Forensic Science International: Genetics 26 (2017) 91–95

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Investigator® HDplex (Qiagen) reference population database for forensic use in Argentina

Gustavo Martínez[a,*], Alicia Borosky[b], Daniel Corach[c], Cintia Llull[d], Laura Locarno[e], Mercedes Lojo[f], Miguel Marino[e], María Cecilia Miozzo[g], Nidia Modesti[h], Carla Pacharoni[h], Juan Pablo Pilili[f], María Isabel Ramella[g], Andrea Sala[c], Cecilia Schaller[a], Carlos Vullo[b,i], Ulises Toscanini[d,*]

[a] Servicio de Genética Forense, Poder Judicial de la Provincia de Entre Ríos, Entre Ríos, Argentina
[b] LIDMO, Córdoba, Argentina
[c] Servicio de Huellas Digitales Genéticas, Facultad de Farmacia y Bioquímica, U.B.A., Buenos Aires, Argentina
[d] PRICAI-Fundación Favaloro, Buenos Aires, Argentina
[e] Laboratorio de Genética Forense, Cuerpo Médico Forense y Criminalístico, Ministerio Público de la Provincia de Mendoza, Mendoza, Argentina
[f] Asesoría Pericial, La Plata, Poder Judicial de la Provincia de Buenos Aires, Buenos Aires, Argentina
[g] Laboratorio Regional de Genética Forense del NOA,Departamento Médico-Poder Judicial de Jujuy, Jujuy, Argentina
[h] Instituto de Genética Forense, Poder Judicial de Córdoba, Córdoba, Argentina
[i] EAAF Forensic Genetic Laboratory, Córdoba, Argentina

ARTICLE INFO

Article history:
Received 8 June 2016
Received in revised form 13 September 2016
Accepted 17 October 2016
Available online 18 October 2016

Keywords:
Investigator® HDplex
STR
Argentina
Forensic genetics
Population genetics
Allele frequencies

ABSTRACT

Currently, autosomal Short Tandem Repeat (STR) markers represent the method of election in forensic human identification. Commercial kits of most common use nowadays –e.g. PowerPlex® Fusion, Promega Corp.; AmpFlSTR GlobalFiler, Thermofisher scientific; Investigator 24Plex QS,Qiagen-, allow the co-amplification of 23 highly polymorphic STR loci providing a high discrimination power in human identity testing. However, in complex kinship analysis and familial database searches involving distant relationships, additional DNA typing is often required in order to achieve well-founded conclusions. The recently developed kit Investigator® HDplex (Qiagen) co-amplify twelve autosomal STRs markers (D7S1517, D3S1744, D12S391, D2S1360, D6S474, D4S2366, D8S1132, D5S2500, D18S51, D21S2055, D10S2325, SE33), nine of which are not present in the above mentioned kits, providing a set of efficient supplementary markers for human identification purposes. In this study we genotyped a sample of 980 individuals from urban areas of ten Argentinean provinces using the Investigator® HDplex kit, aiming to provide forensic estimates for use in forensic casework and parentage testing in Argentina. We report reference allelic frequency databases for each of the provinces studied as well as for the combined samples. No deviation of Hardy-Weinberg equilibrium was observed. A reasonable discrimination capacity and power of exclusion was estimated which allowed predicting an acceptable forensic behavior of this kit, either to be used as the main STR panel for simple cases or as an auxiliary tool in complex cases. Additionally, population comparison tests showed that the studied samples are relatively homogeneous across the country for these STR set.

1. Introduction

Since their insertion in the field of forensic genetics in the early 1990, short tandem repeat (STR) markers rapidly showed a clear potential to be used in human identification, mainly due to the possibility of multiplexing, automation, standardization and a subsequent high power of discrimination. As a consequence of the extensive endorsement of the STR methodology by the international forensic genetic community, the first criminal databases were implemented in the middle 1990s. For instance, the first national DNA database was established in the UK in 1995 [1] – based on a seven-STR core set-, and the Combined DNA Index System (CODIS) was officially launched in USA in 1998 – based on a core set of 13 STRs-, after some pilot projects developed in the previous years [2]. Thus, laboratories tended to adopt a basic set of

* Corresponding authors.
E-mail addresses: gustavo_g_martinez@outlook.com (G. Martínez), utoscanini@pricai.com.ar (U. Toscanini).

http://dx.doi.org/10.1016/j.fsigen.2016.10.009
1872-4973/

markers including the 13 CODIS STRs, usually in combination with a few more additional STRs and the sex marker Amelogenin; or more recently a 23 autosomal STR configuration with the release of the newest commercial formats (PowerPlex® Fusion, Promega Corp.; AmpFlSTR GlobalFiler, Thermofisher scientific; Investigator 24Plex QS, Qiagen). The availability of commercial kits including these markers and the existence of quality control programs (e.g. [3,4]), lead to a high level of standardization through the years.

The massive use of these STR panels by many laboratories served to generate a huge amount of population data necessary for the correct statistical interpretation of the DNA evidence. In Argentina, the forensic performance of these genetic systems was extensively studied across the country and allelic frequency databases of forensic use for the local, regional and national level were produced by different groups (e.g. [5–12]).

Although a combination of 15 to 23 markers proved to be highly efficient in the analysis of simple paternity cases, namely duos and trios, and in forensic casework, the needs for extra discrimination power turned out to be more evident as DNA databases became larger, and as more complex cases began to be assessed by DNA technology. So that, during the last years, new STR markers were developed and validated for forensic use and became commercially available in different grouping formats, some of them in combination with the previously well established set of markers and in other cases, as separate genetic systems, which aided to expand the STR core set of some national DNA databases [13,14].

The Investigator® HDplex kit was released by Qiagen in 2012. It includes twelve autosomal STRs (D7S1517, D3S1744, D12S391, D2S1360, D6S474, D4S2366, D8S1132, D5S2500, D18S51, D21S2055, D10S2325, SE33), three of which (D12S391, D18S51 and SE33) are present in existent commercial kits of common use in forensics. Population studies on this set of markers have already been carried out in different populations [e.g. 15,16] but so far no data are available from Argentinean populations. In this work we genotyped 980 individuals inhabiting urban areas of ten provinces of Argentina for the STRs included in the Investigator® HDplex kit. We aimed to evaluate the forensic performance of this system and to produce reference population databases for its use in the statistical interpretation of the DNA forensic evidence in Argentina. In addition, since it is advisable to built DNA databases with STRs highly discriminative in order to reduce adventitious match, we also aimed to search for new STRs with higher discrimination power in the Argentina population.

2. Materials and methods

2.1. Populations

Urban populations of ten Argentinean provinces encompassing the Northern, Central and Southern regions of the country were studied. In total, 980 non-related individuals were analyzed according to the geographical and numerical distribution shown in Fig. 1. These samples have been previously genotyped for at least one of the commonest STRs kits available, namely AmpFlSTR® Identifiler or AmpFlSTR® Identifiler Plus (Applied Biosystems, USA), PowerPlex® 16, PowerPlex® 16 HS or Powerplex® 21 Systems (Promega Corp, USA). Blood samples and buccal swabs were obtained from the participants after the corresponding written informed consent. Samples were collected and genotyped in eight laboratories of Argentina.

2.2. DNA typing

DNA extraction from blood samples or buccal swabs was done following standard organic (phenol-chloroform) procedures [17] or non-organic procedures, namely Chelex or salting-out extraction methods [18,19].

Amplification of the samples with the Investigator® HDplex kit (Qiagen) was performed according to manufacturer's instructions (Investigator® HDplex handbook) using 1–5 μl of DNA extract containing ~0.5–1 ng as PCR template.

Separation and detection of amplified products were carried out according to manufacturer's instructions using the different ABI platforms available in the different laboratories. These included ABI PRISM 310, ABI PRISM 3100/3100 Avant, ABI PRISM 3130/3130xl, and ABI 3500 Genetic Analyzer. Allele calling was done by comparison with the reference ladder provided with the kit using different versions of the GeneMapper software (Applied Biosystems, USA) and the reference template files available as a download from www.qiagen.com.

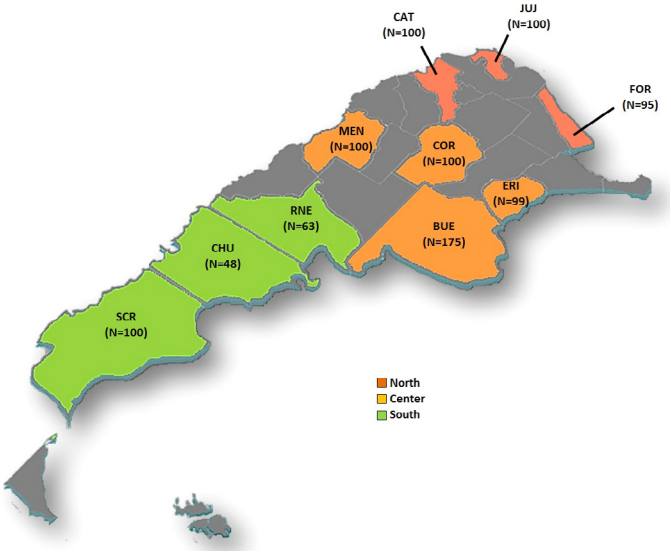

Fig. 1. Geographic location and size of the Argentinean samples analyzed.
JUJ: Jujuy; FOR: Formosa; CAT: Catamarca; MEN: Mendoza; COR: Córdoba; ERI: Entre Ríos; BUE: Buenos Aires; RNE: Río Negro; CHU: Chubut; SCR: Santa Cruz.

2.3. Quality control

The laboratories involved in this study participated in the annual GHEP-ISFG (www.ghep-isfg.org) Intercomparison Exercise and in the annual Quality Control Exercise organized by the Argentinean Society for Forensic Genetics (SAGF, www.sagf.org.ar), having succeeded in all the exercises.

2.4. Statistical analyses

Standard diversity indices, allelic frequency distributions, exact test for Hardy-Weinberg Equilibrium (HWE), and linkage disequilibrium (LD) between every possible pair of loci were computed using PowerMarker software v3.25 [20]. Forensic parameters, namely Matching Probability (MP), Power of Discrimination (PD), Power of Exclusion for duos and trios (PEd and PEt), and Typical Paternity Index (TPI) were estimated using an *ad hoc* MsExcel file. In order to evaluate population substructure, exact tests for population differentiation and genetic distances (*Fst*) between pairs of population were estimated as implemented in Arlequin software v3.5 [21]. Genetic distances were estimated using (a) only the Investigator® HDplex data and (b) using the Investigator® HDplex plus the 13 CODIS STRs genotypic data. Neighbor-Joining (NJ) trees and multidimensional scaling plots (MD) were built based on distance matrices using XLStat [22]. A significance level of 0.05 was considered for all tests, applying Bonferroni's correction for multiple comparisons.

3. Results

All the statistics estimated are provided as Supplementary material. Genetic diversity indices and parameters of forensic interest are summarized in Table S1 for the global sample of Argentina. Full concordance was observed for the genotypes of the loci shared between the Investigator® HDplex and other commercial kits that were previously typed in this dataset, namely D18S51, D12S391 and SE33. The STRs with higher PD were SE33, D7S1517, D21S2055, D10S2325 and D12S391. *P*-values of exact test for HWE are indicated in Table S2. Either for the individual samples or for the grouped data for Argentina, no deviation of HWE was detected, except at locus D3S1744 in Buenos Aires and locus D2S1360 in the global sample, for which the *p*-value of the exact test were slightly below the significance level after applying Bonferroni's correction. Allelic frequency distributions are shown in Table S3.1. (grouped sample for Argentina) and Tables S3.2.–S3.13. (local samples individually analyzed). Minimal allelic frequency calculated as 5 divided by 2 times the number of individuals in each population is included. Three-allele patterns were observed in three samples, two of them in locus D10S2325 (one in Entre Ríos –"12/14/15"– and one in Formosa –"9/14/15"–) and another one in locus D21S2055 in Formosa –"32/33/34"–. These samples were not considered for the allelic frequency estimations nor for the rest of the statistical analyses performed.

Some of the markers in the Investigator® HDplex kit are syntenic with markers from other commercial kits (i.e. they are

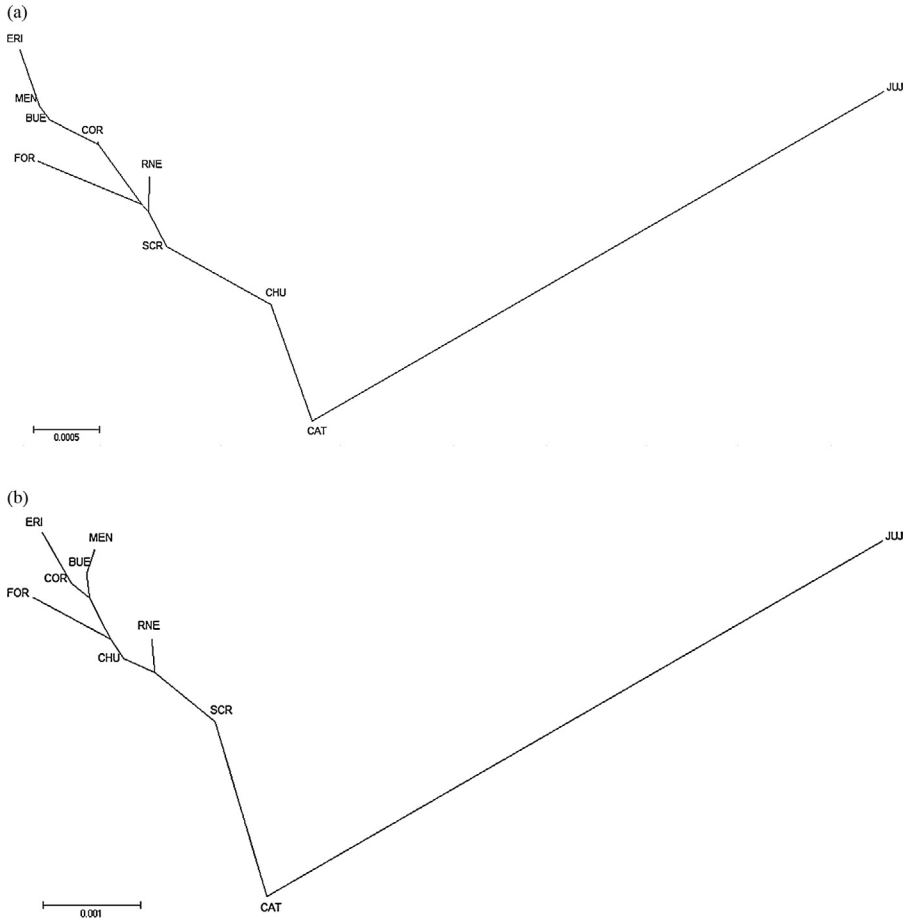

Fig. 2. *NJ tree* based on F_{ST} distances considering (a) only the Investigator® HDplex markers and (b) Investigator® HDplex and CODIS markers.
JUJ: Jujuy; FOR: Formosa; CAT: Catamarca; MEN: Mendoza; COR: Córdoba; ERI: Entre Ríos; BUE: Buenos Aires; RNE: Río Negro; CHU: Chubut; SCR: Santa Cruz.

located in the same chromosome). Genetic distances between pairs of common STR loci including the ones present in the Investigator® HDplex kit can be found in Phillips et al. [23]. LD was tested for every possible pair of 33 loci indicated in Table S5. Results showed no significant *p*-values (significance level = 0.000095 after Bonferroni's correction for 528 comparisons) for any combination of loci. (Table S6).

Regarding population differentiation, exact test yielded non-significant *p*-values for all the pair of populations ($p > 0.1000$), both considering only Investigator® HDplex markers or Investigator® HDplex and CODIS markers (Tables S4.1. and S4.2.). However, significant differences for *Fst* were observed between Jujuy and seven out of the nine remaining populations ($p < 0.0001$) when considering only the Investigator® HDplex kit, and with all of them when considering Investigator® HDplex plus the CODIS loci (Tables S4.3. and S4.4). This is exhibited in the NJ trees in Fig. 2a and b built from the *Fst* matrices in Tables S.4.3. and S4.4 and in the MDS plot in Fig. 3.

4. Discussion

In this study we analyzed a wide-ranging sample of urban origin from ten provinces of Argentina encompassing the North, Centre and South regions of the country, for the twelve markers included in the Investigator® HDplex kit (Qiagen). We aimed to provide reference data for use in the interpretation of parentage testing and forensic casework in Argentina. Allelic frequency tables for the individual samples analyzed and combining the individual datasets in a global sample were produced and parameters of forensic interest were reported. The populations studied appeared to be in HWE, in spite of observing a couple of significant *p*-values in the exact tests, which is not unusual specially when the databases tend to be large [24].

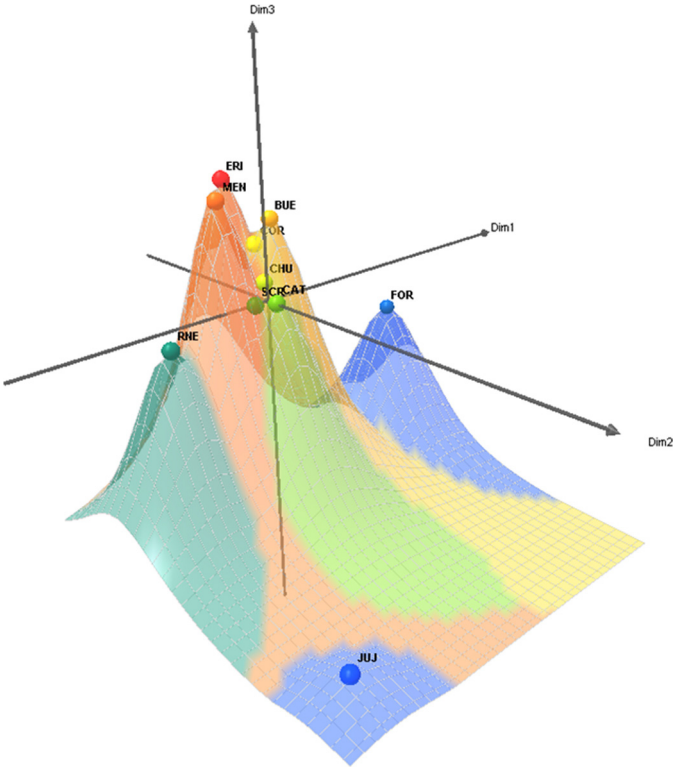

Fig. 3. *MDS* plot based on F_{ST} distances considering the Investigator® HDplex and CODIS markers.
JUJ: Jujuy; FOR: Formosa; CAT: Catamarca; MEN: Mendoza; COR: Córdoba; ERI: Entre Ríos; BUE: Buenos Aires; RNE: Río Negro; CHU: Chubut; SCR: Santa Cruz.

No marked population substructure was evident through the analyses of these STRs across Argentina. Nevertheless, it could be seen that Jujuy presents *Fst* values with respect to the other samples that are higher than those between the remaining pairs of populations. Thus, Jujuy is clearly separated from the other samples in the NJ trees (Fig. 2a, b) and in the MDS plot of Fig. 3. This is not surprising since it is known that the Native American contribution to the gene pool of the extant population of Jujuy, as well as to other northwestern population of Argentina, is higher than in other regions of the country. This was also evidenced through the analyses of Y chromosome STR markers [25,26]. Interestingly, Jujuy is also apart from Catamarca, a nearby province with a similar Native American input. One could also speculate on other ancestral populations contributing to the gene pool of Jujuy (e.g. of African origin), although other studies outside the goal of this work are needed to deeply investigate this hypothesis. Both NJ plots (only Investigator® HDplex STRs and Investigator® HDplex plus CODIS STRs) show similar patterns, which fairly reflect the geographical distribution of the studied samples across the Argentinean territory. This is also mimicked in the MDS graph in Fig. 3.

LD between syntenic loci included in the Investigator® HDplex and those present in other commercial kits was evaluated in this study. No allelic associations were detected between the loci included in the Investigator® HDplex kit with the other loci that co-localize in the same chromosomes available for comparison in this study, namely D7S1517/D7S820, D3S1744/D3S1358, D12S391/vWA, D2S1360/D2S441, D2S1360/D2S1338, D6S474/D6S1043, D4S2366/FGA, D8S1132/D8S1179, D5S2500/CSF1PO, D5S2500/D5S818, D21S2055/D21S11, D10S2325/D10S1248, SE33/D6S1043. However, although this lack of association would make these markers suitable to be used in conjunction with the main STR sets without a strong impact in the estimation of matching probability for unrelated persons on criminal casework [27], it has been shown through simulation experiments that linkage could lead to a considerable overestimation of the LR in specific kinship cases [15,27], and then a correction should be considered as recommended [27].

The forensic parameters analyzed, namely PE (for duos and trios), TPI, and MP, allowed predicting an acceptable performance of this system for routine parentage testing and forensic casework. For instance, PE, MP and TPI values estimated for the twelve markers in the global Argentinean population appeared to be in the same level of magnitude that previously reported values in Argentina for other commercial kits, even with larger sets of markers (e.g. [28–31]). In this way, we calculated the PD, MP and PE (trios and duos) for the 13 CODIS STRs using the same dataset that was analyzed for the Investigator® HDplex and we observed that a higher PD (lower MP) and PE is expected for the Investigator® HDplex system in comparison to the CODIS STR set (Tables S7.1 and S7.2). However, in complex cases, more markers might be needed to get to acceptable results according to the policies of each laboratory. Since at least nine out of the twelve markers included in the Investigator® HDplex kit are not present in other commercial formats, the possibility of adding the analysis of these markers to the results obtained with routinely used STRs can offer the extra power that is needed sometimes, especially in deficient paternity cases. Furthermore, these features place this STR set as a worthy candidate to be considered in the implementation of DNA databases in Argentina, where there are several ongoing database projects, and the forensic parameters produced through this collaborative effort will significantly contribute to the interpretation of the genetic evidence. However, further validations studies that are out of the scope of the present work will be needed in

order to evaluate the performance of the Investigator® HDplex kit in the analyses of forensic samples, specially in those where the DNA quality and quantity might be compromised. The results of such studies should be considered critical for the inclusion of these markers as part of the database loci set.

Finally, although no major differences were observed among the population studies, some of them appeared to be statistically significant, highlighting once again the importance of using local databases for the statistical evaluation of the DNA evidence whenever it is possible.

Conflict of interests

The authors declare no conflict of interests.

Acknowledgement

We thank Qiagen Corporation for providing us with the Investigator® HDplex kits necessary for genotyping all the samples studied in this work.

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.2016.10.009.

References

[1] D.J. Werret, The national DNA database, Forensic Sci. Int. 88 (1997) 33–42.
[2] R. Hoyle, The FBI's national DNA database, Nat. Biotechnol. 16 (1998) 987.
[3] The Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) Proficiency test. www.ghep-isfg.org.
[4] German DNA Profiling (GEDNAP). www.gednap.org.
[5] A. Borosky, U. Toscanini, A. Gómez, M.L. Parolín, N. Basso, C. Vullo, Forensic population data for 20 STR loci in Argentina, Forensic Sci. Int. Genet. 13 (2014) e20–e21.
[6] A. Borosky, L. Catelli, C. Vullo, Analysis of 17 STR loci in different provinces of Argentina, Forensic Sci. Int. 3 (2009) e93–e95.
[7] W.R. Bozzo, M.A. Pena, M.I. Ortiz, M.M. Lojo, Genetic data from Powerplex® 16 system and identifiler™ kits from buenos aires province (Argentina), Legal Med. 9 (2007) 151–153.
[8] D. Corach, O. Lao, C. Bobillo, K. van Der Gaag, S. Zuniga, M. Vermeulen, K. van Duijn, M. Goedbloed, P.M. Vallone, W. Parson, P. de Knijff, M. Kayser, Inferring continental ancestry of argentineans from autosomal, Y-chromosomal and mitochondrial DNA, Ann. Hum. Genet. 74 (2010) 65–76.
[9] M. Marino, A. Sala, D. Corach, Population genetic analysis of 15 autosomal STRs loci in the central region of Argentina, Forencis Sci. Int. 161 (2006) 72–77.
[10] M. Marino, A. Sala, D. Corach, Genetic attributes of 15 autosomal STRs in the population of two patagonian provinces of Argentina, Forensic Sci. Int. 160 (2006) 84–88.
[11] G. Martı'nez, E. Vázquez, C. Schaller, N. Quevedo, Genetic data on 11 STRs (CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, D16S539, D7S820, D13S317, F13B, LPL) in an Argentine northeast population, Forensic Sci. Int. 133 (2003) 254–255.
[12] M.C. Miozzo, C.M. Pacharoni, S.A. Mutal, M.K. Maxzud, A.E. Casanova, N.M. Modesti, STRs data for the loci D3S1385, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 from Cordoba (Argentina), Forensic Sci. Int. 131 (2003) 214–217.
[13] D.R. Hares, Selection and implementation of expanded CODIS core loci in the United States, Forensic Sci. Int. Genet. 17 (2015) 33–34.
[14] P. Gill, L. Fereday, N. Morling, P.M. Schneider, The evolution of DNA databases-Recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.
[15] A. Tillmar, H. Nilsson, D. Kling, K. Montelius, Analysis of investigator HDplex markers in Swedish and Somali populations, Forensic Sci. Int. Genet. 7 (2013) e21–e22.
[16] C. Phillips, L. Fernandez-Formoso, M. Gelabert-Besada, M. García-Magariños, J. Amigo, A. Carracedo, M.V. Lareu, Global population variability in qiagen Investigator® HDplex STRs, Forensic Sci. Int. 8 (2014) 36–43.
[17] C.T. Comey, B.W. Koons, K.W. Presley, J.B. Smerick, C.A. Sobieralski, D.M. Stanley, F.S. Baechtel, DNA extraction strategies for amplified fragment length polymorphism analysis, J. Foresic Sci. 39 (1994) 1254–1269.
[18] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques 10 (1991) 506–513.
[19] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, Nucleic Acids Res. 16 (1988) 1215.
[20] K. Liu, S.V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, Bioinformatics 21 (2005) 2128–2129.
[21] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.
[22] Addinsoft (2007). XLStat —Statistical software for MS Excel. URL: www.xlstat.com.
[23] C. Phillips, D. Ballard, P. Gill, D.S. Court, A. Carracedo, M.V. Lareu, The recombination landscape around forensic STRs: accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data, Forensic Sci. Int. Genet. 6 (2012) 354–365.
[24] National Research Council, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington DC, 1996.
[25] M. Marino, A. Sala, D. Corach, Genetic attributes of the YHRD minimal haplotype in 10 provinces of Argentina, Forensic Sci. Int. Genet. 1 (2007) 129–133.
[26] U. Toscanini, C. Vullo, G. Berardi, C. Llull, A. Borosky, A. Gómez, J. Pardo-Seco, A. Salas, A comprehensive Y-STR portrait of Argentinean populations, Forensic Sci. Int. Genet. 20 (2016) 1–5.
[27] P. Gill, C. Phillips, C. McGovern, J.A. Bright, J. Buckleton, An evaluation of potential allelic association between the STRs vWA and D12S391: Implications in criminal casework and applications to short pedigrees, Forensic Sci. Int. Genet. 6 (2012) 477–486.
[28] U. Toscanini, L. Gusmão, G. Berardi, A. Amorim, A. Carracedo, A. Salas, E. Raimondi, Testing for genetic structure in different urban Argentinian populations, Forensic Sci. Int. 165 (2007) 35–40.
[29] C. Vullo, A. Borosky, C. Romanini, L. Catelli, T. Yamamoto, Frequency data for 12 mini STR loci in Argentina, Forensic Sci. Int. Genet. 4 (2010) e79–e81.
[30] M. Marino, A. Sala, C. Bobillo, D. Corach, Inferring genetic sub-structure in the population of Argentina using fifteen microsatellite loci, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 350–352.
[31] M.C. Miozzo, M.K. Maxzud, A.E. Casanova, S.A. Mutal, C. Pacharoni, N.M. Modesti, Allele frequencies and statistical parameters for Penta E and Penta D STR loci in Córdoba (Argentina) population, J. Forensic Sci. 49 (2004) 407–408.