

Accepted Manuscript

Sequence data management for scientific purposes

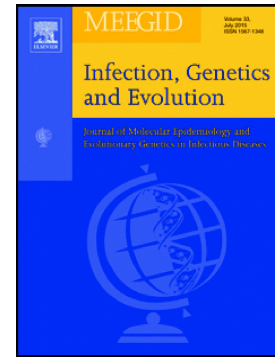
Gabriele Margos, Volker Fingerle, Andreas Sing, Keith Jolley

PII: S1567-1348(17)30228-9
DOI: doi: [10.1016/j.meegid.2017.06.030](https://doi.org/10.1016/j.meegid.2017.06.030)
Reference: MEEGID 3192

To appear in: *Infection, Genetics and Evolution*

Received date: 4 April 2017
Revised date: 6 June 2017
Accepted date: 30 June 2017

Please cite this article as: Gabriele Margos, Volker Fingerle, Andreas Sing, Keith Jolley, Sequence data management for scientific purposes, *Infection, Genetics and Evolution* (2017), doi: [10.1016/j.meegid.2017.06.030](https://doi.org/10.1016/j.meegid.2017.06.030)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Sequence data management for scientific purposes

Gabriele Margos^a, Volker Fingerle^a, Andreas Sing^a, Keith Jolley^b

^aGerman National Reference Centre for *Borrelia*, Bavarian Health and Food Safety Authority, Veterinärstr. 2, 85764 Oberschleissheim, Germany; emails: volker.fingerle@lgl.bayern.de; andreas.sing@lgl.bayern.de

^bDepartment of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK; email: keith.jolley@zoo.ox.ac.uk

corresponding author: Gabriele Margos, Curator of the *Borrelia* MLST database (<https://pubmlst.org/borrelia/>), email: gabriele.margos@lgl.bayern.de

Dear Editor,

In April 2016 an interesting and potentially important article entitled “Imbalanced presence of *Borrelia burgdorferi* s.l. multilocus sequence types in clinical manifestations of Lyme borreliosis” was published in *Infection, Genetics and Evolution* 42 (2016), 66-76. We are writing this letter to make you and readers of the journal aware that the data on which the conclusions of this paper are based are not publicly available in any sequence database.

In this article, the authors characterize a number of *Borrelia* strains by multilocus sequence typing (MLST), a technique now widely used for *Borrelia* spp. Unfortunately, however, although the article apparently gives details of MLST sequence type from the MLST database <https://pubmlst.org/borrelia/>, the sequence type (ST) numbers given for novel STs (i.e. ST numbers of >1,000 in Table 4) do not correspond to numbers listed in that database. Isolate data for these and other newly sequenced strains from that publication are also not available in the database. The authors state that “Each unique allele type of the housekeeping genes received a number” which resembles the method that is normally used for the MLST database, making it impossible for a reader to know whether or not the numbers were given authoritatively via the

MLST database or not. Moreover, the authors fail to provide corresponding GenBank accession numbers of the sequences for the strains they worked with.

Although the authors announce that they have “newly described” 94 new MLST types, there are in fact no descriptions in this paper that other researchers can use, either for typing purposes or comparative analysis. These data are essential for any other researcher to be able to verify and build on the work that is described in the article. It is stated in the article that “Fasta files of newly described sequence types are available on request”. However, when these were requested, a file containing ALL the data used in the study was obtained: newly sequenced strains corresponding to known STs, newly described STs and, intertwined strains that the authors had downloaded from the MLST database. This made it difficult to decide what was new from the study and what already existed in the database. Isolate data of both novel STs and newly sequenced strains from this study, with existing ST numbers, would be a valuable addition to the database for other researchers to use as this allows inferences on the geographic distribution of lineages.

An illustration of the importance of this omission is that in their analysis of *Borrelia* strains the authors have used not only their own sequence data from 440 new isolates, but also existing published sequences from a further 335 isolates, which they have taken from the literature.

Public databases exist for data sharing amongst the scientific community. Sequence typing of microbial pathogens cannot be effective unless such sharing is widespread. We consider it – even for scientists with very little time on their hands - unacceptable to avoid the task of uploading sequence data such as this to the appropriate public resource. This should have been done at the time of submission or at least before publication. Submitting sequence data subsequent to publication is not ideal, but it would at least allow researchers to search in the MLST database via allele or sequence type numbers. Such searching would be facilitated when an addendum or erratum would be added to the article so that the proper numbers could be found in the journal. Without a published erratum or addendum, it would be more difficult to associate those data with the conclusions published in the *Infection, Genetics and Evolution* article, although a subsequent submission to the database may allow interested researchers to theoretically extract the appropriate data and the strains in question from the database via strain designation.

We have several times e-mailed the corresponding author of the article as well as the last named author to ask for subsequent submission of the sequence data to the MLST database. Although it was promised that the data would be submitted this has still not been done. Even upon intervening of the editor, a list of sequences and isolate data for the newly sequenced strains from this study (i.e. sequences for new STs and isolate data for new and known STs) has not been provided.

Now we are taking the step to writing this letter to alert our scientific colleagues that they should not search the MLST database for these data in vain.

Yours sincerely,

Gabriele Margos, Curator of the *Borrelia* MLST database (<https://pubmlst.org/borrelia/>), Bavarian Health and Food Safety Authority

Volker Fingerle, Head of the German National Reference Centre for *Borrelia*, Bavarian Health and Food Safety Authority

Andreas Sing, German National Reference Centre for *Borrelia*, Head of Department GE2, Bavarian Health and Food Safety Authority

Keith Jolley, University of Oxford, Developer and administrator of PubMLST