THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis

Marco Gerdol, Marco Sollitto, Alberto Pallavicini and Immacolata Castellano

Article citation details
Proc. R. Soc. B **286**: 20191812.
http://dx.doi.org/10.1098/rspb.2019.1812

Review timeline

Original submission:	27 March 2019
1st revised submission:	2 August 2019
2nd revised submission:	5 November 2019
Final acceptance:	6 November 2019

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

Review History

RSPB-2019-0683.R0 (Original submission)

Review form: Reviewer 1

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

Scientific importance: Is the manuscript an original and important contribution to its field?
Acceptable

General interest: Is the paper of sufficient general interest?
Acceptable

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis. Gerdol et al.
--

The paper examines the distribution of ovoA type genes in the animals. The authors do a good job of examining 'minor' phyla and report what they describe in the title as 'a complex evolutionary history'. On the whole, it's a solid analysis, but it's hard to judge whether it's reporting a truly unusual phylogenetic scenario. There are a lot of gene families that show a patchy phylogenetic distribution within animals. In particular, the loss events in Ecdysozoa, platyhelminthes and vertebrates don't seem that remarkable. Counterbalancing this, solid evidence of horizontal transfer into the Hydrozoa (or any animal) is definitely interesting (although at the same time, it's not the first reported example of such a transfer).

Specific points

The P.lividus ovoA sequence (uniprot:A0A140GAV9) doesn't have a hit to IPR024775 using Interproscan (http://www.ebi.ac.uk/interpro), at least as far as I can see. Neither with PFAM. I don't doubt these assignments, but the manuscript could be clearer as to what is being done.

As far as I can see, the authors do not discuss whether the diverse ovoAs (i.e. Clade I, II and III) are all likely to catalyze the same chemical reaction. Is there evidence that the key catalytic residues are conserved? Could the authors demonstrate this in the manuscript? This will not be proof, but it does seem a basic test to perform. If I understand correctly Supp. Fig. 1 does this only for metazoan sequences.

P.8 "Even though ovoA was detected in early-branching holozoan groups [such] as Ichthyosporea and choanoflagellates these sequences do not share the same ancestry with metazoan ovoA". Could the authors confirm that they have searched the data reported in Richter et al. Elife 1018:e34226 - doi: 10.7554/eLife.34226 ? To me, admittedly judging only from pairwise similarity

scores, it looks like there are some metazoan clade ovoAs in there. Obviously this is quite an important point for some of the later discussion.

P.10 discussion of convergent evolution. I'm not sure that convergent evolution would be a widely accepted hypothesis for incongruence between a gene tree and species tree. What about inadequacies of the phylogenetic model for dealing with differing lineage specific rates as a null hypothesis? This also comes up in the discussion.

P.14 There is generally no way that template based structural modelling will not give a similar structure to the starting template. That's the point of using a template and not doing ab initio structure prediction. I realise that the literature has a lot of examples like this, but I would raise the same criticism about all of them. I would not let this section stand unless the authors can demonstrate that it reveals something new. It would be better to just report significant sequence similarity scores.

Minor points

The opening paragraph speaks about 'adaptive evolution', 'strategies to cope' and 'environmental adaptation', but the manuscript doesn't really have anything to say about these things - it's a study of a gene family, not an evolutionary process.

P.6 'a few sequences from unicellular prokaryotes (i.e. Ichthyosporea...' the taxa listed are eukaryotes not prokaryotes.

Although the manuscript is entirely comprehensible, if the authors could get a native speaker to go over it, it would benefit. E.g.:

Abstract - 'originated by the fusion between' would be better as 'originating from the fusion of'

Introduction - 'early embryonal phases' -> early embryonic phases

Results - 'OvoA' by itself will be interpreted as singular. So better '[...] OvoA were found in' -> 'OvoAs were found in' or 'OvoA transcripts/orthologs...' or 'OvoA was found in...'

Discussion - 'have played a capital role' -> 'an important role'?

Review form: Reviewer 2

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

Scientific importance: Is the manuscript an original and important contribution to its field?
Acceptable

General interest: Is the paper of sufficient general interest?
Acceptable

Quality of the paper: Is the overall quality of the paper suitable?
Acceptable

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
This is comprehensive analysis of the evolution of the OvaA gene family with a particular emphasis on its evolution in metazoans. For the most part, the research has been undertaken in a thorough and competent manner. However, there are some substantial gaps that if addressed could improve this report.

First, the documentation and discussion of the two putative horizontal gene transfer events in bdelloid rotifers and hydrozoan cnidarians need to be greatly expanded and integrated in the manuscript and its figures. For example, while the hydrozoan HGT is mentioned throughout the Results and Discussion but there is essentially no data shown, except in Figure 1, which is a low-resolution tree of the OvoA genes across life. In the bdelloid example, this needs to be more explicitly compared with the other rotifer clades (monogonts) that have the metazoan-specific form; the same can be said for the hydrozoan genes compared to other cnidarians. In addition, the authors need to demonstrate convincingly that these putative HGTs are integrated into the host genomes and not contaminants – i.e. are the OvoA genes on contigs populated by metazoan/eukaryotic genes. In summary, this is an interesting find but it needs to be much better documented.

Second, the primary tree in the paper (Fig. 3) has a topology that does not match the organismal tree (e.g. OvoA genes from sponges group with Chondrichthyes OvoA genes; there are a number of similar cases on the tree). While the authors note that this may be because of convergence, the details are not provided. Are the residues shared between these widely-separated animal lineages tell us anything about protein function in the marine environment, for instance? Can these be mapped onto the 3-D models? Again a much more detailed analysis is likely to shed light on what parts of the OvoA protein underlie these apparent convergences.

In addition there are a number of minor points that should be addressed:

Fig. 1 – figure and legend need more details about clade, scale bars, etc.

P9 -sponges are not the most ancient animal group – they are the most ancient extant phyletic lineage

P9/10 – ctenophore/sponge debate remains unresolved – suggest that you do not make such a definitive statement about ctenophores being the sister lineage to all other animals. It is well known the ctenophores have lost many genes. There are two alternative evolutionary scenarios that should be considered equally.

P10- Parahoxzoa

P10 – platyhelminths, mesozoans and gastrotrichs are not necessarily basil lophotrochozoans (see Current Biology 29, 312–318, 2019)

P10 - xenoacoelomorphs are not at the 'crossroads' between protostomes and deuterostomes – they are the earliest branching bilaterian phyletic lineage

Fig. 2 – most metazoan trees are drawn as a basal polytomy with sponges, ctenophores and all other animals (parahoxzoans) branching off a common ancestor

Fig.2 spelling mistakes in legend

P13 – CDS?

Fig. 4 – Why draw the exon-intron architecture based on sea urchin? Given the conservation and phylogenetic position of sponges, placozoans and anthozoans, their consensus gene architecture should be the standard for comparison. Why aren't the hydrozoan and the monogont rotifer genes shown?

P17 – Speculation as to the source of hydrozoan HGT needs to be based on strong phylogenetic evidence and not on the presences of modern day symbionts in various cnidarian lineages. HGT events happened in the past and may not reflect modern animal- microbe associations. This part of the discussion needs to be completely reworked.

Decision letter (RSPB-2019-0683.R0)

10-Jun-2019

Dear Dr Castellano:

I am writing to inform you that your manuscript RSPB-2019-0683 entitled "The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis" has, in its current form, been rejected for publication in Proceedings B.

This action has been taken on the advice of the Associate Editor and the referees, who have recommended that substantial revisions are necessary. With this in mind we would be happy to

consider a resubmission, provided the comments of the referees are fully addressed. However please note that this is not a provisional acceptance. I should also remind you that Proceedings B is a broad biological journal, and the abstract in its current form is very specialised, especially in its opening sentences: I would encourage you to make it more broadly accessible.

The resubmission will be treated as a new manuscript. However, we will approach the same reviewers if they are available and it is deemed appropriate to do so by the Editor. Please note that resubmissions must be submitted within six months of the date of this email. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office. Manuscripts submitted after this date will be automatically rejected.

Please find below the comments made by the referees, not including confidential reports to the Editor, which I hope you will find useful. If you do choose to resubmit your manuscript, please upload the following:

1) A 'response to referees' document including details of how you have responded to the comments, and the adjustments you have made.
2) A clean copy of the manuscript and one with 'tracked changes' indicating your 'response to referees' comments document.
3) Line numbers in your main document.

To upload a resubmitted manuscript, log into http://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Resubmission." Please be sure to indicate in your cover letter that it is a resubmission, and supply the previous reference number.

Sincerely,
Professor Loeske Kruuk
Editor
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
I enjoyed reading the manuscript by Castellano et al. This is a robust study on the evolution of a gene family. However, I have some concerns shared with the reviewers. Mainly, the most interesting part of the study (the HGT) is neglected on favour of more traditional (and not very novel) gene family analyses.

I think the manuscript would improve if the HGT events were further explored, and less emphasis made on more generic results (e.g., gene losses in lineages known to have many gene losses). Further inspection of the origin of these genes by BLAST with loose parameters together with gene trees, synteny around these genes, and more analyses on their gene structure would address that.

Importantly, there are a number of evolutionary misconceptions that need to be addressed (sponges as ancient group, acoels as missing links...). I would like to invite the authors to address those, and all the other comments made by the referees. Namely, but not only, the use of convergent evolution to justify gene trees that disagree with species trees (sometimes the phylogenetic signal is just not there), and the presence of these genes in non-metazoan holozoans.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis. Gerdol et al.
--

The paper examines the distribution of ovoA type genes in the animals. The authors do a good job of examining 'minor' phyla and report what they describe in the title as 'a complex evolutionary history'. On the whole, it's a solid analysis, but it's hard to judge whether it's reporting a truly unusual phylogenetic scenario. There are a lot of gene families that show a patchy phylogenetic distribution within animals. In particular, the loss events in Ecdysozoa, platyhelminthes and vertebrates don't seem that remarkable. Counterbalancing this, solid evidence of horizontal transfer into the Hydrozoa (or any animal) is definitely interesting (although at the same time, it's not the first reported example of such a transfer).

Specific points

The P.lividus ovoA sequence (uniprot:A0A140GAV9) doesn't have a hit to IPR024775 using Interproscan (http://www.ebi.ac.uk/interpro), at least as far as I can see. Neither with PFAM. I don't doubt these assignments, but the manuscript could be clearer as to what is being done.

As far as I can see, the authors do not discuss whether the diverse ovoAs (i.e. Clade I, II and III) are all likely to catalyze the same chemical reaction. Is there evidence that the key catalytic residues are conserved? Could the authors demonstrate this in the manuscript? This will not be proof, but it does seem a basic test to perform. If I understand correctly Supp. Fig. 1 does this only for metazoan sequences.

P.8 "Even though ovoA was detected in early-branching holozoan groups [such] as Ichthyosporea and choanoflagellates these sequences do not share the same ancestry with metazoan ovoA". Could the authors confirm that they have searched the data reported in Richter et al. Elife 1018:e34226 - doi: 10.7554/eLife.34226 ? To me, admittedly judging only from pairwise similarity scores, it looks like there are some metazoan clade ovoAs in there. Obviously this is quite an important point for some of the later discussion.

P.10 discussion of convergent evolution. I'm not sure that convergent evolution would be a widely accepted hypothesis for incongruence between a gene tree and species tree. What about inadequacies of the phylogenetic model for dealing with differing lineage specific rates as a null hypothesis? This also comes up in the discussion.

P.14 There is generally no way that template based structural modelling will not give a similar structure to the starting template. That's the point of using a template and not doing ab initio structure prediction. I realise that the literature has a lot of examples like this, but I would raise the same criticism about all of them. I would not let this section stand unless the authors can demonstrate that it reveals something new. It would be better to just report significant sequence similarity scores.

Minor points

The opening paragraph speaks about 'adaptive evolution', 'strategies to cope' and 'environmental adaptation', but the manuscript doesn't really have anything to say about these things - it's a study of a gene family, not an evolutionary process.

P.6 'a few sequences from unicellular prokaryotes (i.e. Ichthyosporea...' the taxa listed are eukaryotes not prokaryotes.

Although the manuscript is entirely comprehensible, if the authors could get a native speaker to go over it, it would benefit. E.g.:

Abstract - 'originated by the fusion between' would be better as 'originating from the fusion of'

Introduction - 'early embryonal phases' -> early embryonic phases

Results - 'OvoA' by itself will be interpreted as singular. So better '[...] OvoA were found in' -> 'OvoAs were found in' or 'OvoA transcripts/orthologs...' or 'OvoA was found in...'

Discussion - 'have played a capital role' -> 'an important role'?

Referee: 2

Comments to the Author(s)
This is comprehensive analysis of the evolution of the OvaA gene family with a particular emphasis on its evolution in metazoans. For the most part, the research has been undertaken in a thorough and competent manner. However, there are some substantial gaps that if addressed could improve this report.

First, the documentation and discussion of the two putative horizontal gene transfer events in bdelloid rotifers and hydrozoan cnidarians need to be greatly expanded and integrated in the manuscript and its figures. For example, while the hydrozoan HGT is mentioned throughout the Results and Discussion but there is essentially no data shown, except in Figure 1, which is a low-resolution tree of the OvoA genes across life. In the bdelloid example, this needs to be more explicitly compared with the other rotifer clades (monogonts) that have the metazoan-specific form; the same can be said for the hydrozoan genes compared to other cnidarians. In addition, the authors need to demonstrate convincingly that these putative HGTs are integrated into the host genomes and not contaminants – i.e. are the OvoA genes on contigs populated by metazoan/eukaryotic genes. In summary, this is an interesting find but it needs to be much better documented.

Second, the primary tree in the paper (Fig. 3) has a topology that does not match the organismal tree (e.g. OvoA genes from sponges group with Chondrichthyes OvoA genes; there are a number of similar cases on the tree). While the authors note that this may be because of convergence, the details are not provided. Are the residues shared between these widely-separated animal lineages tell us anything about protein function in the marine environment, for instance? Can these be mapped onto the 3-D models? Again a much more detailed analysis is likely to shed light on what parts of the OvoA protein underlie these apparent convergences.

In addition there are a number of minor points that should be addressed:

Fig. 1 – figure and legend need more details about clade, scale bars, etc.

P9 -sponges are not the most ancient animal group – they are the most ancient extant phyletic lineage

P9/10 – ctenophore/sponge debate remains unresolved – suggest that you do not make such a definitive statement about ctenophores being the sister lineage to all other animals. It is well known the ctenophores have lost many genes. There are two alternative evolutionary scenarios that should be considered equally.

P10- Parahoxzoa

P10 – platyhelminths, mesozoans and gastrotrichs are not necessarily basil lophotrochozoans (see Current Biology 29, 312–318, 2019)

P10 - xenoacoelomorphs are not at the 'crossroads' between protostomes and deuterostomes – they are the earliest branching bilaterian phyletic lineage

Fig. 2 – most metazoan trees are drawn as a basal polytomy with sponges, ctenophores and all other animals (parahoxzoans) branching off a common ancestor

Fig.2 spelling mistakes in legend

P13 – CDS?

Fig. 4 – Why draw the exon-intron architecture based on sea urchin? Given the conservation and phylogenetic position of sponges, placozoans and anthozoans, their consensus gene architecture should be the standard for comparison. Why aren't the hydrozoan and the monogont rotifer genes shown?

P17 – Speculation as to the source of hydrozoan HGT needs to be based on strong phylogenetic evidence and not on the presences of modern day symbionts in various cnidarian lineages. HGT events happened in the past and may not reflect modern animal- microbe associations. This part of the discussion needs to be completely reworked.

Author's Response to Decision Letter for (RSPB-2019-0683.R0)

See Appendix A.

RSPB-2019-1812.R0

Review form: Reviewer 1

Recommendation
Reject – article is not of sufficient interest (we will consider a transfer to another journal)

Scientific importance: Is the manuscript an original and important contribution to its field?
Acceptable

General interest: Is the paper of sufficient general interest?
Marginal

Quality of the paper: Is the overall quality of the paper suitable?
Marginal

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
General point

In my original review I stated "it's hard to judge whether it's reporting a truly unusual phylogenetic scenario. There are a lot of gene families that show a patchy phylogenetic distribution within animals. In particular, the loss events in Ecdysozoa, platyhelminthes and vertebrates don't seem that remarkable."

I cannot see that any attempt has been made to address this, either in the ms itself or the response letter.

Specific points

Structural modelling concern - the section "OvoA is fused with a unique sulfide-lyase domain in Hydrozoa".

Despite the authors' assertion that the Hydra N-terminal domain "lacks significant primary sequence similarity with any protein deposited in public database." this is not the case. Searching with that sequence reveals hits to structures in the PDB and PFAM (I see this using blastp or phmmer). Phyre2 uses sequence similarity (HMM-HMM comparison, see Figure 1 in their Nature Protocols paper) in its pipeline to detect suitable templates. I-TASSER includes sequence similarity protocols via LOMETS. Once a template has been selected, via sequence similarity searches, there is no way that the modelled protein will not have that overall fold, so stating that the method 'revealed a significant structural match' between the modelled structure and the structure the query was aligned to in order to build the model does not add anything. The authors could report that Phyre or I-Tasser produced a significant similarity score at the template detection stage if they cannot demonstrate it via standard sequence database search, but reporting the significance of the similarity between the structure of a modelled protein and the known crystal structure used for modelling is, I think, actively misleading.

Concern about convergence between sponge and cartilaginous fish.

The authors stick by their hypothesis that the incongruence between gene and species trees is due to convergent evolution. The evidence presented in Figure S2 is far from compelling. Picking a fast-evolving urochordate, which one would expect to be more divergent, does not help. Under such circumstances two more slowly evolving taxa will inevitably appear more similar. The fact that the sponges appear paraphyletic in Fig 3, deuterostomes are unresolved, there is a clade of sponges + cnidarians etc. strongly suggests that the data aren't being modelled adequately (I don't blame the authors for that), so reading too much into the placement of any particular group looks rash. Not even mentioning in the main text that an alternative and, a priori very likely hypothesis, that the result is an artefact, seems wrong to me.

Review form: Reviewer 2

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Good

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
This is an improved manuscript and should be suitable for publication after a number of minor adjustments.

Overall, please check spelling and grammar. There are a number of errors in the manuscript.

The Abstract has been extensively rewritten and is now wordy and less precise than the original. Suggest the authors rework this section to come up with something that highlights the new data but keeps the original structure.

Fig. 1 is much improved, but please state what each clade is (color and taxa) in the legend

Review form: Reviewer 3

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Good

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
No

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
Please see attached. (See Appendix B)

Decision letter (RSPB-2019-1812.R0)

18-Oct-2019

Dear Dr Castellano,

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers have slightly mixed opinions about the manuscript, but we would like to give you the opportunity to revise your manuscript to address them.

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; in particular, please see Reviewer 3's comments, including points raised in the first review. If deemed necessary by the Associate Editor, 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. Please note that we cannot guarantee eventual acceptance of your manuscript at this stage.

To submit your revision please log into http://mc.manuscriptcentral.com/prsb 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.

When submitting your revision please upload a file under "Response to Referees" in the "File Upload" section. This should document, point by point, how you have responded to the reviewers' and Editors' comments, and the adjustments you have made to the manuscript. We

require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Your main manuscript should be submitted as a text file (doc, txt, rtf or tex), not a PDF. Your figures should be submitted as separate files and not included within the main manuscript file.

When revising your manuscript you should also ensure that it adheres to our editorial policies (https://royalsociety.org/journals/ethics-policies/). You should pay particular attention to the following:

Research ethics:
If your study contains research on humans please ensure that you detail in the methods section whether you obtained ethical approval from your local research ethics committee and gained informed consent to participate from each of the participants.

Use of animals and field studies:
If your study uses animals please include details in the methods section of any approval and licences given to carry out the study and include full details of how animal welfare standards were ensured. Field studies should be conducted in accordance with local legislation; please include details of the appropriate permission and licences that you obtained to carry out the field work.

Data accessibility and data citation:
It is a condition of publication that you make available the data and research materials supporting the results in the article. Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the article (https://royalsociety.org/journals/ethics-policies/data-sharing-mining/). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should also be fully cited and listed in the references.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link
http://datadryad.org/submit?journalID=RSPB&manu=(Document not available), which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

For more information please see our open data policy http://royalsocietypublishing.org/data-sharing.

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Please try to submit all supplementary material as a single file.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will

not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

Please submit a copy of your revised paper within three weeks. If we do not hear from you within this time your manuscript will be rejected. If you are unable to meet this deadline please let us know as soon as possible, as we may be able to grant a short extension.

Thank you for submitting your manuscript to Proceedings B; we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes,

Professor Loeske Kruuk
mailto: proceedingsb@royalsociety.org

Associate Editor
Comments to Author:
I'd like to thank the authors for the efforts made to improve the paper. I think that the manuscript has improved much, however I think there are still some important issues standing.

The abstract is now a bit convoluted and hard to follow. I believe it could be streamlined and the content could be reorganised to improve the flow.

I agree with some of the comments from the referees, the authors should take then into consideration, namely further support for LGT (e.g., GC composition and codon bias), the claims on gene loss in specific animal lineages, lack of gene tree resolution vs convergence, evolutionary model selection, and the structural modelling. Please, also consider improving Figure 1 and 3.

Reviewer(s)' Comments to Author:

Referee: 2

Comments to the Author(s).
This is an improved manuscript and should be suitable for publication after a number of minor adjustments.

Overall, please check spelling and grammar. There are a number of errors in the manuscript.

The Abstract has been extensively rewritten and is now wordy and less precise than the original. Suggest the authors rework this section to come up with something that highlights the new data but keeps the original structure.

Fig. 1 is much improved, but please state what each clade is (color and taxa) in the legend

Referee: 3

Comments to the Author(s).
Please see attached.

Referee: 1

Comments to the Author(s).
General point

In my original review I stated "it's hard to judge whether it's reporting a truly unusual phylogenetic scenario. There are a lot of gene families that show a patchy phylogenetic distribution within animals. In particular, the loss events in Ecdysozoa, platyhelminthes and vertebrates don't seem that remarkable."

I cannot see that any attempt has been made to address this, either in the ms itself or the response letter.

Specific points

Structural modelling concern - the section "OvoA is fused with a unique sulfide-lyase domain in Hydrozoa".

Despite the authors' assertion that the Hydra N-terminal domain "lacks significant primary sequence similarity with any protein deposited in public database." this is not the case. Searching with that sequence reveals hits to structures in the PDB and PFAM (I see this using blastp or phmmer). Phyre2 uses sequence similarity (HMM-HMM comparison, see Figure 1 in their Nature Protocols paper) in its pipeline to detect suitable templates. I-TASSER includes sequence similarity protocols via LOMETS. Once a template has been selected, via sequence similarity searches, there is no way that the modelled protein will not have that overall fold, so stating that the method 'revealed a significant structural match' between the modelled structure and the structure the query was aligned to in order to build the model does not add anything. The authors could report that Phyre or I-Tasser produced a significant similarity score at the template detection stage if they cannot demonstrate it via standard sequence database search, but reporting the significance of the similarity between the structure of a modelled protein and the known crystal structure used for modelling is, I think, actively misleading.

Concern about convergence between sponge and cartilaginous fish.

The authors stick by their hypothesis that the incongruence between gene and species trees is due to convergent evolution. The evidence presented in Figure S2 is far from compelling. Picking a fast-evolving urochordate, which one would expect to be more divergent, does not help. Under such circumstances two more slowly evolving taxa will inevitably appear more similar. The fact that the sponges appear paraphyletic in Fig 3, deuterostomes are unresolved, there is a clade of sponges + cnidarians etc. strongly suggests that the data aren't being modelled adequately (I don't blame the authors for that), so reading too much into the placement of any particular group looks rash. Not even mentioning in the main text that an alternative and, a priori very likely hypothesis, that the result is an artefact, seems wrong to me.

Author's Response to Decision Letter for (RSPB-2019-1812.R0)

See Appendix C.

Decision letter (RSPB-2019-1812.R1)

06-Nov-2019

Dear Dr Castellano

I am pleased to inform you that your manuscript entitled "The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

Open Access
You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700. Corresponding authors from member institutions (http://royalsocietypublishing.org/site/librarians/allmembers.xhtml) receive a 25% discount to these charges. For more information please visit http://royalsocietypublishing.org/open-access.

Your article has been estimated as being 10 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Paper charges
An e-mail request for payment of any related charges will be sent out after proof stage (within approximately 2-6 weeks). The preferred payment method is by credit card; however, other payment options are available

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,
Professor Loeske Kruuk
Editor, Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor:
Board Member
Comments to Author:
I think the authors have addressed all the major concerns, I would like to congratulate them for their work and efforts.

Appendix A

RSPB-2019-0683 Resubmission

Dear Editor,

thank you for the Reviewers' Report about our manuscript entitled "The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis" submitted for publication in Proceedings B. We have appreciated a lot the comments received by the Reviewers and yourself and have carefully considered them in preparing a new version of the manuscript for Resubmission.

A point-by-point response to the comments of Referees is attached below.

We believe that the quality of the manuscript is significantly improved thanks to the novel inputs.

We hope that the new version of the typescript deserve publication on *Proceedings of Royal Society B.*

Best regards

Immacolata Castellano & Marco Gerdol

Point by Point response

Referee: 1

Comments to the Author(s)

The paper examines the distribution of ovoA type genes in the animals. The authors do a good job of examining 'minor' phyla and report what they describe in the title as 'a complex evolutionary history'. On the whole, it's a solid analysis, but it's hard to judge whether it's reporting a truly unusual phylogenetic scenario. There are a lot of gene families that show a patchy phylogenetic distribution within animals. In particular, the loss events in Ecdysozoa, platyhelminthes and vertebrates don't seem that remarkable. Counterbalancing this, solid evidence of horizontal transfer into the Hydrozoa (or any animal) is definitely interesting (although at the same time, it's not the first reported example of such a transfer).

Specific points

The P.lividus ovoA sequence (uniprot:A0A140GAV9) doesn't have a hit to IPR024775 using Interproscan (http://www.ebi.ac.uk/interpro), at least as far as I can see. Neither with PFAM. I don't doubt these assignments, but the manuscript could be clearer as to what is being done.

Response:

Thank you for pointing this out. The observation made by the reviewer is correct, as the online web interface of InterProScan and Pfam often do not identify the DinB domain based on standard significance thresholds (which are hidden to the user). This is due to the fact the vast majority of the 134 "seed" sequences used to build the PFAM DinB profile PF12867 are of bacterial origin (based on interPro data, 97.68% of all DinB-like domain containing proteins

are from Bacteria). This leads to a bias towards the detection of the bacterial DinB molecular signature, whereas the few metazoan sequences which show this domain are, in many cases, so divergent from the "canonical" profile that they result in e-values close to (and sometimes lower than) the pre-set thresholds of detection.

When using the standalone version of HMMER however, as shown in one of the files attached, all metazoan proteins display a DinB domain with an e-value lower than the 0.05 e-value significance threshold, even though some of these reach values close to this limit. The sequences most closely matching the canonical signatures usually show e-values in the range of $1e^{-10}$ (the best one being *Exaiptasia pallida*, $2{,}9e^{-11}$). Sequences from cartilaginous fishes were those showing the least significant values (e.g. 0.00016 in *Heterodonthus zebra*). The e-value for DinB in the *P. lividus* sequence is 5.4e-07.

We have updated the text accordingly, specifying the use of HMMER and PFAM profiles with appropriate thresholds in Material and Methods section.

As far as I can see, the authors do not discuss whether the diverse ovoAs (i.e. Clade I, II and III) are all likely to catalyze the same chemical reaction. Is there evidence that the key catalytic residues are conserved? Could the authors demonstrate this in the manuscript? This will not be proof, but it does seem a basic test to perform. If I understand correctly Supp. Fig. 1 does this only for metazoan sequences.

Response:

The referee is correct, the previous supplementary Figure 1 only shows the conservation of metazoan sequences, including those horizontally transferred in rotifers and hydrozoans. Now we have inserted in Supp. Fig.1 also the conservation profile of sequences from Clade II and III, showing that the main residues involved in iron binding motif and in substrate specificity are well conserved. We have also inserted a sentence in the text explaining that the conservation of such residues are predictive of ovothiol production, as already demonstrated in some bacterial and unicellular eukaryotes (refs 5,6,7,13,14,16). Unfortunately, we cannot extend functional inference to the conservation of other residues found in other positions, as their role in the context of the chemical reactions carried out by OvoA are still unknown.

P.8 "Even though ovoA was detected in early-branching holozoan groups [such] as Ichthyosporea and choanoflagellates these sequences do not share the same ancestry with metazoan ovoA". Could the authors confirm that they have searched the data reported in Richter et al. Elife 1018:e34226 - doi: 10.7554/eLife.34226 ? To me, admittedly judging only from pairwise similarity scores, it looks like there are some metazoan clade ovoAs in there. Obviously this is quite an important point for some of the later discussion.

Response:

Thanks to the reviewer for his comments. We really appreciated the suggestion to use this data, as we were not aware of the release of such a complete catalogue of transcriptomes for choanoflagellates. In the previous version of this manuscript, we had considered all the sequences that had been deposited in public sequence databases at the time of submission, which only included the genome-derived sequences from *Monosiga brevicollis* and *Salpingoeca rosetta*,

but not the sequences from Richter et al., which had been deposited in a different database, external to NCBI. As explained in the materials and methods section, "The complexity of the dataset was reduced by clustering sequences sharing pairwise similarity > 55%", in order to keep it within a manageable size and the choanoflagellate representative sequence from *S. rosetta* sequence was included in the final dataset, being placed with high confidence within clade III.

Thanks to the suggestion provided by the reviewer, we extended our analysis to the data from Richter et al., which include the assembled transcriptomes from 19 additional choanoflagellate species. We need to thank the reviewer for this suggestion, as it allowed to identify that some – but not all- choanoflagellates possess metazoan-like OvoA sequences, pushing back the origin of animal OvoA genes compared to our previous estimates.

This finding has an important impact on the conclusions of this manuscript, and we are happy to add these new sequences to the phylogenetic trees and to extensively discuss these results in the text. Figures 1, 2, 3 have been updated, accordingly.

In detail, we found out that choanoflagellates pertaining to the order Acanthoecida possess metazoan-like OvoA genes, whereas the majority of species pertaining to the Craspedida order have OvoA genes falling within clade III, and therefore much more similar to those from dinoflagellates, cyanobacteria, and other non-metazoan organisms. There are some exceptions, though: indeed, both *Codosiga hollandica* and *Salpingoeca dolichothecata* possess metazoan-like OvoA genes. We see that the latter species in particular occupies a particular position in the tree shown in Fig. 3 in the paper by Richter and colleagues, far from the other species of the same genus and basal to all the other Craspedida (this probably indicates that the definition of the genus Salpingoeca is quite outdated). We could not detect any species possessing both types (clade I and clade III) of OvoA enzymes.

P.10 discussion of convergent evolution. I'm not sure that convergent evolution would be a widely accepted hypothesis for incongruence between a gene tree and species tree. What about inadequacies of the phylogenetic model for dealing with differing lineage specific rates as a null hypothesis? This also comes up in the discussion.

Response:

The phylogenetic tree was updated, with the important addition of chanoflagellate sequences as an outgroup. This definitely helped to improve the rooting of the tree, bringing additional information about the ancestral characters of OvoA. As a result, the majority of the incongruences between the species tree and the OvoA tree have been solved, with a major polytomy remaining at the basis of the Bilateria clade, which is consistent with a low number of available phylogenetically informative characters in the MSA file, as well with the rapid radiation of metazoan OvoA sequences during the Cambrian explosion.

However, one major incongruence remained, i.e. the placement of cartilaginous fish and sponges (exluding Calcarea) within the same clade. Since reviewer #2 has provided a similar suggestion, we now provide additional data in the supplementary material, studying more in detail the relative level of conservation of OvoA sequence between these two groups and Urochordata, the evolutionarily closest phyla to sharks and rays, among those with OvoA genes.

This enabled us to detect that, quite surprisingly, cartilaginous fish and sponges share a much higher number of residues compared the urochordates/sharks and urochordate/sponges pairs. In detail, Chondrichthyes and Porifera share more than the double of residues than Urochordata and Porifera (60 vs 27), and show 54% more shared residues than Chondrichthyes and Urochordata (60 vs 39). The most conserved residues were however not equally distributed along the OvoA polypeptide, but they were mostly present in the FGE sulfatase and SAM transferase domains.

Simplified trees, including choanoflagellates as outgroups, and considering the three main functional domains of OvoA separately, show that: (i) the analysis of the DinB domain results in tree topology consistent with species phylogeny; (ii) the analyses based on the FGE sulfatase and SAM transferase domains result in tree topologies which disagree with species phylogeny.

Overall, considering the little knowledge available about the functional importance of specific OvoA residues in the enzymatic activity of this protein in animal species other than echinoderms, and some bacteria (ref. 5 and 14), we cannot fully explain this unexpected high degree of similarity between the sequences from sponges, sharks and rays. Multiple factors might have contributed, with convergent evolution still being one of the possible causes. Indeed, Bayesian trees, despite being less prone to errors related with homoplasy that ML and MP analyses, are not completely exempt from these errors (see Brandley et al., 2009, 1093/sysbio/syp019). Convergent evolution between the sequences of these two distantly related phyla may be possibly linked to the production of the same type of ovothiols, which might be different from those identified so far in other metazoan taxa. With this respect, we expect that the identification of ovothiols from unexplored taxa (including cartilaginous fishes), as well as the biochemical characterization of recombinant OvoA from sponges and cartilaginous fish might clarify the link between apparent sequence convergence and functional differences. Largely different rates of molecular evolution across lineages, combined with inadequacies in the model of molecular evolution suggested by ModelTest, as well as by the low amount of informative characters to resolve the basal nodes of the trees, are other options that need to be taken into account. Chondrichthyes are well-known to be most slowly evolving vertebrates, as shown by whole-genome-scale studies (Venkatesh et al. 2014 and Hara et al., 2018), and the retention of fairly a high number of "ancestral" residues, shared with Porifera, might be consistent with this view.

P.14 There is generally no way that template based structural modelling will not give a similar structure to the starting template. That's the point of using a template and not doing ab initio structure prediction. I realise that the literature has a lot of examples like this, but I would raise the same criticism about all of them. I would not let this section stand unless the authors can demonstrate that it reveals something new. It would be better to just report significant sequence similarity scores.

Response:

We apologize for misunderstanding about this point, probably due to the incorrect reference to Phyre 2 as a "homology modeling" method in the manuscript; this has been amended in the updated version. Neither Phyre2, nor I-TASSER use user-selected templates for modeling, as they are fold-recognition/threading methods. Both approaches screen the entire PDB database of experimentally determined 3D protein structures to identify the most likely templates for

modeling and, if no viable structural template is found, no model can be produced (unless the implementation of *ab initio* modeling is specified by the user).

To date, PDB includes 47254 3D structures, and the templates reported in the manuscript (cyanobacterial L-cystine C-S lyases) have been selected by the algorithms used by the two approaches (and not selected by us) based on strict probabilistic criteria, out of over 45 thousand possible alternatives, in absence of significant primary sequence similarity (detectable by BLAST). This is the point of using threading approaches to predict the structural folding, and both Phyre2 and I-TASSER are precisely intended to work whenever the primary sequence identity between the template and the query sequence falls below the limit of sensitivity of BLAST-based approaches, making the use of homology-modeling approaches inappropriate.

While we agree with the referee about the limitations of this method (e.g. it cannot be used for studying the catalytic mechanisms of enzymes and to investigate ligand-binding surfaces), we need to remark that threading is a widely accepted method to infer protein 3D structures.

However, based on the suggestion provided by the referee, we have shortened this section, keeping it to the most interesting result, which is the striking structural similarity between the N-terminal domain of the hydrozoan OvoA proteins and OvoB, which suggests that the horizontally –transferred genes of these organisms may combine OvoA and OvoB-like catalytic activities in the same polypeptide. We have also added a panel to Fig. 5 to show the peculiar architecture of the protein domains in Hydra.

Minor points

The opening paragraph speaks about 'adaptive evolution', 'strategies to cope' and 'environmental adaptation', but the manuscript doesn't really have anything to say about these things - it's a study of a gene family, not an evolutionary process.

Response:

We agreed. We have modified the opening paragraph accordingly.

P.6 'a few sequences from unicellular prokaryotes (i.e. Ichthyosporea...' the taxa listed are eukaryotes not prokaryotes.

Response:

Thank you. This was a typo, we meant “eukaryotes”

Although the manuscript is entirely comprehensible, if the authors could get a native speaker to go over it, it would benefit. E.g.:

Abstract - 'originated by the fusion between' would be better as 'originating from the fusion of'

Introduction - 'early embryonal phases' -> early embryonic phases

Results - 'OvoA' by itself will be interpreted as singular. So better '[...] OvoA were found in' -> 'OvoAs were found in' or 'OvoAtranscripts/orthologs...' or 'OvoA was found in...'

Discussion - 'have played a capital role' -> 'an important role'?

Response:

We have carefully revised the text and improved it accordingly

Referee: 2

Comments to the Author(s)
This is comprehensive analysis of the evolution of the OvoA gene family with a particular emphasis on its evolution in metazoans. For the most part, the research has been undertaken in a thorough and competent manner. However, there are some substantial gaps that if addressed could improve this report.

First, the documentation and discussion of the two putative horizontal gene transfer events in bdelloid rotifers and hydrozoan cnidarians need to be greatly expanded and integrated in the manuscript and its figures. For example, while the hydrozoan HGT is mentioned throughout the Results and Discussion but there is essentially no data shown, except in Figure 1, which is a low-resolution tree of the OvoA genes across life. In the bdelloid example, this needs to be more explicitly compared with the other rotifer clades (monogonts) that have the metazoan-specific form; the same can be said for the hydrozoan genes compared to other cnidarians. In addition, the authors need to demonstrate convincingly that these putative HGTs are integrated into the host genomes and not contaminants – i.e. are the OvoA genes on contigs populated by metazoan/eukaryotic genes. In summary, this is an interesting find but it needs to be much better documented.

Response:

Thank you for this suggestion. This gave us the opportunity to include additional data to the manuscript, which had originally been removed due to strict space constraint of the journal. While some of the data requested by the referee has been directly added to the main text, other information which could not be included due to lack of space is now reported in detail in the supplementary materials. In general, the two putative HGT events are now discussed more in detail in the text, and additional evidence has been added to exclude the possibility of exogenous contamination, which is actually the first hypothesis we had originally ruled out upon the detection of such sequences.

The first improvement made is the addition of two additional panels to Figure 1, the first one providing a zoom-in on the hydrozoan HGT event in clade II, and the second one providing a zoom-in on the Rotifera HGT event in clade III. Both trees present an improved visualization of the low-resolution Bayesian tree found in Figure 1A, providing details on branching patterns, posterior probability values and enabling a more in depth contextualization of the placement of rotiferan and hydrozoan sequences.

Concerning the possible exogenous contamination as the cause for the finding of an unusual OvoA sequence in hydrozoans, we have no reason to believe that this issue is present in the *Hydra vulgaris* genome assembly published by Chapman and colleagues in 2010, especially considering that the genome of this organism has been entirely resequenced in 2015 using long-range information.

In the *Hydra* genome v.1.0, the OvoA gene was placed in a 157 Kb-long scaffold, whose size increased to 550Kb in Hydra v.2.0 (scaffold 299). The gene is flanked by: (i) the regulator complex protein LAMTOR3-A-like gene, which, based on BLAST homology, displays high homology with other Cnidarian sequences (e.g. *Exaiptasia pallida*, *Pocillopora damicornis*, etc.); (ii) an uncharacterized gene, which appears to be present in several dozen nearly-identical copies in the *Hydra* genome. Another important information is that orthologous genes are found in the two recently sequenced and assembled (but still not annotated and unpublished) genomes of the congeneric species *H. viridissima* and *H. oligactis*. In this case however, the assemblies were too much fragmented (i.e. the genes are located on 13K and 17Kb-long scaffolds, respectively) to assess whether the flanking genes are the same as in *H. vulgaris*.

Overall, we believe this data to provide a solid support to the integration of the OvoA gene in hydrozoan genomes.

As far as rotifers are concerned, bdelloid genomes have been previously demonstrated to contain enormous amount of horizontally transferred genomic material (>8% genes in *Adineta vaga*), so the finding that OvoA genes have been horizontally transferred would not be particularly surprising in this case. All rotiferan genes from different species appear to be placed in relatively large scaffolds, which contain several genes of clear metazoan ancestry. For example, the *A. vaga* gene GSADVG00064524001 is included in a 60 Kb super-scaffold which includes several genes with introns, ruling out the possibility of its exogenous origin from bacterial contaminants. Similarly, OvoA genes from other bdelloid species are placed on very large scaffolds, such as in the case of *Adineta ricciae* (90Kb), *Rotaria magnacalcarata* (60Kb) and *Rotaria macrura* (80Kb).

Second, the primary tree in the paper (Fig. 3) has a topology that does not match the organismal tree (e.g. OvoA genes from sponges group with Chondrichthyes OvoA genes; there are a number of similar cases on the tree). While the authors note that this may be because of convergence, the details are not provided. Are the residues shared between these widely-separated animal lineages tell us anything about protein function in the marine environment, for instance? Can these be mapped onto the 3-D models? Again a much more detailed analysis is likely to shed light on what parts of the OvoA protein underlie these apparent convergences.

Response:

See the response provided to the question made by reviewer #1 above. We have explored this aspect in detail, upon the creation of an improved phylogenetic tree which used Choanoflagellata as outgroups. While most of the incongruences between the gene and the species tree have been solved, the clustering of sequences from sponges (excluding Calcarea), shark and rays remained unchanged.

We noted that most of the unexpectedly high number of residues shared by these two phyla (but not by the closest relatives to cartilaginous fish in the tree, i.e. urochordates) are located within the FGE sulfatase and SAM transferase domains. The reasons for the relatively high similarity between the sequences from these two distantly related phyla is still unexplained, but our data suggest that some key residues are identical in these two important functional domains. While a general functional convergence due to the life in the marine environment can be excluded (many other marine species are included in our analysis), we believe that the two most likely explanations might be found in: (i) the production of similar, but still unexplored, types of

ovothiols, different from all other metazoans; (ii) the slow evolving features of the genomes of cartilaginous fish, which may have contributed to the retention of a high number of ancestral characters in OvoA.

Unfortunately, as no 3D structure is available for OvoA, we could not map any of these residues. Fold recognition approaches could be used, like we did for the PLP-dependent transferase domain of the *Hydra* sequence, but such methods are not sufficiently accurate to perform predictions about the possible function of individual residues, and they can be simply used to infer the overall 3D arrangement of polypeptidic chains.

In addition there are a number of minor points that should be addressed:

Fig. 1 – figure and legend need more details about clade, scale bars, etc.

Response: details have been added.

P9 -sponges are not the most ancient animal group – they are the most ancient extant phyletic lineage

Response: fixed

P9/10 – ctenophore/sponge debate remains unresolved – suggest that you do not make such a definitive statement about ctenophores being the sister lineage to all other animals. It is well known the ctenophores have lost many genes. There are two alternative evolutionary scenarios that should be considered equally.

Response:

We agreed, we have updated the text accordingly (see page 8), by stating that ctenophores are "identified by some authors as the most ancient extant animal taxa". Figure 2 has been modified based on this suggestion, by placing Porifera and Ctenonophora in a polytomic basal node.

P10- Parahoxzoa

Response: fixed

P10 – platyhelminths, mesozoans and gastrotrichs are not necessarily basil lophotrochozoans (see Current Biology 29, 312–318, 2019)

Response:

Thank you for pointing this out, the text has been revised accordingly (see page 8), by referring to these taxa as "marginally studied Spiralian species with debated taxonomical placement".

P10 - xenoacoelomorphs are not at the 'crossroads' between protostomes and deuterostomes – they are the earliest branching bilaterian phyletic lineage.

Response: fixed

Fig. 2 – most metazoan trees are drawn as a basal polytomy with sponges, ctenophores and all other animals (parahoxzoans) branching off a common ancestor

Response:

the tree has been changed to reflect the novel finding about choanoflagellates, which are now used to root the tree. The node with sponges, ctenophores and parahoxzoans is now shown as a polytomy.

Fig.2 spelling mistakes in legend

Response: fixed

P13 – CDS?

Response: we meant coding sequence

Fig. 4 – Why draw the exon-intron architecture based on sea urchin? Given the conservation and phylogenetic position of sponges, placozoans and anthozoans, their consensus gene architecture should be the standard for comparison. Why aren't the hydrozoan and the monogont rotifer genes shown?

Response:

In the previous version of our manuscript we used sea urchin exon-intron architecture because was the first reported in our previous paper (ref. 5). However, based on the new data gathered thanks to the suggestion provided by referee #1, choanoflagellates would serve this purpose even better, but unfortunately no genome is currently available for the species that carry metazoan-like OvoA genes. We updated the picture and now use the sponge *A. queenslandica* as a reference.

Thanks for pointing out the lack of monogonont rotifers in this scheme, which was due to the lack of well-assembled genomes at the time this figure was initially prepared. We have added this information thanks to the analysis of the *Brachionus plicatilis* genome data. The gene architecture in this case is consistent with a shared origin with the other metazoan ovoA genes, as most intron/exon boundaries are conserved, even though the *B. plicatilis* gene only includes 10 exons, due to the loss of several introns (which is consistent with the relevant shrinkage of the genome of this species).

We had originally considered including the hydrozoan gene structure as well, but this option was dropped due to the much different length of the N-terminal region (which includes and extra domain), and the particular rearrangements of the C-terminal domain of this particular sequence. Since the Hydra gene is, most likely, the product of a gene fusion event, its total length does not match with that of the classical metazoan sequences. We have however added an additional supplementary figure (Sup. Fig. 4) which reports this information, together with the gene structure of the chromerid *Vitrella brassicaformis* gene, also answering another request listed below.

P17 – Speculation as to the source of hydrozoan HGT needs to be based on strong phylogenetic evidence and not on the presences of modern day symbionts in various cnidarian lineages. HGT events happened in the past and may not reflect modern animal- microbe associations. This part of the discussion needs to be completely reworked.

R: Thanks for pointing this out. We are now providing a detailed view of the branching patterns of interest in two additional panels of Figure 1. This shows that the three available sequences from hydrozoans (i.e. *Hydractinia symbiolongicarpus*, *Hydra viridissima* and *Hydra vulgaris*) are clustered with 100% posterior probability with the sequence from the chromerid *Vitrella brassicaformis*.

We most definitely agree about the fact that past HGT event might not necessarily reflect current animal-microbe associations. However, cnidarians have evolved in close association with unicellular photosynthetic symbionts for several million years, such as in the case of *Symbiodinium* spp., which is thought to have originated in the early Eocene (Pochon et al., 2006, https://doi.org/10.1016/j.ympev.2005.04.028) or, according to some authors, even earlier (Tchernov et al., 2004, https://doi.org/10.1073/pnas.0402907101).

Only two species of extant chromerids have been described so far, namely *V. brassicaformis* and *Chromera velia* (which also possess an OvoA gene, not included in the present manuscript due to its incompleteness), which are both obligated parasites of stony corals. The current consensus of the scientific community about the evolution of this particular group of apicomplexan- and dinoflagellate-related organisms is that they may have derived from ancient apicomplexans present in symbiotic association with coral reefs (see Janouskovec et al, 2015, https://dx.doi.org/10.1073%2Fpnas.1423790112).

Since

- **(i) the chromerid/cnidarian symbiosis has very ancient roots**
- **(ii) (ii) the three hydrozoan sequences were clustered with high support (posterior probability = 1) with *V. brassicaformis***

we hypothesize that the most likely evolutionary scenario that has led to the acquisition of OvoA in hydrozoans is an HGT event from an apicomplexan/chromerid ancestor that had a symbiotic relationship with a hydrozoan ancestor.

Unfortunately, due to the paucity of genomic data available for chromerids (and the low number of extant species, i.e. 2), we cannot further investigate the details of this hypothetical event, nor provide a precise time estimate. In any case, such event needs to be placed after the split between the Hydrozoan lineage from other Cnidaria (considering the presence/absence patterns identified in our study), but before the split between the two major hydrozoan subclasses, Trachylinae and Hydroidoline (since both subclasses possess OvoA genes).

Given the very different gene architecture of the chromerid and hydrozoan genes, as well as the presence of the additional PLP-dependent transferase in hydrozoans, the HGT event is expected to have been followed by intron gain/loss events and gene fusion with an OvoB-like gene acquired with a different HGT event from an unknown donor of genetic material.

We have added additional information to the supplementary materials, namely:

1) **The aforementioned zoomed-in version of the Bayesian phylogenetic tree**

2) **Supplementary figure 4 showing the exon/intron structure of the *V. brassicaformis* gene, and its comparison with the architecture of the *Hydra* gene. The domain organization of the two encoded proteins is also shown.**

Appendix B

In this revised manuscript, Gerdol et al discuss the evolution of the OvoA protein family across the tree of life with an emphasis on the sequences present in metazoa. I think that the authors have done an adequate job addressing the concerns of the previous two reviewers. I agree with Reviewer 1, that altogether this is a solid analysis however I do not see this as an entirely unexpected phenomenon in eukaryotes in general. That said, these sorts of studies are still important and I suspect that this systematic survey will be an excellent resource for future research.

I just have a few comments/questions for the authors:

1. Given the breadth of metazoan genomes, could the authors consider doing phylogenomic profiling of the lineages with and without the OvoA gene? Perhaps the presence of ovothiols correlates with some other pathways? Although this is definitely beyond scope of the study, but the authors could investigate using CLIME or similar phylogenomic profiling tools. Moreover, is there any link between the oxygen tolerance of these lineages and whether or not they have ovothiols? That is, do worms that experience low and modern oxygen conditions have different types of OvoA than strict aerobes?

2. For the HGT cases, can the authors comment on whether they assess codon usage bias or GC content as an additional line of evidence for NOT being a contamination? That is, if you could compare the sequence characterisitcis of the putative HGT case and the bona fide eukaryotes genes?

3. Can the authors provide the unedited tree files and alignments to a data repository? Perhaps a purely aesthetic comment, but could the authors write out the species names for Figure 3? Its not that many and would make the taxa easier to recognize for non-experts. I was interested in seeing the positions of the non-metazoan sequences but I could not find the full unedited tree.

4. Where are these thiols synthesized in the cell? Is it in the cytoplasm or an organelle?

5. Methods comments:
    a. Typically, people report both maximum likelihood and Bayesian methods. Is the a reason that only Bayesian analysis was performed?
    b. Model Selection – WAG is a rather outdated model and there are plenty of new models that have been reported to better fit biological data. Can the authors comment on why they did not investigate more recent models (e.g., LG or mixture models)?

6. It would be nice to see more discussion about the biology of this molecule and how the ability to synthesize ovothiols relates to the biology of both multicellular and unicellular organisms. Just one or two more sentences in the Discussion.

MINOR SUGGESTIONS

1. Line 145. I do not think that 'Eukaryotes' should be capitalized unless referring to the entirety of eukaryotes.
2. Line 146 'amoebas' – should be amoebozoa? or Amoebae?
3. Use of protozoa vs protists – Some prefer the term protist as 'protozoa' implies 'animal'

Appendix C

RSPB-2019-0683

Dear Editor,

thank you for the Reviewers' Report about our manuscript entitled "The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis" submitted for publication in *Proceedings of Royal Society B*. We have appreciated the comments of the Reviewers and Associate Editor and have carefully considered them in preparing a revised version of the manuscript.

A point-by-point response to the comments is attached below.

We believe that the quality of the manuscript is now significantly improved.

We hope that the new version of the manuscript deserve publication on *Proceedings of Royal Society B*.

Best regards

Immacolata Castellano & Marco Gerdol

Point by Point response

Associate Editor
Comments to Author:
I'd like to thank the authors for the efforts made to improve the paper. I think that the manuscript has improved much, however I think there are still some important issues standing.

The abstract is now a bit convoluted and hard to follow. I believe it could be streamlined and the content could be reorganised to improve the flow.

***Thank you for this observation. We reorganized the original structure of the abstract, including the most relevant results of this work. We hope the text of the abstract now flows better.**

I agree with some of the comments from the referees, the authors should take then into consideration, namely further support for LGT (e.g., GC composition and codon bias), the claims on gene loss in specific animal lineages, lack of gene tree resolution vs convergence, evolutionary model selection, and the structural modelling. Please, also consider improving Figure 1 and 3.

***We followed the suggestions provided by the referees and a detailed response to each point is provided below in response to specific points made by the 3 reviewers. Following the most relevant changes made with respect to the observations made by the editor:**

-We have added an additional supplementary data note which shows that the OvoA genes from *Hydra vulgaris* and *Adineta vaga* display a GC content and codon usage (ENC) very similar to that of the average values of the coding sequences annotated in the respective genomes. This data rule out the possibility that such sequences were included in the genome assemblies due to contamination from exogenous DNA. On the other hand, genes acquired by HGT are usually

subject to strong selection and, if fixed in the new genome, over time they tend to modify their GC content and codon composition to match that of the acquiring genome. Hence, codon usage-based metrics are not always useful indicators of HGT (see https://doi.org/10.1093/oxfordjournals.molbev.a003816 and https://doi.org/10.1007/s00284-012-0205-5).We believe this might be also the case of the rotiferan and hydrozoan OvoA HGT events, which we infer to have occurred in the early stages of metazoan radiation.

-We have improved the discussion of the relationship between OvoA gene loss and the well-documented massive gene loss and genome reduction events that have been documented in some phyla. We believe that referee #1 might have missed some of the modifications that we had already included in the previous version, as this aspect had been already discussed with relevant references, but we have further improved this point by providing additional references and discussion, and by adding some information to Figure 2.

-Reviewer 3 was right about model selection. We realized that due to an error in ModeTest-NG, the LG model of molecular evolution was not among the 8 models selected for testing with the "-T mrbayes" option", even though LG is indeed one of the models implemented in MrBayes. Consequently, we have re-evaluated the best-fitting models, and re-run the phylogenetic analyses with the LG-G+I model. The tree topology was nearly unchanged for Figure 1, but significantly changed for Figure 3. As explained in detail in the response to a comment by reviewer #1, we feel that this figure did not add much to the work and clearly suffered from lack of phylogenetic signal, and we have consequently moved it to the supplementary material, significantly reducing the part of text which dealt with this aspect. Moreover, phylogenetic analyses were also carried out with ML methods, and these results are now presented in the supplementary material.

-We have provided a detailed response about structural modeling, since we believe that our explanation has not been clear enough, and the presence in the text of a wrong statement about the absence of primary sequence homology for the Hydra N-terminal domain certainly contributed to bring further confusion (this has been amended in the present version).
We compared the crystal structure of OvoB, which is the lyase involved in ovothiol biosynthesis, with the predicted 3D model of the Hydra lyase-like domain. Reviewer #1 probably believes that (i) the Hydra N-terminal domain shares primary sequence homology with ovoB and (ii) that its structure was actually modeled using OvoB as a structural template. Indeed, the Hydra lyase-like domain lacks significant primary sequence homology with OvoB, while it shares significant homology with EgtE, the lyase involved in ergothioneine biosynthesis. Therefore, OvoB was not among the selected templates for modeling, neither by Phyre 2, nor by I-Tasser.
Although we have toned down some sentences in the section which concerns structural modeling, we believe that the primary aim of structural modeling approaches is to point out the most likely 3D fold adopted by a sequence whose structure is presently unknown. We are not making any functional assumption at this stage, we would simply like to point out that the model obtained for the Hydra N-terminal domain (whose best template is EgtE) is structurally similar to OvoB, and we believe this is quite relevant, considering that OvoB is the specific lyase involved in ovothiol biosynthesis, and that so far no OvoB-like have been found in metazoans.

-Figure 1 has been implemented as suggested by reviewer #3 and Figure 3, as mentioned above, has been moved to the supplementary materials, with the full names of the species added.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).
General point

In my original review I stated "it's hard to judge whether it's reporting a truly unusual phylogenetic scenario. There are a lot of gene families that show a patchy phylogenetic distribution within animals. In particular, the loss events in Ecdysozoa, platyhelminthes and vertebrates don't seem that remarkable."

I cannot see that any attempt has been made to address this, either in
the ms itself or the response letter.

***We apologize to the reviewer because we forgot to include a specific remark about this point in the response letter, but we had included two statements in the manuscript that we believe the reviewer may have missed:**

1) **"*our analyses allowed the identification of several independent gene loss events, often concurrent with well-documented lineage-specific massive orthologous gene loss events that occurred along animal evolution*", providing a reference to the work by Wyder and colleagues (reporting orthologous gene loss in insects and vertebrates).**
2) ***"On several occasions, large-scale rearrangements of genome structure, accompanied by massive gene loss, can be advocated as the primary cause of the loss of OvoA. These include the genome shrinkage typical of obligated parasites, such as Myxozoa as well as the gene loss events documented in Ecdysozoa"***

We agree with the observation provided by the reviewer, and we do not claim in the text that this distribution pattern is particularly surprising. We also clearly mention that gene loss is a dominating process in genome evolution, adding an important reference: "*the evolution of genomes appears to be dominated by reduction and simplification, punctuated by episodes of complexification*" (Wolf and Koonin, 2013 https://dx.doi.org/10.1002%2Fbies.201300037).

Aside from the aforementioned cases of vertebrates and Ecdysozoa, the loss of OvoA in taxa which mostly include animals with an obligated parasitic lifestyle (e.g. flatworms or myxozoans) or with a highly simplified body plan (e.g. mesozoans) would be consistent with the genome reduction events which typically occur in these organisms. Since the text length constraints for the main text in this journal are quite strict, we could not discuss this aspect in detail. Nevertheless, we added some additional key references to the text, and we have highlighted the documented cases of genome shrinkage with an asterisk (providing relevant references) in Figure 2.

These include, in general, parasitic worms (Jackson, 2015, https://dx.doi.org/10.1017%2FS0031182014001516); Platyhelminthes, Zarowiecki and Berriman 2015, https://dx.doi.org/10.1017%2FS0031182014001449); myxzoans (Chang et al., 2015 https://doi.org/10.1073/pnas.1511468112); mesozoans (Lu et al., 2019 https://doi.org/10.1093/gbe/evz157).

Altogether, we believe that these additions cover quite well the issue and clearly show that several taxa were OvoA is missing have also reported to undergone genome shrinkage.

Specific points

Structural modelling concern - the section "OvoA is fused with a unique sulfide-lyase domain in Hydrozoa".

Despite the authors' assertion that the Hydra N-terminal domain "lacks significant primary sequence similarity with any protein deposited in public database." this is not the case. Searching with that sequence reveals hits to structures in the PDB and PFAM (I see this using blastp or phmmer). Phyre2 uses sequence similarity (HMM-HMM comparison, see Figure 1 in their Nature Protocols paper) in its pipeline to detect suitable templates. I-TASSER includes sequence similarity protocols via LOMETS. Once a template has been selected, via sequence similarity searches, there is no way that the modelled protein will not have that overall fold, so stating that the method 'revealed a significant structural match' between the modelled structure and the structure the query was aligned to in order to build the model does not add anything.
The authors could report that Phyre or I-Tasser produced a significant similarity score at the template detection stage if they cannot demonstrate it via standard sequence database search, but reporting the significance of the similarity between the structure of a modelled protein and the known crystal structure used for modelling is, I think,actively misleading.

***We would like to thank the reviewer for his/her valuable suggestions. The reviewer is right about the presence of significant homology. The reported lack of primary sequence similarity for the N-terminal region was, we believe, the product of setting the maximum number of BLAST hits to 100, paired with the fact that the most significant matches hit the C-terminal region. As suggested by the referee, upon further scrutiny, we found that some matches between the Hydra lyase-like domain and other metazoan sequences were indeed present, even though such similarities were rather modest (20-25% identity over an alignment of 200-300 residues, using a BLOSUM45 matrix, and word size = 2). Matched proteins only contained the PLP-like lyase domain, and shared significant similarity with L-cysteine desulfhydrases from higher plants (>30% sequence identity, with >50% positives), displaying much weaker hits with mitochondrial cysteine desulfurases from other metazoans. We have modified the text accordingly.**

While we understand the points made by the referee concerning template selection, we believe that the referee missed a key point: the OvoB crystal structure was NOT one of the templates selected for modelling the structure of the Hydra lyase-like domain, neither by Phyre2, nor by I-

Tasser. Hence, we did not compare a 3D model with its own structural template (we totally agree that such approach would have not made much sense), but we have rather compared the only known crystal structure for OvoB, which is the lyase involved in ovothiol biosynthesis, with an independently obtained 3D model (whose best template is that of EgtE, which is the lyase involved in ergothioneine biosynthesis). Although EgtE and OvoB belong to the same super-family of lyase, they are not orthologous proteins (they share more or less 25% identity). We are sorry that this was not clear in the previous version of the manuscript, and we have amended this lack of clarity in the revised version, also adding some explaining sentences in the introduction.

The Hydra lyase-like domain and OvoB do not share any significant primary sequence similarity: although, with the aid of structural information, small stretches of amino acids can be manually aligned, such homology cannot be detected as significant neither by blastp, nor by pHMMER. The primary sequence similarity between the Hydra domain and OvoB was so weak that OvoB was not even selected among the top 120 most highly ranked templates for modeling by Phyre2.

As correctly mentioned by the referee, there are several other sequences deposited in the PDB database that do actually show significant similarity with the Hydra sequence domain, and these have been indeed selected as templates for modeling (e.g. the *Neurospora crassa* Etg2).

We have tried to tone down the claims we made in this section of the text, being well aware of the limitations of our analysis and of the drawbacks of structural modeling in general. In our view, the key information that needs to be reported here is that the Hydra lyase-like domain is likely to adopt a structural fold very similar to that of OvoB. Without making any functional assumption, we believe this observation might stimulate further investigations concerning the enzymatic activity carried out by this domain in Hydrozoa in light of the inferred structural similarity with OvoB. Indeed, this would be the first indication of the existence of an OvoB-like lyase in metazoans.

Concern about convergence between sponge and cartilaginous fish.

The authors stick by their hypothesis that the incongruence between gene and species trees is due to convergent evolution. The evidence presented in Figure S2 is far from compelling. Picking a fast-evolving urochordate, which one would expect to be more divergent, does not help.
Under such circumstances two more slowly evolving taxa will inevitably appear more similar. The fact that the sponges appear paraphyletic in Fig 3, deuterostomes are unresolved, there is a clade of sponges +cnidarians etc. strongly suggests that the data aren't being modelled adequately (I don't blame the authors for that), so reading too much into the placement of any particular group looks rash. Not even mentioning in the main text that an alternative and, a priori very likely hypothesis, that the result is an artefact, seems wrong to me.

***Thank you for this observation. Following the suggestion provided by reviewer #3, we believe we may have identified one of the issue behind the lack of an adequate model. In summary, we used Modeltest-NG v.0.1.3 to identify the best-fitting molecular model of evolution, selecting a subset of models suitable for MrBayes analysis with the "-*T mrbayes*" option. We realized that, in spite of the possibility to select the LG model in the most recent released version of MrBayes, Modeltest-**

NG did not include this model among the 8 model tested with the "*-T mrbayes*" option (only WAG, Blosum62, VT, RTREV, CPREV, MTREV, MTMAM and Dayhoff were included). We have re-run the analyses and we found that the LG model (+I+G) was the best-fitting one for both datasets. Consequently, we have replaced the trees generates using the WAG model with new ones generated with the LG model (further implimeneting such analyses by parallel ML analyses performed with RaxML).

Overall, the tree topology and support of the global tree reported in Figure 1 did not change significantly. However, some branching patters of the tree reported in figure 3 changed significantly, affecting in particular the relationship between cartilaginous fish (now grouped, albeit with wek posterior probability support, with urohordates) and sponges. However, we found out that several incongruences still remained (e.g. Porifera were still paraphyletic) and other species were misplaced. Moreover, the fact that a large polytomic node remains evident suggests that our dataset suffers from a lack of phylogenetically informative sites and limited taxonomical breath of sampling to fully resolve the evolutionary history of metazoan OvoA sequences. Hence, we agree about the consideration that the discrepancies between the OvoA gene tree and the species evolution tree should not be linked with a functional interpretation in this case.

For all the reasons mentioned above, we don't believe this tree provides any additional significant information to this work, and we have chosen to move it to the supplementary materials. The text has been modified accordingly and shortened, and the whole paragraph has been reduced to a few sentences to discuss the results obtained.

We found that mentioning the fact that urochordates might be subject to faster evolutionary rates compared with sponges and cartilaginous fish was a fair pojnt, in particular considering the well-known extremely slow evolutionary rates observed in sharks. Hence, the supplementary figure with a comparison between the conserved sites in sponges, urochordates and cartilaginous fish has been removed.

Referee: 2

Comments to the Author(s).
This is an improved manuscript and should be suitable for publication after a number of minor adjustments.

Overall, please check spelling and grammar. There are a number of errors in the manuscript.

***Thank you for this suggestion. The manuscript has been carefully revised for spelling and grammar.**

The Abstract has been extensively rewritten and is now wordy and less precise than the original. Suggest the authors rework this section to come up with something that highlights the new data but keeps the original structure.

***The abstract has been revised, as suggested, with the inclusion of the new data.**

Fig. 1 is much improved, but please state what each clade is (color and taxa) in the legend

***Thank you for this suggestion. This information has been included in the figure caption.**

Referee: 3

Comments to the Author(s).
In this revised manuscript, Gerdol et al discuss the evolution of the OvoA protein family across the tree of life with an emphasis on the sequences present in metazoa. I think that the authors have done an adequate job addressing the concerns of the previous two reviewers. I agree with Reviewer 1, that altogether this is a solid analysis however I do not see this as an entirely unexpected phenomenon in eukaryotes in general. That said, these sorts of studies are still important and I suspect that this systematic survey will be an excellent resource for future research.

***Thank you for this positive assessment. Please see the response to reviewer #1 about the relationship between OvoA loss and documented massive gene loss events in Metazoa.**

I just have a few comments/questions for the authors:
1. Given the breadth of metazoan genomes, could the authors consider doing phylogenomic profiling of the lineages with and without the OvoA gene? Perhaps the presence of ovothiols correlates with some other pathways? Although this is definitely beyond scope of the study, but the authors could investigate using CLIME or similar phylogenomic profiling tools.

***This is an interesting suggestion, and something we would like to pursue in future works. In particular, the co-occurrence of other redox pathways is an aspect of great interest, since compensatory molecular mechanisms might explain some occurrences of gene loss.**
As recognized by the reviewer, a CLIME analysis would be more suited for a separate paper, as the preparation of an appropriate dataset of manually curated orthologs would most likely require several weeks, or even a couple of months. Moreover, we would not be able to select all the species presently included in this manuscript, simply due to the fact that in some cases we had to use transcriptomes as a replacement for genomes, as these resources are still not available for several taxa of key relevance (for example, Acanthoecida choanoflagellates). Whenever a transcriptome is taken into account, the absence of a given sequence from a dataset may simply be linked with a lack of expression rather than to its evolutionary loss. In summary, we are definitely going to consider this suggestion in an upcoming follow-up paper. We plan to select a subset of key taxa with fully-sequenced high-quality genomes available, focusing on a well-defined set of orthologs (most likely, the single copy ortholog metazoan dataset from OrthoDB 10), manually adding other candidate genes of interest involved in redox homeostasis.

Moreover, is there any link between the oxygen tolerance of these lineages and whether or not they have ovothiols? That is, do worms that experience low and modern oxygen conditions have different types of OvoA than strict aerobes?

***We have previously reported in the hemichordate *Saccoglossus kowalevskii* the presence of two genes coding for OvoA and we have discussed about the possibility of a double specialization for the OvoA to adapt to different conditions of oxygen pressure (Castellano et al, Sci Rep, 2016). Indeed, the hemichordate lives in burrows on the sea-bed in anaerobic conditions and during low tide it can be exposed to aerobic conditions. However, there are currently no proofs in support of this hypothesis. Moreover, the actual status of genome assemblies did not enable to discriminate between paralogous gene copies and uncollapsed allelic variants. In the Supplementary Data Note 3 of this manuscript, we report that two similar gene copies in *S. kowalevskii* are on the same genomic scaffold, but transcriptomic evidence from the congeneric species *Saccoglossus mereschkowskii* suggests that only a single locus is expressed, and that the second gene might be a pseudogenic paralog. However, in the absence of functional data we cannot exclude the presence of "specialilezed" OvoA.**

2. For the HGT cases, can the authors comment on whether they assess codon usage bias or GC content as an additional line of evidence for NOT being a contamination? That is, if you could compare the sequence characteristics of the putative HGT case and the bona fide eukaryotes genes?

***Thank you for this suggestion. This is indeed another evidence in support of the fact that HGT OvoA genes are actually encoded by the genomes of hydrozoans and rotifers. We now provide some additional data in the supplementary material concerning codon usage and GC content of the *Hydra vulgaris* and *Adineta vaga* OvoA genes, comparing these metrics with those of the reference genomes. In detail, we have computed the GC content and effective number of codons (ENC) for all the CDS regions annotated in the two genomes, and created scatter plots, evidencing the position of the gene encoding OvoA. In both cases, the OvoA genes displayed GC content and codon usage in line with the average for the species, which clearly supports the fact that both sequences are encoded by hydrozoan and rotiferan genomes and they are not the product of contamination from exogenous genomic DNA.**

3. Can the authors provide the unedited tree files and alignments to a data repository?
Perhaps a purely aesthetic comment, but could the authors write out the species names for Figure 3? Its not that many and would make the taxa easier to recognize for nonexperts.
I was interested in seeing the positions of the non-metazoan sequences but I could not find the full unedited tree.

***We now provide the tree files and alignments as supplementary material. Figure 3 has been edited and now includes full species names (please note that this figure has been moved to the supplementary materials).**

4. Where are these thiols synthesized in the cell? Is it in the cytoplasm or an organelle?

***The exact intracellular localization for the biosynthesis of these compounds is not known yet. Actually, we are performing functional studies on sea urchin larvae and microalge to address this aspect.**

5. Methods comments:

a. Typically, people report both maximum likelihood and Bayesian methods. Is the a reason that only Bayesian analysis was performed?
b. Model Selection – WAG is a rather outdated model and there are plenty of new models that have been reported to better fit biological data. Can the authors comment on why they did not investigate more recent models (e.g., LG or mixture models)?

***We would like to thank the reviewer for mentioning the LG model. Upon further checking, we realized that, even though the LG model was actually supported by the most recent release of MrBayes, it was not included among the 8 models checked by Modeltest-NG v.0.1.3 with the "*-T mrbayes*" option (used to select for testing the subset of models implemented in MrBayes). Consequently, this model was not considered during our previous analyses, and we have therefore re-run the analyses on the two dataset by manually adding LG among the models to be tested.**
We found that the best-fitting molecular model of evolution was LG+I+G for both datasets (the even better scoring LG4X model is still not implemented in MrBayes), and we have re-run the Bayesian analyses and updated the trees accordingly. Overall, changes in tree topology and posterior probability support values were minimal for the tree of Figure 1, but they were quite significant for the tree of Figure 3 (see a detailed discussion about this point in the response to reviewer #1).

About the presentation of Bayesian trees only, this was a deliberate choice made to keep the text as simple as possible due to the strict length constraints of PRSB, also taking into account that (i) we did not conceive this manuscript as a heavily phylogenetic inference-oriented work, and (ii) the tree topology obtained with the two methods was nearly identical.
As requested by the referee, we have included in the supplementary materials of the present version also the ML tree files and the alignments, along with a brief mention of the use of RAxML-NG in the materials and methods section, and a brief note about the similar results obtained with the two methods in the results section.

6. It would be nice to see more discussion about the biology of this molecule and how the ability to synthesize ovothiols relates to the biology of both multicellular and unicellular organisms. Just one or two more sentences in the Discussion.

***Thanks for this suggestion. We improved the discussion addressing this aspect.**

MINOR SUGGESTIONS
1. Line 145. I do not think that 'Eukaryotes' should be capitalized unless referring to the entirety of eukaryotes.
***This was fixed**

2. Line 146 'amoebas' – should be amoebozoa? or Amoebae?
This was changed to Amoebozoa.

3. Use of protozoa vs protists – Some prefer the term protist as 'protozoa' implies 'animal

We have replaced the term "Protozoa" with "protists.