STRSeq:
A Resource for Sequence-Based STR Analysis

Katherine Gettings PhD Research Biologist – NIST USA
Co-Authors – Lisa Borsuk MS – Peter M. Vallone PhD

Disclaimers

Funding: NIJ Inter-Agency Agreement: Forensic DNA Application of Next Generation Sequencing.

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. All work presented has been reviewed and approved by the NIST Human Subjects Protections Office.

A copy of this presentation is available at: http://strbase.nist.gov/NISTpub.htm#Presentations

Inception

RefSeq: NCBI Reference Sequence Database

A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

Locus·Reference·Genomic

LRG sequences provide a stable genomic DNA framework for reporting mutations with a permanent ID and core content that never changes.

Timeline

INTERNATIONAL SOCIETY FOR FORENSIC GENETICS
Internationale Gesellschaft für Forensische Genetik e.V.
http://www.isfg.org

President

Prof. Dr. Walther Parson
Institute of Legal Medicine, Innsbruck Medical University
Müllerstrasse 44, A-6020 Innsbruck

Phone +43 512 9003 70640
Fax +43 512 9003 73640
Email walther.parson@i-med.ac.at

Katherine Butler Gettings, Ph.D.
Research Biologist, Applied Genetics Group
National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg, MD 20899-8314
United States of America

Innsbruck, 04.02.2017

BioProject: Initiative to collect, store and disseminate forensic STR sequence variants on GenBank (NCBI)

Dear Dr. Gettings,

The executive board of the ISFG entirely supports and endorses your initiative towards the planned BioProject, which aims to collect, store and disseminate newly observed STR sequence variation relevant to forensic genetics. We believe that this is an excellent project proposal. The professional environment through NCBI will allow for long-term storage of sequence data representing a valuable resource for forensic scientists and other stakeholders using sequence data in the future.

The BioProject will also be an important resource of STR sequence information for projects that support the forensic community with tools to evaluate STR sequence data, such as STRidER (www.strider.online) offering quality control of published sequences and a curated database of population data for forensic frequency estimates. The editors of the leading forensic genetic journal *Forensic Science International Genetics* (http://www.fsigenetics.com/) require authors of STR population studies to have their data quality controlled by STRidER before submitting a manuscript.

We hope that the BioProject is going to evolve in a fruitful way and wish you success for your work!

With best regards
on behalf of the Executive Board

Prof. Dr. Walther Parson
(ISFG President)

President: W. Parson, Innsbruck, A Vice President: M. Prinz, New York, US Secretary: P.M. Schneider, Cologne, G Treasurer: L. Gusmão, Rio de Janeiro, BR Representative of the Working Parties: J. Butler, Gaithersburg, US

NCBI receives inquiry

NIST-NCBI exploratory meeting

NIST-NCBI agree to move forward

NIST-NCBI planning meeting

ISFG EB Support
NIST announces initiative at AAFS

Partners determined

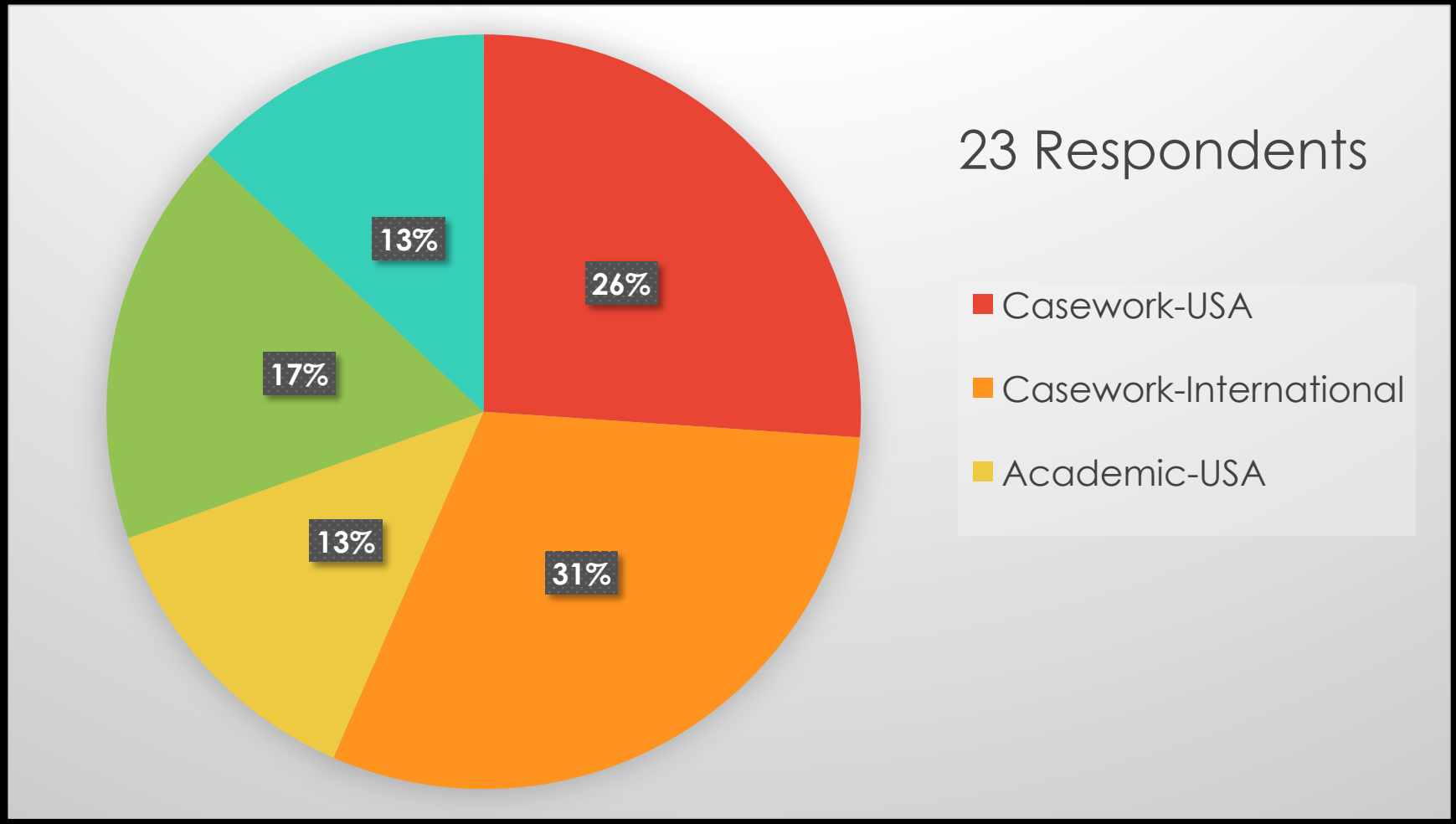
First STRSeq records released

STRSeq manuscript submitted

2016 | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | 2017

Feedback | Planning | Development | Submission

Feedback


13%
26%
17%
13%
31%
23 Respondents
Casework-USA
Casework-International
Academic-USA

Planning

1. GenBank record is created for each unique sequence
2. Non redundant records; number of records per locus varies
3. Records include flanking regions with high confidence sequence
4. Records include length-based allele designations determined by CE
5. Records can expand for future additional flank
6. Records organized into BioProject for improved access

Partners-Roles

UNT HEALTH SCIENCE CENTER

NIST

Population Samples
Project Coordination
Record Submission

KING'S College LONDON

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Population Data
Project Input

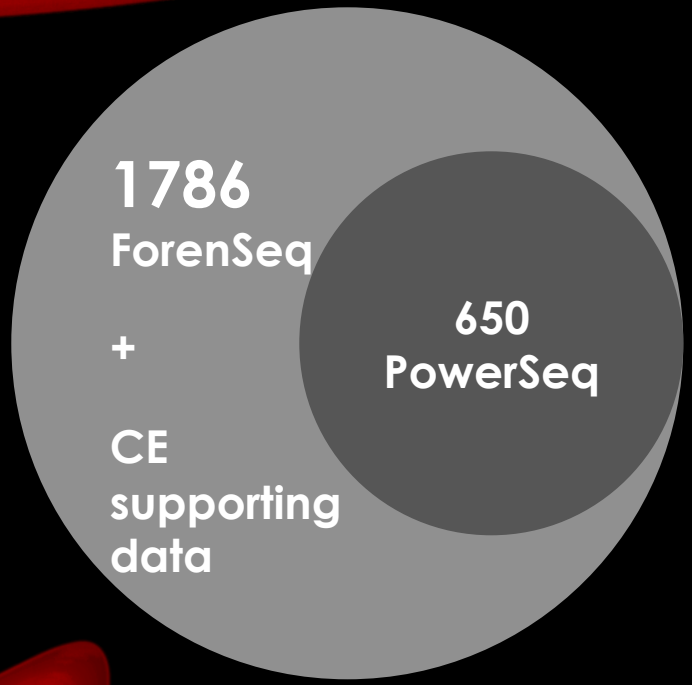
TIROL

Project Input
STRidER Integration

NCBI

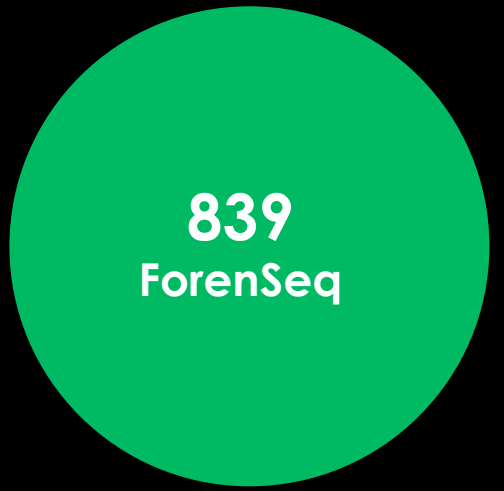
Project Input
Hosting

STRSeq Samples



NIST

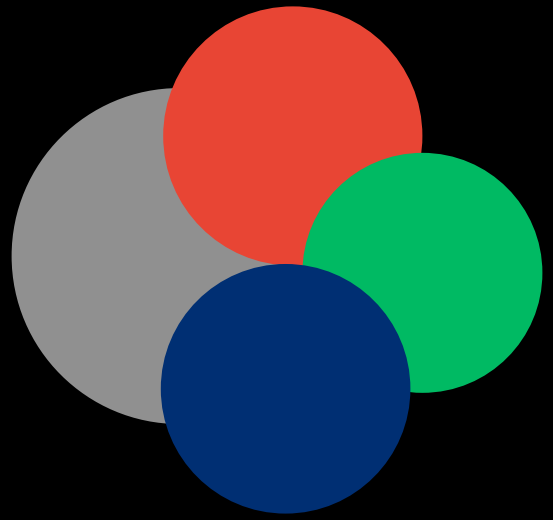
KING'S College LONDON



UNT HEALTH SCIENCE CENTER

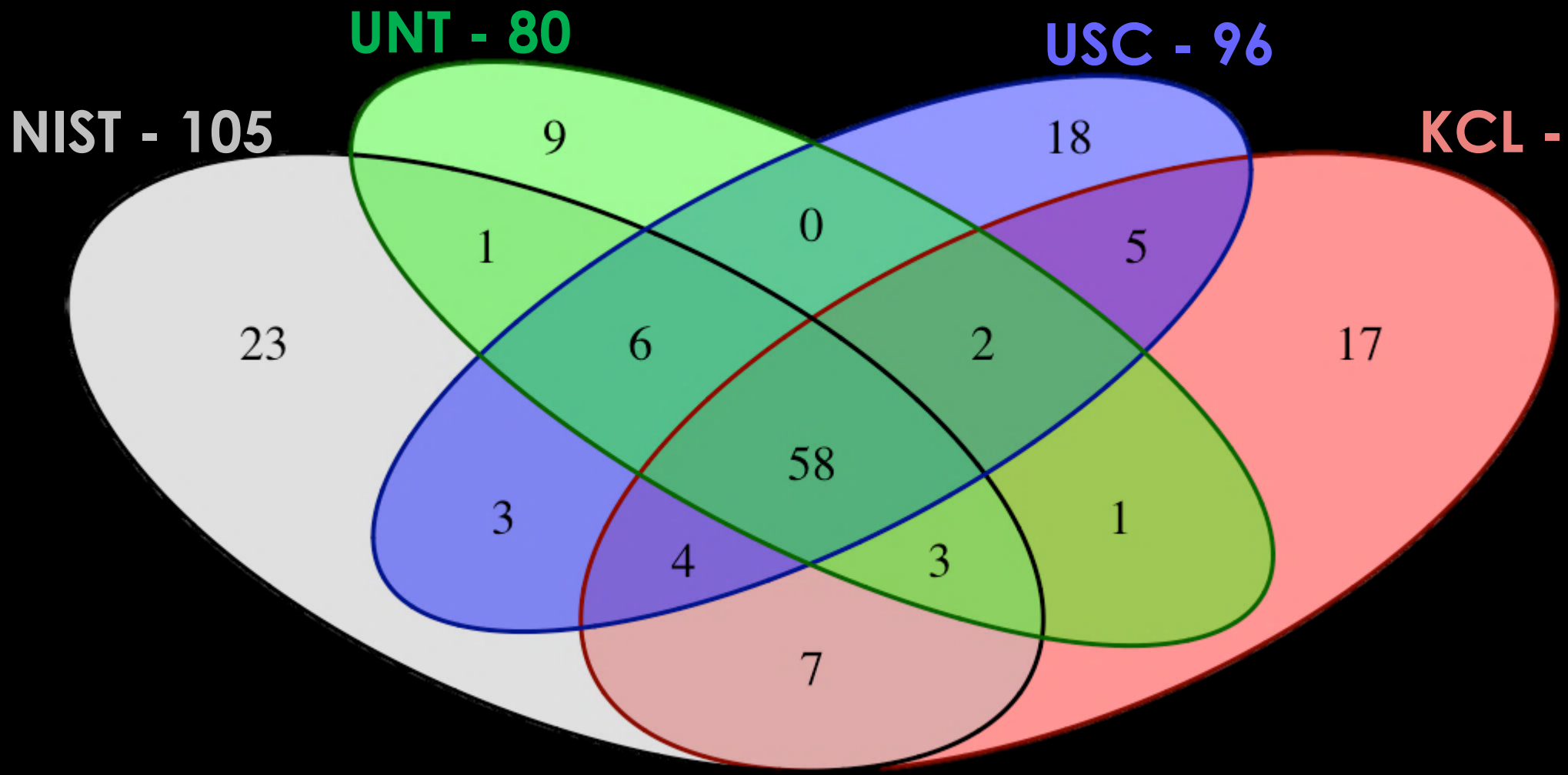
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

STRSeq Samples

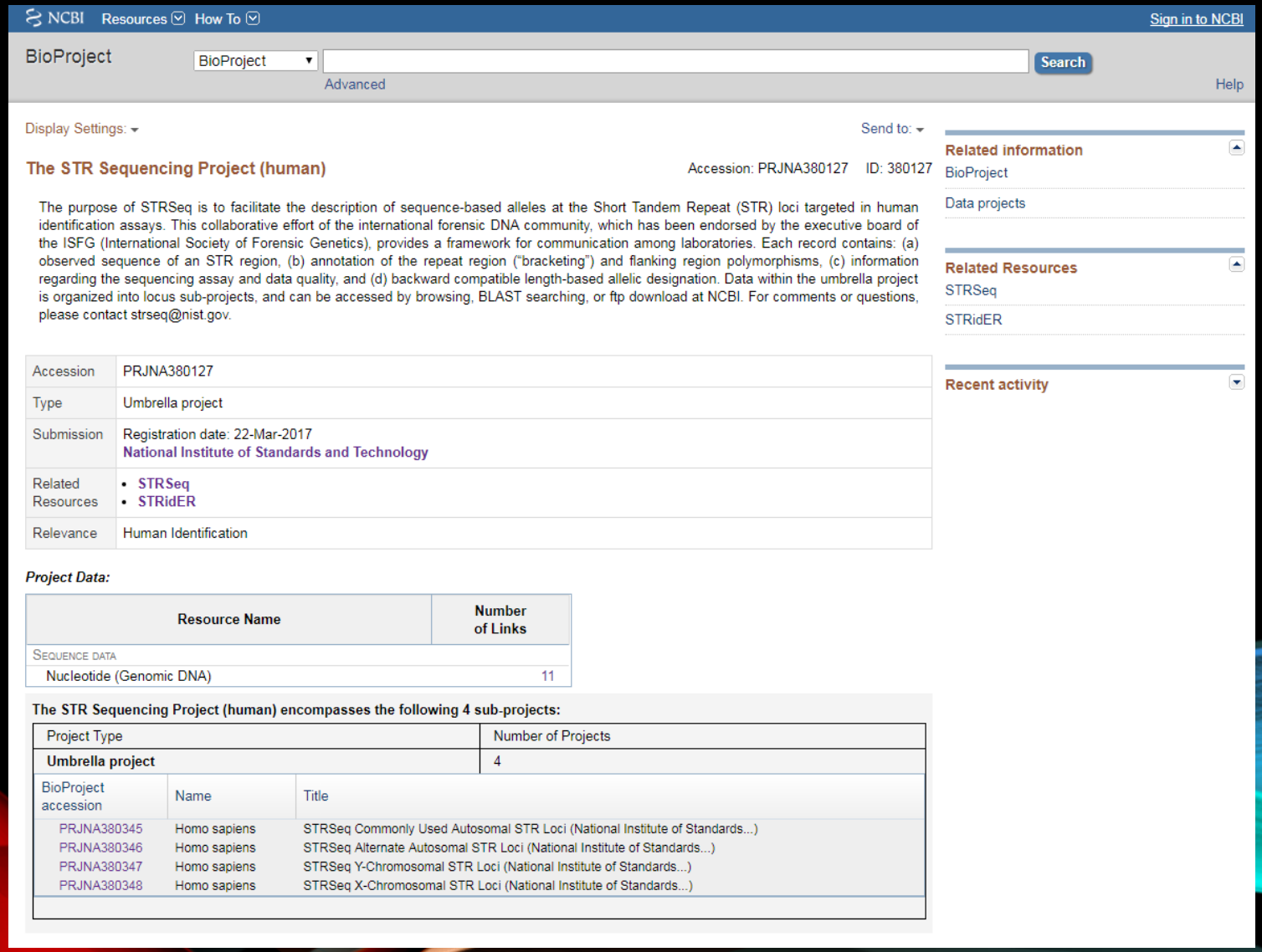


Aggregate alleles from 4612 samples

D12S391 Alleles by Lab



NCBI BioProject: 380127



NCBI Resources How To Sign in to NCBI

BioProject BioProject Search

Advanced Help

Display Settings: Send to:

The STR Sequencing Project (human)

Accession: PRJNA380127 ID: 380127

The purpose of STRSeq is to facilitate the description of sequence-based alleles at the Short Tandem Repeat (STR) loci targeted in human identification assays. This collaborative effort of the international forensic DNA community, which has been endorsed by the executive board of the ISFG (International Society of Forensic Genetics), provides a framework for communication among laboratories. Each record contains: (a) observed sequence of an STR region, (b) annotation of the repeat region ("bracketing") and flanking region polymorphisms, (c) information regarding the sequencing assay and data quality, and (d) backward compatible length-based allelic designation. Data within the umbrella project is organized into locus sub-projects, and can be accessed by browsing, BLAST searching, or ftp download at NCBI. For comments or questions, please contact strseq@nist.gov.

Accession	PRJNA380127
Type	Umbrella project
Submission	Registration date: 22-Mar-2017 National Institute of Standards and Technology
Related Resources	• STRSeq • STRidER
Relevance	Human Identification

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	11

The STR Sequencing Project (human) encompasses the following 4 sub-projects:

Project Type		Number of Projects
Umbrella project		4
BioProject accession	Name	Title
PRJNA380345	Homo sapiens	STRSeq Commonly Used Autosomal STR Loci (National Institute of Standards...)
PRJNA380346	Homo sapiens	STRSeq Alternate Autosomal STR Loci (National Institute of Standards...)
PRJNA380347	Homo sapiens	STRSeq Y-Chromosomal STR Loci (National Institute of Standards...)
PRJNA380348	Homo sapiens	STRSeq X-Chromosomal STR Loci (National Institute of Standards...)

Related information
BioProject
Data projects

Related Resources
STRSeq
STRidER

Recent activity

Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

GenBank: MF044247.1

FASTA Graphics

Go to:

```
LOCUS       MF044247               163 bp    DNA     linear   PRI 30-MAY-2017
DEFINITION  Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence.
ACCESSION   MF044247
VERSION     MF044247.1
DBLINK      BioProject: PRJNA380554
KEYWORDS    STRSeq, STR, TPOX.
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 163)
  AUTHORS   Gettings,K.B., Borsuk,L.A. and Vallone,P.M.
  TITLE     The STR Sequencing Project [manuscript in preparation]
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163)
  AUTHORS   NIST,A.G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2017) Applied Genetics Group, National Institute
            of Standards and Technology, 100 Bureau Drive, MS-8314,
            Gaithersburg, MD 20899, USA
COMMENT     Annotation ('bracketing') of the repeat region is consistent with
            the guidance of the ISFG (International Society of Forensic
            Genetics), PMID: 26844919.  Lower case letters in the 'Bracketed
            repeat' region below denote uncounted bases.   The given
            length-based allele value was determined using the designated
            length-based technology.  Variation in the length-based allele
            between individuals or assays can result from indels in flanking
            regions.  The length of reported sequence is dependent on the assay
            (see 'Sequencing technology') and the quality of the flanking
            sequence. This information is provided as part of the STR
            Sequencing Project (STRseq), a collaborative effort of the
            international forensic DNA community.  The purpose of this project
            is to facilitate the description of sequence-based STR alleles.
            Additional resources can be found at strseq.nist.gov.  For
            questions or feedback, please contact strseq@nist.gov.  Allele
            frequency data can be accessed in the strider.online database.

            ##HumanSTR-START##
            STR locus name        :: TPOX
            Length-based allele   :: 7
            Bracketed repeat      :: [AATG]7
            Sequencing technology :: ForenSeq, MiSeq FGx; PowerSeq Auto, MiSeq
            Coverage              :: >30X
            Length-based tech.    :: PowerPlex Fusion, ABI3500xl
            Assembly              :: GRCh38 (GCF_000001405)
            Chromosome            :: 2
            RefSeq Accession      :: NC_000002.12
            Chrom. Location       :: 1489532..1489698
            Repeat Location       :: 1489653..1489684
            Cytogenetic Location  :: 2p25.3
            ##HumanSTR-END##
```

```
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

```
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1..163
                     /note="Promega PowerSeq Sequence"
     variation       25
                     /note="C/T SNP"
                     /db_xref="dbSNP:rs115644759"
     misc_feature    120..154
                     /note="Illumina ForenSeq Sequence"
     repeat_region   122..149
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

GenBank: MF044247.1

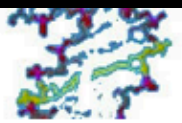
FASTA Graphics

Go to:

```
LOCUS       MF044247               163 bp    DNA     linear   PRI 30-MAY-2017
DEFINITION  Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence.
ACCESSION   MF044247
VERSION     MF044247.1
DBLINK      BioProject: PRJNA380554
KEYWORDS    STRSeq, STR, TPOX.
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 163)
  AUTHORS   Gettings,K.B., Borsuk,L.A. and Vallone,P.M.
  TITLE     The STR Sequencing Project [manuscript in preparation]
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163)
  AUTHORS   NIST,A.G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2017) Applied Genetics Group, National Institute
            of Standards and Technology, 100 Bureau Drive, MS-8314,
            Gaithersburg, MD 20899, USA
COMMENT     Annotation ('bracketing') of the repeat region is consistent with
            the guidance of the ISFG (International Society of Forensic
            Genetics), PMID: 26844919.  Lower case letters in the 'Bracketed
            repeat' region below denote uncounted bases.  The given
            length-based allele value was determined using the designated
            length-based technology.  Variation in the length-based allele
            between individuals or assays can result from indels in flanking
            regions.  The length of reported sequence is dependent on the assay
            (see 'Sequencing technology') and the quality of the flanking
            sequence. This information is provided as part of the STR
            Sequencing Project (STRseq), a collaborative effort of the
            international forensic DNA community.  The purpose of this project
            is to facilitate the description of sequence-based STR alleles.
            Additional resources can be found at strseq.nist.gov.  For
            questions or feedback, please contact strseq@nist.gov.  Allele
            frequency data can be accessed in the strider.online database.

            ##HumanSTR-START##
            STR locus name        :: TPOX
            Length-based allele   :: 7
            Bracketed repeat      :: [AATG]7
            Sequencing technology :: ForenSeq, MiSeq FGx; PowerSeq Auto, MiSeq
            Coverage              :: >30X
            Length-based tech.    :: PowerPlex Fusion, ABI3500xl
            Assembly              :: GRCh38 (GCF_000001405)
            Chromosome            :: 2
            RefSeq Accession      :: NC_000002.12
            Chrom. Location       :: 1489532..1489698
            Repeat Location       :: 1489653..1489684
            Cytogenetic Location  :: 2p25.3
            ##HumanSTR-END##
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1..163
                     /note="Promega PowerSeq Sequence"
     variation       25
                     /note="C/T SNP"
                     /db_xref="dbSNP:rs115644759"
     misc_feature    120..154
                     /note="Illumina ForenSeq Sequence"
     repeat_region   122..149
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

dbSNP
Short Genetic Variations



ar GaP PubMed Nucleotide Protein

ll variations in dbSNP or large structural variations in dbVar

for Go

Reference SNP (refSNP) Cluster Report: rs115644759

RefSNP	
Organism:	human (*Homo sapiens*)
Molecule Type:	Genomic
Created/Updated in build:	132/150
Map to Genome Build:	108/Weight 1
Validation Status:	

Allele	
Variation Class:	SNV: single nucleotide variation
RefSNP Alleles:	C/T (FWD)
Allele Origin:	
Ancestral Allele:	C
Variation Viewer:	VarView
Clinical Significance:	NA
MAF/MinorAlleleCount:	T=0.0112/56 (1000 Genomes) T=0.0156/455 (TOPMED)

HGVS Names
NC_000002.11:g.1493328C>T
NC_000002.12:g.1489556C>T
NG_011581.1:g.81094C>T
NM_000547.5:c.1768+1565C>T
NM_001206744.1:c.1768+1565C>T
NM_001206745.1:c.1598-4246C>T
NM_175719.3:c.1598-4246C>T
NM_175721.3:c.1768+1565C>T
NM_175722.3:c.1249+1565C>T
NT_187529.1:g.171927C>T
XM_005264698.1:c.1804+1565C>T

...more

NP Details are organized in the following sections:

GeneView Map Submission Fasta Resource Diversity Validation

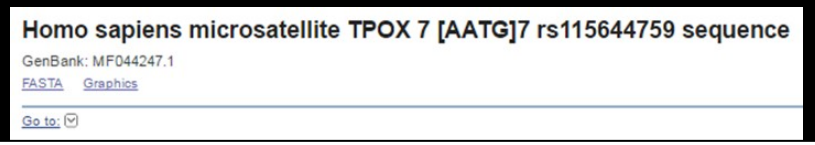
Integrated Maps (Hint: click on 'Chr Pos' to see variant in the new NCBI variation viewer)

Assembly	Annotation Release	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Neighbor SNP	Map Method
GRCh38.p7 (Alt loci)	108	2	NA	NT_187529.1	171927	NA	C	NA	view	mapup
GRCh38.p7	108	2	1489556	NT_005334.17	1479556	Fwd	C	Fwd	view	mapup

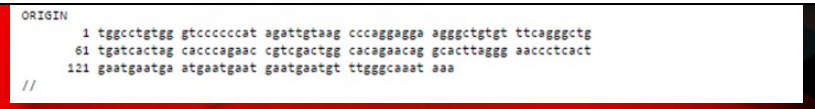
Population Diversity (Alleles in RefSNP orientation) . See additional population frequency from 1000Genome [here]

ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	HWP	Alleles C	Alleles T
ss1295621499	EAS		1008	AF		1.00000000	
	EUR		1006	AF		1.00000000	
	AFR		1322	AF		0.95759994	0.04240000
	AMR		694	AF		1.00000000	
	SAS		978	AF		1.00000000	
ss219021534	pilot_1_YRI_low_coverage_panel		118	AF		0.98305082	0.01694915

Summary	Average Het.+/- std err:	Individual Count	Founders Count	Individual Overlap	Genotype Conflict
	0.022+/-0.103	0	0	0	0




```
##HumanSTR-START##
STR locus name          :: TPOX
Length-based allele     :: 7
Bracketed repeat        :: [AATG]7
Sequencing technology   :: ForenSeq, MiSeq FGx; PowerSeq Auto, MiSeq
Coverage                :: >30X
Length-based tech.      :: PowerPlex Fusion, ABI3500xl
Assembly                :: GRCh38 (GCF_000001405)
Chromosome              :: 2
RefSeq Accession        :: NC_000002.12
Chrom. Location         :: 1489532..1489698
Repeat Location         :: 1489653..1489684
Cytogenetic Location    :: 2p25.3
##HumanSTR-END##
```



Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

GenBank: MF044247.1

FASTA Graphics

Go to:

COMMENT Annotation ('bracketing') of the repeat region is consistent with the guidance of the ISFG (International Society of Forensic Genetics), PMID: 26844919. Lower case letters in the 'Bracketed repeat' region below denote uncounted bases. The given length-based allele value was determined using the designated length-based technology. Variation in the length-based allele between individuals or assays can result from indels in flanking regions. The length of reported sequence is dependent on the assay (see 'Sequencing technology') and the quality of the flanking sequence. This information is provided as part of the STR Sequencing Project (STRseq), a collaborative effort of the international forensic DNA community. The purpose of this project is to facilitate the description of sequence-based STR alleles. Additional resources can be found at strseq.nist.gov. For questions or feedback, please contact strseq@nist.gov. Allele frequency data can be accessed in the strider.online database.

```
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

GenBank: MF044247.1

FASTA Graphics

Go to:

```
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Gettings,K.B., Borsuk,L.A. and Vallone,P.M.
  TITLE     The STR Sequencing Project [manuscript in preparation]
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163)
  AUTHORS   NIST,A.G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2017) Applied Genetics Group, National Institute
            of Standards and Technology, 100 Bureau Drive, MS-8314,
            Gaithersburg, MD 20899, USA
```

```
            Bracketed repeat      :: [AATG]7
            Sequencing technology :: ForenSeq, MiSeq FGx; PowerSeq Auto, MiSeq
            Coverage              :: >30X
            Length-based tech.    :: PowerPlex Fusion, ABI3500xl
            Assembly              :: GRCh38 (GCF_000001405)
            Chromosome            :: 2
            RefSeq Accession      :: NC_000002.12
            Chrom. Location       :: 1489532..1489698
            Repeat Location       :: 1489653..1489684
            Cytogenetic Location  :: 2p25.3
            ##HumanSTR-END##
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1..163
                     /note="Promega PowerSeq Sequence"
     variation       25
                     /note="C/T SNP"
                     /db_xref="dbSNP:rs115644759"
     misc_feature    120..154
                     /note="Illumina ForenSeq Sequence"
     repeat_region   122..149
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

GenBank: MF044247.1

FASTA Graphics

Go to:

```
LOCUS       MF044247                 163 bp    DNA     linear   PRI 30-MAY-2017
DEFINITION  Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence.
ACCESSION   MF044247
VERSION     MF044247.1
DBLINK      BioProject: PRJNA380554
KEYWORDS    STRSeq, STR, TPOX.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
```

```
            Assembly            :: GRCh38 (GCF_000001405)
            Chromosome          :: 2
            RefSeq Accession    :: NC_000002.12
            Chrom. Location     :: 1489532..1489698
            Repeat Location     :: 1489653..1489684
            Cytogenetic Location :: 2p25.3
            ##HumanSTR-END##
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1..163
                     /note="Promega PowerSeq Sequence"
     variation       25
                     /note="C/T SNP"
                     /db_xref="dbSNP:rs115644759"
     misc_feature    120..154
                     /note="Illumina ForenSeq Sequence"
     repeat_region   122..149
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

GenBank: MF044247.1

FASTA Graphics

```
REFERENCE   2  (bases 1 to 163)
  AUTHORS   NIST,A.G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2017) Applied Genetics Group, National Institute
            of Standards and Technology, 100 Bureau Drive, MS-8314,
            Gaithersburg, MD 20899, USA
COMMENT     Annotation ('bracketing') of the repeat region is consistent with
            the guidance of the ISFG (International Society of Forensic
            Genetics), PMID: 26844919.  Lower case letters in the 'Bracketed
            repeat' region below denote uncounted bases.   The given
            length-based allele value was determined using the designated
            length-based technology.  Variation in the length-based allele
            between individuals or assays can result from indels in flanking
            regions.  The length of reported sequence is dependent on the assay
            (see 'Sequencing technology') and the quality of the flanking
            sequence. This information is provided as part of the STR
            Sequencing Project (STRseq), a collaborative effort of the
            international forensic DNA community.  The purpose of this project
            is to facilitate the description of sequence-based STR alleles.
            Additional resources can be found at strseq.nist.gov.  For
            questions or feedback, please contact strseq@nist.gov.  Allele
            frequency data can be accessed in the strider.online database.

            ##HumanSTR-START##
            STR locus name        :: TPOX
            Length-based allele   :: 7
            Bracketed repeat      :: [AATG]7
            Sequencing technology :: ForenSeq, MiSeq FGx; PowerSeq Auto, MiSeq
            Coverage              :: >30X
            Length-based tech.    :: PowerPlex Fusion, ABI3500xl
            Assembly              :: GRCh38 (GCF_000001405)
            Chromosome            :: 2
            RefSeq Accession      :: NC_000002.12
            Chrom. Location       :: 1489532..1489698
            Repeat Location       :: 1489653..1489684
            Cytogenetic Location  :: 2p25.3
            ##HumanSTR-END##
FEATURES             Location/Qualifiers
     source          1..163
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1..163
                     /note="Promega PowerSeq Sequence"
     variation       25
                     /note="C/T SNP"
                     /db_xref="dbSNP:rs115644759"
     misc_feature    120..154
                     /note="Illumina ForenSeq Sequence"
     repeat_region   122..149
                     /rpt_type=tandem
                     /satellite="microsatellite:TPOX"
ORIGIN
        1 tggcctgtgg gtccccccat agattgtaag cccaggagga agggctgtgt ttcagggctg
       61 tgatcactag cacccagaac cgtcgactgg cacagaacag gcacttaggg aaccctcact
      121 gaatgaatga atgaatgaat gaatgaatgt ttgggcaaat aaa
//
```

Homo sapiens microsatellite TPOX 7 [AATG]7 rs115644759 sequence

GenBank: MF044247.1

GenBank FASTA

STRSeq in Bioinformatics

Standalone and API BLAST

Download BLAST
Get BLAST databases and executables

Use BLAST API
Call BLAST from your application

Use BLAST in the cloud
Start an instance at a cloud provider

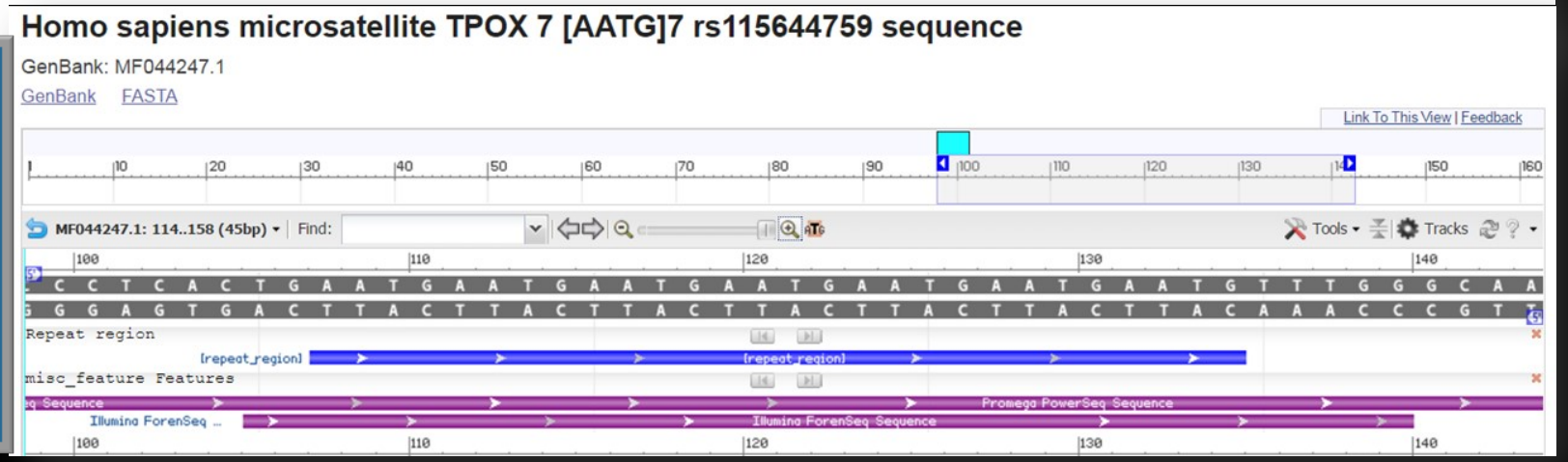
Embedding the NCBI Sequence View in Web Content

Introduction

The NCBI Graphical Sequence Viewer (SV) is a general purpose tool for viewing biological sequence data. The Sequence Viewer has a very rich set of options and can display virtually any sequence. It can be embedded in a wide variety of web pages serving many different needs. This page has examples showing best practice for embedding Sequence Viewer with several different sets of options.

The API Reference should be consulted for the full suite of options and parameters. Additional documentation can be found here:
https://www.ncbi.nlm.nih.gov/tools/sviewer/

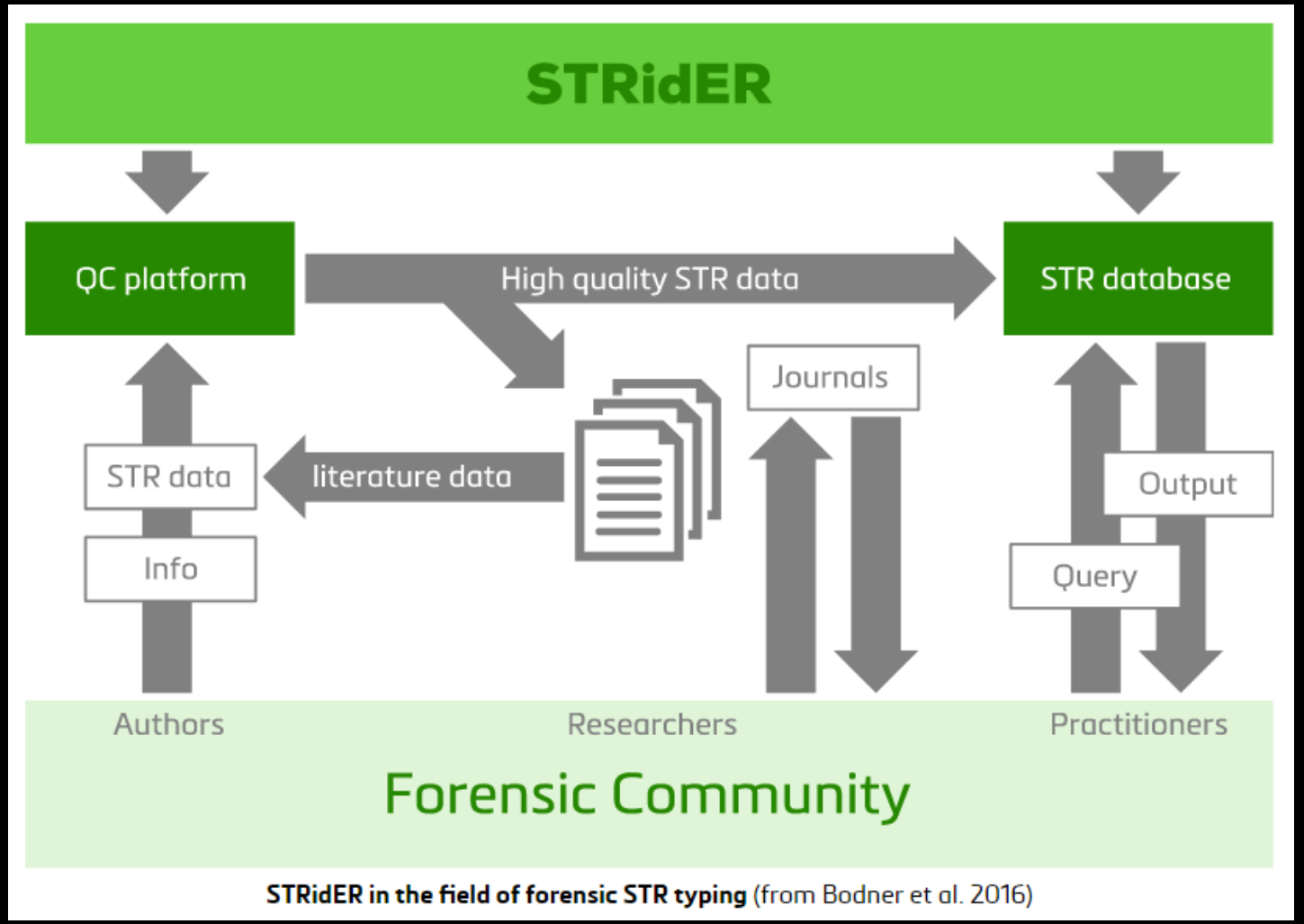
The catalog of FASTA files and GenBank flatfiles is downloadable



STRSeq in Population Data

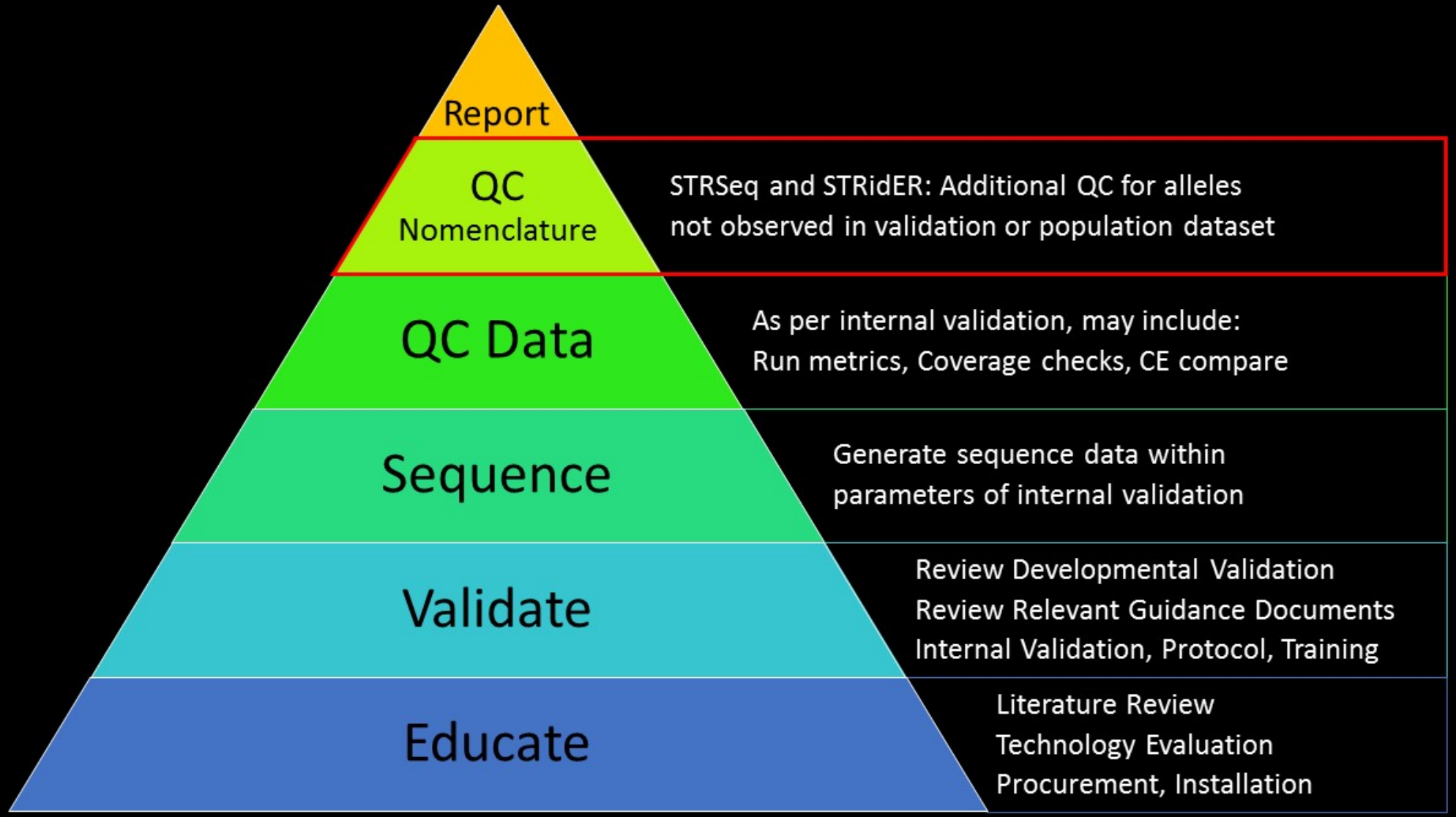
STRSeq

Collaboration in QC and exchange of data



STRidER in the field of forensic STR typing (from Bodner et al. 2016)

STRSeq in Casework



Somewhere something incredible is waiting to be known.
Carl Sagan

Acknowledgements

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA — Dr. Chris Phillips

UNT HEALTH SCIENCE CENTER — Jonathan King, Dr. Bruce Budowle

NIST — Lisa Borsuk, Dr. Peter Vallone

UMIT TIROL — Drs. Martin Bodner, Walther Parson

NCBI — Drs. Lori Black, Melissa Landrum, Ilene Mizrachi, Kim Pruitt, George Riley, Steve Sherry

KING'S College LONDON — Dr. David Ballard, Laurence Devesse

Thanks to the Labs who provided feedback!