Int J Legal Med (2011) 125:101–109
DOI 10.1007/s00414-010-0516-7

TECHNICAL NOTE

Population genetic analyses of the NGM STR loci

Bruce Budowle · Jianye Ge · Ranajit Chakraborty · Arthur J. Eisenberg · Robert Green · Julio Mulero · Robert Lagace · Lori Hennessy

Received: 6 August 2010 / Accepted: 14 September 2010 / Published online: 29 September 2010

Abstract The AmpFlSTR® NGM™ PCR Amplification Kit enables amplification of 15 autosomal short tandem repeat (STR) loci. The loci are the ten STRs in the SGM Plus® Kit plus the EDNAP and ENSFI recommended STRs D10S1248, D22S1045, D2S441, D1S1656, and D12S391. Allele frequency and other forensically relevant statistics data were generated for the NGM loci in three US population groups (African Americans, Caucasians, and Hispanics). The analyses support that the NGM multiplex is one of the most informative STR multiplex kits available to the forensic science community. At the population level, there are no more detectable departures from expectations of the independence of alleles within as well as between loci than would be expected due to chance, even for the two syntenic loci vWA and D12S391; however, linkage analysis in three large pedigree families shows close linkage between these two loci with a recombination fraction of 0.108. Therefore, in contrast to the practices in calculating the rarity of a DNA profile, for kinship analyses independence between the loci, vWA and D12S391 cannot be assumed.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-010-0516-7) contains supplementary material, which is available to authorized users.

B. Budowle (✉) · J. Ge · R. Chakraborty · A. J. Eisenberg
Institute of Investigative Genetics,
University of North Texas Health Science Center,
Fort Worth, TX 76107, USA
e-mail: bbudowle@hsc.unt.edu

B. Budowle · J. Ge · R. Chakraborty · A. J. Eisenberg
Department of Forensic and Investigative Genetics,
University of North Texas Health Science Center,
Fort Worth, TX 76107, USA

R. Green · J. Mulero · R. Lagace · L. Hennessy
Life Technologies Corporation,
Foster City, CA 94404, USA

Keywords Short tandem repeat loci · Next generation multiplex · vWA · D12S391 · Population statistics · Linkage

Introduction

Multiplex amplification of short tandem repeat (STR) loci offers several benefits to the forensic analysis of biological evidence. Primarily, it affords the ability to type a battery of genetic markers while consuming only one aliquot of a sample. Thus, less sample is consumed compared with that which would be required for typing each locus separately. Generally, target amounts of template DNA are at a minimum of 100–250 pg of DNA. The more loci that are contained within a multiplex the more discriminating is that multiplex. Additionally, amplifying all the desired loci in one tube reduces manipulations, which saves on labor and reduces the chances of contamination. Lastly, a multiplex system facilitates automation of the analytical process.

There are seven required STRs for the European Standard Set (ESS) of loci. These are TH01, vWA, FGA, D21S11, D3S1358, D8S1179, and D18S51 [1]. These seven loci form the core loci for DNA databases and hence casework analyses in Europe; however, given the expansion of DNA data sharing in Europe fostered by the Prum Treaty [2], an expansion of the seven core loci was needed to meet the demands for the exchange of DNA profile data across the European Union. To improve the discrimination power, a multiplex kit known as the AmpFlSTR® NGM™ PCR Amplification Kit (NGM kit; Life Technologies, Foster City, CA) was developed. The kit contains the ten STRs in the SGM Plus® Kit (which include the seven ESS loci) and five additional loci, D10S1248, D22S1045, D2S441, D1S1656, D12S391, that were recommended by the European Network of Forensic Science Institutes and the European DNA Profiling group EDNAP [1].

Table 1 African American population data for the NGM STR loci (N=350)

	D10S1248	D12S391	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D8S1179	FGA	TH01	vWA
5														0.004	
6														0.151	
6.1														0.001	
7	0.001													0.373	
8			0.034					0.009		0.001		0.004		0.211	
9	0.001		0.221		0.003	0.001					0.003	0.003		0.159	
9.3														0.086	
10	0.001		0.116	0.003	0.011	0.013		0.040		0.090		0.033		0.014	
10.2				0.001	0.003										
11	0.039		0.293	0.004	0.099	0.054		0.144		0.353		0.057			0.004
11.2					0.003										
11.3										0.033					
12	0.137		0.193	0.063	0.107	0.081		0.063	0.001	0.200	0.001	0.111			
12.1															
12.2					0.039										
12.3										0.001					
13	0.224		0.130	0.040	0.284	0.111		0.003		0.036	0.006	0.183			0.009
13.2				0.003	0.051										
14	0.284		0.013	0.057	0.187	0.246		0.086		0.267	0.091	0.360			0.073
14.2				0.004	0.057										
14.3						0.010									
15	0.189	0.073		0.171	0.063	0.173		0.226	0.003	0.019	0.283	0.177			0.207
15.1		0.001													
15.2					0.041						0.003				
15.3						0.019									
16	0.099	0.050		0.181	0.014	0.101		0.203	0.051		0.329	0.061			0.280
16.1		0.001											0.003		
16.2					0.033										
16.3						0.074									
17	0.021	0.160		0.159		0.026		0.206	0.101		0.220	0.010			0.193
17.1		0.004													
17.2					0.006										
17.3		0.006				0.059									
18	0.003	0.256		0.141		0.004		0.021	0.049		0.059		0.009		0.137
18.2													0.007		
18.3		0.011				0.019									
19		0.140		0.093		0.001			0.163		0.006		0.067		0.070
19.1		0.006													
19.2													0.004		

19.3		0.003				0.006									
20		0.117		0.044					0.100				0.067		0.020
20.2				0.001									0.003		
20.3						0.001									
21		0.070		0.020					0.127				0.119		0.006
21.2													0.001		
21.3		0.001													
22		0.051		0.010					0.124				0.180		
22.2													0.001		
23		0.034		0.003					0.093				0.169		0.001
23.2															
23.3													0.003		
24		0.006							0.087				0.190		
24.2															
25		0.009							0.069				0.094		
26							0.003		0.024				0.040		
27							0.057		0.007				0.027		
27.1							0.001								
28							0.253						0.011		
28.2															
29							0.156								
29.2															
29.3							0.001								
30							0.207						0.001		
30.2							0.016						0.001		
31							0.089								
31.2							0.049								
32							0.013								
32.2							0.071								
33							0.009								
33.2							0.031								
34							0.001								
34.2													0.001		
35							0.035								
36							0.007								
Obs Het	0.169	0.129	0.223	0.106	0.180	0.126	0.134	0.163	0.100	0.243	0.266	0.206	0.134	0.249	0.200
Exp Het	0.196	0.140	0.202	0.126	0.150	0.132	0.151	0.167	0.104	0.245	0.247	0.214	0.130	0.238	0.187
HWE[a]	0.770	0.267	0.145	0.805	0.003	0.281	0.277	0.162	0.806	0.649	0.864	0.768	0.883	0.181	0.568
PD	0.930	0.961	0.928	0.969	0.960	0.966	0.955	0.946	0.977	0.899	0.900	0.925	0.967	0.905	0.938
PE	0.612	0.722	0.600	0.744	0.708	0.738	0.701	0.664	0.787	0.533	0.524	0.595	0.736	0.547	0.630

[a] *P* values for detection of departures from HWE based on exact test

Obs Het observed heterozygosity, *Exp Het* expected heterozygosity

Table 2 Caucasian population data for the NGM STR loci (N=349)

	D10S1248	D12S391	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D8S1179	FGA	TH01	vWA
5														0.001	
6														0.219	
6.1															
7														0.176	
8			0.014									0.020		0.115	
9			0.128							0.006		0.013		0.172	
9.3														0.308	
10			0.039	0.011	0.001	0.003		0.004		0.199		0.107		0.009	
10.2															
11	0.006		0.319	0.009		0.063		0.138		0.337	0.003	0.066			
11.2															
11.3										0.050					
12	0.034		0.309	0.146	0.072	0.159		0.006		0.040		0.150			
12.1					0.001										
12.2					0.001										
12.3										0.003					
13	0.295		0.172	0.120	0.271	0.069		0.010	0.001	0.033	0.001	0.335			0.001
13.2					0.016										
14	0.295		0.017	0.178	0.350	0.062		0.034	0.001	0.284	0.150	0.188			0.086
14.2					0.021										
14.3						0.003									
15	0.195	0.043	0.001	0.153	0.165	0.150		0.364	0.001	0.044	0.269	0.087			0.120
15.1															
15.2					0.034										
15.3															
16	0.135	0.034		0.119	0.059	0.099		0.364	0.042	0.004	0.242	0.029	0.001		0.225
16.1															
16.2					0.003										
16.3						0.047									
17	0.037	0.105		0.109	0.003	0.050		0.075	0.188		0.201	0.004	0.001		0.274
17.1						0.003									
17.2					0.001										
17.3		0.020				0.126									
18	0.003	0.163		0.083	0.001	0.003		0.006	0.083		0.119		0.012		0.182
18.2															
18.3		0.021				0.059									
19		0.123		0.043					0.140		0.014		0.056		0.099
19.1															
19.2															

19.3		0.006				0.017									
20		0.097		0.013					0.156				0.155		0.013
20.2													0.004		
20.3		0.001													
21		0.136		0.010					0.027				0.181		
21.2													0.003		
21.3		0.001													
22		0.106		0.003					0.017				0.193		
22.2													0.009		
23		0.082		0.003					0.100				0.146		
23.2													0.004		
23.3															
24		0.036							0.103				0.143		
24.2															
25		0.020							0.119				0.066		
26		0.003					0.006		0.017				0.019		
27		0.001					0.027		0.003				0.006		
27.1															
28							0.169						0.001		
28.2															
29							0.238								
29.2							0.001								
29.3							0.001								
30							0.234								
30.2							0.027								
31							0.067								
31.2							0.087								
32							0.023								
32.2							0.095								
33															
33.2							0.020								
34															
34.2							0.004								
Obs Het	0.215	0.092	0.209	0.120	0.252	0.095	0.132	0.284	0.123	0.261	0.238	0.192	0.166	0.252	0.172
Exp Het	0.232	0.103	0.245	0.125	0.232	0.101	0.162	0.290	0.123	0.240	0.207	0.194	0.143	0.216	0.189
HWE[a]	0.682	0.380	0.276	0.785	0.448	0.269	0.311	0.213	0.182	0.048	0.467	0.539	0.285	0.429	0.697
PD	0.905	0.977	0.896	0.969	0.915	0.978	0.947	0.868	0.968	0.903	0.925	0.936	0.961	0.920	0.934
PE	0.549	0.789	0.531	0.744	0.562	0.793	0.679	0.468	0.749	0.544	0.587	0.628	0.710	0.574	0.624

[a] *P* values for detection of departures from HWE based on exact test

Table 3 Hispanic population data for the NGM STR loci (N=335)

	D10S1248	D12S391	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D8S1179	FGA	TH01	vWA
5															
6														0.257	
6.1															
7														0.313	
8	0.001		0.019									0.006		0.097	
9	0.001		0.106	0.001		0.001				0.003	0.001	0.003		0.144	
9.3														0.187	
10			0.152	0.004	0.004	0.004		0.006		0.297		0.093		0.013	
10.2															
11	0.003		0.321	0.013	0.015	0.039		0.082		0.315		0.051			0.001
11.2					0.001										
11.3										0.045					
12	0.046		0.242	0.112	0.087	0.099		0.012		0.042	0.001	0.115			0.003
12.1															
12.2					0.015										
12.3										0.004					
13	0.254	0.001	0.143	0.112	0.201	0.078		0.012		0.018	0.004	0.328			0.001
13.2					0.064										
14	0.355	0.001	0.013	0.148	0.297	0.113		0.022		0.236	0.094	0.249			0.067
14.2				0.001	0.042										
14.3						0.003									
15	0.224	0.045	0.003	0.119	0.125	0.151		0.415		0.036	0.331	0.116			0.103
15.1															
15.2				0.001	0.076										
15.3						0.028									
16	0.088	0.055		0.136	0.043	0.154		0.361	0.036	0.004	0.264	0.031			0.297
16.1						0.003									
16.2					0.022										
16.3						0.055									
17	0.027	0.064		0.164	0.001	0.067		0.082	0.182		0.181	0.007			0.270
17.1		0.003													
17.2					0.004										
17.3		0.012				0.149									
18		0.201		0.093		0.006		0.006	0.066		0.112		0.006		0.185
18.2															
18.3		0.022				0.042									
19		0.185		0.037					0.163		0.010		0.075		0.063
19.1															
19.2															
19.3		0.012				0.007									

20		0.161		0.018				0.001	0.142				0.091		0.009
20.2													0.003		
20.3															
21		0.088		0.024					0.042				0.139		
21.2															
21.3															
22		0.075		0.006					0.060				0.142		
22.2													0.006		
23		0.036		0.006					0.142				0.128		
23.2													0.001		
23.3															
24		0.019		0.003					0.094				0.160		
24.2							0.003								
25		0.013							0.058				0.143		
26		0.003					0.003		0.015				0.064		
27		0.001					0.013		0.001				0.031		
27.1															
28							0.116						0.004		
28.2							0.001								
29							0.215						0.004		
29.2															
29.3															
30							0.285						0.001		
30.2							0.016								
31							0.042								
31.2							0.113								
32							0.012								
32.2							0.127								
33							0.001								
33.2							0.045								
34															
34.2							0.001								
35							0.001								
36							0.001								
38							0.001								
Obs Het	0.263	0.167	0.236	0.113	0.179	0.137	0.161	0.313	0.125	0.266	0.266	0.224	0.152	0.218	0.185
Exp Het	0.250	0.125	0.216	0.116	0.165	0.108	0.173	0.316	0.122	0.247	0.233	0.208	0.120	0.225	0.213
HWE[a]	0.267	0.060	0.818	0.528	0.283	0.059	0.676	0.011	0.121	0.422	0.466	0.866	0.032	0.797	0.128
PD	0.891	0.969	0.921	0.972	0.950	0.975	0.947	0.844	0.969	0.898	0.910	0.932	0.971	0.913	0.912
PE	0.523	0.750	0.580	0.761	0.680	0.780	0.660	0.437	0.752	0.526	0.550	0.602	0.754	0.560	0.584

[a] *P* values for detection of departures from HWE based on exact test

With the introduction of this new multiplex, standard population genetic parameters need to be assessed for utility in estimating the rarity of an evidentiary DNA profile. This study describes population statistics on the 15 NGM kit loci in three US population groups (African Americans, Caucasians, and Hispanics). In addition, a new issue needs to be addressed that has been of little relevance for past multiplex kits. That is, the loci vWA and D12S391 are syntenic. The vWA locus is on chromosome 12 at the physical position 6,093 kb and genetic position 17 cM, and the locus D12S391 is on chromosome 12 at the physical position 12,340 kb and genetic position 28.8 cM. Thus, although they are physically approximately 6 Mb apart, their recombination distance is about 12 cM from each other [3]. Linkage analysis was performed on these two loci in three large pedigree CEPH families to assess the utility of both loci in kinship analysis.

Materials and methods

A population study was performed with a set of 1,034 unrelated, healthy, anonymous human DNA samples (350 African Americans, 349 Caucasians, and 335 Hispanics) obtained as whole blood from the Interstate Blood Bank, Inc. (Memphis, TN) or Boca Biolistics (Coconut Creek, FL) and purified on an Applied Biosystems 6100 PrepStation. Quantification of DNA samples was performed using the Quantifiler® Human DNA Quantification Kit (Applied Biosystems, Foster City, CA, USA). DNA samples from three CEPH families (1333; 1340; and 1345) were purchased from the Coriell Medical Institute for Research (http://ccr.coriell.org/sections/collections/nigms/CEPHFamilyKits.aspx?PgId=526&coll=GM). The extracted DNA was amplified using the AmpFlSTR® NGM™ PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA) in the GeneAmp 9700 PCR system (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's recommendations. The analysis of the amplified PCR product was performed on the ABI 3130xL Genetic Analyzer (Applied Biosystems, Foster City, USA). The data analysis and allele identification were performed using GeneMapper® ID (version 3.2) analysis software.

Statistical analyses

Allele designations were determined by comparison of the amplified fragments with those of the allelic ladders. Estimates of allele frequencies for each locus were calculated by the gene count method [4]. Unbiased estimates of expected heterozygosity were computed [5]. The power of discrimination (PD) and power of exclusion (PE) were calculated according to Fisher [6]. Possible deviations from expectations of the Hardy–Weinberg equilibrium (HWE) were tested by the exact test [7], based on 2,000 shuffling experiments. The computer program to perform these tests was developed by R. Chakraborty (University of North Texas Health Science Center, Fort Worth, TX). An inter-class correlation criterion [8] was used for detecting correlations between alleles at any pair of loci for the 15 loci in each sample population. The values for F_{ST} were determined as described by Weir and Cockerham [9]. Recombination rate between the syntenic loci vWA and D12S391 was estimated using the software program LINKAGE [10, 11].

Results and discussion

The allele frequencies and population statistics parameters for the 15 NGM STR loci in African Americans, Caucasians, and Hispanics are shown in Tables 1, 2 and 3, respectively. All loci were highly polymorphic in all three populations. The number of loci departing from HWE was no more than would be expected by chance. The overall PD was greater than 0.99999999 and the overall PE was greater than 0.99999985 in all three populations. Thus, the NGM kit appears to be the most genetically informative of the currently available commercial multiplex kits.

Since there is only one study reporting population data in which all were 15 NGM loci [12], F_{ST} values for the 15 NGM STR loci were estimated using the three US population groups reported herein (Table 4). Typically, African Americans, US Caucasians, and Hispanics would not be combined into a single reference population data set and therefore any F_{ST} estimate based on these three

Table 4 F_{ST} values based on three US population samples

D10S1248	0.0080
D12S391	0.0099
D16S539	0.0131
D18S51	0.0129
D19S433	0.0194
D1S1656	0.0178
D21S11	0.0123
D22S1045	0.0343
D2S1338	0.0114
D2S441	0.0250
D3S1358	0.0059
D8S1179	0.0230
FGA	0.0068
THO1	0.0359
vWA	0.0083
Avg	0.0161

population groups would be an overly conservative value, as observed for the CODIS STR loci [13, 14]. The estimate over all 15 STR loci (F_{ST}=0.0161) is well below the value of 0.03 recommended by the National Research Council [15] and consistent with the results of Phillips et al. [12]. These data support that an F_{ST} value of 0.03 (as recommended in [15]), is conservative and a value of 0.01 is also likely to be conservative for these three population data sets used separately.

The test for evidence of linkage disequilibrium detected no more departures than would be expected by chance (four, six, and two out of 105 pairs of tests, in African Americans, Caucasians, and Hispanics, respectively). Furthermore, there were no departures detected for the two syntenic loci vWA and D12S319 (even before the Bonferroni correction [16]). This observation is consistent with the findings of Phillips et al. [12] and supports that for identity testing multiplying the genotype frequencies is justified for the loci vWA and D12S391. The independence between these two loci at the population level-based statistics is likely the result of the relatively high mutation rates of STR loci; however, for kinship analyses vWA and D12S391 may be linked; they are located on chromosome 12, at positions 6,093 kb and 12,340 kb, respectively [3]. The genetic distance is greater between these two loci at about 12 cM; however, while all physically linked loci are syntenic, physical map distances may not necessarily relate one-to-one with the recombination rate (e.g., due to presence of a recombinational hotspot). Thus, empirical testing is necessary to assess the degree of linkage effect (i.e., extent of co-segregation of alleles) for these two loci.

Three large pedigree CEPH families were typed for the NGM STR loci (ESM Table 5) and a linkage analysis was performed for the two biologically linked loci, vWA and D12S391. Using the software LINKAGE [10, 11], the recombination fraction between the loci vWA and D12S391 was estimated at 0.108. This result is close to the recombination distance (12 cM) and confirms the linkage of the loci vWA and D12S391. The recombination fraction can be used for estimating the co-inheritance of alleles at these two loci. Since the loci vWA and D12S391 are linked, independence cannot be assumed for kinship analyses. Therefore, there are two options recommended for the use of genetic data for kinship analyses: (1) Incorporate the recombination rate and generate maximum likelihood estimates of the haplotype frequencies for these two loci; or (2) use only one of the two loci in a kinship analysis, which would be the more informative of the two in a specific case.

In conclusion, population data for the 15 NGM STR loci in three US population groups are available for forensic testing and genetic studies. The data support that the loci are highly polymorphic and can be multiplied together (following the recommendations of the NRCII Report [15]) to derive an estimate of the rarity of a multilocus DNA profile for direct comparison single source and mixture cases; however, regarding kinship analysis, all loci in the NGM kit are biologically independent except for the two syntenic loci vWA and D12S391; the assumption of independence between these two loci is not valid. For simplicity in kinship analysis cases, we recommend selecting the more informative of the two loci to be combined with the other loci of the multiplex.

References

1. Schneider PM (2009) Expansion of the European standard set of DNA database loci—the current situation. Profiles in DNA 12(1): 6–7, At: http://www.promega.com/profiles/1201/1201_06.html. Accessed 30 May 2010
2. Council of the European Union (2005) Prum Convention, Brussels, 7 July 2005, http://register.consilium.europa.eu/pdf/en/05/st10/st10900.en05.pdf. Accessed 30 May 2010
3. NCBI. NCBI map reviewer. http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606&query=vWA&qchr=12&strain=All; and http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606&query=D12S391&qchr=12&strain=All. Accessed 30 May 2010
4. Li CC (1976) First course in population genetics First Edition. Boxwood Press, Pacific Grove, CA
5. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590
6. Fisher RA (1951) Standard calculations for evaluating a blood-group system. Heredity 5(1):95–102
7. Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372
8. Karlin S, Cameron EC, Williams PT (1981) Sibling and parent–offspring correlation estimation with variable family size. Proc Natl Acad Sci 78(5):2664–2668
9. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
10. Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci 81 (11):3443–3446
11. Mark Lathrop (1993) LINKAGE user's guide. In: An alphabetic list of genetic analysis software. http://linkage.rockefeller.edu/soft/linkage/. Accessed 30 May 2010
12. Phillips C, Fernandez-Formoso L, Garcia-Magariños M, Porras L, Tvedebrink T, Amigo J, Fondevila M, Gomez-Tato A, Alvarez-Dios J, Freire-Aradas A, Gomez-Carballa A, Mosquera-Miguel A, Carracedo A, Lareu MV (2010) Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. Forensic Sci Int Genet doi: 10.1016/j.fsigen.2010.02.003
13. Budowle B, Shea B, Neizgoda S, Chakraborty R (2001) CODIS short tandem repeat loci data for 41 sample populations. J Forens Sci 46(3):453–489
14. Budowle B, Chakraborty R (2001) Population variation at the CODIS core loci in Europeans. Leg Med 3:29–33
15. National Research Council II Report (1996) The evaluation of forensic evidence. National Academy Press, Washington, DC
16. Weir BS (1990) "Multiple tests" in genetic data analysis. Sinauer Associates Inc., Sunderland, Massachusetts, pp 109–110