

## Review for Contextualising samples: Supporting reference genomes for European biodiversity through sample and associated metadata collection

### Summary

This valuable work presents the ERGA initiative, and the results of its pilot project, which ultimately aim at obtaining reference genomes for all eukaryotic species found in Europe. The article explains the rationale behind the submission and processing pipeline offered to research partners by their Sample Processing Committee. The report is detailed, explicitly integrates the FAIR and CARE principles, and is clearly among the most comprehensive initiatives undertaken worldwide.

I have a few comments and questions that I detail below.

### General comments

1) The authors show on Figure 1 the consistency between the research community interest and the clades actually sampled for the ERGA pilot project. I was wondering whether more information could be provided through similar figures. What will be the scientific use for these genomes? Conservation genomics? Fundamental research? Agronomy/species of economic interest? Orphan crops? What is the IUCN threat status of the species considered? Depending on the species and its usage, are there plans to obtain population-level data? More information may reinforce the interest in the initiative.

2) The metadata collection process seems very thorough, but gives the impression of a rather bottom-up, one-way interaction, where users provide a lot of information and put effort into collecting data, while feedback, support and assistance from the Consortium do not appear immediately obvious. Figure 2's design is interesting from this perspective, with the icon of a single researcher repeatedly facing a myriad of protocols and red tape. I do not deny the importance of collecting this information, but wonder whether the whole process could be introduced in a "gentler" way. Is there any risk of bias towards samples provided by larger laboratories which already have expertise in handling genome projects? Figure 4 provides a list of potential issues and solutions, but the concrete way through which solutions are offered to the contributing researchers is not always clear. Are there plans to offer workshops and freely available tutorials? How can the initiative include scientists who have limited experience in genomics, but want to provide samples and learn about analyses? Related to point 1), obtaining a reference genome is of interest to scientists working in sometimes very different fields, each with their own culture and research habits. It may be interesting to provide some examples on how the initiative has tackled such a diversity of aims.

3) The authors mention several other continent-wide (or smaller) initiatives. Are there plans to exchange with other continent-wide initiatives to avoid redundancy for species spanning more than one continent? What would be the process?

### Minor comments

L91: What does ELSI stand for?

L159: For this pilot study, I understand the need to favour feasibility, but are there plans to address the other criteria in the future? The authors mention the need to avoid a biased representation of species, but more information on how this can be achieved might be interesting to their readers.

L173: ITC? Inclusiveness Target Countries? I do not think the acronym is introduced before.

L202 From a concrete perspective, who has the responsibility for providing and controlling such information about Indigenous knowledge and practices? The local researcher, a committee within ERGA? Is there any direct involvement of indigenous communities? Any check from the Consortium? How is the SSP integrated with the ELSI committee? In general, it might be useful to introduce how the SSP may interact with other committees within ERGA.

It might be interesting to provide a few more examples of how the SSP experts assisted users with permits handling (L293), sample processing (L321), or the adoption of guidelines (L381) ? L411 is an interesting example of how the initiative supports the whole community.

Figure 2: Step D panels are hard to read.

Several acronyms are not introduced (COPO, JEDI, ELSI etc.). Maybe consider adding a short glossary and links to websites.