INDITEK-2.0: A Bayesian inverse eco-evolutionary modelling framework for reconstructing Phanerozoic biodiversity

Pedro Cermeno[1,*] (orcid.org/0000-0002-3902-3475), Carmen García-Comas[1] (orcid.org/0000-0001-8054-3918), Gloria Herrero Gascón[1](orcid.org/0009-0007-1089-5230), Michael J. Benton[2] (orcid.org/0000-0002-4323-1824), Thomas Bodin[1] (orcid.org/0000-0002-7393-6112), Khushboo Gurung[3] (orcid.org/0000-0001-8074-7339), Benjamin Mills[3] (orcid.org/0000-0002-9141-0931), Alexandre Pohl[4] (orcid.org/0000-0003-2328-351X)

[1]Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas, Barcelona, 08003, Spain
[2]School of Earth Sciences, University of Bristol, Bristol BS8 1RL, UK
[3]School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
[4]Biogéosciences, UMR 6282 CNRS, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France.

*Corresponding author: Pedro Cermeno (pedrocermeno@icm.csic.es)

Author contributions

P.C. and C.G.C. conceived the main ideas and wrote the manuscript. P.C. and C.G.C developed the INDITEK-2.0 eco-evolutionary model. G.G. conducted model simulations. K.G., B.M., and A.P. performed the Earth System Model simulations (FOAM, cGENIE and FLORA). C.G.C, G.G., and T.B. implemented the Bayesian inverse modelling framework. All authors contributed to the development of the model and revised the final manuscript.

Acknowledgements

This work was funded by research grant PID2023-152076NB-I00 from the Spanish government. We thank Lewis Jones and Thiago Rangel for comments on an early version of the manuscript. T.B. acknowledges the support from the Spanish Ministry of Science and Innovation ATR2023-143517/AEI/10.13039/501100011033. A.P. acknowledges the support of the French *Agence Nationale de la Recherche* (ANR) under reference ANR-22-CE01-0003 (project ECO-BOOST). Calculations were performed using HPC resources from DNUM CCUB (*Centre de Calcul de l'Université de Bourgogne*) and the SMART HPC infrastructure at the Institut de Ciències del Mar. P.C., C.G.C, G.H.G. and T.B. acknowledge the 'Severo Ochoa Centre of Excellence' accreditation (CEX2019-000928-S).

Conflict of interest statement

The authors declare no conflicts of interest.

Abstract

Mechanistic eco-evolutionary models are powerful tools for understanding how life diversified, but estimating their parameters is a major challenge. The vast parameter space of these models makes traditional inverse modelling computationally unfeasible. We address this challenge with INDITEK-2.0, a new, cost-efficient eco-evolutionary model that integrates a Bayesian inverse modelling framework for parameter estimation. To validate our framework, we conducted a proof-of-concept study. We first generated synthetic biodiversity data for marine invertebrates using INDITEK-2.0 with known parameter values. To mimic field data, some realistic random Gaussian errors were added to the data. We then used our Bayesian inverse modelling framework to recover these original parameter values from the dataset of current biodiversity distributions after 500 million years of diversification. Our solution is a probability distribution on the parameter space, and we show that the method successfully recovered the target parameters values. This validates INDITEK-2.0's ability to probabilistically infer the underlying eco-evolutionary processes and reconstruct the evolutionary history of animal biodiversity across Earth's ecosystems and taxonomic groups.

Keywords: Mechanistic eco-evolutionary modelling, biodiversity history, Bayesian inference, inverse modelling, parameter estimation.

1. Introduction

Mechanistic modelling is emerging as a powerful complementary tool to the fossil record and molecular phylogenetics for reconstructing biodiversity history (Beaugrand, 2014; Cermeño et al., 2022; Farnsworth et al., 2023; Hagen, 2023; Lovelace et al., 2020; Rangel et al., 2018). By simulating eco-evolutionary processes under dynamic environmental conditions, mechanistic eco-evolutionary models offer a unique opportunity to explore the drivers and patterns of taxonomic diversification, bridging gaps and testing hypotheses that are difficult to address with traditional methods alone. They notably permit to go beyond the observation of temporal correlations between environmental change and biodiversity dynamics and, through numerical experiments, to address causality and identify the mechanisms underlying observed changes (Hagen, 2023). Mechanistic eco-evolutionary models have been developed to date at two levels of biological organization, at the individual level and the population level (Beaugrand et al., 2018; Garwood et al., 2019; Hagen et al., 2021; Malchow et al., 2021; Rangel et al., 2018; Romero-Mujalli et al., 2019). Individual-based eco-evolutionary models provide a framework for simulating complex ecological and evolutionary processes, with trait variation and demographic stochasticity modelled at the level of individual organisms, each represented as a unique entity with its own set of traits (e.g., size, age, genotype, phenotype). By incorporating genetic and phenotypic trait variation, these can predict how populations respond to environmental changes, including habitat fragmentation, climate change, or the introduction of invasive species (Garwood et al., 2019; Romero-Mujalli et al., 2019; Xuereb et al., 2021). Population-based models, on the other hand, abstract the details of individual organisms into population or species-level traits that influence species' survival, reproduction, and interaction with the environment and other organisms. The distribution and frequency of these traits within the population can change over time through natural selection, genetic drift, and dispersal, reflecting the evolutionary processes of speciation, extirpation, and extinction at play (Beaugrand, 2023; Hagen, 2023; Pontarp et al., 2019).

To capture the details of eco-evolutionary processes, individual- and population-based eco-evolutionary models need to operate at sufficiently high spatial and temporal resolutions. High-resolution modelling more realistically represents habitat heterogeneity and species variability. However, this level of detail creates computational challenges when reconstructing evolutionary histories over vast spatial and temporal scales. This is a problem for both the Earth system models that provide palaeoenvironmental data and the eco-evolutionary models themselves. Likewise, in a mechanistic eco-evolutionary model, each discrete entity or state variable, whether an individual organism or a population of organisms, is governed by its own set of differential equations that describe its state changes over time. These equations account for growth, reproduction, death, migration, interactions with other entities and the environment, genetic drift, heritability, speciation, and extinction. The need to solve a separate set of equations for each discrete entity rapidly escalates the computational demands of these models.

Reconstructing the biodiversity history of a taxonomic group requires estimating parameters that reflect its unique evolutionary characteristics, such as speciation and extinction rates, their environmental dependencies, and carrying capacities. Estimating these crucial parameters is challenging because it requires detailed knowledge of the taxonomic group's biodiversity history—the very history we are trying to reconstruct. Inverse modelling offers a solution: it compares model simulations with empirical data (e.g. fossil records, present-day biodiversity). By finding the simulation that best fits the empirical (or calibration) data, inverse modelling effectively estimates the parameters that most likely generated the observed patterns. Probabilistic inference, particularly Bayesian approaches using Markov chain Monte Carlo (MCMC) algorithms, provides a robust methodological framework to accomplish this objective. MCMC algorithms explore the parameter space by taking random walks guided by the shape of the posterior distribution. By recording the visited locations within the parameter space, MCMC builds a representative sample of the posterior

distribution, which often requires thousands of model evaluations. Therefore, applying Bayesian inversion to these models requires a trade-off between model complexity and computational feasibility.

To address this trade-off, we developed INDITEK, a cost-efficient eco-evolutionary model designed to simplify mechanisms while retaining essential dynamics. INDITEK treats biodiversity as an aggregate property at a taxonomic level greater than the species level (e.g., class, phylum) (Cermeño et al 2022). This fundamental shift moves the analytical focus from individual- or species-level traits to broader attributes at class or phylum level, making biodiversity itself—rather than individuals or species—the state variable. This strategic approach solves computational challenges and allows simulating 500 million years of global scale coastal diversity of marine invertebrates in about thirty seconds on a standard computer. This article presents INDITEK-2.0, an upgraded version of the INDITEK model. INDITEK-2.0 integrates a Bayesian inverse modelling framework, which enables probabilistic estimation of its parameters. Here, we validate the proposed inverse modelling framework, evaluate its applicability to macroevolutionary questions, and provide a discussion of model limitations and future directions.

2. INDITEK-2.0 theoretical framework

2.1. Parameterizing evolutionary rates

As noted, in our model, biodiversity is an aggregate property implicitly reflecting the cumulative effects of ecological and evolutionary processes at or below the species level. These processes give rise to spatial and temporal differences in diversification dynamics, whose parameterization is critical to reconstructing and understanding the evolutionary history of biodiversity.

The metabolic theory of ecology provides a theoretical basis, with substantial empirical support, for understanding how temperature affects evolutionary rates by altering the energy available for growth, reproduction, and survival (Allen et al., 2002; Brown et al., 2004). A greater energy input to the system (or a higher ambient temperature) implies faster metabolic rates and, consequently, shorter generation times (Rohde, 1992). Higher turnover rates allow for quicker accumulation of genetic changes, potentially increasing the rate of evolution (Lanfear et al., 2010). This means that adaptive traits can become prevalent in shorter time spans, allowing populations to respond to selection pressures more rapidly (**Fig. 1**). The metabolic theory of ecology uses the Boltzmann factor, a concept from statistical thermodynamics, to explain how the rates of biological processes, like metabolism, growth and reproduction accelerate with temperature (Gillooly et al., 2001). Its core idea – that probabilities can exponentially change with energy and temperature – can be applied to parameterize how temperature affects speciation rate through its impacts on biological rates. Alternatively, another quantitative expression for parameterising the sensitivity of metabolic rates to temperature changes is the Q10 coefficient, a dimensionless parameter taken from the field of ecophysiology (Mundim et al., 2020). The simplicity of the Q10 model – offering a numerical value to represent how metabolic rates change at a different rate with every 10°C change – makes it particularly adaptable to eco-evolutionary modelling (**Box 1**).

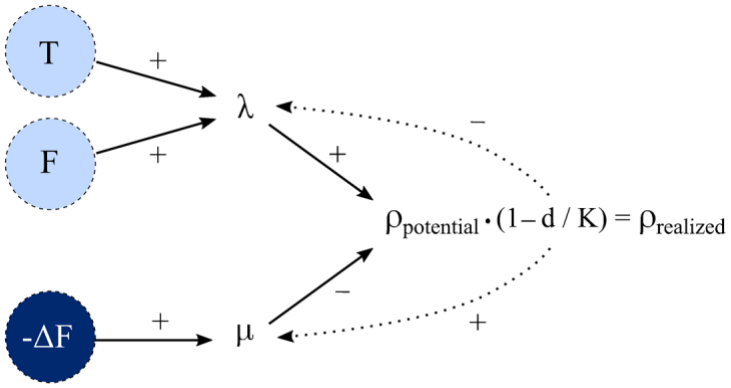

Fig. 1. Parameterization of evolutionary rates. *Speciation rate (λ) is a positive function of temperature (T), food supply (F), and, in terrestrial systems, water availability (W). A reduction in food supply rate (-ΔF) directly lowers the carrying capacity K, implicitly driving an increase in the extinction rate (μ) in systems near K. The difference between λ and μ determines the net diversification rate, $\rho_{potential}$. The actual, or realized, net diversification rate ($\rho_{realized}$), is calculated by multiplying $\rho_{potential}$ by the diversity-dependent limitation term (1-d/K), where d is diversity. Diversity-dependent limitations feedback on evolutionary rates by implicitly decreasing λ, increasing μ, or both (dotted lines) as d approaches K.*

Food supply also plays a pivotal role in controlling speciation rates by affecting population dynamics and energetics (Bambach, 1993; Martin, 1996; Vermeij, 1977). For example, abundant food resources can support larger populations, which tend to maintain higher genetic diversity (Hague & Routman, 2016). Larger, genetically diverse populations provide a broader repertoire for natural selection to act upon, potentially accelerating the speciation process (**Fig. 1**). Similarly, for terrestrial ecosystems, water availability, primarily through precipitation, is a fundamental control on net primary productivity (NPP), which in turn determines the available food supply. As a key component of photosynthesis, water is a limiting resource in many ecosystems, directly influencing the amount of biomass produced and thus acting as a primary driver of speciation rates. As with temperature, quantifying the combined effect of food supply and water availability on speciation rate necessitates the use of generalized mathematical models. These models, ranging from simple linear relationships to more complex nonlinear dynamics, provide a quantitative framework for representing the mechanisms that govern this relationship (**Box 1**). For example, a Michaelis-Menten equation provides a useful quantitative model for describing the dependence of speciation rates on both food and water availability. This approach is analogous to how chemical reaction rates depend on substrate concentration in biochemistry, replicating its wide use to model changes in population abundance as a function of resource availability. It can capture the threshold-like behaviour of resource-driven productivity, where speciation rates increase rapidly with increasing resources in limited environments but plateau in resource-rich conditions where other factors become limiting.

Parameterizing the rate of extinction is considerably more difficult because there is no clear, well-established theoretical framework that comprehensively explains the mechanistic drivers behind it (Raup, 1991). Adaptation is a fundamental evolutionary process allowing species to survive in changing environments. However, significant changes in environmental conditions can outpace a species' ability to adapt, resulting in a mismatch between its physiological requirements and the new environmental conditions (Hallam, 1998; Peters & Foote, 2002; Saupe et al., 2014). When these changes occur rapidly, species may not have enough time to evolve new traits and cope with the new conditions. For example, it has been hypothesized that the latitudinal biodiversity gradient, characterized by the increase in species richness towards the equator, might result from greater environmental instability at higher latitudes (Chown & Gaston, 2000; Willig et al., 2003). This

reasoning posits a direct linkage between the rate of environmental change and the rate of species extinction (Kaiho, 2022; Song et al., 2021). This parameterization requires, however, higher spatial and temporal resolution data than is currently available (see ***Section 4. Model limitations and future directions***). While future Earth system model simulations may provide the necessary inputs, here, we assume a constant background extinction rate for the current model.

2.2. Diversity dependence of net diversification rate

Simulating the effect of ecological interactions, such as competition or predation on diversification rates requires accounting for species interactions, which is not possible in a model operating above the species level. One way to incorporate the effect of biotic factors without getting into the complexities of species interactions is by using the concept of carrying capacity (K) for species richness (Storch & Okie, 2019). The parameter K is intricately tied to the balance between the available resources in the system and the species that depend on these resources for survival. On evolutionary time scales, the dynamic interplay among resource availability and the strategies species employ to exploit and compete for these resources establishes the maximum number of taxonomic units an ecosystem can support (Bambach, 1983; Mondal & Harries, 2016). In the context of evolutionary models, K can be used to approximate the effect of biotic factors on evolutionary rates. As diversity (d) increases (i.e., d/K approaches 1), the strengthening of ecological interactions undermines the probability of species survival, certain populations become smaller, and the diversification process gradually slows down until equilibrium diversity is reached (Levinton, 1979; MacArthur & Wilson, 1967). These diversity-dependent ecological interactions feedback on evolutionary rates by implicitly decreasing λ, increasing μ, or both (Rabosky, 2009) (**Fig. 1**).

In our model, the K of an ecosystem is directly tied to its net primary production (NPP), which we use as a proxy for food supply. We reconstructed marine and terrestrial NPP using an Earth system model of intermediate complexity (see **Supplementary information**) and establish a global minimum (K_{min}) and maximum (K_{max}) values for each realm. The actual K for a specific location and time is then calculated as a linear function of the NPP at that specific point, with a range that encompasses the values between K_{min} and K_{max} (see **BOX 1** for details). This approach hinges on the idea that NPP sets a fundamental limit on the available chemical energy (food) to support consumer populations (Cárdenas & Harries, 2010; Martin, 1996). Should a decrease in NPP occur due to environmental shifts, the ecosystem's K is consequently lowered. In response to this resource limitation, biodiversity is reduced through extinction events until it aligns with the new, lower K (**Fig. 1**). In this way, our model links biodiversity directly to the dynamics of energy input to ecosystems (Wright, 1983).

BOX 1. INDITEK-2.0's model parameterization

To reconstruct biodiversity, we employ a logistic differential equation that describes the change in local diversity (d) over time (t):

$$\partial d / \partial t = (\lambda - \mu)\, d\, [1 - (d/K)], \qquad (1)$$

where d is diversity, t is time, λ is the speciation rate, μ is the background extinction rate, and K is the system's carrying capacity. K varies as a linear function of the net primary production (NPP), ranging between K_{min} and K_{max} corresponding to the 1 and 99 percentiles of the NPP throughout the Phanerozoic.

We model λ according to the following equation:

$$\lambda = \lambda_{max} - (\lambda_{max} - \lambda_{min})\,(1 - Q_{temp}\, Q_{food}), \qquad (2)$$

where λ_{min} and λ_{max} are the range of the maximum potential values of λ. Specifically, λ_{max} is the maximum potential speciation rate under optimal environmental conditions. Q_{temp} and Q_{food} are dimensionless factors, ranging from 0 (maximum limitation) to 1 (no limitation), that modulate λ as a function of temperature and NPP or food supply, respectively.

Q_{temp} is calculated using the equation:

$$Q_{temp} = (Q10^{T-Tmin}) / (Q10^{Tmax - Tmin}) \quad (3)$$

where Q10 is a coefficient that quantifies the sensitivity of λ to temperature, T is the sea surface temperature (°C) at a specific location and time, and T_{min} and T_{max} represent the 1 and 99 percentiles, respectively, of the T frequency distribution within each time interval.

Q_{food} is modelled using a Michaelis-Menten formulation:

$$Q_{food} = F / (k_{food} + F) \quad (4)$$

where F (g C m^{-2} yr^{-1}) is marine export production, that is, the fraction of net primary production (NPP) that escapes the sunlit surface layer, at a given location and time. The parameter k_{food} (g C m^{-2} yr^{-1}) is the value of F at which λ is half of its λ_{max}. The same type of Michaelis-Menten parameterization can be used for modelling Q_{food} in terrestrial systems as its effect on λ is similarly dependent on its influence on NPP.

Within this model, the background extinction rate (μ) is a fixed value that represents the consistent, low-level rate of diversity loss that occurs even in an ideal environment. It is a fundamental component of the potential net diversification rate ($\rho_{potential}$), defined as:

$$\rho_{potential} = \lambda - \mu \quad (5)$$

However, this potential rate is not always realized. The realized net diversification rate ($\rho_{realized}$), which is what is actually observed, is constrained by ecological interactions that intensify as d approaches the system's K:

$$\rho_{realized} = \rho_{potential} (1 - d/K) \quad (6)$$

As d nears K, $\rho_{realized}$ declines toward zero, a reduction implicitly caused by a decrease in λ, an increase in μ, or a combination of both (**Fig. 1**). Crucially, K is not a fixed value; it is directly linked to NPP. Consequently, a decrease in NPP due to environmental shifts lowers the system's K. This resource limitation triggers a decline in d via extinction events, ultimately aligning the value of d with the new, reduced K.

Finally, mass extinctions are simulated by imposing negative net diversification rates (see also Section 4. Model limitations and caveats). Genus-level marine invertebrate fossil diversity curves are used to determine the percentage of diversity loss, starting time, and duration of each mass extinction. To obtain a more robust estimate, independent of biases in individual databases, fossil diversity curves from various sources are averaged into a single dataset.

The eco-evolutionary model code and input data are available at:

https://github.com/glodegloria/indicios/tree/main

2.3. Tracking diversification in space and time

Over geological timescales, plate tectonics has profoundly influenced Earth's surface by controlling the geographic arrangement of continents and ocean basins. As these landmasses and oceans shifted, they moved through different climatic zones, creating diverse environments. Understanding

how Earth's geographic evolution influenced regional diversification dynamics is thus essential to explaining the spatial patterns of biodiversity (Pellissier et al., 2018; Valentine & Moores, 1970; Zaffos et al., 2017). Approaches previous to INDITEK lacked the capacity to track the environmental history of habitats as they moved with continental plates, limiting our ability to link geodynamics to macroevolutionary dynamics. To mechanistically account for the impact of plate tectonics and global climate on macroevolutionary dynamics, INDITEK uses a Lagrangian approach. This approach permits us to track habitats (i.e., discrete points simulated by the palaeogeographic/plate tectonic model) over time as they move across Earth's surface, hence representing both the contribution of continuously-evolving environmental conditions and the legacy of past biodiversity changes. A Lagrangian approach contrasts with the traditional simulation of biodiversity 'snapshots' (Pohl et al., 2023) and provides a more integrated view of how biodiversity evolved in space and time, capturing the continuous nature of evolutionary processes. It allows for the tracking and analysis of slow, large-scale geological processes such as continental drift, thereby providing a framework for investigating the impact of Earth's dynamic surface evolution on the patterns of regional diversification. Today, plate tectonic reconstructions and landscape evolution models offer detailed insights into Earth's palaeogeographic and physiographic evolution throughout the Phanerozoic (Merdith et al., 2021; Salles et al., 2023; Scotese & Wright, 2018; Scotese, 2021; Vérard et al., 2015). Likewise, Earth system models reconstruct past climate conditions and biological productivity (Gurung et al., 2024; Ridgwell et al., 2007; Scotese et al., 2021), providing the environmental context for the unfolding of eco-evolutionary processes in the model (**Fig. 2, top**).

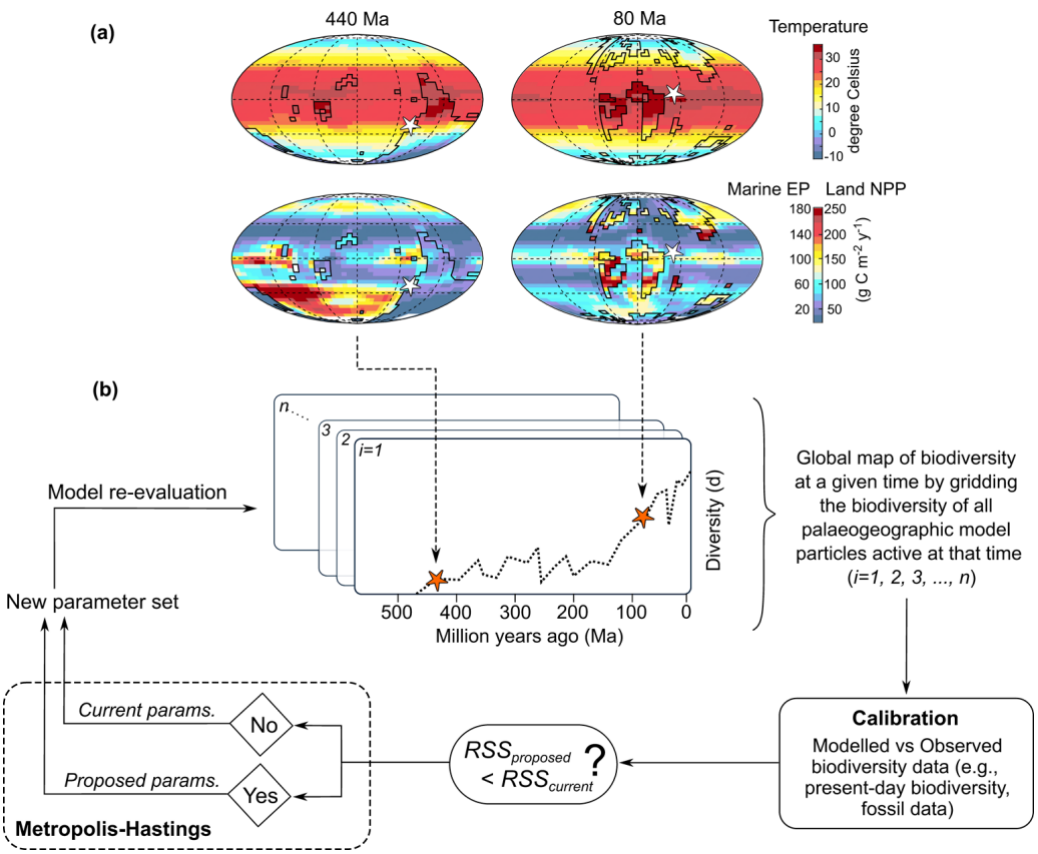

Fig. 2. Conceptual diagram of the Bayesian inverse modelling framework for estimating INDITEK-2.0 parameters. *(a) Global reconstructions of environmental conditions at two distinct time points: 440 Ma (million years ago) and 80 Ma. The top row shows reconstructed temperatures (air on land, sea surface in marine regions), while the row below shows marine export production (EP, $gCm^{-2}y^{-1}$) and land net primary productivity (NPP, $gCm^{-2}y^{-1}$) (**see Supplementary***

***information**). The black stepped outlines show the positions of continents. **(b)** Flow diagram of the Bayesian inverse modelling process. The method estimates model parameters by using a Metropolis-Hastings MCMC algorithm that iteratively compares the modelled biodiversity of moving palaeo-geographic particles (averaged in 0.5°x0.5° grids indexed as i) to a calibration dataset. The algorithm uses the residual sum of squares (RSS), its likelihood function, to quantify the discrepancy between the model output and the calibration data (current global and fossil biodiversity), favouring parameter values that improve the model's fit (i.e., reduce the RSS).*

2.4. Bayesian inverse modelling with Markov chain Monte Carlo

Accurately estimating model parameters that reflect a taxonomic group's unique evolutionary characteristics is essential for reliably reconstructing its biodiversity history. To achieve this, INDITEK-2.0 uses a Bayesian inverse modelling framework with a Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm to efficiently explore the parameter space (Hastings, 1970). The MCMC algorithm is used to generate a series of samples from the posterior probability distribution of the model parameters, which allows us to find the most probable parameter values and quantify their uncertainty. In each iteration, the algorithm proposes a new parameter set by randomly perturbing the current set. It then calculates the Residual Sum of Squares (RSS) as a measure of fit to compare the model's simulated biodiversity data and the observed (or calibration) data. New parameter sets are accepted if they reduce the RSS or according to a probabilistic criterion. This criterion allows for the occasional acceptance of less-optimal sets to avoid getting stuck in a local minimum. This iterative process creates a random walk through the parameter space that eventually converges on the posterior distribution, which is the probabilistic solution. After a sufficient "burn-in" period, the collected samples of parameter values approximate this distribution, providing a comprehensive representation of the probability of different parameter values given the observational data and our prior assumptions. This approach ensures that our parameter estimates are not single point values but a range of plausible values, reflecting the uncertainties inherent in the data. The model's parallelized architecture allows multiple independent chains to run simultaneously, which significantly increases the coverage of the sampled parameter space without requiring additional computational time. INDITEK-2.0 model parameters and their prior distributions are listed in **Table 1**. The priors encode current knowledge about the parameter values. As little is known about them, in this proof of concept we consider priors to be uniform (e.g., equal chance of being selected within a range of values), except for Q10 which is a gaussian because there is ample literature reporting values between 1 and 3 for invertebrates with more probability towards 1.5-2 values.

The code implementing the algorithm is written in Python and available at: https://github.com/glodegloria/indicios/tree/main

Table 1. INDITEK-2.0 model parameters: Prior distributions and posterior results.

Parameter	Units	True	Prior distribution	95% CI	Posterior Mean
K_{min}	#genera	19	Uniform (1, 150)		18.8978
K_{max}	#genera	161	Uniform (80, 1000)		161.0628
λ_{min}	Myr^{-1}	0.002	Uniform (0.0001, 0.05)		0.0020
λ_{max}	Myr^{-1}	0.035	Uniform (0.001, +∞)		0.0349
Q10	dimensionless	1.75	Gaussian (1 – 4)		1.7346
k_{food}	dimensionless	1.5	Fixed	0.5 (fixed)	0.5 (fixed)

3. Results

We conducted a proof-of-concept study to validate the Bayesian inverse modelling framework's ability to estimate key model parameters (see Supporting information for the experiment details). First, we generated a global map of synthetic biodiversity data for present-day marine invertebrates using parameter values from Cermeño et al. 2022. Some random Gaussian noise was added to the synthetic data, to mimic observational and theoretical errors. Next, we applied our Bayesian inverse modelling framework to this simulated data, with the goal of recovering the original, true parameter values. To this end, we ran six independent MCMC chains for each parameter, with each chain starting from a different point in the parameter space. Results showed that the mean of the posterior distribution for each parameter provides a good estimate of the true value (**Fig. 3; Fig. S1**). In particular, the true values are always contained within the 65% credible interval given by the posterior solution.

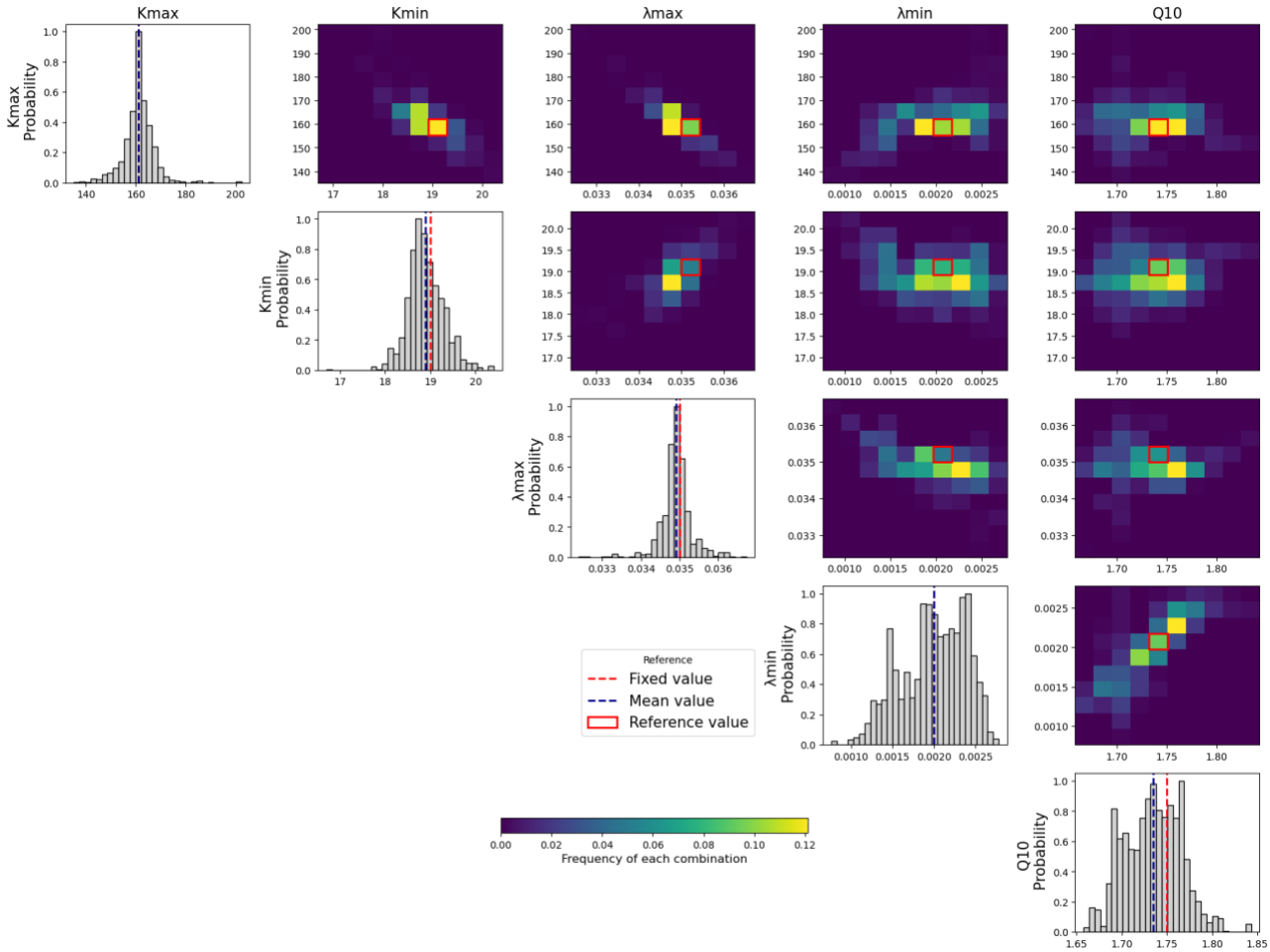

Fig. 3. Proof-of-concept study used to validate the Bayesian inverse modelling framework of INDITEK-2.0. *The diagonal panels show 1D histograms of the posterior distributions for each of the five key parameters (K_{max}, K_{min}, λ_{max}, λ_{min}, and Q10). Histograms represent the frequency of MCMC samples (all six chains pooled together after the burn-in period), while the red and blue dashed lines indicate the "true" parameter value used to produce the synthetic data and the mean of the posterior distribution, respectively. The off-diagonal panels display 2D heatmaps illustrating the joint posterior distributions and correlations between parameter pairs. The colour gradient denotes the density of MCMC samples. The red square included on each heatmap represents the fixed or "true" parameter value used to generate the synthetic biodiversity data. The location of this red square relative to the densest region of the heatmap is a key indicator of the framework's success in recovering the original parameters.*

The two-dimensional joint posterior distributions correctly identified the relationships between parameter pairs. The bivariate correlations among the model parameters offered key insights into the structure of the parameter space and potential interdependencies (**Fig. 3**). A strong positive correlation was observed between λ_{min} and Q10, suggesting that the MCMC algorithm compensated for a lower λ_{min} by exploring higher values for Q10. Conversely, K_{max} and K_{min}, the upper and lower bounds of the carrying capacity showed, respectively, a negative and positive correlation with λ_{max}. This might indicate a trade-off in how the model fits the synthetic biodiversity data. Specifically, to explain biodiversity patterns with a very high λ_{max}, the model must constrain K_{max} to prevent a runaway increase that would overshoot the data. In contrast, a high λ_{max} is positively linked to K_{min}, indicating that rapid diversity growth is necessary to explain a high baseline level of biodiversity. In essence, the model balances a high-speed diversification process against both K_{max} and K_{min} to reproduce the synthetic data. Other parameter pairs showed weak or negligible correlation, as indicated by their generally more circular joint posterior distributions, suggesting that the MCMC algorithm explored these parameters with a high degree of independence.

In line with the posterior-probability results, we observed a close match, without regional imbalances, between the diversity map of the synthetic data and the map of the model simulation with mean parameter values of the posterior distributions (**Fig. 4a-c**). To illustrate the variability among iterations and thus the uncertainty in our reconstructions, we further show the evolutionary histories of two small regions in the Caribbean and the Mediterranean modern biodiversity hotspots, by tracing their mean diversity trajectories back in time (**Fig. 4d-i**). These results illustrate the potential to quantify the uncertainty in our reconstructions, and to identify which macroevolutionary patterns and drivers are most consistently supported by the empirical data, by analysing the posterior distribution of plausible histories.

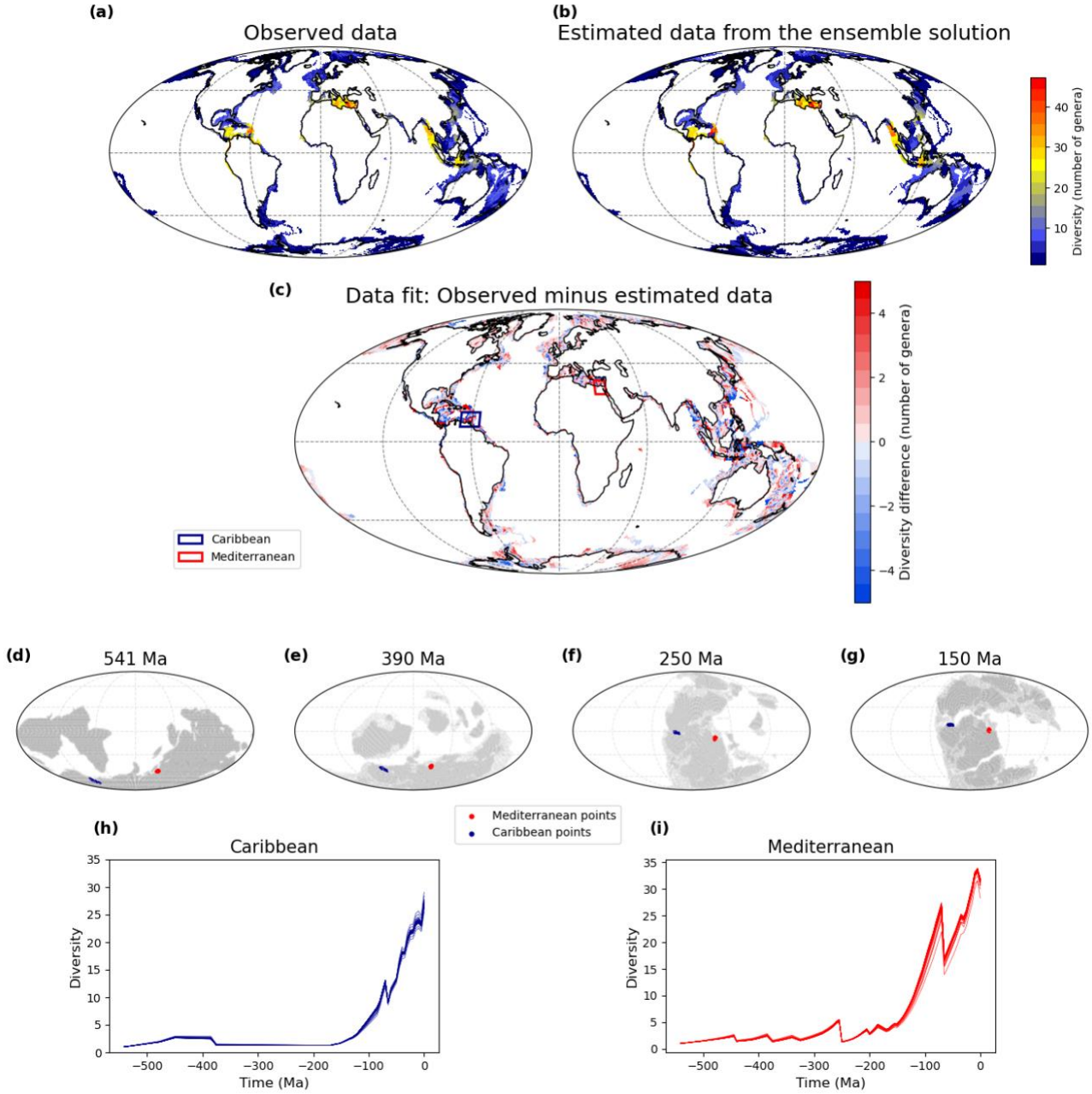

Fig. 4. Global reconstructions of modern biodiversity *obtained with (a) synthetic data and (b) the posterior distributions' mean values; (c) map of their difference (i.e., level of data fit); and the reconstructed evolutionary histories of two small regions in the Caribbean and Mediterranean modern biodiversity hotspots as they drift (d-g). Each evolutionary history (h-i) corresponds to each 50 iterations in the 6 MCMC chains after burn-in (iteration 200), that is 222 histories representing the posterior distribution.*

4. Model limitations and future directions

Our mechanistic modelling tool provides a virtual window into a profoundly complex evolutionary history. Evolutionary processes are shaped by many interacting factors, many of which are beyond the scope and possibilities of our model. Recognizing these limitations is crucial to adequately interpreting the model's results and devising future implementations. Next, we describe some of the model limitations.

1. The extreme environmental changes associated with mass extinctions pose a significant challenge for modelling due to the involvement of multiple stressors, often interacting in nonlinear ways. For example, the Ordovician mass extinction (ca. 445 million years ago) is thought to have been initiated by global glaciation, which was followed by a rapid warming period. These climate fluctuations led to rapid changes in sea level and ocean anoxia with dramatic consequences for marine life (Pohl et al., 2021). Similarly, the end-Permian mass extinction is thought to have been caused by massive volcanic eruptions, leading to severe climatic shifts, ocean acidification, and widespread anoxia (Hülse et al., 2021). Modelling mass extinctions requires sophisticated parameterizations to capture nonlinear dynamics and feedback loops within and between geological, chemical, and biological systems. These systems are intrinsically interconnected, with changes in one domain often leading to cascading effects in others (Hull, 2015; Junium et al., 2022; Schobben et al., 2019). The difficulty in implementing parameterizations that capture the nonlinear dynamics and feedback loops in Earth system models poses significant challenges in recreating the environmental scenarios responsible for mass extinctions.

2. As discussed in ***Section 2.1. Parameterizing evolutionary rates***, adaptation allows species to persist in changing environments. However, rapid environmental changes can outpace the rate of evolutionary adaptation, when these environmental changes occur rapidly, preventing species from developing the necessary traits to cope with new conditions. This reasoning posits a direct linkage between the rate of environmental change, such as temperature change, and the rate of species extinction (Kaiho, 2022; Song et al., 2021). Implementing this parameterization is currently limited by the lack of paleoenvironmental reconstructions with sufficient spatial and temporal resolution. Our current model uses environmental inputs from reconstructions with a temporal resolution of 30 time slices each encompassing 10–20 million-year and a spatial resolution of 36 x 36 (latitude, longitude) equal area grids. Extinction events, however, are not uniform across vast geographic areas or over long timescales; instead, they are often localized and triggered by relatively rapid environmental shifts. As a result, existing coarse-resolution reconstructions may smooth out these crucial, fine-scale variations, resulting in an inaccurate estimation of the extinction rate (μ). Therefore, the successful application of this parameterization is contingent on the availability of high-resolution environmental reconstructions that can accurately capture the nature of extinction dynamics. Current efforts are being made to increase these resolutions to every 1 million years and to 72x72 equal area grids.

3. Extinction selectivity is influenced by countless biological, ecological, and environmental variables, often interacting unpredictably. The ecological niches, life history traits, and geographic distributions of species significantly influence their resilience or vulnerability to environmental perturbations (Cole & Hopkins, 2021; Payne et al., 2016; Pinsky et al., 2019). These species-specific attributes cannot be implemented into a model operating above the species level. However, it is feasible to implement extinction selectivity at a higher taxonomic level by adjusting specific parameter values associated to a class or phylum known for their overall sensitivity or resistance to particular stressors. For example, assigning different values to ectotherms and endotherms for the thermal dependence of speciation rate can reflect their differential responses to temperature changes. Insects' high reproductive rates, short generation

times, and, in many cases, ability to go dormant can confer resistance to environmental stressors. Their evolutionary success could be modelled and parameterized by, for example, lowering their sensitivity to mass extinctions.

4. At or below the species level, mechanistic eco-evolutionary models can be designed to simulate how new traits emerge and evolve. These models can include mechanisms for trait evolution (e.g., trait diffusion), niche construction, and ecological interactions, allowing them to capture the interplay between organisms and their environments that drives speciation (Hagen, 2023).

5. The emergence of innovative evolutionary traits and co-evolutionary relationships has been pivotal in creating novel ecological opportunities and driving significant increases in biodiversity (Benton et al., 2015; Nitecki, 1990; Vermeij, 2017). By partitioning evolutionary history into distinct temporal windows defined by these events, we can independently estimate model parameters for each interval. This approach allows the eco-evolutionary model to use distinct parameter values after a major evolutionary innovation or co-evolutionary relationship begins. This enables the incorporation of specific priors that more accurately reflect the unique evolutionary dynamics of each period.

6. Finally, accurate estimation of model parameters is critically dependent on high-quality calibration data. Without reliable data, the model's parameters will be poorly constrained, leading to uncertain and potentially misleading results. To address the need for robust empirical support, we can integrate multiple, complementary data sources, incuding current biodiversity data, fossil records, and molecular phylogenies. By integrating these diverse datasets, we can overcome the limitations of any single source. For example, we can use fossil data and molecular phylogenies to provide intermediate calibration points or past checkpoints, ensuring modelled biodiversity stays on a realistic track throughout the simulation. In turn, current biodiversity data, available through global databases like OBIS or GBIF, with its detailed spatial resolution, serves as a robust present-day anchor, validating the model's output at the end of the simulation. This multi-faceted approach ensures that the model is calibrated against a solid representation of the true evolutionary history, thereby significantly reducing uncertainty in the parameter estimates and biodiversity reconstructions.

5. Conclusions

Our study introduces and validates INDITEK-2.0, which integrates two innovative approaches. First, a spatially-explicit Lagrangian modelling approach tracks the movement of portions of the Earth's surface over geological time as they drift with plate tectonics. This is crucial for tracing the ancient evolutionary origins of modern regions of interest, such as biodiversity hotspots. Second, INDITEK-2.0 uses a probabilistic Bayesian inverse modelling framework to generate hundreds of potential biodiversity histories, each with a quantified probability. Thus, we can directly link the unique biodiversity of modern locations to their specific palaeoenvironmental contexts—like ancient continental shelves, landmasses, and climates—as they drifted across the globe over millions of years.

INDITEK-2.0 framework elevates our modelling tool into a powerful discovery engine by enabling the probabilistic estimation of eco-evolutionary model parameters to target a specific taxonomic group. This capability opens the door to reconstructing biodiversity histories by comparing model outputs with present-day biodiversity and fossil data. Rather than a single solution, this probabilistic approach yields a range of parameter values, allowing INDITEK-2.0 to generate thousands of potential biodiversity histories, each with a quantified probability. While probabilistic inference has long been used to detect diversification rate shifts from time-calibrated molecular phylogenies (Nee, 2006; Ricklefs, 2007), applying these methods to spatially-explicit mechanistic eco-evolutionary models offers key advantages. This approach not only provides an explicit parameterization and

representation of speciation and extinction histories but, by interrogating the model, also enables the direct investigation of the mechanistic drivers of diversification. Furthermore, the Lagrangian modelling approach implemented here offers an excellent platform to investigate the mechanistic link between Earth's dynamic surface evolution and regional diversification dynamics. Ultimately, integrating mechanistic eco-evolutionary models with probabilistic parameter estimation is essential for reconstructing the evolutionary history of animal biodiversity. This is a major step forward in addressing the origins of the astonishing diversity of modern animals.

Supporting information

Earth system model simulations

This section describes the methods used to simulate the Late Ordovician and Late Cretaceous paleo-environmental conditions shown in **Fig. 2** of the main manuscript. To date, no numerical model offers a coupled representation of climate-carbon-cycle equilibrium, primary productivity on land and marine biogeochemistry in a spatially-resolved, 3D ocean – which together constitute the boundary conditions needed for Phanerozoic-scale mechanistic eco-evolutionary model simulations. This limitation arises from the contrasting time scales at which the underlying mechanisms occur. Carbon cycle equilibrium, on the one hand, results from the balance between (silicate, carbonate and kerogen) weathering on land and (carbonate and organic carbon) burial in the ocean, and operates at the million-year time scale. Oceanic circulation and biogeochemistry, on the other hand, reach equilibrium much faster, in a few 10s of kilo-years. As a result of this major difference in the time constants of the different components of the climate system, the long-term carbon cycle and the ocean-atmosphere system have traditionally been simulated using very different types of models. Long-term carbon cycle models can be run for 100s of million years. They can be used to simulate the evolution of the atmospheric concentration in greenhouse gases (CO_2 in particular) and oxygen, and changes in oceanic nutrient inventories (Mills et al., 2021). Their rapid integration time relies on parameterizations and spatial abstraction – generally representing the ocean and atmosphere as a single box (hence value). Earth system models, on the other hand, offer a spatially-resolved simulation of climate and ocean biogeochemistry, but their resulting computational cost usually does not permit to run them for more than a few 10s of kyrs (Pohl et al., 2022).

To overcome this limitation, we here propose to couple a long-term carbon cycle model (SCION) (Mills et al., 2021) with an Earth system model of intermediate complexity (cGENIE) (Ridgwell et al., 2007). In detail, we propose to first simulate paleo-climatic conditions using an ocean-atmosphere general circulation model (e.g., FOAM) (Jacob, 1997). Paleo-climatic conditions simulated at various atmospheric CO_2 levels can then be used to calculate the long-term climate-carbon-cycle equilibrium using SCION, together with primary productivity on land using the vegetation module FLORA (Gurung et al., 2022, 2024). Atmospheric CO_2 and O_2 concentrations and oceanic nutrient (phosphate) inventories simulated in SCION can be used as boundary conditions in cGENIE, the latter model permitting in turn to simulate ocean temperatures and biogeochemistry (including marine export production) in a 3D ocean. During all steps, the same continental reconstructions are used in every model. The final result is a self-consistent set of paleo-environmental conditions both on land and in the ocean – notably temperatures and primary productivity that are two key drivers of speciation and extinction rates.

Preparation of the proof of concept

In order to validate and demonstrate the resolving potential of our approach, we carried out a synthetic test. First the target parameter values were chosen following Cermeno et al. (2022) results adapted to the continental platform. Then, synthetic data (global biodiversity pattern at a spatial resolution of 0.5x0.5° grid) were calculated with the model with these parameters. In order to mimic observational and theoretical errors, and to enable efficient parameter exploration with the

Metropolis–Hastings algorithm, some random Gaussian random errors were added to the synthetic data set, with a standard deviation of 2 genera. Our inversion approach was then applied to this dataset. Six Markov chains were used to explore the 5-dimensional parameter space and to sample the posterior probability distribution. Each chain was initialized within a range of ±20% of the reference parameter values, with starting points selected using a Latin hypercube sampling strategy to ensure broad coverage of the parameter space. At each iteration, one forward simulation was performed, and the resulting output was compared to the synthetic dataset. Model fit was quantified using the residual sum of squares (RSS), calculated as the squared difference between the simulated and reference data, normalized by the variance of the added noise. As shown in Fig. S1, the different chains were started at different location in the parameter space, and all converged towards the same target distribution.

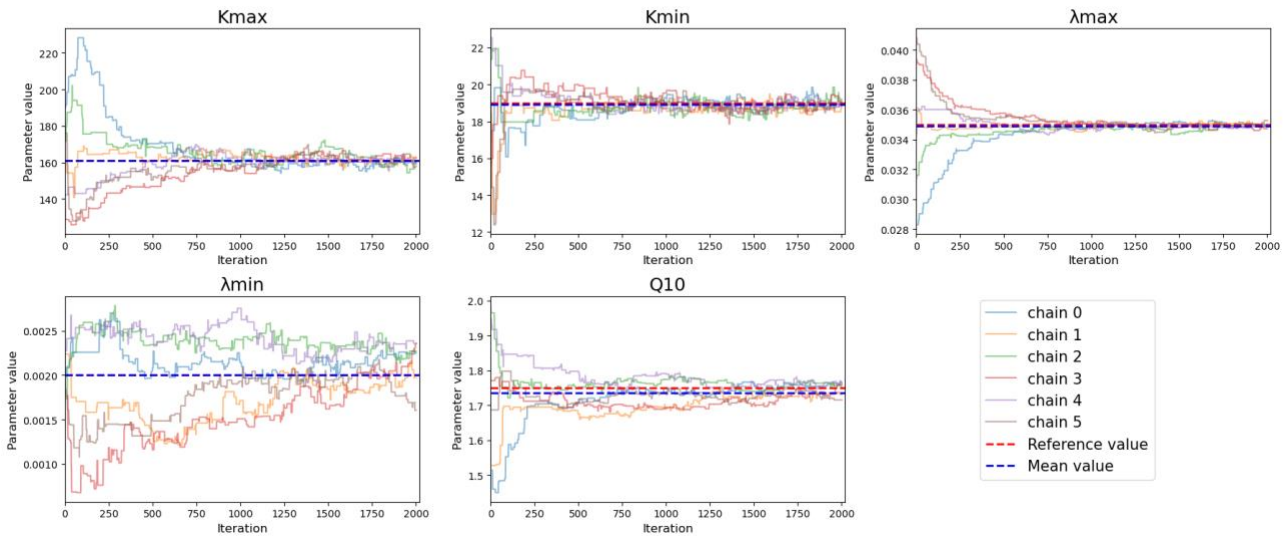

Fig. S1. MCMC Chain Trajectories for parameter inference. *Each line corresponding to a chain and the horizontal dashed lines corresponding to the reference ("true") value (red) and the mean value of the posterior distributions (blue) after the burn-in (from step 200).*

References

Allen, A. P., Brown, J. H., & Gillooly, J. F. (2002). Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, *297*(5586), 1545–1548. https://doi.org/10.1126/science.1072380

Bambach, R. K. (1983). Ecospace utilization and guilds in marine communities through the Phanerozoic. *Biotic Interactions in Recent and Fossil Benthic Communities*, 719–746. https://doi.org/10.1007/978-1-4757-0740-3_15

Bambach, Richard K. (1993). Seafood through time: Changes in biomass, energetics, and productivity in the marine ecosystem. *Paleobiology*, *3*(N/A), 372–397. https://doi.org/10.1017/S0094837300000336

Beaugrand, G. (2014). Marine biodiversity, climatic variability and global change. In *Marine Biodiversity, Climatic Variability and Global Change*. https://doi.org/10.4324/9780203127483

Beaugrand, G. (2023). Towards an Understanding of Large-Scale Biodiversity Patterns on Land and in the Sea. *Biology*, *12*(3), 339. https://doi.org/10.3390/biology12030339

Beaugrand, G., Luczak, C., Goberville, E., & Kirby, R. R. (2018). Marine biodiversity and the chessboard of life. *PLoS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0194006

Benton, M. J., Donoghue, P. C. J., Asher, R. J., Friedman, M., Near, T. J., & Vinther, J. (2015). Constraints on the timescale of animal evolutionary history. In *Palaeontologia Electronica* (Vol. 18, Issue 1). https://doi.org/10.26879/424

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, *85*(7). https://doi.org/10.1890/03-9000

Cárdenas, A. L., & Harries, P. J. (2010). Effect of nutrient availability on marine origination rates throughout the Phanerozoic eon. *Nature Geoscience*, *3*(6), 430–434. https://doi.org/10.1038/ngeo869

Cermeño, P., García-Comas, C., Pohl, A., Williams, S., Benton, M., Le Gland, G., Muller, R. D., Ridgwell, A., & Vallina, S. (2022). Post-extinction recovery of the Phanerozoic oceans and biodiversity hotspots. *Nature*, *607*, 507–511. https://doi.org/10.1038/s41586-022-04932-6

Chown, S. L., & Gaston, K. J. (2000). Areas cradles and museums: The latitudinal gradient in species richness. In *Trends in Ecology and Evolution* (Vol. 15, Issue 8). https://doi.org/10.1016/S0169-5347(00)01910-8

Cole, S. R., & Hopkins, M. J. (2021). Selectivity and the effect of mass extinctions on disparity and functional ecology. *Science Advances*, *7*(19). https://doi.org/10.1126/sciadv.abf4072

Farnsworth, A., Lo, Y. T. E., Valdes, P. J., Buzan, J. R., Mills, B. J. W., Merdith, A. S., Scotese, C. R., & Wakeford, H. R. (2023). Climate extremes likely to drive land mammal extinction during next supercontinent assembly. *Nature Geoscience*, *16*(10). https://doi.org/10.1038/s41561-023-01259-3

Garwood, R. J., Spencer, A. R. T., & Sutton, M. D. (2019). REvoSim: Organism-level simulation of macro and microevolution. *Palaeontology*, *62*(3), 339–355. https://doi.org/10.1111/pala.12420

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., & Charnov, E. L. (2001). Effects of size and temperature on metabolic rate. *Science*, *293*(5538). https://doi.org/10.1126/science.1061967

Gurung, K., Field, K. J., Batterman, S. A., Poulton, S. W., & Mills, B. J. W. (2024). Geographic range of plants drives long-term climate change. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-46105-1

Hagen, O. (2023). Coupling eco-evolutionary mechanisms with deep-time environmental dynamics to understand biodiversity patterns. *Ecography*, *2023*(4), e06132. https://doi.org/10.1111/ecog.06132

Hagen, O., Flück, B., Fopp, F., Cabral, J. S., Hartig, F., Pontarp, M., Rangel, T. F., & Pellissier, L. (2021). gen3sis: A general engine for eco-evolutionary simulations of the processes that shape Earth's biodiversity. *PLoS Biology*, *19*(7). https://doi.org/10.1371/journal.pbio.3001340

Hague, M. T. J., & Routman, E. J. (2016). Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity*, *116*(1), 92–98. https://doi.org/10.1038/hdy.2015.76

Hallam, A. (1998). Mass extinctions in Phanerozoic time. *Geological Society Special Publication*, *140*. https://doi.org/10.1144/GSL.SP.1998.140.01.18

Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Hull, P. (2015). Life in the aftermath of mass extinctions. In *Current Biology* (Vol. 25, Issue 19). https://doi.org/10.1016/j.cub.2015.08.053

Hülse, D., Lau, K. V., van de Velde, S. J., Arndt, S., Meyer, K. M., & Ridgwell, A. (2021). End-Permian marine extinction due to temperature-driven nutrient recycling and euxinia. *Nature Geoscience*, *14*(11). https://doi.org/10.1038/s41561-021-00829-7

Jacob, R. L. (1997). Low frequency variability in a simulated atmosphere-ocean system. In *Dissertation*.

Junium, C. K., Zerkle, A. L., Witts, J. D., Ivany, L. C., Yancey, T. E., Liu, C., & Claire, M. W. (2022). Massive perturbations to atmospheric sulfur in the aftermath of the Chicxulub impact. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(14). https://doi.org/10.1073/pnas.2119194119

Kaiho, K. (2022). Relationship between extinction magnitude and climate change during major marine and terrestrial animal crises. *Biogeosciences*, *19*(14), 3369–3380. https://doi.org/10.5194/bg-19-3369-2022

Lanfear, R., Ho, S. Y. W., Love, D., & Bromham, L. (2010). Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(47), 20423–20428. https://doi.org/10.1073/pnas.1007888107

Levinton, J. S. (1979). A theory of diversity equilibrium and morphological evolution. *Science*, *204*(4390). https://doi.org/10.1126/science.204.4390.335

Lovelace, D. M., Hartman, S. A., Mathewson, P. D., Linzmeier, B. J., & Porter, W. P. (2020). Modeling Dragons: Using linked mechanistic physiological and microclimate models to explore environmental, physiological, and morphological constraints on the early evolution of dinosaurs. *PLoS ONE*, *15*(5). https://doi.org/10.1371/journal.pone.0223872

MacArthur, R. H., & Wilson, E. O. (1967). *The Theory of Island Biogeography.* Princeton University Press.

Malchow, A. K., Bocedi, G., Palmer, S. C. F., Travis, J. M. J., & Zurell, D. (2021). RangeShiftR: an R package for individual-based simulation of spatial eco-evolutionary dynamics and species'

responses to environmental changes. *Ecography*, *44*(10), 1443–1452. https://doi.org/10.1111/ecog.05689

Martin, R. E. (1996). Secular increase in nutrient levels through the Phanerozoic: Implications for productivity, biomass, and diversity of the marine biosphere. *Palaios*, *11*(3), 209–219. https://doi.org/10.2307/3515230

Merdith, A. S., Williams, S. E., Collins, A. S., Tetley, M. G., Mulder, J. A., Blades, M. L., Young, A., Armistead, S. E., Cannon, J., Zahirovic, S., & Müller, R. D. (2021). Extending full-plate tectonic models into deep time: Linking the Neoproterozoic and the Phanerozoic. In *Earth-Science Reviews* (Vol. 214, p. 103477). https://doi.org/10.1016/j.earscirev.2020.103477

Mills, B. J. W., Donnadieu, Y., & Goddéris, Y. (2021). Spatial continuous integration of Phanerozoic global biogeochemistry and climate. In *Gondwana Research* (Vol. 100, pp. 73–86). https://doi.org/10.1016/j.gr.2021.02.011

Mondal, S., & Harries, P. J. (2016). Phanerozoic trends in ecospace utilization: The bivalve perspective. In *Earth-Science Reviews* (Vol. 152, pp. 106–118). https://doi.org/10.1016/j.earscirev.2015.10.005

Mundim, K. C., Baraldi, S., Machado, H. G., & Vieira, F. M. C. (2020). Temperature coefficient (Q10) and its applications in biological systems: Beyond the Arrhenius theory. In *Ecological Modelling* (Vol. 431). https://doi.org/10.1016/j.ecolmodel.2020.109127

Nee, S. (2006). Birth-death models in macroevolution. In *Annual Review of Ecology, Evolution, and Systematics* (Vol. 37, pp. 1–17). https://doi.org/10.1146/annurev.ecolsys.37.091305.110035

Nitecki, M. H. (1990). *Evolutionary innovations*. The University of Chicago Press.

Payne, J. L., Bush, A. M., Chang, E. T., Heim, N. A., Knope, M. L., & Pruss, S. B. (2016). Extinction intensity, selectivity and their combined macroevolutionary influence in the fossil record. *Biology Letters*, *12*(10). https://doi.org/10.1098/rsbl.2016.0202

Pellissier, L., Heine, C., Rosauer, D. F., & Albouy, C. (2018). Are global hotspots of endemic richness shaped by plate tectonics? *Biological Journal of the Linnean Society*, *123*(1). https://doi.org/10.1093/biolinnean/blx125

Peters, S. E., & Foote, M. (2002). Determinants of extinction in the fossil record. *Nature*, *416*(6879), 420–424. https://doi.org/10.1038/416420a

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., Payne, J. L., & Sunday, J. M. (2019). Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature*, *569*(7754). https://doi.org/10.1038/s41586-019-1132-4

Pohl, A., Lu, Z., Lu, W., Stockey, R. G., Elrick, M., Li, M., Desrochers, A., Shen, Y., He, R., Finnegan, S., & Ridgwell, A. (2021). Vertical decoupling in Late Ordovician anoxia due to reorganization of ocean circulation. *Nature Geoscience*, *14*(11). https://doi.org/10.1038/s41561-021-00843-9

Pohl, A., Ridgwell, A., Stockey, R. G., Thomazo, C., Keane, A., Vennin, E., & Scotese, C. R. (2022). Continental configuration controls ocean oxygenation during the Phanerozoic. *Nature*, *608*(7923), 523–527. https://doi.org/10.1038/s41586-022-05018-z

Pohl, A., Stockey, R. G., Dai, X., Yohler, R., Le Hir, G., Hülse, D., Brayard, A., Finnegan, S., & Ridgwell, A. (2023). Why the Early Paleozoic was intrinsically prone to marine extinction. *Science Advances*, *9*(35). https://doi.org/10.1126/sciadv.adg7679

Pontarp, M., Bunnefeld, L., Cabral, J. S., Etienne, R. S., Fritz, S. A., Gillespie, R., Graham, C. H., Hagen, O., Hartig, F., Huang, S., Jansson, R., Maliet, O., Münkemüller, T., Pellissier, L., Rangel, T. F., Storch, D., Wiegand, T., & Hurlbert, A. H. (2019). The Latitudinal Diversity Gradient: Novel Understanding through Mechanistic Eco-evolutionary Models. In *Trends in Ecology and Evolution* (Vol. 34, Issue 3, pp. 211–223). https://doi.org/10.1016/j.tree.2018.11.009

Rabosky, D. L. (2009). Ecological limits and diversification rate: Alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters*, *12*(8). https://doi.org/10.1111/j.1461-0248.2009.01333.x

Rangel, T. F., Edwards, N. R., Holden, P. B., Diniz-Filho, J. A. F., Gosling, W. D., Coelho, M. T. P., Cassemiro, F. A. S., Rahbek, C., & Colwell, R. K. (2018). Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves. *Science*, *361*(6399). https://doi.org/10.1126/science.aar5452

Raup, D. M. (1991). A kill curve for Phanerozoic marine species. *Paleobiology*, *17*(1), 37–48. https://doi.org/10.1017/S0094837300010332

Ricklefs, R. E. (2007). Estimating diversification rates from phylogenetic information. In *Trends in Ecology and Evolution* (Vol. 22, Issue 11, pp. 601–610). https://doi.org/10.1016/j.tree.2007.06.013

Ridgwell, A., Hargreaves, J. C., Edwards, N. R., Annan, J. D., Lenton, T. M., Marsh, R., Yool, A., & Watson, A. (2007). Marine geochemical data assimilation in an efficient Earth system model of global biogeochemical cycling. *Biogeosciences*, *4*(1), 87–104. https://doi.org/10.5194/bg-4-87-2007

Rohde, K. (1992). Latitudinal Gradients in Species Diversity: The Search for the Primary Cause. *Oikos*, *65*(3), 514. https://doi.org/10.2307/3545569

Romero-Mujalli, D., Jeltsch, F., & Tiedemann, R. (2019). Individual-based modeling of eco-evolutionary dynamics: state of the art and future directions. In *Regional Environmental Change* (Vol. 19, Issue 1). https://doi.org/10.1007/s10113-018-1406-7

Salles, T., Husson, L., Lorcery, M., & Hadler Boggiani, B. (2023). Landscape dynamics and the Phanerozoic diversification of the biosphere. *Nature*, *624*(7990). https://doi.org/10.1038/s41586-023-06777-z

Saupe, E. E., Hendricks, J. R., Portell, R. W., Dowsett, H. J., Haywood, A., Hunter, S. J., & Lieberman, B. S. (2014). Macroevolutionary consequences of profound climate change on niche evolution in marine molluscs over the past three million years. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1795). https://doi.org/10.1098/rspb.2014.1995

Saupe, Erin E., Qiao, H., Donnadieu, Y., Farnsworth, A., Kennedy-Asser, A. T., Ladant, J. B., Lunt, D. J., Pohl, A., Valdes, P., & Finnegan, S. (2020). Extinction intensity during Ordovician and Cenozoic glaciations explained by cooling and palaeogeography. *Nature Geoscience*, *13*(1). https://doi.org/10.1038/s41561-019-0504-6

Schobben, M., Van De Schootbrugge, B., & Wignall, P. B. (2019). Interpreting the carbon isotope record of mass extinctions. *Elements*, *15*(4). https://doi.org/10.2138/gselements.15.5.331

Scotese, C.R., & Wright, N. (2018). *PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the Phanerozoic PALEOMAP Project, https://www.earthbyte.org/paleodem-resource-scotese-and-wright-2018/*.

Scotese, Christopher R. (2021). An atlas of phanerozoic paleogeographic maps: The seas come in and the seas go out. In *Annual Review of Earth and Planetary Sciences* (Vol. 49, pp. 679–728). https://doi.org/10.1146/annurev-earth-081320-064052

Scotese, Christopher R., Song, H., Mills, B. J. W., & van der Meer, D. G. (2021). Phanerozoic paleotemperatures: The earth's changing climate during the last 540 million years. In *Earth-Science Reviews* (Vol. 215). https://doi.org/10.1016/j.earscirev.2021.103503

Sepkoski, J. J. (1978). A kinetic model of Phanerozoic taxonomic diversity I. Analysis of marine orders. *Paleobiology*, *4*(3), 223–251. https://doi.org/10.1017/S0094837300005972

Song, H., Kemp, D. B., Tian, L., Chu, D., Song, H., & Dai, X. (2021). Thresholds of temperature change for mass extinctions. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25019-2

Stanley, S. M. (2007). An Analysis of the History of Marine Animal Diversity. *Paleobiology*, *33*(4s), 1–55. https://doi.org/10.1017/S0094837300019217

Storch, D., & Okie, J. G. (2019). The carrying capacity for species richness. *Global Ecology and Biogeography*, *28*(10), 1519–1532. https://doi.org/10.1111/geb.12987

Valentine, J. W., & Moores, E. M. (1970). Plate-tectonic regulation of faunal diversity and sea level: A model. *Nature*, *228*(5272), 657–659. https://doi.org/10.1038/228657a0

Vérard, C., Hochard, C., Baumgartner, P. O., Stampfli, G. M., & Liu, M. (2015). 3D palaeogeographic reconstructions of the Phanerozoic versus sea-level and Sr-ratio variations. *Journal of Palaeogeography*, *4*(1), 64–84. https://doi.org/10.3724/SP.J.1261.2015.00068

Vermeij, G. J. (1977). The mesozoic marine revolution: Evidence from snails, predators and grazers. *Paleobiology*, *3*(3), 245–258. https://doi.org/10.1017/S0094837300005352

Vermeij, G. J. (2017). How the Land Became the Locus of Major Evolutionary Innovations. *Current Biology*, *27*(20). https://doi.org/10.1016/j.cub.2017.08.076

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. In *Annual Review of Ecology, Evolution, and Systematics* (Vol. 34). https://doi.org/10.1146/annurev.ecolsys.34.012103.144032

Wright, D. H. (1983). Species-Energy Theory: An Extension of Species-Area Theory. *Oikos*, *41*(3). https://doi.org/10.2307/3544109

Xuereb, A., Rougemont, Q., Tiffin, P., Xue, H., & Phifer-Rixey, M. (2021). Individual-based eco-evolutionary models for understanding adaptation in changing seas. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 288, Issue 1962). https://doi.org/10.1098/rspb.2021.2006

Zaffos, A., Finnegan, S., & Peters, S. E. (2017). Plate tectonic regulation of global marine animal diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(22), 5653–5658. https://doi.org/10.1073/pnas.1702297114