Forensic Science Regulator

Overseeing Quality

Forensic Science Regulator Guidance

Allele Frequency Databases and Reporting Guidance for the DNA (Short Tandem Repeat) Profiling

FSR-G-213

Issue 2

1. Introduction

1.1.1 With the introduction of the expanded and more sensitive DNA short tandem repeat (STR) multiplex systems being used in casework and the National DNA Database™(NDNAD) this guidance provides the approach for the use of relevant allele frequency population databases for interpreting DNA profiles using statistical data and evaluation for the UK.

2. Purpose and Scope

2.1.1 The statistical approaches to the interpretation of single-source autosomal STRs (with further guidance on DNA mixture interpretation being available in FSR-G-222 'DNA Mixture Interpretation') considering:

a. the use appropriate population frequency database (s),
b. the recommended values of F_{ST},
c. the size bias correction,
d. Using the likelihood ratio (LR) with qualitative or probabilistic interpretation methodology or a combination of the two.

2.1.2 All guidelines should be supported by an organisation's own internal validation study and published scientific literature as appropriate.

2.2 Standards for DNA profile interpretation

2.2.1 National and international standards (ISO/IEC 17025 and ILAC G19) for testing and calibration in laboratories provide guidance on analytical methods. However, there is much less detail for the type of interpretation of analytical results required for DNA analysis.

2.3 Guidelines within the Forensic Science Regulator's Codes

2.3.1 In addition to the Forensic Science Regulator's (FSR's) Codes of Practice and Conduct (the Codes) the following documents are relevant to this topic:

a. FSR-C-108 DNA Analysis;
b. FSR-G-201 Validation;
c. FSR-G-202 The interpretation of DNA evidence (including low-template DNA);
d. FSR-G-222 DNA Mixture Interpretation;

e. FSR-G-223 Software Validation for DNA Mixture Interpretation;

3. Implementation

3.1.1 This appendix is available for incorporation into a provider's quality management system from the date of publication. The Forensic Science Regulator required that the Codes were included in a provider's schedule of accreditation from October 2017. The requirements in this appendix are effective from 01 October 2020.

4. Modification

4.1.1 This is the second issue of this document. It is a major rewrite of the previous version.

4.1.2 The Regulator uses an identification system for all documents. In the normal sequence of documents this identifier is of the form 'FSR-#-###' where (a) the '#' indicates a letter to describe the type or document and (b) '###' indicates a numerical, or alphanumerical, code to identify the document. For example, this document is FSR-G-213. Combined with the issue number this ensures each document is uniquely identified.

4.1.3 If it is necessary to publish a modified version of a document (e.g. a version in a different language), then the modified version will have an additional letter at the end of the unique identifier. The identifier thus becoming FSR-#-####.

4.1.4 In all cases the normal document bearing the identifier FSR-#-###, is to be taken as the definitive version. In the event of any discrepancy between the normal version and a modified version then the text of the normal version shall prevail.

5. Terms and Definitions

5.1.1 The terms and definitions set out in the Forensic Science Regulators (FSR) Codes of Practice and Conduct (the Codes), FSR-C-108 DNA Analysis, FSR-G-202 DNA Interpretation, FSR-G-222 DNA Mixture interpretation and the Glossary at section 19 apply to this document.

6. Population Groups

6.1.1 Where the population group has no impact on the reported LR (for example because the LR is greater than 1 billion for all relevant groups), the population need not be included in the statement or report. However, where the choice of population group affects the LR, the population group(s), and reasons for selection, shall be included in the report or statement. Additional guidance can be found in FSR-G-222 DNA mixture interpretation.

6.1.2 The main population groups that should be used for calculations are set out in Table 1. The overall representation of the population groups might not reflect local conurbations. Witness information on the person of interest should be considered when choosing relevant population groups for calculations.

6.1.3 Allele frequency data available for these populations are set out in Table 2 and published by the Home Office at www.gov.uk/government/statistics/dna-population-data-to-support-the-implementation-of-national-dna-database-dna-17-profiling.

Table 1. UK Population Group Figures collated from the 2011 UK Census

Population Group	Corresponding UK Census groups (proportion of UK population)	Proportion of UK resident population
White	White British (80.5%) Irish (0.9%) Other White (4.4%) [1]	86%
Black African/Caribbean	African (1.8%) Caribbean (1.1%) Other Black (0.5%)	3.4%
South Asian (Indian subcontinent)	Indian (2.5%) Pakistani (2.0%) Bangladeshi (0.8%)	5.3%
East and South East Asian	Chinese (0.7%) Other Asian (1.5%) [2]	0.7–2.2%

Population Group	Corresponding UK Census groups (proportion of UK population)	Proportion of UK resident population
Middle Eastern/North African	Arab (0.4%) Other Asian (1.5%)[2]	0.4–1.9%
Total		97.3%

Notes:

6.1.4 Description of Other White would include Mediterranean European/Hispanic. Both would presumably be predominantly defined as 'White' in the Census data.

6.1.5 People declaring themselves as 'Other Asian' will presumably include those from Central Asia, the Middle East (not identifying themselves as 'Arab') and parts of Russia, as well as those from East and South East Asia. The reported 1.5 per cent in this group is therefore likely to be divided between East/South East Asian and Middle Eastern/North African groups in an unknown proportion.

7. Availability of Data

7.1 UK data collection

7.1.1 Profile data from consenting individuals has been collated by the UK NDNAD and by King's College, London. Based on the population groups in Table 1, the numbers of individuals from which full 16 STR loci genotypes have been generated is set out in Table 2.

Table 2. Data available for UK population allele frequency databases.

Population Group	Number of alleles	Population sources for tested individuals
White	2,550	British [1]
Black African/Caribbean	770	33% Nigeria 10% Other West African 4% Somalia

Population Group	Number of alleles	Population sources for tested individuals
		27% Jamaica 2% Other Caribbean 25% Unknown
South Asian (Indian subcontinent)	400	20% Pakistan 13% India 14% Afghanistan 11% Bangladesh 43% Unknown (UK residents, self-declared South Asian)
East and South East Asian	406	85% China 4% Vietnam 3% Philippines 8% Unknown
Middle Eastern/North African	110	31% Turkey 25% Iraq 13% Iran 4% Egypt 9% Other Middle Eastern 18% Unknown

Notes

7.1.2 Volunteer donors mainly drawn from student populations and police forces in several UK cities.

7.1.3 Individuals with specified countries of origin were sourced from incoming migrants applying for residency in UK. ‘Unknown’ groups are generally those sourced from the UK resident student populations who were not asked for information on their country of origin.

7.1.4 Comparison between the proportions of each grouping within the 2011 Census data (Table 1) and the sourced individuals (Table 2) suggests that the latter are reasonably representative of the known UK population.

7.1.5 For example, the (2011 Census) Black population in the UK comprises approximately 60 per cent African and 40 per cent Caribbean (excluding ‘Other’). The sourced data set (excluding unknowns) is also 60 per cent

African (mainly Nigerian) and 40 per cent Caribbean (mainly Jamaican). It is recognised that the Nigerian and Jamaican populations may not be fully representative of the resident UK Black population. However, from 2011 Census data, these countries of origin do have the largest UK populations of any African and Caribbean countries (excluding South Africa, whose emigrant population is likely to be partly White).

7.1.6 For the South Asian data, the Indian population is under-represented in the available data (30% of the total available India/Pakistan/Bangladesh data set, compared with 47% in the 2011 Census data). However, the data set does represent all of the major constituent groups and it is likely that this deviation from the population proportions will have only a small impact on calculated likelihood ratio values.

7.1.7 The number of alleles in the Middle Eastern/North African data set is significantly lower than the target size of 400 alleles. Data can, and will, continue to be collected for this group and the databases can be updated on a regular basis. Other sources of individuals from this population group could be identified to accelerate this process.

7.1.8 It is noted that a minority of the samples sourced from populations other than White are from non-resident individuals (incoming migrants). It is recognised that within the UK, admixture between resident populations from different geographical origin has and will continue to occur and that sampling from a well-established resident UK population may have helped to account for this unknown. However, the difficulty of obtaining sufficiently large and representative numbers of samples, with informed consent, from these resident populations made this approach impractical. It is believed that the individuals sampled here provide a reasonable approximation for the resident populations, comprising as they do, reasonably representative proportions of the relevant countries of origin of most UK resident populations.

7.1.9 From this overall data set, individual allele counts for each locus can be determined and these data sets form the core allele frequency databases to be made available to and used by UK forensic science providers. Allele counts, as well as calculated proportions, should be made available to allow appropriate probability calculations to be made by individual users.

7.2 Publication

7.2.1 The population data set meets the minimum criteria for publication of allele frequency data in Forensic Science International (FSI):Genetics, the major international repository for such data (Bodner et al. (2016, 2020). These guidelines require a minimum of 16 autosomal short tandem repeat loci, and at least 500 individuals to be typed. It is noted that the representation is lower for some population groups.

8. Use of Allele Frequencies in Calculations

8.1.1 In addition to the provision of allele counts for each locus, the following aspects of the calculation and reporting of likelihood ratios (LRs) need to be considered.

a. Retention of 'in the order of 1 in 1 billion' as the maximum quoted LR in statements and presented evidence.
b. Allowance for population sub-structure (use and value of theta [θ] or fixation index [F_{ST}]) and selection of appropriate population group(s) within the case contexts (i.e. how many populations to consider).
c. Estimation of allele probabilities: allowance for sampling effects (use of size bias correction / pseudo-counting).
d. Consideration of linkage between syntenic loci.
e. Appropriate methods for the interpretation of mixtures.
f. Appropriate methods for the interpretation of low-level profiles where allele drop-out is expected.

8.2 Mutations

8.2.1 Mutations vary between loci and sex and most are single step. Mutation rates from collections of data have been published in the literature by National Institute of Standards and Technology (NIST). The average mutation rate across all loci except SE33 which is 0.0014.

8.3 Silent alleles

8.3.1 'Silent' allele (sometimes referred to as a null allele) describes the apparent absence of an allele, where one may be expected to be present. This could

be due to allelic drop out, or to changes at the primer binding site which affects primer extension.

8.3.2 Drop out of one (allelic drop out) or both alleles (locus drop out), due to a low level of DNA being present, such as at high molecular weight in a degraded sample, effectively results in a partial profile being obtained. A further sample may be required to obtain a better quality, more complete profile.

8.3.3 Changes at the primer binding site could be due to either:

a. a mutation (base pair change) in the DNA template within the primer binding site region, or

b. an insertion or deletion affecting the primer binding site. This can lead to the primer failing to bind correctly, resulting in little or no primer extension.

8.3.4 As the allele does not amplify, or falls unexpectedly below the detection threshold, it goes undetected or unrecorded, and is considered ‘silent’. These are relatively rare events as flanking regions around STR repeats tend to be stable.

8.3.5 Silent alleles which are as a result of primer binding site changes can be confirmed when the same sample is typed using different primer sets that may be employed in alternative PCR kits, (for example, PowerPlex® ESI 17 and NGM SElect™). A relatively common silent allele is D19S433 allele 15 using the PowerPlex® ESI 17 kit, which is observed when using NGM SElect™.

9. Consideration of Linkage Between Syntenic Loci

9.1.1 With the advent of newer, larger multiplex kits, the selection of STR loci by kit designers and policy makers has eschewed a long held principle of kit design that the loci within the kit should be on different chromosomes (or at least on opposite arms of the same chromosome). In particular, in many kits the inclusion of the vWA and D12S391 loci is problematic because they are located on the same arm of chromosome 12 and separated by a physical distance of about 6.3Mb (or a genetic distance of 12cM) (Budowle et al. (2011)).

9.1.2 As the physical distance between these two loci is relatively large, Bright et al (2013) points out: “A range of 10-30 kb for linkage disequilibrium that is useful for association mapping has been suggested for extensively studied northern European populations and less in African populations. … the closest pair are vWA and D12S391 which are reported as being separated by approximately 6.4 mb, which is more than two orders of magnitude larger than the distance of 10-30 kb quoted above.”

9.1.3 A priori therefore, given their distance apart, any linkage disequilibrium exhibited between the vWA and D12S391 loci is expected to be small and the effects therefore relatively weak.

9.1.4 There is a body of literature discussing the potential for the physical linkage between vWA and D12 to cause linkage disequilibrium at the population level. The conclusions reached by a number of these papers (O’Connor and Tillmar (2012), Gill et al. (2012) and Bright et al. (2013)) are that there is no detectable linkage disequilibrium at these loci at the population level and so it is ‘safe’ to use the product rule to estimate likelihood ratios (LRs) when considering unrelated individuals as the alternative source of the DNA.

9.1.5 A method for correction of the LR is described in J.-A. Bright, J.M. Curran, J.S. Buckleton (2013). However, its general application to all DNA profiles would significantly increase computational complexity for single source profiles and (more especially) for mixtures where the alternative source of the DNA includes the proposition of a relative.

9.1.6 An alternative simplification, suggested in Budowle et al (2011) and K. O’Connor and Tillmar (2012) is to drop one locus from the calculation (retaining “the more informative”) where appropriate (i.e. when the alternative contributor in the LR is a close relative [unless a child or parent]). However Gill et al (2012) advises “caution against an approach that does not make use of all available data”.

9.1.7 Two pairs of loci in the DNA-17 set are syntenic (located on the same chromosome). These are D2S1338 and D2S441 (on chromosome 2) and D12S391 and vWA (on chromosome 12). The D2 loci are on separate arms of the chromosome and are not linked.

9.1.8 The sampling correction is sufficient to account for linkage for situations not involving relatedness.

9.1.9 Until a more sophisticated approach is built into software the approach suggested in Budowle et al (2011) and K. O'Connor (2011) to drop a locus is acceptable. However, if interested in which is the more informative locus it may be necessary to carry out both calculations and report the appropriate calculation.

10. Retention of '1 in 1 Billion' as the Maximum quoted Likelihood Ratio

10.1.1 Hopwood et al. (2012) and Bright et al. (2013) calculated that the minimum LR for a full 15-short tandem repeat (STR) profile (minus the SE33 loci) was of the order of 10^{12}, considering three populations corresponding to White, Black African/Caribbean and South Asians. The same calculation can be made for the East/South East Asian and Middle Eastern/North African populations from the above data, and including SE33 for all populations. From this it is clear that it will not be necessary to calculate an LR in situations where there is a full profile match between a crime sample and a suspect as the maximum '1 billion' figure will be exceeded, with the following exception.

10.1.2 It is now known that the LR for the most common full SGM plus™ profile for the East /South East Asian population does not reach a billion. The actual LR has been confirmed for this population group to have a range of 550 to 663 million. As such, it is now recommended that all SGM plus™ DNA matches to a reference DNA profile of the East /South East Asian population should have a LR calculated and that it should no longer be assumed that the LR is a billion.

10.1.3 Hopwood et al. (2012) also calculated the minimum LR for siblings, half-siblings, uncle–nephew, grandparent–grandchild and first cousins (originally reported in Hopwood et al, (2012), but was corrected in Bright et al, (2013) to account for linkage). From these results, for the 16-STR system it is clear that an LR in the order of 1 billion will be obtained for full profiles in cases where the alternative possible source of the DNA has any level of

relatedness with the person of interest beyond the first degree (siblings and parent/child). As noted in 10.1.2, however, a calculation will be required where the East /South East Asian population group is of relevance.

11. Appropriate use of Different Population Groups

11.1.1 Where a LR is calculated (for partial profiles or in cases including mixed profiles), in practice it is simpler to consider the relevant allele frequencies in the major population groups and to report the corresponding LR most favourable to the person of interest (i.e. the smallest LR). For mixtures derived from two or more individuals, this may result in the consideration of different combinations of unknown contributors from different population groups in order to determine the most conservative scenario.

11.1.2 Other practitioners calculate the relevant LR in the population group matching that of the person of interest only. This approach is generally conservative: if the alternative DNA source has a different population group from the person of interest, using the database appropriate for the latter, together with an appropriate F_{ST} adjustment to allow for co-ancestry, tends to give a lower LR than when using the database matching the population group of the alternate source. This approach can be made as conservative as desired by using a sufficiently large value of F_{ST}. A simulation experiment using the White, Black African/Caribbean South Asians and East/South East Asian databases and simulated single-source profiles comprising the 16-STR loci in the DNA-17 locus set found that using $F_{ST} = 0.03$ (3 percent) and the same population group as the person of interest gave an LR that, in over 99.9 per cent of cases, was lower than the LR computed using any of the other three population groups and $F_{ST} = 0$ (zero), irrespective of which database the profile was simulated from. In a similar simulation experiment using 2-person mixtures, this approach was conservative compared with the alternative calculations considered in at least 99.3 per cent of the simulations, and in the few instances that it was not conservative the difference was almost always small.

12. Use of a Stratified Database

12.1.1 As an alternative to the recommended approach of using one of five different population groups to determine allele frequencies, consideration can be given to the use of a stratified database. This single calculation suitably weighted to reflect the proportions of different population groups within the entire UK population (or an appropriate regional sub-set to represent the pool of possible perpetrators for any given crime). Although this has some merits, it is computationally challenging, especially with respect to mixed profiles. It also raises further uncertainties as to the appropriateness of the chosen population of possible perpetrators, whether national or regional. It is not recommended that a stratified database approach be adopted for reporting LRs for general casework matches at this time.

13. Estimation of Probabilities of Alleles: Allowance for Sampling Effects (use of 'Size Bias' or Pseudo-Counting)

13.1.1 The practice of adjusting the unbiased estimate of allele frequencies to account for sampling error has been widely adopted in the UK and elsewhere. The introduction of additional loci does not change this requirement. It is recommended that practitioners and reporting organisations continue to use a method that accounts for sampling errors, such as those described by Balding (1995), Evett and Weir (1998) or Curran et al (2002) and Curran et al (2011) so that no more prescriptive recommendation is required.

14. Allowance for Sampling and Sub Population Effects (use and Value of θ or F_{ST})

14.1.1 The routine use of $F_{ST} = 0.03$ is appropriate, rising up to 0.05 in unusual cases involving small and isolated populations that may be highly differentiated from available databases.

14.1.2 This issue was addressed by Hopwood et al (2012), who concluded that: “An analysis of the population data for the three major populations of the UK, and comparison with other similar populations has provided us with a calculated value for θ, confirming that a value of 0.02 remains conservative in calculating the LR.”

14.1.3 This more conservative F_{ST} value is based on an extensive set of F_{ST} estimates given by Steele, Syndercombe Court and Balding (2010). These analyses use a dataset similar to that described above. It was found that F_{ST} = 0.02 was nearly always conservative, but in some cases a larger value was required, for example, for Latin Americans relative to the White population dataset. It was found that Somali allele frequencies are actually closest to the Middle Eastern/North African population group (smallest F_{ST}), but based on physical appearance and the geographical location of Somalia, it is likely that Somalis will in practice often be compared with the Black population dataset. The use of F_{ST} = 0.03 will ensure that the result tends to be conservative whichever reference population dataset is used.

14.1.4 The F_{ST} has usually been thought of as accounting for the excess allele sharing, relative to databases allele frequencies, for suspected and alternative contributors from the same subpopulation. However, as discussed above, there is another role for the F_{ST}, which is to make the LR sufficiently conservative that it is almost certainly favourable to defendants even allowing for alternative contributors to come from very different ethnic populations.

15. Acknowledgements

15.1.1 This guidance was produced by the Forensic Science Regulator’s DNA Analysis Specialist Group and the Forensic Science Regulation Unit (FSRU).

16. Review

16.1.1 This published guidance will form part of the review cycle as determined by the Forensic Science Regulator.

16.1.2 The Forensic Science Regulator welcomes comments. Please send them to the address as set out at: www.gov.uk/government/organisations/forensic-science-regulator, or email: FSREnquiries@homeoffice.gov.uk

17. References

Balding, D. J. (1995) 'Estimating products in forensic identification using DNA profiles', J. Am. Stat. Assoc., vol. 90, pp 839–844.

Balding, D. J., and Nichols, R.A. (1994) 'DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands', Forensic Science International, vol. 64, pp 125–140.

Bodner, M., Bastisch, I., Butler, J. M., Fimmers, R., Gill, P., Gusmão, L., Morling, N., Phillips, C., Prinz, M., Schneider, P. M. and Parson, W. (2016) 'Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER).', Forensic Science International: Genetics, vol. 24, pp 97–102.

Bodner, M. and Parson, W. (2020) The STRidER Report on Two Years of Quality Control of Autosomal STR Population Datasets. Genes Vol. 11(8), 901; doi:10.3390/genes11080901.

Bright, J.-A., Curran, J. M. and Buckleton, J. S. (2013) 'Relatedness calculations for linked loci incorporating subpopulation effects', Forensic Science International: Genetics, vol. 7, pp 380–383.

Bright, J., Curran, J. M., Hopwood, A. J., Puch-Solis, R. and Buckleton, J. S. (2013b) 'Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts I (corrigendum)', Science & Justice, vol. 53, p 371.

Budowle, B., Ge, J., Chakraborty, R., Eisenberg, A. J., Green, R., Mulero, J., Lagace, R. and Hennessy, L. (2011) 'Population genetic analyses of the NGM STR loci', International Journal of Legal Medicine, vol. 125, pp 101–109.

Curran, J. M. and Buckleton, J. S. (2011) 'An investigation into the performance of methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations', Forensic Science International: Genetics, 5(5):512-6. doi: 10.1016/j.fsigen.2010.11.007.

Curran, J. M., Buckleton, J. S., Triggs, C. M. and Weir, B. S. (2002) 'Assessing uncertainty in DNA evidence caused by sampling effects', Science & Justice, vol. 42, pp 29–37.

Evett, I. W. and Weir, B. S. (1998) Interpreting DNA evidence: Statistical genetics for forensic scientists. Sinaur Associates Inc, ISBN 0 87893 155 4.

Gill, P., Phillips, C., McGovern, C., Bright, J. and Buckleton, J. (2012) 'An evaluation of potential allelic association between the STRs vWA and D12S391: Implications in criminal casework and applications to short pedigrees', Forensic Science International: Genetics, vol. 6, pp 477–486.

Forensic Science Regulator Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System. Available at: www.gov.uk/government/collections/forensic-science-providers-codes-of-practice-and-conduct [Accessed 10/03/2020].

Forensic Science Regulator DNA Analysis, FSR-C-108. Available at: www.gov.uk/government/collections/forensic-science-providers-codes-of-practice-and-conduct#appendices [Accessed 09/07/2018]

Forensic Science Regulator Software validation for DNA mixture interpretation, FSR-G-223. Available at: https://www.gov.uk/government/publications/software-validation-for-dna-mixture-interpretation-fsr-g-223 [Accessed 19/06/2020]

Forensic Science Regulator The interpretation of DNA evidence, FSR-G-202. Available at: www.gov.uk/government/publications/the-interpretation-of-dna-evidence [Accessed 19/06/2020]

Forensic Science Regulator Validation, FSR-G-201. Birmingham: Forensic Science Regulator. Available at: www.gov.uk/government/publications/forensic-science-providers-validation [Accessed 27/05/2020].

Home Office Data to support the implementation of national DNA database DNA-7 profiling. Available at: www.gov.uk/government/statistics/dna-population-data-to-support-the-implementation-of-national-dna-database-dna-17-profiling [Accessed 25/06/2020]

Hopwood, A. J., Puch-Solis, R., Tucker, V. C., Curran, J. M., Skerrett, J., Pope, S. and Tully, G. (2012) 'Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts', Science & Justice, vol. 52, pp 185–190.

International Society for Forensic Genetics (2013) 'New guidelines for the publication of genetic population data', Forensic Science International: Genetics, vol. 7, pp 217–220.

ILAC G19:08/2014 Modules in a Forensic Science Process. International Laboratory Accreditation Cooperation. Available at: http://ilac.org/news/ilac-g19082014-published/ [Accessed 27/05/2020].

ISO/IEC 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories.

NIST Apparent Mutations Observed at STR Loci in the Course of Paternity Testing. Available at: http://www.cstl.nist.gov/biotech/strbase/mutation.htm [Accessed 17/01/2020].

O'Connor, K. L. and Tillmar, A. O. (2012) 'Effect of linkage between vWA and D12S391 in kinship analysis', Forensic Science International: Genetics, vol. 6, issue 6, pp 840–844.

Steele, C. and Balding, D. (2014) 'Choice of population database for forensic DNA profile analysis', Science & Justice, vol. 54, issue 6, pp 487–493. DOI: 10.1016/J.SCIJUS.2014.10.004.

Steele, C., Syndercombe-Court, D. and Balding, D. (2014) 'Worldwide F_{ST} estimates relative to five continental-scale populations', Annals of Human Genetics, vol. 78, pp 468–477.

18. Abbreviations and Acronyms

Abbreviation	Meaning
DNA	Deoxyribonucleic acid
FSI	Forensic Science International
FSR	Forensic Science Regulator
F_{ST}	Fixation Index
LR	Likelihood Ratio
ICCA	The Inns of Court College of Advocacy
NDNAD	National DNA database™
NIST	National Institute of Standards and Technology
RSS	the Royal Statistical Society
STR	Short Tandem Repeat
UK	United Kingdom

19. Glossary

Allele:

A genetic variant at a particular location within an individual's DNA. DNA profiling tests examine a range of alleles that are known to vary widely between individuals. Alleles are represented by peaks in a DNA profile.

Allelic Drop-Out:

Allele(s) missing from a DNA profile, so that it is partially represented.

Autosomal DNA:

Any chromosome that is not a sex-determining chromosome.

Chromosome:

A threadlike structure of nucleic acids in the cell that carries genetic (hereditary) information in the form of genes.

Contamination (Profile):

Spurious DNA profile(s). The contributors are considered to be of no relevance to the case (for example, may be introduced into plastic ware during the manufacturing process, or may have originated from a scientist processing the samples in the laboratory).

DNA-17 System:

Short tandem repeat (STR) multiplex system (kit) with 17 STR loci (including the gender marker amelogenin).

DNA Profile:

This is a format for the representation of an individual's genetic information that can be compared to other profiles, for example stored on a database.

Genotype:

An individual's collection of genes as characterised from the alleles present at each genetic locus.

Likelihood Ratio:

This is the ratio of two probabilities; the probability that the observations would have been obtained if the prosecution proposition were true divided by the probability that the observations would have been obtained if the defence proposition were true.

Locus (Plural Loci):

A specific location or position of an allele on a chromosome. Short tandem repeats (STRs) are examples of loci that are of interest in forensic science because they are polymorphic and are therefore highly discriminatory when several are analysed in combination to generate a DNA profile.

Primer Binding Site Mutation (PBSM):

Occurs when there is a mutation on a DNA strand and the primer is either not able to attach or is unable to attach efficiently for DNA amplification. In

extreme cases this can result in a silent or null allele where a heterozygote locus appears as a homozygote or in less extreme cases as peak imbalance.

Short Tandem Repeat (STR):

A microsatellite consisting of one to six or more nucleotides that is repeated adjacent to each other along the DNA strand.

Syntenic Loci:

When two or more loci are present on the same chromosome.

20. Further Reading

Forensic Science Regulator: Validation – Use of casework material validation, FSR-P-300. Birmingham: Forensic Science Regulator. Available at: www.gov.uk/government/publications/protocol-using-casework-material-for-validation-purposes (Accessed on 20/08/2020)

ICCA/RSS Statistics and probability for advocates. Available at: www.icca.ac.uk/wp-content/uploads/2019/11/RSS-Guide-to-Statistics-and-Probability-for-Advocates.pdf (Accessed on 20/08/20).

Pope, S., Puch-Solis, R., & Roberts, P., Aitken, C. (2012). Practitioner Guide No 2: Assessing the Probative Value of DNA Evidence, Guidance for Judges, Lawyers, Forensic Scientists & Expert Witnesses. Available at: rss.org.uk/RSS/media/News-and-publications/Publications/Reports%20and%20guides/rss-assessing-probative-value.pdf [Accessed 20/08/ 2020]

Sense About Science Making Sense of Forensic Genetics (2017). Available at: senseaboutscience.org/wp-content/uploads/2017/01/making-sense-of-forensic-genetics.pdf [Accessed 20/08/2020]

The Royal Society and the Royal Society of Edinburgh (2017) Forensic DNA Analysis – A primer for courts. Available at: www.royalsociety.org/~/media/about-us/programmes/science-and-law/royal-society-forensic-dna-analysis-primer-for-courts.pdf (Accessed on 20/08/2020)

UKAS® (2016) UKAS Policy on Participation in Proficiency Testing, TPS 47, edition 3, issued November 2016. United Kingdom Accreditation Service. Available at: www.ukas.com/download/publications/Technical%20Policy%20Statements/TPS-47-Participation-in-PT-Edition-4-February-2020.pdf (Accessed on 20/08/2020)

Understanding the use of statistical evidence in courts and tribunals (2017). Available at: juliamortera.files.wordpress.com/2017/12/statistics-and-probability-for-advocatesbooklet-guide-icca-rss1.pdf [Accessed 20/08/ 2020]

Published by:

The Forensic Science Regulator

5 St Philip's Place

Colmore Row

Birmingham

B3 2PW

www.gov.uk/government/organisations/forensic-science-regulator