Forensic Science Regulator

Guidance: The use of allele frequency databases to assign a likelihood ratio (LR) in human DNA cases

FSR-GUI-0012

Issue 1

Publication date February 2026

This document is issued by the Forensic Science Regulator in line with Section 9(1) of the Forensic Science Regulator Act 2021.

Contents

1. Introduction

1.1.1 This guidance provides the recommended approach for the use of relevant allele frequency population databases for interpreting autosomal DNA STR and to report the strength of evidence associated with these systems.

2. Purpose and Scope

2.1.1 The statistical approaches to the interpretation of single-source autosomal STRs considering:

a. the use of appropriate population frequency database (s),
b. the recommended values of F_{ST},
c. the use of appropriate sampling corrections,
d. Using the likelihood ratio (LR) with probabilistic interpretation methodology.

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.

3. Implementation

3.1.1 This guidance document is available for incorporation into a provider's quality management system from the date of publication. The Forensic Science Regulator required that the Code was included in a provider's schedule of accreditation from October 2017. The requirements in this guidance are effective from January 2026.

4. Terms and Definitions

4.1.1 The terms and definitions set out in the Forensic Science Regulator's (FSR) Code of Practice [1] (the Code) and the Glossary at section 22 apply to this document.

5. Allele Frequency data

5.1.1 The statistical evaluation of autosomal STR results in forensic analysis depends on having allele frequency data to compute a genotype probability. Allele frequency data are obtained from collections of STR genotypes taken from random individuals. These individuals represent a convenience sample drawn from a particular population of interest.

5.1.2 In the UK, it is customary to define these populations according to the ethnicity of the subjects – i.e. their ethnic appearance (EA group). Currently, White European individuals are assigned to EA1 (North European) or EA2 (South European), Black African and Caribbean individuals to EA3, South Asian individuals (Indian subcontinent) to EA4, East and Southeast Asian individuals to EA5 and Arab individuals (Middle East/North Africa) to EA6 [2]. These groupings account for most individuals in the UK (see Table 1).

5.1.3 From the genotypes of the collected samples, the number of observations of each allele at each STR locus can be determined (allele count). From the allele counts, an estimate of the frequency of that allele in the relevant population can be made and an allele frequency database compiled for each population. The estimated frequencies can later be used to compute a likelihood ratio for a locus to evaluate the evidential support for alternative propositions relevant to the case.

5.1.4 To ensure representativeness, the individuals sampled for allele frequency data should be unrelated to each other. Whilst it is difficult to exclude the possibility of some distant family relatives being present in any set of random individuals, efforts should be made to avoid sampling from known close relatives. Furthermore, the sample should comprise a minimum of 200 alleles

at each locus i.e. 100 individuals [3]. It is self-evident that the larger the sample size, the more precise will be the estimate of the population allele frequency. It is proposed that, dependent on sample availability, a target of 400 alleles per population group will be the recommended minimum.

6. UK population

6.1.1 The main population groups resident in England and Wales, drawn from the 2021 Census [4] are shown in Table 1. The overall proportions of the population groups might not reflect those in local conurbations.

Table 1. England and Wales Ethnic Group Figures collated from the 2021 UK Census [4]

Population Group	**Corresponding UK Census groups (proportion of UK population)**	**Proportion of UK resident population**
White	British (74.4%) Irish (0.9%) Other White (6.5%)[1]	81.8%
Black African/Caribbean	African (2.5%) Caribbean (1.0%) Other Black (0.5%)	4.0%
South Asian (Indian subcontinent)	Indian (3.1%) Pakistani (2.7%) Bangladeshi (1.1%)	6.9%
East and Southeast Asian	Chinese (0.7%) Other Asian (up to 1.6%)[2]	0.7%
Middle Eastern/North African	Arab (0.6%) Other Asian (up to 1.6%)[2]	0.6%
Total		95.5%

[1] Sum of 3 census categories: White: Gypsy or Irish Traveller (0.2%), White: Roma (0.1%) and White: Other white (6.2%)

[2] People declaring themselves as 'Other Asian' may include those from Central Asia, parts of the Middle East (West Asia), East Asia (excluding China), South East Asia, Sri Lanka, Nepal, Bhutan, the Caucasus, parts of Russia and Mongolia. The reported 1.6% in this group is therefore divided between the East/South East Asian, Middle Eastern/North African and South Asian groups in an unknown proportion. Values for individual population groups in the table are not adjusted for this additional unknown contribution. However, the total 95.5% figure does include the 1.6% "Other Asians".

7. UK DNA17 allele frequency database

7.1.1 Samples from consenting individuals were collected by the UK NDNAD and by King's College, London in 2012/13, prior to implementation of the DNA17 STR systems in the UK NDNAD. Based on the population groups in Table 1, the numbers of individuals from which full 16-locus STR genotypes have been generated to provide allele frequency data are set out in Table 2.

Table 2. Composition of UK DNA17 population allele frequency databases.

Population Group	Number of alleles (n)	Population sources of tested individuals[2]
White	2,550	British [1]
Black African/Caribbean	770	33% Nigeria 10% Other West African 4% Somalia 27% Jamaica 2% Other Caribbean 25% Unknown (UK residents, self-declared)
South Asian (Indian subcontinent)	400	20% Pakistan 13% India 14% Afghanistan 11% Bangladesh 43% Unknown (UK residents, self-declared)

Population Group	Number of alleles (n)	Population sources of tested individuals[2]
East and Southeast Asian	406	85% China 4% Vietnam 3% Philippines 8% Unknown(UK residents, self-declared)
Middle Eastern/North African	110	31% Turkey 25% Iraq 13% Iran 4% Egypt 9% Other Middle Eastern 18% Unknown)UK residents, self-declared)

Notes

[1] White British donors mainly drawn from student populations and police forces in several UK cities.

[2] 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.

7.1.2 Comparison between the proportions of each population within the 2021 Census data (Table 1) and the sourced individuals (Table 2) suggests that the latter are reasonably representative of the known UK population.

7.1.3 For example, the (2021 Census) Black population in the UK comprises approximately 70% African and 30% Caribbean (excluding 'Other'). The sourced data set (excluding unknowns) is approximately 60% African (mainly Nigerian) and 40% 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 2021 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).

7.1.4 It is noted that North East African populations, particularly from Somalia, Ethiopia, Djibouti and Eritrea, form a genetically distinct population group.

According to Hodgson et al. (2014) [5], both the African ancestry (Ethiopic) and the non-African ancestry (Ethio-Somali) in Cushitic speaking populations from this region are significantly differentiated from all neighbouring African and non-African ancestries in East Africa, North Africa, and the Middle-East (the Levant and Arabia). However, at present, no separate UK-published DNA17 population data is available for this population group.

7.1.5 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 45% in the 2021 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.

7.1.6 The number of alleles in the Middle Eastern/North African data set is significantly lower than the target size of 400 alleles. This data set was excluded from the published Home Office data [6]. Data for this population will be published as part of the expansion to DNA24 loci (section 8 of this document).

7.1.7 It is noted that most 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.

7.1.8 From this overall data set, individual allele counts for each locus can be determined and these data sets form the core allele frequency databases for DNA17 multiplexes made available to, and used by, UK forensic science

providers [6]. Total allele counts, as well as calculated proportions, should be made available to allow appropriate probability calculations to be made by individual users.

8. UK allele frequency database – DNA24

8.1.1 The UK National DNA Database was reconfigured to accept DNA17 STR genotypes in 2014 when the DNA17 allele frequency database was published by the Home Office [6]. Since that time, the number of available STR multiplex systems has continued to increase with several multiplexes having 24 or more STR loci (DNA24). A number of these multiplexes are routinely used by UK Forensic Science Providers alongside, or instead of, the DNA17 systems.

8.1.2 It is intended to expand the centrally published frequency databases to include additional loci present in the DNA24 multiplexes. This will enable appropriate and consistent statistical evaluations to be made by UK providers using these systems in casework investigations.

9. Use of individual allele frequencies to compute expected genotype probabilities.

9.1.1 Allele frequency data can be used to compute expected genotype frequencies for any combination of STR alleles at a given locus in a nominated population. For example, the frequency of the TH01 6, 9 genotype in an Asian population can be estimated if the frequencies of the 6 and the 9 allele in that population are both known.

9.1.2 In population genetics, the Hardy-Weinberg (HW) principle states that allele and genotype frequencies will remain constant from generation to generation, and that expected genotype frequencies are p^2 for a homozygote PP and $2pq$ for a heterozygote PQ (where p and q are the allele frequencies of alleles P and Q).

9.1.3 The HW equilibrium is an idealised description of a genetic system and makes certain assumptions:

a. Natural selection is not acting on the locus.
b. Neither mutation nor migration are introducing new alleles into the population.
c. Population size is very large, so there is no impact from genetic drift.
d. Individuals in the population mate randomly.

9.1.4 Under these assumptions, genotype frequencies for all homozygote or heterozygote STR loci in a DNA17 profile could be simply calculated using p^2 and $2pq$. However, not all the above assumptions hold for a real-world population. Nevertheless, the HWE serves as a very useful model and approximates well the genotype probabilities in human populations. For example, some genetic drift is expected in non-idealised-HWE situations, but it is likely to take multiple generations to impact significantly on calculated genotype frequencies. Furthermore, small departures from HWE can be accommodated by including an allowance for population substructure which tends to induce excess homozygosity (known as the Wahlund Effect) – see below.

10. Strength of Evidence

10.1.1 The strength of evidence for DNA analysis in forensic cases can be evaluated by computation of a likelihood ratio which expresses the relative likelihood of the observed genotype results under two alternative propositions. These are commonly referred to as Hp (the “prosecution proposition”) and Hd (the “defence proposition”). To compute the LR, propositions for both Hp and Hd must first be formulated.

10.1.2 Hp is relatively straightforward to define and will include the person of interest (POI). For example, Hp may be that the DNA originates from the POI. For a single source stain yielding a full STR profile designated at all loci, which is also a full match to the genotype of the POI, then this likelihood is assigned to be 1 (i.e., P(E|Hp)=1), where E is the observed evidence (i.e. the genotype observed in the crime stain sample). See below for the case where there are missing alleles under Hp.

10.1.3 The specification of Hd requires deeper consideration. In some cases, it may be specified, for example, the Hd is that the DNA originates from another random person, unrelated to the POI. On many occasions no defence account may be forthcoming and so a 'proxy' proposition is required. It is considered reasonable to assume that a defendant would contest being the source of the DNA and that, unless the circumstances dictated otherwise, the alternative contributor would be a 'random' unknown person unrelated to the POI.

10.1.4 In this context, the concept of unrelatedness is somewhat diffuse in that all humans share a common ancestor possibly as recently as a few thousand years ago. Furthermore, our recent pedigrees extend to include many 2nd, 3rd or more distant cousins, undoubtedly "related" but often outside of our immediate known family groups. In standard forensic casework it has been common practice to either consider a list of nominated close relatives, or a definition based on the expected degree of DNA sharing to define who is considered "related" or "unrelated" when formulating propositions. For example, a common, if arbitrary definition is that anyone sharing less than 12.5% on average of their DNA identical by descent can be considered as unrelated in this context (see Figure 1 for named relationships meeting this criterion). This is based on the expectation that the LR obtained, where a more distant unknown relative is considered as the alternative source of the DNA under Hd, will approach the value obtained where an unrelated individual is considered as the alternative source [7]. Closer relatedness propositions under Hd are considered further in section 15 of this document.

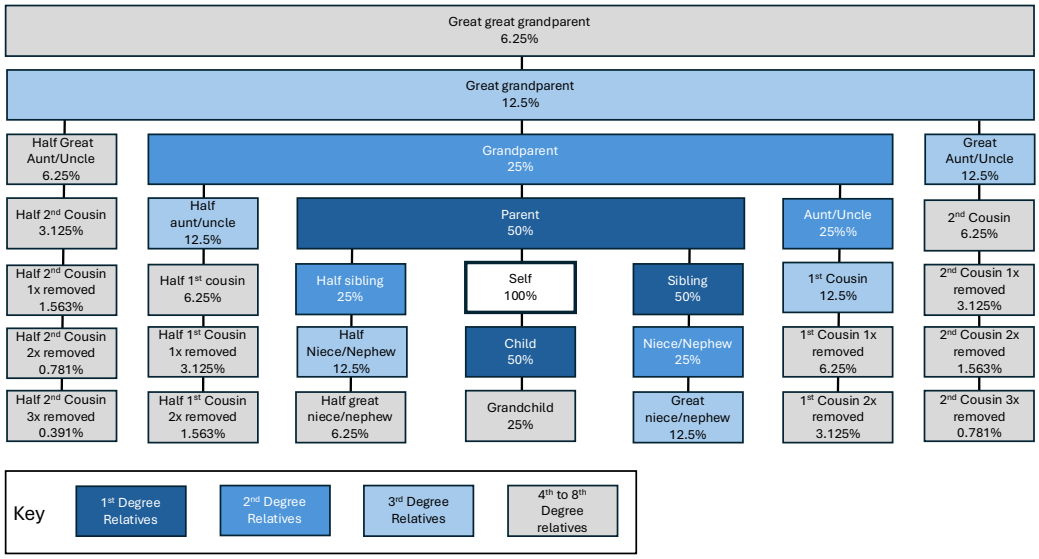

Figure 1 Consanguinity chart showing the expected average degree of DNA sharing for different family relationships.

10.1.5 The expected genotype probability (often referred to as the match probability, mp) for the random individual can be computed using the allele frequencies for the corresponding population sample. For this simple scenario of a single source crime stain profile with a full match to the POI where Hd is that the profile originated from another unrelated individual, then

$$LR = \frac{P(E|Hp)}{P(E|Hd)} = \frac{1}{mp}$$

11. Use of a sampling allowance – multiple methods

11.1.1 The allele counts obtained from a sample of individuals selected from a target population are, by definition, a sample statistic rather than a population parameter. It is legitimate to consider the precision of this statistic especially for those alleles with low counts and, specifically whether to apply some kind of sampling allowance to compensate. Any small sample drawn from a very much larger population will inevitably suffer from sampling variability and this

will be seen most acutely for the rarer alleles (those with few observations in the dataset).

11.1.2 Moreover, where an observed allele is not present in the frequency database and the allele count in the drawn sample is therefore zero, unless a method is deployed to assign a suitable value to the previously unobserved allele, the presence of a zero in the LR calculation leads to an incalculable LR for the locus concerned (as it will include division of the numerator by zero).

11.1.3 Three methods have found favour: minimum default, pseudocounting and Bayesian estimator. No preference is expressed herein.

11.1.4 All three methods lead to inflated counts to compensate for infrequent alleles and the by-product of their application is to ensure that there are no zero frequency alleles involved in LR calculations.

11.2 Minimum default

11.2.1 NRCII – 1996 p148 [8] suggests using a minimum default value of 5/2N (where N=number of individuals in sample database). Using this approach, the minimum default frequency for rare and zero alleles is governed by the size of the database (N).

11.3 Pseudocounting (Size Bias method of Balding and Nichols [9])

11.3.1 This approach comprises temporarily adding the alleles of the person(s) of interest and the crime stain to the database. So, for a single source profile, an identical heterozygous locus genotype in the POI and the crime stain profile would involve adding 2 to each respective allele count and increasing the size of the database by 4 alleles: f = xi+2/2n+4 for each allele (where f – the frequency, n = number of alleles in the sample database and xi = the number of observations of allele i in the database). For a homozygous locus in which the two alleles are identical, this would involve adding 4 to the database count for that allele, leading to a concomitant increase of 4 alleles to the total 2n alleles (f =xi+4/2n+4), again increasing the size of the

database by 4 alleles. Using this approach, the minimum default frequency for rare and zero alleles in a heterozygote is governed by the size of n.

11.4 Bayesian estimator

11.4.1 Bayesian estimator(f_a), with either a uniform or informed prior such as 1/k (where k is the number of alleles in the database with non-zero observations) is given as:

$$f_a = \frac{x_i + \frac{1}{k}}{n+1}$$

(where x_i is the number of observations of allele "i" in the database and n is the total count of alleles for that locus in the sample database) [10])

11.4.2 This may be adapted to a 1/(k+2) prior to allow for unobserved alleles (the Dirichlet prior):

$$f_a = \frac{x_i + \frac{1}{k+2}}{n+1}$$

12. Use of the Product Rule to compute a combined LR

12.1.1 An LR computed for locus 1 may be multiplied by the LR for locus 2 to give a multi-locus LR. This cross-multiplication is known as the Product Rule and relies on an assumption of independence between loci in the multiplex kit(s) in use. This is the assumption of linkage equilibrium (LE).

12.1.2 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 multiplex design that the loci within the multiplex should be located on different chromosomes (or at least on opposite arms of the same chromosome). Many larger multiplex systems now include "syntenic" loci (i.e., loci present on the same chromosome) for which the possibility of linkage disequilibrium should be considered [11] [12].

12.1.3 In many kits the inclusion of the vWA and D12S391 loci could be viewed as problematic because they are located on the same arm of chromosome 12

and separated by a physical distance of about 6.4Mb (or a genetic distance of 13cM) [11] [12]. Other pairs of syntenic loci within the DNA 24 multiplexes are considered to be sufficiently distant for any effects of linkage to be disregarded.

12.1.4 The physical distance between these two loci is relatively large when considering linkage studies. Bright et al [13] point 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.” A priori therefore, given their distance apart, any linkage disequilibrium exhibited between the vWA and D12S391 loci is expected to be minor and the effects relatively weak.

12.1.5 In addition, there is a body of literature discussing the potential for the physical linkage between vWA and D12S391 to cause linkage disequilibrium at the population level. The conclusions reached by a number of these papers [13] [14] [15] are that there is no detectable linkage disequilibrium at these loci at the population level and so it is reasonable to use the product rule to estimate likelihood ratios (LRs) when considering unrelated individuals as the alternative source of the DNA. However, this conclusion may not apply if related individuals are involved, particularly in cases where there may be multiple inheritances from the same individual, such as cases of incestuous paternity.

12.1.6 A simple approach to address this, suggested in Budowle et al [16] and O’Connor and Tillmar [15] is to omit one locus from the calculation (retaining “the more informative” locus) where appropriate (i.e., when the alternative contributor in the LR is a close relative [unless a child or parent]). However, Gill et al [14] advises “caution against an approach that does not make use of all available data”.

12.1.7 A formal mathematical method for correction of the LR using recombination fractions is described in J.-A. Bright, J.M. Curran, J.S. Buckleton [13]. 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 of the POI.

12.1.8 The D2S1338 and D2S441 loci in the DNA-17 set, although syntenic but are on separate arms of the chromosome and therefore considered unlinked.

12.1.9 Even if all markers in a multiplex were physically unlinked, LE is not guaranteed as it may be perturbed by population substructure or by other effects. For this reason, inclusion of an allowance for population substructure has been deemed to be desirable (see next section).

13. Use of a subpopulation allowance – sampling formula and value of F_{ST} (θ)

13.1.1 It is a commonly held misconception that if alleles are physically unlinked on the chromosome they must be in LE. That is not the case as the LE may be perturbed by other effects; most notably by population substructure. The sampling formula proposed by Balding and Nichols [9] has found favour as a method by which a subpopulation allowance may be included in the evidential evaluation.

This relies on the sampling formula below:

$$p(a|k_1, k_2, \ldots) = \frac{n_a\theta+(1-\theta)p_a}{1+(m-1)\theta}$$

where n_a is the number of times allele a appears in profiles k_1, k_2… who are from the same subpopulation and m is the number of alleles in k_1, k_2… and θ is the co-ancestry coefficient (F_{ST}). The sampling formula is applied in a

recursive fashion to each subsequent allele sampled from the relevant subpopulation.

13.1.2 Use of the formula requires the practitioner to specify a numeric value for θ. A routine value of θ = 0.03 has been agreed nationally to be appropriate for criminal casework, rising to 0.05 in unusual cases involving small and isolated populations that may be highly differentiated due to endogamous practices [17].

13.1.3 The value of θ was addressed by Hopwood et al [18], 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 F_{ST}, confirming that a value of 0.02 remains conservative in calculating the LR."

13.1.4 The still more conservative θ value of 0.03 is based on the work of Steele, Syndercombe-Court and Balding [17]. It was found that θ = 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. Allele frequencies from Somalia were closest to the Middle Eastern/North African population group (smallest θ) but based on physical appearance and the geographical location of Somalia, it is likely that Somalis will in practice often be located with the Black population dataset. The use of θ = 0.03 was selected to ensure that the result tended to be conservative whichever reference population dataset had been selected [19].

13.1.5 F_{ST} has traditionally been thought of as accounting for the excess allele sharing, relative to database allele frequencies, for suspected (POI) and alternative contributors from the same subpopulation. However, as used above, the secondary role for the F_{ST} allowance is to make the LR sufficiently conservative that it is almost certainly favourable to defendants, even allowing for alternative contributors to come from different ethnic populations or from a mixed ethnic background.

13.1.6 It is therefore appropriate that, under Hd, the unknown person is deemed to belong to the same population group as the POI. This approach is generally conservative even if the alternative DNA source actually has a different population group from the POI. Using the database appropriate for the POI, 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}.

13.1.7 A simulation experiment [20] using the White, Black African/Caribbean South Asian 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 $\theta = 0.03$ (3%) and the same population group as the person of interest gave an LR that, in over 99.9% of cases, was lower than the LR computed using any of the other three population groups and $\theta = 0$ (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 % of the simulations, and in the few instances that it was not conservative the difference was almost always small.

14. Use of a capped LR

14.1.1 It is recommended that any LRs calculated to be more than one billion (10^9) are not reported as their numerical calculated value but are reported as “in excess of one billion” or “at least one billion”. The phrase “in the order of one billion” has been used historically but does not accurately represent calculated LRs that may exceed one billion by many orders of magnitude so is not preferred.

14.1.2 Hopwood et al. [18] and Bright et al. [21] calculated that the minimum LR for a full single source 15-locus STR profile (i.e. the DNA17 locus set minus the SE33 loci) matching a POI was of the order of 10^{12}, considering three populations corresponding to White, Black African/Caribbean and South

Asian. The same calculation for the East/Southeast Asian and Middle Eastern/North African populations using the UK published data and including SE33 for all populations has a similar outcome. From this it is concluded that it is unnecessary to compute a specific LR in situations where there is a full 16-locus profile (DNA17) match between a crime sample and a suspect (as the capped 'one billion' figure will be comfortably exceeded).

14.1.3 However, it has been calculated that the LR for the most common full 10-locus SGM plus™ profile for the East /Southeast Asian population does not reach a billion. The actual minimum LR for this population group is in the range of 550 to 663 million [22]. As such, it is recommended that all SGM plus™ DNA matches to a POI from the East /Southeast Asian population should have a LR calculated and that it should not be assumed that the LR is greater than one billion.

15. Allowing for relatedness under H_d

15.1.1 The assumption of unrelatedness under Hd may be revised in some circumstances. This might be due to a specific defence account in which the POI suggests that a relative of theirs may be the source of the DNA profile.

15.1.2 Standard formulae have been published [9], [23] based on the number of alleles that are identical by descent (IBD) denoted as Z0 (no alleles shared by descent), Z1 (one allele shared) or Z2 (both alleles shared).

15.1.3 Hopwood et al. [18] also calculated the minimum LR for siblings, halfsiblings, uncle–nephew, grandparent–grandchild and first cousins (originally reported in Hopwood et al, [18] but later corrected in Bright et al, [21] to account for linkage). These results demonstrate that for the 15-STR system, an LR of greater than 1 billion would be obtained for full profile matches in any case where the alternative source of the DNA has any level of relatedness with the person of interest beyond siblings and parent/child.

16. Dealing with allelic drop out

16.1.1 Allelic drop out refers to the non-appearance of an allele where one would be expected if the POI was the source of the DNA. Classically, this is seen as a single allele at a locus where two are expected (heterozygote) were the POI a contributor.

16.1.2 Historically, the invocation of allelic dropout was handled statistically in single source results by applying the 2p rule (or more strictly the $2p\text{-}p^2$ rule) [24]. Such an approach is, however, no longer recommended since it can lead to an overstatement of the evidence (especially when dropout has been invoked but is statistically unlikely based on the height of the surviving allele).

16.1.3 It is possible to probabilistically allow for apparent allelic drop out using probability density functions and computation of the area under the curve that falls below the preset analytical threshold (AT) of the system in use. Several other models for allowing for allelic drop out have been proposed all of which have merit. No preference is stated herein save to discourage use of the 2p rule.

16.1.4 In a single source result, where both alleles are missing at the locus, it may be assigned an LR=1.

17. Acknowledgements

17.1.1 This guidance has been adapted from the previous, non-statutory guidance (FSR-G-213, Allele Frequency databases and reporting guidance for the DNA STR profiling in the UK) and reviewed by the Forensic Science Regulator's DNA Sub-Specialist Group.

18. Modification

18.1.1 This is the first issue of this document.

18.1.2 The PDF is the primary version of this document.

18.1.3 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 first three '#') indicate letters to describe the type of document and (b) (the second four '#') indicates a numerical code to identify the document. For example, this document is FSR-GUI-0012, and the 'GUI' indicates that it is a guidance document. Combined with the issue number this ensures that each document is uniquely identified.

18.1.4 If it is necessary to publish a modified version of a document (for example, 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 - ### - #### - #.

18.1.5 In the event of any discrepancy between the primary version and a modified version then the text of the primary version shall prevail.

19. Review

19.1.1 This document is subject to review by the Forensic Science Regulator at regular intervals.

19.1.2 The Forensic Science Regulator welcomes views on this guidance. Please send any comments to the address as set out at the following web page: www.gov.uk/government/organisations/forensic-science-regulator or send them to the following email address: FSREnquiries@forensicscienceregulator.gov.uk.

20. References

[1] Forensic Science Regulator, "Forensic Science Activities Statutory Code of Practice V2," [Online]. Available: https://www.gov.uk/government/publications/forensic-science-activities-statutory-code-of-practice-version-2. [Accessed 15 07 2025].

[2] K. Bowsher, “The code systems used within the Metropolitan Police Service (MPS) to formally record ethnicity. Metropolitan Police Authority briefing paper,” 07 2007. [Online]. Available: http://policeauthority.org/metropolitan/publications/briefings/2007/0703/index.html. [Accessed 15 07 2025].

[3] M. Bodner, I. Bastisch, J. M. Butler, R. Fimmers, P. Gill, L. Gusmao, N. Morling, C. Phillips, M. Prinz, P. M. Schneider and W. Parson, “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, 2016.

[4] Office for National Statistics, “UK Census 2021,” 2021. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021. [Accessed 14 07 2025].

[5] J. A. Hodgson, C. J. Mulligan, A. Al-Meeri and R. L. Raaum, “Early back to Africa migration into the Horn of Africa,” *PLOS Genetics,* vol. 10, no. 6, 2014.

[6] “Home Office data to support the implementation of the national DNA database DNA-17 profiling,” [Online]. Available: https://www.gov.uk/government/statistics/dna-population-data-to-support-the-implementation-of-national-dna-database-dna-17-profiling. [Accessed 15 07 2025].

[7] J.-A. Bright and M. Coble, Forensic DNA Profiling: A practical guide to assigning likelihood ratios, Boca Raton, FL: CRC Press, 2019.

[8] NRC, DNA Technology in Forensic Science, Washington DC: National Academy Press, 1992.

[9] D. Balding and R. Nicols, “DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands,” *Forensic Science International,* pp. 124-140, 1994.

[10] J. M. Curran and J. S. Buckleton, “An investigation into the performance methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations,” *Forensic Science International: Genetics,* vol. 5, no. 5, pp. 512-516, 2011.

[11] T. C. Matise, F. Chen, W. Chen, F. M. De La Vega, M. Hansen, C. He, F. C. L. Hyland, G. C. Kennedy, X. Kong, S. S. Murray, J. S. Ziegle, W. C. L. Stewart and S. Buyske, “A second-generation combined linkage physical map of the human genome,” *Genome Res.,* vol. 17, no. 12, pp. 1783-1786, 2007.

[12] A. Q. Nato, S. Buyske and T. C. Matise, “The Rutgers map: A third-generation combined linkage-physical map of the human genome,” in *Human Genetics Institute of New Jersey Second Research Day*, Piscataway, New Jersey, Rutgers University, 2018.

[13] J. A. Bright, J. M. Curran and J. S. Buckleton, “Relatedness calculations for linked loci incorporating subpopulation effects.,” *Forensic Science International: Genetics,* vol. 7, pp. 380-383, 2013.

[14] P. Gill, C. Phillips, C. McGovern, J. A. Bright and J. Buckleton, “An evaluation of potential allelic association between STRs vWA and D12S391: Implications in criminal casework and applications to short pedigrees,” *Forensic Science International: Genetics,* vol. 6, no. 4, pp. 477-486, 2012.

[15] K. L. O'Connor and A. O. Tillmar, “Effect of linkage between vWA D12S391 in kinship analysis,” *Forensic Science International: Genetics,* pp. 840-844, 2012.

[16] B. Budowle et al., “Population genetic analysis of the NGM STR loci,” *International Journal of Legal Medicine,* vol. 125, pp. 101-109, 2011.

[17] C. D. Steele, D. Syndercombe-Court and D. J. Balding, “Worldwide FST Estimates Relative to Five Continental-Scale Populations,” *Ann Hum Genetics,* vol. 78, no. 6, pp. 468-477, 2014.

[18] A. J. Hopwood et al., “Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts,” *Science and Justice,* vol. 52, no. 3, pp. 185-190, 2012.

[19] J. Buckleton, J. Curran, J. Goudet, D. Taylor, A. Thiery and B. S. Weir, “Population specific FST values for forensic STR markers: A worldwide survey,” *Forensic Science International: Genetics,* vol. 23, pp. 91-100, 2016.

[20] D. Syndercombe-Court, King's College London, “Personal communication,” 2025.

[21] J. A. Bright, J. M. Curran, A. J. P.-S. R. Hopwood and J. S. Buckleton, “Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts II,” *Science and Justice,* vol. 53, no. 3, p. 371, 2013.

[22] J. Thomson, Eurofins Forensic Services, “Personal communication,” 2025.

[23] A. W. Cotterman, “A calculus for statistico-genetics,” Ohio State University, Columbus, Ohio, 1940.

[24] P. Gill, Forensic Practitioner's Guide to the Interpretation of Complex DNA Profiles, Academic Press, 2020.

21. Abbreviations

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

22. 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.
Allelic Drop-Out	Allele(s) missing from a DNA profile, so that it is partially represented.
Autosomal	Relating to any chromosome that is not sex-determining.
Chromosome	A threadlike structure of nucleic acids in the cell that carries genetic (hereditary) information in the form of genes.
DNA-17 System	Short tandem repeat (STR) multiplex system (kit) with 16 autosomal STR loci (plus 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.
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.

Published by:

The Forensic Science Regulator

23 Stephenson Street

Birmingham

B2 4BJ.

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