# Peer Community In Genomics

# Phased genomes suggest that *L. fluviatilis* and *L. planeri* are two ecotypes of the same species

# Samuel Abalde based on peer reviews by Quentin Rougemont, Ricardo C. Rodríguez de la Vega and 1 anonymous reviewer

Ole K. Tørresen, Benedicte Garmann-Aarhus, Siv Nam Khang Hoff, Sissel Jentoft, Mikael Svensson, Eivind Schartum, Ave Tooming-Klunderud, Morten Skage, Anders Krabberød, Leif Asbjørn Vøllestad, Kjetill S. Jakobsen (2025) Comparison of whole-genome assemblies of European river lamprey (*Lampetra fluviatilis*) and brook lamprey (*Lampetra planeri*). bioRxiv, ver. 5, peer-reviewed and recommended by Peer Community in Genomics. https://doi.org/10.1101/2024.12.06.627158

Submitted: 06 January 2025, Recommended: 06 May 2025

#### Cite this recommendation as:

Abalde, S. (2025) Phased genomes suggest that *L. fluviatilis* and *L. planeri* are two ecotypes of the same species. *Peer Community in Genomics*, 100428. 10.24072/pci.genomics.100428

#### Published: 06 May 2025

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Lampreys are the focus of intense research. Together with hagfishes, they form the Cyclostomata, the sister group of jawed vertebrates, and hence they are a key group for disentangling the early evolution of many vertebrate features (Shimel and Donoghue 2012; McCauley et al. 2015). Ecologically, lamprey species show a diverse array of life modes, including parasitic and non-feeding species, and inhabit freshwater and marine habitats or both (i.e. anadromous species; Docker and Potter 2019). One of these anadromous species, the sea lamprey (*Petromyzon marinus*), took advantage of man-made canals to invade the North American Great Lakes in the early 20th century, decimating many fish populations. Today, the control of these invasive populations is paramount for the survival of the region's fishing industry (Ferreira-Martins et al. 2021). All these research avenues will benefit from the generation of new genomic data, an invaluable resource in evolutionary and conservation biology.

In this manuscript, Tørresen et al. (2025) present phased, chromosome-level assemblies from two lamprey species: the European river lamprey (*Lampetra fluviatilis*) and the brook lamprey (*Lampetra planeri*). These two genome assemblies are of high quality and will undoubtedly become a key resource in lamprey research. In particular, the authors showcase the potential of such genomes from two perspectives. First, comparing their assemblies to the already published genomes from *P. marinus* and another specimen of *L. fluviatilis*, they propose that lamprey genomes are highly conserved and display large syntenic blocks shared among species.

Second, phylogenetic analyses and the annotation of SNPs suggest that *L. fluviatilis* and *L. planeri* should be considered two ecotypes of the same species complex, instead of two separate species. This might not be new for anyone knowledgeable in lamprey biology (Rougemont et al. 2017), but it is surprising given the distinct ecology of the two lampreys: *L. fluviatilis* is a parasitic, anadromous species, whereas *L. planeri* is a non-feeding, freshwater species.

In addition to the biological significance of this manuscript, I would like to acknowledge the robustness of the analytical approaches. These genomes were assembled and annotated following two pipelines recently developed at EBP-Nor, the Norwegian initiative of the Earth BioGenome Project (EBP). These pipelines are designed to be an easy-to-use, end-to-end solution for genomic analyses and are likely to become a standard for the EBP and European Reference Genome Atlas initiatives. There can be no better evidence of their effectiveness than these two phased, chromosome-level, highly complete genome assemblies.

#### **References:**

Docker MF, Potter IC (2019) Life history evolution in lampreys: Alternative migratory and feeding types. In: Docker M (ed) Lampreys: Biology, Conservation and Control. Fish & Fisheries Series, vol 38. Springer, Dordrecht. https://doi.org/10.1007/978-94-024-1684-8\_4

Ferreira-Martins D, Champer J, McCauley DW, Zhang Z, Docker MF (2021) Genetic control of invasive sea lamprey in the Great Lakes. Journal of Great Lakes Research, 47, S764-S775. https://doi.org/10.1016/j.jglr.2021.10.018

McCauley DW, Docker MF, Whyard S, Li W (2015) Lampreys as diverse model organisms in the genomics era. BioScience, 65(11), 1046-1056. https://doi.org/10.1093/biosci/biv139

Rougemont Q, Gagnaire PA, Perrier C, Genthon C, Besnard AL, Launey S, Evanno G (2017) Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Molecular Ecology, 26(1), 142-162. https://doi.org/10.1111/mec.13664

Shimeld SM, Donoghue PC (2012) Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development, 139(12), 2091-2099. https://doi.org/10.1242/dev.074716

Tørresen OK, Garmann-Aarhus B, Hoff SNK, Jentoft S, Svensson M, Schartum E, Tooming-Klunderud A, Skage M, Krabberød A, Vøllestad LA, Jakobsen KS (2025) Comparison of whole-genome assemblies of European river lamprey (*Lampetra fluviatilis*) and brook lamprey (*Lampetra planeri*). bioRxiv, ver. 5 peer-reviewed and recommended by PCI Genomics https://doi.org/10.1101/2024.12.06.627158

## **Reviews**

### **Evaluation round #2**

DOI or URL of the preprint: https://doi.org/10.1101/2024.12.06.627158 Version of the preprint: 4

Authors' reply, 25 April 2025

Download author's reply Download tracked changes file

#### Decision by Samuel Abalde , posted 22 April 2025, validated 22 April 2025

Dear Dr Tørresen,

Thank you for the prompt submission of a revised version of this manuscript. I have shared the manuscript again with one of the reviewers and the feedback was very positive. I will be happy to recommend this manuscript, which I'm sure will become a reference for other genomic studies thanks to the detailed and careful pipeline that you present.

However, I need to ask you to fix the paragraph in lines 384-389 before we move on to the recommendation, as it shows two versions of the same sentence. Besides, I have a few small suggestions and caught some typos. They are very minor things and I normally wouldn't have said anything, but since you are editing the text again, it is a good opportunity to go through them:

- There is a double space in lines 33 ("the L. planeri"), 83 (". A central"), 136 ("steps: incubation"), 186 ("with -5SPM") and 260 (before and after the link to the GenomeEvaluation pipeline).

- There is a space missing in lines 319 ("Omni-Creads") and 339 ("L. fluviatilis,the").

- I would also write the full name of Petromyzon marinus in line 102, as this is the first time this species is named in the main text.

- Change "comparative" to "comparable" in line 472.

- In lines 494 – 495, the sentence "If L. fluviatilis and L. planeri were two clearly differentiated species, we would expect more differences between the species than in a species." is a bit awkward to read as is. I would rephrase it to say: "If L. fluviatilis and L. planeri were two clearly differentiated species, we would expect more differences between them than between the two L. fluviatilis specimens." Or something along those lines.

- There are several paragraphs that read overly repetitive (e.g. 339-349, 504-513, among others). It would be easy to spot them with a careful read, but I leave the decision of rewriting these instances to you. This is a styling decision.

Other than these minor fixes, the manuscript is great, and I will be happy to recommend it as soon as the new version of the text is available.

Sincerely, Samuel Abalde

#### Reviewed by Quentin Rougemont, 17 April 2025

Dear Editors and authors,

I have now read the revised version of the manuscript "Comparison of whole-genome assemblies of European river lamprey (Lampetra fluviatilis) and brook lamprey (Lampetra planeri)" by Tørresen et al.

I am happy to see that the authors have adequately addressed all of the comments from the previous round. I feel that my (minor) concerns were nicely clarified in the reply. The text also reads better and the quality of the manuscript is improved. This will be a nice workflow for new genome assemblies/annotation and look forward to using it.

Regarding the discussion on hifiasm options that influenced the assembly length, I found that increased values (i.e. above 0.7) of the -s parameter, that controls the similarity threshold for purging duplicate haplotigs, were increasing the length of the assembly. Same was true for –hom-cov. Other parameters seemed to have minor effects and across most of my assemblies default values are usually fine.

### **Evaluation round #1**

DOI or URL of the preprint: https://doi.org/10.1101/2024.12.06.627158 Version of the preprint: 3

#### Authors' reply, 08 April 2025

#### Download author's reply Download tracked changes file

#### Decision by Samuel Abalde , posted 01 March 2025, validated 03 March 2025

Dear Dr. Tørresen et al,

Thank you for submitting your manuscript to PCI Genomics. I agree with the three reviewers that this is a very interesting work. Given the complex lifestyle structure of river lampreys, the quality of the genome assemblies presented here will be a great resource for future work on this interesting species.

I sent out the manuscript to three expert reviewers. They all agree on the quality of the manuscript but all point out writing issues that should be amended. In line with this, when I first read the manuscript, I had the impression that it missed one final proofreading before submission. There are some bits of vague sentencing that do not fit in the overall narrative and repetitive sentences that should be removed. These of course are not major concerns, but it is important the text reads clearly before we can proceed any further.

Related to the writing, there are two comments that I find merit special attention during the revision. First, one anonymous reviewer suggests re-focusing the manuscript. They think that the main goal –addressing the species boundaries in the river lampreys– is not aligned with the methods and main results, which are more akin to a genome report than to a diversification study. Please, take it into consideration during the revision. Second, please add more detail to the descriptions of the methods and results.

The comments about the results are more delicate. It seems that the selected genome alignment approach might not have been the best. Given that the SNP calling and the subsequent speciation analyses depend on this decision, I encourage you to look into it and try an alternative. The reviewer suggests specific software and parameters that you can use.

Please check the reviewers's comments below and provide a detailed response. I have also attached some of my own.

Sincerely, Samuel Abalde

# Comments to the authors.

In the methods:

- Since *L. fluviatilis* is a parasitic species, cross-contamination with the host is always a possibility. Is this the reason why blood and not muscle was sampled? Was there anything to avoid contaminations or to filter the relevant reads afterwards?

- This is out of curiosity, but is it possible to know how DeepVariant filter the variants or is this some kind of black box? I think the reader would appreciate some kind of detail on how the method works.

#### Other minor comments:

- Table 2: I understand the table, but I think the caption should include a little more detail to clarify the two columns per species referring to the haplotypes.

Lines 45 – 49: This sentence about species boundaries sounds kind of vague. As it is phrased, I would say it is a given.

Lines 80 – 81 "Investigations of structural variation... for the two sister species": this is repetitive. Already said in the previous sentence.

Lines 119 – 128: I understand that sometimes different methods were applied to the two species and you need to describe them separately, but I'd say this paragraph could be summarised if you explain the sequencing for both species at the same time and just highlight the differences.

Lines 179 – 181: Already said at the beginning of the paragraph.

#### **Reviewed by Quentin Rougemont**, 19 February 2025

#### **Download the review**

#### Reviewed by Ricardo C. Rodríguez de la Vega, 15 February 2025

Comments on Torresen et al. "Comparison of whole-genome assemblies..."

In this genome report, the authors present haplotype-resolved assemblies of one individual of the European river lamprey Lampetra fluviatilis and one individual of the brook lamprey L. planeri. Sequencing data include long-read HiFi and Hi-C contact information. Genome assemblies, annotation, and evaluation were obtained using what the authors describe as "pre-release" pipelines of the Norwegian initiative of the Earth Biogenome Project (EBP-Nor).

This is a flawless genome assembly and annotation report, including generous technical details on all steps and many useful pointers to state-of-the-art analytical pipelines. The biological context is adequately introduced, and the insights provided by the new genome assemblies are reasonably discussed. Compared with genome reports from other similar initiatives, notably outputs from Darwin's Tree of Life, I appreciate the less standardized presentation. Although this may make it harder to systematically parse information in the text, it is balanced by the fact that assembly metrics remain comparable across different studies. I hope the EBP-Nor assembly, annotation, and evaluation pipelines will soon be released in a fully portable form.

I just highlight a list of minor points requiring clarification and conclude by addressing the PCI's questions for reviewers.

#### PCI's questions for reviewers

Title and abstract

Does the title clearly reflect the content of the article? Yes

Does the abstract present the main findings of the study? Yes

Introduction

Are the research questions/hypotheses/predictions clearly presented? Yes

Does the introduction build on relevant research in the field? Yes

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? Yes

Are the methods and statistical analyses appropriate and well described? Yes

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? There ain't no negative results

Are the results described and interpreted correctly? Yes

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? Yes

Are the conclusions adequately supported by the results (without overstating the implications of the findings)? Yes

#### Minor points and proofreading suggestions:

- Marks 119 and 129: Sentences read awkwardly; it was not the individuals that "underwent" sequencing, but rather DNA extracted from them. Consider rewording.

- Marks 237-238: It is unclear if DIAMOND handles the -max-hsps parameter the same way as BLAST's -max\_target\_seq. It may be worth verifying this (see https://academic.oup.com/bioinformatics/

#### article/35/9/1613/5106166).

- Figure 1: Could the authors clarify what the scale insets on the Snail plots represent?

- Marks 267-270: Assembly statistics do not need to be repeated in the main text.

- Table 2: Could the sex of L. planeri be inferred from sequencing data?

- Table 2: One haplotype consistently shows twice as many duplicated BUSCOs; is there a biological explanation for this?

- Marks 290-296: Wording needs polishing, and some clarification is needed (the first comparison appears to contradict the values in Table 2). Consider simplifying, as not all these details may be necessary in the main text.

- Marks 301 and elsewhere: Standardize formatting of bioinformatics tools, which are sometimes italicized and sometimes in normal font.

- Table 3: Define what the values in parentheses represent.

- Marks 370-378: Consider removing the results summary at the start of the discussion, as it largely repeats information from the results section.

#### Reviewed by anonymous reviewer 1, 20 February 2025

The authors of the manuscript entitled "Comparison of whole-genome assemblies of European river lamprey (Lampetra fluviatilis) and brook lamprey (Lampetra planeri)" provide haplotype-resolved chromosome-level reference genome assemblies of two sister species of lampreys. Genomes presented in this work showed high standards in terms of contiguity (chromosome level) and completeness. Overall, the manuscript is well explained and the data generated will be very useful, not only for researchers interested in this group of fish, but any interested in comparative genomics.

However, the framing of the study appears somewhat incongruent with the methodology employed. The authors mention that the main objective is to shed light on the species versus ecotype discussion, but it remains unclear whether whole genome assembly is the proper methodology for this. Given that only one individual per species was sequenced, it is unclear whether the data are sufficiently informative to robustly assess species boundaries. As they mention in the introduction, additional population-level sampling and comparative genomic analyses (e.g., population structure, admixture, or selection scans) could provide more direct insights into this question.

Therefore, a reframing of the primary focus of the study is suggested to emphasize the sequencing and comparative analysis of these genomes. The investigation into whether these taxa constitute separate species or ecotypes could then be presented as a secondary objective, to be explored further with more targeted methodologies.

Furthermore, while the introduction extensively discusses the ecological aspects of these taxa, the discussion section does not seem to fully integrate these considerations when interpreting the genomic findings. This manuscript would benefit greatly from some expansion in the discussion.

Particular issues

Additional minor comments have been attached that could help improve the manuscript: Abstract

ADSILACI

Line 25. Replace "... and the brook lamprey" with "... and two of the brook lamprey"

Introduction

Line 43. How about the large, rapidly diversifying lakes such as those in Africa and Nicaragua?

Lines 47 – 49. The authors mention that there are numerous examples of challenges in determining species in purely marine habitats, maybe they could provide some specific examples and their respective references.

Lines 86 – 87. Is it really the main objective?

Methods

Line 100. Not the same tissues, why?

Line 105. Blood was not mentioned before, right?

Line 128. "M SMRT cells". Above mentioned 8M (line 125), what is it? Are they the same?

Line 137. "Arima Genome-Wide Hi-C+ Kit" Why not the same library protocol?

Line 148. Smudgeplot. Why isn't it included in Table 1?

Line 163. Bedtools version?

Line 173. More details?

Line 179. These assessment tools.

Line 183. Table 1. Maybe this table would be better as supplementary material.

Lines 238 – 239. These codes are not the same as in the rest of the text, tables 2, 3 or supp. table 1. It's understood that that kcLamPlan1.1.hap1 is kcLamPlan1.2.hap1, and the same for the other three. Please clarify this.

Results

Line 249 – 250. Better PacBio instead of Pacific Biosciences

Line 264 and 266. "... of 1073 Mb (Figure 1) and 963 Mb (Table 2)" "...of 1049 Mb (Figure 1) and 960 Mb". The genome assembly sizes differ by approximately 100 Mb between haplotypes in both species. The authors could mention something about the possible causes of these differences.

Line 269 – 270. Replace "... and 12.9 Mb in pseudo-haplotype one and two, respectively" with: "in both one and two pseudo-haplotypes."

Line 273. Code kcLamFluc1 appears instead of kcLamFluv1. Please correct the c.

Table 2. Number of scaffolds. Missing number of chromosomes or pseudochromosomes.

Table 2. Organelles (MT). And maybe they could do something about this if they already have it sequenced, right?

Line 335. "... their support" They all look similar, right?

Line 346. "... sea lamprey" The scientific names should be used, as in the rest of the manuscript, to maintain consistency.

Line 347. Replace rrook with brook.

Line 354. What about comparing different haplotypes (hap1 and hap2) of the same individual for both L.fluviatilis and L.planeri?

Discussion

General: In the introduction, the ecological aspects of these species are extensively discussed; however, this topic is scarcely addressed in the discussion section. The discussion appears somewhat unbalanced.

Line 399. "... most of the gene trees (40.1%)". Referring to 40% as 'most' misleading. Could the authors clarify this?

Lines 412 – 413. "... suggests that the two species rather is a species complex representing two ecotypes". Can this conclusion be suggested based on chromosome-level genome assemblies from two individuals?

Supplementary Material

Supplementary Figure 5. Lines 42 to 44. Please indicate the correspondence of the codes in the figure: kcLamPlan1 (RS), kcLamFluv2 (BS), kcLamFluv1 (RB), and kPerMar (SL).

- I Title and abstract
- o Does the title clearly reflect the content of the article? [X] Yes, [] No (please explain), [] I don't know

o Does the abstract present the main findings of the study? [X] Yes, [] No (please explain), [] I don't know

Introduction

o Are the research questions/hypotheses/predictions clearly presented? [X] Yes, [] No (please explain), [] I don't know

o Does the introduction build on relevant research in the field? [X] Yes, [] No (please explain), [] I don't know

I Materials and methods

o Are the methods and analyses sufficiently detailed to allow replication by other researchers? [X] Yes, [] No (please explain), [] I don't know

o Are the methods and statistical analyses appropriate and well described? [] Yes, [] No (please explain), [X] I don't know.

As mentioned before, it is unclear whether the methods align with the proposed hypothesis.

Results

o In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? [] Yes, [] No (please explain), [X] I don't know

o Are the results described and interpreted correctly? [] Yes, [] No (please explain), [X] I don't know

Could the authors "suggest" that the two species rather is a species complex representing two ecotypes" based on the methodology and the results?

Discussion

o Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? [] Yes, [] No (please explain), [X] I don't know

o Are the conclusions adequately supported by the results (without overstating the implications of the findings)? [] Yes, [] No (please explain), [X] I don't know

As mentioned above, the alignment between the main objective, the methods, and the discussion seems somewhat misaligned. The manuscript would benefit greatly from some expansion in the discussion.

#### Download the review