Hirak Ranjan Dash · Pankaj Shrivastava
Braja Kishore Mohapatra · Surajit Das
Editors

DNA Fingerprinting: Advancements and Future Endeavors

Springer

Hirak Ranjan Dash • Pankaj Shrivastava
Braja Kishore Mohapatra • Surajit Das
Editors

DNA Fingerprinting: Advancements and Future Endeavors

Editors
Hirak Ranjan Dash
DNA Fingerprinting Unit
State Forensic Science Laboratory
Sagar, Madhya Pradesh, India

Pankaj Shrivastava
DNA Fingerprinting Unit
State Forensic Science Laboratory
Sagar, Madhya Pradesh, India

Braja Kishore Mohapatra
Department of Biology and DNA
Fingerprinting Unit
Central Forensic Science Laboratory
New Delhi, Delhi, India

Surajit Das
Department of Life Science, National
Institute of Technology
Rourkela, Odisha, India

ISBN 978-981-13-1582-4 ISBN 978-981-13-1583-1 (eBook)
https://doi.org/10.1007/978-981-13-1583-1

Library of Congress Control Number: 2018959136

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

Chapter 15
Future of DNA Fingerprinting: Application of NGS in Forensic Science

Jahangir Imam, Pankaj Shrivastava, Shivani Dixit, and Amita Shrivastava

Abstract In the relatively short time frame since 2005, NGS has fundamentally altered genomics research and allowed investigators to conduct experiments that were previously not technically feasible or affordable. The various technologies that constitute this new paradigm continue to evolve, and further improvements in technology robustness and process streamlining will pave the path for translation into clinical diagnostics. NGS is no doubt one of most important and noteworthy technological advances in the biological sciences in the last two decades. NGS has also made its mark in the application in forensic sciences. It has overcome the limitations of capillary electrophoresis and also have the potential to provide multi-information like sequence variation detections, differentiating monozygotic twins, STR typing of degraded samples, etc. The best part of NGS is that we can parallel do the typing of CODIS STRs loci and sequencing study to detect the allelic variations simultaneously. Currently many NGS kits are being developed and available which have huge application in forensic field. This chapter reviews the discovery, advancement, applications, and development of new NGS-based forensic kits and highlighted the applications of NGS in the field of forensic science and criminal justice system.

Keywords NGS · CODIS · Genomics · Forensic science · Multi-informative · Monozygotic twins

J. Imam
DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home, Jail and Disaster Management, Government of Jharkhand, Ranchi, India

P. Shrivastava (✉) · S. Dixit
DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

A. Shrivastava
Department of P. G. Studies and Research in Biological Sciences, Rani Durgavati University, Jabalpur, Madhya Pradesh, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_15

15.1 Introduction to Forensic DNA Typing: Discovery

Over the past more than three decades, a huge amount of studies has helped to improve our understanding of unique differences among individuals and also to decipher human origin. The analysis of genetic differences among humans has long been used in courts of law to prove identity with certainty. Forensic DNA typing uses DNA analysis and comparison to resolve legal issues, such as paternity tests and inheritance matters, establish identity in criminal cases where biological evidence is found at crime scenes, and identify victims of mass disasters and missing persons from human remains. Alec Jeffreys at the University of Leicester, in the UK, discovered variable and heritable patterns from repetitive DNA using multi-locus probes, and he named the technique "DNA fingerprinting" [33, 34]. Alec Jeffreys was the first to identify common polymorphisms with the help of RFLP method, and this led to the discovery of different formats of DNA polymorphism study; which has the potential to identify a particular person (except for identical twins). This invention led the stone of a new area of science and further developed in due course of time as the most believed technology in criminal investigations. The technology earlier discovered by Jeffreys is obsolete now. It has gone to a variety of changes to make the technology simpler, user-friendly, and above all more convincing. Besides proving its utility in criminal investigation, this has also proved as multifaceted technology and is now being used in many biological disciplines, namely, in diversity, population genetics, conservation studies, and clinical and anthropological studies. Besides this DNA technology has become the technology which has crossed the boundary to academics, and this has become a socially acceptable technology as well. Forensic DNA typing is based on comparison of the nuclear DNA in a person with that identified in biological material found with the DNA of other person or at the scene of occurrence for the purpose of conclusive inclusion or exclusion. The first case which was solved in March 1985, using Jeffreys techniques, was not a forensic case, but it was a case of immigration [33, 34]. This first application of DNA technology not only saved a young boy from deportation but also established the technology in the public eye as the technology which can save the innocent. The technique identified individuals and helped in individualization of biological evidences, and in its first application, it was based on restriction enzyme digestion, followed by Southern blot analysis. Southern blot technique was laborious and requires good-quality DNA for further analysis. The original technology developed by Jeffreys is now obsolete for forensic use and underwent a huge and continual transformation in the basic technology (Fig. 15.1).

15.2 Capillary Electrophoresis-Based DNA Technology: The Present Technology

Capillary electrophoresis (CE) is one of most important advancements as a part of instrumentation in the field of forensic DNA typing. After PCR invention, scientists consider it as the second most needed development. Already presence of DNA

Fig. 15.1 Technological developmental transformation in forensic DNA typing technique since its inception

repeats and its application were known, and amalgam of capillary electrophoresis in forensic applications not only makes the work easier but also more accurate and authentic which is of paramount importance in forensic DNA analysis [59]. The application of capillary electrophoresis is not only limited for biological samples but has huge importance in the analysis of gunshot residues, explosive residues, and drugs. For forensic DNA analysis, STR profiling (highly polymorphic markers) which is based on fragment analysis is of great value for human identification (HID) due to the single-base resolution capability of CE [30, 31, 59]. Introduction of capillary electrophoresis in STR typing circumvents the tedious and expensive approach of DNA sequencing for STR typing. The approach of CE like precise sizing, its sensitivity for the detection of fluorescence emitted by different dyes, automatic electrophoresis, and data collection software are key factors in the worldwide adoption of CE as the preferred platform for forensic DNA analysis. The most common CE systems used in forensic DNA analysis include the ABI PRISM® 310, 3100, 3100 Avant and 3130, 3130xl, 3500, and 3500xL Genetic Analyzers (GAs). The advanced CE automated machines are developed with advanced features which is useful for forensic scientists [35, 36]. It has many advantages like normalization of peak height, accurate sizing of fragments, sample injection, single-base resolution, high run to run precision, good temperature control and automation, better sensitivity, high throughput, user-friendly, and easy software features to analyze the raw data to the level of precise accuracy [59]. Definitely the incorporation of CE in forensic application must be considered as a milestone for the mankind service.

15.3 NGS: The Upcoming Technology in Forensics

15.3.1 What Is NGS Technology?

Next-generation sequencing (NGS) technology which has overcome the limitations of conventional Sanger sequencing technology has grown rapidly in recent years in the field of genomics research because of its high-throughput capacity and low cost and ancient DNA analysis. So far in the advancement in technology, NGS is no doubt one of most important and noteworthy technological advances in the biological sciences in the last two decades. The evolution of sequencing technology from first to third generation is depicted as shown in Fig. 15.2 [45, 55]. NGS is a highly

Fig. 15.2 Development in NGS technology: from inception to present

sensitive, high-throughput, low-cost, and much more faster technological advancement which makes it the choice for many biological research along with forensic as compared to other sequencing methods. NGS is now progressing toward maturity and expected to replace traditional sequencing method for DNA profiling which is considered gold standard till date.

15.3.2 Multi-informative Potential of NGS Technology

In recent years, as the technology is advancing, NGS is also not lagging behind with the introduction of STR analysis and parallel sequencing with MPS (massively parallel sequencing). This has proved exceptionally advantageous as the time, cost, speed, and improvement in sequence length and accuracy is dramatically improved [5, 17, 21, 22, 24, 50, 53, 57]. Several NGS platforms have become common in forensic research and commercial applications. The technology promises continued improvement and may become the next "gold standard" for forensic genetics [14]. The Illumina MiSeq FGx Forensic Genomics System became available in early 2015 and has been validated by the manufacturer. NGS analysis not only saves a significant amount of time and but can also provide a number of genetic information in a single run (Fig. 15.3). Furthermore, NGS-based methods provide the full sequence data along with the length-based genotypes [25, 61]. The availability of full sequence information from NGS makes it possible to investigate the mutation and/or the true variation (at the nucleotide level) within STR loci and identify previously unknown alleles and mutational events during paternity establishment [7].

15.3.3 NGS Application in STR Sequence Variation Detection

In today's forensic work, STR analysis has been mainly performed by size-based DNA separation using capillary electrophoresis (CE) [13]. This is mostly serving the purpose of forensic scientists in solving the various types of forensic cases. But as far as internal sequence variation is concerned in STR alleles, capillary

Fig. 15.3 Potential fields of application of next-generation sequencing technology in forensic science

electrophoresis fails which is only based on detection of the length of PCR product. This sequence variation in STR alleles is now considered very important in various cases where very low quantity of DNA is present or for mixed samples which makes the interpretation difficult. Few important techniques like mass spectroscopy and next-generation sequencing (NGS) have been now utilized in forensic also to study and identify the potential internal sequence variations in STR alleles [51–53]. STR sequence variation analysis is nowadays becoming more important as this will improve the discrimination between two individuals and also in mixed DNA samples. This also aids in the study of mutation rates in STR alleles and increases our knowledge about the STR mutability rate for that particular allele. NGS is the solution for the detection and exact identification of variations in STR alleles.

There are a lot of variations in STR repeats which always put the forensic scientists to think for better and correct interpretations. These STR loci are characterized on the basis of their STR repeat category like simple repeat, variant allele, compound repeat, and complex repeat [8]. The CE-based STR assay may identify the alleles according to their relative size compared to an allelic ladder containing sequenced alleles; even though internal sequence variation may be present [25]. Another demerit of CE-based STR analysis is allele peak variations with respect to allelic ladder sizing bean if there are insertions or deletions (indels) in the flanking regions of that allele which is not common but potentially present. The PCR-CE-based assay can be done in 1 day for a specific marker. But the big advantage of NGS is its combined STR analysis and NGS assay even if it takes relatively little longer time [8]. Another important advantage of NGS is its capability to analyze

degraded DNA samples which is difficult with CE-based analysis. Among the various NGS methods available, sequencing by ligations has the lowest error rate, and for STR sequencing which is short-read length, sequencing by ligations is the best platform for forensic genetic applications [8].

Many countries are building their STR profile DNA database of convicted and arrested persons concerned with offences, and this seems that STR profiling still valuable tool in forensic analysis work even though NGS will take over. Therefore, development of NGS with incorporated currently used STRs is the need of time as STRs alone contribute 15% of the human genome [4, 25]. Gettings et al. [25] presented the summary of each STR marker (24 autosomal forensic STR loci) which includes classification of STR marker, repeat unit length, the location of chromosome, repetitive STR region physical location in chromosome, and many more. The details about these markers have been presented in Table 15.1 (After Gettings et al. [25]).

Sequence variation study by NGS provides information about the sequence which can be very useful for the in-depth evaluation of STR alleles. The NGS result must be compared with the PCR-CE-length-based genotype. Three categories of discord were observed in comparing the NGS results to the CE data.

(a) *Presence of isoalleles*: If the allele sequence information is available, the loci which contain numerous isoalleles having internal sequence variation within the repeat region provide better resolution.
(b) *Flanking region indels*: In many STR loci, the 5′ and 3′ flanking regions showed variations which are within the range of PCR product. These are mainly SNPs, for example, at the D13S317 locus, where a four-base deletion in the 3′ flanking region resulted in a "9" allele by CE and a "10" allele by NGS [25].
(c) *Bioinformatic null allele*: This is of two types, Type 1, where the deletion of one base sometimes goes undetected as a result of which homozygous result appears (e.g., at Penta D locus), and Type 2, where the allelic ladder bin did not contain the matching allele (e.g., D12S391 locus, where a "17, 17.1 appears as homozygotes) [25].

Among the 24 autosomal STR loci, 9 loci showed the increase in alleles greater than 30% when sequenced with NGS platform as compared to PCR-CE-length-based genotype [25]. The remaining 15 STR loci showed less variation in the repeat region, for example, loci such as D5S818, D7S820, and D13S317. But less variation is also useful which are mainly in the flanking region which helps in the understanding of mutational events with evolution perspectives. The sequencing by NGS reveals the true variation of STR loci as new alleles have been detected by sequencing of simple STRs which is important as it will improve the statistical power of analysis [25]. If the variability in STR loci will be more, better will be the statistical power of investigation which helps in reducing the number of loci which is required in typing. Another advantage of NGS is to differentiate the STR typing homozygous genotype into heterozygous when the individual loci are sequenced.

Table 15.1 Autosomal STR variation study through NGS

Sl. No.	Classification of STR markers	Repeat unit length	Chromosome location	Physical location of the STR repeat region	Repeat motif sequence and size range of sequenced alleles	Primary source of variation (off ladder)	Presence of high-frequency polymorphisms in flanking region (within 150 bp)
1	D1S1656	Tetranucleotide repeat (TAGA)	Chromosome 1	230,769,616 to 230,769,683	8 to 19.3 repeats	[TAGA]n [TAGG] and [TAGA]n TGA [TAGA]n [TAGG]	Three SNPs with high frequencies close to repeat region from the 5′ end
2	TPOX	Tetranucleotide repeat (AATG)	Chromosome 2	1,489,653 to 1,489,684	4 to 14 repeats	Microvariant x.1, x.2, and x.3	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
3	D2S441	Tetranucleotide repeat (TCTA)	Chromosome 2	68,011,947 to 68,011,994	11 to 13 repeats	[TCTA]n [TNNN] [TCTA]n and TCTA]n TCA [TCTA]n	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
4	D2S1338	Tetranucleotide repeat (TGCC) and (TTCC)	Chromosome 2	218,014,859 to 218,014,950	(TGCC) 4 to 9 repeats and (TTCC) 6 to 19 repeats (total 10 to 26 repeats)	[TGCC]n [TTCC]n [GTCC] $[TTCC]_2$	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
5	D3S1358	Tetranucleotide repeat (TCTA)	Chromosome 3	45,540,739 to 45,540,802	7 to 19 repeats (total 11 to 23 repeats)	[TCTA] [TCTG]n [TCTA]n	No SNPs reported at greater than 5% frequency
6	FGA	Tetranucleotide repeat (CTTT)	Chromosome 4	154,587,736 to 154,587,823	13 to 29 repeats	$[TTTC]_3$ [TTTT] TT [CTTT]n [CTCC] $[TTCC]_2$	No SNPs reported at greater than 5% frequency
7	D5S818	Tetranucleotide repeat (AGAT)	Chromosome 5	123,775,556 to 123,775,599	4 to 20 repeats	[AGAT]n [ACAT] [AGAT]n and [AGAT] n [GAT] [AGAT]n	One SNPs with high frequencies close to repeat region from the 5′
8	CSF1PO	Tetranucleotide repeat (AGAT)	Chromosome 5	150,076,324 to 150,076,375	5 to 16 repeats	[AGGT]	No SNPs reported at greater than 5% frequency

(continued)

Table 15.1 (continued)

Sl. No.	Classification of STR markers	Repeat unit length	Chromosome location	Physical location of the STR repeat region	Repeat motif sequence and size range of sequenced alleles	Primary source of variation (off ladder)	Presence of high-frequency polymorphisms in flanking region (within 150 bp)
9	SE33 (most variable STR loci)	Tetranucleotide repeat (AAAG)	Chromosome 6	88,277,144 to 88,277,245	10 to 25 repeats	[AAAG]n AAAAG [AAAG]n and [AAAG]n AG [AAAG]n	No SNPs reported at greater than 5% frequency
10	D6S1043	Tetranucleotide repeat (AGAT)	Chromosome 6	91,740,225 to 91,740,272	9 to 23 repeats	[AGAT]n [ACAT] [AGAT]n	No SNPs reported at greater than 5% frequency
11	D7S820	Tetranucleotide repeat (GATA)	Chromosome 7	84,160,226 to 84,160,277	6 to 14 repeats	No sequence variation within the repeat region. Microvariant x.1 and x.3	Four SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
12	D8S1179	Tetranucleotide repeat (TCTA)	Chromosome 8	124,894,865 to 124,894,916	7 to 19 repeats	[TCTA]n [TCTG] [TCTA]n	No SNPs reported at greater than 5% frequency
13	D10S1248	Tetranucleotide repeat (GGAA)	Chromosome 10	129,294,244 to 129,294,295	7 to 19 repeats	[GGAA]n [AGAA] [GGAA]n	One SNPs with high frequencies close to repeat region from the 3′
14	TH01	Tetranucleotide repeat (AATG)	Chromosome 11	2,171,088 to 2,171,115	3 to 12 repeats	[TGAA]	No SNPs reported at greater than 5% frequency
15	vWA	Tetranucleotide repeat (TCTA)	Chromosome 12	5,983,977 to 5,984,044	10 to 22 repeats	[TCTA]n [TCTG] [TCTA]n	Four SNPs with high frequencies close to repeat region from the 3′
16	D12S391	Tetranucleotide repeat (AGAT)	Chromosome 12	12,297,020 to 12,297,095	------	[AGAT]n [AGAC]n [AGAT]	No SNPs reported at greater than 5% frequency
17	D13S317	Tetranucleotide repeat (TATC)	Chromosome 13	82,148,025 to 82,148,068	5 to 15 repeats	No sequence variation within the repeat region. Microvariant x.1 and x.3	One SNPs with high frequencies close to repeat region from the 3′

18	Penta E	Pentanucleotide repeat (AAAGA)	Chromosome 15	96,831,015 to 96,831,039	5 to 32 repeats	[AAAGA]n [AAATA] n and [AAAGA]n [AAGAA] [AAAGA] n	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
19	D16S539	Tetranucleotide repeat (GATA)	Chromosome 16	86,352,702 to 86,352,745	4 to 17 repeats	Microvariant x.1, x.2, and x.3	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
20	D18S51	Tetranucleotide repeat (AGAA)	Chromosome 18	63,281,667 to 63,281,738	8 to 40 repeats	Three variants [AGCA], [GGAA], between [AGAA]	No SNPs reported at greater than 5% frequency
21	D19S433	Tetranucleotide repeat (AAGG)	Chromosome 19	29,926,235 to 29,926,298	4 to 18 repeats	[TAGG]	No SNPs reported at greater than 5% frequency
22	D21S11	Tetranucleotide repeat (TCTA) and (TCTG)	Chromosome 21	19,181,973 to 19,182,099	24 to 38 repeats	[TA], [TCA], [TCCATA]	No SNPs reported at greater than 5% frequency
23	Penta D	Pentanucleotide repeat (AAAGA)	Chromosome 21	43,636,205 to 43,636,269	5 to 19 repeats	Flanking region deletions	No SNPs reported at greater than 5% frequency
24	D22S1045	Trinucleotide repeat (ATT)	Chromosome 22	37,140,287 to 37,140,337	8 to 19 repeats	[ATT]n [ACT], [ATT]2	No SNPs reported at greater than 5% frequency

Modified from Gettings et al. [25]

15.3.4 Application of NGS in Differentiating Monozygotic Twins

Monozygotic twins, as the word suggests, arise from a single fertilized egg, and thus they are genetically identical as compared to fraternal twins who are not genetically identical. This puts the block in differentiating the monozygotic twins through autosomal STR profiling. Both individuals have exactly the same DNA sequence; conventional genotyping approaches such as STR, SNP, sex-chromosome STR, and mtDNA analyses cannot differentiate between them. In paternity cases and other cases where one of the monozygotic twins are involved, the other twin cannot be excluded through autosomal STR profiling. This scenario can put the legal justice system in a difficult position to exactly match the biological evidence with one of the identical twins conclusively. Monozygotic (MZ) twins are genetically identical, but they must have some genetic differences which have accumulated in their genome generally called as epigenetic changes [11, 39, 41]. These epigenetic or somatic changes are normal but not exception, and if it occurs during the embryonic development of the MZ twins, it will be more prevalent in the tissues of that particular individual [12]. Somatic mutations or epigenetic changes are random, and surely the MZ twins acquire different mutations [12]. May evidence supports that epigenetic markers can be used to distinguish monozygotic (MZ) twins [40], predict tissue type [23], and accurately determine the age of a DNA donor [6].

Epigenetic approaches based on NGS technology include whole-genome bisulfite sequencing [27], methylation beadchips, reduced representation of bisulfite sequencing [44], and methylated DNA immune precipitation sequencing [64]. All these NGS technologies basically work well with larger DNA fragments, but the challenge with forensic samples is having low and highly degraded DNA samples even mixed with other elements. So a highly accurate NGS technology is required for forensic DNA analysis for epigenetic changes. Genome-wide amplification of a bisulfite-modified DNA template, followed by quantitative methylation detection using pyrosequencing, is one of the best NGS methods for extremely low amounts of DNA [47]. Identification of extremely rare mutations to differentiate between the MZ twins using ultra-deep NGS technology is the landmark work for the solution to paternity and forensic cases where MZ twins are involved [65]. Illumina Human Methylation BeadChip NGS technology is also used to study the methylation pattern (CpG sites) as an epigenetic change for differentiating the MZ twins [40]. The high mutation rate of the mitochondrial DNA (mtDNA) has the potential to become a promising biomarker for the differentiation between MZ twins. With the advancement in various NGS technologies, it is now possible to characterize minor differences of mtDNA genomes (mtGenomes) between MZ twins. In the study conducted by Wang et al. [62], nucleotide differences and heteroplasmies of MZ twins' mtGenomes were mapped for from six pairs of adult MZ twins by NGS technology using the Illumina HiSeq 2000 sequencing system. Their experimental evidence suggests that variants of mtGenomes could be a perspective biomarker to distinguish MZ twins from each other [62].

15.4 NGS in Forensic Casework

NGS is a highly sought technology in the field of biological science, and it has been there since over a decade and now also has become an integral part in the field of forensic sciences [61]. Many attempts have been taken such as Y-chromosome sequencing among related male, metagenomic study of biological stains, mtDNA sequencing for forensic applications, and so on with the help of various NGS platforms [5, 10, 32, 42, 46, 48, 49]. Various studies have been undertaken for the sequencing of markers employed for forensic DNA analysis like STRs, SNPs, microRNA, and mtDNA [9, 16, 19, 20, 22, 24, 26, 28, 37, 38, 43, 58, 62, 63, 66–68]. Many of these studies proved the application of NGS in forensic science as it incorporates multiplexing, sequence variation study of STRs, and high throughput. With the promise which NGS holds as a potent technology in forensic, many commercial kits are released by different companies [16, 22, 28, 66, 67].

15.4.1 STR Typing of Degraded Samples Using NGS

CE-based system is commonly used for STR genotyping in forensic, but frequently it is observed that this system fails to generate the DNA profile data from degraded DNA samples and thus put the forensic scientists in a situation where it become impossible to interpret the results [1–3, 15, 29, 54, 65]. Figure 15.4 depicts the CE-based STR genotyping vs NGS which clearly shows the improved results with NGS. Recently few forensic analyses were done for STR sequencing via NGS in routine casework [18, 21, 53, 56, 57, 60].

15.4.2 Massively Parallel Sequencing of Forensic CODIS 13 Autosomal STRs

The limitations of CE-based STR genotyping are known, and many improvements have also been incorporated in this system, but using NGS for all the 13 CODIS STR markers for human DNA profiling is being tried recently, and it is believed in the forensic scientist community that this can be the parallel method with CE-based STR genotyping. Figure 15.5 depicts the NGS analysis of 18 markers (13 CODIS STRs, D2S1338, D19S433, Penta D, Penta E, and amelogenin) which is based on multiplex PCR system and NGS analysis, and this analysis produced consistent results with CE-based STR genotyping. Many additional applications like, detection of sequence variations at target region, generation of STR profiles with degraded DNA samples and even fomo mixed stains is a limitation with only CE-based DNA Profiling [37].

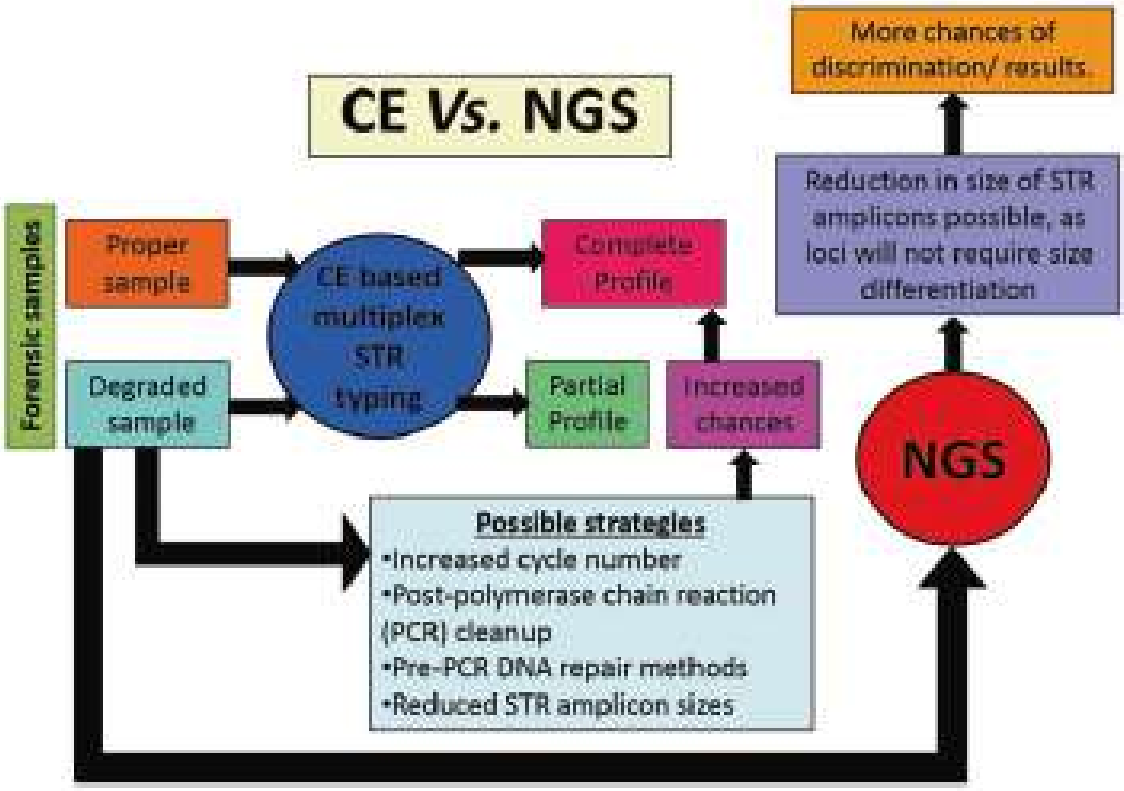

Fig. 15.4 Efforts for improved results in forensic DNA typing CE vs NGS for better results

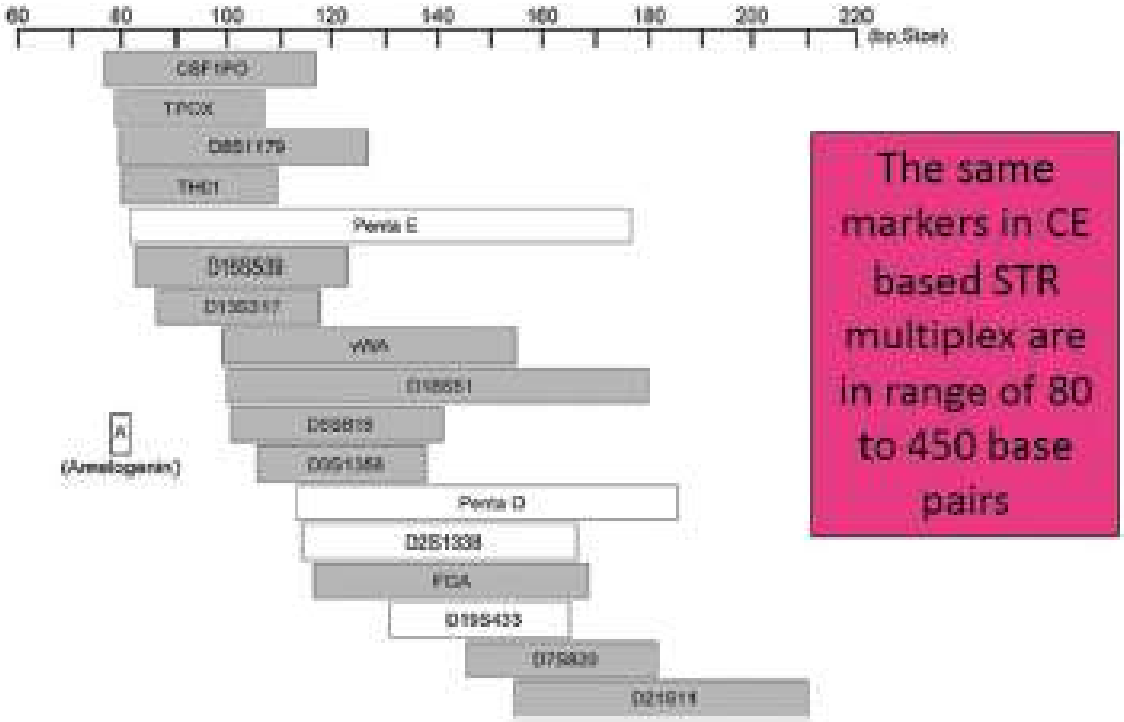

Fig. 15.5 Scheme of 18 forensic markers (13 CODIS shown in gray color) in the developed multiplex PCR system for NGS analysis. (Modified after Kim et al. [37])

Table 15.2 Available NGS-based kits for forensic use

Name of kit	Manufacturer/ supplier	STR loci included
HID-Ion AmpliSeq™ Identity Panel	Thermo Fisher Scientific	124 autosomal SNPs (most of the SNP*for*ID and individual identification SNPs (IISNs)) and 34 Y-chromosome SNPs
HID-Ion AmpliSeq™ Ancestry Panel	Thermo Fisher Scientific	Ancestry informative markers (AIMs)
Ion Torrent™ HID STR 10-Plex panel	Thermo Fisher Scientific	All-in-one solution from amplification of STRs and amelogenin and sequencing to data analysis
Ion Torrent PGM™ platform	Thermo Fisher Scientific	Includes 16 of 20 expanded combined DNA index system (CODIS) core loci and amelogenin
24-plex STR panel	Thermo Fisher Scientific	Cover all recommended combined DNA index system expansion markers (CODIS core loci, ESS markers, DYS391, and amelogenin)
Precision ID GlobalFiler NGS STR panels	Thermo Fisher Scientific	20 autosomal STR CODIS and expanded CODIS loci, 1 Y-chromosome STR locus, 1 autosomal NC02 locus, 3 autosomal low probability of identity (PI) (0.09) STR loci, 5 autosomal next-generation sequencing (NGS) STR loci, 1 indel polymorphic marker on the Y chromosome (Y indel), X and Y amelogenin, the sex determining marker
ForenSeq™ DNA Signature Prep Kit	Illumina	27 autosomal, 24 Y STRs, 7 X STRs and 94 identity, 56 ancestry, and 22 phenotypic SNPs in a single reaction
PowerSeq™ Systems (Auto, Y, and Mito)	Promega	22 autosomal STRs, 23 Y-STRs, and 10 amplicons covering the mitochondrial and amelogenin control region

15.4.3 NGS-Based Other Kits Available for Forensic Purposes

Since the inception and development of the technology, Thermo, Illumina, and Promega have made kits available for forensic purposes (Table 15.2).

15.5 Conclusion

NGS provides a possibility for constructing an all-in-one multiplex with relevant forensic markers that include STRs, SNPs, indels, and mtDNA markers and along with many other information. It is evident from the publications in the last few years that the use of NGS in forensic genetics is presently an important research area and will be in the near future as well. The technology has the enormous potential and could offer the first real alternative to PCR-CE analysis. Future research for development and improvement of NGS will make the technology be applied more smoothly, conveniently, and effectively for forensic DNA typing.

References

1. Abdelkrim J, Robertson B, Stanton JA, Gemmell N (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques 46:185–192
2. Allentoft M, Schuster SC, Holdaway R, Hale M, McLay E, Oskam C et al (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. BioTechniques 46:195–200
3. Allentoft ME, Oskam C, Houston J, Hale ML, Gilbert MT, Rasmussen M et al (2011) Profiling the dead: generating microsatellite data from fossil bones of extinct megafauna – protocols, problems, and prospects. PLoS One 6:e16670
4. Ayres KL, Chaseling J, Balding DJ (2002) Implications for DNA identification arising from an analysis of Australian forensic databases. Forensic Sci Int 129:90–98
5. Berglund EC, Kiialainen A, Syvänen AC (2011) Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet 2:23–37
6. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S et al (2011) Epigenetic predictor of age. PLoS One 6:e14821
7. Bornman DM, Hester ME, Schuetter JM, Kasoji MD, Minard-Smith A, Barden CA, Nelson SC, Godbold GD, Baker CH, Yang B, Walther JE, Tornes IE, Yan PS, Rodriguez B, Bundschuh R, Dickens ML, Young BA, Faith SA (2012) Short-read, high-throughput sequencing technology for STR genotyping. Biotech Rapid Dispatches:1–6
8. Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18:78–89
9. Børsting C, Fordyce SL, Olofsson J, Mogensen HS, Morling N (2013) Evaluation of the ion torrent HID SNP 169-plex: a SNP typing assay developed for human identification by second generation sequencing. Forensic Sci Int Genet 12:144–154
10. Brenig B, Beck J, Schütz E (2010) Shotgun metagenomics of biological stains using ultra-deep DNA sequencing. Forensic Sci Int Genet 4:228–231
11. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, Diaz de Stähl T, Menzel U et al (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy number-variation profile. Amer J Hum Genet 82:763–771
12. Budowle B (2014) Editors pick: Molecular genetic investigative leads to differentiate monozygotic twins. Invest Genet 5:11
13. Butler JM (2012) Advanced topics in forensic DNA typing: methodology. Elsevier, Inc, Waltham
14. Caratti S, Turrina S, Ferrian M, Cosentino E, Leo DD (2015) MiSeqFGx sequencing system: a new platform for forensic genetics. Forensic Sci Int 5:e98. https://doi.org/10.1016/j.fsigss.2015.09.040
15. Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning AP, Fox SE et al (2011) Discovery of highly divergent repeat landscapes in snake genomes using high throughput sequencing. Genome Biol Evol 3:641–653
16. Churchill JD, Schmedes SE, King JL, Budowle B (2016) Evaluation of the Illumina (®) beta version ForenSeq™ DNA signature prep kit for use in genetic profiling. Forensic Sci Int Genet 20:20–29
17. Dalsgaard S, Rockenbauer E, Gelardi C, Børsting C, Fordyce SL, Morling N (2013) Characterization of mutations and sequence variations in complex STR loci by second generation sequencing. Forensic Sci Int Genet Suppl Ser 4:e218–e219
18. Dalsgaard S, Rockenbauer E, Buchard A, Mogensen HS, Frank-Hansen R, Børsting C, Morling N (2014) Non-uniform phenotyping of D12S391 resolved by second generation sequencing. Forensic Sci Int Genet 8:195–199
19. Eduardoff M, Santos C, de la Puente M, Gross TE, Fondevila M, Strobl C et al (2015) Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the ion PGM™. Forensic Sci Int Genet 17:110–121

20. Eduardoff M, Gross TE, Santos C, de la Puente M, Ballard D, Strobl C et al (2016) Inter-laboratory evaluation of the EUROFORGEN global ancestry-informative SNP panel by massively parallel sequencing using the ion PGM™. Forensic Sci Int Genet 23:178–189
21. Fordyce SL, Avila-Arcos MC, Rockenbauer E, Børsting C, Frank-Hansen R, Petersen FT, Willerslev E, Hansen AJ, Morling N, Gilbert MT (2011) High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche genome sequencer FLX platform. BioTechniques 51:127–133
22. Fordyce SL, Mogensen HS, Børsting C, Lagac'e RE, Chang CW, Rajagopalan N, Morling N (2015) Second-generation sequencing of forensic STRs using the ion TorrentTM HID STR 10-plex and the ion PGMTM. Forensic Sci Int Genet 14:132–140
23. Frumkin D, Wasserstrom A, Budowle B, Davidson A (2011) DNA methylation-based forensic tissue identification. Forensic Sci Int Genet 5:517–524
24. Gelardi C, Rockenbauer E, Dalsgaard S, Børsting C, Morling N (2014) Second generation sequencing of three complex STRs D3S1358, D21S11 and D12S391 in Danes and a proposal for nomenclature of sequenced STR alleles. Forensic Sci Int Genet 12:38–41
25. Gettings KB, Aponte RA, Vallone PM, Butler JM (2015) STR allele sequence variation: current knowledge and future issues. Forensic Sci Int 18:118–130
26. Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA et al (2016) Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Sci Int Genet 21:15–21
27. Grunau C, Clark SJ, Rosenthal A (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res 29:E65–E65
28. Guo F, Zhou Y, Liu F, Yu J, Song H, Shen H et al (2016) Evaluation of the early access STR kit v1 on the ion torrent PGM™ platform. Forensic Sci Int Genet 23:111–120
29. Gymrek M, Golan D, Rosset S, Erlich Y (2012) lobSTR: a short tandem repeat profiler for personal genomes. Genome Res 22:1154–1162
30. Hares DR (2012) Expanding the CODIS Core Loci in the United States. Forensic Sci Int Genet 6:e52–e54
31. Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. Forensic Sci Int Genet 17:33–34
32. Holland MM, McQuillan MR, O'Hanlon KA (2011) Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. Croat Med J 52:299–313
33. Jeffreys AJ, Brookfield JF, Semeonoff R (1985a) Positive identification of an immigration test-case using human DNA fingerprints. Nature 317:818–819
34. Jeffreys AJ, Wilson V, Thein SL (1985b) Individual-specific 'fingerprints' of human DNA. Nature 314:67–74
35. Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. Nat Rev Genet 5:739–751
36. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12:179–192
37. Kim EH, Lee HY, Yang IS, Jung SE, Yang WI, Shin KJ (2016) Massively parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons. Forensic Sci Int Genet 22:1–7
38. King JL, LaRue BL, Novroski NM, Stoljarova M, Seo SB, Zeng X et al (2014) High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Sci Int Genet 12:128–135
39. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL et al (2002) Mutations in IRF6 cause van der woude and popliteal pterygium syndromes. Nat Genet 32:285–289
40. Li C, Zhao S, Zhang N, Zhang S, Hou Y (2013) Differences of DNA methylation profiles between monozygotic twins' blood samples. Mol Biol Rep 40:5275–5280