Forensic Science International: Genetics 15 (2015) 2–7

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Forensic massively parallel sequencing data analysis tool: Implementation of MyFLq as a standalone web- and Illumina BaseSpace®-application

Christophe Van Neste [a], Yannick Gansemans [a], Dieter De Coninck [a], David Van Hoofstat [a], Wim Van Criekinge [b], Dieter Deforce [a,1], Filip Van Nieuwerburgh [a,1,*]

[a] Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium
[b] Biobix, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

ARTICLE INFO

Article history:
Received 16 August 2014
Received in revised form 19 September 2014
Accepted 3 October 2014

Keywords:
Illumina
MiSeq
STR
Forensic loci
MPS
NGS

ABSTRACT

Routine use of massively parallel sequencing (MPS) for forensic genomics is on the horizon. The last few years, several algorithms and workflows have been developed to analyze forensic MPS data. However, none have yet been tailored to the needs of the forensic analyst who does not possess an extensive bioinformatics background.

We developed our previously published forensic MPS data analysis framework MyFLq (My-Forensic-Loci-queries) into an open-source, user-friendly, web-based application. It can be installed as a standalone web application, or run directly from the Illumina BaseSpace environment. In the former, laboratories can keep their data on-site, while in the latter, data from forensic samples that are sequenced on an Illumina sequencer can be uploaded to Basespace during acquisition, and can subsequently be analyzed using the published MyFLq BaseSpace application. Additional features were implemented such as an interactive graphical report of the results, an interactive threshold selection bar, and an allele length-based analysis in addition to the sequenced-based analysis.

Practical use of the application is demonstrated through the analysis of four 16-plex short tandem repeat (STR) samples, showing the complementarity between the sequence- and length-based analysis of the same MPS data.

1. Introduction

Obtaining forensic DNA profiles of polymorphic short tandem repeat (STR) loci using PCR followed by capillary electrophoresis (CE) is still the gold standard. However, routine use of massively parallel sequencing (MPS) for forensic genomics is on the horizon. MPS technologies do not rely on size separation and thus relieve the limitation on locus multiplexing that is present in CE [1,2]. MPS therefore creates enhanced possibilities within forensic genomics for analyzing degraded samples, mixed samples, and in dealing with kinship or population substructure [3,4].

Forensic bioinformaticians have been working on several algorithms to process MPS forensic STR data: lobSTR [5], RepeatSeq [6], STRait Razor [7], TSSV [8] and the MyFLq-framework [9]. LobSTR and RepeatSeq are both genome wide STR aligners, and therefore outside of the scope of forensic analysis in its current legal and technological setting, in which targeted sequencing of a limited number of validated loci are investigated.

STRait Razor, TSSV and MyFLq are instead locus-centric, and operate on forensical loci. They require configuration information for each locus in the set, generally consisting of the repeat length of the locus, primer and/or flank sequences, and known alleles for the locus. All three programs have a similar approach to processing the STR data, which is represented in a flowchart in Fig. 1. To date, algorithms in these programs process data to the point of presenting allele candidates (step preceding the dashed red arrow in Fig. 1). It is at this point in the pipeline that data interpretation begins for the forensic analyst.

All current applications, are command-line based and are thus not well suited to be used by forensic analysts that do not have extensive bioinformatics experience. In this report, we present the MyFLq application that we developed into an open-source, web-based application with a user-friendly graphical user interface.

* Corresponding author.
E-mail address: Filip.VanNieuwerburgh@UGent.be (F. Van Nieuwerburgh).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.fsigen.2014.10.006

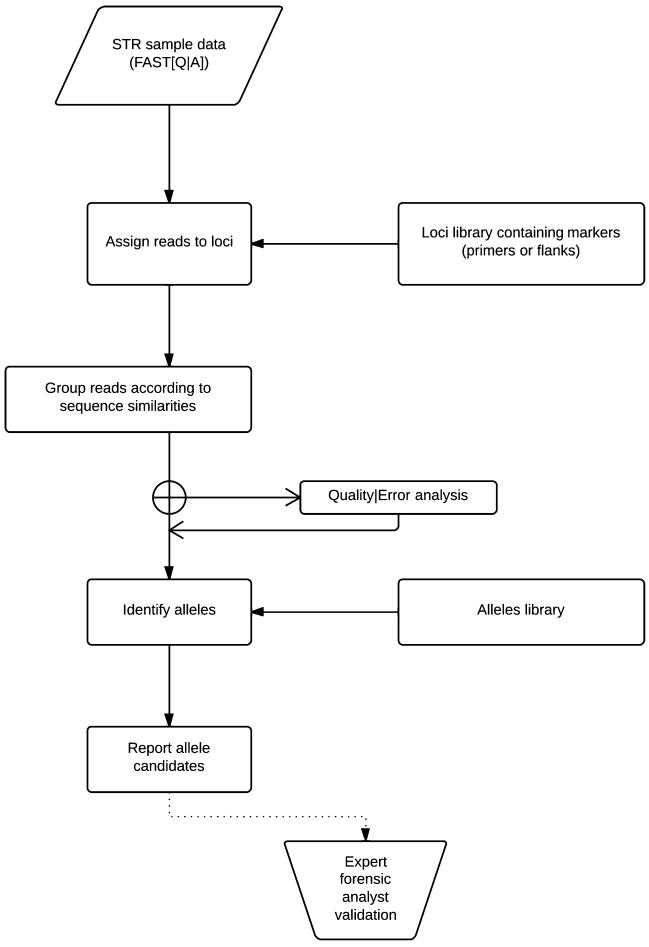


Fig. 1. STR data processing.

Additional features were implemented such as an interactive graphical report of the results, an interactive threshold selection bar, and an allele length-based analysis in addition to the sequenced-based analysis.

2. Materials and methods

MyFLq has been implemented both as a Django web application [10] and an Illumina BaseSpace application. Both implementations run from the same source code and users have access to the latest stable version, no matter the execution preference of the application. The BaseSpace MyFLq application requires no installation from the user. For the Django application, detailed documentation can be found on the MyFLq GitHub repository (https://github.com/beukueb/myflq). A pdf manual can be downloaded from https://gitprint.com/beukueb/myflq, covering both implementations. The development version and previous builds are only available for the Django application.

2.1. Samples

The same data were used as in the MyFLq framework paper [9]. The results presented in this report were obtained with sample 9947A_S1, which is a single contributor control DNA sample (Promega) [11]. This sample was amplified using a 16-plex PCR, based on the PowerPlex® 16 primers (Promega) [12]. The reference profile for 9947A with the 16-plex is shown in Supplementary Table A.1. The MyFLq framework paper [9] also analyzed a second single contributor sample and two multiple person mixtures. Results for these samples are available on BaseSpace, together with the FASTQ data for anyone wishing to experiment with MyFLq.

2.2. Launching MyFLq

To produce the results for this report, MyFLq was launched from http://basespace.illumina.com/apps. A threshold of 0.5% was set to filter read groups with a lower abundance for further analysis. The loci set and the allele database were set to the MyFLq framework paper options, as shown in Fig. 2. The database contained all the alleles from the framework paper's four DNA samples, including sample 9947A [9]. The database consists of all sequences of the Powerplex® 16 alleles present in these four samples. For the other options the default values were used. Detailed information on these settings can be found in Supplementary Table A.2 or the online documentation. A BaseSpace project "FSIG" was made to which the results could be saved. Finally, the analysis was launched by clicking "Continue".

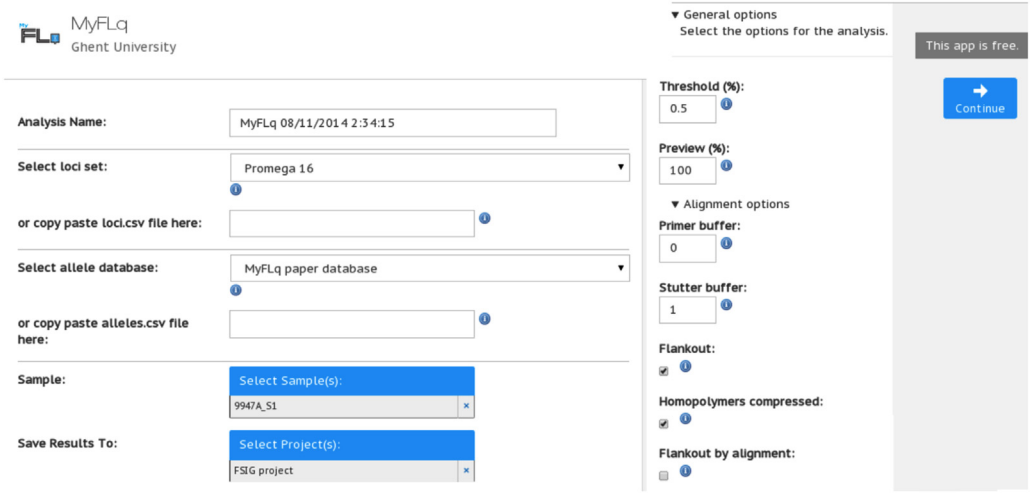
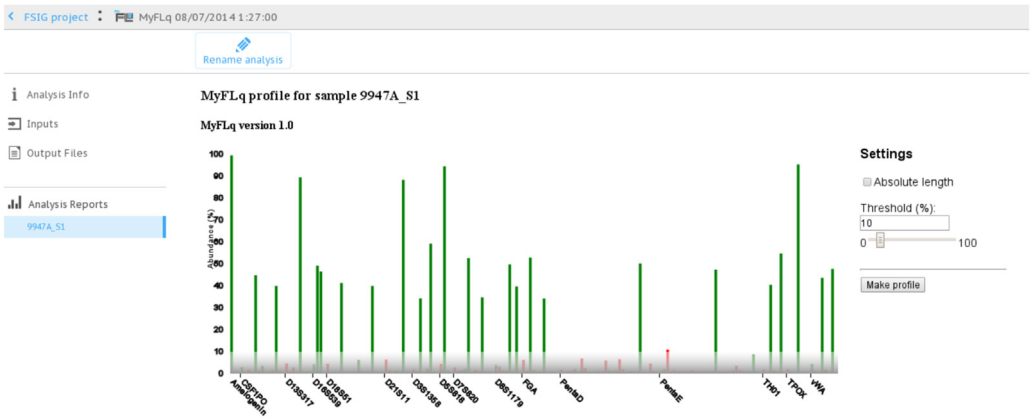
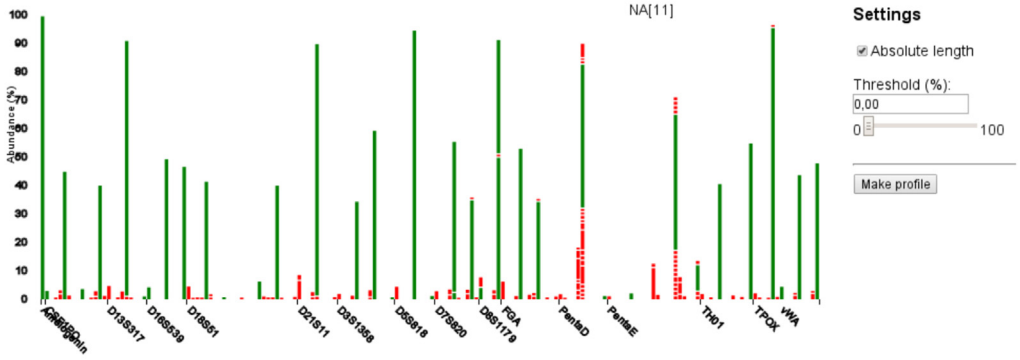


Fig. 2. Primary analysis settings. The setting "Select loci set" controls the configuration of the set of loci that will be analyzed in the dataset. Normally these are the same loci as in the applied STR multiplex PCR. With the "Select allele database" setting, the database with the reference alleles sequences can be configured. Both configurations can be provided as csv-files if they are not available in the dropdown menus. Settings on the right are general and can be used to tweak the results. They are described in Supplementary Table A.2.



(a) Initial display of analysis result.



(b) Alleles proportionally sorted and stacked according to length.

Fig. 3. Initial display and proportionally sorted analysis result.

3. Results

3.1. Initial sequence-based result display

Fig. 3a shows the analysis result page, that can be found under the project folder where the analysis was saved. The initial display shows an interactive visual representation that should be interpreted as a sequence-based analysis rather than a length-based analysis. The different bars represent grouped allele sequences and are sorted according to length. Spacing is however not proportional and allele candidates of the same length are not stacked on top of each other, but rather side-by-side. A green bar is given to sequences that are present in the database, a red bar when not. The vertically adjustable gray transparent zone determines the threshold for which allele candidate bars with a lower abundance will not be withheld in the final profile. By default, it is set to 10%. Note that sequences with an abundance threshold lower than 0.5% (configurable) are already filtered during the analysis.

3.2. Detailed allele candidate information

When hovering over a bar, a detailed block of information is displayed for that allele candidate. An example is shown in Fig. 4. This information can be used to examine if the underlying

Locus D8S1179 → allele candidate: 13[b]

Locus stats		Allele stats	
Total reads	84546	Index	5
Filtered reads	79424	Abundance	39.76%
Total unique	368	Strand distribution	49.17%
Filtered unique	7	Clean flanks	96.3%

In profile: ☑

Relations to other sequences within D8S1179

Allele index	Relation degree	Sequence
5	-	ATCTATCTATCTATCTATCT
1	I'st	ATCTATCTATCTATCT
4	I'st	ATCTATCTATCTATCTATC
6	I'st	ATCTATTTATCTATCTATCT
7	I'st	ATCTGTCTATCTATCTATCT
2	II'nd	ATCTGTCTATCTATCT
3	II'nd	ATCTGTCTATCTATCTATC

Locus D8S1179 → allele candidate: 13[a]

Locus stats		Allele stats	
Total reads	84546	Index	7
Filtered reads	79424	Abundance	49.78%
Total unique	368	Strand distribution	48.75%
Filtered unique	7	Clean flanks	96.2%

In profile: ☑

Relations to other sequences within D8S1179

Allele index	Relation degree	Sequence
7	-	ATCTGTCTATCTATCTATCT
1	I'st	ATCTATCTATCTATCT
2	I'st	ATCTGTCTATCTATCT
3	I'st	ATCTGTCTATCTATCTATC
5	I'st	ATCTATCTATCTATCTATCT
4	II'nd	ATCTATCTATCTATCTATC
6	II'nd	ATCTATTTATCTATCTATCT

Fig. 4. Information blocks for true D8S1179 alleles. Their one-base sequence difference is indicated by dotted lines.

sequence of the bar is either a true allele or erroneous sequence (stutter, sequencing- or PCR error). The title bar of the information block shows the locus name, and the database name of the allele candidate. When the allele is not present in the database, 'NA' together with the number of repeats relative to known alleles is shown between brackets. Locus statistics are summarized in the left column:

- 'Total reads' stands for all reads that are classified under the locus.
- 'Filtered reads' stands for the reads that are retained after the 0.5% abundance threshold filter. 'Filtered reads' are those that are actually shown in the allele bar.
- 'Total unique' and 'Filtered unique' stand for the number of unique sequences, and are a measure of the number of possible allele candidates. In the example in Fig. 4 there are 7 filtered sequences with more than 0.5% of the reads.

Statistics for the current allele candidate are in the right column:

- 'Index' is a unique reference index label assigned to each filtered unique sequence, starting at '1' with the shortest sequence for this locus in the analysis. When two sequences have the same length, the smaller index number is assigned randomly.
- 'Abundance' is the percentage of the filtered reads within the locus that belong to this candidate.
- 'Strand distribution' is the balance between forward and reverse sequences supporting the candidate. Sequences in the report itself are always in forward direction, as determined by the loci set configuration file. In the example in Fig. 4, the strand distribution of the first sequence is 49.17% and of the second 48.75%. They are both close to 50% as is to be expected as a result from normal PCR and sequencing circumstances.
- 'Clean flanks' is the percentage of perfect flanks compared to the allele database.

The bottom part of the information block shows the region of interest of the allele candidate sequence together with related sequences from the same locus. Related sequences with up to two differences are shown; a difference being either one repeat number difference or one base pair difference. One difference is indicated by a relation degree "Ist" and two differences by "IInd".

Fig. 4 shows the two information blocks of the two true alleles from locus D8S1179 in an interesting example that shows the advantage of MPS over CE. For 9947A, CE results show only one peak at locus D8S1179, resulting in a profile with a homozygous allele 13 for D8S1179. Our analysis clearly shows two alleles that have the same length (corresponding to allele 13), but have a different intra-STR sequence when compared to each other. The information blocks support this heterozygous call; only a small portion of the reads are filtered for this locus, the number of unique reads are low and the abundance of the two allele candidates is approximately 50%. The percentage of clean flanks [9] in the candidate alleles sequences is also very high. All these parameters indicate that the sequencing and PCR error rate is low. In the part of the information blocks that shows the related sequences, the G ↔ A difference between the two alleles is shown. The two alleles are related to each other by a "Ist" order degree. Both alleles also have relations to erroneous sequences, in fact all other sequences discovered for this locus have a relation to one or both of the true alleles.

Fig. 5 shows the information block for a candidate allele of locus Penta E. It is the only erroneous sequence that was not

Locus PentaE → allele candidate: NA

Locus stats		Allele stats	
Total reads	446	Index	3
Filtered reads	192	Abundance	10.94%
Total unique	283	Strand distribution	100.00%
Filtered unique	29	Clean flanks	59.5%

In profile:

Relations to other sequences within PentaE

Allele index	Relation degree	Sequence
3	-	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
2	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
4	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGGCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
5	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAA
6	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
7	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCCAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
8	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAGAGAAAAGA
12	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGCAAAGAAAAGAAAAGAAAAGAAAAGA
16	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGCAAAGAAAAGA
17	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGCAAAGAAAAGAAAAGA
19	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGCAAAGAAAAGAAAAGAAAAGA
20	I'st	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGACAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
1	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGA
9	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTAGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
10	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
11	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
13	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACCCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
14	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
15	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
18	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGCCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
21	II'nd	GGGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
22	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAA
23	II'nd	AGTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
24	II'nd	AGTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
27	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
28	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGA
29	II'nd	GTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCGACTGAGCAAGACTCAGTCTCAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGC

Fig. 5. Unknown Penta E allele. The true allele with index 6 is indicated by the dotted line.

automatically filtered by the 10% default threshold. The information supports that this candidate allele should be disregarded. The putative allele length is one STR repeat unit smaller than the high abundant (47.40%) sequence with index 6, indicating that it might be stutter. Apart from this stutter there are no other sequence differences (Ist relation degree). Furthermore, the clean flank percentage is rather low (59.5%), indicating possible low quality sequences. An unexpected strand distribution of 100% implies that there are no complementary reads supporting the presence of this allele candidate. Removing this allele candidate is accomplished by unchecking the "in profile" check-box.

3.3. Allele candidates proportionally sorted according to length

After selecting the "Length-based analysis" check-box, all allele candidates are displayed proportionally, according to their actual length within the locus, as shown in Fig. 3. For each locus, the *x*-axis is adjusted to show the locus length starting from the shortest allele and ending at the longest allele. The threshold bar is no longer displayed because allele candidates with the same length are now stacked on top of each other, which creates one bar that shows the total abundance of all alleles with the same length within each locus. This representation resembles a CE profile. The example of the allele candidate in Fig. 5 now visually looks like a CE stutter peak based on the relative length and abundance difference as compared to the true allele.

3.4. Final profile

After reviewing the profile by setting the threshold to an appropriate value, and removing allele candidates of poor quality, pressing the "Make profile" button yields the final profile. This profile can then be used to query databases or compare to the profile of a sample of interest. Fig. 6 shows the final profile for sample 9947A_S1. Using the threshold of 10%, it has one Penta E allele 13 that is undetected relative to the known genotype (Table A.1). This allele is present in the data at an abundance of 8.85% and its corresponding green bar can be seen clearly in Fig. 3. The sub-optimal results of the pentanucleotide loci, Penta D and Penta E, were previously discussed in detail [9].

Locus	Alleles
Amelogenin	X
CSF1PO	10[a],12[a]
D13S317	11[b]
D16S539	11[a],12[a]
D18S51	15[a],19[a]
D21S11	30[a]
D3S1358	14[a],15[b]
D5S818	11[a]
D7S820	10[b],11[a]
D8S1179	13[a],13[b]
FGA	23[a],24[a]
PentaD	12[a]
PentaE	12[a]
TH01	8[a],9.3[a]
TPOX	8[a]
vWA	17[a],18[a]

Fig. 6. Final profile for sample 9947A_S1 with 10% threshold.

4. Discussion

We show how an MPS data-set can be analyzed using an easy-to-use graphical user interface, requiring a limited number of parameters and almost no bioinformatics expertise. The interactive visual representation of the results shows additional information when hovering over the alleles, allowing for in-depth analysis of the underlying sequences and the related statistics. For clarity of explanation we chose to display and discuss the analysis of a single contributor sample, but the MyFLq framework equally works on mixtures because no assumptions on mixture composition are made to perform the analysis. The main added value of MPS over CE indeed lay in the analysis of mixed and degraded samples [9]. With MPS, sequences can be analyzed more in depth to determine whether they are genuinely from one of the original contributors of a sample, or instead more likely to be the product of a PCR or sequencing error. Additionally, due to the ability to multiplex more loci than CE affords, broader genetic interrogation can be achieved in a single reaction, thus conserving precious samples.

The reported results comprise only 16 loci, but MyFLq can run with any number of loci. When running MyFLq with a custom loci set, the primers of these loci can be imported. The allele database is not strictly necessary to run the program. In exploratory studies, for example if building a database of known alleles, MyFLq can be run with an empty allele database. The GitHub repository contains example files for users that need either a custom locus set or custom allele database.

The used allele database was very small as it only compromised the alleles of the five contributors. Sequences that are currently not in the database are marked as red bars. These bars are very useful to visually monitor the noise level. In the future, with a larger database, it could be that erroneous sequences are nonetheless present in the database, as they could be true alleles for individuals that are not present in the sample. The solution to that problem could be to mark rare alleles (e.g. alleles with a population prevalence $<1\%$) with a different color. The combination of unknown alleles and rare alleles would then indicate the level of noise. A further limitation of the current database is its nomenclature. Currently same-sized alleles get an arbitrary name within the database, which would make it difficult to perform searches in other databases without the original sequence. When an international nomenclature for MPS STR alleles has been established, it will be incorporated in MyFLq.

When all allele candidates have been reviewed, the "Make profile" action generates a report with only the selected alleles. This is the profile that a forensic analyst can use to either store in a database, to query against a database, or for direct comparison to a known sample of interest. Future versions of the software will include possibilities to interact directly with sample databases. New feature requests can be made through the GitHub website.

5. Conclusion

MyFLq is the first open-source, web-based forensic MPS DNA analysis software with an easy-to-use graphical user interface.

It can run natively on Illumina BaseSpace, or independently on a forensic laboratory's server. The possibility to run the program directly from the Illumina BaseSpace environment means no extensive bioinformatics skills are required.

Competing interests

C.V.N. participated in an internship program at Illumina, Inc. to provide feedback on building a native BaseSpace application to the Illumina developers.

Acknowledgments

Funding was provided by the Ghent University Multidisciplinary Research Partnership 'Bioinformatics: from nucleotides to networks'

The authors would also like to thank Illumina, Inc. for providing the data and making MyFLq easily accessible on their BaseSpace platform.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2014.10.006.

References

[1] S.L. Fordyce, M.C. Avila-Arcos, E. Rockenbauer, C. Borsting, R. Frank-Hansen, F.T. Petersen, E. Willerslev, A.J. Hansen, N. Morling, M.T.P. Gilbert, High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform, Biotechniques 51 (2) (2011) 127, http://dx.doi.org/10.2144/000113721.

[2] K.K. Kidd, J.R. Kidd, W.C. Speed, R. Fang, M.R. Furtado, F.C.L. Hyland, A.J. Pakstis, Expanding data and resources for forensic use of SNPs in individual identification, Forensic Sci. Int. Genet. 6 (5) (2012) 646–652. , http://dx.doi.org/10.1016/j.fsigen.2012.02.012.

[3] L.J. McIver, J.W. Fondon III, M.A. Skinner, H.R. Garner, Evaluation of microsatellite variation in the 1000 Genomes Project pilot studies is indicative of the quality and utility of the raw data and alignments, Genomics 97 (4) (2011) 193–199. , http://dx.doi.org/10.1016/j.ygeno.2011.01.001.

[4] J. Weber-Lehmann, E. Schilling, G. Gradl, D.C. Richter, J. Wiehler, B. Rolf, Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing, Forensic Sci. Int. Genet. 9 (2014) 42–46. , http://dx.doi.org/10.1016/j.fsigen.2013.10.015.

[5] M. Gymrek, D. Golan, S. Rosset, Y. Erlich, lobSTR: a short tandem repeat profiler for personal genomes, Genome Res. 22 (6) (2012) 1154–1162. , http://dx.doi.org/10.1101/gr.135780.111.

[6] G. Highnam, C. Franck, A. Martin, C. Stephens, A. Puthige, D. Mittelman, Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles, Nucleic Acids Res. 41 (1) (2013), http://dx.doi.org/10.1093/nar/gks981.

[7] D.H. Warshauer, D. Lin, K. Hari, R. Jain, C. Davis, B. LaRue, J.L. King, B. Budowle, STRait Razor: a length-based forensic STR allele-calling tool for use with second generation sequencing data, Forensic Sci. Int. Genet. 7 (7) (2013) 409–417.

[8] S.Y. Anvar, K.J. van der Gaag, J.W.F. van der Heijden, M.H.A.M. Veltrop, R.H.A.M. Vossen, R.H. de Leeuw, C. Breukel, H.P.J. Buermans, J.S. Verbeek, P. de Knijff, J.T. den Dunnen, J.F.J. Laros, TSSV: a tool for characterization of complex allelic variants in pure and mixed genomes, Bioinformatics 30 (12) (2014) 1651–1659. , http://dx.doi.org/10.1093/bioinformatics/btu068.

[9] C. Van Neste, M. Vandewoestyne, W. Van Criekinge, D. Deforce, F. Van Nieuwerburgh, My-Forensic-Loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing, Forensic Sci. Int. Genet. 9 (2014) 1–8. , http://dx.doi.org/10.1016/j.fsigen.2013.10.012.

[10] J. Forcier, P. Bissex, W. Chun, Python Web Development with Django, 1st ed., Addison-Wesley Professional, 2008.

[11] B. Levin, H. Cheng, D. Reeder, A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection, in: 47th Annual Meeting of the American-Society-of-Human-Genetics, Baltimore, Maryland, October 28–November 01, 1997, Genomics 55 (2) (1999) 135–146. , http://dx.doi.org/10.1006/geno.1998.5513.

[12] A. Masibay, T. Mozer, C. Sprecher, Promega Corporation reveals primer sequences in its testing kits, J. Forensic Sci. 45 (6) (2000) 1360–1362.