DNA & Databases 101: Types of DNA and How We Use Them

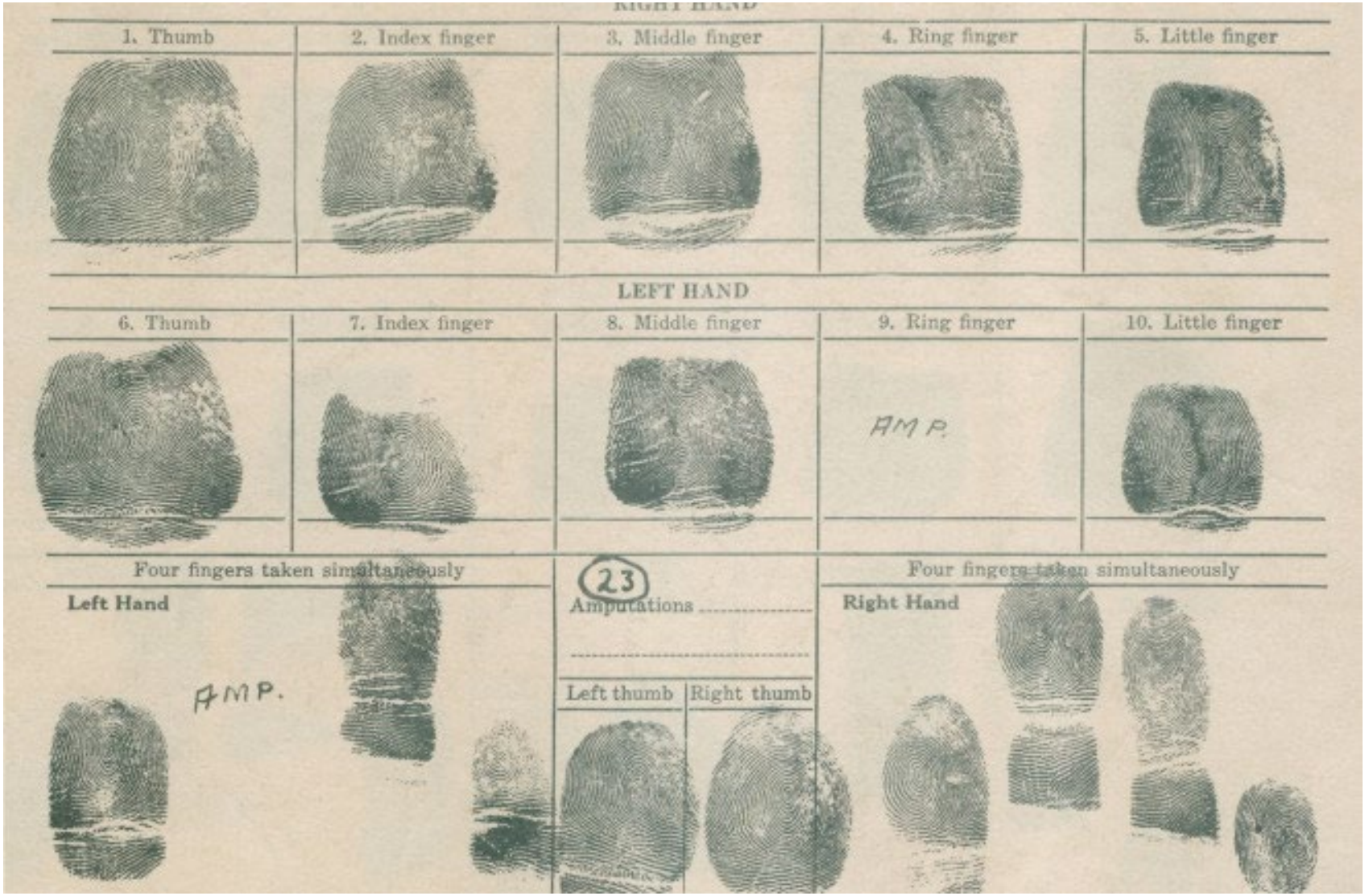
Part 1 of 6 – DNA in Court 101: An Introduction to the Tools and Technologies of the Trade

Presenters:

Dr. Michael Coble, Executive Director, University of North Texas Health Science Center, Center for Human Identification

Mindy Montford, Senior Counsel, Texas AG's Office, Cold Case and Missing Persons Unit

1910: First Fingerprints Admitted as Evidence in Criminal Trial in U.S.



People v. Jennings, 252 Ill. 534 (1911).

- In his opinion, Chief Justice Orrin Carter noted that there was no "case in which this question has been raised" and "we find no statutes or decisions touching the point in this country."

- *We are disposed to hold from the evidence of the four witnesses who testified and from the writings we have referred to on this subject, that there is a scientific basis for the system of finger-print identification and that the courts are justified in admitting this class of evidence; that this method of identification is in such general and common use that the courts cannot refuse to take judicial cognizance of it.* [People v Jennings 1911, pp 9–10]

1985: DNA Profiling Used for the First Time in Criminal Case

Alec Jeffreys British Geneticist

Sir Alec Jeffreys (2019)

1987: First Person Convicted Using DNA Evidence in the United States

Andrews v. Florida (on appeal), 1988

- *"We have found no other appellate decision addressing the admissibility of DNA identification evidence in criminal cases."*

- *The trial court did not abuse its discretion in ruling the test results admissible in this case. In contrast to evidence derived from hypnosis, truth serum and polygraph, evidence derived from DNA print identification appears based on proven scientific principles. Indeed, there was testimony that such evidence has been used to exonerate those suspected of criminal activity. Given the evidence in this case that the test was administered in conformity with accepted scientific procedures so as to ensure to the greatest degree possible a reliable result, appellant has failed to show error on this point. (*533 So.2d 841,13 Fla. L. Weekly 2364)

People v. Jennings, 252 Ill. 534 (1911).

- In his opinion, Chief Justice Orrin Carter noted that there was no "case in which this question has been raised" and "we find no statutes or decisions touching the point in this country."

- *We are disposed to hold from the evidence of the four witnesses who testified and from the writings we have referred to on this subject, that there is a scientific basis for the system of finger-print identification and that the courts are justified in admitting this class of evidence; that this method of identification is in such general and common use that the courts cannot refuse to take judicial cognizance of it.* [People v Jennings 1911, pp 9–10]

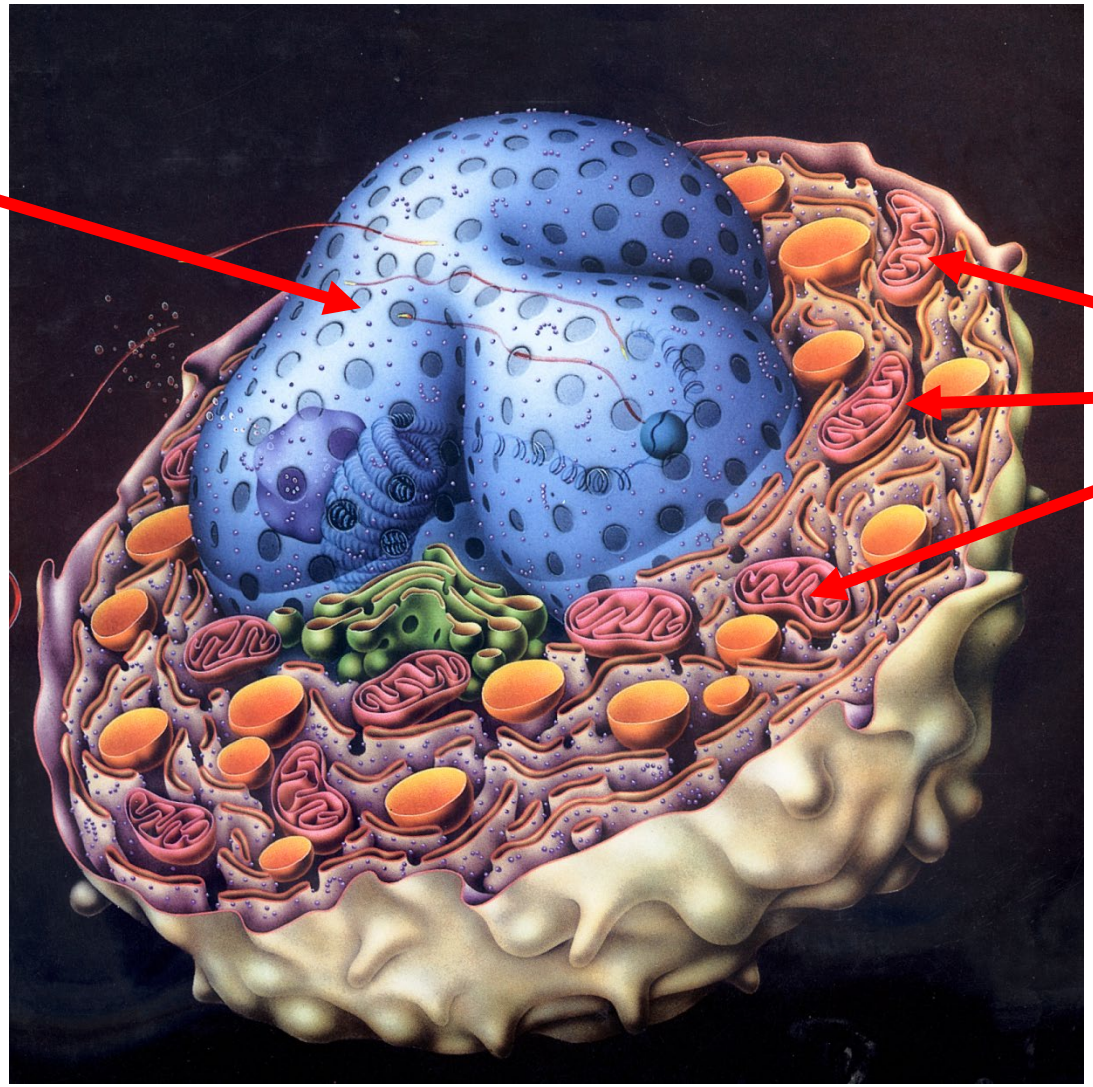
The Human DNA Genome

Nuclear DNA

~3.2 billion bp

High Power of Discrimination

½ of your chromosomes come from each parent



Mitochondrial DNA

16,569 bp

High Copy #

Genetic Variation

GATAGATAGATAGATAGATA

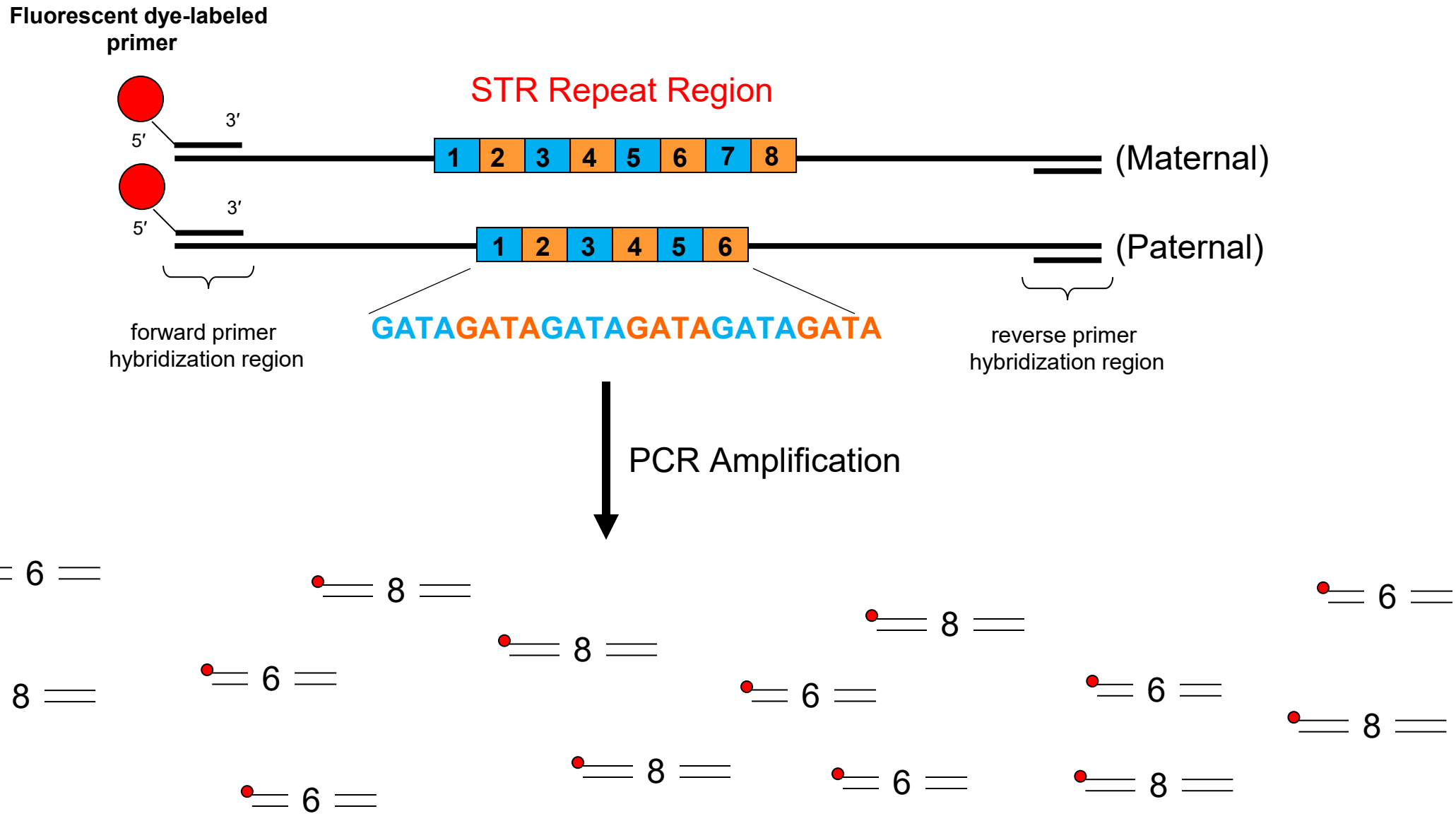
Short Tandem Repeats
(STRs)
Length Based Variation

AGTTCAGTTCCATGG

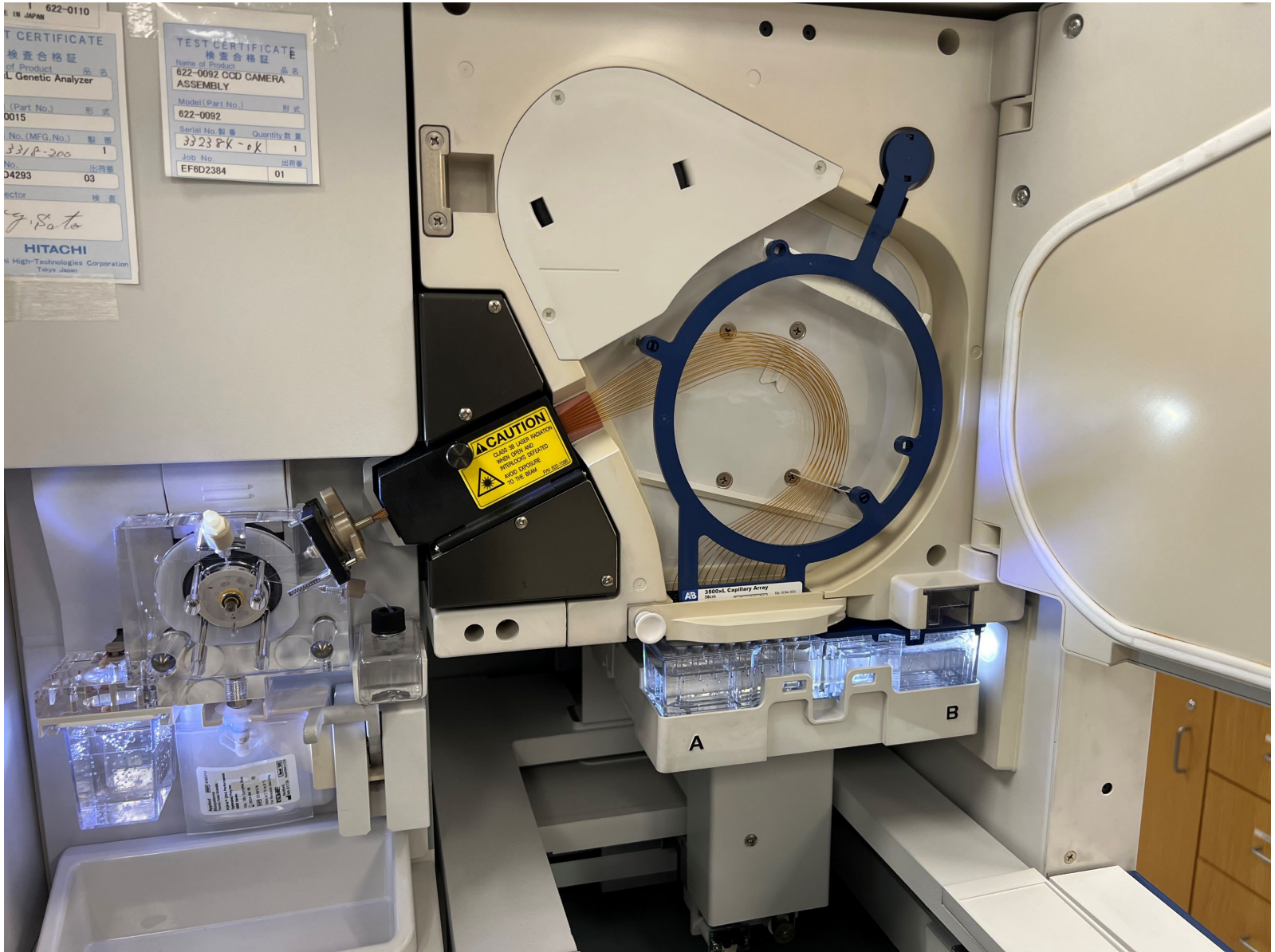
Single Nucleotide Polymorphisms
(SNPs)
Sequence Based Variation

nuclear DNA (STR) Testing

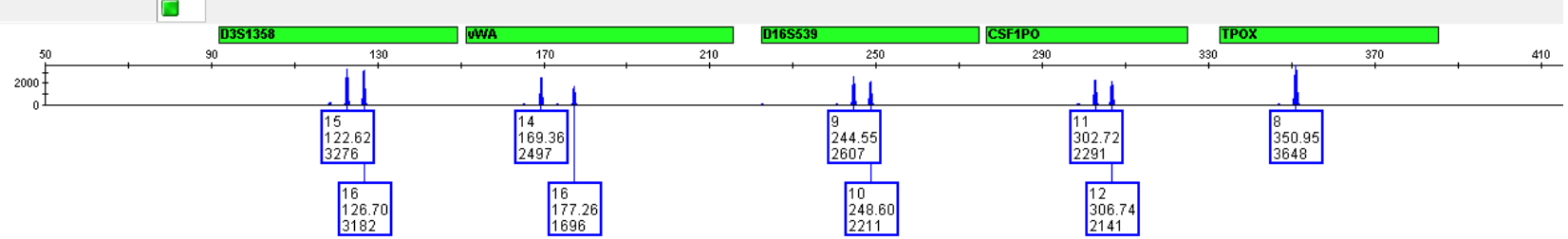
Short Tandem Repeat (STR) Typing

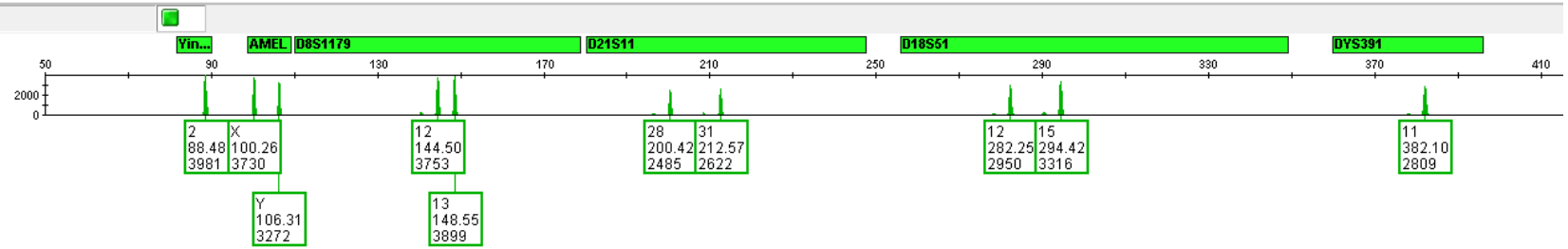


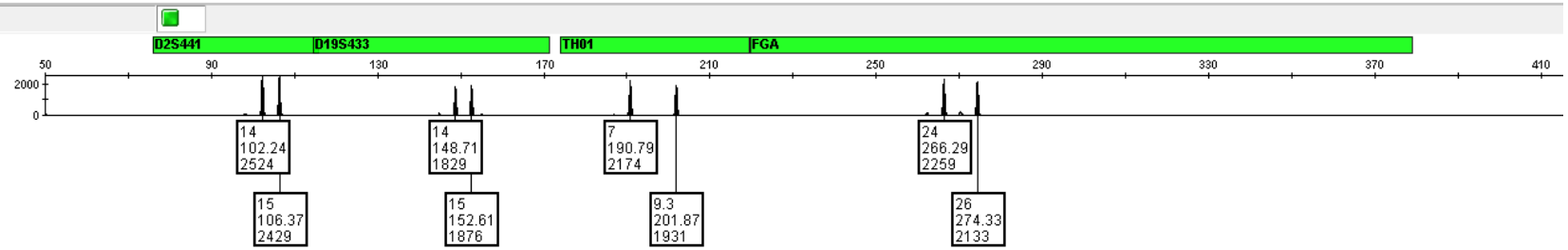
Capillary Electrophoresis

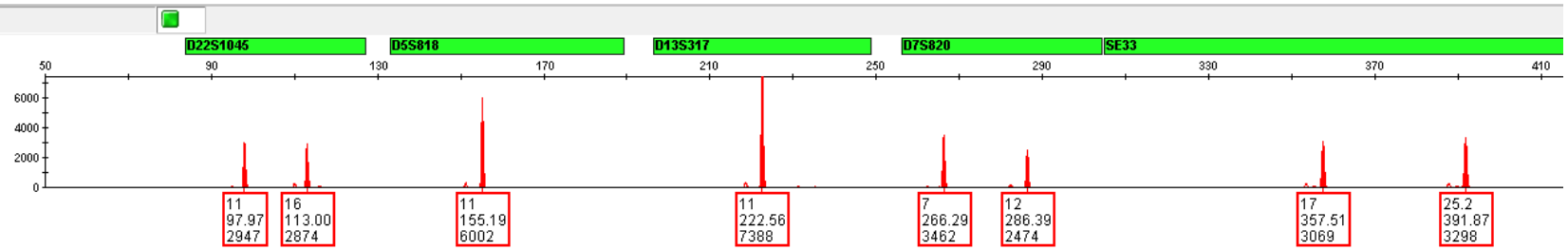


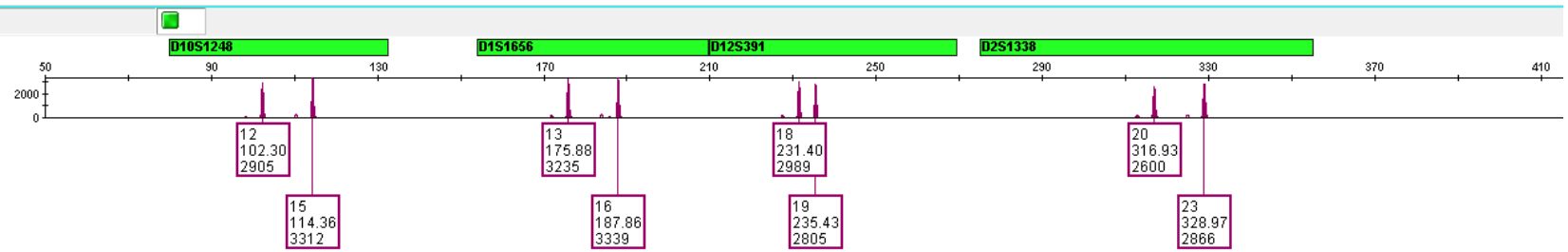
Forensic DNA Profile



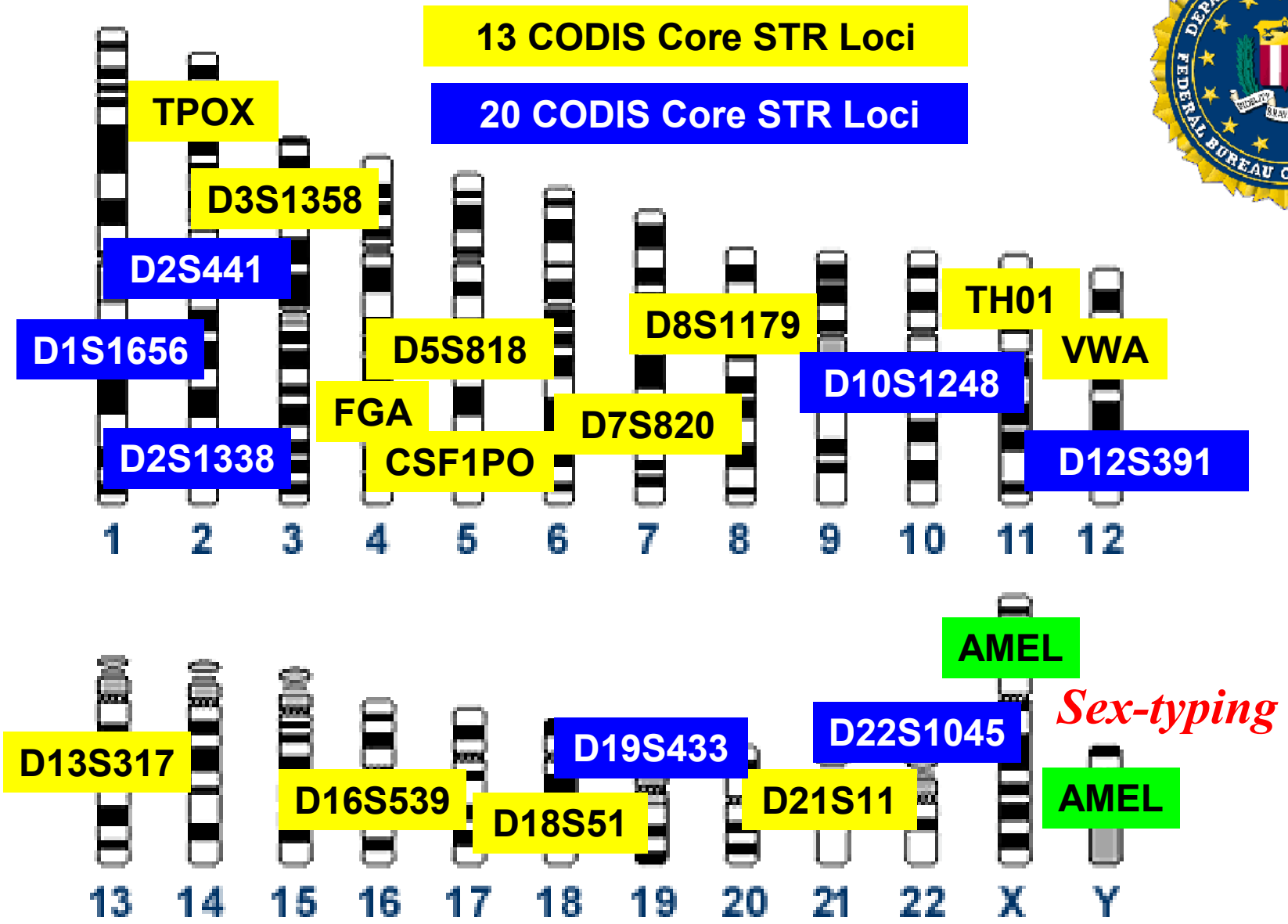








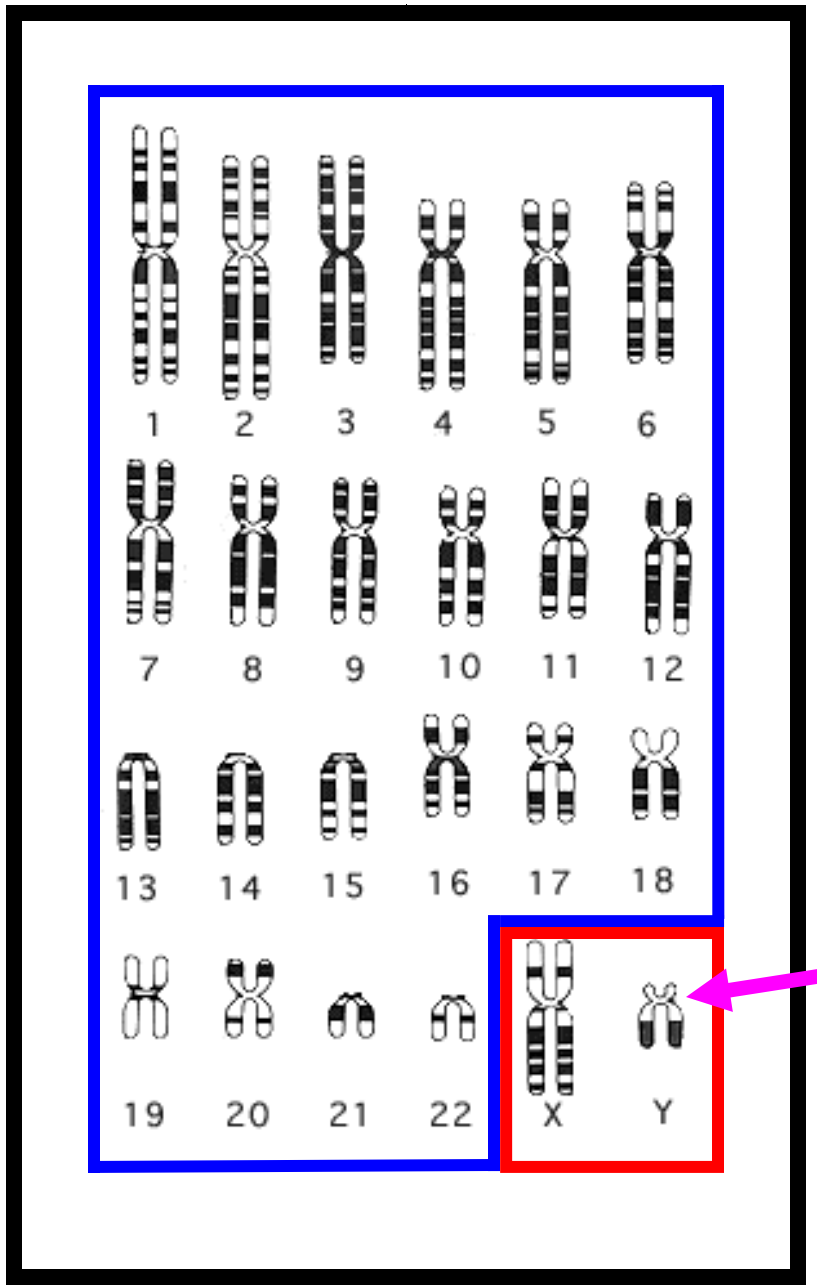
Position of Forensic STR Markers on Human Chromosomes



1997

2017

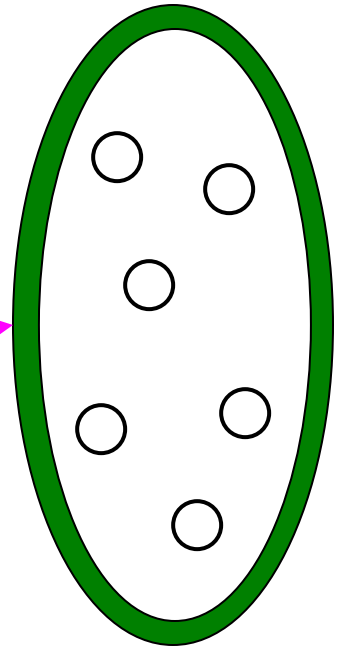
Cell Nucleus – 3.2 billion bp



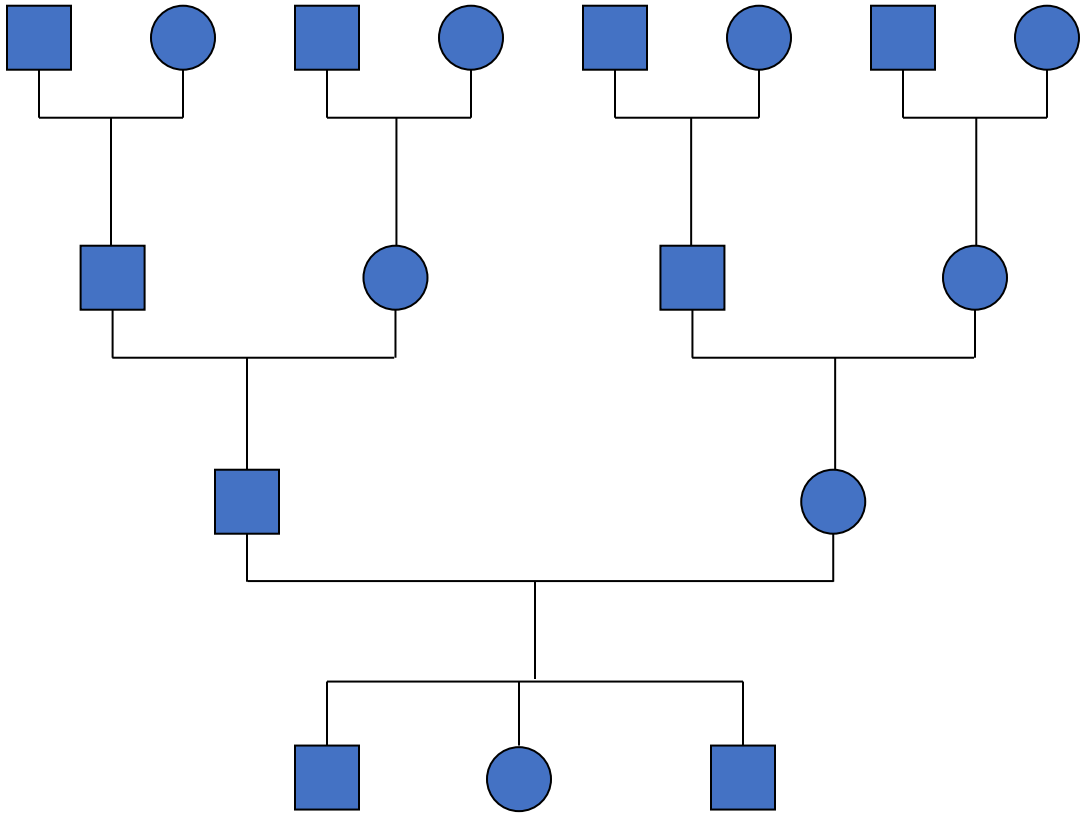
Autosomes – 22 pairs – 2 copies per cell

Sex Chromosomes (XX or XY)

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

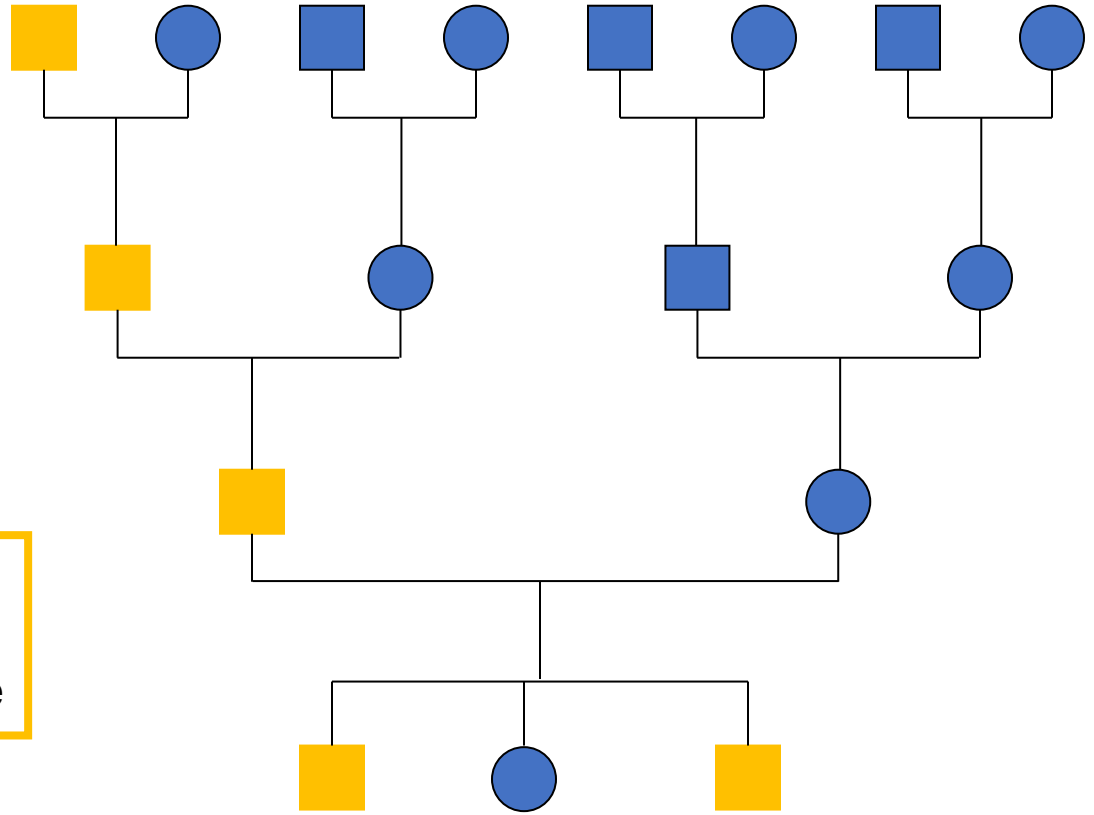


“Autosomal” Markers



Autosomal DNA
1/8 from Great-grandparents

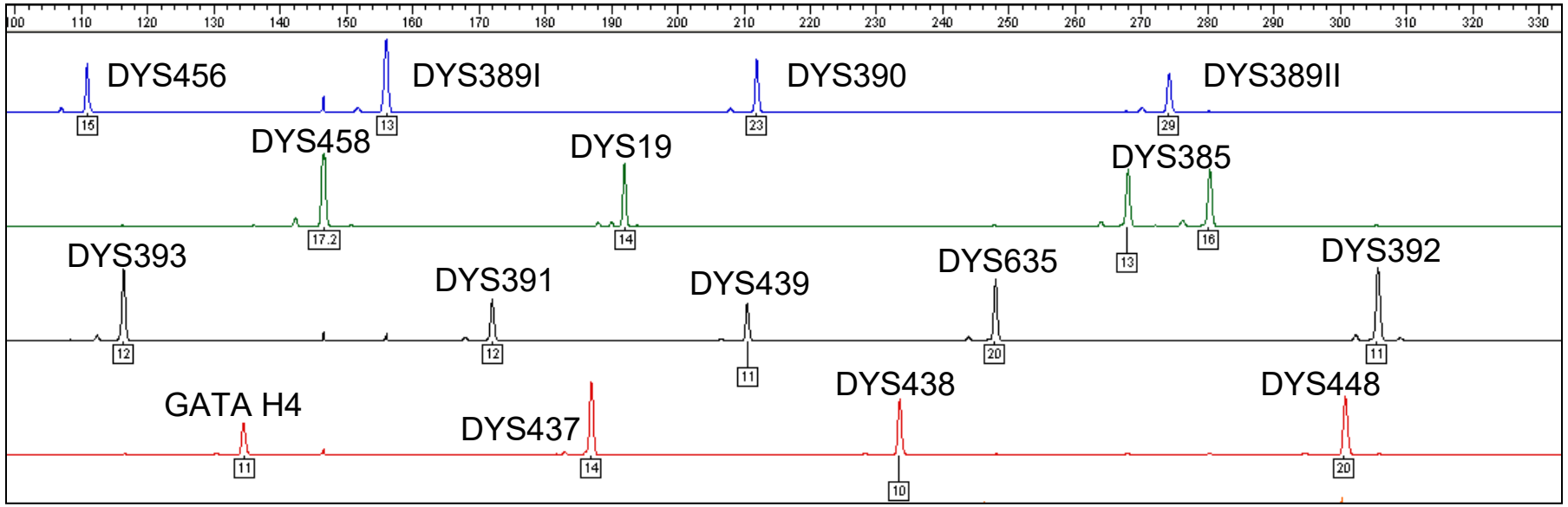
Lineage Markers



Y chromosome passed along paternal lineage

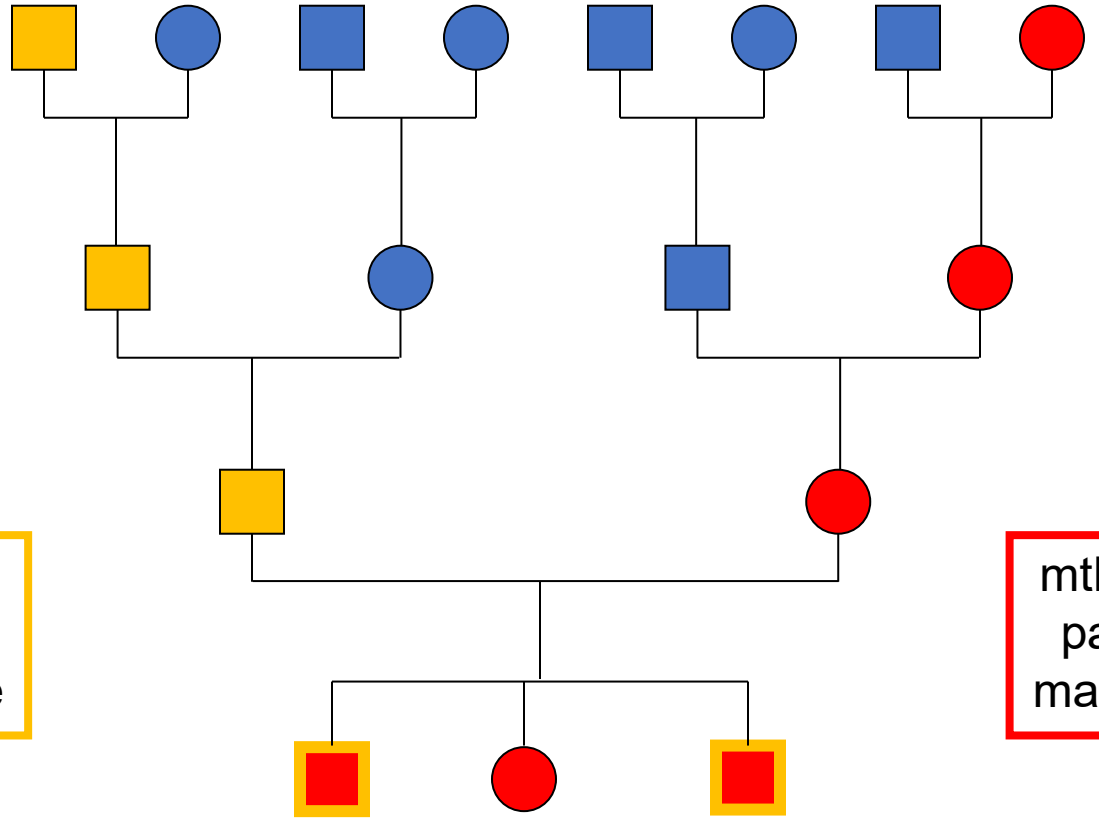
Autosomal DNA
1/8 from Great-grandparents

Forensic Y-STR Testing



Y-filer (Thermo Fisher): 17 Y-STRs

Lineage Markers

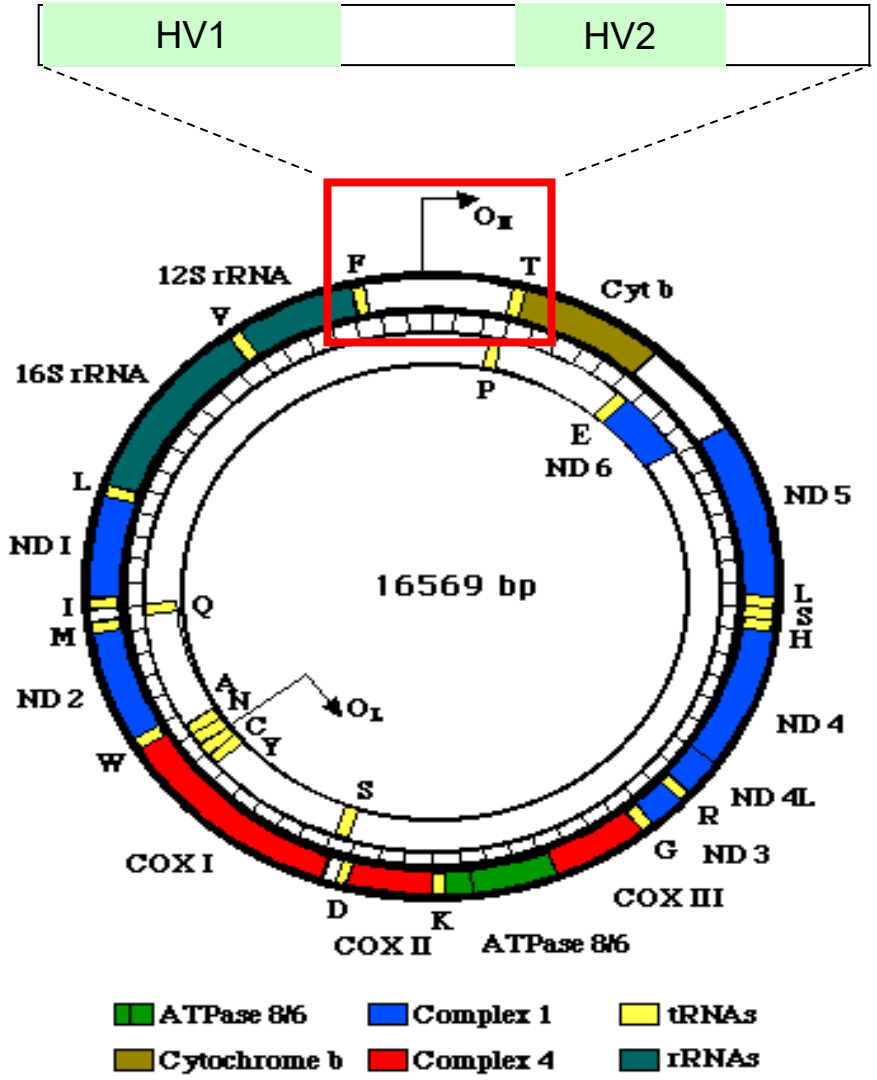


Y chromosome passed along paternal lineage

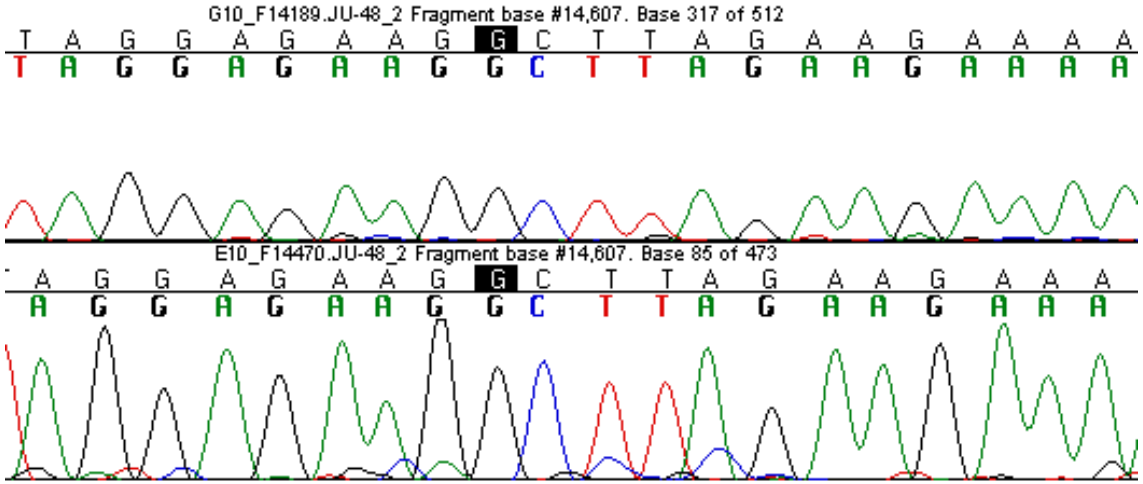
mtDNA genome passed along maternal lineage

Autosomal DNA
1/8 from Great-grandparents

Control region (D-loop)



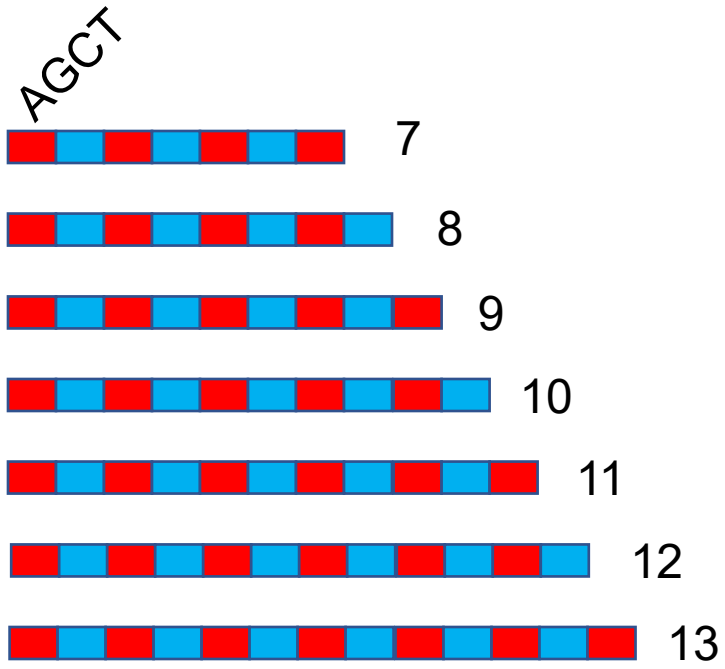
Forensic Scientists have focused on ~600bp of sequence information in the non-coding control region of the mtDNA.



263 A – G
16519 T – C
315.1 C

http://www.mitomap.org/

STRs vs. SNPs



28 possible genotype combinations

7,8 9,13 10,11 12,13
9,9 11,11 8,9 10,13
etc…

C

T

3 possible genotype combinations

CC
CT
TT

STRs vs. SNPs

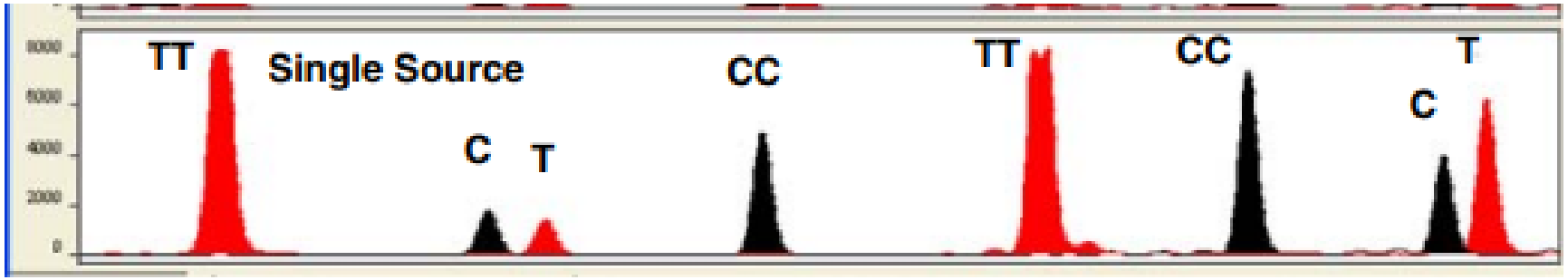
- SNPs – mostly bi-allelic (C or T)
  - Require more SNPs for identification

  ~3-4 SNPs = 1 STR

- SNPs – very low mutation rates compared to STRs – this is great for ***kinship and genealogy***

- SNPs – fewer artifacts than STR data (no stutter!), can provide ancestry and phenotypic (outward appearance) information

Limits of SNP testing (1990s-2010s)



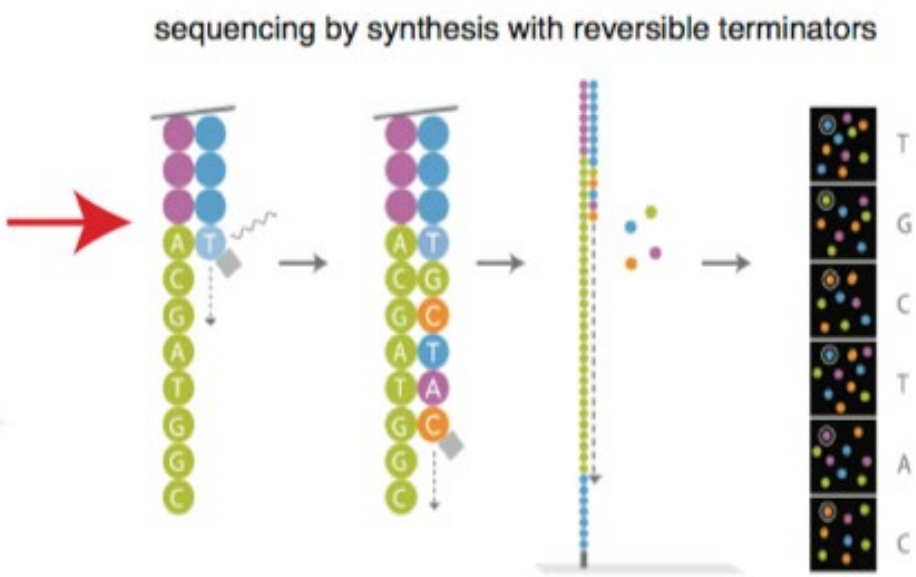
“Single Base Extension” Assay

Fluorescence Dyes/use on CE instruments

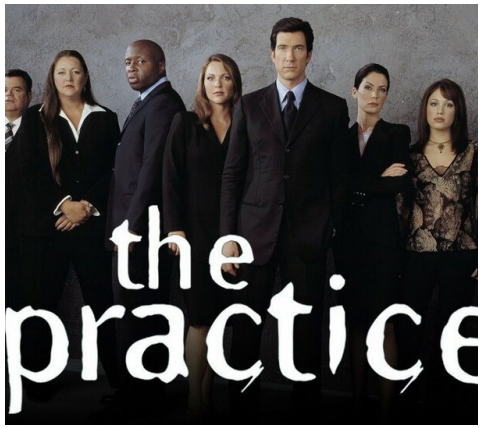
Limited number of SNPs that can be tested

STRs vs. SNPs

- Use of Massive Parallel Sequencing can generate hundreds of thousands to millions of SNPs on a single chip



Traditional Crime Scene Evidence/Testing

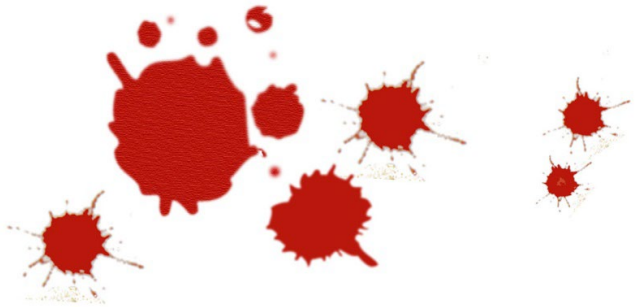


Traditional Crime Scene Evidence Collection/Testing

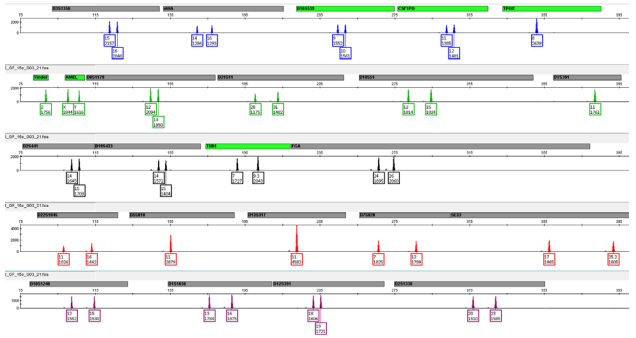
- Collecting blood, semen, saliva, clothing, etc. from crime scene to take to the lab for possible testing

- First possibility is to examine the collected evidence for potential STR (short tandem repeat) testing to hopefully develop a single source suspect profile that can be uploaded into CODIS

Conventional DNA Testing

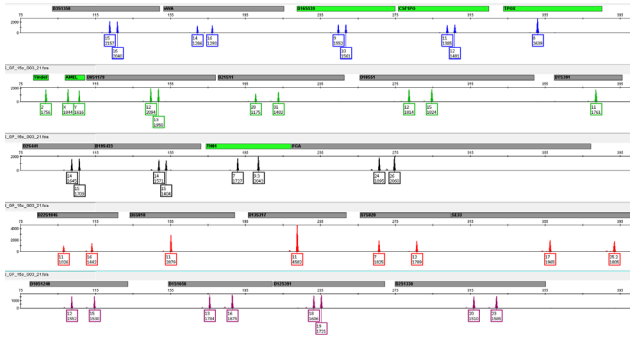


Crime Scene Evidence



=

Person of Interest



Conventional DNA Testing

- With conventional DNA testing, DNA has correctly earned the title of the “Gold Standard” in forensic testing.
- This is based upon the application of scientifically accepted techniques such as PCR and CE.
- We also have over 100 years of population genetic theory to support our statistical conclusions of a match between the stain and the Person of Interest.

1989: Virginia is First State to Create DNA Database of Convicted Offenders

- 1990, the FBI introduces a national DNA database as a pilot project. DNA Identification Act of 1994 gave authority to create national database and it went into effect in 1998.

- CODIS, Combined DNA Index System, is a program that operates local, state, and national databases of DNA profiles from convicted offenders, unsolved crime scene evidence, and missing persons.

CODIS (NDIS/SDIS)

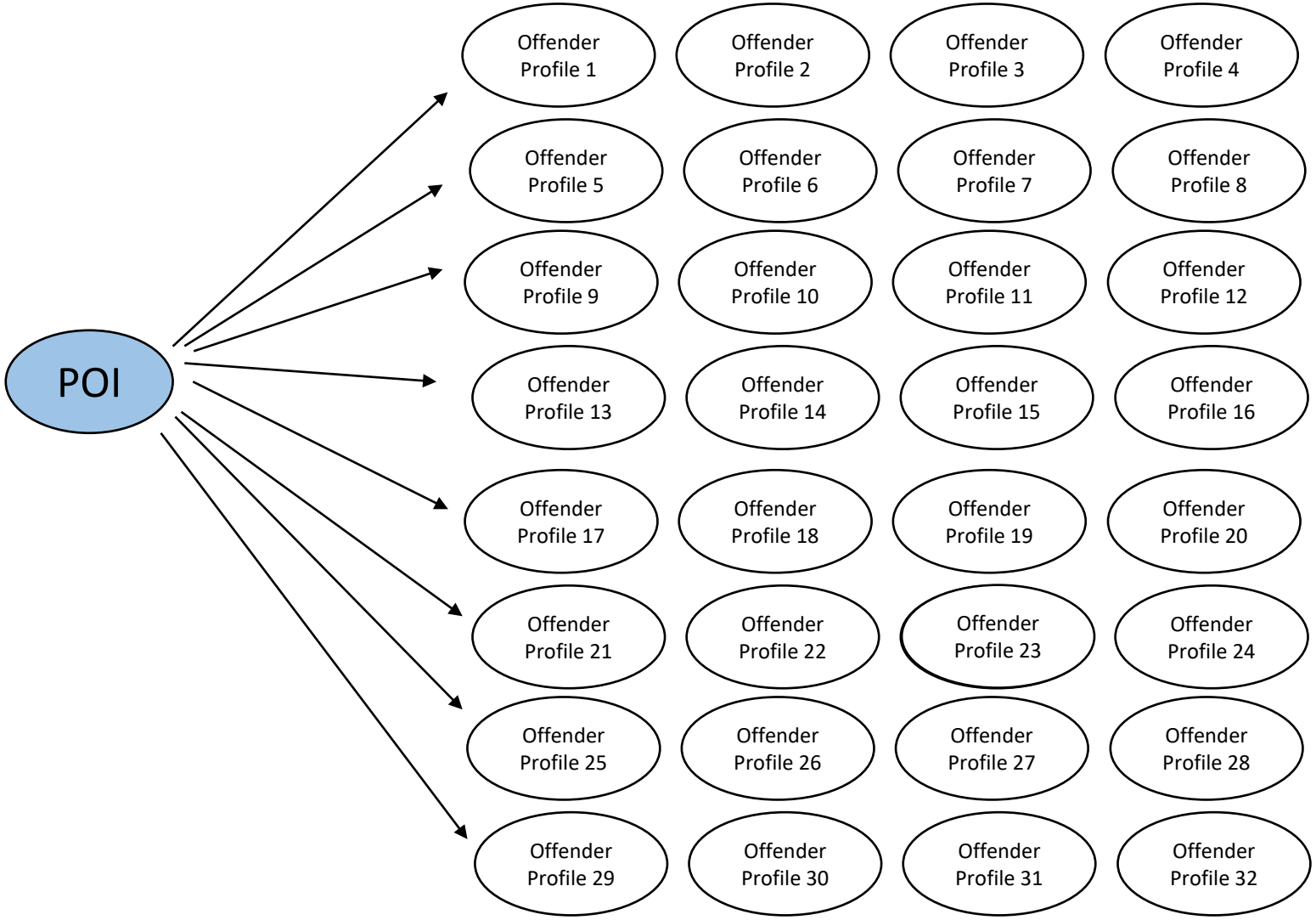
- Each state has different laws that govern who is eligible for DNA collection. In addition, the state maintains its own database (SDIS) with its own requirements for profile uploading.

- The National DNA Index (NDIS) contains over 16,532,335 offender profiles, 5,190,279 arrestee profiles and **1,282,418** forensic profiles as of August 2023. States must follow NDIS requirements for eligible uploads.

Conventional DNA Testing

- What if – we have no suspect?
- Databases – allow the forensic scientist to potentially identify a person that matches the evidence profile

National Database Search



CODIS Hierarchy

National DNA Index System (NDIS)

State DNA Index System (SDIS)

Local DNA Index System (LDIS)

Conventional DNA Testing

- What if – the person of interest is not in the database because he hasn’t committed a qualifying crime that would allow his profile to be entered into the database?

Familial Searching

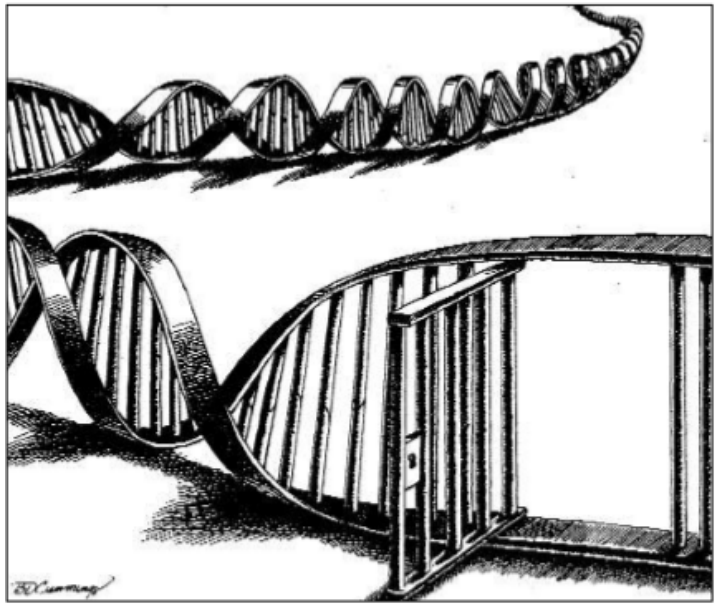
Familial Searching

HUMAN GENETICS

Finding Criminals Through DNA of Their Relatives

Frederick R. Bieber,[1]* Charles H. Brenner,[2] David Lazer[3]

Science (2006) **312:** 1315-1316



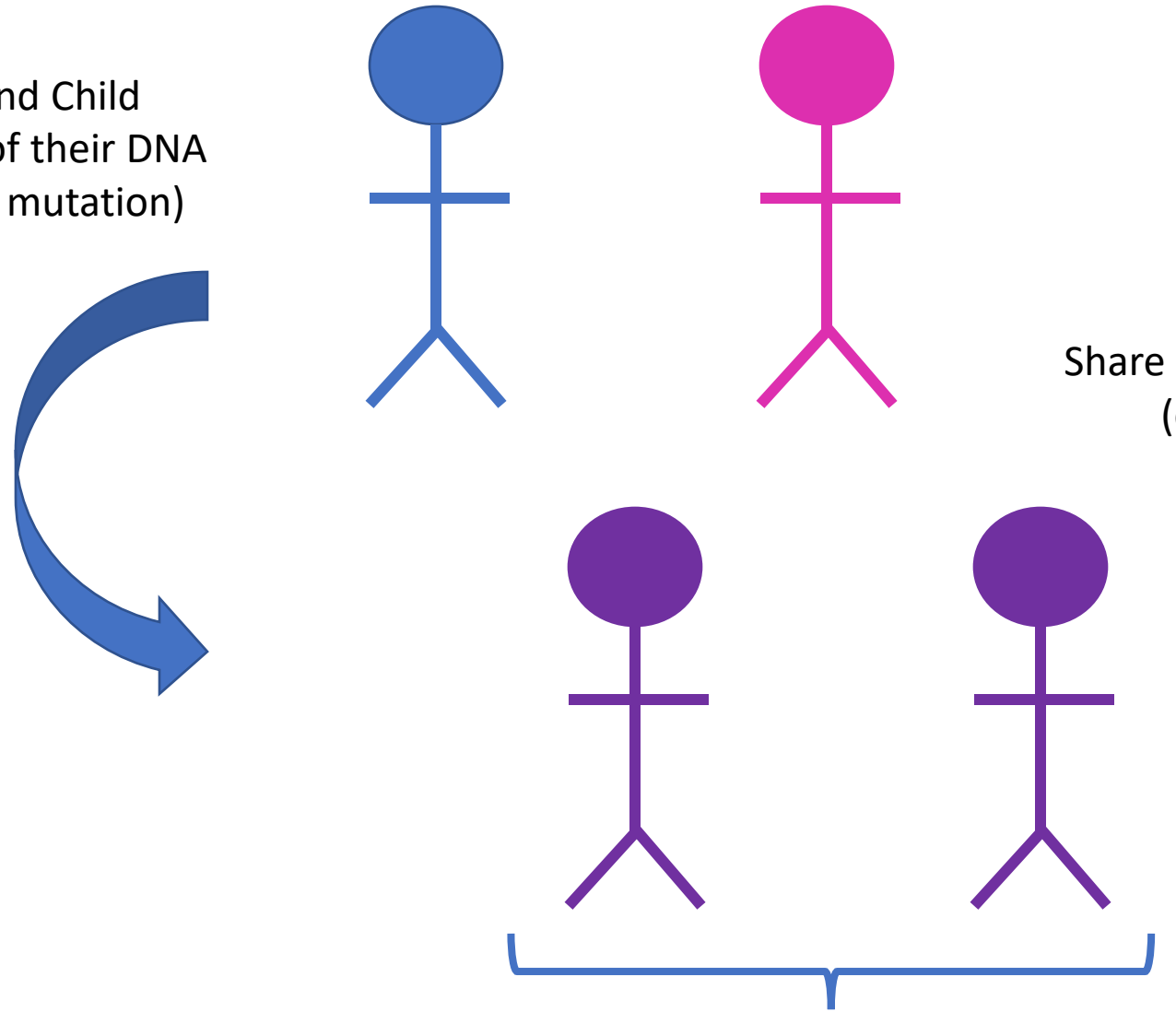
Why look for relatives?

- Statistics show that (in the US) there is a strong probabilistic dependency between the conviction of a parent and their children.

- Surveys have shown that about 46% of prisoners have indicated that they have a close relative also in prison.

Kinship Review

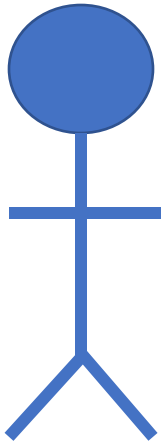
Parent and Child
Share 1/2 of their DNA
(barring a mutation)

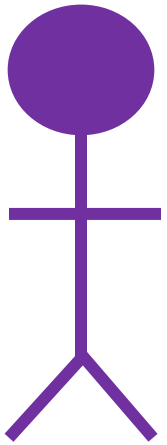


Siblings
Share 1/2 of their DNA
(on average)

Kinship Review

Parent and Child
Share 1/2 of their DNA
(barring a mutation)

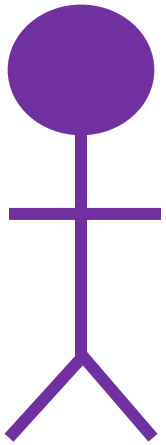


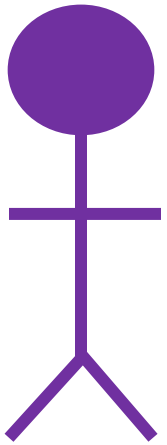


	genotype	genotype
CSF1PO	**10**, 13	**10**, 10
D18S51	17, **19**	15, **19**
D19S433	15, **15.2**	13, **15.2**

Kinship Review

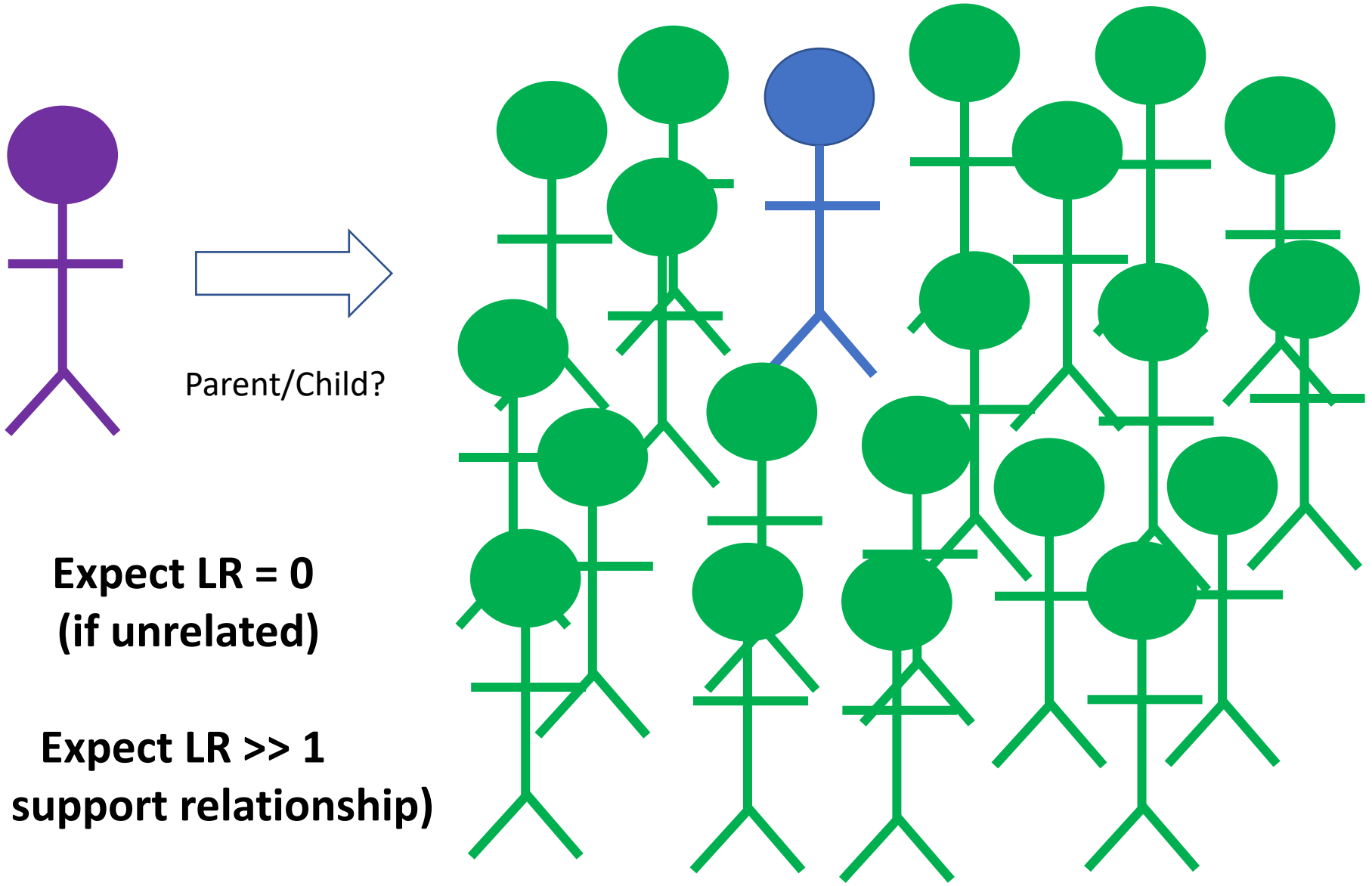
Siblings
Share 1/2 of their DNA
(on average)





	genotype	genotype	
CSF1PO	**10, 13**	**10, 13**	2 alleles in common
D18S51	17, **19**	15, **19**	1 allele in common
D19S433	15, 15.2	14, 16	0 alleles in common

Familial Searching



Expect LR = 0
(if unrelated)

Expect LR >> 1
(to support relationship)

Limits of Familial Searching

- Not performed at a national level (only states)
- Some states prohibit familial testing in the U.S.
- Fortuitous inclusions

Ways to decrease fortuitous inclusions

- Use more autosomal STR loci (kits now have 20+ loci)
- Use ***Y-STRs*** to eliminate false positives – newer Y-STR kits are even more highly discriminating than those in the past.
- “2-step” approach – autosomal STR followed by Y-STR testing.

Limits of Familial Searching

- Not performed at a national level (only states)
- Some states prohibit familial testing in the U.S.
- Fortuitous inclusions
- Females in the database (Y-STRs won’t work)

What if Familial Searching Fails?

- Wait for the POI to commit a qualifying crime for entry into CODIS

- Wait for a relative of the POI to commit a qualifying crime for entry into your SDIS*

*provided your state uses familial searching…

Forensic Investigative Genetic Genealogy

FIGG

IGG

FGG

FIGG/IGG/FGG: What It Is Not

Not CODIS Familial Searching

- different from forensic genetic genealogy
- searching **state** CODIS databases (SDIS) to look for close relatives of an unknown offender using Y-STR profiles (male profiles)
- only allowed in 13 states (including Texas); very costly and time consuming

Genealogy In the News

STR vs SNP

STR uploaded to CODIS

- STR=Short Tandem Repeats
- Crime scene STR profile is compared to the profiles stored in CODIS to find potential perpetrators
- 40% chance of producing a match (lead)
- Uploaded by law enforcement according to statute/controlled by state

SNP uploaded to consumer genealogy database

- SNP=Single Nucleotide Polymorphisms
- Different type of testing (but can take from same extract used from crime scene if enough sample left after STR testing). Profile is uploaded and compared to users in public databases
- 80%-90% chance of producing a lead by matching to a 2nd or 3rd cousin
- Voluntary upload/Private Industry

The Power of the Tool

-credit to Steve Kramer, Indago Solutions, former FBI
and Anne Marie Schubert, former Sacramento District Attorney

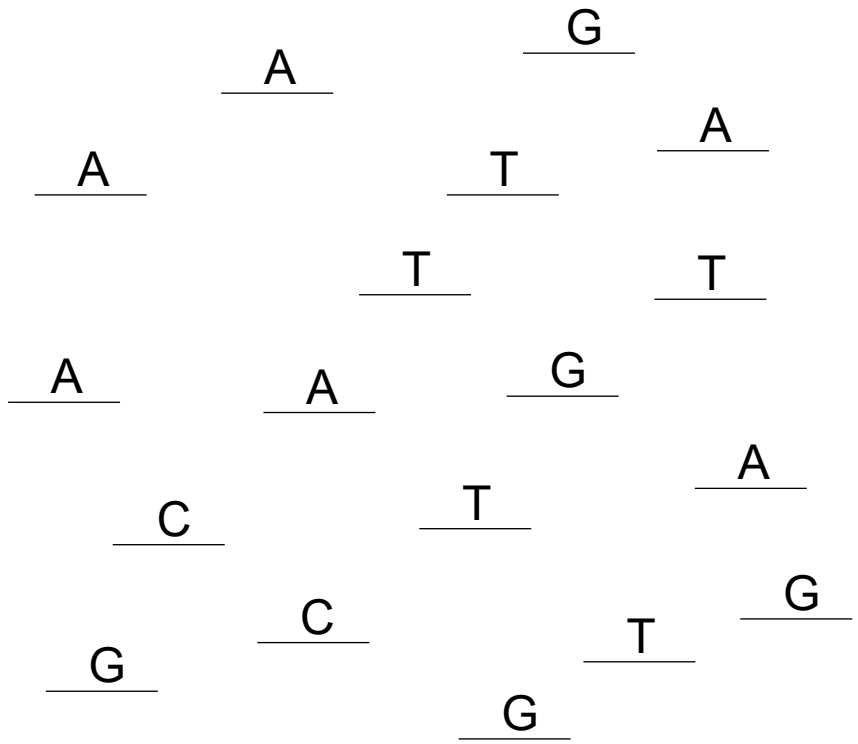
Before Using FIGG

- Took 43 years to solve the Golden State Killer cases
- 650 investigators over the years and 15 LE agencies
- 8000 persons of interest/300 people swabbed for DNA
- 200k man/woman hours=$10 million dollars

After Using FIGG

- Took 63 days to identify Joseph James DeAngelo as the Golden State Killer
- 6 people to work the family tree
- $217 (cost of SNP testing back then)

SNPs for Forensic Identification



ACTTACCGTTCCTGAAGG

ATTTACCGCTCCTGAGGG

ATTTACCGTTCCTGAAGG

ACTTACCGTTCCTGAGGG

~90 ‘random’ SNPs = Full STR Profile

SNPs in close proximity = Genealogy

Thomas Hunt Morgan

1933 Nobel Prize in Physiology or Medicine for discoveries elucidating the role that the chromosome plays in heredity.

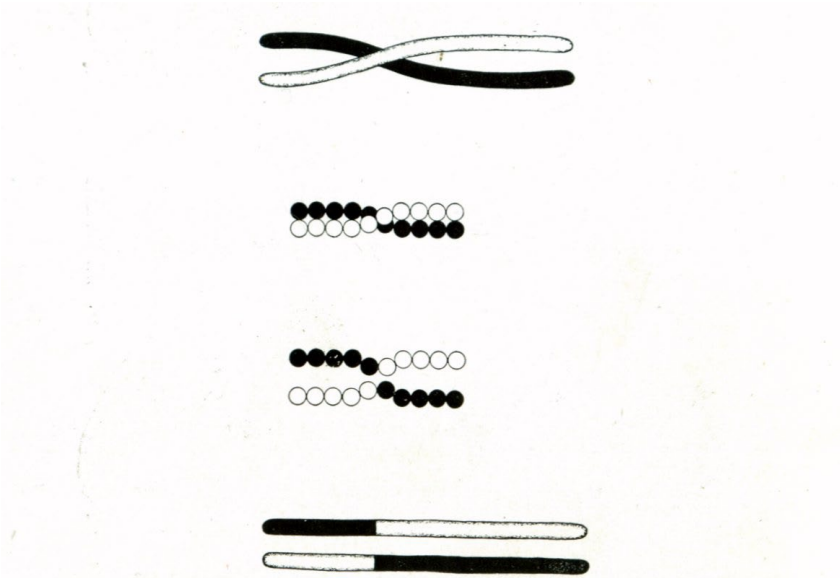


FIG. 64. Scheme to illustrate a method of crossing over of the chromosomes.

CentiMorgan (cM)

- A unit for measuring genetic linkage along chromosomes.
- 1 cM = 1 million bases of DNA (AKA “MegaBase”)
- Two SNPs that are 1 cM apart from one another have only a 1% probability of recombination between them (hence, “centi”)

Large “Blocks” of shared SNPs

108 cM fragment

Me

Likely 2nd or 3rd Cousins

People in a Genealogy Database

1 centiMorgan (cM) = 1 million bases of DNA

My Cousins

Family Finder Matches

Help

All Matches | Detail View | Table View | Exact Search | Search | All

All (9494) | Paternal (0) | Maternal (0) | Both (0) | Filter | Sort by | Export CSV

Y-DNA37

Ancestral Surnames	Haplogroup	Relationship Range	Shared DNA	Longest Block	X Match
Not Provided	Y-DNA: I-M253 mtDNA: N/A	2nd Cousin - 4th Cousin Link on Family Tree	108 cM	37 cM	No Match

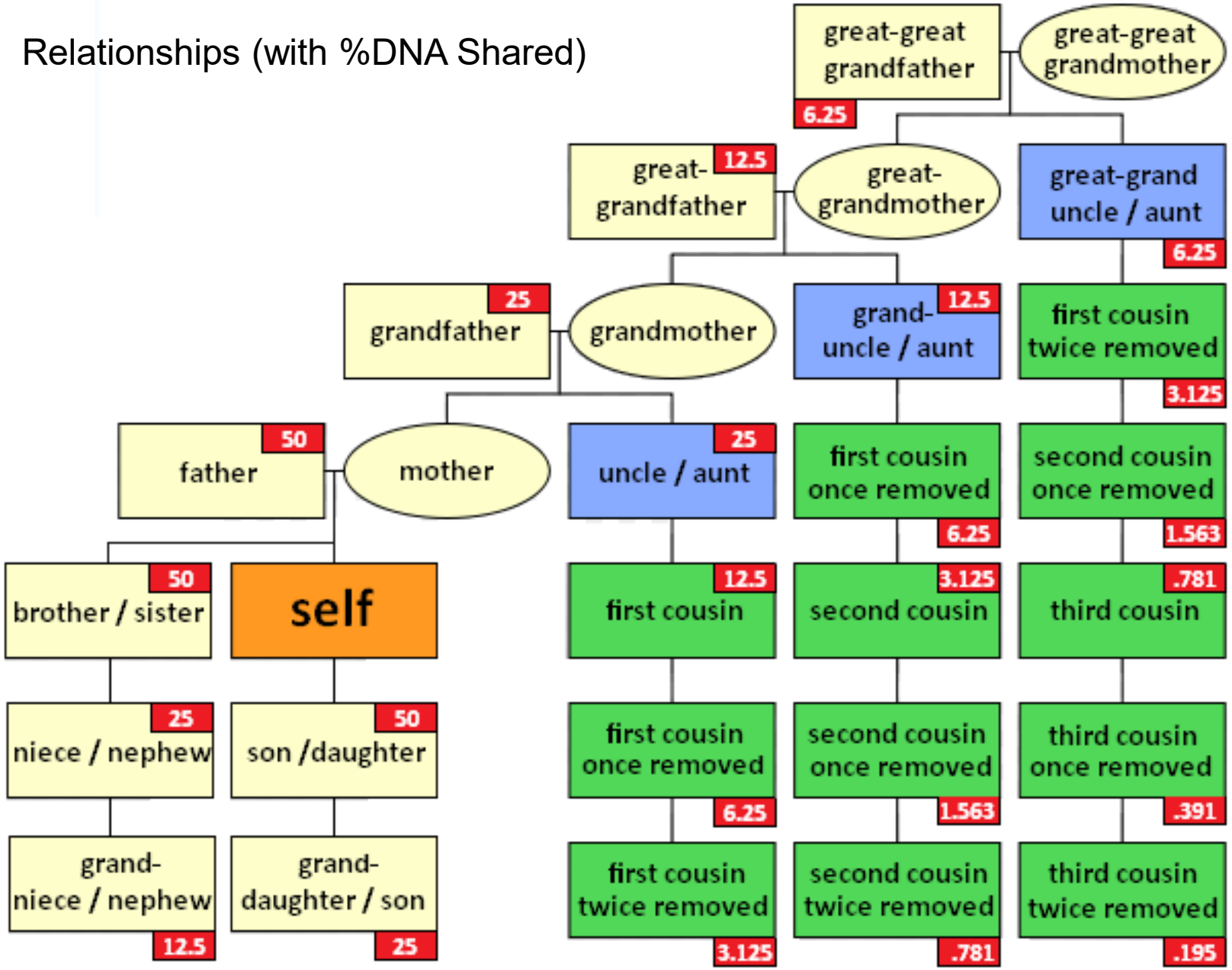
Match date: May 10 2022

Ancestral Surnames	Haplogroup	Relationship Range	Shared DNA	Longest Block	X Match
Not Provided	mtDNA: N/A	2nd Cousin - 4th Cousin Link on Family Tree	81 cM	81 cM	No Match

Match date: May 10 2022

FamilyTreeDNA

Relationships (with %DNA Shared)



ISOGG wiki statistics:

Parent/child: 3539-3748 cMs
1st cousins: 548-1034 cMs
1st cousins 1R: 248-638 cMs
2nd cousins: 101-378 cMs
2nd cousins 2R: 43-191 cMs
3rd cousins: 43-ca 150 cMs
3rd cousins 1R: 11.5-99 cMs
More distant cousins: 5-ca 50 cMs

Company	**23andMe**	**Family Tree DNA's Family Finder test**	**Ancestry.com's AncestryDNA test**
Price (as of January 2021)	$99 for Ancestry and Traits	$79 for the lifetime of the platform	$69 in the U.S.
SNP chip used for testing	Customized Illumina GSA chip	Customized Illumina GSA chip	Customized Illumina chip
Number of autosomal SNPs tested	630,132	612,272	637,639
Number of people in the database (as of 13 August 2023)	**14,000,000**	**1,574,253**	**23,000,000**

Over 46 million people have submitted DNA for testing!

23 and Me  Family Tree DNA  Ancestry.com

https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart

GEDmatch

an open-source personal genomics database and genealogy website

Allows people to search for close relatives across the platforms

23 and Me ⟷ Family Tree DNA ⟷ Ancestry.com

As of February 2024, the GEDmatch database has over 2.0 million genetic profiles

Note: Users of FamilyTree DNA may opt-in for FIGG LE searches

How Effective is this process?

Science

REPORTS

Cite as: Y. Erlich *et al.*, *Science*
10.1126/science.aau4832 (2018).

Identity inference of genomic data using long-range familial searches

Yaniv Erlich[1,2,3,4*], Tal Shor[1], Itsik Pe'er[2,3], Shai Carmi[5]

[1]MyHeritage, Or Yehuda 6037606, Israel. [2]Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York, NY, USA. [3]Center for Computational Biology and Bioinformatics (C2B2), Department of Systems Biology, Columbia University, New York, NY, USA. [4]New York Genome Center, New York, NY, USA. [5]Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel.

***Corresponding author. Email: erlichya@gmail.com**

With ~ I million people in GEDmatch, about 60% of the US European population can be identified (using 3rd cousins).

Facts about FIGG

- Law enforcement does not have access to medical information or identifying information, other than an email, when they use genealogy databases. They only have information on how families are related within centimorgans “cM” so they can begin building and researching family trees and histories.

- FIGG only provides law enforcement with a lead. Detectives must still conduct further investigation by obtaining subsequent samples and confirming leads by using a one-to-one STR test to make sure they have the right individual.

Legal Considerations and FIGG

- **Accreditation**
  - Some states have statutes that govern admissibility of evidence based on whether a lab has received accreditation. Some federal grants require using an accredited lab. Definitely a question you want to ask before using a lab for testing.

- **Discovery**
  - Could be asked to turn over materials used during FIGG research. Databases used, family tree research, identities of relatives and reference samples obtained
  - Recent ruling in Kohberger case out of Idaho (Oct. 2023)
  - Some states have broader discovery rules (Texas-Michael Morton Act, larger standard than Brady)

Legal Considerations: Discovery

California

People v. Waller (Sacramento): Prosecution does not have to disclose:

1. Genealogy company
2. Genealogy data
3. Law enforcement communications with genealogy company

People v. Wilson (Sacramento): Prosecution does not have to disclose:

1. The IGG techniques used in solving the case
2. The IGG related commercial websites/databases used
3. The identities of relatives of defendant
4. The cooperation of any relatives in developing defendant as suspect

Ohio

State v. Bortree (Logan County): Testimony of owner of genealogy company admissible because genealogy merely narrowed the focus of law enforcement

Legal Considerations: Privacy/4th Amendment

Privacy Challenges/3rd Party Doctrine

- Claim that use of genealogy database violates 4th Amendment because right to privacy in shared DNA of relatives. Courts so far have found this an unpersuasive argument based upon a **lack of standing** and no property interest in a 3rd party's genetic material voluntarily uploaded

- **Carpenter v. United States** (2018): Supreme Court expanded privacy concerns with cell site location information (CSLI) claiming it raises greater privacy concerns than GPS location data (citing U.S. v. Jones) because it goes back in time to track a person's every move. Must have a warrant to obtain this data even though it is with a third party (phone company)

  -Says need more than just a court order/subpoena to obtain this data but also states very narrow ruling and case specific exceptions may support a warrantless search

  -Argument that may need a SW to obtain info from genealogy database (3rd party)-is DNA similar enough to CSLI?

Legal Considerations: Abandoned Property/Trash

- ❖ Challenges to covert sample taken from a potential suspect's trash to perform STR testing for purposes of FIGG
- ❖ Claim is that LE requires warrant either to collect the sample or to do the testing (4th Amendment violation)

  - **Oregon v. Lien**: DNA collected from curbside trash suppressed where garbage collector acted as agent of law enforcement
  - **Minnesota v. Carbo**: Law enforcement lawfully collected trash at apartment dumpster after FIGG developed Defendant as potential suspect. Motion to suppress denied
  - **ACLU and Electronic Frontier Foundation**: Have filed several briefs to argue that it's ok to take trash but NOT ok to do DNA testing on trash items without a warrant

DOJ Interim Policy (2019)

DOJ Guidelines

- Interim policy on Forensic Genetic Genealogical DNA Analysis and Searching (FGGS) with specific requirements before FGGS can be used including having an STR profile uploaded into CODIS and getting prosecutorial approval prior to using FGGS

- Must follow if receiving any federal grant funding

https://www.justice.gov/olp/page/file/1204386/download

Considerations with FIGG

Future Legislation

- ❖ Efforts to regulate vary widely from state to state (stems from a lot of misunderstanding about what FIGG is NOT)
- ❖ Familial Searching (NY bill filed Dec. 2023 to prohibit the state from conducting familial searching-other states could follow)
- ❖ Must use responsibly or will lose this valuable tool

Remember: This tool can implicate but can also exonerate

Final Thoughts

- Successful use of FIGG involves a team approach including Prosecutors, Investigators, and the DNA Laboratory.

- There are at least three current strategies for developing SNPs for FIGG – targeted (Kintelligence), microarray, and WGS. Each has their own advantages and disadvantages.

- FIGG should be used as any other tool for solving crime - a lead that needs to be investigated and confirmed before relying on it to make an arrest.

Resources

Season of Justice
https://seasonofjustice.org/

University of North Texas Health Science Center/Center for Human Identification
https://www.unthsc.edu/center-for-human-identification/

Verogen/Qiagen
https://verogen.com/law-enforcement-forensic-investigative-genetic-genealogy/#glossary

National Center for Missing and Exploited Children
https://www.missingkids.org/home

Department of Justice Grant Funding Opportunities
https://bja.ojp.gov/funding

Texas Attorney General's Office Cold Case and Missing Persons Unit
https://www.texasattorneygeneral.gov/criminal-justice/cold-case-and-missing-persons-unit

Thank you to QIAGEN for their support in this project

michael.coble@unthsc.edu

Lorena Vollrath-Bueno

Mindy.Montford@oag.texas.gov