Antonio Alonso[1]
Pedro Alberto Barrio[1]
Petra Müller[2]
Steffi Köcher[4]
Burkhard Berger[2]
Pablo Martin[1]
Martin Bodner[2]
Sascha Willuweit[4]
Walther Parson[2,3]
Lutz Roewer[4]
Bruce Budowle[5]

[1]Department of Biology, National Institute of Toxicology and Forensic Sciences, Madrid, Spain
[2]Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria
[3]Forensic Science Program, The Pennsylvania State University, PA, USA
[4]Institute of Legal Medicine and Forensic Sciences, Charité–Universitätsmedizin Berlin, Berlin, Germany
[5]Center for Human Identification, University of North Texas Health Science Center, TX, USA

Received January 21, 2018
Revised April 21, 2018
Accepted May 2, 2018

Review

Current state-of-art of STR sequencing in forensic genetics

The current state of validation and implementation strategies of massively parallel sequencing (MPS) technology for the analysis of STR markers for forensic genetics use is described, covering the topics of the current catalog of commercial MPS-STR panels, leading MPS-platforms, and MPS-STR data analysis tools. In addition, the developmental and internal validation studies carried out to date to evaluate reliability, sensitivity, mixture analysis, concordance, and the ability to analyze challenged samples are summarized. The results of various MPS-STR population studies that showed a large number of new STR sequence variants that increase the power of discrimination in several forensically relevant loci are also presented. Finally, various initiatives developed by several international projects and standardization (or guidelines) groups to facilitate application of MPS technology for STR marker analyses are discussed in regard to promoting a standard STR sequence nomenclature, performing population studies to detect sequence variants, and developing a universal system to translate sequence variants into a simple STR nomenclature (numbers and letters) compatible with national STR databases.

Keywords:
Capillary electrophoresis / Forensic genetics / Massively parallel sequencing / Short tandem repeats / Validation studies DOI 10.1002/elps.201800030

1 Introduction

Currently, there is an increasing number of forensic genetic institutes and agencies that are investigating and beginning to implement within their laboratories the massively parallel sequencing (MPS) technology for (1) analysis of "classical" forensic DNA markers (i.e., DNA database driven STRs and mitochondrial DNA control region) used worldwide in forensic casework; and (2) study the potential application of other DNA markers less applied in casework, e.g., novel STRs, nuclear SNPs, insertion/deletion (INDEL) markers, and whole genome mitochondrial DNA sequence [1–6]. These additional markers can be used for forensic identity, ancestry, and/or phenotype determinations. One of the advantages of MPS platforms is the ability to incorporate into a single workflow the simultaneous analysis of hundreds or thousands of different DNA markers. Another advantage is that sequence-level variation can be ascertained. It is realized that the extent to which STR profiling (i.e., the gold forensic DNA standard methodology, carried out today with CE) can be performed and forensically validated by using MPS will impact implementation of MPS technology into forensic casework.

The analysis of classical forensic STR markers using MPS offers several advantages over conventional CE analysis namely (1) an increased number of loci that can be analyzed simultaneously, (2) higher discrimination power as a consequence of the increased STR allele sequence diversity and the greater number of loci, and (3) shorter amplicons for a more effective analysis of degraded and/or low quantity forensic biological evidence.

However, to realize these benefits of STR analysis through MPS, there are challenges that must be addressed before this appealing technology can be considered applicable for routine application within and across the range of forensic laboratories worldwide. Some of the challenges have recently been identified in the survey [1] conducted by the DNASEQEX Consortium [7] in collaboration with the European Network of Forensic Science Institutes (ENFSI) DNA Working Group

Correspondence: Dr. Antonio Alonso, Department of Biology, National Institute of Toxicology and Forensic Sciences, José Echegaray 4, 28232 Las Rozas, Madrid, Spain
E-mail: Antonio.alonsoalonso@justicia.es

Abbreviations: **ACR**, allele coverage ratio; **DoC**, depth of coverage; **HLA**, human leukocyte antigens; **INDEL**, insertion/deletion; **MPS**, massively parallel sequencing

Color Online: See the article online to view Figs. 1–3 in color.

(http://enfsi.eu/about-enfsi/structure/working-groups/dna/). These challenges include: a lack of consistent nomenclature and reporting standards, a lack of compatibility with existing National DNA Database infrastructure, and a lack of population data to support statistical calculations. The main obstacle to implementation, perceived by forensic DNA laboratories, is not these scientific challenges per se; but rather it is due to a lack of sufficient funding and the apparent higher cost per run of MPS technology compared with CE technology. Note that this cost is based solely on a run and not the amount of information (i.e., number and types of markers) per run.

Because of the costs and substantial amount of data that need to be analyzed, collaborative strategies have been developed. Multiple forensic DNA scientific societies and working groups [8, 9], various transnational research projects [7, 10], and the industry supplying MPS technologies are undertaking several initiatives to address the challenges this technology presents. The International Society for Forensic Genetics (ISFG) has established minimum criteria for MPS-STR sequence data analyses and addressed the need of a consistent and platform-independent nomenclature system that is backward compatible to the huge body of existing STR data produced by CE [8]. The Scientific Working Group on DNA Analysis Methods (SWGDAM) has provided nominal guidance to consider when performing and attempting to validate MPS in its recently revised version of the Validation Guidelines for DNA analysis methods [9]. The EU funded project DNASEQEX (DNA-STR Massive Sequencing & International Information Exchange) [7] aims to evaluate MPS-based materials (e.g., kits and supporting reagents) in their respective developmental stages using the two established platforms *MiSeq* (Illumina) and Ion *S5*™ (Thermo Fisher Scientific) and then provide feedback to the respective companies to help improve and establish MPS technologies that would meet the needs of forensic purposes. Another deliverable aims at evaluating and/or developing open-source MPS interpretation software that is independent of an analytical platform and translates MPS results into nomenclature that is compatible with established CE-STR data [11–14]. A NIST-led project, endorsed by the ISFG and called STRseq (for STR Sequencing Project), is developing a database to facilitate description of sequence-based alleles of forensically relevant STR loci so that communication among laboratories is facilitated [10]. The database will offer a curated catalog of sequence diversity at forensic STR loci, along with the key elements of nomenclature conforming to current guidelines. The EU funded project STEFA (Steps Towards a European Forensic Science Area; 2018–19) includes the working package *Empowering Forensic Genetic DNA Databases for the Interpretation of Next Generation Sequencing Profiles (dna.bases)* that introduces new alignment and search functions for the established forensic genetic databases EMPOP (empop.online) [15] and STRidER (strider.online) [16]. Finally, commercial companies are developing specific STR multiplex assays for MPS analysis, as well as expert software systems that allow backwards and parallel compatibility of STR data generated with CE systems [17–19]

The main objective of this paper is to review the current state of validation and implementation strategies of MPS technology for the analysis of STR markers for forensic genetics use. The current catalog of commercial STR-MPS panels, leading MPS-platforms, and STR-MPS data analysis tools are described. In addition, the developmental and internal validation studies and population studies carried out to date are summarized. Also, various initiatives developed by several international projects and standardization (or guidelines) groups to facilitate application of MPS technology for STR marker analyses are discussed in regard to promoting a standard STR sequence nomenclature, performing population studies to detect sequence variants, and developing a universal system to translate sequence variants into a simple STR nomenclature (numbers and letters) compatible with national STR databases.

2 MPS-STR panels, platforms, and sequencing data analysis tools

Three commercial STR-MPS panels have been developed to date that include the recommended expanded Combined DNA Index System (CODIS) and the European Standard Set (ESS) core STR loci, as well as additional autosomal STRs (Table 1), and Y-STR markers (Table 2) that enhance discriminatory power. The DNA primer Mix A of the ForenSeq™ DNA Signature Prep Kit [17] includes 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, and the Amelogenin sex marker. In addition, 94 identity informative SNPs are included. The PowerSeq™ Auto/Mito/Y panel [18] combines 23 autosomal STR loci, 23 Y-STR loci, ten subregions covering the whole mitochondrial DNA control region, and the Amelogenin sex marker. The Precision ID Globalfiler™ V2.0 NGS STR Panel [19] includes 20 CODIS STR loci, nine new autosomal Mini-STR loci, two Penta-nucleotide STR markers (Penta D and Penta E), one Y-STR (DYS391), and three sex markers (Amelogenin, SRY and rs2032678).

The Forenseq™ DNA Signature Prep Kit and the PowerSeq™ systems are multiplex systems that are compatible with the MiSeq System (Illumina). A special MPS platform, the MiSeq FGx Forensic Genomics System (Illumina), is available that includes data analysis software only for the ForenSeq™ Signature Prep Kit (see below). ForenSeq™ libraries are generated using a two-step amplification procedure. In the first, the targeted forensic STRs and SNPs are amplified by PCR. The second amplification is performed to attach adapters and unique indices. Incorporated adapters, complementary to immobilized oligos on the surface of the flow cell, allow a library(ies) to bind to the flow cell for bridge amplification. Unique indices are used to label one specific sample and enable the pooling of up to 96 samples in one run. The PowerSeq™ system uses an enzymatic ligation to add adapters and indices to the purified PCR-targets [21]. The recommended DNA input is 1 ng and 500 pg for the

Table 1. Reference information of 34 autosomal STR loci used in commercial STR-MPS systems

Locus	Standard set	Chromosomal location	GRCh38 repeat region start [8]	Repeat motif (forward strand) [8]	Original reading sequence [8]	Amplicon length range (bp)		
						ForenSeq™ DNA signature prep kit	GlobalFiler™ NGS STR panel	PowerSeq™ auto systems
D1S1677	Additional STR	1q23.3	163590026	[TTCC]a	Forward		151–191	
D1S1656	CODIS/ESS	1q42	230769616	[CCTA]a [TCTA]b	Reverse	141–189	167–215	161–208
D2S441	CODIS/ESS	2p14	68011947	[TCTA]a	Forward	144–180	163–195	158–204
TPOX	CODIS	2p25.3	1489653	[AATG]a	Forward	85–145	167–199	196–244
D2S1776	Additional STR	2q24.3	168788893	[AGAT]a	Forward		163–195	
D2S1338	CODIS	2q35	218014859	[GGAA]a [GGCA]b	Reverse	114–182	133–197	197–269
D3S4529	Additional STR	3p12.1	85803484	[GATA]a	Forward		167–195	
D3S1358	CODIS/ESS	3p21.31	45540739	TCTA [TCTG]a [TCTA]b	Forward	138–186	129–177	192–240
D4S2408	Additional STR	4p15.1	31302798	[ATCT]a	Forward	93–117	167–191	
FGA	CODIS/ESS	4q28	154587736	[GGAA]a GGAG [AAAG]b AGAA AAAA [GAAA]c	Reverse	150–306	137–299	176–268
D5S2800	Additional STR	5q11.2	59403132	[GGTA]a [GACA]b [GATA]c [GATT]d	Forward		171–211	
D5S818	CODIS	5q23.2	123775556	[ATCT]a	Reverse	102–150	141–173	191–239
CSF1PO	CODIS	5q33.1	150076324	[ATCT]a	Reverse	85–129	143–183	185–229
D6S1043	Additional STR	6q16.1	91740225	[ATCT]a ATGT 0–1 [ATCT]b	Reverse	163–227	163–227	
D6S474	Additional STR	6q21	112557951	[AGAT]a [GATA]b [GGTA]c [GACA]d	Forward		158–186	
D7S820	CODIS	7q21.11	84160226	[TATC]a	Reverse	135–179	130–166	211–255
D8S1179	CODIS/ESS	8q24.13	124894865	[TCTA]a [TCTG]0-2 [TCTA]b	Forward	86–138	151–199	203–255
D9S1122	Additional STR	9q21.2	77073826	[TAGA]a	Forward	108–140		
D10S1248	CODIS/ESS	10q26.3	129294244	[GGAA]a	Forward	128–172	155–199	135–179
TH01	CODIS/ESS	11p15.5	2171088	[AATG]a ATG 0–1 [AATG]b	Forward	100–148	129–173	220–264
D12S391	CODIS/ESS	12p	12297020	[AGAT]a [AGAC]b AGAT 0–1	Forward	237–281	149–193	202–254
vWA	CODIS/ESS	12p13.31	5983977	[TCTA]a [TCTG]b [TCTA]c	Reverse	132–192	147–207	202–262
D12ATA63	Additional STR	12q23.3	107928590	[TTG]a [TTA]a	Forward		126–146	
D13S317	CODIS	13q31.1	82148025	[TATC]a	Forward	138–186	149–181	209–257
D14S1434	Additional STR	14q32.13	94842054	[CTGT]a [CTAT]b	Forward		163–195	
Penta E	Additional STR	15q26.2	96831015	[TCTTT]a	Reverse	362–467	168–273	179–284
D16S539	CODIS	16q24.1	86352702	[GATA]a	Forward	132–180	139–179	198–253
D17S1301	Additional STR	17q25.1	74684855	[AGAT]a	Forward	114–142		
D18S51	CODIS/ESS	18q21.33	63281667	[AGAA]a	Forward	140–227	156–232	190–277
D19S433	CODIS	19q12	29926235	[CCTT]1 CCTA [CCTT]1 CTTT [CCTT]a	Reverse	154–212	155–195	193–253
D20S482	Additional STR	20p13	4525692	[AGAT]a	Forward	125–165		
D21S11	CODIS/ESS	21q21.1	19181973	[TCTA]a [TCTG]b [TCTA]c TA [TCTA]d TCA [TCTA]e TCCATA [TCTA]f	Forward	158–276	179–245	203–273
Penta D	Additional STR	21q22.3	43636205	[AAAGA]a	Forward	209–293	139–204	192–266
D22S1045	CODIS/ESS	22q12.3	37140287	[ATT]a [ACT]1 [ATT]2	Forward	193–229	178–211	129–176

Table 2. Reference information of 29 Y-STR loci used in commercial STR-MPS systems

Locus	Y-chromosomal location (Mb)*	GRCh38 repeat region start [8]	Repeat motif (forward strand) [8]	Original reading sequence	Amplicon lenght range (bp)	
					ForenSeq™ DNA signature prep kit (Illumina)	PowerSeq™ auto systems (Promega)
DYS393	3.13	3263111	[AGAT]a	Forward		294–256
DYS505	3.68	3772790	[TCCT]a	Forward	154–194	
DYS456	4.27	4402919	[AGAT]a	Forward		141–165
DYS570	6.86	6993190	[TTTC]a	Forward	162–214	157–217
DYS576	7.05	7185318	[AAAG]a	Forward	183–235	155–203
DYS522	7.46	7547585	[ATAG]a	Forward	294–334	
DYS458	7.87	7999821	[GAAA]a	Forward		171–199
DYS481	8.43	8558337	[CTT]a	Forward	102–129	139–184
DYS19	9.52	9684380	[TCTA]a TAGG [TCTA]b	Reverse	261–345	168–294
DYS391	14.10	11982089	[TCTA]a	Forward	123–167	147–178
DYS635	14.38	12258860	[TAGA]a [TACA]b [TAGA]c [TACA]d [TAGA]e [TACA]f [TAGA]g	Reverse	214–306	155–179
DYS437	14.47	12346267	[TCTA]a	Forward	178–210	181–197
DYS439	14.51	12403517	[GATA]a	Forward	199–239	204–224
DYS389I	14.61	12500448	[TAGA]a [CAGA]b N48 [TAGA]c [CAGA]d	Reverse	231–275	258–294
DYS389II	14.61	12500544	[TAGA]a [CAGA]b N48 [TAGA]c [CAGA]d	Reverse	255–299	
DYS438	14.94	12825889	[TTTTC]a	Forward	144–169	202–242
DYS390	17.27	15163067	[TAGA]a [CAGA]b [TAGA]c [CAGA]d	Reverse	242–286	204–248
DYS643	17.43	15314132	[CTTTT]a	Forward	115–215	150–210
DYS533	18.39	16281349	[TATC]a	Forward	198–258	242–284
GATA-H4	18.74	16631673	[TCTA]a	Reverse	151–203	231–251
DYS612	19.32	13640728	[CCT]a [CTT]b [TCT]c [CCT]d [TCT]e	Forward	215–248	
DYS385a	20.8	18680632	[GAAA]a	Forward	316–354	202–303
DYS385b	20.84	18639713	[TTTC]a	Reverse		
DYS460	21.05	18888810	[TATC]a N106 [TATC]b	Reverse	356–380	
DYS549	21.52	19358338	[GATA]a	Forward	214–262	189–230
DYS392	22.63	20471987	[ATA]a	Reverse	346–358	143–164
DYS448	24.36	22218923	[AGAGAT]a N42 [AGAGAT]b	Forward	288–324	213–255
DYF387S1a	25.93	23785361	[AAAG]a [GTAG]b [GAAG]c [AAAG]d [GAAG]e [AAAG]f [GAAG]g [AAAG]h	Forward	123–255	
DYF387S1b	28.03	25884581	[CTTT]a [CTTC]b [CTTT]c [CTTC]d [CTTT]e [CTTC]f [CTAC]g [CTTT]h	Reverse		

initial ForenSeq™ PCR [17] and for the initial PowerSeq™ systems PCR [20], respectively. Prior to sequencing libraries (the number of samples determined based on desired read depth of markers and throughput) are pooled for a run and loaded onto the sequencing cartridge [17, 21]. Bridge amplification generates millions of clonal clusters of individual DNA fragments from purified libraries that are attached to the surface of a flow cell. Parallel sequencing-by-synthesis is carried out by the incorporation of a fluorescently and reversibly terminator-labeled dNTPs, followed by the cleavage of the terminator to allow the incorporation of the next complementary base. During each sequencing cycle, all four dNTPs are present and minimize incorporation bias by natural competition [22]. The simultaneous addition of all four reversibly terminator-labeled nucleotides enables each sequencing cycle to be driven to completion as and minimizes the risk of misincorporation [23]. Since base calling is realized by direct signal intensity measurements raw error rates are widely reduced compared to other MPS methods [24–26].

The Precision ID Globalfiler™ system (Thermo fisher Scientific, Waltham, MA, USA) is a multiplex system that can be analyzed on the Ion Torrent platforms (Ion Torrent PGM/Ion S5). DNA libraries are generated by first amplifying the target forensic STRs and then they are "barcoded" by ligation with ion-code oligonucleotides. Clonal amplification of purified DNA libraries is performed by emulsion PCR and parallel sequencing of each amplicon is carried out by detection of the release of hydrogen ions, as indication of nucleotide incorporation, on a complementary metal-oxide semiconductor (CMOS) chip. Sequential exposure of

individual classes of dNTPs enables determination of the incorporated nucleotide with a concomitant change in pH.

Basically, the analysis of MPS-STR sequence data can be divided into three phases: raw reads, sequence alignment to reference and sorting, and allele calling. The majority of the forensic laboratories uses the analysis software provided by the respective companies for MPS-STR sequence data analysis: The Universal Analysis Software (UAS) (Illumina) [27], the Torrent Suite Software (TSS) [28] and the Converge software (Thermofisher) [29]. These software packages provide background information on quality metrics, read lengths, and alignment and provide standard output files such as BAM (a binary compressed representation of a Sequence Alignment Map (SAM, text format file defining the alignment of each sequence)) and FASTQ files (a text-based format for storing nucleotide sequence and its corresponding quality scores) for compatibility with other STR sequence data analysis software packages.

The UAS software provides semi-automated STR allele and genotype calls, including application of quality indicators that assist in manual data review of loci, tertiary analyses such as automated sample comparisons, and generation of population statistics such as random match probabilities. A STR sample detail table displays length-based allele calls for each marker. Each STR locus detail table provides a bar chart representation of allele calls and number of reads (to mimic presentation of a CE electropherogram) including isometric heterozygotes (alleles of the same fragment length but containing different sequences), and information on the STR sequence, generally excluding flanking region variation, for the alleles at the locus.

The Converge NGS Analysis module is designed to analyze profiles from the Precision ID GlobalFiler NGS STR Panel (and can be applied to other panels as well). NGS data analysis functionality includes graphical representation of STR allele calls and number of reads, including isometric heterozygotes, information on STR sequence motifs, known SNPs in flanking regions, and % of strand (forward or reverse) read depth. User-defined and default analysis settings are provided in the NGS module for flexible data interpretation with an interface to evaluate sequencing data using quality values and flags. Converge Software also allows management of CE profiles for easy comparison with sequence and length-based data obtained with the Precision ID GlobalFiler NGS STR panel.

Table 3 shows the commercial software packages developed to date for STR genotyping from sequence data obtained by MPS [27–32], as well as open-access MPS-STR software systems published recently in the forensic genetics literature [11–14, 33–38]. As an example of open access software, STRait Razor [11,12] is an open-source software tool that runs on all major operating systems including Microsoft Windows and is designed to detect forensically relevant STR alleles in FASTQ sequence data, based on either sequence or allelic/amplicon length. Alleles are detected via matching of the leading and trailing flanking region(s) surrounding the repeat region of a locus. This software is capable of analyzing STR loci with repeat motifs ranging from simple to complex without the need for extensive allelic sequence data and can capture flanking region variation. STRait Razor is designed to interpret both single-end and paired-end data and relies on intelligent parallel processing to reduce analysis time. An ancillary benefit of STRait Razor is that it can analyze SNP and INDEL data as well.

3 Validation studies, quality parameters, and interpretation thresholds

Recently, validation studies have been carried out on the ForenSeq™ system, the first MPS-STR assay commercialized for forensic identification purposes. Jäger et al. [39] published SWGDAM developmental validation studies of the ForenSeq™ system that included species specificity, sensitivity, mixed samples, stability (inhibitors and degradation), accuracy, and precision studies. Default analysis parameters used throughout were a 1.5% analytical threshold (AT) and 4.5% interpretation threshold (IT), for all loci except for DYS389II (>5.0% AT, >15% IT), DYS448 (>3.3% AT, >10% IT), and DYS635 (>3.3% AT, >10% IT). AT is the coverage (% of sequence readings) at and above which alleles can be reliably distinguished from background noise produced by sequencing / PCR errors. The IT is the coverage at or above which it is reasonable to assume that allelic dropout of a heterozygous sister allele has not occurred. AT and IT values were determined for a locus by multiplying the analysis parameter percentage value by the sum of read counts at that locus. In cases of low read depth, a minimum read number of 650 reads was used for the locus in determination of the threshold values. Default stutter filter percentages for autosomal STR, Y-STR, and X-STR markers ranged from 7.5% (D2S441, D4S2408, PentaD) to 50% (DYS481). Sensitivity results for autosomal STR, Y-STR, and X-STR loci with genomic DNA inputs of 1 ng, 500, 250, 125, and 62.5 pg yielded 100% call rates for all loci, with the exception of one amplification where the DXS10103 locus was not detected at 125 pg. Mixture studies demonstrated the ability of the ForenSeq™ system to detect minor contributor alleles at less than 5% of the major donor. Calculations from the STR repeatability and reproducibility studies (1 ng template) indicated 100% accuracy of the ForenSeq™ system in allele calling relative to CE for STRs ($n = 1260$ samples). Results provided support that the ForenSeq™ system meets forensic DNA validation guidelines.

Churchill et al. [40] evaluated the beta version of the ForenSeq™ system for forensic purposes by performing a series of experiments that tested reliability, sensitivity, mixture analysis, concordance, and the ability to analyze challenged samples. Depth of coverage (DoC) (better termed read depth), allele coverage ratio (ACR, calculated as the lower coverage allele divided by the higher coverage allele), and sequence coverage ratio (SCR, calculated as the number of reads used to make nominal repeat length allele calls and the number of reads attributed to stutters divided by the total number

Table 3. Commercial and open-access software packages developed for STR-MPS data analysis

Software	Type	Website	Running platform/programming language	References
Torrent Suite™ Software	Commercial	https://www.thermofisher.com/es/es/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-data-analysis-workflow/ion-torrent-suite-software.html	Web Browser	[28]
HID STR Genotyper Plugin	Commercial	https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0015879_HID_STR_Genotyper_Plugin_UG.pdf	Web Browser	[30]
Converge™	Commercial	https://www.thermofisher.com/es/es/home/industrial/forensics/human-identification/forensic-dna-analysis/forensic-dna-data-interpretation/converge-forensic-analysis-software.html	Web Browser/Desktop	[29]
ForenSeq™ Universal Analysis Software	Commercial	https://www.illumina.com/systems/sequencing-platforms/miseq-fgx/products-services/forenseq-universal-analysis-software.html	Desktop	[27]
NextGENe®	Commercial	http://www.softgenetics.com/NextGENe.php	Desktop	[31]
ExactID®	Commercial	https://www.battelle.org/government-offerings/homeland-security-public-safety/security-law-enforcement/forensic-genomics/exactid	Desktop	[32]
toaSTR	Open-access	www.toastr.de	Web Browser	[33]
STRait Razor v2/v3	Open-access	https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/strait-razor/	Perl/C++	[11, 12]
lobSTR	Open-access	http://lobstr.teamerlich.org/	C/C++	[34]
STRinNGS	Open-access	(Available upon request)	Python	[13]
TSSV	Open-access	https://git.lumc.nl/j.f.j.laros/tssv/blob/master/README.md https://pypi.python.org/pypi/tssvTSSV	Python	[35]
FDSTools	Open-access	https://pypi.python.org/pypi/fdstools/	Python	[36]
MyFLq	Open-access	https://github.com/beukueb/myflq http://forensic.ugent.be/ https://basespace.illumina.com/apps/174174/MyFLq	Web Browser/MySQL	[14]
RepeatSeq	Open-access	https://github.com/adaptivegenome/repeatseq	Python	[37]
SEQ Mapper	Open-access	http://forensic.mc.ntu.edu.tw:9000/SEQMapperWeb/Default.aspx	Web Browser/.NET	[38]

of reads) were used as informative parameters for assessing the quality of the data produced. The average DoC across the STRs was 2104X (range: 68X–13 014X). Thirty-nine of the 40 autosomal-STR and all X-STR markers had an ACR of 0.6–1.0. The SCR compared the number of unique sequence reads used to make allele calls versus the number of unique sequence reads that can be attributed to noise (i.e., sequencing/PCR errors) and ranged from 0.54–0.98. Data were found to be concordant with current CE methods (for markers in common), and mixtures generally up to a 1:19 ratio were resolved accurately. The authors concluded that the beta version of the ForenSeq DNA Signature Prep Kit is a valid tool for forensic DNA typing and demonstrated reproducible results and full profiles with DNA input amounts of 1 ng.

Just et al. [41] investigated the performance of the ForenSeq™ system for autosomal STR and Y-STR typing by examination of 151 sample libraries developed from high quality DNAs amplified at the target 1 ng template by using a STR intralocus balance threshold of 0.5, and the manufacturers default thresholds (analytical threshold of 1.5% of sequence reads and an interpretation threshold of 4.5% of sequence reads) for samples with coverage above 650 reads. They demonstrated that, excluding the D22S1045 and DYS392 loci producing poor results, autosomal STR and Y-STR ForenSeq profiles were 99.96 and 100% concordant, respectively, with CE data.

The prototype PowerSeq Auto System (Promega) containing 23 STR loci and Amelogenin has been evaluated using the MiSeq platform [42]. Zeng et al. [42] showed that the system was reproducible, and complete MPS-STR profiles could be generated using as little as 62 pg of input DNA. The mixture study indicated that partial STR profiles of the minor contributor could be detected up to 1:19 mixtures. The mock forensic casework study showed that full or partial MPS-STR profiles could be obtained from different types of single source and mixture samples. These studies indicated that the PowerSeq Auto System and the MiSeq can generate concordant results with current CE-based methods. Van der Gaag et al. [43] reported reliability and concordance with CE data (except for two Penta D alleles) of PowerSeq autosomal STR results from 297 population samples. The two differences were likely due to a primer-binding variant that caused excessive heterozygote imbalance, which is a phenomenon encountered with CE data as well. A genomic input DNA of

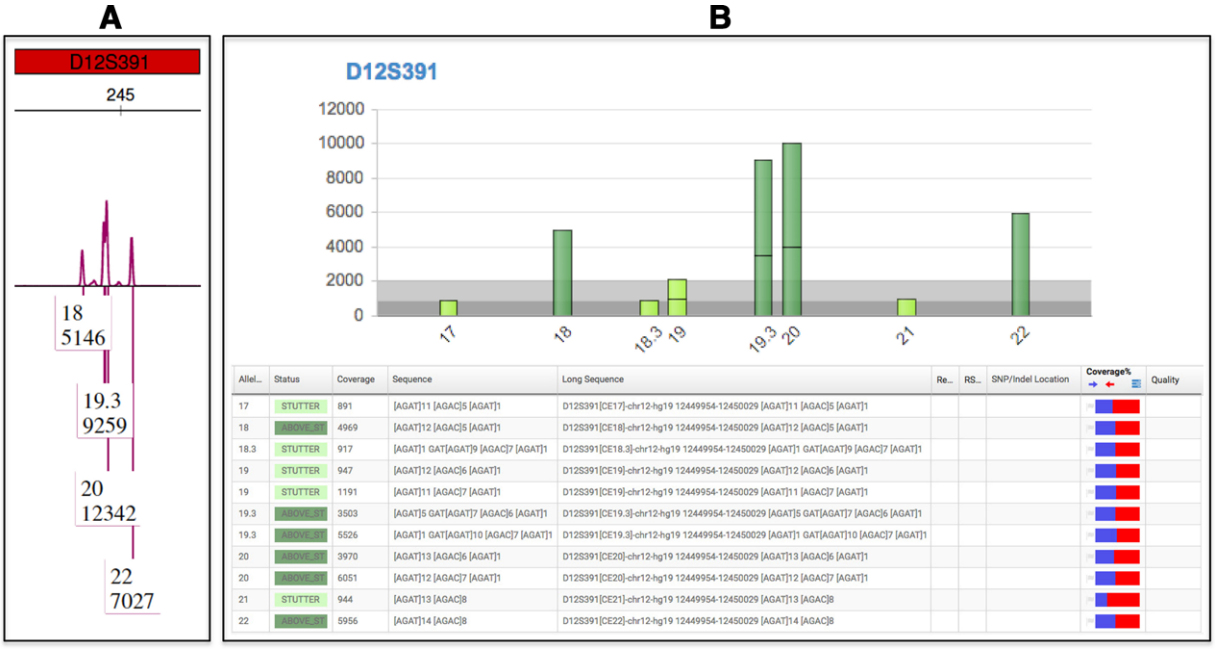

Figure 1. D12S391 results from a three-contributor DNA mixture as analyzed (A) by conventional CE (ABI3500) using the Globalfiler Kit and GeneMapper IDX (Thermo Fisher Scientific) for allele calls (B) by MPS using the Precision ID GlobalFiler NGS STR Panel and Converge V2.0 for allele calls.

500 pg generated high DoC for all alleles (5099–24 269), with an overall average ACR of 0.80 ± 0.15 for all 23 STR loci. With 62 pg of input DNA complete profiles were detected, although some imbalanced ACRs were observed. Mixture analysis down to a minor contribution of 5% was feasible for most STR loci.

Wang et al. [44] evaluated the Precision ID GlobalFiler™ NGS STR Panel using the Ion PGM™ System and the HID STR Genotyper pluginV4.0 for data analysis. Reproducible results were obtained from different types of single source samples (blood stains, muscle samples, hair root samples, semen stains, cigarette butts, and bone samples). Their sensitivity study (dilution series including 1000, 500, 250, 100, and 50 pg DNA template) indicated that single source complete genotypes could be obtained using as little as 100 pg of input DNA. Their mixture study demonstrated that the system could detect partial STR genotypes of the minor contributor up to a 1:19 mixture.

The DNASEQEX consortium has evaluated two approaches available at the time of establishment of the project. First, a prototype of the Precision ID GlobalFiler NGS STR Panel using the Ion Chef platform for automated DNA library and DNA template preparation followed by sequencing on the Ion S5 platform [40] (S5), and second, the Illumina ForenSeq™ DNA Signature Prep Kit using the Illumina MiSeq FGx sequencer (FGx).

For allele calling and sequence data interpretation and reporting on the S5 platform, Converge NGS module V2.0 was used. A genomic input DNA of 1 ng generated high average DoCs for all STR markers (2200–14 000 X). Sensitivity studies were performed with genomic DNA inputs of 500, 250, 125, 62, and 31 pg of the NIST 2372A standard and DNA control 2800 M samples. Results showed reproducibility and generation of complete Precision ID GlobalFiler NGS STR profiles with as little as 62 pg of input DNA. Average stutter percentages determined from 27 single source DNA samples ranged from 5.6% (TH01) to 18.9% (D12ATA63). A 99.9% (1 dropout/1046 alleles) accuracy was obtained with the Precision ID GlobalFiler NGS STR Panel for allele calling relative to that of the CE from single source DNA samples. Mixture studies consisted of analysis of 16 mixed stains of two and three contributors from recent GEDNAP (http://www.gednap.org) proficiency exercises that have been previously analyzed by CE (Globalfiler and PowerPlex Fusion 6C kits). The detection of 25 isoalleles from a total of 315 alleles (7.94%) from the mixed stains, allowed for a more informative characterization of the number of contributors. Figure 1 shows an example of a 3-contributor mixture in which only four alleles were differentiated by CE (compatible with a 2-contributor assumption), while MPS-STR analysis allowed the detection of two additional isoalleles for a total of six alleles. The Precision ID GlobalFiler NGS STR Panel, that included 23 STR markers with amplicons shorter than 200 bp, was very effective for the analysis of degraded bone DNA samples. Figure 2 shows the comparison of STR-CE Profiling and STR-MPS sequencing of a 41-year-old bone sample (*pars petrosa*) from an exhumed skeleton of a newborn in a case of alleged abduction of newborns in Spain. While the D16S539, CSF1PO, and TPOX markers analyzed by CE (CE size ranges in bp: 221–273, 277–325, and

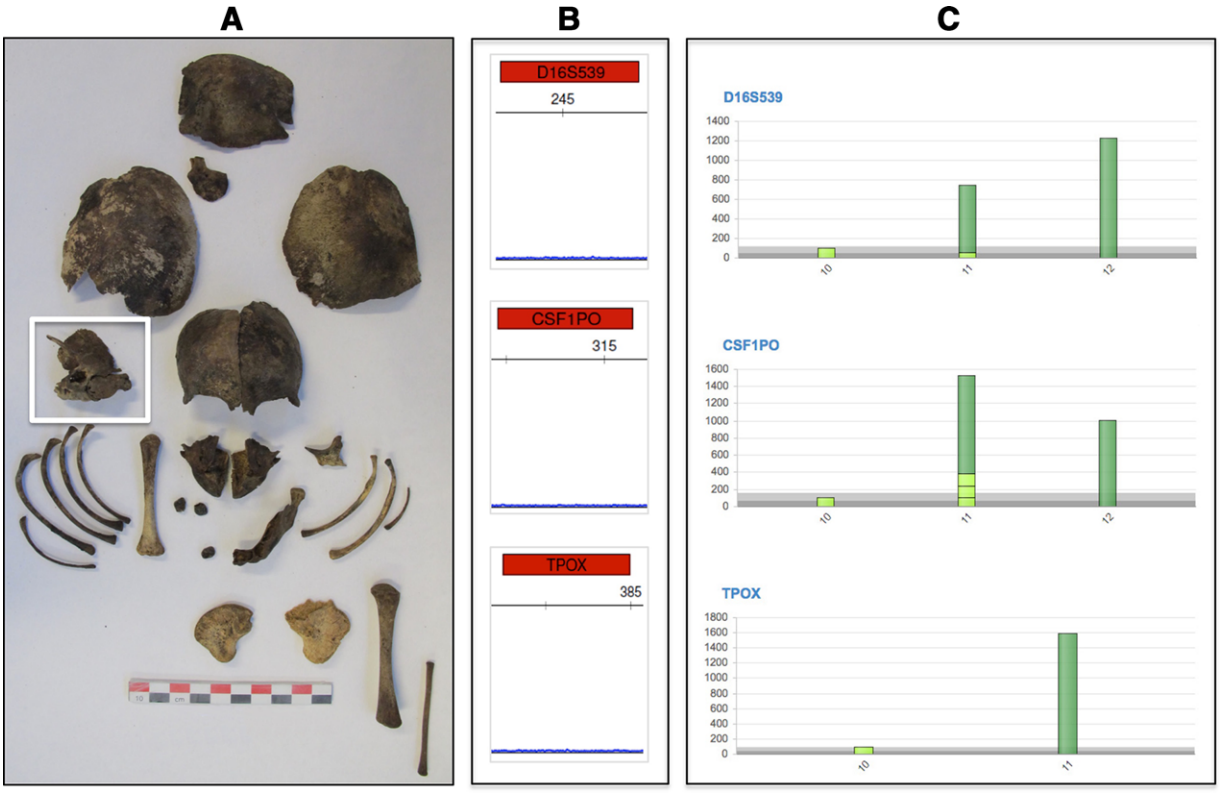

Figure 2. Comparison of STR-CE and STR-MPS results from the exhumed skeleton of a newborn. (A) Exhumed skeleton of a newborn in a case of alleged abduction in Spain. The *pars petrosa*, the bone used to obtain DNA, is highlighted inside a white square. (B) STR-CE negative results (using the HID Globalfiler Kit and HID GeneMapper IDX software for allele calling) obtained from the D16S539, CSF1PO, and TPOX markers. (C) STR-MPS positive results (bar chart representation of allele calls and number of reads as analyzed by Converge V2.0) obtained from the D16S539, CSF1PO, and TPOX markers that have shorter amplicons.

332–384, respectively) yielded negative results, the use of the MPS technology allowed reproducible sequencing results of the three STR markers likely as a result of being shorter amplicons (MPS size ranges in bp: 139–179, 143–183, and 167–199, respectively) (Fig. 2).

For evaluation of the Forenseq™ DNA Signature Prep Kit and the MiSeq FGx Forensic Genomics System the DNASEQEX consortium analyzed the performance of STR markers included in primer mix A. Trimming of adapter sequences, demultiplexing of samples in one pool, quality filtering, alignment, and genotyping of STRs were performed with the UAS software. Similar to the evaluation of the Precision ID GlobalFiler NGS STR Panel concordance, reproducibility, sensitivity, and mixture analyses were performed. The concordance and reproducibility study reference samples were tested in duplicate and compared to CE and between the participating laboratories. Sensitivity studies were performed in the same manner as described above. For mixture studies male–female and male–male mixtures (at ratios of 1:1, 1:5, 1:15, and 1:20, control DNA samples 9947A (Thermo Fisher), 2800M (Promega), and 007 (Thermo Fisher)) were analyzed. Additionally, highly undersatured mixtures (1:100, 1:500, 1:1000) were analyzed to test the limits of detection (manuscript in preparation).

4 Recommendations on STR sequence nomenclature

The adoption of MPS technology in practical forensic work, in which data are to be shared readily, requires an international standardization framework with respect to the nomenclature used in the annotation of each allelic sequence. This nomenclature should be, on the one hand, compatible with the CE-based STR nomenclature used in national DNA databases and population databases (i.e., length-based allele calls), and, on the other hand, should capture all the STR sequence variability and enable future searching of STR sequence data generated by the MPS technology among the different forensic laboratories and databases.

The DNA commission of the ISFG has taken the first step toward this necessary standardization by defining the minimum criteria for MPS-STR sequence data analyses at three hierarchical levels: the full sequence, the alignment of sequences relative to a reference sequence, and the annotation of alleles [8]. The following considerations and recommendations were established:

(1) MPS analysis should be performed with software that allows STR sequences to be exported and stored

in databases as sequence (text) strings to capture the maximum consensus sequence information.

(2) The forward strand direction assigned in the human genome has been constant for all assemblies published since the first draft in 2001 and can be used to align STR sequences.

(3) The choice of reference sequence is crucial for standardizing STR nomenclature systems. At the time of writing, GRCh38 is the most up-to-date sequence assembly and is recommended as the framework with which to define repeat region structure for sequence alignment and for the mapping of sequence features such as SNPs. Software will be required to handle comparisons between multiple reference sequences, particularly in the short term, where sequence variants listed by 1000 Genomes (http://www.internationalgenome.org/1000-genomes-browsers/) currently retain GRCh37 coordinates. Continued discussions are necessary to decide whether or not to adapt to novel genome assemblies.

(4) Further work is needed to translate the nomenclature of STR loci thus far coded relative to the reverse strand and repeat region start and end points. There is a need to strictly define these and other anchor points to specify the repeat regions.

(5) Although simple STR nomenclature systems may be required at some point in the future to facilitate communication and data exchange, comprehensive STR nomenclature systems are preferred for early adopters of STR MPS analysis in order to ensure compatibility with MPS data generated in the future. Backward compatibility to the repeat-based nomenclature derived from CE needs to be maintained to preserve the universal applicability of established national STR databases.

(6) To account for relevant genetic variation outside common repeat regions, STR sequences stored as sequence strings should include flanking sequences as well as the genome coordinates of the sequence read start and end points.

(7) Updated allele frequency databases will be necessary to take full advantage of the increased power of discrimination offered by MPS generated STR data. A unified nomenclature system is needed to ensure compatibility of worldwide population databases.

(8) Future forensic MPS multiplexes would benefit from retention of past markers for backward compatibility and a marker selection process based on population data, molecular biology, sequencing chemistry, and a continued dialogue between the forensic community and commercial suppliers.

The considerations of the ISFG DNA Commission [8] pay special attention to a group of twenty-three forensic STR loci previously aligned relative to the reverse strand (past repeat region sequence), identifying seventeen loci for which a potential frameshift exists when converting to forward strand (future repeat region sequence), and demonstrated potential complications arising from conversion of STR loci to the forward strand by presenting examples at the D19S433, DYS389I/II, and DYS385 a/b loci. Currently, MPS commercial software, such as Converge™ v2.0, perform alignments with the reverse sequence from these 23 STR loci to maintain backward compatibility with previous historical STR length data sets. The ISFG recommendations [8] also illustrated with different STR examples (D18S51, D13S317, and D19S433) the potential difficulties with repeat motif description that can arise from the more detailed characterization of STR sequences that MPS provides, when aligning the sequence generated by MPS to the established repeat motif description of any STR.

Considerations of the ISFG DNA Commission [8] also provide important supplementary reference data including the human genome reference sequence (forward strand, GRCh38 and GRCh37 coordinates) of the repeat regions plus 50 nucleotides of each flanking sequence of 35 autosomal STRs (12 ESS, 20 CODIS markers), 29 Y-STRs, and 7 X-STRs contained nowadays in MPS-STR commercial panels. The SNPs and INDELs currently recorded by 1000 Genomes are also identified in the flanking sequences of STRs. The original file has been expanded, enhanced, and revised as described in Phillips et al. [46] and currently includes 71 autosomal STRs, 48 Y-STRs, 14 X-STRs, and 100 bp of flanking region. The most recent version of this permanently curated and updated STR sequence structure file along with a change log is available at https://strider.online/nomenclature [16].

5 Sequence STR diversity and population studies

Comprehensive MPS-STR datasets from different population samples worldwide will be required to characterize the extent of STR sequence variation for use in STR frequency estimates. The MPS-STR population studies published to date [43, 44, 47–49] showed an increased STR allele diversity when compared with CE length-based analysis. Both internal sequence variations within the repeat region as well as flanking sequence variations (SNPs/INDELs) are the two sources of the additional allele diversity provided by the MPS-STR technology.

One of the most extensive population studies, by Novroski et al. [48], included 777 unrelated individuals from four major population groups (US Caucasian, $N = 210$; Hispanic, $N = 198$; African American, $N = 200$, and Chinese, $N = 169$). An increase in the number of effective alleles at a number of loci was observed due to variation within the repeat region only, the flanking sequence only, or a combination of both. Of the 58 STR markers analyzed, 24 autosomal, 6 X-chromosome, and 14 Y-chromosome loci had an increase in effective alleles greater than 20%. The autosomal loci D2S1338, D12S391, and D21S11 exhibited the largest increase in diversity via sequence variation in the repeat region, while the loci D7S820, D13S317, and D22S1045 exhibited at least a 40% increase in effective number of alleles due to variants within the flanking regions. Only one autosomal locus (TPOX) and eight Y-STR loci (DYS19, DYS389I, DYS391,

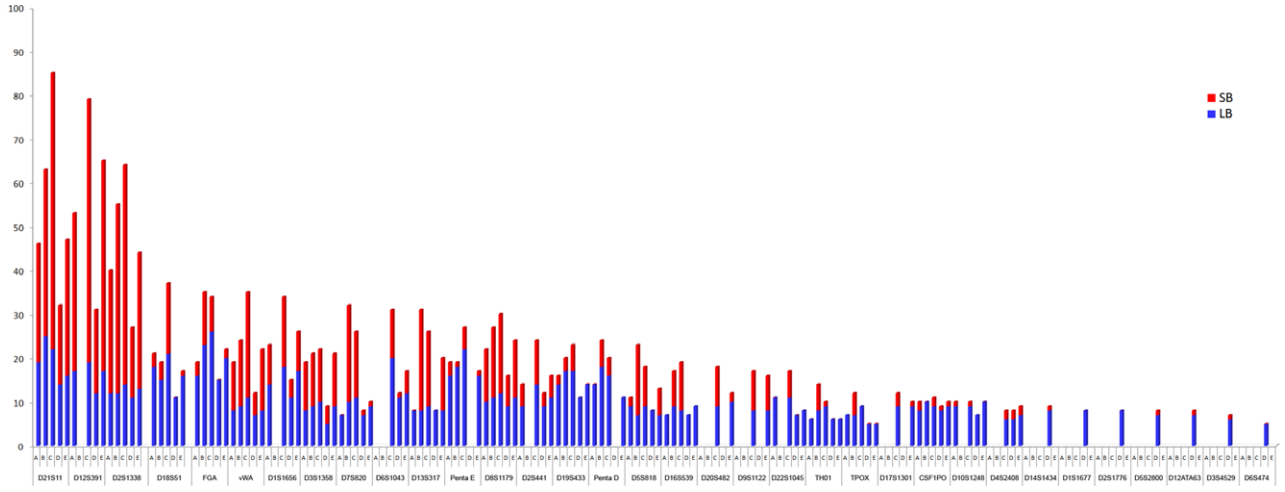

Figure 3. Number of alleles using nominal length-based (LB) as well as the effective increase in observed alleles using sequence-based (SB) variation provided by MPS data for 34 STR loci from the following populations: (A) 183 DNA samples, including African American, Caucasian, and Hispanic individuals [47] (B) 297 DNA samples (101 Dutch samples, 97 samples from Nepal and Bhutan, and 99 Central African Pygmy samples)[43], (C) 777 unrelated individuals from four major population groups (US Caucasian, $N = 210$; Hispanic, $N = 198$; African American, $N = 200$; and East Asian, i.e., Chinese, $N = 169$)[48], (D) 106 unrelated Han Chinese donors [44], and (E) 400 unrelated British individuals (White British, $n = 200$ and British Chinese, $n = 200$)[49].

DYS439, DYS505, DYS549, DYS643, and Y-GATA-H4) did not have any effective increase in alleles by sequencing.

Figure 3 shows the number of alleles using nominal length-based as well as the effective increase in observed alleles using the internal sequence variation provided by MPS data for 34 STRs from five populations studies [43, 44, 47–49].

Another relevant aspect about these first MPS population studies is the high degree of concordance obtained between the STR-MPS profiles and the corresponding STR-CE profiles, which ranged from 99.8 to 100% depending on the STR locus. The reasons for discrepancy were some allele dropouts for certain STR alleles (likely primer-binding variants) and also there were few examples of MPS data that were discordant with operationally defined CE-based data due to INDELs residing in a flanking region or in instances where SNPs reside in the flanking regions immediately proximal to the repeat regions [48]. Software, such as STRait Razor, recalculates a length-based allele incorporating the repeat motif and INDELs in the flanking regions that overcomes inconsistencies due to the latter phenomenon. These findings are good indications that the allele calling data (number of repeat units) provided by the MPS software packages is in high agreement with conventional allele calling provided by CE software expert systems and thus backward compatibility with current database data should not be problematic.

As described above, NIST partnered with the U.S. National Center for Biotechnology Information (NCBI) to launch STRSeq [10] (https://www.ncbi.nlm.nih.gov/bioproject/380127). This resource consists of a curated catalog of sequence diversity at forensic STR loci, conforming to current nomenclature guidelines [8]. The initial data used to populate STRSeq are the aggregate of distinct alleles observed in targeted sequencing studies of single source samples from 4612 individuals from four laboratories: NIST, Kings College London (KCL), University of North Texas Health Science Center–Center for Human Identification (UNTCHI), and University of Santiago de Compostela (USC). This catalog of STR sequences structured in Bio-Projects with stable links to GenBank records contains the following information per allele: the full sequence, the position of the repeat region within the sequence, the position and dbSNP rs number of variations in the flanking regions (when applicable), the subset of sequences that was observed with different commercial assays (when applicable), the bracketed repeat annotation, the sequence technology employed, the minimum threshold of reads observed for the reported sequence, the length-based technology, the given length-based allele, chromosome location, assembly, references, and GenBank accession. STRSeq aims to provide a pathway for submission of newly observed sequence based alleles from laboratories performing population sample studies.

STRidER [16] (https://strider.online/), a publicly available, centrally curated online allele frequency database and quality control platform for autosomal STRs offering reliable STR genotype estimates, has announced plans for storage of nucleotide strings (text strings) in FASTA-like format from population data generated by MPS and to provide software-aided alignment and translation tools between STR nomenclatures. An integrated, seamless process between STRSeq and STRidER has been announced [10] that would strengthen the STRidER quality control function and expand STRSeq, while harmonizing nomenclature between both resources. The quality control aspect of STRidER is of particular importance in forensic genetics, as numerous and diverse errors and pitfalls have been observed by traditional methods with published STR population data and datasets submitted for publication (data not shown). Therefore, the editors of *Forensic Science International Genetics* invited STRidER to perform

quality control of autosomal STR data and require authors to have STR population data quality controlled by STRidER prior to submission of the manuscript to the journal [50].

6 MPS-STR data integration with national DNA databases: NOMAUT

The ISFG recommended simple low-level STR nomenclature systems, that are based on the translation of sequence strings to the operationally defined repeat-based allele designations derived from CE, to make the data directly compatible with those of existing STR databases [8]. In order to capture the additional sequence information, a letter designated nomenclature has been proposed. The use of a simple and CE-compatible nomenclature that collects all sequence variability is also desirable for use in expert reports and in searching local STR databases. However, the transition process will have to be managed by a centralized nomenclature commission to avoid ambiguous or imprecise allele names being adopted, or assigning different names to identical alleles. Given that most countries maintain a national database (historically filled with CE profiles) and still of a continuously increasing size [51], an easy way to compare CE records with MPS results will be needed. Of course, it would be highly unsatisfactory to throw away the MPS information gain (sequence data) and to use only CE-formatted results for comparisons.

The history and naming of human leukocyte antigens (HLA) teaches us how difficult the maintenance of a complex nomenclature catalog can be [52]. When first introduced in 1967 only eight antigens were named (e.g. HL-A LA). Soon the number of new antigens increased rapidly and the centralized naming turned out to be inappropriate (1980s). The World Health Organization (WHO) first had assigned the letter "w" to all antigens that were not yet officially named and abolished it as soon as it had been accepted by the Nomenclature Committee (e.g. HLA-Aw02 and HLA-A2.1). The availability of more modern molecular analysis techniques led to the identification of different alleles within the antigens and highly increased the variation space. As a result, the nomenclature changed from just one or two letters plus one or two numerics to one to four letters, an asterisk and two to eight numerics plus another letter optionally (e.g. HLA-A*92010201L). Because some antigens turned out to have more than the expected 99 alleles and the merge of different levels of granularity in one large numeric (up to eight digits) were very prone to errors and did not reflect the nature the hierarchies, the nomenclature was completely revised in 2010 (e.g. HLA-A*02:101:01:02) [53]. In September 2017, there are almost 4000 different alleles known to the HLA class I gene A (HLA-A). The analogy to the technology-driven infinite increment of STR variation is evident.

To help meet this challenge, the DNASEQEX consortium has proposed the NOMenclature AUThority (NOMAUT) system [54]. NOMAUT was built on a catalog of acquired sequence variants and the ability to grow in a very convenient but safe and robust way. With the catalog as a centralized service, an authoritative "oracle" answering sequence queries with allele calls can be implemented. The basic principle of the catalog is to obtain a compatible CE allele call plus a catalog designation from the MPS sequence allele data. NOMAUT is a self-maintaining system because querying the database will create temporary variants (imminent). Temporary variants (denoted with lowercase letters) become fixed variants (denoted by uppercase letters) by independent observations (new query, new lab). While enacting the underlying nomenclature rules and procedures actually may be rather trivial, maintaining stability, safety, and security on a worldwide scale is very challenging. To ensure reliability and availability NOMAUT was built as a container being easily distributed over web service infrastructures such as Amazon Web Services. NOMAUT is not intended to be desktop software but rather as a service that any (open-source or commercial) software packages for STR-MPS data analysis can include. In this way, consistency, stability, and quality can be maintained on a global scale. In order to use NOMAUT, software producers will need to implement calls against an Application Programme Interface (API) to be published or against an offline version of monthly distributed databases locally.

7 Concluding remarks

Commercial STR-MPS systems developed to date to analyze different sets of autosomal STRs, Y-STRs, and X-STRs have been found to be largely concordant (with previous CE data), reliable, reproducible, and sensitive in several forensic validation studies demonstrating that the STR-MPS technology generally meets forensic DNA validation guidelines [39–45]. The recent STR-MPS population studies have shown two of the fundamental advantages of MPS with respect to conventional CE systems, which are greater multiplexing capabilities and the detection of a large number of new STR sequence variants that increase the overall power of discrimination (and for several currently used loci in particular) [43, 44, 47–49]. Additionally, validation studies have shown that the possibility of using smaller amplicons in MPS compared with CE can provide a more effective analysis of degraded and/or low quantity forensic biological evidence [40].

On the other hand, there are still some limitations of the STR-MPS technology compared to the STR-CE profiling, in addition to the current higher cost of STR-MPS technology: (1) the complex and time-consuming different processes and steps involved in DNA library and DNA template preparation, which make the automation of STR-MPS procedures an indispensable element to guarantee high throughput and reproducibility, (2) the MPS technology requires more powerful bioinformatics tools for the alignment of millions of STR sequences, as well as the availability of MPS data storage servers with higher capacity for the storage of a vast amount of primary and secondary sequence data files, and (3) some limitations to obtain reproducible sequencing results of certain

forensic STRs due to complexities of STR sequence alignments and the current limitation of MPS read length (e.g. SE33) [8]. However, these limitations already are being addressed and appear to be short-termed issues.

Both commercial and open-access software packages developed for STR-MPS data analysis, allow managing STR sequence data rapidly and efficiently and generating allele callings compatible with CE-allele calls. The standardization efforts undertaken by the scientific community [7–10] and scientific collaborations with industry are contributing to the rapid implementation of STR-MPS nomenclature standards supporting the exchange of STR data on a global basis. The development of a worldwide catalogue (as proposed by, e.g., NOMAUT [54] and STRSeq [10]) for the classification and translation of STR-MPS sequence variants into a nomenclature of numbers and letters compatible with the conventional CE nomenclature, will be fundamental to be able to exchange the already identified and new STR-MPS generated data with the historical CE data stored in national STR databases, as well as among sequence-based databases.

Finally, there is high interest demonstrated by a large number of European [1], USA [10], and Asian [44, 55] forensic laboratories in the implementation of MPS technology. These efforts indicate a fast and growing adoption of this new technology for STR genotyping in the forensic genetics workflow.

The DNASeqEx project has been funded with support from the European Commission (grant HOME/2014/ISFP/AG/LAWX/4000007135 under the Internal Security Funding Police programme of the European Commission-Directorate General Justice and Home Affairs).

The authors have declared no conflict of interest.

8 References

[1] Alonso, A., Müller, P., Roewer, L., Willuweit, S., Budowle, B., Parson, W., *Forensic Sci. Int. Genet.* 2017, *29*: e23–e25.

[2] Kidd, K. K., Pakstis, A. J., Speed, W. C., Grigorenko, E. L., Kajuna, S. L., Karoma, N. J., Kungulilo, S., Kim, J. J., Lu, R. B., Odunsi, A., Okonofua, F., Parnas, J., Schulz, L. O., Zhukova, O. V., Kidd, J. R., *Forensic Sci. Int.* 2006, *164*, 20–32.

[3] Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Alvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. V., Carracedo, A., *Forensic Sci. Int. Genet.* 2007, *1*, 273–280.

[4] Kidd, K. K., Speed, W. C., Pakstis, A. J., Furtado, M. R., Fang, R., Madbouly, A., Maiers, M., Middha, M., Friedlaender, F. R., Kidd, J. R., *Forensic Sci. Int. Genet.* 2014. *10*, 23–32.

[5] Kayser, M., *Forensic Sci. Int. Genet.* 2015, *18*, 33–48.

[6] Parson, W., Strobl, C., Huber, G., Zimmermann, B., Gomes, S. M., Souto, L., Fendt, L., Delport, R., Langit, R., Wootton, S., Lagacé, R., Irwin, J., *Forensic Sci. Int. Genet.* 2013, *7*, 543–549.

[7] DNASEQEX. DNA-STR Massive Sequencing & International Information Exchange. https://dnadatabank.forensischinstituut.nl/binaries/dnaseqex-letter-160531_tcm127-629975_tcm37-209493.pdf

[8] Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., Gusmão, L., Hares, D. R., Irwin, J. A., King, J. L., Knijff, P., Morling, N., Prinz, M., Schneider, P. M., Neste, C. V., Willuweit, S., Phillips, C., *Forensic Sci. Int. Genet.* 2016, *22*, 54–63.

[9] Scientific Working Group on DNA Analysis Methods. Validation Guidelines for DNA Analysis Methods. https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf

[10] Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., Vallone, P. M., *Forensic Sci. Int. Genet.* 2017, *31*, 111–117.

[11] King, J. L., Wendt, F. R., Sun, J., Budowle, B., *Forensic Sci. Int. Genet.* 2017 *29*, 21–28.

[12] Woerner, A. E., King, J. L., Budowle, B., *Forensic Sci. Int. Genet.* 2017, *30*, 18–23.

[13] Friis, S. L., Buchard, A., Rockenbauer, E., Børsting, C., Morling, N., *Forensic Sci. Int. Genet.* 2016, *21*, 68–75.

[14] Van Neste, C., Gansemans, Y., De Coninck, D., Van Hoofstat, D., Van Criekinge, W., Deforce, D., Van Nieuwerburgh, F., *Forensic Sci. Int. Genet.* 2015, *15*, 2–7.

[15] Parson, W., Dur, A., *Forensic Sci. Int. Genet.* 2007, *1*, 88–92.

[16] Bodner, M., Bastisch, I., Butler, J. M., Fimmers, R., Gill, P., Gusmão, L., Morling, N., Phillips, C., Prinz, M., Schneider, P. M., Parson, W., *Forensic Sci. Int. Genet.* 2016, *24*, 97–102.

[17] ForenSeq™ DNA Signature Prep Reference Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/forenseq/forenseq-dna-signature-prep-guide-15049528-01.pdf

[18] Massively Parallel Sequencing for Forensic DNA Analysis. https://www.promega.com/-/media/files/promega-worldwide/north-america/promega-us/webinars-and-events/2015/massively-parallel-sequencing-for-forensic-dna-analysis.pdf?la=en

[19] Precision ID GlobalFiler™ NGS STR Panel v2 with the Ion S5™ System. Application guide. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0016129_PrecisionIDSTRIonS5_UG.pdf

[20] Promega, PowerSeqSystems Prototype Protocol, 2016.

[21] Illumina, *TruSeq DNA PCR-Free Library Prep - Reference Guide*, Illumina Inc, San Diego, CA, USA 2015.

[22] Illumina, *An introduction to Next-Generation Sequencing Technology*, Illumina, San Diego 2017 Pub. No. 770-2012-008-B, (Available at: https://emea.illumina.com/science/technology/next-generation-sequencing.html?langsel=/at/).

[23] Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost,

T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M. J., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M. D., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Catenazzi, M. C. E., Chang, S., Cooley, R. N., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fajardo, K. V. F., Furey, W. S., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., Hoschler, K., Hurwitz, S., Ivanov, D. V., Johnson, M. Q., James, T., Jones, T. A. H., Kang, G.-D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ng, B. L., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Pike, A. C., Pinkard, D. C., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Rodriguez, A. C., Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Sohna, J. E. S., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., vandeVondele, S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Ya, J. N., L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, A. J. Smith, *Nature* 2008, *456*, 53–59.

[24] Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., Jaffe, D. B., *Genome Biol.* 2013, *14*, R51.

[25] Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., Gu, Y., *BMC Genomics* 2012, *13*, 341.

[26] Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., *J. Biomed. Biotechnol.* 2012, *2012*, 251364.

[27] Forenseq universal analysis software guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/forenseq-universal-analysis-software/forenseq-universal-analysis-software-guide-15053876-01.pdf

[28] Torrent Suite™ Software 5. Help. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0016409_TorrentSuite5_4Help.pdf

[29] Converge software. https://assets.thermofisher.com/TFS-Assets/LSG/Product-Bulletins/Converge%20Software%20Product%20Bulletin.pdf

[30] HID STR Genotyper Plugin user guide https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0015879_HID_STR_Genotyper_Plugin_UG.pdf

[31] NextGENe user manual. http://www.softgenetics.com/PDF/NextGENe_UsersManual_web.pdf

[32] Battelle ExactID. https://www.battelle.org/government-offerings/homeland-security-public-safety/security-law-enforcement/forensic-genomics/exactid

[33] Ganschow, S., Wiegand, P., Tiemann, C., *Forensic Sci. Int. Genet. Supplement Ser.* 2017, *6*, e119–e121.

[34] Gymrek, M., Golan, D., Rosset, S., Erlich, Y., *Genome Res.* 2012, *22*, 1154–1162.

[35] Anvar, S. Y., van der Gaag, K. J., van der Heijden, J. W., Veltrop, M. H., Vossen, R. H., de Leeuw, R. H., Breukel, C., Buermans, H. P., Verbeek, J. S., de Knijff, P., den Dunnen, J. T., Laros, J. F., *Bioinformatics* 2014, *30*, 1651–1659.

[36] Hoogenboom, J., van der Gaag, K. J., de Leeuw, R. H., Sijen, T., de Knijff, P., Laros, J. F., *Forensic Sci. Int. Genet.* 2017, *27*, 27–40.

[37] Highnam, G., Franck, C., Martin, A., Stephens, C., Puthige, A., Mittelman, D., *Nucleic. Acids Res.* 2013, *41*, e32.

[38] Lee, J. C., Tseng, B., Chang, L. K., Linacre, A., *Forensic Sci. Int. Genet.* 2017, *26*, 66–69.

[39] Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J., Varlaro, J., Stephens, K. M., Holt, C. L., *Forensic Sci. Int. Genet.* 2017, *28*, 52–70.

[40] Churchill, J. D., Schmedes, S. E., King, J. L., Budowle, B., *Forensic Sci. Int. Genet.* 2016, *20*, 20–29.

[41] Just, R. S., Moreno, L. I., Smerick, J. B., Irwin, J. A., *Forensic Sci. Int. Genet.* 2017, *28*, 1–9.

[42] Zeng, X., King, J., Hermanson, S., Patel, J., Storts, D. R., Budowle, B., *Forensic Sci. Int. Genet.* 2015, *19*, 172–179.

[43] van der Gaag, K. J., de Leeuw, R. H., Hoogenboom, J., Patel, J., Storts, D. R., Laros, J. F. J., de Knijff, P., *Forensic Sci. Int. Genet.* 2016, *24*, 86–96.

[44] Wang, Z., Zhou, D., Wang, H., Jia, Z., Liu, J., Qian, X., Li, C., Hou, Y., *Forensic Sci. Int. Genet.* 2017, *31*, 126–134.

[45] Barrio, P., Martin, P., Alonso, A., DNASEQEX Project: Preliminary results of STR markers typing by Massively Parallel Sequencing for Forensic Use. https://www.researchgate.net/publication/318753610_DNASEQEX_Project_Preliminary_results_of_STR_markers_typing_by_Massively_Parallel_Sequencing_MPS_for_Forensic_Use

[46] Phillips, C., Gettings, K. B., King, J. L., Ballard, D., Bodner, M., Borsuk, L., Parson, W., *Forensic Sci. Int. Genet.* 2018, *34*, 162–169.

[47] Gettings, K. B., Kiesler, K. M., Faith, S. A., Montano, E., Baker, C. H., Young, B. A., Guerrieri, R. A., Vallone, P. M., *Forensic Sci. Int. Genet.* 2016, *21*, 15–21.

[48] Novroski, N. M. M., King, J. L., Churchill, J. D., Seah, L. H., Budowle, B., *Forensic Sci. Int. Genet.* 2016, *25*, 214–226.

[49] Devesse, L., Ballard, D., Davenport, L., Riethorst, I., Mason-Buck, G., Syndercombe Court, D., *Forensic Sci. Int. Genet.* 2017, 34, 88–96.

[50] Gusmão, L., Butler, J. M., Linacre, A., Parson, W., Roewer, L., Schneider, P. M., Carracedo, A., *Forensic Sci. Int. Genet.* 2017, *30*, 160–163.

[51] Walsh, S. J., Buckleton, J. S., Ribaux, O., Roux, C., Raymond, T., *Forensic Sci. Int. Genet. Supplement Ser. I* 2008, *1*, 667–668.

[52] Nomenclature for Factors of the HLA System, 2017. http://hla.alleles.org/nomenclature/nomenc_reports.html

[53] Marsh, S. G., Albert, E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., Fernández-Viña, M., Geraghty, D. E., Holdsworth, R., Hurley, C. K., Lau, M., Lee, K. W., Mach, B., Maiers, M., Mayr, W. R., Müller, C. R., Parham, P., Petersdorf, E. W., Sasazuki, T., Strominger, J. L., Svejgaard, A., Terasaki, P. I., Tiercy, J. M., Trowsdale, J., *Bone Marrow Transplant.* 2010, *45*, 846–848.

[54] Willuweit, S., Challenges and Paradigm Shifts by the Adoption of MPS in Forensic Casework. https://www.researchgate.net/publication/318421173_Challenges_and_Paradigm_Shifts_by_the_Adoption_of_MPS_in_Forensic_Casework_-_Lessons_Learned_from_the_Collaborative_DNASeqEx_Project_so_far

[55] Guo, F., Yu, J., Zhang, L., Li, J., *Forensic Sci. Int. Genet.* 2017, *31*, 135–148.

