



IN SILICO ANALYSIS OF DNA PROFILE USED IN FORENSIC SCIENCE

SINGH SATENDRA*, GAUTAM B., YADAV P.K., JAIN P.A., FARMER R., RAM G.D.¹

Department of Computational Biology & Bioinformatics, JSBBE, SHIATS, Allahabad-211007

¹Department of Tissue Engineering, JSBBE, SHIATS, Allahabad-211007

*Corresponding author satendralike@gmail.com

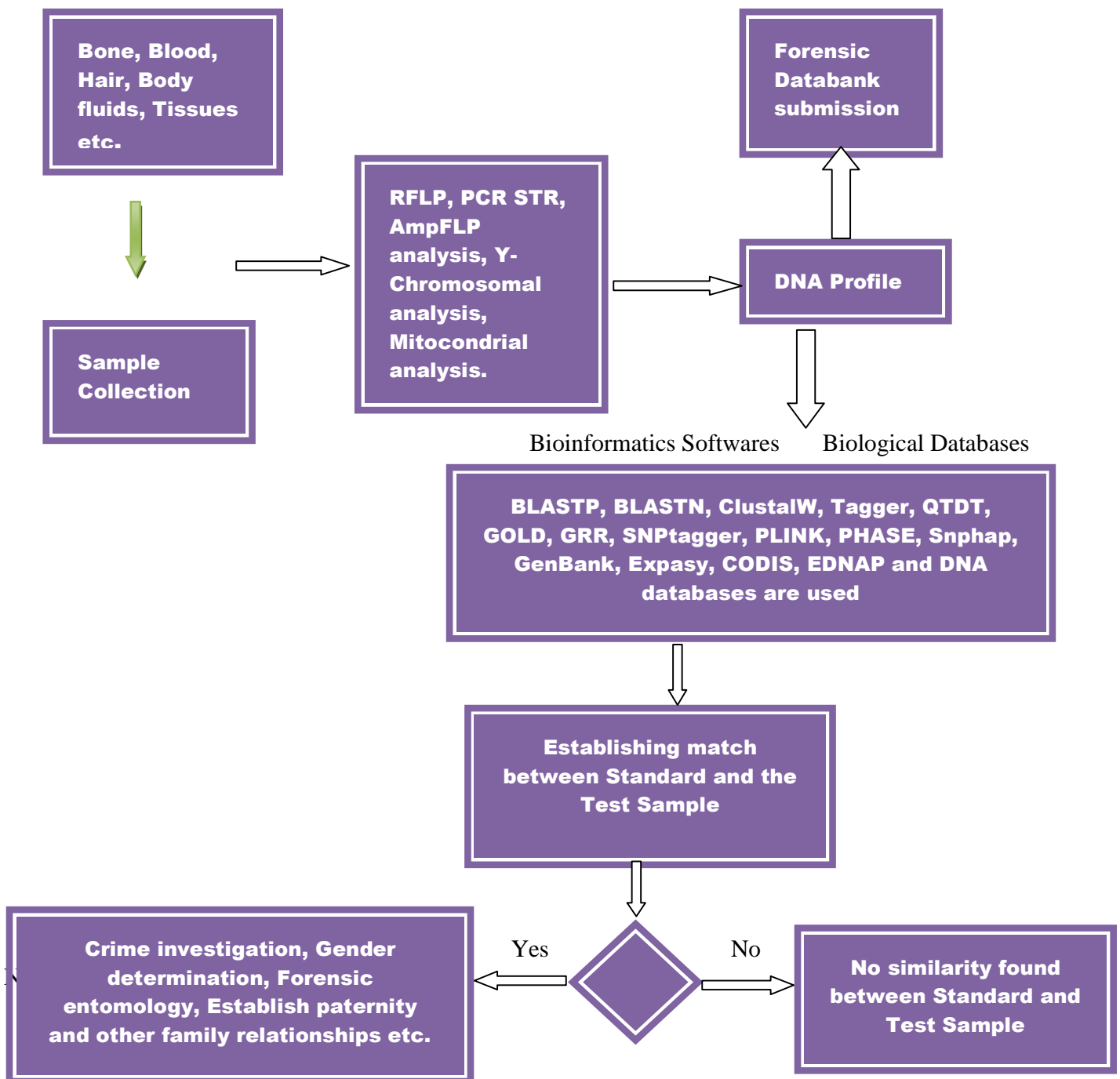
ABSTRACT

Aims: The present work examined the applications of Bioinformatics in carrying out the analysis of DNA Fingerprints, which are used in the forensic sciences to investigate the crime or to establish the genetic relationship. **Methods:** Experiment, includes first to identify and generate DNA profile of the sample under investigation by using different techniques like RFLP analysis, PCR analysis, STR analysis, AmpFLP analysis, Y-Chromosome analysis, and Mitochondrial Analysis. Once the DNA profile is generated different Bioinformatics tools like, BLASTP, BLASTN, ClustalW, Tagger, QTD, GOLD, GRR, SNPtagger, PLINK, PHASE, SnpHap, and different Biological Databases like GenBank, ExPasy, CODIS, EDNAP and DNA databases are used to carry out analysis of DNA profile. **Results:** Analysis of DNA profile helps in establishing the match of the sample DNA profile under investigation with that of the standard one. If a match is found in the sample and standard one then on its bases it can be said that both belongs to the same person and absence of match indicated lack of similarity or relationship of sample with that of the standard one. **Conclusions:** Based on the above modern Computational and Bioinformatics tools and analysis we can address questions on forensic science like forensic examinations on defendant and prosecution positions during crime investigation and criminal court proceedings, Gender determination and Forensic entomology.

KEYWORDS Forensic science; DNA profile; Bioinformatics; DNA Markers.



IN SILICO ANALYSIS OF DNA PROFILE USED IN FORENSIC SCIENCE





IN SILICO ANALYSIS OF DNA PROFILE USED IN FORENSIC SCIENCE

INTRODUCTION

Bioinformatics and forensic DNA are inherently interdisciplinary and draw their techniques from statistics and computer science bringing them to bear on problems in biology and law. Personal identification and relatedness to other individuals are the two major subjects of forensic DNA analysis. Typical contexts for forensic analysis are disputes on kinship; for example paternity disputes, suspected incest case, corpse identification, alimentary frauds, semen detection on underwear for suspected infidelity, insurance company fraud investigations when the actual driver in a vehicle accident is in question, criminal matters, autopsies for human identification following accident investigations. Genetic tests have been widely used for forensic evidences and mass-fatality identification (terrorist attacks, airplane crash, tsunami disaster). Genetic testing results are integrated with information collected by multidisciplinary teams composed of medical examiners, forensic pathologists, anthropologists, forensic dentists, fingerprint specialists, radiologists and experts in search and recovery of physical evidence. [3-5]

DNA profiling (also called DNA testing, DNA typing, or genetic fingerprinting) is a technique employed by forensic scientists to assist in the identification of individuals on the basis of their respective DNA profiles. DNA profiles are encrypted sets of numbers that reflect a person's DNA makeup, which can also be used as the person's identifier.

Although 99.9% of human DNA sequences are the same in every person, enough of the DNA is different to distinguish one individual from another. DNA profiling uses repetitive ("repeat") sequences that are highly variable, called variable number tandem

repeats (VNTR). VNTRs loci are very similar between closely related humans, but so variable that unrelated individuals are extremely unlikely to have the same VNTRs. [10]

Methods RFLP analysis

The first methods for finding out genetics used for DNA profiling involved restriction enzyme digestion, followed by Southern blot analysis. The enzymes and DNA probes were used to analyze VNTR loci. RFLP testing requires larger amounts of DNA and the DNA must be undegraded. [3,9]

PCR analysis

PCR (polymerase chain reaction) greatly amplifies the amounts of a specific region of DNA, even from cigarette ends, toothbrushes, used shavers, tissues on a gun muzzle and on bullets by using oligonucleotide primers and a thermostable DNA polymerase. PCR method is readily adaptable for analyzing VNTR loci. Also, commercial kits are available that analyze single nucleotide polymorphisms (SNPs). PCR has tended to replace RFLP in forensic testing primarily because PCR based tests are faster and more sensitive. [10]



IN SILICO ANALYSIS OF DNA PROFILE USED IN FORENSIC SCIENCE

STR analysis

STRs (Short Tandem Repeat) are used in most of the currently popular forensic DNA tests. STR is a generic term that describes any short, repeating DNA sequence. For example, the DNA sequence ATATATATATAT is an STR that has a repeating motif consisting of two bases, A and T. People vary from one another in the number of repeats they have, at least for some STR loci. PCR is used to help in the analysis of STR [7].

AmpFLP

AmpFLP (Amplified fragment length polymorphism), is faster than RFLP analysis and used PCR to amplify DNA samples. It relied on variable number tandem repeat (VNTR) polymorphisms to distinguish various alleles. Due to its relatively low cost and ease of set-up and operation, AmpFLP remains popular in lower income countries. [7]

Mitochondrial DNA analysis

Mitochondrial DNA (mtDNA) is present in high copy number in human cells. mtDNA analysis of insects uses sequence information in the Cytochrome Oxidase Region (either I, II or both). Sequencing COI or COII will allow investigators to identify the insect by comparing it to known sequences and for phylogenetic analysis. [5]

Gender determination

Testing the amelogenin marker tells us information of the gender [11]. Amelogenin is a matrix protein, which comprises 90% of all the proteins in the tooth enamel. Its purpose is to help distinguish male and female sources of DNA by detecting the X and Y-chromosomes. [6,8]

Using plant, bacteria, pollen and other Bioinformatics data

Forensic entomology allows determining postmortem intervals by analyzing the developmental status of certain species. The use of phylogenetic inference may lead to more precise taxonomic identification of the species, providing geographical information. Similarly, pollen and spores analysis, i.e. palynology, may provide information of a particular place and a certain time frame. [4]

Softwares

There is a large variety of software useful for haplotype analysis. Most of this software comes with example data sets and manuals so it is easy to try different programs and make comparison on the basis of the specific needs and data sets.

Haploview - (www.broad.mit.edu/mpg/haploview) Haploview currently supports the haplotype block analysis, haplotype population frequency estimation, SNP and implementation of Tagger, tag SNP selection algorithm. [2]

Haplofreq- (http://www.cs.princeton.edu/haplo_freq/) estimates the haplotype frequencies over a short genomic region given the genotypes with missing data.[1]

HapMap project- (<http://www.hapmap.org>) focuses on DNA sequence differences among individuals.

Tagger- (www.broad.mit.edu/mpg/tagger/) is a tool for the selection and evaluation of tag SNPs from genotype data such as that from the International HapMap Project.



IN SILICO ANALYSIS OF DNA PROFILE USED IN FORENSIC SCIENCE

QTDT- (<http://bioinformatics.well.ox.ac.uk/project-ld.shtml>) Linkage and association variance components analysis of quantitative traits.

GOLD- (<http://bioinformatics.well.ox.ac.uk/project-ld.shtml>) Graphical Display of Linkage Disequilibrium: color-coding of LD-coefficient matrices. Distribution also includes some handy programs for calculation of LD coefficients.

GRR- (<http://bioinformatics.well.ox.ac.uk/GRR>) Graphical Representation of Relationship errors. Simple representations of observed allele sharing in families to highlight erroneous coding of relationships amongst members.

PLINK- (<http://pngu.mgh.harvard.edu/~purcell/plink/haplo.shtml>) Whole genome association analysis toolset.

PHASE (<http://www.stat.washington.edu/stephens/software.html>) software for haplotype reconstruction, and recombination rate estimation from population data.

Snphap (<http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt>) program for estimating frequencies of haplotypes of large numbers of diallelic markers from unphased genotype data from unrelated subjects.

Arlequin- (<http://anthro.unige.ch/arlequin/software/>) Multipurpose population genetics software implementing a large set of methods such as AMOVA using microsatellite data and dominant markers (RAPDs, AFLPs).

Hapsim- Package for haplotype data simulation.

BLASTN- Search will provide a match (if available) from a laboratory determined sequence to a sequence maintained in GenBank.

CLUSTAL- Facilitates multiple sequence alignment and can generate phylogenetic trees.

Other interesting software and data sets are <http://www.nslj-genetics.org/soft/>, <http://linkage.rockefeller.edu/soft/>, <http://www.animalgenome.org/soft/>, www.hapmap.org/download/encode1.html, <http://www.broad.mit.edu/tools/data.html>.

DATABASES

1. STR database <http://www.cstl.nist.gov/biotech/strbase/index.htm> and [http://www.str-base.org/index.php-descriptions-of-annotated-sequence\[9\]](http://www.str-base.org/index.php-descriptions-of-annotated-sequence[9])
2. In the United States the FBI has standardized a set of 13 VNTR assays for DNA typing, and has organized the CODIS (Combined DNA Index System) database for forensic identification in criminal cases.
3. European DNA Profiling Group (EDNAP)
4. Standardization of DNA Profiling in the EU (STADNAP) European Network of Forensic Science Institutes (ENFSI, <http://www.enfsi.org/>)
5. Various other Biological databases like GenBank, EMBL etc.

REFERENCES

1. Akey J, Jin Li, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *EurJHuman Genet* 2001;9:291–300.



IN SILICO ANALYSIS OF DNA PROFILE USED IN FORENSIC SCIENCE

2. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
3. Balding DJ, Nichols RA. DNA profile match probability calculations: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int* 1994;64:125–40.
4. Byrd, J. H; Castner, J. L. eds. *Forensic Entomology: The Utility of Arthropods in Legal Investigations*. CRC Press: New York, 2001.
5. Califano J. Interview with Amanda Parrish. Mitochondrial genome scan finds cancer mutations in saliva DNA samples. *AffymetrixMicroarray Bull* 2005;1:17
6. Delgado S, Girondot M, Sire J. Molecular evolution of amelogenin in mammals. *JMol Evol* 2005;60:12–30.
7. Mark B. DNA typing in forensic medicine and in criminal investigations: a current survey. *Naturwissenschaften* 1997;84: 181–8.
8. Mitchell RJ, Kreskas M, Baxter E, et al. An investigation of sequence deletions of amelogenin (AMELY), a Y-chromosome locus commonly used for gender determination. *AnnHum Biol* 2006;33:227–40.
9. Piazza F, Lio' P. Statistical analysis of simple repeats in the human genome. *Physica A* 2005;347:472–88.
10. Yancy HF, Mohla A, Farrell DE, Myers MJ. Evaluation of a rapid PCR-based method for the detection of animal material. *J Food Prot* 2005;68:2651–5.