



## STRseq Bioproject

- Human identification STR marker sequence catalog
  - Comprehensive – future public submission capability planned
  - Each STR allele = unique GenBank Record and accession number
    - Non-redundant, well annotated
    - Includes flanking polymorphisms
- Hosted by U.S. National Center for Biotechnology Information (NCBI)
- Organized by STR marker
  - Grouped into Common STR, Alternate STR, Y-STR, X-STR

Modeled after NCBI RefSeq

**RefSeq: NCBI Reference Sequence Database**  
A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

## Why Do We Need STR Nomenclature?

08S1179

90

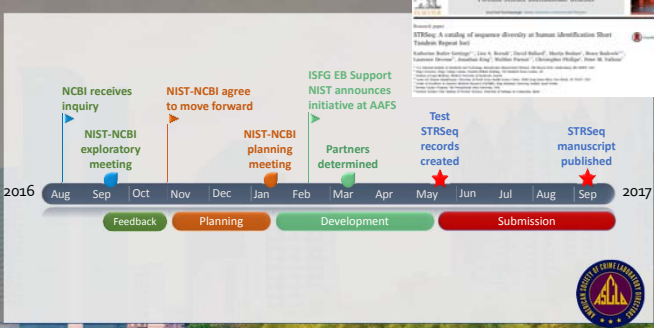
12 9863

No difference by CE

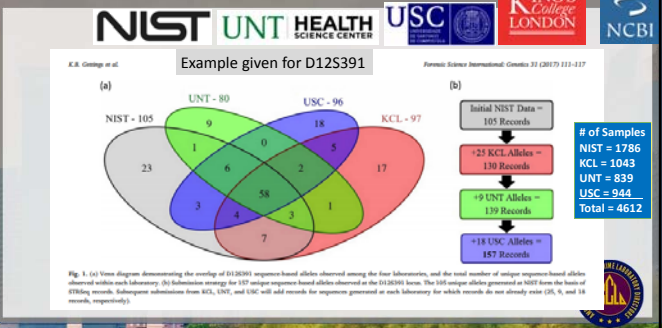
[TCTA]2 TCTG [TCTA]9  
OR  
[TCTA]12

Sequence variation is discriminated by NGS

## STRseq Bioproject Timeline



## STRseq International Collaboration



## STRseq Records – One For Each STR Allele

NCBI RefSeq

Search: TPOX 7 [AATG]7 rs115644759

1 174 bp linear DNA  
Accession: MN042247.1 | GI: 108124208  
BlastLink: BLAST | FASTA: FASTA

2 170 bp linear DNA  
Accession: MN042248.1 | GI: 108124209  
BlastLink: BLAST | FASTA: FASTA

3 166 bp linear DNA  
Accession: MN042249.1 | GI: 108124210  
BlastLink: BLAST | FASTA: FASTA

4 168 bp linear DNA  
Accession: MN042250.1 | GI: 108124211  
BlastLink: BLAST | FASTA: FASTA

GenBank: MF044247.1

FASTA

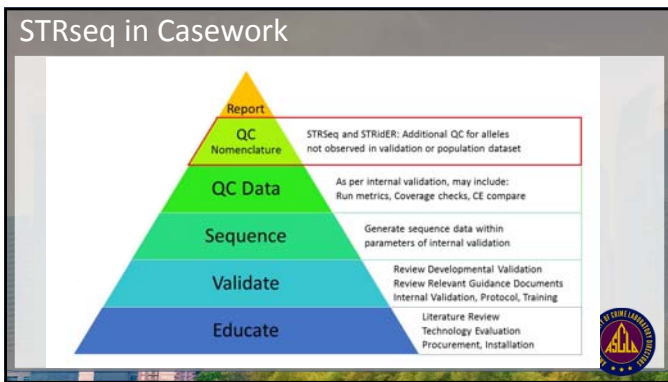
Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

```

>MF044247.1 Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence
TGGCTCTGGGTCCTCCCACTAGATTGAGCCAGGAGGAGGCTGTGTTTCAGGCTGTGATCACTAG
CACCCAGACCGTGCAGTGCAGAGAACAGCCACTTAAAGGAAACCCCTACTGAATGAATGAATGAAT
GAATGAATGTTTGGCAATAA
    
```

bioinformatics report for microsatellite TPOX. The left pane shows a list of alleles with their sizes and frequencies. The right pane shows a chromatogram. A red box highlights the 'Allele' column in the list.

dbSNP Short Genetic Variations. Reference SNP Cluster Report: rs11584739. The image shows the dbSNP website interface with a blue arrow pointing to the 'rs11584739' label.



### Future Directions

Forensic Science International: Genetics. Research Paper | FSIGEN\_2018\_047. Sequence-based U.S. population data for 27 autosomal STR loci. Katherine Gettings, Carolyn Steffen, Kevin Kiesler, Lisa Borsuk, Peter Vallone. Submitted 26 Apr 2018. Reviewer limited 29 Apr 2018. Alert: keep me informed about the submission status for this manuscript.

**SUBMITTED:** Sequence-based population frequency data for 1036 samples, 27 autosomal STRs

NIST engages in collaborative research with assay developers and instrument manufacturers to support competitiveness in American industry.

### Thank You!

- STRseq Collaborators: KING'S College LONDON, UNT HEALTH SCIENCE CENTER, USC
- NIST: Dr. Pete Vallone, Group Leader; Dr. Katherine Gettings; Lisa Borsuk

APPLIED GENETICS  
 Kevin.Kiesler@nist.gov  
 Phone: (301) 975-4306

**Disclaimer:** Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.