Non-uniform phenotyping of D12S391 resolved by second generation sequencing

Dalsgaard, S; Rockenbauer, E; Buchard, A; Mogensen, H S; Frank-Hansen, R; Børsting, Claus; Morling, N

Published in:
Forensic science international. Genetics

DOI:
10.1016/j.fsigen.2013.09.008

Publication date:
2014

Citation for published version (APA):
Case Report

Non-uniform phenotyping of D12S391 resolved by second generation sequencing

S. Dalsgaard, E. Rockenbauer, A. Buchard, H.S. Mogensen, R. Frank-Hansen, C. Børsting*, N. Morling
Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

A R T I C L E   I N F O

Article history:
Received 3 July 2013
Received in revised form 23 September 2013
Accepted 24 September 2013

Keywords:
Forensic genetics
Second generation sequencing
D12S391
STR
Phenotyping accuracy

A B S T R A C T

Non-uniform phenotyping of five case work samples were observed in the D12S391 locus. The samples were typed at least twice with the AmpF/STR® NGM SelectTM PCR Amplification Kit and different alleles were called with GeneMapper® ID-X in the different experiments. Detailed analyses of the electropherograms suggested that the individuals were heterozygous with two alleles that differed in size by one nucleotide. This was confirmed by amplifying the samples with the PowerPlex® ESX 17 system.

D12S391 is a complex STR with variable numbers of AGAT and AGAC repeats. Second generation sequencing revealed that separation of two alleles differing by one nucleotide in length was poor if the number of AGAT repeats in the short allele was higher than in the long allele.

A total of 45 individuals with microvariants or off-ladder alleles in D12S391 were sequenced. Thirty different alleles were detected and sixteen of these were not previously reported.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Short tandem repeats (STRs) with four base pair repeat units are the preferred markers in forensic genetic investigations [1]. They are highly polymorphic and it is possible to identify most individuals and resolve most relationship queries by investigating 15–25 autosomal STRs. The STRs are usually amplified using commercial multiplex PCR kits and the alleles are detected using capillary electrophoresis (CE) by comparing the amplified products with an allelic ladder. Under the recommended CE conditions, the variation in sizing precision is approximately 0.15–0.25 nt [2–5], which is sufficient to allow phenotyping of alleles that differ in length by one nucleotide. However, others have reported variations up to 0.8 for the longest alleles of the FGA locus [6].

Some of the commonly used STRs in forensic genetics are complex STRs with interrupted repeat sequences that consist of 2–3 sub-repeat units (www.cstl.nist.gov/div831/strbase/). The sub-repeats may have different sequences and variations may appear in any of the sub-repeats. Thus, some alleles have different sequences although they have the same lengths and, consequently, same-size alleles may migrate differently in the CE since the nucleotide composition affects the mobility of DNA during electrophoresis [7,8].

D12S391 is one of the new markers in the European Standard Set [9]. It is a complex STR with two variable sub-repeats and a well-known SNP in the last base of the last repeat. Furthermore, microvariants in the form of one base pair deletions are relatively frequent [10–13]. A number of same-size, different-sequence alleles have been identified by sequence analysis [12,13].

In our case work, we have observed non-uniform D12S391 allele calls in duplicate typing of reference samples (Fig. 1A). A close inspection of the electropherograms suggested that the individuals were in fact heterozygous and that the two alleles had almost the same size. The phenomenon was observed for the allele-pairs 19.3/20 and 20.3/21. However, most individuals with 19.3/20 and 20.3/21 allele pairs were identified unambiguously as heterozygous (Fig. 1B).

In this work, we sequenced the D12S391 locus in 45 individuals to elucidate why some allele pairs, differing by one nucleotide, were difficult to separate in CE.

2. Materials and methods

2.1. Samples and DNA extraction

A total of 45 samples from unrelated individuals in Danish paternity cases and individuals in crime cases were selected. All
samples included either a microvariant or an off-ladder variant in D12S391 and 26 individuals were heterozygous for two alleles that differed in length by one nucleotide. Non-uniform allele calls were observed in duplicate typing of five of these samples. The samples were collected as either buccal swabs on FTA cards or as anticoagulated blood. DNA was isolated from blood using the EZ1 Advanced XL BioRobot (Qiagen) as recommended by the manufacturer. A 1.2 mm FTA card punch was washed as previously described [14] prior to PCR amplification with the AmpF/STR® NGM Select™ PCR Amplification Kit (Applied Biosystems/Life Technologies (AB/LT)) and the PowerPlex® ESX 17 system (Promega). The work was approved by the Danish ethical committee (H-1-2011-081).

2.2. STR typing

During the initial casework investigation, the 45 samples were amplified with either the AmpF/STR® NGM Select™ PCR Amplification Kit or the AmpF/STR® NGM Select™ Express Kit (AB/LT) according to the guidelines of the manufacturer. Prior to electrophoresis, 1 μL of the amplified product and 0.5 μL of GeneScan™ 600 LIZ® Size Standard v2.0 (AB/LT) were added to 9.5 μL of deionized Hi-Di™ formamide (AB/LT), denatured at 95 °C for 3 min, and placed on an Eppendorf® PCR-Cooler (Eppendorf) for 3 min. Samples were injected for 10 s at 3 kV and separated by electrophoresis in Performance Optimized Polymer (POP-4™; AB/LT) using the HIDFracntAnalysis36_POP4 Module (AB/LT) and a 1500 s run time on the ABI 3130xl Genetic Analyzer (AB/LT).

A total of 26 samples were typed with the PowerPlex® ESX 17 system (Promega) according to the guidelines of the manufacturer. Prior to electrophoresis, 1 μL of the amplified product and 1 μL of CCS Internal Lane Standard 500 (Promega) were added to 10 μL of deionized Hi-Di™ formamide (AB/LT), denatured at 95 °C for 3 min, and placed on an Eppendorf® PCR-Cooler (Eppendorf) for 3 min. Samples were injected for 3 s at 3 kV and separated by electrophoresis in Performance Optimized Polymer (POP-4™; AB/LT) using the HIDFracntAnalysis36_POP4 Module (AB/LT) and a 1500 s run time on the ABI 3130xl Genetic Analyzer (AB/LT).

Collection files were analyzed with GeneMapper® ID-X Software 1.1.1 (AB/LT) using the default settings except for the following changes: In the Analysis Method Editor, the peak amplitude threshold was 50 RFUs for all colors, the Peak Height Ratio was 0.5 and the Min/Max Peak Height was 150 RFU for homozygotes and 50 RFU for heterozygotes.

2.3. 454 pyrosequencing

A total of 25 fusion primer sets with multiplex identifiers (MIDs) were designed according to the protocol “454 Sequencing System Guidelines for Amplicon Experimental Design” (Roche Diagnostics). The sequences of all fusion primers are presented in Supplemental Table S1. Samples were PCR-amplified in 25-μL reaction volumes using AmpliTaq Gold (AB/LT). Each reaction contained 2.5 μL 1 × AmpliTaq Gold buffer, 2.5 mM MgCl₂, 0.4 μM forward and 0.4 μM reverse primers (DNA Technology A/S), 0.2 mM each dNTP, 0.2 μL AmpliTaq Gold, and 2 μL DNA. The PCR cycling conditions were as follows: 95 °C for 10 min, 30 cycles of 94 °C for 30 s, 62 °C for 30 s, 72 °C for 30 s and a final extension at 60 °C for 10 min. The PCR products were visualized on the LabChip® GX with the HT DNA High Sensitivity LabChip® Kit (Caliper Life Sciences) as recommended by the manufacturer. Excess primers and nucleotides were removed using the Agencourt Ampure XP PCR Purification system (Beckman Coulter) following the recommendations of the manufacturer.

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.foresci.2013.09.008.

The PCR products were quantified using the Qubit® 2.0 Fluorometer with the Qubit® dsDNA HS Assay Kit (AB/LT). Sample concentrations were calculated following the guidelines of the manufacturer.

The samples were sequenced using the GS Junior System (Roche Diagnostics). Prior to sequencing, an emulsion PCR-step was performed according to the manual, emPCR Lib-A (Roche Diagnostics). Sequencing experiments with 23 and 22 samples, respectively, were carried out. The output data was sorted by an algorithm as previously described [15]. This algorithm sorts the sequencing data according to MID-tags, primer sequences and STR-ends, and creates an output file containing sequences for each sample. Incomplete STR sequences were removed. Sequences were aligned toward the D12S391 reference sequence with 19.3 repeats.
(GenBank AP000433) using the alignment software BioEdit (URL: http://www.mbio.ncsu.edu/bioedit/bioedit.html).

3. Results

3.1. Sequencing of D12S391

A total of 45 samples with either a microvariant or an off-ladder variant in D12S391 were selected for sequencing. Two sequencing runs resulted in a total of 401,935 reads. Only sequences that contained primer sequences and the entire STR repeat sequence of D12S391 were analyzed further. An average of 147 reads per sample (range: 56–213 reads) was obtained. Sequences of each allele were grouped according to repeat region lengths and aligned to a reference sequence of D12S391 with 19.3 repeats (GenBank AP000433). Allele calls were determined and compared to those obtained by fragment analysis. There was complete concordance between the allele calls determined by 454 pyrosequencing and CE for 40 of the 45 individuals. However, for five heterozygous individuals where the two sequenced alleles differed by one nucleotide in length (Supplemental Table S2), the automatic CE-analysis failed to detect both alleles when D12S391 were typed with the AmpF/STR® NGM SElect® PCR Amplification Kit and analyzed with GeneMapper® ID-X. Only one allele was identified and, importantly, the allele call was not consistent between duplicate experiments (Supplementary Table S2). Furthermore, some of the allele calls passed the Quality Value Thresholds of the GeneMapper® ID-X and were marked with green flags. The size differences of the two alleles in the five individuals ranged from 0.35 to 0.65 bp when determined manually in GeneMapper® ID-X. In contrast, two alleles were clearly detected by CE in 21 other individuals where the allele size differences ranged from 0.64 to 1.05 bp.

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.fsigen.2013.09.008.

The sequence data revealed that the nucleotide composition seemed to affect the mobility of the PCR products in the CE. Same size alleles with many AGAT repeats migrated slower than alleles with fewer AGAT repeats (exemplified for allele 20 and 21 in Supplementary Table S3). In Fig. 2, the size differences of the two alleles in the 26 heterozygous individuals were plotted against the differences in number of AGAT repeats (long allele minus short allele). Separation of the two alleles was poor if the number of AGAT repeats in the short allele was higher than in the long allele.

3.2. Fragment analysis with the PowerPlex® ESX 17 system

The 26 heterozygous individuals, where the two alleles differed by one nucleotide, were typed with the PowerPlex® ESX 17 system. The D12S391 fragments amplified with the PowerPlex® ESX 17 system were 98 bp shorter than the fragments amplified...

Fig. 2. Comparison of the number of AGAT sub-repeats and the size differences of alleles in 26 heterozygous individuals where the two alleles differed in length by one nucleotide (Supplementary Table S2). The results from 59 CE-experiments are shown. In CE-runs where peaks were not labeled, the sizes were determined manually.

Fig. 3. Duplicate typing of D12S391 of the same samples typed in Fig. 1 using the PowerPlex® ESX 17 system. OL, Off-ladder allele.
Table 1
Novel alleles in D12S391.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Repeat structure</th>
<th>Alleles sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>[AGAT]3[AGAC]1[AGAT]</td>
<td>1</td>
</tr>
</tbody>
</table>

with the AmpF/STR® NGM SElect™ PCR Amplification Kit, and separation of the alleles in CE should be better. All individuals were typed as heterozygous with the PowerPlex® ESX 17 system, including the five individuals that were non-uniformly typed with the AmpF/STR® NGM SElect™ PCR Amplification Kit (Fig. 3). Interestingly, the size differences of the two alleles in these five individuals ranged from 1.25 to 1.52 bp, whereas the size differences of the other 21 heterozygous individuals ranged from 0.88 to 1.33 bp (data not shown). Thus, the two alleles were better separated if the short allele had more AGAT repeats than the long allele, which was the opposite of the observation made with the AmpF/STR® NGM SElect™ PCR Amplification Kit. This suggested that the labeled D12S391 primer in the PowerPlex® ESX 17 system hybridized to the opposite strand of the labeled primer in the AmpF/STR® NGM SElect™ PCR Amplification Kit (see discussion).

3.3. Allele variation in D12S391

A total of 30 different alleles were sequenced in the 45 unrelated individuals (Supplementary Table S4) whereas only 17 different sized alleles were found by fragment analysis. Moreover, 16 of the 30 alleles were not previously reported (Table 1). Incomplete repeat sequences were found most frequently in the second AGAT repeat, but also in the 4th, 5th, 6th and 7th AGAT repeat. No incomplete AGAC repeats were found. Four different alleles with 20 repeats and three different alleles with 20.3 and 21 repeats were observed.

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.fsigen.2013.09.008.

4. Discussion

Poor separation of certain combinations of D12S391 alleles in the CE combined with an expert analysis software that did not register the problem is a matter of concern. The non-uniform phenotyping was identified by typing the samples at least twice and by manual scrutiny of the electropherograms. However, that cannot guarantee that all incidents may be or have been found during analysis. Thus, it is necessary that case officers are aware of the risk of non-uniform allele calling in the D12S391 locus and that special attention is made to cases with a single mismatch in D12S391 between e.g. a trace sample and a reference sample from a suspect or a profile in the DNA database.

We show here that D12S391 allele combinations that were poorly separated after amplification with the AmpF/STR® NGM SElect™ PCR Amplification Kit could be separated when the samples were amplified with the PowerPlex® ESX 17 system. Therefore, it may be advantageous to type all samples with two different kits as proposed previously by others [16,17].

In the AmpF/STR® NGM SElect™ PCR Amplification Kit, the forward D12S391 primer is labeled with NED (Wiljo de Leeuw, AB/LT, personal communication) and the detected strands have the STR sequences [AGAT]1[AGAC]1[AGAT]1. It was previously shown that CA strands migrated faster than GT strands in denaturing polyacrylamide electrophoresis [7]. In this work, we demonstrated that same-size D12S391 alleles with more AGAT repeats (and fewer AGAC repeats) migrated slower than fragments with fewer AGAT repeats (and more AGAC repeats). In effect, more T bases and fewer C bases made the fragments migrate slower, which substantiates the conclusions made by Saitoh and co-workers [7]. This also explained why two alleles that differed by one nucleotide in length were difficult to separate in CE if the short allele had more AGAT repeats than the long allele. If the short allele had three AGAT repeats more than the long allele, the size difference approached the sizing precision under the standard CE-conditions used for the STR kits and it seems likely that it may be impossible to separate the alleles if the short allele has four AGAT repeats more than the long allele (Fig. 2). However, this allele combination was not observed among the relatively few individuals that were sequenced in this work.

When the same heterozygous individuals were typed with the PowerPlex® ESX 17 system, better separation was obtained when the short allele had more AGAT repeats than the long allele. This was the opposite of the observation made with the AmpF/STR® NGM SElect™ PCR Amplification Kit. We were not able to obtain information on the labeled primer of the PowerPlex® ESX 17 system. However, if the reverse primer is labeled, the detected strand in CE would have the STR sequence [ATCT]1[GTCT]1[ATCT]1. Thus, if the allele contains more AGAT repeats and fewer AGAC repeats in the forward sequence, the reverse strand would contain more A bases and fewer G bases. According to Saitoh and co-workers [7], this makes the strands migrate faster and may explain why alleles that differ by one nucleotide in length are better separated if the short allele has more AGAT repeats than the longer allele.

D12S391 is a highly polymorphic locus but only a fraction of the true variation was detected by the standard PCR-CE fragment analysis. Overall, 30 different alleles were identified in 45 individuals by second generation sequencing. Sixteen of the alleles had not previously been reported. We observed a similar level of variation in another complex STR, the D21S11 locus [15,18]. With second generation sequencing techniques, different alleles with the same sizes but with different sequences may be identified and SNPs in nearby positions may add further to the variability as seen for D5S818, D13S317 and D21S1338 [15,19,20]. More variable loci mean more statistical power of the investigations and will reduce the number of loci that needs to be typed to solve a case. It may also be possible to resolve mixtures more easily e.g. if the true allele of the minor contributor has a different sequence than the stutter artifact of the major contributor. Characterization of mutation events in a relationships case will be possible in some STR loci [18], and finally, second generation sequencing will make it possible to create an all-in-one multiplex with relevant forensic markers that include STRs, mtDNA markers and SNPs or indels that may reveal phenotypical traits. Currently, high throughputs STR typing assays and human identification SNP panels for the various second generation sequencing platforms are being developed for forensic genetics, and these assays may provide the forensic community with the first real alternative to PCR-CE analyses.
References


Acknowledgement

We thank Anja L. Jørgensen, Trine L. Hansen and Pernille M. Jensen for technical support.