

Supplementary material for “The evolution of biogeochemical recycling by persistence-based selection”

S1. Supplementary methods and discussion

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. See supplementary table 1 below for a list of all labels used in this section.

Consider some category of measurable large-scale environmental pattern P affected by both biotic and abiotic processes (of which the cycling ratio of a particular nutrient would be one example). Label each specific instance of this pattern P_j . Assume it is comprised of both biotic $P_{bio,j}$ and abiotic/environmental $P_{env,j}$ sub-states, such that the final measurable value of a particular instance of pattern P , P_j , is some sort of interactive function f_P of these biotic and environmental sub-states.

The by-products, feedbacks and ITSNTS cases can be distinguished by the nature of this functional relationship. To say that the abiotic environment is affected by the by-products of life is simply to invoke a one-way relationship f_{BP} between the state of the abiotic environment and that of the biota. To invoke feedbacks between life and the abiotic environment is to combine the by-products view with the idea of consistent mapping f_{FB} back in the other direction, such that the abiotic state of the environment feeds back on biological growth.

The range of ways in which the state of the abiotic environment can affect biological growth encompass physical, climatic and geochemical dynamics, or boundary conditions such as tectonic inputs, availability of key elements etc. Whilst the values of particular climatic/geochemical variables may change over time (e.g. the sun’s linear increase over Earth’s history), *the underlying functional form of basic physical constraints on growth and adaptation are largely time invariant*. Similarly, although genetic innovation and natural selection present new phenotypes to the environment during the course of time, the range of ways in which the biota can affect its environment (growth, metabolism, entropy production, taking up space, etc.) is also broadly consistent over time, subsumable into basic physical constraints upon adaptation.

The ITSNTS case presupposes all the above, with the critical additional claim (discussed at length by Doolittle) that novelty generating biological mutations that would not have otherwise occurred arise in lineages “recruited” via some association with a biogeochemical cycle (or equivalent environmental property), which increasing this cycle’s propensity to persist over time. The biological component of a particular life-environment interaction is necessary for the perpetuation of the abiotic/environmental component and vice versa (as in the feedbacks case). Additionally, a particular novel mutation arises in the context of this interaction, contributing positively to its perpetuation (by deriving some fitness benefit from the environmental state it contributes to producing), within an organism, *the ancestors of which would have gone extinct were it not for the past existence of this same interaction*. The retrospective nature of this claim arguably makes direct testing near impossible, but (as discussed in the main text) we suggest that ITSNTS can be distinguished from the feedbacks case in that it entails that:

- The state of some life-environment interaction is itself an interactive function of the correlation between the biotic and abiotic sub-states of the environmental pattern in question, as opposed to either of these states considered in isolation.
- The long-term future state of the correlation between a biological trait value and an environmental state value is an increasing function of the present state of that correlation, via the addition of qualitatively new functional relationships.

This distinction can be written explicitly as follows (please refer to table 3).

$$P_j \in P_{byproducts} \therefore P_j = f_P \left(P_{bio,j}(t), f_{BP} \left(P_{bio,j}(t-1) \right) \right), \& \exists f_{BP}, f_{BP} \in \overline{F_{BP}}, \frac{d|\overline{F_{BP}}|}{dt} = 0 \quad (32)$$

In words, (32) states that a particular instance of the by-products case $P_j \in P_{byproducts}$, involves the environmental pattern P being subject to a biotic component $P_{bio,j}$ and an abiotic/environmental component which is a function of the biota $P_{env,j}(t) = f_{BP}(P_{bio,j}(t-1))$. We add the suggestion

that the form of these functions is from a fixed set of relationships that reduce to physical constraints on adaptation and therefore do not evolve over time.

$$P_j \in P_{feedbacks} \therefore (P_j \in P_{BP}) \& \left(P_j(t) = f_P \left(f_{FB} \left(P_{env,j}(t-1) \right), f_{BP} \left(P_{bio,j}(t-1) \right) \right) \right), \& \exists f_{FB}, f_{FB} \in \overline{F_{FB}}, \frac{d|\overline{F_{FB}}|}{dt} = 0 \quad (33)$$

All of this is also true for the feedbacks case given in (33), with the additional two-way interaction such the biota is affected by the environmental state that it itself influences $f_{FB}(P_{env,j}(t-1))$.

$$P_j \in P_{ITSNTS}, \therefore (P_j \in P_{feedback}) \& \left(\begin{array}{c} CORR(P_{bio,j}(t), P_{env,j}(t)) \\ = f_I \left(CORR(P_{bio,j}(t-1), P_{env,j}(t-1)) \right) \end{array} \right),$$

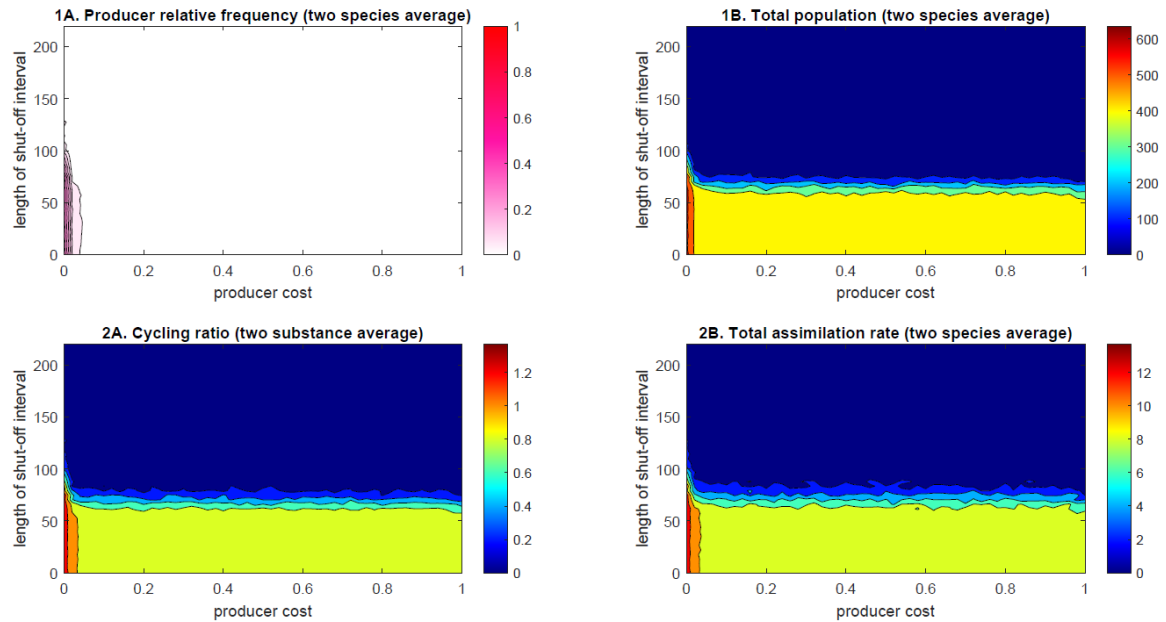
$$\& \exists f_I, f_I \in \overline{F_I}, \frac{d|\overline{F_I}|}{dt} > 0 \quad (34)$$

The ITSNTS case (34) encompasses all of the previous two with the additional point that the size $|\overline{F_I}|$ of the set $\overline{F_I}$ of functional relationships within the life-environment interaction is an increasing function of time. This is a consequence of the persistence-based selection of evolutionary novelties “recruited” by biogeochemical cycles, (discussed at length by Doolittle). As discussed in the main text, we also suggest here that a concrete way in which to test ITSNTS may be long-term directional evolution of correlation patterns between biotic and abiotic states, such that a particular correlation (e.g. between an environmental property and an allele frequency) is a function of its previous value (second bracketed term above). These two ideas are (we suggest) the crux of the distinction between ITSNTS and a more generic by-products/feedbacks view.

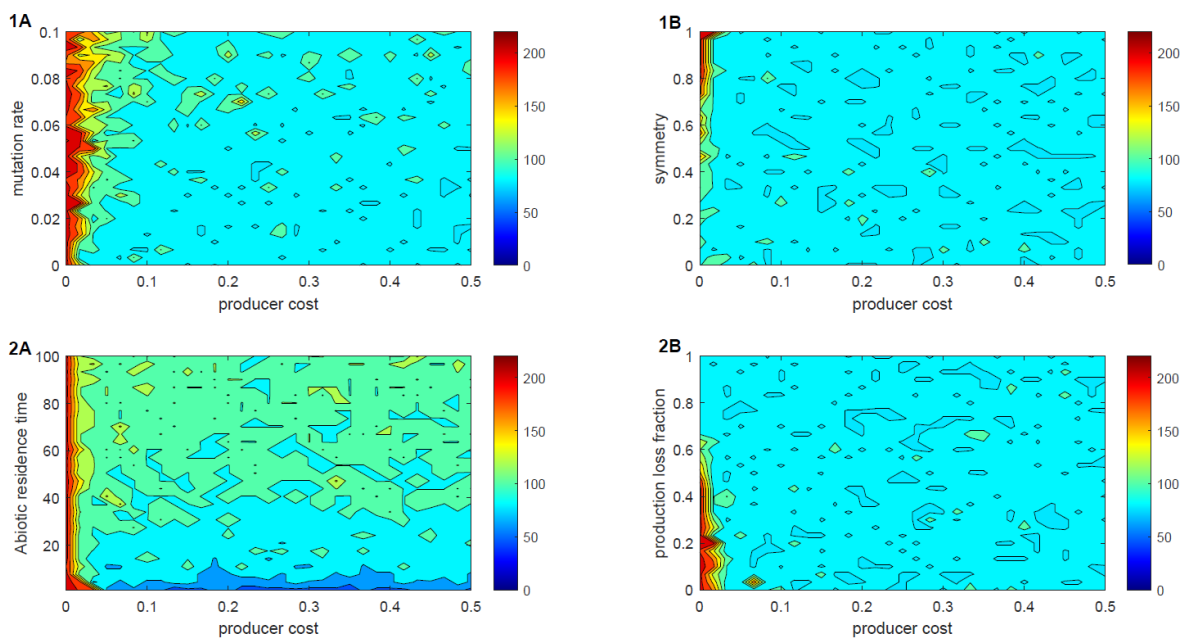
Supplementary Table 1. Labels used in this section	
Label	Meaning
P	Category of large-scale environmental pattern, particular case P_j , with biotic $P_{bio,j}$ and abiotic $P_{env,j}$ components. Divisible into sub-categories $P_{byproducts}$, $P_{feedbacks}$, P_{ITSNTS}
f_P	Functional form of relationship describing how the state value of a particular case of P emerges from its components at time t $P_j(t) = f_P(P_{bio,j}(t), P_{env,j}(t))$
f_{BP}	Functional form of relationship describing how the biotic component of P produces by-products that affect the abiotic component $P_{env,j}(t) = f_{BP}(P_{bio,j}(t-1))$. Total instances collectable into set $\overline{F_{BP}}$ of size $ \overline{F_{BP}} $
f_{FB}	Functional form of relationship describing how abiotic component of feeds back on biotic component $P_{bio,j}(t) = f_{FB}(P_{env,j}(t-1))$. Total instances collectable into set $\overline{F_{FB}}$ of size $ \overline{F_{FB}} $
f_I	Functional form of relationship describing how the interaction between biotic and abiotic components of P , measured by the correlation between the values of these components, relates to previous instances of itself $CORR(P_{bio,j}(t), P_{env,j}(t)) = f_I(CORR(P_{bio,j}(t-1), P_{env,j}(t-1)))$ Total instances collectable into set $\overline{F_I}$ of size $ \overline{F_I} $

S2. Supplementary figures

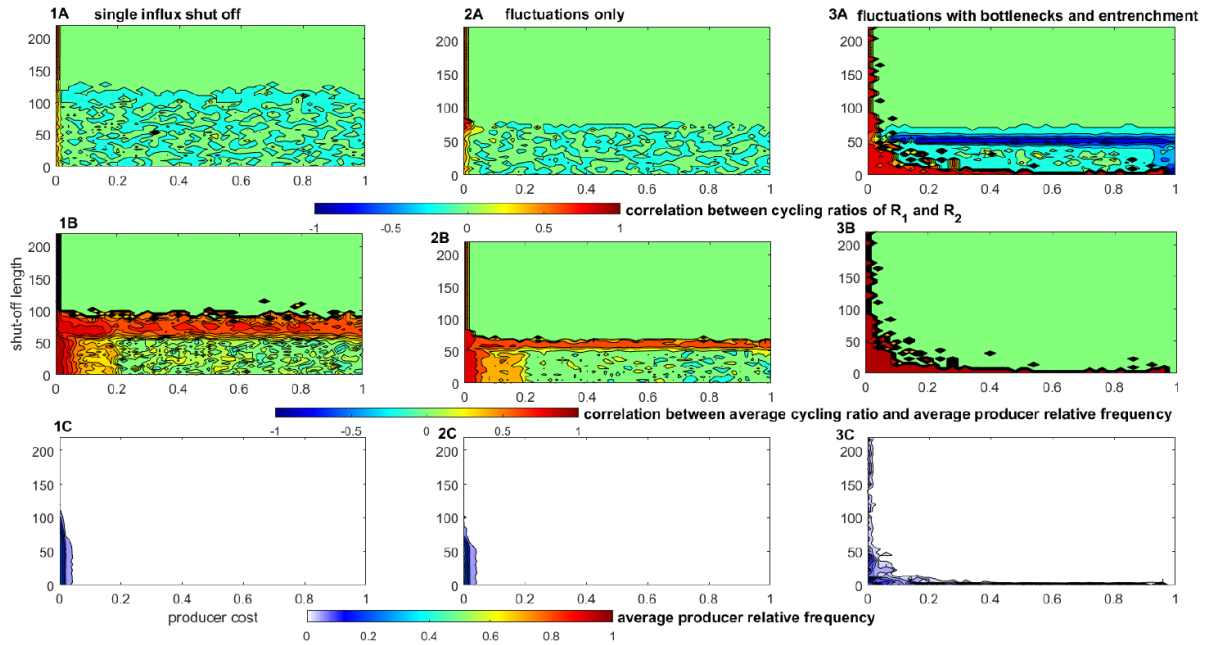
Supplementary figure 1 – Expanded sensitivity analysis to producer cost and influx shut-off interval length. Format identical to figure 4, main text. Note the difference in axis limits.



Supplementary figure 2 – Expanded sensitivity analysis to eco-physiological parameters. Format and sub-figure labels identical to figure 4, main text, with extended X (producer cost) axis.



Supplementary figure 3 – Extension of the replicate-averaged sensitivity analysis shown in figure 8 over an increased parameter range. Format identical to figure 8, main text except with extended axis limits.



Supplementary figure 4 – Distribution of individual simulation results illustrating the differing strength of persistence selection in the different contexts shown in figure 8, main text. Each data point shows the final value of the average cycling ratio (row A) and producer relative frequency (row B) from a single model run in the relevant context. Columns 1, 2 and 3 respectively correspond to single influx shut-off, influx fluctuations alone and the combination of fluctuations, bottlenecks/drift and genetic entrenchment. Colour corresponds to the value of the z-axis variable (i.e. cycling ratio for row A, relative producer frequency for row B). Note the bimodal results distribution distinguishing column 3 from the other model runs, as referred to in the main text.

