Forensic Science International: Genetics 38 (2019) 175–180

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Review article

From next generation sequencing to now generation sequencing in forensics

Peter de Knijff

Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands

ARTICLE INFO

Keywords:
Short tandem repeat (STR)
Massively parallel sequencing (MPS)
Next generation seqeuncing (NGS)

ABSTRACT

In contrast to genetic diagnostic disciplines such as Oncogenetics and Cinical Genetics, where worldwide, since 2010, tens of thousands of DNA samples are routinely screened annually using either targeted genome sequencing or whole genome sequencing using massively parallel sequencing (MPS), the forensic use of MPS is still far from being a routine diagnostic tool. This perspectives focusses on issues that are essential in order to fully understand (i) why MPS of short tandem repeats (STRs) is very different from the capillary electrophoresis (CE) based genotyping of STRs, (ii) what we, DNA experts, should know before explaining MPS-based evidence in court, and (iii) what information should be present in a forensic investigation report that is MPS-based. Here one has to keep in mind that the forensic use of CE was first introduced in 1992–1993 and that it took some time to fully appreciate all intricacies. Obviously, I might be biased in my opinion, having worked on this topic since 2008, but I sincerely hope that MPS will soon be widely accepted and used because, especially in case of mixed-source DNA samples, MPS is much better in the deconvolution of the individual contributors and invariably reveals genetic information that cannot be inferred otherwise.

1. Scope of this paper

In the context of this article August 31, 2011 can be seen as a historical day for the forensic use of next generation sequencing (NGS) or as it is most commonly referred to, massively parallel sequencing (MPS). In a rather hot Vienna, Austria, this day saw the first formal plenary session of the 24th ISFG conference containing three presentations specifically dealing with the forensic application of MPS. Six Years later, during the 27th ISFG meeting in Seoul, South Korea, at least 19 out of 60 (or 32%) oral presentations dealt with MPS related topics. This at least suggests that the forensic use of MPS became a routine diagnostic tool. However, as we speak (unless I am very ill informed), few, if any, forensic investigations involving MPS-based results have been presented in Court. This indicates that MPS is still far from being routinely used in forensics. This is in sharp contrast with other genetic diagnostic disciplines such as Oncogenetics and Cinical Genetics, where worldwide, since 2010, tens of thousands of DNA samples are routinely screened annually, using either targeted genome sequencing or whole genome sequencing using MPS. For this, there are, of course, a number of mitigating factors. First, these other genetic disciplines were never hampered by the availability of invariably minute amounts of degraded DNA and were already used to screen DNA samples for genetic variants at the whole genome level first by means of single nucleotide polymorphism (SNP) array platforms for at least 20 Years [1], using bioinformatics tools to interpret their results. Subsequently, it was relatively small step for these laboratories to enter into the even more complex MPS-era [2,3]. They were also quick to act in terms of ethical guidelines [4] and could rapidly benefit from the already well established and authoritative human genome variation society nomenclature guidelines [5].

In a strict sense, this is not be an extensive review, but a perspectives paper in which I will explore the various reasons why, in forensics, MPS is still not “now”, and what could (or should?) be done to make it “now”. I will not discuss the technical background of the various MPS platforms and methodologies. For that, there are excellent review papers [6,7]. I will further restrict myself to the use of MPS for short tandem repeats - STRs – as other papers in this special issue deal with other genetic MPS targets, including mitochondrial DNA and autosomal microhaplotypes. I will use examples from my own – Dutch – forensic laboratory, where we use exclusively MPS-STR kits from Promega and sequence the PCR product on the Illumina MiSeq platform. Obviously, I would say, I am aware that there are other MPS kits and alternative MPS sequence platforms that perform equally efficient, produce equally complex sequence reads data sets, and are also faced with equally challenging data interpretations issues.

This perspectives is structured according to issues that are essential in order to fully understand (a) why MPS of STRs is very different from the current golden standard: capillary electrophoresis (CE) of STRs, routinely in use since 1995 [8]; (b) what we, DNA experts, if asked, need to be able to explain in Court when presenting MPS-based

E-mail address: Knijff@lumc.nl.

https://doi.org/10.1016/j.fsigen.2018.10.017
Received 12 July 2018; Received in revised form 25 October 2018; Accepted 29 October 2018
Available online 03 November 2018
1872-4973/

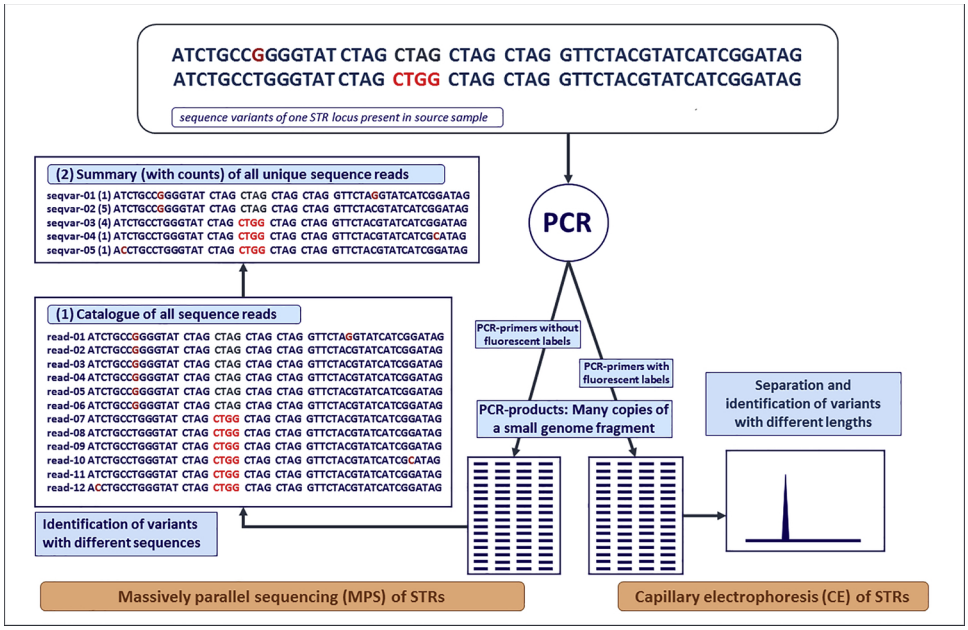

Fig. 1. A simplified explanation of the main differences between CE genotyping of STRs and MPS of STRs.
In this theoretical example, a PCR is designed to reveal the genetic variation of a single STR locus. The DNA sample has two different alleles, both four repeats long. One allele contains a CTGG instead of a CTAG repeat unit. The other allele contains a SNP in the 5′flanking region, a T > G mutation 7 bp. prior to the first repeat unit. When using CE, the PCR uses a set of PCR primers of which one primer has a fluorescent label. After PCR, all PCR products will include this fluorochrome, enabling the detecting of the PCR products on a CE platform. As both alleles are 4 repeats long, only a single peak will be visible. For this locus, this DNA sample will be called a homozygote when using CE. When using MPS, the PCR primers are not labeled with a fluorochrome. All PCR products are simply sequenced on a MPS platform, resulting in a long list of sequence reads. For the sake of simplicity, in this example 12 sequence reads are shown (panel 1). Six reads (01 – 06) contain the T > G mutation in the 5′flanking region. Six other reads (07–12) contain the CTGG repeat. Upon further inspection, three more genetic variants are detected. A single read (01) contains a C > G error 7 bp. 3′of the repeat structure. Another read [10] contains a G > C error 16 bp. 3′of the repeat structure. Finally, read 12 shows a T > C error at the second position of the sequence. This full spectrum of sequence variation is summarized in panel 2. A total of five different sequence variants were detected in this sample. Two variants (02 and 03) were seen multiple times and probably reflect the true alleles. A further three variants (01, 04, and 05) are seen only once; their defining SNP likely representing error reads: reads containing PCR or sequence induced sequence errors.

evidence, and, (c) what information should be present in a MPS-based forensic investigation report. I will also touch upon the issue of storing MPS-based STR genotypes in National/International STR-profile databases. In the following I will frequently describe two different sources of genetic variation: mutation and error. I will use mutation when describing a genetic variant that is carried by the donor of the DNA sample, whereas I will use error to indicate any genetic variant that was caused by technical (predominantly PCR-induced and sequence-induced) inadequacies when processing and analyzing the DNA molecules. When simply sequencing (especially mixed) DNA samples it will not always enable one to differentiate between the two, but one might be in a position to make a (well educated) guess.

2. The difference between the use of MPS and CE to explore STR variation

The most obvious, and in many aspects also the most important difference between MPS results and CE results is the difference in the measured outcome of both technologies (see Fig. 1). In all its seemingly simplicity, for the purpose of STR genotyping, CE translates machine-measured DNA-molecule migration times into DNA fragment lengths [9–11] which, to further aid interpretation, are visualized in peak-profiles and tables with a very simple string of numbers representing these fragment lengths. However, CE does not provide information about the underlying base pair variation of the DNA sample that is studied. This has a major consequence: CE analysis of STRs

underestimates the underlying genetic variation present in the DNA sample. Homoplasy, similar sized DNA fragments with different sequence compositions which display identical fragment sizes in CE, is well known, but cannot be detected. With MPS, irrespective of the underlying sequence technology, the final experimental result is represented as DNA sequences that reveal all underlying sequence variation in the targeted DNA sample. These DNA sequences can be translated, in the case of STRs, into DNA fragment lengths but this is not strictly necessary, unless one wants to compare MPS STR results with CE STR results. Homoplasy is no longer a problem. I would say the opposite: with MPS based detection of STRs, revealing homoplasy is one of the major strengths of this technology.

This fundamental difference in experimental design has two practical consequences, especially when one has to explain STR results in court. First, when CE based, the only annoying experimental error one frequently encounters are “stutter” alleles. These are caused by slippage during DNA-replication in vivo and/or in vitro during a PCR reaction [12,13]. In DNA samples from a single person, genuine alleles and stutter alleles can be easily distinguished. However, the analysis of unbalanced mixtures with low minor contributions is frequently complicated by stutter alleles that cannot be distinguished from genuine alleles of the minor contributors [14]. Furthermore, what is also not revealed by CE analysis of STRs are the erroneous base pair substitutions, mainly due to DNA editing errors occurring during PCR [15], as these do not influence the underlying fragment lengths. In this respect, MPS reveals the entire spectrum of errors: (i) stutters caused by DNA slippage during PCR, (ii) base pair errors due to DNA editing errors during PCR, (iii) strand slippage (mainly at homopolymer stretches) during sequencing, and (iv) base pair errors caused by substitution type miscalls during sequencing [16]. There is sufficient evidence to assume that the latter two sources of error are sequence platform dependent [17,18]. Second, results of CE analysis of STRs are translated into a very simple data file that, essentially, only contains STR locus names, the length(s) of STR alleles, and the fluorescent intensities (or peak-heights) detected by the CE platform. These results can be visualized in easy to explain (but not necessarily easy to understand) peak-profiles, which have been in use for over 20 Years (see Figs. 1 and 2). Results of an MPS experiment can be stored in two relatively simple file formats, each representing all individual DNA sequence reads produced during the MPS analysis: FASTQ [19] and / or FASTA [20]. However, as MPS platforms produce between several tens of millions to many hundreds of millions of sequence reads in a single experiment, one has to rely on specially designed software that translates these millions of reads into an experimental summary that one can understand and present [21–23]. Furthermore, as these DNA sequence files contain all reads produced by the platform, i.e. those representing true alleles and those containing any kind of PCR/sequence error, the MPS software packages used to interpret, summarize and visualize all sequence data need to be

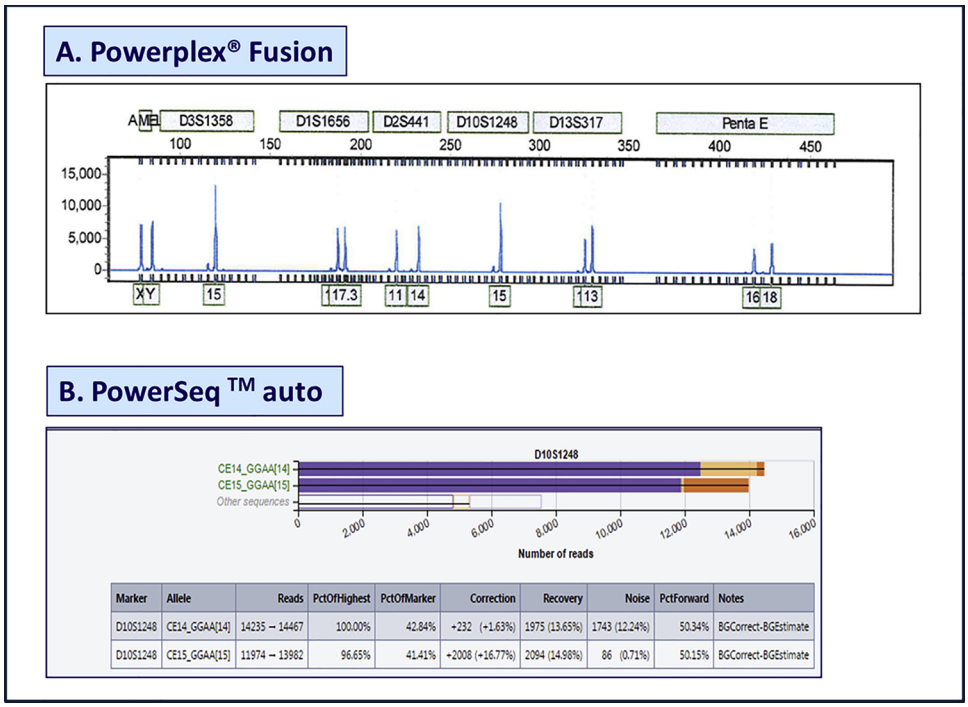

Fig. 2. Graphical outputs of CE detection of STRs and the FDSTools summary of MPS of STRs.
Shown are partial graphical outputs (of two different DNA samples) reflecting the results of CE-based STR genotyping (panel A) and MPS-based STR genotyping (panel B) as produced with FDSTools. As shown in panel A, the electropherogram visualizes the genetic variation as peaks with their heights measured in rfu's (relative fluorescent units), and the alleles designated in terms of number of repeats below each allele. In the case of MPS, for each allele the number of reads with a specific genetic variation is shown as a bar and further specified in a table. For more information see reference [23].

able to distinguish between "chaff and wheat" in such a way that, preferably, one can always, in retrospect, go back to the original sequence data and explore them in alternative ways if requested. In contrast to the CE analysis of STRs, with MPS there is no longer a golden standard with respect to the platform and software. Therefore, it is more difficult to directly compare MPS results among platforms, laboratories and forensic DNA experts. And, perhaps even more important, in order to translate MPS STR results to a format that one can compare with CE STR results, one needs new nomenclature rules that, preferably, have maximum clarity. In short, when used to study the STR variation in a crime-scene sample, CE only informs about DNA fragment length variation, whereas MPS results represent the full spectrum of possible DNA sequence variation. One might be tempted to translate complex MPS STR results into something that is similar to CE STR results, but by doing so, one ignores all additional genetic variation information that might by crucial for a criminal investigation.

2.1. What we should understand and be able to explain when presenting MPS-based evidence

In order to analyze, interpret, and summarize forensic MPS-STR results my lab uses FDSTools, an

open-source software solution that was developed specifically for this purpose [23]. What we had in mind was a tool that would enable us not only to process all raw MPS data, but also to have a simple and portable tool that would allow a DNA expert to summarize, visualize, and explain almost all individual DNA sequences in a flexible fashion in court. When MPS is used, STRs are evaluated as sequence variants that each has particular stutter characteristics which can be precisely determined.

FDSTools uses a database of reference samples to determine stutter and other systemic PCR or sequencing artefacts for each individual allele. In addition, stutter models are created for each repeating element in order to predict stutter artefacts for alleles that are not included in the reference set. This information is subsequently used to recognize and compensate for the noise in a sequence profile. The result is a better representation of the true composition of a sample (Fig. 2).

Since, as stated before, MPS of STRs not only reveals all genuine sequence variants, but also sequences representing the entire spectrum of PCR errors and sequence errors, it is crucial, for a proper interpretation of the results of each experiment, to understand the experimental error-profile of the complete analytical procedure. Where with CE, one needs "only" to keep track of issues as peak-bleed through, peak-shifts, allele-imbalance, unusual high stutters, and new alleles at unexpected locations in the STR profile, with MPS one has to explore the sequence variation among many millions of individual sequence reads. This is best explained with a simple example (Fig. 3). As with CE, where one routinely only considers peaks above a certain fluorescent intensity detection threshold (say 50 rfu's) as genuine peaks, with MPS, at least when using FDSTools, one also has to set an analytical threshold – AT -, in this case the number reads with an identical sequence structure. It strongly depends on the experimental design. If one has pooled many different DNA samples for database purposes into a single MPS run, one expects less reads per sample and per locus (in case of a multiplex STR design), compared to a run with only a few case samples pooled. In our example, taken from our reference population screening we expected roughly 12,000 reads per allele. In this case, (see panel A in Fig. 3), we obtained close to 13,000 reads for CSF1PO allele 11 and over 16,000 reads for allele 10 (for a full explanation how to understand and read this figure see reference [23]. We also detected close to 5000 CSF1PO reads that were labelled by FDSTools as "other sequences". All these sequence reads, since that is how FDSTools works, are sequence reads that align with the full CSF1PO reference sequence but lack the abundancy to pass user-specified thresholds such as coverage or percentage of total locus reads. Whether or not these sequences are visualized as distinct bars in the graphical output depends on the settings of the AT. In panel A of Fig. 3, I put the AT at 12 (or 1/1,000 of the number of reads expected for genuine alleles). As a consequence only four allelic variants were displayed: alleles 10 and 11, an allele 12 (probably representing a +1 forward stutter and categorized as "noise") and the "other sequences". If one reduces the AT to its minimum value, 2, one gets a very different summary (see panel B of Fig. 3). All unique sequences with a minimum of 2 reads are now visualized. I only show a few of these here. All additional unique sequences shown have an identical PCR amplicon length of allele 10 (no. 1 in panel B of Fig. 3), but show the full spectrum of possible genetic variation, varying from purely alternative repeat motifs (no. 2 and 3) to single nucleotide polymorphisms - SNPs – (no. 4, but also others in

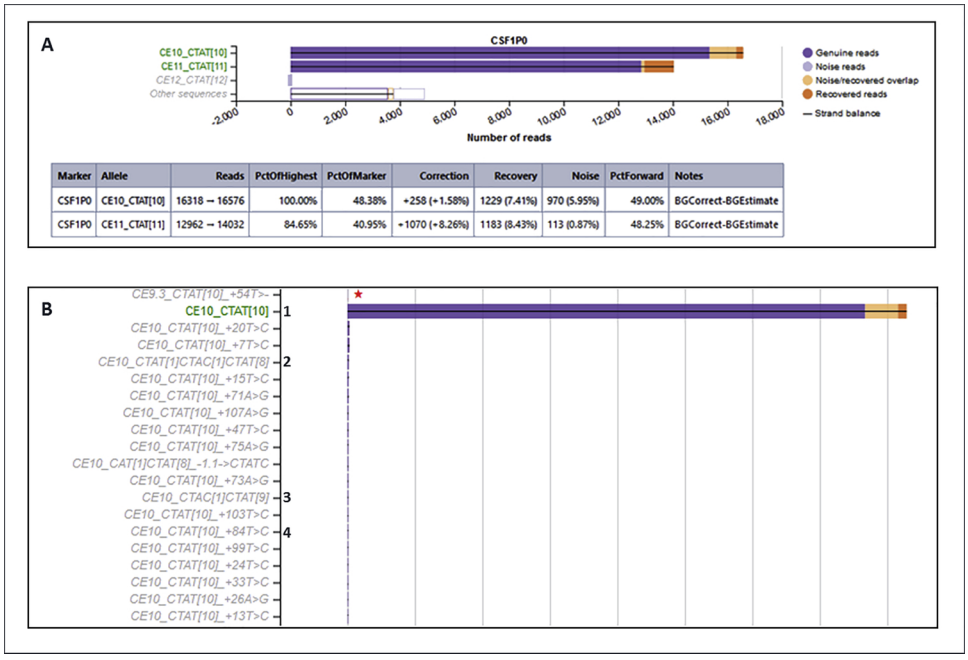

Fig. 3. MPS of STRs without and with error reads revealed.
Panel A shows the output of a single locus (CSF1PO), with the baseline threshold at 12 reads, meaning that only sequence variants with 13 or more reads are visualized and all other reads summarized as "other sequences". Panel B shows the result of the same sample, but now with the baseline threshold set at 2. This reveals a lot more unique sequence structures, all with 12 or less reads, probably reflecting the wide spectrum of PCR and sequence errors. Note that, because of the long list of allele 10 sequence variants, cut-short for graphical purposes, allele 11 is missing from this panel.

panel B of Fig. 3), likely representing error reads: reads containing PCR or sequence induced sequence errors.

Exactly how many error-reads are produced during MPS of STRs, and if certain error patterns are more frequent, is still unknown. We know very little about this issue (but see [16,24]. What we do know is that sequence error patterns can be MPS platform specific (perhaps even individual machine specific) and, as we have learned from ancient DNA research, some sequence errors might be caused by DNA degradation. In addition, we also keep track of the forward / reverse sequence read balance - FRB -. Alleles passing the AT but showing a marked skewed FRB are labelled with a red * (see panel B in Fig. 3). Too many alleles showing a skewed FRB could indicate erroneous MPS conditions. It is important to be aware that many of these error reads are also produced during the CE analysis of STRs but, because their amplicon lengths are identical, they are simply part of the "sequence" peaks that form the "true" alleles in a CE electropherogram. As with CE, each individual laboratory should study the full spectrum of MPS related error issues themselves, as these will vary among MPS platforms, MPS kits, and, more importantly, by the DNA fragments that are targeted. Clear and concise guidelines are not (yet) available.

3. What information (and how) should be present in a MPS-based forensic investigation report

Since with MPS one obtains the full sequence structure of the DNA fragments that were targeted, there will be legal restrictions in many Countries that could prevent the reporting scientists from including the all MPS results of a forensic investigation in the form of e.g. FASTA file or the full description of all the sequence variation. As an example, in the Netherlands, it is legally not allowed to include, in the report, any exact information of the genetic variation underlying the forensic investigation results. Obviously, these genetic data are available and added to the complete investigation file, but both the prosecutor and the defence has to submit a special request to obtain these underlying data. In the case of CE, this additional information is usually in the form of the peak-profiles or electropherograms, and or tables that summarizes all underlying STR data. For MPS, this is, as explained above, a bit more complex. For the time being, that is, until a number of MPS-based forensic DNA investigations have been tried and tested in Dutch Court, we include the MPS-STR results (as pdf-file) in the full case-file (but not in the report) in the form of summarizing bar graphs and tables such as shown in panel B of Fig. 2 and A in Fig. 3. Furthermore, there are legal restrictions which respect to the specific type of locus one can use (and report) in any forensic DNA investigation. All underlying data as FASTQ and FASTA files, including all error reads, an error read summary and other relevant data informing about the quality of the MPS experiment, is only stored digitally and is only made available upon special request. In the case of explaining MPS-STR results in Court, all data, after processing with FDSTools, is available as HTML-format files that can be shown and discussed with the help of a graphical user interface and any browser. I always transfer these HTML files to my IPAD and simply use Safari as the browser. This is a very safe, flexible and stable solution and can be done on any stand-alone computer.

In addition to the genetic data (being error reads or reads reflecting genuine alleles) there is quite a lot of additional information that should be stored. Since a single run on an MPS platform produces millions of sequence reads, it is customary to pool different DNA samples to be sequenced together. Exactly how many one can pool strongly depends on the required number of sequence reads per sample and per allele. This raises the opportunity for a flexible case-by-case and/or sample-by-sample experimental approach. Exactly how many samples one wishes to analyse in a single run is entirely up to the user, although some companies of sequence platforms and sequencing kits do give general guidelines. Obviously, for the purpose of database or reference samples, one can pool much more samples into a single experiment whereas for some mixed-source crime-scene samples with very skewed minor contribution (say 1%–5%) one might use a more substantial part of the sequence run. This aspect of MPS-STR based genotyping clearly needs more -concerted - testing. In the case of my laboratory, as a rule of thumb we aim at a minimum of 100 reads per locus for a single-source reference or database sample and a minimum of 50–100 reads for the lowest contributing allele per locus in a mixture. If one indeed pools different samples into a single MPS run, one has to keep track of the individual barcodes that are used to uniquely label all PCR products of a single DNA sample. For this, again, there are no recommendations or guidelines, and one has, for the time being, to use common sense.

4. Storing MPS-based STR genotypes in National/International STR-profile databases

Storing the results of CE genotyping of STRs in any kind of STR profile database is extremely simple. The sample code, the locus name, and the genotyping result are the three default parameters that can be entered as very simple and short text-strings. Additional information such as individual peak heights and the multiplex STR kit use for creating the profile can be useful to add. For CE, there is a globally accepted uniform standardized allele-calling nomenclature, and companies selling multiplex STR genotyping kits take these standards into consideration. For MPS, there is no standardization at all. One can perform MPS on different platforms. There are no standard nomenclature guidelines yet, although there have been a few initiatives recently [25,26]. Even with a standardized nomenclature it would still be very difficult to squeeze the full spectrum of sequence variation of any MPS identified STR allele into a short and simple text string unless one decides to use an allele code system such as used for the HLA system since 1968 [27]. The main advantage of an allele code system such as with HLA is that the STR sequence result can be recoded as a very short allele designator. I see, however, two major disadvantages of such a STR sequence code. First, one needs a forensic STR nomenclate authority for something (describing sequence variation) that is already in place for a long time [5], and second, with a code one loses the immediate and direct link with the CE-based STR nomenclature. There are no logical and certainly no bioinformatics or ICT reasons why a database should not be able to use a relatively long text string (of say 50–100 characters) as allele designators, which is no unrealistic requirement [28]. Thus, I hope that eventually there will be a general agreement for a fully transparent and fully informative forensic STR uniform nomenclature system one can use for (the automated) calling MPS based STR alleles. At present, at least in The Netherlands, MPS-based STR alleles are recoded to their CE alternative and stored in the National DNA-database with “MPS” as additional remark to flag the availability of more information than simply the CE-based number of repeats.

5. Recommendations and still to do

It will probably be clear by now that before MPS will be, as CE, a routine forensic genetic diagnostic tool, there is still a lot to do. Predominantly this involves a full set of recommendations or guidelines suggesting criteria for all possible technological, interpretive, and reporting issues. Furthermore a number of practical issues need to be solved, including accommodating the various National DNA databases to accept STR alleles identified by means of MPS including the full spectrum of genetic variation detected. Since, in contrast to CE, there is a wide array of MPS platforms and software’s enabling MPS results, developing such recommendations will be more complex as they have to include a much wider spectrum of issues. In my view, these should involve, in random order, at least the following issues:

1 A uniform nomenclature for MPS based STR alleles that allows reconstructing/understanding the full spectrum of genetic variation without the need of back referral, such as in the case of the HLA nomenclature system.
2 Recommendations concerning the minimum number of reads that are required to reliably call an STR allele under various conditions, i.e. a simple reference database sample, a single source crime scene sample or a mixed-source crime-scene sample.
3 Recommendations that can be used to provide information about the full spectrum of non-target (or error) reads. This can also include revealing information about the sample-pool strategy that was used to screen the samples, and documentation of barcode strategies.
4 Recommendations about the MPS strategy that was used. Were the full reads 1 and reads 2 sequenced forward and reverse and subsequently assembled and aligned, or was a less inclusive sequence strategy used? Was the full length of the PCR amplicon sequenced, the length being dependent on the MPS platform or were partially sequenced amplicons assembled aligned?
5 Recommendations about formats necessary to store all MPS results.
6 Minimum requirements of software used to analyses and summarize MPS results. What information should be immediately available?
7 Statistical software packages that were developed for the interpretation of the evidential value of CE based match between STR profiles need to be adjusted to the new allele designations provided by MPS experiments.

This list of issues is far from complete. I know of at least two other sources of error that also might influence the interpretation of MPS-based STR genotyping and for which we even know less in the context of the forensic use of MPS: the formation of chimera molecules during PCR [29] and jumping tags during the sequence library preparation [30]. For all of these issues we still lack sufficient and reliable empirical data. Hence, it might be wise to set up concerted actions / collaborations to gather sufficient data.

The less initiated reader might wonder why she or he should even consider starting using MPS. With such a list of issues still to solve, and confronted with a choice for MPS platforms, a choice of MPS interpretation software’s, and only few commercially available MPS-kits, it seems, at present, prematurely to invest in this new technology. My answer to this reluctance is fairly simple. If hundreds of laboratories worldwide are already using MPS to screen DNA samples for Oncogenetic and/or Clinical Genetic diagnostics, why should we not do the same? As far as I am concerned, it is about time to get a final grip on such issues since, as stated before; MPS is, especially in case of mixed-source DNA samples (and who has had no problems with these using CE?), much better in the deconvolution of the individual contributors and reveals genetic information that cannot be inferred otherwise. Be brave, and you will be rewarded!

Competing interests

The author declares no conflict of interest.

Acknowledgements

I wish to acknowledge the unconditional contribution of my full laboratory staff during the past 10 Years of MPS technology development. In this, I also wish to include Kristiaan van der Gaag en Jerry Hoogenboom, who are now employed elsewhere, but who have played (and still do) a vital role, while working in my lab, in introducing MPS in the Dutch criminal investigation infrastructure.

References

[1] T. LaFramboise, Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances, Nucleic Acids Res. 37 (2009) 4181–4193.
[2] D.C. Koboldt, K. Meltz Steinberg, D.E. Larson, R.K. Wilson, E.R. Mardis, The next-generation sequencing revolution and its impact on genomics, Cell 155 (2013) 27–38.
[3] C. Di Resta, S. Galbiati, P. Carrera, M. Ferrari, Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities, EJIFCC 29 (2018) 4–14.
[4] T. Caulfield, A.L. McGuire, M. Cho, J.A. Buchanan, M.M. Burgess, U. Danilczyk, C.M. Diaz, K. Fryer-Edwards, S.K. Green, M.A. Hodosh, E.T. Juengst, J. Kaye, L. Kedes, B.M. Knoppers, T. Lemmens, E.M. Meslin, J. Murphy, R.L. Nussbaum, M. Otlowski, D. Pullman, P.N. Ray, J. Sugarman, M. Timmons, Research ethics recommendations for whole-genome research: consensus statement, PLoS Biol. 6 (2008) e73.
[5] J.T. den Dunnen, R. Dalgleish, D.R. Maglott, R.K. Hart, M.S. Greenblatt, J. McGowan-Jordan, A.-F. Roux, T. Smith, S.E. Antonarakis, P.E.M. Taschner, On behalf of the human genome variation society (HGVS), the human variome project (HVP), and the human genome organisation (HUGO). HGVS recommendations for the description of sequence variants: 2016 update, Hum. Mutat. 37 (2016)

564–569.
[6] C. Borsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. 18 (2015) 78–89.
[7] J. Shendure, S. Balasubramanian, G.M. Church, W. Gilbert, J. Rogers, J.A. Schloss, R.H. Waterston, DNA sequencing at 40: past, present and future, Nature 550 (2017) 345–353.
[8] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, Nat. Rev. Genet. 5 (2004) 739–751.
[9] K.M. Sullivan, S. Pope, P. Gill, J.M. Robertson, Automated DNA Profiling by fluorescent labeling of PCR products, PCR Methods Appl. 2 (1992) 34–40.
[10] C.P. Kimpton, P. Gill, A. Walton, A. Urquhart, E.S. Millican, M. Adams, Automated DNA profiling employing multiplex amplification of short tandem repeat Loci, PCR Methods Appl. 3 (1993) 13–22.
[11] J.S. Ziegle, Y. Su, K.P. Corcoran, L. Nie, P.E. Mayrand, L.B. Hoff, L.J. McBride, M.N. Kronick, S.R. Diehl, Application of automated DNA sizing technology for genotyping microsatellite loci, Genomics 14 (1992) 1026–1031.
[12] A. Kornberg, Enzymatic synthesis of deoxyribonucleic acid. XVI. Oligonucleotides as templates and the mechanisms of their replication, Proc. Natl. Acad. Sci. U.S.A. 51 (1964) 315–323.
[13] H. Fan, J.Y. Chu, Brief review of short tandem repeat mutation, Geno. Prot. Bioinfo. 5 (2007) No.1.
[14] B. Budowle, A.J. Onorato, T.F. Callaghan, M.A. Della, A.M. Gross, R.A. Guerrieri, J.C. Luttman, D.L. McClure, Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework, J. Forensic Sci. 54 (2009) 810–821.
[15] E. Pienaar, M. Theron, M. Nelson, H.J. Viljoen, A quantitative model of error accumulation during pcr amplification, Comput. Biol. Chem. 30 (2006) 102–111.
[16] M. Schirmer, U.Z. Ijaz, R. D'Amore, N. Hall, W.T. Sloan, C. Quince, Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform, Nucleic Acids Res. 43 (2015) e37.
[17] S.M. Huse, J.A. Huber, H.G. Morrison, M.L. Sogin, D.M. Welch, Accuracy and quality of massively parallel DNA pyrosequencing, Genome Biol. 8 (2007) R143.
[18] M. Kircher, U. Stenzel, J. Kelso, Improved base calling for the Illumina Genome Analyzer using machine learning strategies, Genome Biol. 10 (2009) R83.
[19] P.J.A. Cock, C.J. Fields, N. Goto, M.L. Heuer, P.M. Rice, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res. 38 (2010) 1767–1771.
[20] W.R. Pearson, D.J. lipmant, Improved tools for biological sequence comparison, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 2444–2448.
[21] D.H. Warshauer, J.L. King, B. Budowle, STRait razor v2.0: the improved STR allele identification tool, Forensic Sci. Int. Genet. 14 (2015) 182–186.
[22] S.L. Friis, A. Buchard, E. Rockenbauer, C. Børsting, N. Morling, Introduction of the Python script STRinNGS for analysis of STR regions in FASTQ or BAM files and expansion of the Danish STR sequence database to 11 STRs, Forensic Sci. Int. Genet. 21 (2016) 68–75.
[23] J. Hoogenboom, K.J. van der Gaag, R.H. de Leeuw, T. Sijen, Peter de Knijff, J.F.J. Laros, FDSTools: a software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise, Forensic Sci. Int. Genet. Suppl. Ser. 27 (2017) 27–40.
[24] B. Young, J.L. King, B. Budowle, L. Armogid, A technique for setting analytical thresholds in massively parallel sequencing-based forensic DNA analysis, PLoS One 12 (2018) e0178005.
[25] C. Phillips, K. Butler Gettings, J.L. King, D. Ballard, M. Bodner, L. Borsuk, W. Parson, "The devil's in the detail": release of an expanded, enhanced and dynamically revised forensic STR sequence guide, Forensic Sci. Int. Genet. 34 (2018) 162–169.
[26] W. Parson, D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, L. Gusmão, D.R. Hares, J.A. Irwin, J.L. King, P. de Knijff, N. Morling, M. Prinz, P.M. Schneider, C. Van Neste, S. Willuweit, C. Phillips, Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements, Forensic Sci. Int. Genet. 22 (2016) 54–63.
[27] WHO Nomenclature Committee, Nomenclature for factors of the HL-A system, Bull. World Health Org. 39 (1968) 483–486.
[28] K.J. van der Gaag, P. de Knijff, Forensic nomenclature for short tandem repeats updated for sequencing, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2015).
[29] R.P. Smyth, T.E. Schlub, A. Grimm, V. Venturi, A. Chopra, S. Mallal, M.P. Davenport, J. Mak, Reducing chimera formation during PCR amplification to ensure accurate genotyping, Gene 469 (2010) 45–51.
[30] B.ærholm Schnell, K. Bohmann, M.T.P. Gilbert, Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies, Mol. Ecol. Resour. 15 (2015) 1289–1303.