Forensic Science International: Genetics 30 (2017) 160–163

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Editorial

Revised guidelines for the publication of genetic population data

Since 2007, when the journal was launched, the number of submissions of manuscripts reporting population genetic data to FSI: Genetics has continuously increased. This type of data is very welcome, considering the importance of having accurate estimates of population specific allele/haplotype frequencies of genetic markers relevant to the forensic field. For this reason, in 2010, a section on Forensic Population Genetics was introduced, including three types of formats: Original papers, Short communications and Letters to the Editor [1].

Considering the large amount of population data that is being generated for many populations all over the world, as well as technical developments in the forensic genetic field, we are now updating the minimal requirements that should be observed before a paper is submitted to the journal. This includes the need for (i) a quality check of autosomal STR population data, (ii) new requirements concerning population data generated by massively parallel sequencing (MPS) technologies, (iii) an update on the minimum number of samples and markers required for the publication of autosomal, Y-chromosomal and X-chromosomal population data produced by conventional methodologies, including new requirements concerning SNPs or InDels.

The minimum requirements described in this and previous editorials are not meant to be applied straightforwardly to all population genetic studies, but only to those submissions where content is limited to the description of allele/haplotype frequencies in a population, together with forensically relevant statistical parameters and results of population differentiation analyses. Manuscripts that, apart from a population allele/haplotype frequency database, also include other forensically relevant information (e.g. new methods, mutation rates, recombination rates, relevant case reports), making the work worthy of publication as an original paper in a scientific journal, do not necessarily have to comply strictly with the minimum requirements concerning the number of samples and markers. Also, with these guidelines, we do not intend to exclude studies on small but important populations or ethnic groups. Studies on small populations are indeed very welcome, since these are usually less studied. However, if the total number of samples is far from the minimum number required for a certain set of markers, we expect the authors to try to increase the number of samples (for e.g., by joining data from more than one population) or to use the same samples to study different types of markers.

1. Quality control of population DNA databases

To improve the quality of the population data that are submitted to the journal, in 2010 [1], the editors of the journal and the executive board of the International Society for Forensic Genetics (ISFG) have invited EMPOP (http://www.empop.online) and YHRD (Y Chromosome Haplotype Reference Database; https://yhrd.org) to logistically organize and perform quality control (QC) of mtDNA sequences and Y-STR/Y-SNP data, respectively. These strategies have significantly improved overall data quality in forensic genetics and reduced the amount of errors that would have otherwise been mistakenly introduced in scientific publications [2,3]. Therefore, before manuscripts are put forward to the editors for review, mtDNA and Y-STR/Y-SNP data must be submitted to the above-mentioned databases. The necessary steps for submission data to EMPOP have been described in Carracedo et al. [4] and can be found on EMPOP under 'Contribute' (http://empop.online/contribute). Up-to-date conditions and necessary steps to submit population data to the YHRD can be found in https://yhrd.org/pages/help/contribute.

To also improve the quality of the data generated from autosomal STRs, the ISFG executive board and the editors of FSI: Genetics have now invited STRidER (http://strider.online), a publicly available, centrally curated online allele frequency database and quality control platform for autosomal STRs [5], to logistically organize and perform quality control before autosomal STR manuscripts are put forward for review. Upon successful QC, STRidER accession numbers will be assigned to the submitted population data that serve as indicators of successful QC for the editors and reviewers. The necessary steps for submission of autosomal STR genotypes to STRidER are outlined below.

1.1. Step 1

The minimum requirements for population datasets are 15 autosomal STR loci (in compliance with the present requirements) typed from 500 samples [5]. An STR data file, for which template can be downloaded from the STRidER website, should be prepared. It is imperative that genotypes are compiled individually and unshuffled using a unique identifier for each genotype. The submitted population data need to comply with the format rules and allele nomenclature criteria as described on the STRidER

http://dx.doi.org/10.1016/j.fsigen.2017.06.007
1872-4973/

website. Only complete genotypes are accepted, suspected allelic dropouts must be reported. The data file also specifies the detailed geographic and linguistic/ethnic background of the population, the contact author's name and email address. Further information on the dataset and data generation might be necessary for evaluation and should be submitted in an accompanying file. For details on file preparation see the STRidER website.

1.2. Step 2

The files should be submitted to STRidER by email. A suite of software tools has been developed to scrutinize STR population data and thus increase the quality of datasets to ensure reliable allele frequency estimates as outlined in Bodner et al. [5].

1.3. Step 3

During data evaluation, communication with respect to individual genotypes may follow. Datasets that passed QC will receive a STRidER accession number together with allele frequencies and forensic/population genetic parameters calculated from the datasets.

1.4. Step 4

Data that successfully passed QC will be uploaded onto the STRidER database with the next release. To ensure data protection, only allele frequencies, no individual genotypes, will be made available. The fact that the data quality is scrutinized by STRidER does not relieve the authors from carefully preparing, inspecting and reviewing their entire dataset also beyond the errors announced. Quality control by STRidER should act as a final check on the data.

Highlights: Papers including population data on autosomal STRs will be only considered by the editors of FSI: Genetics for review, after successful quality control to the STRidER database. The quality and reliability of the data will only increase with this additional check.

2. New requirements for population data generated by massively parallel sequencing (MPS) technologies

Reports on population data for forensic marker sets generated by MPS technologies (including, e.g., autosomal, Y-chromosomal and/or X-chromosomal STRs; SNPs/InDels identity, ancestry or/and phenotypic panels; mtDNA Control Region or whole genome data; RNA markers) have recently been submitted for publication in FSI: Genetics. Considering the novelty of the techniques and the lack of frequency data for most populations worldwide, specific requirements are described below in terms of minimum number of samples and quality of MPS population data to be submitted to FSI: Genetics.

In submissions of MPS population data to the journal, the authors should observe the following requirements (Table 1):

1 Results should be submitted for at least 50 unrelated individuals;
2 The full list of sequence strings ('FASTA format') must be submitted as Supplementary information. Following the recommendations of the DNA Commission of the ISFG on MPS of forensic STRs, sequence strings should include flanking sequences and indicate the genome coordinates of the sequence read start and end points, as well as the name of the SNPs (see Parson et al. [6] for nomenclature details);
3 Only full genotype profiles must be submitted, not allowing for dropout;
4 The quality control (QC) procedures followed by the authors must be described in detail, as Supplementary information.

Highlights: At least 50 full genotype profiles are required for the publication of population data generated by MPS. The recommendations of the DNA Commission of the International Society for Forensic

Table 1
Summary of the minimum requirements for the submission of population data to FSI: Genetics.

		Minimum requirements			
		Data[a]	No. samples[c]	Submission to:	Comments:
Data type	Au-STRs	15 STRs	500	STRidER	Full genotypes must be submitted to STRidER for quality control purposes. They will however not be presented on STRidER[d]
	X-STRs	12 STRs	500 (males)	–	QC procedures indicated in Supplemental Materials Haplotypes listed in Supplemental Materials
	Y-STRs	23 STRs	400	YHRD	Haplotypes listed in Supplemental Materials
	Au-SNPs/ InDels	30 SNPs/InDels	500	–	QC procedures indicated in Supplemental Materials Full genotypes are welcome (not mandatory)
	X-SNPs/Indels	20 SNPs/InDels	500 (males)	–	QC procedures indicated in Supplemental Materials Haplotypes listed in Supplemental Materials
	Y-SNPs/Indels	[b]	300	YHRD	At least 23 Y-STR typed in the same 300 samples Haplotypes listed in Supplemental Materials
	mtDNA_Sanger	Full CR	200	EMPOP	Sequences listed in Supplemental Materials
	mtDNA_Sanger	Full mtDNA molecule	100	EMPOP	Sequences listed in Supplemental Materials
	mtDNA-SNPs	SNPs in coding region[b]	200	EMPOP	The full CR must be sequenced in the same 200 samples
	MPS	MPS–strings	50	–	QC procedures indicated in Supplemental Materials

[a] Only samples with full genotype/haplotype (including silent alleles) or sequence profiles are accepted (no missing data or locus/allele droupouts due to a poor quality of the sample should be present in the dataset).
[b] For mitochondrial and Y-chromosomal biallelic markers (Y-SNPs/InDels), a phylogenetic approach can be used in haplogroups' classification and, in this case, not all samples need to be typed for the same group of markers.
[c] In studies on small populations or ethnic groups, it is sometimes impossible to selected large samples of non-closely related individuals. Minimum requirements can be reached by joining data from several population or ethnic groups from the same geographic region, or by studying different type of markers in the same population sample.
[d] For upload, the STR allelic pair is kept intact, but single locus genotypes are shuffled between individuals. This keeps the intra-locus genotype configuration intact, but disrupts the inter-locus genotype association to protect the privacy of sample donors.