

Peer Review File

The evolution of biogeochemical recycling by persistence-based selection



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Web links to the author's journal account have been redacted from the decision letters as indicated to maintain confidentiality.

31st Aug 21

Dear Dr Boyle,

Please allow me to apologise for the long delay in sending a decision on your manuscript titled "The evolution of biogeochemical recycling by persistence-based selection". It has now been seen by 3 reviewers, and I include their comments at the end of this message. They find your work of interest, but some important points are raised. We are interested in the possibility of publishing your study in Communications Earth & Environment, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response that takes into account the points raised. Please highlight all changes in the manuscript text file.

In particular, please ensure that your revised manuscript meets the following editorial thresholds:

****Provide a compelling case that your theoretical model is able to robustly demonstrate the maintenance of nutrient recycling loops even where individual-level selection works against this****

****Ensure your model and mathematical methodology is thoroughly and clearly explained****

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees' comments (which should be in a separate document to any cover letter) and the completed checklist:

[link redacted]

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to receive your revised paper within six weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We understand that due to the current global situation, the time required for revision may be longer than usual. We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning. Of course, if you are unable to estimate, we are happy to accommodate necessary extensions nevertheless.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the

opportunity to review your work.

Best regards,

Joe Aslin

Associate Editor,
Communications Earth & Environment
<https://www.nature.com/commsenv/>
Twitter: @CommsEarth

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

Editorial Policy: [Policy requirements](https://www.nature.com/documents/nr-editorial-policy-checklist.zip)

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist:

[Communications Earth & Environment formatting checklist](https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf)

and also in our style and formatting guide [Communications Earth & Environment formatting guide](https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf) .

***** DATA:** Communications Earth & Environment endorses the principles of the Enabling FAIR data project (<http://www.copdess.org/enabling-fair-data-project/>). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>.

In particular, the Data availability statement should include:

- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at <http://www.nature.com/sdata/policies/repositories>.

If a community resource is unavailable, data can be submitted to generalist repositories such as [figshare](https://figshare.com/) or [Dryad Digital Repository](http://datadryad.org/). Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at <http://www.nature.com/authors/policies/availability.html>.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

In this bracing and thought-provoking article, the authors provide a compelling case of a recent theoretical proposal that aims to extend Evolutionary Theory beyond its traditional sandbox. The biggest benefit from this is for Ecology and Earth Sciences where the intuition that evolutionary explanations would be worthwhile is there, but where the theoretical and experimental justification has not been fully fleshed out. The article aims at remedying this. The article is theoretically ambitious (in the best possible way) and methodologically sound. It will help other researchers generate new models, experiments and hypotheses, and it advances theory. As it is obvious, you have my strongest support for the publication of this article.

“Traditional” Evolutionary Theory is based on differential reproductive success. Some, including the authors, have argued that this view how ever powerful is insufficient to account for all biological adaptations and for other evolutionary processes. They have offered a view of evolution based on differential persistence (differential reproductive success being a large subset of differential persistence strategies). The biggest payoff of such a view is how it allows the inclusion of non-genetic mechanisms or even abiotic processes into evolutionary explanations, not merely as external forces or constraints, but as part of the evolutionary mechanisms themselves. This is an extremely fecund area of research for biological research in general, but for ecological and environmental research in particular because it allows us to treat the dynamics of biogeochemical processes as evolutionary processes (at least in some instances). How ever promising this view is, testing it is an ongoing challenge. This is exactly what the authors do here.

The authors show how a seemingly simple theoretical model can generate correlations and adaptive changes that are responses to selective pressures and changes in the external environment and yet cannot be fully attributed to changes in genotypes. As they correctly point out, this is not a multi-level selection story. Reproduction is insufficient to account for the change. This top-down resource

pressure tied with as they point out “worse-case” Neo-Darwinian population changes still generates positive changes in the system that can lead to stabilization of specific cycles. In itself these are timely and important results, but what makes this contribution especially important is that the model, they argue convincingly and rightly, plausibly resembles a simplification of the conditions that obtained during the assembly of biogeochemical cycles on the Precambrian Earth. This is major stuff. The Gaia Hypothesis has long suffered from the lack of a compelling “origin story”. What the authors provide is a compelling evolutionary story. For readers that may not find the Gaia hypothesis to be their cup of tea, there is still a lot here: even if it may not have played out at the planetary level, localized homeostatic loops may have emerged via the processes described by the authors. Doolittle’s ITSNTS approach demands careful examination and extension and this is what the authors have done. Excellent and important article.

Reviewer #2 (Remarks to the Author):

This paper says that it provides a model to explain how selection processes that operate at the timescale of biological evolution can influence the organization of biogeochemical cycles at geological timescales. Essentially, it attempts to provide a quantitative model of an interesting scenario proposed early by Doolittle.

The problem is important and the first couple of pages of the paper impressively lay out the many interesting issues. I am unable, however, to understand the paper after the authors' statement that "The basic structure of our model is already a feature of evolutionary-ecological thinking on syntrophy [31], symbiosis [32] and multi-level selection [33, 34]. We claim no originality in this regard..." I am not familiar with references 31-34, and the SI of the paper, for me, is not useful: it essentially contains computer code, complete with if-then statements, typeset as a kind of mathematics, but with variables written out as long words. I am unfortunately unable to understand what's written here, so I cannot offer an opinion concerning the paper's correctness and suitability for publication.

Reviewer #3 (Remarks to the Author):

In their paper, Boyle and Lenton develop and use a theoretical model to analyze the evolution of a biogeochemical cycle via persistence-based selection. The main question the authors ask is whether there are conditions under which a certain biogeochemical cycle emerges without invoking natural selection (i.e. selection favoring the cycle itself). Instead, conditions are identified, under which two types that produce or consume two environmental substances can persist despite the presence of two “cheaters” that each consume one of the substances without contributing to the production of the other one. In this scenario, producers generate one of the substance at a cost to themselves. In the presence of non-producers that do not pay the production cost but still benefit from the production of other individuals, natural selection is expected to operate against the cycle (i.e. the two reciprocally interacting individuals). Individuals switch between producer and cheater phenotype depending on the environmental concentration of one of the two substances and the offspring of the corresponding types changes between them with a certain frequency due to mutation. In simulations in which different parameters are systematically varied, the authors identify conditions in which the system can stably maintain the cycle even though individual-level selection

acts against it.

The paper addresses an interesting and relevant question. The provided answers are not only relevant to the emergence of biogeochemical cycles, but in particular also to the evolution of ecological interactions between organisms in general (e.g. gut microbiota). The presented work is founded on a very solid and accurate discussion of the available literature. The paper is an interesting read and represents a valuable extension of the available literature. Even though I could not understand the mathematical side of the model in all detail, I think that presentation of the data needs to be considerably improved to make the paper more accessible.

Main points:

- 1) It would have been nice to introduce a page and line numbering so that reviewers can explain more easily refer to a certain part of the text.
- 2) Page 4, first line: The term “facultative” is misleading and can have different meanings. I suggest to use the term “plastic”. “Phenotypic plasticity” is the phenomenon the authors are talking about.
- 3) In the discussion section, I found the argumentation why genetic drift and environmental fluctuations were likely stronger on early earth difficult to follow. Please explain this more carefully.
- 4) Presentation of data:
 - a) In general I think all graphs need to be reworked in order to improve the accessibility of the data. It took me a long time to understand what the authors wanted to show. This has mainly to do with the fact that the figures appear to be sloppily and inconsistently formatted.

E.g. in Figs. 2 and 3 legends seem to be randomly placed, sometimes overlapping with graphs themselves. “generations” as y-axis label should be removed except for the two panels at the bottom. The axis legends should start consistently with a capital or small letter. The headings of the rows could be placed at the left of the corresponding row.
 - b) Figure 4,7,8: I found it extremely difficult to see any difference in the individual graphs of these figures. I could barely distinguish the colours and see whether a data point is high or low. I think this way of presenting this data is not very reader-friendly and should be changed to a more accessible way. Depicting the data in a plane rather than showing individual data points could be one solution.
 - c) Figure 6: I did not understand why the graphs in panel D1-4 show so low average fixation probabilities when the graphs on the left always have a considerable amount of data points at the upper end of the corresponding distributions. Is this because of the given distribution within replicates at a particular parameter combination?
- 5) I generally dislike the term “cheater”, because it introduces wrong (anthropocentric) associations. A more neutral term such as simply “non-producer” appears to be more apt.

Minor comments:

- Abstract: please remove the word “show”
- I noticed that the authors sometimes did not include the Oxford comma. Please recheck the text.
- Last paragraph of the results section: replace “increases” with “increasing”

Responses to reviewer comments for “The evolution of biogeochemical recycling by persistence based selection”

Reviewer comments in black

Responses in blue

.....

Reviewer #1 (Remarks to the Author):

In this bracing and thought-provoking article, the authors provide a compelling case of a recent theoretical proposal that aims to extend Evolutionary Theory beyond its traditional sandbox. The biggest benefit from this is for Ecology and Earth Sciences where the intuition that evolutionary explanations would be worthwhile is there, but where the theoretical and experimental justification has not been fully fleshed out. The article aims at remedying this. The article is theoretically ambitious (in the best possible way) and methodologically sound. It will help other researchers generate new models, experiments and hypotheses, and it advances theory. As it is obvious, you have my strongest support for the publication of this article.

“Traditional” Evolutionary Theory is based on differential reproductive success. Some, including the authors, have argued that this view however powerful is insufficient to account for all biological adaptations and for other evolutionary processes. They have offered a view of evolution based on differential persistence (differential reproductive success being a large subset of differential persistence strategies). The biggest payoff of such a view is how it allows the inclusion of non-genetic mechanisms or even abiotic processes into evolutionary explanations, not merely as external forces or constraints, but as part of the evolutionary mechanisms themselves. This is an extremely fecund area of research for biological research in general, but for ecological and environmental research in particular because it allows us to treat the dynamics of biogeochemical processes as evolutionary processes (at least in some instances). However promising this view is, testing it is an ongoing challenge. This is exactly what the authors do here. The authors show how a seemingly simple theoretical model can generate correlations and adaptive changes that are responses to selective pressures and changes in the external environment and yet cannot be fully attributed to changes in genotypes. As they correctly point out, this is not a multi-level selection story. Reproduction is insufficient to account for the change. This top-down resource pressure tied with as they point out “worse-case” Neo-Darwinian population changes still generates positive changes in the system that can lead to stabilization of specific cycles. In itself these are timely and important results, but what makes this contribution especially important is that the model, they argue convincingly and rightly, plausibly resembles a simplification of the conditions that obtained during the assembly of biogeochemical cycles on the Precambrian Earth. This is major stuff. The Gaia Hypothesis has long suffered from the lack of a compelling “origin story”.

What the authors provide is a compelling evolutionary story. For readers that may not find the Gaia hypothesis to be their cup of tea, there is still a lot here: even if it may not have played out at the planetary level, localized homeostatic loops may have emerged via the processes described by the authors. Doolittle’s ITSNTS approach demands careful

examination and extension and this is what the authors have done. Excellent and important article.

We thank the reviewer for these supportive remarks, which we find very encouraging. We hope to contribute further to placing biogeochemical processes in a more Darwinian framework in future work.

.....

Reviewer #2 (Remarks to the Author):

This paper says that it provides a model to explain how selection processes that operate at the timescale of biological evolution can influence the organization of biogeochemical cycles at geological timescales. Essentially, it attempts to provide a quantitative model of an interesting scenario proposed early by Doolittle. The problem is important and the first couple of pages of the paper impressively lay out the many interesting issues. I am unable, however, to understand the paper after the authors' statement that "The basic structure of our model is already a feature of evolutionary-ecological thinking on syntrophy [31], symbiosis [32] and multi-level selection [33, 34]. We claim no originality in this regard...

This phrase (which has now been removed and replaced with a re-worded sentence) was a reference to the existence of a rich field within theoretical biology focusing on cases in which there is a difference between the equilibrium solutions at the local and "global" levels. Most obviously, in the case of multi-level selection, the within-group (i.e. individual-level) dynamics favour maximal individual replication rate, whereas the between group dynamics may favour lower rates in conjunction with other phenotypes (two categories sometimes labelled "cheaters" and "altruists" and analogous to our model genotypes). Recycling loops and feedbacks derived from different genotypes assimilating related environmental substances also feature substantially in evolutionary ecology models, as well as other areas such as origin of life research. In other words, we intended to convey the general point that models with roughly the same structure as we use here are already part of the literature, and our purpose in this article was to apply such models to issues relating to natural selection and Gaia.

"I am not familiar with references 31-34, and the SI of the paper, for me, is not useful: it essentially contains computer code, complete with if-then statements, typeset as a kind of mathematics, but with variables written out as long words.

The "IF" statements within equations (8), (19) and (20) are intended to communicate the change in functional form of the relevant fluxes at key threshold values (e.g. the change in the growth rate of the plastic genotype). The "long words" section of the SI to which the reviewer refers to has been completely re-written and substantially simplified. The purpose of this section is to explain in more detail our rationale for the general idea that ITSNTS may be easier to falsify with a focus on directional change in the correlation between particular biological and environmental states, which is consistent with our numerical results. This section now consists of only three equations and all variables are defined in a new table. The key ideas are also stated in words in the early part of this section. We have also stressed the appropriate clauses, for example beginning with:

"The following labelling exercise is intended to make explicit the distinction between ITSNTS and a more general view of feedbacks between life and the abiotic environment. We emphasize that this is not intended as any kind of "proof" of our arguments, merely an attempt to clarify and make explicit our assumptions."

This section could be removed without direct impact on the main modelling results, but we have retained it because we believe it may be of interest to theoretical biologists wishing to frame ITSNTS in terms of conventional population genetics, in that it provides a simple starting framework for expansion in more detail in future work.

I am unfortunately unable to understand what's written here, so I cannot offer an opinion concerning the paper's correctness and suitability for publication.

We have worked hard to make our arguments more accessible and improve the presentation of our results, which we hope the reviewer will now find of an acceptable standard for publication.

.....

Reviewer #3 (Remarks to the Author):

In their paper, Boyle and Lenton develop and use a theoretical model to analyze the evolution of a biogeochemical cycle via persistence-based selection. The main question the authors ask is whether there are conditions under which a certain biogeochemical cycle emerges without invoking natural selection (i.e. selection favoring the cycle itself). Instead, conditions are identified, under which two types that produce or consume two environmental substances can persist despite the presence of two “cheaters” that each consume one of the substances without contributing to the production of the other one. In this scenario, producers generate one of the substance at a cost to themselves. In the presence of non-producers that do not pay the production cost but still benefit from the production of other individuals, natural selection is expected to operate against the cycle (i.e. the two reciprocally interacting individuals). Individuals switch between producer and cheater phenotype depending on the environmental concentration of one of the two substances and the offspring of the corresponding types changes between them with a certain frequency due to mutation. In simulations in which different parameters are systematically varied, the authors identify conditions in which the system can stably maintain the cycle even though individual-level selection acts against it. The paper addresses an interesting and relevant question. The provided answers are not only relevant to the emergence of biogeochemical cycles, but in particular also to the evolution of ecological interactions between organisms in general (e.g. gut microbiota). The presented work is founded on a very solid and accurate discussion of the available literature. The paper is an interesting read and represents a valuable extension of the available literature.

We thank the reviewer for these supportive remarks and the astute summary of the scope of our work. We would just add that we still invoke *natural* selection, just an extended form based on persistence without replication, as discussed by Doolittle and colleagues.

Even though I could not understand the mathematical side of the model in all detail, I think that presentation of the data needs to be considerably improved to make the paper more accessible.

Apologies for any inconvenience caused by the lack of clarity in the previous draft. We have rewritten parts of the model description and the references to it in the main paper in order to address this. We have also recalculated and re-formatted all replicate average results, avoiding 3D scatter plots and plotting only replicate averages rather than raw data.

Main points:

1) It would have been nice to introduce a page and line numbering so that reviewers can explain more easily refer to a certain part of the text.

Done

2) Page 4, first line: The term “facultative” is misleading and can have different meanings. I suggest to use the term “plastic”. “Phenotypic plasticity” is the phenomenon the authors are talking about.

Done, thanks

3) In the discussion section, I found the argumentation why genetic drift and environmental fluctuations were likely stronger on early earth difficult to follow. Please explain this more carefully.

Rephrased, thanks. Line 399: “These distinct lines of evidence converge to give a picture in which growth-permitting resources and nutrients were scarcely and erratically distributed in time and space. We argue that such extreme variability in environmental boundary conditions would have resulted in small populations subject to frequent and extreme bottlenecks.”, which is the essential point that we were attempting to make with this summary of deep time evidence. Essentially, our model is qualitatively analogous to any situation in which the occurrence of environmental conditions limiting to growth fluctuates in time and space. (The absence of growth-permitting conditions being analogous to the “influx-off” interval, their presence the “on” interval).

4) Presentation of data:

a) In general I think all graphs need to be reworked in order to improve the accessibility of the data. It took me a long time to understand what the authors wanted to show. This has mainly to do with the fact that the figures appear to be sloppily and inconsistently formatted.

Sorry about that. We have replotted and reformatted all graphs.

E.g. in Figs. 2 and 3 legends seem to be randomly placed, sometimes overlapping with graphs themselves. “generations” as y-axis label should be removed except for the two panels at the bottom. The axis legends should start consistently with a capital or small letter. The headings of the rows could be placed at the left of the corresponding row.

The reason for the inconsistent placing of legends in the previous draft was to avoid obscuring the key aspects of the plots. We have repositioned these with more care.

The authors should revisit all graphs and format them consistently, especially trying to improve the placement of explanatory elements such as 1A,B,C (upper left corner of graph) and figure legends (e.g. always above all panels).

We have consistently labelled all plots (columns by number, rows by letter). The positioning of legends is a balancing act regarding the most efficient use of space, but we have made an effort to be consistent and (in particular) avoid obscuring any plot lines. Everything except (where relevant) model variables should now be lower case.

b) Figure 4,7,8: I found it extremely difficult to see any difference in the individual graphs of

these figures. I could barely distinguish the colours and see whether a data point is high or low. I think this way of presenting this data is not very reader-friendly and should be changed to a more accessible way. Depicting the data in a plane rather than showing individual data points could be one solution.

Again, sorry for any inconvenience caused by this. All 3D scatter plots have been removed, and replaced with 2D filled contour plots, showing only ensemble average results, not the raw data. We hope this is clearer. The only exception to this is figure 12 within the supplementary material, which we have retained to make a simple point (about the bimodal distribution of the results) and which we only mention in passing in the main text.

c) Figure 6: I did not understand why the graphs in panel D1-4 show so low average fixation probabilities when the graphs on the left always have a considerable amount of data points at the upper end of the corresponding distributions. Is this because of the given distribution within replicates at a particular parameter combination?

This is a consequence of the bimodal nature of the results distribution in the fluctuation runs. Above a certain length of shut-off interval, the overwhelming majority of runs lead to complete population crashes, with the exception of the “lock-in” (producer fixation after entrenchment of the recycling genotype) events. (In other words, there were always far more datapoints on the zero line, which may have been obscured by the formatting issue). This has hopefully been rectified with the new contour plot format.

Conceptually, we are not arguing that the majority of contexts involving fluctuations and drift lead to producer fixation or recycling loop lock-in, merely that some do - and (crucially) those that do not (non-producer fixation) drive themselves extinct via the persistence selection we are talking about.

5) I generally dislike the term “cheater”, because it introduces wrong (anthropocentric) associations. A more neutral term such as simply “non-producer” appears to be more apt.

Done, thank you

Minor comments:

- Abstract: please remove the word “show”
- I noticed that the authors sometimes did not include the Oxford comma. Please recheck the text.
- Last paragraph of the results section: replace “increases” with “increasing”

Done, thanks

17th Dec 21

Dear Dr Boyle,

I am delighted to say that we are happy, in principle, to publish a suitably revised version of your manuscript titled "The evolution of biogeochemical recycling by persistence-based selection" in Communications Earth & Environment under the open access CC BY license (Creative Commons Attribution v4.0 International License).

We therefore invite you to edit your manuscript one last time to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table". Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication under a [CC BY license](http://creativecommons.org/licenses/by/4.0) (Creative Commons Attribution 4.0 International License). This license allows maximum dissemination and re-use of open access materials and is preferred by many research funding bodies.

For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commsenv/article-processing-charges>

At acceptance, you will be provided with instructions for completing this CC BY license on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC). **Please note that that we will not be able to move forward without having first received your payment information and therefore we ask that you please have this information ready when submitting the final version of your manuscript. We will also need your third-party declarations.**

Please use the following link to submit the above items:

[link redacted]

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Joe Aslin

Associate Editor,
Communications Earth & Environment
<https://www.nature.com/commsenv/>
Twitter: @CommsEarth

Responses to reviewer comments for “The evolution of biogeochemical recycling by persistence based selection”

Reviewer comments in black
Responses in blue

.....

Reviewer #1 (Remarks to the Author):

In this bracing and thought-provoking article, the authors provide a compelling case of a recent theoretical proposal that aims to extend Evolutionary Theory beyond its traditional sandbox. The biggest benefit from this is for Ecology and Earth Sciences where the intuition that evolutionary explanations would be worthwhile is there, but where the theoretical and experimental justification has not been fully fleshed out. The article aims at remedying this. The article is theoretically ambitious (in the best possible way) and methodologically sound. It will help other researchers generate new models, experiments and hypotheses, and it advances theory. As it is obvious, you have my strongest support for the publication of this article.

“Traditional” Evolutionary Theory is based on differential reproductive success. Some, including the authors, have argued that this view however powerful is insufficient to account for all biological adaptations and for other evolutionary processes. They have offered a view of evolution based on differential persistence (differential reproductive success being a large subset of differential persistence strategies). The biggest payoff of such a view is how it allows the inclusion of non-genetic mechanisms or even abiotic processes into evolutionary explanations, not merely as external forces or constraints, but as part of the evolutionary mechanisms themselves. This is an extremely fecund area of research for biological research in general, but for ecological and environmental research in particular because it allows us to treat the dynamics of biogeochemical processes as evolutionary processes (at least in some instances). However promising this view is, testing it is an ongoing challenge. This is exactly what the authors do here. The authors show how a seemingly simple theoretical model can generate correlations and adaptive changes that are responses to selective pressures and changes in the external environment and yet cannot be fully attributed to changes in genotypes. As they correctly point out, this is not a multi-level selection story. Reproduction is insufficient to account for the change. This top-down resource pressure tied with as they point out “worse-case” Neo-Darwinian population changes still generates positive changes in the system that can lead to stabilization of specific cycles. In itself these are timely and important results, but what makes this contribution especially important is that the model, they argue convincingly and rightly, plausibly resembles a simplification of the conditions that obtained during the assembly of biogeochemical cycles on the Precambrian Earth. This is major stuff. The Gaia Hypothesis has long suffered from the lack of a compelling “origin story”.

What the authors provide is a compelling evolutionary story. For readers that may not find the Gaia hypothesis to be their cup of tea, there is still a lot here: even if it may not have played out at the planetary level, localized homeostatic loops may have emerged via the processes described by the authors. Doolittle’s ITSNTS approach demands careful

examination and extension and this is what the authors have done. Excellent and important article.

We thank the reviewer for these supportive remarks, which we find very encouraging. We hope to contribute further to placing biogeochemical processes in a more Darwinian framework in future work.

.....

Reviewer #2 (Remarks to the Author):

This paper says that it provides a model to explain how selection processes that operate at the timescale of biological evolution can influence the organization of biogeochemical cycles at geological timescales. Essentially, it attempts to provide a quantitative model of an interesting scenario proposed early by Doolittle. The problem is important and the first couple of pages of the paper impressively lay out the many interesting issues. I am unable, however, to understand the paper after the authors' statement that "The basic structure of our model is already a feature of evolutionary-ecological thinking on syntrophy [31], symbiosis [32] and multi-level selection [33, 34]. We claim no originality in this regard...

This phrase (which has now been removed and replaced with a re-worded sentence) was a reference to the existence of a rich field within theoretical biology focusing on cases in which there is a difference between the equilibrium solutions at the local and "global" levels. Most obviously, in the case of multi-level selection, the within-group (i.e. individual-level) dynamics favour maximal individual replication rate, whereas the between group dynamics may favour lower rates in conjunction with other phenotypes (two categories sometimes labelled "cheaters" and "altruists" and analogous to our model genotypes). Recycling loops and feedbacks derived from different genotypes assimilating related environmental substances also feature substantially in evolutionary ecology models, as well as other areas such as origin of life research. In other words, we intended to convey the general point that models with roughly the same structure as we use here are already part of the literature, and our purpose in this article was to apply such models to issues relating to natural selection and Gaia.

"I am not familiar with references 31-34, and the SI of the paper, for me, is not useful: it essentially contains computer code, complete with if-then statements, typeset as a kind of mathematics, but with variables written out as long words.

The "IF" statements within equations (8), (19) and (20) are intended to communicate the change in functional form of the relevant fluxes at key threshold values (e.g. the change in the growth rate of the plastic genotype). The "long words" section of the SI to which the reviewer refers to has been completely re-written and substantially simplified. The purpose of this section is to explain in more detail our rationale for the general idea that ITSNTS may be easier to falsify with a focus on directional change in the correlation between particular biological and environmental states, which is consistent with our numerical results. This section now consists of only three equations and all variables are defined in a new table. The key ideas are also stated in words in the early part of this section. We have also stressed the appropriate clauses, for example beginning with:

"The following labelling exercise is intended to make explicit the distinction between ITSNTS and a more general view of feedbacks between life and the abiotic environment. We emphasize that this is not intended as any kind of "proof" of our arguments, merely an attempt to clarify and make explicit our assumptions."

This section could be removed without direct impact on the main modelling results, but we have retained it because we believe it may be of interest to theoretical biologists wishing to frame ITSNTS in terms of conventional population genetics, in that it provides a simple starting framework for expansion in more detail in future work.

I am unfortunately unable to understand what's written here, so I cannot offer an opinion concerning the paper's correctness and suitability for publication.

We have worked hard to make our arguments more accessible and improve the presentation of our results, which we hope the reviewer will now find of an acceptable standard for publication.

.....

Reviewer #3 (Remarks to the Author):

In their paper, Boyle and Lenton develop and use a theoretical model to analyze the evolution of a biogeochemical cycle via persistence-based selection. The main question the authors ask is whether there are conditions under which a certain biogeochemical cycle emerges without invoking natural selection (i.e. selection favoring the cycle itself). Instead, conditions are identified, under which two types that produce or consume two environmental substances can persist despite the presence of two “cheaters” that each consume one of the substances without contributing to the production of the other one. In this scenario, producers generate one of the substance at a cost to themselves. In the presence of non-producers that do not pay the production cost but still benefit from the production of other individuals, natural selection is expected to operate against the cycle (i.e. the two reciprocally interacting individuals). Individuals switch between producer and cheater phenotype depending on the environmental concentration of one of the two substances and the offspring of the corresponding types changes between them with a certain frequency due to mutation. In simulations in which different parameters are systematically varied, the authors identify conditions in which the system can stably maintain the cycle even though individual-level selection acts against it. The paper addresses an interesting and relevant question. The provided answers are not only relevant to the emergence of biogeochemical cycles, but in particular also to the evolution of ecological interactions between organisms in general (e.g. gut microbiota). The presented work is founded on a very solid and accurate discussion of the available literature. The paper is an interesting read and represents a valuable extension of the available literature.

We thank the reviewer for these supportive remarks and the astute summary of the scope of our work. We would just add that we still invoke *natural* selection, just an extended form based on persistence without replication, as discussed by Doolittle and colleagues.

Even though I could not understand the mathematical side of the model in all detail, I think that presentation of the data needs to be considerably improved to make the paper more accessible.

Apologies for any inconvenience caused by the lack of clarity in the previous draft. We have rewritten parts of the model description and the references to it in the main paper in order to address this. We have also recalculated and re-formatted all replicate average results, avoiding 3D scatter plots and plotting only replicate averages rather than raw data.

Main points:

1) It would have been nice to introduce a page and line numbering so that reviewers can explain more easily refer to a certain part of the text.

Done

2) Page 4, first line: The term “facultative” is misleading and can have different meanings. I suggest to use the term “plastic”. “Phenotypic plasticity” is the phenomenon the authors are talking about.

Done, thanks

3) In the discussion section, I found the argumentation why genetic drift and environmental fluctuations were likely stronger on early earth difficult to follow. Please explain this more carefully.

Rephrased, thanks. Line 399: “These distinct lines of evidence converge to give a picture in which growth-permitting resources and nutrients were scarcely and erratically distributed in time and space. We argue that such extreme variability in environmental boundary conditions would have resulted in small populations subject to frequent and extreme bottlenecks.”, which is the essential point that we were attempting to make with this summary of deep time evidence. Essentially, our model is qualitatively analogous to any situation in which the occurrence of environmental conditions limiting to growth fluctuates in time and space. (The absence of growth-permitting conditions being analogous to the “influx-off” interval, their presence the “on” interval).

4) Presentation of data:

a) In general I think all graphs need to be reworked in order to improve the accessibility of the data. It took me a long time to understand what the authors wanted to show. This has mainly to do with the fact that the figures appear to be sloppily and inconsistently formatted.

Sorry about that. We have replotted and reformatted all graphs.

E.g. in Figs. 2 and 3 legends seem to be randomly placed, sometimes overlapping with graphs themselves. “generations” as y-axis label should be removed except for the two panels at the bottom. The axis legends should start consistently with a capital or small letter. The headings of the rows could be placed at the left of the corresponding row.

The reason for the inconsistent placing of legends in the previous draft was to avoid obscuring the key aspects of the plots. We have repositioned these with more care.

The authors should revisit all graphs and format them consistently, especially trying to improve the placement of explanatory elements such as 1A,B,C (upper left corner of graph) and figure legends (e.g. always above all panels).

We have consistently labelled all plots (columns by number, rows by letter). The positioning of legends is a balancing act regarding the most efficient use of space, but we have made an effort to be consistent and (in particular) avoid obscuring any plot lines. Everything except (where relevant) model variables should now be lower case.

b) Figure 4,7,8: I found it extremely difficult to see any difference in the individual graphs of

these figures. I could barely distinguish the colours and see whether a data point is high or low. I think this way of presenting this data is not very reader-friendly and should be changed to a more accessible way. Depicting the data in a plane rather than showing individual data points could be one solution.

Again, sorry for any inconvenience caused by this. All 3D scatter plots have been removed, and replaced with 2D filled contour plots, showing only ensemble average results, not the raw data. We hope this is clearer. The only exception to this is figure 12 within the supplementary material, which we have retained to make a simple point (about the bimodal distribution of the results) and which we only mention in passing in the main text.

c) Figure 6: I did not understand why the graphs in panel D1-4 show so low average fixation probabilities when the graphs on the left always have a considerable amount of data points at the upper end of the corresponding distributions. Is this because of the given distribution within replicates at a particular parameter combination?

This is a consequence of the bimodal nature of the results distribution in the fluctuation runs. Above a certain length of shut-off interval, the overwhelming majority of runs lead to complete population crashes, with the exception of the “lock-in” (producer fixation after entrenchment of the recycling genotype) events. (In other words, there were always far more datapoints on the zero line, which may have been obscured by the formatting issue). This has hopefully been rectified with the new contour plot format.

Conceptually, we are not arguing that the majority of contexts involving fluctuations and drift lead to producer fixation or recycling loop lock-in, merely that some do - and (crucially) those that do not (non-producer fixation) drive themselves extinct via the persistence selection we are talking about.

5) I generally dislike the term “cheater”, because it introduces wrong (anthropocentric) associations. A more neutral term such as simply “non-producer” appears to be more apt.

Done, thank you

Minor comments:

- Abstract: please remove the word “show”
- I noticed that the authors sometimes did not include the Oxford comma. Please recheck the text.
- Last paragraph of the results section: replace “increases” with “increasing”

Done, thanks