Forensic Science International: Genetics 7 (2013) 217–220

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

Editorial

New guidelines for the publication of genetic population data

In 2000 a new policy concerning the publication of population genetic data was set up in *Forensic Science International* [1] with the introduction of a new section entitled "Announcement of population data". Subsequently in 2010 [2] a new section on "Forensic Population Genetics" was introduced, and recommendations were redefined.

FSI: Genetics is one of the few journals still considering population genetic data for publication and we strongly believe that this policy has contributed to the dissemination of common standards in the field all over the world and also to motivate labs and people to embark in research in the area of forensic genetics. For this reason it is our intention to continue with this policy, and recently an associate editor exclusively devoted to this topic was appointed to the journal.

Despite having defined a more detailed procedure for acceptance, our journal is still receiving a massive number of submissions of varying quality in this area. Therefore it has become necessary to raise the threshold regarding the acceptance of this type of publication to ensure a high standard of published data. In addition we want to improve the submission, reviewing and publication procedures, and to correct some aspects that we have detected such as the obligation to meet ethical standards in the collection of samples including informed consent and approval by ethical committees.

For this reason, we have decided to publish new guidelines for the publication of population genetic data in the journal.

1. Formats of submission

Manuscripts containing data on *Forensic Population Genetics* can be submitted using three types of formats:

Original paper: In this section full length papers on relevant population genetics issues of forensic interest will be considered for publication. The data should be original, the population genetic analysis must be of the highest quality and the data should have forensic relevance beyond the scope of simply reporting allele or haplotype frequencies.

Short communication: Understanding that both the quality of population data and the relevance of results are crucial, short communications should be submitted in table format. Population data are required to be downloaded as supplementary files the same as for other publication formats (see instructions to authors).

Letter to the editor: If the relevance of the data is not sufficient for an original paper or a short communication, but still worthy of an announcement, the editors can invite authors to submit a letter to the editor. In this case the manuscript must be written in the form of a short letter to the editor summarizing the relevant information while the frequency data must be provided as an electronic supplement, e.g. a spreadsheet table, for online publication in the electronic repository of the journal.

2. Minimum requirements for submission

A minimum of 17 loci are required for autosomal and Y chromosomal STRs, and 12 for X chromosomal STRs. Data from Y chromosomal SNPs should be combined with STR information for the same samples. A minimum of 500 samples are required for autosomal and X chromosomal STRs, and for mtDNA haplotypes/sequences. Because only males can be analyzed, a minimum of 250 samples are required when reporting Y chromosome results. Authors are encouraged to combine population data and not to split in different papers results from (i) a single population using different sets of markers or; (ii) data for the same set of markers in different population groups from a single country or region. If the population is rare but of forensic or anthropological interest and therefore the number of unrelated individuals required is difficult to obtain, data of different markers or populations should be combined to enhance the value of the paper and reach the minimum requirements. For example, if authors want to publish results from a population sample that just includes 125 unrelated individuals: (i) data should include at least 4 different type of markers, namely autosomal STRs, SNPs or Indels, mtDNA, Y chromosomal STRs and SNPs, and X chromosomal STRs; or (ii) data should be presented for two or more populations in order to achieve the minimum required number of samples indicated above. The same combination effort should be made when genetic information for only a few additional loci is being added for samples that have been previously typed. For example, if authors want to publish the results of 5 autosomal STRs in a population sample of 500 individuals that have been previously typed for other 17 STRs: (i) data should also include other types of markers, namely, autosomal SNPs or InDels, mtDNA, Y chromosomal STRs and SNPs, and X chromosomal STRs; (ii) at least 500 new samples from the same population should be typed for the full set of 22 STRs and the results for the previously published markers should be updated; or (iii) at least 500 samples from another population should be typed for the same markers.

Collaborative efforts in order to increase the number of samples and/or populations representing a country, a broad geographic region or continent are strongly encouraged.

1872-4973 © 2013 Published by Elsevier Ireland Ltd. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.fsigen.2013.01.001

3. Information requested

All manuscripts containing Forensic Population Genetic data should always contain information on the description of the population, relevant ethical requirements and quality control as follows:

3.1. Description of the population

With an appropriate length according to the type of paper, a detailed description of the population is essential as well as a description of the interest of that population for population genetics and forensic purposes. Previous population genetic studies should be reported as well as the geographic location, ethnicity, method of sampling, and characteristics of the population.

The description of the population should be documented and supported by reference papers from the scientific literature or well recognized books.

3.2. Ethical requirements

Informed consent and/or specific approval of a recognized ethical committee are required and must be stated in the text. For STRs, the inclusion of the whole genotyping data will not be required due to ethical constraints for the publication of such types of data in some countries, but the authors are requested to provide the anonymized data to interested researchers upon request if not prohibited by ethical constraints. The authors should state in the text that they understand and accept the requirements requested in this editorial.

Any paper not completely fulfilling these ethical requirements will be directly rejected by the editors without sending the manuscript out for review.

3.3. Quality control

For STRs and SNPs the quality of the data must be guaranteed. The QC procedures followed by the authors must be specified. Certification of approval by proficiency testing programs is ideal and encouraged. Authors must state that they have strictly followed ISFG recommendations on the analysis of the DNA polymorphisms used [3], signifying the use of recommended nomenclature and guidelines regarding QC and statistical issues.

Whenever possible, a comparative analysis between the concerned population and neighboring or historically related populations is required. For autosomal and X chromosomal data, comparative population analysis is the only quality control measure of the data, establishing whether or not the results are in accordance with available data for populations in the same geographic region and/or with a shared history/ancestry.

4. mtDNA and Y chromosome polymorphisms

MtDNA and Y chromosome data need special requirements. In this case the importance of high quality population DNA databases justifies a strict publication policy as follows.

4.1. mtDNA

The executive board of the International Society for Forensic Genetics (ISFG) and the editors of *FSI: Genetics* have invited EMPOP[1] to logistically organize and perform quality control (QC) of mtDNA sequences in the course of manuscript preparations for the journal. Before mtDNA papers are put forward to the editors for review, the authors are requested to submit the data to EMPOP. After evaluation, the authors will be contacted by EMPOP, and exemplar raw data may be requested for quality checks. Only original raw data, not recently repeated experiments, are accepted for this purpose. Upon successful QC, the mtDNA sequences will be assigned EMPOP accession numbers that serve as indicators of successful QC for the editors and reviewers. The necessary steps for submission of mtDNA sequences to EMPOP are outlined below.

Important requirement: The presentation of partial control region sequences, such as those of the hypervariable segments (HVS) I and II only, is no any longer state of the art [4]. Only full control region sequences spanning from nucleotide positions 16024-576 with respect to the rCRS [5] will be considered for publication in *FSI: Genetics*. In compliance with earlier recommendations [6,7] the minimum requirement for acceptable data is full double-stranded sequence coverage.

Step 1

The submitted mtDNA population data need to comply with the format indicated in the CONTRIBUTE section of EMPOP (www.empop.org). The following information is required:

- Contact details of the corresponding author (corresponding to the dataset)
- Presentation of individual mtDNA haplotypes annotated relative to the rCRS identified by unambiguous sample names and the corresponding reading frames. Note that the full control region (16024-576) is the minimum analysis requirement; additional coding region information is welcome.
- Haplogroup status of the haplotypes with reference to source (e.g. Phylotree (www.phylotree.org) including build).
- Geographic and linguistic/ethnic information per individual haplotype. Geographic information includes "continent – UN region – country – province – city" and latitude/longitude. A scheme of available geographic/linguistic categories is provided via EMPOP. If relevant, additional information such as ethnic group should be specified using tag words – e.g., "Europe – Western Europe – Austria – Tyrol – Innsbruck (47.265, 11.395)" or "Africa – Middle Africa – Angola (−11,2026920, 17,8738870)" for geographic information, and "Eurasian – Indo-European – Germanic" or "Sub-Saharan – Khoe-San" with the additional tag "Ikomgau" describing an African tribe speaking Khoe-San, respectively. If relevant, please provide additional geographic/ethnical information in additional corresponence or maps.
- Information on sequencing chemistry and sequencing instrument.
- Information on the alignment/sequencing analysis software.
- For templates containing all relevant information see the CONTRIBUTE section on EMPOP.org.

Step 2

Submit your file(s) to EMPOP using the Email address "datasubmission@empop.org". The data will be quality-checked for format, plausibility, clerical errors, sequence range violation, reference errors, indels designation, and phantom mutations using in-house software programs and NETWORK, which is also available through the EMPOP website. Note that tools for sequence data evaluation are continuously added to the EMPOP website to help the authors scrutinize their data before submission.

Step 3

The submission of individual raw data may be necessary. Only original raw data are accepted. Once your data have passed QC you will receive the (corrected) dataset with respective EMPOP numbers. Please provide these EMPOP numbers together with your manuscript to the editor for initiating the review process.

[1] European DNA Profiling (EDNAP) Group's mitochondrial DNA population database project; www.empop.org.

Step 4

Upon successful quality control your data will be uploaded onto EMPOP with the next EMPOP release. The data are not downloadable and they will not be made available to third parties without explicit agreement of the contributing author(s).

Important: The fact that the data quality is scrutinized by EMPOP should not relieve the authors from carefully inspecting their dataset. This data review should not be limited to the announced errors of the EMPOP QC but requests the author(s) to review the whole dataset. Additional errors in a posterior submissions will represent a serious drawback for the future acceptance of the paper. Quality control by EMPOP should act as a final check on the data. The EMPOP staff will be happy to provide help and guidance for the preparation of the population data.

4.2. YSTRs and YSNPs

In the same way as for mtDNA submissions, the executive board of the ISFG and the editors of *FSI: Genetics* have invited YHRD[2] to logistically organize and perform quality control (QC) of YSTR/YSNP data in the course of manuscript preparations for the journal. Before YSTR/YSNP papers are put forward to the editors for review, the authors are required to submit the data to YHRD. After evaluation, the authors will be contacted by YHRD and the YSTR/YSNP data will be assigned to YHRD accession numbers that serve as the indication of successful QC for the editors and reviewers. The necessary steps for submission of YSTR/YSNP data to YHRD are outlined below. Please note, that YHRD can only accept, evaluate and upload Y-STR sets previously validated for forensic purposes.

Step 1

Prepare your YSTR and YSNP data as explained at the website www.yhrd.org/Contribute and in the YHRD manual (www.yhrd.org/downloads/manual.pdf). The YHRD input file is a standard spreadsheet file. The first two columns specify a sample identification number and the origin of the samples. For the latter we request a ternary identifier in the form "region, country [ethnic group]" – e.g., "Berlin, Germany [German]". The geographic background of the samples should be further detailed in an accompanying text or by maps. The other columns list the common YSTR loci and the panel of typed YSNPs, specified by "+" for the derived state and "−" for the ancestral state at a given locus. The last two columns contain the haplogroup designation according to the most updated nomenclature (e.g. [8]), and the final branch marker used for haplogroup assignment, e.g. Q1a3a and M3. Synonymous marker names are allowed. If the haplogroup is unknown, then use a "?" symbol. The YHRD software verifies the correct haplogroup assignment based on the ancestral resp. derived states (encoded "−/+") at the respective YSNP positions

The file should list the individual haplotypes with a single haplotype per line using unique identification numbers. Identical haplotypes should be listed separately. Please note the following format rules

- Alleles at duplicated loci are separated by a comma (e.g., "11,14").
- Alleles containing incomplete repeat motifs are designated by a dot (e.g., 11.2).
- Confirmed "null" alleles are indicated by a "0".

Note that allelic drop-outs at certain YSTR loci may occur due to either molecular mechanisms (e.g., chromosomal rearrangements or deletions, primer site mutations, etc.) or technical problems (e.g., low amounts of DNA template, degradation, etc.). As used here, the term "null allele" refers to allele loss due to molecular mechanisms. These should be reported.

Step 2

The file should be sent as an e-mail attachment to the following addresses "lutz.roewer@charite.de" and "sascha.willuweit@charite.de". The text of the e-mail should contain the title of the study and an author name with an e-mail address for contact. The data will be quality checked for format, clerical errors, allelic range violation using in-house software (e.g. NETWORK, AMOVA).

Step 3

Communication may follow with respect to individual haplotypes/haplogroups. Once your data has passed QC you will receive the yhrd-file of your data listing YHRD accession numbers for all your samples. Please provide these accession numbers together with your manuscript to the editor of the journal.

Step 4

Upon successful quality control your data will be uploaded onto the YHRD with the next release. The data are not downloadable and they will not be made available to third parties without explicit agreement of the contributing author(s).

5. SNPs

Population data of SNPs will be considered but only for SNP sets previously validated for forensic purposes. Check invalid Y-SNPs before your start your analyses at http://www.yhrd.org/Analyse/BYSNP. Invalid SNPs (e.g. the still popular marker P25 which is recurrent within R1b) need to be replaced by upstream and downstream markers which unequivocally resolve the respective phylogenetic branch.

6. Table formats

Genotyping results should follow a standard spreadsheet table format and should be submitted as an electronic supplement file to be published only in the electronic repository of the journal.

Figures and tables will be published only in the electronic repository of the journal in the case of letters to the editor, as well as in the case of short communications.

7. Additional requirement

Authors must state in the paper that they have strictly followed the requirements of this guideline and the ISFG recommendations and also that the study was approved by an ethical committee (please give the name of the ethical committee and the host institutions followed by the approval code or date) and/or written informed consent from all the participants.

References

[1] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, Forensic Sci. Int. 110 (2000) 3–5.
[2] A. Carracedo, J.M. Butler, L. Gusmão, W. Parson, L. Roewer, P.M. Schneider, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147.
[3] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci. Int. 165 (2007) 238–243.
[4] W. Parson, H.J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (2007) 13–19.
[5] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.
[6] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.
[7] G. Tully, W. Bär, B. Brinkmann, A. Carracedo, P. Gill, N. Morling, W. Parson, P. Schneider, Considerations by the European DNA Profiling (EDNAP) group on the

[2] Y Chromosome Haplotype Reference Database; www.yhrd.org.

working practices, nomenclature and interpretation of mitochondrial DNA profiles, Forensic Sci. Int. 124 (2001) 83–91.
[8] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res. 18 (2008) 830–838.

Ángel Carracedo*
Institute of Forensic Science, Genomic Medicine Group, University of Santiago de Compostela, Spain

John M. Butler
National Institute of Standards and Technology, Gaithersburg, MD, USA

Leonor Gusmão[a,b]
[a]*Institute of Pathology and Molecular Immunology of the University of Porto, Portugal*
[b]*University of the State of Rio de Janeiro (UERJ), Brazil*

Adrian Linacre
School of Biological Sciences, Flinders University, Adelaide, Australia

Walther Parson[a,b]
[a]*Institute of Legal Medicine, Innsbruck Medical University, Austria*
[b]*Penn State Eberly College of Science, University Park, PA, USA*

Lutz Roewer
Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin, Berlin, Germany

Peter M. Schneider
Institute of Legal Medicine, Medical Faculty, University of Cologne, Germany

*Corresponding author
E-mail address: angel.carracedo@usc.es (Á. Carracedo)