

Reply to reviewers' comments

We thank the team of PCI for reviewing our manuscript and a very constructive process. We incorporated the helpful comments of the reviewers and hope that you recommend our article with these changes.

Reviewer 1

We thank the reviewer for taking the time to carefully review our manuscript and the helpful 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.

Thank you for the suggestion, we agree that providing more information about the purpose for the generation of the reference genomes included in the pilot provides a better context for their utility. We have now included in Figure 1 information collected by ERGA's Data Analysis committee (DAC) from sample ambassadors in the framework of the flagship paper of the pilot project (Mc Cartney et al. 2023), and believe that the purpose and planned future analyses anchored on the new reference genomes are now clearer.

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.

Thank you very much for making us aware that this section lacked indication of support. It is true that the sample provider is asked to provide a substantial amount of information in this step, however, the SSP is accompanying sample providers here very thoroughly. We provide onboarding meetings, video tutorials and hands on help arranged over email. We have added this information to section "ERGA Manifest for Metadata Collection and Brokering"

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?

This is an excellent point and a major challenge to large scale initiatives. ERGA is establishing as the European Node of the Earth Bio Genome Project (EBP) and hence agrees to exchange information

with other EBP nodes and over EBP-aligned channels. Mainly, ERGA is displaying species that are planned to go into sequencing and those that have entered sequencing over the platform Genomes on a Tree (<https://goat.genomehubs.org/projects/ERGA>) that all EBP-nodes are encouraged to use. For the current ERGA project BGE, sample managers investigate GoaT and in case of overlap detection reach out to the corresponding genome initiative and seek for collaboration. We have added this information to the section “Towards a balanced and strategic prioritisation of species”.

Minor comments

L91: What does ELSI stand for?

Ethical, Legal, and Social Issues, we have added the explanation here and to a glossary

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. We provide a perspective to the future under “Towards a balanced and strategic prioritisation of species” and have extended our explanations therein.

L173: ITC? Inclusiveness Target Countries? I do not think the acronym is introduced before. We have added the full-length wording here and to the new glossary.

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.

We ask from the Genome team lead/the sample providers to provide such information, we have added this information to the paragraph in question. Over the manifest, sample managers can access the sample providers' reply to the question if traditional knowledge labels shall be recognised with the genome. If yes, the information has to be uploaded together with the manifest and is then automatically connected to the genome data. ERGA has deliberately decided to ask the sample provider to adhere to ERGA's sampling code of conduct and take over all legal and ethical responsibility for the samples that are provided to the ERGA infrastructure.

ERGA will not store any Traditional or Indigenous Knowledge associated with the samples or genomics resources we collect or generate. This information will always lie with any Indigenous partners. By including the Traditional Knowledge and Biocultural Labels system articulated in the paper, Indigenous partners can assign an appropriate Label (or Labels) to make visible that there is Traditional Knowledge or Biocultural significance associated with the samples and data collected. They can also provide important provenance information about the Indigenous partner.

In the metadata schema, the Labels ontology expects a PUID that has been generated by the Indigenous partners through the Local Contexts Hub. This PUID then travels within the metadata into public digital repositories (as they are extra-legal human and machine-readable digital tags). This mechanism simply gives any researcher that accesses and uses the data in the future, the opportunity to see the Label, and read the Indigenous rights and interests within the data, and empowers them to act in accordance with what is disclosed in the Label. For example, if the Biocultural Label for “Provenance” is selected by the Indigenous partner and associated with genomics data through the metadata entry, this will travel into digital repositories and show future researchers that if the data is reused there is an expectation that the Indigenous partner should be named and associated with the data in all future uses e.g., fair attribution in peer reviewed publications etc.

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.

[We extended our examples at the requested places.](#)

Figure 2: Step D panels are hard to read.

[We have revised the figure.](#)

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

[We added a glossary](#)

Reviewer 2

We thank the reviewer for taking the time to carefully review our manuscript and the helpful comments.

Title: The title clearly reflects the content of the article. However, I suggest replacement of “for” in the title with “of”

We have modified the title according to the suggestion.

Abstract : The abstract is concise and captures the major points in the article.

55 SSP serves as the sample provider's entry point... I suggest providers' reason is that SSP ought to receive several samples; not one sample

We have revised the sentence structure.

I. The Sampling and Sample Processing committee of ERGA

Introduction clearly demonstrates the motivation for the study.

The introduction builds on relevant recent and past reference research.

76-77 Delete the phrase “one of which is the Sampling and Sample Processing committee (SSP)” It seems to appear slightly early. It can come at the beginning sentence of the next paragraph as follows:

88 The Sampling and Sample Processing committee (SSP) is a working group of volunteer expert ERGA members tasked with developing guidelines

83 to support sampling and sample processing.

We have integrated the suggested change.

Materials and Methods: This section contains sufficient information that can be replicated in similar researches.

Results: Data presented in the article are correct and unambiguously presented.

174Widening countries with 44% and 50% of... I suggest ... Widening countries with 44 and 50 % of and

175 However, only 36% or 42% of the... However, only 36 or 42 % of the...

We have changed the text in question.

The tables and figures (charts) are clear and self-explanatory. However, the texts in Figure 3 could be made more legible for easier reading.

Figure 3 has been revised accordingly.

IV. Sample provision: connecting genome teams with

322 sequencing centres

324 arising from three main categories: biological, logistic, and legal issues. I rather think it should be: 324 arising from **four** main categories: biological, logistic, administrative/policy and legal issues.

We have changed the text and figure in question.

364 Future taxon-specific best-practice guidelines

The approach of having different sampling procedures for different taxa is very commendable as it would eliminates complications arising from structural and functional variations between the taxa.

ERGA's SSP has set this as a priority for the future.

490 References

The listed references are appropriate

General Comment

The article captured very important details associated with an active reference genome community of practice and vividly explained the challenges faced by such a consortium.

[We thank the reviewer for this positive assessment.](#)