



Technology Transition Workshop | *John V. Planz, Ph.D.*

Analysis and Implications of SNP Distributions Within STR Loci

The STR in Forensic Analysis

- **STRs are the mainstay in Human Identity Testing**
- **Allele is described as a nominal repeat number of a sequence motif**
- **Allele distribution typically provides some fairly common alleles and several rarer alleles**
- **Assumptions concerning STR alleles vary among applications**

STR Markers

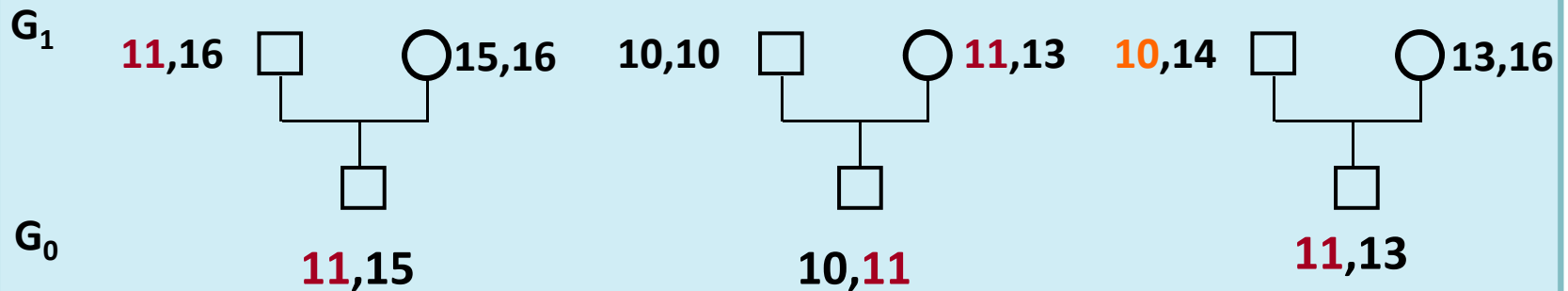
The STR markers we use in routine forensic comparisons are what is considered:

Identical by State

There is no presupposition of ancestral descent attached to them...there doesn't need to be!

STR Markers

History of an 11 Allele



Siblings ?

Parent-Offspring ?

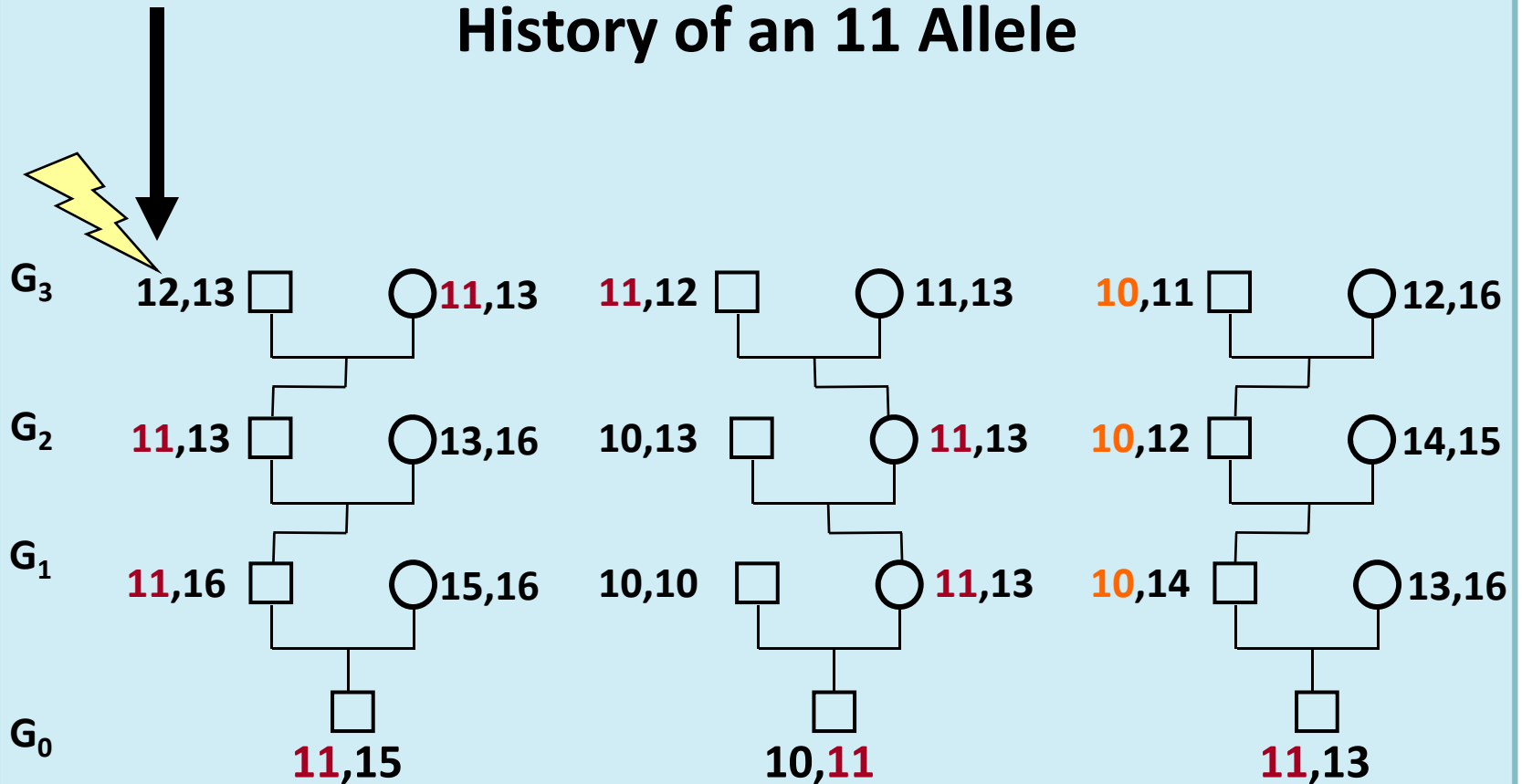
Unrelated ?

Mutation Process for STR Loci

- **STR loci undergo mutational changes at a frequency of approximately 10^{-3}**
- **The allele changes observed between generations is typically one repeat larger or smaller than the allele observed in the parental generation**
- **This process is analogous to the formation of stutter alleles observed when amplifying STR loci**

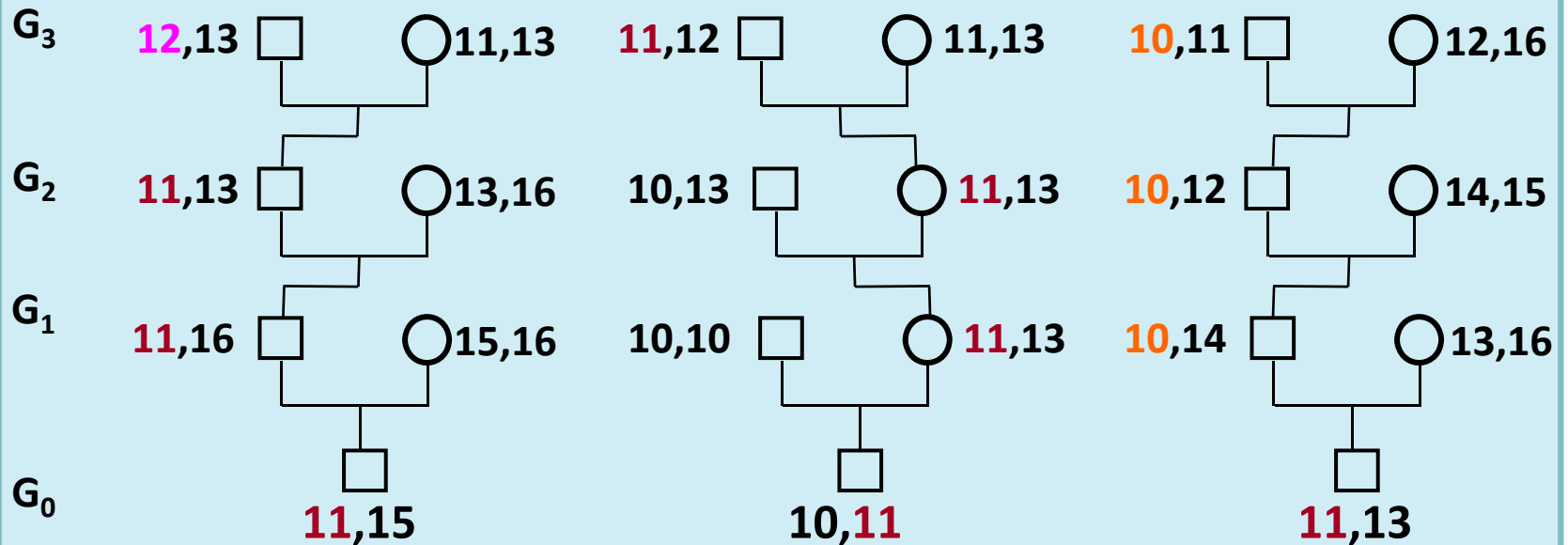
STR Markers

History of an 11 Allele



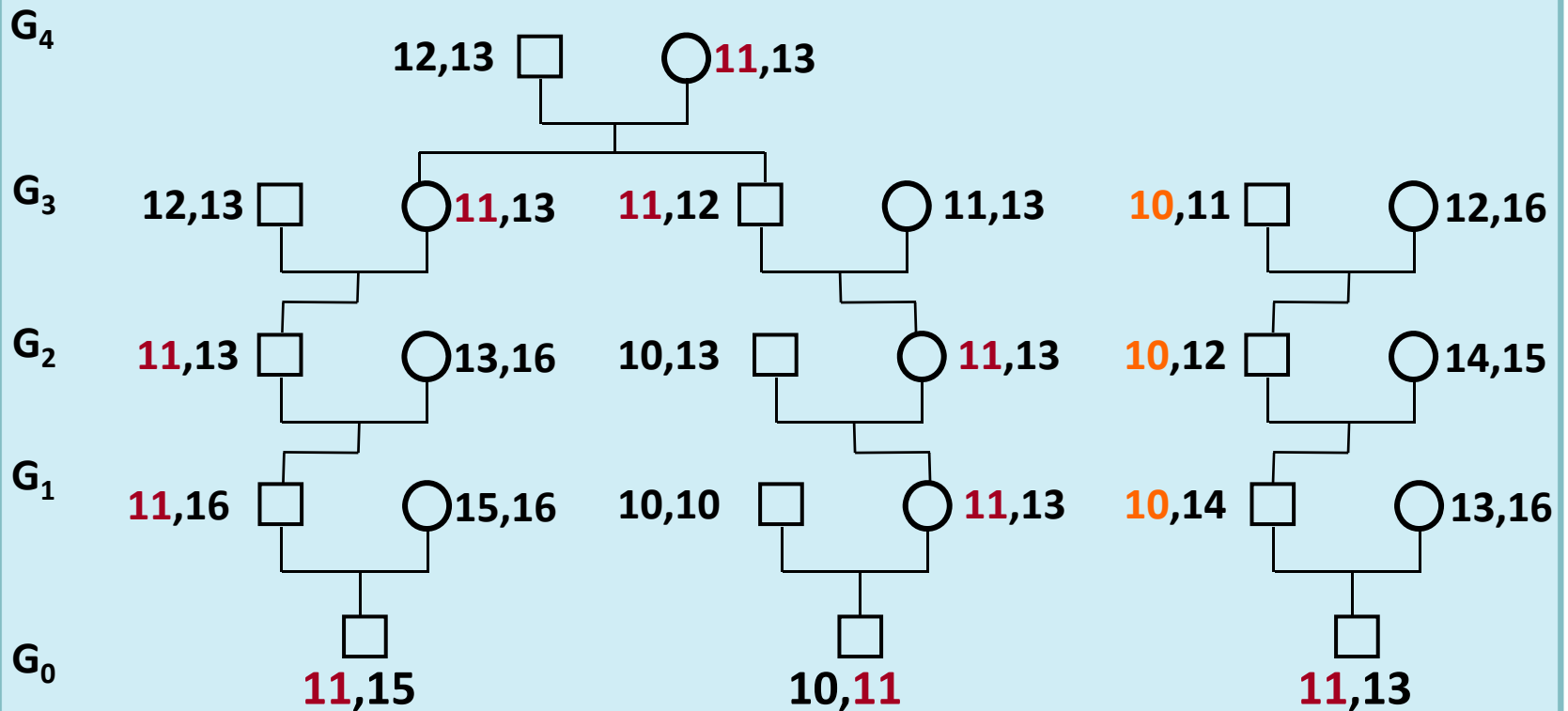
STR Markers

History of an 11 Allele



STR Markers

History of an 11 Allele



Third Cousins!

STR Markers

When comparing alleles between individuals as a result of a database search under the premise of a familial association, we are banking on the shared alleles being:

Identical by Descent

This is clearly a fallacy with regard to the way we use autosomal STR markers when conducting an open database search.

Assumptions/Applications

- **Single nucleotide polymorphisms in and around the STR loci are known**
- **Current analytical methods observe some of these polymorphisms as primer binding mutations or microvariants**
- **Categorizing SNP variants would require sequence analysis**
- **Initial thoughts were that these polymorphisms would be relatively rare**

Assumptions/Applications

- Single nucleotide polymorphisms in and around the STR loci are known
- Current analytical methods observe some of these polymorphisms as primer binding mutations or microvariants
- Categorizing SNP variants would require sequence analysis
- Initial thoughts were that these polymorphisms would be relatively rare



FALSE!

Technology
Transition Workshop



Evaluation of Typing System

- **587 individuals from three US population groups were typed for the 13 CODIS core STR loci**
- **Samples had previously been typed using PowerPlex 16[®]**
- **Data evaluated for allele distribution as well as Hardy Weinberg Equilibrium and Linkage Disequilibrium**
- **An additional 50 individuals belonging to two extensive pedigrees were examined**

SNP Rich STRs

- **Seven of the CODIS core loci contained SNPs within the repeat motifs, which increased the number of commonly observed alleles**
 - **D3S1358, D5S818, D7S820, D8S1179, D13S317, D21S11, vWA**
- **Allele count increased between 4 and 15 alleles dependent on locus and population group**
- **Typing of nominal alleles was concordant with prior PP16[®] results**

Locus	Population	STR-SNP Analysis 0n IBIS T5000				STR Only Analysis on IBIS T5000			
		n	Alleles Detected	H _e	H _o	n	Alleles Detected	H _e	H _o
D13S317	Caucasian	181	12	0.8735	0.9061	182	7	0.7820	0.8297
	African Am.	213	12	0.8345	0.7981	214	7	0.7026	0.7056
	Hispanic	193	13	0.8847	0.9016	193	7	0.8241	0.8290
D21S11	Caucasian	181	23	0.8925	0.9006	182	14	0.8390	0.8681
	African Am.	213	33	0.8688	0.8357	214	20	0.8459	0.8178
	Hispanic	193	25	0.8845	0.8446	193	14	0.8335	0.8083
D3S1358	Caucasian	181	18	0.8641	0.8895	182	8	0.7900	0.8077
	African Am.	213	18	0.8907	0.9061	214	8	0.7485	0.7850
	Hispanic	193	18	0.8101	0.8187	193	8	0.7391	0.7409
D5S818	Caucasian	181	15	0.7870	0.8177	182	9	0.6858	0.6703
	African Am.	213	17	0.8396	0.8310	214	9	0.7449	0.7009
	Hispanic	193	13	0.7471	0.7617	193	9	0.6998	0.6891
D7S820	Caucasian	181	15	0.8525	0.8066	182	8	0.8089	0.7857
	African Am.	213	12	0.8117	0.7887	214	8	0.7957	0.7710
	Hispanic	193	14	0.8271	0.7876	193	9	0.7895	0.7358
D8S1179	Caucasian	181	14	0.8554	0.8785	182	10	0.7970	0.8187
	African Am.	213	19	0.8215	0.8122	214	10	0.7860	0.7523
	Hispanic	193	16	0.8581	0.8860	193	9	0.7983	0.8394
vWA	Caucasian	181	22	0.8471	0.8674	182	10	0.8152	0.8022
	African Am.	213	26	0.8882	0.8920	214	11	0.8166	0.8224
	Hispanic	193	16	0.7955	0.7979	193	7	0.7692	0.7876

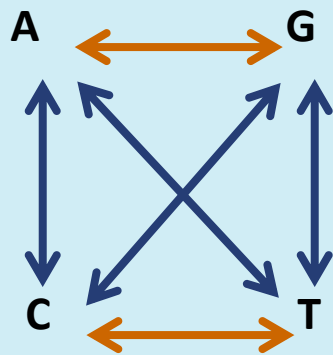
Locus	Population	STR-SNP Analysis On IBIS T5000				STR Only Analysis on IBIS T5000			
		n	Alleles Detected	H _e	H _o	n	Alleles Detected	H _e	H _o
D21S11	Caucasian	181	23	0.8925	0.9006	182	14	0.8390	0.8681
	African Am.	213	33	0.8688	0.8357	214	20	0.8459	0.8178
	Hispanic	193	25	0.8845	0.8446	193	14	0.8335	0.8083
D3S1358	Caucasian	181	18	0.8641	0.8895	182	8	0.7900	0.8077
	African Am.	213	18	0.8907	0.9061	214	8	0.7485	0.7850
	Hispanic	193	18	0.8101	0.8187	193	8	0.7391	0.7409
D5S818	Caucasian	181	15	0.7870	0.8177	182	9	0.6858	0.6703
	African Am.	213	17	0.8396	0.8310	214	9	0.7449	0.7009
	Hispanic	193	13	0.7471	0.7617	193	9	0.6998	0.6891

- **Additional alleles resulting from SNP variations were not observed in D16S539, TH01 and TPOX**
- **CSF1P0, FGA and D18S51 yielded only three to five SNPs among the STR repeats**
 - **However these were too low in frequency to substantially affect overall allele distributions**

Locus	Population	STR-SNP Analysis on IBIS T5000				STR Only Analysis on IBIS T5000			
		n	Alleles Detected	H _e	H _o	n	Alleles Detected	H _e	H _o
CSF1PO	Caucasian	181	7	0.7334	0.7348	182	7	0.7346	0.7363
	African Am.	213	11	0.7826	0.8028	214	9	0.7807	0.8037
	Hispanic	193	9	0.7242	0.7824	193	9	0.7242	0.7824
D16S539	Caucasian	181	7	0.7756	0.7845	182	7	0.7766	0.7802
	African Am.	213	9	0.8037	0.7653	214	9	0.8025	0.7617
	Hispanic	193	8	0.7769	0.7617	193	7	0.7768	0.7617
TH01	Caucasian	181	7	0.7737	0.7845	182	7	0.7732	0.7857
	African Am.	213	6	0.7363	0.7324	214	6	0.7348	0.7290
	Hispanic	193	6	0.7414	0.7306	193	6	0.7401	0.7306
TPOX	Caucasian	181	6	0.6509	0.6575	182	6	0.6524	0.6538
	African Am.	213	7	0.7692	0.7700	214	7	0.7679	0.7757
	Hispanic	193	7	0.6139	0.6114	193	7	0.6118	0.6062
FGA	Caucasian	181	15	0.8632	0.8729	182	13	0.8625	0.8736
	African Am.	213	27	0.8747	0.8920	214	22	0.8729	0.8879
	Hispanic	193	18	0.8806	0.9016	193	15	0.8784	0.9016
D18S51	Caucasian	181	14	0.8788	0.8729	182	13	0.8772	0.8736
	African Am.	213	23	0.8847	0.9249	214	18	0.8818	0.9206
	Hispanic	193	19	0.8755	0.8860	193	15	0.8710	0.8756

Allele Nomenclature

- **The inclusion of various SNP possibilities within the nominal repeat structure of the STR alleles required specific nomenclature to allow for unambiguous naming of the alleles for analysis and databasing**
- **In addition to the nominal allele (based on repeats), a suffix code was added that describes the specific type of polymorphism (Transition or Transversion) and number of SNPs observed in the fragment**



Transitions

Transversions

1 G→A	2 A→G	3 C→T	4 T→C
5 C→G	6 G→C	7 T→G	8 G→T
9 A→T	10 T→A	11 A→C	12 C→A

12 Allele containing a A→G SNP = 12S2

12 Allele containing a A→G and T→C SNP = 12S2S4

12 Allele containing 2 A→G SNPs = 12S2.2

S1 G→A	S2 A→G	S3 C→T	S4 T→C
S5 C→G	S6 G→C	S7 T→G	S8 G→T
S9 A→T	S10 T→A	S11 A→C	S12 C→A

- **Using the mass analysis methodology, allele designations are based on sequence base composition**
- **Unlike fragment analysis with electrophoresis and allelic ladders, allele designations are made against a reference DNA sequence specified for the locus**

D8S1179

Simple TCTA repeat Reference Sequence – G08710 (12 Repeats)

```
LOCUS       G08710                340 bp    DNA        linear    STS 05-FEB-1997
DEFINITION  human STS CHLC.GATA7G07.P6384 clone GATA7G07, sequence tagged site.
ACCESSION   G08710
VERSION     G08710.1  GI:939260
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
              Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 340)
  AUTHORS   Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
  TITLE     Cooperative Human Linkage Center
  JOURNAL   Unpublished (1995)
COMMENT     Synonyms: GATA7G07, CHLC.GATA7G07.#T6383
FEATURES             Location/Qualifiers
     source             1..340
                       /organism="Homo sapiens"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:9606"
     STS                17..193
     primer_bind       17..41
     primer_bind       complement(169..193)
ORIGIN
  1 tggcaactta tatgtatctt tgtatctcat gtgtacattc gtatctatct atctatctat
  61 ctatctatct atctatctat ctatctatct attccccaca gtgaaaataa tctacaggat
 121 aggtaaataa attaaggcat attcacgcaa tgggatacgn tacagtgatg aaaatgaact
 181 aattatagct acgtgaaact atactcatgn acacaatttg gtaaaagaaa ctgggaacaa
 241 qaatacatac qttttttqnc aqctqtqcta ttttacattc ccaacaacaa tqcacaqqqt
 301 ttcagnttct ccacatnctt gtcaacattn tgttatcttg
//
```

G08710 REFERENCE SEQUENCE COURTESY OF THE NATIONAL CENTER FOR BIOTECHNOLOGY
INFORMATION AND THE NATIONAL LIBRARY OF MEDICINE

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?list_uids=939260&db=nucore&dopt=gb

D8S1179

Simple TCTA repeat

Reference Sequence – G08710 (12 Repeats)

```
LOCUS       G08710                340 bp    DNA        linear    STS 05-FEB-1997
DEFINITION human STS CHLC.GATA7G07.P6384 clone GATA7G07, sequence tagged site.
ACCESSION   G08710
VERSION     G08710.1  GI:939260
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
```

```
1 tggcaactta tatgtatttt tgtattttcat gtgtacattc gtatctatct atctatctat
61 ctatctatct atctatctat ctatctatct attccccaca gtgaaaataa tctacaggat
121 aggtaaataa attaaggcat attcacgcaa tgggatacgn tacagtgatg aaaatgaact
181 aattatagct acgtgaaact atactcatgn acacaatttg gtaaaagaaa ctgggaacaa
241 gaatacatac ggtttttqnc agctgtgcta ttttacattc ccaacaacaa tgcacagggt
301 ttcagnttct ccacatnctt gtcaacattn tgttatttttg
```

```
PRIMER_D1179      Complement(109..195)
ORIGIN
1 tggcaactta tatgtatttt tgtattttcat gtgtacattc gtatctatct atctatctat
61 ctatctatct atctatctat ctatctatct attccccaca gtgaaaataa tctacaggat
121 aggtaaataa attaaggcat attcacgcaa tgggatacgn tacagtgatg aaaatgaact
181 aattatagct acgtgaaact atactcatgn acacaatttg gtaaaagaaa ctgggaacaa
241 gaatacatac qtttttqnc agctgtgcta ttttacattc ccaacaacaa tqcacagqqt
301 ttcagnttct ccacatnctt gtcaacattn tgttatttttg
//
```

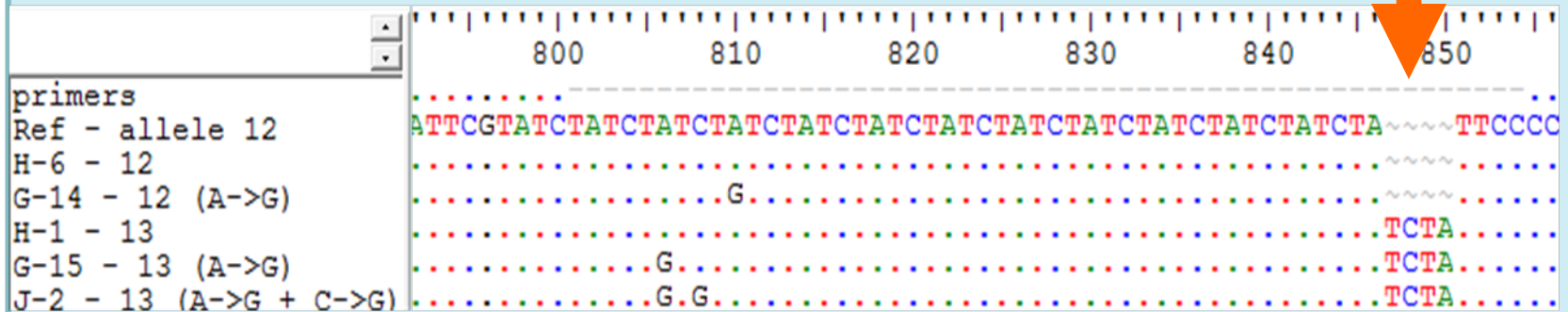
G08710 REFERENCE SEQUENCE COURTESY OF THE NATIONAL CENTER FOR BIOTECHNOLOGY
INFORMATION AND THE NATIONAL LIBRARY OF MEDICINE

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?list_uids=939260&db=nucore&dopt=gb

D8S1179

12 Allele represented by reference

12 and 13 Alleles containing various SNPs

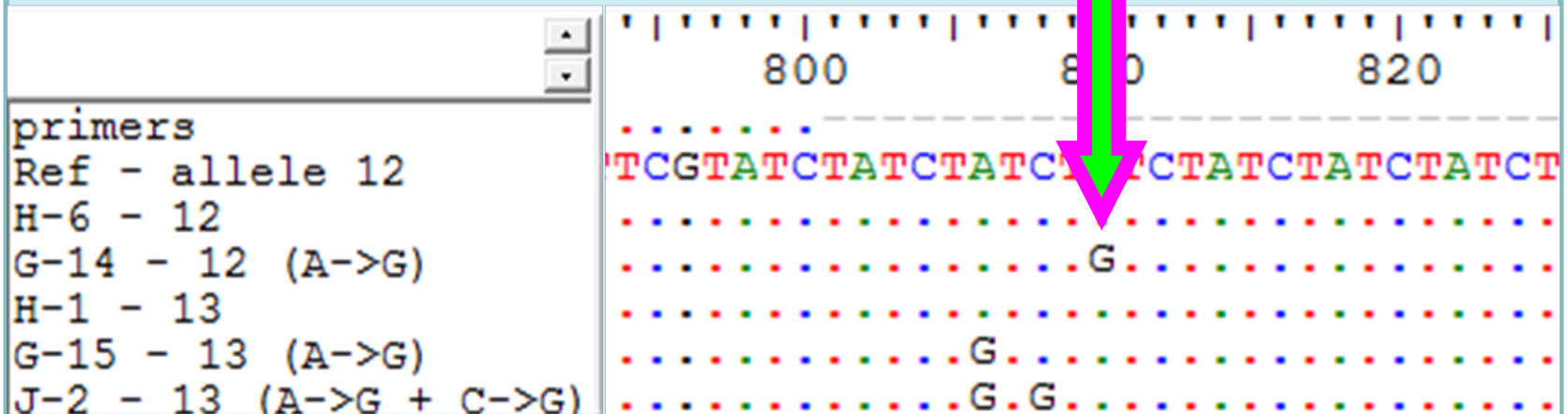
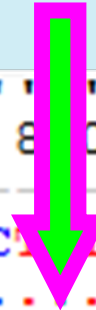


Length variations consist of repeat motifs added to the end of reference sequence

D8S1179

12 Allele containing a A→G SNP = 12S2

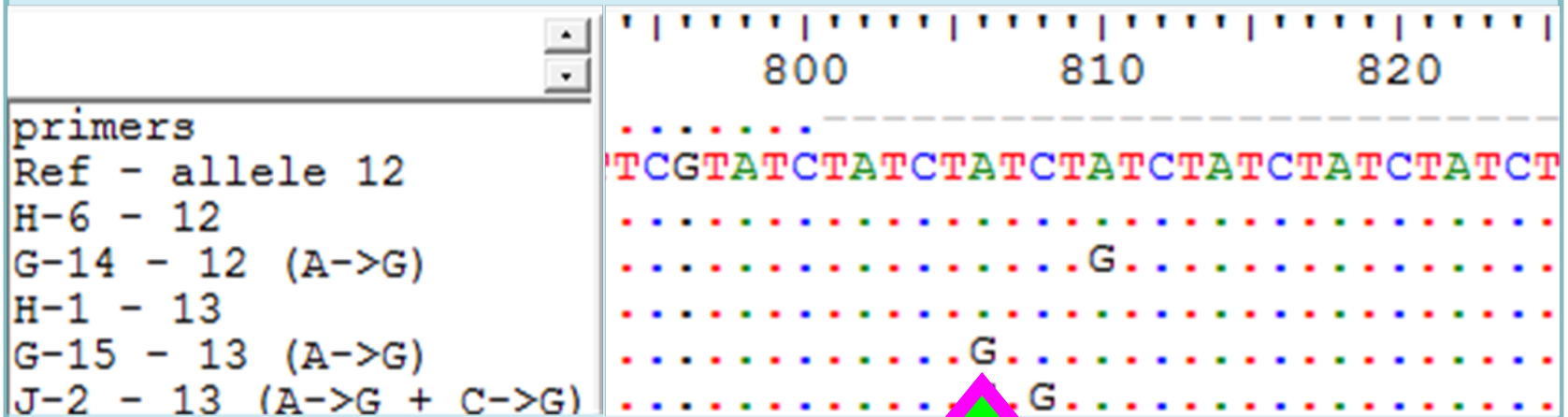
12S2



D8S1179

12 Allele containing a A→G SNP = 12S2

13 Allele containing a A→G SNP = **13S2**



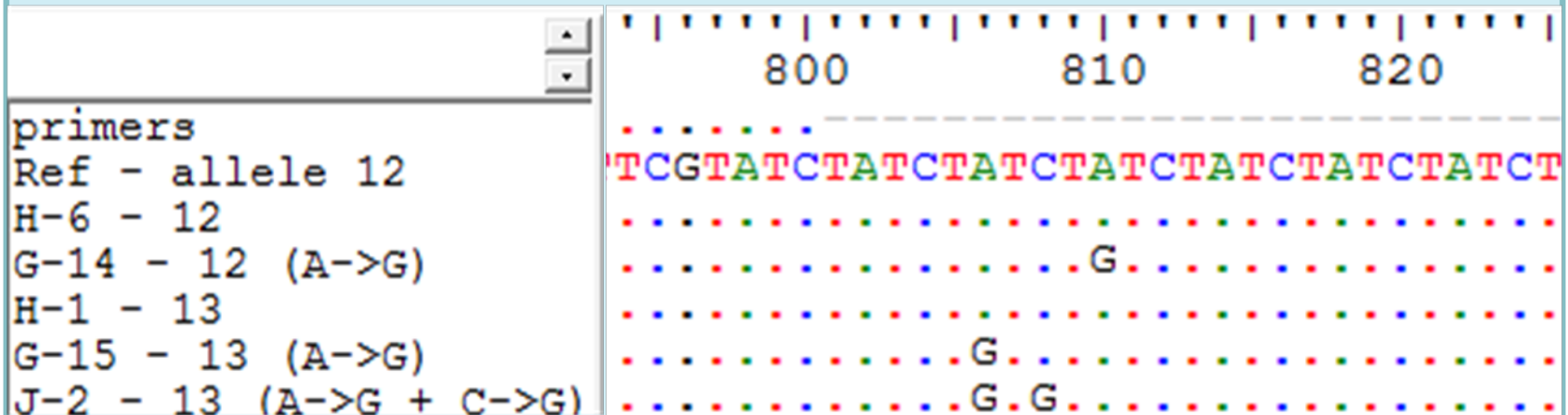
13S2

D8S1179

12 Allele containing a A→G SNP = 12S2

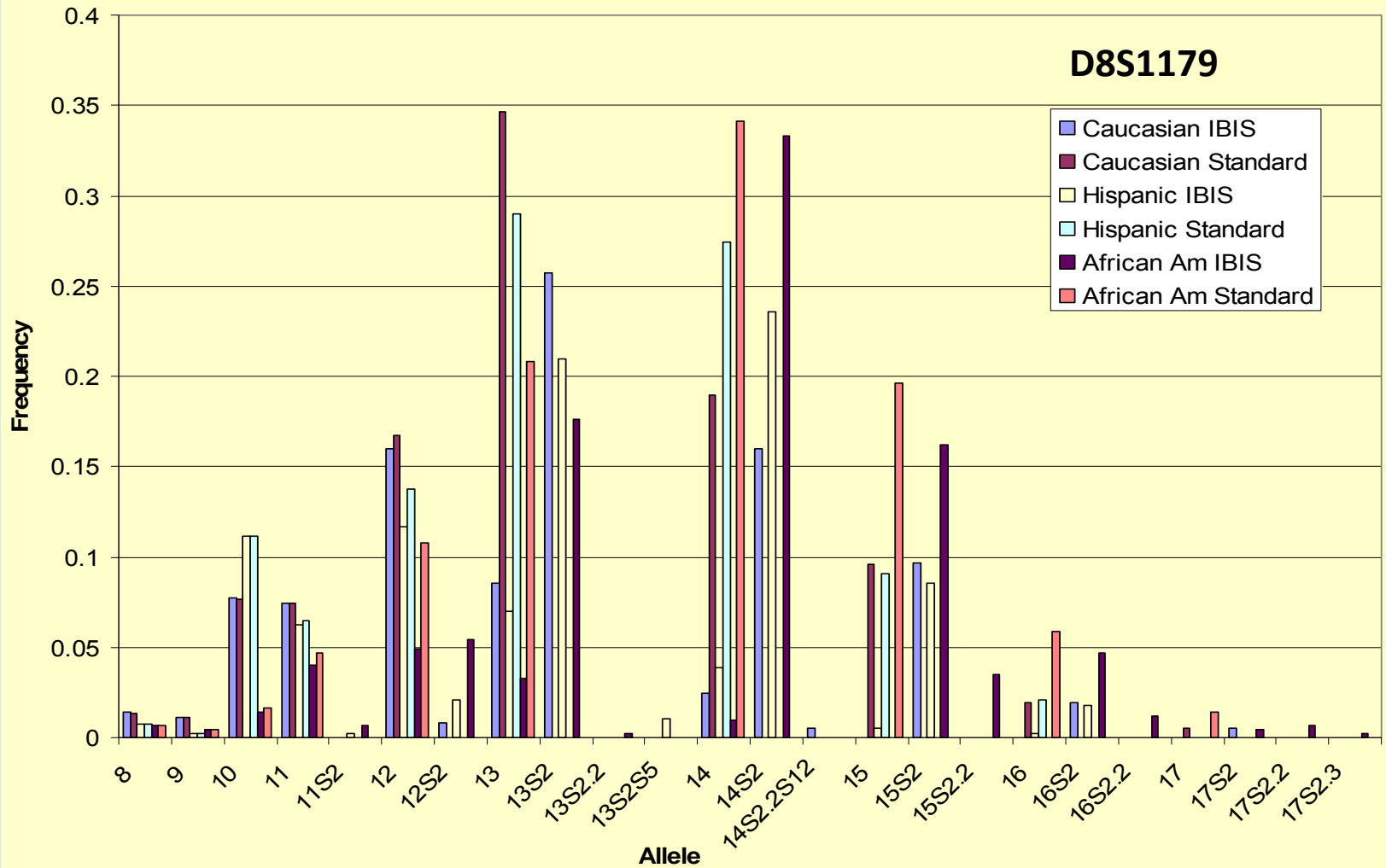
13 Allele containing a A→G SNP = 13S2

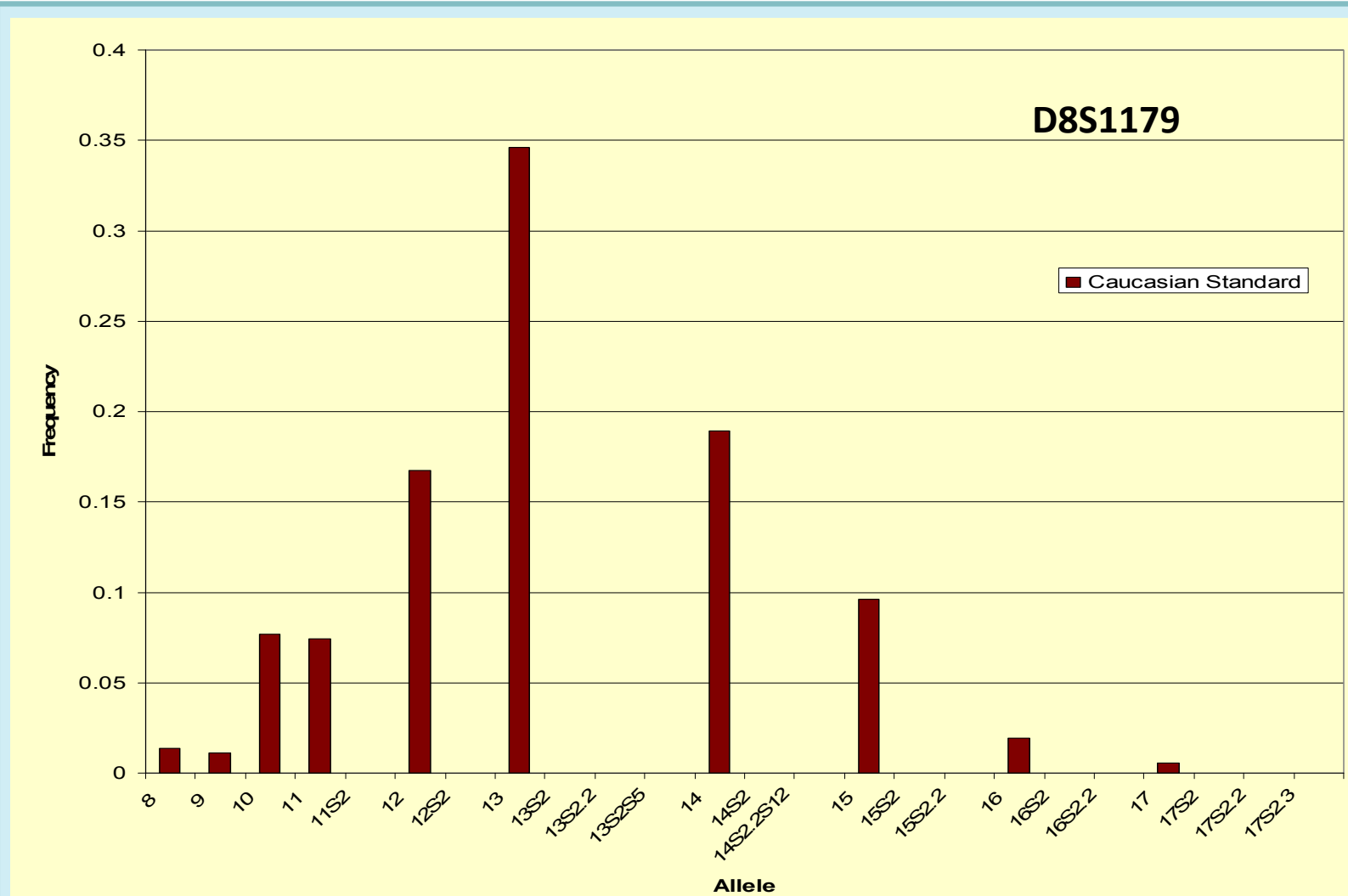
13 Allele containing a A→G and C→G SNP = **13S2S5**



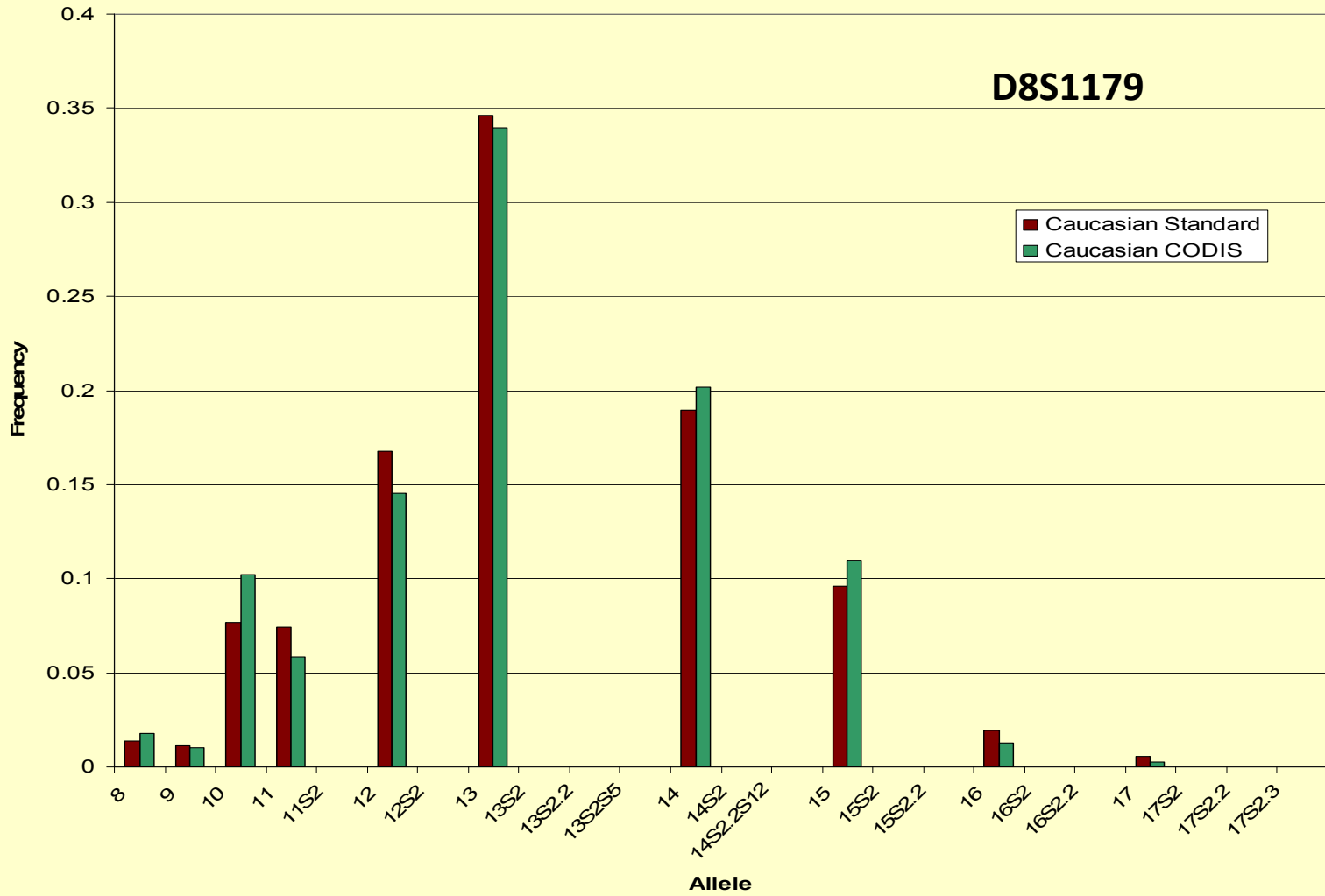
13S2S5

D8S1179

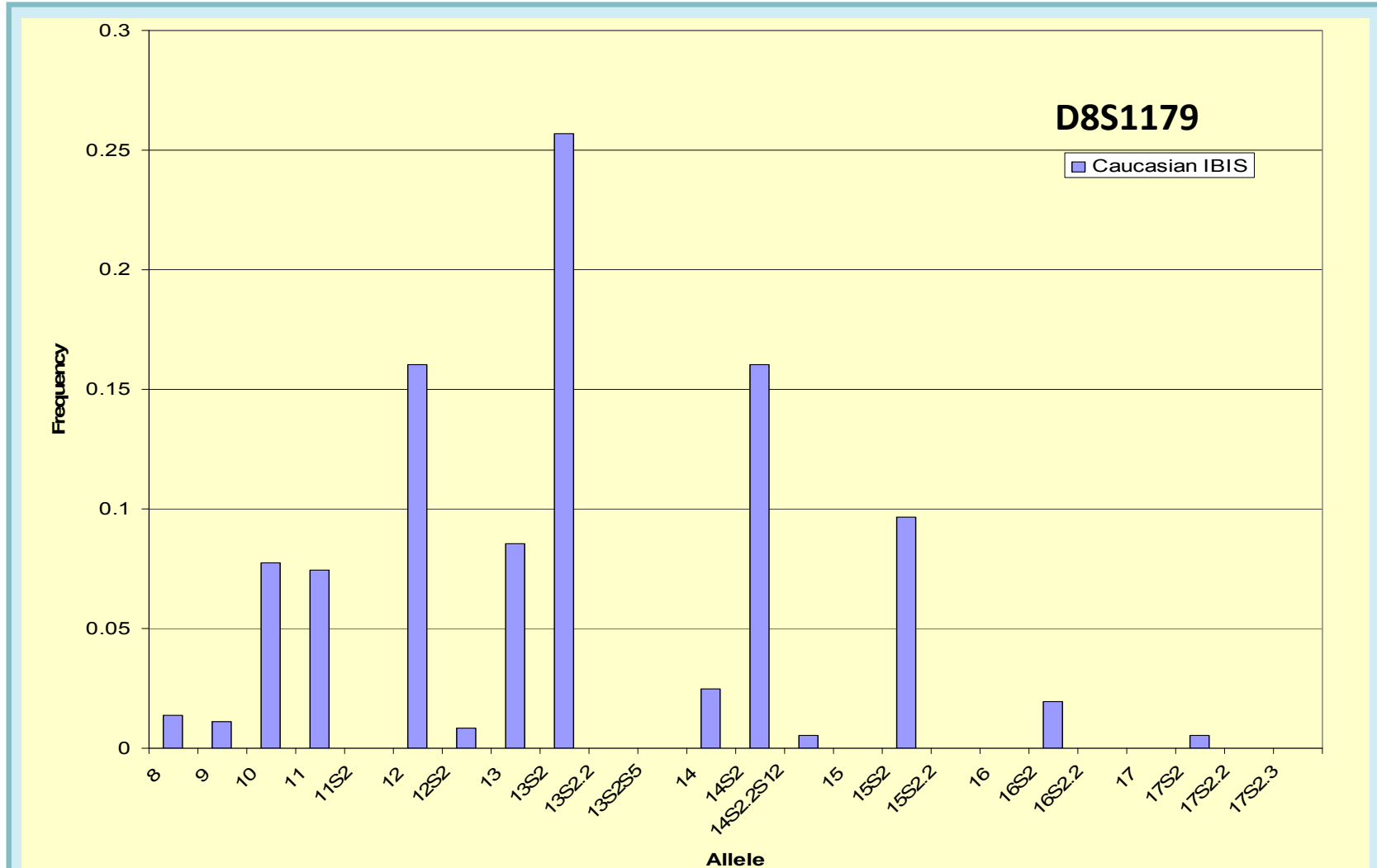




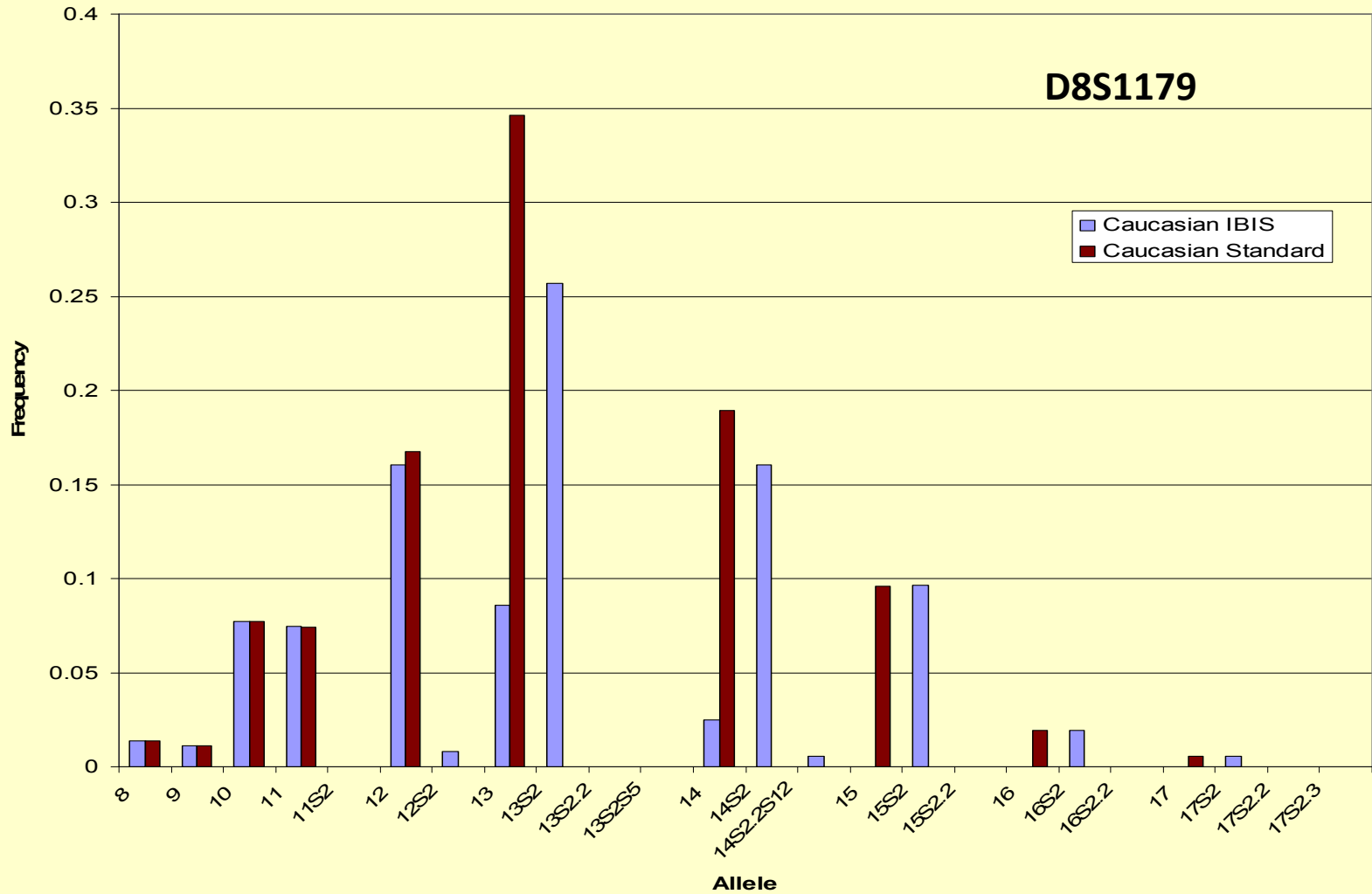
D8S1179



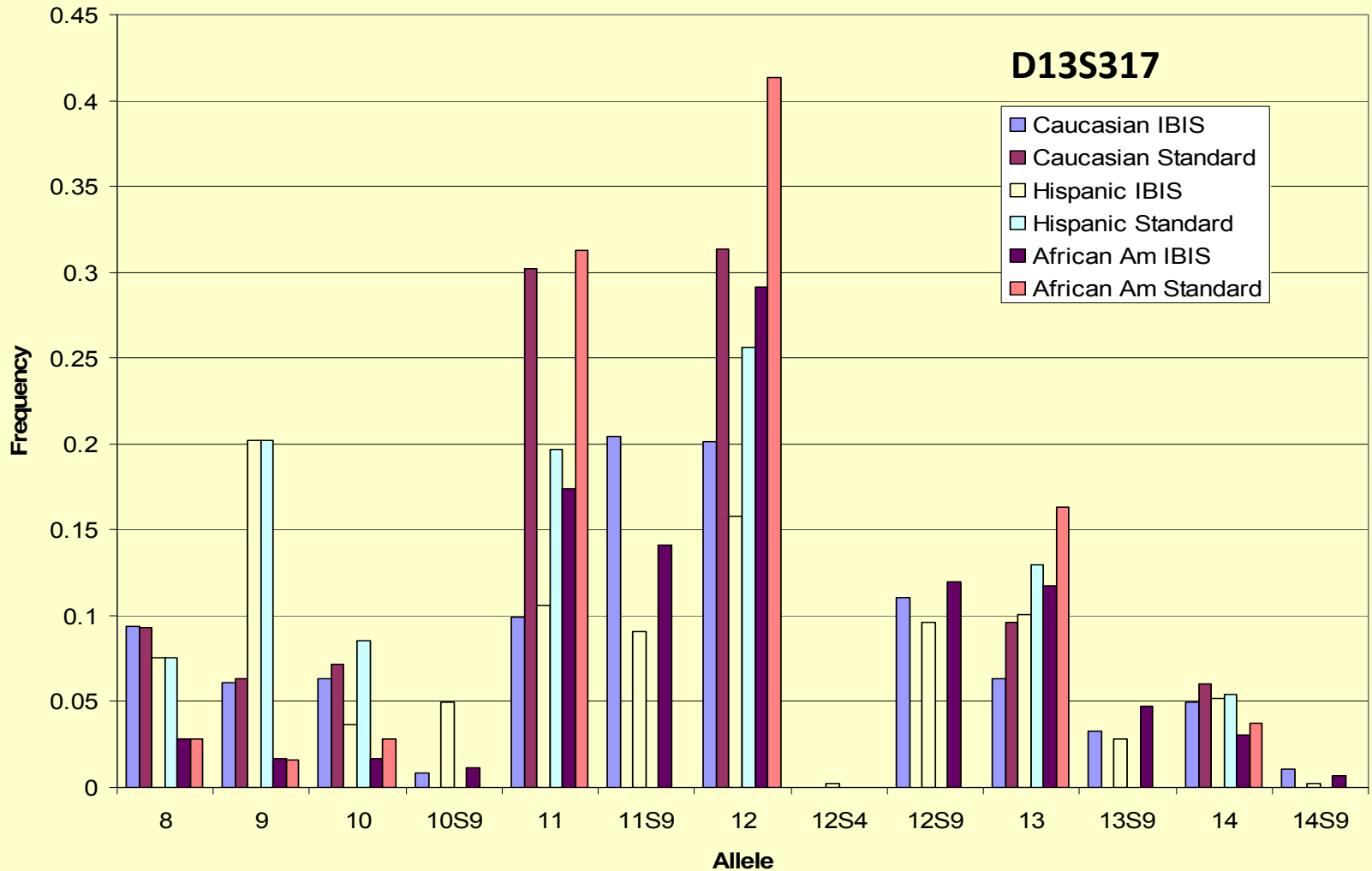
CODIS Data provided by Budowle et al. 1999

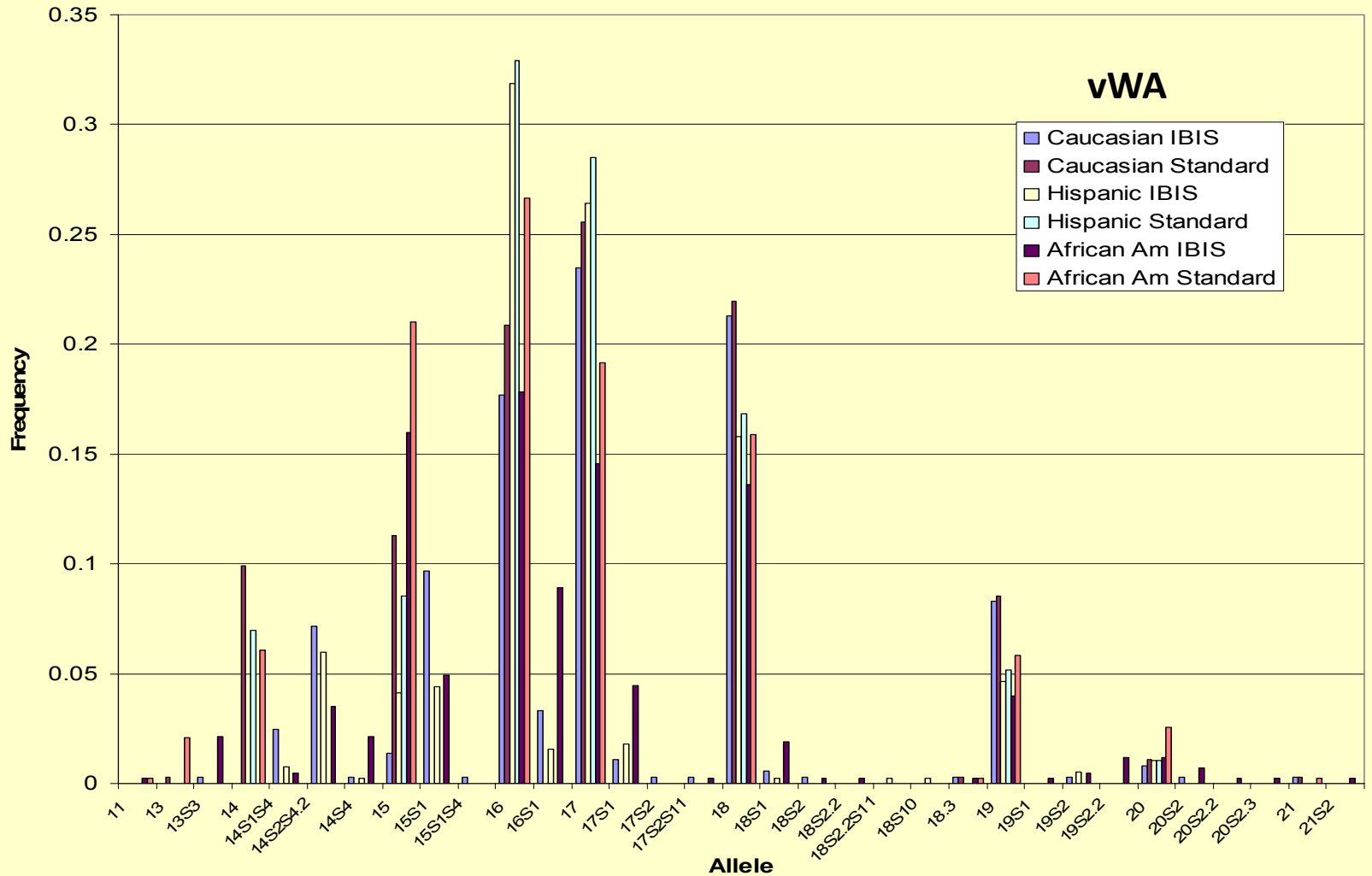


D8S1179

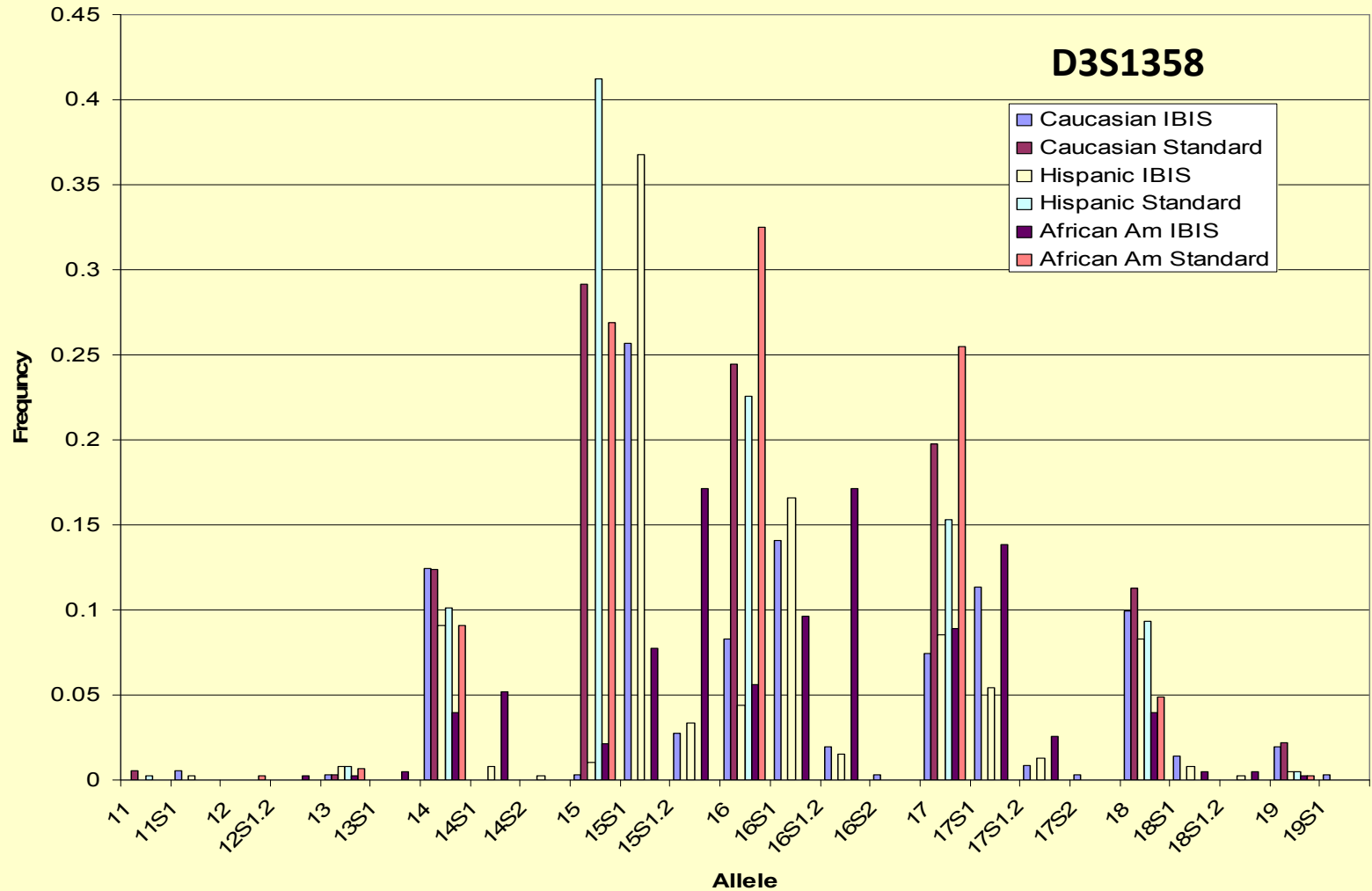


D13S317

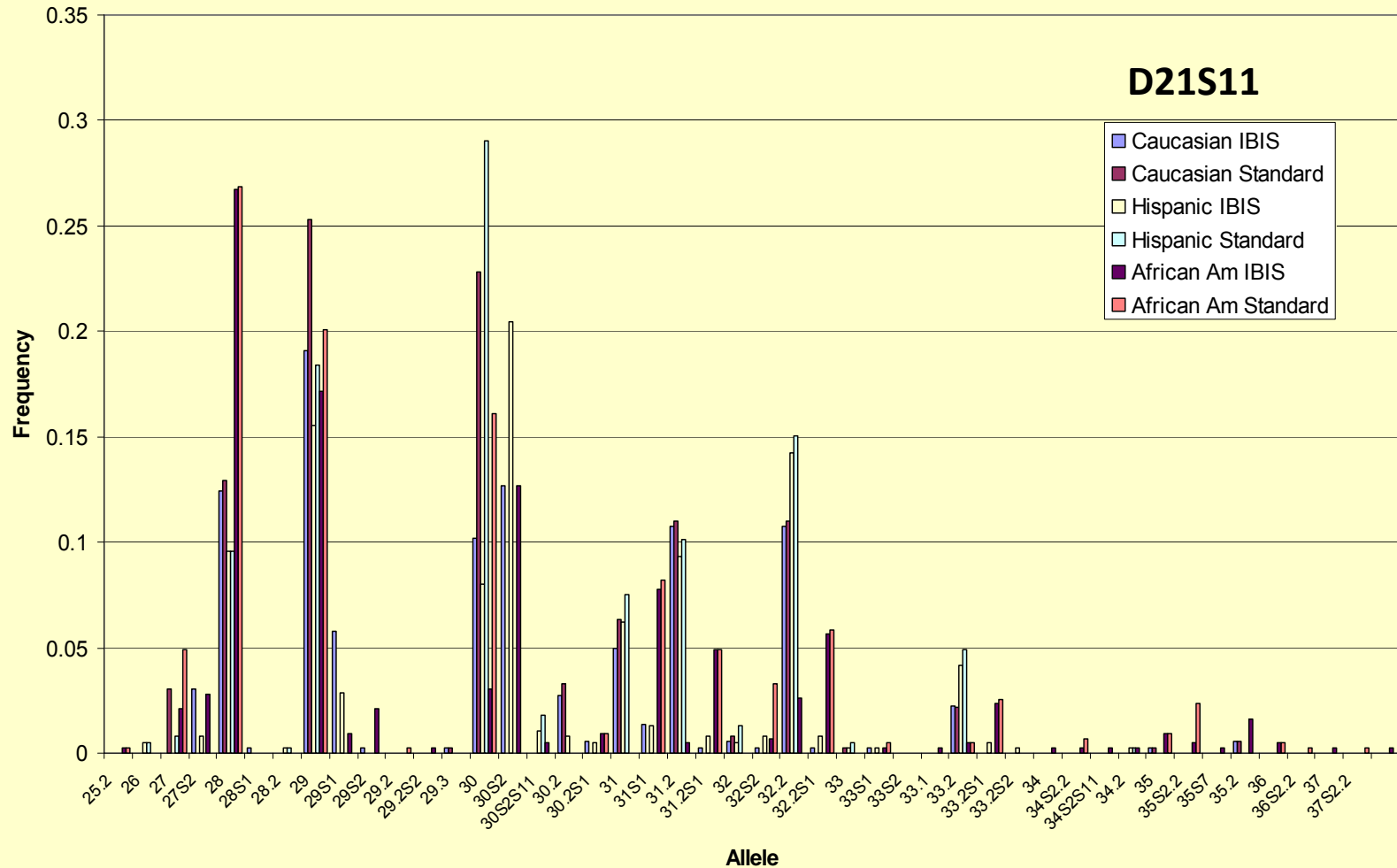




D3S1358



D21S11

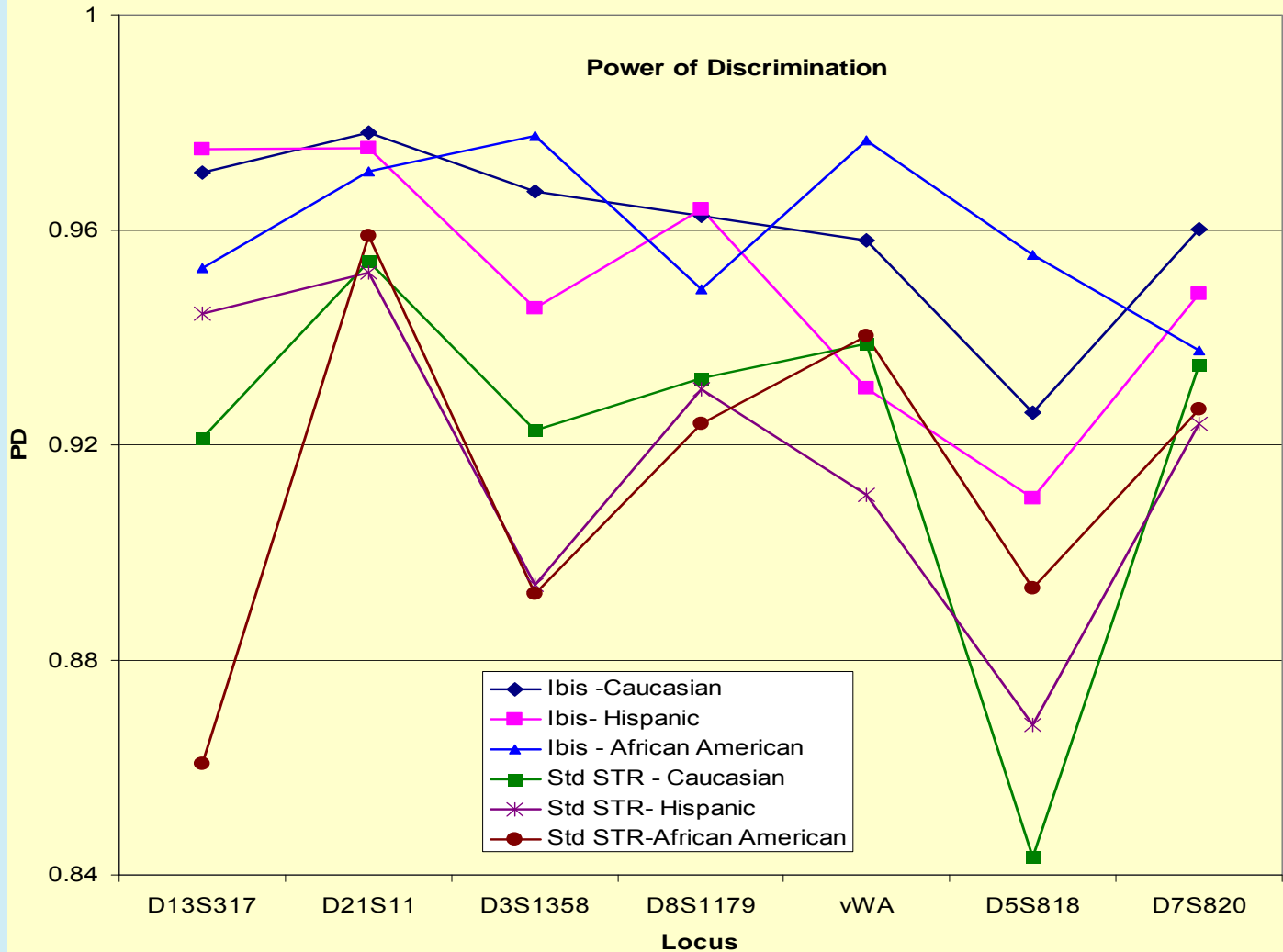


SNPs within STRs Open an Expanded Realm of Applications and Interpretational Power

- **Identity Testing**
 - Increased PE and PD
 - Overall reduction of homozygotes
 - More information without changing the locus panel
 - Complete reverse compatibility with existing databases
- **Relationship Testing**
 - SNPs can be tracked within a lineage
 - STR/SNP combo becomes a haplotype
 - Allele transfer in mutation cases can be verified

Discriminatory Power

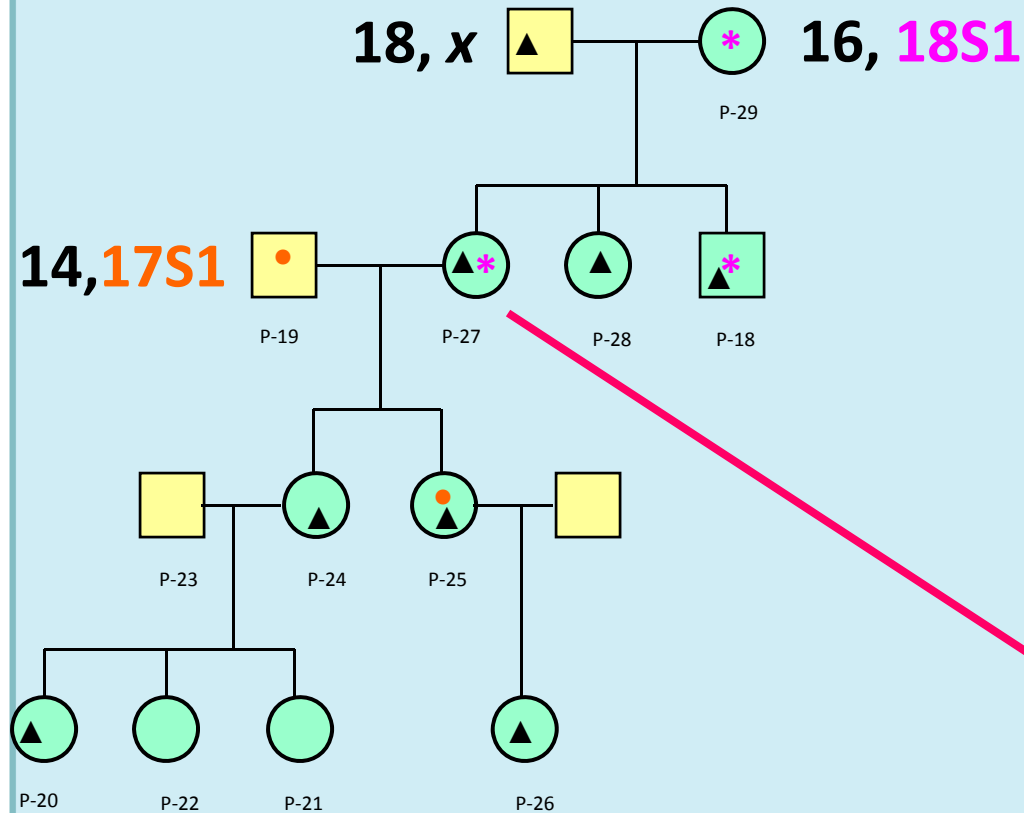
- **Capture of the SNPs within the STRs increases the Combined Power of Discrimination of the seven loci to approximate that of any 10 regularly typed STR loci**
- **Random match probability for 50 individuals selected from each population group decreased on average 10^{-2} to 10^{-4} for the panel of seven loci as compared to standard STR profile statistics**



Uses in Relationship Testing

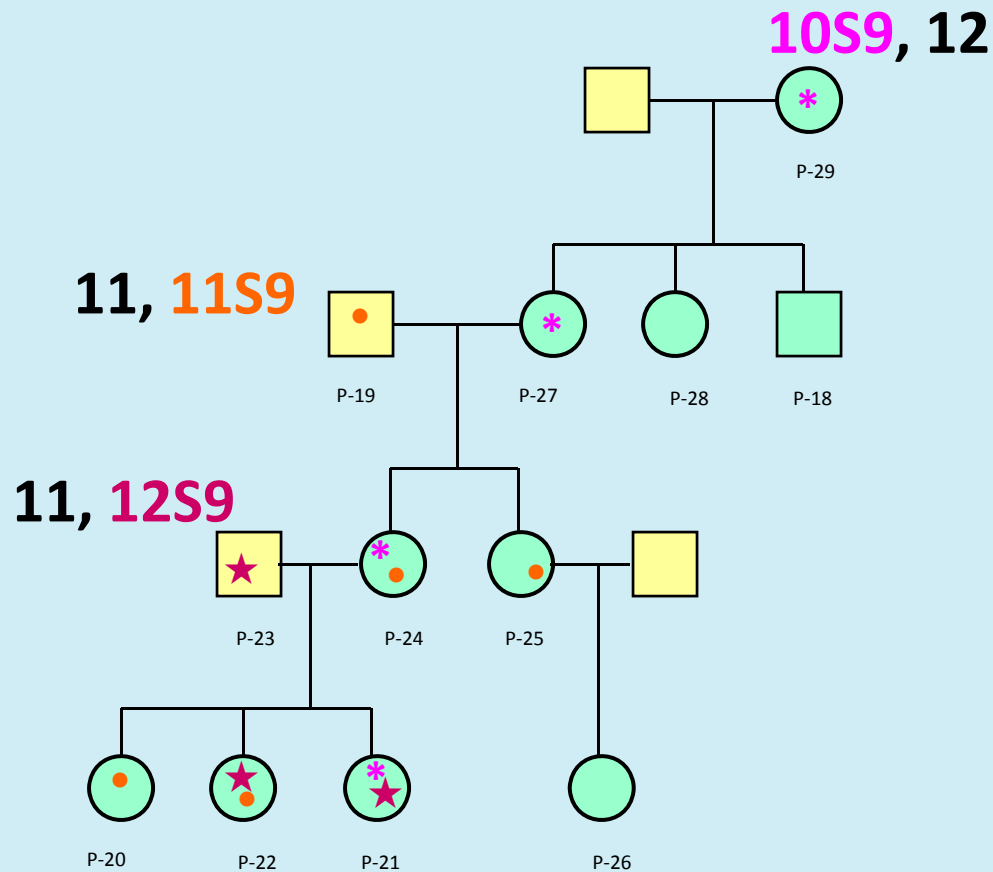
- **SNPs within STR repeats allow individual alleles to be tracked through a pedigree**
- **This enhances the ability to successfully associate relatives in mass disaster and missing persons scenarios**
- **As the loci have increased discriminatory capacity, alleles that match in low stringency “familial” searches have a greater probability of hitting a relative, reducing fortuitous associations obtained in pairwise database searches**

D3S1358



Sample	Genotype
P-18	18, 18S1
P-19	14, 17S1
P-20	14, 18
P-21	14, 14
P-22	14, 14
P-23	14, 15S1
P-24	14, 18
P-25	17S1, 18
P-26	18, 18
P-27	18, 18S1
P-28	16, 18
P-29	16, 18S1

D13S317

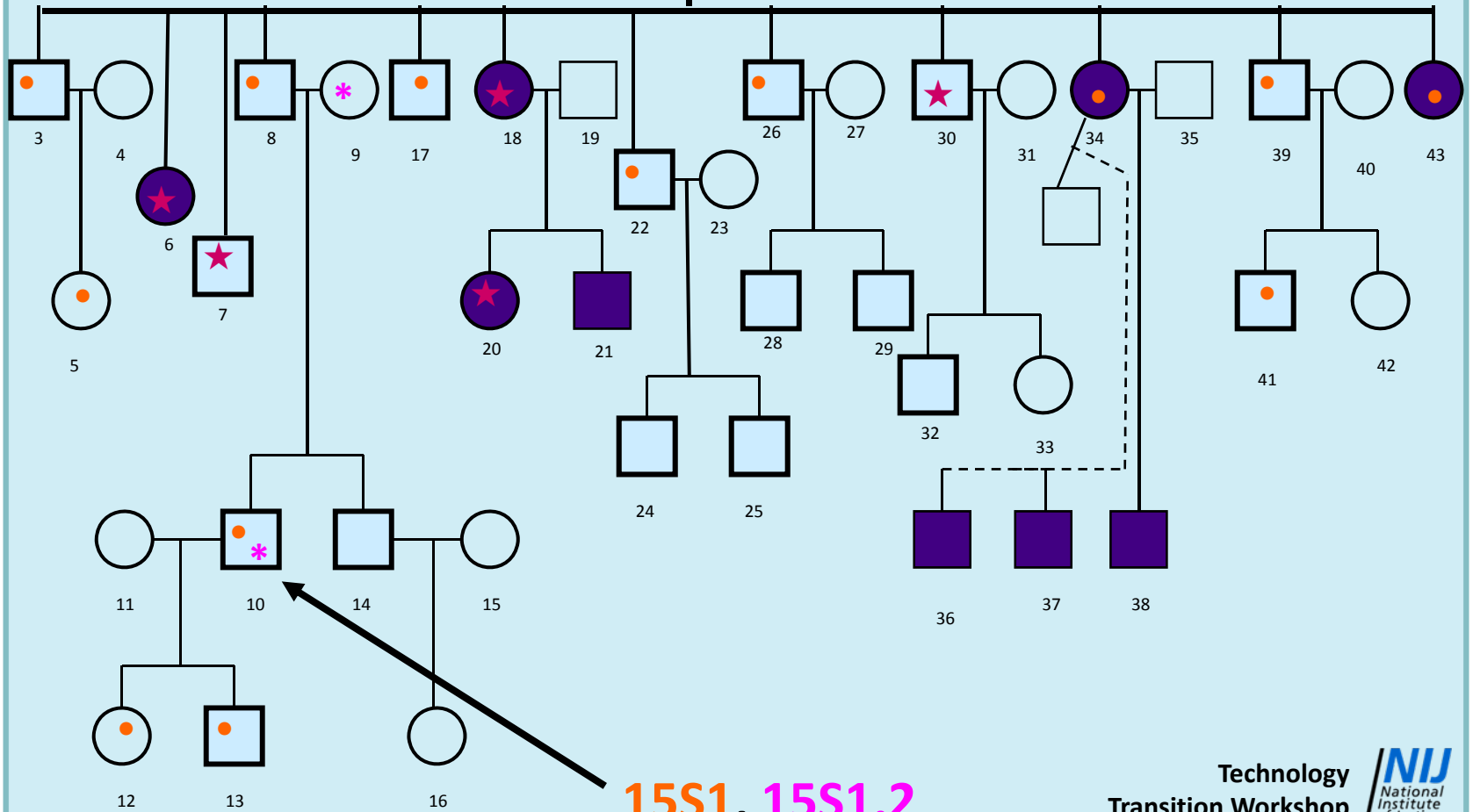


Sample	Genotype
P-18	12, 12
P-19	11, 11S9
P-20	11, 11S9
P-21	10S9, 12S9
P-22	11S9, 12S9
P-23	11, 12S9
P-24	10S9, 11S9
P-25	11S9, 12
P-26	9, 12
P-27	10S9, 12
P-28	12, 12
P-29	10S9, 12

D3S1358

15S1 16S1

1 2 14, 18



Much To Do!

- **Expand database size to 400 to 500 individuals per population group**
- **Sequence a number of individuals sharing a common allele to evaluate effect on population structure**
- **Type and sequence a number of parentage trios exhibiting single non-matching systems in SNP containing loci**
- **Evaluate additional STRs (i.e. NCO's, Y STRs, X STRs) for discriminatory value**

Special Thanks

- **Tom Hall**
- **Kristin Lowery**
- **Suzanne Gonzalez**
- **Pamela Marshall**
- **Elizabeth Feller**
- **Shubra Nandi**
- **Jennifer Thomas**

Questions?

Contact Information:

John V. Planz, Ph.D.

**University of North Texas Health Science Center
Department of Forensic and Investigative Genetics**

Institute of Investigative Genetics

3500 Camp Bowie Blvd., EAD 310

Fort Worth, TX 76107

jplanz@hsc.unt.edu

Note: All images, diagrams, and tables courtesy of
John V. Planz, Ph.D. unless otherwise noted.

Technology
Transition Workshop 