



HAL
open science

Reviews of “Whole-genome screening for near-diagnostic genetic markers for four western European white oak species identification“

Antoine Kremer, Adline Delcamp, Isabelle Lesur Kupin, Stéphanie Wagner, Christian Rellstab, Erwan Guichoux, Thibault Leroy, Alexandru-Lucian Curtu, Marjana Westergren, Erwin Dreyer

► To cite this version:

Antoine Kremer, Adline Delcamp, Isabelle Lesur Kupin, Stéphanie Wagner, Christian Rellstab, et al.. Reviews of “Whole-genome screening for near-diagnostic genetic markers for four western European white oak species identification“. 2024, 10.17180/73QH-D818 . hal-04529452

HAL Id: hal-04529452

<https://hal.inrae.fr/hal-04529452>

Submitted on 14 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Identifiers

Open Peer Review

DOI: 10.17180/73QH-D818

Reviewed Article

DOI: 10.1186/s13595-024-01236-9

History

Reviews: Feb. 2024

Rebuttal: Feb. 26, 2024

Acceptance: March 21, 2024

Licence

CC BY 4.0

©The authors

Reviews of “Whole-genome screening for near-diagnostic genetic markers for four western European white oak species identification”

Authors: Antoine Kremer^{1*§}, Adline Delcamp², Isabelle Lesur³, Stefanie Wagner⁴, Christian Rellstab⁵, Erwan Guichoux^{2,*}, Thibault Leroy^{1,6§}

Reviewers: Anonymous^R, Alexandru Lucian Curtu^{7,R}

Editors: Marjana Westergren^{8, HE}, Erwin Dreyer^{9, EIC}

* Corresponding author

§ These authors contributed equally to the research.

R Reviewer

HE Handling Editor

EIC Editor in chief

1. UMR BIOGECO, INRAE, Université de Bordeaux, 69 Route d'Arcachon, 33612 Cestas, France
2. UMR BIOGECO, INRAE, Université de Bordeaux, PGTB, 69 Route d'Arcachon, 33612 Cestas, France
3. Helix Venture, 33700 Merignac, France
4. UMR CAGT, CNRS, Université Paul Sabatier, 37 Allées Jules Guesde, 31000 Toulouse, France
5. Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland
6. GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France
7. University of Transilvania, Bulevardul Eroilor 29, Braşov 500036, Roumania
8. Slovenian Forestry Institute, Večna pot 2, 1000 Ljubljana, Slovenia
9. UMR SILVA, INRAE, Université de Lorraine, AgroParisTech, rue d'Amance, 54280 Champenoux, France

Original submission – R0

R0 - Authors

Cover letter R0

Cestas, November 25 2023

Editors of Annals of Forest Science,

On behalf of all the co-authors, I am submitting the manuscript entitled «Whole-genome screening for near-diagnostic genetic markers for white oak species identification in Europe» for publication in Annals of Forest Science

The manuscript addresses a long-standing concern in forest science and forest management concerning the taxonomic assignment of species that exhibit extremely low genetic differentiation, as the white oaks. We made use of whole genome sequences to explore diagnostic genomic imprints of species and found a limited number of markers that are close to full diagnosticity. We discuss the evolutionary processes that may have shaped diagnosticity in a very few numbers of genomic regions. We think that our results offer a new invaluable tool not only in science (evolutionary biology, paleobotany, taxonomy, introgression...) but also in operational forestry (forensics, traceability, certification...).

The manuscript has not been submitted to any other journal.

I thank you for considering the manuscript for submission to Annals of Forest Science

Sincerely yours,

Antoine Kremer

RO - Reviewer 1 (Anonymous)

Reviewer Recommendation Term: Accept with minor modifications.

Level of interest: An article of importance in the field.

Quality of written English: Acceptable.

Declaration of competing interests: I declare that I have no competing interests.

The submitted article is a useful and original contribution as it offers an easy-to-use and efficient tool for species identification in European white oaks. The study makes use of extensive genomic data from a large number of populations, which are stretched over large parts of the species' distribution, compared to earlier research. In order to identify highly informative loci, an ad hoc metric, diagnosticity, is introduced, whose calculation is very straightforward, in my opinion. In addition, on the extensive genomic data used and the good coverage of the species' ranges, a further strength of the study is the ease of application of the proposed marker set consisting of only 38 SNPs. Moreover, the inclusion of *Quercus pubescens* and *Q. pyrenaica* will be also welcome by many researchers, as these two species have been more rarely included in similar studies. The study is methodologically sound, the results are presented properly with meaningful and informative Tables and Figures. The study is put in a larger context of previous research citing significant related papers. Finally, the Discussion covers important issues like species delimitation, but also genome permeability, as well as current and past interspecific gene flow. Therefore, I think that the study is a valuable contribution to publish in *Annals of Forest Science*.

I only have a few minor comments which I summarize here:

L. 85: "to correctly assign to the correct species": I think "correctly" can be deleted here.

L. 124: the Trimmomatic options should be written in text, not with abbreviations which may not be familiar to many readers of *Ann. For. Sci.*

L. 169: even if it becomes understandable from the context, I needed to re-read the text 1-2 times and check the cited table to understand that a sample of a certain number of individuals (and not one tree per population) is meant here. Maybe this could be stated more clearly with some minor additions to this text.

L. 169 onwards: "Table 5 in Appendix" is OK, but a bit unusual for me. I would usually opt for Table / Figure 1, 2, 3 in the manuscript and Table / Fig. S1, S2, ... in the Appendix (and then there is not even need to repeat "in Appendix").

L. 257: I would cite Figure 2 again when talking about the clustering of near-diagnostic markers on a chromosome (even if it's cited in the previous sentence), as this is easily visible on the Figure.

L. 313: Do the authors mean Table 7 in Appendix here?

L. 351-352: This is not very clear. I would rather expect the authors to stress the high frequency of other species's near-diagnostic markers in DP populations instead the low frequencies of such markers in EP populations. I think re-writing this sentence would help the reader follow the text better (the next paragraph, 355-360, for example, explains better (in my opinion) the same phenomenon in other species.

L. 407: I think, prevented from gene flow is not fitting here. I would say "less affected by" or "less permeable to"

RO - Reviewer 2 (Alexandru Lucian Curtu)

Reviewer Recommendation Term: Accept with minor modifications

Level of interest: An article of importance in the field.

Quality of written English: Acceptable.

Declaration of competing interests: I declare that I have no competing interests.

This manuscript is a very good example on how genetics can help operational forestry and forest research distinguish easily between four closely related European white oak species and their hybrids, which co-occur frequently across Europe. A series of highly discriminating SNPs with species-specific alleles for one of the four studied oak species have been identified and their analysis optimised. The methodological approach is sound (e.g., the use of Gregorius measure for genetic differentiation instead of F_{ST} for the newly defined diagnosticity index). The authors also

explained very well the variation of diagnosticity within the genome and throughout the natural range of the four target species.

I have only some minor suggestions:

Title needs to reflect that only four European white oak species were considered. As it was mentioned in the manuscript, there are other white oak species in Europe.

Abstract: Ln. 12 "Aims: Our aim was to identify diagnostic single nucleotide polymorphisms" and in the text, Ln.95-100 " Our main goal was to a new set of near-diagnostic SNPs". Diagnostic or near-diagnostic? The aim should be consistent throughout the manuscript, for example, using "near-diagnostic SNPs", as fully diagnostic markers rarely exists (Ln. 81).

Ln 55-56: "morphological characteristics" At line 45 there were "morphological characters". Try to be consistent across the manuscript and use the same terminology (e.g., characteristics).

Ln. 189: the samples were collected by local collectors. Did the authors check themselves the leaf morphology to make sure the species was correctly identified in the field?

Discussion: the possible impact of genetic drift on the variation of the allele frequency profiles across geographic regions within species should be mentioned.

Ln. 824: Table 5 in Appendix. Geographic origins of training and validation samples. Adding a line with the total number of samples would be helpful.

Figure captions

Figure 1, a,b,c, d - please mention the oak species for each section (e.g. 1a - *Q. petraea*, 1b ...)

Figure 5 - Figure 5a not "1a"

RO - Handling Editor (Marjana Westergren)

Recommendation R0: Accept with minor modifications

Dear Authors,

Thank you for submitting your manuscript entitled Whole-genome screening for near-diagnostic genetic markers for white oak species identification in Europe to *Annals of Forest Science*. The manuscript is a well-written original contribution and presents an easy-to-use and affordable tool consisting of 38 diagnostic SNP markers for operational forestry, conservation and forestry research, allowing easy discrimination between four closely related European white oaks, *Q. petraea*, *Q. robur*, *Q. pubescens* and *Q. pyrenaica*. The study is based on extensive genomic data from a large number of populations across the distribution ranges of the four species and defines a new informative metric, diagnosticity, which is easy to calculate.

I firmly believe that your contribution will be welcomed by researchers, foresters and conservationists working with oaks and that the presented metric of diagnosticity has the potential to be used for species identification of other sympatric, hybridising (tree) species.

Please consider simplifying (and perhaps slightly extending) your key message; while it is easy to understand for a geneticist, it might be challenging for other profiles. The key message will be disseminated as a stand-alone text to draw attention to the paper once published.

Please address the minor comments raised by the two reviewers. Also, once you have uploaded the data to publicly accessible data repository of INRAE and obtained a DOI, you can cite it in the reference list of your paper, increasing the impact and the visibility of your work.

RO - Editor in chief (Erwin Dreyer)

Decision letter R0: Accept with minor modifications

Dear Dr Antoine Kremer,

We have received the reports from our advisors on your manuscript (ANFS-D-23-00177), 'Whole-genome screening for near-diagnostic genetic markers for white oak species identification in Europe', submitted to *Annals of Forest Science*.

The two external reviewers as well as the handling editor are very positive to the content of the manuscript, and only require some minor modifications before the manuscript may be published. I concur with their opinion, and believe this will be a very interesting contribution for diagnosing the species identity in this large syngamon of four white oak species in western Europe. The

reviewers and the handling editor provided a few quite relevant suggestions of improvement of this very clear and well-structured manuscript. I have a few very minor additional suggestions in the attached copy of the manuscript:

1. provide access to the data for this last round of revision and cite the data set in the reference list;
2. expand the key message for a broader audience;
3. renumber the figures and tables in the appendix: they should be in continuity with the main text;
4. correct a few minor typos along the text.

Below, please find the reviewers' comments for your perusal. You are kindly requested to also check the website for possible reviewer attachment(s).

In the frame of the test of our "Optional Open Review Process", we might combine the external reports and the decisions into a peer review report and publish it with a DOI in a devoted section of the HAL platform. Do you still agree with this? We will also require the agreement of the two reviewers (they were already consulted) before any further move from our side. This will only happen after final publication.

With kind regards, and many thanks for this very interesting and well-structured contribution to Annals of Forest Science.

Revised Version R1

R1 - Authors

Cover letter R1

Cestas, February 25, 2024

Editors of Annals of Forest Science

On behalf of all the co-authors, I am submitting the revised version of manuscript ANFS-D-23-00177 entitled « Whole-genome screening for near-diagnostic genetic markers for four western European white oak species identification » for publication in Annals of Forest Science.

We addressed all the comments and suggestions made by the editors and reviewers, and attached our responses in a separate file. We amended the manuscript accordingly. Notably we changed the title and extended the key message. The data were stored on the INRAE data repository and we provided the doi of the data set. We also attach all the dois of the references.

We agree with your suggestion to publish the review report with a DOI in a devoted section of the HAL platform, if you intend to proceed as such.

I thank you for considering the manuscript for submission to Annals of Forest Science.

Sincerely yours,

Antoine Kremer

Answers to editors

1. *provide access to the data for this last round of revision and cite the data set in the reference list;*
2. *expand the key message for a broader audience;*
3. *renumber the figures and tables in the appendix: they should be in continuity with the main text;*
4. *correct a few minor typos along the text.*

Data are available at the INRAE data repository and are accessible via a doi url address (<https://doi.org/10.57745/0JYLZU>). The doi was added in the reference list.

The key message was extended within the editorial length allowed (100 words). The extension provides clarifications of the term «near-diagnostic» and highlights the potential and promising uses that can be made by this method in operational forestry and research.

Figures and Tables were renumbered according to the editorial rules. And typos were corrected in the manuscript. According to the recommendations to the authors, the numbering of the tables

and figures in the Appendix should follow sequentially the numbers used in the main text. We followed these rules.

In the frame of the test of our "Optional Open Review Process", we might combine the external reports and the decisions into a peer review report and publish it with a DOI in a devoted section of the HAL platform. Do you still agree with this? We will also require the agreement of the two reviewers (they were already consulted) before any further move from our side. This will only happen after final publication.

We agree with this option.

Answers to Reviewer 1

L. 85: "to correctly assign to the correct species": I think "correctly" can be deleted here.

Correction done.

L. 124: the Trimmomatic options should be written in text, not with abbreviations which may not be familiar to many readers of Ann. For. Sci.

We clarified the meaning of the different options used in Trimmomatic and added the following text. « Raw reads were then trimmed using Trimmomatic (v. 0.33, Bolger et al, 2014). Bases off the start and off the end of each read were removed if their quality fell below 3 (LEADING:3 TRAILING:3). Then, a sliding window was performed using SLIDINGWINDOW:4:15. Starting at the 5' end, the scanning cut the read once the average quality within the 4 bp window falls below 15. By considering multiple bases, a single quality base below the threshold of 15 will not cause the removal of high quality data later in the read. Following these steps, only reads longer than 50bp were kept for the analysis (MINLEN:50). »

L. 169: even if it becomes understandable from the context, I needed to re-read the text 1-2 times and check the cited table to understand that a sample of a certain number of individuals (and not one tree per population) is meant here. Maybe this could be stated more clearly with some minor additions to this text.

The composition of the training set was rephrased as follows:

«The candidate diagnostic SNPs of the discovery panel were first tested on a limited number of oak individuals (training set), randomly sampled in 7 to 14 populations per species (in total 19 to 48 sampled trees per species) ».

L. 169 onwards: "Table 5 in Appendix" is OK, but a bit unusual for me. I would usually opt for Table / Figure 1, 2, 3 in the manuscript and Table / Fig. S1, S2, ... in the Appendix (and then there is not even need to repeat "in Appendix").

We followed the editorial comments of the journal. Annals of Forest Science does not add supplemental material, but recommends to insert additional tables and figures as Appendix material. The numbering of Tables and Figures in Appendix should follow those cited in the main text. We followed these recommendations and numbered the Tables and Figures accordingly.

L. 257: I would cite Figure 2 again when talking about the clustering of near-diagnostic markers on a chromosome (even if it's cited in the previous sentence), as this is easily visible on the Figure.

Correction done.

L. 313: Do the authors mean Table 7 in Appendix here?

Correct. Table 7 is more relevant to be cited here than Table 6.

L. 351-352: This is not very clear. I would rather expect the authors to stress the high frequency of other species's near-diagnostic markers in DP populations instead the low frequencies of such markers in EP populations. I think re-writing this sentence would help the reader follow the text better (the next paragraph, 355-360, for example, explains better (in my opinion) the same phenomenon in other species.

This section was rephrased, putting more emphasis on the deviation of allele frequencies of the DP populations. The DP *Q. petraea* populations exhibited elevated frequencies of near-diagnostic alleles of *Q. pubescens* (Pomieri and Aspromonte), or *Q. robur* (Killarney and Montejo). Thus, they deviated also in respect to the expected frequencies of near-diagnostic alleles of the three other species, contrary to the EP populations (Fig. 4).

L. 407: I think, prevented from gene flow is not fitting here. I would say "less affected by" or "less permeable to".

Correction made.

Answers to Reviewer 2

Title needs to reflect that only four European white oak species were considered. As it was mentioned in the manuscript, there are other white oak species in Europe.

We changed the title. « Whole-genome screening for near-diagnostic genetic markers for four western European white oak species identification »

Abstract: Ln. 12 "Aims: Our aim was to identify diagnostic single nucleotide polymorphisms" and in the text, Ln.95-100 " Our main goal was to a new set of near-diagnostic SNPs". Diagnostic or near-diagnostic? The aim should be consistent throughout the manuscript, for example, using "near-diagnostic SNPs", as fully diagnostic markers rarely exists (Ln. 81).

We checked throughout the manuscript to be consistent. We changed when relevant « diagnostic » into « near-diagnostic ».

Ln 55-56: "morphological characteristics" At line 45 there were "morphological characters". Try to be consistent across the manuscript and use the same terminology (e.g., characteristics).

We checked throughout the manuscript and changed « characters » into « characteristics » when relevant.

Ln. 189: the samples were collected by local collectors. Did the authors check themselves the leaf morphology to make sure the species was correctly identified in the field?

We did not check the leaves at reception of the lots. We found very few misclassifications after the molecular assay in the following populations: Killarney, Dalkeith and Arbalan.

Discussion: the possible impact of genetic drift on the variation of the allele frequency profiles across geographic regions within species should be mentioned.

Given the large population sizes recurrently mentioned in oak genetic surveys, we did not mention the potential effect of genetic drift. However, in marginal isolated populations, for example (Pomieri, Aspromonte, Montejo...), genetic drift may have impacted changes in near-diagnostic allele frequencies.

We mentioned explicitly the putative effect of genetic drift in the latter case. « In populations located at the southern edge of distribution (Pomieri, Aspromonte, and Montejo for *Q. petraea*, Fig. 1 and Fig. 4), the lower diagnosticity may have resulted from more ancient genetic exchanges with *Q. pubescens* and *Q. robur*, **not excluding the potential role of genetic drift in isolated populations**. Indeed, the two italian populations (Pomieri and Aspromonte) in Sicilia and Calabria consist today in almost pure **isolated** stands, where *Q. pubescens* is extremely rare, if not absent (Bagnato et al, 2012; Modica, 2001), while our results indicated introgression of *Q. pubescens* into *Q. petraea* (Fig. 4). »

Ln. 824: Table 5 in Appendix. Geographic origins of training and validation samples. Adding a line with the total number of samples would be helpful.

This table was amended and completed and total numbers added.

Figure captions

Figure 1, a,b,c, d - please mention the oak species for each section (e.g. 1a - *Q. petraea*, 1b ...)

Figure 5 - Figure 5a not "1a"

These indications are added in the captions of the Figures. In the final publishing they will most likely be added by the publisher.

R1 - Handling Editor (Marjana Westergren)

Recommendation R1: Accept

R1 – Editor in chief (Erwin Dreyer)

Decision R1: Accept