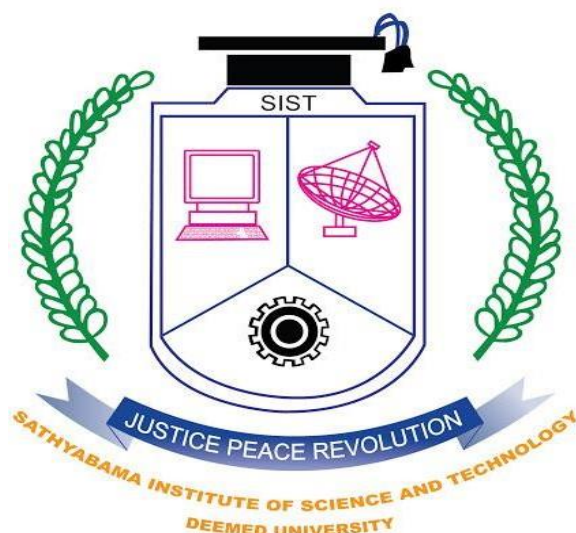


STUDIES ON THE ROLE OF SHORT TANDEM REPEATS ON Y CHROMOSOME IN FORENSIC SCIENCE

Submitted in partial fulfillment of the requirements
for the award of Master of Science Degree in
BIOINFORMATICS AND DATA SCIENCE

By

RAJKUMAR R (Register No.40739004)



DEPARTMENT OF BIOINFORMATICS

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI
- 600 119**

APRIL 2022



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119
www.sathyabama.ac.in



DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **RAJKUMAR R** (Reg. No. 40739004) who carried out the project entitled "**STUDIES ON THE ROLE OF SHORT TANDEM REPEATS ON Y CHROMOSOME IN FORENSIC SCIENCE**" under our supervision from December 2021 to April 2022.

Dr. DANIEL ALEX ANAND
Internal Guide

Dr. JEMMY CHRISTY
Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

We **RAJKUMAR R** (Reg. No. 40739004) hereby declare that the Project Report entitled “**STUDIES ON THE ROLE OF SHORT TANDEM REPEATS IN FORENSIC SCIENCE**” done by us under the guidance of **Dr. DANIEL ALEX ANAND** is submitted in partial fulfilment of the requirements for the award of Master of Science degree in Bioinformatics.

1.

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to Board of Management of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr. JEMMY CHRISTY “ Head of the Department”** and **Dr. SWETHA SUNKAR, Dept. of Bioinformatics** for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **Dr. DANIEL ALEX ANAND, Dept of Bioinformatics** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of our project work.

We wish to express our thanks to all Teaching and Non-teaching staff members of the **Department of Bioinformatics** who were helpful in many ways for the completion of the project.

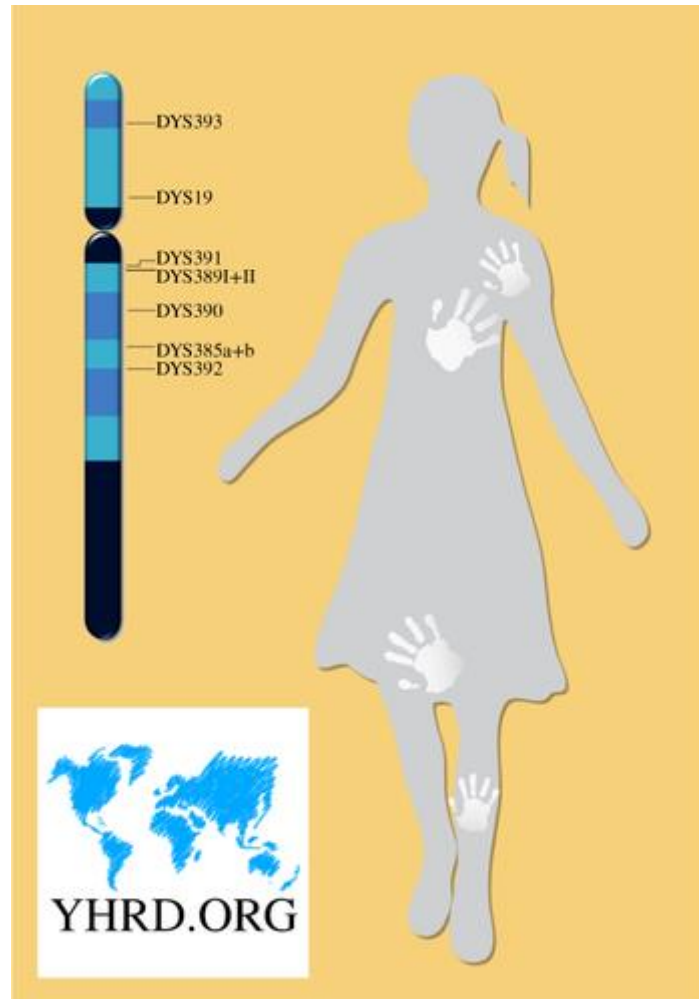
ABSTRACT

STR markers located on the Y chromosome and the Y Chromosome Haplotype Reference Database (YHRD) are used to analyze and interpret DNA evidence in sexual assault investigations.

The analysis of short tandem repeat (STR) DNA sequences is of fundamental importance to forensic science because they have become the recognized standard in constructing national public databases. Consequently, considerable effort has been expended in developing multiplexed (one tube) reactions that analyze several loci in combination. The implementation of STRs in casework cannot take place without a full understanding of the systems used. The purpose of validation is to characterize multiplexes when one is challenged with forensic samples. For example, mixtures are often encountered that may be particularly difficult to interpret against a background of allelic artifacts. By increasing the number of PCR amplification cycles, it is possible to dramatically boost the sensitivity of the system so that just a handful of cells may be successfully analyzed. However, interpretation is much more complex because the origin of DNA profiles may be less certain and complicated by issues such as contamination, the potential for innocent transfer, and a predominance of mixtures. This review provides a brief historical background of the development of STRs in forensic casework that culminated in the creation of national DNA databases. The development of guidelines to interpret complex DNA profiles, such as mixtures, is outlined. Finally, the recent innovation of low copy number DNA profiling is explained along with the special considerations needed to report in court.

Short tandem repeats (STRs) are short tandemly repeated DNA sequences that involve a repetitive unit of 1–6 bp. Because of their polymorphisms and high mutation rates, STRs are widely used in biological research. Strand-slippage replication is the predominant mutation mechanism of STRs, and the stepwise mutation model is regarded as the main mutation model. STR mutation rates can be influenced by many factors. Moreover, some trinucleotide repeats are associated with human neurodegenerative diseases. In order to deepen our knowledge of these diseases and broaden STR application, it is essential to understand the STR mutation

process in detail. In this review, we focus on the current known information about STR.



Python in Forensic science:

Python has built-in capabilities to support digital investigation and protect the integrity of evidence during an investigation. In this tutorial, we will explain the fundamental concepts of applying Python in computational (digital) forensics that includes extracting evidence, collecting basic data, and encryption of passwords as required.

Keywords: Y Chromosome Haplotype Reference Database (YHRD) , Short tandem repeat (STR)

INDEX

S.NO	TITLE	PAGE NO
1	CHAPTER 1- INTRODUCTION 1.1.	1-4
2	CHAPTER 2- LITERATURE SURVEY	5-1
3	CHAPTER 3- AIM AND OBJECTIVES	
4	CHAPTER 4- METHODOLOGY	
5	CHAPTER 5-RESULT AND DISCUSSION	
6	CHAPTER 6- SUMMARY AND CONCLUSION	
6	REFERENCES	54-55

LIST OF FIGURES

FIGURE No.	FIGURE NAME	PAGE
------------	-------------	------

2.1	Allele	2
2.2	Computational forensics	8

ABBREVIATION

CHAPTER-1

INTRODUCTION

1.1. OUTLINE OF THE PROJECT

1.1.1. SHORT TANDEM REPEATS (STR):

Short tandem repeats (STRs) are short repeated sequences of DNA (2–6 bp) that account for approximately 3% of the human genome ([Lander et al., 2001](#)). The number of repeat units is highly variable among individuals, which offers a high power of discrimination when analyzed for identification purposes. It is a widely accepted notion that STRs are non-coding in nature and are therefore not implicated in gene expression ([Tautz and Schlotterer, 1994](#); [Ramel, 1997](#); [Butler, 2006](#); [Biscotti et al., 2015](#)). There is increasing evidence, however, that non-coding DNA sequences such as STRs may be involved in gene regulation via various mechanisms, hence being associated with phenotype ([Sawaya et al., 2013](#); [Chen et al., 2016](#)).

1.1.2. STR MARKERS:

A short tandem repeat (STR or microsatellite) is a pattern of two or more nucleotides that are repeated directly adjacent to each other.

The repeats can range in length from 2 to 6 base pairs/repeat. A short tandem repeat polymorphism occurs when homologous STR loci differ in the number of repeats between individuals. By identifying repeats of a specific sequence at specific locations in the genome, it is possible to create a genetic profile of an individual.

When running a QF-PCR assay, STRs are amplified by PCR using the unique sequences of flanking regions as primers. By using fluorescently labelled primers amplified STRs can be identified and quantified on a capillary electrophoresis instrument.

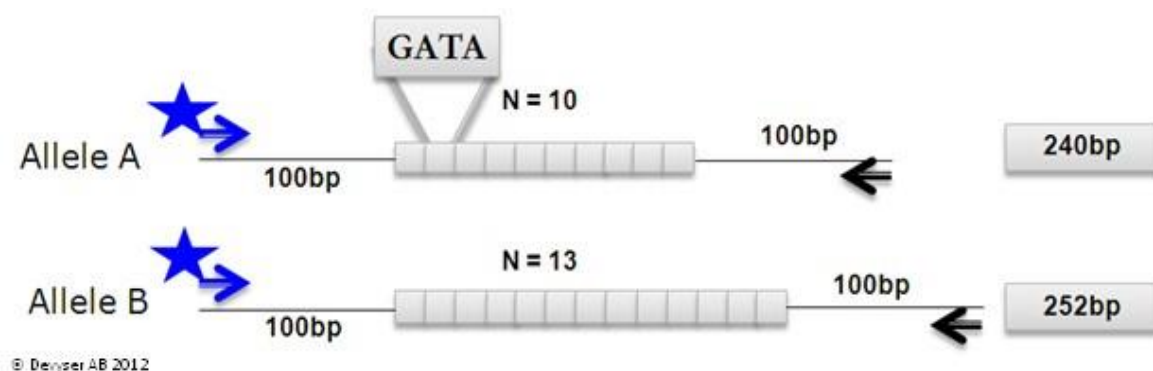


Figure – 2.1

1.1.2.1. QF – PCR ASSAY:

Quantitative fluorescent PCR (QF-PCR) has been used by many laboratories for prenatal diagnosis of the most common aneuploidies. QF - PCR is rapid, cost-effective, and suitable for automation and can detect most abnormalities diagnosed by conventional karyotyping. Whether QF-PCR should be used alone in most of the samples and in which karyotyping should also be offered is currently a topic of debate. We evaluated and compared the results obtained from 7679 prenatal samples in which conventional karyotype and QF-PCR had been performed, including 1243 chorionic villi and 6436 amniotic fluid samples. Concordant QF-PCR and karyotype results were obtained in 98.75% of the samples. An abnormal karyotype associated with adverse clinical outcome undetected by QF-PCR was found in 0.05% of samples. Therefore, QF-PCR can be used alone in a large number of samples studied in a prenatal laboratory, thereby reducing both the workload in cytogenetic laboratories and parental anxiety when awaiting results.

1.1.2.2. PCR:

Polymerase chain reaction (PCR) is a laboratory technique used to amplify DNA sequences. The method involves using short DNA sequences called primers to

select the portion of the genome to be amplified. The temperature of the sample is repeatedly raised and lowered to help a DNA replication enzyme copy the target DNA sequence. The technique can produce a billion copies of the target sequence in just a few hours.

1.1.3. Y CHROMOSOME:

The Y chromosome is one of the two sex chromosomes in people. The Y chromosome traverses in excess of 59 million structure squares of DNA and speaks to just about 2 percent of the complete DNA in cells. Every individual ordinarily has one set of sex chromosomes in every cell. The Y chromosome is available in Men, who have one X and one Y chromosome, while females have two X chromosomes. Distinguishing qualities on every chromosome is a functioning territory of hereditary exploration. Since analysts utilize various ways to deal with foresee the quantity of qualities on every chromosome, the assessed number of qualities changes. The Y chromosome probably contains 50 to 60 qualities that give directions to making proteins. Sex is dictated by the SRY quality, which is liable for the improvement of an embryo into a male.. Numerous qualities are remarkable to the Y chromosome, yet qualities in territories known as pseudoautosomal areas are available on both sex chromosomes. Accordingly, people each have two useful duplicates of these qualities. Numerous qualities in the pseudoautosomal districts are basic for typical turn of events.

1.1.4 FORENSIC SCIENCE:

Forensic science is a critical element of the criminal justice system. Forensic scientists examine and analyze evidence from crime scenes and elsewhere to develop objective findings that can assist in the investigation and prosecution of perpetrators of crime or absolve an innocent person from suspicion.

Common forensic science laboratory disciplines include forensic molecular biology (DNA), forensic chemistry, trace evidence examination (hairs and fibers, paints and polymers, glass, soil, etc.), latent fingerprint examination, firearms and toolmarks examination, handwriting analysis, fire and explosives examinations, forensic toxicology, and digital evidence. Some forensic disciplines practiced outside forensic laboratories include forensic pathology, forensic nursing, forensic psychiatry, forensic entomology, and forensic engineering. Practitioners of these disciplines are most often found in medical examiner or coroner offices, in universities, or in private practices.

The Department of Justice maintains forensic laboratories at the Bureau of Alcohol, Tobacco, Firearms, and Explosives, the Drug Enforcement Administration, and the Federal Bureau of Investigation. The Department, through the National Institute of Justice, is a sponsor of cutting-edge research. Its labs serve as a model for government forensic agencies at the federal, state and local

levels. The Department strives to set the global standard for excellence in forensic science and to advance the practice and use of forensic science by the broader community.

This website contains information of value to the forensic science community, as well as stakeholders engaged in the criminal justice system with interests in forensic science.

Y CHROMOSOME IN FORENSIC SCIENCE:

Measurable examination of Y-chromosome DNA is particularly valuable when a modest quantity of male DNA is blended in with a lot of female DNA, for example, happens in rape cases. Clarifying this proof in court, notwithstanding, is troublesome in light of the fact that the Y chromosome goes down generally unaltered from fathers to children, so a solitary Y-chromosome profile can be shared by many men in a populace. Rather than a match likelihood or information base tally it's usually recommended that courts be told about the conceivable number of coordinating guys in the populace, and the potential outcomes of their relatedness, which is regularly more far off than uncle or cousin however a lot nearer than for an arbitrary man. They additionally show how the appropriation of coordinating Men can be influenced by information base data, and propose approaches to introduce this data in court to clarify that Y-chromosome proof can't absolutely distinguish the guilty party, however can drastically decrease the quantity of potential wellsprings of the DNA.. After the presentation of DNA profiling utilizing non-sex chromosomes, the system had issues that, when tended to, made profiling an incredible asset that has altered scientific science. **Presently, Y chromosome profiling must go through similar cycle to evaluate the outcomes in a manner that is substantial and straightforwardly interpretable to courts.** The new programming introduced in this examination could be utilized to improve the precision of Y chromosome proof and to build its comprehension by judges and members of the jury.

1.2. OUTSOURCE:

1.2.1. TOOLS AND SOFTWARE USED TO PREDICT:

The tool here used for "STUDIES ON THE ROLE OF SHORT TANDEM REPEATS ON Y CHROMOSOME USING FORENSIC SCIENCE" are
UNIPRO UGENE AND PYTHON .

1.2.2.UGENE:

UGENE is computer [software](#) for [bioinformatics](#). It works on [personal computer](#) operating systems such as [Windows](#), [macOS](#), or [Linux](#). It is released as [free and open-source software](#), under a [GNU General Public License](#) (GPL) version 2.

UGENE helps biologists to analyze various [biological genetics](#) data, such as [sequences](#), annotations, [multiple alignments](#), [phylogenetic trees](#), [NGS assemblies](#), and others. The data can be stored both locally (on a personal computer) and on a shared storage (e.g., a lab database).

UGENE integrates dozens of well-known biological tools, algorithms, and original tools in the context of [genomics](#), [evolutionary biology](#), [virology](#), and other branches of life science. UGENE provides a [graphical user interface](#) (GUI) for the pre-built tools so biologists with no [computer programming](#) skills can access those tools more easily.

Using UGENE Workflow Designer, it is possible to streamline a multi-step analysis. The workflow consists of blocks such as data readers, blocks executing embedded tools and algorithms, and data writers. Blocks can be created with command line tools or a script. A set of sample workflows is available in the Workflow Designer, to annotate sequences, convert data formats, analyze NGS data, etc.

Beside the graphical interface, UGENE also has a [command-line interface](#). Workflows may also be executed thereby.

To improve performance, UGENE uses [multi-core processors](#) (CPUs) and [graphics processing units](#) (GPUs) to optimize a few algorithms

1.2.3.PYTHON:

Python is a [high-level](#), [general-purpose programming language](#). Its design philosophy emphasizes [code readability](#) with the use of [significant indentation](#). Its [language constructs](#) and [object-oriented](#) approach aim to help [programmers](#) write clear, logical code for small- and large-scale projects.

Python is [dynamically-typed](#) and [garbage-collected](#). It supports multiple [programming paradigms](#), including [structured](#) (particularly [procedural](#)), object-oriented and [functional programming](#). It is often described as a "batteries included" language due to its comprehensive [standard library](#).

[Guido van Rossum](#) began working on Python in the late 1980s as a successor to the [ABC programming language](#) and first released it in 1991 as Python 0.9.0.^[33] Python 2.0 was released in 2000 and introduced new features such as [list comprehensions](#), [cycle-detecting](#) garbage collection, [reference counting](#), and [Unicode](#) support. Python 3.0, released in 2008, was a major revision that is not completely [backward-compatible](#) with earlier versions. Python 2 was discontinued with version 2.7.18 in 2020.

Python consistently ranks as one of the most popular programming languages.

CHAPTER-2

LITERATURE SURVEY

2.1. REVIEW-LITERATURE WORK:

2.1.1. *What are STR:*

Short tandem repeat (STR) typing methods are widely used today for human identity testing applications including forensic DNA analysis. Short Tandem Repeats composing strategies are broadly utilized today for human character testing applications including legal DNA examination. Following multiplex PCR enhancement, DNA tests containing the length-variation STR alleles are commonly isolated by slim electrophoresis and genotyped by correlation with an allelic stepping stool provided with a business unit.

Short Tandem Repeats, which are here and there alluded to as microsatellites or straight forward succession rehashes, are accordion-like stretches of DNA containing centre recurrent units of somewhere in the range of two and seven nucleotides long that are tandem rehashed from roughly about six to a few dozen times. In spite of the fact that the human genome contains heaps of STR markers, just a little centre arrangement of loci has been chosen for use in legal DNA and human personality testing. Like utilizing solitary, normal cash from a monetary perspective, centre loci grant proportional hereditary data to be shared and thought about. Business packs are presently accessible to create DNA profiles containing these centre STR loci. A large number of STR profiles are created worldwide every year by government, college, and private research facilities performing different types of human character testing, including DNA data basing, measurable casework, missing people/mass fiasco casualty distinguishing proof, or parentage testing.

2.1.2. Why str:

Short tandem repeats, also known as microsatellites or simple sequence repeats, are short tandem repeated DNA sequences that involve a repetitive unit of 1-6 BP, forming series with lengths of up to 100 nucleotides. STRs are widely found in prokaryotes and eukaryotes, including humans. They appear scattered more or less evenly throughout the human genome, accounting for about 3% of the entire genome. However, their distribution within chromosomes is not quite uniform—they appear less frequently in subtelomeric regions. Most STRs are found in the non coding regions, while only about 8% locate in the coding regions. On the basis of different repeat units, STRs can be classified into different types. On the one hand, according to the length of the major repeat unit, STRs are classified into mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats. The total number of each type decreases as the size of the repeat unit increases. The most common STRs in the human genome are dinucleotide repeats. **On the other hand, according to the repeat structure, STRs are classified into perfect repeats, containing only one repetitive unit, and imperfect repeats, consisting of different composition repeats.**

These proteins are not necessary for the viability of the bacteria, but they can help bacteria to evade the human immune system. Some STRs may take part in regulating the transcription. From yeasts to humans, many proteins involved in transcriptional regulation contain glutamine-rich domains and trinucleotide repeats encoding series of polyglutamine. Moreover, researches have shown that some STRs can regulate the transcription of the epidermal growth factor gene, the tyrosine hydroxylase gene, and the PIG3 gene.

2.1.3. Study on python forensics:

Python is a general-purpose programming language with easy, readable code that can be easily understood by both professional developers as well as novice programmers. Python comprises of many useful libraries that can be used with any stack framework. Many laboratories rely on Python to build basic models for predictions and to run experiments. It also helps to control critical operational systems.

Python has built-in capabilities to support digital investigation and protect the integrity of evidence during an investigation. In this tutorial, we will explain the fundamental concepts of applying Python in digital or computation forensics.

2.1.4. What is computational forensics:

Computational Forensics is an emerging research domain. It deals with solving forensic problems using digital methods. It uses computational science to study digital evidence.

Computation Forensics includes a broad range of subjects which has objects, substances, and processes investigated, mainly based on pattern evidence, such as toolmarks, fingerprints, shoeprints, documents etc., and also includes physiological and behavioral patterns, DNA, and digital evidence at crime scenes.

The following diagram shows the broad range of subjects covered under Computational Forensics.

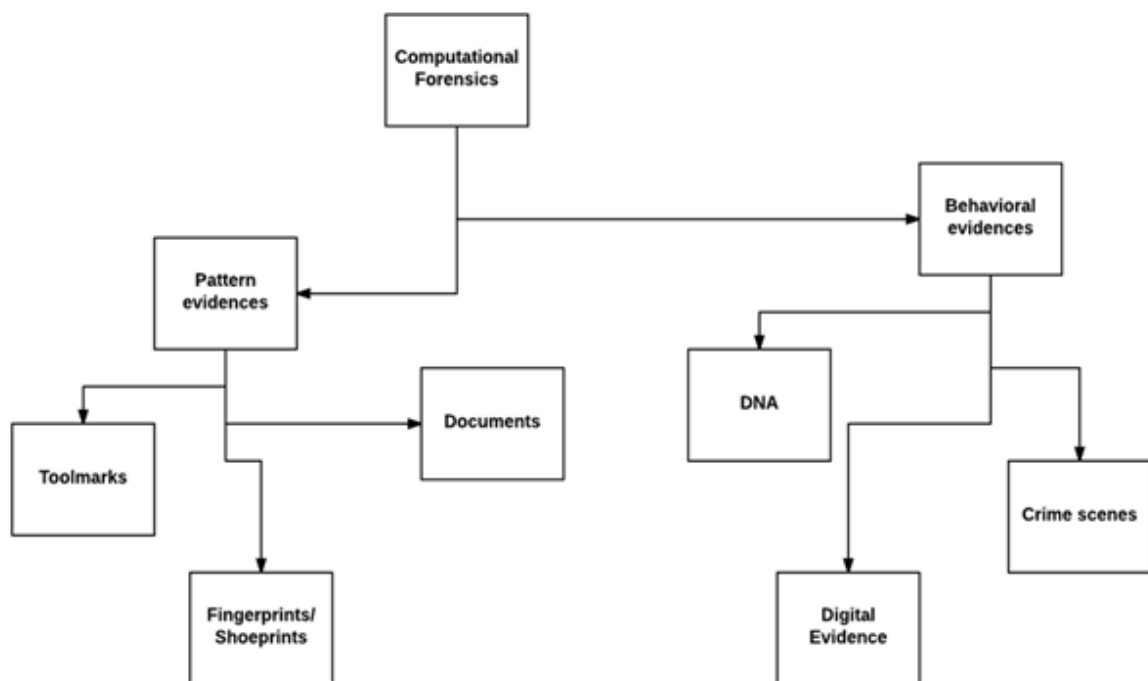


Figure - 2.2

Computational forensics is implemented with the help of some algorithms. These algorithms are used for signal and image processing, computer vision and graphics. It also includes data mining, machine learning, and robotics.

Computational forensics involves diverse digital methods. The best solution to ease all digital methods in forensics is to use a general-purpose programming language like Python.

CHAPTER-3

AIM AND OBJECTIVES

OBJECTIVE:

Understanding and analysing STRS on the Y chromosome in forensic crime solving using UNIPRO UGENE and also with the help of BIOPYTHON AND C PROGRAMS.

Listed are the prominent objectives of the study:

- To study all the short tandem repeats in the y chromosome.
- To study STR dataset – international as well as national.
- To create a worksheet based on the STR database.
- To work-on finding the repeats of STR using bio python and c program.
- To work-on UNIPRO UGENE DATABASE the given dataset.
- To analyze the MARKERS IN Y CHROMOSOME.
- To execute and find the STR using PYTHON code.
- To run the c program in c compiler using the Program.

CHAPTER-4

METHODOLOGY

4.1. SHORT TANDEM REPEATS IN FORENSICS DATABASES USED

4.1.1.UNIPRO UGENE:

Unipro UGENE is a multiplatform open-source software with the main goal of assisting molecular biologists without much expertise in bioinformatics to manage, analyze and visualize their data. UGENE integrates widely used bioinformatics tools within a common user interface. The toolkit supports multiple biological data formats and allows the retrieval of data from remote data sources. It provides visualization modules for biological objects such as annotated genome sequences, Next Generation Sequencing (NGS) assembly data, multiple sequence alignments, phylogenetic trees and 3D structures. Most of the integrated algorithms are tuned for maximum performance by the usage of multithreading and special processor instructions. UGENE includes a visual environment for creating reusable workflows that can be launched on local resources or in a High Performance Computing (HPC) environment. UGENE is written in C++ using the Qt framework. The built-in plugin system and structured UGENE API make it possible to extend the toolkit with new functionality. Figure – 4.1 shows the source of the dataset.

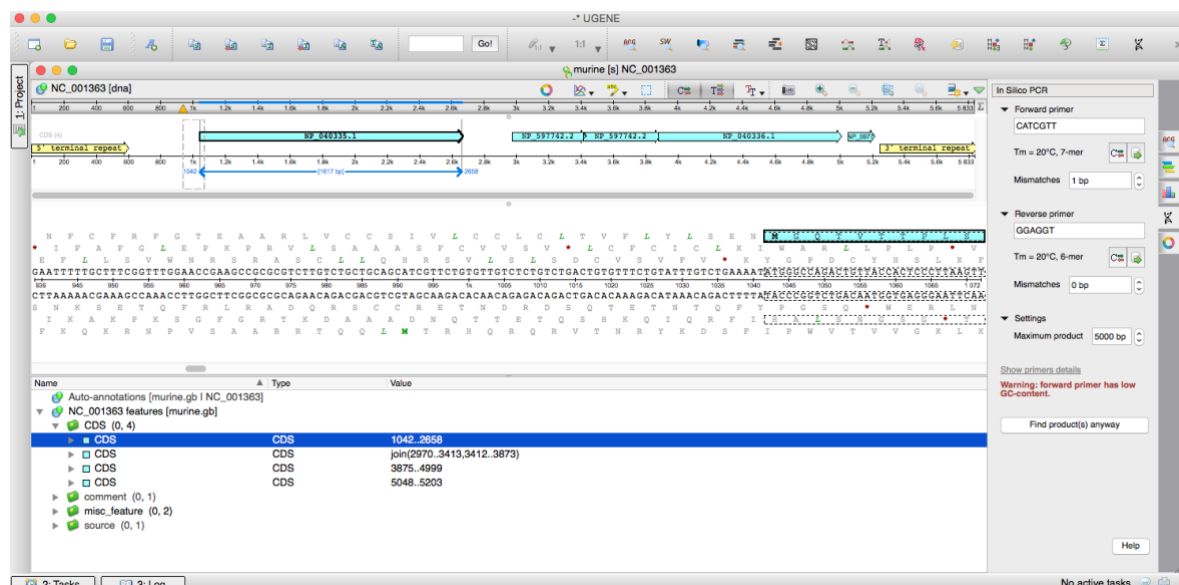
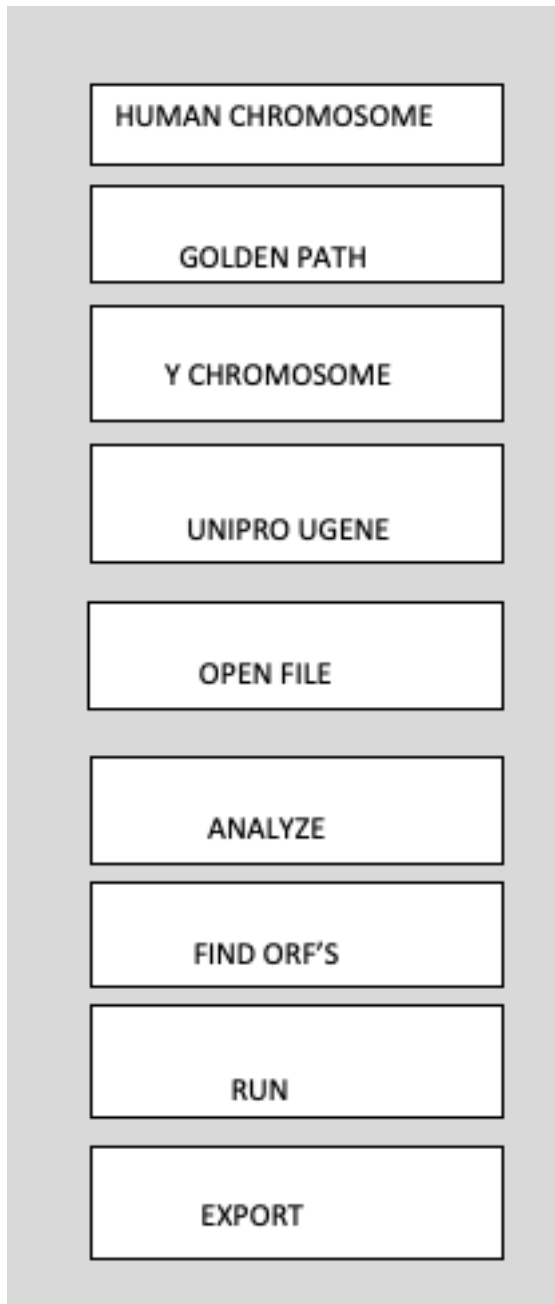


Figure – 4.1: Source of the dataset

work flow



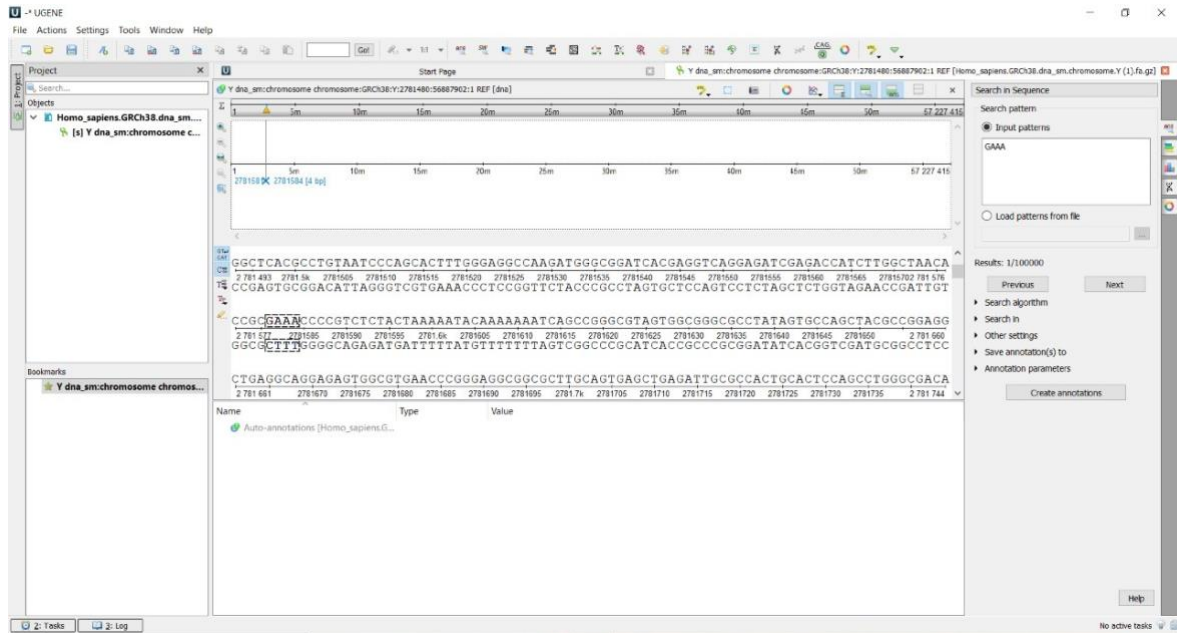
DATABASES:

STR DATA BASE

- This database is an information resource for the forensic DNA typing community with details on commonly used short tandem repeat (STR) DNA

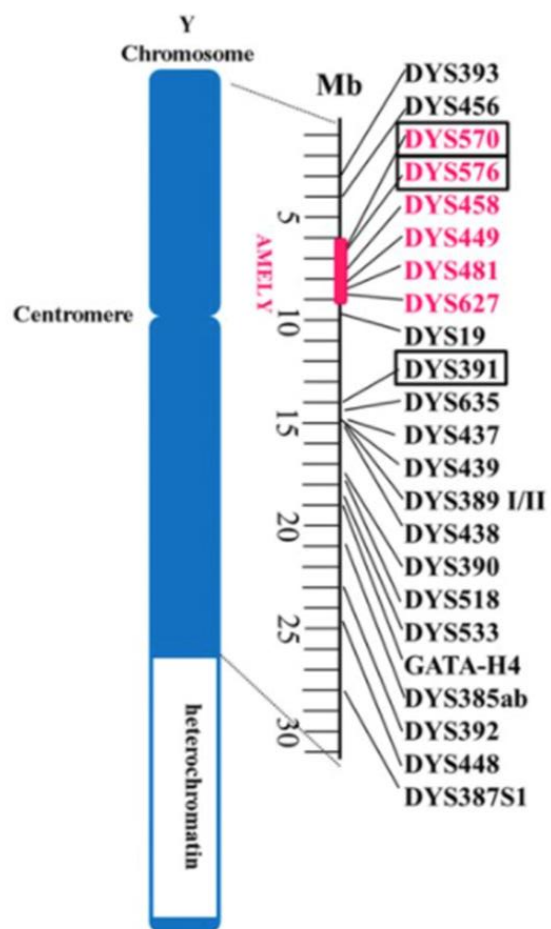
markers. STRBase consolidates and organizes the abundant literature on this subject to facilitate on-going efforts in DNA typing.

THE TOOL IMPORTED WITH THE FILE AND ANALYSED



The screenshot shows the UGENE software interface. The main window displays a DNA sequence alignment. The sequence is shown in a grid format with columns representing positions. A search for 'GAAA' is highlighted in the sequence. The left sidebar shows the project structure, including 'Homo_sapiens.GRCh38.dna_sm...' and 'Y dna_sm.chromosome c...'. The right sidebar shows search results and options, including 'Search in Sequence', 'Input patterns', and 'Load patterns from file'. The bottom status bar indicates 'No active tasks'.

Y chromosome



DATA COLLECTION

TABLE 1 A list of frequently used Y-STR panels in forensic practice

Marker	Regular copy number ^a	17 marker (CE) ^b	23 marker (CE)	27 marker (CE)	23 marker (Seq) ^c	26 marker (Seq)	RM panel ^d
DYS19	1	*	*	*	*	*	
DYS385	2	*	*	*	*	*	
DYS389I	1	*	*	*	*	*	
DYS389II	1	*	*	*	*	*	
DYS390	1	*	*	*	*	*	
DYS391	1	*	*	*	*	*	
DYS392	1	*	*	*	*	*	
DYS393	1	*	*	*	*	*	
DYS437	1	*	*	*	*	*	
DYS438	1	*	*	*	*	*	
DYS439	1	*	*	*	*	*	
DYS448	1	*	*	*	*	*	
DYS449	1			*			*
DYS456	1	*	*	*	*		
DYS458	1	*	*	*	*		
DYS460	1			*		*	
DYS481	1		*	*	*	*	
DYS505	1					*	
DYS518	1			*			*
DYS522	1					*	
DYS526	1						*
DYS533	1		*	*	*	*	
DYS547	1						*

DATA COLLECTION FROM STR BASE

S. No	Database	Marker	Pattern	Pattern Elaborated	No of repeats in the ref genome
1	STR base	DYS19	GAAA	GAAA	1
2	STR base	DYS385	[TAGA] ₃ tagg[TAGA] _n	TAGA TAGA TAGATAGGTAGA	4
3	STR base	DYS389I	[TCTG] ₃ [TCTA] ₆	5'-CCAACTCTCATCTGTATTATCTAT-3' 5'-TCTTATCTCCACCCAGA-3'	2
4	STR base	DYS390	[TCTG] ₈ [TCTA] ₉ [TCTG] ₁ [TCTA] ₄	5'-TATATTTACACATTTTGGGCC-3' 5'-TGACAGTAAATGAACACATTGC-3'	1
5	STR base	DYS391	[TCTA] ₆	5'-CTATTCATTCATCATACA CCCA-3' 5'-GATTCITTTGTGGTGGGTC TG-3'	2

REFERENCE: https://strbase.nist.gov/str_y385.htm

- Python forensics
- Python is a general-purpose programming language with easy, readable code that can be easily understood by both professional developers as well as novice programmers. Python comprises of many useful libraries that can be used with any stack framework. Many laboratories rely on Python to build basic models for predictions and to run experiments. It also helps to control critical operational systems.
- Python has built-in capabilities to support digital investigation and protect the integrity of evidence during an investigation. In this tutorial, we will explain the fundamental concepts of applying Python in digital or computation forensics.

Python String find()

In this tutorial, we will learn about the Python String find() method with the help of examples.

The `find()` method returns the index of first occurrence of the substring (if found). If not found, it returns -1.

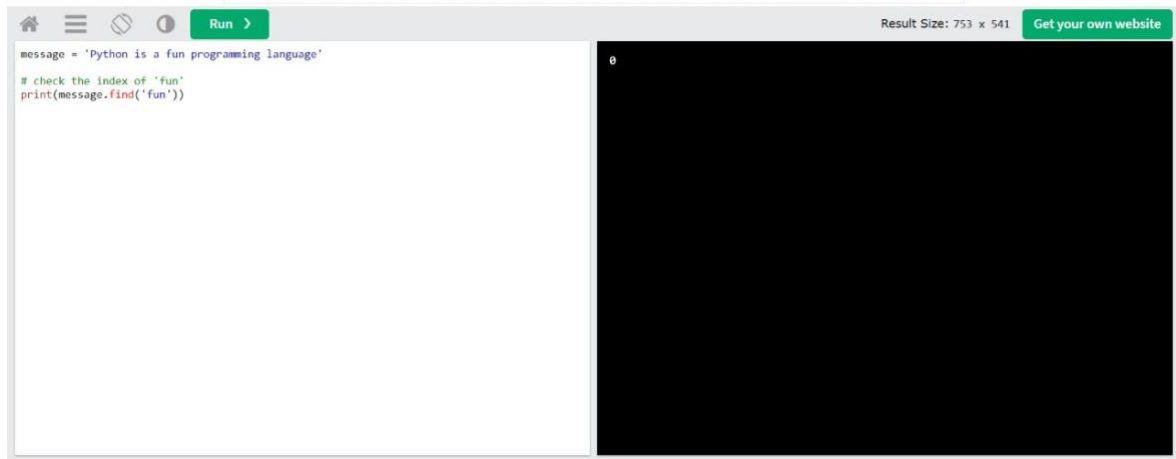
Example

```
message = 'Python is a fun programming language'

# check the index of 'fun'
print(message.find('fun'))

# Output: 12
```

- What is Computational Forensics?
- Computational Forensics is an emerging research domain. It deals with solving forensic problems using digital methods. It uses computational science to study digital evidence.
- Computation Forensics includes a broad range of subjects which has objects, substances, and processes investigated, mainly based on pattern evidence, such as toolmarks, fingerprints, shoeprints, documents etc., and also includes physiological and behavioral patterns, DNA, and digital evidence at crime scenes.



SUMMARY AND CONCLUSION

- Y chromosome markers are used to assist forensic investigations in a broad range of casework, either for rapid sexing of traces, for the identification and profiling of masked male DNA component(s) in mixtures with an excess of female DNA, for tracing male lineages in relationship testing or for analyzing the deep-rooting ancestry of unidentified corpses or unknown trace donors. Like any forensic DNA analysis, judging the significance of a match depends upon the population frequency of the profile. As has been discussed in this review, Y chromosomes are highly geographically structured and there is the further complication that all members of a patrilineage are expected to share Y haplotypes. The forensic community has been proactive in establishing large quality-assured and publicly accessible databases of Y-haplotypes to address these issues. The international cooperation in the YHRD project is a blueprint for collaboration and community building in the field of forensic genetics.

REFERENCES

- Alghafri, R., Goodwin, W., Ralf, A., Kayser, M., & Hadi, S. (2015). A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. *Forensic Science International Genetics*, **17**, 91– 98.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Ambers, A., Votrubova, J., Vanek, D., Sajantila, A., & Budowle, B. (2018). Improved Y-STR typing for disaster victim identification, missing persons investigations, and historical human skeletal remains. *International Journal of Legal Medicine*, **132**(6), 1545– 1553.

[CrossrefPubMedWeb of Science®Google Scholar](#)

- Andersen, M. M., & Balding, D. J. (2017). How convincing is a matching Y-chromosome profile? *PLoS Genetics*, **13**(11), e1007028.

[CrossrefPubMedWeb of Science®Google Scholar](#)

- Andersen, M. M., Curran, J., de Zoete, J., Taylor, D., & Buckleton, J. (2018). Modelling the dependence structure of Y-STR haplotypes using graphical models. *Forensic Science International Genetics*, **37**, 29– 36.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Andersen, M. M., Eriksen, P. S., & Morling, N. (2013). The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. *Journal of Theoretical Biology*, **329**, 39– 51.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Ayub, Q., Mohyuddin, A., Qamar, R., Mazhar, K., Zerjal, T., Mehdi, S. Q., & Tyler-Smith, C. (2000). Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Research*, **28**(2), e8– e88.

[CrossrefCASPubMedGoogle Scholar](#)

- Baker, J. L., Rotimi, C. N., & Shriner, D. (2017). Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Scientific Reports*, **7**(1), 1572.
-

[CrossrefPubMedWeb of Science®Google Scholar](#)

- Balaesque, P., Poulet, N., Cussat-Blanc, S., Gerard, P., Quintana-Murci, L., Heyer, E., & Jobling, M. A. (2015). Y-chromosome descent clusters and male differential reproductive success: Young lineage expansions dominate Asian pastoral nomadic populations. *European Journal of Human Genetics*, **23**(10), 1413– 1422.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., ... Kayser, M. (2010). Mutability of Y-chromosomal microsatellites: Rates, characteristics, molecular bases, and forensic implications. *American Journal of Human Genetics*, **87**(3), 341– 353.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Ballantyne, K. N., Keerl, V., Wollstein, A., Choi, Y., Zuniga, S. B., Ralf, A., ... Kayser, M. (2012). A new future of forensic Y-chromosome analysis: Rapidly mutating-STRs for differentiating male relatives and paternal lineages. *Forensic Science International Genetics*, **6**(2), 208– 218.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Ballantyne, K. N., Ralf, A., Aboukhalid, R., Achakzai, N. M., Anjos, M. J., Ayub, Q., ... Kayser, M. (2014). Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. *Human Mutation*, **35**(8), 1021– 1032.
-

[Wiley Online LibraryCASPubMedWeb of Science®Google Scholar](#)

- Batini, C., & Jobling, M. A. (2017). Detecting past male-mediated expansions using the Y chromosome. *Human Genetics*, **136**(5), 547– 557.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- de Knijff, P. (2000). Y-chromosomes shared by descent or by state. In C. Renfrew & K. Boyle (Eds.), *Archaeogenetics: DNA and the population prehistory of Europe* (pp. 301– 304). Cambridge, England: The McDonald Institute.
-

[Google Scholar](#)

- Dettlaff-Kakol, A., & Pawlowski, R. (2002). First Polish DNA “manhunt”—An application of Y-chromosome STRs. *International Journal of Legal Medicine*, **116**(5), 289– 291.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Diaz-Lacava, A., Walier, M., Willuweit, S., Wienker, T. F., Fimmers, R., Baur, M. P., & Roewer, L. (2011). Geostatistical inference of main Y-STR-haplotype groups in Europe. *Forensic Science International Genetics*, **5**(2), 91– 94.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**(2), 479– 491.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Ge, J., Sun, H., Li, H., Liu, C., Yan, J., & Budowle, B. (2014). Future directions of forensic DNA databases. *Croatian Medical Journal*, **55**, 163– 166.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Ge, J., Yan, J., Xie, Q., Sun, H., Zhou, H., & Li, B. (2013). Development of Chinese forensic Y-STR DNA database. *Fa Yi Xue Za Zhi*, **29**, 212– 221. (In Chinese language).
-

[PubMedGoogle Scholar](#)

-
- Geppert, M., & Roewer, L. (2012). SNaPshot® minisequencing analysis of multiple ancestry-informative Y-SNPs using capillary electrophoresis. *Methods in Molecular Biology*, **830**, 127– 140.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., ... Vallone, P. M. (2017). STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. *Forensic Science International Genetics*, **31**, 111– 117.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Gopinath, S., Zhong, C., Nguyen, V., Ge, J., Lagacé, R. E., Short, M. L., & Mulero, J. J. (2016). Developmental validation of the Yfiler(®) Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Science International Genetics*, **24**, 164– 175.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Hallast, P., Batini, C., Zadik, D., Maisano Delser, P., Wetton, J. H., Arroyo-Pardo, E., ... Jobling, M. A. (2015). The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Molecular Biology and Evolution*, **32**(3), 661– 673.

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Hanson, E. K., & Ballantyne, J. (2014). A Y-short tandem repeat specific DNA enhancement strategy to aid the analysis of late reported (≥6 days) sexual assault cases. *Medicine, Science and the Law*, **54**(4), 209– 218.

[CrossrefPubMedWeb of Science®Google Scholar](#)

- Hanson, E. K., & Ballantyne, J. (2016). Enhanced DNA profiling of the semen donor in late reported sexual assaults: Use of Y-chromosome-targeted pre-amplification and next generation Y-STR amplification systems. In W. Goodwin (Ed.), *Forensic DNA typing protocols Methods in molecular biology series* (Vol. **1420**, pp. 185– 200). New York, NY: Humana Press.

[CrossrefGoogle Scholar](#)

- Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. *Nature Reviews Genetics*, **4**(8), 598– 612.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Just, R. S., Moreno, L. I., Smerick, J. B., & Irwin, J. A. (2017). Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens. *Forensic Science International Genetics*, **28**, 1– 9.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Kayser, M. (2017). Forensic use of Y-chromosome DNA: A general overview. *Human Genetics*, **136**(5), 621– 635.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., ... Roewer, L. (1997). Evaluation of Y-chromosomal STRs: A multicenter study. *International Journal of Legal Medicine*, **110**(3), 125– 133 141-9.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., ... de Knijff, P. (2001). An extensive analysis of Y chromosomal microsatellite haplotypes in globally dispersed human populations. *American Journal of Human Genetics*, **68**(4), 990– 1018.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Kimura, M., & Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the United States of America*, **75**(6), 2868– 2872.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

-
- Köcher, S., Müller, P., Berger, B., Bodner, M., Parson, W., Roewer, L., ... DNASEqEx Consortium. (2018). Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International Genetics*, **36**, 77– 85.
-

[CrossrefPubMedWeb of Science®Google Scholar](#)

- Kwon, S. Y., Lee, H. Y., Kim, E. H., Lee, E. Y., & Shin, K. J. (2016). Investigation into the sequence structure of 23 Y chromosomal STR loci using massively parallel sequencing. *Forensic Science International Genetics*, **25**, 132– 141.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Larmuseau, M. H., Van Geystelen, A., Kayser, M., van Oven, M., & Decorte, R. (2015). Towards a consensus Y-chromosomal phylogeny and Y-SNP set in forensics in the next-generation sequencing era. *Forensic Science International Genetics*, **15**, 39– 42.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Millstein, R. L. (2010). The concepts of population and metapopulation in evolutionary biology and ecology. In M. A. Bell, D. J. Futuyma, W. F. Eanes, & J. S. Levinton (Eds.), *Evolution since Darwin: The first 150 years*. Sunderland, MA: Sinauer.
-

[Google Scholar](#)

- Mulero, J. J., Chang, C. W., Calandro, L. M., Green, R. L., Li, Y., Johnson, C. L., & Hennessy, L. K. (2006). Development and validation of the AmpFISTR Yfiler PCR amplification kit: A male specific, single amplification 17 Y-STR multiplex system. *Journal of Forensic Sciences*, **51**(1), 64– 75.
-

[Wiley Online LibraryCASPubMedWeb of Science®Google Scholar](#)

- Murdock, G. P. (1967). *Ethnographic atlas*. Pittsburgh, PA: University of Pittsburgh Press.
-

[Google Scholar](#)

-
- National Institute of Justice. (2017). National best practices for sexual assault kits: A multidisciplinary approach. Retrieved from <https://nij.gov/safer-report>
-

[Google Scholar](#)

- National Research Council. (2009). Strengthening forensic science in the United States: A path forward. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>
-

[Google Scholar](#)

- Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., & Stoneking, M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nature Genetics*, **29**(1), 20– 21.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Pinotti, T., Bergström, A., Geppert, M., Bawn, M., Ohasi, D., Shi, W., ... Tyler-Smith, C. (2018). Y chromosome sequences reveal a short Beringian standstill, rapid expansion, and early population structure of Native American Founders. *Current Biology*, **29**(1), 149– 157.
-

[CrossrefPubMedWeb of Science®Google Scholar](#)

- Purps, J., Geppert, M., Nagy, M., & Roewer, L. (2015). Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases. *Forensic Science International Genetics*, **19**, 238– 242.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., ... Roewer, L. (2014). A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International Genetics*, **12**, 12– 23.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

-
- Qian, X., Hou, J., Wang, Z., Ye, Y., Lang, M., Gao, T., ... Hou, Y. (2017). Next Generation Sequencing Plus (NGS+) with Y-chromosomal Markers for Forensic Pedigree Searches. *Scientific Reports*, **7**(1), 11324.
-

[CrossrefPubMedWeb of Science®Google Scholar](#)

- Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S. H., Hammer, M. F., Mehdi, S. Q., ... McElreavey, K. (2001). Y chromosome lineages trace diffusion of people and languages in southwestern Asia. *American Journal of Human Genetics*, **68**(2), 537– 542.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Redd, A. J., Agellon, A. B., Kearney, V. A., Contreras, V. A., Karafet, T., Park, H., ... Hammer, M. F. (2002). Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Science International*, **130**(2-3), 97– 111.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Redd, A. J., Chamberlain, V. F., Kearney, V. F., Stover, D., Karafet, T., Calderon, K., ... Hammer, M. F. (2006). Genetic structure among 38 populations from the United States based on 11 US core Y chromosome STRs. *Journal of Forensic Sciences*, **51**(3), 580– 585.
-

[Wiley Online LibraryCASPubMedWeb of Science®Google Scholar](#)

- Roewer, L. (2016). The Y-chromosome haplotype reference database (YHRD)—Publicly available reference and research datasets for the forensic interpretation of Y-chromosome STR profiles. In *Handbook of forensic genetics* (pp. 231– 248). London, England: World Scientific Publishing Europe Lda.
-

[CrossrefGoogle Scholar](#)

- Roewer, L., Croucher, P. J., Willuweit, S., Lu, T. T., Kayser, M., Lessig, R., ... Krawczak, M. (2005). Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Human Genetics*, **116**(4), 279– 291.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

-
- Roewer, L., & Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Science International*, **53**, 163– 171.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., & de Knijff, P. (1996). Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Human Molecular Genetics*, **5**(7), 1029– 1033. Erratum in: *Hum Mol Genet* 1997, 6(5):828.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., ... Kayser, M. (2001). Online reference database of EuropeanY-chromosomal short tandem repeat (STR) haplotypes. *Forensic Science International*, **118**(2-3), 106– 113.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Roewer, L., & Willuweit, S. (2018). Y-chromosomal STR analysis in forensic practice. *Rechtsmedizin*, **28**(2), 149– 164. In German language.
-

[CrossrefWeb of Science®Google Scholar](#)

- Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., ... Jobling, M. A. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *American Journal of Human Genetics*, **67**(6), 1526– 1543.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Schiermeier, Q. (2001). Y-chromosome analysis urged for sex crimes. *Nature*, **413**, 6.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

-
- Shewale, J. G., Nasir, H., Schneida, E., Gross, A. M., Budowle, B., & Sinha, S. K. (2004). Y-chromosome STR system, Y-PLEX 12, for forensic casework: Development and validation. *Journal of Forensic Sciences*, **49**(6), 1278– 1290.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Taylor, D., Curran, J., & Buckleton, J. (2018). Likelihood ratio development for mixed Y-STR profiles. *Forensic Science International Genetics*, **35**, 82– 96.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Thompson, J. M., Ewing, M. M., Frank, W. E., Pogemiller, J. J., Nolde, C. A., Koehler, D. J., ... Storts, D. R. (2013). Developmental validation of the PowerPlex® Y23 system: A single multiplex Y-STR analysis system for casework and database samples. *Forensic Science International Genetics*, **7**(2), 240– 250.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Todd, E. (2011). *L'origine des systèmes familiaux*. Gallimard, Paris: NRF Essais. In French language.
-

[Web of Science®Google Scholar](#)

- van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R., & Larmuseau, M. H. (2014). Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome. *Human Mutation*, **35**(2), 187– 191.
-

[Wiley Online LibraryPubMedWeb of Science®Google Scholar](#)

- Wei, W., Ayub, Q., Xue, Y., & Tyler-Smith, C. (2013). A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping. *Forensic Science International Genetics*, **7**(6), 568– 572.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Willuweit, S., Anslinger, K., Bäßler, G., Eckert, M., Fimmers, R., Hohoff, C., ... Roewer, L. (2018). Joint recommendations of the project group “Statistical analysis of DNA”

and the German Stain Commission on the statistical analysis of Y-chromosomal DNA typing results. *Rechtsmedizin*, **28**(2), 138– 142. In German language.

[CrossrefWeb of Science®Google Scholar](#)

- Willuweit, S., & Roewer, L. (2015). The new Y Chromosome Haplotype Reference Database. *Forensic Science International Genetics*, **15**, 43– 48.
-

[CrossrefCASPubMedWeb of Science®Google Scholar](#)

- Willuweit, S., Weirich, V., Schneider, P. M., & Roewer, L. (2018). Gemeinsame Empfehlungen der Projektgruppe "Biostatistische DNA-Berechnungen" und der Spurenkommission zur biostatistischen Bewertung von Y-chromosomalen DNA-Befunden. *Neue Zeitschrift für Strafrecht*, **8**, 437– 440. In German language.
-

[Google Scholar](#)

- Yao, H. B., Wen, S. Q., Tong, X., Zhou, B., Du, P., Shi, M., ... Li, H. (2016). Y chromosomal clue successfully facilitated the arrest of Baiyin serial killer. *Scientific Bulletin*, **61**, 1715– 1717.
-

[Google Scholar](#)