Old lake versus young taxa: a comparative phylogeographic perspective on the evolution of Caspian Sea gastropods (Neritidae: *Theodoxus*)

Arthur F. Sands, Thomas A. Neubauer, Saeid Nasibi, Majid Fasihi Harandi, Vitaliy V. Anistratenko, Thomas Wilke and Christian Albrecht

Article citation details
*R. Soc. open sci.* 6: 190965.
http://dx.doi.org/10.1098/rsos.190965

Review timeline
Original submission: 3 June 2019
1st revised submission: 7 August 2019
2nd revised submission: 29 August 2019
Final acceptance: 18 September 2019

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History
RSOS-190965.R0 (Original submission)

Review form: Reviewer 1

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
No

Is the language acceptable?
Yes

Do you have any ethical concerns with this paper?
No
Have you any concerns about statistical analyses in this paper?
Yes

Recommendation?
Major revision is needed (please make suggestions in comments)

Comments to the Author(s)
This study provides a comparative phylogeographic analyses of Theodoxus gastropods from the Pontocaspian region and Southern Iran. The study is based on a good dataset with samples from quite a few distinct localities well spread across the study area. Although the manuscript is generally well written and most of the analyses/conclusions drawn from the analyses are ok, I do have some comments that need to be considered before this ms is ready for publication:

Major comments on analyses:

1. General comment on LTT-plots: LTT plots are typically NOT used for assessing demographic histories, but rather to infer/visualise the accumulation of distinct species/phylogenetic lineages through time. LTTs are not very useful for inferring intraspecific patterns of diversification as here LTTs would only make sense if you had all (or nearly all) the species' haplotypes included. As this is something you cannot be sure of, I strongly recommend removing LTTs from your ms.

2. Bayesian Skyline Plots: You have lots of phylogeographic structure in your data, in particular in the Iranian dataset, but also in the Pontocaspian data (even if you state that there is hardly any population genetic structuring). It is well known that BSPs might be heavily affected by population structure and there are already some studies available that tested the effect of population structure and sampling design on the performance of BSPs. I strongly recommend to check the relevant literature and adapt your analysis strategy and/or interpretation of the BSP patterns accordingly.

Minor comments:

1. Summary & Materials and methods: You write that your study is based on three amplified gene fragments. It's never mentioned you also sequenced them ...

2. I strongly suggest to provide some more information on the biology/ecology of the study species (habitat preference, specialist or generalist, ...) and the hydrological systems(s) (potential habitat discontinuities in the Caspian Sea that are potential barriers to dispersal, hydrological connection to other systems, ...). This info is important for interpreting the patterns you find.

3. Summary, Results & Discussion: Is the resolution (mutation rate) of your markers good enough to test for a potential influence of Pleistocene lake level fluctuations on population structure?

4. Introduction, line 47: How old is the Caspian Sea?

5. Introduction, line 52: In what sense is the Caspian Sea a unique model system to study the effect of palaeo-environmental changes on the phylogeographic structure in long-lived lake biota? It's certainly an excellent model system, but I can think of quite a few other systems that are equally well (or maybe even better) suited - e.g., just think of the East African Great Lakes.

6. M&M, page 2, line 48: Why did you use the BD tree prior? Any specific reasons for doing so? I'm asking because you are looking at a mixture of inter- and (mainly) intraspecific data and tree prior choice might not be a trivial issue here.
7. M&M, page 2, lines 67-68: I guess you mean 95\% HPD.

8. M&M, page 3, line 16+: Did you do a BSP (as stated in the ms) or an EPSP analysis? I'm asking because as far as I know the standard BSP analysis does not allow for multilocus data. Please clarify.

8. M&M, page 3, line 25: Did you really use the 95\% confidence intervals or rather the 95\% HPD intervals (which would be the standard way).

9. Results, page 3, lines 33-34: Three (in Iran) and on (Pontocaspian) divergence events? You can't say it this way, especially since there is quite some intralineage divergence in the main Pontocaspian haplogroups. Please rephrase.

10. Results, page 3, lines 38+: It would be good to indicate the different morphospecies you refer to here also in the tree.

11. Discussion, page 4, line 1: Where is the evidence for the proposed catastrophic bottleneck? This is not based on your data, or is it? Please clarify.

12. Discussion, page 4, line2 & Conclusions, page 5, line 30: no phylogeographic structure in the Pontocaspian lineage? I disagree here. Judging from the network and Table S1, there's quite some structure, within the three main clades/refugial lineages. I think you can and should put much more emphasis on the phylogeographic patterns in the discussion, e.g. how haplogroup V got distributed over such a large area, from the Black Sea over rivers northwest of the Caspian Sea and the Caspian Sea to rivers southeast of the Caspian Sea, whereas the two subclades of haplogroup VI have very narrow distributions. By the way, why don't you consider these two subclades different haplogroups? This would make much more sense.

13. Discussion: Please discuss your findings also in light of what is known on habitat preferences of the species. E.g., are Theodoxus habitat specialists or rather generalists? And are there any obvious habitat discontinuities that might act as dispersal barriers?

Review form: Reviewer 2

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
Yes

Is the language acceptable?
Yes

Do you have any ethical concerns with this paper?
No

Have you any concerns about statistical analyses in this paper?
Yes

Recommendation?
Accept with minor revision (please list in comments)
Comments to the Author(s)

The manuscript reports on a phylogenetic reconstruction of the neritid genus *Theodoxus* from the Caspian Sea region. The authors have a range of samples from across the region, and used three genes (two mitochondrial, one nuclear) for their inferences. There was evidence of a general lack of reciprocal monophyly among species, but instead isolation by distance was present. The authors also undertook reconstructions of historic population sizes using their data, and speculate on potential links between climate and present day diversity. Overall this is a nice paper based on a hard-won dataset. The analyses appear appropriate to the questions being asked, and the results are fairly clear. Most of my comments relate to relatively minor issues of presentation.

P2, L40-43. I’m not convinced there is any evidence here of effects of low lake levels on population genetic structure. At best the evidence is that high stands may have provided opportunities for genetic homogenisation.

P2, L58. use “only a small number”

P3, L23. Use “should possess similar”

P3, L35. I’m not convinced you give any insights into plasticity here (it doesn’t appear to be mentioned elsewhere in the paper). Indeed, these data alone do not tell us anything about the plasticity, or otherwise, of these phenotypes.

P3, L41. I am not familiar with the tests you used for saturation (it would be helpful if you named and cited the specific test used), but typically a p < 0.05 is considered statistically significant.

P3, L46. You say taxa, but this is a pleural of taxon (a group of one or more populations of an organism). Surely you mean individual.

P4, L3. No further burn-in removed.

P4, L32. It is not clear how your “phylogroups” were delimited.

P5, L1. I’m not sure you found any evidence for a catastrophic bottleneck. The data are, I feel, being overinterpreted.

P6, L6. RE synonomy. Your analysis is based on effectively 2 independent loci (both mtDNA genes are linked). Could such patterns arise from parallel evolution?

P6, L35. “contrasts with the fossil evidence”

P9, L37. Zookeys volume absent.

P11, Table. All your p-values for your isolation by distance tests are identical? It is not clear to me how you tested for isolation by distance, but since Mantel tests are now suggested to be flawed you should consider distance-based redundancy analysis in the vegan R package.

P13-15, Figures. Your colours are indistinguishable. I suggest you reconsider these, perhaps use letter and colour combinations.

P17, L37. What is “intraspecific diversification” in this sense? Why should it only start at this point? It doesn’t make any logical sense to me.

Decision letter (RSOS-190965.R0)

28-Jun-2019

Dear Mr Sands,

The editors assigned to your paper (“Old lake vs. young taxa: a comparative phylogeographic perspective on the evolution of Caspian Sea gastropods (Neritidae: *Theodoxus*)”) have now received comments from reviewers. We would like you to revise your paper in accordance with the referee and Associate Editor suggestions which can be found below (not including
confidential reports to the Editor). Please note this decision does not guarantee eventual acceptance.

Please submit a copy of your revised paper before 21-Jul-2019. Please note that the revision deadline will expire at 00.00am on this date. If we do not hear from you within this time then it will be assumed that the paper has been withdrawn. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office in advance. We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Editors, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available, we may invite new reviewers.

To revise your manuscript, log into http://mc.manuscriptcentral.com/rsos and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision. Revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you must respond to the comments made by the referees and upload a file "Response to Referees" in "Section 6 - File Upload". Please use this to document how you have responded to the comments, and the adjustments you have made. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response.

In addition to addressing all of the reviewers' and editor's comments please also ensure that your revised manuscript contains the following sections as appropriate before the reference list:

• Ethics statement (if applicable)
If your study uses humans or animals please include details of the ethical approval received, including the name of the committee that granted approval. For human studies please also detail whether informed consent was obtained. For field studies on animals please include details of all permissions, licences and/or approvals granted to carry out the fieldwork.

• Data accessibility
It is a condition of publication that all supporting data are made available either as supplementary information or preferably in a suitable permanent repository. The data accessibility section should state where the article's supporting data can be accessed. This section should also include details, where possible of where to access other relevant research materials such as statistical tools, protocols, software etc can be accessed. If the data have been deposited in an external repository this section should list the database, accession number and link to the DOI for all data from the article that have been made publicly available. Data sets that have been deposited in an external repository and have a DOI should also be appropriately cited in the manuscript and included in the reference list.

If you wish to submit your supporting data or code to Dryad (http://datadryad.org/), or modify your current submission to dryad, please use the following link: http://datadryad.org/submit?journalID=RSOS&manu=RSOS-190965

• Competing interests
Please declare any financial or non-financial competing interests, or state that you have no competing interests.
• Authors’ contributions
All submissions, other than those with a single author, must include an Authors’ Contributions section which individually lists the specific contribution of each author. The list of Authors should meet all of the following criteria; 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published.

All contributors who do not meet all of these criteria should be included in the acknowledgements.

We suggest the following format:
AB carried out the molecular lab work, participated in data analysis, carried out sequence alignments, participated in the design of the study and drafted the manuscript; CD carried out the statistical analyses; EF collected field data; GH conceived of the study, designed the study, coordinated the study and helped draft the manuscript. All authors gave final approval for publication.

• Acknowledgements
Please acknowledge anyone who contributed to the study but did not meet the authorship criteria.

• Funding statement
Please list the source of funding for each author.

Once again, thank you for submitting your manuscript to Royal Society Open Science and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Kind regards,
Lianne Parkhouse
Editorial Coordinator
Royal Society Open Science
openscience@royalsociety.org

on behalf of Dr Kristina Sefc (Associate Editor) and Kevin Padian (Subject Editor)
openscience@royalsociety.org

Subject Editor’s comments (Professor Kevin Padian):

Thanks for your submission. We hope you find the reviewers’ comments constructive as you revise, and we look forward to the next version. Best wishes.

Associate Editor’s comments (Dr Kristina Sefc):

The reviewers of the manuscript are both impressed by the remarkable phylogeographic sampling underlying the study, and are mostly in agreement with the presented work. Please respond to the concerns (as raised by reviewer 1) regarding the use of lineage-through-time plots and Bayesian skyline plots in your analysis. I agree with the reviewer that it is preferable to omit analyses if assumptions are violated or if there’s a risk to arrive at spurious results. Regarding the BSP, if a literature survey (e.g. simulation studies testing effects of population structure on BSP analyses) suggests that the structure in your data might pose a problem, you could consider to
run analyses for the individual haplogroups separately and omit divergent lineages that include only few haplotypes. Please also provide the additional information and discussions as suggested by reviewer 1, and follow the comments of reviewer 2.

Reviewers' Comments to Author:

Reviewer: 1

This study provides a comparative phylogeographic analyses of Theodoxus gastropods from the Pontocaspian region and Southern Iran. The study is based on a good dataset with samples from quite a few distinct localities well spread across the study area. Although the manuscript is generally well written and most of the analyses/conclusions drawn from the analyses are ok, I do have some comments that need to be considered before this ms is ready for publication:

Major comments on analyses:

1. General comment on LTT-plots: LTT plots are typically NOT used for assessing demographic histories, but rather to infer/visualise the accumulation of distinct species/phylogenetic lineages through time. LTTs are not very useful for inferring intraspecific patterns of diversification as here LTTs would only make sense if you had all (or nearly all) the species' haplotypes included. As this is something you cannot be sure of, I strongly recommend removing LTTs from your ms.

2. Bayesian Skyline Plots: You have lots of phylogeographic structure in your data, in particular in the Iranian dataset, but also in the Pontocaspian data (even if you state that there is hardly and population genetic structuring). It is well known that BSPs might be heavily affected by population structure and there are already some studies available that tested the effect of population structure and sampling design on the performance of BSPs. I strongly recommend to check the relevant literature and adapt your analysis strategy and/or interpretation of the BSP patterns accordingly.

Minor comments:

1. Summary & Materials and methods: You write that your study is based on three amplified gene fragments. It's never mentioned you also sequenced them ...

2. I strongly suggest to provide some more information on the biology/ecology of the study species (habitat preference, specialist or generalist, ...) and the hydrological systems(s) (potential habitat discontinuities in the Caspian Sea that are potential parriers to dispersal, hydrological connection to other systems, ...). This info is important for interpreting the patterns you find.

3. Summary, Results & Discussion: Is the resolution (mutation rate) of your markers good enough to test for a potential influence of Pleistocene lake level fluctuations on population structure?

4. Introduction, line 47: How old is the Caspian Sea?

5. Introduction, line 52: In what sense is the Caspian Sea a unique model system to study the effect of palaeo-environmental changes on the phylogeographic structure in long-lived lake biota? It's certainly an excellent model system, but I can think of quite a few other systems that are equally well (or maybe even better) suited - e.g., just think of the East African Great Lakes.
6. M&M, page 2, line 48: Why did you use the BD tree prior? Any specific reasons for doing so? I'm asking because you are looking at a mixture of inter- and (mainly) intraspecific data and and tree prior choice might not be a trivial issue here.

7. M&M, page 2, lines 67-68: I guess you mean 95%HPD.

8. M&M, page 3, line 16+: Did you do a BSP (as stated in the ms) or an EPSP analysis? I'm asking because as far as I know the standard BSP analysis does not allow for multilocus data. Please clarify.

8. M&M, page 3, line 25: Did you really use the 95% confidence intervals or rather the 95% HPD intervals (which would be the standard way).

9. Results, page 3, lines 33-34: Three (in Iran) and on (Pontocaspian) divergence events? You cant' say it this way, especially since there is quite some intralineage divergence in the main Pontocaspian haplogroups. Please rephrase.

10. Results, page 3, lines 38+: It would be good to indicate the different morphospecies you refer to here also in the tree.

11. Discussion, page 4, line 1: Where is the evidence for the proposed catastrophic bottleneck? This is not based on your data, or is it? Please clarify.

12. Discussion, page 4, line2 & Conclusions, page 5, line 30: no phylogeographic structure in the Pontocaspian lineage? I disagree here. Judging from the network and Table S1, there's quite some structure, within the three main clades/refugial lineages. I think you can and should put much more emphasis on the phylogeographic patterns in the discussion, e.g. how haplogroup V got distributed over such a large area, from the Black Sea over rivers northwest of the Caspian Sea and the Caspian Sea to rivers southeast of the Caspian Sea, whereas the two subclades of haplogroup VI have very narrow distribtutions. By the way, why don't you consider these two subclades different haplogroups? This would make much more sense.

13. Discussion: Please discuss your findings also in light of what is known on habitat preferences of the species. E.g., are Theodoxus habitat specialists or rather generalists? And are there any obvious habitat discontinuities that might act as dispersal barriers?

Reviewer: 2

The manuscript reports on a phylogenetic reconstruction of the neritid genus Theodoxus from the Caspian Sea region. The authors have a range of samples from across the region, and used three genes (two mitochondrial, one nuclear) for their inferences. There was evidence of a general lack of reciprocal monophyly among species, but instead isolation by distance was present. The authors also undertook reconstructions of historic population sizes using their data, and speculate on potential links between climate and present day diversity. Overall this is a nice paper based on a hard-won dataset. The analyses appear appropriate to the questions being asked, and the results are fairly clear. Most of my comments relate to relatively minor issues of presentation.

P2, L40-43. I'm not convinced there is any evidence here of effects of low lake levels on population genetic structure. At best the evidence is that high stands may have provided opportunities for genetic homogenisation.
Author's Response to Decision Letter for (RSOS-190965.R0)

See Appendix A.

RSOS-190965.R1 (Revision)

Review form: Reviewer 1

Is the manuscript scientifically sound in its present form?
Yes

Are the interpretations and conclusions justified by the results?
No

Is the language acceptable?
Yes

Do you have any ethical concerns with this paper?
No
Have you any concerns about statistical analyses in this paper?
Yes

Recommendation?
Accept with minor revision (please list in comments)

Comments to the Author(s)
Most of my previous concerns have been addressed satisfactorily. Two issues, however, remain:

1. lines 175-177 (and comment 10 of the original review): ... no, I didn't misread the sentence. Note that each node in the tree represents a divergence event. In fact, you can delete this entire sentence as it essentially conveys the same information as the previous sentence (lines 174-175).

2. EBSPs: Yes, doing separate analyses for the two Pontocaspian phylogroups (now presented as Supplementary info) is the right way to deal with these data. And yes, for most of the Iranian phylogroups sample size/diversity is too small to allow for doing EBSP analyses. This, however, is no justification for pooling all these four pretty divergent phylogroups (cryptic species?) for a single EBSP analysis. You have virtually no intraphylogroup diversity in the Iranian samples and with the large interphylogroup divergence you inevitably get a signature of a drastic recent population decline in your EBSP. It's simply not possible to do reliable EBSP analyses with these Iranian samples. I suggest you have a look at previous studies that evaluated the performance of (E)BSPs in the presence of population structure. Specifically, in highly structured populations, drastic recent population size declines were observed (similar to what you found with your data) even though the sequences were simulated under a constant population size scenario. Two relevant studies that come to my mind are Heller et al. 2013 PLoS One & Grant 2015 J Hered. I think you’ll have to have a more critical look at your EBSPs, especially since these data are essential for your discussion/conclusions on conservation aspects in your gastropods.

Decision letter (RSOS-190965.R1)

27-Aug-2019

Dear Mr Sands:

Manuscript ID RSOS-190965.R1 entitled "Old lake vs. young taxa: a comparative phylodographic perspective on the evolution of Caspian Sea gastropods (Neritidae: Theodoxus)" which you submitted to Royal Society Open Science, has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter.

Please submit a copy of your revised paper before 19-Sep-2019. Please note that the revision deadline will expire at 00.00am on this date. If we do not hear from you within this time then it will be assumed that the paper has been withdrawn. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office in advance. We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Editors, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers.

To revise your manuscript, log into http://mc.manuscriptcentral.com/rsos and enter your
When submitting your revised manuscript, you must respond to the comments made by the referees and upload a file "Response to Referees" in "Section 6 - File Upload". Please use this to document how you have responded to the comments, and the adjustments you have made. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response.

In addition to addressing all of the reviewers' and editor's comments please also ensure that your revised manuscript contains the following sections before the reference list:

• Ethics statement
If your study uses humans or animals please include details of the ethical approval received, including the name of the committee that granted approval. For human studies please also detail whether informed consent was obtained. For field studies on animals please include details of all permissions, licences and/or approvals granted to carry out the fieldwork.

• Data accessibility
It is a condition of publication that all supporting data are made available either as supplementary information or preferably in a suitable permanent repository. The data accessibility section should state where the article's supporting data can be accessed. This section should also include details, where possible of where to access other relevant research materials such as statistical tools, protocols, software etc can be accessed. If the data have been deposited in an external repository this section should list the database, accession number and link to the DOI for all data from the article that have been made publicly available. Data sets that have been deposited in an external repository and have a DOI should also be appropriately cited in the manuscript and included in the reference list.

• Competing interests
Please declare any financial or non-financial competing interests, or state that you have no competing interests.

• Authors’ contributions
All submissions, other than those with a single author, must include an Authors’ Contributions section which individually lists the specific contribution of each author. The list of Authors should meet all of the following criteria; 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published.

All contributors who do not meet all of these criteria should be included in the acknowledgements.

We suggest the following format:
AB carried out the molecular lab work, participated in data analysis, carried out sequence alignments, participated in the design of the study and drafted the manuscript; CD carried out the statistical analyses; EF collected field data; GH conceived of the study, designed the study, coordinated the study and helped draft the manuscript. All authors gave final approval for publication.
• Acknowledgements
Please acknowledge anyone who contributed to the study but did not meet the authorship criteria.

• Funding statement
Please list the source of funding for each author.

Once again, thank you for submitting your manuscript to Royal Society Open Science and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Kind regards,
Alice Power
Editorial Coordinator
Royal Society Open Science
openscience@royalsociety.org

on behalf of Dr Kristina Sefc (Associate Editor) and Kevin Padian (Subject Editor)
openscience@royalsociety.org

Associate Editor Comments to Author (Dr Kristina Sefc):

Dear authors,
I’d like to thank you for addressing most of the concerns raised in the first round of review. The issue of pooling divergent lineages for the EBSPs still remains and actually represents a rather serious one since an important conclusion - regarding the decline in population size - may be based on an artifact in the analysis. Please see the reviewer’s comments for details. If the sampling / the data don't allow to analyse the demographic history of the Iranian populations, then you’ll have to consider dropping this part from the manuscript (or interpret the network structures of the mtDNA and ncDNA sequences verbally - less diversity in the mt than nc genomes within each phylogroup may indeed point to a recent bottleneck); or at least discuss the problem associated with pooling the divergent lineages and make clear that there’s a risk of a spurious result.
Sincerely, Kristina Sefc

Subject Editor Comments to Authors:

Thank you for addressing previous comments. Please note that the reviewer and the AE still feel there is a major issue to address. In your resubmission please make the necessary edits and address the comments. If the AE does not feel these are sufficiently addressed we will not be able to consider the manuscript further. Best wishes.

Reviewer comments to Author:
Reviewer: 1

Most of my previous concerns have been addressed satisfactorily. Two issues, however, remain:

1. lines 175-177 (and comment 10 of the original review): ... no, I didn't misread the sentence. Note that each node in the tree represents a divergence event. In fact, you can delete this entire sentence as it essentially conveys the same information as the previous sentence (lines 174-175).

2. EBSPs: Yes, doing separate analyses for the two Pontocaspian phylogroups (now presented as
Supplementary info) is the right way to deal with these data. And yes, for most of the Iranian phylogroups sample size/diversity is too small to allow for doing EBSP analyses. This, however, is no justification for pooling all these four pretty divergent phylogroups (cryptic species?) for a single EBSP analysis. You have virtually no intraphylogroup diversity in the Iranian samples and with the large interphylogroup divergence you inevitably get a signature of a drastic recent population decline in your EBSP. It's simply not possible to do reliable EBSP analyses with these Iranian samples. I suggest you have a look at previous studies that evaluated the performance of (E)BSPs in the presence of population structure. Specifically, in highly structured populations, drastic recent population size declines were observed (similar to what you found with your data) even though the sequences were simulated under a constant population size scenario. Two relevant studies that come to my mind are Heller et al. 2013 PLoS One & Grant 2015 J Hered. I think you'll have to have a more critical look at your EBSPs, especially since these data are essential for your discussion/conclusions on conservation aspects in your gastropods.

Author's Response to Decision Letter for (RSOS-190965.R1)

See Appendix B.

Decision letter (RSOS-190965.R2)

18-Sep-2019

Dear Mr Sands,

I am pleased to inform you that your manuscript entitled "Old lake vs. young taxa: a comparative phylogeographic perspective on the evolution of Caspian Sea gastropods (Neritidae: Theodoxus)" is now accepted for publication in Royal Society Open Science.

You can expect to receive a proof of your article in the near future. Please contact the editorial office (openscience_proofs@royalsociety.org and openscience@royalsociety.org) to let us know if you are likely to be away from e-mail contact -- if you are going to be away, please nominate a co-author (if available) to manage the proofing process, and ensure they are copied into your email to the journal.

Due to rapid publication and an extremely tight schedule, if comments are not received, your paper may experience a delay in publication.

Royal Society Open Science operates under a continuous publication model (http://bit.ly/cpFAQ). Your article will be published straight into the next open issue and this will be the final version of the paper. As such, it can be cited immediately by other researchers. As the issue version of your paper will be the only version to be published I would advise you to check your proofs thoroughly as changes cannot be made once the paper is published.
On behalf of the Editors of Royal Society Open Science, we look forward to your continued contributions to the Journal.

Kind regards,
Lianne Parkhouse
Royal Society Open Science
openscience@royalsociety.org

on behalf of Dr Kristina Sefc (Associate Editor) and Kevin Padian (Subject Editor)
openscience@royalsociety.org

Comments to the Author:

I’d like to thank the authors for the revised discussion of the results of the demographic analysis, which now include reference to potential methodological caveats.

Follow Royal Society Publishing on Twitter: @RSocPublishing
Follow Royal Society Publishing on Facebook:
https://www.facebook.com/RoyalSocietyPublishing.FanPage/
Read Royal Society Publishing's blog: https://blogs.royalsociety.org/publishing/
Appendix A

To the Chief Editor, Royal Society Open Science

Dear Prof. Jeremy Sanders CBE FRS
Cc: Prof. Kristina Sefc (Associate Editor), Prof. Kevin Padian (Subject Editor)

Reference Number: RSOS-190965

We very much appreciate the time and effort of the reviewers and the editors in assessing our paper and we are grateful for their useful comments. Below we list all critical comments and suggestions of the reviewers (in italics), together with our point-by-point replies (in blue). Attached at the very end is the revised version of the research paper with the 'Track Changes – All Markup' enabled, for easy reference of where changes have been made.

ASSOCIATE EDITOR (Dr Kristina Sefc):

Please respond to the concerns (as raised by reviewer 1) regarding the use of lineage-through-time plots and Bayesian skyline plots in your analysis. I agree with the reviewer that it is preferable to omit analyses if assumptions are violated or if there’s a risk to arrive at spurious results. Regarding the BSP, if a literature survey (e.g. simulation studies testing effects of population structure on BSP analyses) suggests that the structure in your data might pose a problem, you could consider to run analyses for the individual haplogroups separately and omit divergent lineages that include only few haplotypes.

We have removed the LTT plots and improved our Bayesian skyline analyses as suggested. Please see replies to reviewer 1: major comments 1 & 2 and minor comment 8.

REVIEWER: 1

MAJOR COMMENTS

1) General comment on LTT-plots: LTT plots are typically NOT used for assessing demographic histories, but rather to infer/visualise the accumulation of distinct species/phylogenetic lineages through time. LTTs are not very useful for inferring intraspecific patterns of diversification as here LTTs would only make sense if you had all (or nearly all) the species’ haplotypes included. As this is something you cannot be sure of, I strongly recommend removing LTTs from your ms.

We agree with the suggestion and have remove the LTT plots and all content related to them.

2) Bayesian Skyline Plots: You have lots of phylogeographic structure in your data, in particular in the Iranian dataset, but also in the Pontocaspian data (even if you state that there is hardly and population genetic structuring). It is
well known that BSPs might be heavily affected by population structure and there are already some studies available that tested the effect of population structure and sampling design on the performance of BSPs. I strongly recommend to check the relevant literature and adapt your analysis strategy and/or interpretation of the BSP patterns accordingly.

We are interested in the overall trends in both groups rather than the phylogroup specific trends. As most phylogroups have overlapping ranges (and thus likely have experienced similar environmental histories), the impact bias of a specific phylogroup on the overall pattern should be limited. Moreover, some phylogroups (specifically those in Iran) are too small to construct reliable skyline plots individually. Nevertheless we note the reviewer’s concern and the associate editor’s suggestions and have constructed EBS plots (also see minor comment 8 of reviewer 1) for all phylogroups where ESS values < 200 could be established (specifically phylogroups V and VI). These show little deviation from the overall trend from the Pontocaspian group. We have added the new additional results to the supplementary information and made alterations to the methods and results chapters. See: L150–168 & L199–208 and Appendix A, figure S1.

MINOR COMMENTS

1) Summary & Materials and methods: You write that your study is based on three amplified gene fragments. It’s never mentioned you also sequenced them ...

We have now added sequencing information in the ‘materials and methods’ section. See: L100-101.

2) I strongly suggest to provide some more information on the biology/ecology of the study species (habitat preference, specialist or generalist, ...) and the hydrological systems(s) (potential habitat discontinuities in the Caspian Sea that are potential barriers to dispersal, hydrological connection to other systems, ...). This info is important for interpreting the patterns you find.

Added as suggested. See: L65–68.

3) Summary, Results & Discussion: Is the resolution (mutation rate) of your markers good enough to test for a potential influence of Pleistocene lake level fluctuations on population structure?

This largely depends on the duration of the lake level fluctuation and is critically discussed (see L222–224 & L237–239). Moreover, there are a number of recently published studies on *Theodoxus*, using these three genes, to assess phylogeographic patterns and evolutionary history on the intraspecific level with success (i.e. showed to have sufficient resolution to detect a variety of major influences) [1–5]. As most *Theodoxus* intraspecific diversification occurs after the
Pliocene [6], this suggest they variable enough to detect the effect of major lake level changes during this period.

4) Introduction, line 47: How old is the Caspian Sea?

Added the age of the Caspian Sea as suggested. See: L37–38.

5) Introduction, line 52: In what sense is the Caspian Sea a unique model system to study the effect of palaeo-environmental changes on the phylogeographic structure in long-lived lake biota? It's certainly an excellent model system, but I can think of quite a few other systems that are equally well (or maybe even better) suited - e.g., just think of the East African Great Lakes.

We have changed 'unique', to 'suitable'. See: L44.

6) M&M, page 2, line 48: Why did you use the BD tree prior? Any specific reasons for doing so? I'm asking because you are looking at a mixture of inter- and (mainly) intraspecific data and tree prior choice might not be a trivial issue here.

While the 'coalescent population (constant and exponential) tree priors' are more orientated towards population level studies, we selected the birth death-prior due to having more than one species and not wanting to assume all Caspian Sea morphospecies belonged to the same phylogenetic species. Moreover, according to the latest published research [7], the choice of tree prior and molecular clock does not substantially affect phylogenetic patterns or diversification rates.

7) M&M, page 2, lines 67-68: I guess you mean 95%HPD.

Changed as suggested. Added ‘95%’ in front of HPD. See: L123–124.

8) M&M, page 3, line 16+: Did you do a BSP (as stated in the ms) or an EBSP analysis? I'm asking because as far as I know the standard BSP analysis does not allow for multi locus data. Please clarify.

Changed as suggested. All BS plots have now been reconstructed using the EBSP method and content/references adapted. See: L26, L150–168, L199–208 & L298–307, figure 5 + caption and Appendix A; figure S1.

9) M&M, page 3, line 25: Did you really use the 95% confidence intervals or rather the 95% HPD intervals (which would be the standard way).

This should have been 'HPD'. Changed as suggested. See: L120.
10) Results, page 3, lines 33-34: Three (in Iran) and one (Pontocaspian) divergence events? You can’t say it this way, especially since there is quite some intralineage divergence in the main Pontocaspian haplogroups. Please rephrase.

We feel the reviewer may have misread the sentence (See L174–176). We specifically state ‘between phylogroups’ and not ‘within and between phylogroups’. We are specifically describing the divergence events of I, II, III and IV from each other; and V and VI from each other.

11) Results, page 3, lines 38+: It would be good to indicate the different morphospecies you refer to here also in the tree.

Changed as suggested. We have adapted figure 3 to also indicate species names. See: figure 3 + caption.

12) Discussion, page 4, line 1: Where is the evidence for the proposed catastrophic bottleneck? This is not based on your data, or is it? Please clarify.

We partly agree with the reviewer, in that the early bottleneck discussed is somewhat speculative as the extended Bayesian skyline plots cannot cover this period of time to adequately test it. The reasoning is based on the presence of diverse fossil data in the absence of early molecular diversification in the phylogeny (all diversity stems from a single lineage over this period) and low nucleotide diversity (indicating a young species flock). This is a reasonable explanation given the data available and the results we have. However we do critically discuss our interpretation and note the speculativeness thereof. See: L212, L249, L256 & L327–328.

13) Discussion, page 4, line 2 & Conclusions, page 5, line 30: no phylogeographic structure in the Pontocaspian lineage? I disagree here. Judging from the network and Table S1, there’s quite some structure, within the three main clades/refugial lineages. I think you can and should put much more emphasis on the phylogeographic patterns in the discussion, e.g. how haplogroup V got distributed over such a large area, from the Black Sea over rivers northwest of the Caspian Sea and the Caspian Sea to rivers southeast of the Caspian Sea, whereas the two subclades of haplogroup VI have very narrow distributions. By the way, why don’t you consider these two subclades different haplogroups? This would make much more sense.

We agree with the reviewer that more discussion needs to be included regarding the broad distribution of phylogroups V and VI. As such we have added some discussion about the effects of high stands and how they may have driven homogeneity of
haplotypes across the system (also see response to reviewer 2, minor comment 1). However the two ‘subclades’ in phylogroup VI are not supported by posterior probabilities in the phylogeny and nor are the differences conserved across the AMOVAs of all genes. As such, discussion of ‘narrow ranges’ or the general patterns within subclades of phylogroup V or VI would be very dubious.

14) Discussion: Please discuss your findings also in light of what is known on habitat preferences of the species. E.g., are Theodoxus habitat specialists or rather generalists? And are there any obvious habitat discontinuities that might act as dispersal barriers?

We have added discussion on this topic. See: L241–247.

REVIEWER: 2

MINOR COMMENTS

1) P2, L40-43. I’m not convinced there is any evidence here of effects of low lake levels on population genetic structure. At best the evidence is that high stands may have provided opportunities for genetic homogenisation.

We agree with the reviewer that there is little evidence to support low stands having any effect on population structure; indeed that is primarily our outcome. However, following the suggestion, we have adapted our text to put more focus on the homogenising effects of high stands. See: L30–31, L240–247 & L322–324.

2) P2, L58. use “only a small number”

Changed ‘low’ to ‘small’ as suggested. See: L51.

3) P3, L23. Use “should possess similar”

Changed ‘mimic’ to ‘possess’ as suggested. See: L85.

4) P3, L35. I’m not convinced you give any insights into plasticity here (it doesn’t appear to be mentioned elsewhere in the paper). Indeed, these data alone do not tell us anything about the plasticity, or otherwise, of these phenotypes.

We agree. We have removed the sub-aim to ‘provide perspective on phenotypic plasticity’. See: L86.
5) **P3, L41.** I am not familiar with the tests you used for saturation (it would be helpful if you named and cited the specific test used), but typically a $p < 0.05$ is considered statistically significant.

There is no specific name for the test for saturation, however we have added in an additional reference as requested. Regarding the level of significance: we gave the $p$ values as found by the analyses (i.e. $p < 0.001$) rather than the level of significance. See: L103–105.

6) **P3, L46.** You say taxa, but this is a pleural of taxon (a group of one or more populations of an organism). Surely you mean individual.

Changed ‘taxa’ to ‘individuals’ as suggested. See: L110.

7) **P3, L~60.** 50% burn-in ‘removed’?

Added ‘removed’ as suggested. See: L128.

8) **P4, L3.** No further burn-in ‘removed’.

Added ‘removed’ as suggested. See: L131.

9) **P4, L32. It is not clear how your “phylogroups” were delimited.**

We have now included a small explanation to the methods of how the phylogroups were delimited. See: L131–132.

10) **P5, L1. I’m not sure you found any evidence for a catastrophic bottleneck. The data are, I feel, being over interpreted.**

See reply to reviewer 1, minor comment 12.

11) **P6, L6. RE synonymy.** Your analysis is based on effectively 2 independent loci (both mtDNA genes are linked). Could such patterns arise from parallel evolution?

Parallel evolution is unlikely given its rarity, especially considering that the amount of mutations that have occurred and the extended geological time over which these took place. Moreover, although not discussed or shown, the majority of key morphological characters are conserved, such as those related to the operculum (which is often used to distinguish species). As this is not really in the scope of this study, it will be addressed in a follow-up article currently in preparation by the authors reviewing the taxonomy and morphology of all extant *Theodoxus* spp.
12) P6, L35. “contrasts with the fossil evidence”

Added ‘with’ as suggested. See: L325

13) P9, L37. Zookeys volume absent.

Volume number now added. See: reference 70.

14) P11, Table. All your p-values for your isolation by distance tests are identical?

It is not clear to me how you tested for isolation by distance, but since Mantel tests are now suggested to be flawed you should consider distance-based redundancy analysis in the vegan R package. We have incorporated the distance-based redundancy analysis in R as suggested and adjusted the text accordingly. See: L25, L142–147 & L194–196, table 2 and Appendix B.

15) P13-15, Figures. Your colours are indistinguishable. I suggest you reconsider these, perhaps use letter and colour combinations.

Changed as suggested. See: figures 2-4.

16) P17, L37. What is “intraspecific diversification” in this sense? Why should it only start at this point? It doesn’t make any logical sense to me.

‘Intraspecific diversification’ refers here to the date at which the skyline plots begin (i.e. the starting point of the observed diversification in the group). We have adapted the caption of figure 5 to make this clear.

REFERENCE LIST FOR REPLY COMMENTS

1. Bunje PME. 2007 Fluvial range expansion, allopatry, and parallel evolution in a Danubian snail lineage (Neritidae: Theodoxus). Biol. J. Linn. Soc. 90, 603–617. (doi:10.1111/j.1095-8312.2007.00750.x)

2. Bunje PME. 2005 Pan-European phylogeography of the aquatic snail Theodoxus fluviatilis (Gastropoda: Neritidae). Mol. Ecol. 14, 4323–4340. (doi:10.1111/j.1365-294X.2005.02703.x)

3. Fehér Z, Zettler ML, Bozsó M, Szabó K. 2009 An attempt to reveal the systematic relationship between Theodoxus prevostianus (C. Pfeiffer, 1828) and Theodoxus danubialis (C. Pfeiffer, 1828) (Mollusca, Gastropoda, Neritidae). Mollusca 27, 95–107.

4. Fehér Z, Albrecht C, Major Á, Sereda S V., Krízsik V. 2012 Extremely low genetic diversity in the endangered striped nerite, Theodoxus transversalis
Mollusca, Gastropoda, Neritidae) – a result of ancestral or recent effects? *North. West. J. Zool.* 8, 300–307.

5. Gergs R, Koester M, Grabow K, Schöll F, Thielsh A, Martens A. 2015 *Theodoxus fluviatilis* re-establishment in the River Rhine: A native relict or a cryptic invader? *Conserv. Genet.* 16, 247–251. (doi:10.1007/s10592-014-0651-7)

6. Sands AF, Sereda SV, Stelbrink B, Neubauer TA, Lazarev S, Wilke T, Albrecht C. 2019 Contributions of biogeographical functions to species accumulation may change over time in refugial regions. *J. Biogeogr.* 46, 1274–1286.

7. Sarver BAJ, Pennell MW, Brown JW, Keeble S, Hardwick KM, Sullivan J, Harmon LJ. 2015 The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. *PeerJ* 7, e6334. (doi:10.7717/peerj.6334)

8. Krijgsman W et al. 2019 Quaternary time scales for the Pontocaspian domain: Interbasinal connectivity and faunal evolution. *Earth-Science Rev.* 188, 1–40. (doi:10.1016/j.earscirev.2018.10.013)

9. Chen JL, Pekker T, Wilson CR, Tapley BD, Kostianoy AG, Cretaux J-F, Safarov ES. 2017 Long-term Caspian Sea level change. *Geophys. Res. Lett.* 44, 6993–7001. (doi:10.1002/2017GL073958)

10. Kosarev AN. 2005 Physico-geographical conditions of the Caspian Sea. In *The Caspian Sea Environment* (eds AG Kostianoy, AN Kosarev), pp. 5–31. Springer, Berlin, Heidelberg. (doi:10.1007/698_5_002)

11. Forte AM, Cowgill E. 2013 Late Cenozoic base-level variations of the Caspian Sea: A review of its history and proposed driving mechanisms. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 386, 392–407. (doi:10.1016/J.PALAEO.2013.05.035)

12. Wesselingh FP et al. 2019 Mollusc species from the Pontocaspian region – an expert opinion list. *Zookeys* 827, 31–124. (doi:10.3897/zookeys.827.31365)

13. Audzijonyte A, Baltroñaitė L, Väinõlá R, Arbačiauskas K. 2015 Migration and isolation during the turbulent Ponto-Caspian Pleistocene create high diversity in the crustacean *Paramysis lacustris*. *Mol. Ecol.* 24, 4537–4555. (doi:10.1111/mec.13333)

14. Audzijonyte A, Wittmann KJ, Ovcarenko I, Väinõlá R. 2009 Invasion phylogeography of the Ponto-Caspian crustacean *Limnornysis benedeni* dispersing across Europe. *Divers. Distrib.* 15, 346–355. (doi:10.1111/j.1472-4642.2008.00541.x)

15. Audzijonyte A, Daneliya ME, Väinõlá R. 2006 Comparative phylogeography of Ponto-Caspian mysid crustaceans; isolation and exchange among dynamic inland sea basins. *Mol. Ecol.* 15, 2969–2984. (doi:10.1111/j.1365-294X.2006.03018.x)

16. Brown JE, Stepien CA. 2008 Ancient divisions, recent expansions: phylogeography and population genetics of the round goby *Apollonia melanostoma*. *Mol. Ecol.* 17, 2598–2615. (doi:10.1111/j.1365-294X.2008.03777.x)

17. Cristescu MEA, Hebert PDN, Onciu TM. 2003 Phylogeography of Ponto-Caspian crustaceans: a benthic-planktonic comparison. *Mol. Ecol.* 12, 985–996. (doi:10.1046/j.1365-294X.2003.01801.x)

18. Kotlik P, Marková S, Choleva L, Bogutskaya NG, Ekmeççi FG, Ivanova PP. 2008 Divergence with gene flow between Ponto-Caspian refugia in an anadromous cyprinid *Rutilus frisii* revealed by multiple gene phylogeography.
19. Levin BA et al. 2017 Phylogeny and phylogeography of the roaches, genus Rutilus (Cyprinidae), at the Eastern part of its range as inferred from mtDNA analysis. Hydrobiologia 788, 33–46. (doi:10.1007/s10750-016-2984-3)

20. Parvizi E, Naderloo R, Keikhosravi A, Solhjouy-Fard S, Schubart CD. 2018 Multiple Pleistocene refugia and repeated phylogeographic breaks in the southern Caspian Sea region: Insights from the freshwater crab Potamon ibericum. J. Biogeogr. 45, 1234–1245. (doi:10.1111/jbi.13195)

21. Ruban GI, Khолодова М. V., Калмыков VA, Сорокин PA. 2008 Morphological and molecular genetic study of the Persian sturgeon Acipenser persicus Borodin (Acipenseridae) taxonomic status. J. Ichthyol. 48, 891–903. (doi:10.1134/S0032945208100068)

22. Neubauer TA, Harzhauser M, Kroh A, Georgopoulou E, Mandic O. 2015 A gastropod-based biogeographic scheme for the European Neogene freshwater systems. Earth-Science Rev. 143, 98–116. (doi:10.1016/J.EARSCIREV.2015.01.010)

23. Neubauer TA, Harzhauser M, Georgopoulou E, Kroh A, Mandic O. 2015 Tectonics, climate, and the rise and demise of continental aquatic species richness hotspots. Proc. Natl. Acad. Sci. U. S. A. 112, 11478–83. (doi:10.1073/pnas.1503992112)

24. Van Bocxlaer B. 2017 Paleoecological insights from fossil freshwater mollusks of the Kanapoi Formation (Omo-Turkana Basin, Kenya). J. Hum. Evol. (doi:10.1016/J.JHEVOL.2017.05.008)

25. von Rintelen T, Stelbrink B, Marwoto RM, Glaubrecht M. 2014 A snail perspective on the biogeography of Sulawesi, Indonesia: Origin and intra-island dispersal of the viviparous freshwater gastropod Tylomelania. PLoS One 9, e98917. (doi:10.1371/journal.pone.0098917)

26. Wilke T, Albrecht C, Anistratenko VV, Sahin SK, Yildirim MZ. 2007 Testing biogeographical hypotheses in space and time: faunal relationships of the putative ancient Lake Egirdir in Asia Minor. J. Biogeogr. 34, 1807–1821. (doi:10.1111/j.1365-2699.2007.01727.x)

27. Bandel K. 2001 The history of Theodoxus and Neritina connected with description and systematic evaluation of related Neritimorpha (Gastropoda). Mitteilungen aus dem Geol. Inst. der Univ. Hambg. 85, 65–164.

28. Karpinsky MG. 2002 Ecology of the Benthos of the Middle and Southern Caspian. Moscow: VNIRO.

29. Anistratenko VV, Zettler ML, Anistratenko OY. 2017 On the taxonomic relationship between Theodoxus pallasi and T. astrachanicus (Gastropoda: Neritidae) from the Ponto-Caspian region. Arch. für Molluskenkd. 146, 213–226. (doi:10.1127/arch.moll/146/213-226)

30. Palumbi SR, Martin A, Romano S, McMillan WO, Stice L, Grabowski G. 1991 The Simple Fool’s Guide to PCR. Privately published document compiled by S. Palumbi. Dept. Zoology, University of Hawaii, Honolulu, HI, 96822, USA.

31. Jarman SN, Ward RD, Elliott NG. 2002 Oligonucleotide primers for PCR amplification of coelomate introns. Mar. Biotechnol. 4, 347–355. (doi:10.1007/s10126-002-0029-6)

32. Biomatters Ltd. 2017 Geneious.

33. Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003 An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26, 1–7. (doi:10.1016/S1055-7903(02)00326-3)
34. Xia X. 2013 DAMBE5: A comprehensive software package for Data Analysis in Molecular Biology and Evolution. Mol. Biol. Evol. 30, 1720–1728. (doi:10.1093/molbev/mst064)

35. Xia X, Xie Z. 2001 DAMBE: Software package for Data Analysis in Molecular Biology and Evolution. J. Hered. 92, 371–373. (doi:10.1093/jhered/92.4.371)

36. Bouckaert RR, Drummond AJ. 2017 bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17, 42. (doi:10.1186/s12862-017-0890-6)

37. Bouckaert RR, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014 BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10, e1003537. (doi:10.1371/journal.pcbi.1003537)

38. Miller MA, Pfeiffer W, Schwartz T. 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In 2010 Gateway Computing Environments Workshop, GCE 2010, pp. 1–8. IEEE. (doi:10.1109/GCE.2010.5676129)

39. Rambaut A, Drummond AJ, Suchard M. 2018 Trancer v1.7.

40. Clement M, Posada D, Crandall KA. 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1659. (doi:10.1046/j.1365-294x.2000.01020.x)

41. Stephens M, Donnelly P. 2003 A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73, 1162–1169. (doi:10.1086/379378)

42. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017 DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302. (doi:10.1093/molbev/msx248)

43. Excoffier L, Lischer HEL. 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. (doi:10.1111/j.1755-0998.2010.02847.x)

44. Holm S. 1979 A simple sequentially rejective multiple test procedure on JSTOR. Scand. J. Stat. 6, 65–70.

45. Legendre P, Anderson MJ. 1999 Distance-based redundancy analysis: testing multiple species responses in multifactorial ecological experiments. Ecol. Monogr. 69, 1–24. (doi:10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2)

46. Legendre P, Fortin M-J. 2010 Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Mol. Ecol. Resour. 10, 831–844. (doi:10.1111/j.1755-0998.2010.02866.x)

47. Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D. 2019 vegan: Community ecology R package.

48. R Core Team. 2018 R: A language and environment for statistical computing.

49. Hindley JA, Graham BA, Pulgarin-R. PC, Burg TM. 2018 The influence of latitude, geographic distance, and habitat discontinuities on genetic variation in a high latitude montane species. Sci. Rep. 8, 11846. (doi:10.1038/s41598-018-29982-7)

50. Peakall R, Smouse PE. 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28, 2537–2539. (doi:10.1093/bioinformatics/bts460)
51. Andrusov N. 1923 Apsheronskiy yarus. Tr. Geol. Kom. novaya seriya **110**, 1–294.

52. Nalivkin DV. 1914 Mollyuski gory bakinskogo yarusa. Tr. Geol. Kom. novaya seriya **116**, 1–32.

53. Neubauer TA, van de Velde S, Yanina T, Wesselingh FP. 2018 A late Pleistocene gastropod fauna from the northern Caspian Sea with implications for Pontocaspian gastropod taxonomy. *Zookeys*, 43–103. (doi:10.3897/zookeys.770.25365)

54. van de Velde S, Yanina TA, Neubauer TA, Wesselingh FP. 2019 The Late Pleistocene mollusk fauna of Selitrennoye (Astrakhan province, Russia): A natural baseline for endemic Caspian Sea faunas. *J. Great Lakes Res.* (doi:10.1016/J.JGLR.2019.04.001)

55. D’Hont A, Gittenberger A, Hendriks AJ, Leuven RSEW. 2018 Drivers of dominance shifts between invasive Ponto-Caspian dreissenids *Dreissena polymorpha* (Pallas, 1771) and *Dreissena rostriformis bugensis* (Andrusov, 1897). *Aquat. Invasions* **13**, 449–462. (doi:10.3391/ai.2018.13.4.03)

56. Pauli N-C, Briski E. 2018 Euryhalinity of Ponto-Caspian invaders in their native and introduced regions. *Aquat. Invasions* **13**, 439–447. (doi:10.3391/ai.2018.13.4.02)

57. Aladin NV, Filipov AA, Plotnikov IS, Orlova MI, Williams WD. 1998 Changes in the structure and function of biological communities in the Aral Sea, with particular reference to the northern part (Small Aral Sea), 1985–1994: A review. *Int. J. Salt Lake Res.* **7**, 301–343. (doi:10.1023/A:1009009924839)

58. Kangas P, Skoog G. 1978 Salinity tolerance of *Theodoxus fluviatilis* (Mollusca, Gastropoda) from freshwater and from different salinity regimes in the Baltic Sea. *Estuar. Coast. Mar. Sci.* **6**, 409–416. (doi:10.1016/0302-3524(78)90131-7)

59. Birstein VJ, Ruban G, Ludwig A, Doukakis P, DeSalle R. 2005 The enigmatic Caspian Sea Russian sturgeon: How many cryptic forms does it contain? *Syst. Biodivers.* **3**, 203–218. (doi:10.1017/S1477200005001647)

60. Ruban GI, Kholodova M V., Kalmykov VA, Sorokin PA. 2008 Morphological and molecular genetic study of the Persian sturgeon *Acipenser persicus* Borodin (Acipenseridae) taxonomic status. *J. Ichthyol.* **48**, 891–903. (doi:10.1134/S0032945208100068)

61. Parr TD, Tait RD, Maxon CL, Newton FC, Hardin JL. 2007 A descriptive account of benthic macrofauna and sediment from an area of planned petroleum exploration in the southern Caspian Sea. *Estuar. Coast. Shelf Sci.* **71**, 170–180. (doi:10.1016/J.ECSS.2006.07.018)

62. Dumont HJ. 1998 The Caspian Lake: History, biota, structure, and function. *Limnol. Oceanogr.* **43**, 44–52. (doi:10.4319/lo.1998.43.1.0044)

63. Svitoch AA. 2012 The Caspian Sea shell during the Pleistocene regressive epochs. *Oceanology* **52**, 526–539. (doi:10.1134/S0001437012030113)

64. Nevesskaja LA. 2001 History of Lymnocardiinae (Bivalvia, Cardiidae). *Paleontol. J.* **35**, S147–S217.

65. Tagieva EN, Aleskerov BD, Veliev SS, Kuliev MY. 2013 Transformation of vegetation cover on the territory of Azerbaijan during the Pleistocene. *Geogr. Nat. Resour.* **34**, 395–400. (doi:10.1134/S187537281304015X)

66. Zastrozhnov A et al. 2018 Biostratigraphical investigations as a tool for palaeoenvironmental reconstruction of the Neopleistocene (Middle-Upper Pleistocene) at Kosika, Lower Volga, Russia. *Quat. Int.*
67. Abdullayev N. 2000 Seismic stratigraphy of the Upper Pliocene and Quaternary deposits in the South Caspian Basin. *J. Pet. Sci. Eng.* 28, 207–226. (doi:10.1016/S0920-4105(00)00079-6)

68. Eynoddin EH et al. 2017 Assessment of relative active tectonics in the Bozgoush basin (SW of Caspian Sea). *Open J. Mar. Sci.* 07, 211–237. (doi:10.4236/ojms.2017.72016)

69. Javadi Mosavi E, Arian M. 2015 Tectonic geomorphology of Atrak River, NE Iran. *Open J. Geol.* 5, 106–114. (doi:10.4236/ojg.2015.53010)

70. Sharkov E, Lebedev V, Chugayev A, Zabarinyska L, Rodnikov A, Sergeeva N, Safonova I. 2015 The Caucasian-Arabian segment of the Alpine-Himalayan collisional belt: Geology, volcanism and neotectonics. *Geosci. Front.* 6, 513–522. (doi:10.1016/J.GSF.2014.07.001)

71. Sokhadze G, Floyd M, Godoladze T, King R, Cowgill ES, Javakhishvili Z, Hahubia G, Reiling R. 2018 Active convergence between the Lesser and Greater Caucasus in Georgia: Constraints on the tectonic evolution of the Lesser–Greater Caucasus continental collision. *Earth Planet. Sci. Lett.* 481, 154–161. (doi:10.1016/J.EPSL.2017.10.007)

72. Salzburger W, Van Bocxlaer B, Cohen AS. 2014 Ecology and evolution of the African Great Lakes and their faunas. *Annu. Rev. Ecol. Evol. Syst.* 45, 519–545. (doi:10.1146/annurev-ecolsys-120213-091804)

73. Schultheiß R, Wilke T, Jörgensen A, Albrecht C. 2011 The birth of an endemic species flock: demographic history of the *Bellamya* group (Gastropoda, Viviparidae) in Lake Malawi. *Biol. J. Linn. Soc.* 102, 130–143. (doi:10.1111/j.1095-8312.2010.01574.x)

74. Schultheiß R, Van Bocxlaer B, Wilke T, Albrecht C. 2009 Old fossils–young species: evolutionary history of an endemic gastropod assemblage in Lake Malawi. *Proc. R. Soc. B Biol. Sci.* 276, 2837–2846. (doi:10.1098/rspb.2009.0467)

75. Starmühlner F, Edlauer A. 1957 Ergebnisse der Österreichischen Iran-Expedition 1949/50 (Mit Berücksichtigung der Ausbeute der Österreichischen Iran-Expedition 1956): Beiträge zur Kenntnis der Molluskenfauna des Iran. *Sitzungsberichte der österreichischen Akad. der Wissenschaften, Math. Klasse, Abteilung /166*, 435–494.

76. Glöer P, Pešić V. 2012 The freshwater snails (Gastropoda) of Iran, with descriptions of two new genera and eight new species. *Zookeys* 219, 11–61. (doi:10.3897/zookeys.219.3406)

77. Dunker W. 1861 Beschreibung neuer Mollusken. *Malakozool. Blätter* 8, 35–45.

78. Sepehr M, Kharazian A, Nasab HV, Jayez M, Hashemi S-M, Nateghi A, Abdollahi A, Kharazian MA, Berillon G. 2019 Anzo: The first evidence of Paleolithic cave sites in the northern margin of the Iranian Central Desert, Semnan, Iran. *Archaeology* 2019, 1–5. (doi:10.5923/j.archaeology.20190701.01)

79. Vahdati Nasab H, Hashemi M. 2016 Playas and Middle Paleolithic settlement of the Iranian Central Desert: The discovery of the Chah-e Jam Middle Paleolithic site. *Quat. Int.* 408, 140–152. (doi:10.1016/J.QUAINT.2015.11.117)

80. Wang X, Wei H, Taheri M, Khormali F, Danukalova G, Chen F. 2016 Early Pleistocene climate in western arid central Asia inferred from loess-palaeosol sequences. *Sci. Rep.* 6, 20560. (doi:10.1038/srep20560)

81. Carolin S, Walker RT, Henderson GM, Maxfield L, Ersek V, Sloan A, Talebian
M. Fattahi M, Nezamdoust J. 2015 Decadal-scale climate variability on the Central Iranian Plateau spanning the so-called 4.2 ka BP drought event. Am. Geophys. Union, Fall Meet. 2015, Abstr. id. PP21D-07

82. Zobeidi T, Yazdanpanah M, Forouzani M, Khosravipour B. 2016 Climate change discourse among Iranian farmers. Clim. Change 138, 521–535. (doi:10.1007/s10584-016-1741-y)

83. Lattuada M, Albrecht C, Wilke T. 2019 Differential impact of anthropogenic pressures on Caspian Sea ecoregions. Mar. Pollut. Bull. 142, 274–281. (doi:10.1016/J.MARPOLBUL.2019.03.046)

84. Sands AF, Neubauer TA, Nasibi S, Fasihi Harandi M, Anistratenko V V., Wilke T, Albrecht C. 2019 Data from: Old lake vs. young taxa: a comparative phylogeographic perspective on the evolution of Caspian Sea gastropods (Neritidae: Theodoxus). Dryad Digital Repository. (doi:10.5061/dryad.mn15f80)
Old lake vs. young taxa: a comparative phylogeographic perspective on the evolution of Caspian Sea gastropods (Neritidae: Theodoxus)

Arthur F. Sands1*, Thomas A. Neubauer1,2, Saeid Nasibi1, Majid Fashi Harandi3, Vitaliy V. Anistratenko4, Thomas Wilke1, Christian Albrecht1

1 Department of Animal Ecology and Systematics, Justus Liebig University, Heinrich-Buff-Ring 26–32 (IFZ), 35392 Giessen, Germany
2 Naturalis Biodiversity Centre, P.O. Box 9517, 2300 RA Leiden, The Netherlands
3 Research Center for Hydatid Disease in Iran, School of Medicine, Kerman University of Medical Sciences, Kerman 76169-14115, Iran
4 I.I. Schmalhausen Institute of Zoology of the National Academy of Sciences of Ukraine, B. Khmelnitsky St., 15, Kiev, 01030, Ukraine

Keywords: Ancient lakes, Pontocaspian, evolution, lake level fluctuations, salinity, molluscs.

1. Summary
The Caspian Sea has been a highly dynamic environment throughout the Quaternary and witnessed major oscillations in lake level, which were associated with changes in salinity and habitat availability. Such environmental pressures are considered to drive strong phylogeographic structure in species by forcing populations into suitable refugia. However, little is actually known on the effect of lake level fluctuations in the Caspian Sea on its aquatic biota. We compared the phylogeographic patterns of the aquatic neritid snail genus Theodoxus across the Pontocaspian region with refugial populations in southern Iran, where populations have persisted in refugia. Three amplified gene fragments were used to determine relationships and divergence times between sampled populations in both groups. A dated phylogeny and statistical haplotype networks were generated in conjunction with analyses of molecular variance and calculations of isolation by distance, using distance-based redundancy analyses. Extended Bayesian skyline and lineage-through-time plots were constructed to assess demographic history. Compared to the Iranian populations, we found little phylogeographic structure for the Pontocaspian Theodoxus group, with more recent diversification, homogeneity of haplotypes across the Pontocaspian region and a relatively stable demographic history since the Middle Pleistocene. Our results argue against a strong influence of lake level fluctuations in Caspian Sea low stands on population structure, at least post the Last Glacial Maximum, whereas high stands may have increased the dispersal possibilities and homogenisation of haplotypes across the Pontocaspian region during this time. However, prior to this period a more dramatic low stand in the Caspian Sea around a million years ago may have had caused reduction of Theodoxus diversity to a single lineage in the region. In addition, our results provide new insights into Theodoxus taxonomy and outlooks for regional conservation.

2. Introduction
The Caspian Sea is one of the largest and oldest lakes on the planet. Barring episodic overflow events, the Caspian Sea has been an isolated endorheic basin since at least the early Pliocene, around 5.3 million years ago (Ma) [8]. Today, the Caspian Sea and its catchment cover an area of approximately 3.500.000 km² [9,10]. It is renowned for dramatic historical lake level fluctuations and salinity shifts, primarily as a consequence of glacial cycles during the Quaternary [8,11]. Moreover, temporary connections occurred between the Caspian Sea and other Pontocaspian basins during high stands (i.e. the Aral, Azov and Black seas [8]) during high stands. These changes are hypothesized to have had a major effect on the evolutionary history of its aquatic biota, such as the highly endemic gastropod fauna [12]. Such fluctuations make the Caspian Sea a unique unuitable model system to study the effect of palaeo-environmental change on phylogeographic structure in long-lived lake biota.

The recurrent fluctuations in salinity and lake level during the Quaternary may have created periodic temporal refugia. Isolated Caspian sub-basins during low stands or spring sources and riverine systems during periods of increased salinity are potential candidate refugia. The restricted gene flow across such refugial populations are expected to generate strong phylogeographic structure within the species as a consequence of isolated evolutionary histories. Compared to the large extent of the Caspian drainage basin, surprisingly only a low number of phylogeographic and population level studies exist for aquatic taxa. These studies primarily focus on fish and crustaceans in a broader Pontocaspian context (see [13–21]). Interestingly, while many of these studies indicate

*Author for correspondence (Arthur.F.Sands@allzool.bio.uni-giessen.de).
strong geographic structure across the entire Pontocaspian region, most show a lack of visible intraspecific structure in the Caspian drainage basin [14,16–18,21]. Although this may be a genuine evolutionary pattern (e.g. due to a lack of genetic breaks), there may be a bias. Often, taxa were selected which are easily dispersed, or the sampling design was rather narrow across the Caspian basin itself.

Gastropods can make good candidates for documenting responses to environmental changes as they often have limited dispersal capabilities, fast generation turnover, good fossil preservation and are sensitive to ecological shifts [22–26]. Selecting a widespread Pontocaspian gastropod taxon, such as the neritid genus *Theodoxus*, which is both diocious and lacks a planktotrophic larval stage that facilitates dispersal [27], may not only improve our understanding of the effects of major salinity and lake level changes on the evolutionary history of Pontocaspian gastropods, but also improve our understanding of the phylogeography and the existence of refugia in the Caspian system as a whole. Most *Theodoxus* spp. are generalists and occur in lakes, rivers, estuaries and springs, although they depend on well oxygenated water and environments with hard substrate, where they can graze on algae (i.e. rocks and stones, shell beds or aquatic plants) [27]. Due to the abundance of these environments in the Pontocaspian region, *Theodoxus* has become a common component of its malaco fauna [12,28]. Four species have been considered endemic to the Pontocaspian region and share a presence in the Caspian basin, *T. astrachanicus* Storobogatov in Storobogatov et al., 1994, *T. major* Isotel, 1865, *T. pallasi* Lindholm, 1924 and *T. schultzi* (Grimm, 1877) (figure 1) [6,12]. Despite considerable phenotypic variability (figure 1), recent studies based on morphology [29] and molecular analyses [6] have led to speculations that these four may represent a single species [12]. Here, we use this group as a model taxon to determine the effects of historical lake level and salinity fluctuations on major limnological events in the Caspian Sea on the phylogeographic structure of Pontocaspian gastropods. We compare their phylogeographic structure to a potential sister group of *Theodoxus* occurring in isolated spring systems in southern Iran, feeding the remnants of endorheic basins in southern Iran (as a model for gastropods that have persisted in refugia). To reach this objective we 1) sampled *Theodoxus* across the Pontocaspian and southern Iranian drainages (i.e. lakes, springs, rivers, streams) to carry out phylogenetic analyses to determine population structure, and 3) tested for demographic expansion/contraction events through time, which may help to identify bottleneck events.

We hypothesise that repeated fluctuations in lake level and Caspian Sea low stands with increased salinity (reflecting in the phenotypic variability, figure 1) resulted in multiple isolated refugia across the Caspian Sea and its catchments, and thus a high degree of population structure. As such, Pontocaspian and Iranian *Theodoxus* should possess similar population structures. Additionally, as the taxonomy of some of these species is poorly resolved, we provide a molecular perspective on species identities and phenotypic plasticity. The outcomes of this research should identify the timing and extent of how major salinity and lake level changes may have affected the evolutionary histories of native Pontocaspian aquatic gastropods and identify Pontocaspian refugia (particularly across the Caspian Sea drainage network). Moreover, we discuss how our results may help resolve taxonomic uncertainties and what the implications may mean for the conservation of Pontocaspian taxa.

### 3. Materials and Methods

#### 3.1 Sample collection and laboratory protocols

*Theodoxus* specimens were collected and stored following Sands et al. [6]. We included the genetic data of either five or ten specimens per location to allow for robust analyses (table 1; figure 2). Sands’ et al. [6] protocols were again followed to extract and amplify two mtDNA fragments; cytochrome c oxidase subunit I (COI) and 16S rRNA (16Sr) and one nDNA intron fragment, ATP synthetase subunit alpha (ATPα). Total genomic DNA was extracted from foot tissue of the snails using a DNeasy Blood and Tissue kit (QIAGEN, Hilden, Germany) and amplified using the primers TheoS1F and TheoS1R for COI [6], 16Sr-L and 16Sr-H for 16S [30], and ATPs1aF and ATPs1aR for ATPα [31] – for primer sequences and PCR conditions see Sands et al. [6].

Purification and bidirectional *Sequencing* of the amplified gene fragments were carried out by LGC Ltd. (Berlin, Germany). Where needed, we incorporated published GenBank sequences of individuals where all three DNA fragments were available. Sequence ends were trimmed and aligned in Genious 10.1.2 [32] using the Genious alignment algorithm. To test for saturation over 1,000 replicates [33], we used DAMBE 5.0.52 [34,35]. None of the three molecular datasets showed any significant degree of saturation (COI, 16S and ATPα; p < 0.001).

#### 3.2 Phylogenetic analyses

To investigate the relationships among all *Theodoxus* specimens, *delimit phylogenies* and provide divergence date estimates between and within the Pontocaspian and southern Iranian groups, a dated Bayesian phylogeny was constructed. The phylogeny combined all amplified gene fragments (1,609 base pairs in total) for 222 taxa (individuals) (211 sequenced for the first time; table 1).

First, a log-norm relaxed clock was set for all gene datasets and a birth-death tree prior was selected. *bModelTest v1.3.2* [36], as implemented in BEAST v2.5.2 [37], was used to determine the best-fit model for each gene set. *TN93 (121131)* was determined the optimal model for COI and ATPα and a variant of the HKY (with an additional group for the rates τd and τp; 121323) for 16S. Second, sequences derived from *T. jordani* (Sowerby, 1836) and the more distantly related *T. transversalis* (Pfeiffer, 1828) were used to root the phylogeny (see [6]). As fossil dating is challenging for *Theodoxus*, the phylogeny was *calibrated* using the molecular clock rates and secondary dating of internal nodes as established by Sands et al. [6]. For congruency, clock rates and secondary dates were set to the
3.3 Phylogeographic structure

Evolutionary relationships among COI, 16S and ATPα haplotypes of the Pontocaspian and southern Iranian Theodoxus groups were established independently for each group and through statistical haplotype networks in TCS v1.2.1 [40]. Several ambiguities associated with heterozygote states in the ATPα sequences (sDNA) were resolved by determining 90% probability score alleles in Phase v2.1 [41] as implemented in DnaSP v6.11.01 [42] under default settings. To further test for differentiation within each group, among geographical sampling localities, analyses of molecular variance (AMOVAs) were performed in Arlequin v3.5.2.2 [43]. P-values were subjected to Holm’s sequential Bonferroni correction [44]. Furthermore, Arlequin was used to calculate haplopytic (h) and nucleotide diversities (n and final), Lastly, isolation by distance (IBD) was tested for using the distance-based redundancy analysis (db-RDA) [45,46] with the package vegan v2.5.2 [47] in the R statistical environment v3.5.2 [48]. Since longitude, latitude and geographic distance among localities were strongly correlated, we only included the latter parameter in the analyses. Principal coordinates analyses was performed on the matrices of pairwise geographic distances and the first principal component was used as single continuous variable in the db-RDA [49].

All input matrices of genetic and geographic distance were constructed in GenAIEx v6.5 [50].

3.4 Evaluation of demographic history

To provide a temporal perspective on population demographics, extended Bayesian skyline (EBS) and lineage through time (LTT) trees was plotted with Tracer v1.7.1 independently for the Pontocaspian and southern Iranian Theodoxus groups. To eliminate potential bias of our interpretations, caused by individual phylogenetic trees on the overall trends, BEAST v2.5.2 was also constructed for individual phylogroups (Appendix A, figure S1). To generate the needed input data and log files, the BEAST v2.5.2 package and CIPRES Science Gateway were again used. The settings and process followed similarly to the dated phylogeny described above (see 3.2 Phylogenetic analyses). However, ‘Coalescent Extended Bayesian Skyline’ was set at the tree prior, only in-group taxa of each group were used and each EBS plot was generated through a single run of 600,000,000 MCMC generations (saving one tree and log file every 20,000 generations and one EBSP log in every 5,000 generations). The best-fit models were again determined by bModelTest III (Pontocaspian group: COI = TN93, 16S = a variant of K81 with additional groups for the rate rT (12,1122123232), ATPα = a variant of K81 with additional groups for the rates rT (12,1122123232), southern Iranian group: COI = a variant of HKY with additional groups for the rate rT (12,1122123232), 16S and ATPα = TN93). Dating followed a slightly different process: Coalescent Bayesian nodes of each in-group were set according to 95% confidence (HPD) intervals determined for the date of divergence of the most common recent ancestor (MCRA), as established in the newly generated phylogeny below using a normally distributed prior (see figure 3).

Last trees were assessed in Tracer v1.7.1 and only runs where all ESS values were > 200 were converted to EBSP plots. To construct the EBS plots, the output EBSP log files were imported into the BEAST package tool EBSPAnalyzer v2.5.2 [37]. Linear reconstruction was selected and the burn-in removed was set as determined by parameter convergence in Tracer v1.7.1.

4. Results

4.1 Phylogenetic structure

Based on the combined analyses, the monophyly of Pontocaspian and southern Iranian Theodoxus groups was well supported (Bayesian posterior probability (PP) = 1.00; figure 3). Within the two groups, six phylogroups were identified (four in southern Iran, termed I–IV herein, and two in the Caspian system, V–VI; figure 3). Three divergence events occurred between southern Iranian phylogroups, while only a single divergence event occurred splitting the two Caspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another started much earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3). The phylogeny suggests that the divergence of southern Iranian phylogroups from one another begun earlier than the two Pontocaspian phylogroups (figure 3).

However, there is both polyphyly and some paraphyly between the morphospecies T. pallidus (Dunker, 1861) and T. dornae Issl, 1865 (table 2; figure 3). In comparison, while there are haplotypes confined to specific locations among the Pontocaspian Theodoxus, there are no location-specific phylogroups. Both supported phylogroups contain individuals from various locations and similarly suggest paraphyly and polyphyly among morphospecies commonly identified as T. major, T. schultzi, T. pallasi, and T. astrachanicus, both within and between phylogroups (table 1; figure 3).
4.2 Phylogeographic structure

The EBS plots suggest the effective population size (in respect to the generation time) in the Pontocaspian
population has been around 1000 at 5 thousand years ago (ka), while the
southern Iranian group entered a steady demographic decline earlier, around 0.1 Ma/ka (figure S3). The LTT plots
demonstrate a relatively constant, nearly exponential pattern of lineage accumulation in the Pontocaspian group.
Compared to the more gradual and periodic, reverse sigmoidal, accumulation observed in the southern Iranian
region. Individual EBS plots for the different phylogroups (I–VI; figure 3) of each of the two major groups did not all find
suitable ESS values to warrant construction (potentially as a result of limited specimens in certain phylogroups). (figure S1). However, those that were supported (figure S1), show strong similarity to the combined EBS plots in
figure 5. This similarity suggests individual phylogroups have had limited prejudice against the overall trends
established for each group (figure S4–5) and by combining data has only aided to increased support.

5. Discussion

5.1 Pontocaspian biogeography

We found some indication for an early catastrophic bottleneck reducing *Theodoxus* diversity to a single lineage in the
Pontocaspian region during the Early Pleistocene. However, post this proposed bottleneck there is little
intraspecific structure within the Pontocaspian *Theodoxus* group to warrant divergence events being specifically
linked to environmental changes (figures 3, 4, and 5). In contrast to strong intraspecific phylogeographic structure
and IBD observed in the southern Iranian *Theodoxus*, haplotypes belonging to the Pontocaspian *Theodoxus*
group are shared among a variety of locations and IBD was either far weaker or not supported at all (table 2; figures 3 and 4,
tables S1 and S2; figures S1 and S2). Therefore, our hypothesis that major low stands drove strong phylogeographic
structure in the Pontocaspian *Theodoxus* group (as a result of populations that were confined to isolated refugia) is
inconclusive prior to the Middle Pleistocene, but can be rejected from this point onward.

Supported intraspecific divergence events in Pontocaspian *Theodoxus* do not necessarily correspond with Caspian
Sea low stands. Although the 95% confidence intervals on divergence events overlap with several low stands [8], they
remain too broad to be attributed to a specific lake level event (figure 3). However, the EBS plots indicate a relatively stable population and LTT plots show a nearly exponential accumulation of lineages post the
mid-Bakunian until at least the Novocaspian stage (Holocene; figure 5a–5). If increased salinity and lake level
changes at low stands had indeed driven species into refugia, we would have expected to see indications of population
bottlenecks and more gradual accumulation during this period. Perhaps Pontocaspian species (or *Theodoxus*) in particular are more tolerant to such environmental changes than expected or the changes during this period were not dramatic enough (both in duration and intensity) to cause major bottlenecking and force taxa into refugia. A
number of Caspian relict molluscs persist in the Caspian Sea [12.51–54], which would suggest it has not been totally
inhospitable. Furthermore, some Pontocaspian molluscs have recently been documented to have wide environmental
tolerances [12,55,56]. Some *Theodoxus* spp. (including *T. pallasi*) can withstand salinities around 20 psu [27,57,58].

The weak phylogeographic structure observed in the Pontocaspian *Theodoxus* group is also found among other
aquatic taxa in the Caspian Sea [13,15,19,59,60], although some are more easily dispersed. Our findings do not
mitigate the role lake level and salinity may have in general, but suggest *low stands* may have had little effect
during this period post the mid-Bakunian on Pontocaspian endemics. Moreover, salinity values during low stands
probably did not exceed the tolerance of *Theodoxus*. However, Caspian Sea high stands likely increased dispersal
possibilities and promoted the homogenisation of haplotypes across the Pontocaspian region. As generalists,
biogeographic barriers to *Theodoxus* dispersal within the Caspian Sea are limited, other than by the extreme deep water
anoxic environments [12,61,62]. High stands may have increased deep water anoxic environments, but they also
increased the littoral and sublittoral zones [8], broadening dispersal pathways. This would explain the overlapping
and well distributed phylogroups in the system (phylogroups V and VI; figures 3 and 4). Certain Caspian Sea high
stands have also been linked to overflow events into the Sea of Azov [8] and would explain the presence of the
phylogroup V in the Ustvolskiy Liman (figures 2 and 3).

Although difficult to test, lake level changes in Caspian Sea low stand may still have had an effect earlier, towards the
end of the Aphanomerian (Late Early Pleistocene). Fossils closely mimicking extant Pontocaspian *Theodoxus* species
have been recorded from Aphanomerian deposits in Azerbaijan and Turkmenistan [51]. Given this fossil record, one
would have expected multiple older lineages in the Pontocaspian (similar to those seen in the southern Iranian group;
largely unstructured Pontocaspian conservation status.

However, the same causes could have been responsible for the apparent population increase, such as by reducing competition. Consequently, species monitoring and distribution modelling are required to better assess its conservation status.

6. Conclusion

Our analyses were able to detect strong phylogeographic structure in the southern Iranian group, compared to a largely unstructured Pontocaspian Theodoxus group. This comparison indicates that the drivers of diversification in
Pontocaspian *Theodoxus* are unlikely related to changes in lake level and refugial persistence as in southern Iran. While low lake level and increased salinity cannot be totally excluded from driving intraspecific divergence events in the Pontocaspian region, we found little evidence to suggest low stands have had a significant effect on phylogeographic structure or demographic history of Pontocaspian *Theodoxus* since the Middle Pleistocene. Rather, high stands and the lack of barriers to gene flow may have caused a homogenising effect given the haplotype sharing unique specimens from diverse localities. However, it was surprising that all intraspecific diversity stemmed from a single lineage during the Middle Pleistocene, which contrasts with the available fossil evidence. We argue that lake level and salinity changes as a consequence of the Turgayyan low stand 1.05–0.95 Ma, being one of the longest and most dramatic regressions in the Caspian Sea, coupled with hydrological, climatic and geological changes, may have caused an earlier catastrophic bottleneck event in Pontocaspian *Theodoxus* with only a single lineage persisting. Moreover, our phylogenetic analyses demonstrate extensive paraphyly and polyphyly across several phylogroups of traditionally recognized morphospecies and suggest the presence of only two species, *T. major* and *T. pallidus*, which are endemic to the Pontocaspian region and southern Iran, respectively.

These results not only broaden our understanding of the effects of lake level and salinity changes in the context of Pontocaspian taxa and provide insights into the evolution and outlooks on taxonomy and conservation of *Theodoxus*. Firstly, they highlight that the Pontocaspian region may lack geographic barriers to gene flow, as even among faunas with limited dispersal capabilities such as *Theodoxus*. Thus, the Caspian Sea may have acted as a single refugium. Secondly, they suggest Pontocaspian gastropods may be more tolerant to the salinity and lake level regimes and changes than previously thought. Thirdly, phenotypic variation in *Theodoxus* is obviously high. This could have large implications for regional stratigraphy and environmental reconstruction (which often use molluscs as environmental indicators) as well as invasion biology (as more adaptable species often have higher invasive potential). Finally, species monitoring and modelling in the Pontocaspian region are essential for assessing the need for conservation.

Future research would best be directed to investigate the ecological flexibility, distribution and abundance and taxonomy of key Pontocaspian aquatic mollusc species.

 Acknowledgments

We would like to thank the Institute of Hydroecology and Ichthyology for the Armenian National Academy of Science (Bardukh Gabrielyan), Institute of Zoology for the Azerbaijan National Academy of Sciences (Elman Yusifov and Konul Ahmadova), Kazakhstan Agency for Applied Ecology (Feodor Klimov), Southern Research Center of the Russian Academy of Sciences (Vadim Titov), Laboratory of Macro-Ecology and Biogeography of Invertebrates - St. Petersburg State University (Maxim Vinarski) and Laboratory of Pleistocene Palaeoecography – Moscow State University (Tamara Yanina) for their logistical assistance. Moreover we are specifically grateful to Konul Ahmadova, Olga Anistratenko, Shebnem Feteliyeva Farzali, Maksim Khaliman, Feodor Klimov, Matteo Latuada, Sergei Sereda, Maxim Vinarski and Frank Wesselingh for their help collecting material in the field. Torsten Hauffe for collecting material and analytical assistance and Conrad Matthee for his assistance interpretation of some results. Finally, the constructive comments of two anonymous reviewers and the associate editor Kristina Sefc greatly improved the paper.

 Ethical Statement

Nagoya protocol-relevant materials were collected based on ABS agreement NBC-KAP5-270417 with the Kazakhstan Agency of Applied Ecology. Other material derived from Armenia, Azerbaijan, Iran, Russia and Ukraine were sampled through bilateral agreements among collaborating institutes.

 Funding Statement

The research herein has received funding from the European Commission’s Horizon 2020 research and innovation programme under the grant agreement no. 642973: Drivers of Pontocaspian Biodiversity Rise and Demise (PRIDE). A.F.S. was supported by a Marie Skłodowska-Curie Action Fellowship and T.A.N. by an Alexander von Humboldt scholarship.

 Data Accessibility

Raw sequence information was deposited on GenBank, accession numbers: still to add – will upload and amend internal tables with accession numbers upon first acceptance of the research article, see temporary Appendix B (DNA sequence data) for the purposes of the review process.

 Voucher specimens of all processed material are stored in the Justus Liebig University Giessen, Systematics and Biodiversity collection, Germany. DNA sequences were deposited on GenBank, accession numbers: MN168547, MN168757, MN174926, MN175136 and MN180417–MN180627. All nucleotide sequence alignment files used as input data for the various analyses, output data used to construct the graphs of the EBS plot analyses and dRDA input files with R script are available through Dryad (https://doi.org/10.5061/dryad.mn15f80).[84]

 Competing Interests

The authors declare no competing interests.

 Authors’ Contributions

All authors contributed to the conceptualisation of the research article. In particular A.F.S. conceptualised the study, A.F.S., S.N., M.F.H. and V.V.A. conducted the fieldwork and organised logistics, A.F.S. and S.N. performed the laboratory work, A.F.S. and T.A.N. analysed the data and A.F.S., T.A.N., T.W. and C.A. leading the writing.

 References

1. Bunje PME. 2007 Fluvial range and snail faunas (Neritidae): a case study: 392–399 [snail faunas (Neritidae): a case study: 392–399](doi:10.1111/j.1095-8312.2007.00750.x). Parallel evolution in a Danubian stream, Bih. J. Linn. Soc.393 (8312.2007.00750.x).

2. Bunje PME. 2005 Pan–
733 730 725 722 712 703 700 696 693 689 682 678 675 672 671 670 668 666 662 659 650 644.

1720–1728. (doi:10.1093/mbe/sto64) 754

760 758 756 753 751 750 747 746 743 739 738 735 732 730 727 724 721 718 715 712 709 706 703 700 697 694 691 688

58. Kangas P, Skog G, 1978 Salinity tolerance of Theodoxus fluviatilis (Mollusca, Gastropoda) from freshwater and from different salinity regimes in the Baltic Sea. Estuar. Coast. Mar. Sci. 6, 409–416. (doi:10.1016/0257-8916(78)90031-7)

59. Briston VJ, Ruban G, Ludwig A, Doulakis P, DeSalle R. 2005 The enigmatic Caspian Sea Russian sturgeon: How many cryptic forms does it contain? Syst. Biodivers. 3, 203–218. (doi:10.1017/S1477200005000511) 647

60. Ruban GI, Khodolova M.V., Kalmykov VA, Sordoni PA. 2008 Morphological and molecular genetic study of the Persian sturgeon Aipenser persicus Boréon (Acipenseridae) taxonomic status. J. Ichthyol. 48, 891–903. (doi:10.1134/s0047940508090069) 608

61. Parr TD, Tid RF, Maxon CL, Newton FC, Hardin JL. 2007 A descriptive account of benthic macrofauna and sediment from an area of planned petroleum exploration in the southern Caspian Sea. Estuar. Coast. Shelf Sci. 71, 170–180. (doi:10.1016/j.ecss.2006.07.018)

62. Dumont HJ. 1998 The Caspian Sea: History, biology, structure, and function. Limno. Oceanogr. 43, 44–52. (doi:10.4319/lo.1998.43.1.0044)

63. Svitoch AA. 2012 The Caspian Sea shell during the Pleistocene-regressive epochs. Oceanology 52, 526–539. (doi:10.1134/S0049947012030133)

64. Nuevesskaja LA. 2001 History of Lymnocardinae (Bivalvia, Cardidae). Paleontol. J. 35, S147–S211.

65. Tagaya EN, Aleskov BD, Veliev SS, Kuliev MY. 2013 Transformation of vegetation cover on the territory of Azerbaijan during the Pleistocene. Geogr. Nat. Resour. 34, 395–400. (doi:10.1134/s1875372813013040)
66. Zastrozhnov A et al. 2018. Biostratigraphical investigation as a tool for discriminating palaeoenvironmental changes. Neopliocene (Middle-Upper) deposits in the South Caspian Sea. J. Pet. Sci. Eng. 166, 435–494.

67. Abdullahayev N. 2000 Seismic stratigraphy of the Upper Pliocene and Quaternary deposits in the South Caspian Sea. Geosci. Front. 28, 1005–207. (doi:10.1016/S0920-0190(00)00079-6)

68. Eyvindottir EH et al. 2017. Assessment of relative dating techniques in the Bozioneh Badlands. Geosci. Front. 8, 1051–1066. (doi:10.1016/S0920-0190(17)30131-9)

69. Javadi Mosavi E, Asli M. 2013. Tectonic geomorphology of the Atrak River, NE Iran. Open J. Earth Sci. 3, 106–114. (doi:10.4236/ojes.2013.33018)

70. Shanko E, Lebedev V. 2019. Chugaev A, Zabarnikaya L. 2020. Rodnina A, Sergeeva N. 2022. Salanova I. 2015. The Caucasian-Arabian segment: a tectonic perspective. Geol. Soc. Lond. Spec. Publ. 423, 105–124. (doi:10.1144/SP423.1)

71. Sokhadze G, Floyd M. 2019. Godolazde T, King R, Cowgill3. 2013. ES, Javakhishvili Z. 2012. Reillon R. 2018 Active 1051–1056. (doi:10.1016/j.cuto.2018.04.019)

72. Salzburger W, Van Bocxlaer B. 2019. Change in western arid central Asia inferred from fossil-palaeosoil sequences. Sci. Rep. 9, 20560. (doi:10.1038/s41598-019-42600-9)

73. Cohen A, 2014. Evolution and distribution of the African Great Lakes. Rev. Ecol. Evol. Syst. 45, 81–94. (doi:10.1146/annurev-ecolsys-122012-091104)

74. Schulteth R, Wilke T. 2019. The birth of an endemic species flock: demographic history of the Belamy group (Gastropoda, Viviparidae) in the Lower Lake Malawi. Biol. J. Linn. Soc. 103, 130–143. (doi:10.1111/j.1095-8312.2010.01574.x)

75. Starmüller F, Edlauer A. 2009. Ergebnisse der Expedition 1956: Beiträge zur Kenntnis der Molluskenfauna des Iran. Sitzungsberichte der österreichischen Akad. der Wissenschaften, Math. Klasse, Abteilung 166, 435–494. Glück P, Petri V. 2012. The freshwater snails (Gastropoda) of Iran, with descriptions of two new genera and eight new species. Zookeys 219, 11–61. (doi:10.3897/zookeys.219.3406)
| Theodoxus species | Number of specimens | Locality | Country | GenBank accession numbers |
|-------------------|--------------------|----------|---------|--------------------------|
| T. transversalis var. major | 10 | Vardar River | Macedonia | MN186947, MN186956, MK754724, MK754725, MK754736, MK754737 |
| T. transversalis var. major | 5 | Bath of Aphrodite, Aspas | Cyprus | MN186947, MN186956, MK754724, MK754725, MK754736 |
| T. jordani | 10 | Blue Waters Beach | Azerbaijan | MK754733, MK754734, MK754735, MK754736, MK754737, MK754738 |
| T. doriae | 5 | Pirallahi Island | Iran | MN186947, MN186956, MK754724, MK754725, MK754736 |
| T. schultzii | 5 | Shakhly Spring | Kazakhstan | MN186947, MN186956, MK754724, MK754725, MK754736 |
| T. pallasi | 10 | Bolu Amos Spring | Iran | MN186947, MN186956, MK754724, MK754725, MK754736 |
| T. doriae | 5 | Shakhly Spring | Kazakhstan | MN186947, MN186956, MK754724, MK754725, MK754736 |
| T. pallasi | 5 | Shakhly Spring | Kazakhstan | MN186947, MN186956, MK754724, MK754725, MK754736 |
| T. doriae | 10 | Shakhly Spring | Kazakhstan | MN186947, MN186956, MK754724, MK754725, MK754736 |
| T. pallasi | 10 | Shakhly Spring | Kazakhstan | MN186947, MN186956, MK754724, MK754725, MK754736 |

Note: Numbers of GenBank accession numbers are specific to the table. More information is required to fill in the GenBank accession numbers.
| Population statistics | Pontocaspian group | Southern Iranian group |
|-----------------------|-------------------|-----------------------|
|                       | COI   | 16S   | ATPα  | COI   | 16S   | ATPα  |
| Number of specimens n | 175   | 175   | 45    | 45    | 45    |
| Number of sequences n | 175   | 175   | 350   | 45    | 45    | 90    |
| Haplotypic diversity h| 0.793 | 0.706 | 0.791 | 0.680 | 0.610 | 0.612 |
| Nucleotide diversity π| 0.007 | 0.002 | 0.003 | 0.018 | 0.007 | 0.006 |
| Isolation by distance (IBD) | R² | 0.00572 | 0.01052 | 0.01015 | 0.00017 | 0.044417 |
| p value               | 0.44088 | 0.00728 | 0.00781 | 0.98030 | 0.94806 | 0.21423 |
Figures
Table 1. Species details and GenBank accession numbers for Pontocaspian, southern Iranian and outgroup *Theodoxus* spp. Locality names correspond to those in figure 1.

Table 2. Comparative summary of population statistics for COI, 16S and ATPα datasets between Pontocaspian and southern Iranian *Theodoxus* groups. Note IBD was calculated using db-RDA (see 3.3 Phylogeographic structure).

Figure 1. Representative phenotypes of the Pontocaspian and southern Iranian *Theodoxus* species studied herein. Pontocaspian: (a, b) *T. pallasi* (UGSB 20712); (c, d) *T. astrachanicus* (UGSB 18130); (e, f) *T. pallasi* (23435); (g, h) *T. major* (24895); (i, j) *T. major* (24909); (k, l) *T. schultzii* (25116); (m, n). Southern Iranian: *T. doriae* (26038); (o, p) *T. pallidus* (26429). Scale bar = 1 mm.

Figure 2. Map depicting the locations of the sampling sites around the Pontocaspian system and southern Iran. Colours of dots correspond to the locations as indicated in the key. Dashed lines encircle (I) the Pontocaspian and (II) the southern Iranian *Theodoxus* sampling localities. The size of the dots represent the sample size at each location (larger = 10; smaller = 5 specimens).

Figure 3. Dated phylogeny of Pontocaspian and southern Iranian *Theodoxus* spp. constructed in BEAST based on COI, 16S and ATPα sequence data. Supported phylogroups of Pontocaspian and southern Iranian *Theodoxus* are labelled I to VI. Node labels among these phylogroups and outgroup species denote divergence time in millions of years ago (Ma), with the 95% credibility interval given in parentheses and as grey bars for in–group taxa. Small red squares at nodes (with darkened node bars and, in some instances, dates) indicate significant posterior probabilities of divergence events. Parallel to each supported phylogroup, coloured bars indicate the localities and respective morphospecies of the included specimens as defined in the key on the left (also see figure 2). Caspian Sea lake level variations over the last 1.5 Ma (relative to absolute sea level) and regional stratigraphy (following the “short–Akchagylian” option) are adapted from Krijgsman et al. [8] (Khv. = Khvalynian).

Figure 4. Statistical haplotype networks for COI, 16S and ATPα sequence data for Pontocaspian and southern Iranian *Theodoxus* groups. The total number of sequences in each network are demarcated by ‘n’. The circle sizes represent relative frequency of sequences per haplotype. The number of site changes separating haplotypes are indicated by blank dots. Colours correspond to the sampling locations as indicated in the key and in figure 2. Haplotype groupings are boxed and labelled according to phylogroups determined through the dated phylogeny (I–VI; figure 3).

Figure 5. a) Extended Bayesian skyline (EBS) plots indicating population trends for the Pontocaspian (blue) and southern Iranian (red) *Theodoxus* groups. b) Lineage-through-time (LTT) plot indicating the accumulation of lineages through time for Pontocaspian and southern Iranian *Theodoxus*. In the EBS plots, the central line of each plot represents the median value, while in the LTT plots the line rather constitutes the mean. In both plots and the shaded area indicates the 95% confidence interval and floating grey bars indicate the 95% confidence interval associated with the onset of intraspecific diversification. The EBS plots depict marginally younger intraspecific diversification for each group when compared to the phylogeny. Importantly however, there is strong overlap of EBS plot starting dates with the 95% confidence intervals established among the analyses for the onset of intraspecific diversification in each group as shown in the phylogeny (figure 3).
To the Chief Editor, Royal Society Open Science

Dear Prof. Jeremy Sanders CBE FRS
Cc: Prof. Kristina Sefc (Associate Editor), Prof. Kevin Padian (Subject Editor)

Reference Number: RSOS-190965

We very much appreciate the time and effort of the reviewer and the editors in assessing our revised paper and we are grateful for their additional comments. Below we list all critical comments and suggestions of the associate editor and reviewer (in italics), together with our point-by-point replies (in blue).

ASSOCIATE EDITOR (Dr Kristina Sefc):

I’d like to thank you for addressing most of the concerns raised in the first round of review. The issue of pooling divergent lineages for the EBSPs still remains and actually represents a rather serious one since an important conclusion - regarding the decline in population size - may be based on an artefact in the analysis. Please see the reviewer's comments for details. If the sampling / the data don't allow to analyse the demographic history of the Iranian populations, then you'll have to consider dropping this part from the manuscript (or interpret the network structures of the mtDNA and ncDNA sequences verbally - less diversity in the mt than nc genomes within each phylogroup may indeed point to a recent bottleneck); or at least discuss the problem associated with pooling the divergent lineages and make clear that there’s a risk of a spurious result.

We note the feedback of the reviewer and the associate editor regarding the EBSPs. We understand there may still be a potential bias from pooling the Iranian data without being able to assess the EBSP trends in individual Iranian phylogroups due to the ESS values lacking support (potentially as a result of small sample sizes of the phylogroups). We have thus followed the associate editor’s suggestions:

1) We have added further interpretation of our networks and haplotypic diversity to add support for a bottleneck within the Iranian species as observed in the EBSP (see L171-174, L190-191 & L284-286).

2) We now note that “caution” should be taken with the Iranian group’s EBSP when interpreting the trend of the pooled data (figure 5) given the bias that may be caused (L190-192, L282-284 & L289-290). We reference the papers noted by the reviewer in this regard (L135 & L283-284). However, we also note that Heller et al. (2013) indicates that at least in part some of these concerns may be mitigated by a balanced sampling strategy such as our own (L284). Moreover we note that even neglecting the EBSP (from a conservation perspective), given current climate shifts and anthropogenic threats to spring habitats in the region, conservation efforts may still be needed to protect this species (L290-293).
1. Lines 175-177 (and comment 10 of the original review): Note that each node in the tree represents a divergence event. In fact, you can delete this entire sentence as it essentially conveys the same information as the previous sentence (lines 174-175).

We have removed this sentence as suggested.

2. EBSPs: Yes, doing separate analyses for the two Pontocaspian phylogroups (now presented as Supplementary info) is the right way to deal with these data. And yes, for most of the Iranian phylogroups sample size/diversity is too small to allow for doing EBSP analyses. This, however, is no justification for pooling all these four pretty divergent phylogroups (cryptic species?) for a single EBSP analysis. You have virtually no intraphylogroup diversity in the Iranian samples and with the large interphylogroup divergence you inevitably get a signature of a drastic recent population decline in your EBSP. It's simply not possible to do reliable EBSP analyses with these Iranian samples. I suggest you have a look at previous studies that evaluated the performance of (E)BSPs in the presence of population structure. Specifically, in highly structured populations, drastic recent population size declines were observed (similar to what you found with your data) even though the sequences were simulated under a constant population size scenario. Two relevant studies that come to my mind are Heller et al. 2013 PLoS One & Grant 2015 JHered. I think you'll have to have a more critical look at your EBSPs, especially since these data are essential for your discussion/conclusions on conservation aspects in your gastropods.

Please see response to the associate editor's comment.