Next-generation DNA Sequencing in the Mirror of Law Enforcement

Balázs Kocsis[1], Norbert Mátrai[2], Balázs Egyed[3]

Next-generation sequencing (NGS) also known as massively parallel sequencing (MPS) has been increasingly important in recent years not exclusively in clinical applications, but also in criminal investigations. Several kits (chemical packages) validated for forensics are now available from manufacturers.

Since 2022, one of those kits examining a wide range of STR markers has been introduced into the portfolio of accredited analytical methods in the Hungarian Institute for Forensic Sciences (HIFS). During the internal laboratory validation of the Precision ID GlobalFiler NGS STR Panel v2 kit all quality measures and setting parameters were determined to make the technology available for routine forensic investigations. In the Institute a comprehensive population genetic analysis has been launched to create a representative allele frequency database of the Hungarian population to support the statistical interpretation, in court, of the results generated by the new genetic method. The other aim of the study was to assess the backward compatibility and divergence of the new technology to the traditional capillary electrophoresis.

In order to build up the population database 200 volunteer samples were analysed by the new MPS-STR method until now to create a first version allele frequency database. The developed population database is crucial for the biostatistical calculation and interpretation to measure the weight of evidence in forensic cases. Using the new technology, the increased variability of the observed alleles in the database will elevate the power of discrimination of each genetic evidence. According to recent experience the new MPS technology can be applied to analyse highly degraded and low copy number DNA samples as well.

Keywords: Next-generation sequencing, forensic application, accreditation, criminal cases, weight of evidence

I.   Introduction

The multifaceted possibilities given by genetics, especially by forensic genetics, has been widely used during the investigation of real crime cases from the early 90s. The traditional DNA analysis method based on capillary electrophoresis (CE) has long been known and used for forensic DNA analysis in criminal investigations. A lot of experience is already available regarding the possibilities and limitations of these methods and there is no doubt about the validity and usability of the results obtained during these tests. Therefore, forensic DNA analysis has become one of the most important tools of law enforcement today[4].

Thanks to the continuous development of technology, the success rate of examinations among the increasingly challenging on-site traces are also high. The biggest challenge for forensic genetics is clearly the high demand for testing low copy number (LCN) DNA samples and

---

[1] Forensic geneticist, Hungarian Institute for Forensic Sciences, Department of Genetics; PhD student: Eötvös Loránd University, Department of Genetics, e-mail: kocsisba@nszkk.gov.hu, orcid.org/0000-0001-8612-0529

[2] Forensic geneticist, Hungarian Institute for Forensic Sciences, Department of Genetics, e-mail: matrain@nszkk.gov.hu, orcid.org/0000-0003-0962-6205

[3] Senior lecturer, Eötvös Loránd University, Department of Genetics, e-mail: egyed.balazs@ttk.elte.hu, orcid.org/0000-0003-3960-2052

[4] Pádár et al., '*Genetics and Law Enforcement - The First Quarter of a Century of Forensic DNA-testing in Hungary I.*', (2019) DOI: 10.38146/BSZ.2019.12.1

degraded- or mixture samples originating from multiple individuals. From the beginning of the 2000s the manufacturers started to provide validated multiplex PCR/STR kits for forensic DNA analyses. With the first kits, only 3-4 loci could be analysed at the same time with one spectral channel on the first versions of capillary electrophoresis systems. The advent of multi-spectral devices made it possible to analyse more loci at the same time on multiple spectral channels, which lead to the early versions of these multiplex PCR kits. The newer 6 dye panel-STR kits containing sixteen to twenty-three autosomal loci are quite widespread; based on this type of multiplexes the DNA databases also expanded their upload options to include the *ESS*[5] loci to increase the potential matches with the older and smaller panels. Increasing the number of analysed loci in the human genome was the first step of this development line. However, capillary electrophoresis is limited in the length of amplicons that can be analysed; the longest amplicons in these kits are approximately 450-500 base pairs in length.

In addition, the manufacturers have started to produce panels that use smaller amplicons or more dyes to get valuable information from highly degraded challenging forensic samples. Nowadays there are kits under development with more than 30 loci and as many as 8 dyes, or multiplex PCR systems containing more than twenty markers within 400 bp PCR fragment length[6]. Other kits are designed specifically for degraded samples, which contain less markers but the amplicon sizes of the alleles are much shorter[7] (not longer than 250-300 bp). However, the developments did not end here.

At the same time when CE was evolving, a new method based on massive parallel sequencing of DNA (MPS) was arisen that is now known as next-generation sequencing (NGS). Over the last decade this methodology has added a new dimension to the field of forensic genetics. It is now possible to examine not only short tandem repeats (STRs), but also single nucleotide polymorphisms (SNPs) and insertion/deletion variations (indels) simultaneously for the identification of individuals or missing persons. Thanks to this capability, MPS offers an advantage over CE systems, because in addition to the length polymorphism, the difference in nucleotide sequences (sequence polymorphism) could be also considered, which has led to a significant increase in the discrimination power of STR genotyping. Moreover, MPS uses shorter amplicons to generate complete STR profiles, making it easier to analyse highly degraded samples.

With every further development, like the introduction of an alternative method, or the development of a new tool, the applied STR markers must be carefully selected. The reasons for this can be found in the existing national criminal DNA databases and in their information content. The Hungarian National Criminal DNA Database is based on the first *CODIS Core-loci*[8] *and European Standard Set loci (ESS)*, which has been continuously expanded with the addition of new recommended *CODIS Core-loci and ESS loci*. The original *CODIS Core-loci* are the following 13 markers*: CSF1PO, FGA, THO1, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11*, which have been used until December 31, 2016. As a result, the Hungarian DNA database contains the highest proportion of information about these markers. From January 1, 2017 – almost at the same time as the new STR kits appeared – 7 more markers have been added to the recommended marker list: *D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, D22S1045*. One point to be considered is that the Hungarian Institute for Forensic Sciences (HIFS) started to use and register the loci into the database earlier than the recommendation was published. Other markers can also be registered in the databases, as our institute does for additional autosomal markers, like *SE33* locus, which is widely used and analysed in European populations.

---

[5] ENFSI DNA Working Group, '*DNA database management review and recommendations*. (2022)
[6] Promega PowerPlex 35GY System; Cat. number: DC3520.
[7] AB NGM Detect PCR Amplification Kit, Cat. number: A31832.
[8] Combined DNA Index System

In addition, two other very important aspects must be considered before using new methods for forensic genetic testing: the existence of developmental validation of the new kits/techniques and proof of their usability (verification) to obtain an accredited laboratory process.

The Department of Genetics at the Hungarian Institute for Forensic Sciences (HIFS) has been accredited for the ISO/IEC 17025 standard since the end of 2013 and places great emphasis on continuously improving its quality management. According to the specifications of the standard system the entire institute - including the Department of Genetics - always strives to use validated (by developers, or manufacturers) methods for forensic examinations which aims to include in the range of accredited test methods as soon as possible.

In 2017, the NGS technology was introduced to the department with purchasing the Ion Torrent Next-Generation sequencing[9] platform (*ThermoFisher Scientific*). With this new technology, the first studies were carried out in HIFS on the mitochondrial genome and was successfully used in real crime cases – at that time there were no available validated autosomal STR genotyping kits on this platform. Shortly afterwards, the first autosomal STR genotyping kits appeared on the market for the new NGS platform, which set the direction of the development. The researcher needed to implement the new method: this was supported by doctoral grants and scholarships[10], like that awarded to the first author of this article.

The aims of our study were (1) to introduce the NGS technology as an accredited method in our routine laboratory testing; (2) to create the first allele frequency database to support the statistical interpretation of the results generated by the new genetic method in court; and (3) to highlight the power of the new technology in the light of justification through two sample cases.

II.   Laboratory criteria for the substantiated expert opinion

To ensure that genetic experts deliver the best possible results, the following points are crucial:
- Well-developed and - where possible - validated and accredited (according to ISO/IEC 17025 standard) testing methods.
- Appropriate reference database for biostatistical calculations.
- Appropriate statistical software and established formulas and settings.

In our recent study, CE and NGS technology based STR profiling was performed on 200 volunteer samples to generate a comprehensive allele frequency database for the general Hungarian population. Before collection of the buccal swab samples, written informed consent was obtained from each sample donor. All test steps were carried out according to the accredited process of Department of Genetics at HIFS[11] as follows:
- The DNA was extracted from buccal swabs using EZ1 Advanced XL instrument with EZ1&2 DNA Investigator Kit[12], accreditation ID: *VM-G-EZ-R*.
- Genomic DNA was quantified using the Quantifiler Trio DNA Quantification Kit in the 7500 Real-Time PCR System applying the HID Real-Time PCR Analysis Software[13]; accreditation ID: *VM-G-QF-R*.
- Capillary Electrophoresis (PCR/STR); accreditation ID: *VM-G-MP-R*:
  o samples were amplified using the Promega PowerPlex Fusion 6C System[14] and

---

[9] ThermoFisher Scientific Ion Torrent next-generation sequencing system
[10] Ministry of Culture and Innovation of Hungary from the National Research, Development, and Innovation Fund, financed under the KDP-2020 funding scheme
[11] Accreditation Certificate, ID NAH-1-1826/2022, https://www.nah.gov.hu/hu/szervezet/nemzeti-szakertoi-es-kutato-kozpont-bunugyi-igazsagugyi-szakertoi-igazgatosag-genetikai-szakertoi-intezet-1/
[12] Qiagen EZ1&2 DNA Investigator Kit
[13] ThermoFisher Scientific Quantifiler Trio DNA Quantification Kit
[14] Promega PowerPlex Fusion 6C System, Cat. no.: DC2705

were genotyped on ABI 3500[15] genetic analyzer.
- For the automated Next-Generation Sequencing IonChef Instrument was used together with Ion GeneStudio S5 System which is a semiconductor-based sequencer system[13]:
  - Library preparation was performed using the Precision ID GlobalFiler NGS STR Panel v2[16] kit and the Precision ID Chef DL8 Kits.
  - The pooled libraries were quantified using the Ion Library TaqMan Quantification Kit in the 7500 Real-Time PCR System and were analysed using the HID Real-Time PCR Analysis Software.
  - For template preparation and sequencing Ion S5 Precision ID Sequencing Kit together with Ion 530 Chip were used.

At the beginning of the research the applied NGS method had not yet been accredited: accreditation was in progress and was gained in 2022 with ID: *VM-G-NGS-R*.

Since 2017, the Promega PowerPlex Fusion 6C System is accredited for both 25µl and 12.5 µl final PCR volume in the laboratory, and mostly measurements are carried out using automated pipetting for better security and sample tracking.

Between 2017 and 2022, a lot of energy was invested to meet all the critical quality parameters of the NGS system, and to understand how does it work. During this period, in addition to carrying out the sensitivity-, mixture-, repeatability-, reproducibility- and stability studies, the analytical threshold, heterozygous imbalance, and stutter ratios were also set. The backward compatibility and convergence of the new (NGS) technology to the traditional capillary electrophoresis systems were also assessed through the 200 volunteer samples on 22 overlapping STR markers plus *Amelogenin* and on the *DYS391* Y chromosome marker (Table 1.). During the concordance analysis only one difference was found[17]. This concordance analysis is essential for the confirmation of the equivalence between the two technologies. By the way, there are some differences in the marker list of the Promega PowerPlex Fusion 6C System, and of Precision ID GlobalFiler NGS STR Panel v2 on NGS, however both kits contain all the 21 *CODIS Core-loci* (Table 1).

Table 1: Distribution and overlapping of the analyzed STR loci on the different typing systems and loci sets. Dark gray: locus is not implemented in the typing system/set.

SE33	AMEL-X	AMEL-Y	D1S1656	D2S441	D3S1358	FGA	D8S1179	D10S1248	TH01	vWA	D12S391	D18S51	D21S11	D22S1045	D5S818	CSF1PO	D7S820	D13S317	D16S539	D2S1338	TPOX	D19S433	PENTA D	PENTA E	DYS391	D1S1677	D2S1776	D3S4529	D4S2408	D5S2800	D6S1043	D6S474	D12ATA63	D14S1434	rs2032678
			ESS (European Standard Set) dated from 2009																																
			CODIS Core loci dated from 01. September 2017																																
			PowerPlex Fusion 6C System/capillary electrophoresis dated from 2017 (in HIFS)																																
			Precision ID GlobalFiler NGS STR Panel v2/IonTorrent S5 NGS system dated from 2020 (in HIFS)																																

As a result of the efforts made in the institute during this period, the established NGS STR typing method has been added to the department's list of accredited test methods by the mid-2022.

III   The appropriate reference database for biostatistical calculations

The first national allele-frequency database was created and published in the late 90's and in early 2000s. This was created on limited forensic STR markers by traditional PCR and CE technology which was at that time the cutting-edge method in forensic genetics.

---

[15] ThermoFisher Scientific, 3500xL Genetic Analyzer for Human Identification
[16] ThermoFisher Scientific, Precision ID GlobalFile NGS STR Panel v2
[17] Kocsis et al., '*Forensic Implications of the Discrepancies Caused between NGS and CE Results by New Microvariant Allele at Penta E Microsatellite.*' (2023) https://doi.org/10.3390/genes14051109

That database was developed on the STR data of 220 unrelated volunteer samples analysed by forensically informative autosomal STR markers[18,19]. Later this database was extended with additional markers – basically with the new ESS markers – in the early 2010s. Due to the development of the recent NGS based STR profiling technology, there is an urgent need to review the earlier databases, in order to add new markers to the set that contain new genetic information for the allelic discrimination, like nucleotide sequence polymorphism. We are now at this stage.

The number of sequence reads generated by the NGS process elevated so radically that the bioinformatics power requirements are significantly higher compared to previous CE based studies. The results of the NGS sequencing analysis are stored in large informatic files which could be further analysed and visualised using dedicated software packages (like *TorrentBrowser, HID Genotyper Plugin, Converge* software). One of the most challenging parts of the technology is to store this large amount of data over long periods in time, and to make further biostatistics like LR (likelihood ratio) calculation. The bioinformatic research and development has resulted in the creation of software packages that utilize NGS results to determine the probability of derivation from an individual, and to ascertain the strength of evidence.

In addition to dedicated software packages, a well-established allele frequency database is also a prerequisite for this type of testing. According to the latest recommendation of the professional community, the minimum number of samples required from unrelated individuals in the case of NGS genetic population data publication is 50 samples[20]. The Hungarian Institute for Forensic Sciences Department of Genetics aims to reach a higher number (i.e. 500 unrelated people) which is required in case of the traditional STR data produced by capillary electrophoresis. To date, we have already analysed 300 volunteer samples using the new NGS method in the laboratory, of which more than 200 have already been evaluated.

From the generated data the allele frequency table for statistical calculation has been prepared in two formats. One was based on the *RU* (*R*epeat *U*nit) calculation, another in the *RU_LUS* (*R*epeat *U*nit_*L*ongest *U*ninterrupted *S*tretch) format. In the first case, the allele designation is determined only by repeat unit numbers (the same as in CE), while in the second case, the number of the longest uninterrupted stretch (LUS[21]) - which is derivate of the bracketed form of the allele sequences (to take into consideration the isoalleles which are caused by sequence difference) - is also taken into account (Table 2).

[18] Egyed et al., '*Population Genetic Data on the STR Loci D2S1338, D19S433 and SE33 in Hungary.*' (2005) DOI:10.1520/JFS2004479

[19] Egyed et al., '*Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers.*' (2006) https://doi.org/10.1016/j.forsciint.2005.07.004

[20] Editors of Forensic Science International: Genetics '*Revised guidelines for the publication of genetic population data.*' (2017) http://dx.doi.org/10.1016/j.fsigen.2017.06.007

[21] Justet al., "*Use of the LUS in sequence allele designations to facilitate probabilistic genotyping of NGS-based STR typing results.*" (2018) https://doi.org/10.1016/j.fsigen.2018.02.016

Table 2: Example of the two types of allelic determination, designation and frequencies. The RU (Repeat Unit) and RU_LUS (Repeat Unit_Longest Uninterrupted Stretch) based allele frequencies were established on the 200 Hungarian individuals population database.

Locus	Bracketed form of the example alleles (sequencing results)	RU allele designation	RU allele frequency	RU_LUS allele designation	RU_LUS allele frequencies
D8S1179	[TCTA]**2** TCTG[TCTA]**11**	14	0.2525	14_11	0.045
	[TCTA]**14**	14		14_14	0.0275
	[TCTA]**1** TCTG[TCTA]**11**	14		14_12	0.18

As expected, during the analysis of 200 volunteer samples, several specific structures could be observed at the majority of sequenced STR loci (Figure 1.) due to sequence polymorphisms, therefore the number of allele counts increased significantly in more than the half of the loci.

Figure 1.: Number of observed alleles per locus (31 autosomal loci from Precision ID GlobalFiler NGS STR v2 panel) in relation to RU and RU_LUS formats [22]
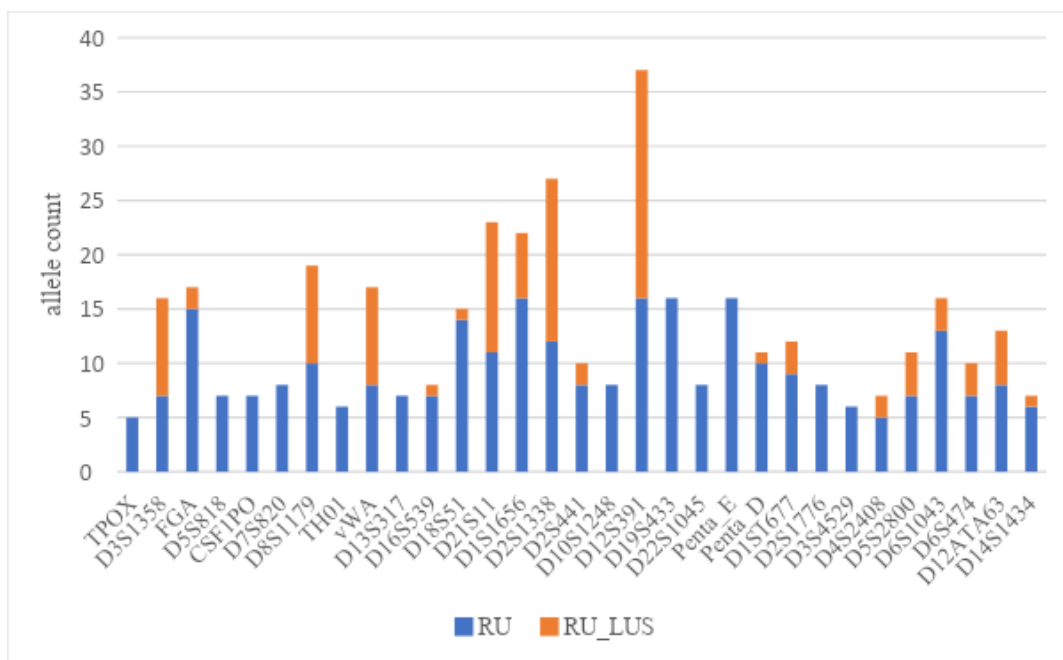
What is the relevance and the importance of the two different forms of genotype determination? It is the occasional sequence polymorphism of the determined allele (like allele 14 in Table 2), which could harbour differing sequences between two, otherwise seemingly identical alleles. For example in Table 2, two persons can have the similar allele 14 in a locus by traditional RU based typing, however it could be differentiated in their sequence motif: first person's allele could be determined as 14_11, while the second as 14_12 allele in the *RU_LUS* format.

---

[22] The orange bars present the increased allele counts.

In addition, this type of diversity which could not be detected technically by traditional capillary electrophoresis, is also considered during the likelihood ratio calculations (LR[23,24]), which uses the previously described and prepared allelic-frequency databases.

The basic principle of allele frequency databases is that alleles which are rarely observed in a given population correspond to lower frequencies, and vice versa. If the frequency of a given allele is low, the LR value calculated by the software for that allele will be high, thus increasing the overall weight of evidence (Table 2). For instance in Table 2. the RU-based allele 14 at locus D18S1179 has a frequency of 0.2525. However, in the RU_LUS based allele frequency table, due to the appearance of three isoalleles, the allele frequency of 0.2525 will be fragmented into the following frequency values for the four RU_LUS alleles: 14_11: 0.045; 14_12: 0.18; 14_14: 0.0275. This finding will have considerable relevance to the probability of the derivation from an individual, and to the weight of evidence (LR calculation).

Thanks to the fact that the first version of the allele-frequency database has been completed, which is based on less but still a sufficient amount of data, there is a possibility to start to fulfil the first statistical calculations and collecting experience. However, for a better resolution of calculations there is a constant need to elevate the number up to 500 samples on which the allele-frequency database is based. After reaching this point, the whole database could be published and used routinely not only for kinship testing (identification of deceased persons) but also for several types of crime cases (the derivation of traces from an unknown person).

III.   Appropriate statistical software and well-designed settings

Several statistical calculation programs are already available, many of which have been validated for forensic genetic purposes and are available either in open-source format or for purchasing. To analyse low copy number or mixed DNA traces the applied software can operate the following statistical models:
- Consensus method: A consensus genetic profile is created by keeping the alleles that are present in at least half of the replicates. Calculations based only on the presence/absence of the alleles (*LRmixStudio*[25]).
- Semi-continuous method: all replicates are considered separately without creating a consensus profile. Calculations based only on the presence/absence of the alleles. (*LRmixStudio*).
- Continuous method: all replicates are considered separately without creating a consensus profile, but the RFU intensity (in case of CE) or the number of sequence reads (in case of NGS) are applied to the calculation in addition to the presence/absence of the alleles (e.g. *DNAxs*[26], or *EuroForMix*[27]).

According to the ISFG recommendation published in 2018[28], *"probabilistic models are now preferred instead of the consensus approach"* (Recommendation 6); where drop-in, drop-out,

[23] Gillet al., '*DNA commission of the International Society of Forensic genetics: Recommendations on the interpretation of mixtures.'* (2006) https://doi.org/10.1016/j.forsciint.2006.04.009

[24] The ISFG recommends the likelihood ratio (LR) calculation method for analysing and interpreting forensic DNA mixtures to evaluate the evidential support for the identification hypothesis that a suspect contributed their DNA to the biological evidence.

[25] https://www.isfg.org/Software#dna021

[26] Benschop et al., "*DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles.*" (2019) https://doi.org/10.1016/j.fsigen.2019.06.015

[27] Bleka et al., '*EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts.'* (2016) https://doi.org/10.1016/j.fsigen.2015.11.008

[28] Gill et al., '*DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions.'* (2020) https://doi.org/10.1016/j.fsigen.2019.102186

peak height variability or relationships can aid a scientific approach, using continuous testing and model validation software.

Until present for NGS data interpretation *EuroForMix* software was tested and used in our department. The software can calculate genotype data in both RU and RU_LUS format, therefore the read numbers per allele are integrated in the calculations, furthermore these calculations could be performed with or without stutter-models. The software can specify a potential degradation model for DNA profiles based on the quality of the uploaded DNA profile. In the case of Precision ID GlobalFiler NGS STR results this model could not have been used until now due to the absence of the software developer settings related to the panel. This gap will be filled through collaborative work between the developer and our Institute.

The strength of the technology will be illustrated through the following two cases.

IV.   Case 1

An incomplete neurocranium missing the entire facial region was dug out during a construction work (Picture 1).

Picture 1: Frontal picture from the incomplete brain skull.

No other bones, clothes or personal documents were found. For the genetic analyses the petrous bone of the skull has been cut out. The petrous bone was milled and powdered in a *Retch Cryomill* instrument using our accredited method (*VM-G-CSP:2017[8]*) and two DNA extraction were made in a Automate Express Instrument with the Prepfiler BTA Kit (*ThermoFisher Scientific*) according to the accredited method of the laboratory (*VM-G-AE-R[8]*).

The quantity of the two DNA isolates were measured on 7500 Real-Time PCR system with Quantifiler™ Trio DNA Quantification Kit (*VM-G-QF-R[8]*). Both DNA isolates contained DNA in a very low concentration and in a highly degraded state. In addition, DNA genotyping was attempted by the traditional multiplex PCR/STR technology (altogether four PCR/STR experiments with two different type of kits) on CE, and with NGS technology (three replicates on two different libraries). The results are shown in Table 3., however the real genotypes cannot be presented here, because of the special personal data.

Table 3: The genotyping results of the os petrosa of the cranium by traditional PCR based CE and by the new NGS methods[29] Genotype data cannot be presented due to the personal genetic data.

Method	Capillary Electrophoresis	NGS
Kit/Locus	Promega PowerPlex Fusion 6C + GlobalFiler kit (ThermoFisher Scientific)	Precision ID GlobalFiler NGS STR Panel v2 (ThermoFisher Scientific)
AMEL	X	X
D10S1248	failed	failed
vWA	successful	successful
D16S539	successful	successful
D2S1338	failed	successful
D8S1179	successful	successful
D21S11	failed	failed
D18S51	failed	successful
D22S1045	successful	failed
D19S433	successful	successful
TH01	successful	successful
FGA	failed	failed
D2S441	successful	successful
D3S1358	successful	successful
D1S1656	successful	failed
D12S391	successful	successful
Penta E	failed	failed
Penta D	failed	failed
SE33	failed	Not included in the kit
D7S820	failed	failed
CSF1PO	failed	successful
D13S317	failed	successful
TPOX	failed	successful
D5S818	failed	failed
D12ATA63	Not included in the kit	successful
D14S1434	Not included in the kit	successful
D1S1677	Not included in the kit	successful
D2S1776	Not included in the kit	successful
D3S4529	Not included in the kit	successful
D4S2408	Not included in the kit	failed
D5S2800	Not included in the kit	successful
D6S1043	Not included in the kit	successful
D6S474	Not included in the kit	successful

Only 10 loci plus Amelogenin (sex determining locus) were successfully determined by CE method (from which 6 out of 10 were homozygous) – that result was enough only for a restricted database search. Nonetheless, in case of NGS typing 21 STR loci plus Amelogenin were successfully evaluated, which is a significant increase in the number of determined loci. According to the DNA quantitation the DNA was highly degraded, so the increase in evaluated loci could be due to the fact that NGS works with much shorter fragments, and with more sensitive detection technology.

Unfortunately, there were no reference samples available for a comparison at that time, therefore no direct probabilistic statistical calculation could be performed. However it was

---

[29]Genotype data cannot be presented due to the personal genetic data.

concluded that the probability of a possible direct match within the database is much higher with NGS results (21 loci) than only with CE genotypes (10 loci). The updated DNA profile has been uploaded to the national criminal DNA database (CODIS) and is under permanent search.

V.   Case 2

Demonstrative tests were used during the validation of the Precision ID GlobalFiler NGS STR system, where female/male DNA mixture samples were created in different ratios for the representation of a typical sexual assault case sample. The analytical results of *19:1 female:male DNA ratio* (475 pg female DNA plus 25 pg male DNA mixed as input to the PCR reaction) sample were used for the further statistical calculation and interpretation. This type of mixed biological traces are typical in case of sexual assaults, where in addition to the dominant female-, the minor male biological material is only present to a lesser extent.

To identify the perpetrator through statistical calculations is a big challenge for these types of samples. Typically, the presence of the major female donor (usually the victim from whom the sample was provided) could be fixed for the biostatistical calculations under the applied hypotheses (H_p - hypothesis of prosecution and the H_d - hypothesis of defence) so the statistical calculations can be done according to Table 4.

Table 4: Results of the statistical calculations[30]

Female:male ratio 19:1= 475:25 pg input DNA	CE	CE	NGS RU data	NGS RU_LUS data
Type of the methods	Semi-continuous replicates, drop-in, drop-out rates are considered	Continuous replicates, drop-in, drop-out estimation model and peak-heights are considered, stutter-models used		
Used software for biostatistics	LRmixStudio	EuroForMix		
Number of replicates taken into account	3			
Number of analysed autosomal loci	22		31	
LR (likelihood ratio) value	$1.07*10^{-6}$	$2.5*10^5$	$2.19*10^{17}$	$1.07*10^{19}$
Oral interpretation	Exclusive	Supportive	Extremely supportive	Extremely supportive

No allele drop-out was observed in case of the major donor, however in case of the minor donor several were found (25%). The number of analysed loci were 31 in case of NGS and 22 in case of CE kits, so proportionally more loci results can be used for statistics in case of NGS method (Table 4.). Furthermore, the possible nucleotide polymorphism – which can only be analysed by NGS not by CE kits – can elevate the statistical value of the DNA evidence.

The likelihood ratios have been steadily increased according to the type of method used. According to the CE results using the semi-continuous method the presence of the minor contributor can be excluded. However, using the continuous method, the presence of the minor contributor could be slightly supported in the 22 loci.

---

[30] The alleles of the major (female) donor were fixed during the calculation. CE=Capillary Electrophoresis; NGS=Next Generation Sequencing; RU: calculation made by sequencing results in Repeat Unit data format; RU_LUS: calculation made by sequencing results in Repeat Unit_Longest Uninterrupted Stretch data format.

For NGS-based calculations, the weight of evidence was extremely supportive in both types of data (*RU* and *RU_LUS*) where the highest LR rate was obtained for the *RU_LUS* NGS data. This may be due to the elevated number of analysed loci and the allele frequencies of the observed isoalleles. An instance of the latter can be found in the calculation of the LR value for locus D8S1179. The software-derived LR value based on RU data was 4.36, in contrast, the LR value calculated on RU_LUS allele designation was 109.2, which was about 20 times elevation in LR value. The minor (male) donor had RU-based allele 14 and RU_LUS allele 14_14 at this locus. For more details, see Table 2.

VI.   Summary

In this study maybe the most important technological developments and trends were presented that have characterised the field of forensic genetics over the past 20 years. The pace of technological development has accelerated during the last decade to such an extent that it is often challenging to keep up for the users and experts too.

The main goal of the Hungarian Institute for Forensic Sciences is to follow this model line with introducing the new technologies into the forensic practice. However, these new technologies require a lot of investments, namely: infrastructure, laboratory equipments, reagents, instruments, automation, good laboratory practice, staff and ongoing education. The NGS based STR genotyping technology together with the appropriate biostatistical and interpretation method can put a new tool into the hands of forensic experts and for the law enforcement. Together with an established allele-frequency database the Hungarian Institute for Forensic Sciences covered a milestone where NGS DNA technology can contribute to both solving of criminal cases, or identifying missing persons. Of course, the work cannot stop here. There are still much to be done: the size of the database has to be elevated to a minimum of 500 individuals; population comparison studies must be done to establish the population structuring effects that have to be considered during biostatistical calculations.

Using the new NGS technology, the increased variability of the observed alleles in the database will elevate the power of genetic evidence. In addition, according to recent experience, the new NGS technology can be applied to analyse highly degraded and low copy number DNA samples. At the same time, there is an ongoing need for the continuation of population genetic studies to refresh the allele-frequency databases alongside the laboratory developments, to make the laboratory results more tangible and comprehensible for the participants in judicial proceedings.

VII.   Funding

VIII.   Acknowledgment

IX.   References

Benschop, C.G. Corina; Hoogenboom, Jerry; Hovers, Pauline; Slagter, Martin; Kruise, Dennis; Parag, Raymond; Steensma, Kristy; Slooten, Klaas; Nagel, H.A. Jord; Dieltjes, Patrick; Marion, Vincent van; Paassen, Heidi van; Jong, Jeroen de; Creeten, Christophe; Sijen, Titia and Kneppers, Alexander, '*DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles.*' (2019) Forensic Science International Genetics, Volume 42, 81-89. https://doi.org/10.1016/j.fsigen.2019.06.015

Bleka, Øyvind; Storvik, Geir and Gill, Peter, '*EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts.*' (2016) Forensic Science International: Genetics, Volume 21, 35-44. https://doi.org/10.1016/j.fsigen.2015.11.008

Combined DNA Index System, Online: https://www.fbi.gov/how-we-can-help-you/dna-fingerprint-act-of-2005-expungement-policy/codis-and-ndis-fact-sheet

Editors of Forensic Science International: Genetics, '*Revised guidelines for the publication of genetic population data.*' (2017) Forensic Science International: Genetics, Volume 30, 160–163. http://dx.doi.org/10.1016/j.fsigen.2017.06.007

Egyed, Balázs; Csikai, Mátyás; Füredi, Sándor and Pádár, Zsolt, '*Population Genetic Data on the STR Loci D2S1338, D19S433 and SE33 in Hungary.*' (2005) Journal of Forensic Sciences, Vol. 50, No. 3., Paper ID JFS200447, DOI:10.1520/JFS2004479

Egyed, Balázs; Füredi, Sándor; Angyal, Miklós, Balogh, István; Kalmár, Lajos; Pádár, Zsolt, '*Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers.*' (2006) Forensic Science International, Volume 158, Issues 2–3, 244-249. https://doi.org/10.1016/j.forsciint.2005.07.004

ENFSI DNA Working Group, '*DNA database management review and recommendations.*' Online: https://enfsi.eu/news/guideline-dna-database-management-review-and-recommendations/

Gill, P., Brenner, C.H., Buckleton, J.S., Carracedo, A., Krawczak, M., Mayr, W.R., Morling, N., Prinz, M., Schneider, P.M. and Weir, B.S., '*DNA commission of the International Society of Forensic genetics: Recommendations on the interpretation of mixtures.*' (2006). Forensic Science International, Volume 160 (2–3) 90-101. https://doi.org/10.1016/j.forsciint.2006.04.009

Gill, Peter; Hicks, Tacha; Butler, M. John; Connolly, Ed; Gusmão, Leonor; Kokshoorn, Bas; Morling, Niels; Oorschot, Roland A.H. van; Parson, Walther; Prinz, Mechthild; Schneider, M. Peter; Sijen, Titia and Taylor, Duncan, '*DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions.*' (2020) Forensic Science International: Genetics, Volume 44, 102186., https://doi.org/10.1016/j.fsigen.2019.102186

Just, S. Rebecca and Irwin, A Jodi, '*Use of the LUS in sequence allele designations to facilitate probabilistic genotyping of NGS-based STR typing results.*' (2018) Forensic Science International: Genetics, Vol 34, 197–205. https://doi.org/10.1016/j.fsigen.2018.02.016

Kocsis, Balázs; Mátrai, Norbert and Egyed, Balázs, '*Forensic Implications of the Discrepancies Caused between NGS and CE Results by New Microvariant Allele at Penta E Microsatellite.*' (2023) Genes, 14, 1109. https://doi.org/10.3390/genes14051109

Pádár, Zsolt; Kovács, Gábor; Nógel, Mónika; Czebe, András; Zenke, Petra and Kozma, Zsolt, '*Genetics and Law Enforcement - The First Quarter of a Century of Forensic DNA-testing in Hungary I.*' (2019) Belügyi Szemle: 67 (12). 7-34., DOI: 10.38146/BSZ.2019.12.1

Promega, PowerPlex Fusion 6C System, Cat. no.: DC2705, Onlone: https://worldwide.promega.com/products/forensic-dna-analysis-ce/str-amplification/powerplex-fusion-6c-system/?catNum=DC2705

Promega PowerPlex 35GY System, Online: https://worldwide.promega.com/products/forensic-dna-analysis-ce/str-amplification/powerplex-35gy-system/?catNum=DC3520

Qiagen EZ1&2 DNA Investigator Kit, Online: https://www.qiagen.com/us/products/human-id-and-forensics/investigator-solutions/ez1-2-dna-investigator-kit

ThermoFisher Scientific NGM Detect PCR Amplification Kit, Online: https://www.thermofisher.com/order/catalog/product/A31832

ThermoFisher Scientific, Quantifiler Trio DNA Quantification Kit, Online: https://www.thermofisher.com/order/catalog/product/4482910

ThermoFisher Scientific Ion Torrent next-generation sequencing system, Online: https://www.thermofisher.com/hu/en/home/brands/ion-torrent.html?SID=fr-iontorrent-main

ThermoFisher Scientific, 3500xL Genetic Analyzer for Human Identification, Online: https://www.thermofisher.com/order/catalog/product/4406016

ThermoFisher Scientific, Precision ID GlobalFiler NGS STR Panel v2 https://www.thermofisher.com/order/catalog/product/A33114