Research Article
Volume 8:2, 2023

Journal of Forensic Medicine

ISSN: 2472-1026 Open Access

Allelic Frequencies and Statistical Parameters Determination, for 16 Autosomal STR Markers Included in the NGM Detect™ PCR Amplification Kit, for the Bulgarian Population

Plamen Iliev[1], Vera Djeliova[2], Ekaterina Angelova[1], Bogdan Mirchev[1], Milka Mileva[3], Atanas Hristov[1], Dimo Krastev[4] and Aleksandar Apostolov[1]*

[1]*Department of Forensic Medicine and Deontology, DNA Laboratory, University Hospital "Alexandrovska"-1, "St. Georg Sofiiski" blvd., Sofia 1431, Bulgaria*
[2]*Department of the Molecular Biology of the Cell Cycle, Institute of Molecular Biology "Roumen Tsanev", Bulgarian Academy of Sciences, Akad. G. Bonchev Street, Bl. 21, Sofia 1113, Bulgaria*

[3]*Department of Virology, Laboratory of Biological Response Modifiers and Pathogenesis of Viral Infections, The Stephan Angeloff Institute of Microbiology, Bulgarian Academy of Sciences, 26, Academic Georgi Bonchev Street, Sofia 1113, Bulgaria*

[4]*South-West University "Neofit Rilski"-Blagoevgrad, Bulgaria*

Abstract

In the presented study, the allelic frequencies and statistical parameters for 500 unrelated individuals from the 28 administrative regions of the Republic of Bulgaria were determined for 16 autosomal genetic markers (STR), included in the NGM Detect™ PCR Amplification Kit (Applied Biosystems), for the Bulgarian population. No deviations from Hardy-Weinberg equilibrium for the 16 studied loci were observed, except for a deviation for the D8S1179 and FGA loci. The most polymorphic and informative for the Bulgarian population is the SE marker, with the next most informative markers being: D1S1656, D12S391 and D18S51. For all 16 loci, the combined value Match Probability (MP) is 1.20e-21 and the power of exclusion (PE) is 99.999999022%.

The results obtained from the DNA analysis have a very high identification value. Their interpretation is particularly important when a DNA profile found at a crime scene matches that of a suspect or an injured person. Such statistical estimates are also valuable in detecting mixed DNA profiles, as well as in calculating the probability of kinship between two or more individuals. For these calculations, it is necessary to perform a statistical analysis of the occurrence frequency of the specified STR variants among a given population group of individuals. So far, no such research representative of the Bulgarian population has been published.

Keywords: DNA analysis • STR • Bulgarian population data • STR population database • NGM Detect™ PCR amplification kit

Introduction

Short tandem repeats (STRs) form approximately 3% of the total human genome and occur on average once in every 10,000 nucleotides [1]. When applying a DNA-based identification method, determining the allelic frequencies for the target STR loci and their statistical parameters for each population, is essential to provide a more accurate reference database for the forensic investigation. Estimating the STR allele frequencies is an essential prerequisite for this type of analysis to be applicable in forensic practice. This is because each population has its own composition of allele distribution. STR allele frequencies estimation of the target population ensures correct weight-of-evidence calculations in court practice, including when deciding cases of disputed parentage.

The goal of the present study is to determine the allele frequencies and their statistical parameters for the Bulgarian population for 16 autosomal genetic markers included in the NGM Detect™ PCR Amplification Kit: D2S1338, SE33, D16S539, D18S51, TH01, D12S391, D3S1358, FGA, vWA, D21S11, D1S1656, D2S441, D8S1179, D19S433, D22S1045, D10S124, by deriving the genetic profiles of 500 unrelated individuals from the 28 administrative districts of the Republic of Bulgaria.

With an increasing number of laboratories using STR marker technology, more and more STR data of the population is being reported in Europe and around the world, allowing regional genetic comparison of the population. The diversity of STRs in the genome is the result of DNA polymerase slippage during the replication of the DNA, or mismatch of the basic group of slippage strand and complementary strand in the process of DNA replication and repair, resulting in one or more missing repetitives or inserted units [2-4].

Materials and Methods

In the present study 500 unrelated individuals from the Bulgarian population, from the 28 administrative regions of the Republic of Bulgaria, were examined. The analyses were conducted in the DNA laboratory at the University Hospital "Alexandrovska"-Sofia and covered the period from 2017 to 2022. Declarations of informed consent were signed from the examined living persons for the samples obtained. In the declaration of informed consent, only the regional origin of the volunteer is disclosed, in compliance with the legislative requirements for control when working with personal data of by the Republic of Bulgaria and the EU. When expert samples were provided to us by the investigating authorities for performing forensic examinations for the

***Address for correspondence:** Aleksandar Apostolov, Department of Forensic Medicine and Deontology, DNA Laboratory, University Hospital "Alexandrovska"-1, "St. Georg Sofiiski" blvd., Sofia 1431, Bulgaria, E-mail: alexa2000@mail.bg*

Received: 15 March, 2023, Manuscript No. JFM-23-91785; **Editor assigned:** 17 March, 2023, PreQC No. P-91785; **Reviewed:** 29 March, 2023, QC No. Q-91785; **Revised:** 04 April, 2023, Manuscript No. R-91785; **Published:** 12 April, 2023, DOI: 10.37421/2472-1026.2023.8.192

identification of the perpetrator of a criminal act, or the establishment of kinship or the identification of corpses of unknown identity, these forms were not applicable. The research has been conducted in accordance with the Helsinki declaration and approved by the ethics committee of the Medical University of Sofia (Protocol No 06/16.05.2022).

Sixteen autosomal genetic markers (STR) included in the NGM Detect™ PCR Amplification Kit (Applied Biosystems) were used to determine the allelic frequencies and statistical parameters of the 500 said individuals: D2S1338, SE33, D16S539, D18S51, TH01, D12S391, D3S1358, FGA, vWA, D21S11, D1S1656, D2S441, D8S1179, D19S433, D22S1045, D10S124.

The DNA profiles were isolated by Chelex100 extraction [5] from buccal mucosa samples of living individuals and from biological samples taken during routine forensic practice.

Multiplex PCR amplification was performed with 0.5 ng of genomic DNA for 16 somatic loci and the Amelogenin sex-determining system, according to the manufacturer's instructions, performed in two steps at:

- Real-Time PCR system 7500 (Life Technologies) with PC Notebook, for HID analyses-with Quantifiler™ Trio DNA Quantification Kit and quantitative evaluation of the DNA present in the samples by HID Real-Time PCR Analysis Software v1.2.
- PCR apparatus SimpliAmp™ Thermal Cycler, 96 × 0.2 ml (Life Technologies) in a volume of 25 µl with NGM Detect™ PCR Amplification Kit (Applied Biosystems) for the analysed autosomal and sex-defining STR markers.

Fragment analysis was performed on a Genetic Analyzer model 3500 Series Genetic Analyzers for Human Identification (Life Technologies) by 8 capillary electrophoresis (with 3500 POP- 4™ Polymer) with laser detection of the fragments and computer analysis using Gene MapperTM v1.2 Full Software (Life Technologies) for HID analysis.

The control and standardization of analyzes was carried out by:

- Positive control-DNA Control 007; negative control-NS; Matrix Standart DS-33 3500 Series (6- FAM™, VIC™, NED™, PET™, LIZ™ dyes); internal standard-GeneScan™ 600 LIZ ™ Size Standart v2.0; internal quality control markers-IQCS and IQCL; an allelic witness (NGM Detect™ Allelic Ladder) for the relevant STR markers, validated and embedded in the NGM Detect™ Kit (Applied Biosystems).

Statistical analyzes of the generated DNA profiles were performed with analysis programs: FORSTAT- Forensic statistics analysis toolbox (https://fdl-uwc.shinyapps.io/forstat) [6] and STRAF software Version 1.0.5. http://cmpg.unibe.ch/shiny/STRAF/ [7,8].

The following were analysed for: expected heterozygosity (H exp), observed heterozygosity (H obs), homozygosity (Ho), power of discrimination (PD), polymorphism information content (PIC), match probability (MP), exclusion power (PE), typical paternity index (TPI) and for a test of Hardy-Weinberg equilibrium (pHW).

Results

For the statistical analysis, 500 STR profiles were used, taken from samples of buccal mucosa of living persons and expert samples taken during routine forensic practice from the 28 administrative regions of the Republic of Bulgaria. The research was carried out in the DNA laboratory at the Department of Forensic medicine and Deontology at the University Hospital "Aleksandrovska"- Sofia. In order to enter the generated DNA profiles into an information array, the Familias software, v.3.2.8 (https://familias.no) [7] was used.

The allele frequencies of 16 autosomal genetic markers (STR) for the Bulgarian population (n= 500), calculated using statistical programs, are presented in tabular and graphical form, respectively, in Figure 1 and Figure 2 and the calculated statistical parameters are summarized in Figure 3.

When analyzing the established allelic variants of the analyzed markers for the Bulgarian population, it was established that markers SE33, FGA, D21S11, D12S391, D19S433 and D1S1656 show greater variability. Marker SE33 was found to present 47 different alleles. FGA and D21S11 presented 18 allelic variants and markers D12S391, D19S433 and D1S1656 presented 17 respectively. On the other hand, those with the lowest number of alleles detected in the population were the loci: TH01 with six alleles and D16S539 with eight allelic variants.

The average value of Hobs is 0.8291 and the value of Hesp is 0.8236. The highest Hobs index was obtained in marker SE33 (0.952), followed by D18S51 (0.9) and D1S1656 (0.896). With the highest value of Hesp was SE33 (0.9459), followed by D1S1656 (0.898) and D12S391 (0.8835). As for the lowest indices obtained, in the Hobs index these are markers D2S441 and D22S1045 (0.752) and the same markers in the Hesp index: D2S441 (0.7433) and D22S1045 (0.7311).

PD is characterized by the probability that two unrelated and randomly selected individuals can be genetically differentiated by an analysis of a marker or set of markers. Analyzing the results shown in Figure 1, it can be seen that the PD index for marker SE33 is 0.9927, which is the highest, followed by D1S1656 (0.9795) and D12S391 (0.9738). The markers with the lowest PD index were D2S441 (0.891) and D22S1045 (0.8889). The average PD of the markers is 0.939875.

The PE parameter allows us to determine the proportion of individuals falsely included in an expert assessment. The most informative marker was found to be SE33 (0.9023), followed by D18S51 (0.7954) and D1S165 (0.7873), while the least informative markers were found to be D22S1045 and D2S441 (0.5132). The combined Power of Exclusion (PE) is 0.99999999022.

The three most informative STRs for the Bulgarian population, due to their PIC, were found to be SE33 (0.9433), followed by D1S1656 (0.889) and D12S391 (0.8725). The markers with the lowest index were D22S1045 (0.6898) and D2S441 (0.7025). Allele frequencies of all 16 studied autosomal STR markers have PIC values higher than 0.5, which meets the condition of being sufficiently informative for the studied population [9].

Match Probability (MP) values ranged from 0.0073 (SE) to 0.1111 (DS221045) with a combined value of 1.20e-21. The marker with the highest match probability index was D22S1045 (0.1111), followed by D2S441 (0.1090) and D10S1248 (0.0897), while the marker with the lowest index was SE33 (0.0073), followed by D1S1656 (0.0205) and D12S391 (0.0262).

Typical Paternity Index (PItypical) ranges from 2.0161 (DS441 and DS221045) to 10.4167 (SE33).

The Hardy-Weinberg equilibrium test (p-value>0.05) showed a bias in equilibrium in two genetic markers: D8S1179 (p=0.003) and FGA (p=0.011). However, this deviation was not significant after applying the Bonferroni correction [10].

Discussion

After the analysis of the 16 STR markers contained in the NGM Detect™ PCR Amplification Kit, it was found that the most polymorphic and informative locus for Forensic Identification of the Bulgarian population is SE33. This marker has the highest PIC, PD, PE, PI and the lowest frequency of MP and homozygosity. For the Bulgarian population, the next most informative markers are: D12S391, D1S1656 and D18S51. These markers have the highest PIC, PD, PE, PI and the lowest value of MP and homozygosity. The lowest levels of PIC, PD, PE, PI and the highest values of MP and homozygosity were found for two markers: D2S441 and D22S1045. All studied loci except D8S1179 and FGA were in the Hardy-Weinberg equilibrium.

The results obtained from the DNA analysis have a very high identification value and their interpretation is particularly important when a DNA profile found and seized from a scene of a crime matches that of a suspect or injured person. In these cases, it is mandatory to calculate what is the statistical probability that another, randomly selected person from the population,

Allele	D2S1338	SE33	D16S539	D18S51	TH01	D12S391	D3S1358	FGA	vWA	D21S11	D1S1656	D2S441	D8S1179	D19S433	D22S1045	D10S1248
6					0.273											
7					0.15											
8			0.019		0.118								0.014			
9			0.153		0.192								0.015	0.001		
9.3					0.257											
10		0.001	0.082	0.005	0.01						0.006	0.16	0.064	0.001		
10.2		0.001														
11			0.301	0.012							0.106	0.329	0.085	0.003	0.143	0.002
11.3												0.093				
12			0.258	0.134			0.002		0.001		0.117	0.033	0.106	0.095	0.013	0.02
12.2														0.002		
12.3												0.001				
13		0.006	0.153	0.139			0.001		0.002		0.132	0.017	0.339	0.295	0.002	0.202
13.2														0.011		
14		0.026	0.027	0.173			0.084		0.119		0.067	0.335	0.225	0.306	0.065	0.321
14.2		0.003												0.024		
14.3											0.002			0.002		
15	0.001	0.041	0.007	0.134		0.039	0.257		0.105		0.129	0.022	0.114	0.133	0.393	0.248
15.2		0.001												0.05		
15.3											0.051					
16	0.06	0.038		0.157		0.016	0.257	0.002	0.193		0.136	0.01	0.032	0.039	0.286	0.154
16.2		0.001												0.03		
16.3		0.002									0.051					
17	0.216	0.057		0.117		0.072	0.226	0.001	0.309		0.038		0.004	0.001	0.088	0.048
17.2														0.005		
17.3		0.002				0.01					0.105					
18	0.09	0.085		0.051		0.189	0.163	0.005	0.19		0.004		0.002		0.009	0.004
18.2		0.002												0.002		
18.3						0.021					0.047					
19	0.118	0.088		0.035		0.155	0.009	0.086	0.067		0.001				0.001	0.001
19.2		0.001														
19.3		0.001				0.019					0.007					
20	0.151	0.066		0.026		0.123	0.001	0.107	0.013							
20.2		0.005														
20.3						0.001					0.001					
21	0.035	0.038		0.009		0.109		0.177	0.001							
21.2		0.006						0.004								
21.3								0.002								
22	0.03	0.008		0.006		0.124		0.187								
22.2		0.033						0.005								
23	0.103	0.002		0.002		0.072		0.142								
23.1								0.001								
23.2		0.035						0.003								
24	0.097	0.002				0.034		0.162								
24.2		0.033														
25	0.083					0.012		0.08		0.001						
25.2		0.035														
26	0.014	0.001				0.003		0.028		0.008						
26.2		0.032														
27		0.001				0.001		0.007		0.032						
27.2		0.055														
28	0.001							0.001		0.146						
28.2		0.086														
29	0.001	0.006								0.234						
29.2		0.083								0.007						
29.3										0.001						
30										0.179						
30.2		0.043								0.043						
31		0.001								0.06						
31.1										0.001						
31.2		0.028								0.116						
31.3										0.001						
32										0.003						
32.2		0.018								0.12						
33		0.006								0.003						
33.2		0.007								0.035						
34		0.003														
34.2		0.003								0.01						
35		0.002														
35.2		0.003														
36		0.002														
36.2		0.001														

Figure 1. Allelic frequencies of 16 STR loci in the Bulgarian population (n=500).

apart from the specified one, has the same profile. Such statistical estimates are also valuable in the detection of mixed DNA profiles from two or three individuals, the identification of unidentified individuals and the calculation of the probability of kinship. For these calculations, it is necessary to perform a statistical analysis of the occurrence frequency of the specified variants of the short tandem repeats among a given population group of individuals. As until now, such representative study of the Bulgarian population has not yet been published.

Allele	D2S1338	SE33	D16S539	D18S51	TH01	D12S391	D3S1358	FGA	vWA	D21S11	D1S1656	D2S441	D8S1179	D19S433	D22S1045	D10S1248
N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
N alleles	14	47	8	14	6	17	9	18	10	18	17	9	11	17	9	9
He exp	0.8757	0.9459	0.7882	0.8717	0.7860	0.8835	0.7831	0.8611	0.8013	0.8561	0.8980	0.7433	0.7974	0.7870	0.7311	0.7682
He obs	0.8940	0.9520	0.7820	0.9000	0.7860	0.8780	0.8080	0.8680	0.7840	0.8700	0.8960	0.7520	0.8000	0.7820	0.7520	0.7620
Ho	0.1060	0.0480	0.2180	0.1000	0.2140	0.1220	0.1920	0.1320	0.2160	0.1300	0.1040	0.2480	0.2000	0.2180	0.2480	0.2380
PD	0.9701	0.9927	0.9234	0.9677	0.9198	0.9738	0.9148	0.9619	0.9324	0.9618	0.9795	0.891	0.9256	0.9243	0.8889	0.9103
PIC	0.8635	0.9433	0.7571	0.858	0.7522	0.8725	0.7482	0.8454	0.774	0.8403	0.889	0.7025	0.7725	0.7583	0.6898	0.7311
MP	0.0299	0.0073	0.0766	0.0323	0.0802	0.0262	0.0852	0.0381	0.0676	0.0382	0.0205	0.1090	0.0744	0.0757	0.1111	0.0897
PE	0.7832	0.9023	0.5661	0.7954	0.5733	0.7507	0.614	0.7306	0.5697	0.7346	0.7873	0.5132	0.599	0.5661	0.5132	0.5305
TPI	4.717	10.4167	2.2936	5	2.3364	4.0984	2.6042	3.7879	2.3148	3.8462	4.8077	2.0161	2.5	2.2936	2.0161	2.1008
pHW	0.2010	0.8400	0.7890	0.9550	0.7430	0.2950	0.6810	0.0110	0.1920	0.3650	0.4280	0.7270	0.0030	0.5700	0.4700	0.8660

Figure 2. Statistical parameters of 16 STR loci in the Bulgarian population (n=500). N: number of alleles for 500 individuals in locus, N all: number of observed alleles in locus, Hexp: Expected Heterozygosity Ho: Observed Heterozygosity; Ho: Homozygosity; PD: Power of Discrimination, PIC: Polymorphism Information Content, MP: Matching Probability; PE: Power of Exclusion; TPI: Typical Paternity Index, pHW: Hardy-Weinberg equilibrium exact test (number of permutations: 1000).

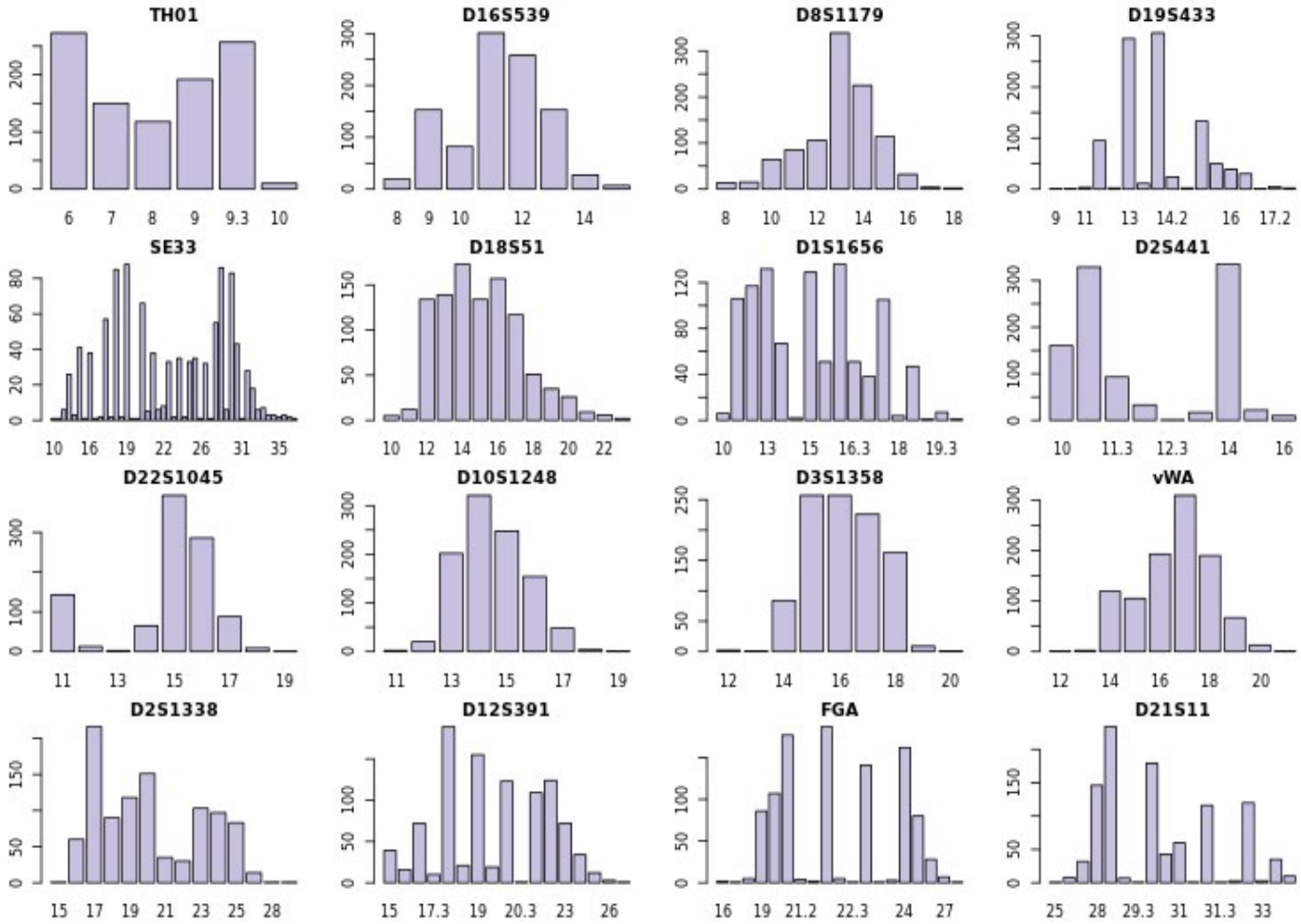

Figure 3. Allelic frequencies of 16 STR loci in the Bulgarian population (n=500).

With the growing number of laboratories that use STR technology, more and more data collected in Europe and around the world is being reported on the ST distribution among the population, which allows comparison between regional populations, including the Bulgarian one.

Conclusion

Of the 16 genetic markers examined, the most polymorphic and informative locus for forensic identification was found to be SE33. For the Bulgarian population, the next most informative markers are: D1S1656, D12S391 and D18S51. The lowest levels of PIC, PD, PE, PI and the highest values of MP and homozygosity were found for two markers: D2S441 and D22S1045. All studied loci except D8S1179 and FGA were in Hardy-Weinberg equilibrium. However, this deviation was not significant after applying the Bonferroni correction.

We believe that by carrying out a DNA population study for the Bulgarian population for the frequency of occurrence of alleles and calculation of statistical parameters in established genetic loci, as established in world practice, that that significally contributes to the collection and construction of the world DNA characteristics library of the population. Such library is of great value for the purposes of forensic and criminological investigations for the identification of corpses with unknown identity, as well as for the resolution of cases of disputed parentage. Thus, this data should be popularized and be used by other countries when conducting comparative studies of DNA profiles of persons from the Bulgarian population. This data is especially valuable due to the possibility of free movement of the Bulgarian citizens and when combined with the data from relatively significant Bulgarian diasporas in various countries around the world, that would, in cases of need, be of essential help when comparative identification studies must be conducted. The availability of DNA

population data will ensure the accuracy of the mathematical calculations of the results of these studies, as well as their probative value.

Funding

This study was realized thanks to a program for financing scientific research - competition "GRAND- 2022" of the Medical University - Sofia, Bulgaria /Contract № D- 176/14.06.2022/.

Conflicts of Interest

The authors declare that they have no competing interests. The funders had no role in the study design; in the collection, analysis or interpretation of data; in writing the manuscript; or in the decision to publish the results.

Acknowledgements

None.

References

1. Butler, John M. “Forensic DNA typing: Biology, technology and genetics of STR markers.” *Elsevier* (2005).
2. Fan, Hao and Jia-You Chu. "A brief review of short tandem repeat mutation." *Genom, Proteom Bioinformat* 5 (2007): 7-14.
3. Wang, Li-Feng, Ying Yang, Xiao-Nan Zhang and Xiao-Liang Quan, et al. "Tri-allelic pattern of short tandem repeats identifies the murderer among identical twins and suggests an embryonic mutational origin." *Foren Sci Internat: Genet* 16 (2015): 239-245.
4. Gharesouran, Jalal, Hassan Hosseinzadeh, Soudeh Ghafouri-Fard and Mohammad Taheri, et al. "STRs: Ancient architectures of the genome beyond the Sequence." *J Mol Neurosci* 71 (2021): 2441-2455.
5. Walsh, P Sean, David A Metzger and Russell Higuchi. "Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material." *Biotechn* 10 (1991): 506-513.
6. Ristow, Peter G. and Maria E D’amato. "Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics." *Foren Sci Internat: Genet Suppl Ser* 6 (2017): e52-e54.
7. Kling, Daniel, Andreas O Tillmar and Thore Egeland. "Familias 3–extensions and new functionality." *Foren Sci Internat: Genet* 13 (2014): 121-127.
8. Gouy, Alexandre and Martin Zieger. "STRAF-a convenient online tool for STR data evaluation in forensic genetics." *Foren Sci Internat: Genet* 30 (2017): 148-151.
9. Crow, J.F. and M Kimura. "An Introduction to Population Genetics Theory. New York: Harper and Row." (1970): 591.
10. Bonferroni, Carlo. "Teoria statistica delle classi e calcolo delle probabilita." *Pub del R Ist Sup di Sci Econom e Commer di Firenze* 8 (1936): 3-62.

How to cite this article: Iliev, Plamen, Vera Djeliova, Ekaterina Angelova and Bogdan Mirchev, et al. “Allelic Frequencies and Statistical Parameters Determination, for 16 Autosomal STR Markers Included in the NGM Detect™ PCR Amplification Kit, for the Bulgarian Population.” *J Forensic Med* 8 (2023): 192.