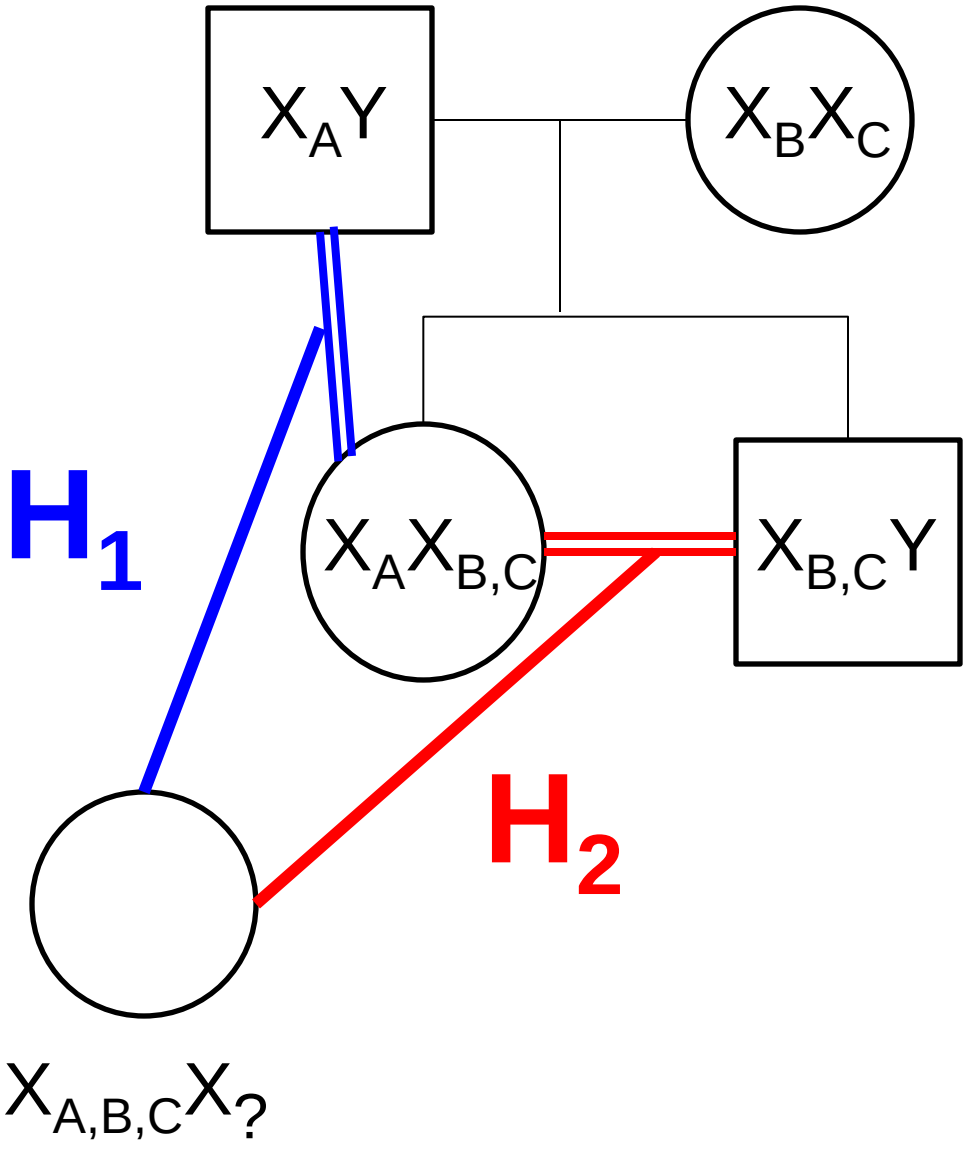
Paternal grandmother-granddaughter

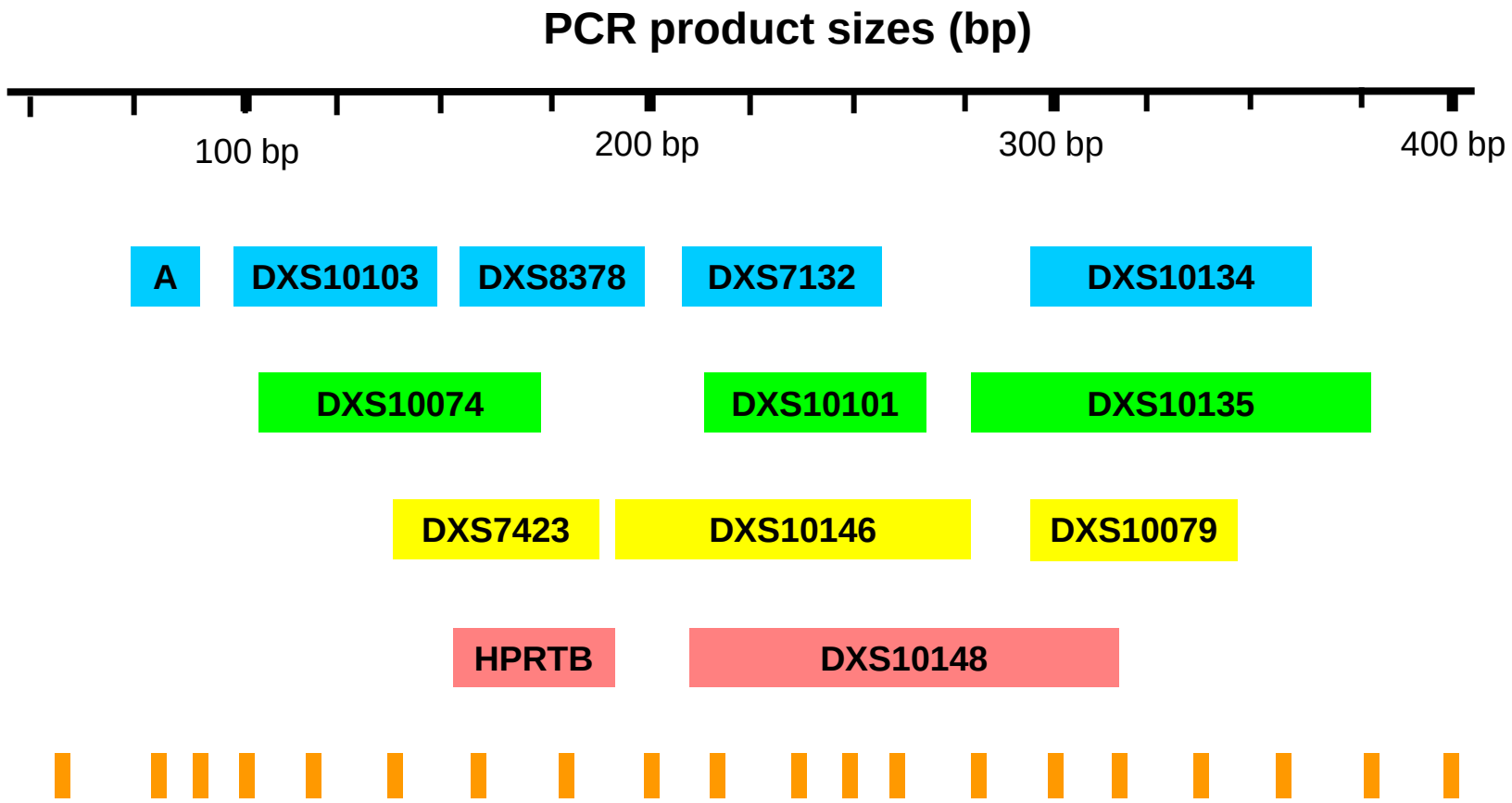


Deficiency cases



Half-sisters testing


X_AY
X_BX_C
H_1
$X_AX_{B,C}$
$X_{B,C}Y$
H_2
$X_{A,B,C}X_?$



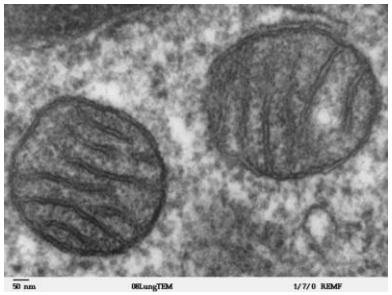
Investigator Argus X-12 Kit

X-STR Summary

- ChrX analysis has potential forensic and human identity testing applications due to its inheritance pattern compared to other genetic markers
- As with the rest of the human genome, STR markers are prevalent along the X-chromosome with comparable density to autosomal STRs
- A number of X-STR assays and kits are available
- Population studies are regularly published with X-STR data

Mitochondrial DNA (mtDNA)

Why Mitochondrial DNA?



- Mitochondria are organelles within cells
  - Produce energy via Krebs Cycle
- Separate genome from the nucleus (≈ 16,569 bp)
- Human cells have hundreds of mitochondria
- Between 2 – 10 genome copies per mitochondrion
  ≈ 1000 genome copies per cell
- A single cell's mtDNA can be amplified by PCR
  - 6 pg of DNA = 1 nuclear genome = 1000 mtDNA copies
  - When nuclear DNA fails to amplify, can often obtain mtDNA results
- In forensic samples quantity of evidence is sometimes a limitation
  - Trace evidence (hair, blood, bone)

http://remf.dartmouth.edu/images/mammalianLungTEM/source/8.html

Primary mtDNA Characteristics

- High copy number of mtDNA
- Maternal inheritance of mtDNA
- Lack of recombination
- High mutation rate compared to single copy nucDNA

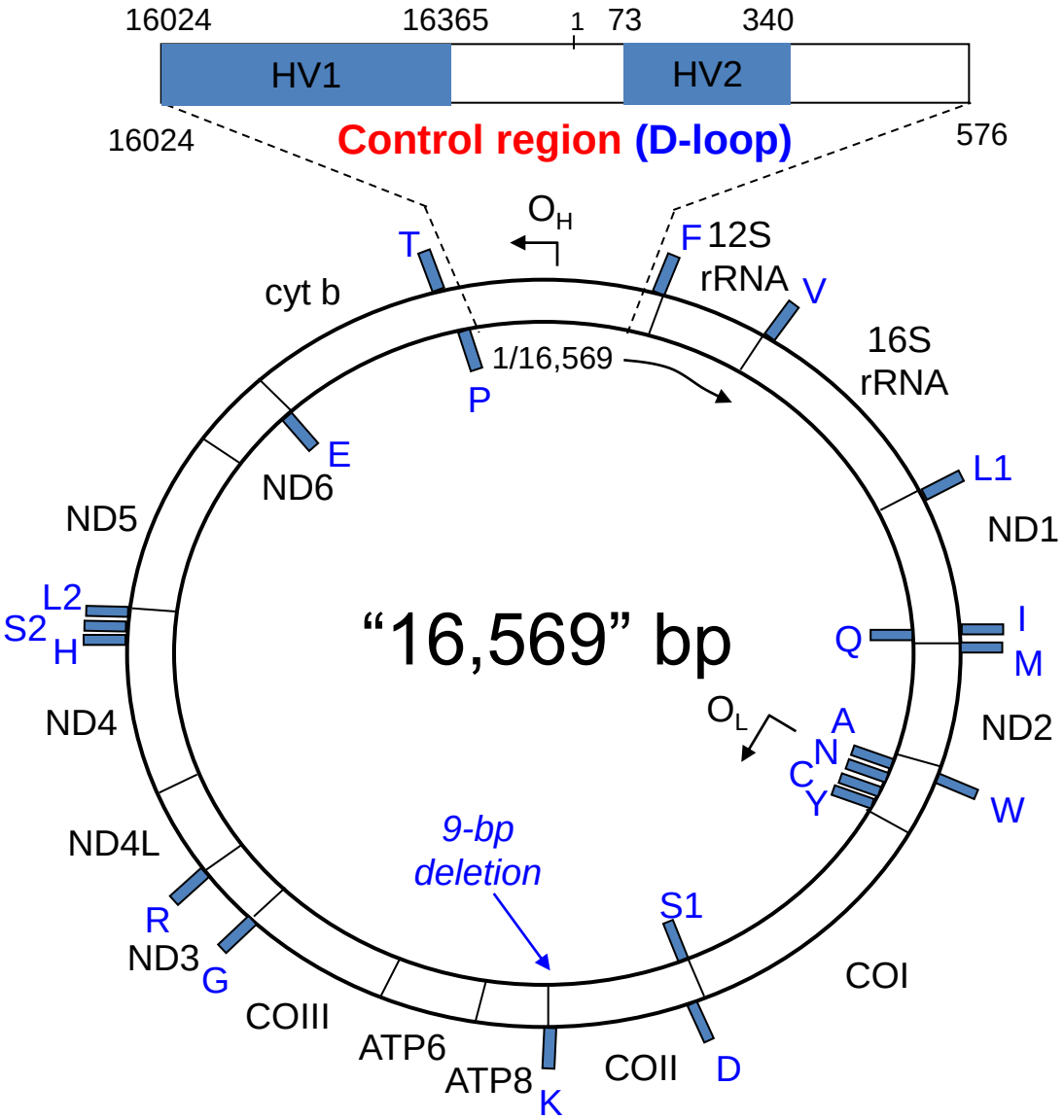
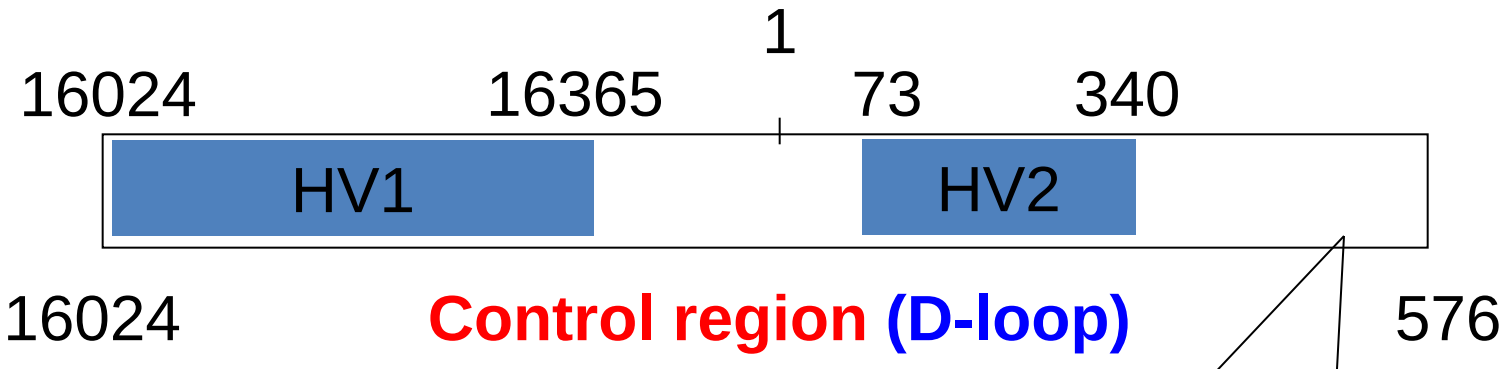


Figure 10.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Control Region (16024-576)

1,122 nucleotide positions

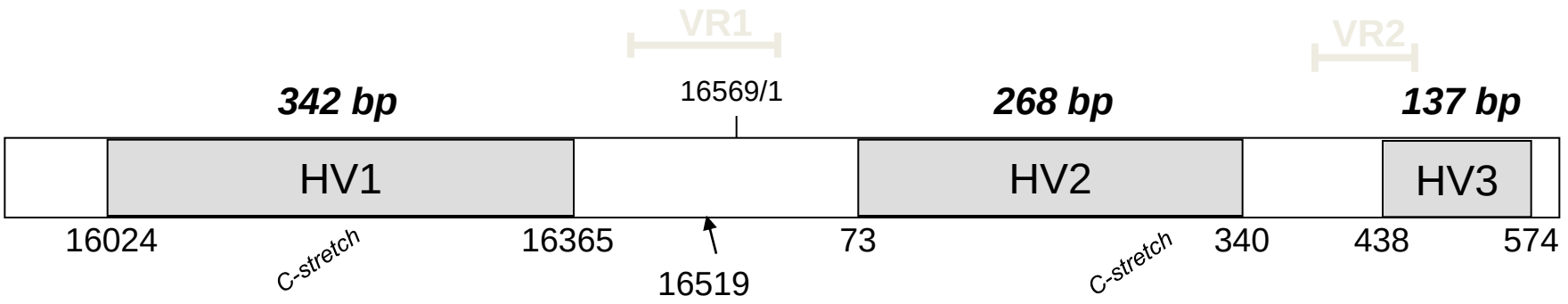


$(AC)_3$
$(AC)_4$
$(AC)_5$
$(AC)_6$
$(AC)_7$

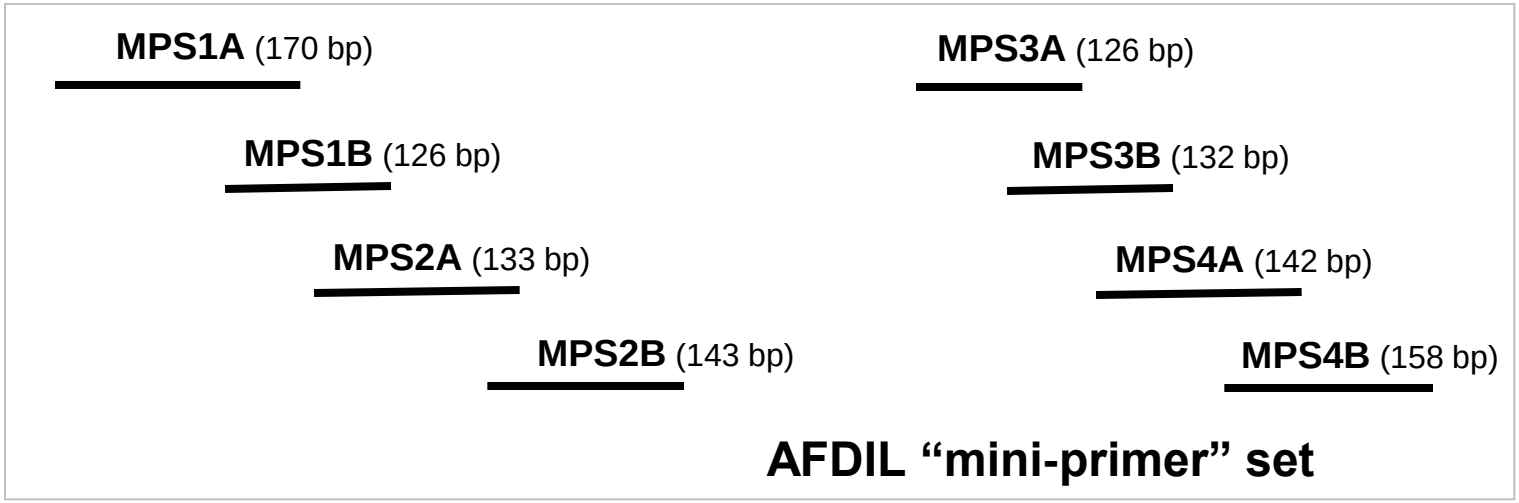
Forensic Focus

Typically only **610 bases examined**

– (HVI: 16024-16365; HVII: 73-340)

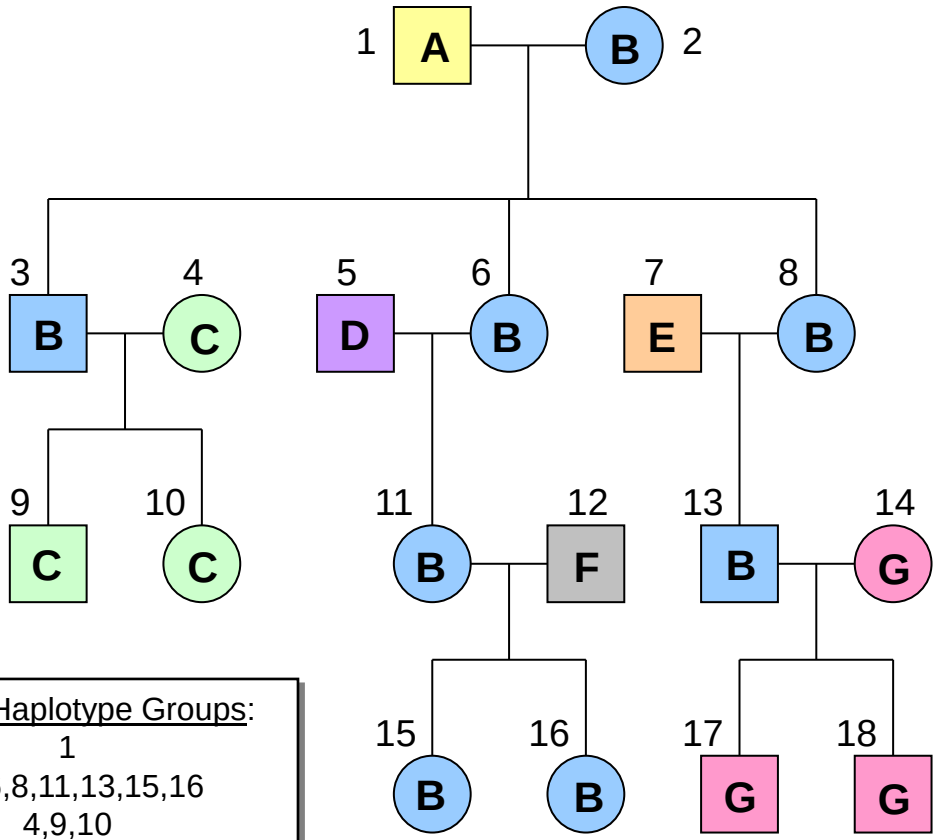

VR1
VR2
342 bp
16569/1
268 bp
137 bp
HV1
HV2
HV3
16024
C-stretch
16365
16519
73
C-stretch
340
438
574


F15989
PSI (263 bp)
R16251
F16190
PSII (221 bp)
R16410
F15
PSIII (271 bp)
R285
F155
PSIV (227 bp)
R381
AFDIL primer set


MPS1A (170 bp)
MPS1B (126 bp)
MPS2A (133 bp)
MPS2B (143 bp)
MPS3A (126 bp)
MPS3B (132 bp)
MPS4A (142 bp)
MPS4B (158 bp)
AFDIL “mini-primer” set

Maternal Inheritance of mtDNA

- Fertilizing sperm contributes only nuclear DNA
- Cellular components including the mitochondria in the cytoplasm come from the mother's ovum
- Any sperm mitochondria that may enter a fertilized egg are selectively destroyed due to a ubiquitin tag added during spermatogenesis
- Barring mutation, a mother passes her mtDNA type on to her children



MtDNA Haplotype Groups:

1

2,3,6,8,11,13,15,16

4,9,10

5

7

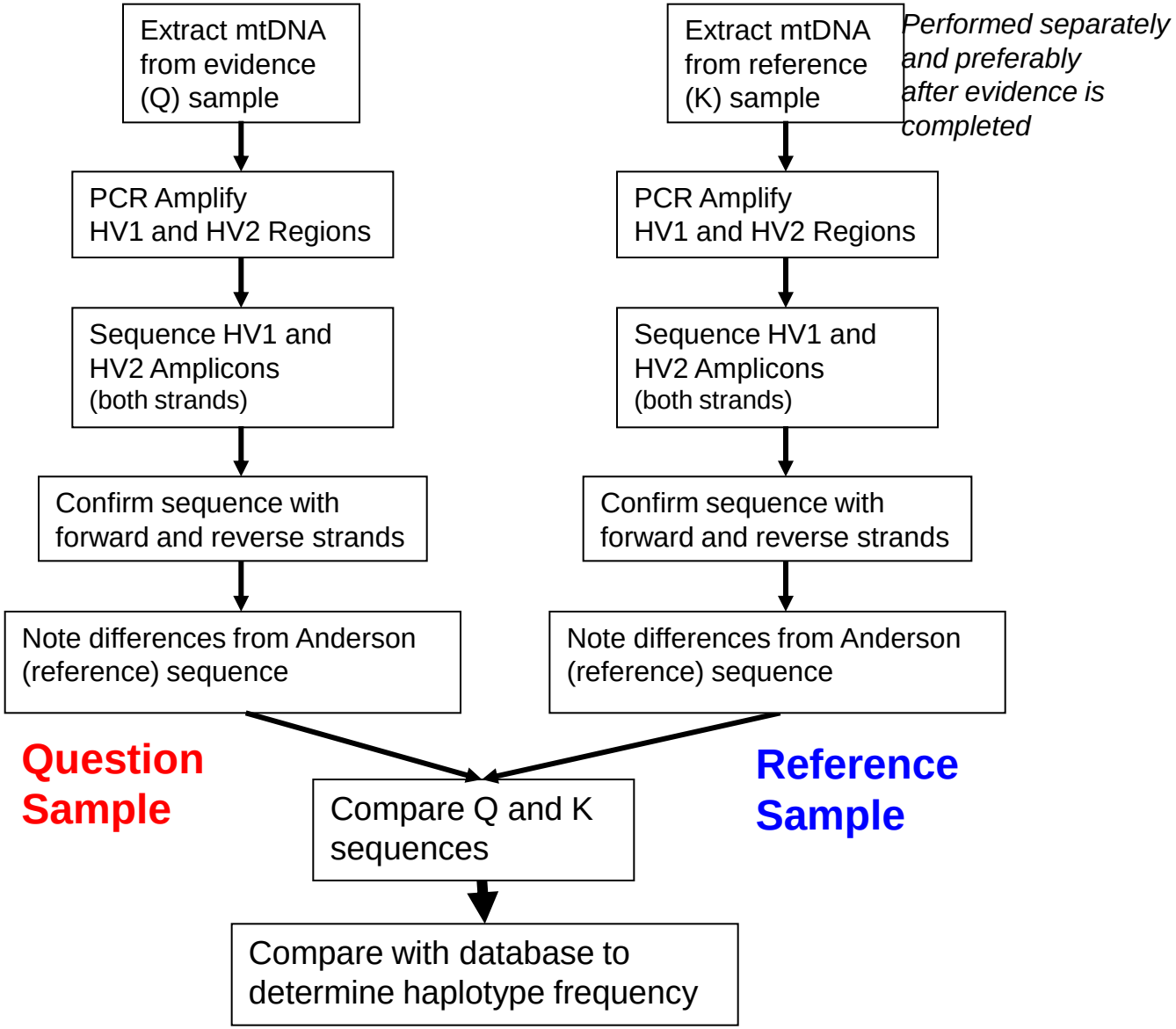
12

14,17,18

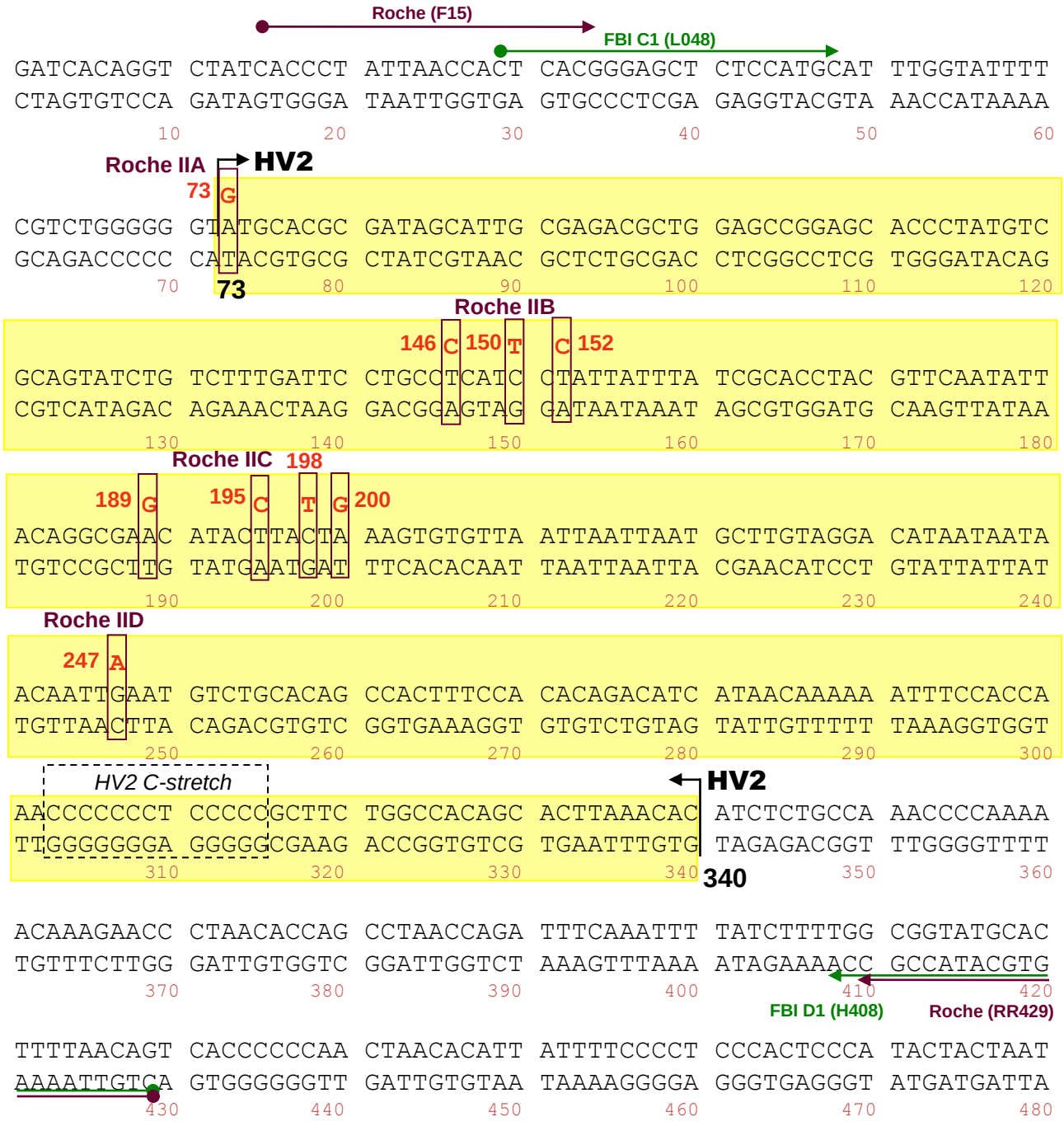
Candidates for mtDNA Testing

- Shed hairs lacking root bulb or attached tissue
- Fragments of hair shafts
- Aged bones or teeth that have been subjected to long periods of exposure
- Crime scene stains or swabs that were unsuccessful for nuclear DNA testing
- Tissues (muscle, organ, skin) that were unsuccessful for nuclear DNA testing

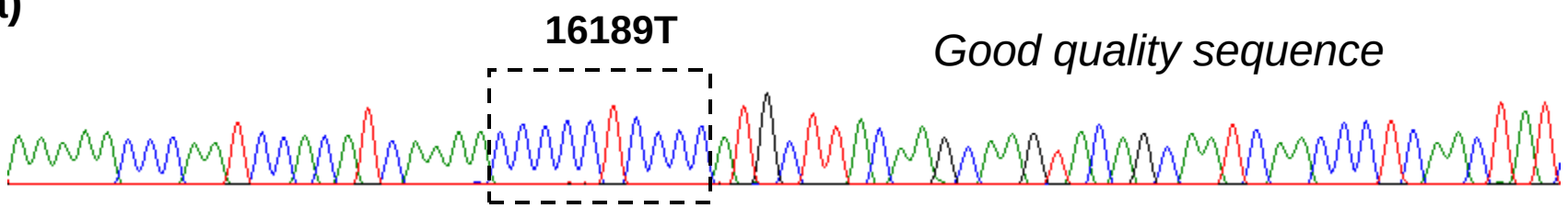
Process for Evaluation of mtDNA Samples



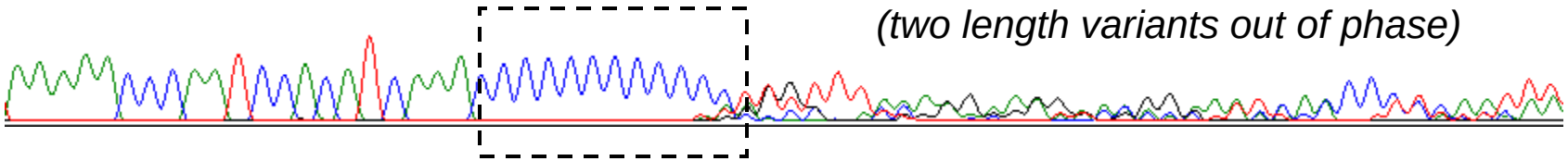

Roche (F15975)
FBI A1 (L15997)
GAAAAAGTCT TTAACTCCAC CATTAGCACC CAAAGCTAAG ATTCTAATTT AAACTATTCT
CTTTTTCAGA AATTGAGGTG GTAATCGTGG GTTTCGATTC TAAGATTAAA TTTGATAAGA
15970 15980 15990 16000 16010 16020
HV1
CTGTTCTTTC ATGGGGAAGC AGATTTGGGT ACCACCCAAG TATTGACTCA CCCATCAACA
GACAAGAAAG TACCCCTTCG TCTAAACCCA TGGTGGGTTC ATAACTGAGT GGGTAGTTGT
16024 16030 16040 16050 16060 16070 16080
Roche IA
16093 C
16126 C A 16129
ACCGCTATGT ATTTCGTACA TTACTGCCAG CCACCATGAA TATTGTACGG TACCATAAAT
TGGCGATACA TAAAGCATGT AATGACGGTC GGTGGTACTT ATAACATGCC ATGGTATTTA
16090 16100 16110 16120 16130 16140
HVI C-stretch
ACTTGACCAC CTGTAGTACA TAAAAACCCA ATCCACATCA AAACCCCCTC CCCATGCTTA
TGAACTGGTG GACATCATGT ATTTTTGGGT TAGGTGTAGT TTTGGGGGAG GGGTACGAAT
16150 16160 16170 16180 16190 16200
CAAGCAAGTA CAGCAATCAA CCCTCAACTA TCACACATCA ACTGCAACTC CAAAGCCACC
GTTCGTTCAT GTCGTTAGTT GGGAGTTGAT AGTGTGTAGT TGACGTTGAG GTTTCGGTGG
16210 16220 16230 16240 16250 16260
Roche IE
Roche IC
T T C G C
CCTCACCCAC TAGGATACCA ACAAACCTAC CCACCCTTAA CAGTACATAG TACATAAAGC
GGAGTGGGTG ATCCTATGGT TGTTTGGATG GGTGGGAATT GTCATGTATC ATGTATTTCG
16270 16280 16290 16300 16310 16320
Roche ID
C
HV1
CATTTACCGT ACATAGCACA TTACAGTCAA ATCCCTTCTC GTCCCCATGG ATGACCCCCC
GTAAATGGCA TGTATCGTGT AATGTCAGTT TAGGGAAGAG CAGGGGTACC TACTGGGGGG
16330 16340 16350 16360 16365 16370 16380
TCAGATAGGG GTCCCTTGAC CACCATCCTC CGTGAAATCA ATATCCCGCA CAAGAGTGCT
AGTCTATCCC CAGGGAACTG GTGGTAGGAG GCACTTTAGT TATAGGGCGT GTTCTCACGA
16390 16400 16410 16420 16430 16440
FBI B1 (H16391)
Roche (R16418)


Roche (F15)
FBI C1 (L048)
GATCACAGGT CTATCACCCT ATTAACCACT CACGGGAGCT CTCCATGCAT TTGGTATTTT
CTAGTGTCCA GATAGTGGGA TAATTGGTGA GTGCCCTCGA GAGGTACGTA AACCATAAAA
10 20 30 40 50 60
Roche IIA
HV2
73 G
CGTCTGGGGG GTATGCACGC GATAGCATTG CGAGACGCTG GAGCCGGAGC ACCCTATGTC
GCAGACCCCC CATACGTGCG CTATCGTAAC GCTCTGCGAC CTCGGCCTCG TGGGATACAG
70 80 90 100 110 120
73
Roche IIB
146 C 150 T C 152
GCAGTATCTG TCTTTGATTC CTGCCTCATC CTATTATTTA TCGCACCTAC GTTCAATATT
CGTCATAGAC AGAAACTAAG GACGGAGTAG GATAATAAAT AGCGTGGATG CAAGTTATAA
130 140 150 160 170 180
Roche IIC
198
189 G 195 C T G 200
ACAGGCGAAC ATACTTACTA AAGTGTGTTA ATTAATTAAT GCTTGTAGGA CATAATAATA
TGTCCGCTTG TATGAATGAT TTCACACAAT TAATTAATTA CGAACATCCT GTATTATTAT
190 200 210 220 230 240
Roche IID
247 A
ACAATTGAAT GTCTGCACAG CCACTTTCCA CACAGACATC ATAACAAAAA ATTTCCACCA
TGTTAACTTA CAGACGTGTC GGTGAAAGGT GTGTCTGTAG TATTGTTTTT TAAAGGTGGT
250 260 270 280 290 300
HV2 C-stretch
HV2
AACCCCCCCT CCCCCGCTTC TGGCCACAGC ACTTAAACAC ATCTCTGCCA AACCCCAAAA
TTGGGGGGGA GGGGGCGAAG ACCGGTGTCG TGAATTTGTG TAGAGACGGT TTGGGGTTTT
310 320 330 340 350 360
340
ACAAAGAACC CTAACACCAG CCTAACCAGA TTTCAAATTT TATCTTTTGG CGGTATGCAC
TGTTTCTTGG GATTGTGGTC GGATTGGTCT AAAGTTTAAA ATAGAAAACC GCCATACGTG
370 380 390 400 410 420
FBI D1 (H408)
Roche (RR429)
TTTTAACAGT CACCCCCCAA CTAACACATT ATTTTCCCCT CCCACTCCCA TACTACTAAT
AAAATTGTCA GTGGGGGGTT GATTGTGTAA TAAAAGGGGA GGGTGAGGGT ATGATGATTA
430 440 450 460 470 480

(a)

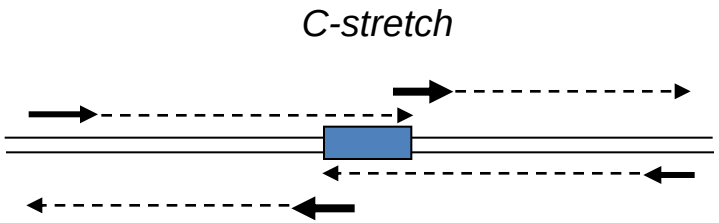


(b)

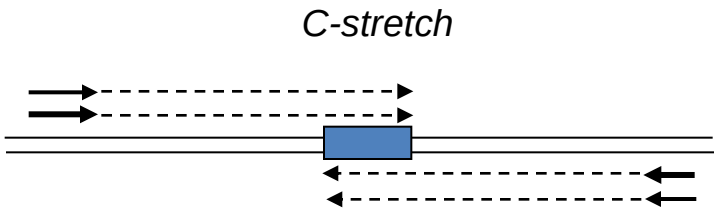


HV1 C-stretch

(c) Primer strategies typically used with C-stretch containing samples

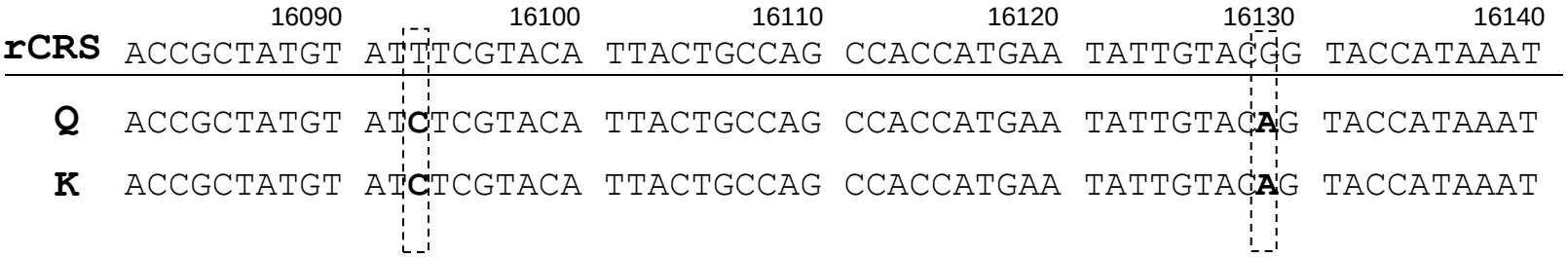


Use of internal primers



Double reactions from the same strand

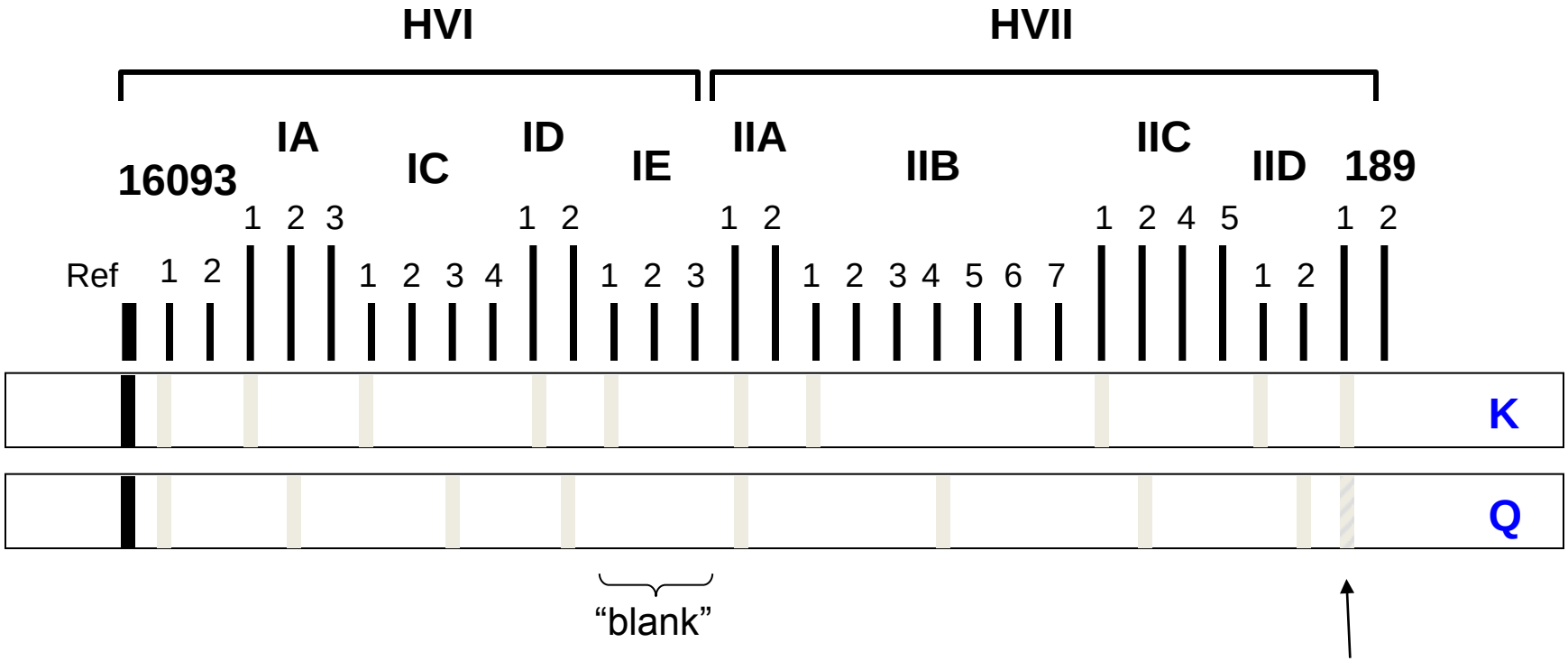
(a) mtDNA Sequences Aligned with rCRS (positions 16071-16140)



(b) Reporting Format with Differences from rCRS

Sample Q	**Sample K**
16093C	16093C
16129A	16129A

(a)



(b) Reported Types

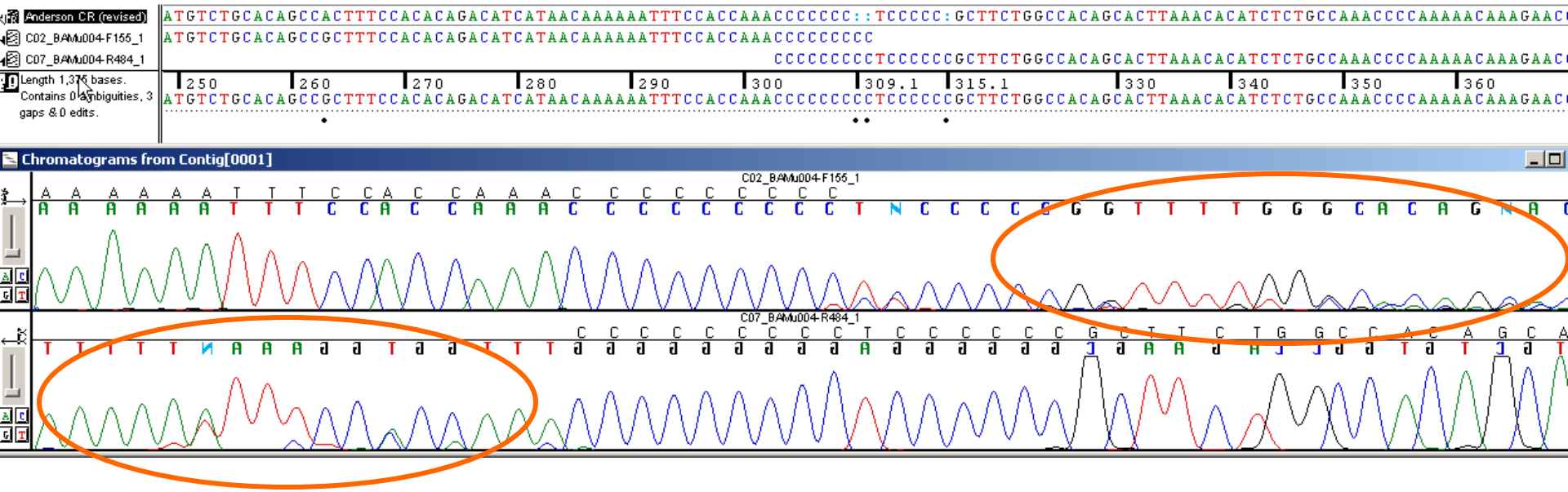
K: 1-1-1-1-1-1-1-1-1-1-1

Q: 1-2-3-2-0-1-4-2-2-*w*1

Interpretational Issues - Heteroplasmy

- Heteroplasmy – the presence of more than one mtDNA type in an individual
- Once thought to be rare, heteroplasmy exists (at some level) in all tissues
- Especially important in forensic mtDNA analysis of hair

HV2 Length Heteroplasmy



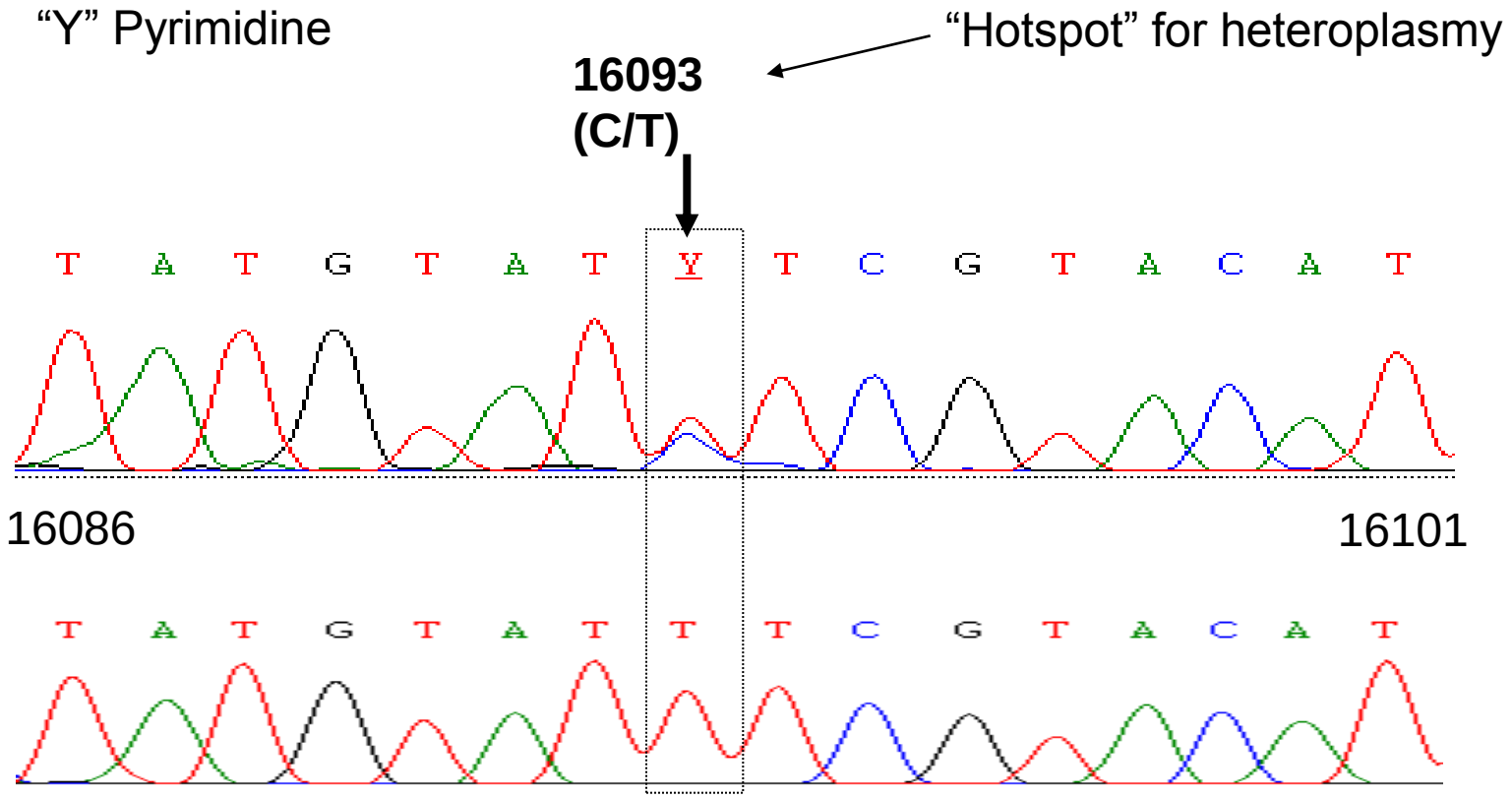
"Out of phase!"

Sequence 1 AAACCCCCCCCTCCCCCCGCTTC

Sequence 2 AAACCCCCCCCCTCCCCCCGCTTC

Sequence 3 AAACCCCCCCCCCTCCCCCCGCTTC

Point Heteroplasmy



Origination of Heteroplasmy

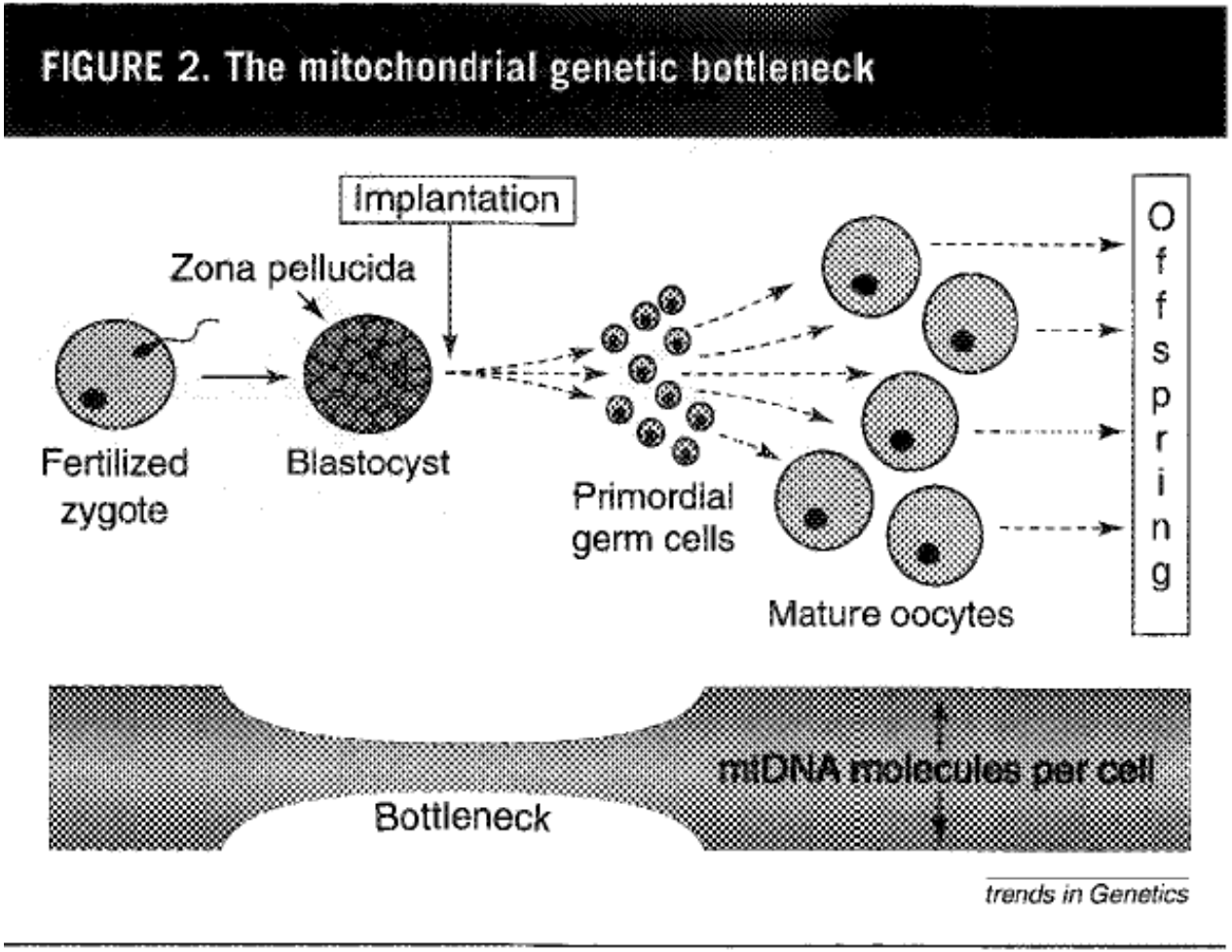
Ovum – 100K mitochondria

↓

Very little mito growth until implantation

↓

Females – produce ~7 million ova during fetal development only a few hundred become mature oocytes



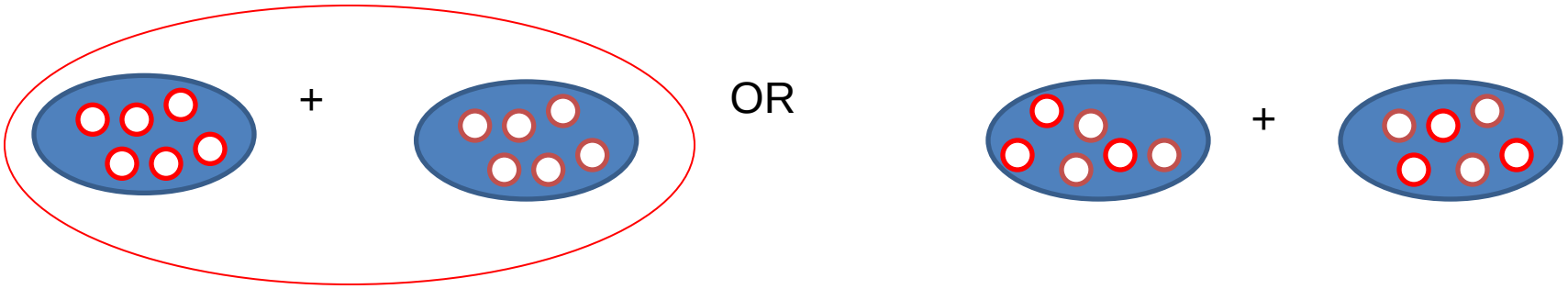
Chinnery *et al.* (2000) Trends in Genetics

Int J Legal Med (2008) 122:189–197
DOI 10.1007/s00414-007-0190-6

ORIGINAL ARTICLE

Single lymphocytes from two healthy individuals with mitochondrial point heteroplasmy are mainly homoplasmic

Sabine Lutz-Bonengel · Timo Sänger · Walther Parson · Helena Müller · Joachim W. Ellwart · Marie Follo · Bernhard Bonengel · Harald Niederstätter · Marielle Heinrich · Ulrike Schmidt



OPEN ACCESS Freely available online

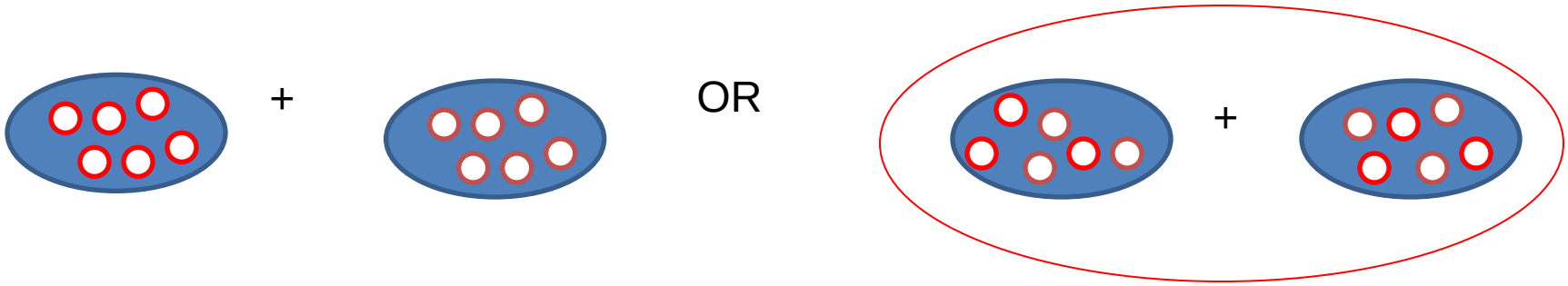
PLoS one

Detection of Heteroplasmic Mitochondrial DNA in Single Mitochondria

Joseph E. Reiner[1]*, Rani B. Kishore[1], Barbara C. Levin[2], Thomas Albanetti[3], Nicholas Boire[3], Ashley Knipe[3], Kristian Helmerson[1], Koren Holland Deckman[3]

1 Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland, United States of America, **2** Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland, United States of America, **3** Department of Chemistry, Gettysburg College, Gettysburg, Pennsylvania, United States of America

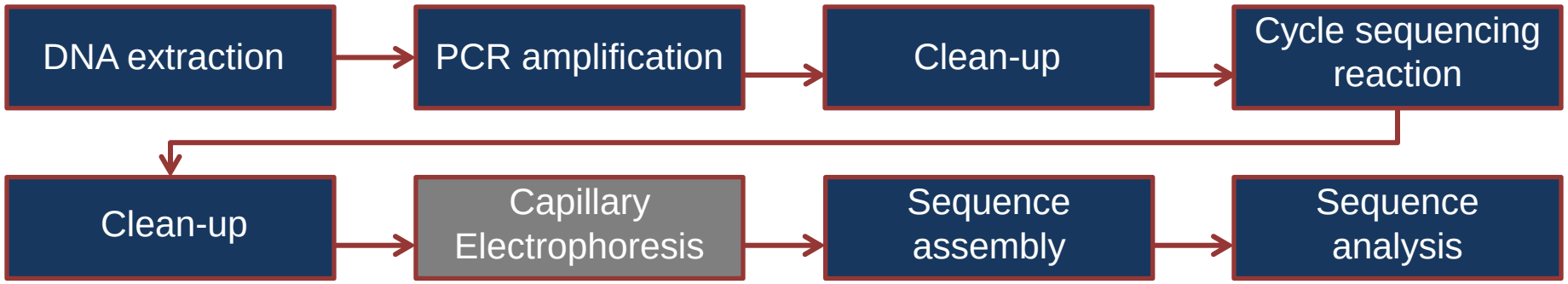
December 2010 | Volume 5 | Issue 12 | e14359



mtDNA Base Composition Analysis by Mass Spectrometry

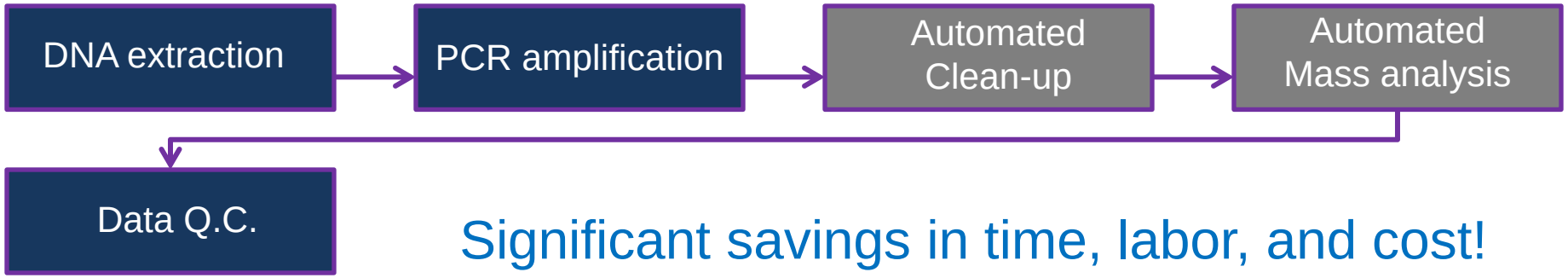
Kevin Kiesler provided slides and performed NIST work

Sequencing vs. Mass Spec



Sanger sequencing Cost: **$300** per sample Time: **10** hours

Plex-ID system Cost: **$185** per sample Time: **5** hours



Significant savings in time, labor, and cost!

Key:

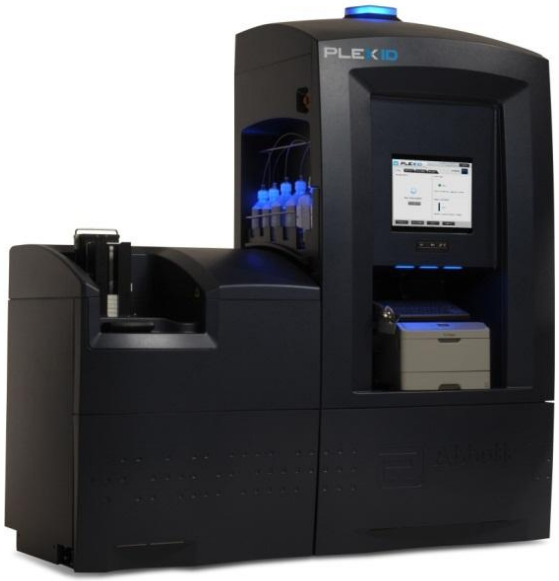
Manual step

Automated step

Base Composition by Mass Spectrometry

- Electrospray is a soft ionization method
  - Does not fragment molecules
  - Dissociates strands of DNA (PCR product)
  - 5 kV ionization negatively charges DNA (multiple charge states)
- Masses of forward and reverse strands measured
  - Time of flight analyzer
  - Mass/charge ratio (m/z) is result

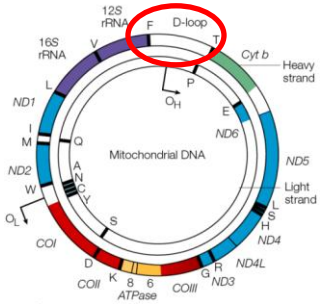
Plex-ID Instrument

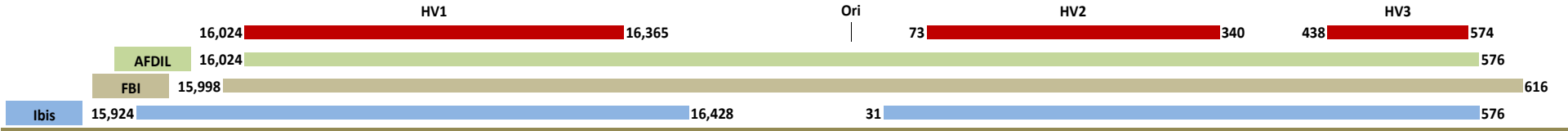


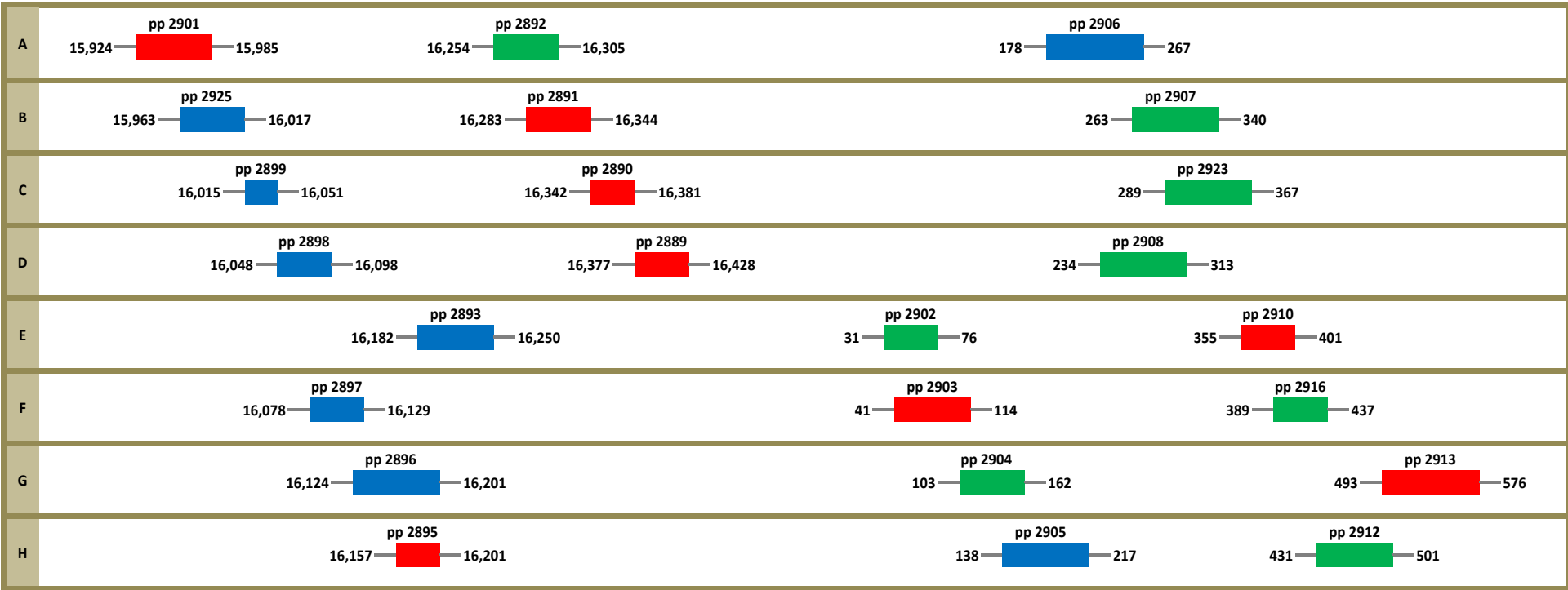
- Automated DNA cleanup
- Closed system

mtDNA 2.0 Triplex PCR Reactions

- The control region is amplified by 24 PCR primer pairs in eight triplex PCR reactions
  - Tiled over HV1, HV2, and HV3
  - Each nucleotide position is assayed at least once







mtDNA 2.0 Assay from Ibis Biosciences

	1	2	3	4	5	6	7	8	9	10	11	12
A	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892	2906 2901 2892
B	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907	2925 2891 2907
C	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923	2899 2890 2923
D	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908	2898 2889 2908
E	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902	2893 2910 2902
F	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916	2897 2903 2916
G	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904	2896 2913 2904
H	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10	Negative Control	Positive Control

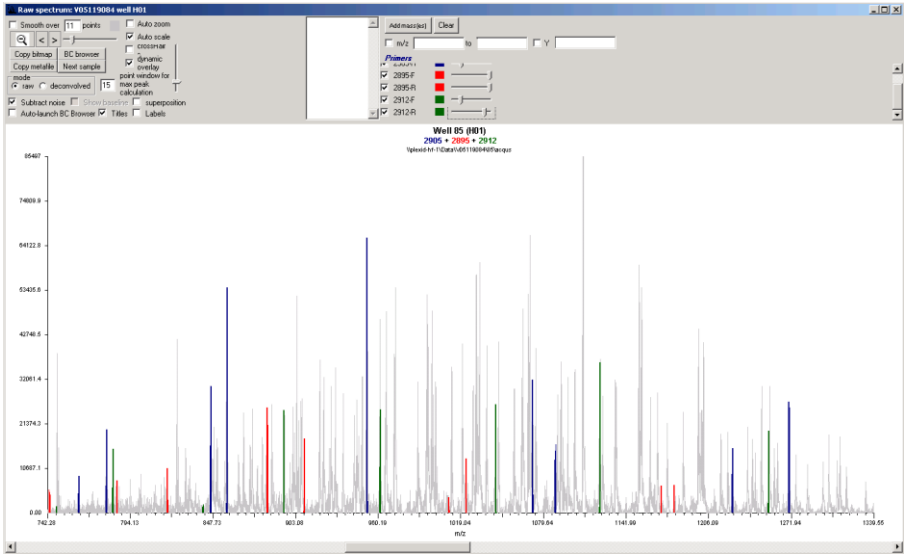
mtDNA 2.0 Assay from Ibis Biosciences

	1	2	3	4	5	6	7	8	9	10	11	12
A	2906 2901 2892	2906 2901	2906 2901	2906 2901	2906 2901	2906 2901	2906 2901	2906 2901	2906 2901	2906 2901	2906 2901	2906 2901
B	2925 2891 2907											
C	2899 2890 2923											
D	2898 2889 2908											
E	2893 2910 2902											
F	2897 2903 2916											
G	2896 2913 2904	2904	2904	2904	2904	2904	2904	2904	2904	2904	2904	2904
H	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912	2905 2895 2912
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10	Negative Control	Positive Control

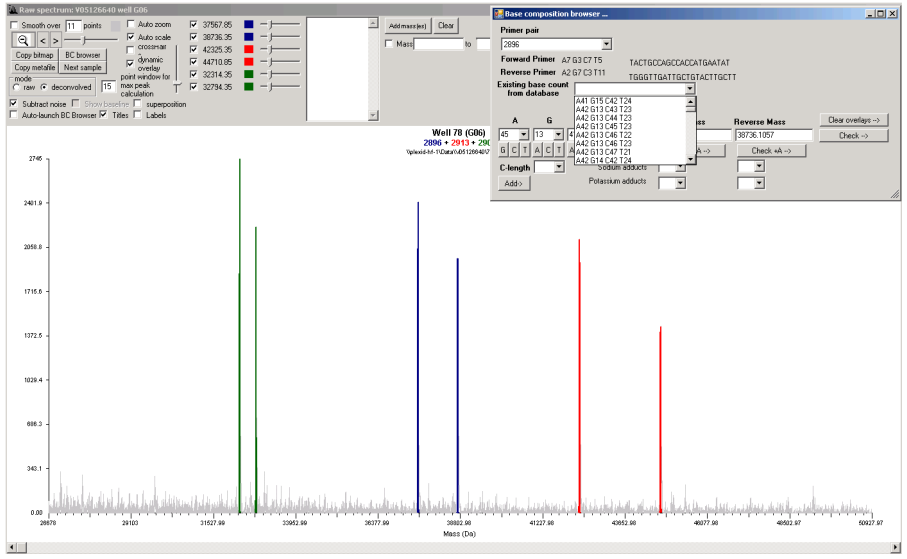
- Prefabricated 96-well plate contains all reagents for amplification
- Each well in plate contains 3 PCR primer pairs (24 amplicons total)
- For each sample, 5 µL is placed in 8 wells in a column (A - H)
- PCR is cycled off-platform then loaded onto the PlexID input stacker
- PCR products are desalted with magnetic bead-based purification
- Automated cleanup and injection is performed by the instrument
- Up to 15 plates can be processed in a single, fully automated run

Results – Mass Spectra

- Complex spectrum of multiple charge states of DNA are deconvolved into simplified spectrum

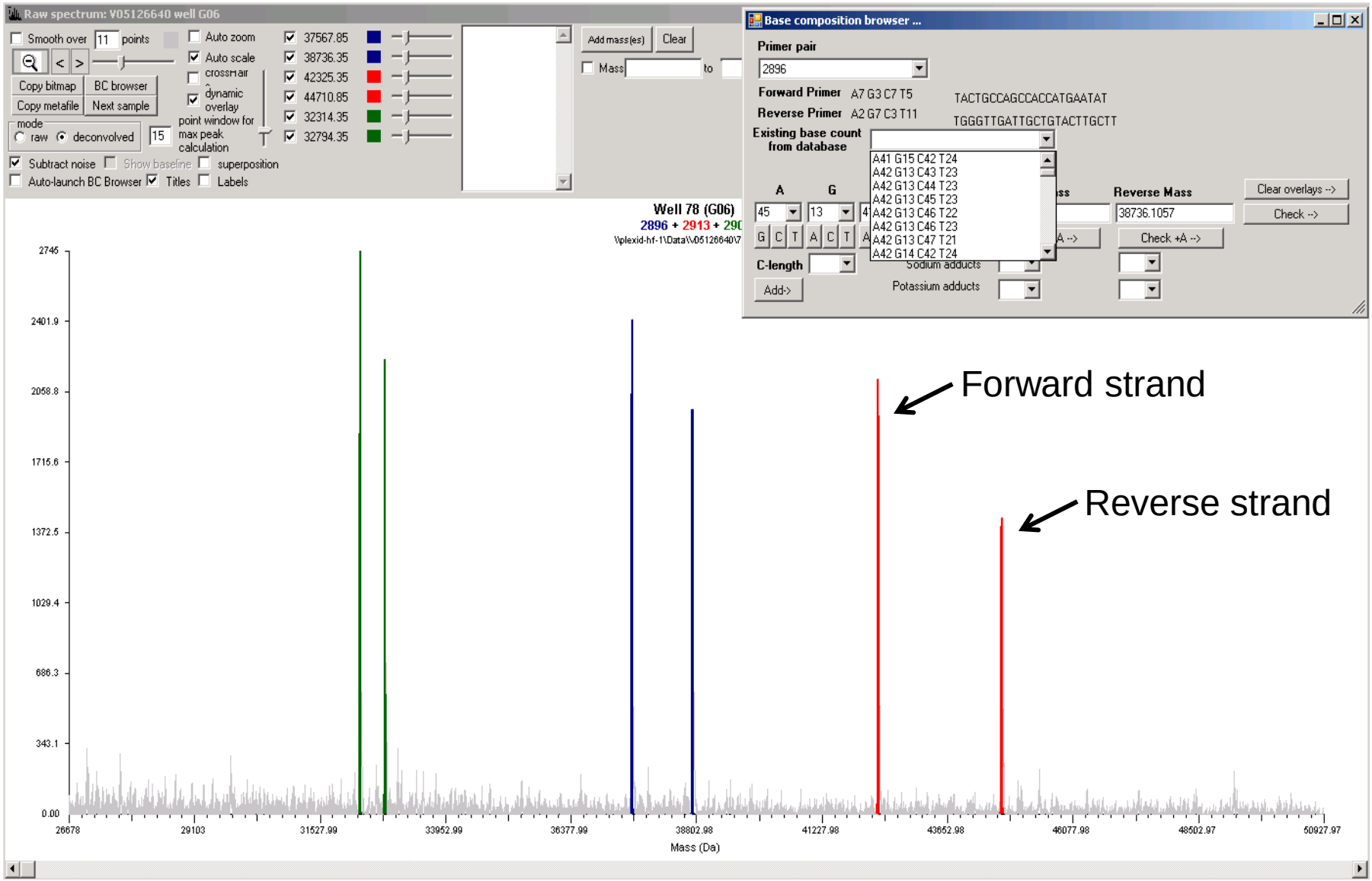


Raw Data



Processed Data

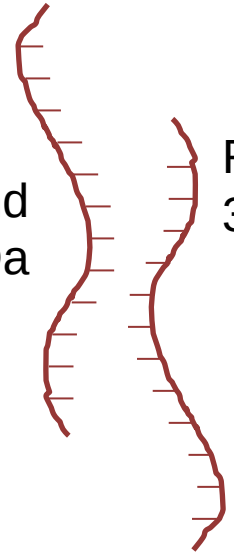
Results – Mass Spectra



Assigning Base Compositions to Mass Measurements

- Forward and reverse strands of PCR amplicons are measured independently (1)
- Watson-Crick reverse complement rules are applied to confirm that the two measured strands are from the same PCR product (2)
- Masses are correlated to a reference database of known base compositions in order to arrive at a measured base composition (3)
  - Combined base compositions of 24 amplicons are the “profile”
  - Can be used to search databases for matching profiles

1



Fwd strand
32,282 Da

Rev strand
32,744 Da

2

Verify
Reverse
Complement
A=T
G=C
C=G
T=A

3

Reference Database

A34 G14 C24 T31
A35 G14 C24 T31
A35 G14 C24 T32 = 32,282 Da
A35 G14 C25 T31
A35 G15 C24 T31
A35 G15 C25 T31

DNA Mass Spectrometry Limitation: Masses of Natural Nucleotides

- When A -> G & T -> C polymorphisms are present within an amplicon, the mass difference is +1 Dalton compared to the reference sequence
  - Cannot be differentiated by mass spec
  - **A -> G** (329.2 – 313.2 = +16 Da)
  - **T -> C** (289.2 – 304.2 = - 15 Da)
  - **tagctagctgacgatcgatgctag** mass = 7455 Da
  - **tagctagctgGcgatcgaCgctag** mass = 7456 Da

- To resolve this limitation, the assay uses a G nucleotide labeled with heavy carbon isotope, ^{13}C
  - Adds 10 Da to the mass of the nucleotide
  - Eliminates the ambiguity in combinations of nucleotide masses
  - **tagctagctgacgatcgatgctagctag** mass = **7525 Da**
  - **tagctagctgGcgatcgaCgctagctag** mass = **7536 Da**

Base Composition vs. Sequence

- Sequencing results in an ordered string of bases
  - AAGAGGTTTCACCCTGGTT
- Base composition yields an empirical formula of bases without knowing the order
  - $A_4G_5C_4T_6$
- The signature of 24 amplicons' base comp signature is almost as unique as sequence
  - Difference: cannot resolve reciprocal base changes within one amplicon
  - Example: C -> T + T -> C = no change in mass

Improved extraction protocols for mtDNA testing

Slides from Mike Coble

and work performed at AFDIL

Current Extraction Protocols – Forensic mtDNA Labs

DNA Extraction

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 1 (2007) 191–195
FSI GENETICS
www.elsevier.com/locate/fsig

Short communication

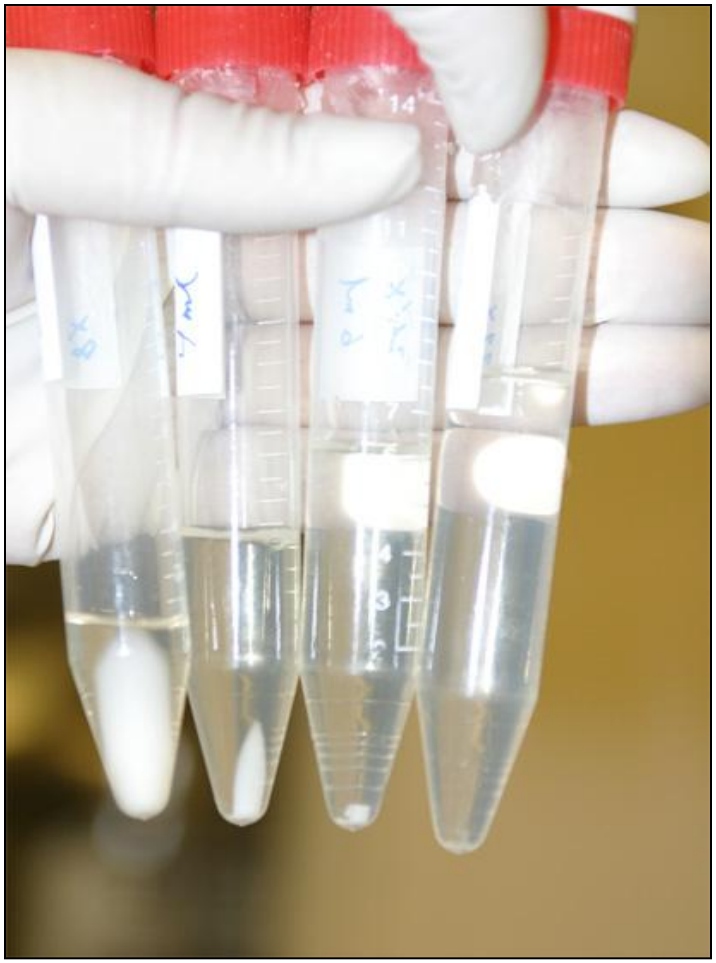
High efficiency DNA extraction from bone by total demineralization ☆

Odile M. Loreille *, Toni M. Diegoli, Jodi A. Irwin, Michael D. Coble, Thomas J. Parsons [1]

Armed Forces DNA Identification Laboratory, 1413 Research Blvd., Bldg. 101, Rockville, MD 20850, United States

Received 24 January 2007; accepted 3 February 2007

Demineralization protocol



- EDTA 0.5M, pH 8.5
- Detergent
- Proteinase K
- 1g powder

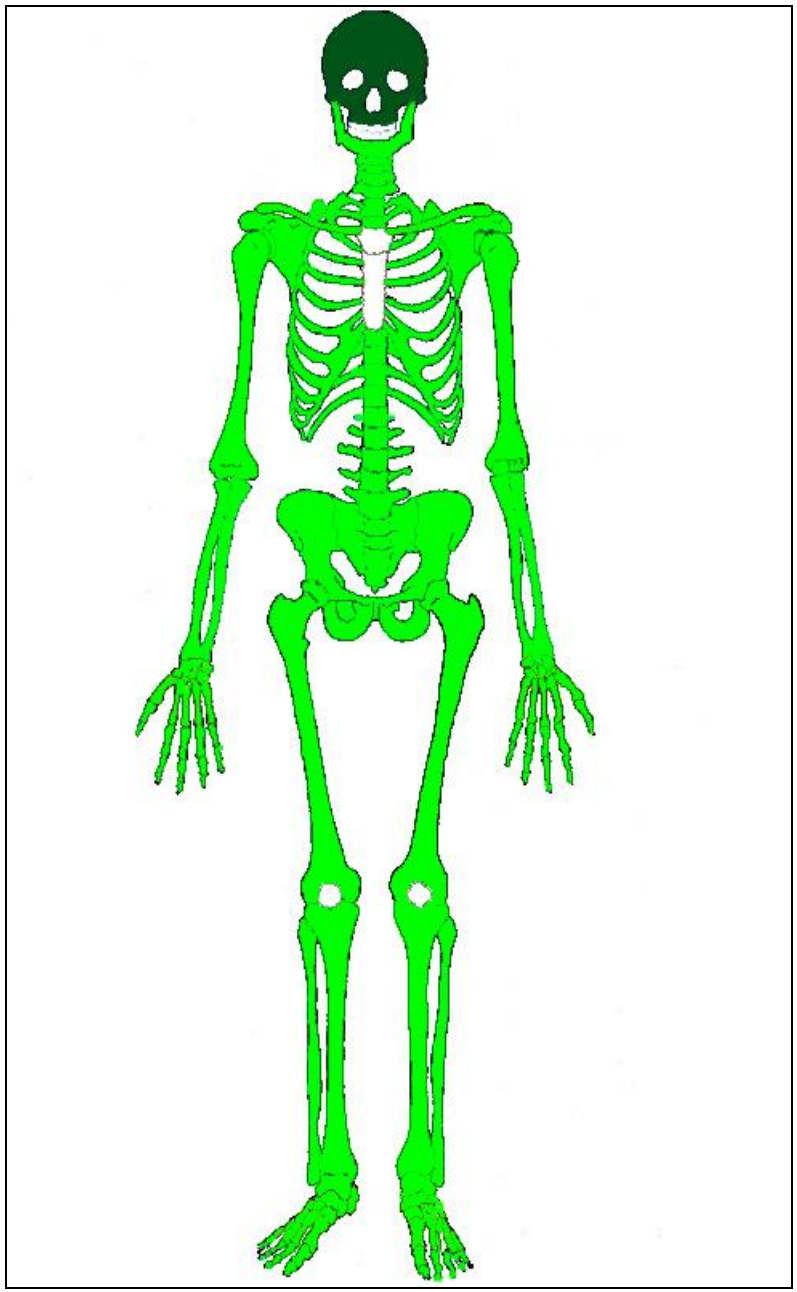
15ml extraction buffer

- Organic extraction (phenol-chloroform)
- Concentration and washes in filtration devices.

Casework SOP

Demineralization protocol

10mM Tris, pH 8.0, 100mM NaCl, 50mM EDTA, pH 8.0, 0.5% SDS; ProK



- 90-100% success
- 80-89.9% success
- 70-79.9% success
- 60-69.9% success
- 50-59.9% success
- 40-49.9% success
- 30-39.9% success
- 0-20% success

Demineralization success

736 samples processed

Courtesy of Suzie Barritt-Ross

Acknowledgments

- Thanks to Mike Coble (NIST) and Kevin Kiesler (NIST) for many of the slides

- For more information, see *Advanced Topics in Forensic DNA Typing* (2012)
  - Chapter 13 Y-Chromosome DNA Testing
  - Chapter 14 Mitochondrial DNA Analysis
  - Chapter 15 X-Chromosome Analysis