International Journal of Engineering Research and Science & Technology

ISSN : 2319-5991
Vol. 4, No. 3
August 2015

www.ijerst.com

Email: editorijerst@gmail.com or editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 4, No. 3, August 2015

Research Paper

EVALUATING EMERGING TECHNOLOGIES APPLIED IN FORENSIC ANALYSIS

Halima Bano[1]*, Ola Mohamed[1], Fatemah Bensaheb[1], Shalini Behl[1] and Mohammad Nazir[1]

*Corresponding Author: **Halima Bano** ✉ s20111002@ums.ae

This review presents the advancement in technologies used in forensic analysis. In forensic analysis, DNA evidence obtained from variety of sources is used to differentiate one cell, individual from another. The insufficiency of DNA material available for analysis is a problem as much information cannot be deduced. The scientific advancement has solved critical problem to certain extent. The innovations of new technologies have proved to be more efficient in forensic analysis. This review elaborates information about the source of DNA evidence, the genetic markers (Next Generation Sequencing, Mitochondrial DNA, Single Nucleotide Polymorphism (SNP), Tandem repeats, Polymerase Chain Reaction, and Restriction Fragment Length Polymorphism) and how automated technologies provides platform for simultaneous analysis of multiple genetic markers. In addition to discussion of techniques advancement, this article provides application of these techniques in forensic analysis along with their pros and cons.

Keywords: Genetic Markers, Next Generation Sequencing, Forensic Analysis, SNPs, Tandem repeats

INTRODUCTION

A genetic marker is an effortlessly identifiable piece of genetic material, usually DNA, which can be used in the laboratory to differentiate one from the other cells, individuals, populations, or species. The utilization of genetic marker starts with extraction of DNA from tissues of plant/ animals. Genetic markers, therefore, allows us to describe genetic diversity. The repertoire of genetic markers used for characterization of biological materials has evolved substantially within the forensic field with each marker set and concomitant technology platform, augmenting resolution and/or sensitivity of detection (Budowle and Daal, 2008). The extensive studies of the genome, DNA fingerprinting technique was found to be reliable and useful in forensic analysis. DNA profiling has now succeeded the method “DNA fingerprint” in the field of forensics. DNA profiling is a technique which helps the forensic scientists finds the identity of a person for their DNA.

[1] University of Modern Sciences, Al Twar-3, 7A Street, United Arab Emirates.

In forensic analysis, DNA evidence is obtained from a variety of sources. The most suitable sources include blood, semen, vaginal fluid, nasal secretion, and hair with roots. Prior to any analysis, it is essential to extract DNA from samples. The specific framework of this protocol has its specific problems, ranging from technical approach, to ethical issues, through statistical analysis. The technical approach implemented has to take into account the possible problems associated with DNA profiling. The major difficulty, at this point, is related to scarcity and/or degradation of the analyzable genetic material (Ziêtkiewicz *et.al.,* 2011). The insufficiency of DNA available for the analysis may reflect on the critically small amount of material available for DNA extraction. The incompetent data is because, the source material mostly comes in the form of contact traces, for example, on personal items of a victim (to be used as reference samples) or on forensic samples (meant to reveal the identity of a perpetrator) (Ziêtkiewicz *et al.,* 2011). This problem affects the choice of extraction methods, of DNA profiling techniques and, to some extent, of markers to be considered.

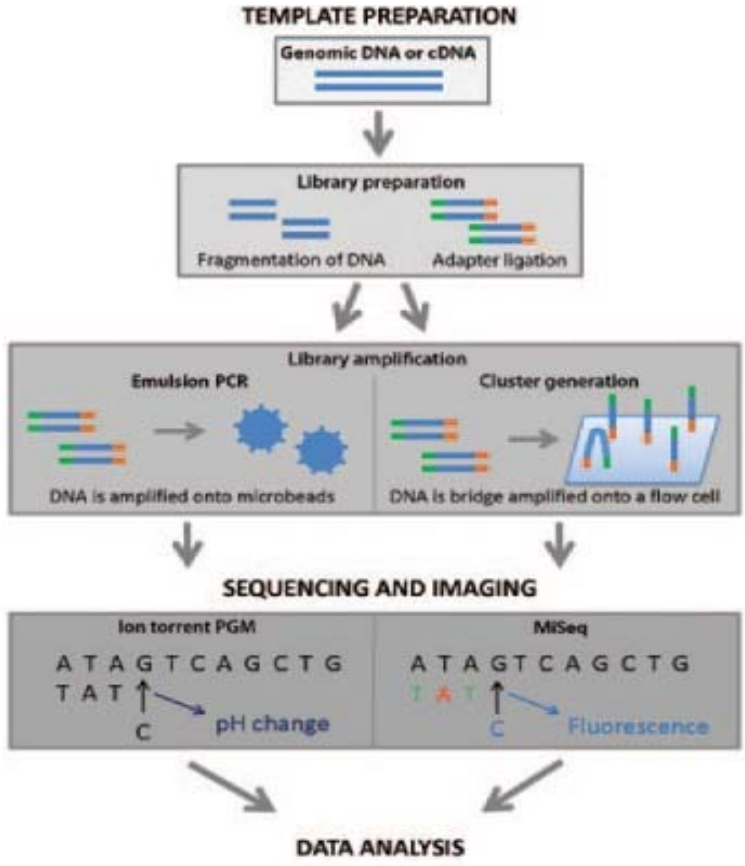
The scientific progress has solved these critical problems to certain extent. The introduction of modern technology has performed human identity testing with greater efficiency and improved abilities. It has made individual identification possible which is desirable in number of situation, For instance, where the remains are highly fragmented, the use of fingerprints or dental records is a vain attempt and DNA analysis often becomes the only method to bring closure to the chaos of such an event. The new technological innovations in every aspect of our life have helped in the advancement of forensic analysis technologies. These advancements will lead to a more efficient approach in understanding of the methods of Next generation sequencing and analyzing the forensic evidence (Figure 1).

Sequencing is most important technique in many fields, ranging from anthropology, forensic sciences, biotechnology, molecular biology, genetics, evolutionary biology, medicine, metagenomics and many more. The remarkable innovation in sequencing has revolutionized the fundamentals of many fields. Some of most applications of sequencing includes Criminal Investigation, Victim Identifications**,** Paternity/ Maternity testing, Disease identification, Missing person investigations, and Mass casualty incident. The Mitochondrial DNA (MtDNA) analysis is robust as it works with good quality DNA and also with difficult samples when all possible extra steps for recovering a profile are applied. Thus, this technique is extremely convenient in analysis of even degraded samples.

Another powerful technique of forensic analysis is by genetic typing, i.e., Single Nucleotide Polymorphism (SNPs). They are the most stable genetic markers with low mutation rate. Short Tandem Repeats (STRs) allow the recovery of information even with small amount of biological material. But, However, Variable Number of Tandem Repeats (VNTRs) is more commonly used in linkage analysis of diploid genome. And Restriction Fragment Length Polymorphism (RFLP) allows analysis of different length of DNA fragment and also plays a vital role in mapping genome. The Polymerase Chain Reaction (PCR) is a major tool used in genetic fingerprinting and many other techniques of forensic analysis.

In this review, we will discuss the applications of developing technologies in Next generation

Figure 1: Next-Generation Sequencing Methodology



Source: Grada A and Weinbrecht K (2013), Next-Generation Sequencing: Methodology and Application, Journal of Investigative Dermatology, 133 (11)

sequencing in section 2, along with Mitochondrial DNA in section 3, Single Nucleotide Polymorphism (SNPS) in section 4, tandem repeats in section 5, Restriction Fragment Length Polymorphism (RFLP) in section 6, and PCR in section 7. The mechanism, pros, and limitation of each technique will be discussed. Finally, in section 8 will have concluding remarks.

NEXT GENERATION SEQUENCING

Next Generation Sequencing (NGS), also known as high throughput sequence was developed as incentive for cheaper and faster sequencing method. After the completion of Human Genome Project with first generation sequencing, known as Sanger sequencing. A fundamentally different approach of sequence to overcome barriers, a new technology emerged that elicited numerous revolutionary discoveries and ignited a revolt in genomic science. NGS innovation have launch to distinctive quality, with progressively boundless appropriation of a few stages that separately execute distinctive kinds of immensely parallel cyclic-array sequencing. NGS platforms perform massively parallel sequencing, during which millions of fragments of DNA from a single sample are sequenced in unison. Massively parallel

sequencing technology facilitates high-throughput sequencing, which allows an entire genome to be sequenced in less than one day. In the past decade, several NGS platforms have been developed that provide low-cost, high-throughput sequencing (Grada, and Weinbrecht, 2013). The most frequently used platforms in research and clinical labs today are: the Life Technologies Ion Torrent Personal Genome Machine (PGM) and the Illumina MiSeq. The creation of these and other NGS platforms has made sequencing accessible to more labs, rapidly increasing the amount of research and clinical diagnostics being performed with nucleic acid sequencing (Grada and Weinbrecht, 2013).

WHAT NEXT GENERATION SEQUENCING DOES?

NGS does a lot of DNA sequencing, which provides much cheaper, high throughput alternative to traditional Sanger Sequencing and now whole genome can be sequenced in a day. Next-generation DNA sequencing has the potential to dramatically accelerate biological and biomedical research, by enabling the comprehensive analysis of genomes, transcriptomes and interactomes to become inexpensive, routine and widespread, rather than requiring significant production-scale efforts (Shendure and Ji, 2008) (Table 1). Next

Table 1: Comparing Characteristics of NGS Technologies

Instruments	Feature Generation	Sequencing Method	Read Length (nucleotide)	No of reads[a]	Output (Gb)[a]	No of samples[a,b]	Run time	Raw Accuracy	Pros	Cons	References
Roche 454	Emulsion PCR	Pyrosequencing	400[c]	$1x10^5$	0.035	132	9 h	99.5%	Long reads	Homopolymer errors, Expensive	Shendure, J., & Ji, H. (2008), Ziętkiewicz, E et.al (2011), Berglund et al (2011), Bosch JR, Grody WW (2008), Tucker T, Marra M, Friedman JM (2009),
Life Technologies SOLID	Emulsion PCR	Ligase (octamers with two-base encoding)	75[e]	$1.5x10^9$	180	1,152	14 days[d]	99.94%	Inherent error correction	Short reads	Shendure, J., & Ji, H. (2008) , Pareek CS, Smoczynski R, Tretyn A (2011)
Polonator	Emulsion PCR	Ligase (nonamers)	135bp	26 read/run	10 per run	180 million beads	4 days	98%	High throughput, inexpensive instrument	Lower efficiency,	Shendure, J., & Ji, H. (2008), http://www.polonator.org/protocols/polony.aspx
Heliscope	Single Molecules	asynchronous extensions	35[f]	$1x10^9$	35	4800	8 days	>99%	RNA Sequencing, Single Molecule Sequencing	Short reads High error rate	Shendure, J., & Ji, H. (2008), Pareek CS, Smoczynski R, Tretyn A (2011)
Illumina MiSeq	Bridge-PCR on a solid surface	Reversible dye terminators	150	$5x10^6$	1.5	96	27h	>98.5%	Short run time Ease of use	Expensive per base	Shendure, J., & Ji, H. (2008), Pareek CS, Smoczynski R, Tretyn A (2011)
Ion Torrent PGM	Emulsion PCR	Physicochemical	>100[f]	$1x10^6$	1	16	2 h	98%	Low cost instrument upgraded through disposable clear trajectory to improved performance	Higher error rate More hand-on time	Shendure, J., & Ji, H. (2008), http://www.moleculargecologist.com/next-gen-table-4-2013/
SMRT	Single molecule real time sequencing by synthesis	Sequencing by synthesis	Longer than 1000	$1x10^5$	0.1	1	90 Minutes	99.99%	Single Molecule Sequencing long reads, short run time	High error rate, low yield	Pareek CS, Smoczynski R, Tretyn A (2011), Shendure, J., & Ji, H. (2008),http://www.pacificbiosciences.com/pdf/Perspective_UnderstandingAccuracySMRTSequencing.pdf

Generation Sequencing (NGS) refers to technologies that do not rely on traditional Dideoxy-nucleotide (Sanger) sequencing where labeled DNA fragments are physically resolved by electrophoresis, but relies on different strategy (Richardson, (2010).

NGS accelerates the detection of genes and regulatory elements allied with disease by its high throughput sequencing of the human genome. With a growing variation in molecular method, a broad range of biological phenomena such as genetic variation, RNA expression, Protein-DNA interactions and chromosome conformation can be assessed by high-throughput DNA sequencing. The subjects of metagenomics (environmental sequencing) and sequencing of ancient DNA samples are two additional areas where NGS platforms have provided unforeseen opportunities for generating previously unattainable levels of understanding from these samples (Richardson, 2010).

NGS can be used for both genome and transcriptome analysis. In genome analysis, it permits high-quality variant calling for SNPs, insertions and deletions, and allows the analysis of Copy Number Variants (CNVs) and other structural rearrangements. In transcriptome analysis, NGS can be used to annotate coding SNPs, discover transcript isoforms, identify regulatory RNAs, characterize splice junctions and determine the relative abundance of transcripts. It allows generating full sequence information from any poly (A)-tailed RNA, analyzing gene expression and identifying novel or rare transcripts in one experiment (Ziêtkiewicz, *et al.,* 2011).

NGS allows the identification of disease-causing mutations for diagnosis of pathological conditions by targeting sequence. RNA-seq can provide information on the entire transcriptome of a sample in a single analysis without requiring previous knowledge of the genetic sequence of an organism. This technique offers a strong alternative to the use of microarrays in gene expression studies (Grada and Weinbrecht, 2013).

The increase in throughput and reduction in cost achieved by massively parallel sequencing are a result of two factors: (1) many thousands or millions of sequencing reactions are performed in parallel rather than just 1 to 96 at a time, as in conventional sequencing machines; and (2) cloning or template amplification of the DNA fragments that are being sequenced is either unnecessary (in single molecule sequencing) or fully automated within the same instrument that performs massively parallel sequencing. Another advantage of massively parallel sequencing is the ability to detect minor alleles accurately (Tucker *et al.,* 2009).

NGS METHODOLOGIES

NGS technologies allows the simultaneous analysis of hundreds of loci or even of the whole genome, without being dependent on genome annotation. NGS produces massively parallel sequence data, in some cases in excess of one billion short reads per instrument run (Tucker *et al.,* 2009; Li *et al.,* 2009a; Harismendy *et al.,* 2009; Pareek *et al.,* 2011). Cloning or template amplification of the DNA fragments that are being sequenced is fully automated and intimately coupled with the sequence reading process. NGS provides an increased depth of sequence coverage, which leads to a higher accuracy of the reads and allows inferences on the copy number of analyzed segments (Tucker *et al.,* 2009; Ziêtkiewicz *et.al*., 2011).

The three best-known NGS platforms are: Illumina Genome Analyzer, ABI SOLID and Roche 454 Genome Sequencer. All of them involve the template amplification step, but they use different sequencing chemistries and amplification approaches (Tucker *et al.,* 2009). Illumina uses bridge-PCR on a solid surface, while SOLID and 454 rely on PCR in emulsion; the sequencing chemistry is polymerase-based (with fluorescently labelled nucleotides), ligation-based and pyrosequencing, respectively. Illumina and SOLID use fluorescent-based signal detection and 454 are based on the detection of light emitted upon pyrophosphate disintegration. The length of the reads is 75 bp for Illumina GA, up to 50 bp for ABI SOLID and up to 400 bp for Roche 454, while the raw base accuracy is 99.5% for Illumina and Roche, and 99.94% for ABI Solid (Bosch *et al.,* 2008; Tucker *et al.m* 2009; Ziêtkiewicz *et al.,* 2011).

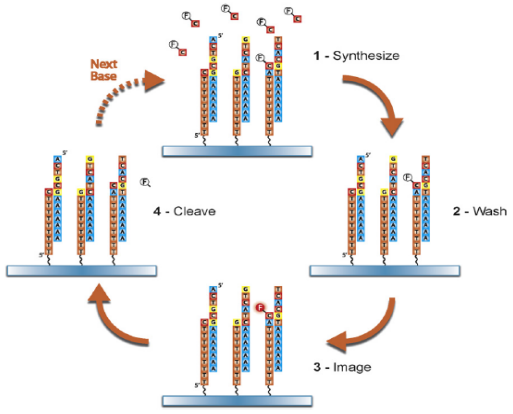
Most of the NGS platforms (except Roche 454) provide very short reads (50–100 bp). While the possibility to read very short sequences is an important advantage for application to degraded DNA material, this characteristics makes sequence assembly of repetitive and rearranged DNA segments more difficult. Short length of the reads also complicates interpretation when it is necessary to determine the phase of variants, as in the case of mixed DNA samples. Currently, NGS is more expensive in comparison to the array-based technology. As this may change in the future, a search for potential applications of NGS to DVI is fully justified (Ziêtkiewicz *et al.,* 2011).

Massively parallel NGS of forensically relevant loci could play a significant role in solving criminal cases and the identification of human remains. Targeted sequencing of candidate genes and genomic intervals in four individuals, using Illumina GA, SOLID and 454, demonstrated the suitability of NGS platforms for population-based targeted sequencing studies, although it also indicated the need for further optimization (Harismendy *et al*., 2009). The biggest disadvantage of NGS is a necessity of performing PCR amplification that may introduce base sequence errors or favor certain sequences over the other. Third-generation HT-Technology, also known as sequencing from a single DNA molecule, overcomes this problem because it is based on sequencing by synthesis without any amplification step. The available on the market HT-NGS platforms, offered by a number of companies, represent a different technical approach to sequencing by synthesis.

The newer Helicos technology is a polymerase-based method, which is unique in that it does not include the amplification step and uses a single DNA molecule as a template. The accuracy of the reads (30–35 bp) using this platform is 99% (Braslavsky *et al.,* 2003). A highly sensitive fluorescence detection system is used to interrogate each nucleotide directly as it is synthesized (Tucker *et al.,* 2009).

As shown in the Figure 2, Fragments are captured by poly-T oligomers tethered to an array.

Figure 2: Helicos Heliscope Sequencer Workflow

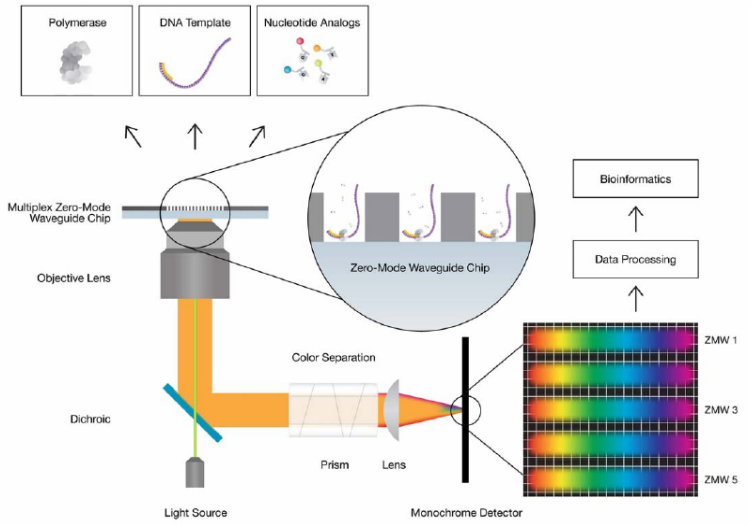


At the each sequencing cycle, polymerase and single fluorescently labeled nucleotide are added and the array is imaged. The fluorescent tag is then removed and the cycle is repeated. Reprinted with permission from Helicos BioSciences Corporation (Tucker *et al.,* 2009).

The Multiplex Polony Technology (MPT) is developed, introduced, and lead by Prof. G Church's research group and run by the privately-funded Personal Genome Project (PGP) [www.personalgenomes.org]. In this technique, several hundred sequencing templates are deposited onto thin agarose layers and sequences are determined in parallel. This method presents increase of several orders of magnitudes in the number of samples which can be analyzed simultaneously. It has the advantage, in terms of large reduction of the reaction volumes, requiring smaller amounts of reagents and the resulting at a lower cost. The designed instrument, i.e., Danaher Motion Polonator model G.007, is capable of 10 to 35 Gbp per module per 2.5 day run. Instrument can couple with 200 of these modules to collect 100 diploid genomes at 30X coverage in 5 days, with the remaining 5 days used for repeating any weak runs to assure 98% coverage at 1E-5 accuracy. With the significant reduced volume of reagents, the cost per unit volume is lower about 10-fold and the company hopes to meet the goal of $1000 per genome soon (Pareek *et al.,* 2011).

SMRT sequencing technology is based on a single molecule, real-time sequencing. The process of enzymatic incorporation of labelled nucleotides is performed on Zero-Mode Waveguides (ZMWs) located on the SMRT chip. SMRT technology enables the observation of DNA synthesis as it occurs in real time. To overcome the challenges inherent in observing the an enzyme that is 15 nanometers, or nm, in diameter running in real time, pacific bioscience developed three key innovations: (1) The SMRT cell; (2) Phospho-linked nucleotides; and (3) The PacBio RSII. The whole process is captured by a CCD camera (Figure 3).

Figure 3: SMRT Sequencing Technology



Source: Figure taken from http://www.pacificbiosciences.com/products/smrttechnology/

A different approach is represented by the Nanopore DNA sequencer and Ion Torrent sequencing technology. Both platforms are based on physicochemical processing that occurs during the DNA sequencing process rather than nucleotide labelling detection. In the Nanopore sequencer, the DNA sequence is determined by the disruption of ionic current in a protein nanopore, while Ion Torrent technology is based on pH change caused by nucleotide incorporation into a DNA strand. In contrast to all of the mentioned platforms, PGM (Ion Personal Genome Machine) sequencer is laser-capture-independent because it uses ion sensors densely spaced on the bottom of a high-density array of micro-machined wells. This new approach allows producing more sequence reads on time and faster gathering of information than light-based technology platforms. It also enables to perform a wider range of sequencing applications, such as multiplexing amplicon, small RNA sequencing, paired-end sequencing and methylation analysis [http://www.iontorrent.com/] (Pareek *et al.,* 2011).

The Ion Proton™ System is the first benchtop sequencing system capable of human-scale genome, exome, or transcriptome sequencing in a few hours—with DNA-to-variants called in a single day. The system combines semiconductor sequencing technology with natural biochemistry to directly translate chemical information into digital data. By leveraging the exponential improvements in the semiconductor industry (known as Moore's Law), the Ion Proton™ Sequencing System provides an unprecedented level of scalability and flexibility to support a broad range of high throughput sequencing applications, ranging from human-scale genome to exome to transcriptome sequencing. The system's use of the simplest natural sequencing chemistry eliminates the need for expensive optics and complex sequencing chemistries, resulting in a highly affordable sequencing system to own and operate. Real-time, direct, electrical detection of sequencing, combined with the enormous amount of computing power in both the Ion Proton™ Sequencer and Ion Proton™ Torrent Server, enables generation of high quality sequencing results from DNA library to variants in a single day (https://www.lifetechnologies.com/order/catalog/product/4476610)

Features of the Ion Proton™ System

Fastest high-throughput next-generation sequencing workflow with the fastest sequencing run-times of 2-4 h on the Ion PI™ Chip. Highest number of complete sequencing runs per week, with a simple and automated workflow (when used with the Ion OneTouch™ 2 System). Semiconductor sequencing workflow is simple with robust and simple hardware. Scalable, high throughput sequencing capabilities, allowing runs for either Ion PI™ or Ion PII™* chips, along with up to 200-base single reads and flexible library options (e.g., fragment gDNA, targeted/exome, RNA). Two Ion Proton™ systems per rack, delivers the only benchtop genome center with mounting capability of optional rack. Low-cost platform and attractively priced semiconductor chips and reagents for a range of applications. Proven Ion semiconductor sequencing requires no complex optics and employs natural nucleotides to deliver highly accurate variant detection, uniformity of coverage, and sensitivity to detect low frequency variants. A Range of fast and simple library solutions and kits with low-input requirements is provided for a variety of applications, such as genomic sequencing, exome sequencing, sequencing sets of genes, and RNA sequencing. Complete end-to-end

solution from base calls to variants with the Proton™ Torrent Server and Torrent Suite Software v3.0. The DNA variation analysis across single, paired, or trio samples by simple and integrated tools for tertiary data analysis with Ion Reporter™ Software. With the availability of the groundbreaking Ion Proton™ System, rapid, high throughput sequencing is finally accessible by all laboratories (https://www.lifetechnologies.com/order/catalog/product/4476610)

The nanopore technologies achieve sequencing by single-molecule sequencing without amplification, real-time sequencing without repeated cycles and synthesis can be eliminated (Venkatesan *et al.,* 2013). Nanopore sequencing possesses a number of fruitful advantages over existing commercialized next-generation sequencing technologies. Firstly, it potentially reaches long read length >5 kbp with speed 1 bp/ns (Timp *et al*., 2010). Moreover, detection of bases is fluorescent tag-free. Thirdly, except the use of exonuclease for holding up ssDNA and nucleotide cleavage (Clarke *et al*., 2009), involvement of enzyme is remarkably obviated in nanopore sequencing (Branton *et al.,* 2010). This implies that nanopore sequencing is less sensitive to temperature throughout the sequencing reaction and reliable outcome can be maintained. Fourthly, instead of sequencing DNA during polymerization, single DNA strands are sequenced through nanopore by means of DNA strand depolymerization. Hence, hand-on time for sample preparation such as cloning and amplification steps can be shortened significantly (Liu and Li, 2012).

NGS Limitation

NGS, although much less costly in time and money in comparison to first-generation sequencing, is still too expensive for many labs. Although NGS technology appears to have an important role in future forensic studies, more work is required to fully achieve this goal, which includes overcoming problems with low-template library preparation, error rate, type estimations and issues with NGS data processing and mining. Guidelines for the application of NGS in forensic science also need to be generated. With the technical advances of NGS technology and continuous translational efforts of forensic scientists, we believe that NGS technology is likely to become an easily accessible routine method in forensic practice (Yang *et al.,* 2014).

One of the biggest consequences with NGS technologies is how to deal with all of the data produced by these platforms (Richardson, 2010). Inaccurate sequencing of homopolymer regions (spans of repeating nucleotides) on certain NGS platforms, including the Ion Torrent PGM, and short-sequencing read lengths (on average 200-500 nucleotides) can lead to sequence errors (Grada and Weinbrecht, 2013). It may require special knowledge of bioinformatics to gather accurate information from sequenced data and data analysis can be time-consuming.

The overall error rate is reduced because of the high degree of sequencing depth, typically 40-fold for a diploid genome that is necessary to achieve complete coverage with massively parallel sequencing. Higher coverage is especially important when looking for mutations or sequence variants in repetitive or massively rearranged regions (Srivatsan *et al.,* 2008; Thomas *et al*., 2006). However, greater depth means more sequencing, thereby reducing the advantages of using massively parallel sequencing (Tucker *et al.,* 2009).

Mitochondrial DNA

The mitochondrion organelle is found within cells that contain mitochondrial DNA and is involved in the production of cell energy. It is a circular molecule of DNA 16,569 base pairs in size, first referred to as the Anderson sequence. There can be between 100 and 1000 mitochondria within a cell, each contain abundant copies of mitochondrial genome. There is very little variation between individuals, as this sequence is entirely functional and highly conserved. Though there is a 1000 base pair long non-coding D-loop, known as the control region, which contains two hypervariable regions referred to as Hypervariable 1 (HV1) and Hypervariable 2 (HV2). Single Nucleotide polymorphisms (SNPs) are generally the variation within these regions. The length of mtDNA is not altered by SNPs, and these are regions that are focused on in forensic analysis of mtDNA. MtDNA is often subject to a comparatively high amount of mutation due to its privation of DNA reparation, causing variation between individuals. HV1 and HV2 differ by 1-3% between non-related individuals; the variation within this small portion is itself not especially significant. It contains its own genome distinct from the genome found in the nucleus (nuclear DNA) due to many features, including: how it is inherited; how it is replicated; its copy number; and its size. Mitochondrial DNA is circular, double stranded, and inherited maternally [http://www.encyclopedia.com].

In the MtDNA analysis, the DNA is extracted and the PCR amplifies HV1 and HV2 regions. DNA sequencing establishes the base pair sequence of these regions. This is then compared with Cambridge Reference Sequence and differences noted (http://forensicsciencecentral.co.uk/dna.shtml). To establish potential similarities, other samples can then also be analyzed and comparisons can be made. MtDNA method is generally used when other methods such as STR analysis have failed. This is frequently in case of badly degraded bodies, in cases of disaster or accidents where an individual is too badly damaged to recognize, and sometimes in classification to determine species using cytochrome b gene.

One of the most important advantages of the use of MtDNA is the possibility of analyzing even highly degraded samples. It is possible to extract mtDNA if a specimen is severely decomposed to the point that it is not possible to extract nuclear DNA. Furthermore, only a very small amount of sample is required. Conversely, the use of MtDNA does have its drawbacks. MtDNA is inherited from mother only, this cannot form a full DNA fingerprint of the individual, thus this technique is only useful if the DNA profiles of maternal families are available, such as the individual mother or biological siblings. Due to this reason, MtDNA is significantly less discriminatory than, for instance, Short tandem repeats. Identifying sequencing dissimilarity is also comparatively time consuming and costly.

DNA typing is a method used by forensics scientists to match DNA from an unknown sample to a sample collected at a crime scene. It is ideally used in special cases where the DNA is degraded or the source of the sample doesn't contain enough genomic nuclear DNA for analysis. As it is maternally inherited, the DNA from siblings and all maternal relatives should be identical (in the absence of spontaneous mutations). For this reason, the remains of missing persons can be rapidly identified by using mitochondrial DNA analysis of relatives. Additionally, there is generally a lack of recombination, an event that takes place

during nuclear DNA cell division in which two stands of DNA cross over and exchange information, thereby creating greater sequence diversity. Therefore, even matriarchal relatives separated by several generations can serve as reference samples. Nuclear DNA samples cannot provide this function, due to multiple recombination events that take place throughout the nuclear DNA genome [http://www. encyclopedia.com].

The two genomes are not mutually exclusive; instead they rely on each other for survival. The nuclear DNA can encode roughly 1,000 proteins that are targeted for the mitochondria and play a role in oxidative phosphorylation, or energy production, while the mitochondrial DNA produces energy by producing ATP as well as several other functions. All other DNA typing systems use nuclear DNA analysis.

There are several advantages to studying the mitochondrial DNA of a sample. The application of mitochondrial DNA analysis in forensic sciences stems from characteristics of he mitochondrial DNA genome, including its copy number within the cell, its hypervariable region, its size, and its sequence variations. The mitochondrial genome is roughly 16,569 base pairs in size (compared to the 3 billion base pairs in the nuclear DNA). Whereas nuclear DNA has only two copies of each gene, tightly woven into chromosomes, mitochondrial DNA can be copied 2-10 times per mitochondrion and there can be hundreds to even thousands of mitochondria per cell. With the mitochondria's role as an energy provider, different tissues contain different amounts of mitochondrial DNA, depending on the energy requirements of the cell. A higher copy number equates to greater sensitivity. This is particularly important if the DNA sample is significantly degraded, or the DNA is present only in a very small quantity. The likelihood of recovering mitochondrial DNA from a small or degraded sample is, therefore, greater in mitochondrial DNA samples compared to nuclear DNA samples since the mitochondrial DNA has a larger copy number [http://www. encyclopedia. com].

The low fidelity of DNA repair mechanisms to correct specific mitochondrial DNA mutations has led to a 5-10 fold higher mutation rate, and, in turn, a higher rate of evolution. Human identity testing employs these regions where there is hypervariability as a consequence of a higher mutation rate. Two hypervariable (HV1 and HV2) regions are part of a control region. On average, there are roughly 8 nucleotide differences between Caucasians and 15 differences between individuals with African decent in these two hypervariable regions. Mitochondrial DNA typing using HV1 and HV2 can be readily performed by using a mitochondrial DNA-specific polymerase chain reaction and amplification of genomic mitochondrial DNA. This is followed by direct DNA sequencing and identification of sequence variations [http://www.encyclopedia.com].

The sample source can often determine which DNA typing system represents the ideal approach. For example, if a hair is left at the scene of the crime, nuclear DNA can only be analyzed if the root is intact. However, mitochondrial DNA can be analyzed from anywhere along the hair follicle, including the shaft. Bones and teeth also contain mitochondrial DNA and can be used in mitochondrial DNA analysis [http://www. encyclopedia.com].

There are several disadvantages of using mitochondrial DNA typing in forensics in lieu of

nuclear DNA markers. As all individuals of the same maternal lineage are virtually indistinguishable by mitochondrial DNA analysis, identification of the remains of an individual would not be possible without comparing it to maternally-related relatives. Additionally, using mitochondrial DNA analysis to match a suspect to a sample by comparing different genomic locations might reveal a similar profile. Mitochondrial DNA should not be viewed as a unique identifier, since seemingly unrelated individuals might have an unknown shared maternal relative in their distant past. If this is the case, a mistaken match might be suggested. Finally, using a more sophisticated (multi-locus) nuclear DNA analysis will provide far greater discriminatory power [http://www.encyclopedia.com].

The Analytical Process

An mtDNA analysis begins when total genomic DNA is extracted from biological material, such as a tooth, blood sample, or hair. The Polymerase Chain Reaction (PCR) is then used to amplify, or create many copies of, the two hypervariable portions of the non-coding region of the mtDNA molecule, using flanking primers. Primers are small bits of DNA that identify and hybridize to or adhere to the ends of the region one wishes to PCR amplify, therefore targeting a region for amplification and subsequent analysis.

Care is taken to eliminate the introduction of exogenous DNA during both the extraction and amplification steps via methods such as the use of pre-packaged sterile equipment and reagents, aerosol-resistant barrier pipette tips, gloves, masks, and lab coats, separation of pre- and post-amplification areas in the lab using dedicated reagents for each, ultraviolet irradiation of equipment, and autoclaving of tubes and reagent stocks. In casework, questioned samples are processed at different times than known samples and they are usually processed in different laboratory rooms.

When adequate amounts of PCR product are amplified to provide all the necessary information about the two hypervariable regions, sequencing reactions are performed. These chemical reactions use each PCR product as a template to create a new complementary strand of DNA in which some of the As, Ts, Cs, and Gs (nucleotide bases) that make up the DNA sequence are labeled with dye. The strands created in this stage are then separated according to size by an automated sequencing machine that uses a laser to "read" the sequence, or order, of the nucleotide bases. Where possible, the sequences of both hypervariable regions are determined on both strands of the double-stranded DNA molecule, with sufficient redundancy to confirm the nucleotide substitutions that characterize that particular sample.

At least two forensic analysts independently assemble the sequence and then compare it to a standard, commonly used, reference sequence. The entire process is then repeated with a known sample, usually a blood or saliva collected from a known individual. Finally, in the event of an inclusion or match, the EMPOP mtDNA database is searched for the mitochondrial sequence that has been observed for the samples. The analysts can then report the number of observations of this type based on the nucleotide positions that have been read. A written report is provided to the submitting agency.

Non-Forensic Uses

While mtDNA is useful for forensic examinations, it has also been used extensively in two other major scientific realms. There are a number of

serious human diseases caused by deleterious mutations in gene-coding regions of the mtDNA molecule, which have been studied by the medical profession to understand their mode of inheritance. In addition, molecular anthropologists have been using mtDNA for two decades to examine both the extent of genetic variation in humans and the relatedness of populations all over the world. Because of its unique mode of maternal inheritance it can reveal ancient population histories, which might include migration patterns, expansion dates, and geographic homelands. Recently mtDNA was extracted and sequenced from a Neanderthal skeleton. These results suggested that modern humans do not share a close relationship with Neanderthals in the human evolutionary tree. While all the applications of mtDNA, including forensic, are relatively recent, the general methods for performing an mtDNA analysis are identical to those used in molecular biology laboratories all over the world for studying DNA from any living organism. There are several thousand published articles regarding mtDNA.

Advantages and Disadvantages

MtDNA has advantages and disadvantages as a forensic typing locus, especially compared to the more traditional nuclear DNA markers that are typically used. As mentioned above, mtDNA is maternally inherited, so that any maternally related individuals would be expected to share the same mtDNA sequence. This fact is useful in cases where a long deceased or missing individual is not available to provide a reference sample but any living maternal relative might do so. Because of meiotic recombination and the diploid (bi-parental) inheritance of nuclear DNA, the reconstruction of a nuclear profile from even first degree relatives of a missing individual is rarely this straightforward. However, the maternal inheritance pattern of mtDNA might also be considered problematic. Because all individuals in a maternal lineage share the same mtDNA sequence, mtDNA cannot be considered a unique identifier. In fact, apparently unrelated individuals might share an unknown maternal relative at some distant point in the past.

Interpretation of Results

At the present time the available forensic database of human mitochondrial DNA sequences has around 5000 sequences available for a search of a casework sequence. The current convention in the event of an inclusion (a match between questioned and reference sample sequences) is for the analyst to report the number of times the observed sequence is present in the database to provide some idea of its relative frequency in the database. A frequency statistic may also be used, and a 95% or 99% confidence interval is placed around the calculated frequency to account for the inherent uncertainty in the frequency calculation.

While most types appear to be rare or at least infrequent in each of the ethnic databases (African or African-origin, Asian or Asian-origin, Caucasian or European-origin, and Hispanic), there is one type which is seen in around 7% of Caucasians. However, almost two thirds of newly-typed samples have novel sequences, so we have not yet uncovered all the variation present in the general human population. For novel types, a 95% or 99% upper bound frequency calculation may be performed.

In general, the pattern observed in most populations around the world, with the exception of a few populations of anthropological interest, is that the vast majority of sequences is

uncommon, and relatively few types present at frequencies greater than 1% in the databases. Because of this fact, it will be possible to exclude greater than 99% of the population as potential contributors of a sample in most cases, except where one is dealing with a more "common" type.

In contrast, a multi-locus nuclear DNA typing profile provides vastly superior discriminatory power, such that we can now approach the possibility that a typed individual has a unique profile with respect to any other person in the world. Therefore, mtDNA can never provide the resolution of individuality that nuclear typing can. For this primary reason, it should be reserved for cases or samples for which nuclear typing is simply not possible.

Samples Typically Chosen for mtDNA Typing

Candidates for mtDNA typing analyses would most likely be: (1) shed hairs with no follicle, tissue, or root bulb attached; (2) hair shaft fragments; (3) bones or teeth which have been subjected to long periods of high acidity, high temperature, or high humidity; (4) stain or swab material that has been previously unsuccessfully typed for nuclear markers; and (5) tissue (skin, muscle, organ) that has been previously unsuccessfully typed for nuclear markers.

Hair roots, when available, should be removed from the shaft and processed separately for nuclear DNA markers prior to attempting mtDNA analysis on the hair shaft. Hair shafts or fragments are only suitable for mtDNA analysis as they can contain fewer than 100 copies of the mtDNA molecule and virtually no nuclear DNA. The same is generally true for older skeletal remains. While mtDNA typing of blood stains is possible, it is more likely that mixtures will be obtained, due to the extreme sensitivity of this form of typing in samples that unlike hairs and bones are difficult to clean before DNA extraction.

Duration and Throughput of mtDNA Analyses

Finally, it must be noted that mtDNA analyses are the most rigorous and time-consuming of DNA forensic analyses. Based on informal statistics available from all laboratories performing these typing, the rate of throughput is approximately 3-4 cases/analyst/month. The reasons for this include: (1) small/degraded samples requiring numerous PCR reactions to obtain sufficient DNA template for sequencing; (2) exhaustive procedures to control for contamination; and (3) sequencing analyses of both strands of DNA in both hypervariable regions.

In addition, for some types of samples, especially hairs, mtDNA analysis is more likely to consume the whole sample than nuclear DNA typing. For example, a single mtDNA analysis could be performed on a 0.2-2 cm hair fragment. A 4 cm fragment could have duplicate testing for confirmation of the sequence. In both cases the fragment would be totally consumed. However, a root ball, follicle, or skin tissue attached to a hair would also be consumed in a nuclear typing effort.

For both mtDNA and nuclear DNA testing there is a possibility that sufficient extracted DNA might remain for duplicate testing in another lab. Swatch, swab, stain, bone, and tooth analyses are less likely to consume all material, as these samples can often be divided, although the difficulties of obtaining enough DNA for analysis could result in consumption of these materials as well. For the reasons above, pre-analysis documentation (microscopy, photography) is desirable. (http://www.mitotyping.com)

Single Nucleotide Polymorphism (SNPs)

The most common type of variation among individuals is single nucleotide polymorphism, frequently called SNPs (pronounced as "snips") and composing around 90% variation in humans. Each SNP denotes a difference in a single nucleotide. For instance, a SNP may replace the nucleotide Cytosine (c) with the nucleotide thymine (T) in a certain section of DNA. SNPs occur in every 300 nucleotide on average, which means there are roughly 10 million SNPs in the human genome. They act as biological markers, helping scientists locate genes that are associated with disease. They play a more direct role in disease by affecting the gene's function. Researchers have found SNPs that may help predict an individual's response to certain drug, susceptibility to environmental factors such as toxins, and risk of developing particular diseases. SNPs can also be used to track the inheritance of disease genes within families. Future studies will work to identify SNPs associated with complex diseases such as heart disease, diabetes, and cancer [http://ghr.nlm.nih.gov/handbook/genomicresearch/snp].

A map of 1.42 million Single Nucleotide Polymorphisms (SNPs) distributed throughout the human genome, providing an average density on available sequence of one SNP every 1.9 kilobases. These SNPs were primarily discovered by two projects: The SNP Consortium and the analysis of clone overlaps by the International Human Genome Sequencing Consortium. The map integrates all publicly available SNPs with described genes and other genomic features. We estimate that 60,000 SNPs fall within exon (coding and untranslated regions), and 85% of exons are within 5 kb of the nearest SNP. Nucleotide diversity varies greatly across the genome, in a manner broadly consistent with a standard population genetic model of human history. This high-density SNP map provides a public resource for defining haplotype variation across the genome, and should help to identify biomedically important genes for diagnosis and therapy (Sachidanandam *et al.,* 2000).

A powerful means of establishing identity in criminal cases where biological evidence is found at crime scenes—paternity testing, inheritance matters, identification of victims in mass disasters, and identification of missing persons from human remains is by Genetic typing. The genetic evidence can be obtained from any biological material, including blood, semen, bone, hair, teeth, muscle tissue, and saliva, and can be used for classification of animals, plants, and microorganisms. The catalogue of genetic markers used for characterization of biological materials has evolved extensively within the forensic field with each marker set and affiliated technology platform, amplifying resolution and/or sensitivity of detection. SNPs offer advance opportunities for the forensic scientist to genetically illustrate an evidentiary sample or identify an unknown person.

DNA level technology helps to overcome limitations to a much greater level. First, there is a great amount of variation at the DNA level to exploit for individuality testing. Second, any biological material that encloses nucleated cells possibly can be typed for DNA polymorphisms. Third, DNA tends to be more stable in forensic samples than proteins. Given the current DNA typing techniques and the series of available genetic markers, typing of human polymorphisms at the DNA level is more subtle, more specific, and more informative than the classical protein genetic markers. Furthermore, DNA technology

gives the forensic scientist the greatest prospective to exclude individuals who have been deceptively associated with a biological sample and to reduce the number of individuals potentially included as contributors of the sample to a few (if not one) individuals. The first genetic marker systems analyzed using PCR-based systems were based on SNP variation (Budowle and Daal, 2008).

A class of genetic markers that proves useful, mainly for typing degraded samples and for increasing the amount of genetic information assembled from challenged forensic samples, is SNPs (SNPs are base substitutions, insertions, or deletions that occur at single positions in the genome of any organism. Most SNPs are biallelic and thus are not as informative on a per-locus basis as the forensically selected STR loci. Because all forensic DNA indices, particularly convicted felon DNA databases, are well-established and based on STR loci, it is unlikely that SNPs will become the primary forensic markers. It would take a substantial reduction in cost, increased throughput, and enhanced capabilities to resolve mixed samples for the forensic community to replace the STR loci. Therefore, for the foreseeable future, it is unlikely that SNPs will replace STR loci as the predominate genetic markers for human forensic identification (Gill *et al.,* 2004). Nevertheless, there is a wealth of genetic information that can be tapped, since approximately 85% of human variation is derived from SNPs (Gill, *et al.,* 2004; Cooper *et al*., 1985; Wang *et al.,* 1998). SNPs, however, offer advantages for forensic analyses in some situations, such as use in mass disasters and missing person's cases where the DNA may be substantially fragmented. If the resultant degraded fragments are smaller than the required length for STR typing, then no result will be obtained. It is possible to design the PCR for SNP amplification to reduce the required fragment length to only 60-80 bp (24, 26, and 20). Therefore, in highly degraded biological materials, SNP testing should provide more results than STR typing. Another advantage is that substantial research and development currently is underway to improve analytical capabilities, possibly making large multiplex assays and complete automation feasible. With automation, rapid typing of many samples with consistent quality control can be achieved; SNP-based technology may enhance current capabilities in this regard. Lastly, SNPs have relatively low mutation rates and thus are stable genetic markers for lineage-based analyses, such as inheritance cases, missing person cases, and situations where no direct reference sample may be available (Budowle and Daal, 2008).

SNPs do have a few limitations. Since most SNPs are biallelic, they are less informative for identity testing than STR analyses. Thus, more SNPs will be needed to achieve the same level of discrimination afforded with the core 13 STR loci used in the United States. To obtain the same power of discrimination that existing STR multiplex systems provide a panel of at least 50-100 SNP autosomal loci would be required (Chakraborty *et al.,* 1999). Creating such large multiplexes that are validated for forensic analysis is a challenging task. To type such a large battery of SNPs may require far more sample DNA than is needed for several multiplex STR systems. Also SNPs, because of the limited number of alleles per locus, will be more problematic for interpretation in situations involving mixtures, that is, multiple donor samples. Deconvolution of a mixed sample into its individual contributors may be problematic

for many scenarios using SNPs. Lastly, the low mutation rate, will contribute to population substructure effects requiring more judicious selection of SNPs for identity testing purposes. SNPs for forensic applications should not be discussed generically. Specialized applications are feasible, such as mtDNA and Y chromosome lineage analyses, characterization of highly degraded DNA samples, migration and lineage biographical ancestry analysis, high-throughput rapid screening of DNA samples prior to conventional STR analysis, and potential determination of physical traits. The SNPs for forensic analyses can be divided into four categories (Budowle and Daal, 2008):

1. Identity-testing SNPs—for individualization; requiring high heterozygosity and a low coefficient of inbreeding (Fst) (i.e., low population heterogeneity) (Budowle and Daal, 2008).
2. Lineage informative SNPs—sets of tightly linked SNPs that function as haplotype markers to identify missing persons through kinship analyses (Budowle and Daal, 2008).
3. Ancestry informative SNPs—for establishing high probability of an individual's biogeographical ancestry to indirectly infer some phenotypic characteristics for investigative lead value; requiring low heterozygosity and high Fst values (Budowle and Daal, 2008).
4. Phenotype informative SNPs—for establishing high probability that an individual has a particular phenotypic characteristic, such as skin color, hair color, or eye color for investigative lead value (Budowle and Daal, 2008).

SNPs are unlikely to replace the battery of STR loci in the foreseeable future. However, SNP markers will serve an important role in analyzing challenging forensic samples, such as those that are much degraded, for augmenting the power of kinship analyses and family reconstructions for missing persons and unidentified human remains, as well as for providing investigative lead value in some cases without a suspect (and no STR match in CODIS). There are four different classes of SNPs that apply to forensic analyses: identity-testing SNPs, lineage informative SNPs, ancestry informative SNPs, and phenotypic SNPs. The identity-testing SNPs are likely the first group to be implemented for routine work, with lineage based systems to follow. The ancestry SNPs (which already have been used in some cases) and phenotypic SNPs will be used in more specialized applications that may provide useful investigative leads. While more accurate than AIMs for predicting phenotype, phenotypic SNPs require considerably more research and development efforts and to date have been most applicable to hair and eye color. We anticipate that more efforts will be dedicated to the development of SNPs to enhance the forensic science toolbox for genetically characterizing biological evidence with the most challenging of forensic samples while generating valid, objective, and reliable results in a timely fashion (Budowle and Daal, 2008).

Tandems Repeat

Two important classifications of tandem repeat have been used commonly in forensic genetics are: Minisatellites, also referred to as Variable Number Tandem Repeats (VNTRs); and microsatellites, also referred to as Short Tandem Repeats (STRs). The general structure of VNTRs and STRs is the same. Variation between

different alleles is caused by a difference in the number of repeat units that results in alleles that are of different lengths and hence tandem repeat polymorphisms are known as length polymorphisms [http://www.councilforresponsible genetics.org/geneticprivacy/DNA_forensics_ 1.html]

Microsatellite or Short Tandem Repeats (STRs)

Short Tandem Repeats (STRs), or Microsatellites or Simple Sequence Repeats (SSRs), are accordion-like stretches of DNA comprising core repeat units of between 2 and 7 nucleotides in length (Butler, 2007) that are tandemly repeated approximately 17 times. STRs are found on 22 autosomal chromosomes as well as both X and Y sex chromosomes, although those on Y chromosome differ less due to lack of recombination. The human genome contain thousands of STR markers, but only a small core set of loci have been certain for use in forensic DNA and human identity testing (Butler, 2006). STRs express a high degree of polymorphism, making them of certain use to the forensic scientist. The Variability in STRs is caused by the inaccuracy of DNA polymerase in copying the region. As STR region are non-coding, there is no selective pressure against the high mutation rate, resulting in high variation between different people [http://forensicsciencecentral.co.uk/ dna.shtml].

Millions of STR profiles are generated worldwide each year by government, university, and private laboratories performing various forms of human identity testing, including DNA data basing, forensic casework, missing persons/ mass disaster victim identification, or parentage testing. With STR typing, PCR is used to recover information from small amounts of available biological material. The relatively short PCR product sizes of approximately 100-500 bp generated with STR testing are generally compatible with degraded DNA that may be present due to environmental insults on the evidentiary biological material found at a crime scene. PCR amplification of multiple STR loci simultaneously, or multiplexing, is possible with different colored fluorescent dyes and different sized PCR products. Use of multiple loci enables a high power of discrimination in a single test without consuming much DNA (e.g., 1 ng or less of starting material). It is worth noting that these core STR loci occur in between genes in which a high degree of variability is tolerated and are thus not directly responsible for physical traits such as hair color or eye color or genetic diseases (Butler, 2007). Short Tandem Repeat (STR) loci are widely considered to be effective for variety of applications including forensic applications, phylogenetic reconstruction and chimerism based post Haematopoietic Stem Cell Transplantation (HSCT) graft monitoring. For each application, specific sets of STR loci are used (Agrawal *et al.,* 2004).

DNA Separations

The overall length of the STR amplicon is measured to determine the number of repeats present in each allele found in the DNA profile, following PCR amplification. This length measurement is made via a sized-based separation involving gel or Capillary Electrophoresis (CE). Each STR amplicon has been fluorescently labeled during PCR, since either the forward or reverse locus-specific primer contains a fluorescent dye. Thus, by recording the dye color and migration time of each DNA fragment relative to an internal size standard, the

size for each STR allele may be determined following its separation from other STR alleles (Butler, 2007). Commonly used instruments for STR allele separation and sizing include the ABI PRISM 310 and ABI PRISM 3100 genetic analyzers (Applied Biosystems) (Butler *et al.*, 2004; Butler, 2007).

There are number of biological and instrumental objects that must be sorted completely in order to generate a complete and accurate STR profile. Biological artifacts include stutter products, split peaks from incomplete adenylation, triallelic patterns, and variant alleles containing mutations in the repeat or flanking regions that cause an allele to be off-ladder. Instrumental artifacts arise from voltage spikes, dye blobs, and bleed-through between dye colors (Butler, 2007).

While multicolor fluorescence detection CE instrumentation, such as the ABI PRISM 3100 genetic analyzer, presently dominate the field, efforts are ongoing to develop microchip CE platforms (Yeung *et al.,* 2006) to perform high-resolution DNA separations with eventual integration of the PCR amplification and CE separations (Liu *et al*., 2007). In addition, Mass Spectrometry (MS) with Matrix-Assisted Laser Desorption/ionization (MALDI) and Electrospray Ionization (ESI) techniques have been used for STR typing without allelic ladders (Butler *et al*., 1998; Oberacher, *et al.,* 2008; Butler, 2007).

There are 13 STR loci that are typically used to create a genetic fingerprint of an individual in forensic laboratories. The profile is kept in DNA databases. In the United Kingdom (UK), the profile is kept in National DNA Database and in the United Stated, the genetic fingerprint are kept in various Combined DNA Index System (CODIS) database ranging from the smaller, local levels to the national level.

Benefits of STR Analysis

STR analysis is still somewhat of a new technology in forensic science. In fact, it was only towards the end of the 1980s that the technique started to become more widely used. Currently, it is a popular method for genetic profiling and STR analysis used today involves four to five nucleotide repeats, which allows scientists to obtain relatively precise and accurate information. In addition, this type of approach allows for somewhat 'mistake-proof' information to be gathered even in poor conditions. If a repeat sequence is shorter, there are problems that occur. A number of genetic conditions are linked to three nucleotide repeats. For example, Huntington's disease is associated with a shorter repeat sequence.

Another benefit of this type of analysis is that it can be successfully used to monitor patients following transplant therapy. It can allow for the prediction of problems associated with transplants, such as graft rejection or a relapse of the disease. This proactive type of testing can significantly aid the patient in obtaining recovery from their condition and transplant surgery.

While there are many techniques to choose from with regards to DNA analysis, STR analysis has many benefits to offer for several different applications. Perhaps most pressing are the concerns regarding the storage of information obtained by STR analysis, which entails databases at local and national levels. Clearly, the ethical aspects of DNA analysis need to be addressed and resolved as techniques such as STR analysis continue to be refined or replaced with more effective and efficient ways of analyzing

DNA [http://www.exploredna.co.uk/str-analysis.html].

Minisatellites or Variable Number of Tandem Repeats (VNTRs)

Variable number tandem repeats (VNTR; Minisatellites DNA): A family of genetic loci found in eukaryotes consisting of short (15-100 bp) sequences of DNA repeated in tandem arrays; in humans these arrays are typically 1-5 kb long. The alleles for any particular locus all have the same sequence but differ as to how many times the sequence is repeated. VNTR loci contribute to repetitive DNA and have proved valuable in DNA fingerprinting, especially in human forensic science. The VNTR sequences can be released intact from a DNA sample using restriction endonuclease enzymes and identified using gene probes with Southern blotting. Alternatively, the VNTR alleles can be amplified by the polymerase chain reaction, separated by gel electrophoresis, and the resultant patterns compared without the need for special gene probes. Since each VNTR locus typically has many different alleles, the likelihood of two individuals having identical sets of alleles for even a few such loci is very remote. Hence, a DNA sample can be ascribed to a particular person with a high degree of certainty [http://www.encyclopedia.com]. VNTRs are not generally used in forensics, as this technique is often a more expensive process and take longer due to VNTRs having a greater than, compared to, STRs.

Mechanism by which Variation in Number of Tandem Repeats is Generated

The most likely way variation in tandem repeat regions is generated is so called replication slippage. During replication the DNA polymerase copies the template and sometimes stutters in areas where tandem repeats are located. As a result the number of repeats is increasing. This is called backward replication slippage. Particularly if a tandem repeat region contains many repeats a mechanism called forward replication slippage may occur. In that case during replication when the DNA is single stranded the repeat region often is forming loops. The DNA polymerase may accidentally skip this looped region and as a result the replicated stand contains a decreased number of tandem repeats.

Multiple-Locus VNTR Analysis

Multiple-Locus Variable number tandem repeat Analysis (MLVA) is a method used to perform molecular typing of particular microorganisms. It utilizes the naturally occurring variation in the number of tandem repeated DNA sequences found in many different loci in the genome of a variety of organisms. The method originates from forensic science where it is used for DNA fingerprinting in samples from human origin. MLVA is also widely used to assess the molecular fingerprint of micro-organisms such as bacteria. The molecular typing profiles are used to study transmission routes, to assess sources of infection and also to assess the impact of human intervention such as vaccination and use of antibiotics on the composition of bacterial populations [http://www.mlva.net]. The use of VNTR loci originates from forensic science where it utilized for DNA fingerprinting in samples from human origin, e.g., to characterize DNA traces on crimes scenes, but also for paternity testing. VNTRs are also widely used to assess the molecular fingerprint of micro-organisms such as bacteria.

The MLVA Method PCR of VNTR Loci

In MLVA the variation in the number of repeats in a set of VNTR loci is assessed. This is achieved

by performing PCR of the VNTR loci and subsequent sizing of the PCR products on agarose gels and capillary systems or on automated DNA sequencers.

Separation of PCR Products and Sizing

VNTR PCR products can be separated on agarose gels which will reveal the size variation of the PCR products. If a DNA sequencer is utilized for separation, fluorescently labeled primers may be used in a multiplex PCR that amplifies at least 4 different VNTR loci. An internal marker which is present in each sample and the high resolution separation of the sequencer enables accurate and high throughput analyses.

Assessing the Number of Tandem Repeats

The assessed PCR product size is used to calculate the number of repeat units in each locus. The size of the PCR product is assessed. The flanking regions with known sizes are subtracted from the PCR product size which results in the net size of the repeat region. This size is divided by the repeat unit size and this will reveal the number of repeat. This example shows there is a small inaccuracy in sizing but mostly do not prevent the assessment of the number of repeat.

Advantages of MLVA

- MLVA can connect suspected, fast-evolving bacterial strains to an outbreak when they might look the same using other methods of DNA fingerprinting, such as PFGE.
- PulseNet USA uses MLVA as a complementary technique to PFGE, allowing scientists to see more detailed differences between bacteria that have similar PFGE patterns.

Limitations of MLVA

- Requires a trained and skilled technician
- Few standardized protocols available

Restriction Fragment Length Polymorphism (RFLP)

The first technique used for DNA analysis in forensic science and several other fields is RFLP but it is not widely used now because it is slow and cumbersome. RFLP refers to a variation in DNA sequence that is detected through gel electrophoresis.

Performing RFLP Analysis

In RFLP, the different lengths of DNA fragments are analyzed. These fragments are from the digestion of a sample of DNA with a restriction endonuclease enzyme. The enzyme chops DNA in a certain way called a restriction endonuclease recognition site. Whether or not particular recognition sites are present will provide different lengths of DNA fragments, which are then divided up through electrophoresis. DNA probes then serve to hybridize the fragments through complementary binding (Murnaghan, 2014).

Benefits of RFLP Analysis

There are numerous benefits of RFLP analysis, despite the fact that it is not broadly used now. It plays main role by letting scientists to map the human genome, provide information on genetic diseases, and to find where a specific gene lies on a chromosome. This is accomplished by looking at the DNA from a set of family members who suffer from the disease and then searching for RFLP alleles that share the same type of inheritance pattern for the condition. By using RFLP analysis, scientists are then able to determine others who might be at risk for the disease or a carrier of the mutated gene (Murnaghan, 2014).

RFLP was also one of the first methods used for genetic typing - also known as genetic fingerprinting, profiling or testing, in addition to its benefits for genetic disease testing. A genetic profile of a person, which could then be used to compare to samples at a crime scene or to find out a person's paternity, was provided by using this application. Not only that, but RFLP analysis also had roles in aiding our understanding of the genetic aspect of different breeding patterns in various animals (Murnaghan, 2014).

Challenges of RFLP Analysis

Despite its many benefits and useful earlier applications, RFLP analysis is still a slow and more tedious process compared to some of the newer DNA analysis techniques (Murnaghan, 2014). Also, it requires significantly larger sample sizes than any other forms of analysis. In RFLP, the sample would usually need to be approximately the size of a one-pound coin. While that may sound small, it is large relative to other techniques such as PCR analysis that require only a few cells for successful sequencing (Murnaghan, 2014). The period of the process itself is a lengthy one; as well it may take up to a month to complete.

Many of the previous applications of RFLP have been replaced by Polymerase Chain Reaction (PCR). Currently, RFLP has variations such as Terminal Restriction Fragment Length 5 Polymorphism (TRFLP), which inclines to have applications associated to the characterization of bacteria and related communities. It is somewhat a blend of PCR and RFLP analysis. The technique uses PCR amplification of the DNA with primers that have a fluorescent label. After the products of PCR are digested from RFLP enzymes, the patterns that occur can be seen with a DNA sequence, which are then analyzed (Murnaghan, 2014).

RFLP analysis is still been important in establishing our understanding of DNA analysis, while also spurring the development of new, more efficient techniques, although it is no longer used widely. It is likely that our current techniques will follow suit in the future by their refinement and replacement with even more advanced types of DNA analysis (Murnaghan, 2014).

Polymerase Chain Reaction (PCR)

The PCR is a technique developed by Kary Mullis in 1983, that is used to amplify single or few copies of pieces of DNA across several order of magnitude, generating thousands to millions of copies of a particular piece of DNA. PCR is now more commonly used in medical and research laboratories for various applications. These includes DNA cloning for sequencing, DNA based phylogeny, functional analysis of gene, diagnosis of hereditary diseases, identification of genetic fingerprint for forensic and paternity testing, and detection and diagnosis of infectious diseases. PCR is efficient and cost effective way to amplify small pieces of DNA or RNA.

A basic set up of PCR requires several components and reagents. DNA template that contain DNA target to amplify, Two sets of primers, DNA polymerase, Deoxynucleoside triphosphates (DNTPs), buffer solution, monovalent and bivalent cations. The PCR is commonly carried out in a volume of 10-200 µL in a reaction tube (0.2-0.5 mL) in a thermal cycler. Before initiation of PCR, DNA is extracted from a sample through a multi-step procedure. Following sample preparation, the three step PCR process is initiated.

Denaturation – Separating the target DNA: In this step, the reaction is heated at 94-98°C for 20-30 s. It causes melting of the DNA template by disrupting the Hydrogen bond between complementary bases, yielding single stranded DNA molecules.

Annealing – *Binding primers to the DNA Sequence:* In this step, the primers anneal to the single-stranded DNA template. The reaction temperature is lowered to 50-65°C for 20-40 s. This low temperature allows the specific hybridization of the primers to the strand. When the primer sequence closely matches with template sequence, stable DNA-DNA hydrogen bonds are formed. The DNA polymerase binds to the primer-template hybrid and begins DNA formation.

Extension – Elongating to create copies: In this step, the temperature is dependent on the DNA polymerase used. A new DNA strand complementary to the DNA template strand is synthesized by adding DNTPs in 5' to 3' direction by DNA polymerase. The extension time is also dependent on type of DNA polymerase used and on the length of the DNA fragment to amplify. After completion of extension, two identical copies of the original DNA have been made.

After making two copies of DNA through PCR, the cycle begins again; this time uses the duplicated DNA. Each cycle creates the double of last cycle. After approximately 30 or 40 PCR cycles, more than one billion copies of the original DNA segments have been made. The PCR process is automated and it can be completed in just few hours.

Benefits and Drawbacks of PCR

PCR have numerous benefits. It is a technique that is simple to understand and use, and it generate results rapidly. PCR reaction is highly sensitive as it can make million copies of specific DNA sequences, but the DNA polymerase used can lead mutation and it is also prone to errors in the fragment generated. PCR is sensitive technique that can generate specific fragments amplifications. This specificity can be altered by nonspecific binding of the primers to other similar sequences in the DNA template. Also to generate PCR product, some prior sequences information are usually required to design primers.

CONCLUSION

The rapid development in Bio-technologies provides the perfect tool for identification of victims using various appropriate genetic markers. The advantage of one technique over the other in terms of identification is evident. The technologies with complicated methodologies are still underdevelopment, serves the particular required role, and they are becoming critical in forensic analysis. There are some major technical problems that should be over-come, in order to use the techniques to full of its potentials. Since there are different genetic markers that serves different role, and each technique is optimized to obtain best results for the test being conducted.

Standard STR typing provides sufficient discrimination power for most applications, and most countries have already established large-scale forensic DNA databases for resolving crimes based on STR technology in forensic science. In applied forensic science, DNA samples are usually limited and often cannot fulfill the requirements of simultaneous analysis. This resulted in difficulties in providing sufficient information and can limit their use as legal evidence. NGS technology not only meets these requirements but can also potentially be applied

in many areas of research, including DNA database construction, ancestry and phenotypic inference, body fluid and species identification, and forensic microbiological analysis.

Although NGS technology seems to have an significant role in future forensic studies, more work is required to fully achieve this goal, which includes overcoming problems with low-template library preparation, error rate, type estimations and issues with NGS data processing and mining. Strategies for the application of NGS in forensic science also need to be generated. With the technical advances of NGS technology and continuous translational efforts of forensic scientists, we believe that NGS technology is likely to become an easily accessible routine method in forensic practice (Yang *et al.,* 2014).

REFERENCES

1. Agrawal S, Khan F, Talwa S and Nityanand S (2004), "Short tandem repeat technology has diverse applications: individual identification, phylogenetic reconstruction and chimerism based post haematopoietic stem cell transplantation graft monitoring", *Indian journal of medical sciences, Vol. 58*, No. 7.
2. Berglund *et al.* (2011), "Next-generation sequencing technologies and applications for human genetic history and forensics", *Investigative Genetics*, Vol. 2, p. 23.
3. Bosch J R and Grody W W (2008), "Keeping up with the next generation: massively parallel sequencing in clinical diagnostics", *J Mol Diagn,* Vol. 10, pp. 484-492.
4. Branton D, Deamer D W, Marziali A *et al*. (2010), "The potential and challenges of nanopore sequencing," *Nature Biotechnology,* Vol. 26, No. 10, pp. 1146–1153, 2008, pp. 281–294, 2010.
5. Braslavsky I, Hebert B, Kartalov E and Quake S R (2003), "Sequence information can be obtained from single DNA molecules", *Proc Natl Acad Sci USA,* Vol. 100, pp. 3960–3964.
6. Budowle B and Daal A (2008), "Forensically relevant SNP classes", *Biotechniques,* Vol. 44, pp. 603-610.
7. Budowle B (2004), "SNP typing strategies", *Forensic Sci. Int.,* Vol. 146, pp. S139-S142.
8. Budowle B, Planz J, Campbell R and Eisenberg A (2004), "SNPs and microarray technology in forensic genetics: development and application to mitochondrial DNA", *Forens. Sci. Rev.,* Vol. 16, pp. 22-36.
9. Butler J (2007), "Short Tandem Repeat Typing Technologies Used In Human Identity Testing", *Biotechniques,* Sii-Sv.
10. Butler J (2006), "Genetics And Genomics Of Core Short Tandem Repeat Loci Used In Human Identity Testing", *Journal of Forensic Sciences,* pp. 253-265.
11. Butler J M, Buel E, Crivellente F and McCord B R (2004), "Forensic DNA typing by capillary electrophoresis: using the ABI Prism 310 and 3100 Genetic Analyzers for STR analysis", *Electrophoresis,* Vol. 25, pp. 1397-1412.
12. Butler J M, Li J, Shaler T A, Monforte J A and Becker C H (1998), "Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry.], *Int. J. Legal Med.,* Vol. 112, pp. 45-49.

13. Chakraborty R, Stivers D N, Su B, Zhong Y and Budowle B (1999), "The utility of STR loci beyond human identification: implications for the development of new DNA typing systems", *Electrophoresis,* Vol. 20, pp. 1682-1696.

14. Clarke J, Wu H C, Jayasinghe L, Patel A, Reid S and Bayley H (2009), "Continuous base identification for single-molecule nanopore DNA sequencing," *Nature Nanotechnology,* Vol. 4, No. 4, pp. 265–270.

15. Cooper D N, Smith B A, Cooke H J, Niemann S, and Schmidtke J (1985), "An estimate of unique DNA sequence heterozygosity in the human genome", *Hum. Genet.,* Vol. 69, pp. 201-205.

16. Divne A M and Allen M (2005), "A DNA microarray system for forensic SNP analysis", *Forens. Sci Int*., Vol. 154, pp. 111-121.

17. Gill P, Werrett D J, Budowle B and Guerrieri R (2004), "An assessment of whether SNPs will replace STRs in national DNA databases#joint consider- joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM)", *Sci. Justice,* Vol. 44, pp. 51-53.

18. Grada A and Weinbrecht K (2013), "Next-Generation Sequencing: Methodology and Application", *Journal of Investigative Dermatology,* Vol. 133, No. 11.

19. Harismendy O, Ng P C, Strausberg R L, Wang X, Stockwell T B, Beeson K Y, Schork NJ, Murray S S, Topol E J, Levy S and Frazer K A (2009), "Evaluation of next generation sequencing platforms for population targeted sequencing studies", *Genome Biol.,* Vol. 10, p. R32.

20. Ku C and Roukos D (2013), "From next-generation sequencing to nanopore sequencing technology: Paving the way to personalized genomic medicine", *Expert Review of Medical Devices,* pp. 1-6.

21. Li C, Qi B, Ji A, Xu X and Hu L (2009a), "The combination of single cell micromanipulation with LV-PCR system and its application in forensic science", *Forensic Sci Int Genet Suppl Ser,* Vol. 2, pp. 516–517.

22. Liu L and Li Y (2012), "Comparison of Next-Generation Sequencing System", *Journal of Biomedicine and Biotechnology,* pp. 11-11.

23. Liu P, Seo T S, Beyor N, Shin K J, Scherer J R and Mathies R A (2007), "Integrated portable polymerase chain reaction-capillary electrophoresis microsystem for rapid forensic short tandem repeat typing", *Anal. Chem.,* Vol. 79, pp. 1881-1889.

24. Murnaghan I (2014, November 1), RFLP Analysis. Retrieved April 20, 2015, from http://www.exploredna.co.uk/rflp-analysis.html

25. "Mitochondrial DNA Typing" World of Forensic Science. 2005.*Encyclopedia.com.* 31 Mar. 2015 <http://www.encyclopedia.com>.

26. Oberacher H, Pitterl F, Huber G, Niederstatter H, Steinlechner M and Parson W (2008), "Increased forensic efficiency of DNA fingerprints through simultaneous resolution of length and nucleotide variability by high-performance mass spectrometry", *Hum. Mutation.*

27. Pareek C S, Smoczynski R and Tretyn A (2011), "Sequencing technologies and genome sequencing", *J Appl Genet* (in press). doi:10.1007/s13353-011-0057-x

28. Richardson P (2010), "Special Issue: Next Generation DNA Sequencing", *Genes,* pp. 385-387.

29. Sachidanandam R, Weissman D, Schmidt S, Kako J, Stein L, Marth G and Altshuler D (2000), " A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms", *Nature,* Vol. 409, pp. 928-933. http://www.nature.com/nature/journal/v409/n6822/full/409928a0.html

30. Shendure J and Ji H (2008), "Next-generation DNA sequencing", *Nature Biotechnology,* Vol. 26, No. 10, pp. 1135-1145.

31. Shendure J, Porreca G J, Reppas N B, Lin X, McCutcheon J P *et al.* (2005), "Accurate multiplex polony sequencing of an evolved bacterial genome", *Science,* Vol. 309, pp. 1728–1732.

32. Srivatsan A, Han Y, Peng J, Tehranchi A K, Gibbs R, Wang J D and Chen R (2008), "High-precision, wholegenome sequencing of laboratory strains facilitates genetic studies", *PLoS Genet,* 4, e1000139.

33. Tucker T, Marra M, Friedman J M (2009), "Massively parallel sequencing: the next big thing in genetic medicine", *Am J Hum Genet.,* Vol. 85, pp. 142–154.

34. Thomas R K, Nickerson E, Simons J F, Janne P A, Tengs T, Yuza Y, Garraway L A, LaFramboise T, Lee J C, Shah K *et al*. (2006), "Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing", *Nat. Med*., Vol. 12, pp. 852–855.

35. Venkatesan B M and Bashir R (2011), "Nanopore sensors for nucleic acid analysis", *Nat. Nanotechnol.,* Vol. 6, pp. 615–624 (2011)."variable number tandem repeats." A Dictionary of Biology. 2004.*Encyclopedia.com.* 1 Apr. 2015 <http://www.encyclopedia.com>.

36. Timp W, Mirsaidov U M, Wang D, Comer J, Aksimentiev A and Timp G (2010), "Nanopore sequencing: electrical measurements of the code of life," *IEEE Transactions on Nanotechnology*, Vol. 9, No. 3.

37. Wang D G., Fan J B, Siao C J, Berno A, Young P, Sapolshy R, Chandona G, Perkins N, *et al.* (1998), "Large scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome", *Science,* Vol. 280, pp. 1077-1082.

38. Yang Y, Xie B and Yan J (2014), "Application of Next-generation Sequencing Technology in Forensic Science", *Genomics Proteomics Bioinformatics,* Vol. 12, pp. 190-197.

39. Yeung S H, Greenspoon S A, McGuckian A, Crouse C A, Emrich C A, BanJ, and Mathies R A (2006), "Rapid and high-throughput forensic short tandem repeat typing using a 96-lane microfabricated capillary array electrophoresis microdevice}, *J. Forensic Sci*., Vol. 51, pp. 740-747.

40. Ziêtkiewicz E, Witt M, Daca P, ¯ebracka-Gala J, Goniewicz M, Jarz¹b B and Witt M (2011), "Current genetic methodologies in the identification of disaster victims and in forensic analysis", *Journal of Applied Genetics,* pp. 41-60.

International Journal of Engineering Research and Science & Technology
Hyderabad, INDIA. Ph: +91-09441351700, 09059645577
E-mail: editorijlerst@gmail.com or editor@ijerst.com
Website: www.ijerst.com