Assessing the potential of germplasm collections for the management of genetic diversity: the case of the French National Cryobank

Revision round #1

- **Recommender**

Major comments

Dears Authors,
We already have the feedback of two independent reviewers. I agree that the manuscript is well-written and presents a valuable methodology for characterizing the genetic diversity of cryopreserved collections in gene banks.
However, before recommending it, the authors must consider the few major comments of both reviewers.
After that, I'm confident that the manuscript can be recommended for publication.
Best regards,
Yuliaxis
by Yuliaxis Ramayo-Caldas, 27 Oct 2023 10:06
Manuscript: https://doi.org/10.1101/2023.07.19.549644
version: 1

*****

Authors' response

Dear recommender,

We thank you for your feedback on our manuscript.
Please find below our answers to the comments of the two reviewers.

Kind regards.

*****

- **Review 1 by Roy Costilla, 25 Oct 2023 05:16**

This study proposes a methodological framework and two novel statistics, a Gini-like index and the index of diversity impact (IDI), to assess the management of genetic diversity in germplasm collections and apply them to the National Cryobank in France. The manuscript is well-written, and the methodology, data and results are clearly explained and detailed.
I congratulate the authors for their nice work, including calculating uncertainty for their novel statistics. Please see below my comments/suggestions.
Major comments
·   The abstract mentions that IDI values can also be calculated using genomic data. However, this is never detailed in the manuscript. It'd be great to expand on the authors' suggestion, outlining how could this be achieved.
·   The term "kindship" is used frequently throughout the manuscript but never adequately defined. Readers will benefit from a proper description of this term and a brief explanation of the method(s) behind the PEDIG software to calculate it.
Minor comments
L62 More recent work from FAO, Ajmone-Marsan et. al (2023), is missing here.
L69 Less common breeds -> Non-commercial breeds?

L153 Regression trees are mentioned for the first time without any explanation (this comes later in L170)
Typos/Omissions
L180 "Nb_donors" not defined in the text
L185 Nb_doses not defined in the text
L238, L238 and others Define variable "phi"
References
Ajmone-Marsan, P., Colli, L., Ginja, C, Kantanen J. & Lenstra, J.A., eds. 2023. Genomic characterization of animal genetic resources. FAO Animal Production and Health Guidelines No. 32. Rome, FAO.

*****

Authors' response

We would like to thank the reviewer for his comments on the article. We have taken his comments/suggestions into account below.
Regards.

- <u>Major comments</u>

The abstract mentions that IDI values can also be calculated using genomic data. However, this is never detailed in the manuscript. It'd be great to expand on the authors' suggestion, outlining how could this be achieved. → OK
We have added some suggestions on how to compute IDI values from molecular data(L592-604).

The term "kindship" is used frequently throughout the manuscript but never adequately defined. Readers will benefit from a proper description of this term and a brief explanation of the method(s) behind the PEDIG software to calculate it → OK
We have added a sentence to define the term "kinship" and we've added the references (L235-237).
We have added an explanation of the method behind the *colleau* module of PEDIG and make the link with the article which described the complete method (L233-234).

- <u>Minor comments</u>

L62 More recent work from FAO, Ajmone-Marsan et. al (2023), is missing here. → OK
The reference has been added at the end of the discussion. The FAO reference (2015) provides an assessment of biodiversity loss in domestic animal populations. The FAO reference (2023) is a guideline for the use of genomic tools to characterize animal genetic resources. So, we have therefore added this relevant reference to L591.

L69 Less common breeds -> Non-commercial breeds? → OK
Not really, some less common breeds can still be economically valuable (especially for local breeds valorized through designation of origin). We prefer to retain the term "rare breeds" (L69) which remains more flexible for characterizing the diversity of breeds of each species.
L153 Regression trees are mentioned for the first time without any explanation (this comes later in L170) → OK
We have corrected the sentence (L153).

L180 "Nb_donors" not defined in the text → OK
L185 Nb_doses not defined in the text → OK
We have added the definition in the text (L171).

L238, L238 and others Define variable "phi" → OK
This is already described in all the text from lines 237 to 245 for each cohort, but we have added definition in L235.

*****

- **Review 2 by anonymous reviewer 1, 16 Oct 2023 10:40**

The paper is well-written and presents a valuable methodology for characterizing the genetic diversity of cryopreserved collections in gene banks. It employs the French National Cryobank as a case study and introduces new biodiversity metrics, such as the Index of Diversity Impact (IDI), for effective management and assessment. I have just few general comments.

The paper lacks a comprehensive discussion on potential methodological limitations that influence the assessment of factors such as the number of donors per breed and doses per donor. In countries with extensive data recording or efficient communication with farmers, controlling for variables like pedigrees might be more straightforward. Therefore, in my opinion, the study falls short in addressing the quality of pedigrees in the interpretation of genetic diversity metrics. This omission compromises the reliability of the study's findings. In situations where pedigrees are unavailable or unreliable—particularly for local breeds in extensive or challenging conditions—alternative methods such as molecular markers or genomic analyses should be considered

The term "quickly," used to describe the loss of genetic diversity, is vague and needs quantitative clarification. Different species can lose diversity at differing rates, so establishing a time-scale would provide useful context.

*****

Authors' response

We thank the reviewer for his appreciation and additional comments.
You'll find our answers below.
Regards.

- Major comments

The paper lacks a comprehensive discussion on potential methodological limitations that influence the assessment of factors such as the number of donors per breed and doses per donor. In countries with extensive data recording or efficient communication with farmers, controlling for variables like pedigrees might be more straightforward. Therefore, in my opinion, the study falls short in addressing the quality of pedigrees in the interpretation of genetic diversity metrics. This omission compromises the reliability of the study's findings. In situations where pedigrees are unavailable or unreliable—particularly for local breeds in extensive or challenging conditions—alternative methods such as molecular markers or genomic analyses should be considered. → OK
Pedigree quality has been calculated for all breeds analyzed (L237-240) and the results are available in Table 3 in order that the reader can examine the results knowing the accuracy of the genealogies.
We have completed the discussion on this subject, in particular we stated that this question of pedigrees completeness is more critical for contemporary individuals than for cryopreserved ones in order to retrieve potential relatedness between those two sets of individuals (L581-587).
We have also developed the discussion concerning the calculation of indicators with molecular data (L592-604)

The term "quickly," used to describe the loss of genetic diversity, is vague and needs quantitative clarification. Different species can lose diversity at differing rates, so establishing a time-scale would provide useful context. → OK
We have added some corrections to the text (L560 and L607).

*****