Concordance study between the ParaDNA® Intelligence Test, a Rapid DNA profiling assay, and a conventional STR typing kit (AmpFlSTR® SGM Plus®)

G. Ball, N. Dawnay, B. Stafford-Allen, M. Panasiuk, P. Rendell, S. Blackman, N. Duxbury, S. Wells

LGC, ParaDNA Business Unit, F4 Culham Science Centre, Abingdon, Oxfordshire, OX14 3ED, United Kingdom

Abstract

The ParaDNA® Intelligence Test enables STR profiling directly from human biological samples and evidence items collected from crime scene in 75 min. Designed for non-expert use this system allows DNA information to be available to investigators before it would typically be available from a laboratory. The ParaDNA Intelligence Test system amplifies D3S1358, D8S119, D16S539, D18S1358 and TH01 STR loci and the gender typing locus amelogenin and detects the alleles present with HyBeacon1 probes. Individual DNA samples from 381 UK Caucasian individuals were analysed using AmpFlSTR1 SGM Plus® and the ParaDNA Intelligence Test with the derived STR profiles compared. Here we describe the high level of concordance demonstrated between the two systems and discuss this with reference to allele frequencies and the discriminatory power offered by the ParaDNA Intelligence Test.

Keywords: Short tandem repeat, STR, DNA typing, Rapid DNA, ParaDNA, Direct PCR

Introduction

Short tandem repeat (STR) profiling remains the standard method for forensic identification of individuals and many commercial kits are now available [1]. The ParaDNA Intelligence Test is a novel STR profiling kit developed specifically for rapid DNA profiling by DNA melt analysis [2]. It is designed to enable rapid DNA profiling in the hands of non-expert users for intelligence driven investigation using the same platform as the recently launched ParaDNA Screening System [3].

The ParaDNA Intelligence Test system amplifies D3S1358, D8S119, D16S539, D18S1358 and TH01 STR loci and the gender typing locus amelogenin across 4 independent PCR reaction tubes. Template material from human biological samples and evidence items is transferred into the four reaction

tubes using the ParaDNA Sample Collector. DNA is released by direct lysis of cellular material and the five STR loci amplified in a series of PCR reactions performed on the ParaDNA System. The alleles present are detected by DNA melting analysis with HyBeacon probes which bind the STR amplicons. The temperature is then raised and the probes de-nature from their STR targets at a melting temperature (TM) specific for the number of repeats, signalled by a drop in fluorescence [2]. STR profiles are called automatically and displayed by the ParaDNA Intelligence System software, which also provides profile searching and comparison functions.

Discounting the occurrence of null alleles, the allele range at each locus detected and reported by the ParaDNA Intelligence Test is anticipated to cover 94.13% expected alleles at these loci, based on allele frequencies from sample populations [4–6]. A further 5.76% more rare alleles with repeat numbers above these ranges are also detected but cannot be resolved. Therefore the highest allele reported in each range is described as a number of repeats (n) plus any larger alleles which might be present (hence n+).

Concordance studies between STR typing kits are conducted during the developmental validation to confirm the validity of data produced by each with reference to others [7,8], in particular to detect the presence of mutations such as single nucleotide polymorphisms(SNPs)or deletions that might affect primer binding, thereby causing null alleles [9]. Concordance studies are typically conducted on a number (typically several hundred or more) of samples from individuals in a reference population or population subgroup using one profiling kit for which profiling data has been obtained using another product, thus establishing concordance between the profiles generated by the two products [4,10,11]. In addition, population studies can provide data on allele frequencies anticipated in the reference population or populations [5,6,12].

This report describes the high level of concordance demonstrated when DNA samples from 381 UK Caucasian individuals were profiled using the ParaDNA Intelligence Test compared with AmpFlSTR SGM Plus.

Materials and methods

Anonymous DNA samples were obtained from 381 UK Caucasian individuals in the European Collection of Cell Cultures (ECACC) Human Random Control (HRC) panel supplied by Public Health England (PHE: Cat# 60 41302-5). Further samples representing some previously characterised examples of sequence microvariants were analysed from the ECACC Ethnic Diversity Panel (Cat# 7020701p). The ECACC panel is comprised of purified DNA from EBV transformed lymphoblastoid cell lines derived from unrelated, random, UK Caucasian blood donors. STR profiles were obtained using AmpFlSTR SGM Plus PCR Amplification Kit (Life Technologies: Cat# 4307133) by PHE following

manufacturers conditions with confirmation of certain samples by LGC. A single ambiguous sample was profiled for Y chromosome STR (PowerPlex1 Y23 System, Promega: Cat# DC2305) and autosomal and Y contributions quantified (Plexor1 HY System, Promega: Cat# DC1001) to confirm gender identification.

This population was chosen as a relevant, well characterised population group upon which to conduct this concordance study [5,6]. Similar studies have sampled similar numbers of individuals from single [13,14] or multiple ethnic groups [4,8,9] represented within their geographical location.

Purified DNA was added directly to AmpFlSTR SGM Plus PCR and profiles analysed as per the manufacturers recommended and PHE or LGC standard protocols. Sample DNA (4 ng total) in TE buffer was applied to the ParaDNA Sample Collector and dried in air for 1 h before addition to the ParaDNA Intelligence Test reagents in the 4 well tube reaction (1 ng per reaction tube). ParaDNA Sample Collectors were then sealed into ParaDNA Reaction Plates and loaded onto the ParaDNA System. Amplification, data collection, allele calling and profile generation were performed on the unit under the control of ParaDNA Intelligence Software (Version 1.1). Further analyses were performed when required using custom development software packages for the visualisation and analysis of melt curves. Samples that did not yield a full ParaDNA profile of 12 alleles or that gave a profile that was discordant with that obtained from the AmpFlSTR SGM Plus profile were repeated to confirm the result with the consensus profile used to compare concordance.
Loci were tested for departures from Hardy Weinberg expectations (HWE) using Genepop software [15,16]. Polymorphism Information Content (PIC) was calculated in Excel following [17]. Probability of Identity (PI) was calculated using ApiCalc [18]. Power of Discrimination (PD) and Power of Exclusion (PE) were calculated following [19].

Results

Of the 381 UK Caucasian samples tested more than 99.8% of consensus allele calls were concordant with the AmpFlSTR SGM Plus profile, as displayed in Table 1 below. Two loci, D16S539 and amelogenin had allele calls which were completely concordant while others had a low number (<1%) of discordant allele calls. In total 98.4% of samples gave a consensus five STR profile concordant with that produced by AmpFlSTR SGM Plus.

Of the concordant allele calls 6%, fell within the n+ designation used by the ParaDNA system to identify alleles above the measurable range. The majority of calls in this designation were at the TH01 locus (9.3+) for which the frequency of alleles in this population with 9.3 or more repeats was approximately 34%.

Repeated typing of 41 (11%) samples was performed to confirm instances of allelic dropout. Of these a discordant result was confirmed in 6 samples as detailed in Table 2. The remaining samples returned concordant profiles on repeated typing. In total 0.15% of allele calls produced by the ParaDNA Intelligence Test were discordant with AmpFlSTR SGM Plus. The observed discordance and underlying causes of the variations are listed in Table 2.

One reason for the discordance between the ParaDNA Intelligence Test and AmpFlSTR SGM Plus is the occurrence of microvariants. Only a single sample with a microvariant was observed in the 381 UK Caucasian samples tested for this study, a TH01 8.3 allele, which indicated a 9 allele (Table 2). However further examples of microvariant samples were analysed from the ECACC Ethnic Diversity Panel (data not shown), with a similar result for another sample with an 8.3 allele at TH01. Another sample with 14.2,18 alleles at D18S51 (data not shown) produced a "not called" result for the variant allele (18,), or indicated a potential mixed profile depending on the signal strength. Microvariant sequences were confirmed by DNA sequencing (data not shown).

A further cause for discordance observed in these data was the occurrence of one sample with a D3S1358 profile of 11, 17. As the 11 allele is below the range detected by the assay the 11 acts as a null allele and the profile displayed was 17,-- , as the signal strength for the 17 allele was below the threshold level to call a homozygote.

Four further samples gave ParaDNA Intelligence Test profiles that were discordant with those generated by AmpFlSTR SGM Plus. Three were due to samples being called with dropout by the ParaDNA Intelligence Test software due to large heterozygote imbalances (data not shown). The fourth sample was called as a mixture due to the presence of a large stutter peak (data not shown).

A single sample indicated an AMEL profile X, -- with the ParaDNA Intelligence Test when an X, Y profile had been expected according to the profile supplied by PHE). Re-profiling at LGC with AmpFlSTR SGM Plus indicated a profile of X, X. Further typing with a Y chromosome STR profiling system (PowerPlex Y23) and Y chromosome quantification (Plexor HY) confirmed only trace male DNA contribution (data not shown). This suggests that this sample was a male exhibiting a high degree of allelic dropout of the Y chromosome (possibly due to chromosomal abnormalities arising during cell culture) or a female sample with a low level of male DNA contamination (Table 2).

Allele frequency data generated by the ParaDNA Intelligence Test (Table 3) indicated a similar distribution of alleles to those measured using the AmpFlSTR SGM Plus [5,6]. The multi-locus probability of identity (system P_I), calculated by multiplying the individual P_I values of each locus, is often used as a measure of the discriminatory ability of a profiling system [18]. The PI for the

ParaDNA Intelligence Test system was calculated at 3.71 $x10^{-7}$ suggesting the average probability of observing a random matching profile (PM) from this population was about 1 in 2.7 million. The P_I for the same samples profiled at these loci using AmpFlSTR SGM Plus was 3.39 $x\ 10^{-7}$ or 1 in 2.95 million, with the increase in discrimination due to the ability to identify more, rarer alleles. With the full range of eleven loci profiled using AmpFlSTR SGM Plus the P_I was calculated as 1.11×10^{-13}. Deviations from HWE were observed at the D18S51 and D3S1358 loci which were not observed when the corresponding AmpFlSTR SGM Plus data was analysed. This is largely due to the reduced confidence when calling homozygotes leading to a reduction in the observed number of homozygote calls compared to the expected frequency calculated from HWE.

Discussion

More than 99.8% of consensus allele calls produced by the ParaDNA Intelligence test were concordant with the allele calls produced by AmpFlSTR SGM Plus, as displayed in Table 1. This high level of concordance in allele calls is similar to the levels of concordance observed when other STR typing assays are compared, for example the published concordance of 99.7% between AmpFlSTR Identifiler and AmpFlSTR Minifiler [10], 99.77% between AmpFlSTR Identifiler and a custom Mini STR assay [11] and 99.97% between PowerPlex ESI 16 and AmpFlSTR SGM Plus [13]. The ParaDNA Intelligence Assay can therefore be considered to have high concordance of allele calling with a UK standard STR typing kit.

Of the concordant allele calls, 6% fell within the n+ designation. These allele calls are considered concordant with AmpFlSTR SGM Plus allele calls as the n+ designation encompasses any alleles at that locus with a higher number of repeats when analysed and displayed within the ParaDNA Intelligence Software. Work to further resolve alleles within this designation is planned.

The data also included a number of samples (85 or 22.3% of samples; 2.3% of alleles) which displayed “uncalled” (--) at one or more loci as they contained homozygous alleles which were below the fluorescence threshold at which they could be confidently called as homozygotes. These were not treated as discordant as they did not display an incorrect allele call, rather not displaying the second (homozygous) allele which had been called by AmpFlSTR SGM Plus. These thresholds were set conservatively ie. to avoid incorrectly calling a homozygote in the case of a sample which might show dropout of a heterozygous allele. This hence led to some instances of uncalled alleles ( ) in the allele frequency data in Table 3. As the homozygote cannot be confirmed in these instances some data is lost, contributing to the increased likelihood of observing a random match (PM) compared to

AmpFlSTR SGM Plus observed. The ParaDNA Intelligence Software searching and comparison functions use a series of rules to widen the search criteria (wildcards) in this instance [20].

Seven allele calls (0.15%) were discordant with the AmpFlSTR SGM Plus profile as indicated in Table 2. One reason for discordance between the ParaDNA Intelligence Test and AmpFlSTR SGM Plus is the occurrence of microvariants. The most common microvariant, the 9.3 variant in TH01 [4–6], is correctly called as 9.3+ by the ParaDNA Intelligence Test. However due to the lower resolution of melting curve analysis compared with capillary electrophoresis (CE), other microvariants may not be distinguished from full repeat variants by the Para DNA Intelligence Test. This is the case in the example in Table 2 of a sample with a 8.3 allele in TH01 which was called as a 9 allele.

Further examples of microvariant samples the ECACC Ethnic Diversity Panel were analysed (data not shown), with similar results for another sample with a 8.3 allele at TH01. Such microvariants are likely to result in a discordant allele call upon direct comparison with the AmpFlSTR SGM Plus allele call. In contrast, a sample with 14.2, 18 alleles at D18S51 (data not shown) produced a “not called” result for the variant allele (18, --) or indicated a potential mixed profile depending on the signal strength. This is due to the 14.2 variation falling between the 14 and 15 allele bins whereas a 14.3 allele would lie closer to the larger allele bin. Thus in some cases a microvariant will cause a “not called” result.

The occurrence of microvariants for these five loci (other than TH01 9.3) is generally low in the population. In a review of studies containing allele frequencies for several of the major ethnic population groups in the UK and US the total frequency of microvariants in all five loci used in the ParaDNA Intelligence Test was less than 0.1% [4–6]. In addition the ParaDNA Intelligence Software searching/comparison functions use a series of rules to widen the search criteria, thereby mitigating this potential issue [20].

One other cause for discordance observed in this data was the occurrence of one sample with a D3S1358 profile of 11, 17. As the D3S1358 11 allele is below the range detected by the assay it is not observed (a null allele) and the profile displayed was 17--, as the signal strength for the 17 allele was below the threshold level to call a homozygote. The occurrence of alleles below the range of the assay is generally low in the population. The expected frequency of null alleles due to this cause in all five loci used in the ParaDNA Intelligence Test combined was less than 0.05% [4–6]. Work to extend the working ranges of the ParaDNA Intelligence Test assays and detect alleles below the current test ranges is planned.

Four further samples gave ParaDNA Intelligence Test profiles that were potentially discordant with those generated by AmpFlSTR SGM Plus. These were due to samples being called with dropout or as mixtures by the ParaDNA Intelligence Test software due to large heterozygote imbalances or stutter peaks (data not shown). It is likely that in a small number of cases this effect is difficult to avoid, again due to the resolution of the melting analysis. However allele calling in the ParaDNA Intelligence software is set conservatively to avoid incorrect calls being made with the result that some information is lost rather than incorrect calls being made.

A similar distribution of allele frequencies was measured using the ParaDNA Intelligence Test as AmpFlSTR SGM Plus. These were also very similar to previously published allele frequencies data from a UK Caucasian population [5,6]. Similar distributions of allele frequencies at these loci have also been reported from UK Afro-Caribbean and Asian populations [5,6] and US population groups [4].

The P_I was estimated for the population based on the allele frequencies for the ParaDNA Intelligence Test [15]. The likelihood of an adventitious match was estimated at a mean of approximately 1 in 2.7 million. This is considered to provide a useful discriminatory capability for intelligence led investigation.

Acknowledgements

This work was funded by LGC. The authors would like to thank all who assisted with this publication, in particular Jim Thomson for reviewing the manuscript before submission and the following staff members for their contributions to the development of the ParaDNA System: David French, David Moore, Nick Tribble, Romana Ahmed, Sarah Naif, Daniel Leonard, Daren Clark, Aaron Batterby, Martin Pascoe, Thane Gill, Doug Sharp, Shaun Dowson, Mario Andreou, Peter Johnson, Peter Turton, Alaa Saleh, Rachel Scott, Paul Debenham and Mark Dearden.

References

[1] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, Academic Press, Elsevier, Waltham, 2012.

[2] D.J. French, R.L. Howard, N. Gale, T. Brown, D.G. McDowell, P.G. Debenham, Interrogation of short tandem repeats using fluorescent probes and melting curve analysis: a step towards rapid DNA identity screening, Forensic Sci. Int. Genet. 2 (2008) 333–339.

[3] N. Dawnay, B. Stafford-Allen, D. Moore, S. Blackman, P. Rendell, E.K. Hanson, J. Ballantyne, B. Kallifatidis, J. Mendel, D.K. Mills, R. Naggy, S. Wells, Developmental validation of the ParaDNA1

Screening System – a presumptive test for the detection of DNA on forensic evidence items, Forensic Sci. Int. Genet. 11 (2014) 73–79.
[4] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations, J. Forensic Sci. 48 (2003) 908–911.
[5] I.W. Evett, P.D. Gill, J.A. Lambert, N. Oldroyd, R. Frazier, S. Watson, S. Panchal, A. Connolly, C. Kimpton, Statistical analysis of data for three British ethnic groups from a new STR multiplex, Int. J. Legal Med. 110 (1997) 5–9.
[6] L.A. Foreman, I.W. Evett, Statistical analysis to support forensic interpretation for a new ten-locus STR profiling system, Int. J. Legal Med.114 (2001) 147–155.
[7] J.J. Mulero, C.W. Chang, R.E. Lagacé, D.Y. Wang, J.L. Bas, T.P. McMahon, L.K. Hennessy, Development and validation of the AmpFlSTR MiniFiler PCR amplification kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA, J. Forensic Sci. 53 (2008) 838–852.
[8] R.L. Green, R.E. Lagacé, N.J. Oldroyd, L.K. Hennessy, J.J. Mulero, Developmental validation of the AmpFlSTR1 NGM SElectTM PCR amplification kit: a next generation STR multiplex with the SE33 locus, Forensic Sci. Int. Genet. 7 (2013) 41–51.
[9] C.R. Hill, M.C., Kline, D.L., Duewer, J.M. Butler, Strategies for Concordance Testing, [Internet] 2010.[cited:2013,Nov19th].Available from: http://www.promega.co.uk/resources/profiles-in-dna/2010/strategies-for-concordance-testing/.
[10] C.R. Hill, M.C. Kline, J.J. Mulero, R.E. Lagacé, C.W. Chang, L.K. Hennessy, J.M. Butler, Concordance study between the AmpFlSTR1 MiniFiler™ PCR amplification kit and conventional STR typing kits, J. Forensic Sci. 52 (2007) 870–873.
[11] J. Drábek, D.T. Chung, J.M. Butler, B.R. McCord, Concordance study between Miniplex assays and a commercial STR typing kit, J. Forensic Sci. 49 (2004) 859–860.
[12] AmpFlSTR SGM Plus PCR. Amplification Kit User Guide, Publication Number 4309589, Rev. H, Revision Date August 2012.
[13] V.C. Tucker, A.J. Hopwood, C.J. Sprecher, R.S. McLaren, D.R. Rabbach, M.G. Ensenberger, J.M. Thompson, D.R. Storts, Developmental validation of the PowerPlex1 ESX 16 and PowerPlex1 ESX 17 Systems, Forensic Sci. Int. Genet. 6 (2012) 124–131.
[14] D.Y. Wang, C.W. Chang, R.E. Lagacé, N.J. Oldroyd, L.K. Hennessy, Development and validation of the AmpF'STR1 Identifiler1 Direct PCR amplification kit: a multiplex assay for the direct amplification of single-source samples, J. Forensic Sci. 56 (2011) 835–845.
[15] M. Raymond, F. Rousset, GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism, J. Heredity 86 (1995) 248–249.

[16] F. Rousset, Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux, Mol. Ecol. Resour. 8 (2008) 103–106.
[17] J.A. Anderson, G.A. Churchill, J.E. Autrique, S.D. Tanksley, M.E. Sorrells, Optimizing parental selection for genetic linkage maps, Genome 36 (1993) 181 –186.
[18] K.L. Ayres, A.D.J. Overall, API-CALC 1.0: a computer program for calculating the average probability of identity allowing for substructure, inbreeding and the presence of close relatives, Mol. Ecol. Notes 4 (2004) 315–318.
[19] J.M. Butler, Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, Academic Press, Elsevier, Waltham, 2012, pp. 490.
[20] ParaDNA Application Note. Speculatively searching an external database with exported ParaDNA profiles. [Internet] 2014. [cited: 2014, Aug 14th]. Available from: http://paradna.lgcforensics.com/faqa/.

Table 1: Concordance of ParaDNA Intelligence Test allele calls with AmpFlSTR SGM Plus allele calls across the 5 shared loci.

	D16S539	D18S51	TH01	D8S1179	D3S1358	Amelogenin	Total
Concordant	100.00% (762)	99.87% (761)	99.87% (761)	99.48% (758)	99.87% (761)	100.00% (762)	99.85% (4565)
n+[a]	0.13% (1)	0.52% (4)	33.73% (257)	0.52% (4)	0.92% (7)	(–)	5.97% (273)
Discordant	0.00% (0)	0.13% (1)	0.13% (1)	0.52% (4)	0.13% (1)	0.00% (0)	0.15% (7)

[a] Of the concordant allele calls, around 6%, fell within the n+ designation used by the ParaDNA system to identify alleles above the measurable range. As these alleles were not misidentified they are still deemed concordant.

Table 2: Summary of potential discordances observed between ParaDNA Intelligence Test and AmpFlSTR SGM Plus allele calls. Two instances of discordance (1 and 2) were caused by microvariants and an allele below the detection range. Four potential discordances were caused by heterozygote imbalances or high stutter in particular samples (repeated measure confirmed result).

	Locus	SGM+	ParaDNA Intelligence Test	Cause of variation
1	TH01	6, 8.3	6, 9	Microvariant is called as larger allele
2	D3S1358	11, 17	–, 17	11 allele is below measured range
3	D18S51	19, 24	19, – (dropout)	Heterozygote imbalance
4	D8S1179	14, 15	15, – (dropout)	Heterozygote imbalance
5	D8S1179	14, 15	15, – (dropout)	Heterozygote imbalance
6	D8S1179	11, 14	–, – (mix)	High stutter
7	AMEL[a]	X,X (on re-test)	X, X	Incorrect sample label or null allele

[a] single sample (nominally from a male donor) indicated a X, X profile using the ParaDNA Intelligence Test. Re-testing at LGC with AmpFlSTR SGM Plus confirmed a profile of X, X.

Table 3: Allele frequencies for the five STR loci in the ParaDNA Intelligence Test measured from 381 UK Caucasian individuals.

Allele	THO1	D8S1179	D18S51	D16S539	D3S1358
5	0.001				
6	0.193				
7	0.192				
8	0.109	0.009		0.018	
9	0.134	0.012		0.117	
9.3(+)	0.346				
10		0.100	0.005	0.076	
11		0.080	0.014	0.302	
12		0.134	0.168	0.265	
13		0.339	0.140	0.180	0.005
14		0.169	0.129	0.033	0.131
15(+)		0.092	0.122	0.001	0.251
16		0.039	0.122		0.234
17(+)		0.005	0.135		0.186
18			0.073		0.138
19(+)			0.055		0.009
20			0.014		
21(+)			0.005		
Uncalled[a]	0.025	0.021	0.016	0.008	0.046
Ho	0.760	0.766	0.914	0.734	0.781
He	0.723	0.767	0.842	0.764	0.714
p	0.682	0.342	0.002	0.576	0.010
PIC	0.766	0.804	0.875	0.781	0.791
P_I	0.091	0.061	0.029	0.081	0.076
PD	0.909	0.939	0.971	0.919	0.924
PE	0.498	0.506	0.776	0.462	0.529

[a] All descriptive statistics were obtained from allele frequency data when treating the uncalled data as missing data. The proportion of uncalled data is also provided. Ho: Observed Heterozygosity, He: Expected Heterozygosity, p: probability that values depart from HWE expectations, PIC: Polymorphism Information Content, P_I: Probability of Identity, PD: Power of Discrimination, PE: Power of Exclusion.